

hexagon:2610.00611v1

Continuity, Testing Regions and Fragility of Quantum Channel Divergences

M. Jones

October 2026

Abstract

Before a quantum gate is used in a computation or a sensing protocol it is characterized, and converse bounds on discriminating or estimating it, which limit every strategy including adaptive ones, are then evaluated on the characterized model. Is their sensitivity to characterization error a defect of the bounds or a property of the tasks they bound? The book opens with this question. Its first chapter shows that the supports of the Choi matrices separate genuine fragility, where the operational exponent itself jumps, from spurious fragility, where only single-letter converses jump. Near a noisy gate the sensitivity of the converses is governed by the noise floor, the smallest eigenvalue of the normalized Choi matrix: at first order it belongs to the task, whereas at second order the converses can overshoot the adaptive Stein and Chernoff exponents by factors that grow as the noise decreases. Certifying a converse from tomographic data costs a number of channel uses set by the same floor.

The rest of the book develops the theory behind these results. Converse bounds limit how well quantum channels can be discriminated, estimated or used to transmit information, and they are evaluated on a model of the channels that is known only approximately. We study when such bounds, and the operational exponents they bound, depend continuously on the channels, and what sets the size of the error when they do.

The divergences built from Kubo–Ando means, among them the geometric fidelity, the geometric Rényi divergences and the Belavkin–Staszewski relative entropy, are functions of a single positive contraction: the Arveson–Radon–Nikodym derivative of one channel with respect to the sum of the pair. In this coordinate they are continuous with dimension-free moduli. In the diamond norm, however, they are continuous at a pair of completely positive maps exactly when the Choi matrix of the sum is invertible or the fidelity vanishes, so ideal unitary gates are points of discontinuity. Near pairs whose Choi matrices are bounded below by λ , the fidelity is Lipschitz with a relative constant of the sharp order λ^{-1} .

Divergences built from hypothesis tests behave differently. The Umegaki relative entropy and the hockey-stick f -divergences are functionals of the testing region of a pair, and those with finite boundary values are uniformly continuous in every dimension. For channels, the adaptive Stein exponent is continuous exactly when the Choi matrix of the second channel is invertible or the exponent is infinite. We also study when the hockey-stick Rényi quasi-entropies are super-multiplicative under tensor products, and prove several cases of the corresponding conjecture.

For adaptive discrimination, every protocol is a controlled random walk on the spectrum of the derivative, which gives a geometric Chernoff converse with an explicit second-order term. Single-letter converses of this kind can be infinite where the adaptive exponent tends to zero, as for amplitude-damping channels with nearby damping rates. A chapter of proofs then returns to noisy gates and establishes the results of the first chapter, and a final chapter collects the open problems.

Acknowledgements

The author thanks their advisor, Dr. Javier Alejandro Chávez-Domínguez, for discussions of some of the results of Chapter 4 and for first encouraging the author's interest in this area of mathematics. The author also thanks Hunter Hamby for discussions of some of the results. This book was prepared independently: neither of them directed or reviewed it, and neither is responsible for its contents. Any errors are the author's alone. This research was originally motivated by understanding the papers of Okayasu [100] and Kian [76], but then sprawled into a general study of the Arveson–Radon–Nikodym theorem and how this can be used to study norm estimates of channel means. Throughout this process, the works of Bhatia [16] and Hiai [59] have proven invaluable. Once the author noticed the problem with the geometric channel fidelity, stemming from the work of [29], the author decided to push just how far this problem could be taken: the continuity estimates, discrimination, and the extension of the results to the hockey-stick divergences. Hence, during the preparation of this work the author used Claude (Anthropic; models Opus 5.5 and Fable 5.1). The purpose of this at first was to help to build a roadmap of the current literature which exists on the subject. Once trained on the material and on what the author wanted to do using the ARN theorem, these models were used in studying the background literature, searching for references, suggesting proofs and statements, and suggesting edits to the exposition. This book would not exist without the fundamental work of Arveson [10], Raginsky [111], Kian [76], Okayasu [100], Cree and Sikora [29], Hiai and Mosonyi [61], and Watrous [127]. Chapter 4 of this book is the author's arXiv submission [69]; the rest of this book followed from the author's work with Claude over several months.

Contents

Abstract	i
Acknowledgements	ii
1 Genuine and spurious fragility of quantum channel divergences near noisy gates	1
1.1 Introduction	1
1.2 Setting	3
1.3 Three regimes	4
1.4 First-order sensitivity: the Choi-floor law	5
1.5 Second-order sensitivity: monotone metrics and the multi-qubit window	7
1.6 Genuine and spurious parts of the fragility	9
1.7 Certification from finite data	14
1.8 Discussion	15
1.9 Plan of the book	17
2 Overview and reading guide	18
2.1 From noisy gates to the general questions	18
2.2 Three shadows of a pair	19
2.3 Results by chapter	19
2.4 How the chapters depend on each other	22
2.5 A reading guide	23
2.6 About this book	23
3 Notation and conventions	24
3.1 Spaces, operators and maps	24
3.2 Vectorisation and the Choi matrix	24
3.3 Norms on maps	25
3.4 Kubo–Ando connections and means	25
3.5 The Arveson–Radon–Nikodym derivative and the shadows	26
3.6 Fidelities	26
3.7 Divergences	27
3.8 The Thompson metric	28
3.9 Operational exponents	29
3.10 Logarithms	29
3.11 Differences between the chapters	29

4	Continuity of certain quantum channel divergences	32
4.1	Introduction	32
4.2	Preliminaries	34
4.2.1	The Arveson–Radon–Nikodym theorem	35
4.2.2	The effect point of view	36
4.2.3	Kubo–Ando means of operators	38
4.2.4	Constructing operator connections	39
4.2.5	The mean of completely positive maps	40
4.3	The mean of channels and geometric channel fidelity	42
4.3.1	Matsumoto fidelity	42
4.3.2	Geometric channel fidelity	42
4.3.3	An alarming property	43
4.4	Norm and continuity estimates	45
4.5	A general approach	50
4.5.1	Geometric Rényi relative entropy	50
4.5.2	The meta problem	52
4.5.3	Operator convex functions and divergences	53
4.6	Unbounded maximal f -divergences	58
4.6.1	The operator perspective on supports	58
4.6.2	Closed form and continuity	60
4.6.3	The relative Rényi entropies continued	63
4.7	All operator convex functions	67
4.7.1	The Lebesgue decomposition of a pair	67
4.7.2	Boundary values	67
4.7.3	The closed form	68
4.7.4	Continuity in the derivative	70
4.8	Continuity in the completely bounded trace norm	72
4.9	Explicit moduli for all operator convex functions and all means, with sharp or near-sharp constants	77
4.9.1	Operator convex functions	78
4.9.2	Operator Lipschitz constants and Schur multipliers	78
4.9.3	Sharp Lipschitz constants	86
4.9.4	All Kubo–Ando means	90
4.10	Pairs with different sums	91
4.11	Sums with different Choi supports	98
4.11.1	The Lebesgue decomposition of the sums	99
4.11.2	The estimate	101
4.11.3	Continuity in the derivative across support changes	104
4.12	Where continuity fails: a dichotomy	105
4.13	Conclusion and outlook	114
4.14	Norm bounds and a variational characterisation of means of completely positive maps	116
4.14.1	Norms of means	116
4.14.2	A variational characterisation of the geometric mean	117
4.14.3	Cauchy–Schwarz bounds	118

5	Shadows of quantum divergences: testing regions, modular lifts and continuity	120
5.1	Introduction	120
5.2	Three shadows of a pair of states	123
5.3	Structure of the testing-region shadow	126
5.4	Continuity on states	128
5.5	The modular shadow: Petz divergences	130
5.6	Tensor products: what replaces multiplicativity	131
5.7	Channels: the Thompson metric and the adaptive Stein exponent	133
5.8	The classification after this chapter	143
5.9	Numerical checks	145
5.10	Open problems	146
6	Super-multiplicativity of hockey-stick Rényi quasi-entropies	147
6.1	Introduction	147
6.2	Preliminaries from Chapter 5	149
6.3	Tensor products and the conjecture	150
6.4	Super-multiplicativity: three cases that can be proved	153
6.5	Integer orders as a correlation inequality on a simplex	158
6.6	Order three: the defect as an integral of Lieb–Thirring gaps	163
6.7	Pure partners, a cycle inequality and negative cycles	168
6.8	What the conjecture is a shadow of: the Frenkel measure and the convex order	176
6.9	Why channels cannot close the gap: the shadow discrepancy and restricted tests	179
6.10	Modular clocks, dephasing flows and two qubit pairs	181
6.11	Numerical checks	184
6.12	Open problems	189
7	Adaptive quantum channel discrimination as a controlled random walk	192
7.1	Introduction	192
7.1.1	Background and motivation	192
7.1.2	Contributions	193
7.1.3	Related work	194
7.1.4	Notation	194
7.1.5	Organization	194
7.2	The level decomposition	195
7.3	Every adaptive protocol is a controlled random walk	196
7.4	The geometric Chernoff converse	197
7.5	When the bound is attained	200
7.6	Continuity of the converse	202
7.6.1	Amplitude damping	206
7.7	Numerical illustration	211
7.8	Concluding remarks	213
8	Noisy gates: proofs and further results	215
8.1	Notation and conventions	215
8.2	Choi matrices and the diamond norm	216
8.3	The Choi-floor law (Theorem 1.4.1)	216
8.4	Closed forms for Pauli channels	218
8.5	The multi-qubit window	219

8.6	The monotone-metric hierarchy	220
8.7	Amplitude damping	221
8.8	The local Stein exponent	222
8.9	First-order fragility: task versus bound	223
8.10	The Chernoff exponent at a noisy gate	227
8.11	The Stein exponent at finite miscalibration	229
8.12	Biased Pauli noise and correlated miscalibrations	231
8.13	Certification	236
8.14	Numerical methods	239
8.15	Refinements of results from Chapters 4, 5 and 7	240
9	Open problems	244
9.1	Sharp constants for the Kubo–Ando family	244
9.2	Singular pairs and the dichotomy within channels	245
9.3	Beyond operator convex functions and finite dimensions	245
9.4	Regularised channel divergences and one-shot formulas	246
9.5	Super-multiplicativity of testing-region quasi-entropies	246
9.6	Adaptive discrimination of channels	247
9.7	Noisy gates and certification	249
	Bibliography	250

Chapter 1

Genuine and spurious fragility of quantum channel divergences near noisy gates

Summary. Converse bounds on discriminating and estimating noisy quantum gates are evaluated on a characterized model of the gate. We ask whether their sensitivity to characterization error belongs to the task or only to the bound. The answer is governed by the noise floor $\tilde{\lambda}$, the smallest eigenvalue of the normalized Choi matrix. At first order, a diamond-norm error ϵ moves the geometric-fidelity and Belavkin–Staszewski converses for adaptive strategies by at most order $\epsilon/\tilde{\lambda}$, but at the worst case the adaptive exponents move equally, so this sensitivity is genuine. At second order, a coherent miscalibration δ of a gate with error probability p moves these converses by order δ^2/p , the Umegaki relative entropy by $\delta^2 \ln(1/p)$, and the Uhlmann fidelity only by δ^2 , which is not amplified by the floor. Choi supports separate genuine jumps of the exponent from spurious ones, where converses are infinite while the exponent vanishes, as for the Chernoff exponent of amplitude damping. At a depolarized qubit gate the converses exceed the adaptive Stein and Chernoff exponents by at most a factor $3/2$ as $\delta \rightarrow 0$, but by factors growing as $1/(\sqrt{p} \ln(1/p))$ near $\delta \approx 2\sqrt{p}$. For n qubits the fragile window of the converses scales as $p^{w/2}$ for an error of weight w ; under biased noise it is set by the rate of the error itself. Certification through diamond-norm regions costs order $1/(\eta\tilde{\lambda})^2$ uses for incoherent learners, variance-adaptive regions order $1/(\eta^2\tilde{\lambda})$ for Pauli channels. This opening chapter states the problem in physical terms and announces the results; the general theory behind them is developed in Chapters 4–7, and the proofs specific to noisy gates are in Chapter 8.

1.1 Introduction

Before a quantum gate is used in a computation or a sensing protocol it is characterized. Randomized benchmarking, gate set tomography and process tomography produce a model of the gate together with error bars [23, 39, 78], and many statements about what the gate can do are then computed from that model. Among the most useful are converse bounds. They limit how well two gates can be told apart, or a parameter of a gate estimated, by any strategy, including adaptive ones that interleave the unknown gate with quantum memory and error correction [42, 53, 56, 72, 107, 131]. A claim that an error-corrected sensor operates at the precision limit set by its gates, or that an experiment demonstrates an advantage in discriminating two gates, is checked against such a bound. These bounds are single-letter: they are evaluated on one use of the channel, often as a semidefinite program,

and they hold for any number of uses. A converse is therefore only as reliable as the model it is evaluated on, and its response to characterization error is a practical question.

For the converses that hold against adaptive strategies the response is poor. The geometric channel fidelity and the Belavkin–Staszewski relative entropy [14, 42, 72, 91] are discontinuous at ideal unitary gates: the geometric fidelity of a gate with itself is one, but it vanishes against every miscalibrated copy, however close (Corollary 4.12.6 in Chapter 4). A real gate has an incoherent noise floor, which restores continuity, but the modulus of continuity then depends on the floor. For a depolarized single-qubit gate with error probability p , the δ^2 coefficient of these divergences under a small coherent over-rotation δ grows as $1/p$, that of the Umegaki relative entropy as $\ln(1/p)$, and that of the Uhlmann fidelity stays of order one (Table 1.1). This is a different effect from the known gap between average infidelity and diamond distance [80], which concerns the size of an error rather than the response of a bound to it. The floor is also the quantity that tomography resolves worst: for a two-qubit gate with $p = 10^{-3}$ under local noise the normalized floor is about 6×10^{-8} , and a two-qubit ZZ over-rotation of only 5×10^{-4} rad already reduces the geometric fidelity to 0.9.

Is this sensitivity a defect of the bounds or a property of the tasks they bound? The question is not trivial, because the operational quantities are fragile too. Distinct unitaries can be discriminated perfectly with finitely many adaptive uses [35], so adaptive error exponents jump at ideal gates. For a rotation under depolarizing noise, the Hamiltonian lies in the span of the products of Kraus operators, so error correction cannot restore Heisenberg scaling [31, 32, 131, 132]; it does, however, raise the asymptotic quantum Fisher information per use to order $1/p$ [36, 73, 131]. A converse valid for adaptive strategies must reproduce these features, but it need not be more sensitive than the task itself.

In this first chapter we separate the two effects. Part of the mathematical groundwork is developed later in the book, in Chapters 4, 5 and 7, whose results we use by forward reference, and the remaining proofs are collected in Chapter 8. “Sensitivity” has two meanings here, which we keep apart: the first-order change of a converse per unit of diamond-norm error in the model, and the second-order curvature of a converse in a coherent miscalibration δ . Our main findings are the following.

1. The supports of the Choi matrices decide whether a jump is genuine (the exponent itself jumps), spurious (only the converse jumps) or absent. Amplitude damping, the standard model of energy relaxation, is genuine for the Stein exponent and spurious for the Chernoff exponent; a program-simulation converse stays finite there.
2. At first order, the geometric fidelity, the geometric Chernoff divergence and the Belavkin–Staszewski divergence obey a Choi-floor law: a diamond-norm error ϵ moves them by at most $\ln[1/(1-r)]$ with $r = \epsilon/(2\tilde{\lambda})$, where $\tilde{\lambda}$ is the floor of the normalized Choi state. The resulting slope $1/(2\tilde{\lambda})$ is sharp, and at the worst case the adaptive Stein and Chernoff exponents equal the converses and move at the same rate, so in this first-order sense the sensitivity belongs to the task.
3. At second order, the coefficients form the hierarchy of monotone metrics, and for n qubits the fragile window is $\delta \sim p^{w/2}$ for a miscalibration of weight w . At a depolarized qubit gate the $1/p$ curvature of the Stein and Chernoff converses is genuine for $\delta \lesssim p \ln(1/p)$, where the ratio of converse to exponent is at most $3/2$ as $\delta \rightarrow 0$, and spurious beyond, where the converse overshoots by a factor that grows as $p \rightarrow 0$. Under biased single-qubit noise the local ratio stays below two; for correlated two-qubit errors it is open whether the exponent shares the window of the converse.

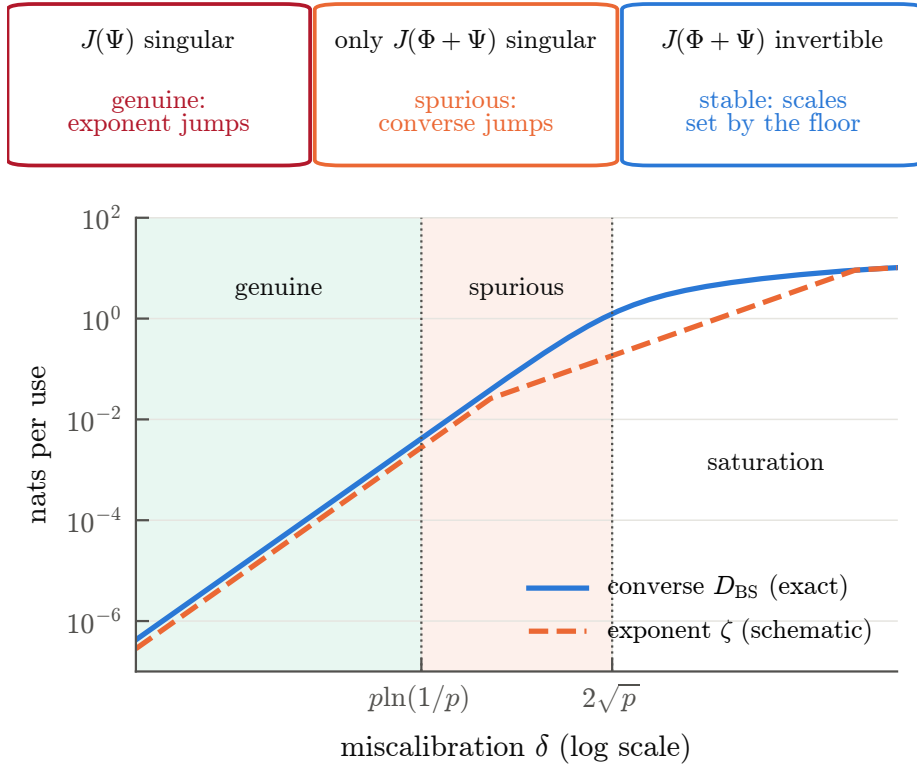


Figure 1.1: Schematic of the three regimes. Top: the supports of the Choi matrices decide whether a discontinuity is genuine (the operational exponent jumps), spurious (only single-letter converse jump) or absent. Bottom: at a stable pair, a depolarized qubit gate with $p = 10^{-4}$ and a coherent miscalibration δ , the distinction reappears on scales set by p . The solid line is the exact Belavkin–Staszewski converse; the dashed line is a schematic exponent that grows as $\delta^2/(3p)$ and then as $\delta \ln(1/p)$, illustrating the growth laws of Sec. 1.6 (the certified bracket is in Fig. 1.5). The converse follows the exponent for $\delta \lesssim p \ln(1/p)$, overshoots it up to $\delta \approx 2\sqrt{p}$, and saturates beyond.

4. Certifying a geometric converse through a diamond-norm confidence region costs of order $1/(\eta\tilde{\lambda})^2$ channel uses for accuracy η with incoherent learners, whereas variance-adaptive confidence regions need only $O(1/(\eta^2\tilde{\lambda}))$ uses for Pauli channels.

Figure 1.1 summarizes the classification and the scales on which it reappears at a noisy gate.

1.2 Setting

Let Φ, Ψ be quantum channels from an input space of dimension d_X to an output space of dimension d_Y , with Choi matrices $J(\Phi) = \sum_{ij} \Phi(|i\rangle\langle j|) \otimes |i\rangle\langle j|$, so that $\text{Tr } J(\Phi) = d_X$. The quantity that controls everything below is the *Choi floor* $\lambda(\Psi) = \lambda_{\min}(J(\Psi))$; the floor of the normalized Choi state is $\tilde{\lambda} = \lambda/d_X$. Distances between channels are measured in the unnormalized diamond norm $\|\cdot\|_{\diamond}$, so that $\|\Phi - \Psi\|_{\diamond} \leq 2$.

The geometric fidelity is $\hat{F}(\Phi, \Psi) = \lambda_{\min} \text{Tr}_Y[J(\Phi) \# J(\Psi)]$, where $A \# B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$ is the Kubo–Ando geometric mean [79] and Tr_Y the partial trace over the output [72] (Proposition 4.3.1). For singular arguments the mean is defined by continuity from above, $A \# B = \lim_{\varepsilon \downarrow 0} (A + \varepsilon \mathbf{1}) \# (B + \varepsilon \mathbf{1})$.

The geometric Chernoff divergence $\hat{C}$ replaces $\#$ by the weighted means $\#_s$ and minimizes over $s \in [0, 1]$, and the Belavkin–Staszewski (BS) divergence is $D_{\text{BS}}(\Phi\|\Psi) = \lambda_{\max} \text{Tr}_Y[J_\Phi^{1/2} \ln(J_\Phi^{1/2} J_\Psi^{-1} J_\Phi^{1/2}) J_\Phi^{1/2}]$ with $J_\Phi = J(\Phi)$ [14, 42].

The operational quantities are the adaptive Stein exponent $\zeta(\Phi\|\Psi)$, the decay rate of the type-II error at vanishing type-I error, and the adaptive Chernoff exponent $\xi(\Phi, \Psi)$ of the symmetric error. Both allow arbitrary adaptive strategies with quantum memory, which can outperform parallel ones [53, 116]; ζ equals the regularized, or amortized, channel relative entropy [41, 125, 128]. The converses we study are

$$\zeta(\Phi\|\Psi) \leq D_{\text{BS}}(\Phi\|\Psi), \quad (1.2.1)$$

$$\xi(\Phi, \Psi) \leq \hat{C}(\Phi, \Psi) \leq -2 \ln \hat{F}(\Phi, \Psi), \quad (1.2.2)$$

from the chain rule of Fang and Fawzi [42] and from Ref. [72] and Theorem 7.4.1 and Corollary 7.4.5 in Chapter 7. For comparison we use the Umegaki channel divergence $D(\Phi\|\Psi)$ and the Uhlmann channel root fidelity $F_U(\Phi, \Psi)$, optimized over inputs with a reference system. Neither gives an adaptive converse. The Umegaki divergence satisfies $D(\Phi\|\Psi) \leq \zeta(\Phi\|\Psi)$ [41, 125], so it is a lower bound on the exponent, and F_U stays close to one for close unitaries, which are perfectly distinguishable.

Our gate model is an n -qubit unitary U with a coherent miscalibration $R_P(\delta) = e^{-i\delta P/2}$ about a Pauli string P of weight w , followed by local depolarizing noise,

$$\mathcal{N}_{p,\delta} = \mathcal{D}_p^{\otimes n} \circ \text{Ad}_{R_P(\delta)} \circ \text{Ad}_U, \quad (1.2.3)$$

where $\mathcal{D}_p(\rho) = (1-p)\rho + p \text{Tr}(\rho) \mathbf{1}/2$, $\text{Ad}_V(\rho) = V\rho V^\dagger$ and $\mathcal{N}_p = \mathcal{N}_{p,0}$. Its Choi floor is $\lambda = (p/2)^n$, i.e., $\tilde{\lambda} = (p/4)^n$, independently of U .

1.3 Three regimes

Whether a divergence-based statement is continuous at a pair (Φ, Ψ) is decided by the supports of Choi matrices, and the operational exponents and their converses are decided by different ones. At the level of states, the analogous phenomenon is the discontinuity of the quantum Fisher information and of the Bures metric where the rank of the state changes [118, 123].

Genuine fragility. The adaptive Stein exponent is continuous at (Φ, Ψ) if and only if $J(\Psi)$ is invertible or $\zeta(\Phi\|\Psi) = \infty$ (Corollary 5.7.5(c) in Chapter 5). At an ideal gate U , $\zeta(U\|U) = 0$ but $\zeta(U'\|U) = \infty$ for every $U' \neq U$, and the same holds for ξ [35]. Every valid converse must therefore jump at ideal gates. The same applies to zero-temperature amplitude damping $\mathcal{A}_\gamma$, whose Choi matrix has rank two: for $\gamma \neq \gamma'$ the outputs ρ_γ and $\rho_{\gamma'}$ of $\mathcal{A}_\gamma$ and $\mathcal{A}_{\gamma'}$ on a maximally entangled input satisfy $\text{supp } \rho_\gamma \not\subseteq \text{supp } \rho_{\gamma'}$, so $\zeta(\mathcal{A}_\gamma\|\mathcal{A}_{\gamma'}) = \infty$. A model with pure T_1 decay thus predicts, for any two damping rates, a type-II error that decays faster than any exponential. In this case the discontinuity lies in the model itself.

Spurious fragility. As functions on pairs of completely positive maps, the geometric quantities are continuous at (Φ, Ψ) if and only if $J(\Phi + \Psi)$ is invertible or $\hat{F}(\Phi, \Psi) = 0$ (Theorem 4.12.4 in Chapter 4). Within channels, this dichotomy is proved at the diagonal pairs (Φ, Φ) , which are continuity points if and only if $J(\Phi)$ is invertible (Corollary 4.12.10); for other channel pairs with $J(\Phi + \Psi)$ singular and $\hat{F} > 0$ it is open. The divergence $\hat{C}$ is continuous wherever $J(\Phi + \Psi)$ is invertible (Theorem 7.6.4(c) in Chapter 7 and Proposition 8.15.1 in Chapter 8). When $J(\Phi + \Psi)$ is singular but the true exponent is finite, a converse can jump while the exponent does not. Amplitude damping provides the physical

example [114]. For $\gamma \neq \gamma'$ one has $\hat{F}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) = 0$, so $\hat{C}$, $-2 \ln \hat{F}$ and the barycentric converse C_{bar} of Ref. [66] are all infinite (Proposition 7.6.8), whereas the true Chernoff exponent tends to zero as $\gamma' \rightarrow \gamma$. The infinite value is a property of the Choi-level geometric mean and is witnessed by the input $|0\rangle$, on which both channels act identically, so it carries no physical information. With $\Delta = \gamma - \gamma'$ we find (see Chapter 8, Section 8.7)

$$\begin{aligned} \xi_{\text{prod}} &\leq \xi(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) \leq -2 \ln \mu_*, \\ \mu_* &= \frac{\sqrt{\gamma\gamma'}}{1 - \sqrt{(1-\gamma)(1-\gamma')}}, \end{aligned} \tag{1.3.1}$$

where ξ_{prod} is the classical Chernoff information of the outputs $\text{diag}(\gamma, 1-\gamma)$ and $\text{diag}(\gamma', 1-\gamma')$ on the product input $|1\rangle$. To leading order in Δ , $\xi_{\text{prod}} \simeq \Delta^2/[8\gamma(1-\gamma)]$ and $-2 \ln \mu_* \simeq \Delta^2/[4\gamma^2(1-\gamma)]$. The upper bound comes from a converse that applies to any pair of channels: μ_* is the largest overlap of two pure program states from which a common channel simulates both channels, as in the programmable simulations of Refs. [106, 107]; it is a semidefinite program for any pair, and the error of any adaptive strategy is at least $\frac{1}{4}\mu_*^{2n}$ after n uses. For $n \leq 4$ uses, exact semidefinite programs give the minimal root fidelity of the two final states as f^n , where f is the Bhattacharyya coefficient of the two damping probabilities. This suggests the sharper bound $\xi \leq -2 \ln f$ [Fig. 1.2(a)].

Replacement channels, which output a fixed state, give a second example. There are pairs Φ_t, Ψ_t at diamond distance $O(\sqrt{t})$ from a pair of identical channels for which $\hat{C}$, $-2 \ln \hat{F}$ and C_{bar} are infinite for every $t > 0$, while $\xi = 2 \ln[(1+t)/(1-t)] \approx 4t$ exactly [Theorem 8.15.6 in Chapter 8 and the example after it; Fig. 1.2(b)]; the parameter t controls the overlap of the two output states. Here the program-simulation converse is tight, $-2 \ln \mu_* = \xi$, so the jump is an artifact of $\hat{C}$, $-2 \ln \hat{F}$ and C_{bar} alone.

Stable pairs. When $J(\Phi + \Psi)$ is invertible, all the quantities above are continuous (Theorem 4.12.4(a), Theorem 7.6.4(c), Proposition 8.15.1 and Remark 8.15.3 in Chapter 8; for D_{BS} , Remark 8.15.4, which derives it from Theorem 4.10.4(e) and Corollary 4.7.7). Every channel of full Kraus rank, and so every gate model with a nonzero depolarizing or thermal floor, falls in this class. Divergences built directly from hypothesis tests, such as the hockey-stick divergences and their integrals [58, 63], are even uniformly continuous with dimension-free moduli (Theorem 5.4.1 and Corollary 5.4.2 in Chapter 5), but they do not give adaptive converses. Near realistic gates the classification is therefore only qualitative, and the rest of this chapter makes it quantitative.

1.4 First-order sensitivity: the Choi-floor law

The basic step converts a diamond-norm error into a relative error of Choi matrices. For two channels, $\|J(\Phi) - J(\Phi')\|_\infty \leq \frac{d_X}{2} \|\Phi - \Phi'\|_\diamond$, with equality attained in every dimension (Lemma 8.2.1 in Chapter 8). Hence a channel Ψ' within diamond distance ϵ of Ψ satisfies $(1-r)J(\Psi) \leq J(\Psi') \leq (1+r)J(\Psi)$ in the positive semidefinite order, with relative radius

$$r = \frac{d_X \epsilon}{2\lambda(\Psi)} = \frac{\epsilon}{2\tilde{\lambda}(\Psi)}. \tag{1.4.1}$$

Combined with the monotonicity and homogeneity of Kubo–Ando means, and with the operator monotonicity of the logarithm, this gives the following result.

Theorem 1.4.1 (Choi-floor law). *Let Ψ, Ψ' be channels with $\|\Psi' - \Psi\|_\diamond \leq \epsilon$ and $r < 1$ as*

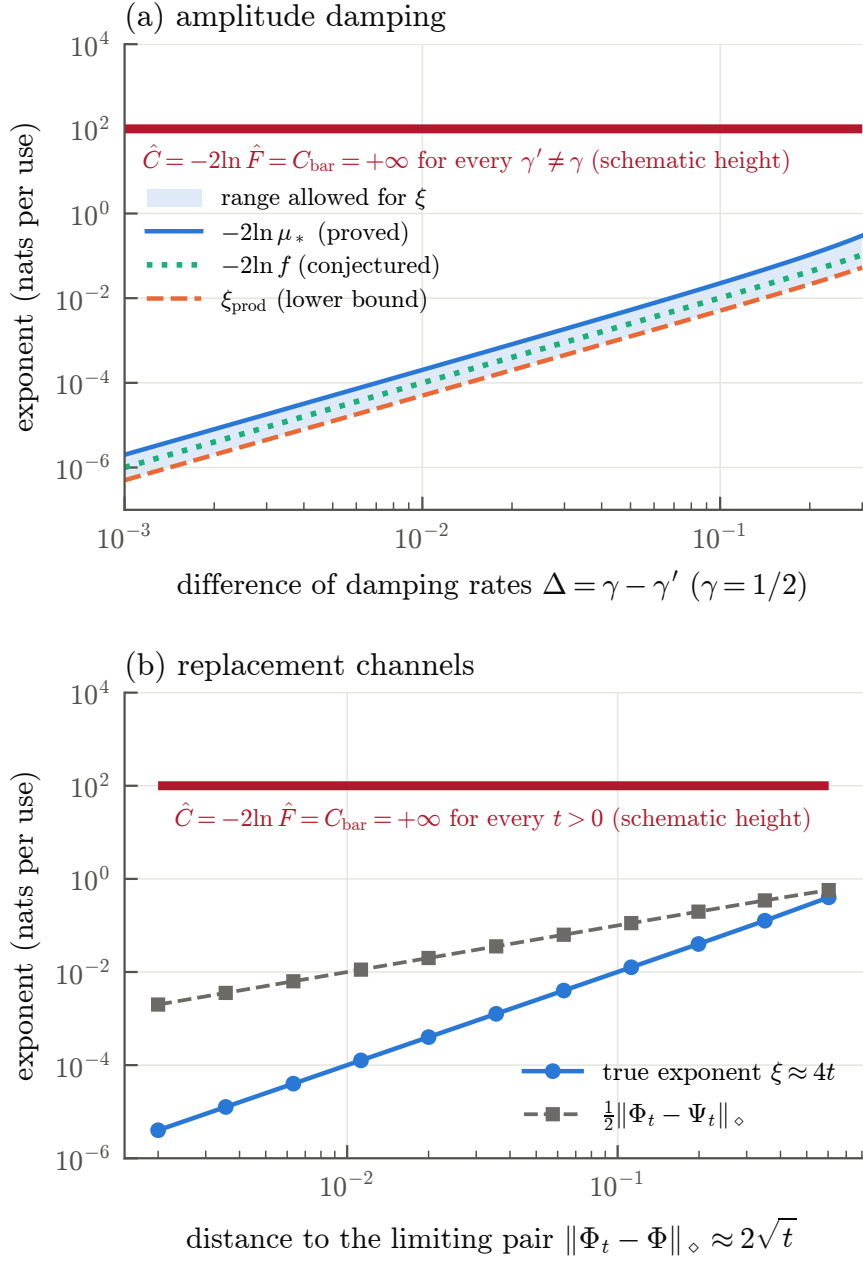


Figure 1.2: Spurious fragility of single-letter Chernoff converses. (a) Amplitude damping $\mathcal{A}_{\gamma}$ against $\mathcal{A}_{\gamma'}$ at $\gamma = 1/2$, as a function of $\Delta = \gamma - \gamma'$. The converses $\hat{C}$, $-2\ln \hat{F}$ and C_{bar} are infinite for every $\Delta \neq 0$ (thick line, drawn at a schematic height). The true adaptive Chernoff exponent ξ lies in the shaded range between the product-input lower bound ξ_{prod} (dashed) and the program-simulation upper bound $-2\ln \mu_*$ of Eq. (1.3.1) (solid); the dotted line is the conjectured bound $-2\ln f$. (b) Replacement channels Φ_t, Ψ_t with outputs ϕ_t, ψ_t , which converge to a common state as $t \rightarrow 0$, against the diamond distance to the limiting pair. The converses are again infinite, whereas the true exponent, the quantum Chernoff bound of ϕ_t and ψ_t , is $\xi = 2\ln[(1+t)/(1-t)] \approx 4t$ (circles), which here equals the program-simulation bound $-2\ln \mu_*$; squares show $\frac{1}{2}\|\Phi_t - \Psi_t\|_{\diamond}$ for comparison. In both cases the converses are infinite while the exponent tends to zero.

in Eq. (1.4.1). For every completely positive map Φ ,

$$\begin{aligned} \sqrt{1-r} \hat{F}(\Phi, \Psi) &\leq \hat{F}(\Phi, \Psi') \leq \sqrt{1+r} \hat{F}(\Phi, \Psi), \\ |\hat{C}(\Phi, \Psi') - \hat{C}(\Phi, \Psi)| &\leq \ln \frac{1}{1-r}, \end{aligned} \quad (1.4.2)$$

and for every channel Φ , $-\ln(1+r) \leq D_{\text{BS}}(\Phi\|\Psi') - D_{\text{BS}}(\Phi\|\Psi) \leq \ln[1/(1-r)]$. If $d_X = d_Y$ and the λ -eigenspace of $J(\Psi)$ contains a maximally entangled vector, as for every Pauli channel composed with a unitary, the constants are sharp. For $\ln \hat{F}$,

$$\sup_{\Phi} \limsup_{\epsilon \rightarrow 0} \sup_{\|\Psi' - \Psi\|_{\diamond} \leq \epsilon} \frac{|\ln \hat{F}(\Phi, \Psi') - \ln \hat{F}(\Phi, \Psi)|}{\epsilon} = \frac{1}{4\tilde{\lambda}}, \quad (1.4.3)$$

and at the same Φ the bounds for $\hat{C}$ and D_{BS} are attained to first order.

The worst case is a unitary $\Phi = \text{Ad}_W$ whose Choi vector lies in the floor eigenspace, perturbed towards any unitary W' with $\text{Tr } W^\dagger W' = 0$. For local noise, W differs from the gate by a correlated error of full weight. Only the floor of the perturbed argument enters; a two-sided version in which both channels move holds with $1 \pm r$ in place of $\sqrt{1 \pm r}$ (Theorem 4.8.8 in Chapter 4; Proposition 8.15.2 in Chapter 8 allows separate floors). At this worst case $\hat{C} = D_{\text{BS}} = -2 \ln \hat{F} = \ln(1/\tilde{\lambda})$ (Proposition 1.6.1 below), so the three bounds of Theorem 1.4.1 are attained together, with slope $1/(2\tilde{\lambda})$ for $\hat{C}$, D_{BS} and $-2 \ln \hat{F}$. At a depolarized qubit gate, $1/(4\tilde{\lambda}) = 1/p$, so for the worst-case Φ a diamond-norm error ϵ can move $-2 \ln \hat{F}$ by $2\epsilon/p$. In absolute terms the largest slope of $\hat{F}$ itself lies between $[2/(3\sqrt{3}p)][1 + o(1)] \approx 0.385/p$, attained by a unitary Φ , and $1/p$. We conjecture that unitary Φ are the worst case; a numerical search over mixed Φ found nothing larger. The maximizing unitary sits at a pair where the converse already overshoots the Stein exponent, by a certified factor of at least 5 at $p = 10^{-3}$ (Table 8.4), so this absolute slope is a property of the converse.

The Stein exponent behaves differently in the channel under test. Errors of size $\epsilon' = \frac{1}{2}\|\Phi - \Phi'\|_{\diamond}$ in Φ change ζ by at most $2\epsilon' \ln d_Y + (1 + \epsilon')h(\epsilon'/(1 + \epsilon')) + \epsilon' \ln \kappa(\Psi)$, where h is the binary entropy and $\kappa(\Psi)$ the condition number of $J(\Psi)$ (Corollary 5.7.5 in Chapter 5). Near a gate $\kappa \simeq (4/p)^n$, so this modulus grows only as $\ln(1/p)$. Errors in the reference channel Ψ enter ζ through the Thompson metric between $J(\Psi)$ and $J(\Psi')$, which is at most $\ln[1/(1-r)]$ [120] (Theorem 5.7.2, and Eq. (8.2.1) in Chapter 8). In fact, every quantity considered here, the exponents ζ and ξ included, is 1-Lipschitz in the Thompson metric, so none of them moves by more than $\ln[1/(1-r)]$ under such an error (Proposition 8.9.2). Section 1.6 shows that the exponents attain this rate at the worst case of Theorem 1.4.1.

1.5 Second-order sensitivity: monotone metrics and the multi-qubit window

Theorem 1.4.1 is a worst case over pairs. At a given pair, such as a gate and its miscalibrated copy, the response is governed by curvature. The hierarchy in Table 1.1 has a simple interpretation. To second order, each divergence is a monotone Riemannian metric [104] evaluated on the Choi state: the right logarithmic derivative (RLD) metric for $\hat{F}$ and D_{BS} , the Bogoliubov–Kubo–Mori (BKM) metric for D , and the symmetric logarithmic derivative (SLD) metric for F_U [24, 71, 104]. A coherent error couples the top eigenvector of the Choi state, with eigenvalue $\mu \approx 1$, to a floor eigenvector with eigenvalue $\tilde{\lambda}$. The metric defined by an operator monotone function f weights such a coherence by $c_f(\mu, \tilde{\lambda}) = 1/[\mu f(\tilde{\lambda}/\mu)]$,

Table 1.1: Second-order response of channel divergences to a coherent over-rotation δ at a depolarized qubit gate, $\mathcal{N}_{p,\delta}$ against $\mathcal{N}_p$ with $P = Z$: coefficients of δ^2 . Each coefficient is a Fisher information per use of one monotone metric; the last row is the asymptotic Fisher information with adaptive control and error correction. All entries are exact. D and F_U use the maximally entangled input, which is optimal.

quantity	coefficient of δ^2	$p \rightarrow 0$
$1 - \hat{F}$	$F_{\text{RLD}}/8$	$1/(8p)$
$D_{\text{BS}}(\mathcal{N}_{p,\delta} \mathcal{N}_p)$	$F_{\text{RLD}}/2$	$1/(2p)$
$D(\mathcal{N}_{p,\delta} \mathcal{N}_p)$	$F_{\text{BKM}}/2$	$\ln(4/p)/4$
$1 - F_U(\mathcal{N}_{p,\delta}, \mathcal{N}_p)$	$F_{\text{SLD}}/8$	$1/8$
F_{RLD} (RLD)	$2(1-p)^2(2-p)/[p(4-3p)]$	
F_{BKM} (BKM)	$\frac{1}{2}(1-p)\ln[(4-3p)/p]$	
F_{SLD} (SLD)	$2(1-p)^2/(2-p)$	
F_∞ (adaptive)	$2(1-p)^2/[p(3-2p)]$	

which is $(\mu + \tilde{\lambda})/(2\mu\tilde{\lambda})$ for RLD, $\ln(\mu/\tilde{\lambda})/(\mu - \tilde{\lambda})$ for BKM and $2/(\mu + \tilde{\lambda})$ for SLD. The weight stays bounded as $\tilde{\lambda} \rightarrow 0$ exactly when $f(0^+) > 0$. Two caveats matter in practice. The Uhlmann fidelity is not amplified by the floor for coherent errors, but an incoherent change of a floor eigenvalue costs $1/\tilde{\lambda}$ in every metric; at the worst case of Theorem 1.4.1 the Umegaki divergence and the Uhlmann fidelity equal the converses and have the same first-order sensitivity (Proposition 1.6.1). The hierarchy is therefore a statement about the second-order response to coherent errors only. And the closed forms extend to any Pauli channel, any Pauli-string miscalibration and any n (Theorem 8.4.1 in Chapter 8).

For n qubits the floor $\tilde{\lambda} = (p/4)^n$ is exponentially small, and which errors reach it depends on their weight. For $\mathcal{N}_{p,\delta}$ of Eq. (1.2.3),

$$\left(1 - \frac{3p}{4}\right)^n g(x) \leq 1 - \hat{F}(\mathcal{N}_{p,\delta}, \mathcal{N}_p) \leq g(x), \quad (1.5.1)$$

where $g(x) = 1 - (1+x)^{-1/2}$, $x = (\kappa_w - 1)^2 \sin^2(\delta/2)/(4\kappa_w)$, and $\kappa_w = [(4-3p)/p]^w$ is the ratio of the two Choi eigenvalues coupled by the error; for $w = n$ it is the condition number of $J(\mathcal{N}_p)$. The geometric infidelity therefore reaches a level η at the window

$$\delta_\eta = 4\sqrt{x_\eta} \left(\frac{p}{4}\right)^{w/2} [1 + O(np)], \quad (1.5.2)$$

with $x_\eta = (1 - \eta)^{-2} - 1$, and D_{BS} obeys the same scaling. The window depends on the weight of the error, not on the number of qubits: a weight-one error on any number of qubits has $\delta_{0.1} \approx 0.97\sqrt{p}$. A full-weight error has $\delta_\eta \propto p^{n/2}$. At $p = 10^{-3}$ and $\eta = 0.1$ this is $\delta_\eta \approx 4.8 \times 10^{-4}$ for a two-qubit ZZ error and 1.2×10^{-7} for a four-qubit error of full weight. Under global depolarizing noise of the same strength the window is $\delta_\eta \simeq 4\sqrt{x_\eta p}/d$, with $d = 2^n$, which is 1.5×10^{-2} for two qubits. By contrast $D \leq w \ln[(4-3p)/p] \sin^2(\delta/2)$ and $1 - F_U \leq \sin^2(\delta/2)$, independently of n . The worst-case modulus of Theorem 1.4.1 at this gate, taken over the other argument, is $\frac{1}{4}(4/p)^n$. Figure 1.3 shows the collapse of exact evaluations for $n = 1, 2, 3$.

Biased noise. The closed forms cover any Pauli noise (Chapter 8, Section 8.12). For a miscalibration about P the window is set by the coupled rate q_P , the probability of the Pauli error P itself, as $\delta_\eta \simeq 4\sqrt{x_\eta q_P/q_1}$, whereas the floor $\tilde{\lambda} = \min_Q q_Q$ sets the worst-case modulus of Theorem 1.4.1; the two are set by different rates. Under dephasing-dominated noise with bias $b = p_z/(p_x + p_y)$, as in bias-preserving architectures [109, 121],

a Z miscalibration keeps the wide window $1.94\sqrt{p_z}$ for every b , whereas the window of an X miscalibration narrows to $1.94\sqrt{p_z/(2b)}$, set by the small transverse rate.

1.6 Genuine and spurious parts of the fragility

We now compare the converses with the exponents they bound, first at first order and then at second order.

First order. At the worst case of Theorem 1.4.1 the converses are exact.

Proposition 1.6.1 (Genuine first-order fragility). *Let $d_X = d_Y$, let Ψ be a channel with $J(\Psi) > 0$, and let W be a unitary whose Choi vector is an eigenvector of $J(\Psi)$ with eigenvalue $d_X \tilde{\mu}$. Then*

$$\zeta = \xi = D_{\text{BS}} = \hat{C} = -2 \ln \hat{F} = D = -2 \ln F_U = \ln(1/\tilde{\mu}), \quad (1.6.1)$$

all evaluated at (Ad_W, Ψ) . All seven quantities have the same derivative in every direction of Ψ , so for a floor eigenvector ($\tilde{\mu} = \tilde{\lambda}$) each of them changes at the rate $1/(2\tilde{\lambda})$ per unit of diamond distance, the sharp rate of Theorem 1.4.1.

The optimal test is a Bell measurement in the frame of W , which detects every error other than W ; for a Pauli channel $\tilde{\mu} = q_W$, and Eq. (1.6.1) holds exactly at every Pauli channel in the same frame. The sharp constant of Theorem 1.4.1 is therefore a property of the task: the floor sets the error budget for any statement about discriminating a gate from the model Ψ . More generally, wherever a converse is tight and differentiable, the exponent is at least as sensitive at first order as the converse, so a first-order spurious gap can occur only where the converse already overestimates the exponent (Lemma 8.9.5 in Chapter 8). For Pauli-diagonal pairs the single-letter converses are exact (Proposition 8.9.6), whereas at the slope maximizer of Sec. 1.4 they are not.

Second order. We compare the BS converse with the exponent it bounds, $\zeta(\mathcal{N}_{p,\delta} \| \mathcal{N}_p)$, for the qubit gate ($n = 1$, $P = Z$). At small δ the comparison can be made exactly. Any fixed k -use strategy gives $\zeta \geq \frac{1}{k} D(\rho_k^\delta \| \rho_k^0)$, where ρ_k^δ are its output states [41, 125]. Because $J(\mathcal{N}_{p,\delta}) \geq (p/2)\mathbf{1}$, these outputs have δ -independent support, and the relative entropy is, to second order, a BKM metric, which dominates the SLD metric [104]. Taking the error-corrected strategies of Zhou and Jiang [131] gives $\liminf_{\delta \rightarrow 0} \zeta/\delta^2 \geq F_\infty/2$, where F_∞ is the asymptotic Fisher information per use. For this gate the Hamiltonian lies in $\text{span}\{K_i^\dagger K_j\}$, where K_i are the Kraus operators, so the scaling is linear, and $F_\infty = 2(1-p)^2/[p(3-2p)]$; this is the closed form of Ref. [131], and a primal–dual pair of the corresponding semidefinite program, evaluated in 50-digit arithmetic, agrees with it to a relative gap below 2×10^{-47} (Chapter 8, Section 8.8). Since $D_{\text{BS}} = F_{\text{RLD}}\delta^2/2 + O(\delta^4/p^2)$, our result for the local regime is

$$\begin{aligned} 1 \leq \liminf_{\delta \rightarrow 0} \frac{D_{\text{BS}}}{\zeta} &\leq \limsup_{\delta \rightarrow 0} \frac{D_{\text{BS}}}{\zeta} \\ &\leq \frac{F_{\text{RLD}}}{F_\infty} = \frac{(2-p)(3-2p)}{4-3p} < \frac{3}{2}. \end{aligned} \quad (1.6.2)$$

In the local regime the $1/p$ curvature is therefore genuine: $\zeta \geq [1 + O(p)] \delta^2/(3p) + o(\delta^2)$, and the BS converse is tight to within a factor $3/2$. The local coefficient itself is known only to lie between $F_\infty/2$ and $F_{\text{RLD}}/2$, i.e., $1/3 \lesssim p\zeta/\delta^2 \leq 1/2$ for small p . The upper bound in Eq. (1.6.2) would be attained if the local Stein coefficient were the SLD rate $F_\infty/2$:

Conjecture 1.6.2 (Local tightness of the BS converse). *For the depolarized qubit gate, $\lim_{\delta \rightarrow 0} D_{\text{BS}}(\mathcal{N}_{p,\delta} \| \mathcal{N}_p)/\zeta(\mathcal{N}_{p,\delta} \| \mathcal{N}_p) = F_{\text{RLD}}/F_\infty$, which tends to $3/2$ as $p \rightarrow 0$.*

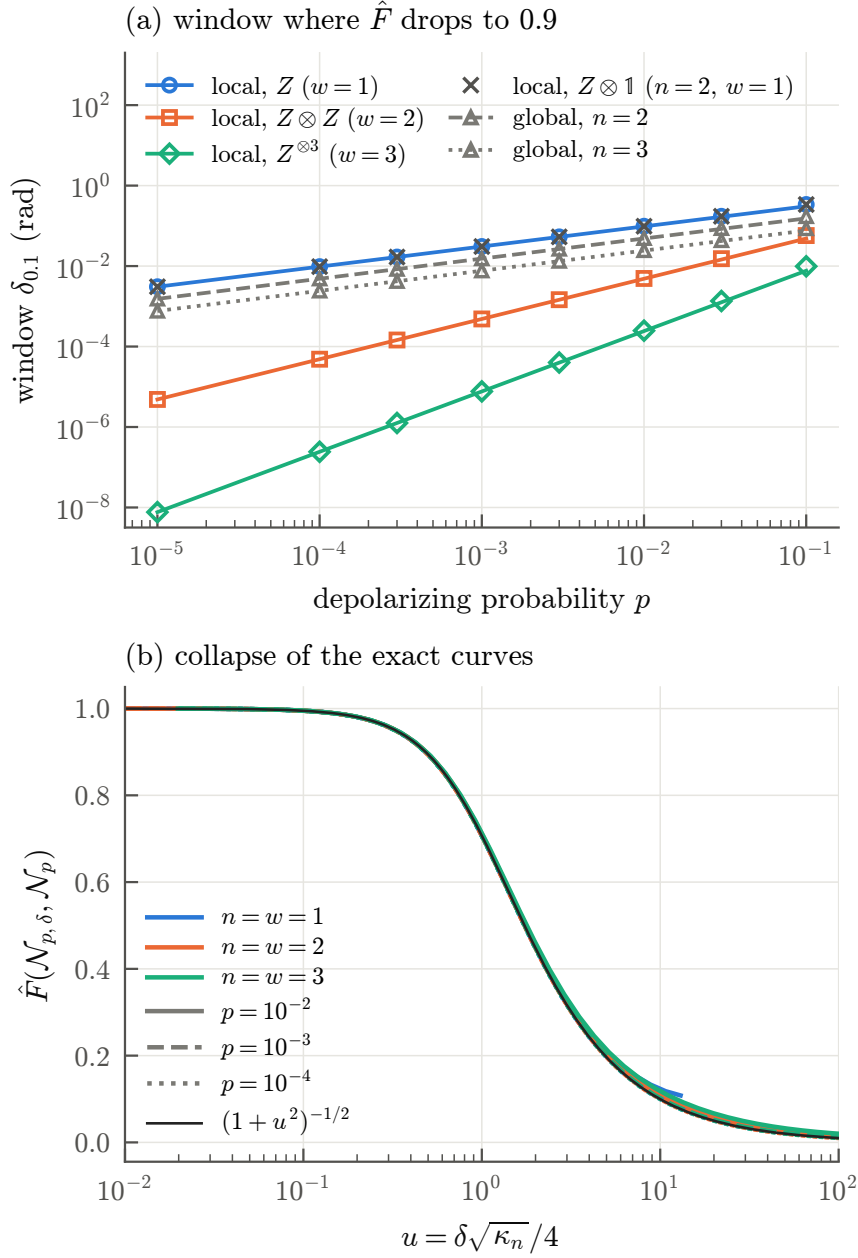


Figure 1.3: Fragility window of the geometric fidelity for $n = 1, 2, 3$ qubits. (a) Miscalibration $\delta_{0.1}$ at which $\hat{F}(\mathcal{N}_{p,\delta}, \mathcal{N}_p)$ drops to 0.9, from exact 50-digit evaluations (symbols), for local depolarizing noise with a miscalibration of weight w (colors) and for global depolarizing noise (grey). Lines are the asymptotic laws $1.94(p/4)^{w/2}$ and $1.94\sqrt{p}/d$; fitted slopes for $p \leq 10^{-3}$ are 0.500, 1.000 and 1.500 for $w = 1, 2, 3$. The weight-one error $Z \otimes \mathbf{1}$ on two qubits (crosses) has the same window as Z on one qubit. (b) The exact curves for $w = n$ and $p = 10^{-2}, 10^{-3}, 10^{-4}$, plotted against $u = \delta\sqrt{\kappa_n}/4$, collapse onto $(1 + u^2)^{-1/2}$; here $\kappa_n = [(4 - 3p)/p]^n$ is both the ratio of the two coupled Choi eigenvalues and the condition number of $J(\mathcal{N}_p)$. The fragile window of an n -qubit gate is set by the weight of the error and is exponentially small in it.

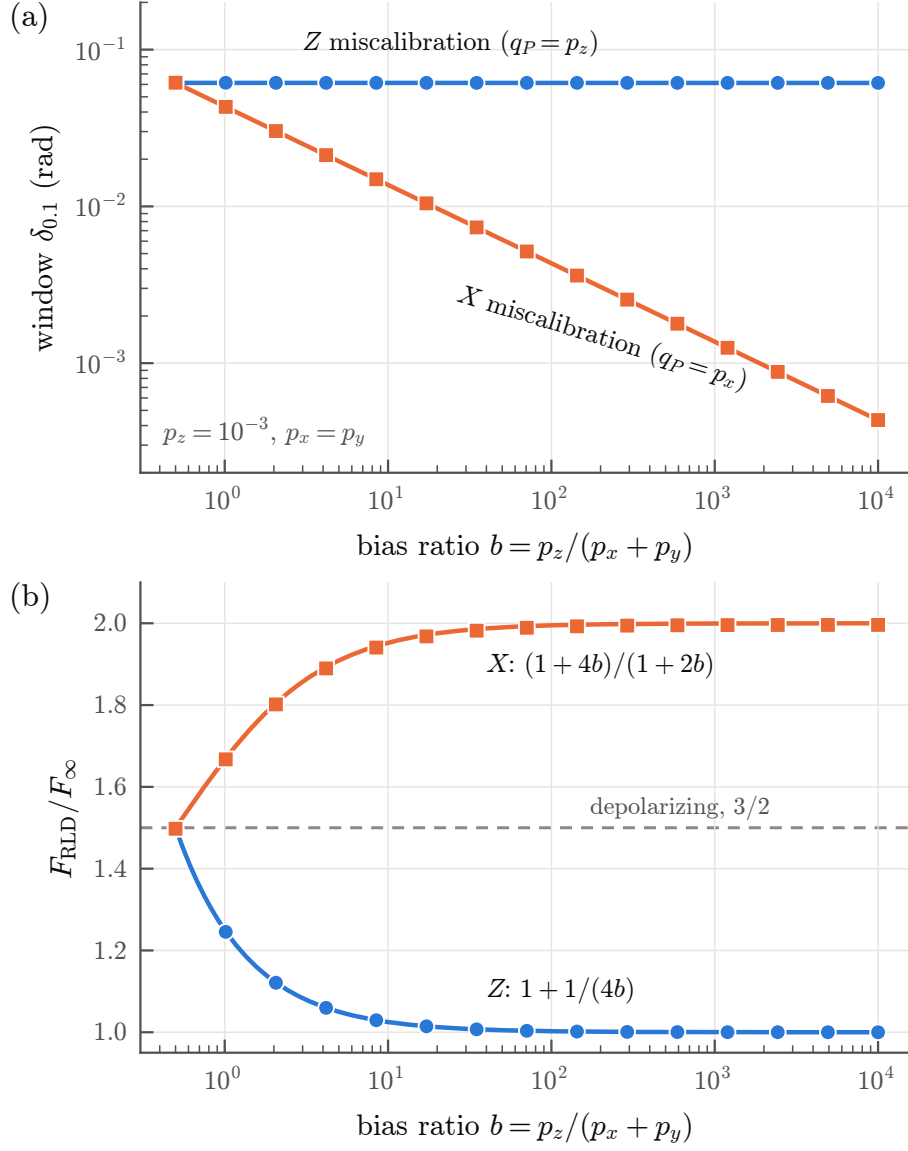


Figure 1.4: Biased qubit noise with $p_z = 10^{-3}$ and $p_x = p_y$, against the bias ratio $b = p_z / (p_x + p_y)$; $b = \frac{1}{2}$ is depolarizing noise. (a) Window $\delta_{0.1}$ at which $\hat{F}(\mathcal{N}_\delta, \mathcal{N}_0) = 0.9$, for Z miscalibrations (circles) and X miscalibrations (squares); markers are exact, lines the law $\delta_{0.1} = 1.937\sqrt{q_P/q_1}$. (b) The ratio $F_{\text{RLD}}/F_{\infty}$, which bounds $\limsup_{\delta \rightarrow 0} D_{\text{BS}}/\zeta$; lines are the leading-order laws $1 + 1/(4b)$ (Z) and $(1 + 4b)/(1 + 2b)$ (X), and the dashed line is the depolarizing value $3/2$. Errors aligned with the bias keep a wide window and a converse within $3/2$ of the exponent; misaligned errors have a narrow window and a gap that approaches two.

We do not know whether Conjecture 1.6.2 holds; for the explicit strategies we evaluated, the relative-entropy rate exceeds the SLD rate of the same strategy, so the limit could lie strictly below $F_{\text{RLD}}/F_{\infty}$.

The same argument applies beyond the depolarized qubit. For every Pauli channel and Pauli-string miscalibration the asymptotic adaptive Fisher information is $F_{\infty} = (1 - \Gamma)/\Gamma$, with $\Gamma = \sum_{\{Q, QP\}} 4q_Q q_{QP} / (q_Q + q_{QP})$ summed over the pairs of Pauli strings coupled by the error, which extends the qubit formula of Ref. [131] (Theorem 8.12.4 in Chapter 8). So the local bracket is $1 \leq D_{\text{BS}}/\zeta \leq F_{\text{RLD}}/F_{\infty}$ in general. Under Z -biased noise it tightens to $1 + 1/(4b)$ for a Z miscalibration, and under pure dephasing the local Stein coefficient is known exactly, $\lim_{\delta \rightarrow 0} \zeta/\delta^2 = F_{\infty}/2 = F_{\text{RLD}}/2$. For an X miscalibration it widens towards 2 [Fig. 1.4(b)], so the narrow X window is still genuine up to a factor two, and for general single-qubit Pauli noise the ratio is unbounded at leading order (Eq. (8.12.15)). The gap between converse and exponent therefore depends on how the error couples to the noise. For a two-qubit ZZ miscalibration under local depolarizing noise, weight-one errors limit $F_{\infty} \simeq 2/p$ while $F_{\text{RLD}} \simeq 4/p^2$, and the bracket widens to $2/p$; for weight w it widens as $p^{-\lfloor w/2 \rfloor}$. Whether ζ inherits the $\delta \sim p$ window of D_{BS} for correlated errors, or whether that window is a spurious fragility, is open.

At finite δ we bracketed ζ numerically. Lower bounds ζ_- are exact relative entropies of explicit sequential strategies with finitely many uses, including one that uses a two-qubit code to herald bit-flip errors. Upper bounds ζ_+ are the smallest of D_{BS} , the continuity bound (8.3.2) of Chapter 8 applied at the pair $(\mathcal{N}_p, \mathcal{N}_p)$, the bound $\zeta \leq \ln(4/p)$, and the convex-optimization divergences $D^{\#}$ of Fawzi and Fawzi [43] for one and two copies, whose feasible points we verified in high-precision arithmetic. The upper and lower bounds on ζ differ by at most a factor 4.1. Figure 1.5 shows the result against $\bar{\delta} = \delta/[p \ln(1/p)]$. For $\bar{\delta} \lesssim 1$ the converse coincides with the certified upper bound, $\zeta_+ = D_{\text{BS}}$, and the numerical bracket on D_{BS}/ζ is $[1, 3.16]$ for all three values of p . Its upper end lies above the proved limit of Eq. (1.6.2) because the explicit strategies reach only $p\zeta_-/\delta^2 \rightarrow 0.158$, rather than the asymptotic $pF_{\infty}/2 \rightarrow 1/3$. Above $\bar{\delta} \approx 1$ the certified ratio D_{BS}/ζ_+ grows. It peaks at 1.81, 3.74 and 9.15 for $p = 10^{-2}$, 10^{-3} and 10^{-4} , at $\delta \approx 2\sqrt{p}$ ($\delta = 0.23, 0.069$ and 0.023), where $\hat{F} \approx 0.66\text{--}0.68$; the peak value scales as $0.8/(\sqrt{p} \ln(1/p))$. These peak values are set by the Fawzi–Fawzi bounds alone; the continuity bound (8.3.2) enters ζ_+ only at the largest δ .

Chernoff exponent. Symmetric discrimination behaves in the same way. Here $\hat{C} = -\ln \hat{F} \simeq F_{\text{RLD}}\delta^2/8$, and the argument behind Eq. (1.6.2), with the quantum Chernoff bound in place of the relative entropy, gives $F_{\infty}/8 \leq \liminf_{\delta \rightarrow 0} \xi/\delta^2 \leq \limsup_{\delta \rightarrow 0} \xi/\delta^2 \leq F_{\text{RLD}}/8$, so $1 \leq \hat{C}/\xi \leq F_{\text{RLD}}/F_{\infty} < 3/2$ as $\delta \rightarrow 0$ (Proposition 8.10.2 in Chapter 8). Single-use strategies reach only $\xi \approx \delta^2/4$, without the factor $1/p$, so the $1/p$ curvature again comes from adaptive accumulation with error detection. At finite δ we bound ξ by refereed one-copy converses evaluated at the midpoint $\delta/2$. The certified overshoot $\hat{C}/\xi$ then reaches 1.13, 2.48 and 6.00 for $p = 10^{-2}$, 10^{-3} and 10^{-4} , near $\delta \approx 2.9\sqrt{p}$, and as $p \rightarrow 0$ it is at least $0.27/(\sqrt{p} \ln(1/p))$.

Table 1.2 collects the verdicts. The crossover has a simple origin. The exponent grows at least as $\delta^2/(3p)$, and at most as $\delta^2/(2p)$, only until it reaches its own continuity bound, which is of order $\delta \ln(1/p)$; this happens at $\delta \sim p \ln(1/p)$. Beyond that point ζ grows at most linearly in δ , whereas D_{BS} continues to grow as $\delta^2/(2p)$ until $\hat{F}$ approaches saturation at the window $\delta \sim \sqrt{p}$ of Eq. (1.5.2). The three regimes of Sec. 1.3 thus reappear at a noisy gate on three scales: genuine fragility for $\delta \lesssim p \ln(1/p)$, a spurious overshoot that grows up to its peak near $\delta \approx 2\sqrt{p}$, and saturation beyond, where the overshoot recedes. For estimation, which probes the local regime, the $1/p$ curvature of the RLD-type bounds is physical. For discrimination of gates that differ by more than $p \ln(1/p)$, it is mostly an artifact of single-

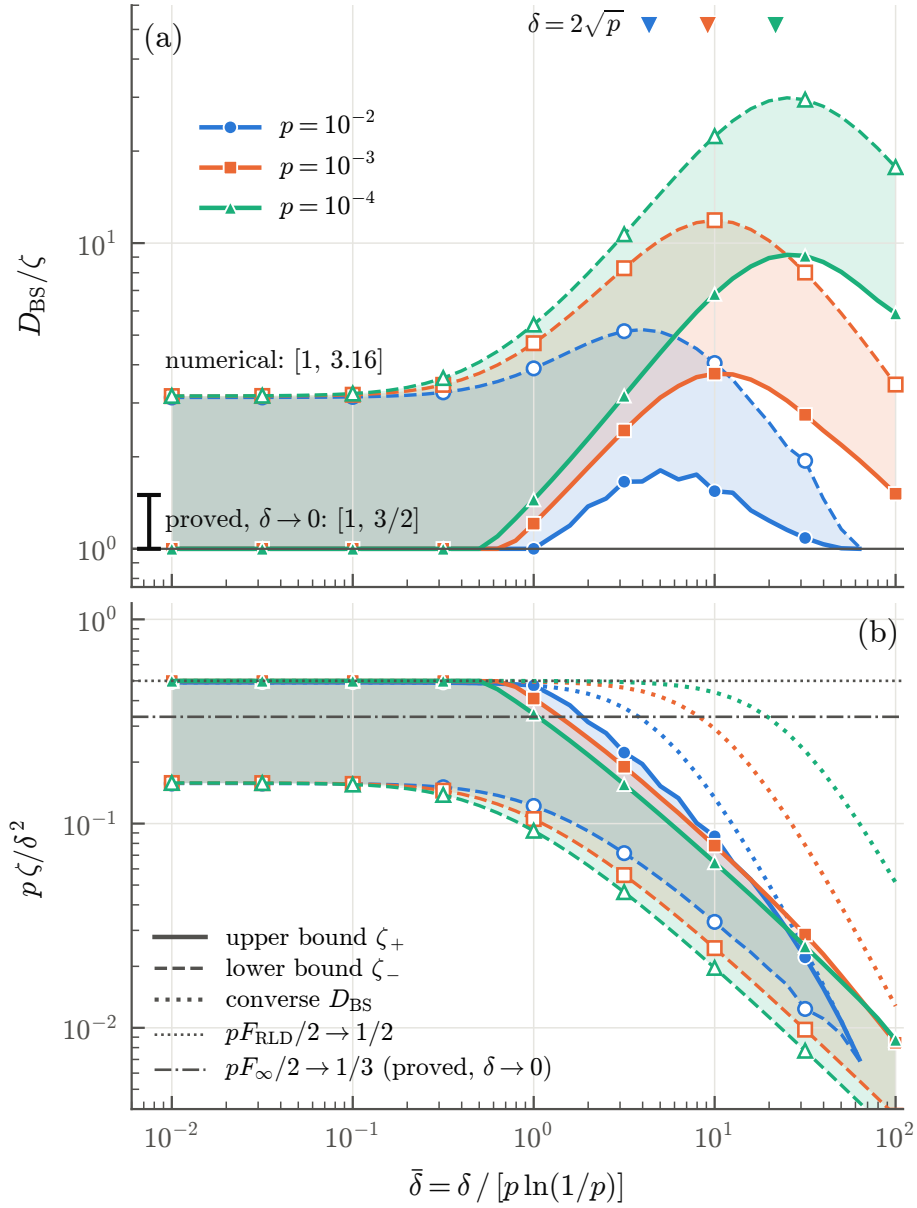


Figure 1.5: Crossover from genuine to spurious fragility for the adaptive Stein exponent $\zeta(\mathcal{N}_{p,\delta} \| \mathcal{N}_p)$ of a depolarized qubit gate, for $p = 10^{-2}, 10^{-3}, 10^{-4}$, against $\bar{\delta} = \delta / [p \ln(1/p)]$. Here $\zeta_- \leq \zeta \leq \zeta_+$ are lower bounds from explicit finite-length strategies and certified upper bounds. (a) Ratio of the Belavkin–Staszewski converse to ζ : solid lines divide by ζ_+ and are certified lower bounds on the ratio; dashed lines divide by ζ_- and are upper bounds. For $\bar{\delta} \lesssim 1$ the numerical bracket is $[1, 3.16]$; the bar marks the narrower bracket $[1, F_{RLD}/F_\infty] \subset [1, 3/2]$ that Eq. (1.6.2) proves in the limit $\delta \rightarrow 0$. Arrows mark $\delta = 2\sqrt{p}$, near which the certified overshoot peaks. (b) $\zeta_{\pm}$ and D_{BS} in units of δ^2/p , with the local limits $pF_\infty/2 \rightarrow 1/3$ (a proved lower bound on $p\zeta/\delta^2$ as $\delta \rightarrow 0$) and $pF_{RLD}/2 \rightarrow 1/2$. Below $\delta \approx p \ln(1/p)$ the converse follows ζ ; above it, the converse exceeds ζ by a factor that grows as p decreases.

Table 1.2: Genuine and spurious fragility of the converses, by sense of sensitivity and task. “Genuine” means that the operational exponent has the same sensitivity as the converse (up to the stated factor); “spurious” means that only the converse has it. The first row concerns amplitude damping, the second the worst case of Theorem 1.4.1, and the last two the depolarized qubit gate $(\mathcal{N}_{p,\delta}, \mathcal{N}_p)$; under biased noise the local ratio stays below 2 for single-qubit errors, and for correlated two-qubit errors the local bracket widens to $2/p$ (Sec. 1.6).

regime	Stein: ζ vs. D_{BS}	Chernoff: ξ vs. $\hat{C}$ and $-2 \ln \hat{F}$
Choi supports (Sec. 1.3)	genuine: $\zeta(\mathcal{A}_\gamma \ \mathcal{A}_{\gamma'}) = \infty$	spurious: converses = ∞ while $\xi \rightarrow 0$
first order, per unit of $\ \Psi' - \Psi\ _\diamond$	genuine: both at most $1/(2\tilde{\lambda})$, equal at the worst case	genuine: both at most $1/(2\tilde{\lambda})$, equal at the worst case
second order, $\delta \lesssim p \ln(1/p)$	genuine: $1 \leq D_{\text{BS}}/\zeta \leq F_{\text{RLD}}/F_\infty < 3/2$ as $\delta \rightarrow 0$	genuine: $1 \leq \hat{C}/\xi \leq F_{\text{RLD}}/F_\infty < 3/2$ as $\delta \rightarrow 0$
second order, $p \ln(1/p) \lesssim \delta \lesssim \sqrt{p}$	spurious: D_{BS}/ζ up to 1.81, 3.74, 9.15 for $p = 10^{-2}, 10^{-3}, 10^{-4}$; $\approx 0.8/(\sqrt{p} \ln \frac{1}{p})$	spurious: $\hat{C}/\xi$ up to 1.13, 2.48, 6.00 for $p = 10^{-2}, 10^{-3}, 10^{-4}$; $\approx 0.5/(\sqrt{p} \ln \frac{1}{p})$

letter bounds. A better target is then ζ itself, which is computable [117, 124] although no convergence rate is known, or its continuity bound.

1.7 Certification from finite data

A converse computed from a tomographic estimate is a statement about the estimate. To make it a statement about the device, one needs a confidence region and a bound valid on all of it. Equation (1.4.1) provides this.

Theorem 1.7.1 (Certified converse). *Suppose that, with probability at least $1 - \alpha$, the true channels lie within diamond distance ϵ of channel estimates $\hat{\Phi}, \hat{\Psi}$ (completely positive and trace preserving) whose Choi floors are at least $\hat{\lambda}$, and let $r = d_X \epsilon / (2\hat{\lambda}) < 1$. Then, with the same probability, the true floors are at least $(1 - r)\hat{\lambda}$ and*

$$\xi(\Phi, \Psi) \leq \hat{C}(\Phi, \Psi) \leq \hat{C}(\hat{\Phi}, \hat{\Psi}) + \ln \frac{1}{1 - r}. \quad (1.7.1)$$

An analogous bound for ζ adds the continuity bound of Sec. 1.4 to $D_{\text{BS}}(\hat{\Phi} \| \hat{\Psi})$ (Chapter 8, Section 8.13). Both bounds are one-sided in the direction that keeps the converse valid, which is the direction that matters for certification. Plug-in values carry no such guarantee. In our simulations the plug-in $\hat{F}$ exceeded the true value, which makes the converse claim a stronger limit than is true, in up to 32% of runs for linear inversion, depending on p and the sample size, and in up to 96% of runs for a hedged maximum-likelihood estimate [22] at small sample sizes (Section 8.13).

By the sharpness in Eq. (1.4.3), a diamond ball of radius ϵ contains pairs whose values of $\ln \hat{F}$ differ by $\epsilon/(4\tilde{\lambda})$, so an additive accuracy η requires $\epsilon \lesssim 4\eta\tilde{\lambda}$. Diamond-norm tomography [93] reaches this with $O((d_X d_Y k + d_X \ln \frac{1}{\alpha})/(\eta\tilde{\lambda})^2)$ uses for channels of Kraus rank k ; this learner is coherent, since it processes the channel outputs jointly. For learners that measure each channel output separately, possibly choosing later inputs from earlier outcomes, no strategy reaches diamond accuracy ϵ on the gate family of Eq. (1.2.3) with fewer than order $1/\epsilon^2$ uses (Proposition 8.13.1(b) in Chapter 8; see also Ref. [102]), hence order $1/(\eta\tilde{\lambda})^2$. For such incoherent learners, certification through a diamond ball therefore

costs order $1/(\eta\tilde{\lambda})^2$ uses. Whether coherent learners can do better is open. The two-point argument behind the incoherent bound does not extend to them: with error-corrected coherent probing the Fisher information per use for the rotation angle is $F_\infty \approx 2/(3p)$, and the same argument then gives only order $p/\epsilon^2 \approx 1/(\eta^2 p)$ uses, no more than the intrinsic cost below.

This cost is not intrinsic to the task. For Pauli channels the relevant information sits in the error rate q of a floor sector, and a relative change η of q carries information of order $\eta^2 q$ per use. For local depolarizing noise, any strategy, adaptive or coherent, needs at least $0.0246/(\eta^2 \tilde{\lambda})$ uses to estimate $\hat{F}(\text{Ad}_{PU}, \Psi)$, the fidelity of the noisy gate with a correlated error of full weight, to relative accuracy better than $\eta \leq 0.1$. Bell-basis estimation of the Pauli error rates in a known frame [45, 52] achieves $O(\ln(1/\alpha)/(\eta^2 \tilde{\lambda}))$. The separation is therefore between variance-blind confidence regions, whose radius is set by the largest outcome variances, and variance-adaptive ones, whose width in the floor directions shrinks with the floor; it is not specific to the diamond norm. We conjecture that $\tilde{O}(\text{poly}(d)/(\eta^2 \tilde{\lambda}))$ uses suffice for all channel pairs.

Simulations of single-qubit Pauli tomography illustrate both scalings (Fig. 1.6). For the pair with $\delta = 2\sqrt{p}$, where $\hat{F} \approx 0.71$, we certified a lower bound on $\hat{F}$, hence an upper bound on the converse, to accuracy $\eta = 0.05$ in $\ln \hat{F}$ at 95% confidence. In the first procedure, Theorem 1.7.1 is applied to a diamond-norm ball around the linear-inversion estimate, whose Hoeffding radius does not adapt to the small outcome probabilities in the floor directions. Its cost, 3.6×10^{13} uses per channel at $p = 10^{-3}$, follows from that radius by construction and is proportional to p^{-2} . In the second procedure, the Pauli error rates are estimated directly in the known frame with exact binomial intervals, and $\hat{F}$ is minimized exactly over the resulting confidence polytope; the cost grows as $p^{-1.00 \pm 0.04}$ and is 1.2×10^7 uses at $p = 10^{-3}$. The certified bounds held in all 1000 repetitions at every p . Through the same diamond ball, the hockey-stick divergence and the Uhlmann fidelity are certified with a number of uses independent of p . Bounds of the opposite type, which limit distinguishability from below, show the same separation: variance-blind isotropic regions cost order p^{-2} , whereas a box of exact binomial intervals on the setting probabilities, which is variance-adaptive and needs no model assumption, already costs order p^{-1} (Chapter 8, Section 8.13). The second procedure assumes Pauli noise in the known frame; whether model-free tomography attains the $1/\tilde{\lambda}$ law on the converse side is open. For a two-qubit CNOT gate under local noise, plug-in (not certified) estimates of $\hat{F}$ needed a number of shots growing approximately as $p^{-4} \propto 1/\lambda^2$, with $\lambda = p^2/4$, in line with the multi-qubit floor (three points on a one-decade grid).

1.8 Discussion

The sensitivity of geometric channel divergences near noisy gates has two senses. At first order, a diamond-norm error in the model moves the converses by up to order $1/\tilde{\lambda}$ per unit of distance, with the sharp constant of Theorem 1.4.1. At the worst case the adaptive exponents move at the same rate (Proposition 1.6.1), so in this sense the sensitivity belongs to the task, and a first-order gap between converse and exponent requires a converse that is already loose. At second order, the curvature of the converses in a coherent miscalibration grows as $1/p$. In this sense the sensitivity has two sources. One is physical: adaptive strategies can exploit the noise floor, and the Stein and Chernoff exponents and the error-corrected Fisher information inherit a $1/p$ scale for small miscalibrations. The other is an artifact of single-letter bounds, which keep the $1/p$ curvature well beyond the point where the exponent has crossed over to its $\ln(1/p)$ continuity bound, and which can be

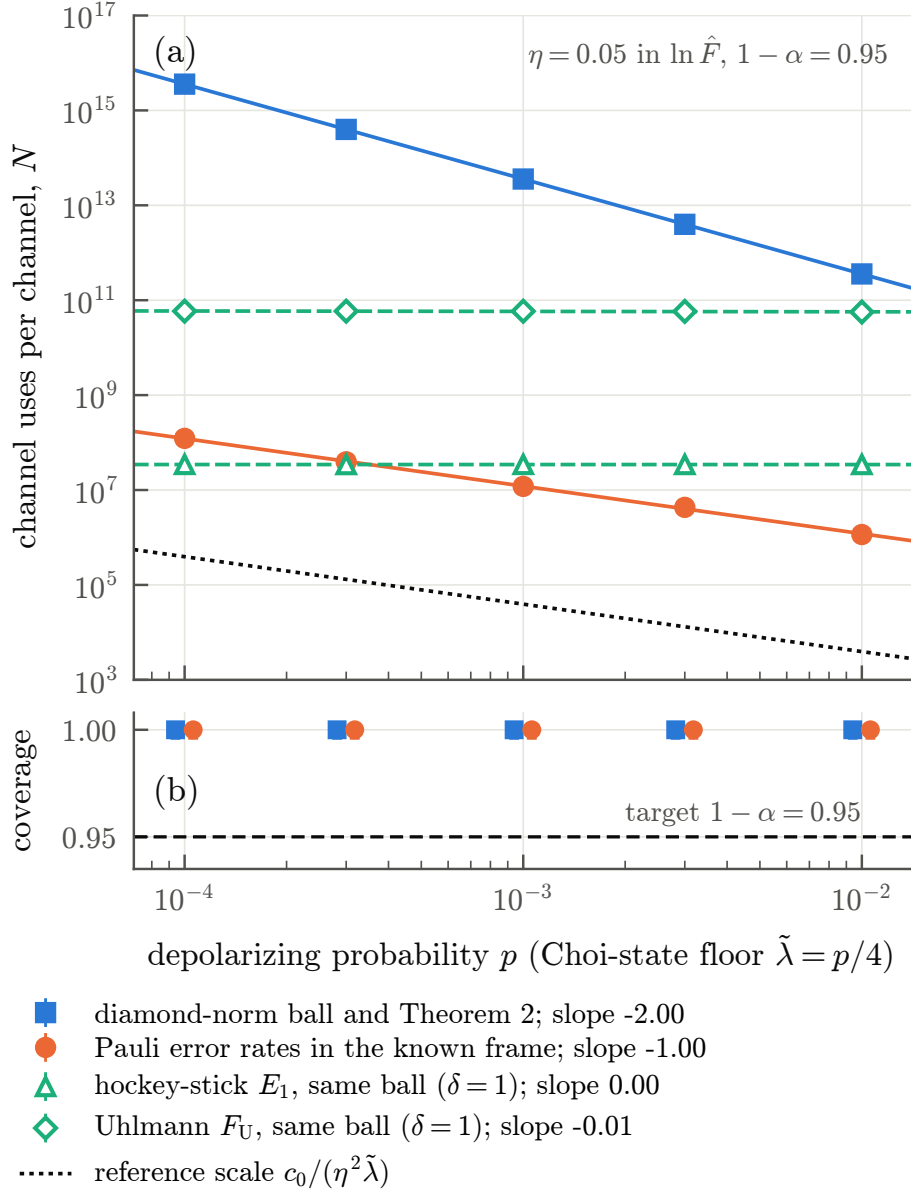


Figure 1.6: Cost of certifying converses from simulated single-qubit Pauli tomography. (a) Channel uses per channel for a certified lower bound on $\hat{F}(\mathcal{N}_{p,\delta}, \mathcal{N}_p)$ at $\delta = 2\sqrt{p}$, i.e., an upper bound on the converse $-\ln \hat{F}$, to accuracy $\eta = 0.05$ at confidence $1 - \alpha = 0.95$ (median run; error bars are bootstrap intervals, lines are fits). A variance-blind diamond-norm ball combined with Theorem 1.7.1 costs order p^{-2} by construction; variance-adaptive estimation of the Pauli error rates in the known frame costs order p^{-1} . Through the same ball, the hockey-stick divergence and the Uhlmann fidelity at $\delta = 1$ cost the same at every p . The dotted line is the reference scale $c_0/(\eta^2 \tilde{\lambda})$, proved as a lower bound for the related estimand $\hat{F}(\text{Ad}_{PU}, \Psi)$, not for the pair simulated here. (b) Empirical coverage of the certified bounds in 1000 runs per point; the target is 0.95.

infinite where the exponent vanishes. For an experimentalist the practical consequences are concrete. A converse computed from a pure- T_1 or noiseless model can be infinite for reasons that have nothing to do with the device; the residual thermal population or depolarizing floor sets its value and must be part of the model. Error budgets for geometric converses must be set relative to the Choi floor, which for n qubits under local noise is $(p/4)^n$, and, for a given miscalibration, relative to the rate of the corresponding Pauli error, which under biased noise can be far smaller than the dominant error rate. And confidence regions should be variance-adaptive rather than isotropic in the diamond norm, which for Pauli channels saves a factor of order $1/\tilde{\lambda}$ in data.

Several questions remain open. We do not know whether Conjecture 1.6.2 holds, where in its bracket the local Chernoff coefficient lies, whether unitary channels give the worst-case slope of $\hat{F}$, whether the exponent inherits the narrow window of the converses for correlated multi-qubit errors, whether coherent learners can certify through a diamond ball with fewer than order $1/(\eta\tilde{\lambda})^2$ uses, or whether the intrinsic certification cost is of order $1/(\eta^2\tilde{\lambda})$ for all channels. For amplitude damping, the conjectured bound $\xi \leq -2\ln f$ would pin ξ within a factor two. A single-letter converse that stays finite at spurious pairs, such as $\min\{\hat{C}, C_{\text{bar}}, -2\ln \mu_*\}$, and the extension of these laws to bosonic loss channels with energy constraints, are natural next steps.

1.9 Plan of the book

The rest of the book develops the theory behind this chapter. Chapter 2 is an overview: it explains the three classical “shadows” of a pair of states or channels that organize the theory, summarizes the main results of each chapter, shows how the chapters depend on each other and suggests reading orders. Chapter 3 fixes the notation and records where the conventions of the chapters differ. Chapter 4 introduces the Arveson–Radon–Nikodym derivative of a pair of channels and proves the continuity dichotomy and the Choi-floor estimates used in Sections 1.3 and 1.4. Chapter 5 treats testing regions and hockey-stick divergences and proves the continuity properties of the adaptive Stein exponent used in Sections 1.3 and 1.4; Chapter 6 studies the super-multiplicativity of the hockey-stick Rényi quasi-entropies under tensor products. Chapter 7 describes adaptive discrimination as a controlled random walk and proves the geometric Chernoff converse of Eq. (1.2.2) and the fragility of single-letter Chernoff converses discussed in Section 1.3. Chapter 8 contains the proofs of the results stated in this chapter, together with closed forms for Pauli channels, the analysis of amplitude damping, the certification bounds and the numerical methods. Chapter 9 collects the open problems of all chapters, including those listed above.

Chapter 2

Overview and reading guide

This book is about one question: when does a quantum divergence between two channels, or an operational quantity that it bounds, depend continuously on the channels, and how large is the error when it does? Chapter 1 has introduced the question through its physical motivation and stated the answers for characterized gates. Chapters 4–7 develop the general theory, and Chapter 8 contains the proofs of the results of Chapter 1. This chapter explains the organising idea of three classical “shadows” of a pair, the main results of each chapter and how the chapters depend on each other, and suggests reading orders. Chapter 3 fixes the notation and records where the chapters differ in their conventions, and Chapter 9 collects the open problems.

2.1 From noisy gates to the general questions

The single-letter converses of Chapter 1, the geometric channel fidelity $\hat{F}$, the geometric Rényi divergences $\hat{D}_\alpha$ and the Belavkin–Staszewski relative entropy D_{BS} , are each built from a Kubo–Ando mean of the Choi matrices of the two channels [42, 72]. That chapter showed how badly they can depend on the model of the channels. Their discontinuities are not limited to ideal gates: two pairs of replacement channels with the same sum can be arbitrarily close in the diamond norm and have geometric fidelities 0 and 1 (Corollary 4.3.3), just as an ideal unitary gate has geometric fidelity 1 with itself and 0 with every miscalibrated copy (Corollary 4.12.6).

This raises several general questions, which organise the rest of the book.

- (a) In which coordinates are these divergences continuous, and with which moduli? Where exactly are they discontinuous in the diamond norm?
- (b) Which divergences are continuous everywhere, and why?
- (c) Do the operational quantities, such as the Stein and Chernoff exponents of adaptive discrimination, share the discontinuities of their converses? When a converse jumps, does the task jump with it?
- (d) Near realistic noisy gates, where continuity holds, how large is the modulus, and how does it compare with the sensitivity of the task itself?

Chapter 1 answered (d) and the physical side of (c). The answers to (a)–(c) depend on the family of divergences, and the main structural message is that each family is a functional of a different classical object attached to the pair.

2.2 Three shadows of a pair

For a pair of positive operators, or of completely positive maps, there are several ways to extract a classical pair of measures, and each of the standard families of quantum divergences is a classical divergence of one of them. We call these classical objects the *shadows* of the pair. Three of them carry most of this book.

(I) *The reverse test.* For states ρ, σ with $\omega = \rho + \sigma$, the operator $T = \omega^{-1/2} \rho \omega^{-1/2}$ (on the support of ω) is a positive contraction, and its spectral measure in the state ω gives a classical pair. Its classical divergences are the maximal f -divergences of Matsumoto [91] and the Kubo–Ando means; in particular $\hat{F}$, $\hat{D}_\alpha$ and D_{BS} belong to this family. For completely positive maps the role of T is played by the Arveson–Radon–Nikodym (ARN) derivative of Φ with respect to $\Omega = \Phi + \Psi$: if Ω has linearly independent Kraus operators $V_1, \dots, V_r$, then $\Phi(X) = \sum_{k,l} T_{kl} V_k X V_l^*$ for a unique $0 \leq T \leq \mathbf{1}$ (Theorem 4.2.2). Its transpose is the effect, on the environment of a Stinespring dilation of Ω , that selects the branch Φ (Theorem 4.2.5). Every Kubo–Ando mean of the Choi matrices is $W g_\sigma(T) W^*$ for a scalar function g_σ and the matrix W of Kraus vectors (Theorem 4.2.15). This is the shadow of Chapters 4 and 7.

(II) *The testing region.* The hockey-stick profile $e(\gamma) = \text{Tr}(\rho - \gamma\sigma)_+$, $\gamma \geq 0$, encodes all the binary tests between ρ and σ . Its second derivative $\mu = e''$ is a measure on $(0, \infty)$, completed by boundary atoms at 0 and ∞ . The hypothesis-testing quantities, the Hirche–Tomamichel f -divergences [63] and, by Frenkel’s formula [46], the Umegaki relative entropy are functionals of this shadow. It is the shadow of Chapters 5 and 6.

(III) *The modular shadow.* The Nussbaum–Szkoła pair of the relative modular operator [99] gives the Petz f -divergences. Through the modular derivative $\tilde{T} = \Delta(\mathbf{1} + \Delta)^{-1}$ it can be treated with the same tools as the reverse test (Section 5.5).

Two further families appear as limits or suprema of these: the measured divergences, a supremum over many classical pairs, and the sandwiched and regularised divergences, which appear only after tensor powers. Section 5.8 contains a table that places all five families side by side.

The three shadows are ordered. The testing-region shadow is a garbling of the reverse test (Theorem 5.3.2), and the modular shadow has the boundary masses of the testing region, not those of the reverse test (Theorem 5.5.1). This ordering explains the difference in continuity. The reverse test can be destroyed by an $O(\sqrt{t})$ perturbation that couples the kernel of the sum to its range, so the obstruction to continuity for family (I) is the invertibility of the sum $J(\Phi + \Psi)$. The testing region does not see that mechanism, and the only obstruction left for family (II) is a leak of support, which is controlled by the invertibility of the second argument.

2.3 Results by chapter

The results of Chapter 1 are listed in its introduction and are not repeated here. The other chapters contain the following main results.

Chapter 4: continuity of Kubo–Ando channel divergences

Chapter 4 develops family (I) for completely positive maps. Since every quantity of the family is a scalar function of the ARN derivative, read through the Kraus operators of the sum, it is continuous in T with a dimension-free constant. For the fidelity,

$$|\hat{F}(\Phi_1, \Psi_1) - \hat{F}(\Phi_2, \Psi_2)| \leq \|\Omega^*(\mathbf{1}_Y)\|_\infty \|T_1 - T_2\|_\infty^{1/2}$$

for two pairs with the same sum Ω , and the constant and the exponent are sharp (Theorem 4.4.6). Analogous bounds hold for all means, for $\widehat{Q}_\alpha$ with $\alpha \in (0, 1)$ and for bounded maximal f -divergences (Theorems 4.4.5, 4.5.5 and 4.5.13), and, under a bound on $D_{\max}$ within each pair, for D_{BS} and $\widehat{Q}_\alpha$ with $\alpha \in (1, 2]$ (Theorem 4.6.7). Every maximal f -divergence of an operator convex f has a closed form as the largest eigenvalue of a partial trace and is continuous in T with values in $(-\infty, +\infty]$ (Theorems 4.7.5 and 4.7.6). Section 4.9 gives explicit moduli for every mean and every operator convex f .

The passage to the diamond norm goes through the quantity $\nu(\Omega)$, the smallest operator norm of a unit combination of the Kraus operators of the sum, which bounds the derivative distance by the diamond distance (Theorem 4.8.4, Corollary 4.8.6). Pairs with different sums are compared through the Thompson metric (Theorem 4.10.4, Corollary 4.10.7), and support changes through the Lebesgue decomposition of the sums (Theorem 4.11.3). The main structural result is a dichotomy (Theorem 4.12.4): on pairs of completely positive maps, $\widehat{F}$ is continuous at (Φ, Ψ) if and only if $J(\Phi + \Psi)$ is invertible or $\widehat{F}(\Phi, \Psi) = 0$, with analogues for the other quantities. Within channels the dichotomy holds at diagonal pairs (Φ, Φ) , where output rotations already produce the jump (Theorem 4.12.9, Corollary 4.12.10), and zero-temperature amplitude damping jumps from 1 to 0 within its own family (Example 4.12.8). Quantitatively, if the Choi matrices are bounded below by $\lambda > 0$, the fidelity is Lipschitz in the diamond norm with a relative constant of the sharp order λ^{-1} (Theorem 4.8.8, Example 4.8.11), and at mutually singular pairs it tends to zero at the sharp rate $d^{1/2}/\sin \theta$ (Theorem 4.12.11).

Chapter 5: testing regions, modular lifts and continuity

Chapter 5 carries the programme of Chapter 4 through families (II)–(V). The testing-region shadow has atoms exactly at the relative spectrum, dominated by those of the reverse test, and an absolutely continuous part whose density is the coupling of σ across the Helstrom test (Theorems 5.3.1 and 5.3.2). One consequence is an explicit nonnegative integral formula for $D_{\text{BS}} - D$, and another is that the measured, testing-region and maximal divergences of a strictly convex function coincide only for commuting pairs (Corollary 5.3.3). Every testing-region f -divergence with finite boundary values is uniformly continuous with a dimension-free modulus, so these divergences have no dichotomy (Theorem 5.4.1, Corollary 5.4.2), and the Umegaki relative entropy is continuous at (ρ, σ) exactly when σ is invertible or $D(\rho \parallel \sigma) = \infty$ (Theorem 5.4.3). Bounded Petz divergences are continuous everywhere, with a dimension-free Hölder modulus (Theorem 5.5.2).

For channels the chapter replaces the trace norm by the Thompson metric of the Choi matrices and the relative entropy by a conditional entropy. Every n -use adaptive quantity of a suitable divergence is $n d_{\text{T}}$ -Lipschitz in the second channel (Theorem 5.7.2), and the first argument is controlled uniformly over references (Theorem 5.7.4). Together these give an explicit modulus for the adaptive Stein exponent ζ , and the analogue of the dichotomy: ζ is continuous at (Φ, Ψ) if and only if $J(\Psi)$ is invertible or $\zeta(\Phi, \Psi) = \infty$ (Corollary 5.7.5). The chapter also treats the strong-converse exponent and the regularised sandwiched and Petz divergences (Theorems 5.7.8 and 5.7.11, Corollary 5.7.9), and gives an exact one-shot dual formula for testing-region channel divergences together with their regularised values (Theorems 5.7.13 and 5.7.16).

Chapter 6: super-multiplicativity of hockey-stick Rényi quasi-entropies

The reverse test and the modular shadow factorise under tensor products, but the testing region does not. Chapter 6 studies what replaces multiplicativity for the testing-region

Rényi quasi-entropies Q_α^{HT} , the Mellin moments of the hockey-stick profile. Conjecture 6.3.5 asserts that they are super-multiplicative, so that D_α^{HT} is superadditive for $\alpha > 1$ and subadditive for $\alpha < 1$. The chapter proves the conjecture at $\alpha = 2$ (Theorem 6.4.1), at every integer order for pure states and at $\alpha = 3$ for sign-free pairs (Theorem 6.5.3), at every integer order against pure partners (Theorem 6.7.1), at $\alpha = 3$ whenever one reference state has at most two distinct eigenvalues (Corollary 6.6.3), for one measured factor (Theorem 6.4.2), to first order as $\alpha \rightarrow \infty$ (Theorem 6.4.3) and for tensor powers with enough copies (Proposition 6.8.3). Structurally, the conjecture is the family of moment inequalities of a convex-order domination between two Frenkel measures (Theorem 6.8.2). One-way LOCC tests never beat the classical product of the shadows (Proposition 6.9.2), so a proof must use entangled tests. The chapter also gives an elementary proof of $D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$ for $\alpha \geq 2$ (Theorem 6.3.2). Unlike the other chapters, it is a chapter of matrix analysis: continuity plays no role, and quantum channels appear only as garblings between a pair and its shadow (Section 6.9).

Chapter 7: adaptive discrimination as a controlled random walk

Chapter 7 turns the ARN derivative into a converse for adaptive discrimination. Diagonalising $T = \sum_\ell \tau_\ell P_\ell$ writes the pair as $\Phi = \sum_\ell \tau_\ell \Omega_\ell$ and $\Psi = \sum_\ell (1 - \tau_\ell) \Omega_\ell$ with common completely positive level maps Ω_ℓ (Lemma 7.2.1). Each use of the unknown channel therefore selects a level and then acts in a way that does not depend on the hypothesis, and every adaptive protocol is a post-processing of a controlled random walk whose step values are fixed (Proposition 7.3.1). The result is the geometric Chernoff converse: every adaptive strategy errs with probability at least $\frac{1}{4} \exp(-n\hat{C} - \sqrt{2 \ln 4} \Lambda \sqrt{n})$, where $\hat{C} = -\ln \min_s \lambda_{\min}(B_s)$ is computable from one eigendecomposition (Theorem 7.4.1). No other common decomposition gives a better bound of this kind (Proposition 7.4.3); $\hat{C}$ lies between one half of the fidelity converse and the fidelity converse and is not comparable with the barycentric converse (Corollary 7.4.5, Proposition 7.4.6). If the levels are orthogonal at an optimal input, product strategies attain $\hat{C}$ (Theorems 7.5.1 and 7.5.3).

The continuity section applies Chapter 4 to the converse itself. Under perturbations that are small relative to the pair, $e^{-\hat{C}}$ is Lipschitz (Theorem 7.6.2), which gives a converse valid for a whole neighbourhood of a model pair (Corollary 7.6.3); under general perturbations it is only lower semicontinuous, and continuous where $J(\Phi + \Psi)$ is invertible (Theorem 7.6.4). Near replacement channels with rank-deficient outputs, the fidelity, barycentric and geometric Chernoff converses are all infinite while the adaptive exponent tends to zero (Theorem 7.6.5). The same happens for amplitude-damping channels with nearby damping rates, where a pure-program simulation gives a finite converse of the right order (Theorem 7.6.9, Corollary 7.6.10).

Chapter 8: noisy gates, proofs and further results

Chapter 8 proves the results that Chapter 1 states for characterized gates: the classification of discontinuities into genuine, spurious and absent by the supports of Choi matrices, the Choi-floor law $\ln[1/(1-r)]$ with $r = \epsilon/(2\tilde{\lambda})$ and its sharpness (Section 8.3), the equality of converses and exponents at the worst case (Section 8.9), and the comparison of the converses with the adaptive Stein and Chernoff exponents under a coherent miscalibration, locally and at finite miscalibration (Sections 8.8, 8.10 and 8.11). It also contains closed forms of the divergences and of the asymptotic adaptive Fisher information for all Pauli channels and Pauli-string miscalibrations (Sections 8.4, 8.5 and 8.12), the analysis of amplitude damping and of the program-simulation converse (Section 8.7), the certification bounds

(Section 8.13) and the numerical methods (Section 8.14). Section 8.15 points to the results of Chapters 4, 5 and 7 that these proofs use and proves the refinements of them that are needed.

2.4 How the chapters depend on each other

Figure 2.1 shows the logical dependencies. Chapter 1 comes first because it frames the theory, but it states its results with forward references: they are proved in Chapter 8, with the help of Chapters 4, 5 and 7. Chapter 4 is the foundation: it introduces the ARN derivative, the Thompson metric between sums, the relative distance of two pairs and the dichotomy. Chapter 5 uses the reverse test of Chapter 4 as the reference point for the testing region and the modular shadow, and compares its continuity results with the dichotomy. Chapter 6 uses the definitions and the tensor-product sandwiches of Chapter 5. Chapter 7 uses the derivative, the level decomposition it induces, and the continuity estimates for pairs with different sums from Chapter 4. Chapter 8 draws on all three of Chapters 4, 5 and 7: the Choi-floor estimates and the dichotomy, the continuity of the Stein exponent, and the geometric Chernoff converse with its fragility. Together, Chapters 1 and 8 are largely self-contained. Chapters 6 and 7 are linked only through an open problem, which asks whether the adaptive Chernoff exponent is a regularised testing-region quantity (Problem 9.6.6).

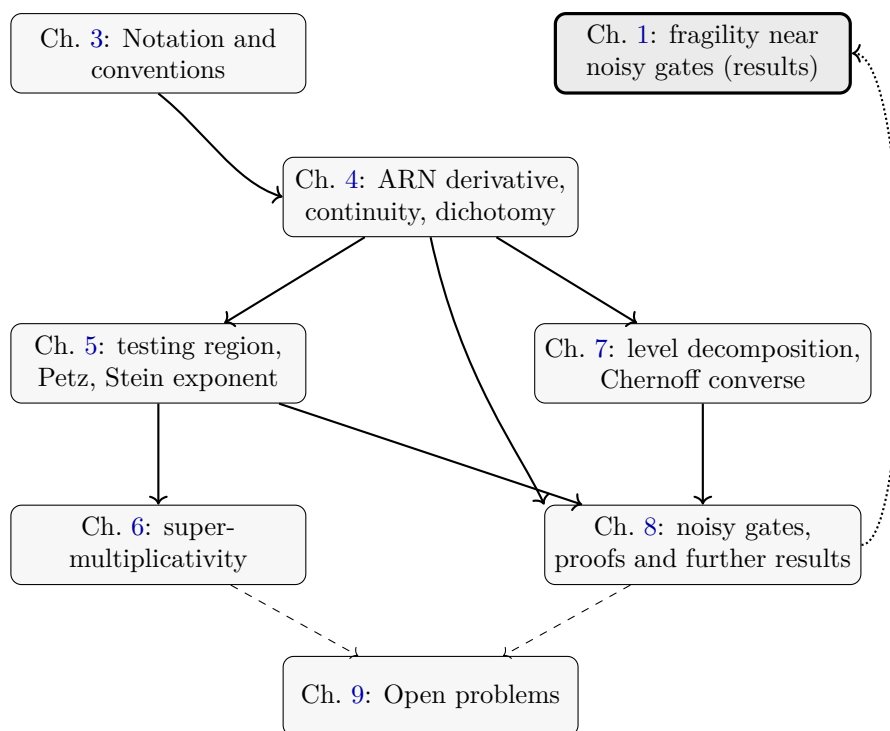


Figure 2.1: Dependencies between the chapters. A solid arrow means that the chapter at its head uses definitions or results of the chapter at its tail. The book opens with Chapter 1 (thick frame), which states results proved in Chapter 8 (dotted arrow). Dashed arrows indicate that the open problems draw on every chapter.

There are also some overlaps that we have kept. Chapter 6 recalls the definitions of the testing region in a compact form, and Chapter 1 restates the results it needs from Chapters 4, 5 and 7 so that it can be read on its own, before them; in both cases the restatements point to the original results.

2.5 A reading guide

The chapters after this one can be read in several orders.

Mathematicians interested in operator theory and matrix analysis can read Chapter 4 in order: Sections 4.2–4.4 contain the ARN derivative, the Kubo–Ando background and the basic continuity estimate, and Section 4.12 the dichotomy. Chapter 5 follows naturally, with Sections 5.2–5.5 on states. Chapter 6 can be read directly after Section 5.2; it is the most self-contained chapter and needs no channels.

Information theorists interested in converse bounds for discrimination can continue with Chapter 7, which needs only the ARN theorem (Theorem 4.2.2) and, in its continuity section, the estimates of Section 4.10. Section 5.7 contains the corresponding results for the Stein exponent and the strong-converse exponent.

Physicists interested in noisy gates and their characterization will find the results in Chapter 1, which states them operationally and defines every symbol it uses, and their proofs in Chapter 8. The remarks headed “Physical interpretation” in Chapter 4 explain the derivative, the quantity $\nu(\Omega)$ and the dichotomy in terms of instruments and environments, and Section 4.8 contains the Choi-floor estimates in their original form.

Readers who move between chapters should consult Chapter 3. Each chapter keeps its own notation, and the conventions differ in a few places, most importantly in the ordering of tensor factors in the Choi matrix, in whether the fidelity of states is squared, and in the notation for the (natural) logarithm (Table 3.1).

2.6 About this book

Chapter 4 is a revised and extended version of [69]. Theorems are numbered by chapter and section, and every cross-reference between chapters is internal.

Numerical verification

The results of this book have been checked numerically. Every result of Chapter 4 went through an independent, computer-assisted audit: its proofs were re-derived, and every formula and constant was tested numerically, in 50–110-digit arithmetic near singular points and with diamond norms and Schur-multiplier norms computed exactly by semidefinite programming. No statement was found to be false and no proof was found to have a gap; the minor corrections the audit produced (missing hypotheses, edge cases and one display) are incorporated in Chapter 4. Every numerical value, table and figure in Chapters 1 and 8 and in Section 7.6.1, and the hockey-stick estimate of Chapter 5, is reproduced by a script, and the proved statements of those chapters were cross-checked numerically. The code, the data, the audit scripts and the audit reports are in the `code/` directory of the repository that accompanies this book, <https://github.com/mjones1298/Continuity-of-divergences-python->: `code/README.md` maps each claim to the script that produces it, and `code/audit_ch04/` contains the audit of Chapter 4. The numerical checks reported in Sections 5.9, 6.11 and 7.7 were computed with the author’s own code, which is not included.

Chapter 3

Notation and conventions

Each of Chapters 1 and 4–8 keeps its own notation; Chapters 1 and 8, on noisy gates, share one notation. This chapter collects the definitions that are used throughout, in a common notation close to that of Watrous [127], and Section 3.11 lists the places where the chapters differ. A reader who moves from one chapter to another should check Tables 3.1 and 3.2 first. All Hilbert spaces are finite-dimensional.

3.1 Spaces, operators and maps

Complex Euclidean spaces are denoted by calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{R}, \dots$, and we write $d_{\mathcal{X}} = \dim \mathcal{X}$ (also d_X). The standard basis of $\mathbb{C}^n$ is $e_1, \dots, e_n$, and $E_{ab} = e_a e_b^*$. We write $L(\mathcal{X}, \mathcal{Y})$ for the linear operators from $\mathcal{X}$ to $\mathcal{Y}$, $L(\mathcal{X}) = L(\mathcal{X}, \mathcal{X})$, $\text{Pos}(\mathcal{X})$ for the positive semidefinite operators and $D(\mathcal{X})$ for the density operators (states). The identity operator is $\mathbf{1}$ or $\mathbf{1}_{\mathcal{X}}$, and the identity map on $L(\mathcal{X})$ is id or $\mathbf{1}_{L(\mathcal{X})}$. The adjoint of an operator or a map is A^* ; physics notation $A^\dagger$ is used in Chapters 1 and 8.

For Hermitian A, B , $A \leq B$ means that $B - A$ is positive semidefinite (the Loewner order). The extreme eigenvalues are $\lambda_{\min}$ and $\lambda_{\max}$, the support projection of $A \geq 0$ is Π_A , and $\text{range } A$ is its range. The trace is tr or Tr , and $\text{Tr}_{\mathcal{Y}}$ is the partial trace over $\mathcal{Y}$. The Schatten norms are $\|\cdot\|_p$, with $\|\cdot\|_1$ the trace norm and $\|\cdot\|_\infty$ the operator norm. For Hermitian A , A_+ is its positive part.

A linear map $\Phi : L(\mathcal{X}) \rightarrow L(\mathcal{Y})$ is completely positive if $\Phi \otimes \text{id}_{L(\mathcal{Z})}$ is positive for every $\mathcal{Z}$. We write $\text{CP}(\mathcal{X}, \mathcal{Y})$ for the completely positive maps and $\text{C}(\mathcal{X}, \mathcal{Y})$ for the channels, the completely positive trace-preserving maps; a map is trace preserving exactly when $\Phi^*(\mathbf{1}_{\mathcal{Y}}) = \mathbf{1}_{\mathcal{X}}$. For maps, $\Phi \leq \Psi$ (also written $\Phi \leq_{\text{cp}} \Psi$) means that $\Psi - \Phi$ is completely positive. Every $\Phi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ has a Kraus decomposition $\Phi(X) = \sum_{k=1}^r V_k X V_k^*$ with $V_k \in L(\mathcal{X}, \mathcal{Y})$, and the smallest possible r , attained exactly when the V_k are linearly independent, is the Kraus rank. For a channel Φ and a state ρ on $\mathcal{R} \otimes \mathcal{X}$ we write $\Phi_\rho = (\text{id}_{L(\mathcal{R})} \otimes \Phi)(\rho)$, or $(\Phi \otimes \text{id})(\rho)$ when the reference system is written second.

3.2 Vectorisation and the Choi matrix

The vectorisation $\text{vec} : L(\mathcal{X}, \mathcal{Y}) \rightarrow \mathcal{Y} \otimes \mathcal{X}$ is the linear map with $\text{vec}(E_{ab}) = e_a \otimes e_b$; it stacks the rows of a matrix. Hence $\text{vec}(uv^*) = u \otimes \bar{v}$ and

$$(A \otimes B) \text{vec}(X) = \text{vec}(AXB^\top).$$

The *Choi matrix* of a linear map $\Phi : L(\mathcal{X}) \rightarrow L(\mathcal{Y})$ is

$$J(\Phi) = \sum_{a,b=1}^{d_{\mathcal{X}}} \Phi(E_{ab}) \otimes E_{ab} \in L(\mathcal{Y} \otimes \mathcal{X}), \quad (3.2.1)$$

with the output factor first. Choi's theorem [28] says that Φ is completely positive if and only if $J(\Phi) \geq 0$, and J is a linear bijection. The map $X \mapsto AXB^*$ has Choi matrix $\text{vec}(A) \text{vec}(B)^*$, so a Kraus decomposition by linearly independent $V_1, \dots, V_r$ gives $J(\Phi) = WW^*$ with the injective matrix $W = [\text{vec}(V_1) \cdots \text{vec}(V_r)] \in L(\mathbb{C}^r, \mathcal{Y} \otimes \mathcal{X})$, and the Kraus rank is $\text{rank } J(\Phi)$. The partial trace over the output is $\text{Tr}_{\mathcal{Y}} J(\Phi) = \Phi^*(\mathbf{1}_{\mathcal{Y}})^{\top}$, so $\text{Tr}_{\mathcal{Y}} J(\Phi) = \mathbf{1}_{\mathcal{X}}$ and $\text{Tr } J(\Phi) = d_{\mathcal{X}}$ for a channel. The Choi matrix in (3.2.1) is unnormalized; the normalized *Choi state* is $J(\Phi)/d_{\mathcal{X}}$, the output of $\Phi \otimes \text{id}$ on the maximally entangled state $\text{vec}(\mathbf{1}_{\mathcal{X}}) \text{vec}(\mathbf{1}_{\mathcal{X}})^*/d_{\mathcal{X}}$.

The *Choi floor* of a channel is $\lambda(\Phi) = \lambda_{\min}(J(\Phi))$, and $\tilde{\lambda}(\Phi) = \lambda(\Phi)/d_{\mathcal{X}}$ is the floor of the Choi state. A channel has an invertible Choi matrix exactly when its Kraus rank is $d_{\mathcal{X}}d_{\mathcal{Y}}$, the largest possible. A unitary channel has a Choi matrix of rank one, and a qubit depolarizing channel $\rho \mapsto (1-p)\rho + p \text{Tr}(\rho)\mathbf{1}/2$ has Choi floor $p/2$.

Swapping the two tensor factors in (3.2.1) gives the input-first Choi matrix $\sum_{a,b} E_{ab} \otimes \Phi(E_{ab})$, which is unitarily equivalent to $J(\Phi)$ through the swap of $\mathcal{X}$ and $\mathcal{Y}$. All spectral statements (positivity, ranks, extreme eigenvalues, the Thompson metric, invertibility, partial traces up to transposition) are therefore the same in the two conventions. Chapter 5 uses the input-first convention (Table 3.1).

3.3 Norms on maps

The diamond norm, or completely bounded trace norm, of a linear map $\Delta : L(\mathcal{X}) \rightarrow L(\mathcal{Y})$ is

$$\|\Delta\|_{\diamond} = \max \{ \|(\Delta \otimes \text{id}_{L(\mathcal{X})})(X)\|_1 : X \in L(\mathcal{X} \otimes \mathcal{X}), \|X\|_1 \leq 1 \}$$

[77, 127]. It is unnormalized in every chapter, so $\|\Phi - \Psi\|_{\diamond} \leq 2$ for channels, and for completely positive Φ it equals $\|\Phi^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty}$ [103]. Chapter 4 writes $\|\cdot\|_1$ for this norm. The Choi matrix and the diamond norm are related by

$$\|J(\Phi) - J(\Phi')\|_{\infty} \leq \frac{d_{\mathcal{X}}}{2} \|\Phi - \Phi'\|_{\diamond}$$

for channels, with equality attained in every dimension (Lemma 4.8.7). If $\lambda(\Psi) > 0$ and $\|\Psi' - \Psi\|_{\diamond} \leq \epsilon$, this gives $(1-r)J(\Psi) \leq J(\Psi') \leq (1+r)J(\Psi)$ with $r = d_{\mathcal{X}}\epsilon/(2\lambda(\Psi)) = \epsilon/(2\tilde{\lambda}(\Psi))$, which is the starting point of the Choi-floor estimates in Chapters 1, 4 and 8.

3.4 Kubo–Ando connections and means

A binary operation $(A, B) \mapsto A \sigma B$ on positive semidefinite operators is a (Kubo–Ando) *connection* if it is monotone in both arguments, satisfies the transformer inequality $C^*(A \sigma B)C \leq (C^*AC) \sigma (C^*BC)$, and is continuous from above [79]. The Kubo–Ando theorem identifies the connections with the operator monotone functions $f : (0, \infty) \rightarrow [0, \infty)$ through

$$A \sigma B = A^{1/2} f(A^{-1/2} B A^{-1/2}) A^{1/2} \quad (A > 0),$$

extended to singular arguments by $A \sigma B = \lim_{\eta \downarrow 0} (A + \eta \mathbf{1}) \sigma (B + \eta \mathbf{1})$. A connection is a *mean* if $f(1) = 1$, that is, $\mathbf{1} \sigma \mathbf{1} = \mathbf{1}$. Examples are the weighted arithmetic means ($f(t) = 1 - s + st$), the harmonic mean, and the weighted geometric means

$$A \#_s B = A^{1/2} (A^{-1/2} B A^{-1/2})^s A^{1/2}, \quad s \in [0, 1], \quad A \# B = A \#_{1/2} B.$$

We use $A \#_s B = B \#_{1-s} A$, $(cA) \#_s (c'B) = c^{1-s} c'^s (A \#_s B)$ for $c, c' > 0$, and $\text{range}(A \#_s B) \subseteq \text{range } A \cap \text{range } B$ for $0 < s < 1$. In particular $A \#_s B = 0$ whenever $\text{range } A \cap \text{range } B = \{0\}$. In Chapter 4 the letter σ denotes a connection; in all other chapters it denotes a state.

For completely positive maps, a connection of the Choi matrices defines a completely positive map $\Phi \sigma \Psi$ by $J(\Phi \sigma \Psi) = J(\Phi) \sigma J(\Psi)$ [76]; weighted geometric means of completely positive maps were first introduced in [47]. If T is the derivative of Section 3.5, then

$$J(\Phi) \sigma J(\Psi) = W g_\sigma(T) W^*, \quad g_\sigma(t) = t \sigma(1-t),$$

a scalar function of T (Theorem 4.2.15). For the weighted geometric means this reads $J(\Psi) \#_s J(\Phi) = W g_s(T) W^*$ with $g_s(t) = t^s(1-t)^{1-s}$, the function that Chapter 4 calls h_s (Eq. (4.5.1)), Chapter 7 calls g_s and Chapter 8 calls b_s . It is $g_{\#_{1-s}}$, not the function $g_{\#_s}$ attached to the mean $\#_s$.

3.5 The Arveson–Radon–Nikodym derivative and the shadows

Let $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, let $\Omega = \Phi + \Psi$ have a Kraus decomposition by linearly independent operators $V_1, \dots, V_r$, and let W be as in Section 3.2, so that $J(\Omega) = WW^*$. For $K \in \text{L}(\mathbb{C}^r)$ put

$$\Psi_K(X) = \sum_{k,l=1}^r K_{kl} V_k X V_l^*, \quad J(\Psi_K) = W K W^*.$$

By Arveson’s Radon–Nikodym theorem in its Kraus form [10, 111] (Theorem 4.2.2) there is a unique T with $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$ and $\Phi = \Psi_T$; then $\Psi = \Psi_{\mathbf{1}-T}$. We call T the *ARN derivative* of Φ with respect to Ω , and Ψ_K is K read through the Kraus operators of Ω . A different choice of Kraus operators replaces T by $\bar{U} T U^\top$ for a unitary U , so the spectrum of T and all unitarily invariant norms of differences $T_1 - T_2$ do not depend on the choice. The spectral decomposition $T = \sum_\ell \tau_\ell P_\ell$ gives the *level decomposition* $\Phi = \sum_\ell \tau_\ell \Omega_\ell$, $\Psi = \sum_\ell (1 - \tau_\ell) \Omega_\ell$ with $\Omega_\ell = \Psi_{P_\ell}$ (Lemma 7.2.1); a level is ambiguous if $0 < \tau_\ell < 1$.

Chapter 4 also uses the quantity

$$\nu(\Omega) = \min \left\{ \left\| \sum_{k=1}^r c_k V_k \right\|_\infty : c \in \mathbb{C}^r, \|c\|_2 = 1 \right\} > 0,$$

which converts derivative distances into diamond distances (Theorem 4.8.4).

For states ρ, σ with $\omega = \rho + \sigma$, the same construction gives $T = \omega^{-1/2} \rho \omega^{-1/2}$ on the support of ω ; its spectral measure in the state ω is the *reverse test* (shadow (I) of Section 2.2). The *testing-region shadow* (shadow (II)) is the measure $\mu = e''$ of the profile $e(\gamma) = \text{Tr}(\rho - \gamma\sigma)_+$, completed by the boundary masses $m_0 = \text{Tr} \Pi_\rho^\perp \sigma$ at 0 and $m_\infty = \text{Tr} \Pi_\sigma^\perp \rho$ at ∞ (Definition 5.2.4). The *modular shadow* (shadow (III)) is the Nussbaum–Szkola pair of the relative modular operator Δ , and $\tilde{T} = \Delta(\mathbf{1} + \Delta)^{-1}$ is the modular derivative (Section 5.5).

3.6 Fidelities

Several fidelities appear, and two of them come in a root and a squared form.

- The *Uhlmann fidelity* of states [122]. In its root form, $F(\rho, \sigma) = \|\sqrt{\rho} \sqrt{\sigma}\|_1 = \text{tr} \sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}}$; this is the form used in Chapters 1, 4 and 8. Chapter 7 uses the

squared form $F(\rho, \sigma) = \|\rho^{1/2}\sigma^{1/2}\|_1^2$. The *Uhlmann channel root fidelity* $F_U(\Phi, \Psi)$ of Chapters 1 and 8 is the minimum over inputs (with a reference system) of the root fidelity of the two outputs.

- The *Matsumoto or geometric fidelity* of states, $\hat{F}(\rho, \sigma) = F_M(\rho, \sigma) = \text{tr}(\rho \# \sigma)$ [29, 90]. It satisfies $F_M \leq F$, with F the root Uhlmann fidelity.
- The *geometric channel fidelity*

$$\hat{F}(\Phi, \Psi) = \inf_{\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})} \text{tr}(\Phi_\rho \# \Psi_\rho) = \lambda_{\min}(\text{Tr}_Y(J(\Phi) \# J(\Psi)))$$

(Proposition 4.3.1). It is used in root form in every chapter, so $\hat{F}(\Phi, \Phi) = 1$ for a channel Φ . Katariya and Wilde [72] call its square the geometric fidelity. In Chapter 7 it appears as $\hat{Q}_{1/2}(\Phi \|\Psi)$, and the *fidelity converse* for the Chernoff exponent is $-2 \ln \hat{F}(\Phi, \Psi)$.

- The *Bhattacharyya coefficient* of probability vectors p, q is $\sum_i \sqrt{p_i q_i}$; for amplitude damping, Chapters 1, 7 and 8 write $f(\gamma, \gamma') = \sqrt{\gamma \gamma'} + \sqrt{(1-\gamma)(1-\gamma')}$.
- The *testing-region fidelity* $Q_{1/2}^{\text{HT}}$ of Chapter 5 is the testing-region Rényi quasi-entropy of order $\frac{1}{2}$ (Section 3.7).

3.7 Divergences

Divergences of states

For $\rho, \sigma \in \mathcal{D}(\mathcal{X})$ (or positive semidefinite operators) we use the following quantities; each is $+\infty$ when the support condition that makes it finite fails.

- The *Umegaki relative entropy* $D(\rho \|\sigma) = \text{Tr} \rho (\log \rho - \log \sigma)$, finite if $\Pi_\rho \leq \Pi_\sigma$.
- The *max-relative entropy* $D_{\max}(\rho \|\sigma) = \log \inf\{c > 0 : \rho \leq c\sigma\}$; for completely positive maps, $D_{\max}(\Phi \|\Psi) = \log \inf\{c > 0 : \Phi \leq_{\text{cp}} c\Psi\}$.
- The *maximal f -divergence* of a convex function f [61, 91], which for operator convex f and $\sigma > 0$ is $\hat{D}_f(\rho \|\sigma) = \text{Tr} \sigma^{1/2} f(\sigma^{-1/2} \rho \sigma^{-1/2}) \sigma^{1/2}$. Its boundary values $f(0^+)$ and $f'(\infty) = \lim_{t \rightarrow \infty} f(t)/t$ decide finiteness.
- The *Belavkin–Staszewski relative entropy* $D_{\text{BS}}(\rho \|\sigma) = \text{Tr} \rho \log(\rho^{1/2} \sigma^{-1} \rho^{1/2})$ [14], the maximal f -divergence of $f(t) = t \log t$.
- The *geometric Rényi quasi-entropy and divergence* $\hat{Q}_\alpha(\rho \|\sigma) = \text{Tr}(\sigma \#_\alpha \rho)$ and $\hat{D}_\alpha = \frac{1}{\alpha-1} \log \hat{Q}_\alpha$, for $\alpha \in (0, 1) \cup (1, 2]$ (for $\alpha > 1$ through the formula $\sigma^{1/2}(\sigma^{-1/2} \rho \sigma^{-1/2})^\alpha \sigma^{1/2}$ when $\Pi_\rho \leq \Pi_\sigma$). The geometric fidelity is $\hat{Q}_{1/2}$.
- The *Petz quasi-entropy* $Q_\alpha^{\text{P}}(\rho \|\sigma) = \text{Tr} \rho^\alpha \sigma^{1-\alpha}$ and the *sandwiched quasi-entropy* $\tilde{Q}_\alpha(\rho \|\sigma) = \text{Tr}(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}})^\alpha$, with divergences D_α^{P} and $\tilde{D}_\alpha$ defined by $\frac{1}{\alpha-1} \log$.
- The *hockey-stick divergence* $E_\gamma(A \|\sigma) = \text{Tr}(A - \gamma \sigma)_+ = \max_{0 \leq M \leq 1} \text{Tr} M(A - \gamma \sigma)$, $\gamma \geq 0$, and the *testing-region* (Hirche–Tomamichel) f -divergence [63]

$$D_f^{\text{HT}}(\rho \|\sigma) = \int_1^\infty (E_\gamma(\rho \|\sigma) f''(d\gamma) + E_\gamma(\sigma \|\rho) \tilde{f}''(d\gamma)), \quad \tilde{f}(x) = x f(1/x),$$

for convex f with $f(1) = 0$, with Rényi members $Q_\alpha^{\text{HT}} = 1 + (\alpha-1) D_{f_\alpha}^{\text{HT}}$, $f_\alpha(x) = \frac{x^\alpha - 1}{\alpha - 1}$, and $D_\alpha^{\text{HT}} = \frac{1}{\alpha-1} \log Q_\alpha^{\text{HT}}$ (Definition 5.2.3). By Frenkel's formula [46], $D_{x \log x}^{\text{HT}} = D$.

- The *measured* divergences D_f^M , the supremum of the classical f -divergences of the outcome distributions of all measurements.

The maximal divergences dominate the others: for operator convex f the measured, testing-region and maximal f -divergences satisfy $D_f^M \leq D_f^{\text{HT}} \leq \widehat{D}_f$, and the three coincide for a strictly convex f only on commuting pairs (Corollary 5.3.3). In particular $D \leq D_{\text{BS}}$.

Divergences of channels

For a divergence $\mathbb{D}$ of states, the channel divergence is

$$\mathbb{D}(\Phi \parallel \Psi) = \sup_{\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})} \mathbb{D}(\Phi_\rho \parallel \Psi_\rho),$$

and for the quasi-entropies with $\alpha < 1$ and the fidelities the supremum becomes an infimum. The geometric quantities have closed forms in the Choi matrices [42, 72]:

$$\begin{aligned} \widehat{Q}_\alpha(\Phi \parallel \Psi) &= \lambda_{\min}(\text{Tr}_{\mathcal{Y}}(J(\Psi) \#_\alpha J(\Phi))) \quad (0 < \alpha < 1), \\ D_{\text{BS}}(\Phi \parallel \Psi) &= \lambda_{\max}\left(\text{Tr}_{\mathcal{Y}}\left[J(\Phi)^{1/2} \log(J(\Phi)^{1/2} J(\Psi)^{-1} J(\Phi)^{1/2}) J(\Phi)^{1/2}\right]\right), \end{aligned}$$

the latter when $J(\Psi)$ is invertible, and more generally $\widehat{D}_f(\Phi \parallel \Psi) = \lambda_{\max}(\text{Tr}_{\mathcal{Y}}(W h_f(T) W^*))$ for operator convex f (Theorem 4.7.5). For a channel divergence we also use its n -use adaptive version $\mathbb{D}^{(n)}$, the supremum over adaptive strategies of the divergence of the two final states, its regularisation $\mathbb{D}^{\text{reg}}(\Phi \parallel \Psi) = \lim_n \frac{1}{n} \mathbb{D}(\Phi^{\otimes n} \parallel \Psi^{\otimes n})$ when the limit exists, and its amortised version $\mathbb{D}^A$ [128] (Section 5.7).

Three converses for the Chernoff exponent appear in Chapters 1, 7 and 8: the fidelity converse $-2 \ln \widehat{F}(\Phi, \Psi)$ [72], the barycentric converse C_{bar} of Ji, Mishra, Mosonyi and Wilde [66], and the *geometric Chernoff divergence*

$$\widehat{C}(\Phi, \Psi) = -\ln \min_{0 \leq s \leq 1} \lambda_{\min}(B_s) = -\ln \inf_{0 < s < 1} \widehat{Q}_s(\Phi \parallel \Psi), \quad B_s = \Psi_{g_s(T)}^*(\mathbf{1}_{\mathcal{Y}}),$$

where $g_s(t) = t^s(1-t)^{1-s}$ on $(0, 1)$ and $g_s(0) = g_s(1) = 0$. Chapter 7 uses the first expression and Chapter 8 the second; they agree because $s \mapsto \lambda_{\min}(B_s)$ is continuous on $[0, 1]$ and equals $\widehat{Q}_s(\Phi \parallel \Psi)$ for $0 < s < 1$. The *program overlap* $\mu_*(\Phi, \Psi)$ of Theorem 7.6.9 gives a fourth converse, $-2 \ln \mu_*$.

3.8 The Thompson metric

For $A, B \geq 0$ the Thompson metric is

$$d_T(A, B) = \inf\{t \geq 0 : e^{-t}A \leq B \leq e^t A\} = \max\{D_{\max}(A \parallel B), D_{\max}(B \parallel A)\},$$

which is finite exactly when A and B have the same support. For completely positive maps, $d_T(\Phi, \Psi) = d_T(J(\Phi), J(\Psi))$, equivalently the same expression with the completely positive order. All chapters use this metric with the same normalisation; Chapter 4 writes $d_T(\Omega_1, \Omega_2)$, Chapter 5 writes d_T , and Chapters 7 and 8 write d_T . The metric enters in two ways: through $\tanh(d_T/2)$, which measures the change of a sum in Chapters 4 and 7 (Lemma 4.10.1), and as a Lipschitz constant per channel use for adaptive quantities in Chapters 1, 5 and 8 (Theorem 5.7.2).

3.9 Operational exponents

An adaptive strategy with n uses of an unknown channel prepares a state on the input and a memory, interleaves the n uses with channels acting on the output and the memory, and ends with a measurement; parallel and product strategies are special cases. Let $\rho_\Phi^{(n)}$ and $\rho_\Psi^{(n)}$ be the final states under the two hypotheses.

- The *adaptive Stein exponent* ζ is the best exponential decay rate of the type-II error when the type-I error is required to vanish asymptotically. It equals the regularised and the amortised Umegaki channel divergence [41, 125]. Chapter 5 writes $\zeta(\Phi, \Psi)$ and Chapters 1 and 8 write $\zeta(\Phi \| \Psi)$.
- The *adaptive Chernoff exponent* is $\xi = \limsup_n -\frac{1}{n} \ln p(n)$, where $p(n)$ is the smallest equal-prior error $\frac{1}{2} - \frac{1}{4} \|\rho_\Phi^{(n)} - \rho_\Psi^{(n)}\|_1$ over adaptive strategies. Chapter 7 writes ξ_{adapt} , and ξ_{prod} for the exponent of product strategies. The quantum Chernoff divergence of two operators is $\xi_{\text{QC}}(A, B) = -\ln \min_{0 \leq s \leq 1} \text{tr}(A^s B^{1-s})$ [11, 99].
- The *strong-converse exponent* is the rate at which the success probability must vanish when the type-II error decays faster than $e^{-n\zeta}$; it is governed by the regularised sandwiched divergences (Corollary 5.7.9).
- The Fisher informations of a one-parameter family of states or channels are written F_{SLD} , F_{BKM} and F_{RLD} (symmetric, Kubo–Mori and right logarithmic derivative), and F_∞ is the asymptotic adaptive Fisher information per channel use [131]. They appear in Chapters 1, 7 and 8, and must not be confused with the fidelities, which carry no such subscript.

3.10 Logarithms

All logarithms in this book are natural, and entropies are measured in nats. Chapters 4, 5 and 6 write $\log$, and Chapters 1, 7 and 8 write $\ln$. Chapter 4 does not state the base, but its estimates use the natural logarithm, for instance in the factor $e^{\pm d}$ attached to the Thompson metric. The binary entropy $h(x) = -x \log x - (1-x) \log(1-x)$ in the continuity bounds for the Stein exponent is also in nats.

3.11 Differences between the chapters

The chapters can be read independently, and their conventions differ in the places listed in Table 3.1. None of these differences affects a statement that is invariant under the relevant change; for example, every result about the spectrum of a Choi matrix holds in both tensor orderings. Table 3.2 lists the symbols that have different meanings in different chapters. Within a chapter, every symbol has one meaning unless the chapter says otherwise.

Table 3.1: Conventions of Chapters 1 and 4–8. Chapters 1 and 8, on noisy gates, share the last column. Chapter 6 works with states only and follows Chapter 5 where the two overlap.

	Ch. 4 (continuity)	Ch. 5–6 (shadows)	Ch. 7 (Chernoff)	Ch. 1 and 8 (noisy gates)
Choi ordering	output first: $J(\Phi) = \sum_{ab} \Phi(E_{ab}) \otimes E_{ab}$ on $\mathcal{Y} \otimes \mathcal{X}$ (as in [127])	<i>input first</i> : $J(\Phi) = \sum_{ij} i\rangle\langle j \otimes \Phi(i\rangle\langle j)$ on $\mathcal{X} \otimes \mathcal{Y}$ (Ch. 5); no Choi matrices in Ch. 6	output first: $J(\Phi) = \sum_{ij} \Phi(e_i e_j^*) \otimes e_i e_j^*$	output first: $J(\Phi) = \sum_{ij} \Phi(i\rangle\langle j) \otimes i\rangle\langle j $
Choi normalisation	unnormalized; $\text{Tr}_{\mathcal{Y}} J(\Phi) = \Phi^*(\mathbf{1})^\top$	unnormalized; $J(\Phi) \leq d_{\mathcal{X}} \mathbf{1}$ for channels	unnormalized	unnormalized, $\text{Tr } J = d_X$; also the Choi state J/d_X with floor $\bar{\lambda} = \lambda/d_X$
Vectorisation	$\text{vec}(E_{ab}) = e_a \otimes e_b$ (rows stacked), $\text{vec} : \mathcal{L}(\mathcal{X}, \mathcal{Y}) \rightarrow \mathcal{Y} \otimes \mathcal{X}$; $J(A \cdot B^*) = \text{vec}(A) \text{vec}(B)^*$	not fixed; the frame W is taken from Ch. 4 (Proposition 5.7.7); with input-first Choi matrices the matching vectorisation stacks columns	$\text{vec}(yx^*) = y \otimes \bar{x}$, i.e. $\text{vec}(e_i e_j^*) = e_i \otimes e_j$, $\text{vec} : \mathcal{L}(\mathcal{X}, \mathcal{Y}) \rightarrow \mathcal{Y} \otimes \mathcal{X}$ (rows stacked, as in Ch. 4)	$\text{vec}(A) = \sum_{ij} A_{ij} i\rangle\langle j $ (rows stacked, as in Ch. 4); used only in Ch. 8
Uhlmann fidelity of states	root: $F(\rho, \xi) = \ \sqrt{\rho}\sqrt{\xi}\ _1$	not used (the testing-region fidelity is $Q_{1/2}^{\text{HT}}$)	<i>squared</i> : $F(\rho, \sigma) = \ \rho^{1/2} \sigma^{1/2}\ _1^2$	root: F_U (channel version, minimised over inputs)
Geometric fidelity	$\hat{F}$, root form (Katariya–Wilde use its square); also $F_M = \text{tr}(\rho \# \xi)$ for states	$\hat{F}$ as in Ch. 4	$\hat{F} = \hat{Q}_{1/2}$, root form; converse $-2 \ln \hat{Q}_{1/2}$	$\hat{F}$, root form; converse $-2 \ln \hat{F}$
Logarithm	log, base not stated; natural	log, natural (stated)	ln	ln
Diamond norm	$\ \cdot\ _1$ (cb trace norm), d_{cb} for distances; unnormalized	$\ \cdot\ _\diamond$, unnormalized; moduli in $\varepsilon = \frac{1}{2} \ \Phi - \Phi'\ _\diamond$	$\ \cdot\ _\diamond$, unnormalized	$\ \cdot\ _\diamond$, unnormalized; Choi-floor law in $\epsilon = \ \Psi - \Psi'\ _\diamond$, Stein modulus in $\epsilon' = \frac{1}{2} \ \Phi - \Phi'\ _\diamond$
Derivative	T : ARN derivative of Φ w.r.t. $\Omega = \Phi + \Psi$; Ψ_K ; frame W with r columns; g_σ , h_f ; relative derivatives T_i for different sums	states: $T = \omega^{-1/2} \rho \omega^{-1/2}$ (reverse test), modular derivative $\tilde{T} = \Delta(\mathbf{1} + \Delta)^{-1}$; channels: T , W as in Ch. 4	$T = \sum_\ell \tau_\ell P_\ell$; Ψ_K ; levels $\Omega_\ell = \Psi_{P_\ell}$, $S_\ell = \Omega_\ell^*(\mathbf{1})$; g_s ; steps $\lambda_\ell = \ln \frac{\tau_\ell}{1-\tau_\ell}$	in Ch. 8 only: T with frame W , $WW^\dagger = J(\Phi + \Psi)$; b_s for g_s ; levels τ_ℓ , S_ℓ , B_s and Λ as in Ch. 7
$\hat{C}$	—	—	$-\ln \min_s \lambda_{\min}(B_s)$, minimum over $s \in [0, 1]$	$-\ln \inf_s \hat{Q}_s$, infimum over $s \in (0, 1)$, in Ch. 8; minimum over $s \in [0, 1]$ in Ch. 1 (same value)
Exponents	—	$\zeta(\Phi, \Psi)$	ξ_{adapt} , ξ_{prod} , ξ_{QC}	$\zeta(\Phi \ \Psi)$, $\xi(\Phi, \Psi)$, ξ_{prod}

continued on the next page

	Ch. 4	Ch. 5–6	Ch. 7	Ch. 1 and 8
Identity, trace, adjoint	$\mathbf{1}_{\mathcal{V}}$ (space always indicated), $\mathbf{1}_{L(\mathcal{X})}$ for the identity map; tr , $\text{Tr}_{\mathcal{Y}}$; A^*	$\mathbf{1}$, id ; Tr ; A^*	$\mathbf{1}$, id ; tr , $\text{Tr}_{\mathcal{Y}}$; A^*	$\mathbf{1}$, id ; Tr , $\text{Tr}_{\mathcal{Y}}$; $A^\dagger$
Spaces and dimensions	$L(\mathcal{X})$, $D(\mathcal{X})$, $\text{CP}(\mathcal{X}, \mathcal{Y})$, $C(\mathcal{X}, \mathcal{Y})$; $\dim \mathcal{X}$, $n_{\mathcal{Y}} = \dim \mathcal{Y}$	$\mathcal{B}(\mathcal{X})$ for the operators; $d_{\mathcal{X}}$, $d_{\mathcal{Y}}$	L , D , $C(\mathcal{X}, \mathcal{Y})$; $d = \dim \mathcal{X}$	d_X , d_Y ; $d = 2^n$ for n qubits
Spelling	British	British	American	American

Table 3.2: Symbols with different meanings in different chapters.

Symbol	Meanings
σ	a Kubo–Ando connection or mean in Ch. 4 (states there are ρ, ξ); a state in Ch. 1 and 5–8
ξ	a state in Ch. 4 (and occasionally in proofs of Ch. 5); the adaptive Chernoff exponent in Ch. 1, 7 and 8
λ	the Choi floor in Section 4.8 and Ch. 1 and 8; eigenvalues in the structure theorem of Ch. 5 (Theorem 5.3.1); the log-likelihood steps λ_ℓ in Ch. 7
ν	$\nu(\Omega)$, the least norm of a unit combination of Kraus operators, in Ch. 4 and 7; the Frenkel measure in Ch. 6; the laws ν_Φ, ν_Ψ in Theorem 7.5.3
τ	the in-pair derivative distance in Ch. 4 (pairs with different sums) and Theorem 7.6.2; the eigenvalues (levels) τ_ℓ of T in Ch. 7 and 8
δ	the relative distance of two pairs in Ch. 4 and 7; the distance between nearby channels in the local comparisons of Ch. 7; a coherent miscalibration angle in Ch. 1 and 8
η	the size of a Choi perturbation in Theorem 4.8.8 and a regularisation parameter in Ch. 4; the error profile $\eta(\gamma) = 1 - e(\gamma)$ in Ch. 5–6; a target accuracy in Ch. 1 and 8
κ	the condition number of $J(\Psi)$ in Ch. 1, 5 and 8; a derivative radius in Corollary 7.6.3; a parameter of the operator layer-cake inequality in Ch. 6
μ	the testing-region shadow $\mu = e''$ in Ch. 5–6; the program overlap μ_* in Ch. 1, 7 and 8
r	the number of Kraus operators of the sum in Ch. 4 and 7; the relative radius $r = d_X \epsilon / (2\lambda)$ in Ch. 1 and 8
d	a diamond distance in Ch. 4 (e.g. Theorem 4.12.11); $\dim \mathcal{X}$ in Ch. 7; the dimension 2^n of n qubits in Ch. 1 and 8
h	the functions h_f , h_α of the derivative in Ch. 4 (h_α is the g_α of Ch. 7 and the b_α of Ch. 8); the binary entropy in Ch. 1, 5 and 8
F	a fidelity (root or squared, see Table 3.1); with a subscript RLD, SLD, BKM or ∞ , a Fisher information (Ch. 1, 7 and 8)
T	the ARN derivative; T_1 also denotes the energy-relaxation time in the phrase “pure T_1 decay” in Ch. 1 and 8
Λ	the largest absolute log-likelihood step over ambiguous levels in Ch. 7 and 8; a channel (a simulator or a post-processing) elsewhere in the same chapters

Chapter 4

Continuity of certain quantum channel divergences

Summary. The geometric channel fidelity, the geometric Rényi divergences and the Belavkin–Staszewski relative entropy give converse bounds for channel capacities, discrimination and estimation. Channels are known only approximately, yet two pairs of channels with equal sums can be arbitrarily close in the diamond norm and have fidelities 0 and 1. We show that continuity is recovered when pairs are compared through the Arveson–Radon–Nikodym derivative T of one map with respect to the sum. Its transpose is the environment effect that selects this map in a Stinespring dilation of the sum. In T , the fidelity, the geometric Rényi quasi-entropies of order in $(0, 1)$ and the bounded maximal f -divergences are continuous with explicit dimension-free moduli. So are the Belavkin–Staszewski relative entropy and the other maximal f -divergences of operator convex f while the spectrum of T stays away from the points 0 or 1 at which they become infinite, and every maximal f -divergence of operator convex f is continuous in T with values in $(-\infty, +\infty]$. For two pairs of channels with the same sum, the fidelities differ by at most $2\|T_1 - T_2\|_\infty^{1/2}$, which gives Hölder continuity in the diamond norm on such splittings. The fidelity is continuous at a pair of completely positive maps exactly when the Choi matrix of the sum is invertible or the fidelity vanishes. Thus pairs containing a noisy channel of full Kraus rank are continuity points, whereas the pair formed by an ideal unitary gate and itself is not. Within channels, the fidelity is continuous at a diagonal pair (Φ, Φ) exactly when the Choi matrix of Φ is invertible; if it is singular, the discontinuity already occurs along the unitary orbit of Φ . If the Choi matrices are bounded below by $\lambda > 0$, a perturbation of the Choi matrices of size η changes the fidelity by a relative amount of at most η/λ , and the order λ^{-1} is sharp.

4.1 Introduction

The geometric channel fidelity of $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ is $\widehat{F}(\Phi, \Psi) = \inf_\rho \text{tr}(\Phi_\rho \# \Psi_\rho)$, where $\Phi_\rho = (\Phi \otimes \mathbf{1}_{L(\mathcal{X})})(\rho)$ for states ρ on $\mathcal{X} \otimes \mathcal{X}$ and $\#$ is the matrix geometric mean (Katariya and Wilde [72] use its square); the geometric Rényi divergences and the Belavkin–Staszewski relative entropy arise in the same way from maximal f -divergences [91]. These channel divergences give converse bounds for channel capacities [42], discrimination and estimation [72]. In practice a channel is known only approximately, and the relevant worst-case error is the diamond distance $\|\cdot\|_1$ [80], so such bounds need continuity in this norm, and this continuity fails. If $\dim \mathcal{Y} \geq 2$, then for every $\eta > 0$ there are replacement channels

This chapter is a revised and extended version of [69].

with $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$ and $|||\Phi_1 - \Phi_2|||_1 = |||\Psi_1 - \Psi_2|||_1 < \eta$, but $\widehat{F}(\Phi_1, \Psi_1) = 0$ and $\widehat{F}(\Phi_2, \Psi_2) = 1$ (Corollary 4.3.3); on the same pairs every bounded maximal f -divergence with $f(1) = 0$ jumps by $f(0^+) + f'(\infty)$, which is positive unless f is affine (Corollary 4.5.14).

Continuity is recovered by recording how a pair splits its sum. Let $\Omega = \Phi + \Psi$ have linearly independent Kraus operators $V_1, \dots, V_r$, and let W have columns $\text{vec}(V_k)$. By Arveson's Radon–Nikodym theorem (Theorem 4.2.2), $\Phi(X) = \sum_{k,l} T_{kl} V_k X V_l^*$ for a unique $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$ on $\mathbb{C}^r$, and $\mathbf{1}_{\mathbb{C}^r} - T$ gives Ψ . Up to normalisation, (Φ, Ψ) is a two-outcome instrument implementing Ω , with which-branch effect $T^\top$ on the environment of its Stinespring dilation (Theorem 4.2.5). By an identity essentially due to Kian [76], $J(\Phi)\sigma J(\Psi) = W g_\sigma(T) W^*$ with $g_\sigma(t) = t\sigma(1-t)$ for every Kubo–Ando mean σ (Theorem 4.2.15), so all quantities above are functions of T . Hence, by Theorem 4.4.6,

$$|\widehat{F}(\Phi_1, \Psi_1) - \widehat{F}(\Phi_2, \Psi_2)| \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \|T_1 - T_2\|_\infty^{1/2}, \quad (4.1.1)$$

with no dimensional factor, where T_1 and T_2 are computed in the same Kraus decomposition $V_1, \dots, V_r$ of Ω ; the constant and the exponent are sharp (see the remark following Theorem 4.4.6). Here $\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty = 2$ if Φ_i and Ψ_i are channels, and it equals 1 if Ω itself is a channel, as for an instrument. In the counterexample T_1 is a projection, on which $g_\#$ vanishes, while $T_2 = \frac{1}{2}\mathbf{1}_{\mathbb{C}^r}$. Analogous bounds hold for all means, the geometric Rényi quasi-entropies $\widehat{Q}_\alpha$, $\alpha \in (0, 1)$, and bounded maximal f -divergences (Theorems 4.4.5, 4.5.5, 4.5.13). As $\nu(\Omega)^2 \|T_1 - T_2\|_\infty \leq |||\Phi_1 - \Phi_2|||_1$, with $\nu(\Omega)$ the least operator norm of a unit combination of the V_k (Theorem 4.8.4), $\widehat{F}$ is Hölder continuous in the diamond norm on the splittings of Ω , with constant $\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty / \nu(\Omega)$ (Corollary 4.8.6); in the counterexample $\nu(\Omega) \rightarrow 0$.

On pairs of completely positive maps, Theorem 4.12.4 shows exactly where continuity survives:

$$\widehat{F} \text{ is continuous at } (\Phi, \Psi) \iff J(\Phi + \Psi) \text{ is invertible or } \widehat{F}(\Phi, \Psi) = 0. \quad (4.1.2)$$

The same holds for $\widehat{Q}_\alpha$, and there are analogues for all means other than the weighted arithmetic means and, under a rank condition, for bounded maximal f -divergences. Pairs with a channel of full Kraus rank, e.g. generalised amplitude damping with damping $p \in (0, 1]$ and thermal parameter $N \in (0, 1)$ [74], are continuity points, whereas an ideal gate $U(\cdot)U^*$ has $\widehat{F} = 1$ with itself and $\widehat{F} = 0$ with every mis-rotated $U'(\cdot)U'^*$, $U' \notin \mathbb{C}U$, however close (Corollary 4.12.6).

For unbounded divergences, if $D_{\max}(\Phi_i \| \Psi_i) \leq \log((1-\epsilon)/\epsilon)$ for both pairs, as for a detector of efficiency $1-\epsilon$, the Belavkin–Staszewski relative entropy and $\widehat{Q}_\alpha$, $\alpha \in (1, 2]$, obey continuity estimates in T whose constants blow up as $\epsilon \downarrow 0$ (Theorem 4.6.7, Corollaries 4.6.11 and 4.6.12). Every maximal f -divergence with operator convex f is continuous in T with values in $(-\infty, +\infty]$ (Theorem 4.7.6). Section 4.9 gives explicit moduli for every mean and every operator convex f , obtained from exact operator Lipschitz constants of one-pole kernels; for operator convex f on $[\epsilon, 1-\epsilon]$ the Lipschitz constants obtained are within a factor 3 of optimal, and in many cases exact.

If the two pairs have different sums, the change of the sum enters through the Thompson metric; for arbitrary pairs of channels, $|\widehat{F}(\Phi_1, \Psi_1) - \widehat{F}(\Phi_2, \Psi_2)| \leq 5.06 \delta^{1/2}$, with δ their distance relative to $\Phi_1 + \Psi_1 + \Phi_2 + \Psi_2$ (Theorem 4.10.4, Corollary 4.10.7). Support changes, as under leakage [130], enter through the Lebesgue decomposition of the sums (Theorem 4.11.3).

We also consider perturbations within the set of channels. If $J(\Phi)$ and $J(\Psi)$ are bounded below by $\lambda > 0$, then

$$|\widehat{F}(\Phi', \Psi') - \widehat{F}(\Phi, \Psi)| \leq \frac{\dim \mathcal{X}}{2\lambda} d \widehat{F}(\Phi, \Psi)$$

for all channels Φ', Ψ' at diamond distance $d < 2\lambda/\dim \mathcal{X}$, and the order λ^{-1} of this Lipschitz bound cannot be improved (Theorem 4.8.8, Example 4.8.11). At every channel Φ with singular Choi matrix, output rotations $\Phi' \rightarrow \Phi$ keep $\hat{F}(\Phi, \Phi')$ bounded away from 1, and within the zero-temperature amplitude damping family $\hat{F}$ jumps from 1 to 0 (Theorem 4.12.9, Corollary 4.12.10, Example 4.12.8). At mutually singular pairs $\hat{F} = O(d^{1/2}/\sin \theta)$, where θ is the smallest angle between the Choi ranges; the exponent and the angle dependence are sharp (Theorem 4.12.11, Example 4.12.12).

We now relate our results to the literature. Arveson [10] proved the Radon–Nikodym theorem for completely positive maps; Raginsky [111] gave its Kraus form, the upper bound of Lemma 4.4.1 and, for derivatives taken with respect to a fixed reference map, a lower bound with the dimensional factor $(\dim \mathcal{H})^{-2}$ for maps on $B(\mathcal{H})$ [111, eq. (17)]. Weighted Kubo–Ando geometric means of completely positive maps, together with a superoperator perspective function, were first introduced by Frenkel, Mosonyi, Vrana and Weiner [47] in their study of error exponents for composite quantum hypothesis testing and channel discrimination. Kian [76] later defined general Kubo–Ando means of such maps via these derivatives, and Okayasu studied their geometric mean and Lebesgue decomposition [100]. Okayasu [101] also introduced operator-valued maximal f -divergences of completely positive maps, with criteria for a dense domain and for boundedness [101, Proposition 4.11] and joint lower semicontinuity [101, Theorem 4.7]; in finite dimensions our closed form is the largest eigenvalue of his divergence of the adjoint maps (Section 4.7). Matsumoto [90, 91] introduced his fidelity and maximal f -divergences; Katariya and Wilde [72] and Fang and Fawzi [42] gave closed forms for geometric channel divergences. Maximal f -divergences and their lower semicontinuity are studied in [60, 61], continuity of Rényi divergences in [95], and dimension-dependent continuity bounds for Belavkin–Staszewski quantities of states in [21]. To our knowledge, the quantitative continuity of these divergences and means in T , the dimension-free lower bound in the two-sided estimate, the results for different sums and supports, and the dichotomy are new.

The chapter is organised as follows. Section 4.2 collects preliminaries on the Arveson–Radon–Nikodym theorem and on Kubo–Ando means of completely positive maps, and Section 4.3 presents the counterexample. Sections 4.4–4.7 prove continuity in T : first for the means and the fidelity, then for the geometric Rényi divergences and bounded maximal f -divergences, and finally for unbounded ones and all operator convex f . Section 4.8 passes to the diamond norm and treats pairs with a Choi floor, and Section 4.9 gives explicit moduli. Sections 4.10 and 4.11 treat pairs with different sums and with different supports, Section 4.12 proves the dichotomy and studies perturbations within channels, and Section 4.13 lists open questions. Section 4.14 contains norm bounds and a variational characterisation of means of completely positive maps. Remarks headed “Physical interpretation” explain the main notions operationally.

4.2 Preliminaries

Notation. We follow [127]. Complex Euclidean spaces are denoted by calligraphic letters, $\mathbf{1}_{\mathcal{V}}$ is the identity operator on $\mathcal{V}$, and $\mathbf{1}_{L(\mathcal{V})}$ is the identity map on $L(\mathcal{V})$; the space is always indicated, for instance $\mathbf{1}_{\mathbb{C}^r}$ for the identity on the space $\mathbb{C}^r$ of an ARN derivative. We write $\mathbf{j} = (1, \dots, 1)^\top$ for the all-ones vector, of the length dictated by the context, so that $\mathbf{j}\mathbf{j}^\top$ is the matrix all of whose entries equal 1; it should not be confused with an identity operator. Finally, δ_x denotes the unit point mass at a point x .

4.2.1 The Arveson–Radon–Nikodym theorem

Given two completely positive maps, [10] showed that one can construct a kind of Radon–Nikodym derivative analogue. This is used in [76] to construct the mean of completely positive maps. We will repeat the construction here in the finite-dimensional setting, adjusting the notation to be compatible with Watrous [127]. For now we recall the important facts for the Arveson–Radon–Nikodym (ARN) theorem. The following follows from [10] with some objects changed to fit the conventions of [76]. We show the proof in the finite-dimensional case. This version uses the Kraus form of the ARN theorem [111]. The following lemma is a special case of the Douglas factorisation lemma [34]; we include a proof for completeness.

Lemma 4.2.1. *Let $W \in L(\mathbb{C}^r, \mathcal{V})$ be injective and let $Y \in L(\mathcal{V})$ satisfy $0 \leq Y \leq WW^*$. Then there is a unique $T \in L(\mathbb{C}^r)$ with $Y = WTW^*$, and it satisfies $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$.*

Proof. Since W has full column rank, W^*W is invertible, and the Moore–Penrose pseudoinverse $W^+ = (W^*W)^{-1}W^*$ satisfies $W^+W = \mathbf{1}_{\mathbb{C}^r}$. The operator $P = WW^+$ is a self-adjoint orthogonal projection onto $\text{range}(W)$.

If $W^*x = 0$, then

$$0 \leq x^*Yx \leq x^*WW^*x = \|W^*x\|^2 = 0,$$

so $Y^{1/2}x = 0$ and therefore $Yx = 0$. Hence $(\text{range } W)^\perp = \ker W^* \subseteq \ker Y$. This implies $Y(\mathbf{1}_{\mathcal{V}} - P) = 0$, that is, $Y = YP$. Taking adjoints and using that P and Y are self-adjoint, we obtain $Y = PY$, so that $Y = PYP$.

Let $T = W^+Y(W^+)^*$. This satisfies

$$WTW^* = (WW^+)Y(WW^+)^* = PYP = Y.$$

Since conjugation preserves the Loewner order, $0 \leq Y \leq WW^*$ implies

$$0 \leq T \leq W^+WW^*(W^+)^* = (W^+W)(W^+W)^* = \mathbf{1}_{\mathbb{C}^r}.$$

Finally, suppose that $T, T' \in L(\mathbb{C}^r)$ satisfy $WTW^* = WT'W^*$. Multiplying by W^+ on the left and by $(W^+)^*$ on the right, and using $W^+W = \mathbf{1}_{\mathbb{C}^r}$, we obtain $T = T'$. $\square$

Theorem 4.2.2 (Arveson’s Radon–Nikodym theorem, Kraus form). *Let $\Phi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ with $\Phi(X) = \sum_{k=1}^r V_k X V_k^*$, where $V_1, \dots, V_r \in L(\mathcal{X}, \mathcal{Y})$ are linearly independent. For $T \in L(\mathbb{C}^r)$ define the linear map*

$$\Psi_T(X) = \sum_{k,l=1}^r T_{kl} V_k X V_l^*.$$

A linear map $\Psi : L(\mathcal{X}) \rightarrow L(\mathcal{Y})$ satisfies

$$0 \leq \Psi \leq \Phi$$

in the completely positive ordering (that is, Ψ and $\Phi - \Psi$ are completely positive) if and only if $\Psi = \Psi_T$ for some $T \in L(\mathbb{C}^r)$ satisfying $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$. This T is unique. Moreover, with the injective operator $W = [\text{vec}(V_1) \ \cdots \ \text{vec}(V_r)]$, we have $J(\Psi_T) = WTW^$ for every $T \in L(\mathbb{C}^r)$.*

Proof. Let $W = [\text{vec}(V_1) \ \cdots \ \text{vec}(V_r)]$. Since vec is a linear bijection and the V_k are linearly independent, W has full column rank. As the map $X \mapsto AXB^*$ has Choi matrix $\text{vec}(A)\text{vec}(B)^*$ [127, Section 2.2.2], linearity of J gives $J(\Psi_T) = WTW^*$ for every T , and in particular $J(\Phi) = WW^*$. If $0 \leq \Psi \leq \Phi$, then by Choi’s theorem [28] $0 \leq J(\Psi) \leq$

$J(\Phi) = WW^*$, so Lemma 4.2.1 yields a unique T with $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$ and $J(\Psi) = WTW^* = J(\Psi_T)$; injectivity of J gives $\Psi = \Psi_T$, and uniqueness of T follows from the uniqueness in Lemma 4.2.1. Conversely, if $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$, then $J(\Psi_T) = WTW^* \geq 0$ and $J(\Phi - \Psi_T) = W(\mathbf{1}_{\mathbb{C}^r} - T)W^* \geq 0$, so Ψ_T and $\Phi - \Psi_T$ are completely positive by Choi's theorem, that is, $0 \leq \Psi_T \leq \Phi$. $\square$

This theorem really says that we have an order preserving bijection

$$\{T \in L(\mathbb{C}^r) : 0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}\} \leftrightarrow \{\Psi \in \text{CP}(\mathcal{X}, \mathcal{Y}) : \Psi \leq \Phi\}.$$

We make a definition which differs slightly from the general C^* -algebra version, this is purely out of convenience because in the finite-dimensional case, the corresponding commutant elements are decided by positive semidefinite operators [100].

Definition 4.2.3. Let Φ and Ψ be as above. We say that the matrix T is the ARN derivative of Ψ with respect to Φ (relative to the Kraus decomposition $V_1, \dots, V_r$ of Φ).

Remark 4.2.4. Two Kraus decompositions $(V_k)_{k=1}^r$ and $(V'_j)_{j=1}^r$ of Φ by linearly independent operators are related by a unitary $U \in L(\mathbb{C}^r)$, $V'_j = \sum_k U_{jk} V_k$ [127, Section 2.2.2]. A direct computation shows that the corresponding ARN derivatives of a fixed $\Psi \leq \Phi$ are related by $T' = \bar{U} T U^\top$. In particular T and T' have the same spectrum, and all unitarily invariant norms of T , and of differences $T_1 - T_2$ of derivatives computed in the same Kraus decomposition, are independent of the decomposition chosen.

For a nonzero completely positive map Ω with Kraus decomposition $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$ by linearly independent operators $V_1, \dots, V_r$, let

$$\nu(\Omega) = \min \left\{ \left\| \sum_{k=1}^r c_k V_k \right\|_\infty : c \in \mathbb{C}^r, \|c\|_2 = 1 \right\} > 0.$$

This quantity relates the operator norm of a derivative to the completely bounded trace norm of the corresponding map; its properties are collected in Lemma 4.8.2 and Theorem 4.8.4 in Section 4.8.

4.2.2 The effect point of view

One may represent a completely positive map as the compression of an operator which acts on the tensor product of the underlying space and some ancillary space. One can read the ARN derivative as saying that if $\Psi \leq \Phi$, then Ψ can be understood as applying an effect on the ancilla space with respect to which the Stinespring dilation [119] of Φ is written. Let $\Phi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ have Kraus decomposition $\Phi(X) = \sum_{k=1}^r V_k X V_k^*$ with linearly independent V_k , let $\mathcal{Z} = \mathbb{C}^r$ with standard basis $e_1, \dots, e_r$, and let $S \in L(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be the Stinespring operator written in the style of [127, Section 2.2.2]

$$Sx = \sum_{k=1}^r V_k x \otimes e_k,$$

so that $\Phi(X) = \text{Tr}_{\mathcal{Z}}(SXS^*)$. We can then consider the complementary completely positive map [33] $\Phi^c(X) = \text{Tr}_{\mathcal{Y}}(SXS^*)$.

Theorem 4.2.5. For Hermitian $T \in L(\mathcal{Z})$ we have

1. $\Psi_T(X) = \text{Tr}_{\mathcal{Z}}((\mathbf{1}_{\mathcal{Y}} \otimes T^\top)SXS^*)$ for all $X \in L(\mathcal{X})$;

$$2. \operatorname{tr}(\Psi_T(X)) = \operatorname{tr}(T^\top \Phi^c(X));$$

$$3. \Psi_T^*(\mathbf{1}_Y) = S^*(\mathbf{1}_Y \otimes T^\top)S = (\Phi^c)^*(T^\top).$$

If $0 \leq T \leq \mathbf{1}_Z$, one can consider Ψ_T as the outcome of applying Φ and measuring an effect $T^\top$ on the ancillary environment generated by Φ . In particular, one can decompose each channel Φ as a sum $\sum_i \Psi_{T_i}$ where the $T_i^\top$ form a POVM on the environment Z . Moreover, for $T \geq 0$ the completely bounded trace norm (which coincides with the induced trace norm for completely positive maps [103]) satisfies

$$|||\Psi_T|||_1 = \max_{\rho \in \mathcal{D}(\mathcal{X})} \operatorname{tr}(T^\top \Phi^c(\rho)).$$

Proof. We have $SXS^* = \sum_{kl} V_k X V_l^* \otimes e_k e_l^*$, hence

$$\operatorname{Tr}_Z \left((\mathbf{1}_Y \otimes T^\top) SXS^* \right) = \sum_{kl} \operatorname{tr}(T^\top e_k e_l^*) V_k X V_l^* = \sum_{kl} T_{kl} V_k X V_l^* = \Psi_T(X),$$

since $\operatorname{tr}(T^\top e_k e_l^*) = (T^\top)_{lk} = T_{kl}$. This proves (1).

Taking the full trace in (1) and using

$$\operatorname{tr}((\mathbf{1}_Y \otimes T^\top) SXS^*) = \operatorname{tr}(T^\top \operatorname{Tr}_Y(SXS^*))$$

gives (2).

For (3), the adjoint of

$$X \mapsto (\mathbf{1}_Y \otimes T^\top) SXS^*$$

is

$$R \mapsto S^*(\mathbf{1}_Y \otimes \bar{T})RS$$

and the adjoint of Tr_Z is

$$Y \mapsto Y \otimes \mathbf{1}_Z.$$

Hence we have that

$$\Psi_T^*(Y) = S^*(\mathbf{1}_Y \otimes \bar{T})(Y \otimes \mathbf{1}_Z)S = S^*(Y \otimes \bar{T})S = S^*(Y \otimes T^\top)S,$$

where we have used that $\bar{T} = T^\top$ for Hermitian T , and letting $Y = \mathbf{1}_Y$ and comparing with $(\Phi^c)^*(Z) = S^*(\mathbf{1}_Y \otimes Z)S$. Finally, if $T \geq 0$ then Ψ_T is completely positive, so we may apply the Russo–Dye theorem [115] in the form [103, Proposition 3.6] to conclude

$$|||\Psi_T|||_1 = \|\Psi_T^*(\mathbf{1}_Y)\|_\infty = \max_{\rho \in \mathcal{D}(\mathcal{X})} \operatorname{tr}(\Psi_T^*(\mathbf{1}_Y)\rho) = \max_{\rho \in \mathcal{D}(\mathcal{X})} \operatorname{tr}(\Psi_T(\rho)),$$

and (2) gives the last formula. $\square$

Remark 4.2.6 (Physical interpretation). Applied to $\Omega = \Phi + \Psi$, the theorem shows that $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}_{\mathcal{C}^*} - T}$, where T is the derivative of Φ with respect to Ω , arise by running a Stinespring dilation of Ω and measuring the POVM $\{T^\top, \mathbf{1}_Z - T^\top\}$ on its environment. So (Φ, Ψ) is a two-outcome instrument implementing Ω (when Ω is a channel), and $T^\top$ is the which-branch effect hidden in the environment. Classically, $t = d\mathbb{P}/d(\mathbb{P} + \mathbb{Q})$ is the posterior probability of $\mathbb{P}$ under a uniform prior and $D_f(\mathbb{P}||\mathbb{Q}) = \int h_f(t) d(\mathbb{P} + \mathbb{Q})$, with h_f as in Lemma 4.7.2; by analogy, Theorem 4.7.5(b) is its channel version.

4.2.3 Kubo–Ando means of operators

Means of channels were first considered by Frenkel, Mosonyi, Vrana and Weiner [47], who introduced weighted Kubo–Ando geometric means of completely positive maps; in [76], one considers general Kubo–Ando means of channels. Hence we must take a quick detour and list some background facts.

In order to consider any kind of notion of means of channels using Kubo–Ando theory, as performed in [76], we must first recall the Kubo–Ando theory of means of positive semidefinite operators. The following is an amalgamation of facts from [16], [59], [62], and the original paper [79] and none of these results are claimed as ours. In order to make later steps more compatible with quantum information theory related texts, we rephrase some statements or change notation to be consistent with [127]. This will be consistent across the rest of the text.

Matrix means are usually studied through operator connections. Although operator connections need only satisfy a few general properties, they admit several useful classification theorems. We will be denoting complex Euclidean spaces using calligraphic letters, $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$, etc. Many of the philosophies in this section carry over to the analogous operator convex setting, where one defines relations between positive semidefinite matrices using operator convex functions (see [62], [16]). However, by shifting the point of view appropriately, we can obtain the desired results for some divergences without considering the techniques of Section 4.6.

Definition 4.2.7 (Loewner order). Given a complex Euclidean space $\mathcal{X}$, the Loewner order is the partial order on the Hermitian operators given by

$$A \leq B \iff B - A \geq 0$$

in the sense that $B - A$ is positive semidefinite.

Definition 4.2.8 (Operator connection [79]). A family of binary operations $(A, B) \mapsto A \sigma B$ on $\text{Pos}(\mathcal{X})$, one for each complex Euclidean space $\mathcal{X}$, is an operator connection if it satisfies the following on each $\mathcal{X}$

- (monotonicity) $0 \leq A \leq A'$ and $0 \leq B \leq B'$ imply $A \sigma B \leq A' \sigma B'$;
- (transformer inequality) for $A, B \geq 0$ and arbitrary C ,

$$C^*(A \sigma B)C \leq (C^*AC) \sigma (C^*BC);$$

- (continuity from above) if $A_n \downarrow A$ and $B_n \downarrow B$ then

$$(A_n \sigma B_n) \downarrow (A \sigma B),$$

where $A_n \downarrow A$ indicates that $A_n \rightarrow A$ in the operator norm and $A_1 \geq A_2 \geq \dots$

In particular, for $A, B \in \text{Pos}(\mathcal{X})$ one has

$$A \sigma B = \lim_{\eta \downarrow 0} A_\eta \sigma B_\eta,$$

where $A_\eta = A + \eta \mathbf{1}_{\mathcal{X}}$ and $B_\eta = B + \eta \mathbf{1}_{\mathcal{X}}$, so that a connection is determined by its values on positive definite operators.

4.2.4 Constructing operator connections

We will see that we can construct operator connections in a standard way. We need to introduce several definitions before we continue.

Definition 4.2.9 (Operator monotone function [16, Chapter V]). Let (t_-, t_+) be an open interval. A function $f : (t_-, t_+) \rightarrow \mathbb{R}$ is said to be operator monotone (or matrix monotone) on (t_-, t_+) if $f(A) \leq f(B)$ whenever $A \leq B$ are Hermitian operators, of any size, with spectra contained in (t_-, t_+) .

Definition 4.2.10 (Pick function [16, Chapter V]). Let $\mathbb{C}^+ = \{z \in \mathbb{C} : \text{Im}(z) > 0\}$. A function $f : \mathbb{C}^+ \rightarrow \mathbb{C}$ is called a Pick function if it is analytic and $\text{Im} f(z) \geq 0$ for all $z \in \mathbb{C}^+$ (by the open mapping theorem, a nonconstant Pick function maps $\mathbb{C}^+$ into $\mathbb{C}^+$). We denote the set of all Pick functions by $\mathcal{P}$. Given a real interval (t_-, t_+) , we denote by $\mathcal{P}(t_-, t_+)$ the set of all Pick functions which extend analytically across (t_-, t_+) with real values on (t_-, t_+) .

Finding operator monotone functions may appear to be a tricky business; however, they are characterised by a theorem of Loewner [85]; see also [16, Chapter V].

Theorem 4.2.11 (Loewner [85]). *Let (t_-, t_+) be a real interval (possibly all of $\mathbb{R}$) and f a real valued function on (t_-, t_+) . Then f is operator monotone on (t_-, t_+) if and only if f is the restriction to (t_-, t_+) of a function in $\mathcal{P}(t_-, t_+)$.*

The next theorem tells us that operator connections correspond to operator monotone functions, and the general theory of operator means defined in this way shares the name Kubo–Ando with this theorem.

Theorem 4.2.12 (Kubo–Ando theorem [79]). *For each operator connection σ there exists a unique operator monotone function $f_\sigma : (0, \infty) \rightarrow [0, \infty)$, the representing function of σ , such that*

$$f_\sigma(t)\mathbf{1}_X = \mathbf{1}_X \sigma(t\mathbf{1}_X)$$

for every $t > 0$, and, for $A > 0$ and $B \geq 0$,

$$A \sigma B = A^{1/2} f_\sigma(A^{-1/2} B A^{-1/2}) A^{1/2} = f_\sigma(B A^{-1}) A$$

via the functional calculus, with $f_\sigma(0) = f_\sigma(0^+)$ when B is singular. If A is not invertible, one perturbs it and takes limits from above. Conversely, every operator monotone $f : (0, \infty) \rightarrow [0, \infty)$ defines, by this formula, an operator connection whose representing function is f .

Definition 4.2.13. If also $f_\sigma(1) = 1$, that is $\mathbf{1}_X \sigma \mathbf{1}_X = \mathbf{1}_X$, we call σ an operator mean.

Remark 4.2.14. The term operator mean varies in the literature, and one must take care here. We do not impose on operator means that they be commutative unlike other authors.

For example, the representing functions $\sqrt{t}$, t^α with $0 < \alpha < 1$, and $t/(1+t)$ give, respectively, the geometric mean [79, 110], the weighted geometric mean [79] and the parallel sum [3],

$$\begin{aligned} A \# B &= A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}, \\ A \#_\alpha B &= A^{1/2} (A^{-1/2} B A^{-1/2})^\alpha A^{1/2}, \\ A : B &= (A^{-1} + B^{-1})^{-1}, \end{aligned}$$

for positive definite A, B , extended to positive semidefinite operators by the regularisation above.

We then see by combining Theorems 4.2.12 and 4.2.11 that every nonnegative function in $\mathcal{P}(0, \infty)$ gives rise to an operator connection.

4.2.5 The mean of completely positive maps

We now come to the construction developed by [76]. Given two completely positive maps $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, let $\Omega = \Phi + \Psi$. If Ω has Kraus decomposition $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$ with linearly independent V_k , then

$$(\Phi \sigma \Psi)(X) = \sum_{k,l=1}^r (T_1 \sigma T_2)_{kl} V_k X V_l^*$$

where T_1 and T_2 are the ARN derivatives of Φ and Ψ with respect to the Kraus decomposition of Ω . Kian defines the mean of two maps dominated by a common map through their Arveson–Radon–Nikodym derivatives, so that the derivative of the mean is the mean of the derivatives [76, Proposition 3.5], and the formula above is, in our notation, the finite-dimensional Kraus form of this definition, which [76, Proposition 3.13] gives in the Heisenberg picture. The choice $\Omega = \Phi + \Psi$ gives what Kian calls the intrinsic Kubo–Ando mean [76, Remark 3.3]. In general the relative mean depends on the dominating map [76, Remark 3.3], but not in our setting: the relative geometric mean does not depend on it [76, Remark 4.2], and for maps between matrix algebras the same holds for every Kubo–Ando mean [76, Theorem 5.1]. One observation which is easily made is that under the ARN bijection, since the derivative of Ω with respect to itself is $\mathbf{1}_{\mathbb{C}^r}$, we must have $T_2 = \mathbf{1}_{\mathbb{C}^r} - T_1$. We now formulate the fundamental theorem of means of completely positive maps which will inform the rest of the chapter. Let σ be a Kubo–Ando mean, so that its representing function f_σ is positive on $(0, \infty)$ and $f_\sigma(1) = 1$. For scalars $s, t > 0$ we have $s \sigma t = s f_\sigma(t/s)$. We define the function

$$g_\sigma(t) = t \sigma (1 - t) = t f_\sigma\left(\frac{1-t}{t}\right), \quad 0 < t < 1.$$

For a function f on $(0, \infty)$ we write $f(0^+) = \lim_{t \downarrow 0} f(t)$ and $f'(\infty) = \lim_{t \rightarrow \infty} f(t)/t$ whenever these limits exist, following [101]. For $f = f_\sigma$ both limits exist in $[0, \infty)$, since f_σ is non-decreasing and concave [16, Theorem V.2.5], and we set $g_\sigma(0) = f'_\sigma(\infty)$ and $g_\sigma(1) = f_\sigma(0^+)$, which are the limits of g_σ at the endpoints. The following theorem is essentially due to Kian: the Choi-level identity [76, Theorem 5.1], together with the Kraus form [76, Proposition 3.13], gives

$$J(\Phi) \sigma J(\Psi) = J(\Phi \sigma \Psi) = W(T \sigma (\mathbf{1}_{\mathbb{C}^r} - T)) W^*$$

in the notation of the theorem, and for the geometric mean the identity $J(\Phi \# \Psi) = J(\Phi) \# J(\Psi)$ is also [100, Proposition 4.1]. We include a proof for completeness in the present notation; what we add is the identity $T \sigma (\mathbf{1}_{\mathbb{C}^r} - T) = g_\sigma(T)$, through the functional calculus, and the resulting description of the spectrum.

Theorem 4.2.15. *Let σ be a Kubo–Ando mean with representing function f_σ , let $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, let $\Omega = \Phi + \Psi$ have a Kraus decomposition with linearly independent operators V_k , let $W = [\text{vec}(V_1) \cdots \text{vec}(V_r)]$, and let T be the ARN derivative of Φ with respect to Ω . Then at the Choi level*

$$J(\Phi) \sigma J(\Psi) = W g_\sigma(T) W^* = W \left(T^{1/2} f_\sigma(T^{-1/2} (\mathbf{1}_{\mathbb{C}^r} - T) T^{-1/2}) T^{1/2} \right) W^*,$$

where in the case of singular T or $\mathbf{1}_{\mathbb{C}^r} - T$ one takes decreasing nonsingular regularisations from above, using the continuity from above of operator connections. Consequently the mean $\Phi \sigma \Psi$ is a submap of $\Phi + \Psi$ with ARN derivative $g_\sigma(T)$, whose spectrum is $g_\sigma(\text{spec}(T))$.

Proof. Let $W^+ = (W^*W)^{-1}W^*$ and $P = WW^+$ be as in the proof of Lemma 4.2.1, so that $W^+W = \mathbf{1}_{\mathbb{C}^r}$ and P is the orthogonal projection onto $\text{range}(W)$. We first show that

$$(WAW^*)\sigma(WBW^*) = W(A\sigma B)W^* \quad \text{for all } A, B \in \text{Pos}(\mathbb{C}^r). \quad (4.2.1)$$

Let $M = (WAW^*)\sigma(WBW^*)$. We first note that the transformer inequality also holds for rectangular $C \in L(\mathcal{V}_1, \mathcal{V}_2)$, by embedding in $\mathcal{V}_1 \oplus \mathcal{V}_2$ with zero extension. This uses that $(A \oplus 0)\sigma(B \oplus 0) = (A\sigma B) \oplus 0$, by the Kubo–Ando formula for positive definite operators and regularisation. The transformer inequality with $C = W^*$ gives $W(A\sigma B)W^* \leq M$. With $c_{A,B} = \max\{\|A\|_\infty, \|B\|_\infty\}$, monotonicity gives

$$M \leq (c_{A,B}WW^*)\sigma(c_{A,B}WW^*) = c_{A,B}WW^*,$$

since $X\sigma X = X$ for a mean [79]; hence $\ker W^* \subseteq \ker M$, and $M = PMP$. The transformer inequality with $C = (W^+)^*$ gives

$$W^+M(W^+)^* \leq (W^+WAW^*(W^+)^*)\sigma(W^+WBW^*(W^+)^*) = A\sigma B,$$

and conjugating by W yields $M = PMP \leq W(A\sigma B)W^*$, which proves (4.2.1).

By the proof of Theorem 4.2.2, $J(\Phi) = WTW^*$ and $J(\Psi) = W(\mathbf{1}_{\mathbb{C}^r} - T)W^*$, so (4.2.1) gives $J(\Phi)\sigma J(\Psi) = W(T\sigma(\mathbf{1}_{\mathbb{C}^r} - T))W^*$, and the second expression in the statement is the Kubo–Ando formula when T and $\mathbf{1}_{\mathbb{C}^r} - T$ are invertible. It remains to show that $T\sigma(\mathbf{1}_{\mathbb{C}^r} - T) = g_\sigma(T)$. For $\eta > 0$ the operators $T + \eta\mathbf{1}_{\mathbb{C}^r}$ and $(\mathbf{1}_{\mathbb{C}^r} - T) + \eta\mathbf{1}_{\mathbb{C}^r}$ are positive definite and commute, so by Theorem 4.2.12 and the functional calculus,

$$(T + \eta\mathbf{1}_{\mathbb{C}^r})\sigma(\mathbf{1}_{\mathbb{C}^r} - T + \eta\mathbf{1}_{\mathbb{C}^r}) = g_{\sigma,\eta}(T)$$

with $g_{\sigma,\eta}(t) = (t + \eta)f_\sigma(\frac{1-t+\eta}{t+\eta})$. As $\eta \downarrow 0$ the left-hand side decreases to $T\sigma(\mathbf{1}_{\mathbb{C}^r} - T)$ by continuity from above, while $g_{\sigma,\eta}(t) \rightarrow g_\sigma(t)$ for every $t \in [0, 1]$: for $0 < t < 1$ by continuity of f_σ , and at $t = 0$ and $t = 1$ by the definitions of $f'_\sigma(\infty)$ and $f_\sigma(0^+)$. Since T is a positive semidefinite operator and σ is a Kubo–Ando mean, $g_{\sigma,\eta}(T) \rightarrow g_\sigma(T)$, and hence $T\sigma(\mathbf{1}_{\mathbb{C}^r} - T) = g_\sigma(T)$.

Finally, for $0 < t < 1$ monotonicity gives

$$0 \leq g_\sigma(t) = t\sigma(1-t) \leq \mathbf{1}\sigma\mathbf{1} = 1,$$

and the same holds at the endpoints by taking limits. Thus $0 \leq g_\sigma(T) \leq \mathbf{1}_{\mathbb{C}^r}$, and $\Phi\sigma\Psi = \Psi T\sigma(\mathbf{1}_{\mathbb{C}^r} - T) = \Psi_{g_\sigma(T)}$ is a submap of Ω with ARN derivative $g_\sigma(T)$ by Theorem 4.2.2. The statement about the spectrum is the spectral mapping theorem. $\square$

By Theorem 4.2.15 and injectivity of J , the map $\Phi\sigma\Psi$ is determined by $J(\Phi)\sigma J(\Psi)$, so it does not depend on the Kraus decomposition of Ω used to define it. It is completely positive, but even for channels Φ and Ψ it is in general only trace non-increasing: concavity of f_σ gives $f_\sigma(x) \leq 1-w+wx$ with $w = f'_\sigma(1)$, which lies in $[0, 1]$ because f_σ is nondecreasing and $f_\sigma(0^+) \geq 0$; hence $\Phi\sigma\Psi \leq (1-w)\Phi + w\Psi$ in the completely positive order, for instance $\Phi\#\Psi \leq \frac{1}{2}(\Phi + \Psi)$, and $\Phi\#\Psi = 0$ for the replacement channels of Proposition 4.3.2 below. Thus a mean of channels is a completely positive map, but in general not a channel.

Remark 4.2.16. The proof of Theorem 4.2.15 uses only monotonicity, the transformer inequality, continuity from above and the identity $X\sigma X = f_\sigma(1)X$. It therefore applies verbatim to every operator connection σ , with representing function f_σ as in Theorem 4.2.12. Put

$$\varphi_\sigma(t) = t\sigma(1-t) = tf_\sigma\left(\frac{1-t}{t}\right) \quad (0 < t < 1), \quad \varphi_\sigma(0) = f'_\sigma(\infty), \quad \varphi_\sigma(1) = f_\sigma(0^+),$$

which are the limits of φ_σ at the endpoints, since f_σ is nonnegative, nondecreasing and concave. Then:

1. (4.2.1) holds for the connection σ , that is, $(WAW^*)\sigma(WBW^*) = W(A\sigma B)W^*$ for all $A, B \in \text{Pos}(\mathbb{C}^r)$;
2. $T\sigma(\mathbf{1}_{\mathbb{C}^r} - T) = \varphi_\sigma(T)$ for $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$, in the functional calculus.

If σ is a mean, then $\varphi_\sigma = g_\sigma$.

4.3 The mean of channels and geometric channel fidelity

In this section, we present a concrete application of the means of completely positive maps in terms of the ARN derivative. In quantum information theory, one can consider many kinds of divergence. One of these is the geometric channel fidelity. This quantity has some annoying properties. In this section, we will build up the definition through the Matsumoto fidelity [90]. We will retranslate some definitions in [90] and [72] in the language of Watrous [127].

4.3.1 Matsumoto fidelity

Motivation of the definition

In quantum information theory, one often wants to consider how much two quantum states “overlap”. One ubiquitous example of this is the quantum (or Uhlmann) fidelity [122] of two states. Given a complex Euclidean space $\mathcal{X}$ and density operators $\rho, \xi \in \text{D}(\mathcal{X})$, their fidelity is defined as

$$F(\rho, \xi) = \text{tr} \left(\sqrt{\sqrt{\xi} \rho \sqrt{\xi}} \right) = \|\sqrt{\rho} \sqrt{\xi}\|_1.$$

The quantum fidelity is the analogue of the Bhattacharyya coefficient in classical probability (see, for example, [127, Section 3.2]). The fidelity is closely related to the geometric mean: for invertible ξ one checks directly that

$$F(\rho, \xi) = \text{tr} (\xi (\xi^{-1} \# \rho)).$$

In [90], Matsumoto considers flipping the definition for quantum fidelity, and instead takes the trace of the geometric mean of the operators themselves. The Matsumoto fidelity is defined on states $\rho, \xi \in \text{D}(\mathcal{X})$ by

$$F_M(\rho, \xi) = \text{tr}(\rho \# \xi).$$

The following interpretation is given in [29]. Two classical probability distributions $\{p_i\}, \{q_i\}$ prepare states ρ and ξ if there exists a quantum channel Φ such that $\Phi(\text{diag}(\{p_i\})) = \rho$ and $\Phi(\text{diag}(\{q_i\})) = \xi$. The Matsumoto fidelity is the maximal possible classical fidelity between classical probability distributions that can prepare states ρ and ξ . The quantum fidelity is the minimum classical fidelity over quantum-to-classical measurement procedures; the Matsumoto fidelity is the maximum classical fidelity over classical-to-quantum preparations. In this sense, it is a kind of dual to the quantum fidelity [29]. Comprehensive properties of the Matsumoto fidelity can be found in [29] and the original paper by Matsumoto [90].

4.3.2 Geometric channel fidelity

Given two channels $\Phi, \Psi \in \text{C}(\mathcal{X}, \mathcal{Y})$, one can consider the maximum output fidelity [127, Section 3.3] given by

$$F_{\max}(\Phi, \Psi) = \max_{\rho, \xi \in \text{D}(\mathcal{X})} F(\Phi(\rho), \Psi(\xi)).$$

The geometric channel fidelity as defined in [72] (see also [98, eq. (19)]) is a kind of opposite to this, using the Matsumoto fidelity. They define the geometric channel fidelity as the square of the following quantity:

$$\widehat{F}(\Phi, \Psi) = \inf_{\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})} F_M((\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho), (\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho)).$$

We will not be using the squared quantity, and will refer to the above as the geometric channel fidelity. We notice in this definition that there is only one input state, shared by both channels. This is because if we allowed two, it would be very easy to make this quantity zero.

In this section and the next we work under the assumption that we are considering completely positive maps, not necessarily channels. The definition of the geometric channel fidelity still makes sense, because complete positivity guarantees that $(\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho)$ and $(\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho)$ are positive semidefinite, which is all we require for the geometric mean to make sense; by Choi's theorem [28] the same holds for the Choi matrices $J(\Phi)$ and $J(\Psi)$.

Proposition 4.3.1. *For $\Phi, \Psi \in \mathcal{CP}(\mathcal{X}, \mathcal{Y})$, $\widehat{F}(\Phi, \Psi) = \lambda_{\min}(\text{Tr}_{\mathcal{Y}}(J(\Phi) \# J(\Psi)))$.*

Proof. For channels this is a consequence of [72, Proposition 44] at $\alpha = \frac{1}{2}$ (see also the semidefinite program for the square root of the geometric fidelity of channels in [98, eq. (20)]); we include the short argument since we need it for completely positive maps. Since F_M is jointly concave on pairs of positive semidefinite operators (for states see [29]; in general, F_M is positively homogeneous, and the transformer inequality with $C: v \mapsto v \oplus v$, applied to $A \oplus A'$ and $B \oplus B'$, together with $(A \oplus A') \# (B \oplus B') = (A \# B) \oplus (A' \# B')$, gives $(A \# B) + (A' \# B') \leq (A + A') \# (B + B')$ [79]) and $\rho \mapsto (\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho)$ is linear, the function being minimised is concave on the compact convex set $\mathcal{D}(\mathcal{X} \otimes \mathcal{X})$, so the infimum may be taken over pure states uu^* . Every unit vector $u \in \mathcal{X} \otimes \mathcal{X}$ is of the form $u = \text{vec}(A) = (\mathbf{1}_{\mathcal{X}} \otimes A^\top) \text{vec}(\mathbf{1}_{\mathcal{X}})$ with $\|A\|_2 = 1$ [127, Section 1.1.2], whence

$$(\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(uu^*) = (\mathbf{1}_{\mathcal{Y}} \otimes A^\top) J(\Phi) (\mathbf{1}_{\mathcal{Y}} \otimes A^\top)^*, \quad (4.3.1)$$

and similarly for Ψ . If A is invertible, the transformer equality for invertible operators [79] gives

$$\begin{aligned} \text{tr}((\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(uu^*) \# (\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(uu^*)) &= \text{tr}((\mathbf{1}_{\mathcal{Y}} \otimes \overline{A} A^\top) J(\Phi) \# J(\Psi)) \\ &= \text{tr}((A A^*)^\top \text{Tr}_{\mathcal{Y}}(J(\Phi) \# J(\Psi))), \end{aligned}$$

since $(\mathbf{1}_{\mathcal{Y}} \otimes A^\top)^*(\mathbf{1}_{\mathcal{Y}} \otimes A^\top) = \mathbf{1}_{\mathcal{Y}} \otimes (A A^*)^\top$. As A ranges over invertible operators with $\|A\|_2 = 1$, $(A A^*)^\top$ ranges over the invertible density operators, so the infimum over such u is

$$\lambda_{\min}(\text{Tr}_{\mathcal{Y}}(J(\Phi) \# J(\Psi))).$$

Finally, F_M is upper semicontinuous (it is a decreasing limit of the continuous functions $\text{tr}(\rho_\eta \# \xi_\eta)$), so the infimum over all pure states equals the infimum over the dense set of pure states with invertible A . $\square$

4.3.3 An alarming property

The Matsumoto fidelity is the trace of a geometric mean of positive semidefinite operators. It is well-known that if one takes the geometric mean of operators whose ranges have intersection $\{0\}$, then this mean is zero. This observation is stated in [29] as the distinct image property. This distinct image property extends to Choi matrices of completely positive maps. This leads to the following proposition.

Proposition 4.3.2. *Suppose $\dim \mathcal{Y} \geq 2$. There exist pairs of channels (Φ_1, Ψ_1) , (Φ_2, Ψ_2) from $\mathcal{X}$ to $\mathcal{Y}$ such that $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$ but $\widehat{F}(\Phi_1, \Psi_1) = 0$ and $\widehat{F}(\Phi_2, \Psi_2) = 1$.*

Proof. Let $\phi = y_1 y_1^*$ and $\psi = y_2 y_2^*$ be distinct pure states on $\mathcal{Y}$, which exist since $\dim \mathcal{Y} \geq 2$, and let $\Phi_1 = \text{tr}(\cdot) \phi$ and $\Psi_1 = \text{tr}(\cdot) \psi$ be the corresponding replacement channels. Their Choi matrices are

$$J(\Phi_1) = \phi \otimes \mathbf{1}_{\mathcal{X}}, \quad J(\Psi_1) = \psi \otimes \mathbf{1}_{\mathcal{X}},$$

with ranges $\text{span}\{y_1\} \otimes \mathcal{X}$ and $\text{span}\{y_2\} \otimes \mathcal{X}$. Since $\phi \neq \psi$, the unit vectors y_1 and y_2 are linearly independent, so these ranges intersect trivially. Hence $J(\Phi_1) \# J(\Psi_1) = 0$ by the distinct image property [29], and Proposition 4.3.1 gives $\widehat{F}(\Phi_1, \Psi_1) = \lambda_{\min}(0) = 0$.

Let $\Omega = \Phi_1 + \Psi_1$ and $\Phi_2 = \Psi_2 = \frac{1}{2}\Omega$, so that $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$ and $J(\Phi_2) = J(\Psi_2) = \frac{1}{2}(\phi + \psi) \otimes \mathbf{1}_{\mathcal{X}}$. Since $A \# A = A$, we have $J(\Phi_2) \# J(\Psi_2) = J(\Phi_2)$, and therefore

$$\widehat{F}(\Phi_2, \Psi_2) = \lambda_{\min}\left(\text{Tr}_{\mathcal{Y}}\left(\frac{1}{2}(\phi + \psi) \otimes \mathbf{1}_{\mathcal{X}}\right)\right) = \lambda_{\min}\left(\frac{1}{2} \text{tr}(\phi + \psi) \mathbf{1}_{\mathcal{X}}\right) = 1,$$

since ϕ and ψ are states. $\square$

Proposition 4.3.2 suggests something alarming. One may have two pairs of completely positive maps which sum to the same completely positive map, with each pair having opposite geometric fidelities. Stronger still, by rotating ϕ slightly, the two splittings can be made arbitrarily close in the completely bounded trace norm.

Corollary 4.3.3 (The completely bounded trace norm is ineffective for geometric fidelity). *Suppose $\dim \mathcal{Y} \geq 2$. For any $\eta > 0$, there exist channels $\Phi_1, \Psi_1, \Phi_2, \Psi_2 \in C(\mathcal{X}, \mathcal{Y})$ such that $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$,*

$$|||\Phi_1 - \Phi_2|||_1 = |||\Psi_1 - \Psi_2|||_1 < \eta$$

and

$$\widehat{F}(\Phi_1, \Psi_1) = 0, \quad \widehat{F}(\Phi_2, \Psi_2) = 1.$$

Proof. Choose $\theta \in (0, \pi/2)$ with $\sin \theta < \eta$, and let $\phi = y_1 y_1^*$ and $\psi = y_2 y_2^*$ be pure states on $\mathcal{Y}$ with $|\langle y_1, y_2 \rangle| = \cos \theta$; they are distinct since $\theta > 0$. Define $\Phi_1, \Psi_1, \Phi_2, \Psi_2$ as in the proof of Proposition 4.3.2. That proof gives $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$, $\widehat{F}(\Phi_1, \Psi_1) = 0$ and $\widehat{F}(\Phi_2, \Psi_2) = 1$.

It remains to compute the distances. We have

$$\Phi_1 - \Phi_2 = \frac{1}{2}(\Phi_1 - \Psi_1) = -(\Psi_1 - \Psi_2).$$

For $X_1, X_2 \in L(\mathcal{X})$,

$$((\Phi_1 - \Psi_1) \otimes \mathbf{1}_{L(\mathcal{X})})(X_1 \otimes X_2) = \text{tr}(X_1) (\phi - \psi) \otimes X_2,$$

so by linearity, for every $\rho \in D(\mathcal{X} \otimes \mathcal{X})$,

$$((\Phi_1 - \Psi_1) \otimes \mathbf{1}_{L(\mathcal{X})})(\rho) = (\phi - \psi) \otimes \rho_R, \quad \rho_R = (\text{tr} \otimes \mathbf{1}_{L(\mathcal{X})})(\rho) \in D(\mathcal{X}).$$

Since $\|A \otimes B\|_1 = \|A\|_1 \|B\|_1$ and $\|\rho_R\|_1 = 1$, this has trace norm

$$\|\phi - \psi\|_1 = 2\sqrt{1 - |\langle y_1, y_2 \rangle|^2} = 2 \sin \theta$$

for every ρ . As $\Phi_1 - \Psi_1$ is Hermitian-preserving, its completely bounded trace norm is the supremum of this quantity over $\rho \in D(\mathcal{X} \otimes \mathcal{X})$ [127, Section 3.3], so $|||\Phi_1 - \Psi_1|||_1 = 2 \sin \theta$ and

$$|||\Phi_1 - \Phi_2|||_1 = |||\Psi_1 - \Psi_2|||_1 = \sin \theta < \eta. \quad \square$$

Remark 4.3.4. The mechanism behind Proposition 4.3.2 and Corollary 4.3.3 is known at the level of states: the Matsumoto fidelity of two distinct pure states vanishes [29], and maximal f -divergences are lower semicontinuous [60, Section 5], [95, Theorem A.4] (see also [91]) but in general not continuous. Section 4.1.2 determines exactly where the quantities considered here are continuous.

We seek to address this problem through the ARN derivative, in particular to explain why it is allowed to happen. First we must take a detour and develop some norm inequalities.

4.4 Norm and continuity estimates

Throughout this section $||| \cdot |||_1$ denotes the completely bounded trace norm [77], [127, Section 3.3.2]. For completely positive maps it coincides with the induced trace norm and equals $\|\Phi^*(\mathbf{1}_Y)\|_\infty$; this follows from the Russo–Dye theorem [115], in the form [103, Proposition 3.6], together with the duality between the completely bounded trace norm and the completely bounded operator norm [68, Theorem 3].

Lemma 4.4.1 (Sharp Lipschitz inequality in ARN derivatives). *Let $0 \leq \Psi_1, \Psi_2 \leq \Phi$ with ARN derivatives T_1, T_2 with respect to a fixed Kraus decomposition of Φ by linearly independent operators. Then*

$$|||\Psi_{T_1} - \Psi_{T_2}|||_1 \leq |||\Psi_{|T_1 - T_2|}|||_1 \leq |||\Phi|||_1 \|T_1 - T_2\|_\infty.$$

Remark 4.4.2. The outer inequality was shown by Raginsky in [111].

Proof. Let $\Theta = \Psi_{T_1} - \Psi_{T_2}$ and write $T_1 - T_2 = A_+ - A_-$ with $A_\pm \geq 0$ and $A_+ A_- = 0$ (the Jordan decomposition), so that $|T_1 - T_2| = A_+ + A_-$. By linearity of $T \mapsto \Psi_T$ we have

$$\Psi_{T_1} - \Psi_{T_2} = \Psi_{A_+} - \Psi_{A_-}.$$

The map Θ is Hermitian-preserving, so by [127, Section 3.3], the completely bounded trace norm is attained on a pure state uu^* with $u \in \mathcal{X} \otimes \mathcal{X}$ a unit vector. Since uu^* is positive semidefinite, and each $\Psi_{A_\pm}$ is completely positive, we have

$$(\Theta \otimes \mathbf{1}_{L(\mathcal{X})})(uu^*) = X_+ - X_-$$

where $X_+ = (\Psi_{A_+} \otimes \mathbf{1}_{L(\mathcal{X})})(uu^*) \geq 0$ and $X_- = (\Psi_{A_-} \otimes \mathbf{1}_{L(\mathcal{X})})(uu^*) \geq 0$. Let ρ be the reduced state of uu^* on the first tensor factor, so that $\text{tr}((\Xi \otimes \mathbf{1}_{L(\mathcal{X})})(uu^*)) = \text{tr}(\Xi^*(\mathbf{1}_Y)\rho)$ for every map Ξ . By the triangle inequality and the fact that the trace norm of a positive operator is its trace, we obtain

$$\begin{aligned} |||\Theta|||_1 &= \|X_+ - X_-\|_1 \leq \text{tr}(X_+) + \text{tr}(X_-) = \text{tr}((\Psi_{A_+} + \Psi_{A_-})^*(\mathbf{1}_Y)\rho) \\ &= \text{tr}(\Psi_{|T_1 - T_2|}^*(\mathbf{1}_Y)\rho) \leq \|\Psi_{|T_1 - T_2|}^*(\mathbf{1}_Y)\|_\infty, \end{aligned}$$

the last step because $\Psi_{|T_1 - T_2|}^*(\mathbf{1}_Y) \geq 0$ and ρ is a state. Since $\Psi_{|T_1 - T_2|}$ is completely positive, [103, Proposition 3.6] and [68, Theorem 3] give

$$\|\Psi_{|T_1 - T_2|}^*(\mathbf{1}_Y)\|_\infty = |||\Psi_{|T_1 - T_2|}|||_1,$$

which proves the first inequality. For the second, Theorem 4.2.5 gives

$$\Psi_{|T_1 - T_2|}^*(\mathbf{1}_Y) = S^*(\mathbf{1}_Y \otimes |T_1 - T_2|^\top)S.$$

Since $0 \leq |T_1 - T_2|^\top \leq \|T_1 - T_2\|_\infty \mathbf{1}_Z$ and conjugation preserves the Loewner order,

$$0 \leq S^*(\mathbf{1}_Y \otimes |T_1 - T_2|^\top)S \leq \|T_1 - T_2\|_\infty S^*S = \|T_1 - T_2\|_\infty \Phi^*(\mathbf{1}_Y).$$

By the Weyl monotonicity principle [16, Section III.2], this inequality between positive operators carries over to the operator norm, so

$$\|S^*(\mathbf{1}_Y \otimes |T_1 - T_2|^\top)S\|_\infty \leq \|T_1 - T_2\|_\infty \|\Phi^*(\mathbf{1}_Y)\|_\infty = \|\Phi\|_1 \|T_1 - T_2\|_\infty,$$

where we have applied [103, Proposition 3.6] again in the last equality. $\square$

Proposition 4.4.3. 1. For a complex Euclidean space Z and $0 \leq E, E' \leq \mathbf{1}_Z$,

$$\|(E - E^2) - (E' - E'^2)\|_\infty \leq \|E - E'\|_\infty.$$

2. For E, E' as above, with $g_\#(t) = \sqrt{t(1-t)}$, we have

$$\|g_\#(E) - g_\#(E')\|_\infty \leq \|E - E'\|_\infty^{1/2}.$$

Hence, for completely positive pairs $(\Phi_1, \Psi_1), (\Phi_2, \Psi_2)$ with the same sum Ω and derivatives T_1 and T_2 of Φ_1 and Φ_2 with respect to Ω , relative to one fixed Kraus decomposition of Ω by linearly independent operators, we have that

$$\|\Phi_1 \# \Psi_1 - \Phi_2 \# \Psi_2\|_1 \leq \|\Omega\|_1 \|T_1 - T_2\|_\infty^{1/2}.$$

Proof. 1. Let $A = E - E'$ and $B = \mathbf{1}_Z - E - E'$. Since

$$E^2 - E'^2 = \frac{1}{2}(A(E + E') + (E + E')A),$$

we have

$$(E - E^2) - (E' - E'^2) = \frac{1}{2}(AB + BA).$$

Since $0 \leq E + E' \leq 2 \cdot \mathbf{1}_Z$, we have $-\mathbf{1}_Z \leq B \leq \mathbf{1}_Z$, so $\|B\|_\infty \leq 1$. By the triangle inequality and submultiplicativity of the operator norm,

$$\|(E - E^2) - (E' - E'^2)\|_\infty \leq \|A\|_\infty \|B\|_\infty \leq \|A\|_\infty = \|E - E'\|_\infty.$$

2. Since E and $\mathbf{1}_Z - E$ commute, we have $g_\#(E) = (E - E^2)^{1/2} \geq 0$. Since the square root is operator monotone on $[0, \infty)$ and $\sqrt{0} = 0$, we have by [7] (see also [16, Theorem X.1.1]) that $\|\sqrt{X} - \sqrt{Y}\|_\infty \leq \|X - Y\|_\infty^{1/2}$ for $X, Y \geq 0$, and therefore, using part (1),

$$\|g_\#(E) - g_\#(E')\|_\infty \leq \|(E - E^2) - (E' - E'^2)\|_\infty^{1/2} \leq \|E - E'\|_\infty^{1/2}.$$

Suppose now that $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2 = \Omega$ and let T_1 and T_2 be the respective ARN derivatives of Φ_1 and Φ_2 with respect to Ω . By Theorem 4.2.15, $\Phi_i \# \Psi_i = \Psi_{g_\#(T_i)}$, so Lemma 4.4.1 (applied with $\Phi = \Omega$) gives

$$\|\Phi_1 \# \Psi_1 - \Phi_2 \# \Psi_2\|_1 \leq \|\Omega\|_1 \|g_\#(T_1) - g_\#(T_2)\|_\infty.$$

Using part (2), we obtain

$$\|\Phi_1 \# \Psi_1 - \Phi_2 \# \Psi_2\|_1 \leq \|\Omega\|_1 \|T_1 - T_2\|_\infty^{1/2}.$$

$\square$

Remark 4.4.4. Both the constant and the exponent in part (2) are sharp: for 1×1 matrices $E = \delta$ and $E' = 0$ one has $\|g_{\#}(E) - g_{\#}(E')\|_{\infty} = \sqrt{\delta(1-\delta)}$, so the ratio to $\|E - E'\|_{\infty}^{1/2} = \sqrt{\delta}$ tends to 1 as $\delta \rightarrow 0$. The same example, with $\Omega = V(\cdot)V^*$ for a single Kraus operator $V \neq 0$, $T_1 = \delta$ and $T_2 = 0$, shows that the constant and the exponent in the last display of Proposition 4.4.3 are sharp as well.

Theorem 4.4.5. *Let $\Omega = \Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$ be completely positive maps and let T_i be the ARN derivative of Φ_i with respect to Ω , relative to one fixed Kraus decomposition of Ω by linearly independent operators, the same for $i = 1, 2$. For a continuous function $g : [0, 1] \rightarrow \mathbb{R}$ let*

$$\omega_g(\delta) = \sup \{ \|g(E) - g(E')\|_{\infty} : 0 \leq E, E' \leq \mathbf{1}_{\mathcal{Z}}, \|E - E'\|_{\infty} \leq \delta \}, \quad \delta \geq 0,$$

the supremum being over all finite-dimensional $\mathcal{Z}$. Then, for every Kubo–Ando mean σ ,

$$\| \Phi_1 \sigma \Psi_1 - \Phi_2 \sigma \Psi_2 \|_1 \leq \| \Omega \|_1 \omega_{g_{\sigma}}(\|T_1 - T_2\|_{\infty}).$$

Proof. By Theorem 4.2.15,

$$\Phi_1 \sigma \Psi_1 - \Phi_2 \sigma \Psi_2 = \Psi_K$$

where $K = g_{\sigma}(T_1) - g_{\sigma}(T_2)$. By definition of $\omega_{g_{\sigma}}$,

$$\|K\|_{\infty} \leq \omega_{g_{\sigma}}(\|T_1 - T_2\|_{\infty}).$$

Since $g_{\sigma}(T_1)$ and $g_{\sigma}(T_2)$ are effects, Lemma 4.4.1 gives $\|\Psi_K\|_1 \leq \|\Omega\|_1 \|K\|_{\infty}$, which implies the claim. $\square$

We now come to the main result, which suggests that in order to resolve this problematic behaviour of the geometric channel fidelity, one has to look at the ARN derivatives of the completely positive maps.

Theorem 4.4.6. *Suppose we are given completely positive maps $\Phi_1, \Psi_1, \Phi_2, \Psi_2$ such that $\Omega = \Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$. Let $\hat{F}_1 = \hat{F}(\Phi_1, \Psi_1)$ and $\hat{F}_2 = \hat{F}(\Phi_2, \Psi_2)$. Let T_1 and T_2 be the respective ARN derivatives of Φ_1 and Φ_2 with respect to Ω , relative to one fixed Kraus decomposition of Ω by linearly independent operators, the same for both pairs. Then*

$$|\hat{F}_1 - \hat{F}_2| \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty} \|T_1 - T_2\|_{\infty}^{1/2}.$$

Proof. By Weyl's inequality [16, Section III.2], $\lambda_{\min}$ is 1-Lipschitz with respect to the operator norm on Hermitian operators. Let $K = g_{\#}(T_1) - g_{\#}(T_2)$, which is Hermitian. By Proposition 4.3.1 and Theorem 4.2.15,

$$\begin{aligned} |\hat{F}_1 - \hat{F}_2| &\leq \| \text{Tr}_{\mathcal{Y}}(J(\Phi_1) \# J(\Psi_1) - J(\Phi_2) \# J(\Psi_2)) \|_{\infty} \\ &= \| \text{Tr}_{\mathcal{Y}}(W K W^*) \|_{\infty}. \end{aligned}$$

Since

$$-\|K\|_{\infty} W W^* \leq W K W^* \leq \|K\|_{\infty} W W^*$$

and the partial trace is a positive map,

$$-\|K\|_{\infty} \text{Tr}_{\mathcal{Y}}(W W^*) \leq \text{Tr}_{\mathcal{Y}}(W K W^*) \leq \|K\|_{\infty} \text{Tr}_{\mathcal{Y}}(W W^*),$$

and for a Hermitian H and positive M with $-sM \leq H \leq sM$ one has $\|H\|_{\infty} \leq s\|M\|_{\infty}$. Moreover

$$\text{Tr}_{\mathcal{Y}}(W W^*) = \text{Tr}_{\mathcal{Y}}(J(\Omega)) = \Omega^*(\mathbf{1}_{\mathcal{Y}})^{\top},$$

whose operator norm is $\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty}$. Hence, by Proposition 4.4.3(2),

$$|\hat{F}_1 - \hat{F}_2| \leq \|K\|_{\infty} \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty} \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty} \|T_1 - T_2\|_{\infty}^{1/2}. \quad \square$$

What Theorem 4.4.6 says is that if two pairs of completely positive maps have the same sum, then their respective geometric channel fidelities are controlled by the operator norm of the difference of their ARN derivatives. Coming back to Corollary 4.3.3, in the Kraus decomposition $\{V_i = y_1 e_i^*, V'_i = y_2 e_i^*\}_{i=1}^n$ of Ω , where $n = \dim \mathcal{X}$ and $e_1, \dots, e_n$ is an orthonormal basis of $\mathcal{X}$ (these $2n$ operators are linearly independent because y_1 and y_2 are), the derivative of Φ_1 is the projection $T_1 = \mathbf{1}_{\mathbb{C}^n} \oplus 0_{\mathbb{C}^n}$, while the splitting $(\frac{\Omega}{2}, \frac{\Omega}{2})$ has ARN derivative $T_2 = \frac{1}{2} \mathbf{1}_{\mathbb{C}^{2n}}$. Hence $\|T_1 - T_2\|_\infty = \frac{1}{2}$ for every $\theta > 0$, however small. The adjoint of Ω evaluated at $\mathbf{1}_{\mathcal{Y}}$ is

$$\Omega^*(\mathbf{1}_{\mathcal{Y}}) = (\text{tr}(\phi) + \text{tr}(\psi))\mathbf{1}_{\mathcal{X}} = 2 \cdot \mathbf{1}_{\mathcal{X}},$$

so the right-hand side of Theorem 4.4.6 equals

$$\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \|T_1 - T_2\|_\infty^{1/2} = 2 \cdot (\frac{1}{2})^{1/2} = \sqrt{2} \geq 1 = |\hat{F}_1 - \hat{F}_2|.$$

The completely bounded trace norm between the splittings tends to zero, but the distance between the derivatives does not, which is exactly what allows the fidelity to jump.

Remark (Sharpness of Theorem 4.4.6). The constant $\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty$ and the exponent $\frac{1}{2}$ in Theorem 4.4.6 are both attained asymptotically, already by channels. Let $\Phi_1, \Psi_1 \in \text{CP}(\mathcal{X}, \mathcal{Y})$ be nonzero with $\text{range } J(\Phi_1) \cap \text{range } J(\Psi_1) = \{0\}$, as for the replacement channels of Proposition 4.3.2, and for $0 < s \leq \frac{1}{2}$ put

$$\Phi_2 = (1 - s)\Phi_1 + s\Psi_1, \quad \Psi_2 = s\Phi_1 + (1 - s)\Psi_1,$$

so that $\Phi_2 + \Psi_2 = \Phi_1 + \Psi_1 = \Omega$. Since $J(\Phi_1) = WT_1W^*$ and $J(\Psi_1) = W(\mathbf{1}_{\mathbb{C}^r} - T_1)W^*$ have ranges $W \text{range } T_1$ and $W \text{range } (\mathbf{1}_{\mathbb{C}^r} - T_1)$ and W is injective, an eigenvector of T_1 with eigenvalue in $(0, 1)$ would give a nonzero vector in both ranges; hence T_1 is a projection, different from 0 and $\mathbf{1}_{\mathbb{C}^r}$ because $\Phi_1, \Psi_1 \neq 0$. By uniqueness in Theorem 4.2.2, $T_2 = (1 - s)T_1 + s(\mathbf{1}_{\mathbb{C}^r} - T_1)$, so

$$\|T_1 - T_2\|_\infty = \|s(2T_1 - \mathbf{1}_{\mathbb{C}^r})\|_\infty = s$$

and $\text{spec}(T_2) = \{s, 1 - s\}$, whence $g_\#(T_1) = 0$ and $g_\#(T_2) = \sqrt{s(1 - s)} \mathbf{1}_{\mathbb{C}^r}$. Theorem 4.2.15 and Proposition 4.3.1 give $\hat{F}_1 = 0$ and $\hat{F}_2 = \sqrt{s(1 - s)} \lambda_{\min}(\Omega^*(\mathbf{1}_{\mathcal{Y}}))^\top$, so

$$\frac{|\hat{F}_1 - \hat{F}_2|}{\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \|T_1 - T_2\|_\infty^{1/2}} = (1 - s)^{1/2} \frac{\lambda_{\min}(\Omega^*(\mathbf{1}_{\mathcal{Y}}))}{\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty}.$$

If Φ_1 and Ψ_1 are channels, then $\Omega^*(\mathbf{1}_{\mathcal{Y}}) = 2 \cdot \mathbf{1}_{\mathcal{X}}$ and the ratio is $(1 - s)^{1/2} \rightarrow 1$ as $s \rightarrow 0$, while

$$|||\Phi_1 - \Phi_2|||_1 = s |||\Phi_1 - \Psi_1|||_1 \leq 2s.$$

Remark 4.4.7 (Physical interpretation). By Remark 4.2.6, pairs with the same sum are one device Ω with different which-branch effects. A classical readout that reports the outcome of Φ as that of Ψ with probability p_{01} , and conversely with probability p_{10} , gives the pair

$$((1 - p_{01})\Phi + p_{10}\Psi, p_{01}\Phi + (1 - p_{10})\Psi).$$

The sum is unchanged, and T becomes $p_{10}\mathbf{1}_{\mathbb{C}^r} + (1 - p_{01} - p_{10})T$, a change of norm at most $\max\{p_{01}, p_{10}\}$ because $\text{spec}(T) \subseteq [0, 1]$. So, for every pair, Theorem 4.4.6 bounds the change of $\hat{F}$ by $\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \max\{p_{01}, p_{10}\}^{1/2}$. Such classical noise was found to dominate readout errors on superconducting devices [87]. If $p_{01} + p_{10} \leq 1$, the new spectrum lies in $[p_{10}, 1 - p_{01}]$: imperfect readout gives the two-sided bounds of Sections 4.7 and 4.9.

We now verify the modulus of continuity does indeed behave properly under the right circumstances.

Lemma 4.4.8. *For a compact interval $I \subseteq \mathbb{R}$ and a continuous function $g : I \rightarrow \mathbb{R}$ put*

$$\omega_g^I(\delta) = \sup \{ \|g(E) - g(E')\|_\infty : E, E' \text{ Hermitian with spectra in } I, \\ \|E - E'\|_\infty \leq \delta \}, \quad \delta \geq 0,$$

the supremum being over all finite dimensions, so that $\omega_g = \omega_g^{[0,1]}$ is the modulus of Theorem 4.4.5; for $\epsilon \in (0, 1)$ we also write $\omega_g^{(\epsilon)} = \omega_g^{[0, 1-\epsilon]}$. Then $\omega_g^I(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. In particular this holds for $g = g_\sigma$, where σ is any Kubo–Ando mean, and for $g = h_f$, where f is as in Lemma 4.5.12.

Proof. Let $\eta > 0$ and $R = \max\{|x| : x \in I\}$. By the Weierstrass approximation theorem there is a polynomial $\tilde{g}(t) = \sum_{k=0}^n c_k t^k$ with $\sup_{t \in I} |g(t) - \tilde{g}(t)| \leq \eta$. If E, E' are Hermitian with spectra in I , the spectral theorem gives $\|g(E) - \tilde{g}(E)\|_\infty \leq \eta$ and $\|g(E') - \tilde{g}(E')\|_\infty \leq \eta$. Since $\|E\|_\infty, \|E'\|_\infty \leq R$ and

$$E^k - E'^k = \sum_{j=0}^{k-1} E^j (E - E') E'^{k-1-j},$$

we have $\|E^k - E'^k\|_\infty \leq kR^{k-1}\|E - E'\|_\infty$, and hence $\|\tilde{g}(E) - \tilde{g}(E')\|_\infty \leq L_{\tilde{g}}\|E - E'\|_\infty$ with $L_{\tilde{g}} = \sum_k k|c_k|R^{k-1}$, a constant independent of the dimension. Therefore

$$\|g(E) - g(E')\|_\infty \leq 2\eta + L_{\tilde{g}}\|E - E'\|_\infty,$$

so $\omega_g^I(\delta) \leq 2\eta + L_{\tilde{g}}\delta$ and $\limsup_{\delta \rightarrow 0} \omega_g^I(\delta) \leq 2\eta$. Since $\eta > 0$ was arbitrary, $\omega_g^I(\delta) \rightarrow 0$.

For the two special cases it remains to check continuity on $[0, 1]$. The function g_σ is continuous on $(0, 1)$, and at the endpoints its values were defined as the limits $g_\sigma(0) = f'_\sigma(\infty)$ and $g_\sigma(1) = f_\sigma(0^+)$. The function h_f is continuous on $[0, 1]$ by the first paragraph of the proof of Theorem 4.5.13, up to the sentence “Hence h_f is continuous on $[0, 1]$ ”, which does not use the present lemma. $\square$

Continuity results regarding maximal divergences have also been studied by [95], and more recently in the von Neumann algebra context via lower semicontinuity by Okayasu in [101].

We conclude this section with a lower bound for the maps themselves.

Proposition 4.4.9. *Let $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$ with linearly independent $V_1, \dots, V_r$, let $W = [\text{vec}(V_1) \cdots \text{vec}(V_r)]$ and $G = W^* W$. For Hermitian $T_1, T_2 \in \mathcal{L}(\mathbb{C}^r)$,*

$$\frac{\lambda_{\min}(G)}{\dim \mathcal{X}} \|T_1 - T_2\|_1 \leq \| |\Psi_{T_1} - \Psi_{T_2}| \|_1 \\ \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \|T_1 - T_2\|_\infty.$$

Proof. The upper bound follows from the proof of Lemma 4.4.1 with $\Phi = \Omega$, which uses only that $T_1 - T_2$ is Hermitian, together with $\| |\Omega| \|_1 = \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty$. For the lower bound put $K = T_1 - T_2$, so that $\Psi_{T_1} - \Psi_{T_2} = \Psi_K$ and $J(\Psi_K) = W K W^*$. Since

$$J(\Psi_K) = (\Psi_K \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\text{vec}(\mathbf{1}_{\mathcal{X}}) \text{vec}(\mathbf{1}_{\mathcal{X}})^*)$$

and $\text{vec}(\mathbf{1}_{\mathcal{X}})\text{vec}(\mathbf{1}_{\mathcal{X}})^*/\dim \mathcal{X}$ is a density operator, $\|WKW^*\|_1 \leq \dim \mathcal{X} \|\Psi_K\|_1$. Write $W = UG^{1/2}$ with U an isometry (polar decomposition); then $\|WKW^*\|_1 = \|G^{1/2}KG^{1/2}\|_1$, and since G is invertible, by Hölder's inequality [16, Chapter IV],

$$\begin{aligned} \|K\|_1 &= \|G^{-1/2}(G^{1/2}KG^{1/2})G^{-1/2}\|_1 \\ &\leq \|G^{-1/2}\|_\infty^2 \|G^{1/2}KG^{1/2}\|_1 = \lambda_{\min}(G)^{-1} \|G^{1/2}KG^{1/2}\|_1. \end{aligned}$$

Combining these estimates gives the lower bound. $\square$

We record the estimate used in the proof of Theorem 4.4.6 in the form in which it is used later.

Lemma 4.4.10. *Let Ω , $V_1, \dots, V_r$ and W be as in Proposition 4.4.9. For Hermitian $N_1, N_2 \in L(\mathbb{C}^r)$,*

$$\|\text{Tr}_{\mathcal{Y}}(WN_1W^*) - \text{Tr}_{\mathcal{Y}}(WN_2W^*)\|_\infty \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \|N_1 - N_2\|_\infty,$$

and the same bound holds for

$$|\lambda_{\min}(\text{Tr}_{\mathcal{Y}}(WN_1W^*)) - \lambda_{\min}(\text{Tr}_{\mathcal{Y}}(WN_2W^*))|$$

and for the corresponding difference of largest eigenvalues.

Proof. Put $K = N_1 - N_2$. Since $-\|K\|_\infty \mathbf{1}_{\mathbb{C}^r} \leq K \leq \|K\|_\infty \mathbf{1}_{\mathbb{C}^r}$ and the partial trace is positive,

$$-\|K\|_\infty \text{Tr}_{\mathcal{Y}}(WW^*) \leq \text{Tr}_{\mathcal{Y}}(WKW^*) \leq \|K\|_\infty \text{Tr}_{\mathcal{Y}}(WW^*),$$

and

$$\text{Tr}_{\mathcal{Y}}(WW^*) = \text{Tr}_{\mathcal{Y}} J(\Omega) = \Omega^*(\mathbf{1}_{\mathcal{Y}})^\top \geq 0.$$

For a Hermitian H and positive M with $-sM \leq H \leq sM$ one has $\|H\|_\infty \leq s\|M\|_\infty$, which gives the first claim; the second follows from Weyl's inequality [16, Section III.2]. $\square$

4.5 A general approach

4.5.1 Geometric Rényi relative entropy

The problem presented in Corollary 4.3.3 is not unique to the geometric channel fidelity. In general, any channel divergence which can be expressed through a Choi-level Kubo–Ando mean that vanishes on pairs with trivially intersecting ranges will have this problem as well; Theorem 4.5.6 below characterises these means.

In [72], the following definitions are made which we recall (see also [42]; up to sign, the underlying state quantity $\hat{Q}_\alpha$ is the maximal f -divergence of [91] for the power function $f(t) = t^\alpha$).

Definition 4.5.1 (Geometric Rényi relative entropy). Let $\rho \in D(\mathcal{V})$ be a quantum state and let $\xi \in \text{Pos}(\mathcal{V})$ be a positive semidefinite operator. Let $\alpha \in (0, 1) \cup (1, \infty)$. The geometric Rényi relative quasi-entropy is defined as

$$\hat{Q}_\alpha(\rho\|\xi) = \lim_{\eta \downarrow 0} \text{tr} (\xi_\eta^{1/2} (\xi_\eta^{-1/2} \rho \xi_\eta^{-1/2})^\alpha \xi_\eta^{1/2})$$

where $\xi_\eta = \xi + \eta \mathbf{1}_{\mathcal{V}}$. The geometric Rényi relative entropy is then defined as

$$\hat{D}_\alpha(\rho\|\xi) = \frac{1}{\alpha - 1} \log \hat{Q}_\alpha(\rho\|\xi).$$

We note that in the above definition, we do not even require ρ to be a state for the definition to make sense. We consequently do not require that ρ is a state in the rest of this text.

If $\alpha \in (0, 1)$, then $t \mapsto t^\alpha$ is operator monotone on $(0, \infty)$ and takes the value 1 at $t = 1$; it is the representing function of the weighted geometric mean $\#_\alpha$. In this notation

$$\widehat{Q}_\alpha(\rho \parallel \xi) = \text{tr}(\xi \#_\alpha \rho) = \text{tr}(\rho \#_{1-\alpha} \xi),$$

with the regularisation from above built into the definition of the mean. This will prove important shortly. The next definition we recall from [72] is the following; again we drop the requirement that the maps be channels and replace it with complete positivity.

Definition 4.5.2 (Geometric Rényi relative entropy of completely positive maps). Let $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ be completely positive. The geometric Rényi relative entropy is defined for $\alpha \in (0, 1) \cup (1, \infty)$ as

$$\widehat{D}_\alpha(\Phi \parallel \Psi) = \sup_{\rho \in \text{D}(\mathcal{X} \otimes \mathcal{X})} \widehat{D}_\alpha((\Phi \otimes \mathbf{1}_{\text{L}(\mathcal{X})})(\rho) \parallel (\Psi \otimes \mathbf{1}_{\text{L}(\mathcal{X})})(\rho)).$$

For $\alpha \in (0, 1)$ this quantity admits a closed form in terms of the Choi matrices [72], and for $\alpha \in (1, 2]$ one is given in [42, Lemma 5]. We state it for $\alpha \in (0, 1)$, where, since $\frac{1}{\alpha-1} < 0$, the supremum of $\widehat{D}_\alpha$ corresponds to an infimum of $\widehat{Q}_\alpha$.

Lemma 4.5.3 ([72, Proposition 44]). *Let $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ and $\alpha \in (0, 1)$, and define*

$$\widehat{Q}_\alpha(\Phi \parallel \Psi) = \inf_{\rho \in \text{D}(\mathcal{X} \otimes \mathcal{X})} \widehat{Q}_\alpha((\Phi \otimes \mathbf{1}_{\text{L}(\mathcal{X})})(\rho) \parallel (\Psi \otimes \mathbf{1}_{\text{L}(\mathcal{X})})(\rho)).$$

Then

$$\widehat{D}_\alpha(\Phi \parallel \Psi) = \frac{1}{\alpha-1} \log \widehat{Q}_\alpha(\Phi \parallel \Psi), \quad \widehat{Q}_\alpha(\Phi \parallel \Psi) = \lambda_{\min}(\text{Tr}_{\mathcal{Y}}(J(\Psi) \#_\alpha J(\Phi))),$$

where the weighted geometric mean is understood with the regularisation from above.

Proof. The first identity follows from the definitions, since $t \mapsto \frac{1}{\alpha-1} \log t$ is decreasing for $\alpha \in (0, 1)$. For the second, the proof of Proposition 4.3.1 applies verbatim with $\#$ replaced by $\#_\alpha$: joint concavity of $(\rho, \xi) \mapsto \text{tr}(\xi \#_\alpha \rho)$ [79] reduces the infimum to pure states, the transformer equality [79] handles invertible A , and upper semicontinuity handles the closure; compare [42, Lemma 5] for the case $\alpha \in (1, 2]$. $\square$

The careful reader will notice that $\widehat{D}_\alpha(\Phi \parallel \Psi)$ may take the value $+\infty$ for $\alpha \in (0, 1)$. This is not a problem; it corresponds to $\widehat{Q}_\alpha(\Phi \parallel \Psi) = 0$, which occurs for instance when the Choi matrices have ranges with trivial intersection.

It is not much of a stretch to see that the problem of Proposition 4.3.2 extends to $\widehat{Q}_\alpha$, because this quantity is induced by a Kubo–Ando mean of Choi matrices; indeed the same replacement maps work, since $A \#_\alpha B = 0$ whenever the ranges of A and B intersect trivially (Theorem 4.5.6 below). We will again demonstrate that we can control such problems through the ARN derivative. Since $J(\Psi) \#_\alpha J(\Phi) = J(\Phi) \#_{1-\alpha} J(\Psi)$, Theorem 4.2.15 applied to the mean $\#_{1-\alpha}$ shows that $J(\Psi) \#_\alpha J(\Phi) = Wh_\alpha(T)W^*$, where T is the ARN derivative of Φ with respect to $\Omega = \Phi + \Psi$ and

$$h_\alpha(t) = g_{\#_{1-\alpha}}(t) = t \#_{1-\alpha} (1-t) = t^\alpha (1-t)^{1-\alpha}, \quad t \in [0, 1]. \quad (4.5.1)$$

Lemma 4.5.4. *Let $\mathcal{Z}$ be a complex Euclidean space, let $0 \leq E, E' \leq \mathbf{1}_{\mathcal{Z}}$ and let $\alpha \in (0, 1)$. Then*

$$\|h_\alpha(E) - h_\alpha(E')\|_\infty \leq \|E - E'\|_\infty^\alpha + \|E - E'\|_\infty^{1-\alpha}.$$

Proof. Since E and $\mathbf{1}_Z - E$ commute, we have

$$h_\alpha(E) = E^\alpha(\mathbf{1}_Z - E)^{1-\alpha} \geq 0,$$

and therefore

$$h_\alpha(E) - h_\alpha(E') = (E^\alpha - E'^\alpha)(\mathbf{1}_Z - E)^{1-\alpha} + E'^\alpha((\mathbf{1}_Z - E)^{1-\alpha} - (\mathbf{1}_Z - E')^{1-\alpha}).$$

By the triangle inequality and submultiplicativity of $\|\cdot\|_\infty$,

$$\begin{aligned} \|h_\alpha(E) - h_\alpha(E')\|_\infty &\leq \|E^\alpha - E'^\alpha\|_\infty \|(\mathbf{1}_Z - E)^{1-\alpha}\|_\infty \\ &\quad + \|E'^\alpha\|_\infty \|(\mathbf{1}_Z - E)^{1-\alpha} - (\mathbf{1}_Z - E')^{1-\alpha}\|_\infty. \end{aligned}$$

Since $\|(\mathbf{1}_Z - E)^{1-\alpha}\|_\infty \leq 1$ and $\|E'^\alpha\|_\infty \leq 1$,

$$\|h_\alpha(E) - h_\alpha(E')\|_\infty \leq \|E^\alpha - E'^\alpha\|_\infty + \|(\mathbf{1}_Z - E)^{1-\alpha} - (\mathbf{1}_Z - E')^{1-\alpha}\|_\infty.$$

Now apply the operator-norm Hölder continuity of positive power functions [7], [16, Theorem X.1.1]: for $A, B \geq 0$ and $\beta \in (0, 1]$,

$$\|A^\beta - B^\beta\|_\infty \leq \|A - B\|_\infty^\beta.$$

Taking $\beta = \alpha$, $A = E$, $B = E'$ gives $\|E^\alpha - E'^\alpha\|_\infty \leq \|E - E'\|_\infty^\alpha$, and taking $\beta = 1 - \alpha$, $A = \mathbf{1}_Z - E$, $B = \mathbf{1}_Z - E'$ gives

$$\|(\mathbf{1}_Z - E)^{1-\alpha} - (\mathbf{1}_Z - E')^{1-\alpha}\|_\infty \leq \|(\mathbf{1}_Z - E) - (\mathbf{1}_Z - E')\|_\infty^{1-\alpha} = \|E - E'\|_\infty^{1-\alpha}.$$

Combining the two bounds,

$$\|h_\alpha(E) - h_\alpha(E')\|_\infty \leq \|E - E'\|_\infty^\alpha + \|E - E'\|_\infty^{1-\alpha},$$

as claimed. $\square$

Theorem 4.5.5. *Let $\alpha \in (0, 1)$ and let $\Phi_1, \Psi_1, \Phi_2, \Psi_2 \in \text{CP}(\mathcal{X}, \mathcal{Y})$ satisfy $\Omega = \Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$. Let T_1 and T_2 denote the ARN derivatives of Φ_1 and Φ_2 with respect to Ω , relative to one fixed Kraus decomposition of Ω by linearly independent operators, the same for both pairs. Then*

$$|\widehat{Q}_\alpha(\Phi_1 \| \Psi_1) - \widehat{Q}_\alpha(\Phi_2 \| \Psi_2)| \leq \|\Omega^*(\mathbf{1}_Y)\|_\infty (\|T_1 - T_2\|_\infty^\alpha + \|T_1 - T_2\|_\infty^{1-\alpha}).$$

Proof. By Lemma 4.5.3, $\widehat{Q}_\alpha(\Phi_i \| \Psi_i) = \lambda_{\min}(\text{Tr}_Y(W h_\alpha(T_i) W^*))$. The proof of Theorem 4.4.6 now applies verbatim with $g_\#$ replaced by h_α and Proposition 4.4.3(2) replaced by Lemma 4.5.4. $\square$

4.5.2 The meta problem

In this subsection we describe in general terms when the problem of Proposition 4.3.2 occurs. Let σ be a Kubo–Ando mean with representing function f_σ . By the integral representation of operator connections [62, Theorem 5.11], for all positive semidefinite A, B ,

$$A \sigma B = a_\sigma A + b_\sigma B + \int_{(0, \infty)} \frac{1 + \lambda}{\lambda} ((\lambda A) : B) dm_\sigma(\lambda),$$

where m_σ is a finite positive measure, $a_\sigma = f_\sigma(0^+) = 1 \sigma 0$ and $b_\sigma = f'_\sigma(\infty) = 0 \sigma 1$. Evaluating the representing function at 1 gives

$$a_\sigma + b_\sigma + m_\sigma((0, \infty)) = f_\sigma(1) = 1.$$

We also use that $\text{range}(A : B) = \text{range}(A) \cap \text{range}(B)$ for positive semidefinite A, B [3].

Theorem 4.5.6. *Let σ be a Kubo–Ando mean with representing function f_σ .*

1. *If $f_\sigma(0^+) = f'_\sigma(\infty) = 0$, then $\text{range}(A \sigma B) = \text{range}(A) \cap \text{range}(B)$ for all positive semidefinite A, B . In particular, $A \sigma B = 0$ if and only if $\text{range}(A) \cap \text{range}(B) = \{0\}$.*
2. *Conversely, if $A \sigma B = 0$ for all positive semidefinite A, B on a space of dimension at least 2 with $\text{range}(A) \cap \text{range}(B) = \{0\}$, then $f_\sigma(0^+) = f'_\sigma(\infty) = 0$.*

Proof. (1) Here $a_\sigma = b_\sigma = 0$ and $m_\sigma((0, \infty)) = 1$. Since $\text{range}(\lambda A) = \text{range}(A)$ for $\lambda > 0$, each of the positive semidefinite operators $(\lambda A) : B$ has range $R = \text{range}(A) \cap \text{range}(B)$, and hence kernel $R^\perp$. If $v \in R^\perp$ then $\langle v, (A \sigma B)v \rangle = 0$. If $v \notin R^\perp$ then $\langle v, ((\lambda A) : B)v \rangle > 0$ for every $\lambda > 0$, and since $m_\sigma \neq 0$ the integral is strictly positive. Thus $\ker(A \sigma B) = R^\perp$, that is, $\text{range}(A \sigma B) = R$.

(2) Let e_1, e_2 be orthonormal and $A = e_1 e_1^*, B = e_2 e_2^*$. Then $\text{range}(A) \cap \text{range}(B) = \{0\}$, so every $(\lambda A) : B$ vanishes and $A \sigma B = a_\sigma A + b_\sigma B$. If this is zero, then $a_\sigma = b_\sigma = 0$. $\square$

Consequently, suppose that a real-valued function D on $\text{CP}(\mathcal{X}, \mathcal{Y}) \times \text{CP}(\mathcal{X}, \mathcal{Y})$ has the form $D(\Phi, \Psi) = H(J(\Phi) \sigma J(\Psi))$ for a function H with $H(0) = 0$, where σ satisfies the hypothesis of Theorem 4.5.6(1). Then $D(\Phi, \Psi) = 0$ whenever $\text{range}(J(\Phi)) \cap \text{range}(J(\Psi)) = \{0\}$. By Proposition 4.3.1 this applies to $\hat{F}$, with $H(X) = \lambda_{\min}(\text{Tr}_{\mathcal{Y}} X)$, and by Lemma 4.5.3 it applies to $\hat{Q}_\alpha$ with the roles of Φ and Ψ exchanged. The converse implication fails in general, because $\lambda_{\min}(\text{Tr}_{\mathcal{Y}} X)$ can vanish for nonzero $X \geq 0$. For instance, if $\Phi = \Psi$ and $\Phi^*(\mathbf{1}_{\mathcal{Y}})$ is singular, then $\hat{F}(\Phi, \Phi) = \lambda_{\min}(\Phi^*(\mathbf{1}_{\mathcal{Y}})^\top) = 0$, although the Choi ranges coincide.

We can consider this in relation to the ARN derivative. Let $\Omega = \Phi + \Psi$ and let T be the ARN derivative of Φ with respect to Ω . The condition $f_\sigma(0^+) = f'_\sigma(\infty) = 0$ says exactly that $g_\sigma(0) = g_\sigma(1) = 0$, while $g_\sigma(t) = t f_\sigma(\frac{1-t}{t}) > 0$ for $0 < t < 1$. Since W is injective, Theorem 4.2.15 shows that $J(\Phi) \sigma J(\Psi) = W g_\sigma(T) W^*$ vanishes if and only if $\text{spec}(T) \subseteq \{0, 1\}$, that is, if and only if T is a projection. For the geometric mean this vanishing criterion was obtained by Kian [76], in the form that $\Phi \# \Psi = 0$ exactly when Φ and Ψ have no nonzero common completely positive submap; submaps whose derivative is a projection are the sharp submaps studied in [75]. The replacement maps of Proposition 4.3.2 are of exactly this kind: in the Kraus decomposition described after Theorem 4.4.6, the ARN derivative of Φ_1 is $\mathbf{1}_{\mathbb{C}^n} \oplus 0_{\mathbb{C}^n}$.

4.5.3 Operator convex functions and divergences

In this subsection, we show that the Kubo–Ando framework from above extends further to other channel level divergences. Many channel divergences are defined through operator convex functions instead of Kubo–Ando means.

Definition 4.5.7 (Operator convex function [16, Chapter V]). Let (t_-, t_+) be an open interval. A function $f : (t_-, t_+) \rightarrow \mathbb{R}$ is said to be operator convex on (t_-, t_+) if

$$f(tA + (1-t)B) \leq t f(A) + (1-t) f(B)$$

for all $t \in [0, 1]$ and all Hermitian operators A, B of the same (arbitrary) size with spectra contained in (t_-, t_+) . It is operator concave if $-f$ is operator convex.

Multiple quantum channel divergences can be framed in a similar way to those defined by Kubo–Ando means. A broad class of these channel divergences are the maximal f -divergences, introduced by Matsumoto [91] and studied systematically in [60, 61].

Definition 4.5.8 (Maximal f -divergence [61, 91]). Let $f : (0, \infty) \rightarrow \mathbb{R}$ be operator convex; the limits $f(0^+)$ and $f'(\infty)$ (notation as before Theorem 4.2.15) exist in $(-\infty, +\infty]$ (Lemma 4.7.2). For positive definite $A, B \in \mathcal{L}(\mathcal{V})$ the *maximal f -divergence* is

$$\widehat{D}_f(A\|B) = \operatorname{tr}\left(B^{1/2}f(B^{-1/2}AB^{-1/2})B^{1/2}\right),$$

and for positive semidefinite A, B it is defined by

$$\widehat{D}_f(A\|B) = \lim_{\eta \downarrow 0} \widehat{D}_f(A + \eta \mathbf{1}_{\mathcal{V}}\|B + \eta \mathbf{1}_{\mathcal{V}}).$$

We call f *bounded* if $f(0^+) < \infty$ and $f'(\infty) < \infty$.

Definition 4.5.9 (Maximal f -divergence of completely positive maps; cf. [42, 72]). For $\Phi, \Psi \in \mathcal{CP}(\mathcal{X}, \mathcal{Y})$,

$$\widehat{D}_f(\Phi\|\Psi) = \sup_{\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})} \widehat{D}_f(\Phi_\rho\|\Psi_\rho),$$

where $\Phi_\rho = (\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho)$ and $\Psi_\rho = (\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho)$.

Definition 4.5.10 (Examples). For positive semidefinite A, B and $0 < \alpha < 1$:

1. *Hellinger divergence* [20, 108]: $f_{\text{H}}(t) = (1 - \sqrt{t})^2$, and

$$\widehat{D}_{\text{H}}(A\|B) = \operatorname{tr} A + \operatorname{tr} B - 2 \operatorname{tr}(A \# B).$$

2. *Hellinger divergence of order α* [82]: $f_{\text{H},\alpha}(t) = \frac{1 - t^\alpha}{1 - \alpha}$, and

$$\widehat{D}_{\text{H},\alpha}(A\|B) = \frac{\operatorname{tr} B - \operatorname{tr}(B \#_\alpha A)}{1 - \alpha}.$$

3. *Vincze–Le Cam (triangular) divergence* [82]: $f_{\text{LC}}(t) = \frac{(t-1)^2}{t+1}$, and

$$\widehat{D}_{\text{LC}}(A\|B) = \operatorname{tr} A + \operatorname{tr} B - 4 \operatorname{tr}(A : B).$$

4. *Maximal Jensen–Shannon divergence* [83]: $f_{\text{JS}}(t) = t \log t - (1+t) \log \frac{1+t}{2}$, and

$$\widehat{D}_{\text{JS}}(A\|B) = \operatorname{tr}\left(B^{1/2}f_{\text{JS}}(B^{-1/2}AB^{-1/2})B^{1/2}\right).$$

Each f above is bounded, with $(f(0^+), f'(\infty))$ equal to $(1, 1)$, $(\frac{1}{1-\alpha}, 0)$, $(1, 1)$ and $(\log 2, \log 2)$ respectively.

Remark 4.5.11. If $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}_{\mathbb{C}^r} - T}$, where T is the ARN derivative of Φ with respect to $\Omega = \Phi + \Psi$, then the maximal f -divergence of the pair is governed by $h_f(T)$, where $h_f(t) = (1-t)f(\frac{t}{1-t})$ on $(0, 1)$, extended by $h_f(0) = f(0^+)$ and $h_f(1) = f'(\infty)$. For the examples above,

$$h_{\text{H}}(T) = \mathbf{1}_{\mathbb{C}^r} - 2\sqrt{T(\mathbf{1}_{\mathbb{C}^r} - T)}, \quad h_{\text{H},\alpha}(T) = \frac{(\mathbf{1}_{\mathbb{C}^r} - T) - T^\alpha(\mathbf{1}_{\mathbb{C}^r} - T)^{1-\alpha}}{1 - \alpha},$$

$$h_{\text{LC}}(T) = (2T - \mathbf{1}_{\mathbb{C}^r})^2, \quad h_{\text{JS}}(T) = (\log 2) \mathbf{1}_{\mathbb{C}^r} + T \log T + (\mathbf{1}_{\mathbb{C}^r} - T) \log(\mathbf{1}_{\mathbb{C}^r} - T).$$

Under certain circumstances, the problem of Proposition 4.3.2 reappears. We first decompose a bounded maximal f -divergence into an affine part and a Kubo–Ando connection. Divergences of this form, built from Kubo–Ando means, appear in [108]; the closed-form expression in Theorem 4.5.13 extends those of [42, 72], and an operator-valued analogue is the divergence of [101, Definition 3.1].

Lemma 4.5.12. *Let $f : (0, \infty) \rightarrow \mathbb{R}$ be operator convex with $f(0^+)$ and $f'(\infty)$ finite. Set*

$$h(t) = f(0^+) + f'(\infty)t - f(t), \quad t > 0.$$

Then h is a nonnegative operator monotone function on $(0, \infty)$ with $h(0^+) = 0$ and $h(t)/t \rightarrow 0$ as $t \rightarrow \infty$. Writing σ_h for the Kubo–Ando connection with representing function h , we have, for all positive semidefinite $A, B \in \mathcal{L}(\mathcal{V})$,

$$\widehat{D}_f(A\|B) = f(0^+) \operatorname{tr} B + f'(\infty) \operatorname{tr} A - \operatorname{tr}(B \sigma_h A).$$

Proof. Extend f to $[0, \infty)$ by $f(0) = f(0^+)$; the extension is convex, so $t \mapsto (f(t) - f(0))/t$ is nondecreasing on $(0, \infty)$ with limit $f'(\infty)$. Hence $f(t) \leq f(0^+) + f'(\infty)t$, that is, $h \geq 0$. As an affine function minus an operator convex function, h is operator concave, and extended by $h(0) = 0$ it is continuous, nonnegative and operator concave on $[0, \infty)$; it is therefore operator monotone [16, Theorem V.2.5]. The two limits follow from the definitions of $f(0^+)$ and $f'(\infty)$. For positive definite A, B , applying the functional calculus to $f = f(0^+) + f'(\infty)t - h$ and using $B^{1/2}(B^{-1/2}AB^{-1/2})B^{1/2} = A$ and $B \sigma_h A = B^{1/2}h(B^{-1/2}AB^{-1/2})B^{1/2}$ gives

$$B^{1/2}f(B^{-1/2}AB^{-1/2})B^{1/2} = f(0^+)B + f'(\infty)A - B \sigma_h A.$$

For positive semidefinite A, B , replace A, B by $A + \eta \mathbf{1}_{\mathcal{V}}$, $B + \eta \mathbf{1}_{\mathcal{V}}$, take traces and let $\eta \downarrow 0$: the left-hand side converges to $\widehat{D}_f(A\|B)$ by definition, and $(B + \eta \mathbf{1}_{\mathcal{V}}) \sigma_h (A + \eta \mathbf{1}_{\mathcal{V}})$ decreases to $B \sigma_h A$ [79]. $\square$

Theorem 4.5.13. *Let f be as in Lemma 4.5.12. Let $\Omega \in \mathcal{CP}(\mathcal{X}, \mathcal{Y})$ have a Kraus decomposition $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$ with linearly independent $V_1, \dots, V_r$, and let $W = [\operatorname{vec}(V_1) \ \cdots \ \operatorname{vec}(V_r)]$. Define $h_f : [0, 1] \rightarrow \mathbb{R}$ by*

$$h_f(t) = (1-t)f\left(\frac{t}{1-t}\right) \quad (0 < t < 1), \quad h_f(0) = f(0^+), \quad h_f(1) = f'(\infty).$$

(a) *If $0 \leq T \leq \mathbf{1}_{\mathcal{C}^r}$, $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}_{\mathcal{C}^r} - T}$, then*

$$\widehat{D}_f(\Phi\|\Psi) = \lambda_{\max}\left(\operatorname{Tr}_{\mathcal{Y}}(W h_f(T) W^*)\right).$$

(b) *If $0 \leq T_1, T_2 \leq \mathbf{1}_{\mathcal{C}^r}$, then*

$$\left| \widehat{D}_f(\Psi_{T_1}\|\Psi_{\mathbf{1}_{\mathcal{C}^r} - T_1}) - \widehat{D}_f(\Psi_{T_2}\|\Psi_{\mathbf{1}_{\mathcal{C}^r} - T_2}) \right| \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty} \omega_{h_f}(\|T_1 - T_2\|_{\infty}),$$

where ω_{h_f} is as in Theorem 4.4.5, and $\omega_{h_f}(\delta) \rightarrow 0$ as $\delta \rightarrow 0$.

Proof. The argument follows the strategy of [42, Lemma 5]. With h as in Lemma 4.5.12, put $k(t) = (1-t)h(\frac{t}{1-t})$ for $0 < t < 1$ and $k(0) = k(1) = 0$. Substituting $f = f(0^+) + f'(\infty)t - h$ into the definition of h_f gives

$$h_f(t) = f(0^+)(1-t) + f'(\infty)t - k(t), \quad 0 \leq t \leq 1.$$

The function k is continuous on $[0, 1]$: as $t \downarrow 0$ we have $k(t) \rightarrow h(0^+) = 0$, and as $t \uparrow 1$, writing $x = t/(1-t) \rightarrow \infty$, we have $k(t) = t h(x)/x \rightarrow 0$. Hence h_f is continuous on $[0, 1]$, and $\omega_{h_f}(\delta) \rightarrow 0$ by Lemma 4.4.8.

The ARN derivatives of Ψ and Φ with respect to Ω are $\mathbf{1}_{\mathcal{C}^r} - T$ and T , and the scalar connection satisfies $s \sigma_h(1-s) = s h(\frac{1-s}{s})$, which equals $k(t)$ at $s = 1-t$. Theorem 4.2.15, applied to the pair (Ψ, Φ) , therefore gives $J(\Psi) \sigma_h J(\Phi) = W k(T) W^*$. (If $h \not\equiv 0$ then $h(1) > 0$ and $\sigma_h = h(1) \sigma_{h/h(1)}$ with $\sigma_{h/h(1)}$ a mean; if $h \equiv 0$ there is nothing to prove.) Since $J(\Phi) = W T W^*$ and $J(\Psi) = W(\mathbf{1}_{\mathcal{C}^r} - T) W^*$, we then have

$$Z_f = f(0^+) J(\Psi) + f'(\infty) J(\Phi) - J(\Psi) \sigma_h J(\Phi) = W h_f(T) W^*.$$

Let $M = \text{Tr}_{\mathcal{Y}}(W h_f(T) W^*)$, a Hermitian operator on $\mathcal{X}$. Let $u = \text{vec}(A)$ with $\|A\|_2 = 1$ and A invertible, and let $C = \mathbf{1}_{\mathcal{Y}} \otimes A^\top$, which is invertible. As in the proof of Proposition 4.3.1, $(\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(uu^*) = C J(\Phi) C^*$ and similarly for Ψ . By the transformer equality for invertible operators [79],

$$(C J(\Psi) C^*) \sigma_h (C J(\Phi) C^*) = C (J(\Psi) \sigma_h J(\Phi)) C^*,$$

so Lemma 4.5.12 gives

$$\begin{aligned} \widehat{D}_f((\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(uu^*) \| (\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(uu^*)) &= \text{tr}(C Z_f C^*) = \text{tr}((\mathbf{1}_{\mathcal{Y}} \otimes (AA^*)^\top) Z_f) \\ &= \text{tr}((AA^*)^\top M), \end{aligned}$$

using $C^* C = \mathbf{1}_{\mathcal{Y}} \otimes (AA^*)^\top$ and the previous part. Let

$$d(\rho) = \widehat{D}_f((\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho) \| (\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho))$$

for $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})$. By Lemma 4.5.12, d is a linear function minus

$$\rho \mapsto \text{tr}((\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho) \sigma_h (\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho)),$$

which is concave because σ_h is jointly concave [79] and $\rho \mapsto ((\Phi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho), (\Psi \otimes \mathbf{1}_{\mathcal{L}(\mathcal{X})})(\rho))$ is linear. Thus d is convex, and writing $\rho = \sum_i p_i u_i u_i^*$ in its spectral decomposition gives $d(\rho) \leq \max_i d(u_i u_i^*)$; so the supremum defining $\widehat{D}_f(\Phi \| \Psi)$ may be taken over pure states. Moreover d is lower semicontinuous: for $\eta > 0$ the map

$$(A, B) \mapsto \text{tr}((B + \eta \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}) \sigma_h (A + \eta \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}))$$

is continuous and decreases to $\text{tr}(B \sigma_h A)$ as $\eta \downarrow 0$, so $(A, B) \mapsto \text{tr}(B \sigma_h A)$ is upper semicontinuous. Every unit vector $u = \text{vec}(A)$ is a limit of unit vectors $u_n = \text{vec}(A_n)$ with A_n invertible, and then $d(uu^*) \leq \liminf_n d(u_n u_n^*)$. Hence the supremum may be taken over $u = \text{vec}(A)$ with A invertible. As A ranges over such operators, $(AA^*)^\top$ ranges over the invertible density operators on $\mathcal{X}$, which are dense in $\mathcal{D}(\mathcal{X})$. By the previous parts,

$$\widehat{D}_f(\Phi \| \Psi) = \sup \{ \text{tr}(\xi M) : \xi \in \mathcal{D}(\mathcal{X}) \text{ invertible} \} = \max_{\xi \in \mathcal{D}(\mathcal{X})} \text{tr}(\xi M) = \lambda_{\max}(M),$$

which is (a).

Let $M_i = \text{Tr}_{\mathcal{Y}}(W h_f(T_i) W^*)$ and $K = h_f(T_1) - h_f(T_2)$. By (a) and Weyl's inequality [16, Section III.2], the left-hand side of (b) is

$$|\lambda_{\max}(M_1) - \lambda_{\max}(M_2)| \leq \|M_1 - M_2\|_\infty.$$

Since $-\|K\|_\infty \mathbf{1}_{\mathbb{C}^r} \leq K \leq \|K\|_\infty \mathbf{1}_{\mathbb{C}^r}$, conjugating by W and applying the positive map Tr_Y gives

$$-\|K\|_\infty \text{Tr}_Y(WW^*) \leq M_1 - M_2 \leq \|K\|_\infty \text{Tr}_Y(WW^*),$$

and

$$\text{Tr}_Y(WW^*) = \text{Tr}_Y J(\Omega) = \Omega^*(\mathbf{1}_Y)^\top \geq 0.$$

Hence

$$\|M_1 - M_2\|_\infty \leq \|\Omega^*(\mathbf{1}_Y)\|_\infty \|K\|_\infty \leq \|\Omega^*(\mathbf{1}_Y)\|_\infty \omega_{h_f}(\|T_1 - T_2\|_\infty),$$

which is (b). $\square$

Theorem 4.5.13 also explains how the problem of Proposition 4.3.2 arises for maximal f -divergences. By Lemma 4.5.12, the connection σ_h inside $\hat{D}_f$ has representing function h with $h(0^+) = 0$ and $h'(\infty) = 0$, so, by Theorem 4.5.6, it vanishes on Choi matrices whose ranges intersect trivially.

Corollary 4.5.14. *Let f be as in Lemma 4.5.12 and let $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$. If $\text{range}(J(\Phi)) \cap \text{range}(J(\Psi)) = \{0\}$, then (cf. [101])*

$$\hat{D}_f(\Phi\|\Psi) = \lambda_{\max}(f(0^+) \Psi^*(\mathbf{1}_Y) + f'(\infty) \Phi^*(\mathbf{1}_Y)).$$

In particular, if $f(1) = 0$, then for the maps of Corollary 4.3.3 we have $\hat{D}_f(\Phi_1\|\Psi_1) = f(0^+) + f'(\infty)$ and $\hat{D}_f(\Phi_2\|\Psi_2) = 0$, although $\|\Phi_1 - \Phi_2\|_1 = \|\Psi_1 - \Psi_2\|_1 < \eta$.

Proof. Apply Theorem 4.5.13 with $\Omega = \Phi + \Psi$. Its proof shows that $\hat{D}_f(\Phi\|\Psi) = \lambda_{\max}(\text{Tr}_Y Z_f)$, where

$$Z_f = f(0^+)J(\Psi) + f'(\infty)J(\Phi) - J(\Psi)\sigma_h J(\Phi).$$

If $h \not\equiv 0$, then $\sigma_h = h(1)\sigma_{h/h(1)}$ and $\sigma_{h/h(1)}$ is a mean satisfying the hypothesis of Theorem 4.5.6(1); if $h \equiv 0$ then $\sigma_h = 0$. In either case $J(\Psi)\sigma_h J(\Phi) = 0$, so

$$\text{Tr}_Y Z_f = (f(0^+)\Psi^*(\mathbf{1}_Y) + f'(\infty)\Phi^*(\mathbf{1}_Y))^\top,$$

using $\text{Tr}_Y J(\Xi) = \Xi^*(\mathbf{1}_Y)^\top$. Transposition does not change eigenvalues, which proves the formula.

For the maps of Corollary 4.3.3, the Choi ranges of Φ_1 and Ψ_1 intersect trivially by the proof of Proposition 4.3.2, and $\Phi_1^*(\mathbf{1}_Y) = \Psi_1^*(\mathbf{1}_Y) = \mathbf{1}_X$, so $\hat{D}_f(\Phi_1\|\Psi_1) = f(0^+) + f'(\infty)$. For $\Phi_2 = \Psi_2$ we have $J(\Phi_2)\sigma_h J(\Phi_2) = h(1)J(\Phi_2)$, hence

$$Z_f = (f(0^+) + f'(\infty) - h(1))J(\Phi_2) = f(1)J(\Phi_2) = 0$$

and $\hat{D}_f(\Phi_2\|\Psi_2) = 0$. $\square$

Thus every bounded maximal f -divergence with $f(1) = 0$ and f not affine exhibits the phenomenon of Corollary 4.3.3: the jump $f(0^+) + f'(\infty)$ equals $h(1)$, which vanishes only if $h \equiv 0$, since a nonnegative concave function on $[0, \infty)$ vanishing at 0 and 1 vanishes identically, that is, only if f is affine. Theorem 4.5.13(b) shows that the divergence is controlled by the ARN derivative.

Remark 4.5.15. The normalisation $f(1) = 0$ is used only in the last statement of Corollary 4.5.14, where it is assumed. Lemma 4.5.12, Theorem 4.5.13 and the first statement of Corollary 4.5.14 hold for every bounded operator convex f , that is, every operator convex f with $f(0^+)$ and $f'(\infty)$ finite.

4.6 Unbounded maximal f -divergences

Theorem 4.5.13 requires both $f(0^+)$ and $f'(\infty)$ to be finite. Several of the most widely used maximal f -divergences have $f'(\infty) = +\infty$: the Belavkin–Staszewski relative entropy, with $f(t) = t \log t$ [14]; the geometric Rényi quasi-entropies with $\alpha \in (1, 2]$, with $f(t) = t^\alpha$ [42, 72]; and the maximal χ^2 -divergence, with $f(t) = (t - 1)^2$. For these the function h_f of Theorem 4.5.13 is continuous on $[0, 1)$ but unbounded near $t = 1$, which is where the ARN derivative $\mathbf{1}_{\mathbb{C}^r} - T$ of Ψ degenerates. In this section we show that the closed form and the continuity estimates of Theorem 4.5.13 persist under a bound on the max-relative entropy, which keeps the spectrum of T away from 1, and we compute explicit moduli of continuity for the first two examples. We first make a note of why the max-relative entropy is the natural object to discuss. Let Φ and Ψ be completely positive maps and suppose that $\Psi \leq \Phi$. Let W be the column matrix of a Kraus decomposition of Φ by linearly independent operators, as in Theorem 4.2.2, and let T be the ARN derivative of Ψ with respect to Φ . Then the top eigenvalue t_1 of T is the quantity

$$\min\{\lambda \geq 0 : \Psi \leq \lambda \Phi\}.$$

This is because by Choi’s theorem, we have

$$\Psi \leq \lambda \Phi \iff J(\Psi) \leq \lambda J(\Phi).$$

By transferring to the positive semidefinite picture by the ARN correspondence, this means that

$$\Psi \leq \lambda \Phi \iff W(\lambda \mathbf{1}_{\mathbb{C}^r} - T)W^* \geq 0.$$

But this is only true if and only if $\lambda \mathbf{1}_{\mathbb{C}^r} - T \geq 0$. Since T is positive semidefinite, we then have $\lambda \mathbf{1}_{\mathbb{C}^r} - T \geq 0$ if and only if $\lambda \geq t$ for every eigenvalue t of T , that is, if and only if $\lambda \geq t_1$. The max-relative entropy $D_{\max}(\Psi \| \Phi)$ is the logarithm of this value. So this is a natural object with which to control things.

Throughout this section $f : (0, \infty) \rightarrow \mathbb{R}$ is operator convex with $f(0^+) < \infty$. We do not assume that $f(1) = 0$ or that $f'(\infty) < \infty$, and we extend f continuously to $[0, \infty)$ by $f(0) = f(0^+)$. The maximal f -divergences $\hat{D}_f(A \| B)$ and $\hat{D}_f(\Phi \| \Psi)$ are defined by the formulas of Definitions 4.5.8 and 4.5.9; Lemma 4.6.4 below shows that the limit in Definition 4.5.8 exists in $(-\infty, +\infty]$. For $A \geq 0$ and positive definite B we write

$$\mathcal{P}_f(A, B) = B^{1/2} f(B^{-1/2} A B^{-1/2}) B^{1/2}$$

for the operator perspective of f , so that $\hat{D}_f(A \| B) = \text{tr } \mathcal{P}_f(A, B)$ in that case.

4.6.1 The operator perspective on supports

In this subsection, we show similar results to the Kubo–Ando case for operator convex f . This is necessary for the unbounded case, as we cannot exploit the rewriting of unbounded maximal f -divergences in the same way as the bounded case through the mean. We present a few well-known facts. The first is the invariance under congruences of operator perspective functions (the analogous objects to the connections induced by operator monotone functions).

Lemma 4.6.1. *Let $A, B, R \in \mathcal{L}(\mathcal{V})$ with $A \geq 0$, $B > 0$ and R invertible. Then*

$$\mathcal{P}_f(RAR^*, RBR^*) = R \mathcal{P}_f(A, B) R^*.$$

Proof. Let $Y = RB^{1/2}$, which is invertible, and let $Y = (YY^*)^{1/2}U$ be its polar decomposition, with U unitary. Since $YY^* = RBR^*$ and $RAR^* = YXY^*$ with $X = B^{-1/2}AB^{-1/2}$, we have

$$(RBR^*)^{-1/2}RAR^*(RBR^*)^{-1/2} = UXU^*,$$

and hence

$$f((RBR^*)^{-1/2}RAR^*(RBR^*)^{-1/2}) = Uf(X)U^*.$$

Therefore

$$\begin{aligned} \mathcal{P}_f(RAR^*, RBR^*) &= (YY^*)^{1/2}Uf(X)U^*(YY^*)^{1/2} \\ &= Yf(X)Y^* = R\mathcal{P}_f(A, B)R^*. \end{aligned}$$

□

Definition 4.6.2. Let $A, B \in \text{Pos}(\mathcal{V})$ with $\text{range}(A) \subseteq \text{range}(B)$ and $B \neq 0$. Let $m = \dim \text{range}(B)$, let $U \in \text{L}(\mathbb{C}^m, \mathcal{V})$ be an isometry with $\text{range}(U) = \text{range}(B)$, and put $A_0 = U^*AU$ and $B_0 = U^*BU$, so that $B_0 > 0$, $A = UA_0U^*$ and $B = UB_0U^*$. We set

$$\mathcal{P}_f(A, B) = U\mathcal{P}_f(A_0, B_0)U^*,$$

and $\mathcal{P}_f(0, 0) = 0$.

The identity $A = UA_0U^*$ holds because A is Hermitian with $\text{range}(A) \subseteq \text{range}(U)$, so that $A = UU^*AUU^*$. Two isometries with $\text{range}(U) = \text{range}(B)$ differ by a unitary on $\mathbb{C}^m$, so by Lemma 4.6.1 the definition does not depend on the choice of U ; for $B > 0$ we may take $U = \mathbf{1}_{\mathcal{V}}$, and the definition agrees with the previous one.

Lemma 4.6.3. Let $A, B \in \text{Pos}(\mathcal{V})$ with $\text{range}(A) \subseteq \text{range}(B)$, and let $C \in \text{L}(\mathcal{V}, \mathcal{V}')$ be injective on $\text{range}(B)$. Then $\text{range}(CAC^*) \subseteq \text{range}(CBC^*)$ and

$$\mathcal{P}_f(CAC^*, CBC^*) = C\mathcal{P}_f(A, B)C^*.$$

Proof. We may assume $B \neq 0$. With U , A_0 and B_0 as in Definition 4.6.2, the operator $Y = CU$ is injective. Its polar decomposition is $Y = U'R$, where $R = (Y^*Y)^{1/2}$ is invertible and $U' = YR^{-1}$ is an isometry with $\text{range}(U') = \text{range}(Y)$. Then $CBC^* = U'(RB_0R)U'^*$ and $CAC^* = U'(RA_0R)U'^*$ with $RB_0R > 0$, so

$$\text{range}(CBC^*) = \text{range}(U') \supseteq \text{range}(CAC^*).$$

Using Definition 4.6.2 with the isometry U' and then Lemma 4.6.1,

$$\begin{aligned} \mathcal{P}_f(CAC^*, CBC^*) &= U'\mathcal{P}_f(RA_0R, RB_0R)U'^* = U'R\mathcal{P}_f(A_0, B_0)RU'^* \\ &= Y\mathcal{P}_f(A_0, B_0)Y^* = C\mathcal{P}_f(A, B)C^*. \end{aligned}$$

□

Lemma 4.6.4. Let $A, B \in \text{Pos}(\mathcal{V})$ and $n_{\mathcal{V}} = \dim \mathcal{V}$.

1. The function

$$\varphi_{\eta}(A, B) = \text{tr} \mathcal{P}_f(A + \eta \mathbf{1}_{\mathcal{V}}, B + \eta \mathbf{1}_{\mathcal{V}}) - \eta n_{\mathcal{V}} f(1)$$

is nonincreasing in $\eta \in (0, \infty)$. Consequently the limit defining $\hat{D}_f(A\|B)$ exists in $(-\infty, +\infty]$, and $\hat{D}_f(A\|B) = \sup_{\eta > 0} \varphi_{\eta}(A, B)$.

2. The map $(A, B) \mapsto \hat{D}_f(A\|B)$ is jointly convex and lower semicontinuous on $\text{Pos}(\mathcal{V}) \times \text{Pos}(\mathcal{V})$.

3. If $\text{range}(A) \subseteq \text{range}(B)$, then $\hat{D}_f(A\|B) = \text{tr} \mathcal{P}_f(A, B)$; in particular, $\hat{D}_f(A\|B)$ is finite.

Proof. (1) On pairs of positive definite operators, $\mathcal{P}_f$ is jointly operator convex [37] (see also [38]) and positively homogeneous, $\mathcal{P}_f(\lambda A, \lambda B) = \lambda \mathcal{P}_f(A, B)$ for $\lambda > 0$. It is therefore subadditive:

$$\mathcal{P}_f(A_1 + A_2, B_1 + B_2) = 2 \mathcal{P}_f\left(\frac{1}{2}(A_1 + A_2), \frac{1}{2}(B_1 + B_2)\right) \leq \mathcal{P}_f(A_1, B_1) + \mathcal{P}_f(A_2, B_2).$$

For $0 < \eta' < \eta$, apply this with $(A_1, B_1) = (A + \eta' \mathbf{1}_V, B + \eta' \mathbf{1}_V)$ and $A_2 = B_2 = (\eta - \eta') \mathbf{1}_V$, and use $\mathcal{P}_f(c \mathbf{1}_V, c \mathbf{1}_V) = cf(1) \mathbf{1}_V$; taking traces gives $\varphi_\eta(A, B) \leq \varphi_{\eta'}(A, B)$. Hence $\varphi_\eta(A, B)$ increases as $\eta \downarrow 0$, and since $\eta n_V f(1) \rightarrow 0$, the limit of $\text{tr} \mathcal{P}_f(A + \eta \mathbf{1}_V, B + \eta \mathbf{1}_V)$ exists in $(-\infty, +\infty]$ and equals $\sup_{\eta > 0} \varphi_\eta(A, B)$.

(2) For fixed $\eta > 0$, the map φ_η is continuous, by continuity of square roots, inverses and the continuous functional calculus on positive definite operators, and it is jointly convex, as the trace of a jointly operator convex map plus a constant. A pointwise supremum of continuous jointly convex functions is lower semicontinuous and jointly convex, so (1) gives (2).

(3) We may assume $B \neq 0$. Decompose $V = \text{range}(B) \oplus \ker(B)$. Since $\text{range}(A) \subseteq \text{range}(B)$ and A is Hermitian, $\ker(B) \subseteq \ker(A)$, so $A = A_0 \oplus 0$ and $B = B_0 \oplus 0$ with $B_0 > 0$, identifying $\text{range}(B)$ with $\mathbb{C}^m$ as in Definition 4.6.2. The functional calculus respects direct sums, so

$$\text{tr} \mathcal{P}_f(A + \eta \mathbf{1}_V, B + \eta \mathbf{1}_V) = \text{tr} \mathcal{P}_f(A_0 + \eta \mathbf{1}_{\mathbb{C}^m}, B_0 + \eta \mathbf{1}_{\mathbb{C}^m}) + \eta f(1) \dim \ker(B).$$

As $\eta \downarrow 0$, the operators

$$X_\eta = (B_0 + \eta \mathbf{1}_{\mathbb{C}^m})^{-1/2} (A_0 + \eta \mathbf{1}_{\mathbb{C}^m}) (B_0 + \eta \mathbf{1}_{\mathbb{C}^m})^{-1/2}$$

converge to $X_0 = B_0^{-1/2} A_0 B_0^{-1/2}$, and for $\eta \leq 1$ their spectra lie in a fixed compact interval $I_0 \subseteq [0, \infty)$. Since f is continuous on I_0 , $f(X_\eta) \rightarrow f(X_0)$, and therefore

$$\mathcal{P}_f(A_0 + \eta \mathbf{1}_{\mathbb{C}^m}, B_0 + \eta \mathbf{1}_{\mathbb{C}^m}) \rightarrow \mathcal{P}_f(A_0, B_0).$$

The second term tends to 0, so

$$\widehat{D}_f(A \| B) = \text{tr} \mathcal{P}_f(A_0, B_0) = \text{tr} \mathcal{P}_f(A, B).$$

□

4.6.2 Closed form and continuity

We now come to a statement further relating the max-relative entropy and the ARN derivative. We notice that this then relates to the Lebesgue decomposition of Proposition 4.7.1.

Lemma 4.6.5. *Let Ω and W be as in Theorem 4.5.13, let $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$, and put $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}_{\mathbb{C}^r} - T}$. For $c \geq 0$, we have $\Phi \leq c\Psi$ in the completely positive order if and only if $T \leq \frac{c}{1+c} \mathbf{1}_{\mathbb{C}^r}$. Consequently,*

$$D_{\max}(\Phi \| \Psi) := \log \inf\{c \geq 0 : \Phi \leq c\Psi\} = \log \frac{\|T\|_\infty}{1 - \|T\|_\infty},$$

with $\log 0 = -\infty$ and $D_{\max}(\Phi \| \Psi) = +\infty$ when $\|T\|_\infty = 1$. In particular, for $\epsilon \in (0, 1)$ we have $D_{\max}(\Phi \| \Psi) \leq \log \frac{1-\epsilon}{\epsilon}$ if and only if $T \leq (1 - \epsilon) \mathbf{1}_{\mathbb{C}^r}$.

Proof. By linearity of $T \mapsto \Psi_T$, $c\Psi - \Phi = \Psi_{c\mathbf{1}_{\mathbb{C}^r} - (1+c)T}$, whose Choi matrix is

$$W(c\mathbf{1}_{\mathbb{C}^r} - (1+c)T)W^*$$

by Theorem 4.2.2. By Choi's theorem, $c\Psi - \Phi$ is completely positive if and only if this operator is positive semidefinite, and this happens if and only if $c\mathbf{1}_{\mathbb{C}^r} - (1+c)T \geq 0$.

Reasoning about the spectrum of T , we then obtain that this translates to

$$\frac{\|T\|_\infty}{1 - \|T\|_\infty} \leq c < \infty.$$

This then implies that

$$\{c \geq 0 : \Phi \leq c\Psi\} = \left[\frac{\|T\|_\infty}{1 - \|T\|_\infty}, \infty \right)$$

if $\|T\|_\infty < 1$, and is empty otherwise. This gives the formula for $D_{\max}$. The last statement follows because $t \mapsto t/(1-t)$ is increasing on $[0, 1)$. $\square$

Remark 4.6.6 (Physical interpretation). For channels and $c > 1$, $\Phi \leq c\Psi$ holds if and only if $\Psi = c^{-1}\Phi + (1 - c^{-1})\Xi$ with $\Xi = (c\Psi - \Phi)/(c-1)$ a channel, so $e^{-D_{\max}(\Phi\|\Psi)}$ is the largest probability with which Ψ secretly runs Φ . Equivalently, by complete positivity of $c\Psi - \Phi$, on every input, ancilla allowed, every measurement outcome is at most $e^{D_{\max}(\Phi\|\Psi)}$ times likelier under Φ than under Ψ , as for states [112]. Detector inefficiency $\epsilon \in (0, 1)$ gives such bounds: an instrument (Φ, Ψ) is observed as $((1-\epsilon)\Phi, \Psi + \epsilon\Phi)$, the case $p_{01} = \epsilon$, $p_{10} = 0$ of Remark 4.4.7, with derivative $(1-\epsilon)T \leq (1-\epsilon)\mathbf{1}_{\mathbb{C}^r}$, so

$$D_{\max}((1-\epsilon)\Phi\|\Psi + \epsilon\Phi) \leq \log((1-\epsilon)/\epsilon)$$

by the lemma, whatever the ideal pair.

For $\epsilon \in (0, 1)$ and a continuous function $g : [0, 1-\epsilon] \rightarrow \mathbb{R}$ we use the modulus $\omega_g^{(\epsilon)} = \omega_g^{[0, 1-\epsilon]}$ of Lemma 4.4.8, which tends to 0 as $\delta \rightarrow 0$.

Theorem 4.6.7. *Let $\Omega \in \text{CP}(\mathcal{X}, \mathcal{Y})$ have linearly independent Kraus operators $V_1, \dots, V_r$, so that $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$, let W be the matrix with columns $\text{vec}(V_1), \dots, \text{vec}(V_r)$, and define the continuous function $h_f : [0, 1) \rightarrow \mathbb{R}$ by*

$$h_f(t) = (1-t) f\left(\frac{t}{1-t}\right), \quad \text{so that } h_f(0) = f(0^+).$$

Let $\epsilon \in (0, 1)$.

(a) *If $0 \leq T \leq (1-\epsilon)\mathbf{1}_{\mathbb{C}^r}$, $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}_{\mathbb{C}^r} - T}$, then $\hat{D}_f(\Phi\|\Psi)$ is finite and*

$$\hat{D}_f(\Phi\|\Psi) = \lambda_{\max}\left(\text{Tr}_{\mathcal{Y}}(W h_f(T) W^*)\right).$$

(b) *If $0 \leq T_1, T_2 \leq (1-\epsilon)\mathbf{1}_{\mathbb{C}^r}$, then*

$$\left| \hat{D}_f(\Psi_{T_1}\|\Psi_{\mathbf{1}_{\mathbb{C}^r} - T_1}) - \hat{D}_f(\Psi_{T_2}\|\Psi_{\mathbf{1}_{\mathbb{C}^r} - T_2}) \right| \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \omega_{h_f}^{(\epsilon)}(\|T_1 - T_2\|_\infty),$$

and $\omega_{h_f}^{(\epsilon)}(\delta) \rightarrow 0$ as $\delta \rightarrow 0$.

By Lemma 4.6.5, the hypotheses say exactly that the pairs involved satisfy $D_{\max} \leq \log \frac{1-\epsilon}{\epsilon}$.

Proof. (a) The structure of the first part is that we aim to use the conditions to transform into the absolute continuity form of Lemma 4.6.4(3). Let $c = (1 - \epsilon)/\epsilon$. Then $\Phi \leq c\Psi$ by Lemma 4.6.5. We proceed in a similar way to the bounded case by letting $\rho \in D(\mathcal{X} \otimes \mathcal{X})$ write $\Phi_\rho = (\Phi \otimes \mathbf{1}_{L(\mathcal{X})})(\rho)$ and $\Psi_\rho = (\Psi \otimes \mathbf{1}_{L(\mathcal{X})})(\rho)$, and let $d(\rho) = \widehat{D}_f(\Phi_\rho \| \Psi_\rho)$.

Since $c\Psi - \Phi$ is completely positive, $\Phi_\rho \leq c\Psi_\rho$; if $\Psi_\rho v = 0$ then $0 \leq v^* \Phi_\rho v \leq c v^* \Psi_\rho v = 0$, so $\ker(\Psi_\rho) \subseteq \ker(\Phi_\rho)$ and $\text{range}(\Phi_\rho) \subseteq \text{range}(\Psi_\rho)$. By Lemma 4.6.4(3), $d(\rho) = \text{tr } \mathcal{P}_f(\Phi_\rho, \Psi_\rho)$ is finite for every ρ .

We now need to inspect these quantities at the Choi level, following that we can reduce the maximal f -divergences to partial traces of Choi relations. By Theorem 4.2.2, $J(\Phi) = WTW^*$ and $J(\Psi) = W(\mathbf{1}_{\mathbb{C}^r} - T)W^*$. Since $\mathbf{1}_{\mathbb{C}^r} - T \geq \epsilon \mathbf{1}_{\mathbb{C}^r}$ is invertible and W is injective, Lemma 4.6.3 with $C = W$ gives $\text{range}(J(\Phi)) \subseteq \text{range}(J(\Psi))$ and

$$\mathcal{P}_f(J(\Phi), J(\Psi)) = W \mathcal{P}_f(T, \mathbf{1}_{\mathbb{C}^r} - T) W^*.$$

As T and $\mathbf{1}_{\mathbb{C}^r} - T$ commute,

$$(\mathbf{1}_{\mathbb{C}^r} - T)^{-1/2} T (\mathbf{1}_{\mathbb{C}^r} - T)^{-1/2} = T (\mathbf{1}_{\mathbb{C}^r} - T)^{-1},$$

and the functional calculus for operator perspectives gives

$$\mathcal{P}_f(T, \mathbf{1}_{\mathbb{C}^r} - T) = (\mathbf{1}_{\mathbb{C}^r} - T) f(T(\mathbf{1}_{\mathbb{C}^r} - T)^{-1}) = h_f(T).$$

Hence $\mathcal{P}_f(J(\Phi), J(\Psi)) = W h_f(T) W^*$.

We now show that the supremum over input states in the definition of $\widehat{D}_f(\Phi \| \Psi)$ is the largest eigenvalue of the partial trace of this Choi-level operator, by the same line of reasoning as in the Kubo–Ando case. We again look at invertible operators first.

Let $M = \text{Tr}_{\mathcal{Y}}(W h_f(T) W^*)$, a Hermitian operator on $\mathcal{X}$. Let $u = \text{vec}(A)$ with $\|A\|_2 = 1$ and A invertible, and put $C = \mathbf{1}_{\mathcal{Y}} \otimes A^\top$, which is invertible. By (4.3.1), $\Phi_{uu^*} = C J(\Phi) C^*$ and $\Psi_{uu^*} = C J(\Psi) C^*$. By Lemma 4.6.4(3), Lemma 4.6.3 and the previous step,

$$\begin{aligned} d(uu^*) &= \text{tr}(C \mathcal{P}_f(J(\Phi), J(\Psi)) C^*) \\ &= \text{tr}((\mathbf{1}_{\mathcal{Y}} \otimes (AA^*)^\top) W h_f(T) W^*) \\ &= \text{tr}((AA^*)^\top M), \end{aligned}$$

using $C^*C = \mathbf{1}_{\mathcal{Y}} \otimes (AA^*)^\top$.

By Lemma 4.6.4(2), d is convex and lower semicontinuous on $D(\mathcal{X} \otimes \mathcal{X})$, being the composition of a jointly convex, lower semicontinuous function with the linear map $\rho \mapsto (\Phi_\rho, \Psi_\rho)$. Writing $\rho = \sum_i p_i u_i u_i^*$ in its spectral decomposition, convexity gives $d(\rho) \leq \max_i d(u_i u_i^*)$, so the supremum defining $\widehat{D}_f(\Phi \| \Psi)$ may be taken over pure states. Every unit vector $u = \text{vec}(A)$ is a limit of unit vectors $u_n = \text{vec}(A_n)$ with A_n invertible, and lower semicontinuity gives $d(uu^*) \leq \liminf_n d(u_n u_n^*)$, so the supremum may be taken over $u = \text{vec}(A)$ with A invertible. As A ranges over such operators, $(AA^*)^\top$ ranges over the invertible density operators on $\mathcal{X}$, which are dense in $D(\mathcal{X})$. Hence

$$\widehat{D}_f(\Phi \| \Psi) = \sup \{ \text{tr}(\xi M) : \xi \in D(\mathcal{X}) \text{ invertible} \} = \max_{\xi \in D(\mathcal{X})} \text{tr}(\xi M) = \lambda_{\max}(M).$$

(b) Let $M_i = \text{Tr}_{\mathcal{Y}}(W h_f(T_i) W^*)$ and $K = h_f(T_1) - h_f(T_2)$, so that $\|K\|_\infty \leq \omega_{h_f}^{(\epsilon)}(\|T_1 - T_2\|_\infty)$ by definition. By (a), the left-hand side of (b) is $|\lambda_{\max}(M_1) - \lambda_{\max}(M_2)|$, which is at most $\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \|K\|_\infty$ by Lemma 4.4.10. Finally, h_f is continuous on $[0, 1 - \epsilon]$, so $\omega_{h_f}^{(\epsilon)}(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. $\square$

Remark 4.6.8 (Physical interpretation). For states, $\hat{D}_f$ is the least classical f -divergence of a pair of distributions that prepares them through one classical-to-quantum channel [91]. For channels and an input $u = \text{vec}(A)$ with A invertible, let $T = \sum_i t_i P_i$ spectrally, $\widetilde{W} = (\mathbf{1}_{\mathcal{V}} \otimes A^\top)W$ and $w_i = \text{tr}(\widetilde{W} P_i \widetilde{W}^*)$. The distributions $p_i = t_i w_i$, $q_i = (1 - t_i) w_i$ and the preparation $\delta_i \mapsto \widetilde{W} P_i \widetilde{W}^* / w_i$ give the outputs $\widetilde{W} T \widetilde{W}^*$ and $\widetilde{W}(\mathbf{1}_{\mathbb{C}^r} - T) \widetilde{W}^*$ (see (4.3.1)), and $\sum_i q_i f(p_i/q_i) = \sum_i w_i h_f(t_i)$ is the quantum value by Lemma 4.7.4, so they are optimal. Their likelihood ratios $t_i/(1 - t_i)$ do not depend on the input, and $\hat{D}_f(\Phi\|\Psi)$ is the worst case over the weights w_i . The bound $D_{\max} \leq \log((1 - \epsilon)/\epsilon)$ means that all ratios lie in $[0, (1 - \epsilon)/\epsilon]$, and the constants of Corollaries 4.6.11 and 4.6.12 measure how steep h_f is on $[0, 1 - \epsilon]$.

4.6.3 The relative Rényi entropies continued

In Section 4.5, we considered the bounds for the Rényi relative entropies for $\alpha \in (0, 1)$. We now have the machinery in place to compute explicit bounds for $\alpha \in (1, 2]$. We first need to list some useful bounds. Using the subsequent results, we will also show continuity bounds for the Belavkin–Staszewski divergence. In the following lemma, integrals of matrices are understood in the functional calculus sense.

Lemma 4.6.9. *Let $\mathcal{V}$ be a complex Euclidean space, $\delta \in (0, 1]$, $\beta \in (0, 1]$ and $\epsilon > 0$.*

1. *If $X_1, X_2 \geq \epsilon \mathbf{1}_{\mathcal{V}}$, then $\|\log X_1 - \log X_2\|_\infty \leq \|X_1 - X_2\|_\infty / \epsilon$.*

2. *If $X_1, X_2 \geq \epsilon \mathbf{1}_{\mathcal{V}}$, then*

$$\|X_1^{-\beta} - X_2^{-\beta}\|_\infty \leq \beta \epsilon^{-\beta-1} \|X_1 - X_2\|_\infty.$$

3. *If $0 \leq A, B \leq \mathbf{1}_{\mathcal{V}}$ and $\|A - B\|_\infty \leq \delta$, then*

$$\|A \log A - B \log B\|_\infty \leq \delta \log \frac{1}{\delta} + (1 + \delta) \log(1 + \delta) \leq \delta \log \frac{1}{\delta} + 2 \log 2 \cdot \delta$$

4. *If $0 \leq A, B \leq \mathbf{1}_{\mathcal{V}}$, then $\|A^{1+\beta} - B^{1+\beta}\|_\infty \leq (3 - \beta) \|A - B\|_\infty$.*

Proof. (1) For $X > 0$, the scalar identity

$$\log x = \int_0^\infty \left(\frac{1}{1+s} - \frac{1}{x+s} \right) ds$$

and the functional calculus give

$$\log X = \int_0^\infty ((1+s)^{-1} \mathbf{1}_{\mathcal{V}} - (X+s\mathbf{1}_{\mathcal{V}})^{-1}) ds.$$

Hence

$$\log X_1 - \log X_2 = \int_0^\infty (X_1 + s\mathbf{1}_{\mathcal{V}})^{-1} (X_1 - X_2) (X_2 + s\mathbf{1}_{\mathcal{V}})^{-1} ds,$$

and $\|(X_i + s\mathbf{1}_{\mathcal{V}})^{-1}\|_\infty \leq (\epsilon + s)^{-1}$ gives the bound

$$\|X_1 - X_2\|_\infty \int_0^\infty (\epsilon + s)^{-2} ds = \|X_1 - X_2\|_\infty / \epsilon.$$

(2) For $\beta = 1$ this follows from $X_1^{-1} - X_2^{-1} = X_1^{-1}(X_2 - X_1)X_2^{-1}$. For $0 < \beta < 1$ we use the representation

$$x^{-\beta} = \frac{\sin \beta \pi}{\pi} \int_0^\infty s^{-\beta} (x+s)^{-1} ds,$$

valid for $x > 0$ [16, Chapter V], which gives

$$X_1^{-\beta} - X_2^{-\beta} = \frac{\sin \beta \pi}{\pi} \int_0^\infty s^{-\beta} (X_1 + s \mathbf{1}_V)^{-1} (X_2 - X_1) (X_2 + s \mathbf{1}_V)^{-1} ds.$$

Since $\|(X_i + s \mathbf{1}_V)^{-1}\|_\infty \leq (\epsilon + s)^{-1}$ and

$$\int_0^\infty s^{-\beta} (\epsilon + s)^{-2} ds = \epsilon^{-\beta-1} \Gamma(1-\beta) \Gamma(1+\beta) = \epsilon^{-\beta-1} \frac{\beta \pi}{\sin \beta \pi}$$

the norm of the right-hand side is at most $\beta \epsilon^{-\beta-1} \|X_1 - X_2\|_\infty$ where Γ is the gamma function.

(3) The case $A = B$ is trivial, and since the right-hand side is increasing in δ we may assume $\delta = \|A - B\|_\infty > 0$. For $s > 0$ let $g_s(t) = \frac{t}{1+s} - \frac{t}{t+s}$. The scalar identity $t \log t = \int_0^\infty g_s(t) ds$, valid for $t \geq 0$, gives

$$A \log A - B \log B = \int_0^\infty (g_s(A) - g_s(B)) ds.$$

We claim

$$\|g_s(A) - g_s(B)\|_\infty \leq \min \left\{ \frac{\delta}{s(1+s)}, \frac{1}{1+s} \right\}.$$

Indeed, the first bound is $\text{Lip}^{[0,1]}(g_s) = \frac{1}{s(1+s)}$ from Corollary 4.9.7(d), whose proof does not use the present lemma, and the second holds because, by the same corollary, $g_s(A)$ and $g_s(B)$ have spectra in $[-\frac{1}{1+s}, 0]$. Integrating,

$$\|A \log A - B \log B\|_\infty \leq \int_0^\delta \frac{ds}{1+s} + \int_\delta^\infty \frac{\delta ds}{s(1+s)} = \log(1+\delta) + \delta \log \frac{1+\delta}{\delta},$$

which is the first bound. The second follows from $(1+\delta) \log(1+\delta) \leq 2 \log 2 \cdot \delta$ for $\delta \in [0, 1]$, by convexity.

(4) For $\beta = 1$ the claim is $\|A^2 - B^2\|_\infty \leq 2\|A - B\|_\infty$, which follows from

$$A^2 - B^2 = \frac{1}{2}((A+B)(A-B) + (A-B)(A+B)).$$

For $0 < \beta < 1$, the representation

$$x^\beta = \frac{\sin \beta \pi}{\pi} \int_0^\infty s^{\beta-1} x(x+s)^{-1} ds$$

for $x \geq 0$ [16, Chapter V] gives

$$A^{1+\beta} = \frac{\sin \beta \pi}{\pi} \int_0^\infty s^{\beta-1} A^2 (A + s \mathbf{1}_V)^{-1} ds$$

and similarly for B . Writing

$$A^2 (A + s \mathbf{1}_V)^{-1} = A - s \mathbf{1}_V + s^2 (A + s \mathbf{1}_V)^{-1}$$

and

$$Y_s = A(A + s \mathbf{1}_V)^{-1}$$

$$Z_s = B(B + s \mathbf{1}_V)^{-1}$$

so that

$$s(A + s \mathbf{1}_V)^{-1} = \mathbf{1}_V - Y_s$$

and

$$s(B + s\mathbf{1}_{\mathcal{V}})^{-1} = \mathbf{1}_{\mathcal{V}} - Z_s$$

we obtain with $H = A - B$

$$A^2(A + s\mathbf{1}_{\mathcal{V}})^{-1} - B^2(B + s\mathbf{1}_{\mathcal{V}})^{-1} = H - (\mathbf{1}_{\mathcal{V}} - Y_s)H(\mathbf{1}_{\mathcal{V}} - Z_s) = Y_s H + H Z_s - Y_s H Z_s.$$

Since $0 \leq A, B \leq \mathbf{1}_{\mathcal{V}}$, we have $\|Y_s\|_{\infty}, \|Z_s\|_{\infty} \leq (1+s)^{-1}$, so the norm of this operator is at most $(\frac{2}{1+s} + \frac{1}{(1+s)^2})\|H\|_{\infty}$. As $\int_0^{\infty} \frac{s^{\beta-1}}{1+s} ds = \frac{\pi}{\sin \beta \pi}$ and

$$\int_0^{\infty} \frac{s^{\beta-1}}{(1+s)^2} ds = \Gamma(\beta)\Gamma(2-\beta) = \frac{(1-\beta)\pi}{\sin \beta \pi},$$

integrating gives $\|A^{1+\beta} - B^{1+\beta}\|_{\infty} \leq (2 + 1 - \beta)\|H\|_{\infty}$. $\square$

Remark 4.6.10. For $f(t) = t^{\alpha}$, Bhatia and Sinha [19] showed that the Fréchet derivative of $A \mapsto A^{\alpha}$ at a positive operator $A \in \mathcal{L}(\mathcal{V})$ has norm $\|Df(A)\| = \|\alpha A^{\alpha-1}\|_{\infty}$ when $\alpha < 0$ or $\alpha \geq 2$, extending the case $0 \leq \alpha \leq 1$ obtained earlier, but that this fails when $1 < \alpha < \sqrt{2}$. Since

$$\|Df(A)\| \geq \|Df(A)(\mathbf{1}_{\mathcal{V}})\|_{\infty} = \|\alpha A^{\alpha-1}\|_{\infty}$$

and $Df(cA) = c^{\alpha-1}Df(A)$ for $c > 0$, for each $1 < \alpha < \sqrt{2}$ there is a positive A with $\|A\|_{\infty} = 1$ and $\|Df(A)\| > \alpha$; as the best constant in Lemma 4.6.9(4) is $\text{Lip}^{[0,1]}(t^{1+\beta}) \geq \|Df(A)\|$ for such A and $\alpha = 1 + \beta$ (Lemma 4.9.3 and the Daleckiĭ–Krein formula), the constant $3 - \beta$ there cannot be lowered to $1 + \beta$, the maximum of the derivative of $t^{1+\beta}$ on $[0, 1]$, when $0 < \beta < \sqrt{2} - 1$. For $\beta = 1$ the constant $3 - \beta = 2$ is optimal, since $Df(\mathbf{1}_{\mathcal{V}})(H) = 2H$ for $f(t) = t^2$.

Using these bounds, we now obtain continuity bounds for the Belavkin–Staszewski relative entropy for channels with the same sum with max-relative entropy upper bounds on the ARN derivative.

Corollary 4.6.11. *Let $\hat{D}_{\text{BS}} = \hat{D}_f$ with $f(t) = t \log t$, the Belavkin–Staszewski relative entropy [14], and let $\Omega, W, \epsilon, T_1, T_2$ be as in Theorem 4.6.7(b). Let $\Phi_i = \Psi_{T_i}$, $\Psi_i = \Psi_{\mathbf{1}_{\mathcal{C}^r} - T_i}$ and $\delta = \|T_1 - T_2\|_{\infty}$. Then*

$$|\hat{D}_{\text{BS}}(\Phi_1\|\Psi_1) - \hat{D}_{\text{BS}}(\Phi_2\|\Psi_2)| \leq \|\Omega^*(\mathbf{1}_{\mathcal{V}})\|_{\infty} \delta \left(\log \frac{4}{\epsilon \delta} + \frac{1}{\epsilon} \right),$$

the right-hand side being interpreted as 0 for $\delta = 0$.

Proof. Here $f(0^+) = 0$ and

$$h_f(t) = t \log \frac{t}{1-t} = t \log t - t \log(1-t).$$

With $L_i = \log(\mathbf{1}_{\mathcal{C}^r} - T_i)$,

$$h_f(T_1) - h_f(T_2) = (T_1 \log T_1 - T_2 \log T_2) - (T_1 - T_2)L_1 - T_2(L_1 - L_2).$$

Since $\delta \leq 1$, the first term has ∞ -norm at most $\delta \log \frac{1}{\delta} + 2 \log 2 \cdot \delta$ by Lemma 4.6.9(3). The spectrum of $\mathbf{1}_{\mathcal{C}^r} - T_1$ lies in $[\epsilon, 1]$, so $\|L_1\|_{\infty} \leq \log \frac{1}{\epsilon}$, and Lemma 4.6.9(1) with $X_i = \mathbf{1}_{\mathcal{C}^r} - T_i$ gives $\|L_1 - L_2\|_{\infty} \leq \delta/\epsilon$; also $\|T_2\|_{\infty} \leq 1$. Thus

$$\begin{aligned} \|h_f(T_1) - h_f(T_2)\|_{\infty} &\leq \delta \log \frac{1}{\delta} + 2 \log 2 \cdot \delta + \log \frac{1}{\epsilon} \cdot \delta + \frac{\delta}{\epsilon} \\ &= \delta \left(\log \frac{4}{\epsilon \delta} + \frac{1}{\epsilon} \right), \end{aligned}$$

and the proof of Theorem 4.6.7(b) gives the claim. $\square$

Corollary 4.6.12. *Let $\alpha \in (1, 2]$ and $\beta = \alpha - 1$, and let $\Omega, W, \epsilon, T_1, T_2, \Phi_i, \Psi_i$ and δ be as in Corollary 4.6.11. Define $\hat{Q}_\alpha(\Phi\|\Psi) = \sup_{\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})} \hat{Q}_\alpha(\Phi_\rho\|\Psi_\rho)$, so that $\hat{D}_\alpha(\Phi\|\Psi) = \frac{1}{\alpha-1} \log \hat{Q}_\alpha(\Phi\|\Psi)$. Then*

$$\hat{Q}_\alpha(\Phi_i\|\Psi_i) = \lambda_{\max}\left(\mathrm{Tr}_Y(W h_\alpha(T_i) W^*)\right), \quad h_\alpha(t) = t^\alpha(1-t)^{1-\alpha},$$

and

$$|\hat{Q}_\alpha(\Phi_1\|\Psi_1) - \hat{Q}_\alpha(\Phi_2\|\Psi_2)| \leq \|\Omega^*(\mathbf{1}_Y)\|_\infty \left((4-\alpha)\epsilon^{1-\alpha} + (\alpha-1)\epsilon^{-\alpha} \right) \delta.$$

If moreover $\Phi_1, \Psi_1, \Phi_2, \Psi_2$ are channels, then $|\hat{D}_\alpha(\Phi_1\|\Psi_1) - \hat{D}_\alpha(\Phi_2\|\Psi_2)|$ is at most $\frac{1}{\alpha-1}$ times the right-hand side.

Proof. The function $f(t) = t^\alpha$ is operator convex on $(0, \infty)$ for $\alpha \in [1, 2]$ [16, Chapter V], with $f(0^+) = 0$, and $h_f = h_\alpha$ is given by the formula (4.5.1), now considered on $[0, 1]$. The definition of $\hat{Q}_\alpha(\rho\|\xi)$ regularises only ξ , but when $\mathrm{range}(\rho) \subseteq \mathrm{range}(\xi)$ the direct-sum argument in the proof of Lemma 4.6.4(3) shows that it also equals $\mathrm{tr} \mathcal{P}_f(\rho, \xi)$, since $0^\alpha = 0$. By the proof of Theorem 4.6.7(a), $\mathrm{range}(\Phi_{i,\rho}) \subseteq \mathrm{range}(\Psi_{i,\rho})$ for every ρ , so $\hat{Q}_\alpha(\Phi_i\|\Psi_i) = \hat{D}_f(\Phi_i\|\Psi_i)$, and Theorem 4.6.7(a) gives the closed form.

For the estimate, put $X_i = \mathbf{1}_{\mathbb{C}^r} - T_i \geq \epsilon \mathbf{1}_{\mathbb{C}^r}$, so that $h_\alpha(T_i) = T_i^\alpha X_i^{-\beta}$ and

$$h_\alpha(T_1) - h_\alpha(T_2) = (T_1^\alpha - T_2^\alpha)X_1^{-\beta} + T_2^\alpha(X_1^{-\beta} - X_2^{-\beta}).$$

By Lemma 4.6.9(4) and (2), together with $\|X_1^{-\beta}\|_\infty \leq \epsilon^{-\beta}$ and $\|T_2^\alpha\|_\infty \leq 1$, the norm of this operator is at most

$$(3-\beta)\epsilon^{-\beta}\delta + \beta\epsilon^{-\beta-1}\delta = ((4-\alpha)\epsilon^{1-\alpha} + (\alpha-1)\epsilon^{-\alpha})\delta,$$

and the proof of Theorem 4.6.7(b) gives the estimate. If all four maps are channels, then $\Phi_{i,\rho}$ and $\Psi_{i,\rho}$ are states, so $\hat{Q}_\alpha(\Phi_{i,\rho}\|\Psi_{i,\rho}) \geq 1$ by the data-processing inequality for $\hat{D}_\alpha$ under the trace [72]; hence $\hat{Q}_\alpha(\Phi_i\|\Psi_i) \geq 1$, and the last claim follows from $|\log x - \log y| \leq |x - y|$ for $x, y \geq 1$. $\square$

Remark 4.6.13. The same argument applies to the maximal χ^2 -divergence, $f(t) = (t-1)^2$, for which $h_f(t) = (2t-1)^2(1-t)^{-1}$. Writing

$$\begin{aligned} h_f(T_1) - h_f(T_2) &= ((2T_1 - \mathbf{1}_{\mathbb{C}^r})^2 - (2T_2 - \mathbf{1}_{\mathbb{C}^r})^2)X_1^{-1} \\ &\quad + (2T_2 - \mathbf{1}_{\mathbb{C}^r})^2(X_1^{-1} - X_2^{-1}) \end{aligned}$$

with $X_i = \mathbf{1}_{\mathbb{C}^r} - T_i$, the identity

$$A^2 - B^2 = \frac{1}{2}((A+B)(A-B) + (A-B)(A+B))$$

gives $\omega_{h_f}^{(\epsilon)}(\delta) \leq (4\epsilon^{-1} + \epsilon^{-2})\delta$. By Corollary 4.9.7(b), this improves to the optimal $\omega_{h_f}^{(\epsilon)}(\delta) \leq (\epsilon^{-2} - 4)\delta$ for $\epsilon \leq \frac{1}{3}$.

Remark 4.6.14. The hypothesis $D_{\max} \leq \log \frac{1-\epsilon}{\epsilon}$ cannot simply be dropped when $f'(\infty) = +\infty$, since then $\hat{D}_f(A\|B) = +\infty$ whenever $\mathrm{range}(A) \not\subseteq \mathrm{range}(B)$ (the case $\dim \mathcal{X} = 1$ of Theorem 4.7.5(a); for the standard divergence see [61, Corollary 3.4]). In the example of Corollary 4.3.3, the derivative T_1 of Φ_1 is a projection, so $D_{\max}(\Phi_1\|\Psi_1) = +\infty$ by Lemma 4.6.5, and $\hat{D}_f(\Phi_1\|\Psi_1) = +\infty$, whereas $\hat{D}_f(\Phi_2\|\Psi_2) = f(1)$ by Theorem 4.6.7(a) with $T = \frac{1}{2}\mathbf{1}_{\mathbb{C}^r}$. The constants in Corollaries 4.6.11 and 4.6.12 blow up as $\epsilon \downarrow 0$, that

is, as the max-relative entropy within the pairs grows, and the impossibility example is the limiting case. If instead $f(0^+) = +\infty$ and $f'(\infty) < \infty$, the transpose $\tilde{f}(t) = tf(1/t)$ is operator convex with $\tilde{f}(0^+) = f'(\infty)$, and $\hat{D}_f(A\|B) = \hat{D}_{\tilde{f}}(B\|A)$ (immediate from the reverse-test definition [91]; for the standard divergence see [61, Proposition 3.7]); the results of this section then apply with the roles of Φ and Ψ exchanged, under the hypothesis $T \geq \epsilon \mathbf{1}_{\mathbb{C}^r}$.

4.7 All operator convex functions

Section 4.6 assumed $f(0^+) < \infty$, and the remark at its end treated the mirror case $f'(\infty) < \infty$. Several natural divergences have $f(0^+) = f'(\infty) = +\infty$. Examples are the Jeffreys divergence, with $f(t) = (t-1) \log t$; the symmetrised χ^2 -divergence, with $f(t) = (t-1)^2(1+t^{-1})$; and $f(t) = t^\alpha + t^{-\beta}$ with $\alpha \in (1, 2]$, $\beta \in (0, 1]$. In this section we assume only that f is operator convex. The closed form of Theorem 4.6.7 then holds verbatim, with h_f extended to $[0, 1]$ by the boundary values $f(0^+)$ and $f'(\infty)$. As we will show, finiteness determines when values of ∞ are attained.

4.7.1 The Lebesgue decomposition of a pair

For $A, B \in \text{Pos}(\mathcal{V})$, Ando's Lebesgue decomposition [5] writes $A = A_{\text{ac}} + A_s$. Here A_{ac} is the largest positive operator below A that is dominated by a multiple of B ; in finite dimensions it is the shorted operator of A to $\text{range}(B)$ [2, 4], and $\text{range}(A_s) \cap \text{range}(B) = \{0\}$. For completely positive maps we apply this to Choi matrices. Okayasu has recently considered the Lebesgue decomposition in [100].

Proposition 4.7.1. *Let Ω , $V_1, \dots, V_r$ and W be as in Theorem 4.2.15, let $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$, $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}_{\mathbb{C}^r} - T}$, and let P_0 and P_1 be the spectral projections of T for the eigenvalues 0 and 1 (zero if these are not eigenvalues). Relative to $J(\Psi)$, the Lebesgue decomposition of $J(\Phi)$ is $J(\Phi)_{\text{ac}} = WT(\mathbf{1}_{\mathbb{C}^r} - P_1)W^*$ and $J(\Phi)_s = WP_1W^*$. Symmetrically, relative to $J(\Phi)$, $J(\Psi)_s = WP_0W^*$. We write $\Phi_s = \Psi_{P_1}$ and $\Psi_s = \Psi_{P_0}$. In particular, the following are equivalent: $\Phi_s = 0$; $1 \notin \text{spec}(T)$; $\text{range } J(\Phi) \subseteq \text{range } J(\Psi)$; and $D_{\max}(\Phi\|\Psi) < \infty$.*

Proof. We repeat a similar argument to the Douglas factorisation used to prove Theorem 4.2.2. If $0 \leq C \leq WTW^*$, then $\text{range } C \subseteq \text{range } W$, so $C = WC'W^*$ with $C' = W^+C(W^+)^*$ and $0 \leq C' \leq T$. Since W is injective, the short of $J(\Phi)$ to $\text{range } J(\Psi) = W \text{range}(\mathbf{1}_{\mathbb{C}^r} - T)$ is W times the short of T to $\text{range}(\mathbf{1}_{\mathbb{C}^r} - T) = \text{range}(\mathbf{1}_{\mathbb{C}^r} - P_1)$, times W^* . Write $P_1^\perp = \mathbf{1}_{\mathbb{C}^r} - P_1$; it commutes with T . If $0 \leq C' \leq T$ and $\text{range } C' \subseteq \text{range } P_1^\perp$, then $C' = P_1^\perp C' P_1^\perp \leq P_1^\perp T P_1^\perp = T P_1^\perp$. Moreover $T P_1^\perp \leq T$, because $T - T P_1^\perp = T P_1 \geq 0$. So the short is $T P_1^\perp$, and $J(\Phi)_s = W T P_1 W^* = W P_1 W^*$. The last equivalences follow from

$$\text{range } J(\Phi) \subseteq \text{range } J(\Psi) \iff \text{range } T \subseteq \text{range}(\mathbf{1}_{\mathbb{C}^r} - T) \iff P_1 = 0$$

and from Lemma 4.6.5. $\square$

4.7.2 Boundary values

Lemma 4.7.2. *Let $f : (0, \infty) \rightarrow \mathbb{R}$ be operator convex. The limits $f(0^+) = \lim_{t \downarrow 0} f(t)$ and $f'(\infty) = \lim_{t \rightarrow \infty} f(t)/t$ exist in $(-\infty, +\infty]$. The function $h_f(t) = (1-t)f(\frac{t}{1-t})$ on $(0, 1)$, extended by $h_f(0) = f(0^+)$ and $h_f(1) = f'(\infty)$, is convex and continuous as a map $[0, 1] \rightarrow (-\infty, +\infty]$. In particular $m_f = \min_{[0, 1]} h_f$ is finite.*

Proof. For $t \in (0, 1)$, the convex function f satisfies $f(t) \geq f(1) + f'_-(1)(t-1)$, so $\lim_{t \downarrow 0} f(t)$ exists in $(-\infty, +\infty]$. The chord slopes $(f(t) - f(1))/(t-1)$ increase in t , so their limit as $t \rightarrow \infty$, which is also the limit of $f(t)/t$, exists in $(-\infty, +\infty]$. On $(0, 1)$ we have $h_f(t) = g(t, 1-t)$, where $g(x, y) = yf(x/y)$ is the perspective of f , which is jointly convex on $(0, \infty)^2$; hence h_f is convex on $(0, 1)$. As $t \downarrow 0$, $h_f(t) \rightarrow f(0^+)$. As $t \uparrow 1$, $h_f(t) = tf(x)/x$ with $x = t/(1-t) \rightarrow \infty$, so $h_f(t) \rightarrow f'(\infty)$. A convex function on $(0, 1)$ with limits in $(-\infty, +\infty]$ at both ends is bounded below and extends continuously to $[0, 1]$ with values in $(-\infty, +\infty]$. $\square$

For Hermitian T with spectrum in $[0, 1]$, we say that $h_f(T)$ is finite if h_f is finite on $\text{spec}(T)$; then $h_f(T)$ is defined by the functional calculus. By Proposition 4.7.1, $h_f(T)$ is finite exactly when $f'(\infty) < \infty$ or $\Phi_s = 0$, and $f(0^+) < \infty$ or $\Psi_s = 0$.

4.7.3 The closed form

We want to derive a closed form for all operator convex functions with appropriate conditions. Lemma 4.6.4(3) says that we have a closed form when $\text{range}(A) \subseteq \text{range}(B)$. This is not always true, but we can perform congruence transformations to fix this issue as the below lemma shows.

Lemma 4.7.3. *Let f be operator convex. Parts (1) and (2) of Lemma 4.6.4 hold for f , with the same proofs. If $A, B \in \text{Pos}(\mathcal{V})$ satisfy $\text{range}(A) = \text{range}(B)$, then, with U, A_0, B_0 as in Definition 4.6.2, both A_0 and B_0 are invertible and $\widehat{D}_f(A\|B) = \text{tr } \mathcal{P}_f(A, B)$. If moreover C is injective on $\text{range}(B)$, then*

$$\widehat{D}_f(CAC^* \| CBC^*) = \text{tr } (C \mathcal{P}_f(A, B) C^*).$$

Proof. Parts (1) and (2) of Lemma 4.6.4 use only the joint operator convexity and positive homogeneity of $\mathcal{P}_f$ on positive definite pairs and the continuity of φ_η , which hold for every operator convex f . For the second statement we follow the proof of part (3): now $X_0 = B_0^{-1/2} A_0 B_0^{-1/2}$ is invertible, so for small η the spectra of X_η lie in a fixed compact interval $I_0 \subseteq [\frac{1}{2}\lambda_{\min}(X_0), \infty)$, where f is continuous, and

$$\mathcal{P}_f(A_0 + \eta \mathbf{1}_{\mathbb{C}^m}, B_0 + \eta \mathbf{1}_{\mathbb{C}^m}) \rightarrow \mathcal{P}_f(A_0, B_0).$$

For the last statement, CAC^* and CBC^* again have equal ranges. The proofs of Lemmas 4.6.1 and 4.6.3 use f only on the spectrum of $B_0^{-1/2} A_0 B_0^{-1/2}$, which is now contained in $(0, \infty)$, so $\mathcal{P}_f(CAC^*, CBC^*) = C \mathcal{P}_f(A, B) C^*$, and the second statement applies. $\square$

Since T is positive semidefinite, and congruence transformations are injective, we obtain a specific closed form for the above by applying the positive semidefinite functional calculus.

Lemma 4.7.4. *Let f be operator convex, let $\widetilde{W} \in L(\mathbb{C}^r, \mathcal{V})$ be injective, and let $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$ have spectral decomposition $T = \sum_\lambda \lambda P_\lambda$. Then $\text{tr}(\widetilde{W} P_\lambda \widetilde{W}^*) > 0$ for every $\lambda \in \text{spec}(T)$, and*

$$\widehat{D}_f(\widetilde{W} T \widetilde{W}^* \| \widetilde{W} (\mathbf{1}_{\mathbb{C}^r} - T) \widetilde{W}^*) = \sum_{\lambda \in \text{spec}(T)} h_f(\lambda) \text{tr}(\widetilde{W} P_\lambda \widetilde{W}^*),$$

with values in $(-\infty, +\infty]$. In particular the left-hand side is finite if and only if $h_f(T)$ is finite, and then it equals $\text{tr}(\widetilde{W} h_f(T) \widetilde{W}^)$.*

Proof. Since $\widetilde{W}$ is injective, $\text{tr}(\widetilde{W}P_\lambda\widetilde{W}^*) = \|\widetilde{W}P_\lambda\|_2^2 > 0$. Put $A = \widetilde{W}T\widetilde{W}^*$, $B = \widetilde{W}(\mathbf{1}_{\mathbb{C}^r} - T)\widetilde{W}^*$ and, for $\theta \in (0, 1]$, $T_\theta = (1 - \theta)T + \frac{\theta}{2}\mathbf{1}_{\mathbb{C}^r}$. Then T_θ and $\mathbf{1}_{\mathbb{C}^r} - T_\theta$ are invertible, so $A_\theta = \widetilde{W}T_\theta\widetilde{W}^*$ and $B_\theta = \widetilde{W}(\mathbf{1}_{\mathbb{C}^r} - T_\theta)\widetilde{W}^*$ both have range $\text{range } \widetilde{W}$. Lemma 4.7.3, applied with $C = \widetilde{W}$ to the invertible pair T_θ , $\mathbf{1}_{\mathbb{C}^r} - T_\theta$, and the commuting positive semidefinite calculus as in the proof of Theorem 4.6.7(a), give

$$\begin{aligned}\widehat{D}_f(A_\theta\|B_\theta) &= \text{tr}(\widetilde{W}\mathcal{P}_f(T_\theta, \mathbf{1}_{\mathbb{C}^r} - T_\theta)\widetilde{W}^*) = \text{tr}(\widetilde{W}h_f(T_\theta)\widetilde{W}^*) \\ &= \sum_\lambda h_f((1 - \theta)\lambda + \frac{\theta}{2}) \text{tr}(\widetilde{W}P_\lambda\widetilde{W}^*).\end{aligned}$$

As $\theta \downarrow 0$, the right-hand side tends to $\sum_\lambda h_f(\lambda) \text{tr}(\widetilde{W}P_\lambda\widetilde{W}^*)$ in $(-\infty, +\infty]$, by Lemma 4.7.2. On the other hand, $(A_\theta, B_\theta) = (1 - \theta)(A, B) + \theta(N, N)$ with $N = \frac{1}{2}\widetilde{W}\widetilde{W}^*$, and $\widehat{D}_f(N\|N) = f(1) \text{tr } N$ because $\mathcal{P}_f(X, X) = f(1)X$ for $X > 0$. Joint convexity (Lemma 4.6.4(2), valid by Lemma 4.7.3) gives

$$\widehat{D}_f(A_\theta\|B_\theta) \leq (1 - \theta)\widehat{D}_f(A\|B) + \theta f(1) \text{tr } N,$$

and lower semicontinuity gives $\liminf_{\theta \downarrow 0} \widehat{D}_f(A_\theta\|B_\theta) \geq \widehat{D}_f(A\|B)$. Hence $\widehat{D}_f(A_\theta\|B_\theta) \rightarrow \widehat{D}_f(A\|B)$ in $(-\infty, +\infty]$, and the two limits agree. $\square$

We now combine the previous two lemmas in the following formula.

Theorem 4.7.5. *Let $f : (0, \infty) \rightarrow \mathbb{R}$ be operator convex, let Ω , $V_1, \dots, V_r$ and W be as in Theorem 4.2.15, let $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$, and put $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}_{\mathbb{C}^r} - T}$.*

(a) *$\widehat{D}_f(\Phi\|\Psi) < \infty$ if and only if $h_f(T)$ is finite, that is, if and only if*

$$[f'(\infty) < \infty \text{ or } \Phi_s = 0] \quad \text{and} \quad [f(0^+) < \infty \text{ or } \Psi_s = 0].$$

(b) *In that case*

$$\widehat{D}_f(\Phi\|\Psi) = \lambda_{\max}\left(\text{Tr}_Y(W h_f(T) W^*)\right).$$

(c) *In that case, with $P_a = \mathbf{1}_{\mathbb{C}^r} - P_0 - P_1$, $\Phi_{ac} = \Psi_{TP_a}$ and $\Psi_{ac} = \Psi_{(\mathbf{1}_{\mathbb{C}^r} - T)P_a}$,*

$$\text{Tr}_Y(W h_f(T) W^*) = \text{Tr}_Y \mathcal{P}_f(J(\Phi_{ac}), J(\Psi_{ac})) + f'(\infty) \text{Tr}_Y J(\Phi_s) + f(0^+) \text{Tr}_Y J(\Psi_s),$$

where a term with an infinite coefficient is absent because the corresponding singular part vanishes.

Proof. We prove (a) and (b) together in a very similar manner to the Kubo–Ando case.

Let $u = \text{vec}(A)$ with $\|A\|_2 = 1$ and A invertible, put

$$C = \mathbf{1}_Y \otimes A^\top$$

and $\widetilde{W} = CW$, which is injective. By (4.3.1) and Theorem 4.2.2, $\Phi_{uu^*} = \widetilde{W}T\widetilde{W}^*$ and

$$\Psi_{uu^*} = \widetilde{W}(\mathbf{1}_{\mathbb{C}^r} - T)\widetilde{W}^*$$

If $h_f(T)$ is not finite, Lemma 4.7.4 gives $\widehat{D}_f(\Phi_{uu^*}\|\Psi_{uu^*}) = +\infty$, hence $\widehat{D}_f(\Phi\|\Psi) = +\infty$. If $h_f(T)$ is finite, Lemma 4.7.4 gives

$$\widehat{D}_f(\Phi_{uu^*}\|\Psi_{uu^*}) = \text{tr}(\widetilde{W} h_f(T) \widetilde{W}^*) = \text{tr}((AA^*)^\top M)$$

with $M = \text{Tr}_Y(W h_f(T) W^*)$, as in the proof of Theorem 4.6.7(a). The rest of that proof uses only the convexity and lower semicontinuity of $\rho \mapsto \widehat{D}_f(\Phi_\rho \| \Psi_\rho)$, which hold for every operator convex f by Lemma 4.7.3, and gives $\widehat{D}_f(\Phi \| \Psi) = \lambda_{\max}(M)$. The reformulation in (a) is Proposition 4.7.1.

(c)

$$h_f(T) = h_f(T)P_a + f'(\infty)P_1 + f(0^+)P_0$$

and $J(\Phi_s) = WP_1W^*$, $J(\Psi_s) = WP_0W^*$. On range P_a the operators T and $\mathbf{1}_{\mathbb{C}^r} - T$ are invertible. Lemma 4.7.3, with $C = W$ restricted to range P_a , gives

$$\mathcal{P}_f(WTP_aW^*, W(\mathbf{1}_{\mathbb{C}^r} - T)P_aW^*) = W h_f(T)P_a W^*.$$

□

Part (c) is the channel analogue of the classical formula [82]

$$D_f(\mathbb{P} \| \mathbb{Q}) = \int f\left(\frac{d\mathbb{P}_{\text{ac}}}{d\mathbb{Q}_{\text{ac}}}\right) d\mathbb{Q}_{\text{ac}} + f'(\infty)\mathbb{P}_s(\mathcal{S}) + f(0)\mathbb{Q}_s(\mathcal{S}),$$

where $\mathbb{P} = \mathbb{P}_{\text{ac}} + \mathbb{P}_s$ and $\mathbb{Q} = \mathbb{Q}_{\text{ac}} + \mathbb{Q}_s$, with $\mathbb{P}_{\text{ac}}$ and $\mathbb{Q}_{\text{ac}}$ mutually absolutely continuous, $\mathbb{P}_s \perp \mathbb{Q}$ and $\mathbb{Q}_s \perp \mathbb{P}$. Theorem 4.7.5 contains Theorem 4.5.13(a), where both boundary values are finite, and Theorem 4.6.7(a). It also contains the mirror case discussed after Corollary 4.6.12. When both boundary values are infinite, a case not covered by those results, part (a) requires $\Phi_s = \Psi_s = 0$, that is, $\text{range } J(\Phi) = \text{range } J(\Psi)$. The operator-valued counterpart of part (a) is [101, Proposition 4.11], which characterises when the divergence of [101, Definition 3.1] has dense domain and gives sufficient conditions for it to be bounded; when $f(0^+) = f'(\infty) = +\infty$, a dense domain requires Φ and Ψ to be mutually absolutely continuous, that is $\Phi \ll \Psi$ and $\Psi \ll \Phi$ [101, Proposition 4.11(3), Definition 4.10], and boundedness holds if each is dominated by a multiple of the other, which in our setting is again the condition $\text{range } J(\Phi) = \text{range } J(\Psi)$, by Proposition 4.7.1. In finite dimensions the two are directly related. In the Stinespring representation $Y \mapsto S^*(Y \otimes \mathbf{1}_{\mathcal{Z}})S$ of Ω^* given by Theorem 4.2.5, the Radon–Nikodym derivatives of Φ^* and Ψ^* are $\mathbf{1}_Y \otimes T^\top$ and $\mathbf{1}_Y \otimes (\mathbf{1}_{\mathbb{C}^r} - T)^\top$, so, when $h_f(T)$ is finite, the divergence of [101, Definition 3.1] of the adjoint maps is $S^*(\mathbf{1}_Y \otimes h_f(T)^\top)S = \Psi_{h_f(T)}^*(\mathbf{1}_Y)$, the transpose of $\text{Tr}_Y(W h_f(T) W^*)$. Part (b) thus identifies $\widehat{D}_f(\Phi \| \Psi)$ with the largest eigenvalue of Okayasu’s operator-valued divergence of Φ^* and Ψ^* ; to our knowledge this scalar identification, and the closed form when both boundary values are infinite, are new. Quantitatively, $\epsilon \mathbf{1}_{\mathbb{C}^r} \leq T \leq (1 - \epsilon) \mathbf{1}_{\mathbb{C}^r}$ is equivalent to

$$\max\{D_{\max}(\Phi \| \Psi), D_{\max}(\Psi \| \Phi)\} \leq \log \frac{1 - \epsilon}{\epsilon}$$

by Lemma 4.6.5, so the natural parameter for such f is the Thompson distance between Φ and Ψ .

4.7.4 Continuity in the derivative

We now show continuity results in this more general setting described in the previous theorem.

Theorem 4.7.6. *Let f be operator convex and let Ω and W be as in Theorem 4.2.15. Write $\mathcal{D}(T) = \widehat{D}_f(\Psi_T \| \Psi_{\mathbf{1}_{\mathbb{C}^r} - T})$ for $0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}$.*

(a) *Let $I \subseteq [0, 1]$ be a compact interval on which h_f is finite. If $\text{spec}(T_1), \text{spec}(T_2) \subseteq I$, then*

$$|\mathcal{D}(T_1) - \mathcal{D}(T_2)| \leq \|\Omega^*(\mathbf{1}_Y)\|_\infty \omega_{h_f}^I(\|T_1 - T_2\|_\infty).$$

(b) For every T and every $\lambda \in \text{spec}(T)$,

$$\mathcal{D}(T) \geq \nu(\Omega)^2(h_f(\lambda) - m_f) + \min\{m_f, 0\} \|\Omega^*(\mathbf{1}_Y)\|_\infty.$$

(c) The map $T \mapsto \mathcal{D}(T)$ is continuous from $\{0 \leq T \leq \mathbf{1}_{\mathbb{C}^r}\}$ to $(-\infty, +\infty]$.

Proof. (a) By Theorem 4.7.5(b) and Lemma 4.4.10, as in the proof of Theorem 4.6.7(b).

(b) We may assume $\mathcal{D}(T) < \infty$. Let v be a unit eigenvector of T for λ . Then

$$h_f(T) \geq m_f \mathbf{1}_{\mathbb{C}^r} + (h_f(\lambda) - m_f)vv^*$$

so, since conjugation by W and the partial trace are positive,

$$M = \text{Tr}_Y(W h_f(T) W^*) \geq m_f \text{Tr}_Y(WW^*) + (h_f(\lambda) - m_f) \text{Tr}_Y(Wvv^*W^*).$$

With $B = \sum_k v_k V_k$ we have $Wv = \text{vec}(B)$ and $\text{Tr}_Y(\text{vec}(B) \text{vec}(B)^*) = (B^*B)^\top$, whose largest eigenvalue is $\|B\|_\infty^2 \geq \nu(\Omega)^2$. Moreover $\text{Tr}_Y(WW^*) = \Omega^*(\mathbf{1}_Y)^\top$. The inequality

$$\lambda_{\max}(X + Y) \geq \lambda_{\max}(X) + \lambda_{\min}(Y)$$

and Theorem 4.7.5(b) give the claim.

(c) Let $\mathcal{F}_f = \{t \in [0, 1] : h_f(t) < \infty\}$, which is $[0, 1]$ minus at most the endpoints. If $h_f(T)$ is finite, choose a compact interval $I \subseteq \mathcal{F}_f$ whose relative interior in $[0, 1]$ contains $\text{spec}(T)$. By Weyl's inequality, $\text{spec}(T') \subseteq I$ for T' close to T , and (a) together with Lemma 4.4.8 gives continuity at T . If $h_f(T)$ is not finite, then $\mathcal{D}(T) = +\infty$, and some $\lambda \in \text{spec}(T)$ is an endpoint with $h_f(\lambda) = +\infty$. Every T' close to T has an eigenvalue λ' close to λ , and $h_f(\lambda') \rightarrow +\infty$ by Lemma 4.7.2. So $\mathcal{D}(T') \rightarrow +\infty$ by (b), since $\nu(\Omega) > 0$. $\square$

Thus the ARN derivative is a coordinate in which every maximal f -divergence is continuous, finite or not. This stands in contrast with the completely bounded trace norm (Corollary 4.3.3). Part (b) also shows how fast the divergence grows with the max-relative entropy, since $\lambda_{\max}(T) = g(D_{\max}(\Phi\|\Psi))$ with $g(x) = e^x/(1 + e^x)$ by Lemma 4.6.5.

Corollary 4.7.7. *For every operator convex f and all $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ with $\Omega = \Phi + \Psi \neq 0$, with the conventions $g(-\infty) = 0$ and $g(+\infty) = 1$, and with the right-hand side read as $+\infty$ where h_f is $+\infty$,*

$$\widehat{D}_f(\Phi\|\Psi) \geq \nu(\Omega)^2 h_f(g(D_{\max}(\Phi\|\Psi))) - |m_f|(\nu(\Omega)^2 + \|\Omega^*(\mathbf{1}_Y)\|_\infty),$$

and the same holds with $h_f(1 - g(D_{\max}(\Psi\|\Phi)))$. For the Belavkin–Staszewski relative entropy, $h_f(g(x)) = g(x)x \geq x/2$ for $x \geq 0$, and $|m_f| < 0.28$, so

$$\widehat{D}_{\text{BS}}(\Phi\|\Psi) \geq \frac{1}{2}\nu(\Omega)^2 D_{\max}(\Phi\|\Psi) - 0.28(\nu(\Omega)^2 + \|\Omega^*(\mathbf{1}_Y)\|_\infty)$$

whenever $D_{\max}(\Phi\|\Psi) \geq 0$.

Proof. Apply Theorem 4.7.6(b) with $\lambda = \lambda_{\max}(T)$, respectively $\lambda = \lambda_{\min}(T) = 1 - g(D_{\max}(\Psi\|\Phi))$, and use

$$-\nu(\Omega)^2 m_f + \min\{m_f, 0\} \|\Omega^*(\mathbf{1}_Y)\|_\infty \geq -|m_f|(\nu(\Omega)^2 + \|\Omega^*(\mathbf{1}_Y)\|_\infty).$$

For $f(t) = t \log t$, $h_f(t) = t \log \frac{t}{1-t}$, and its minimum on $[0, 1]$ is $-0.2785\dots$ $\square$

4.8 Continuity in the completely bounded trace norm

All estimates so far are stated in terms of ARN derivatives. Fault-tolerance thresholds, however, refer to additive worst-case errors, for which the completely bounded trace (diamond) norm is the standard measure [80]. In this section we show that, for a fixed Ω , the operator norm distance between derivatives and the completely bounded trace norm distance between the corresponding submaps are equivalent, with constants that do not depend on any dimension (Theorem 4.8.4). The continuity estimates in T of Sections 4.4–4.7 therefore hold in the completely bounded trace norm on the splittings of a fixed map (Corollary 4.8.6), and the impossibility example of Corollary 4.3.3 is identified as the case in which the constant degenerates. Pairs with different sums are treated in Section 4.10.

Let $\Omega \in \text{CP}(\mathcal{X}, \mathcal{Y})$ be nonzero with Kraus decomposition $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$, $V_1, \dots, V_r$ linearly independent, and $W = [\text{vec}(V_1) \cdots \text{vec}(V_r)]$. Recall from Section 4.2 the quantity $\nu(\Omega)$, the smallest operator norm of a normalised linear combination of the Kraus operators. The minimum is attained and positive, since the unit sphere is compact and the V_k are linearly independent. We first record the intrinsic meaning of the operator norm of a derivative.

Lemma 4.8.1. *For Hermitian $K \in \text{L}(\mathbb{C}^r)$ and every unitarily invariant norm $\|\cdot\|$,*

$$\|K\| = \|J(\Omega)^{+1/2} J(\Psi_K) J(\Omega)^{+1/2}\|,$$

where $J(\Omega)^{+1/2}$ is the square root of the Moore–Penrose inverse of $J(\Omega)$. Moreover, for $d \geq 0$ we have $\|K\|_\infty \leq d$ if and only if $-d\Omega \leq \Psi_K \leq d\Omega$ in the completely positive order. In particular, if $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2 = \Omega$ with derivatives T_1, T_2 , then $\|T_1 - T_2\|_\infty$ is the smallest $d \geq 0$ such that $-d\Omega \leq \Phi_1 - \Phi_2 \leq d\Omega$.

Proof. Write $W = UG^{1/2}$ with $G = W^*W$ invertible and U an isometry (polar decomposition). Then $J(\Omega) = WW^* = UGU^*$, so $J(\Omega)^{+1/2} = UG^{-1/2}U^*$, and

$$J(\Psi_K) = WKW^* = UG^{1/2}KG^{1/2}U^*.$$

Hence $J(\Omega)^{+1/2}J(\Psi_K)J(\Omega)^{+1/2} = UKU^*$, which has the same singular values as K . For the second statement, $d\Omega \pm \Psi_K = \Psi_{d\mathbf{1}_{\mathbb{C}^r} \pm K}$ has Choi matrix $W(d\mathbf{1}_{\mathbb{C}^r} \pm K)W^*$. By Choi’s theorem [28], both maps are completely positive if and only if $W(d\mathbf{1}_{\mathbb{C}^r} \pm K)W^* \geq 0$, and conjugating by $W^+ = (W^*W)^{-1}W^*$ and by W shows that this holds if and only if $d\mathbf{1}_{\mathbb{C}^r} \pm K \geq 0$, that is, $\|K\|_\infty \leq d$. The last statement follows with $K = T_1 - T_2$, since $\Phi_1 - \Phi_2 = \Psi_{T_1 - T_2}$. $\square$

One of the principal pains with continuity results for channel divergences is dimensional constraints. The next lemma shows that in the same-sum case, we have a natural candidate which depends on the sum being considered. The primary application of this will be control from below of quantum channel divergences which consider two different decompositions of the same completely positive map.

Lemma 4.8.2. *Let Ω be as above and $J = J(\Omega)$, with $\lambda_{\min}^+(J)$ its smallest nonzero eigenvalue.*

1. *We have*

$$\nu(\Omega)^2 = \min \{ \|\text{vec}^{-1}(w)\|_\infty^2 / \langle w, J^+ w \rangle : w \in \text{range}(J) \setminus \{0\} \},$$

where J^+ is the Moore–Penrose inverse of J . In particular $\nu(\Omega)$ does not depend on the Kraus decomposition.

2. $\lambda_{\min}^+(J)/\min\{\dim \mathcal{X}, \dim \mathcal{Y}\} \leq \nu(\Omega)^2 \leq \lambda_{\min}^+(J)$.
3. If $\Omega' \in \mathcal{CP}(\mathcal{X}, \mathcal{Y})$ satisfies $\Omega' \geq \Omega$ in the completely positive order and $\text{range } J(\Omega') = \text{range } J(\Omega)$, then $\nu(\Omega') \geq \nu(\Omega)$.

Proof. (1) By linearity of vec , $\sum_k c_k V_k = \text{vec}^{-1}(Wc)$. As c ranges over the unit sphere of $\mathbb{C}^r$, the vector $w = Wc$ ranges over $\text{range}(W) = \text{range}(WW^*) = \text{range}(J)$, and $c = W^+w$ with $W^+ = (W^*W)^{-1}W^*$, so that

$$\|c\|_2^2 = \langle w, (W^+)^* W^+ w \rangle = \langle w, J^+ w \rangle,$$

because

$$(W^+)^* W^+ = W(W^*W)^{-2}W^* = (WW^*)^+.$$

Hence $\nu(\Omega)^2$ is the minimum of $\|\text{vec}^{-1}(w)\|_\infty^2$ over $w \in \text{range}(J)$ with $\langle w, J^+ w \rangle = 1$, and the formula follows by homogeneity. It involves only $J(\Omega)$.

(2) For $w \in \text{range}(J)$ we have

$$\|w\|_2^2 / \lambda_{\max}(J) \leq \langle w, J^+ w \rangle \leq \|w\|_2^2 / \lambda_{\min}^+(J),$$

while

$$\frac{\|w\|_2^2}{\min\{\dim \mathcal{X}, \dim \mathcal{Y}\}} \leq \frac{\|w\|_2^2}{\text{rank}(\text{vec}^{-1}(w))} \leq \|\text{vec}^{-1}(w)\|_\infty^2 \leq \|\text{vec}^{-1}(w)\|_2^2 = \|w\|_2^2.$$

The lower bound follows from (1). For the upper bound take w to be an eigenvector of J for the eigenvalue $\lambda_{\min}^+(J)$.

(3) Put $\mathcal{R} = \text{range } J(\Omega) = \text{range } J(\Omega')$. By Choi's theorem [28], $J(\Omega') \geq J(\Omega)$; both operators are positive definite on $\mathcal{R}$ and vanish on $\mathcal{R}^\perp$, so operator monotonicity of inversion on $\mathcal{R}$ gives $J(\Omega')^+ \leq J(\Omega)^+$, hence $\langle w, J(\Omega')^+ w \rangle \leq \langle w, J(\Omega)^+ w \rangle$ for $w \in \mathcal{R}$, and (1) gives $\nu(\Omega')^2 \geq \nu(\Omega)^2$. $\square$

Remark 4.8.3 (Physical interpretation). Since $\Omega^c(\rho) = \sum_{k,l} \text{tr}(V_k \rho V_l^*) e_k e_l^*$ (proof of Theorem 4.2.5), $\langle u, \Omega^c(\rho) u \rangle = \text{tr}(V_u \rho V_u^*)$ with $V_u = \sum_k \bar{u}_k V_k$, whose maximum over states ρ is $\|V_u\|_\infty^2$. Hence

$$\nu(\Omega)^2 = \min_{\|u\|_2=1} \max_{\rho} \langle u, \Omega^c(\rho) u \rangle,$$

the excitability of the least excitable environment direction. As $\Psi_{\bar{u}\bar{u}^*} = V_u(\cdot)V_u^*$, changing T by $\bar{u}\bar{u}^*$ for a minimising u , a change of norm 1, changes the submaps by only $\nu(\Omega)^2$ in $|||\cdot|||_1$: this is the mechanism behind Corollary 4.3.3, where $\nu(\Omega) \rightarrow 0$ as $\theta \rightarrow 0$. Noise helps: if Φ and Ψ each contain a fraction p of the completely depolarising channel, whose Choi matrix is $\mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}/n_{\mathcal{Y}}$ with $n_{\mathcal{Y}} = \dim \mathcal{Y}$, then $J(\Omega) \geq (2p/n_{\mathcal{Y}})\mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}$, and Lemma 4.8.2(2) gives $\nu(\Omega)^2 \geq 2p/(n_{\mathcal{Y}} \min\{\dim \mathcal{X}, n_{\mathcal{Y}}\})$; in particular Theorem 4.12.4(a) applies.

Theorem 4.8.4. Let $\Omega, V_1, \dots, V_r$ be as above. For every Hermitian $K \in \mathcal{L}(\mathbb{C}^r)$,

$$\nu(\Omega)^2 \|K\|_\infty \leq |||\Psi_K|||_1 \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \|K\|_\infty.$$

The upper bound is attained for $K = \mathbf{1}_{\mathbb{C}^r}$, and the lower bound is attained for the completely dephasing map $\Omega(X) = \sum_{i=1}^n E_{ii} X E_{ii}$ on $\mathbb{C}^n$ and $K = \frac{1}{n} \mathbf{j} \mathbf{j}^*$, where $\mathbf{j}$ is the all-ones vector, one has $\nu(\Omega)^2 = \frac{1}{n}$, $\|K\|_\infty = 1$ and $|||\Psi_K|||_1 = \frac{1}{n}$.

Proof. The upper bound is the proof of Lemma 4.4.1 with $\Phi = \Omega$, which uses only that K is Hermitian, together with $\|\Omega\|_1 = \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty$.

For the lower bound let $v \in \mathbb{C}^r$ be a unit eigenvector of K with $Kv = \lambda v$ and $|\lambda| = \|K\|_\infty$. On the subspace $\mathcal{S} = \text{span}\{V_1, \dots, V_r\}$ of $L(\mathcal{X}, \mathcal{Y})$ define the linear functional $\varphi_0(\sum_k c_k V_k) = \sum_k c_k \overline{v_k}$, which is well defined by linear independence. By the Cauchy–Schwarz inequality and the definition of $\nu(\Omega)$,

$$|\varphi_0(\sum_k c_k V_k)| \leq \|c\|_2 \|v\|_2 = \|c\|_2 \leq \nu(\Omega)^{-1} \|\sum_k c_k V_k\|_\infty,$$

so φ_0 has norm at most $\nu(\Omega)^{-1}$ with respect to the operator norm. By the Hahn–Banach theorem it extends to a functional on $L(\mathcal{X}, \mathcal{Y})$ of the same norm, and by the duality between the operator norm and the trace norm [127, Section 1.1.3] there is $Z \in L(\mathcal{X}, \mathcal{Y})$ with $\|Z\|_1 \leq \nu(\Omega)^{-1}$ and $\text{tr}(Z^* V_k) = \overline{v_k}$ for $k = 1, \dots, r$.

Let $Z = U|Z|$ be a polar decomposition and put $A = |Z|^{1/2} \in L(\mathcal{X})$ and $M = U|Z|^{1/2} \in L(\mathcal{X}, \mathcal{Y})$, so that $Z = MA^*$, $Z^* = AM^*$ and $\|A\|_2^2 = \|M\|_2^2 = \text{tr}|Z| = \|Z\|_1$. Consider the density operator $\rho = \text{vec}(A) \text{vec}(A)^* / \|A\|_2^2$ on $\mathcal{X} \otimes \mathcal{X}$ and the operator $N = \text{vec}(M) \text{vec}(M)^* / \|M\|_2^2$ on $\mathcal{Y} \otimes \mathcal{X}$, which satisfies $\|N\|_\infty = 1$. Since $\text{vec}(A) = (\mathbf{1}_{\mathcal{X}} \otimes A^\top) \text{vec}(\mathbf{1}_{\mathcal{X}})$ and $(\mathbf{1}_{\mathcal{Y}} \otimes A^\top) \text{vec}(V_k) = \text{vec}(V_k A)$ [127, Section 1.1.2], the identity $J(\Psi_K) = WKW^*$ from the proof of Theorem 4.2.2 gives

$$(\Psi_K \otimes \mathbf{1}_{L(\mathcal{X})})(\text{vec}(A) \text{vec}(A)^*) = \sum_{k,l=1}^r K_{kl} \text{vec}(V_k A) \text{vec}(V_l A)^*,$$

and therefore

$$\begin{aligned} & \text{tr} \left(N (\Psi_K \otimes \mathbf{1}_{L(\mathcal{X})})(\text{vec}(A) \text{vec}(A)^*) \right) \|M\|_2^2 \\ &= \sum_{k,l} K_{kl} \langle \text{vec}(M), \text{vec}(V_k A) \rangle \overline{\langle \text{vec}(M), \text{vec}(V_l A) \rangle}. \end{aligned}$$

Here

$$\begin{aligned} \langle \text{vec}(M), \text{vec}(V_k A) \rangle &= \text{tr}(M^* V_k A) = \text{tr}(V_k A M^*) \\ &= \text{tr}(V_k Z^*) = \text{tr}(Z^* V_k) = \overline{v_k}, \end{aligned}$$

so the right-hand side equals $\sum_{k,l} K_{kl} \overline{v_k} v_l = \langle v, Kv \rangle = \lambda$. Since $\|N\|_\infty = 1$, the duality between the trace norm and the operator norm gives

$$\|(\Psi_K \otimes \mathbf{1}_{L(\mathcal{X})})(\rho)\|_1 \geq \frac{|\lambda|}{\|A\|_2^2 \|M\|_2^2} = \frac{\|K\|_\infty}{\|Z\|_1^2} \geq \nu(\Omega)^2 \|K\|_\infty,$$

and the left-hand side is at most $\|\Psi_K\|_1$ because ρ is a density operator.

For the example, $\|\sum_i c_i E_{ii}\|_\infty = \max_i |c_i| \geq n^{-1/2}$ for unit c , with equality for $c = n^{-1/2} \mathbf{j}$, so $\nu(\Omega)^2 = \frac{1}{n}$; and $\Psi_K(X) = \frac{1}{n} \sum_{i,j} E_{ii} X E_{jj} = \frac{1}{n} X$, so $\|\Psi_K\|_1 = \frac{1}{n}$, while K is a rank-one projection. $\square$

Remark 4.8.5. By Lemma 4.8.2(2), Theorem 4.8.4 implies

$$\|K\|_\infty \leq \frac{\min\{\dim \mathcal{X}, \dim \mathcal{Y}\}}{\lambda_{\min}^+(J(\Omega))} \|\Psi_K\|_1,$$

which is never weaker than the operator norm consequence of Proposition 4.4.9, and is attained, up to the value of $\nu(\Omega)^2$, in the dephasing example. In practice $\lambda_{\min}^+(J(\Omega))$ is computable from the Choi matrix, whereas $\nu(\Omega)^2$ is the minimum of a nonconvex function

on a sphere; Lemma 4.8.2(2) provides the computable substitute. An earlier lower bound of this kind, with the dimensional factor $(\dim \mathcal{H})^{-2}$ in place of $\nu(\Omega)^2$, is due to Raginsky [111, eq. (17)], for maps on $B(\mathcal{H})$ and derivatives taken with respect to a fixed reference channel.

Corollary 4.8.6. *Let $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2 = \Omega \neq 0$ with $\Phi_i, \Psi_i \in \text{CP}(\mathcal{X}, \mathcal{Y})$, let T_i be the ARN derivative of Φ_i with respect to Ω , in one fixed Kraus decomposition of Ω , and put*

$$\delta = |||\Phi_1 - \Phi_2|||_1 / \nu(\Omega)^2 = |||\Psi_1 - \Psi_2|||_1 / \nu(\Omega)^2.$$

Then $\|T_1 - T_2\|_\infty \leq \delta$, and consequently

$$|\hat{F}(\Phi_1, \Psi_1) - \hat{F}(\Phi_2, \Psi_2)| \leq \frac{\|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty}{\nu(\Omega)} |||\Phi_1 - \Phi_2|||_1^{1/2},$$

while

$$|||\Phi_1 \sigma \Psi_1 - \Phi_2 \sigma \Psi_2|||_1 \leq \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \omega_{g_\sigma}(\delta)$$

for every Kubo–Ando mean σ . More generally, each of the estimates of Theorems 4.4.5, 4.4.6, 4.5.5, 4.5.13(b), 4.6.7(b) and 4.7.6(a) and of Corollaries 4.6.11 and 4.6.12 remains valid, under its hypotheses, with $\|T_1 - T_2\|_\infty$ replaced by $\min\{\delta, 1\}$.

Proof. Since $\Phi_1 - \Phi_2 = \Psi_{T_1 - T_2}$ and $\Psi_1 - \Psi_2 = -(\Phi_1 - \Phi_2)$, Theorem 4.8.4 gives $\|T_1 - T_2\|_\infty \leq \delta$, and $\|T_1 - T_2\|_\infty \leq 1$ because $0 \leq T_i \leq \mathbf{1}_{\text{Cr}}$. The right-hand side of each cited estimate is a nondecreasing function of $\|T_1 - T_2\|_\infty \in [0, 1]$: the moduli $\omega_g, \omega_g^{(\epsilon)}$ and ω_g^I are suprema over sets that increase with δ , the functions $\delta \mapsto \delta^{1/2}$ and $\delta \mapsto \delta^\alpha + \delta^{1-\alpha}$, $\alpha \in (0, 1)$, are increasing, the bound of Corollary 4.6.12 is linear in δ , and that of Corollary 4.6.11 has derivative $\log \frac{4}{\epsilon\delta} - 1 + \frac{1}{\epsilon} > 0$ for $\delta \in (0, 1]$. The displayed bound is Theorem 4.4.6 with $\|T_1 - T_2\|_\infty^{1/2} \leq \delta^{1/2}$. $\square$

Pairs with a Choi floor

Corollary 4.8.6 and Theorem 4.12.4(a) give Hölder continuity of order $\frac{1}{2}$. If the Choi matrices of both maps are bounded below, the fidelity is in fact Lipschitz continuous, with a constant of the sharp order $\lambda_{\min}^{-1}$. We first record the comparison between the operator norm of a Choi matrix and the completely bounded trace norm.

Lemma 4.8.7. *Let $\Delta: L(\mathcal{X}) \rightarrow L(\mathcal{Y})$ be Hermitian-preserving and $n = \dim \mathcal{X}$. Then*

$$\|J(\Delta)\|_\infty \leq \frac{n}{2} |||\Delta|||_1 + \frac{1}{2} |\text{tr } J(\Delta)| \leq n |||\Delta|||_1.$$

In particular $\|J(\Delta)\|_\infty \leq \frac{n}{2} |||\Delta|||_1$ if $\text{tr} \circ \Delta = 0$, for instance if Δ is the difference of two channels. Both constants are attained for every n : the second by $\Delta = \epsilon \mathbf{1}_{L(\mathcal{X})}$, $\epsilon > 0$, and the first, when $\mathcal{Y} = \mathcal{X}$, by $\Delta = \mathbf{1}_{L(\mathcal{X})} - V(\cdot)V^$ with V unitary and $\text{tr } V = 0$.*

Proof. Let $\tau = \frac{1}{n} \text{vec}(\mathbf{1}_{\mathcal{X}}) \text{vec}(\mathbf{1}_{\mathcal{X}})^*$, a density operator on $\mathcal{X} \otimes \mathcal{X}$, so that $J(\Delta) = n(\Delta \otimes \mathbf{1}_{L(\mathcal{X})})(\tau)$ [127, Section 2.2.2]. The operator $Z = (\Delta \otimes \mathbf{1}_{L(\mathcal{X})})(\tau)$ is Hermitian, with $\|Z\|_1 \leq |||\Delta|||_1$ and $\text{tr } Z = \frac{1}{n} \text{tr } J(\Delta)$. Writing $Z = Z_+ - Z_-$ with $Z_\pm \geq 0$ and $Z_+ Z_- = 0$, we get

$$\|Z\|_\infty = \max\{\|Z_+\|_\infty, \|Z_-\|_\infty\} \leq \max\{\text{tr } Z_+, \text{tr } Z_-\} = \frac{1}{2}(\|Z\|_1 + |\text{tr } Z|),$$

which is the first inequality after multiplication by n . The second follows from $|\text{tr } Z| \leq \|Z\|_1$. If $\text{tr} \circ \Delta = 0$, then $\text{tr } J(\Delta) = \sum_i \text{tr } \Delta(E_{ii}) = 0$. For $\Delta = \epsilon \mathbf{1}_{L(\mathcal{X})}$, $J(\Delta) =$

$\epsilon \text{vec}(\mathbf{1}_{\mathcal{X}}) \text{vec}(\mathbf{1}_{\mathcal{X}})^*$ has norm $n\epsilon = n \|\Delta\|_1$. For $\Delta = \mathbf{1}_{\mathcal{L}(\mathcal{X})} - V(\cdot)V^*$ with $\text{tr } V = 0$, the vectors $\text{vec}(\mathbf{1}_{\mathcal{X}})$ and $\text{vec}(V)$ are orthogonal, since $\langle \text{vec}(\mathbf{1}_{\mathcal{X}}), \text{vec}(V) \rangle = \text{tr } V$, and both have squared norm n . Hence $\|J(\Delta)\|_{\infty} = n$, and Z is the difference of two orthogonal rank-one projections, so $\|\Delta\|_1 \geq \|Z\|_1 = 2$. As $\|\Delta\|_1 \leq 2$ for a difference of channels, $\|J(\Delta)\|_{\infty} = \frac{n}{2} \|\Delta\|_1$. $\square$

Theorem 4.8.8. *Let $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ with $\lambda = \min\{\lambda_{\min}(J(\Phi)), \lambda_{\min}(J(\Psi))\} > 0$, let $\Phi', \Psi' \in \text{CP}(\mathcal{X}, \mathcal{Y})$, and put*

$$\eta = \max\{\|J(\Phi') - J(\Phi)\|_{\infty}, \|J(\Psi') - J(\Psi)\|_{\infty}\}.$$

(a) *For every operator connection σ , $J(\Phi')\sigma J(\Psi') \leq (1 + \frac{\eta}{\lambda})J(\Phi)\sigma J(\Psi)$, and if $\eta \leq \lambda$ also $J(\Phi')\sigma J(\Psi') \geq (1 - \frac{\eta}{\lambda})J(\Phi)\sigma J(\Psi)$. Consequently $\|\Phi'\sigma\Psi' - \Phi\sigma\Psi\|_1 \leq \frac{\eta}{\lambda} \|\Phi\sigma\Psi\|_1$ if $\eta \leq \lambda$.*

(b) *$\hat{F}(\Phi', \Psi') \leq (1 + \frac{\eta}{\lambda})\hat{F}(\Phi, \Psi)$, and if $\eta < \lambda$,*

$$(1 - \frac{\eta}{\lambda})\hat{F}(\Phi, \Psi) \leq \hat{F}(\Phi', \Psi'), \quad |\log \hat{F}(\Phi', \Psi') - \log \hat{F}(\Phi, \Psi)| \leq -\log(1 - \frac{\eta}{\lambda}).$$

In particular $|\hat{F}(\Phi', \Psi') - \hat{F}(\Phi, \Psi)| \leq \frac{\eta}{\lambda} \hat{F}(\Phi, \Psi)$. The same bounds hold for $\hat{Q}_{\alpha}$, $\alpha \in (0, 1)$. If $\Psi' = \Psi$, the factors $1 \pm \frac{\eta}{\lambda}$ for $\hat{F}$ may be replaced by $(1 \pm \frac{\eta}{\lambda})^{1/2}$.

(c) *With $d_{\text{cb}} = \max\{\|\Phi' - \Phi\|_1, \|\Psi' - \Psi\|_1\}$ we have $\eta \leq \dim \mathcal{X} d_{\text{cb}}$, and $\eta \leq \frac{1}{2} \dim \mathcal{X} d_{\text{cb}}$ if all four maps are channels.*

Proof. (a) Since $J(\Phi) \geq \lambda \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}$,

$$(1 - \frac{\eta}{\lambda})J(\Phi) \leq J(\Phi) - \eta \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}} \leq J(\Phi') \leq J(\Phi) + \eta \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}} \leq (1 + \frac{\eta}{\lambda})J(\Phi),$$

and likewise for Ψ . Monotonicity of σ in both arguments and $(cA)\sigma(cB) = c(A\sigma B)$ for $c \geq 0$ [79] give the two operator inequalities. Put $\Xi = \Phi\sigma\Psi$ and $G = J(\Xi) = J(\Phi)\sigma J(\Psi)$ (Theorem 4.2.15 and Remark 4.2.16); we may assume $\Xi \neq 0$. Then $-\frac{\eta}{\lambda}G \leq J(\Phi'\sigma\Psi') - G \leq \frac{\eta}{\lambda}G$, so the range of $J(\Phi'\sigma\Psi') - G$ lies in that of G . With respect to a Kraus decomposition of Ξ by linearly independent operators, $\Phi'\sigma\Psi' - \Xi = \Psi_K$ for a Hermitian K , and Lemma 4.8.1 gives $\|K\|_{\infty} \leq \frac{\eta}{\lambda}$. The upper bound of Theorem 4.8.4, applied to Ξ , gives

$$\|\Psi_K\|_1 \leq \|\Xi^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty} \|K\|_{\infty} = \|\Xi\|_1 \|K\|_{\infty}.$$

(b) By Proposition 4.3.1, $\hat{F} = \lambda_{\min} \circ \text{Tr}_{\mathcal{Y}}$ applied to the geometric mean of the Choi matrices, and $Z \mapsto \lambda_{\min}(\text{Tr}_{\mathcal{Y}} Z)$ is monotone and positively homogeneous; apply (a) with $\sigma = \#$. By Lemma 4.5.3 the same argument applies to $\hat{Q}_{\alpha}$, with the weighted geometric mean. Moreover

$$J(\Phi)\#J(\Psi) \geq (\lambda \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}})\#(\lambda \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}) = \lambda \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}},$$

so $\hat{F}(\Phi, \Psi) \geq \lambda \dim \mathcal{Y} > 0$ and the logarithmic form is defined; it follows from $\log(1+x) \leq -\log(1-x)$ for $0 \leq x < 1$. If $\Psi' = \Psi$, use $(cA)\#B = c^{1/2}(A\#B)$ instead.

(c) This is Lemma 4.8.7 applied to $\Phi' - \Phi$ and $\Psi' - \Psi$. $\square$

Remark 4.8.9. Theorem 4.8.8 complements Theorem 4.12.4(a). If $J(\Phi), J(\Psi) \geq \lambda \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}$, then $J(\Omega) \geq 2\lambda \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}$, and Lemma 4.8.2(2) gives $\nu(\Omega)^2 \geq 2\lambda / \min\{\dim \mathcal{X}, \dim \mathcal{Y}\}$. For channels, Theorem 4.12.4(a) therefore bounds $|\hat{F}(\Phi, \Psi) - \hat{F}(\Phi', \Psi')|$ by

$$5.06 (\min\{\dim \mathcal{X}, \dim \mathcal{Y}\} d_{\text{cb}} / (2\lambda))^{1/2},$$

whereas Theorem 4.8.8 gives $\frac{\dim \mathcal{X}}{2\lambda} d_{\text{cb}} \widehat{F}(\Phi, \Psi) \leq \frac{\dim \mathcal{X}}{2\lambda} d_{\text{cb}}$. The second bound is smaller as soon as

$$d_{\text{cb}} \leq 51 \lambda \min\{\dim \mathcal{X}, \dim \mathcal{Y}\} / (\dim \mathcal{X})^2;$$

it is linear rather than of order $\frac{1}{2}$ in d_{cb} . Theorem 4.8.8 also controls the relative error, and so the error of $-\log \widehat{F}$, which is the quantity entering converse bounds.

Corollary 4.8.10 (Noisy gates). *Let $\dim \mathcal{X} = \dim \mathcal{Y} = n$, let $\mathcal{D}_p(\rho) = (1-p)\rho + p \text{tr}(\rho) \mathbf{1}_{\mathcal{Y}}/n$ be the depolarising channel with $p \in (0, 1]$, let Φ_0, Ψ_0 be channels (for instance unitary gates), and put $\Phi = \mathcal{D}_p \circ \Phi_0$, $\Psi = \mathcal{D}_q \circ \Psi_0$ and $\bar{p} = \min\{p, q\}$. For all channels Φ', Ψ' with*

$$d_{\text{cb}} = \max\{\|\Phi' - \Phi\|_1, \|\Psi' - \Psi\|_1\} < 2\bar{p}/n^2,$$

$$|\widehat{F}(\Phi', \Psi') - \widehat{F}(\Phi, \Psi)| \leq \frac{n d_{\text{cb}}}{2\lambda} \widehat{F}(\Phi, \Psi),$$

$$|\log \widehat{F}(\Phi', \Psi') - \log \widehat{F}(\Phi, \Psi)| \leq -\log \left(1 - \frac{n d_{\text{cb}}}{2\lambda}\right),$$

with $\lambda = \bar{p}/n$, so that the first factor is $n^2 d_{\text{cb}} / (2\bar{p})$, and $2d_{\text{cb}}/\bar{p}$ for a qubit. For $n = 2^k$ and k qubits with local noise $\mathcal{D}_p^{\otimes k}$ and $\mathcal{D}_q^{\otimes k}$ in place of $\mathcal{D}_p$ and $\mathcal{D}_q$, the same holds with $\lambda = (\bar{p}/2)^k$ and $d_{\text{cb}} < 2\lambda/n$.

Proof. We have

$$J(\mathcal{D}_p \circ \Phi_0) = (1-p)J(\Phi_0) + \frac{p}{n} \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}} \geq \frac{p}{n} \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}},$$

because the Choi matrix of $\text{tr}(\cdot) \mathbf{1}_{\mathcal{Y}}/n$ is $\frac{1}{n} \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}$. Apply Theorem 4.8.8(b),(c) with $\lambda = \bar{p}/n$ and $\eta \leq \frac{n}{2} d_{\text{cb}} < \lambda$. For local noise, $\mathcal{D}_p^{\otimes k} - p^k \mathcal{D}_1^{\otimes k}$ is completely positive, so

$$J(\mathcal{D}_p^{\otimes k} \circ \Phi_0) \geq p^k J(\mathcal{D}_1^{\otimes k} \circ \Phi_0) = (p/2)^k \mathbf{1}_{\mathcal{Y} \otimes \mathcal{X}}.$$

□

Example 4.8.11. The order λ^{-1} in Theorem 4.8.8 cannot be improved. Let $\mathcal{Y} = \mathbb{C}^2$, $0 < \lambda < \frac{1}{2}$, $\rho = \text{diag}(\lambda, 1-\lambda)$, $\xi = \text{diag}(1-\lambda, \lambda)$, and let $\Phi = \text{tr}(\cdot)\rho$ and $\Psi = \text{tr}(\cdot)\xi$ be replacement channels in $\text{CP}(\mathcal{X}, \mathcal{Y})$. Then $J(\Phi) = \rho \otimes \mathbf{1}_{\mathcal{X}}$ and $J(\Psi) = \xi \otimes \mathbf{1}_{\mathcal{X}}$ have smallest eigenvalue λ , and since $(A \otimes C) \# (B \otimes C) = (A \# B) \otimes C$ [79], $\widehat{F}(\Phi, \Psi) = \text{tr}(\rho \# \xi) = 2\sqrt{\lambda(1-\lambda)}$. For $0 < e < \lambda$ let $\Phi'_e = \text{tr}(\cdot) \text{diag}(\lambda+e, 1-\lambda-e)$, so that $\eta = e$ and $\Psi' = \Psi$. Then

$$\widehat{F}(\Phi'_e, \Psi) = \sqrt{(\lambda+e)(1-\lambda)} + \sqrt{(1-\lambda-e)\lambda}, \quad \lim_{e \rightarrow 0} \frac{\widehat{F}(\Phi'_e, \Psi) - \widehat{F}(\Phi, \Psi)}{e \widehat{F}(\Phi, \Psi)} = \frac{1-2\lambda}{4\lambda(1-\lambda)},$$

while Theorem 4.8.8(b) with $\Psi' = \Psi$ bounds this quotient by $\lim_{e \rightarrow 0} e^{-1}((1+e/\lambda)^{1/2} - 1) = \frac{1}{2\lambda}$. The ratio of the two is $\frac{1-2\lambda}{2(1-\lambda)} \rightarrow \frac{1}{2}$ as $\lambda \rightarrow 0$.

4.9 Explicit moduli for all operator convex functions and all means, with sharp or near-sharp constants

Lemma 4.4.8 shows that $\omega_g^I(\delta) \rightarrow 0$ for every continuous g , but gives no rate. Explicit moduli were computed above only for individual functions: for the geometric mean in Proposition 4.4.3, for the weighted geometric means in Lemma 4.5.4, and for the Belavkin–Staszewski relative entropy and the geometric Rényi divergences with $\alpha \in (1, 2]$ in Section 4.6. In this section we obtain explicit, dimension-free moduli for every operator convex

f and for every Kubo–Ando mean from their integral representations. The kernels of these representations are rational functions with a single pole. We compute their operator Lipschitz constants exactly (Theorem 4.9.6), and this makes the resulting constants optimal in many cases and within a factor 3 of optimal in general (Theorem 4.9.9).

4.9.1 Operator convex functions

We first recall some well-known facts about operator convex functions, many of which can be found in [61]. They form the backbone of the following subsections.

Every operator convex $f : (0, \infty) \rightarrow \mathbb{R}$ has a representation [61, eq. (2.2)]

$$f(x) = a + b(x - 1) + c(x - 1)^2 + \int_{[0, \infty)} \frac{(x - 1)^2}{x + s} d\mu(s), \quad x > 0, \quad (4.9.1)$$

with $a = f(1)$, $b = f'(1)$, $c \geq 0$ and a positive measure μ on $[0, \infty)$ with $\int (1 + s)^{-1} d\mu(s) < \infty$; conversely every such function is operator convex. Very similarly to the Kubo–Ando case, the value c is the limit $c = \lim_{x \rightarrow \infty} f(x)/x^2$. We call (a, b, c, μ) the *Loewner data* of f . Substituting $x = t/(1 - t)$ gives, for $0 < t < 1$,

$$h_f(t) = a(1 - t) + b(2t - 1) + c\kappa_\infty(t) + \int_{[0, \infty)} \kappa_s(t) d\mu(s), \quad (4.9.2)$$

where

$$\kappa_s(t) = \frac{(2t - 1)^2}{s(1 - t) + t} \quad (s \geq 0), \quad \kappa_\infty(t) = \frac{(2t - 1)^2}{1 - t}.$$

Since $\kappa_s(0) = 1/s$ and $\kappa_s(1) = 1$, monotone convergence gives $f(0^+) = a - b + c + \int s^{-1} d\mu(s)$ and $f'(\infty) = b + \mu([0, \infty))$ if $c = 0$, while $f'(\infty) = +\infty$ if $c > 0$. Thus f is bounded in the sense of Definition 4.5.8 if and only if $c = 0$, $\int_{[0, 1]} s^{-1} d\mu(s) < \infty$ (in particular $\mu(\{0\}) = 0$) and μ is finite. The affine part of (4.9.2) has slope $2b - a$. Every kernel κ_s , $s \in [0, \infty]$, is convex on its domain, vanishes to second order at $t = \frac{1}{2}$, and for $s \neq 1$ has a single pole, at $t = s/(s - 1) \notin (0, 1)$ (at $t = 1$ for $s = \infty$), while $\kappa_1(t) = (2t - 1)^2$.

4.9.2 Operator Lipschitz constants and Schur multipliers

We will produce concrete, explicit bounds in a similar manner to the methods of Section 4.5.

We first recall some useful facts which can be found in any foundational textbooks on matrix analysis (e.g. [62], [16]).

Definition 4.9.1 ([16, Section 1.6]). The Schur product or Hadamard product of two matrices A and B is defined to be the matrix $A \circ B$ whose (i, j) -entry is $A_{ij}B_{ij}$. This has the following significant properties.

1. $\|A \circ B\|_\infty \leq \|A\|_\infty \|B\|_\infty$ for all A, B .
2. If $A, B \geq 0$, then $A \circ B \geq 0$.

Theorem 4.9.2 (Daleckiĭ–Krein formula [62, Theorem 3.25]). Assume that f is a C^1 -function on an open interval containing $t_1, t_2, \dots, t_n$, and let $A = \text{diag}(t_1, t_2, \dots, t_n)$. If $B = B^*$, then the derivative of $t \mapsto f(A + tB)$ at $t = 0$ is a Schur product:

$$\left. \frac{d}{dt} f(A + tB) \right|_{t=0} = M \circ B, \quad (4.9.3)$$

where M is the divided difference matrix:

$$M_{ij} = \begin{cases} \frac{f(t_i) - f(t_j)}{t_i - t_j} & \text{if } t_i - t_j \neq 0, \\ f'(t_i) & \text{if } t_i - t_j = 0. \end{cases}$$

For a compact interval I of positive length and a function g that is continuously differentiable on I , with one-sided derivatives at the endpoints, put

$$\text{Lip}^I(g) = \sup \left\{ \frac{\|g(X) - g(Y)\|_\infty}{\|X - Y\|_\infty} : X \neq Y \in \mathcal{L}(\mathcal{V}) \text{ Hermitian with spectra in } I \right\},$$

the supremum being over all finite dimensions, so that $\omega_g^I(\delta) \leq \text{Lip}^I(g) \delta$. For a finite set $\Lambda \subseteq I$ let $g_\Lambda^{[1]} = [g^{[1]}(\lambda, \lambda')]_{\lambda, \lambda' \in \Lambda}$ be the divided difference matrix, with $g^{[1]}(\lambda, \lambda) = g'(\lambda)$. For a real matrix M let

$$\|M\|_S = \sup \{ \|M \circ Z\|_\infty : \|Z\|_\infty \leq 1 \}$$

be the norm of the Schur multiplier $Z \mapsto M \circ Z$, where Z ranges over all complex matrices of the size of M .

We can express the Lipschitz continuity modulus in terms of the divided difference matrix.

Lemma 4.9.3. $\text{Lip}^I(g) = \sup_\Lambda \|g_\Lambda^{[1]}\|_S$, the supremum being over finite subsets $\Lambda \subseteq I$. In particular $\text{Lip}^I(g) \geq \max_I |g'|$.

Remark 4.9.4 (Schur products and multipliers). The following lemma and proof uses four standard facts. First, the Schur product depends on the choice of basis, but operator norms do not: if Ξ is a linear map on matrices and U is unitary, then $X \mapsto U^* X U$ is an isometric bijection for the operator norm, so the induced norm of Ξ is the same whether Ξ is written in the standard basis or in the basis given by the columns of U . Second (Grothendieck–Haagerup), $\|M\|_S \leq C$ if and only if there are vectors x_i, y_j in a Hilbert space with $M_{ij} = \langle x_i, y_j \rangle$ and $\sup_i \|x_i\| \sup_j \|y_j\| \leq C$, and $\|M\|_S$ is the least such C . Third, if $M'_{ij} = M_{\iota_1(i)\iota_2(j)}$ for surjective maps ι_1, ι_2 , so that M' is obtained from M by repeating rows and columns, then $\|M'\|_S = \|M\|_S$. Indeed, a factorization $M_{ij} = \langle x_i, y_j \rangle$ gives $M'_{ij} = \langle x_{\iota_1(i)}, y_{\iota_2(j)} \rangle$ with the same bound, so $\|M'\|_S \leq \|M\|_S$. Conversely, choose one index in each fibre of ι_1 and of ι_2 ; for $\|Z\|_\infty \leq 1$, let Z' be Z placed in the chosen rows and columns and zero elsewhere, so that $\|Z'\|_\infty = \|Z\|_\infty$ and $M' \circ Z'$ is $M \circ Z$ placed in the same positions, whence $\|M\|_S \leq \|M'\|_S$. Fourth, if M is real symmetric and H is Hermitian, then $M \circ H$ is Hermitian, since $\overline{M_{ji} H_{ji}} = M_{ij} H_{ij}$.

Proof of Lemma 4.9.3. Both sides depend only on the restriction of g to I , so we may extend g beyond the endpoints of I by its tangent lines there and assume that g is continuously differentiable on a neighbourhood of I . Let $Z = \sum_\lambda \lambda P_\lambda$ be Hermitian with spectrum $\Lambda \subseteq I$. The Daleckiĭ–Krein formula [16, Theorem V.3.3] (see also [62, Theorem 3.25]) gives

$$Dg(Z)[H] = \sum_{\lambda, \lambda'} g^{[1]}(\lambda, \lambda') P_\lambda H P_{\lambda'}.$$

Choose an orthonormal basis adapted to the P_λ , and let λ_i be the eigenvalue of Z belonging to the i -th basis vector. The (i, j) entry of $Dg(Z)[H]$ is then $g^{[1]}(\lambda_i, \lambda_j) H_{ij}$, with the convention $g^{[1]}(\lambda, \lambda) = g'(\lambda)$, that is,

$$Dg(Z)[H] = M \circ H, \quad M_{ij} = g^{[1]}(\lambda_i, \lambda_j).$$

The matrix M is obtained from $g_\Lambda^{[1]} = (g^{[1]}(\lambda, \lambda'))_{\lambda, \lambda' \in \Lambda}$ by repeating each row and column according to the multiplicity of the corresponding eigenvalue. By the repetition and basis-independence facts of Remark 4.9.4, $\|M\|_S = \|g_\Lambda^{[1]}\|_S$ and the choice of adapted basis is irrelevant, so

$$\|Dg(Z)\| \leq \|g_\Lambda^{[1]}\|_S.$$

The set of Hermitian operators with spectra in I is convex, so for X, Y in this set the segment $Y + u(X - Y)$, $u \in [0, 1]$, stays in it, and

$$g(X) - g(Y) = \int_0^1 Dg(Y + u(X - Y))[X - Y] du.$$

Therefore

$$\|g(X) - g(Y)\|_\infty \leq \sup_\Lambda \|g_\Lambda^{[1]}\|_S \|X - Y\|_\infty,$$

which gives $\text{Lip}^I(g) \leq \sup_\Lambda \|g_\Lambda^{[1]}\|_S$.

For the reverse inequality, let $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ lie in the interior of I , and put $Z = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $M = g_\Lambda^{[1]}$, which is real symmetric. For Hermitian H of the same size and small u , the spectrum of $Z + uH$ lies in I , and by [62, Theorem 3.25], $Dg(Z)[H] = M \circ H$, so

$$M \circ H = \lim_{u \rightarrow 0} u^{-1}(g(Z + uH) - g(Z)).$$

Hence

$$\|M \circ H\|_\infty \leq \text{Lip}^I(g) \|H\|_\infty \quad \text{for all Hermitian } H. \quad (4.9.4)$$

To pass from Hermitian H to arbitrary matrices, let N be any $n \times n$ matrix and set

$$\widehat{M} = \begin{bmatrix} M & M \\ M & M \end{bmatrix}, \quad \widehat{N} = \begin{bmatrix} 0 & N \\ N^* & 0 \end{bmatrix}.$$

Then $\widehat{N}$ is Hermitian with $\|\widehat{N}\|_\infty = \|N\|_\infty$, and

$$\widehat{M} \circ \widehat{N} = \begin{bmatrix} 0 & M \circ N \\ M \circ N^* & 0 \end{bmatrix} = \begin{bmatrix} 0 & M \circ N \\ (M \circ N)^* & 0 \end{bmatrix},$$

where the lower-left block satisfies $M \circ N^* = \overline{M} \circ N^* = (M \circ N)^*$ because $M = M^\top = \overline{M}$. This matrix is Hermitian with norm $\|M \circ N\|_\infty$. Moreover $\widehat{M}$ is the matrix associated in the same way with $\text{diag}(Z, Z)$, whose eigenvalues are those of Z repeated twice and still lie in the interior of I , so the argument leading to (4.9.4) applies to it and gives

$$\|M \circ N\|_\infty = \|\widehat{M} \circ \widehat{N}\|_\infty \leq \text{Lip}^I(g) \|\widehat{N}\|_\infty = \text{Lip}^I(g) \|N\|_\infty.$$

Taking the supremum over $\|N\|_\infty \leq 1$ yields $\|g_\Lambda^{[1]}\|_S \leq \text{Lip}^I(g)$ for Λ in the interior of I , and continuity of $g^{[1]}$, and hence of $\Lambda \mapsto \|g_\Lambda^{[1]}\|_S$, extends this to all finite $\Lambda \subseteq I$.

Finally, take $\Lambda = \{t\}$ with $t \in I$. Then $g_\Lambda^{[1]}$ is the 1×1 matrix $(g'(t))$, whose Schur norm is $|g'(t)|$, so $\text{Lip}^I(g) \geq |g'(t)|$ for every $t \in I$, that is, $\text{Lip}^I(g) \geq \max_I |g'|$. $\square$

Lemma 4.9.5. *Let M be a real matrix.*

- (a) $\|M\|_S \geq \max_{i,j} |M_{ij}|$, and $\|M\|_S = \max_i M_{ii}$ if $M \geq 0$. For vectors x, y , $\|xy^\top\|_S = \max_i |x_i| \max_j |y_j|$.

(b) (Compression.) If M is $k \times k$ and Π is a real $n \times k$ matrix whose rows have ℓ^1 -norm at most 1, then $\|\Pi M \Pi^\top\|_S \leq \|M\|_S$.

(c) (2×2 matrices.) For real q_1, q, q_2 put $L = \max\{|q_1|, |q_2|\}$ and, for $t \geq L$, $A(t) = \sqrt{(t - q_1)(t - q_2)}$ and $B(t) = \sqrt{(t + q_1)(t + q_2)}$. Then $\left\| \begin{bmatrix} q_1 & q \\ q & q_2 \end{bmatrix} \right\|_S$ is the smallest $t \geq L$ with $A(t) + B(t) \geq 2|q|$.

(d) If $q_1 \leq q \leq q_2$, then $\left\| \begin{bmatrix} q_1 & q \\ q & q_2 \end{bmatrix} \right\|_S \leq \sqrt{2} \max\{|q_1|, |q_2|\}$.

Proof. We use Haagerup's characterisation [103, Chapter 8]: $\|M\|_S \leq t$ if and only if there are vectors x_i, y_j in a Hilbert space with $M_{ij} = \langle x_i, y_j \rangle$ and $\|x_i\|^2, \|y_j\|^2 \leq t$; equivalently, if and only if there are positive semidefinite Y_1, Y_2 with diagonal entries at most t such that $\begin{bmatrix} Y_1 & M \\ M^\top & Y_2 \end{bmatrix} \geq 0$.

(a) The first claim follows from $|\langle x_i, y_j \rangle| \leq t$. If $M \geq 0$, write $M_{ij} = \langle g_i, g_j \rangle$ with $\|g_i\|^2 = M_{ii}$. For $xy^\top$ use one-dimensional vectors, rescaled.

(b) Put $x'_i = \sum_l \Pi_{il} x_l$ and $y'_j = \sum_{l'} \Pi_{jl'} y_{l'}$. Then $\langle x'_i, y'_j \rangle = (\Pi M \Pi^\top)_{ij}$ and $\|x'_i\| \leq \sum_l |\Pi_{il}| \|x_l\| \leq \sqrt{t}$.

(c) Since M is real and complex conjugation preserves positivity, the entrywise real parts $\frac{1}{2}(Y_1 + \overline{Y_1})$, $\frac{1}{2}(Y_2 + \overline{Y_2})$ of an admissible pair form an admissible pair, so we may assume that Y_1 and Y_2 are real. If (Y_1, Y_2) is an admissible pair for the block decomposition from M , so is (Y_2, Y_1) , by conjugation with the flip of the two blocks, and hence so is $(\frac{Y_1 + Y_2}{2}, \frac{Y_1 + Y_2}{2})$. Moreover

$$\begin{pmatrix} Y_1 & M \\ M & Y_1 \end{pmatrix} \geq 0$$

if and only if $Y_1 \pm M \geq 0$, and raising the diagonal of Y_1 to t preserves both conditions. Hence $\|M\|_S \leq t$ if and only if there is a real $\tilde{q}$ with

$$\begin{pmatrix} t \pm q_1 & \tilde{q} \pm q \\ \tilde{q} \pm q & t \pm q_2 \end{pmatrix} \geq 0$$

for both signs, that is, $t \geq L$, $|\tilde{q} - q| \leq A(t)$ and $|\tilde{q} + q| \leq B(t)$. Such a $\tilde{q}$ exists if and only if $2|q| \leq A(t) + B(t)$. Since A and B are nondecreasing on $[L, \infty)$, the claim follows.

(d) Replacing M by $-M$ and exchanging the two indices preserves the Schur multiplier norm and maps (q_1, q, q_2) to $(-q_2, -q, -q_1)$, so we may assume $q \geq 0$, and by scaling $L = 1$; then $0 \leq q \leq q_2 \leq 1$ and $-1 \leq q_1 \leq q_2$. Let $t = \sqrt{2}$. Since

$$\partial_{q_1}(A + B) = \frac{1}{2}(\sqrt{(t + q_2)/(t + q_1)} - \sqrt{(t - q_2)/(t - q_1)}) \geq 0$$

for $q_1 \leq q_2$, we have

$$A(t) + B(t) \geq \varphi(q_2) := \sqrt{(t + 1)(t - q_2)} + \sqrt{(t - 1)(t + q_2)}.$$

The function φ is concave, $\varphi(0) > 0$ and $\varphi(1) = 2\sqrt{t^2 - 1} = 2$, so $\varphi(q_2) \geq 2q_2 \geq 2q$ on $[0, 1]$, and (c) gives the claim. $\square$

We now combine the previous lemmas into a theorem which will begin to allow us to express the continuity moduli for the operator convex functions in terms of their integral representations. In the following theorem, the function has the structure of (4.9.1).

Theorem 4.9.6. *Let $I = [t_-, t_+]$ with $t_- < t_+$, let $p \in \mathbb{R} \setminus I$, $\psi(t) = |t - p|^{-1}$, and $g = C + \beta t + w\psi$ with $C, \beta, w \in \mathbb{R}$. Then:*

(a)

$$\text{Lip}^I(g) = \|g_{\{t_-, t_+\}}^{[1]}\|_S = \left\| \begin{bmatrix} g'(t_-) & g^{[1]}(t_-, t_+) \\ g^{[1]}(t_-, t_+) & g'(t_+) \end{bmatrix} \right\|_S,$$

which is given explicitly by Lemma 4.9.5(c);

(b) $\max_I |g'| \leq \text{Lip}^I(g) \leq \sqrt{2} \max_I |g'|$;(c) $\text{Lip}^I(g) = \max_I |g'|$ if $\beta w(p - t_-) \geq 0$, or if $g'(t_0) = 0$ for some $t_0 \in I$ with $\psi(t_0) \leq \frac{1}{2}(\psi(t_-) + \psi(t_+))$. The latter holds whenever t_0 lies on the far side of the midpoint of I from p , the midpoint included;(d) the constant $\sqrt{2}$ in (b) is optimal: $k_z(t) = t + 2z^2/(t + z)$ satisfies $\max_{[0,1]} |k'_z| = 1$ and $\text{Lip}^{[0,1]}(k_z) \rightarrow \sqrt{2}$ as $z \downarrow 0$.

Proof. Let $\zeta = +1$ if $p > t_+$ and $\zeta = -1$ if $p < t_-$. We then have $\psi^{[1]}(x, y) = \zeta \psi(x) \psi(y)$ for $x, y \in I$. This is because if $p \notin I$, the sign of $t - p$ is constant on I , equal to $-\zeta$, so $\psi(t) = -\zeta/(t - p)$ and $\psi(x)\psi(y) = 1/((x - p)(y - p))$. For $x \neq y$ in I ,

$$\psi(x) - \psi(y) = -\zeta \left(\frac{1}{x - p} - \frac{1}{y - p} \right) = \zeta \frac{x - y}{(x - p)(y - p)},$$

so $\psi^{[1]}(x, y) = \zeta \psi(x) \psi(y)$. For $x = y$, $\psi'(x) = \zeta/(x - p)^2 = \zeta \psi(x)^2$, so the formula holds on the diagonal too.

$$g^{[1]}(x, y) = \beta + \zeta w \psi(x) \psi(y).$$

(a) The function ψ is continuous and monotone on I , so for $x \in I$ there is $\lambda(x) \in [0, 1]$ with $\psi(x) = \lambda(x)\psi(t_-) + (1 - \lambda(x))\psi(t_+)$. Put $\lambda_{t_-} = \lambda$ and $\lambda_{t_+} = 1 - \lambda$. Since $\lambda_{t_-} + \lambda_{t_+} = 1$ and the kernel $\beta + \zeta w u_1 u_2$ is affine in u_1 and in u_2 ,

$$g^{[1]}(x, y) = \sum_{t, t' \in \{t_-, t_+\}} \lambda_t(x) \lambda_{t'}(y) g^{[1]}(t, t'), \quad x, y \in I.$$

Thus $g_\Lambda^{[1]} = \Pi g_{\{t_-, t_+\}}^{[1]} \Pi^\top$, where Π has the nonnegative entries $\lambda_t(x)$ and rows summing to 1. Lemma 4.9.5(b) and Lemma 4.9.3 give $\text{Lip}^I(g) \leq \|g_{\{t_-, t_+\}}^{[1]}\|_S$, and the reverse inequality is Lemma 4.9.3 with $\Lambda = \{t_-, t_+\}$.

(b) The first inequality is in Lemma 4.9.3. Since g'' has constant sign on I , $g^{[1]}(t_-, t_+)$ lies between $g'(t_-)$ and $g'(t_+)$, and g' is monotone, so $\max_I |g'| = \max\{|g'(t_-)|, |g'(t_+)|\}$. Lemma 4.9.5(d), applied to g or to $-g$, gives the second inequality.

(c) If $\beta \zeta w \geq 0$, then $\zeta \text{sgn}(w) g_\Lambda^{[1]} = |\beta| \mathbf{j} \mathbf{j}^\top + |w| \psi \psi^\top \geq 0$, where ψ also denotes the vector $(\psi(x))_{x \in \Lambda}$, so Lemma 4.9.5(a) gives $\|g_\Lambda^{[1]}\|_S = \max_\Lambda |g'|$. Otherwise put $c_0 = \sqrt{-\beta/(\zeta w)} > 0$, so that $g' = \zeta w(\psi^2 - c_0^2)$ and, with $v = \psi - c_0 \mathbf{j}$,

$$g_\Lambda^{[1]} = \zeta w(\psi \psi^\top - c_0^2 \mathbf{j} \mathbf{j}^\top) = \zeta w(c_0 v \mathbf{j}^\top + c_0 \mathbf{j} v^\top + v v^\top).$$

By Lemma 4.9.5(a),

$$\|g_\Lambda^{[1]}\|_S \leq |w|(2c_0 N + N^2) = |w|((c_0 + N)^2 - c_0^2)$$

with $N = \max_\Lambda |\psi - c_0|$. If $c_0 \leq \frac{1}{2}(\psi_{\min} + \psi_{\max})$, where $\psi_{\min}, \psi_{\max}$ are the values of ψ at the endpoints, then $N \leq \psi_{\max} - c_0$ and the bound is $|w|(\psi_{\max}^2 - c_0^2) = |g'(t_1)|$, where t_1 is the endpoint closest to p . If $g'(t_0) = 0$ then $\psi(t_0) = c_0$, which gives the stated condition.

Finally, ψ is convex and decreases away from p ; if t_0 lies on the far side of the midpoint t_{mid} , then

$$\psi(t_0) \leq \psi(t_{\text{mid}}) \leq \frac{1}{2}(\psi(t_-) + \psi(t_+)).$$

(d) Here $k'_z(t) = 1 - 2z^2/(t+z)^2$ increases from -1 to $1 - 2z^2/(1+z)^2$, and $k_z^{[1]}(0, 1) = 1 - 2z/(1+z)$. The endpoint matrix tends to $\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$, whose Schur multiplier norm is $\sqrt{2}$ by Lemma 4.9.5(c), since $A(t) + B(t) = 2\sqrt{t^2 - 1}$. The norm in Lemma 4.9.5(c) depends continuously on (q_1, q, q_2) , and (a) gives the claim. $\square$

Theorem 4.9.6(d) shows that operator convexity does not imply $\text{Lip}^I(g) = \max_I |g'|$, even for a single pole. The operator Lipschitz constant can exceed the scalar one because g' changes sign on I , from -1 at the endpoint near the pole to almost 1 on the rest of the interval.

We now construct sharp constants for the kernels described in (4.9.2). This will allow us to define continuity constants for operator divergences defined through convex perspectives.

Corollary 4.9.7 (Sharp constants for the kernels). *Let $I \subseteq [0, 1]$ be a compact interval and $r_I = \max_{t \in I} |2t - 1|$.*

- (a) *Let $s \geq 0$ and $\ell_s(t) = s(1-t) + t$, and assume $d_s = \min_{t \in I} \ell_s(t) > 0$ (for $t \in [0, 1]$ the function $\ell_s(t) = (1-t)s + t$ is a convex combination of s and 1, so this fails only if $s = 0$ and $0 \in I$). Then for $t \in I$,*

$$|\kappa'_s(t)| = |2t - 1| \left(\frac{2}{\ell_s(t)} + \frac{1+s}{\ell_s(t)^2} \right),$$

the bracket is positive, and therefore

$$|\kappa'_s(t)| \leq r_I \left(\frac{2}{d_s} + \frac{1+s}{d_s^2} \right)$$

and

$$\text{Lip}^I(\kappa_s) \leq r_I \left(\frac{2}{d_s} + \frac{1+s}{d_s^2} \right),$$

with equality, and equal to $\max_I |\kappa'_s|$, if I is symmetric about $\frac{1}{2}$. The same holds for κ_∞ , with $\ell_\infty(t) = 1 - t$ and $1 + s$ replaced by 1. In addition

$$\|\kappa_s(X) - \kappa_s(Y)\|_\infty \leq \text{osc}_I \kappa_s = \max_I \kappa_s - \min_I \kappa_s.$$

- (b) $\text{Lip}^{[0, 1-\epsilon]}(\kappa_\infty) = \epsilon^{-2} - 4$ for $0 < \epsilon \leq \frac{1}{3}$.

- (c) For $s > 0$, the kernel $\eta_s(t) = (1+s)t(1-t)/(1+(s-1)t)$ of (4.9.5) below satisfies

$$\text{Lip}^{[0, 1]}(\eta_s) = \max_{[0, 1]} |\eta'_s| = (1+s) \max\{1, s^{-1}\}.$$

- (d) For $s > 0$, $g_s(t) = \frac{t}{1+s} - \frac{t}{t+s}$ satisfies $\text{Lip}^{[0, 1]}(g_s) = \frac{1}{s(1+s)}$ and $-\frac{1}{1+s} \leq g_s \leq 0$ on $[0, 1]$.

- (e) $d_s \geq (1+s) \min_{t \in I} \min\{t, 1-t\}$.

Proof. (a) The identity

$$(1-s)(2t-1) = 2\ell_s(t) - (1+s)$$

gives

$$\kappa'_s = (2t-1)(2\ell_s + (1+s))/\ell_s^2,$$

which is the formula. For the bound, with $X, Y \in L(\mathcal{V})$ Hermitian with spectra in I , put

$$\ell_s(X) = s(\mathbf{1}_{\mathcal{V}} - X) + X$$

and

$$R(X) = (2X - \mathbf{1}_{\mathcal{V}})\ell_s(X)^{-1},$$

so that

$$\kappa_s(X) = (2X - \mathbf{1}_{\mathcal{V}})R(X), \quad \|\ell_s(X)^{-1}\|_{\infty} \leq d_s^{-1},$$

and $\|R(X)\|_{\infty} \leq r_I/d_s$. Expanding,

$$(2X - \mathbf{1}_{\mathcal{V}})\ell_s(Y) - \ell_s(X)(2Y - \mathbf{1}_{\mathcal{V}}) = (1+s)(X - Y),$$

since the terms in XY cancel. Hence

$$R(X) - R(Y) = (1+s)\ell_s(X)^{-1}(X - Y)\ell_s(Y)^{-1},$$

using that $\ell_s(X)^{-1}$ commutes with $2X - \mathbf{1}_{\mathcal{V}}$, and

$$\kappa_s(X) - \kappa_s(Y) = 2(X - Y)R(Y) + (1+s)R(X)(X - Y)\ell_s(Y)^{-1},$$

which gives

$$\|\kappa_s(X) - \kappa_s(Y)\|_{\infty} \leq r_I \left(\frac{2}{d_s} + \frac{1+s}{d_s^2} \right) \|X - Y\|_{\infty}.$$

If I is symmetric about $\frac{1}{2}$, then at the endpoint where ℓ_s is minimal both $|2t-1| = r_I$ and $\ell_s = d_s$, so $\max_I |\kappa'_s|$ equals the bound, and Lemma 4.9.3 gives equality. For κ_{∞} use

$$\ell_{\infty}(X) = \mathbf{1}_{\mathcal{V}} - X$$

and

$$(2X - \mathbf{1}_{\mathcal{V}})(\mathbf{1}_{\mathcal{V}} - Y) - (\mathbf{1}_{\mathcal{V}} - X)(2Y - \mathbf{1}_{\mathcal{V}}) = X - Y.$$

The last statement holds because $\kappa_s(X)$ and $\kappa_s(Y)$ have spectra in $[\min_I \kappa_s, \max_I \kappa_s]$.

(b) We have

$$\kappa_{\infty}(t) = (1-t)^{-1} - 4t$$

a one-pole function with $p = 1$, $\beta = -4$ and $w = 1$, and κ'_{∞} vanishes at $t_0 = \frac{1}{2}$. On $[0, 1-\epsilon]$, $\psi(t) = (1-t)^{-1}$ takes the values 1 and ϵ^{-1} at the endpoints and $\psi(t_0) = 2$, so the condition of Theorem 4.9.6(c) is

$$2 \leq \frac{1}{2}(1 + \epsilon^{-1})$$

that is, $\epsilon \leq \frac{1}{3}$. Then

$$\text{Lip} = \max\{|\kappa'_{\infty}(0)|, \kappa'_{\infty}(1-\epsilon)\} = \max\{3, \epsilon^{-2} - 4\} = \epsilon^{-2} - 4$$

(c) For $s = 1$, $\eta_1(t) = 2t(1-t)$, and for every quadratic g the matrix $g_{\Lambda}^{[1]}$ is a compression of $g_{\{t_-, t_+\}}^{[1]}$ as in the proof of Theorem 4.9.6(a), with λ affine; here

$$g_{\{0,1\}}^{[1]} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

has Schur multiplier norm 2, since $\|\text{diag}(d)\|_S = \max_i |d_i|$, as $\text{diag}(d) \circ Z = \text{diag}(d_i Z_{ii})$.

For $s > 1$, $-\eta_s$ is a one-pole function with pole $p = -1/(s-1) < 0$; then

$$\psi(t) = (s-1)/(1+(s-1)t)$$

with $\psi(0) = s-1$ and $\psi(1) = (s-1)/s$. The derivative η'_s vanishes at $t_0 = 1/(1+\sqrt{s})$, where $\psi(t_0) = (s-1)/\sqrt{s}$. The condition $\psi(t_0) \leq \frac{1}{2}(\psi(0) + \psi(1))$ reads $2\sqrt{s} \leq 1+s$, so Theorem 4.9.6(c) gives

$$\text{Lip} = \max\{|\eta'_s(0)|, |\eta'_s(1)|\} = \max\{1+s, (1+s)/s\} = 1+s$$

For $s < 1$ use $\eta_s(1-t) = \eta_{1/s}(t)$.

(d) We have

$$g_s(t) = \frac{t}{1+s} - 1 + \frac{s}{t+s}$$

which is a one-pole function with $p = -s$, $\beta = \frac{1}{1+s}$ and $w = s$. Its derivative

$$g'_s(t) = \frac{1}{1+s} - \frac{s}{(t+s)^2}$$

vanishes at

$$t_0 = \sqrt{s(1+s)} - s \in (0, 1)$$

where $\psi(t_0) = (s(1+s))^{-1/2}$. The condition of Theorem 4.9.6(c) is

$$(s(1+s))^{-1/2} \leq \frac{1}{2} \left(\frac{1}{s} + \frac{1}{1+s} \right)$$

that is, $2\sqrt{s(1+s)} \leq 1+2s$, which holds. Hence

$$\text{Lip} = \max\{|g'_s(0)|, |g'_s(1)|\} = \max\left\{\frac{1}{s(1+s)}, \frac{1}{(1+s)^2}\right\} = \frac{1}{s(1+s)}$$

Finally $|g_s(t)| = \frac{t(1-t)}{(1+s)(t+s)} \leq \frac{1-t}{1+s}$.

(e) $s(1-t) + t \geq (1+s) \min\{t, 1-t\}$. □

We finally come to the explicit modulus data for the integral representative of the operator perspective channel divergences.

Theorem 4.9.8. *Let f be operator convex with Loewner data (a, b, c, μ) , and let $I \subseteq [0, 1]$ be a compact interval on which h_f is finite. With r_I , d_s and d_∞ as in Corollary 4.9.7, for $\delta \geq 0$,*

$$\omega_{h_f}^I(\delta) \leq |2b-a|\delta + cr_I \left(\frac{2}{d_\infty} + \frac{1}{d_\infty^2} \right) \delta + \int_{[0, \infty)} \min \left\{ r_I \left(\frac{2}{d_s} + \frac{1+s}{d_s^2} \right) \delta, \text{osc}_I \kappa_s \right\} d\mu(s),$$

where the c -term is absent if $c = 0$. Consequently, Theorems 4.5.13(b), 4.6.7(b) and 4.7.6(a) hold with this explicit modulus.

Proof. If $1 \in I$ then $c = 0$, and if $0 \in I$ then $\mu(\{0\}) = 0$, since h_f is finite on I ; so all constants are finite where they occur. For $X \in \mathcal{L}(\mathcal{V})$ with spectrum in I , (4.9.2) and the spectral theorem give

$$h_f(X) = (a-b)\mathbf{1}_{\mathcal{V}} + (2b-a)X + c\kappa_\infty(X) + \int \kappa_s(X) d\mu(s),$$

the integral being a finite sum of convergent scalar integrals times spectral projections. Subtract the same identity for Y , take norms inside the integral, and apply Corollary 4.9.7(a). □

4.9.3 Sharp Lipschitz constants

On the intervals $[\epsilon, 1 - \epsilon]$ that correspond to a two-sided bound on the max-relative entropy, the kernel constants of Corollary 4.9.7 are attained. For $0 < \epsilon \leq \frac{1}{2}$ put $r_\epsilon = 1 - 2\epsilon$,

$$M_s(\epsilon) = r_\epsilon \left(\frac{2}{d_s} + \frac{1+s}{d_s^2} \right), \quad d_s = \epsilon(1+s) + r_\epsilon \min\{1, s\}, \quad M_\infty(\epsilon) = \frac{1 - 4\epsilon^2}{\epsilon^2},$$

so that $M_s(\epsilon) = \max_{[\epsilon, 1-\epsilon]} |\kappa'_s|$ for $s \in [0, \infty]$ by Corollary 4.9.7(a), and

$$L_f(\epsilon) = |2b - a| + c M_\infty(\epsilon) + \int_{[0, \infty)} M_s(\epsilon) d\mu(s).$$

Theorem 4.9.9. *Let f be operator convex with Loewner data (a, b, c, μ) , let $0 < \epsilon < \frac{1}{2}$ and $I = [\epsilon, 1 - \epsilon]$.*

- (a) $\max_I |h'_f| \leq \text{Lip}^I(h_f) \leq L_f(\epsilon) \leq 3 \max_I |h'_f|$.
- (b) $\text{Lip}^I(h_f) = L_f(\epsilon) = h'_f(1 - \epsilon)$ if μ is carried by $[1, \infty)$ and $2b \geq a$; and $\text{Lip}^I(h_f) = L_f(\epsilon) = -h'_f(\epsilon)$ if $c = 0$, μ is carried by $[0, 1]$ and $2b \leq a$.
- (c) If $h_f(1 - t) = h_f(t)$ for all t , equivalently $f(x) = xf(1/x)$, then $\text{Lip}^I(h_f) = \max_I |h'_f| = h'_f(1 - \epsilon)$.

Proof. (a) The first inequality is Lemma 4.9.3, and the second follows from (4.9.2) as in the proof of Theorem 4.9.8.

For the third, write $t_- = \epsilon$ and $t_+ = 1 - \epsilon$. Each κ_s , $s \in [0, \infty]$, is convex with $\kappa'_s(\frac{1}{2}) = 0$, so

$$\kappa'_s(t_-) \leq 0 \leq \kappa'_s(t_+)$$

and $M_s(\epsilon) = \max\{|\kappa'_s(t_-)|, \kappa'_s(t_+)\}$.

Writing

$$\Sigma_+ = c \kappa'_\infty(t_+) + \int \kappa'_s(t_+) d\mu(s)$$

and

$$\Sigma_- = c |\kappa'_\infty(t_-)| + \int |\kappa'_s(t_-)| d\mu(s)$$

so that

$$L_f(\epsilon) \leq |2b - a| + \Sigma_+ + \Sigma_-$$

$$h'_f(t_+) = (2b - a) + \Sigma_+$$

and

$$h'_f(t_-) = (2b - a) - \Sigma_-.$$

Since h_f is convex,

$$N := \max_I |h'_f| = \max\{|h'_f(t_-)|, |h'_f(t_+)|\}.$$

Hence

$$N \geq (2b - a) + \Sigma_+$$

and $N \geq \Sigma_- - (2b - a)$, so $\Sigma_+ + \Sigma_- \leq 2N$. Moreover $N \geq |2b - a|$: if $2b \geq a$ this follows from the first inequality, and otherwise from the second. Thus $L_f(\epsilon) \leq 3N$.

(b) For $s \geq 1$ the function ℓ_s is nonincreasing, so $\max_I |\kappa'_s|$ is attained at t_+ , by the formula in Corollary 4.9.7(a); the same holds for κ_∞ . If μ is carried by $[1, \infty)$ and $2b \geq a$, then

$$L_f(\epsilon) = (2b - a) + c \kappa'_\infty(t_+) + \int \kappa'_s(t_+) d\mu(s) = h'_f(t_+) \leq \max_I |h'_f|,$$

and (a) gives equality throughout. The second case is the mirror image.

(c) Write $\tilde{g}(t) = g(1 - t)$. Since $h_f = \frac{1}{2}(h_f + \tilde{h}_f)$ and the affine part of the right-hand side has slope 0, (4.9.2) gives

$$h_f = \text{const} + \frac{c}{2} \kappa_\infty^{\text{sym}} + \frac{1}{2} \int_{[0, \infty)} \kappa_s^{\text{sym}} d\mu(s), \quad \kappa_s^{\text{sym}}(t) = \kappa_s(t) + \kappa_s(1 - t).$$

We claim that $\text{Lip}^I(\kappa_s^{\text{sym}}) = (\kappa_s^{\text{sym}})'(t_+)$ for every $s \in [0, \infty]$. Then

$$\text{Lip}^I(h_f) \leq \frac{c}{2} (\kappa_\infty^{\text{sym}})'(t_+) + \frac{1}{2} \int (\kappa_s^{\text{sym}})'(t_+) d\mu(s) = h'_f(t_+),$$

and Lemma 4.9.3 gives equality. For $s = 1$, $\kappa_1^{\text{sym}} = 2\kappa_1$ is quadratic, and the claim follows as in the proof of Corollary 4.9.7(c). For $s \neq 1$, $\kappa_s = C + \beta t + w\psi_p$ with one pole $p \notin (0, 1)$ and $\psi_p(t) = |t - p|^{-1}$; then $\kappa_s(1 - t) = C' + \beta't + w\psi_{1-p}$, with the mirrored pole on the other side of $[0, 1]$, and the slopes cancel because κ_s^{sym} is symmetric. By the computation in the proof of Theorem 4.9.6, with $u = (\psi_p(x))_{x \in \Lambda}$ and $\tilde{u} = (\psi_{1-p}(x))_{x \in \Lambda}$,

$$\pm(\kappa_s^{\text{sym}})_\Lambda^{[1]} = w(uu^\top - \tilde{u}\tilde{u}^\top) = \frac{w}{2}((u - \tilde{u})(u + \tilde{u})^\top + (u + \tilde{u})(u - \tilde{u})^\top).$$

By Lemma 4.9.5(a) its Schur multiplier norm is at most $|w| \max_\Lambda |u - \tilde{u}| \max_\Lambda (u + \tilde{u})$. On the symmetric interval I , the function $\psi_p - \psi_{1-p}$ is monotone and antisymmetric about $\frac{1}{2}$, and $\psi_p + \psi_{1-p}$ is convex and symmetric, so both maxima are attained at t_+ , and the bound equals

$$|w| |\psi_p(t_+)^2 - \psi_{1-p}(t_+)^2| = |(\kappa_s^{\text{sym}})'(t_+)|.$$

Finally $(\kappa_s^{\text{sym}})'(t_+) \geq 0$, because κ_s^{sym} is convex and symmetric about $\frac{1}{2}$. $\square$

Part (c) applies to the Jeffreys, symmetrised χ^2 , Le Cam, Hellinger and Jensen–Shannon divergences, and part (b) to the Pearson and Neyman χ^2 -divergences. For these, Table 4.2 gives the exact Lipschitz constants. For the Belavkin–Staszewski and Rényi divergences the Loewner data are two-sided and the theorem gives only $\max_I |h'_f| \leq \text{Lip}^I(h_f) \leq L_f(\epsilon)$.

The next proposition shows that if the identity $\text{Lip}^I(g) = \|g_{\{t_-, t_+\}}^{[1]}\|_s$ of Theorem 4.9.6(a) holds for all combinations of two kernels κ_s , then it holds for every h_f .

Proposition 4.9.10. *Let $I = [t_-, t_+] \subseteq (0, 1)$. If $\text{Lip}^I(g) = \|g_{\{t_-, t_+\}}^{[1]}\|_s$ for every $g = \beta t + w_1 \kappa_{s_1} + w_2 \kappa_{s_2}$ with $\beta \in \mathbb{R}$, $w_1, w_2 \geq 0$ and $s_1, s_2 \in [0, \infty]$, then $\text{Lip}^I(h_f) = \|(h_f)_{\{t_-, t_+\}}^{[1]}\|_s$ for every operator convex f .*

Proof. Suppose first that $h = \beta t + \sum_{k=1}^N w_k \kappa_{s_k}$ with $w \in [0, \infty)^N$; the c -term is the case $s_k = \infty$, and constants do not affect divided differences. Put $u_k = \kappa_{s_k}^{[1]}(t_-, t_+) - \kappa'_{s_k}(t_-)$ and $u'_k = \kappa'_{s_k}(t_+) - \kappa_{s_k}^{[1]}(t_-, t_+)$, which are positive by strict convexity. The endpoint data $(h'(t_-), h^{[1]}(t_-, t_+), h'(t_+))$ are determined by $h'(t_-)$, $\sum_k w_k u_k$ and $\sum_k w_k u'_k$. The set $\mathcal{P}$ of $w' \in [0, \infty)^N$ with $\sum_k w'_k u_k = \sum_k w_k u_k$ and $\sum_k w'_k u'_k = \sum_k w_k u'_k$ is a compact polytope containing w , whose vertices have at most two nonzero coordinates. Write $w = \sum_j \lambda_j v^{(j)}$ as a convex combination of vertices, and let $h_j = \beta_j t + \sum_k v_k^{(j)} \kappa_{s_k}$, with β_j chosen so that $h'_j(t_-) = h'(t_-)$. Then each h_j has the same endpoint data as h , and $h - \sum_j \lambda_j h_j$ is affine with derivative 0 at t_- , hence constant. For every finite Λ ,

$$\|h_\Lambda^{[1]}\|_s \leq \sum_j \lambda_j \text{Lip}^I(h_j) = \|h_{\{t_-, t_+\}}^{[1]}\|_s.$$

In general, $\int \kappa_s d\mu(s) = \int (1+s)\kappa_s d\tilde{\mu}(s)$ with the finite measure $d\tilde{\mu} = (1+s)^{-1}d\mu$, and $s \mapsto (1+s)\kappa_s$ is continuous from $[0, \infty]$ into $C^1(I)$, with value κ_∞ at $s = \infty$, because $I \subseteq (0, 1)$. Approximating $\tilde{\mu} + c\delta_\infty$ by finitely supported positive measures along finer and finer partitions of $[0, \infty]$, uniform continuity gives functions h_n of the above form with $h_n \rightarrow h_f$ in $C^1(I)$. Then $(h_n)^{[1]} \rightarrow (h_f)^{[1]}$ uniformly on $I \times I$, and for every finite Λ ,

$$\|(h_f)_\Lambda^{[1]}\|_S = \lim_n \|(h_n)_\Lambda^{[1]}\|_S \leq \lim_n \|(h_n)_{\{t_-, t_+\}}^{[1]}\|_S = \|(h_f)_{\{t_-, t_+\}}^{[1]}\|_S.$$

□

Corollary 4.9.11 (Lipschitz moduli under a Thompson bound). *Let f be operator convex with Loewner data (a, b, c, μ) and $\epsilon \in (0, \frac{1}{2})$; for $\epsilon = \frac{1}{2}$ the interval $[\epsilon, 1 - \epsilon]$ is a point and the modulus vanishes. Then $\omega_{h_f}^{[\epsilon, 1-\epsilon]}(\delta) \leq L_f(\epsilon)\delta$, where $L_f(\epsilon)$ is within a factor 3 of the optimal constant $\text{Lip}^{[\epsilon, 1-\epsilon]}(h_f)$, and*

$$L_f(\epsilon) \leq |2b - a| + \frac{1 - 4\epsilon^2}{\epsilon^2} \left(c + \int_{[0, \infty)} \frac{d\mu(s)}{1+s} \right).$$

Hence, if $\Phi_i = \Psi_{T_i}$ and $\Psi_i = \Psi_{\mathbf{1}_{\mathbb{C}^r} - T_i}$ satisfy

$$\max\{D_{\max}(\Phi_i \| \Psi_i), D_{\max}(\Psi_i \| \Phi_i)\} \leq \log \frac{1 - \epsilon}{\epsilon}$$

for $i = 1, 2$, then

$$|\hat{D}_f(\Phi_1 \| \Psi_1) - \hat{D}_f(\Phi_2 \| \Psi_2)| \leq \|\Omega^*(\mathbf{1}_Y)\|_\infty L_f(\epsilon) \|T_1 - T_2\|_\infty.$$

Proof. The first statement is Theorem 4.9.9(a). Corollary 4.9.7(e) gives $d_s \geq \epsilon(1+s)$, hence $M_s(\epsilon) \leq r_\epsilon \frac{1+2\epsilon}{\epsilon^2(1+s)} = \frac{1-4\epsilon^2}{\epsilon^2(1+s)}$. The last statement follows once we identify the hypothesis with $\epsilon \mathbf{1}_{\mathbb{C}^r} \leq T_i \leq (1 - \epsilon) \mathbf{1}_{\mathbb{C}^r}$, and Theorem 4.7.6(a). □

Every operator convex f therefore gives a divergence that is Lipschitz in the derivative under a two-sided bound on the max-relative entropy. The factor ϵ^{-2} is attained for the χ^2 -divergences, for which $L_f(\epsilon)$ is exact. For logarithmic f the order is ϵ^{-1} (Table 4.2).

Corollary 4.9.12 (Bounded f : rates). *If f is bounded, then for all $\delta \geq 0$,*

$$\omega_{h_f}(\delta) \leq |2b - a|\delta + \int_{(0, 1]} \min\left\{\frac{(1+3s)\delta}{s^2}, \frac{1}{s}\right\} d\mu(s) + \int_{(1, \infty)} \min\{(3+s)\delta, 1\} d\mu(s),$$

with the exact kernel constants of Corollary 4.9.7(a), and the right-hand side tends to 0 as $\delta \rightarrow 0$. If moreover $d\mu(s) \leq Cs^\lambda ds$ on $(0, 1]$ and $d\mu(s) \leq Cs^{-1-\lambda'} ds$ on $(1, \infty)$ for some $\lambda, \lambda' > 0$, then $\omega_{h_f}(\delta) \leq C'\delta^{\min\{\lambda, \lambda', 1\}}$, with an extra factor $\log \frac{1}{\delta}$ if $\min\{\lambda, \lambda'\} = 1$. Here C' depends only on $a, b, C, \lambda, \lambda'$.

Proof. Take $I = [0, 1]$ in Theorem 4.9.8, so that $r_I = 1$ and $d_s = \min\{1, s\}$. For $s \leq 1$, $\frac{2}{s} + \frac{1+s}{s^2} = \frac{1+3s}{s^2}$, and for $s > 1$, $2 + 1 + s = 3 + s$. Since κ_s is convex and nonnegative,

$$\text{osc } \kappa_s \leq \max\{\kappa_s(0), \kappa_s(1)\} = \max\left\{\frac{1}{s}, 1\right\}.$$

The limit is dominated convergence, since $\int_{[0, 1]} s^{-1} d\mu + \mu((1, \infty)) < \infty$. For the rate, use $(1+3s)s^{-2} \leq 4s^{-2}$ on $(0, 1]$ and $3+s \leq 4s$ on $(1, \infty)$, and split $\int_0^1 \min\{4\delta s^{-2}, s^{-1}\} s^\lambda ds$ at $s = 4\delta$, and $\int_1^\infty \min\{4s\delta, 1\} s^{-1-\lambda'} ds$ at $s = 1/(4\delta)$. □

For $f(t) = -t^\alpha$, $\alpha \in (0, 1)$, one has $d\mu = \frac{\sin \pi \alpha}{\pi} \frac{s^\alpha}{(1+s)^2} ds$, so $\lambda = \alpha$ and $\lambda' = 1 - \alpha$. Corollary 4.9.12 then recovers the order $\delta^{\min\{\alpha, 1-\alpha\}}$ of Lemma 4.5.4, with a worse constant.

Example 4.9.13. Table 4.1 collects the operator perspective divergences and their Loewner data, along with the measures their integrals are taken with respect to in their integral representations, as well as the function emerging from the operator perspective version of Theorem 4.2.15. Table 4.2 is its continuation, collecting where the divergences are finite (in terms of the spectrum of the ARN derivatives) and where in this chapter the moduli can be found, along with the optimal Lipschitz constants in the case of bounds in terms of the max-relative entropy.

divergence	$f(x)$	(a, b, c)	$d\mu(s)$	$h_f(t)$
Belavkin–Staszewski	$x \log x$	$(0, 1, 0)$	$\frac{s}{(1+s)^2} ds$	$t \log \frac{t}{1-t}$
reverse BS	$-\log x$	$(0, -1, 0)$	$\frac{1}{(1+s)^2} ds$	$(1-t) \log \frac{1-t}{t}$
Jeffreys	$(x-1) \log x$	$(0, 0, 0)$	$\frac{1}{1+s} ds$	$(2t-1) \log \frac{t}{1-t}$
Pearson χ^2	$(x-1)^2$	$(0, 0, 1)$	0	$\frac{(2t-1)^2}{1-t}$
Neyman χ^2	$\frac{(x-1)^2}{x}$	$(0, 0, 0)$	δ_0	$\frac{(2t-1)^2}{t}$
symmetric χ^2	$(x-1)^2(1 + \frac{1}{x})$	$(0, 0, 1)$	δ_0	$\frac{(2t-1)^2}{t(1-t)}$
Le Cam	$\frac{(x-1)^2}{x+1}$	$(0, 0, 0)$	δ_1	$(2t-1)^2$
Hellinger	$(\sqrt{x} - 1)^2$	$(0, 0, 0)$	$\frac{2}{\pi} \frac{\sqrt{s}}{(1+s)^2} ds$	$(\sqrt{t} - \sqrt{1-t})^2$
Jensen–Shannon	$x \log x - (1+x) \log \frac{1+x}{2}$	$(0, 0, 0)$	$\frac{\min\{s, 1\}}{(1+s)^2} ds$	$\log 2 - H_{\text{bin}}(t)$
Rényi, $\alpha \in (1, 2)$	x^α	$(1, \alpha, 0)$	$\frac{\sin \pi(\alpha-1)}{\pi} \frac{s^\alpha}{(1+s)^2} ds$	$t^\alpha(1-t)^{1-\alpha}$
Rényi, $\alpha \in (-1, 0)$	x^α	$(1, \alpha, 0)$	$-\frac{\sin \pi \alpha}{\pi} \frac{s^\alpha}{(1+s)^2} ds$	$t^\alpha(1-t)^{1-\alpha}$

Table 4.1: Loewner data and extended h_f for the standard examples. For the Jensen–Shannon divergence, H_{bin} is the binary entropy in nats.

divergence	finite on	modulus on that interval	$\text{Lip}^{[\epsilon, 1-\epsilon]}(h_f)$
Belavkin–Staszewski	$[0, 1]$	Cor. 4.6.11	$\log \frac{1-\epsilon}{\epsilon} + \frac{1}{\epsilon} \uparrow$
reverse BS	$(0, 1]$	mirror of Cor. 4.6.11	$\log \frac{1-\epsilon}{\epsilon} + \frac{1}{\epsilon} \uparrow$
Jeffreys	$(0, 1)$	last column	$2 \log \frac{1-\epsilon}{\epsilon} + \frac{1-2\epsilon}{\epsilon(1-\epsilon)}$
Pearson χ^2	$[0, 1)$	$(\epsilon^{-2} - 4)\delta$ if $\epsilon \leq \frac{1}{3}$	$\epsilon^{-2} - 4$
Neyman χ^2	$(0, 1]$	$(\epsilon^{-2} - 4)\delta$ if $\epsilon \leq \frac{1}{3}$	$\epsilon^{-2} - 4$
symmetric χ^2	$(0, 1)$	last column	$\epsilon^{-2} - (1-\epsilon)^{-2}$
Le Cam	$[0, 1]$	4δ	$4(1-2\epsilon)$
Hellinger	$[0, 1]$	$2\sqrt{\delta}$	$\frac{1-2\epsilon}{\sqrt{\epsilon(1-\epsilon)}}$
Jensen–Shannon	$[0, 1]$	$2\delta \log \frac{1}{\delta} + 4 \log 2 \cdot \delta$	$\log \frac{1-\epsilon}{\epsilon}$
Rényi, $\alpha \in (1, 2)$	$[0, 1)$	Cor. 4.6.12	$(\alpha - 1 + \epsilon) \frac{(1-\epsilon)^{\alpha-1}}{\epsilon^\alpha} \uparrow$
Rényi, $\alpha \in (-1, 0)$	$(0, 1]$	mirror of Cor. 4.6.12	$(\epsilon - \alpha) \frac{(1-\epsilon)^{-\alpha}}{\epsilon^{1-\alpha}} \uparrow$

Table 4.2: Moduli and optimal Lipschitz constants. In the third column, ϵ is the distance of the spectrum of T from the infinite endpoints, and $\delta = \|T_1 - T_2\|_\infty$. The last column is the exact operator Lipschitz constant of h_f on $[\epsilon, 1-\epsilon]$, except for the entries marked $\uparrow$. These are lower bounds.

4.9.4 All Kubo–Ando means

We now repeat the above process using the integral representation of the Kubo–Ando means, as for the operator perspectives. Every operator connection σ has a representing function of the form [79]

$$f_\sigma(x) = a_\sigma + b_\sigma x + \int_{(0,\infty)} \frac{x(1+s)}{x+s} dm_\sigma(s), \quad a_\sigma, b_\sigma \geq 0,$$

with a finite positive measure m_σ ; then $a_\sigma = f_\sigma(0^+)$, $b_\sigma = f'_\sigma(\infty)$, and σ is a mean if and only if $a_\sigma + b_\sigma + m_\sigma((0, \infty)) = 1$. The function $\varphi_\sigma(t) = t f_\sigma((1-t)/t)$ is therefore

$$\varphi_\sigma(t) = a_\sigma t + b_\sigma (1-t) + \int_{(0,\infty)} \eta_s(t) dm_\sigma(s), \quad (4.9.5)$$

where

$$\eta_s(t) = \frac{(1+s)t(1-t)}{1+(s-1)t} = \left(\frac{1}{1+s} t^{-1} + \frac{s}{1+s} (1-t)^{-1} \right)^{-1},$$

a mixture of weighted harmonic means of t and $1-t$.

By Corollary 4.9.7(c),

$$\text{Lip}^{[0,1]}(\eta_s) = \max_{[0,1]} |\eta'_s| = (1+s) \max\{1, s^{-1}\}$$

for $s > 0$.

Lemma 4.9.14. *For $s > 0$, $0 \leq \eta_s \leq \frac{1+s}{(1+\sqrt{s})^2} \leq 1$ on $[0, 1]$.*

Proof. We have $\eta_s(0) = \eta_s(1) = 0$ and

$$\eta'_s(t) = \frac{(1+s)(1-2t-(s-1)t^2)}{(1+(s-1)t)^2}.$$

The numerator is positive at $t = 0$, negative at $t = 1$, and a polynomial of degree at most 2, so its only zero in $(0, 1)$ is $t_0 = 1/(1+\sqrt{s})$. Hence $\max_{[0,1]} \eta_s = \eta_s(t_0) = (1+s)/(1+\sqrt{s})^2 \leq 1$. $\square$

Theorem 4.9.15. *For every operator connection σ with representation as above,*

$$\begin{aligned} \omega_{\varphi_\sigma}(\delta) &\leq |a_\sigma - b_\sigma| \delta + \int_{(0,1]} \min\left\{\frac{(1+s)\delta}{s}, 1\right\} dm_\sigma(s) \\ &\quad + \int_{(1,\infty)} \min\{(1+s)\delta, 1\} dm_\sigma(s) \xrightarrow{\delta \rightarrow 0} 0. \end{aligned}$$

Consequently Theorem 4.4.5 and Theorem 4.10.4(a) hold with this explicit modulus for every mean. If $dm_\sigma(s) \leq Cs^{\lambda-1}ds$ on $(0, 1]$ and $dm_\sigma(s) \leq Cs^{-\lambda'-1}ds$ on $(1, \infty)$ with $\lambda, \lambda' \in (0, 1)$, then $\omega_{\varphi_\sigma}(\delta) \leq C'\delta^{\min\{\lambda, \lambda'\}}$.

Proof. Insert (4.9.5) into $\varphi_\sigma(X) - \varphi_\sigma(Y)$; the affine part has slope $a_\sigma - b_\sigma$. Use Corollary 4.9.7(c), together with the bound $\|\eta_s(X) - \eta_s(Y)\|_\infty \leq \max \eta_s \leq 1$ of Lemma 4.9.14. The rate follows as in Corollary 4.9.12. $\square$

4.10 Pairs with different sums

We now come to the more general and noncommutative setting. In general, two pairs of completely positive maps need not have the same sum. In this section we use the sum of all four maps as the dominating map. In this noncommutative setting the inequalities are weaker than before, and an extra term in the Thompson metric appears. This term behaves well enough as a continuity constant.

Let $\Phi_1, \Psi_1, \Phi_2, \Psi_2 \in \text{CP}(\mathcal{X}, \mathcal{Y})$, put $\Omega_i = \Phi_i + \Psi_i$, $\Omega' = \Omega_1 + \Omega_2$ and $\bar{\Omega} = \frac{1}{2}\Omega'$, and assume $\Omega' \neq 0$. Let $\Omega'(X) = \sum_{k=1}^r V_k X V_k^*$ with linearly independent $V_1, \dots, V_r$ and $W = [\text{vec}(V_1) \cdots \text{vec}(V_r)]$. Each of the four maps lies below Ω' in the completely positive order, so by Theorem 4.2.2 there are unique $0 \leq A_i, B_i \leq \mathbf{1}_{\mathbb{C}^r}$ with $\Phi_i = \Psi_{A_i}$ and $\Psi_i = \Psi_{B_i}$, all taken with respect to $V_1, \dots, V_r$. We put

$$C_i = A_i + B_i, \quad \gamma = \|C_1 - \frac{1}{2}\mathbf{1}_{\mathbb{C}^r}\|_\infty, \quad \delta = \max\{\|A_1 - A_2\|_\infty, \|B_1 - B_2\|_\infty\}.$$

Since $\Psi_{A_1+B_1+A_2+B_2} = \Omega' = \Psi_{\mathbf{1}_{\mathbb{C}^r}}$, uniqueness in Theorem 4.2.2 gives $C_1 + C_2 = \mathbf{1}_{\mathbb{C}^r}$; in particular C_1 and C_2 commute and

$$\gamma = \|C_2 - \frac{1}{2}\mathbf{1}_{\mathbb{C}^r}\|_\infty = \frac{1}{2}\|C_1 - C_2\|_\infty.$$

Note that $\|\Omega'^*(\mathbf{1}_{\mathcal{Y}})\|_\infty = 2\|\bar{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty$. By Lemma 4.8.1, applied to Ω' with $K = A_1 - A_2$ and $K = B_1 - B_2$, δ is the smallest $d \geq 0$ such that $-d\Omega' \leq \Phi_1 - \Phi_2 \leq d\Omega'$ and $-d\Omega' \leq \Psi_1 - \Psi_2 \leq d\Omega'$.

Thus δ measures the distance between the two pairs *relative* to their total Ω' , in the completely positive order. The second parameter is the distance between the two sums. With $D_{\max}$ as in Lemma 4.6.5, we define the Thompson distance

$$d_T(\Omega_1, \Omega_2) = \max\{D_{\max}(\Omega_1\|\Omega_2), D_{\max}(\Omega_2\|\Omega_1)\} \in [0, +\infty].$$

This is the natural object to introduce, as in the commuting case, the infinite behaviour of the divergences was controlled by the max-relative entropy. Physically (Remark 4.6.6), $d_T(\Omega_1, \Omega_2) \leq d$ says that every outcome probability of the device with its outcome forgotten, on every input with ancilla, is reproduced within a factor $e^{\pm d}$: a multiplicative calibration error. Theorem 4.10.4 separates the in-pair change τ from this drift, and Corollary 4.10.8 and Lemma 4.10.5 bound both by the additive errors $\|\Phi_1 - \Phi_2\|_1$ and $\|\Psi_1 - \Psi_2\|_1$ (through δ , when $\delta < \frac{1}{2}$).

Lemma 4.10.1. *The following are equivalent:*

- (i) $d_T(\Omega_1, \Omega_2) < \infty$;
- (ii) $\text{range } J(\Omega_1) = \text{range } J(\Omega_2)$;
- (iii) C_1 and C_2 are invertible;
- (iv) $\gamma < \frac{1}{2}$.

When they hold,

$$2\gamma = \|C_1 - C_2\|_\infty = \tanh\left(\frac{1}{2}d_T(\Omega_1, \Omega_2)\right).$$

Proof. Let λ_- and λ_+ be the smallest and largest eigenvalues of C_1 , so that $\gamma = \max\{\lambda_+ - \frac{1}{2}, \frac{1}{2} - \lambda_-\}$ and the spectrum of $C_2 = \mathbf{1}_{\mathbb{C}^r} - C_1$ is $\{1 - \lambda : \lambda \in \text{spec}(C_1)\}$. As in the proof of Lemma 4.6.5, $\Omega_1 \leq c\Omega_2$ if and only if $C_1 \leq cC_2 = c(\mathbf{1}_{\mathbb{C}^r} - C_1)$, that is, $(1+c)C_1 \leq c\mathbf{1}_{\mathbb{C}^r}$,

that is, $\lambda_+ \leq c/(1+c)$. Hence $D_{\max}(\Omega_1\|\Omega_2) = \log \frac{\lambda_+}{1-\lambda_+}$ if $\lambda_+ < 1$ and $+\infty$ if $\lambda_+ = 1$; similarly $D_{\max}(\Omega_2\|\Omega_1) = \log \frac{1-\lambda_-}{\lambda_-}$ if $\lambda_- > 0$ and $+\infty$ if $\lambda_- = 0$. Thus (i) holds if and only if $0 < \lambda_- \leq \lambda_+ < 1$, which is (iii), and (iii) is equivalent to (iv) because $\gamma < \frac{1}{2}$ means exactly $\lambda_- > 0$ and $\lambda_+ < 1$.

For (ii), Theorem 4.2.2 gives

$$J(\Omega_i) = WC_i W^* = (WC_i^{1/2})(WC_i^{1/2})^*,$$

whose range is $W(\text{range } C_i)$; since W is injective, (ii) holds if and only if $\text{range } C_1 = \text{range } C_2$. If C_1 and C_2 are invertible this is clear. Conversely, if $\text{range } C_1 = \text{range } C_2 = \mathcal{R}$, then $\mathcal{R} \supseteq \text{range}(C_1 + C_2) = \mathbb{C}^r$, so both are invertible. Finally, $\tanh(\frac{1}{2} \log x) = \frac{x-1}{x+1}$ gives $\tanh(\frac{1}{2} \log \frac{\lambda_+}{1-\lambda_+}) = 2\lambda_+ - 1$ and $\tanh(\frac{1}{2} \log \frac{1-\lambda_-}{\lambda_-}) = 1 - 2\lambda_-$; since $\tanh$ is increasing,

$$\tanh(\frac{1}{2} d_T) = \max\{2\lambda_+ - 1, 1 - 2\lambda_-\} = 2\gamma,$$

and

$$\|C_1 - C_2\|_\infty = \|2C_1 - \mathbf{1}_{\mathbb{C}^r}\|_\infty = 2\gamma.$$

□

For the rest of this section we assume that $\text{range } J(\Omega_1) = \text{range } J(\Omega_2)$, that is, that the two sums have the same Choi support. By Lemma 4.10.1, C_1 and C_2 are then invertible, and we put

$$a_1 = C_1^{1/2}, \quad a_2 = C_2^{1/2}, \quad T_i = C_i^{-1/2} A_i C_i^{-1/2}, \quad \tau = \|T_1 - T_2\|_\infty.$$

Lemma 4.10.2. *Under this assumption, $0 \leq T_i \leq \mathbf{1}_{\mathbb{C}^r}$ and $\mathbf{1}_{\mathbb{C}^r} - T_i = C_i^{-1/2} B_i C_i^{-1/2}$. Moreover, T_i is the ARN derivative of Φ_i with respect to Ω_i , relative to the Kraus decomposition of Ω_i whose column matrix is $WC_i^{1/2}$. Finally, $a_1^2 + a_2^2 = \mathbf{1}_{\mathbb{C}^r}$, the operators a_1 and a_2 commute, and*

$$\|a_1 - a_2\|_\infty \leq 2\gamma, \quad \|a_1 + a_2\|_\infty \leq \sqrt{2}.$$

Proof. Positivity of A_i and B_i gives $T_i \geq 0$ and $C_i^{-1/2} B_i C_i^{-1/2} \geq 0$, and these two operators add up to $C_i^{-1/2} C_i C_i^{-1/2} = \mathbf{1}_{\mathbb{C}^r}$. The columns of $WC_i^{1/2}$ are $\text{vec}(\tilde{V}_j)$, where $\tilde{V}_j = \sum_k (C_i^{1/2})_{kj} V_k$; these are linearly independent because $C_i^{1/2}$ is invertible. By Theorem 4.2.2,

$$J(\Omega_i) = WC_i W^* = (WC_i^{1/2})(WC_i^{1/2})^*,$$

so the $\tilde{V}_j$ form a Kraus decomposition of Ω_i , and

$$J(\Phi_i) = WA_i W^* = (WC_i^{1/2}) T_i (WC_i^{1/2})^*$$

identifies T_i as the derivative of Φ_i in this decomposition. The operators a_1 and a_2 are functions of C_1 , hence commute, and $a_1^2 + a_2^2 = C_1 + C_2 = \mathbf{1}_{\mathbb{C}^r}$. If λ runs over the spectrum of C_1 , the eigenvalues of $a_1 \mp a_2$ are $\sqrt{\lambda} \mp \sqrt{1-\lambda}$, and

$$|\sqrt{\lambda} - \sqrt{1-\lambda}| = \frac{|2\lambda - 1|}{\sqrt{\lambda} + \sqrt{1-\lambda}} \leq |2\lambda - 1| \leq 2\gamma, \quad \sqrt{\lambda} + \sqrt{1-\lambda} \leq \sqrt{2},$$

by $\sqrt{\lambda} + \sqrt{1-\lambda} \geq 1$ and the Cauchy–Schwarz inequality.

□

Lemma 4.10.3. *Let $\varphi : [0, 1] \rightarrow \mathbb{R}$ be continuous, with $\|\varphi\|_\infty = \max_{[0,1]} |\varphi|$. Then*

$$\|a_1 \varphi(T_1) a_1 - a_2 \varphi(T_2) a_2\|_\infty \leq \frac{1}{2} \omega_\varphi(\tau) + \frac{2\sqrt{2}\gamma}{(1+s_\gamma)^{1/2}} \|\varphi\|_\infty, \quad s_\gamma = \sqrt{1-4\gamma^2},$$

and in particular

$$\|a_1 \varphi(T_1) a_1 - a_2 \varphi(T_2) a_2\|_\infty \leq \frac{1}{2} \omega_\varphi(\tau) + 2\sqrt{2}\gamma \|\varphi\|_\infty.$$

The coefficient of the second term is sharp as $\gamma \rightarrow 0$. If $T_1, T_2 \leq (1-\epsilon)\mathbf{1}_{\mathbb{C}^r}$ for some $\epsilon \in (0, 1)$ and φ is only assumed continuous on $[0, 1-\epsilon]$, the same holds with ω_φ replaced by $\omega_\varphi^{(\epsilon)}$ and $\|\varphi\|_\infty$ by $\max_{[0,1-\epsilon]} |\varphi|$. More generally, if $\text{spec}(T_1), \text{spec}(T_2) \subseteq I$ for a compact interval I on which φ is continuous, the same holds with ω_φ^I and $\max_I |\varphi|$.

Proof. Put $H_- = \varphi(T_1) - \varphi(T_2)$ and $H_+ = \varphi(T_1) + \varphi(T_2)$, both Hermitian, with $\|H_-\|_\infty \leq \omega_\varphi(\tau)$ and $\|H_+\|_\infty \leq 2\|\varphi\|_\infty$. Then

$$\begin{aligned} a_1 \varphi(T_1) a_1 - a_2 \varphi(T_2) a_2 &= \frac{1}{2} (a_1 H_- a_1 + a_2 H_- a_2) \\ &\quad + \frac{1}{4} \left((a_1 - a_2) H_+ (a_1 + a_2) + (a_1 + a_2) H_+ (a_1 - a_2) \right), \end{aligned}$$

as one checks by expanding the right-hand side. Since $-\|H_-\|_\infty \mathbf{1}_{\mathbb{C}^r} \leq H_- \leq \|H_-\|_\infty \mathbf{1}_{\mathbb{C}^r}$ and $a_1^2 + a_2^2 = \mathbf{1}_{\mathbb{C}^r}$, conjugating by a_1 and by a_2 and adding gives

$$-\|H_-\|_\infty \mathbf{1}_{\mathbb{C}^r} \leq a_1 H_- a_1 + a_2 H_- a_2 \leq \|H_-\|_\infty \mathbf{1}_{\mathbb{C}^r},$$

so the first term has norm at most $\frac{1}{2}\|H_-\|_\infty \leq \frac{1}{2}\omega_\varphi(\tau)$. For the second term $Y = \frac{1}{4}(vH_+u + uH_+v)$, with $u = a_1 + a_2$ and $v = a_1 - a_2$, let $e_1, \dots, e_r$ be an orthonormal eigenbasis of C_1 with $C_1 e_i = \lambda_i e_i$, and write $\lambda_i = \cos^2 \beta_i$ with $\beta_i \in [0, \frac{\pi}{2}]$ and $\psi_i = \frac{\pi}{4} - \beta_i \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. Then $a_1 e_i = \cos \beta_i e_i$ and $a_2 e_i = \sin \beta_i e_i$, so $u e_i = \sqrt{2} \cos \psi_i e_i$ and $v e_i = \sqrt{2} \sin \psi_i e_i$, and

$$\langle e_i, Y e_j \rangle = \frac{1}{4} (H_+)_{ij} (2 \sin \psi_i \cos \psi_j + 2 \cos \psi_i \sin \psi_j) = \frac{1}{2} (H_+)_{ij} \sin(\psi_i + \psi_j).$$

Let $m = \max_i |\sin \psi_i|$; if $m = 0$ then $Y = 0$. Otherwise put $x_i = (m^{-1/2} \sin \psi_i, m^{1/2} \cos \psi_i)$ and $y_j = (m^{1/2} \cos \psi_j, m^{-1/2} \sin \psi_j)$ in $\mathbb{R}^2$, so that $\langle x_i, y_j \rangle = \sin(\psi_i + \psi_j)$ and $\|x_i\|^2, \|y_j\|^2 \leq m + m = 2m$. With the operators $X, Y': \mathbb{C}^r \rightarrow \mathbb{C}^r \otimes \mathbb{C}^2$ given by $X e_i = e_i \otimes x_i$ and $Y' e_j = e_j \otimes y_j$ we have $Y = \frac{1}{2} X^* (H_+ \otimes \mathbf{1}_{\mathbb{C}^2}) Y'$, and $\|X\|_\infty, \|Y'\|_\infty \leq (2m)^{1/2}$ because the vectors $e_i \otimes x_i$ are pairwise orthogonal. Hence $\|Y\|_\infty \leq m \|H_+\|_\infty \leq 2m \|\varphi\|_\infty$. Finally $\sin 2\psi_i = \cos 2\beta_i = 2\lambda_i - 1$, so $|\sin 2\psi_i| \leq 2\gamma$; as $|2\psi_i| \leq \frac{\pi}{2}$, this gives $\cos 2\psi_i \geq s_\gamma$ and

$$\sin^2 \psi_i = \frac{1}{2} (1 - \cos 2\psi_i) \leq \frac{1}{2} (1 - s_\gamma) = \frac{2\gamma^2}{1 + s_\gamma}.$$

Thus $m \leq \sqrt{2}\gamma(1+s_\gamma)^{-1/2}$, which gives the first inequality, and the second follows from $s_\gamma \geq 0$. For $\varphi \equiv 1$ and $T_1 = T_2$ the left-hand side is $\|C_1 - C_2\|_\infty = 2\gamma$, while the bound is $2\gamma(\frac{2}{1+s_\gamma})^{1/2} = 2\gamma(1+O(\gamma^2))$; this is the sharpness assertion. The other statements are proved in the same way. $\square$

Theorem 4.10.4. *Let $\Phi_1, \Psi_1, \Phi_2, \Psi_2 \in \text{CP}(\mathcal{X}, \mathcal{Y})$ be such that $\text{range } J(\Phi_1 + \Psi_1) = \text{range } J(\Phi_2 + \Psi_2)$, and let $\bar{\Omega}, \tau$ and $d_T = d_T(\Phi_1 + \Psi_1, \Phi_2 + \Psi_2)$ be as above.*

(a) *For every Kubo–Ando mean σ ,*

$$\|\Phi_1 \sigma \Psi_1 - \Phi_2 \sigma \Psi_2\|_1 \leq \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \left(\omega_{g_\sigma}(\tau) + 2\sqrt{2} \|g_\sigma\|_\infty \tanh \frac{d_T}{2} \right).$$

(b) For the geometric channel fidelity,

$$|\hat{F}(\Phi_1, \Psi_1) - \hat{F}(\Phi_2, \Psi_2)| \leq \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \left(\tau^{1/2} + \sqrt{2} \tanh \frac{d_T}{2} \right).$$

(c) For $\alpha \in (0, 1)$,

$$|\hat{Q}_\alpha(\Phi_1 \| \Psi_1) - \hat{Q}_\alpha(\Phi_2 \| \Psi_2)| \leq \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \left(\tau^\alpha + \tau^{1-\alpha} + 2\sqrt{2} \tanh \frac{d_T}{2} \right).$$

(d) For f as in Lemma 4.5.12,

$$|\hat{D}_f(\Phi_1 \| \Psi_1) - \hat{D}_f(\Phi_2 \| \Psi_2)| \leq \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \left(\omega_{h_f}(\tau) + 2\sqrt{2} \|h_f\|_\infty \tanh \frac{d_T}{2} \right).$$

(e) For f as in Section 4.6, if moreover $D_{\max}(\Phi_i \| \Psi_i) \leq \log \frac{1-\epsilon}{\epsilon}$ for $i = 1, 2$ and some $\epsilon \in (0, 1)$, then (d) holds with ω_{h_f} replaced by $\omega_{h_f}^{(\epsilon)}$ and $\|h_f\|_\infty$ by $\max_{[0, 1-\epsilon]} |h_f|$.

(f) For every operator convex f and every compact interval $I \subseteq [0, 1]$ on which h_f is finite, if $\text{spec}(T_i) \subseteq I$ for $i = 1, 2$, then

$$|\hat{D}_f(\Phi_1 \| \Psi_1) - \hat{D}_f(\Phi_2 \| \Psi_2)| \leq \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \left(\omega_{h_f}^I(\tau) + 2\sqrt{2} \max_I |h_f| \tanh \frac{d_T}{2} \right).$$

For $I = [\epsilon, 1 - \epsilon]$ with $\epsilon \in (0, \frac{1}{2})$ the hypothesis says

$$\max\{D_{\max}(\Phi_i \| \Psi_i), D_{\max}(\Psi_i \| \Phi_i)\} \leq \log \frac{1 - \epsilon}{\epsilon},$$

and Corollary 4.9.11 gives $\omega_{h_f}^I(\tau) \leq L_f(\epsilon)\tau$.

If $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$, then $d_T = 0$, $\bar{\Omega} = \Phi_1 + \Psi_1$, and $\tau = \|T_1 - T_2\|_\infty$ for the derivatives T_i of Φ_i with respect to this common sum, so that (a)–(e) reduce to Theorems 4.4.5, 4.4.6, 4.5.5, 4.5.13(b) and 4.6.7(b) with the same constants.

Proof. By Lemma 4.10.2, $A_i = a_i T_i a_i$ and $B_i = a_i (\mathbf{1}_{\mathbb{C}^r} - T_i) a_i$ for $i = 1, 2$. By Lemma 4.10.1, $2\gamma = \tanh \frac{d_T}{2}$, so Lemma 4.10.3 reads

$$\|a_1 \varphi(T_1) a_1 - a_2 \varphi(T_2) a_2\|_\infty \leq \frac{1}{2} \omega_\varphi(\tau) + \sqrt{2} \|\varphi\|_\infty \tanh \frac{d_T}{2}. \quad (4.10.1)$$

Let σ' be an operator connection. By (4.2.1), which holds for connections by Remark 4.2.16, $J(\Phi_i) \sigma' J(\Psi_i) = W(A_i \sigma' B_i) W^*$. The transformer equality for the invertible operator a_i [79], followed by Remark 4.2.16(2) for the commuting pair $T_i, \mathbf{1}_{\mathbb{C}^r} - T_i$, gives

$$A_i \sigma' B_i = a_i (T_i \sigma' (\mathbf{1}_{\mathbb{C}^r} - T_i)) a_i = a_i \varphi_{\sigma'}(T_i) a_i,$$

and in the same way $B_i \sigma' A_i = a_i \varphi_{\sigma'}(\mathbf{1}_{\mathbb{C}^r} - T_i) a_i$. For a mean σ we have $\varphi_\sigma = g_\sigma$.

The partial-trace estimate. If $N_1, N_2 \in L(\mathbb{C}^r)$ are Hermitian, Lemma 4.4.10, applied to Ω' , gives

$$\begin{aligned} \|\text{Tr}_Y(W N_1 W^*) - \text{Tr}_Y(W N_2 W^*)\|_\infty &\leq \|\Omega'^*(\mathbf{1}_Y)\|_\infty \|N_1 - N_2\|_\infty \\ &= 2\|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \|N_1 - N_2\|_\infty, \end{aligned}$$

and the same bound holds for the difference of the smallest, or of the largest, eigenvalues. The factor 2 combines with the factor $\frac{1}{2}$ in (4.10.1).

(a) By definition of the mean and the first step,

$$J(\Phi_i \sigma \Psi_i) = J(\Phi_i) \sigma J(\Psi_i) = W(A_i \sigma B_i)W^*,$$

so $\Phi_i \sigma \Psi_i = \Psi_{N_i}$ with respect to $V_1, \dots, V_r$, where $N_i = a_i g_\sigma(T_i) a_i$. Theorem 4.8.4, applied to Ω' , gives

$$\|\Psi_{N_1} - \Psi_{N_2}\|_1 \leq \|\Omega'^*(\mathbf{1}_Y)\|_\infty \|N_1 - N_2\|_\infty.$$

Now apply (4.10.1) with $\varphi = g_\sigma$.

(b) By Proposition 4.3.1 and the first step, $\hat{F}(\Phi_i, \Psi_i) = \lambda_{\min}(\text{Tr}_Y(W N_i W^*))$ with $N_i = a_i g_\#(T_i) a_i$. Apply the partial-trace estimate and (4.10.1) with $\varphi = g_\#$, using $\omega_{g_\#}(t) \leq t^{1/2}$ (Proposition 4.4.3(2)) and $\|g_\#\|_\infty = \frac{1}{2}$.

(c) By Lemma 4.5.3,

$$\hat{Q}_\alpha(\Phi_i \| \Psi_i) = \lambda_{\min}(\text{Tr}_Y(J(\Psi_i) \#_\alpha J(\Phi_i))),$$

and since $g_{\#1-\alpha} = h_\alpha$, the first step gives

$$J(\Psi_i) \#_\alpha J(\Phi_i) = J(\Phi_i) \#_{1-\alpha} J(\Psi_i) = W a_i h_\alpha(T_i) a_i W^*.$$

Apply the partial-trace estimate and (4.10.1) with $\varphi = h_\alpha$, using $\omega_{h_\alpha}(t) \leq t^\alpha + t^{1-\alpha}$ (Lemma 4.5.4) and $\|h_\alpha\|_\infty \leq 1$.

(d) By Lemma 4.10.2, T_i is the derivative of Φ_i with respect to Ω_i , relative to the Kraus decomposition with column matrix $W a_i$. Theorem 4.5.13(a), applied to Ω_i with this matrix in place of W , gives

$$\hat{D}_f(\Phi_i \| \Psi_i) = \lambda_{\max}(\text{Tr}_Y(W a_i h_f(T_i) a_i W^*)).$$

Apply the partial-trace estimate and (4.10.1) with $\varphi = h_f$.

(e) Put $c = (1 - \epsilon)/\epsilon$. The set of $c' \geq 0$ with $\Phi_i \leq c' \Psi_i$ is closed and upward closed, so the definition of $D_{\max}$ in Lemma 4.6.5 gives $\Phi_i \leq c \Psi_i$, that is, $W(c B_i - A_i)W^* \geq 0$ by Choi's theorem; conjugating by W^+ gives $c B_i \geq A_i$, and conjugating by a_i^{-1} gives $c(\mathbf{1}_{\mathbb{C}^r} - T_i) \geq T_i$, that is, $T_i \leq (1 - \epsilon)\mathbf{1}_{\mathbb{C}^r}$. In particular B_i is invertible, so Lemma 4.6.3 with $C = W$, Lemma 4.6.1 with $R = a_i$, and the computation in the proof of Theorem 4.6.7(a) give

$$\mathcal{P}_f(J(\Phi_i), J(\Psi_i)) = W \mathcal{P}_f(A_i, B_i) W^* = W a_i h_f(T_i) a_i W^*.$$

The remainder of the proof of Theorem 4.6.7(a) uses only $\Phi_i \leq c \Psi_i$ and this representation, and shows that

$$\hat{D}_f(\Phi_i \| \Psi_i) = \lambda_{\max}(\text{Tr}_Y(W a_i h_f(T_i) a_i W^*)).$$

Apply the partial-trace estimate and the second statement of Lemma 4.10.3, in the form (4.10.1), with $\varphi = h_f$.

(f) By Lemma 4.10.2, T_i is the derivative of Φ_i with respect to Ω_i , relative to the Kraus decomposition with column matrix $W a_i$. Theorem 4.7.5(b), applied to Ω_i with this matrix in place of W , gives

$$\hat{D}_f(\Phi_i \| \Psi_i) = \lambda_{\max}(\text{Tr}_Y(W a_i h_f(T_i) a_i W^*)),$$

since $h_f(T_i)$ is finite. Apply the partial-trace estimate and the last statement of Lemma 4.10.3 with $\varphi = h_f$. The identification of the hypothesis for $I = [\epsilon, 1 - \epsilon]$ is Lemma 4.6.5 applied within each pair, in the frame $W a_i$, to T_i and to $\mathbf{1}_{\mathbb{C}^r} - T_i$.

Finally, if $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2 = \Omega$, then $\Omega' = 2\Omega$ and $d_T = 0$; the derivatives with respect to Ω' are $A_i = \frac{1}{2}T_i$ and $B_i = \frac{1}{2}(\mathbf{1}_{\mathbb{C}^r} - T_i)$, where T_i is the derivative of Φ_i with respect to Ω , so $C_1 = C_2 = \frac{1}{2}\mathbf{1}_{\mathbb{C}^r}$ and the operators T_i defined above coincide with these derivatives. $\square$

Since $2\gamma = \tanh \frac{d_T}{2}$ and $(1 - 4\gamma^2)^{1/2} = \operatorname{sech} \frac{d_T}{2}$, the first inequality of Lemma 4.10.3 shows that each of the estimates (a)–(f) of Theorem 4.10.4 remains valid with $\tanh \frac{d_T}{2}$ replaced by $\tanh \frac{d_T}{2} (1 + \operatorname{sech} \frac{d_T}{2})^{-1/2}$, which is smaller by a factor between $2^{-1/2}$ (as $d_T \rightarrow 0$) and 1.

Lemma 4.10.5. *Without any assumption on the supports, $\gamma \leq \delta$. If $\delta < \frac{1}{2}$, then $\operatorname{range} J(\Omega_1) = \operatorname{range} J(\Omega_2)$ and*

$$\tau \leq \frac{2\sqrt{2}\delta}{\sqrt{1-4\gamma^2}} \leq \frac{2\sqrt{2}\delta}{\sqrt{1-4\delta^2}}.$$

The constant $2\sqrt{2}$ is optimal: in Example 4.10.6, $\tau/\delta \rightarrow 2\sqrt{2}$ as $\delta \rightarrow 0$.

Proof. We have

$$2\gamma = \|C_1 - C_2\|_\infty \leq \|A_1 - A_2\|_\infty + \|B_1 - B_2\|_\infty \leq 2\delta.$$

If $\delta < \frac{1}{2}$ then $\gamma < \frac{1}{2}$, and Lemma 4.10.1 gives the support condition. Put $u = a_1 + a_2$ and $v = a_1 - a_2$, which commute with C_1 , and

$$\bar{T} = \frac{1}{2}(T_1 + T_2), \quad \bar{T}' = \mathbf{1}_{\mathbb{C}^r} - \bar{T}, \quad M = \bar{T} - \bar{T}', \quad T_- = T_1 - T_2.$$

Since $0 \leq T_i \leq \mathbf{1}_{\mathbb{C}^r}$, also $0 \leq \bar{T}, \bar{T}' \leq \mathbf{1}_{\mathbb{C}^r}$. By Lemma 4.10.2, $A_i - B_i = a_i(2T_i - \mathbf{1}_{\mathbb{C}^r})a_i$, and $2T_1 - \mathbf{1}_{\mathbb{C}^r} = M + T_-$, $2T_2 - \mathbf{1}_{\mathbb{C}^r} = M - T_-$, so

$$(A_1 - A_2) - (B_1 - B_2) = (a_1 M a_1 - a_2 M a_2) + (a_1 T_- a_1 + a_2 T_- a_2).$$

For every X , $a_1 X a_1 + a_2 X a_2 = \frac{1}{2}(u X u + v X v)$ and $a_1 X a_1 - a_2 X a_2 = \frac{1}{2}(u X v + v X u)$, as one checks by expanding. Moreover

$$uv = a_1^2 - a_2^2 = C_1 - C_2 = (A_1 - A_2) + (B_1 - B_2).$$

Put

$$\tilde{E} = u^{-1}(A_1 - A_2)u^{-1}, \quad \tilde{F} = u^{-1}(B_1 - B_2)u^{-1}, \quad N = u^{-1}v.$$

Since u and v commute, N is Hermitian and $N = u^{-1}(uv)u^{-1} = \tilde{E} + \tilde{F}$. Conjugating the displayed identity by u^{-1} gives

$$\tilde{E} - \tilde{F} = \frac{1}{2}(NM + MN) + \frac{1}{2}(T_- + NT_-N).$$

Since $M = \bar{T} - \bar{T}'$ and $\mathbf{1}_{\mathbb{C}^r} = \bar{T} + \bar{T}'$, this is the exact identity

$$T_- + NT_-N = 2Y, \quad Y = \tilde{E}\bar{T}' + \bar{T}'\tilde{E} - \tilde{F}\bar{T} - \bar{T}\tilde{F}. \quad (4.10.2)$$

Write

$$Y = \begin{bmatrix} \bar{T}' & -\bar{T} \end{bmatrix} \begin{bmatrix} \tilde{E} \\ \tilde{F} \end{bmatrix} + \begin{bmatrix} \tilde{E} & -\tilde{F} \end{bmatrix} \begin{bmatrix} \bar{T}' \\ \bar{T} \end{bmatrix}$$

as a sum of products of a row and a column of operators. We have

$$\|[\bar{T}' \quad -\bar{T}]\|_\infty^2 = \|\bar{T}'^2 + \bar{T}^2\|_\infty \leq \|\bar{T}' + \bar{T}\|_\infty = 1,$$

and likewise for the column $(\bar{T}', \bar{T})$. With $\tilde{\delta} = \max\{\|\tilde{E}\|_\infty, \|\tilde{F}\|_\infty\}$, the column $(\tilde{E}, \tilde{F})$ and the row $[\tilde{E} \quad -\tilde{F}]$ have norm at most $\sqrt{2}\tilde{\delta}$. Hence $\|Y\|_\infty \leq 2\sqrt{2}\tilde{\delta}$.

Let λ run over the spectrum of C_1 , which lies in $[\frac{1}{2} - \gamma, \frac{1}{2} + \gamma]$, and put $\theta_\lambda = 2\sqrt{\lambda(1-\lambda)}$ and $s_\gamma = \sqrt{1-4\gamma^2}$. Since $4\lambda(1-\lambda) = 1 - (2\lambda - 1)^2$, we have $\theta_\lambda \geq s_\gamma$. The eigenvalues of $u^2 = \mathbf{1}_{\mathbb{C}^r} + 2a_1a_2$ are $1 + \theta_\lambda \geq 1 + s_\gamma$, so $\tilde{\delta} \leq \delta/(1 + s_\gamma)$. Those of $N^2 = v^2u^{-2}$ are $(1 - \theta_\lambda)/(1 + \theta_\lambda) \leq (1 - s_\gamma)/(1 + s_\gamma)$. By (4.10.2), $\|T_-\|_\infty \leq 2\|Y\|_\infty + \|N\|_\infty^2\|T_-\|_\infty$, and therefore

$$\tau = \|T_-\|_\infty \leq \frac{2\|Y\|_\infty}{1 - \|N\|_\infty^2} \leq 2 \cdot \frac{2\sqrt{2}\delta}{1 + s_\gamma} \cdot \frac{1 + s_\gamma}{2s_\gamma} = \frac{2\sqrt{2}\delta}{s_\gamma}.$$

The second inequality of the statement follows from $\gamma \leq \delta$. $\square$

Example 4.10.6. Let $r = 2$, $Z = \text{diag}(1, -1)$, $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and let $U_0 = (Z + X)/\sqrt{2}$ and $U_1 = (Z - X)/\sqrt{2}$, which are symmetric unitaries. For $0 < \delta < \frac{1}{2}$ put $T_0 = \text{diag}(\delta, 1 - \delta)$ and

$$A_{1,2} = \frac{1}{2}(T_0 \pm \delta U_0), \quad B_{1,2} = \frac{1}{2}(\mathbf{1}_{\mathbb{C}^2} - T_0 \pm \delta U_1).$$

These are positive semidefinite, since all four have positive diagonal entries and determinant at least $\frac{\sqrt{2}-1}{4\sqrt{2}}\delta(1-2\delta) > 0$. They sum to $\mathbf{1}_{\mathbb{C}^2}$, and $\|A_1 - A_2\|_\infty = \|B_1 - B_2\|_\infty = \delta$; by Theorem 4.2.2 they are the derivatives of four completely positive maps with respect to any linearly independent pair V_1, V_2 . Here $C_1 = \frac{1}{2}(\mathbf{1}_{\mathbb{C}^2} + \sqrt{2}\delta Z)$, so $u = \sqrt{2}\mathbf{1}_{\mathbb{C}^2} + O(\delta)$, $N = O(\delta)$, $\tilde{E} = \frac{\delta}{2}U_0 + O(\delta^2)$, $\tilde{F} = \frac{\delta}{2}U_1 + O(\delta^2)$ and $\bar{T} = \text{diag}(0, 1) + O(\delta)$. A direct computation gives

$$U_0\bar{T}'_0 + \bar{T}'_0U_0 - U_1\bar{T}_0 - \bar{T}_0U_1 = \sqrt{2}\mathbf{j}\mathbf{j}^\top, \quad \bar{T}_0 = \text{diag}(0, 1), \quad \bar{T}'_0 = \text{diag}(1, 0),$$

so $Y = \frac{\delta}{\sqrt{2}}\mathbf{j}\mathbf{j}^\top + O(\delta^2)$, and (4.10.2) gives $\tau = 2\sqrt{2}\delta + O(\delta^2)$. Since $2\sqrt{2} > 2$, we get $\tau > 2\delta$ for small δ , so the sharp bound 2δ of the commuting case (see the remark following Corollary 4.10.8) does not hold in general.

Corollary 4.10.7. For all $\Phi_1, \Psi_1, \Phi_2, \Psi_2 \in \text{CP}(\mathcal{X}, \mathcal{Y})$ with $\Omega' \neq 0$,

$$|\hat{F}(\Phi_1, \Psi_1) - \hat{F}(\Phi_2, \Psi_2)| \leq 2.528 \|\bar{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \delta^{1/2},$$

where δ is the relative distance of Lemma 4.8.1. For channels the constant is 5.06.

Proof. Put $\delta_0 = 0.1566$ and $s_x = (1 - 4x^2)^{1/2}$, so that $s_{\delta_0} > 0.94968$. Suppose first that $\delta \leq \delta_0$. By Lemma 4.10.5,

$$\tau \leq 2\sqrt{2}\delta/s_\delta \leq 2\sqrt{2}\delta/s_{\delta_0} < 2.9783\delta,$$

so $\tau^{1/2} < 1.7258\delta^{1/2}$. The proof of Theorem 4.10.4(b), with the first inequality of Lemma 4.10.3 in place of (4.10.1), gives

$$|\hat{F}(\Phi_1, \Psi_1) - \hat{F}(\Phi_2, \Psi_2)| \leq \|\bar{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \left(\tau^{1/2} + \frac{2\sqrt{2}\gamma}{(1 + s_\gamma)^{1/2}} \right).$$

The function $x \mapsto x(1 + s_x)^{-1/2}$ is increasing on $[0, \frac{1}{2})$, and $\gamma \leq \delta \leq \delta_0$ by Lemma 4.10.5, so the second term in the bracket is at most

$$2\sqrt{2}\delta_0^{1/2}(1 + s_{\delta_0})^{-1/2}\delta^{1/2} < 0.8017\delta^{1/2}.$$

This gives the constant $1.7258 + 0.8017 < 2.528$. If $\delta > \delta_0$, we use that $0 \leq \hat{F}(\Phi_i, \Psi_i) \leq \|\bar{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty$: indeed,

$$J(\Phi_i) \# J(\Psi_i) \leq \frac{1}{2}(J(\Phi_i) + J(\Psi_i)) = \frac{1}{2}J(\Omega_i)$$

by the arithmetic–geometric mean inequality [79], so Proposition 4.3.1 gives

$$\widehat{F}(\Phi_i, \Psi_i) \leq \frac{1}{2} \lambda_{\min}(\Omega_i^*(\mathbf{1}_{\mathcal{Y}})^\top) \leq \frac{1}{2} \|\Omega_i^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \leq \|\overline{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty,$$

using $\Omega_i \leq \Omega' = 2\overline{\Omega}$. Hence the left-hand side is at most $\|\overline{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty < \delta_0^{-1/2} \|\overline{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \delta^{1/2}$, and $\delta_0^{-1/2} < 2.5270$. For channels, $\overline{\Omega}^*(\mathbf{1}_{\mathcal{Y}}) = 2 \cdot \mathbf{1}_{\mathcal{X}}$, and $2 \cdot 2.528 < 5.06$. The threshold δ_0 is essentially optimal for this argument, since the two constants $(2\sqrt{2}/s_\delta)^{1/2} + 2\sqrt{2}\delta^{1/2}(1 + s_\delta)^{-1/2}$ and $\delta^{-1/2}$ coincide at $\delta = 0.15657$, where both equal 2.5273. (With the weaker bound $2\sqrt{2}\gamma\|\varphi\|_\infty$ for the second term of Lemma 4.10.3, the same optimisation gives 2.7437.) $\square$

Corollary 4.10.8. *For all $\Phi_1, \Psi_1, \Phi_2, \Psi_2 \in \text{CP}(\mathcal{X}, \mathcal{Y})$ with $\Omega' \neq 0$,*

$$\delta \leq \frac{\max\{\|\Phi_1 - \Phi_2\|_1, \|\Psi_1 - \Psi_2\|_1\}}{\nu(\Omega')^2},$$

and if $\text{range } J(\Phi_2 + \Psi_2) \subseteq \text{range } J(\Phi_1 + \Psi_1)$, then $\nu(\Omega') \geq \nu(\Phi_1 + \Psi_1)$. Consequently, in the latter case,

$$|\widehat{F}(\Phi_1, \Psi_1) - \widehat{F}(\Phi_2, \Psi_2)| \leq \frac{2.528 \|\overline{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty}{\nu(\Phi_1 + \Psi_1)} \max\{\|\Phi_1 - \Phi_2\|_1, \|\Psi_1 - \Psi_2\|_1\}^{1/2}.$$

Proof. With respect to the Kraus decomposition of Ω' we have $\Phi_1 - \Phi_2 = \Psi_{A_1 - A_2}$ and $\Psi_1 - \Psi_2 = \Psi_{B_1 - B_2}$, so the first inequality is Theorem 4.8.4 applied to Ω' . If $\text{range } J(\Phi_2 + \Psi_2) \subseteq \text{range } J(\Phi_1 + \Psi_1)$, then, since the range of a sum of positive semidefinite operators is the sum of their ranges, $\text{range } J(\Omega') = \text{range } J(\Phi_1 + \Psi_1)$, and $\Omega' \geq \Phi_1 + \Psi_1$ in the completely positive order, so Lemma 4.8.2(3) gives $\nu(\Omega') \geq \nu(\Phi_1 + \Psi_1)$. The last statement follows from Corollary 4.10.7. $\square$

Remark 4.10.9. Corollary 4.10.8 shows that the geometric channel fidelity is locally Hölder continuous of order $\frac{1}{2}$ in the completely bounded trace norm at every pair (Φ_1, Ψ_1) , along perturbations that do not enlarge the Choi support of the sum, with a constant that depends only on $\nu(\Phi_1 + \Psi_1)$ and on $\|\overline{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_\infty \leq \frac{1}{2}(\|\Omega_1^*(\mathbf{1}_{\mathcal{Y}})\|_\infty + \|\Omega_2^*(\mathbf{1}_{\mathcal{Y}})\|_\infty)$, and not on any dimension. The same argument applies to the other quantities in Theorem 4.10.4, with the corresponding moduli. The support condition is needed here only to control τ by δ : if Φ_2 adds a small multiple $t\Xi$ of a map whose Choi matrix has range outside $\text{range } J(\Phi_1 + \Psi_1)$, then $\nu(\Omega')^2$ may be of order t , and Corollary 4.3.3 shows that some dependence on ν is unavoidable. Section 4.11 extends Theorem 4.10.4 itself to arbitrary supports. In the commuting case $\|T_1 - T_2\|_\infty \leq 2\delta$, and this is sharp. Indeed, on a common eigenvector let t_i, x_i, y_i and $c_i = x_i + y_i$ be the eigenvalues of T_i, A_i, B_i and C_i , so that $x_i = c_i t_i$ and $y_i = c_i(1 - t_i)$, write $c_1 = \frac{1}{2} + \gamma'$ and $c_2 = \frac{1}{2} - \gamma'$, and assume that $d = t_1 - t_2 \geq 0$. If $\gamma' \geq 0$, then $c_1 d = (x_1 - x_2) - 2\gamma' t_2 \leq \delta$; if $\gamma' < 0$, then $c_2 d = 2\gamma'(1 - t_1) - (y_1 - y_2) \leq \delta$. In both cases the coefficient of d is at least $\frac{1}{2}$, so $d \leq 2\delta$. Equality holds for equal sums with $t_2 = 0$ and $t_1 = 2\delta$. Example 4.10.6 shows that noncommutativity costs a factor $\sqrt{2}$ in the leading constant of Lemma 4.10.5, and no more.

4.11 Sums with different Choi supports

We now drop the hypothesis $\text{range } J(\Omega_1) = \text{range } J(\Omega_2)$ of Section 4.10. Its only role there was to make C_1 and C_2 invertible. Lemma 4.10.3 then measures the change of the sums by $\gamma = \|C_1 - \frac{1}{2}\mathbf{1}_{\mathbb{C}^r}\|_\infty$, a supremum over the spectrum of C_1 . A single direction that is

present in one sum and absent from the other gives $\gamma = \frac{1}{2}$, however small its weight, and the estimate becomes trivial. We replace this supremum by weighted estimates, organised by the Lebesgue decomposition of Section 4.7.

We keep the notation of Section 4.10: $\Omega', V_1, \dots, V_r, W, A_i, B_i, C_i, a_1 = C_1^{1/2}$ and $a_2 = C_2^{1/2}$. Put

$$\Delta(Z) = \text{Tr}_{\mathcal{Y}}(WZW^*), \quad Z \in L(\mathbb{C}^r),$$

a completely positive map. For $K \geq 0$, $\Delta(K) = \text{Tr}_{\mathcal{Y}} J(\Psi_K)$ is the transpose of $\Psi_K^*(\mathbf{1}_{\mathcal{Y}})$, so

$$\|\Delta(\mathbf{1}_{\mathbb{C}^r})\|_{\infty} = \|\Omega'^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty} = 2\|\bar{\Omega}^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty}$$

and $\|\Delta(C_i)\|_{\infty} = \|\Omega_i^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty}$. We use two elementary facts. If Z is Hermitian and $-K \leq Z \leq K$, then

$$\|\Delta(Z)\|_{\infty} \leq \|\Delta(K)\|_{\infty}, \quad \|\Psi_Z\|_1 \leq \|\Psi_K^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty}. \quad (4.11.1)$$

The first holds because Δ is positive. For the second, $\Psi_K \pm \Psi_Z = \Psi_{K \pm Z}$ is completely positive, so

$$-(\Psi_K \otimes \mathbf{1}_{L(\mathcal{X})})(\rho) \leq (\Psi_Z \otimes \mathbf{1}_{L(\mathcal{X})})(\rho) \leq (\Psi_K \otimes \mathbf{1}_{L(\mathcal{X})})(\rho)$$

for every state ρ , and therefore

$$\|(\Psi_Z \otimes \mathbf{1}_{L(\mathcal{X})})(\rho)\|_1 \leq \text{tr}(\Psi_K \otimes \mathbf{1}_{L(\mathcal{X})})(\rho) \leq \|\Psi_K^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty}.$$

4.11.1 The Lebesgue decomposition of the sums

Let Q_1, Q_R and Q_2 be the spectral projections of C_1 for $\{1\}$, $(0, 1)$ and $\{0\}$. They commute with a_1 and a_2 , and $P_1 = Q_1 + Q_R$ and $P_2 = Q_R + Q_2$ are the range projections of C_1 and $C_2 = \mathbf{1}_{\mathbb{C}^r} - C_1$. By Proposition 4.7.1, applied to the pair (Ω_1, Ω_2) , whose derivative with respect to Ω' is C_1 , the singular part of Ω_1 relative to Ω_2 is $(\Omega_1)_s = \Psi_{Q_1}$, and that of Ω_2 relative to Ω_1 is $(\Omega_2)_s = \Psi_{Q_2}$. The absolutely continuous parts $\Omega_i^{\text{ac}} = \Omega_i - (\Omega_i)_s = \Psi_{Q_R C_i}$ have the same Choi range, and the supports of the sums agree if and only if $Q_1 = Q_2 = 0$.

Lemma 4.11.1. *Let $T_1 = a_1^+ A_1 a_1^+$ and $T_2 = a_2^+ A_2 a_2^+$, where $^+$ denotes the Moore–Penrose inverse, regarded as operators on $\text{range } P_1$ and $\text{range } P_2$.*

- (a) $0 \leq T_i \leq P_i$, $A_i = a_i T_i a_i$ and $B_i = a_i(P_i - T_i)a_i$.
- (b) If U_i is an isometry with $\text{range } U_i = \text{range } P_i$, then $U_i^* T_i U_i$ is the derivative of Φ_i with respect to Ω_i , relative to the Kraus decomposition of Ω_i whose column matrix is $W a_i U_i$.
- (c) For every operator connection σ' , $A_i \sigma' B_i = a_i \varphi_{\sigma'}(T_i) a_i$, where $\varphi_{\sigma'}(T_i)$ is formed by the functional calculus in $L(\text{range } P_i)$ and extended by 0.
- (d) Let $N_i = a_i \varphi(T_i) a_i$ in this sense. Then $J(\Phi_i) \sigma J(\Psi_i) = W N_i W^*$ for a mean σ and $\varphi = g_{\sigma}$, so that $\widehat{F}(\Phi_i, \Psi_i) = \lambda_{\min}(\Delta(N_i))$ for $\varphi = g_{\#}$ and $\widehat{Q}_{\alpha}(\Phi_i \| \Psi_i) = \lambda_{\min}(\Delta(N_i))$ for $\varphi = h_{\alpha}$. For every operator convex f with $h_f(T_i)$ finite, $\widehat{D}_f(\Phi_i \| \Psi_i) = \lambda_{\max}(\Delta(N_i))$ for $\varphi = h_f$.

Proof. (a) Since $0 \leq A_i \leq C_i = a_i^2$, we have $\text{range } A_i \subseteq \text{range } C_i = \text{range } P_i$, so $A_i = P_i A_i P_i = a_i a_i^+ A_i a_i^+ a_i = a_i T_i a_i$. Moreover $T_i \leq a_i^+ C_i a_i^+ = P_i$, and $B_i = C_i - A_i = a_i(P_i - T_i)a_i$.

(b)

$$J(\Omega_i) = W C_i W^* = (W a_i U_i)(W a_i U_i)^*,$$

and Wa_iU_i is injective because W is injective and a_i is injective on $\text{range } P_i$. By (a),

$$J(\Phi_i) = WA_iW^* = (Wa_iU_i)U_i^*T_iU_i(Wa_iU_i)^*,$$

and Theorem 4.2.2 gives the claim.

(c) The operators A_i , B_i and a_i vanish on $\ker P_i$ and leave $\text{range } P_i$ invariant. A connection respects such direct sums, by its defining formula for invertible arguments and continuity from above, and $0\sigma'0 = 0$. On $\text{range } P_i$ the operator a_i is invertible, so the transformer equality and Remark 4.2.16(2), applied in $L(\text{range } P_i)$, give

$$A_i\sigma'B_i = a_i(T_i\sigma'(P_i - T_i))a_i = a_i\varphi_{\sigma'}(T_i)a_i.$$

(d) The first statements follow from (c) and (4.2.1), exactly as in the proof of Theorem 4.10.4, and from Proposition 4.3.1 and Lemma 4.5.3. For the last, apply Theorem 4.7.5(b) to Ω_i with the Kraus decomposition of (b). $\square$

We now define the quantities that replace τ and γ . Put

$$\begin{aligned}\tau_{ac} &= \|Q_R(T_1 - T_2)Q_R\|_\infty, & \kappa_i &= \|Q_iT_iQ_R\|_\infty \ (i = 1, 2), \\ \gamma_{ac} &= \|Q_R(C_1 - \tfrac{1}{2}\mathbf{1}_{C^r})Q_R\|_\infty.\end{aligned}$$

These are the in-pair distance on the common part of the supports, the coupling inside each pair between the singular and the common directions, and the half-distance of C_1 from $\frac{1}{2}$ on the common part. Since Q_R is the spectral projection for $(0, 1)$ and $r < \infty$, we have $\gamma_{ac} < \frac{1}{2}$. If $Q_R \neq 0$, the proof of Lemma 4.10.1, applied to Ω_1^{ac} and Ω_2^{ac} , whose derivatives with respect to their sum Ψ_{Q_R} are the restrictions of C_1 and C_2 to $\text{range } Q_R$, gives

$$\tanh(\tfrac{1}{2}d_{ac}) = 2\gamma_{ac}, \quad d_{ac} = d_T(\Omega_1^{\text{ac}}, \Omega_2^{\text{ac}}) < \infty.$$

If $Q_R = 0$, that is, if the Choi ranges of the two sums intersect trivially (for instance if one sum is 0), then $\Omega_1^{\text{ac}} = \Omega_2^{\text{ac}} = 0$ and $\gamma_{ac} = 0$, and we set $d_{ac} = 0$, so that $\tanh(\frac{1}{2}d_{ac}) = 2\gamma_{ac}$ in all cases. Let $\Omega_1 \# \Omega_2$ be the completely positive map with Choi matrix $J(\Omega_1) \# J(\Omega_2)$. By (4.2.1) this Choi matrix is $W(C_1 \# C_2)W^* = Wa_1a_2W^*$, since C_1 and C_2 commute, and $a_1a_2 = Q_Ra_1a_2Q_R$. Hence $\Omega_1^{\text{ac}} + \Omega_2^{\text{ac}} - 2\Omega_1 \# \Omega_2 = \Psi_{Q_R(a_1 - a_2)^2}$ is completely positive, and we put

$$h_{ac} = \|(\Omega_1^{\text{ac}} + \Omega_2^{\text{ac}} - 2\Omega_1 \# \Omega_2)^*(\mathbf{1}_Y)\|_\infty^{1/2}, \quad m_s = \|((\Omega_1)_s + (\Omega_2)_s)^*(\mathbf{1}_Y)\|_\infty.$$

Thus h_{ac} is a Hellinger-type distance between the absolutely continuous parts of the sums, and m_s is the size of their singular parts.

Lemma 4.11.2. (a) If $\text{range } J(\Omega_1) = \text{range } J(\Omega_2)$, then $m_s = \kappa_1 = \kappa_2 = 0$, $\tau_{ac} = \tau$, $\gamma_{ac} = \gamma$ and $d_{ac} = d_T$.

(b) $\kappa_1, \kappa_2 \leq \frac{1}{2}$.

(c)

$$h_{ac}^2 \leq \min \{ \|\Delta(Q_R|C_1 - C_2|)\|_\infty, 4\gamma_{ac}^2 \|\Delta(Q_R)\|_\infty \},$$

and $4\gamma_{ac}^2 \|\Delta(Q_R)\|_\infty \leq 8\gamma_{ac}^2 \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty$. Here $\Delta(Q_R|C_1 - C_2|)$ is the transpose of $\Psi_{Q_R|C_1 - C_2|}^*(\mathbf{1}_Y)$, the total variation of the absolutely continuous parts measured in the frame of Ω' .

Proof. (a) is immediate from the definitions. (b) Let $u \in \text{range } Q_1$ and $v \in \text{range } Q_R$ be unit vectors with $|\langle u, T_1 v \rangle| = \kappa_1$. The compression of T_1 to $\text{span}\{u, v\}$ is a 2×2 matrix $\begin{bmatrix} s & x \\ x & t \end{bmatrix}$ with eigenvalues in $[0, 1]$ and $|x| = \kappa_1$. Then $|x|^2 \leq st$ and $|x|^2 \leq (1-s)(1-t)$, so $|x|^4 \leq s(1-s)t(1-t) \leq \frac{1}{16}$. The same argument applies to κ_2 . (c) For commuting $a_1, a_2 \geq 0$, $|a_1 - a_2| \leq a_1 + a_2$, so

$$(a_1 - a_2)^2 \leq |a_1 - a_2|(a_1 + a_2) = |a_1^2 - a_2^2| = |C_1 - C_2|.$$

Also $|\sqrt{\lambda} - \sqrt{1-\lambda}| \leq |2\lambda - 1| \leq 2\gamma_{\text{ac}}$ for λ in the spectrum of C_1 on $\text{range } Q_R$, so $Q_R(a_1 - a_2)^2 \leq 4\gamma_{\text{ac}}^2 Q_R$. Apply Δ , and use

$$\|\Delta(Q_R)\|_\infty \leq \|\Delta(\mathbf{1}_{C^r})\|_\infty = 2\|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty.$$

□

4.11.2 The estimate

Theorem 4.11.3. *Let $\Phi_1, \Psi_1, \Phi_2, \Psi_2 \in \text{CP}(\mathcal{X}, \mathcal{Y})$ be arbitrary, with $\Omega' \neq 0$, and let T_1, T_2 be as in Lemma 4.11.1. Let $I \subseteq [0, 1]$ be a compact interval containing $\text{spec}(T_1) \cup \text{spec}(T_2)$, let $\varphi : I \rightarrow \mathbb{R}$ be continuous, $\|\varphi\|_I = \max_I |\varphi|$, and $N_i = a_i \varphi(T_i) a_i$. Then $\|\Delta(N_1) - \Delta(N_2)\|_\infty$ and $\|\Psi_{N_1} - \Psi_{N_2}\|_1$ are both at most*

$$\begin{aligned} R_\varphi &= \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \omega_\varphi^I(\tau_{\text{ac}}) \\ &\quad + \|\varphi\|_I \left(m_s + \min \left\{ 2\sqrt{2} \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \tanh \frac{d_{\text{ac}}}{2}, 2\|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty^{1/2} h_{\text{ac}} \right\} \right) \\ &\quad + \|\Omega_1^*(\mathbf{1}_Y)\|_\infty \omega_\varphi^I(\kappa_1) + \|\Omega_2^*(\mathbf{1}_Y)\|_\infty \omega_\varphi^I(\kappa_2). \end{aligned}$$

Consequently:

- (a) $\|\Phi_1 \sigma \Psi_1 - \Phi_2 \sigma \Psi_2\|_1 \leq R_{g_\sigma}$ for every Kubo–Ando mean σ , with $I = [0, 1]$;
- (b) $|\hat{F}(\Phi_1, \Psi_1) - \hat{F}(\Phi_2, \Psi_2)| \leq R_{g_\#}$, and with $I = [0, 1]$, $\omega_{g_\#}(t) \leq t^{1/2}$ and $\|g_\#\|_\infty = \frac{1}{2}$,
$$R_{g_\#} \leq \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \tau_{\text{ac}}^{1/2} + \frac{1}{2} m_s + \sum_{i=1,2} \|\Omega_i^*(\mathbf{1}_Y)\|_\infty \kappa_i^{1/2} \\ + \min \left\{ \sqrt{2} \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \tanh \frac{d_{\text{ac}}}{2}, \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty^{1/2} h_{\text{ac}} \right\};$$
- (c) $|\hat{Q}_\alpha(\Phi_1 \| \Psi_1) - \hat{Q}_\alpha(\Phi_2 \| \Psi_2)| \leq R_{h_\alpha}$ for $\alpha \in (0, 1)$, with $I = [0, 1]$;
- (d) $|\hat{D}_f(\Phi_1 \| \Psi_1) - \hat{D}_f(\Phi_2 \| \Psi_2)| \leq R_{h_f}$ for every operator convex f such that h_f is finite on I .

If $\text{range } J(\Omega_1) = \text{range } J(\Omega_2)$, then, taking the first entry of the minimum, R_φ equals the bound of Theorem 4.10.4 for the corresponding quantity.

Proof. Let $T_1^{\text{d}} = Q_1 T_1 Q_1 + Q_R T_1 Q_R$ and $T_2^{\text{d}} = Q_R T_2 Q_R + Q_2 T_2 Q_2$ be the block-diagonal parts of T_i with respect to the spectral decomposition of C_1 . Compressions of an operator with spectrum in the interval I have spectra in I , so $\varphi(T_i^{\text{d}})$ is defined, and $\|T_i - T_i^{\text{d}}\|_\infty = \kappa_i$, since $T_1 - T_1^{\text{d}} = Q_1 T_1 Q_R + Q_R T_1 Q_1$ and similarly for T_2 . Put $N_i^{\text{d}} = a_i \varphi(T_i^{\text{d}}) a_i$. By (4.11.1), all three steps below give the same bound for $\|\Delta(\cdot)\|_\infty$ and for $\|\Psi(\cdot)\|_1$.

Coupling. Since $\|T_1 - T_1^{\text{d}}\|_\infty \leq \omega_\varphi^I(\kappa_1)$, we have

$$\pm(N_1 - N_1^{\text{d}}) \leq \omega_\varphi^I(\kappa_1) a_1^2 = \omega_\varphi^I(\kappa_1) C_1,$$

and

$$\|\Delta(C_1)\|_\infty = \|\Psi_{C_1}^*(\mathbf{1}_Y)\|_\infty = \|\Omega_1^*(\mathbf{1}_Y)\|_\infty.$$

The same holds for the second pair.

Singular parts. On range Q_1 we have $a_1 = \mathbf{1}_{\text{range } Q_1}$ and $a_2 = 0$, and on range Q_2 we have $a_1 = 0$ and $a_2 = \mathbf{1}_{\text{range } Q_2}$. Since T_i^d is block diagonal,

$$\begin{aligned} N_1^d - N_2^d &= Q_1 \varphi(Q_1 T_1 Q_1) Q_1 - Q_2 \varphi(Q_2 T_2 Q_2) Q_2 + Z_R, \\ Z_R &= a_1 \varphi(Q_R T_1 Q_R) a_1 - a_2 \varphi(Q_R T_2 Q_R) a_2, \end{aligned}$$

where in Z_R the functional calculus is taken in $L(\text{range } Q_R)$. The first two terms lie between $\mp \|\varphi\|_I (Q_1 + Q_2)$, and $\|\Delta(Q_1 + Q_2)\|_\infty = m_s$.

Common part. On range Q_R put $H_- = \varphi(Q_R T_1 Q_R) - \varphi(Q_R T_2 Q_R)$ and $H_+ = \varphi(Q_R T_1 Q_R) + \varphi(Q_R T_2 Q_R)$, so that $\|H_-\|_\infty \leq \omega_\varphi^I(\tau_{ac})$ and $\|H_+\|_\infty \leq 2\|\varphi\|_I$. As in the proof of Lemma 4.10.3,

$$Z_R = \frac{1}{2}(a_1 H_- a_1 + a_2 H_- a_2) + Y, \quad Y = \frac{1}{4}((a_1 - a_2)H_+(a_1 + a_2) + (a_1 + a_2)H_+(a_1 - a_2)),$$

with all operators restricted to range Q_R . Since $a_1^2 + a_2^2 = Q_R$ there,

$$\pm \frac{1}{2}(a_1 H_- a_1 + a_2 H_- a_2) \leq \frac{1}{2}\|H_-\|_\infty Q_R,$$

and $\|\Delta(Q_R)\|_\infty \leq 2\|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty$. For Y we give two bounds. First, as in Lemma 4.10.3, on range Q_R we have

$$\|Y\|_\infty \leq \frac{1}{2}\|a_1 - a_2\|_\infty \|H_+\|_\infty \|a_1 + a_2\|_\infty \leq \frac{1}{2} \cdot 2\gamma_{ac} \cdot 2\|\varphi\|_I \cdot \sqrt{2},$$

so $\pm Y \leq 2\sqrt{2}\gamma_{ac}\|\varphi\|_I Q_R$. Second, let X, X', H_+ be Hermitian and $\eta > 0$. Expanding

$$(\eta^{1/2}X \mp \eta^{-1/2}X'H_+)(\eta^{1/2}X \mp \eta^{-1/2}H_+X') \geq 0$$

gives

$$\pm(XH_+X' + X'H_+X) \leq \eta X^2 + \eta^{-1}X'H_+^2X' \leq \eta X^2 + \eta^{-1}\|H_+\|_\infty^2 X'^2.$$

With $X = a_1 - a_2$ and $X' = a_1 + a_2$, and since $(a_1 + a_2)^2 = Q_R + 2a_1a_2 \leq 2Q_R$ on range Q_R , this gives

$$\pm 4Y \leq \eta Q_R(a_1 - a_2)^2 + 2\eta^{-1}\|H_+\|_\infty^2 Q_R.$$

By definition $\|\Delta(Q_R(a_1 - a_2)^2)\|_\infty = h_{ac}^2$, so optimising over η , the bound is

$$\frac{1}{4} \cdot 2\|H_+\|_\infty h_{ac} (2\|\Delta(Q_R)\|_\infty)^{1/2} \leq 2\|\varphi\|_I \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty^{1/2} h_{ac}.$$

Since $2\gamma_{ac} = \tanh \frac{d_{ac}}{2}$, the first bound contributes at most $2\sqrt{2}\|\varphi\|_I \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \tanh \frac{d_{ac}}{2}$.

Adding the three steps gives R_φ . Parts (a)–(d) follow from Lemma 4.11.1(d) and

$$|\lambda_{\min}(M) - \lambda_{\min}(M')|, |\lambda_{\max}(M) - \lambda_{\max}(M')| \leq \|M - M'\|_\infty;$$

for (a), $\Phi_i \sigma \Psi_i = \Psi_{N_i}$ with respect to $V_1, \dots, V_r$. If the supports agree, Lemma 4.11.2(a) shows that R_φ with the first entry of the minimum is

$$\|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \left(\omega_\varphi(\tau) + 2\sqrt{2}\|\varphi\|_I \tanh \frac{d_T}{2} \right),$$

which is the bound of Theorem 4.10.4. □

Theorem 4.11.3 separates three effects. The singular parts of the sums enter linearly, through their size m_s . The change of the common part enters through the in-pair distance τ_{ac} and through either the Thompson or the Hellinger distance of the absolutely continuous parts, whichever is smaller. The Hellinger form is better when some common directions carry very different relative weights in the two sums but little absolute weight. The couplings κ_i measure how strongly each pair mixes the directions it does not share with the other sum into the common ones. The next corollary is the cleanest case, in which only the singular part survives.

Corollary 4.11.4 (Adding an independent branch). *Let $\Phi_1, \Psi_1, \Theta_\Phi, \Theta_\Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, put $\Omega_1 = \Phi_1 + \Psi_1$ and $\Theta = \Theta_\Phi + \Theta_\Psi$, and assume $\text{range } J(\Omega_1) \cap \text{range } J(\Theta) = \{0\}$. Let $\Phi_2 = \Phi_1 + \Theta_\Phi$ and $\Psi_2 = \Psi_1 + \Theta_\Psi$. Then for every operator connection σ' ,*

$$J(\Phi_2) \sigma' J(\Psi_2) = J(\Phi_1) \sigma' J(\Psi_1) + J(\Theta_\Phi) \sigma' J(\Theta_\Psi),$$

and consequently

$$\widehat{F}(\Phi_1, \Psi_1) + \widehat{F}(\Theta_\Phi, \Theta_\Psi) \leq \widehat{F}(\Phi_2, \Psi_2) \leq \widehat{F}(\Phi_1, \Psi_1) + \frac{1}{2} \|\Theta^*(\mathbf{1}_\mathcal{Y})\|_\infty.$$

For every mean σ ,

$$\|\Phi_2 \sigma \Psi_2 - \Phi_1 \sigma \Psi_1\|_1 = \|\Theta_\Phi \sigma \Theta_\Psi\|_1 \leq \|\Theta^*(\mathbf{1}_\mathcal{Y})\|_\infty.$$

In the notation of Theorem 4.11.3, $(\Omega_1)_s = 0$, $(\Omega_2)_s = \Theta$, and $\tau_{ac} = \kappa_1 = \kappa_2 = d_{ac} = 0$, where $d_{ac} = 0$ holds by the convention above if $\Omega_1 = 0$.

Proof. Let $\mathcal{V}_1 = \text{range } J(\Omega_1)$ and $\mathcal{V}_2 = \text{range } J(\Theta)$. Since $\mathcal{V}_1 \cap \mathcal{V}_2 = \{0\}$, there is an invertible Z on $\mathcal{Y} \otimes \mathcal{X}$ with $Z\mathcal{V}_1 \perp Z\mathcal{V}_2$. The Choi matrices of Φ_1, Ψ_1 have ranges in $\mathcal{V}_1$, and those of Θ_Φ, Θ_Ψ in $\mathcal{V}_2$, because $J(\Phi_1), J(\Psi_1) \leq J(\Omega_1)$ and $J(\Theta_\Phi), J(\Theta_\Psi) \leq J(\Theta)$. After conjugation by Z the two pairs live on orthogonal subspaces, where the connection acts blockwise. The transformer equality for the invertible Z gives the identity. Taking partial traces, $\text{Tr}_\mathcal{Y}(J(\Phi_2) \# J(\Psi_2)) = M_1 + M_\Theta$ with $M_1 = \text{Tr}_\mathcal{Y}(J(\Phi_1) \# J(\Psi_1))$ and $M_\Theta = \text{Tr}_\mathcal{Y}(J(\Theta_\Phi) \# J(\Theta_\Psi))$. Then

$$\lambda_{\min}(M_1) + \lambda_{\min}(M_\Theta) \leq \lambda_{\min}(M_1 + M_\Theta) \leq \lambda_{\min}(M_1) + \|M_\Theta\|_\infty,$$

and $J(\Theta_\Phi) \# J(\Theta_\Psi) \leq \frac{1}{2} J(\Theta)$ by the arithmetic–geometric mean inequality, so $\|M_\Theta\|_\infty \leq \frac{1}{2} \|\Theta^*(\mathbf{1}_\mathcal{Y})\|_\infty$; Proposition 4.3.1 gives the inequalities. For a mean, $0 \leq J(\Theta_\Phi) \sigma J(\Theta_\Psi) \leq J(\Theta)$ in the completely positive sense, which gives the trace-norm bound. For the last statement, choose the column matrix $W = [W_1 \ W_\Theta]$ of $\Omega' = 2\Omega_1 + \Theta$ with $W_1 W_1^* = 2J(\Omega_1)$ and $W_\Theta W_\Theta^* = J(\Theta)$; its columns are linearly independent because $\mathcal{V}_1 \cap \mathcal{V}_2 = \{0\}$. Then, splitting $\mathbb{C}^r = \mathbb{C}^{r_1} \oplus \mathbb{C}^{r_\Theta}$ according to the columns of W_1 and W_Θ , $C_1 = \frac{1}{2} \mathbf{1}_{\mathbb{C}^{r_1}} \oplus 0$, $C_2 = \frac{1}{2} \mathbf{1}_{\mathbb{C}^{r_1}} \oplus \mathbf{1}_{\mathbb{C}^{r_\Theta}}$, $A_2 = A_1 \oplus T_\Theta$ for the derivative T_Θ of Θ_Φ , and $T_1 = 2A_1$, $T_2 = 2A_1 \oplus T_\Theta$. $\square$

Remark 4.11.5 (Physical interpretation). Corollary 4.11.4 covers heralded erasure. Regard channels Φ_1, Ψ_1 as maps into $L(\mathcal{Y} \oplus \mathbb{C}v_0)$, and let them output a flag $v_0 v_0^*$ with probability p : $\Phi_2 = (1-p)\Phi_1 + p \text{tr}(\cdot) v_0 v_0^*$ and likewise Ψ_2 . The corollary applies to $((1-p)\Phi_1, (1-p)\Psi_1)$ and $\Theta_\Phi = \Theta_\Psi = p \text{tr}(\cdot) v_0 v_0^*$, since $\text{range } J(\Theta) = \mathbb{C}v_0 \otimes \mathcal{X}$ is orthogonal to $\mathcal{Y} \otimes \mathcal{X} \supseteq \text{range } J(\Phi_1 + \Psi_1)$. As

$$\widehat{F}(\Theta_\Phi, \Theta_\Psi) = p = \frac{1}{2} \|\Theta^*(\mathbf{1}_{\mathcal{Y} \oplus \mathbb{C}v_0})\|_\infty$$

and $\widehat{F}$ is positively homogeneous, its two bounds coincide and give $\widehat{F}(\Phi_2, \Psi_2) = (1-p)\widehat{F}(\Phi_1, \Psi_1) + p$: an incoherent new branch enters exactly with its weight. Leakage coherently coupled to existing directions [130] enters instead through the couplings κ_i , and can cost a constant at vanishing weight (Example 4.11.7).

Without the independence assumption, the quantities of Theorem 4.11.3 cannot be bounded by the completely bounded distance between the pairs alone. Take $\dim \mathcal{X} = 1$, so that completely positive maps are positive semidefinite matrices, and linearly independent unit vectors y_1, y_2 with $|\langle y_1, y_2 \rangle|$ close to 1. Let $\Phi_1 = \Psi_1 = y_1 y_1^*$ and $\Phi_2 = y_1 y_1^*, \Psi_2 = y_2 y_2^*$. Then $\hat{F}(\Phi_1, \Psi_1) = 1$ and $\hat{F}(\Phi_2, \Psi_2) = 0$, because the geometric mean of two distinct rank-one projections is 0, although $\|\Psi_1 - \Psi_2\|_1$ is small. In the frame $W = [\sqrt{3} y_1 \ y_2]$ of $\Omega' = 3y_1 y_1^* + y_2 y_2^*$ one finds $C_1 = \text{diag}(\frac{2}{3}, 0)$, $T_1 = \frac{1}{2}$ and $T_2 = \text{diag}(1, 0)$. Hence $(\Omega_2)_s = y_2 y_2^*$ has size $m_s = 1$, however close y_2 is to y_1 , and $\tau_{ac} = \frac{1}{2}$. The jump is therefore recorded by the theorem, through τ_{ac} and m_s , and not hidden. It is the support-changing counterpart of the impossibility example of Corollary 4.3.3.

4.11.3 Continuity in the derivative across support changes

Theorem 4.11.3 is a continuity statement in which the relative distance δ is replaced by the quantities of the Lebesgue decomposition. When the supports differ, $\delta \geq \frac{1}{2}$ by Lemma 4.10.5, so no statement in terms of δ alone can cover this case.

Corollary 4.11.6 (Continuity across support changes). *Fix (Φ_1, Ψ_1) , let $(\Phi_2^{(n)}, \Psi_2^{(n)})$ be pairs with $\sup_n \|\Omega_2^{(n)*}(\mathbf{1}_{\mathcal{Y}})\|_\infty < \infty$, and let $\tau_{ac}^{(n)}, \kappa_i^{(n)}, m_s^{(n)}, d_{ac}^{(n)}, h_{ac}^{(n)}$ be the quantities of Theorem 4.11.3 for the pairs (Φ_1, Ψ_1) and $(\Phi_2^{(n)}, \Psi_2^{(n)})$. If*

$$\tau_{ac}^{(n)} \rightarrow 0, \quad \kappa_1^{(n)}, \kappa_2^{(n)} \rightarrow 0, \quad m_s^{(n)} \rightarrow 0, \quad \min\{d_{ac}^{(n)}, h_{ac}^{(n)}\} \rightarrow 0,$$

then $\hat{F}(\Phi_2^{(n)}, \Psi_2^{(n)}) \rightarrow \hat{F}(\Phi_1, \Psi_1)$, $\hat{Q}_\alpha(\Phi_2^{(n)} \|\Psi_2^{(n)}) \rightarrow \hat{Q}_\alpha(\Phi_1 \|\Psi_1)$ for $\alpha \in (0, 1)$, and $\|(\Phi_2^{(n)} \sigma \Psi_2^{(n)} - \Phi_1 \sigma \Psi_1)\|_1 \rightarrow 0$ for every mean σ . The same holds for $\hat{D}_f$ for every operator convex f such that h_f is finite on a compact interval containing the spectra of all the in-pair derivatives. If the supports of the sums agree for all n , the hypothesis reduces to $\tau^{(n)} \rightarrow 0$ and $\min\{d_T^{(n)}, h_{ac}^{(n)}\} \rightarrow 0$, which holds in particular if $\tau^{(n)} \rightarrow 0$ and $d_T^{(n)} \rightarrow 0$.

Proof. Every term of R_φ is a constant times ω_φ^I at one of $\tau_{ac}^{(n)}, \kappa_1^{(n)}, \kappa_2^{(n)}$, or a constant times $m_s^{(n)}, \tanh \frac{1}{2} d_{ac}^{(n)}$ or $h_{ac}^{(n)}$, and the constants are bounded in n . Lemma 4.4.8 gives $\omega_\varphi^I(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. The last statement is Lemma 4.11.2(a). $\square$

None of the hypotheses can be dropped. For τ_{ac} , take the pairs of Corollary 4.3.3: they have the same sum, so all other quantities vanish, while $\tau_{ac} = \frac{1}{2}$ and $\hat{F}$ jumps from 0 to 1. For m_s , take Corollary 4.11.4 with $\Theta_\Phi = \Theta_\Psi$ and $\hat{F}(\Theta_\Phi, \Theta_\Psi) > 0$, for instance the flag of Remark 4.11.5: then $\tau_{ac} = \kappa_1 = \kappa_2 = d_{ac} = h_{ac} = 0$, while $\hat{F}$ increases by at least $\hat{F}(\Theta_\Phi, \Theta_\Psi)$. For $\min\{d_{ac}, h_{ac}\}$, let $\dim \mathcal{X} = 1$, $\mathcal{Y} = \mathbb{C}^2$, $J(\Phi_1) = J(\Psi_1) = \frac{1}{2} \mathbf{1}_{\mathbb{C}^2}$ and $J(\Phi_2) = J(\Psi_2) = \frac{1}{2} \text{diag}(1, s)$ with $0 < s < 1$. The supports agree and $\tau_{ac} = \kappa_1 = \kappa_2 = m_s = 0$, while $d_{ac} = \log \frac{1}{s}$, $h_{ac} = 1 - \sqrt{s}$ and $\hat{F}(\Phi_1, \Psi_1) - \hat{F}(\Phi_2, \Psi_2) = \frac{1-s}{2}$. In the example following Corollary 4.11.4, $m_s = 1$ and $\tau_{ac} = \frac{1}{2}$ while the pairs are arbitrarily close. The next example shows that the coupling alone can carry a jump, with every other quantity tending to zero.

Example 4.11.7 (Coupling). Let $\dim \mathcal{X} = 1$ and $\mathcal{Y} = \mathbb{C}^2$, let P be the projection onto $(e_1 + e_2)/\sqrt{2}$, and put $Z_t = \text{diag}(\sqrt{t}, 1)$. Let $J(\Phi_t) = Z_t P Z_t$, $J(\Psi_t) = Z_t (\mathbf{1}_{\mathbb{C}^2} - P) Z_t$, and $J(\Phi_0) = J(\Psi_0) = \frac{1}{2} e_2 e_2^*$. Then $\|J(\Phi_t) - J(\Phi_0)\|_1$ and $\|J(\Psi_t) - J(\Psi_0)\|_1$ are $O(\sqrt{t})$. Now $J(\Phi_t)$ and $J(\Psi_t)$ are distinct rank-one operators, so their geometric mean is 0 and $\hat{F}(\Phi_t, \Psi_t) = 0$, whereas $\hat{F}(\Phi_0, \Psi_0) = \frac{1}{2}$. Take (Φ_t, Ψ_t) as the first pair and (Φ_0, Ψ_0) as the second. Then $J(\Omega') = \text{diag}(t, 2)$, $W = \text{diag}(\sqrt{t}, \sqrt{2})$ and $C_1 = \text{diag}(1, \frac{1}{2})$, so $Q_1 = e_1 e_1^*$,

$Q_R = e_2 e_2^*$ and $Q_2 = 0$. Moreover $T_1 = P$, and $T_2 = \frac{1}{2}$ on $\text{range } Q_R$. Hence $\tau_{ac} = 0$, $d_{ac} = h_{ac} = 0$, $\kappa_2 = 0$ and $m_s = t \rightarrow 0$, but $\kappa_1 = |\langle e_1, P e_2 \rangle| = \frac{1}{2}$ for every t . The jump of size $\frac{1}{2}$ is carried by the coupling, and the term $\|\Omega_1^*(\mathbf{1}_{\mathcal{Y}})\|_{\infty} \kappa_1^{1/2} = (1+t)/\sqrt{2}$ of Theorem 4.11.3(b) accounts for it.

4.12 Where continuity fails: a dichotomy

Example 4.11.7 suggests a general mechanism. A pair whose sum has a non-invertible Choi matrix can be approximated by pairs that use the missing Choi directions with vanishing weight but with full coupling, and this changes the derivative in the limit. We show that this is the only obstruction to continuity, and we compute the size of the jumps. The answer is a dichotomy: for the fidelity, $\hat{Q}_\alpha$ and every mean other than a weighted arithmetic mean, the quantity is continuous at a pair exactly when the Choi matrix of the sum is invertible or the quantity takes an extreme value there. For bounded maximal f -divergences this holds under the rank condition of Theorem 4.12.4(d)(iii).

Throughout this section, $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, $\Omega = \Phi + \Psi \neq 0$ has column matrix $W \in \text{L}(\mathbb{C}^r, \mathcal{Y} \otimes \mathcal{X})$ as in Theorem 4.2.15, T is the derivative of Φ with respect to Ω , and

$$c_\Omega = \dim(\mathcal{Y} \otimes \mathcal{X}) - r = \dim \ker J(\Omega)$$

is the Choi codimension of Ω . Pairs converge in any norm on the finite-dimensional space of Hermitian-preserving maps, for instance in the completely bounded trace norm. Every operator connection σ has a representing function

$$f_\sigma(x) = a_\sigma + b_\sigma x + \int_{(0,\infty)} \frac{x(1+s)}{x+s} dm_\sigma(s)$$

with $a_\sigma, b_\sigma \geq 0$ and a finite positive measure m_σ [79]. By (4.9.5),

$$\varphi_\sigma = \ell_\sigma + \eta_\sigma, \quad \ell_\sigma(t) = a_\sigma t + b_\sigma(1-t), \quad \eta_\sigma = \int_{(0,\infty)} \eta_s dm_\sigma(s) \geq 0,$$

where η_σ vanishes at 0 and 1, and is strictly positive on $(0,1)$ unless $m_\sigma = 0$. By Remark 4.2.16, $J(\Phi)\sigma J(\Psi) = W\varphi_\sigma(T)W^*$ and $a_\sigma J(\Phi) + b_\sigma J(\Psi) = W\ell_\sigma(T)W^*$. For an operator convex f with $f(0^+)$ and $f'(\infty)$ finite we write $\ell_f(t) = f(0^+)(1-t) + f'(\infty)t$ for the chord of h_f , so that $h_f \leq \ell_f$ on $[0,1]$.

We call Φ and Ψ *mutually singular* if $\text{range } J(\Phi) \cap \text{range } J(\Psi) = \{0\}$. This holds if and only if T is a projection: an eigenvector u of T with eigenvalue in $(0,1)$ gives $Wu \in W \text{range } T \cap W \text{range } (\mathbf{1}_{\mathbb{C}^r} - T)$, and conversely W is injective. By Proposition 4.7.1, it means that $\Phi = \Phi_s$ and $\Psi = \Psi_s$. In that case $J(\Phi)\sigma J(\Psi) = a_\sigma J(\Phi) + b_\sigma J(\Psi)$, and in particular $J(\Phi)\#J(\Psi) = 0$. The equivalence of mutual singularity with $J(\Phi)\#J(\Psi) = 0$ is the criterion following Theorem 4.5.6, which for the geometric mean is due to Kian [76, Theorem 4.3]; for states it is the distinct image property of [29].

Lemma 4.12.1 (Envelope). (a) For positive semidefinite $A, B \in \text{L}(\mathcal{V})$ and every connection σ , $a_\sigma A + b_\sigma B \leq A \sigma B$.

(b) If $A_n \rightarrow A$ and $B_n \rightarrow B$, then every limit point L of $(A_n \sigma B_n)$ satisfies $a_\sigma A + b_\sigma B \leq L \leq A \sigma B$.

(c) $\hat{F}$ and $\hat{Q}_\alpha$, $\alpha \in (0,1)$, are upper semicontinuous on $\text{CP}(\mathcal{X}, \mathcal{Y}) \times \text{CP}(\mathcal{X}, \mathcal{Y})$. For every operator convex f , $\hat{D}_f$ is lower semicontinuous. If $f(0^+)$ and $f'(\infty)$ are finite, then

$$\hat{D}_f(\Phi\|\Psi) \leq \lambda_{\max}(\text{Tr}_{\mathcal{Y}}(f(0^+)J(\Psi) + f'(\infty)J(\Phi))).$$

Proof. (a) For $A > 0$, the inequality $f_\sigma(x) \geq a_\sigma + b_\sigma x$ gives

$$A \sigma B = A^{1/2} f_\sigma(A^{-1/2} B A^{-1/2}) A^{1/2} \geq a_\sigma A + b_\sigma B.$$

The general case follows by continuity from above.

(b) Since $x(1+s)/(x+s) \leq 1+x$, we have $f_\sigma(x) \leq (a_\sigma + m_0) + (b_\sigma + m_0)x$ with $m_0 = m_\sigma((0, \infty))$, so

$$A_n \sigma B_n \leq (a_\sigma + m_0)A_n + (b_\sigma + m_0)B_n$$

is bounded. The lower bound follows from (a), since the positive cone is closed. For $\lambda > 0$, monotonicity gives $A_n \sigma B_n \leq (A_n + \lambda \mathbf{1}_\mathcal{Y}) \sigma (B_n + \lambda \mathbf{1}_\mathcal{Y})$. The right-hand side converges to $(A + \lambda \mathbf{1}_\mathcal{Y}) \sigma (B + \lambda \mathbf{1}_\mathcal{Y})$, because the defining formula is continuous on positive definite pairs. Hence $L \leq (A + \lambda \mathbf{1}_\mathcal{Y}) \sigma (B + \lambda \mathbf{1}_\mathcal{Y})$ for every $\lambda > 0$, and continuity from above gives $L \leq A \sigma B$.

(c) By Proposition 4.3.1, $\hat{F}(\Phi, \Psi) = \lambda_{\min}(\text{Tr}_\mathcal{Y}(J(\Phi) \# J(\Psi)))$, and $\lambda_{\min} \circ \text{Tr}_\mathcal{Y}$ is continuous and monotone. Applying (b) along a subsequence that realises the lim sup gives upper semicontinuity, and Lemma 4.5.3 gives the same for $\hat{Q}_\alpha$. $\hat{D}_f$ is a supremum over ρ of the functions $(\Phi, \Psi) \mapsto \hat{D}_f(\Phi_\rho \| \Psi_\rho)$, which are lower semicontinuous by Lemmas 4.6.4(2) and 4.7.3; see also [60, Section 5], [95, Theorem A.4] and [91]. For the last inequality, $h_f \leq \ell_f$ gives

$$Wh_f(T)W^* \leq f(0^+)J(\Psi) + f'(\infty)J(\Phi);$$

apply Theorem 4.7.5(b). $\square$

Remark 4.12.2. Parts (a) and (b) of Lemma 4.12.1 are elementary consequences of the Kubo–Ando representation and of continuity from above, and the upper semicontinuity of the Matsumoto fidelity was already used in the proof of Proposition 4.3.1. The lower semicontinuity of maximal f -divergences is noted by Matsumoto [91] as a consequence of his dual formula, and is proved in [60, Section 5] and [95, Theorem A.4]. Part (b) also shows that one-sided perturbations cannot produce jumps: if $A_n \geq A$ and $B_n \geq B$ for all n , then monotonicity gives $A_n \sigma B_n \geq A \sigma B$, so every limit point equals $A \sigma B$, and $A_n \sigma B_n \rightarrow A \sigma B$.

Lemma 4.12.3 (Dilation into the Choi kernel). *Suppose $c_\Omega \geq 1$, let K be a matrix whose c_Ω columns form a basis of $\ker J(\Omega)$, and let*

$$\hat{T} = \begin{bmatrix} T & X \\ X^* & Z \end{bmatrix} \quad \text{on } \mathbb{C}^r \oplus \mathbb{C}^{c_\Omega}, \quad 0 \leq \hat{T} \leq \mathbf{1}_{\mathbb{C}^r \oplus \mathbb{C}^{c_\Omega}}.$$

For $t > 0$ put $W_t = [W \quad \sqrt{t}K]$, and let $\Phi_t = \Psi_{\hat{T}}$ and $\Psi_t = \Psi_{\mathbf{1}_{\mathbb{C}^r \oplus \mathbb{C}^{c_\Omega}} - \hat{T}}$ with respect to the Kraus operators given by the columns of W_t . Then $\|J(\Phi_t) - J(\Phi)\|_\infty$ and $\|J(\Psi_t) - J(\Psi)\|_\infty$ are $O(\sqrt{t})$. For every connection σ ,

$$J(\Phi_t) \sigma J(\Psi_t) = W_t \varphi_\sigma(\hat{T}) W_t^* \rightarrow W[\varphi_\sigma(\hat{T})]_{11} W^*$$

as $t \downarrow 0$, where $[\cdot]_{11}$ is the compression to $\mathbb{C}^r$. For every operator convex f with $h_f(\hat{T})$ finite,

$$\hat{D}_f(\Phi_t \| \Psi_t) \rightarrow \lambda_{\max}(\text{Tr}_\mathcal{Y}(W[h_f(\hat{T})]_{11} W^*)).$$

Proof. The columns of W_t are linearly independent, because

$$\text{range } K \cap \text{range } W = \ker J(\Omega) \cap \text{range } J(\Omega) = \{0\}.$$

Since $W_t W_t^* = J(\Omega) + t K K^*$, W_t is the column matrix of $\Phi_t + \Psi_t$, and $\hat{T}$ is the derivative of Φ_t . For any block operator $M = [M_{ij}]$ on $\mathbb{C}^r \oplus \mathbb{C}^{c_\Omega}$,

$$W_t M W_t^* = W M_{11} W^* + \sqrt{t} (W M_{12} K^* + K M_{21} W^*) + t K M_{22} K^*.$$

With $M = \hat{T}$ and $M = \mathbf{1}_{\mathbb{C}^r \oplus \mathbb{C}^{c_\Omega}} - \hat{T}$ this gives the first statement. With $M = \varphi_\sigma(\hat{T})$, Remark 4.2.16 gives the second. With $M = h_f(\hat{T})$, Theorem 4.7.5(b) gives the third, since $\lambda_{\max} \circ \text{Tr}_Y$ is continuous. $\square$

The dilations we use are projections, after Halmos [50]. If $\mathcal{U} \subseteq \mathbb{C}^r$ is invariant under T , write $T = T_{\mathcal{U}} \oplus T_{\mathcal{U}^\perp}$, $R = (T_{\mathcal{U}}(\mathbf{1}_{\mathcal{U}} - T_{\mathcal{U}}))^{1/2}$, $k = \text{rank } R$, and let R_0 be a partial isometry from $\mathcal{U}$ into $\mathbb{C}^{c_\Omega}$ with $R_0^* R_0$ the range projection of R , which requires $k \leq c_\Omega$. Put

$$\hat{T}_{\mathcal{U}} = T_{\mathcal{U}^\perp} \oplus \begin{bmatrix} T_{\mathcal{U}} & R R_0^* \\ R_0 R & R_0 (\mathbf{1}_{\mathcal{U}} - T_{\mathcal{U}}) R_0^* \end{bmatrix} \quad \text{on } \mathcal{U}^\perp \oplus (\mathcal{U} \oplus \mathbb{C}^{c_\Omega}). \quad (4.12.1)$$

Since $T_{\mathcal{U}}$, R and $R_0^* R_0$ commute, a direct computation shows that the second summand $\hat{P}_{\mathcal{U}}$ is a projection, and its compression to $\mathcal{U}$ is $T_{\mathcal{U}}$. Hence for every continuous φ on $[0, 1]$ with chord $\ell_\varphi(t) = \varphi(0)(1-t) + \varphi(1)t$, we have $\varphi(\hat{P}_{\mathcal{U}}) = \ell_\varphi(\hat{P}_{\mathcal{U}})$, and

$$[\varphi(\hat{T}_{\mathcal{U}})]_{11} = \varphi(T_{\mathcal{U}^\perp}) \oplus \ell_\varphi(T_{\mathcal{U}}) = \varphi(T) - P_{\mathcal{U}}(\varphi - \ell_\varphi)(T)P_{\mathcal{U}}, \quad (4.12.2)$$

where $P_{\mathcal{U}}$ denotes the orthogonal projection onto $\mathcal{U}$, which commutes with T .

Theorem 4.12.4 (Continuity dichotomy). *Let Φ, Ψ, Ω, W, T and c_Ω be as above.*

- (a) *If $J(\Omega)$ is invertible, then for every pair (Φ', Ψ') , with $d_{\text{cb}} = \max\{\|\Phi - \Phi'\|_1, \|\Psi - \Psi'\|_1\}$ and $\bar{\Omega} = \frac{1}{2}(\Omega + \Phi' + \Psi')$,*

$$|\hat{F}(\Phi, \Psi) - \hat{F}(\Phi', \Psi')| \leq 2.528 \|\bar{\Omega}^*(\mathbf{1}_Y)\|_\infty \nu(\Omega)^{-1} d_{\text{cb}}^{1/2}.$$

Theorem 4.10.4 and Corollary 4.10.8 give local moduli for every mean, for $\hat{Q}_\alpha$ and for $\hat{D}_f$ with bounded f . In particular all these quantities are continuous at (Φ, Ψ) .

- (b) *Suppose $c_\Omega \geq 1$. Let $\mathcal{U} \subseteq \mathbb{C}^r$ be invariant under T , with $\text{rank } T_{\mathcal{U}}(\mathbf{1}_{\mathcal{U}} - T_{\mathcal{U}}) \leq c_\Omega$, and let (Φ_t, Ψ_t) be the pairs of Lemma 4.12.3 with $\hat{T} = \hat{T}_{\mathcal{U}}$. Then*

$$\|J(\Phi_t) - J(\Phi)\|_\infty, \|J(\Psi_t) - J(\Psi)\|_\infty = O(\sqrt{t}),$$

and simultaneously for all connections σ ,

$$J(\Phi_t) \sigma J(\Psi_t) \longrightarrow W(\varphi_\sigma(T) - P_{\mathcal{U}} \eta_\sigma(T) P_{\mathcal{U}}) W^*,$$

and for every operator convex f with $f(0^+), f'(\infty) < \infty$,

$$\hat{D}_f(\Phi_t \| \Psi_t) \longrightarrow \lambda_{\max} \left(\text{Tr}_Y (W(h_f(T) + P_{\mathcal{U}}(\ell_f - h_f)(T)P_{\mathcal{U}})W^*) \right).$$

- (c) *If $c_\Omega \geq 1$ and $c_\Omega \geq \text{rank } T(\mathbf{1}_{\mathbb{C}^r} - T)$, the choice $\mathcal{U} = \text{range } T(\mathbf{1}_{\mathbb{C}^r} - T)$ in (b) makes Φ_t and Ψ_t mutually singular. (If $c_\Omega = 0$, the condition $\text{rank } T(\mathbf{1}_{\mathbb{C}^r} - T) \leq c_\Omega$ forces T to be a projection; then Φ and Ψ are already mutually singular, and the assertions below hold with $\Phi_t = \Phi$ and $\Psi_t = \Psi$.) Then*

$$\Phi_t \sigma \Psi_t = a_\sigma \Phi_t + b_\sigma \Psi_t \rightarrow a_\sigma \Phi + b_\sigma \Psi,$$

which is the smallest limit allowed by Lemma 4.12.1. Moreover $\hat{F}(\Phi_t, \Psi_t) = \hat{Q}_\alpha(\Phi_t \| \Psi_t) = 0$, and for bounded f ,

$$\hat{D}_f(\Phi_t \| \Psi_t) \rightarrow \lambda_{\max} (\text{Tr}_Y (f(0^+)J(\Psi) + f'(\infty)J(\Phi))),$$

which is the largest value allowed by Lemma 4.12.1.

(d) Consequently:

- (i) for a connection σ with $m_\sigma \neq 0$, the map $(\Phi', \Psi') \mapsto \Phi' \sigma \Psi'$ is continuous at (Φ, Ψ) if and only if $J(\Omega)$ is invertible or Φ and Ψ are mutually singular;
- (ii) $\widehat{F}$ is continuous at (Φ, Ψ) if and only if $J(\Omega)$ is invertible or $\widehat{F}(\Phi, \Psi) = 0$, and the same holds for $\widehat{Q}_\alpha$, $\alpha \in (0, 1)$;
- (iii) for bounded operator convex f , $\widehat{D}_f$ is continuous at (Φ, Ψ) if $J(\Omega)$ is invertible. If $1 \leq c_\Omega$ and $\text{rank } T(\mathbf{1}_{\mathbb{C}^r} - T) \leq c_\Omega$, it is continuous if and only if

$$\widehat{D}_f(\Phi \| \Psi) = \lambda_{\max}(\text{Tr}_{\mathcal{Y}}(f(0^+)J(\Psi) + f'(\infty)J(\Phi)));$$

- (iv) if $f'(\infty) = +\infty$ or $f(0^+) = +\infty$, and $c_\Omega \geq 1$, then every neighbourhood of (Φ, Ψ) contains pairs with $\widehat{D}_f = +\infty$.

Proof. (a) Since $J(\Omega)$ is invertible, $\text{range } J(\Phi' + \Psi') \subseteq \text{range } J(\Omega)$, and Corollary 4.10.8 applies with $(\Phi_1, \Psi_1) = (\Phi, \Psi)$. For the other quantities, the same corollary gives $\delta \leq d_{\text{cb}}/\nu(\Omega)^2$. When $\delta < \frac{1}{2}$, Lemma 4.10.5 gives $\tau \leq 2\sqrt{2}\delta/\sqrt{1-4\delta^2}$ and $\tanh \frac{d_T}{2} = 2\gamma \leq 2\delta$, and Theorem 4.10.4 applies.

(b) Since $\text{rank } R = \text{rank } T_{\mathcal{U}}(\mathbf{1}_{\mathcal{U}} - T_{\mathcal{U}}) \leq c_\Omega$, the dilation (4.12.1) exists and $0 \leq \widehat{T}_{\mathcal{U}} \leq \mathbf{1}_{\mathbb{C}^r \oplus \mathbb{C}^\Omega}$. By Lemma 4.12.3 and (4.12.2) with $\varphi = \varphi_\sigma$, whose chord is ℓ_σ because $\varphi_\sigma(1) = a_\sigma$ and $\varphi_\sigma(0) = b_\sigma$, the means converge to $W(\varphi_\sigma(T) - P_{\mathcal{U}}\eta_\sigma(T)P_{\mathcal{U}})W^*$. With $\varphi = h_f$, whose chord is ℓ_f , the divergences converge as stated.

(c) On $\mathcal{U}^\perp$, $T(\mathbf{1}_{\mathbb{C}^r} - T) = 0$, so $T_{\mathcal{U}^\perp}$ is a projection, and so is $\widehat{T}_{\mathcal{U}}$. Since W_t is injective, Φ_t and Ψ_t are mutually singular, and $\varphi_\sigma(\widehat{T}_{\mathcal{U}}) = \ell_\sigma(\widehat{T}_{\mathcal{U}})$ gives $J(\Phi_t)\sigma J(\Psi_t) = a_\sigma J(\Phi_t) + b_\sigma J(\Psi_t)$. The geometric mean and $\#_{1-\alpha}$ have $a_\sigma = b_\sigma = 0$, so $\widehat{F}(\Phi_t, \Psi_t) = \widehat{Q}_\alpha(\Phi_t \| \Psi_t) = 0$ by Proposition 4.3.1 and Lemma 4.5.3. Similarly $h_f(\widehat{T}_{\mathcal{U}}) = \ell_f(\widehat{T}_{\mathcal{U}})$ gives

$$\widehat{D}_f(\Phi_t \| \Psi_t) = \lambda_{\max}(\text{Tr}_{\mathcal{Y}}(f(0^+)J(\Psi_t) + f'(\infty)J(\Phi_t))),$$

which converges as stated.

(d)(i) If $J(\Omega)$ is invertible, use (a) and Theorem 4.10.4(a). If T is a projection, then $\eta_\sigma(T) = 0$, so the two bounds in Lemma 4.12.1(b) coincide, and every limit point equals $J(\Phi)\sigma J(\Psi)$; since the sequence is bounded, it converges. Conversely, let $c_\Omega \geq 1$ and let u be a unit eigenvector of T with eigenvalue $q \in (0, 1)$. Then $\eta_\sigma(q) > 0$ because $m_\sigma \neq 0$. With $\mathcal{U} = \mathbb{C}u$, (b) produces the limit $W(\varphi_\sigma(T) - \eta_\sigma(q)uu^*)W^*$, which differs from $W\varphi_\sigma(T)W^*$ because W is injective.

(ii) If $J(\Omega)$ is invertible, use (a). If $\widehat{F}(\Phi, \Psi) = 0$, upper semicontinuity (Lemma 4.12.1(c)) and $\widehat{F} \geq 0$ give continuity. Conversely, let $c_\Omega \geq 1$ and $\widehat{F}(\Phi, \Psi) > 0$. Put $M = \text{Tr}_{\mathcal{Y}}(Wg_\#(T)W^*)$ and let x be a unit vector with $\langle x, Mx \rangle = \lambda_{\min}(M) = \widehat{F}(\Phi, \Psi)$. Write $g_\#(T) = \sum_k g_\#(q_k)u_k u_k^*$, the sum over the eigenvalues $q_k \in (0, 1)$ of T , since $g_\#(0) = g_\#(1) = 0$. Since $\langle x, Mx \rangle > 0$, some k has

$$\beta_k := g_\#(q_k)\langle x, \text{Tr}_{\mathcal{Y}}(Wu_k u_k^* W^*)x \rangle > 0.$$

With $\mathcal{U} = \mathbb{C}u_k$, (b) gives

$$\widehat{F}(\Phi_t, \Psi_t) \rightarrow \lambda_{\min}(\text{Tr}_{\mathcal{Y}}(W(g_\#(T) - g_\#(q_k)u_k u_k^*)W^*)) \leq \langle x, Mx \rangle - \beta_k < \widehat{F}(\Phi, \Psi).$$

The same argument applies to $\widehat{Q}_\alpha$, with h_α in place of $g_\#$.

(iii) The first statement is (a), with Theorem 4.10.4(d). Let

$$\mathcal{E}_f(\Phi, \Psi) = \lambda_{\max}(\text{Tr}_{\mathcal{Y}}(f(0^+)J(\Psi) + f'(\infty)J(\Phi))),$$

a continuous function of the pair. If $1 \leq c_\Omega$ and $\text{rank } T(\mathbf{1}_{\mathbb{C}^r} - T) \leq c_\Omega$, then by (c) there are pairs converging to (Φ, Ψ) along which $\hat{D}_f \rightarrow \mathcal{E}_f(\Phi, \Psi)$, so continuity forces $\hat{D}_f(\Phi \| \Psi) = \mathcal{E}_f(\Phi, \Psi)$. Conversely, if equality holds, lower semicontinuity gives $\liminf \hat{D}_f(\Phi' \| \Psi') \geq \hat{D}_f(\Phi \| \Psi)$, and Lemma 4.12.1(c) gives

$$\limsup \hat{D}_f(\Phi' \| \Psi') \leq \lim \mathcal{E}_f(\Phi', \Psi') = \mathcal{E}_f(\Phi, \Psi) = \hat{D}_f(\Phi \| \Psi).$$

(iv) If $\hat{D}_f(\Phi \| \Psi) = +\infty$ there is nothing to prove. Otherwise apply Lemma 4.12.3 with $\hat{T} = T \oplus \mathbf{1}_{\mathbb{C}^{c_\Omega}}$ if $f'(\infty) = +\infty$, and with $\hat{T} = T \oplus 0_{\mathbb{C}^{c_\Omega}}$ if $f(0^+) = +\infty$. Then $J(\Phi_t) = J(\Phi) + tKK^*$ and $J(\Psi_t) = J(\Psi)$ in the first case, and $J(\Phi_t) = J(\Phi)$ and $J(\Psi_t) = J(\Psi) + tKK^*$ in the second, so $(\Phi_t, \Psi_t) \rightarrow (\Phi, \Psi)$. Since $\hat{T}$ has the eigenvalue 1, respectively 0, and W_t is injective, Proposition 4.7.1 shows that Φ_t , respectively Ψ_t , has a nonzero singular part, and Theorem 4.7.5(a) gives $\hat{D}_f(\Phi_t \| \Psi_t) = +\infty$. $\square$

The approximants in (c) separate the Choi supports of the two maps at cost $O(\sqrt{t})$, and every mean then collapses to its affine part. Theorem 4.11.3 records exactly this: for (Φ_t, Ψ_t) against (Φ, Ψ) one has $\tau_{ac} = 0$, $d_{ac} = 0$ and $m_s = O(t)$, but the couplings do not tend to zero. The condition $c_\Omega \geq \text{rank } T(\mathbf{1}_{\mathbb{C}^r} - T)$ holds whenever the Kraus rank of Ω is at most $\frac{1}{2} \dim \mathcal{X} \dim \mathcal{Y}$. When it fails, (b) still gives a nonzero jump whenever (d) asserts discontinuity. Part (iv) is the channel form of the familiar fact that divergences with $f'(\infty) = \infty$ or $f(0^+) = \infty$ require a support condition. Parts (i)–(iii) have no classical counterpart: if all Choi matrices involved commute, $\hat{F}$ is continuous, since $\sqrt{xy}$ is.

The ingredients of Theorem 4.12.4 are classical, but we are not aware of an earlier description of the points of continuity. Discontinuities at singular pairs are known: the Matsumoto fidelity of two distinct pure states vanishes [29], which is the mechanism of Corollary 4.3.3, and the Belavkin–Staszewski conditional entropy is discontinuous on singular states, with continuity bounds of order $\sqrt{\delta}/(\lambda_-(d^{-1} - \lambda_-))$ on states on a d -dimensional space whose smallest eigenvalue is at least λ_- , where $0 < \lambda_- < d^{-1}$ [21]. Theorem 4.12.4(a) is an analogue for completely positive maps, with a constant governed by $\nu(\Omega)$, which satisfies $\lambda_{\min}(J(\Omega))/\min\{\dim \mathcal{X}, \dim \mathcal{Y}\} \leq \nu(\Omega)^2 \leq \lambda_{\min}(J(\Omega))$ for invertible $J(\Omega)$ by Lemma 4.8.2(2). The ways of approaching a singular pair that are used in the literature are all compatible with Theorem 4.12.4: regularisation from above [8, 79] and positive semidefinite perturbations, as in the perturbation bounds for parallel sums of [86], converge by Remark 4.12.2, and, as observed in the introduction of [91], maximal f -divergences of states are continuous along line segments into the interior of their domain. By Corollary 4.10.8 and Remark 4.12.2, a jump requires perturbations (Φ_n, Ψ_n) that enlarge the Choi support of the sum and do not satisfy both $\Phi_n \geq \Phi$ and $\Psi_n \geq \Psi$.

Corollary 4.12.5 (States). *For positive semidefinite ρ_1, ρ_2 on $\mathbb{C}^d$ with $\rho_1 + \rho_2 \neq 0$, the Matsumoto fidelity $\text{tr}(\rho_1 \# \rho_2)$ is continuous at (ρ_1, ρ_2) if and only if $\rho_1 + \rho_2$ is invertible or $\text{tr}(\rho_1 \# \rho_2) = 0$. If $\text{rank}(\rho_1 + \rho_2) \leq d/2$, there are mutually singular pairs $(\rho_{1,t}, \rho_{2,t}) \rightarrow (\rho_1, \rho_2)$, which can be normalised to states when ρ_1, ρ_2 are states, and $\text{tr}(\rho_{1,t} \# \rho_{2,t}) = 0$.*

Proof. Take $\dim \mathcal{X} = 1$ in Theorem 4.12.4. Then $\hat{F}(\rho_1, \rho_2) = \text{tr}(\rho_1 \# \rho_2)$, and

$$\text{rank } T(\mathbf{1}_{\mathbb{C}^r} - T) \leq \text{rank}(\rho_1 + \rho_2) \leq d - \text{rank}(\rho_1 + \rho_2) = c_\Omega.$$

Normalising each map by its trace preserves mutual singularity and convergence. $\square$

For example, $\rho_1 = \rho_2 = \frac{1}{2}(e_1 e_1^* + e_2 e_2^*)$ on $\mathbb{C}^4$ has $\text{tr}(\rho_1 \# \rho_2) = 1$. Coupling e_1 to e_3 as in (4.12.1) gives $\rho_{1,t} = \frac{1}{2}(v v^* + e_2 e_2^*)$ and $\rho_{2,t} = \frac{1}{2}(w w^* + e_2 e_2^*)$ with $v, w = e_1 \pm \sqrt{t} e_3$, which are not unit vectors, so that $\text{tr} \rho_{1,t} = \text{tr} \rho_{2,t} = 1 + \frac{t}{2}$. These are at trace distance

$O(\sqrt{t})$ from ρ_1 , and $\text{tr}(\rho_{1,t}\#\rho_{2,t}) = \frac{1}{2}$ (after normalisation to states, $\frac{1}{2}(1 + \frac{t}{2})^{-1}$), because the mean splits over $\mathbb{C}e_2 \oplus \text{span}\{e_1, e_3\}$ and $vv^*\#ww^* = 0$. Coupling e_2 to e_4 as well gives fidelity 0.

Corollary 4.12.6 (Channels). *Let Φ be a channel. Then $\widehat{F}(\Phi, \Phi) = 1$, and if $J(\Phi)$ is invertible, $\widehat{F}$ is continuous at (Φ, Φ) . If there are channels $\Phi_n \rightarrow \Phi$ with $\text{range } J(\Phi_n) \cap \text{range } J(\Phi) = \{0\}$, then $\widehat{F}(\Phi, \Phi_n) = 0$ for all n . This is the case for every isometric channel $\Phi = V(\cdot)V^*$ with $\dim \mathcal{Y} \geq 2$: take $\Phi_n = U_n V(\cdot)V^* U_n^*$ with unitaries $U_n \rightarrow \mathbf{1}_{\mathcal{Y}}$ such that $U_n V \notin \mathbb{C}V$.*

Proof. We have

$$\widehat{F}(\Phi, \Phi) = \lambda_{\min}(\text{Tr}_{\mathcal{Y}} J(\Phi)) = \lambda_{\min}(\Phi^*(\mathbf{1}_{\mathcal{Y}})^{\top}) = 1.$$

Continuity is Theorem 4.12.4(a). If the Choi ranges intersect trivially, the pair (Φ, Φ_n) is mutually singular, so $J(\Phi)\#J(\Phi_n) = 0$. For an isometric channel the Choi matrices are $|V\rangle\langle V|$ and $|U_n V\rangle\langle U_n V|$, whose ranges intersect trivially when $U_n V \notin \mathbb{C}V$. Choose a Hermitian H with $HV \notin \mathbb{C}V$, which is possible because $\dim \mathcal{Y} \geq 2$; for instance $H = yy^*$ with $y^*V \neq 0$, and with $y \notin \mathbb{C}V$ if $\dim \mathcal{X} = 1$. Since $s^{-1}(e^{isH}V - V) \rightarrow iHV$ as $s \rightarrow 0$ and $\mathbb{C}V$ is closed, $e^{isH}V \notin \mathbb{C}V$ for all sufficiently small $s \neq 0$. Hence $U_n = e^{iH/n}$ has the required property for all $n \geq n_0$, for some n_0 , and we use these n . $\square$

Thus the geometric channel fidelity, viewed as a function of its second argument, is discontinuous at every isometric channel, although it equals 1 there. This is the distinct image property of [29], applied to the rank-one Choi matrices of isometric channels. By Theorem 4.12.9 and Corollary 4.12.10 below, $\widehat{F}(\Phi, \cdot)$ is discontinuous along channel-valued approximants at every channel Φ with non-invertible Choi matrix. Output rotations of Φ suffice, although the limiting value of the fidelity need not be 0.

Remark 4.12.7 (Physical interpretation). Corollary 4.12.6 is a caution: an ideal gate $U(\cdot)U^*$ and a mis-rotated gate $U'(\cdot)U'^*$, $U' \notin \mathbb{C}U$, have $\widehat{F} = 0$ however close they are in $\|\cdot\|_1$, so near unitaries $\widehat{F}$ is not a usable plug-in figure of merit. Noisy devices often have invertible Choi matrices, and as $J(\Phi + \Psi) \geq J(\Phi)$, full Kraus rank of one of the maps suffices. The generalised amplitude damping channel with damping p and thermal population N , a qubit in a bath at nonzero temperature, has four Kraus operators [74], linearly independent for $p \in (0, 1]$ and $N \in (0, 1)$ (two are off-diagonal in different positions, two are diagonal and independent as $p \neq 0$), so pairs containing it are continuity points by Theorem 4.12.4(a). At zero temperature ($N = 0 < p < 1$) the Kraus rank is 2, and as $\widehat{F}(\Phi, \Phi) = 1$, Theorem 4.12.4(d)(ii) shows that $\widehat{F}$ is discontinuous at (Φ, Φ) under completely positive perturbations. Example 4.12.8 below shows that the jump persists within the zero-temperature amplitude damping family itself, where it is a jump from 1 to 0.

Perturbations within channels

Example 4.12.8 (Amplitude damping). For $\gamma \in [0, 1]$ let $\mathcal{A}_{\gamma} \in \text{CP}(\mathbb{C}^2, \mathbb{C}^2)$ be the amplitude damping channel with Kraus operators $K_0 = e_0 e_0^* + \sqrt{1-\gamma} e_1 e_1^*$ and $K_1 = \sqrt{\gamma} e_0 e_1^*$, that is, the channel of Remark 4.12.7 with $N = 0$ and damping γ . For $\gamma \neq \gamma'$,

$$\widehat{F}(\mathcal{A}_{\gamma}, \mathcal{A}_{\gamma'}) = 0, \quad \|\mathcal{A}_{\gamma} - \mathcal{A}_{\gamma'}\|_1 \leq |\gamma - \gamma'| + \left((\gamma - \gamma')^2 + 4(\sqrt{1-\gamma} - \sqrt{1-\gamma'})^2 \right)^{1/2},$$

while $\widehat{F}(\mathcal{A}_{\gamma}, \mathcal{A}_{\gamma}) = 1$. Thus, at every $\gamma \in [0, 1]$, $\widehat{F}(\mathcal{A}_{\gamma}, \cdot)$ jumps from 1 to 0 within the amplitude damping family. For $\gamma < 1$ the distance is $O(|\gamma - \gamma'|)$.

Proof. Write $e_{ab} = e_a \otimes e_b \in \mathcal{Y} \otimes \mathcal{X}$ and $c = \sqrt{1-\gamma}$, $c' = \sqrt{1-\gamma'}$. Then $\text{vec}(K_0) = e_{00} + c e_{11}$ and $\text{vec}(K_1) = \sqrt{\gamma} e_{01}$, so $\text{range } J(\mathcal{A}_\gamma) \subseteq \text{span}\{e_{00} + c e_{11}, e_{01}\}$. A vector

$$\alpha(e_{00} + c e_{11}) + \beta e_{01} = \alpha'(e_{00} + c' e_{11}) + \beta' e_{01}$$

has $\alpha = \alpha'$ and $\alpha c = \alpha' c'$, hence $\alpha = 0$ since $c \neq c'$. So $\text{range } J(\mathcal{A}_\gamma) \cap \text{range } J(\mathcal{A}_{\gamma'}) \subseteq \mathbb{C} e_{01}$, and Theorem 4.5.6(1) gives $J(\mathcal{A}_\gamma) \# J(\mathcal{A}_{\gamma'}) = \kappa e_{01} e_{01}^*$ for some $\kappa \geq 0$. Its partial trace $\kappa e_{11} e_1^*$ is singular on $\mathcal{X} = \mathbb{C}^2$, so Proposition 4.3.1 gives $\widehat{F} = 0$. The value 1 is Corollary 4.12.6. For the distance, let Δ be Hermitian-preserving. Its completely bounded trace norm is attained at a pure state uu^* , $u = \text{vec}(A)$ with $\|A\|_2 = 1$ [127, Section 3.3], and (4.3.1), which is linear in the map, gives

$$\|(\Delta \otimes \mathbf{1}_{L(\mathcal{X})})(uu^*)\|_1 \leq \|A\|_\infty^2 \|J(\Delta)\|_1 \leq \|J(\Delta)\|_1.$$

Here

$$J(\mathcal{A}_\gamma) - J(\mathcal{A}_{\gamma'}) = (\gamma - \gamma') e_{01} e_{01}^* + (c - c')(e_{00} e_{11}^* + e_{11} e_{00}^*) + (\gamma' - \gamma) e_{11} e_{11}^*,$$

and the block

$$\begin{pmatrix} 0 & x \\ x & y \end{pmatrix}$$

on $\text{span}\{e_{00}, e_{11}\}$, with $x = c - c'$ and $y = \gamma' - \gamma$, has eigenvalues $\frac{1}{2}(y \pm (y^2 + 4x^2)^{1/2})$ of opposite signs, so its trace norm is $(y^2 + 4x^2)^{1/2}$. $\square$

For $0 < \gamma, \gamma' \leq 1$, Proposition 7.6.8(b) in Chapter 7 proves $\widehat{F}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) = 0$ by a different argument, from the level decomposition of the pair, and deduces that the geometric Chernoff and barycentric converses are infinite there.

The mechanism of Example 4.12.8, a perturbation that moves the Choi range of a channel off itself, is available at every channel with singular Choi matrix. We use the short $[A]_{\mathcal{S}}$ of a positive semidefinite operator A to a subspace $\mathcal{S}$, the largest operator Z with $0 \leq Z \leq A$ and $\text{range } Z \subseteq \mathcal{S}$ [2, 4].

Theorem 4.12.9. *Let $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, write $\mathcal{R}_\Phi = \text{range } J(\Phi)$, $\mathcal{R}_\Psi = \text{range } J(\Psi)$, $\mathcal{R}_\Omega = \text{range } J(\Phi + \Psi) = \mathcal{R}_\Phi + \mathcal{R}_\Psi$ and $G = J(\Phi) \# J(\Psi)$. For Hermitian $H \in L(\mathcal{Y})$ let $U_t = e^{-itH}$, $\Psi_t = U_t \Psi(\cdot) U_t^*$, and*

$$\mathcal{S}_H = \{w \in \mathcal{R}_\Phi \cap \mathcal{R}_\Psi : (H \otimes \mathbf{1}_{\mathcal{X}})w \in \mathcal{R}_\Omega\}.$$

(a) $\|\Psi_t - \Psi\|_1 \leq 2t\|H\|_\infty \|\Psi\|_1$, and

$$\limsup_{t \rightarrow 0} \widehat{F}(\Phi, \Psi_t) \leq \lambda_{\min}(\text{Tr}_{\mathcal{Y}}[G]_{\mathcal{S}_H}).$$

In particular $\widehat{F}(\Phi, \cdot)$ is discontinuous at Ψ if some unit vector x with $\langle x, \text{Tr}_{\mathcal{Y}}(G)x \rangle = \widehat{F}(\Phi, \Psi)$ satisfies $\langle x, \text{Tr}_{\mathcal{Y}}(G - [G]_{\mathcal{S}_H})x \rangle > 0$.

(b) *Let $\Phi = \Psi$ be a channel with $J(\Phi)$ singular, $\mathcal{R} = \text{range } J(\Phi)$, and $\Phi_t = U_t \Phi(\cdot) U_t^*$. Then*

$$\mathcal{S}_H = \mathcal{R}_H = \{w \in \mathcal{R} : (H \otimes \mathbf{1}_{\mathcal{X}})w \in \mathcal{R}\}$$

and

$$\limsup_{t \rightarrow 0} \widehat{F}(\Phi, \Phi_t) \leq 1 - \lambda_{\max}(\text{Tr}_{\mathcal{Y}}(J(\Phi) - [J(\Phi)]_{\mathcal{R}_H})).$$

The Hermitian H with $\mathcal{R}_H = \mathcal{R}$ form a proper real subspace of the Hermitian operators on $\mathcal{Y}$, and for every other H the right-hand side is strictly less than 1. If $\mathcal{R}_H = \{0\}$, it is 0.

Proof. (a) For a unitary U and any X ,

$$(U \otimes \mathbf{1})X(U \otimes \mathbf{1})^* - X = ((U - \mathbf{1}) \otimes \mathbf{1})X(U \otimes \mathbf{1})^* + X((U - \mathbf{1}) \otimes \mathbf{1})^*,$$

so $\|U(\cdot)U^* - \mathbf{1}_{\mathcal{L}(\mathcal{Y})}\|_1 \leq 2\|U - \mathbf{1}_{\mathcal{Y}}\|_\infty$, and $\|e^{-itH} - \mathbf{1}_{\mathcal{Y}}\|_\infty \leq t\|H\|_\infty$ by the spectral theorem; the first claim follows by submultiplicativity of $\|\cdot\|_1$ under composition. Choose $t_n \rightarrow 0$ such that $\widehat{F}(\Phi, \Psi_{t_n})$ converges to the limit superior and, using the boundedness in the proof of Lemma 4.12.1(b), such that $G_n = J(\Phi) \# J(\Psi_{t_n})$ converges to some L . Since $J(\Psi_{t_n}) \rightarrow J(\Psi)$, Lemma 4.12.1(b) gives $L \leq G$. We claim that $\text{range } L \subseteq \mathcal{S}_H$. Let $w = Lx$ and $w_n = G_n x$, so $w_n \rightarrow w$. With $\mathbf{U}_n = U_{t_n} \otimes \mathbf{1}_{\mathcal{X}}$ we have $J(\Psi_{t_n}) = \mathbf{U}_n J(\Psi) \mathbf{U}_n^*$, so Theorem 4.5.6(1) gives $w_n \in \mathcal{R}_\Phi \cap \mathbf{U}_n \mathcal{R}_\Psi$. Hence $w = \lim w_n \in \mathcal{R}_\Phi$ and $w = \lim \mathbf{U}_n^* w_n \in \mathcal{R}_\Psi$. Moreover

$$t_n^{-1}(\mathbf{U}_n^* - \mathbf{1})w_n = t_n^{-1}(\mathbf{U}_n^* w_n - w_n) \in \mathcal{R}_\Psi + \mathcal{R}_\Phi = \mathcal{R}_\Omega,$$

and as $t^{-1}(e^{itH} - \mathbf{1}_{\mathcal{Y}}) \rightarrow iH$ these vectors converge to $i(H \otimes \mathbf{1}_{\mathcal{X}})w$, so $(H \otimes \mathbf{1}_{\mathcal{X}})w \in \mathcal{R}_\Omega$. This proves the claim. As $0 \leq L \leq G$ and $\text{range } L \subseteq \mathcal{S}_H$, maximality of the short gives $L \leq [G]_{\mathcal{S}_H}$. By Proposition 4.3.1 and continuity and monotonicity of $\lambda_{\min} \circ \text{Tr}_{\mathcal{Y}}$, the limit superior equals $\lambda_{\min}(\text{Tr}_{\mathcal{Y}} L) \leq \lambda_{\min}(\text{Tr}_{\mathcal{Y}} [G]_{\mathcal{S}_H})$. For the last statement put $D = \text{Tr}_{\mathcal{Y}}(G - [G]_{\mathcal{S}_H}) \geq 0$; then

$$\lambda_{\min}(\text{Tr}_{\mathcal{Y}} G - D) \leq \langle x, \text{Tr}_{\mathcal{Y}}(G)x \rangle - \langle x, Dx \rangle < \widehat{F}(\Phi, \Psi).$$

(b) Here $\mathcal{R}_\Phi = \mathcal{R}_\Psi = \mathcal{R}_\Omega = \mathcal{R}$ and $G = J(\Phi) \# J(\Phi) = J(\Phi)$, so $\mathcal{S}_H = \mathcal{R}_H$. Since $\text{Tr}_{\mathcal{Y}} J(\Phi) = \Phi^*(\mathbf{1}_{\mathcal{Y}})^\top = \mathbf{1}_{\mathcal{X}}$, we have

$$\lambda_{\min}(\text{Tr}_{\mathcal{Y}} [J(\Phi)]_{\mathcal{R}_H}) = 1 - \lambda_{\max}(\text{Tr}_{\mathcal{Y}} (J(\Phi) - [J(\Phi)]_{\mathcal{R}_H})),$$

and (a) gives the bound. The set $\mathcal{I}$ of Hermitian H with $(H \otimes \mathbf{1}_{\mathcal{X}})\mathcal{R} \subseteq \mathcal{R}$, that is, with $\mathcal{R}_H = \mathcal{R}$, is a real linear subspace. If it contained every Hermitian operator, $\mathcal{R}$ would be invariant under $\mathcal{L}(\mathcal{Y}) \otimes \mathbf{1}_{\mathcal{X}}$, so its projection would commute with this $*$ -algebra and hence be of the form $\mathbf{1}_{\mathcal{Y}} \otimes P'$. Then $J(\Phi) = (\mathbf{1}_{\mathcal{Y}} \otimes P')J(\Phi)(\mathbf{1}_{\mathcal{Y}} \otimes P')$ and

$$\mathbf{1}_{\mathcal{X}} = \text{Tr}_{\mathcal{Y}} J(\Phi) = P' \text{Tr}_{\mathcal{Y}} J(\Phi) P' = P',$$

so $\mathcal{R} = \mathcal{Y} \otimes \mathcal{X}$, contradicting the singularity of $J(\Phi)$. For $H \notin \mathcal{I}$ we have $\mathcal{R}_H \neq \mathcal{R}$, so $D = J(\Phi) - [J(\Phi)]_{\mathcal{R}_H}$ is positive semidefinite and nonzero, since $[J(\Phi)]_{\mathcal{R}_H}$ has range in $\mathcal{R}_H$. Then $\text{tr Tr}_{\mathcal{Y}} D = \text{tr } D > 0$ and $\lambda_{\max}(\text{Tr}_{\mathcal{Y}} D) > 0$. If $\mathcal{R}_H = \{0\}$, then $[J(\Phi)]_{\mathcal{R}_H} = 0$ and the bound is $1 - \lambda_{\max}(\mathbf{1}_{\mathcal{X}}) = 0$. $\square$

Corollary 4.12.10. *For a channel $\Phi \in \text{CP}(\mathcal{X}, \mathcal{Y})$ the following are equivalent:*

- (i) $J(\Phi)$ is invertible;
- (ii) $\widehat{F}$, restricted to pairs of channels, is continuous at (Φ, Φ) ;
- (iii) $\widehat{F}(\Phi, U\Phi(\cdot)U^*) \rightarrow 1$ as the unitary U tends to $\mathbf{1}_{\mathcal{Y}}$.

In particular the dichotomy of Theorem 4.12.4 holds within channels at all diagonal pairs.

Proof. (i) $\Rightarrow$ (ii) is Theorem 4.12.4(a). (ii) $\Rightarrow$ (iii) holds because $\widehat{F}(\Phi, \Phi) = 1$ (Corollary 4.12.6) and $U\Phi(\cdot)U^*$ is a channel converging to Φ . (iii) $\Rightarrow$ (i) is the contrapositive of Theorem 4.12.9(b). $\square$

For zero-temperature amplitude damping and $H = \sigma_x$, one checks directly that $\mathcal{R}_H = \{0\}$: the rotated vector

$$(\sigma_x \otimes \mathbf{1})(\alpha(e_{00} + c e_{11}) + \beta e_{01}) = \alpha(e_{10} + c e_{01}) + \beta e_{11}$$

lies in $\mathcal{R}$ only if $\alpha = \beta = 0$. Hence output rotations also drive the fidelity of Example 4.12.8 to 0. If $2 \operatorname{rank} J(\Phi) > \dim \mathcal{X} \dim \mathcal{Y}$, then

$$\dim \mathcal{R}_H \geq 2 \operatorname{rank} J(\Phi) - \dim \mathcal{X} \dim \mathcal{Y} > 0$$

for every H , since the preimage of $\mathcal{R}$ under $H \otimes \mathbf{1}_{\mathcal{X}}$ has dimension at least $\dim \mathcal{R}$; then the bound of Theorem 4.12.9(b) need not vanish.

By Lemma 4.12.1(c), the fidelity is continuous wherever it vanishes. We now turn to the rate of continuity there; at mutually singular pairs it is governed by the smallest angle between the Choi ranges.

Theorem 4.12.11. *Let $\Phi, \Psi \in \operatorname{CP}(\mathcal{X}, \mathcal{Y})$ be nonzero and mutually singular, let P and Q be the projections onto $\operatorname{range} J(\Phi)$ and $\operatorname{range} J(\Psi)$, let $\theta \in (0, \frac{\pi}{2}]$ be given by $\cos \theta = \|PQ\|_{\infty}$, and put $a = \max\{\|J(\Phi)\|_{\infty}, \|J(\Psi)\|_{\infty}\}$. For $\Phi', \Psi' \in \operatorname{CP}(\mathcal{X}, \mathcal{Y})$ with*

$$\eta = \max\{\|J(\Phi') - J(\Phi)\|_{\infty}, \|J(\Psi') - J(\Psi)\|_{\infty}\} \leq a,$$

$$\widehat{F}(\Phi', \Psi') \leq \frac{\operatorname{tr}(J(\Phi') \# J(\Psi'))}{\dim \mathcal{X}} \leq \frac{3\sqrt{2} \dim \mathcal{Y} (a\eta)^{1/2}}{\sin \theta}.$$

If all four maps are channels, then with $d_{\text{cb}} = \max\{\|\Phi' - \Phi\|_1, \|\Psi' - \Psi\|_1\}$,

$$\widehat{F}(\Phi', \Psi') \leq \frac{3 \dim \mathcal{X} \dim \mathcal{Y} d_{\text{cb}}^{1/2}}{\sin \theta}.$$

Proof. Since $\operatorname{range} J(\Phi) \cap \operatorname{range} J(\Psi) = \{0\}$, $\|PQ\|_{\infty} < 1$, so $\theta > 0$. As $J(\Phi) \leq aP$, we have $J(\Phi') \leq aP + \eta \mathbf{1}$ and likewise $J(\Psi') \leq aQ + \eta \mathbf{1}$, so $J(\Phi') \# J(\Psi') \leq A \# B$ with $A = aP + \eta \mathbf{1}$ and $B = aQ + \eta \mathbf{1}$, by monotonicity. By Halmos' two-subspace theorem [50], in finite dimensions, $\mathcal{Y} \otimes \mathcal{X}$ is an orthogonal direct sum of subspaces reducing both P and Q : the subspaces $\operatorname{range} P \cap \ker Q$, $\ker P \cap \operatorname{range} Q$ and $\ker P \cap \ker Q$ (here $\operatorname{range} P \cap \operatorname{range} Q = \{0\}$), and two-dimensional subspaces $\mathcal{V}_j$ on which $P = u_j u_j^*$ and $Q = v_j v_j^*$ for unit vectors with

$$|\langle u_j, v_j \rangle| = \cos \theta_j \leq \|PQ\|_{\infty} = \cos \theta.$$

The mean $A \# B$ splits accordingly, and we bound its trace on each summand. On $\operatorname{range} P \cap \ker Q$ and $\ker P \cap \operatorname{range} Q$ it is

$$(a + \eta) \# \eta = (\eta(a + \eta))^{1/2} \leq (2a\eta)^{1/2}$$

per dimension, and on $\ker P \cap \ker Q$ it is $\eta \leq (a\eta)^{1/2}$ per dimension. On $\mathcal{V}_j$ the compressions A_j, B_j have $\det A_j = \det B_j = \eta(a + \eta) =: \alpha^2$, and

$$A_j \# B_j = \frac{\alpha(A_j + B_j)}{\det(A_j + B_j)^{1/2}}.$$

Indeed, $N = A_j^{-1/2} B_j A_j^{-1/2}$ has determinant 1, so its eigenvalues are μ, μ^{-1} , and $(\mathbf{1} + N) \det(\mathbf{1} + N)^{-1/2}$ has eigenvalues $(1 + \mu)/((1 + \mu)(1 + \mu^{-1}))^{1/2} = \mu^{1/2}$ and $\mu^{-1/2}$, so it equals $N^{1/2}$; conjugating by $A_j^{1/2}$ and using $\det(A_j + B_j) = \alpha^2 \det(\mathbf{1} + N)$ gives the formula. Now $\operatorname{tr}(A_j + B_j) = 2a + 4\eta$, and the eigenvalues $1 \pm \cos \theta_j$ of $u_j u_j^* + v_j v_j^*$ give

$$\det(A_j + B_j) = a^2 \sin^2 \theta_j + 4a\eta + 4\eta^2 \geq a^2 \sin^2 \theta.$$

Hence, using $\eta \leq a$,

$$\mathrm{tr}(A_j \# B_j) \leq \frac{2(\eta(a + \eta))^{1/2}(a + 2\eta)}{a \sin \theta} \leq \frac{6\sqrt{2}(a\eta)^{1/2}}{\sin \theta},$$

that is, at most $3\sqrt{2}(a\eta)^{1/2}/\sin \theta$ per dimension. As $(2a\eta)^{1/2}$ and $(a\eta)^{1/2}$ are smaller, each summand contributes at most this amount per dimension, and

$$\mathrm{tr}(J(\Phi') \# J(\Psi')) \leq \mathrm{tr}(A \# B) \leq \frac{3\sqrt{2} \dim \mathcal{X} \dim \mathcal{Y} (a\eta)^{1/2}}{\sin \theta}.$$

Finally

$$\widehat{F}(\Phi', \Psi') = \lambda_{\min}(\mathrm{Tr}_{\mathcal{Y}}(J(\Phi') \# J(\Psi'))) \leq \mathrm{tr}(J(\Phi') \# J(\Psi')) / \dim \mathcal{X}.$$

For channels, $J(\Phi) = \dim \mathcal{X} (\Phi \otimes \mathbf{1}_{L(\mathcal{X})})(\tau)$ with τ as in the proof of Lemma 4.8.7, so $a \leq \dim \mathcal{X}$, and $\eta \leq \frac{1}{2} \dim \mathcal{X} d_{\mathrm{cb}}$ by Lemma 4.8.7. If $\eta \leq a$, then $(a\eta)^{1/2} \leq \dim \mathcal{X} (d_{\mathrm{cb}}/2)^{1/2}$ and the first bound gives the second. If $\eta > a$, then

$$\frac{1}{2} \dim \mathcal{X} d_{\mathrm{cb}} > a \geq \|J(\Phi)\|_{\infty} \geq \mathrm{tr} J(\Phi) / (\dim \mathcal{X} \dim \mathcal{Y}) = 1 / \dim \mathcal{Y},$$

so the right-hand side exceeds $3(2 \dim \mathcal{X} \dim \mathcal{Y})^{1/2} > 1$, while $\widehat{F}(\Phi', \Psi') \leq 1$ for channels, as in the proof of Corollary 4.10.7. $\square$

Example 4.12.12. The exponent $\frac{1}{2}$ and the factor $(\sin \theta)^{-1}$ in Theorem 4.12.11 are sharp. Let $\mathcal{Y} = \mathbb{C}^2$, let b, f, g be unit vectors with $f \perp g$ and $|\langle f, b \rangle| = \cos \theta$, $\theta \in (0, \frac{\pi}{2}]$, and let $\Phi = \mathrm{tr}(\cdot) f f^*$ and $\Psi = \mathrm{tr}(\cdot) b b^*$. These are mutually singular, with range $J(\Phi) = \mathbb{C} f \otimes \mathcal{X}$, range $J(\Psi) = \mathbb{C} b \otimes \mathcal{X}$ and $\|PQ\|_{\infty} = |\langle f, b \rangle| = \cos \theta$. For $0 < \epsilon \leq \sin^2 \theta$ put $\Phi' = \mathrm{tr}(\cdot) \rho_{\epsilon}$ with $\rho_{\epsilon} = (1 - \epsilon) f f^* + \epsilon g g^*$, so that

$$\|\Phi' - \Phi\|_1 = \|\rho_{\epsilon} - f f^*\|_1 = 2\epsilon =: d_{\mathrm{cb}},$$

as in the proof of Corollary 4.3.3. For $A > 0$ the Kubo–Ando formula gives $A \# b b^* = \langle b, A^{-1} b \rangle^{-1/2} b b^*$, and, as in Example 4.8.11, $\widehat{F}(\Phi', \Psi) = \mathrm{tr}(\rho_{\epsilon} \# b b^*)$. Hence

$$\widehat{F}(\Phi', \Psi) = \left(\frac{\cos^2 \theta}{1 - \epsilon} + \frac{\sin^2 \theta}{\epsilon} \right)^{-1/2} \geq \left(\frac{\epsilon}{2 \sin^2 \theta} \right)^{1/2} = \frac{d_{\mathrm{cb}}^{1/2}}{2 \sin \theta},$$

using $\epsilon \cos^2 \theta \leq (1 - \epsilon) \sin^2 \theta$. As $\epsilon \rightarrow 0$ with θ fixed, $\widehat{F}(\Phi', \Psi) \sin \theta / d_{\mathrm{cb}}^{1/2} \rightarrow 2^{-1/2}$.

4.13 Conclusion and outlook

Channel divergences of the form $H(J(\Phi) \sigma J(\Psi))$ with $H(0) = 0$ and $f_{\sigma}(0^+) = f'_{\sigma}(\infty) = 0$, such as those built from the geometric mean, vanish on pairs with trivially intersecting Choi ranges (Theorem 4.5.6); the fidelity, $\widehat{Q}_{\alpha}$ with $\alpha \in (0, 1)$ and the bounded maximal f -divergences with f not affine are not continuous in the diamond norm, even for pairs with the same sum (Corollaries 4.3.3 and 4.5.14).

The Arveson–Radon–Nikodym derivative T of Φ with respect to $\Omega = \Phi + \Psi$ is the coordinate in which these quantities behave well. Its transpose is an effect on the environment of Ω , so it records how an instrument splits a fixed device into two branches, and $\|T_1 - T_2\|_{\infty}$ measures perturbations relative to Ω in the completely positive order. Every quantity considered is a scalar function of T , namely g_{σ} or h_f , read through the Kraus operators of Ω . Continuity in T therefore holds with the dimension-free constant

$\|\Omega^*(\mathbf{1}_Y)\|_\infty$. The integral representations of means and of operator convex functions give explicit moduli, which are exact for the one-pole kernels and, for operator convex f on $[\epsilon, 1 - \epsilon]$, within a factor 3 of optimal (Section 4.9). Quantitative estimates for unbounded divergences need a bound on $D_{\max}$ within each pair (without it they are still continuous in T with values in $(-\infty, +\infty]$), and for a general operator convex f the Lebesgue decomposition of the pair decides finiteness. Theorem 4.8.4 converts these estimates into continuity in the diamond norm on the splittings of a fixed Ω ; for the fidelity this is Hölder continuity of order $\frac{1}{2}$ with constant $\|\Omega^*(\mathbf{1}_Y)\|_\infty/\nu(\Omega)$, and in the counterexample $\nu(\Omega) \rightarrow 0$. When the sums differ, their change enters through the Thompson metric or, if their Choi supports differ, through the singular parts, the absolutely continuous parts and the couplings of the Lebesgue decomposition of the sums.

The dichotomy of Theorem 4.12.4 says where continuity holds. The fidelity, $\hat{Q}_\alpha$ with $\alpha \in (0, 1)$, the means and the bounded maximal f -divergences are continuous at every pair whose sum has an invertible Choi matrix. At a pair whose sum has a singular Choi matrix each of them, as a function on pairs of completely positive maps, is continuous exactly when it takes an extreme value there (0 for $\hat{F}$ and $\hat{Q}_\alpha$, $a_\sigma J(\Phi) + b_\sigma J(\Psi)$ for a mean, $\lambda_{\max}(\text{Tr}_Y(f(0^+)J(\Psi) + f'(\infty)J(\Phi)))$ for $\hat{D}_f$); for means this requires the mean not to be a weighted arithmetic mean, and for bounded maximal f -divergences the ‘only if’ needs a rank condition. For channels this separates noisy devices of full Kraus rank, which are continuity points, from isometric gates, at which $\hat{F}(\Phi, \cdot)$ jumps from 1 to 0 (Corollary 4.12.6). Within channels, (Φ, Φ) is a continuity point exactly when $J(\Phi)$ is invertible (Corollary 4.12.10), and for zero-temperature amplitude damping the fidelity jumps from 1 to 0 within the family itself (Example 4.12.8). Quantitatively, near pairs whose Choi matrices are bounded below by λ the fidelity is Lipschitz with relative constant of the sharp order λ^{-1} (Theorem 4.8.8), and at mutually singular pairs it tends to 0 at the sharp rate $d^{1/2}/\sin \theta$, where d is the diamond distance and θ the smallest angle between the Choi ranges (Theorem 4.12.11). Our estimates complement the joint lower semicontinuity of the operator-valued maximal f -divergences of [101, Theorem 4.7]: in finite dimensions and in the coordinate T , continuity is two-sided and quantitative. They also complement the continuity results of [95] for Rényi divergences of states and channels. They are upper bounds: the two-sided estimate of Theorem 4.8.4 concerns the maps, not the divergences.

The program is continued in later chapters: Chapter 5 treats divergences that are not functions of T through testing regions and hockey-stick divergences, Chapter 7 uses T to prove a geometric Chernoff converse for adaptive channel discrimination, and Chapter 8 proves the results on the sensitivity of converse bounds near noisy gates announced in Chapter 1.

We close with some open questions.

1. *Constants.* Lemma 4.10.5 is sharp as $\delta \rightarrow 0$. Is its factor $(1 - 4\delta^2)^{-1/2}$ optimal for finite δ ? The constants of Corollaries 4.6.11 and 4.6.12 are not optimised. The bound of Theorem 4.4.6 is attained asymptotically (see the remark following it), and the coefficient of the second term of Lemma 4.10.3 is sharp as $\gamma \rightarrow 0$. Can the second term be replaced by $2\gamma\|\varphi\|_\infty$ for all γ ? By the proof of Lemma 4.10.3, this would follow if the Schur multiplier norm of $[\sin(\psi_i + \psi_j)]_{i,j}$, for $\psi_i \in [-\frac{\pi}{4}, \frac{\pi}{4}]$, were $\max_i |\sin 2\psi_i|$. Finally, Example 4.8.11 leaves a factor 2 between the constants of Theorem 4.8.8 and the best examples; what is the exact local Lipschitz constant of $\log \hat{F}$ at noisy gates?
2. *Singular pairs.* Let $c_\Omega = \dim \ker J(\Omega)$. When $1 \leq c_\Omega < \text{rank } T(\mathbf{1}_{C^r} - T)$, what is the exact set of limit points? At mutually singular pairs the fidelity is continuous at the sharp rate $d^{1/2}/\sin \theta$ (Theorem 4.12.11). At what rate is it continuous at pairs

with $\hat{F} = 0$ that are not mutually singular, such as $(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'})$ in Example 4.12.8? Within channels the dichotomy holds at the diagonal pairs (Corollary 4.12.10). Does it hold at every pair of channels with $J(\Phi + \Psi)$ singular and $\hat{F}(\Phi, \Psi) > 0$? By Theorem 4.12.9(a), it would suffice to find H for which the short $[J(\Phi) \# J(\Psi)]_{S_H}$ lowers $\lambda_{\min} \circ \text{Try}$. Is the bound of Theorem 4.12.9 an equality?

3. *Convex f .* For merely convex f the maximal f -divergence is still defined through reverse tests [91], but the perspective formula need not hold. Is it still a function of T , and of which function?
4. *Operational consequences.* The geometric Rényi divergences with $\alpha \in (1, 2]$ give converse bounds for channel capacities [42], and geometric distinguishability measures limit channel discrimination and estimation [72]. Our estimates show how such bounds respond to relative perturbations, and Theorem 4.12.4 shows that continuity fails at pairs with singular sum unless the quantity takes its extreme value there. Can this be made quantitative for a specific task?
5. *Other norms and algebras.* Proposition 4.4.3(1) holds in every unitarily invariant norm $\|\cdot\|$, since $\|AB\| \leq \|A\| \|B\|_\infty$, and, by Ando's inequality [7], Proposition 4.4.3(2) becomes $\|g_\#(E) - g_\#(E')\|_p \leq \|E - E'\|_{p/2}^{1/2}$ for $p \geq 2$, but statements for maps need norms adapted to the Schatten classes. The framework of Okayasu [100, 101] suggests extensions to normal completely positive maps between von Neumann algebras, with T in the commutant of a Stinespring representation; our estimates are already dimension-free.

4.14 Norm bounds and a variational characterisation of means of completely positive maps

Throughout this section $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, $\Omega = \Phi + \Psi$ has a Kraus decomposition $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$ with linearly independent $V_1, \dots, V_r$, $W = [\text{vec}(V_1) \cdots \text{vec}(V_r)]$, and T is the ARN derivative of Φ with respect to Ω , so that $\Psi = \Psi_{\mathbf{1}_{\mathcal{Y}} - T}$. For a linear map $\Xi : \text{L}(\mathcal{X}) \rightarrow \text{L}(\mathcal{Y})$ and $p \in [1, \infty]$ we write

$$\|\Xi\|_{p \rightarrow p} = \sup \{ \|\Xi(X)\|_p : X \in \text{L}(\mathcal{X}), \|X\|_p \leq 1 \},$$

its completely bounded version $\|\Xi\|_{\text{cb}, p \rightarrow p} = \sup_{\mathcal{V}} \|\Xi \otimes \mathbf{1}_{\text{L}(\mathcal{V})}\|_{p \rightarrow p}$, the supremum being over all complex Euclidean spaces $\mathcal{V}$, so that $\|\Xi\|_{\text{cb}, 1 \rightarrow 1} = \|\Xi\|_1$, and we write $\|\cdot\|_{\text{ui}}$ for an arbitrary unitarily invariant norm [16, Chapter IV]. We use the following fact repeatedly: if Ξ is completely positive, then the supremum defining $\|\Xi\|_{p \rightarrow p}$ may be restricted to positive semidefinite X [126].

4.14.1 Norms of means

Proposition 4.14.1. *Let σ be a Kubo–Ando mean. For every $p \in [1, \infty]$,*

$$\|\Phi \sigma \Psi\|_{p \rightarrow p} \leq \|g_\sigma(T)\|_\infty \|\Omega\|_{p \rightarrow p}, \quad \|g_\sigma(T)\|_\infty = \max_{t \in \text{spec}(T)} g_\sigma(t),$$

and for $p = 1$,

$$\|\Phi \sigma \Psi\|_{1 \rightarrow 1} = \|\Phi \sigma \Psi\|_1 = \max_{\rho \in \text{D}(\mathcal{X})} \text{tr} (g_\sigma(T)^\top \Omega^c(\rho)),$$

where Ω^c is the complementary map of Ω , that is, Theorem 4.2.5 applied with Ω in place of Φ .

Proof. By Theorem 4.2.15, $\Phi \sigma \Psi = \Psi_{g_\sigma(T)}$ with $g_\sigma(T) \geq 0$. Put $c = \|g_\sigma(T)\|_\infty$. Since $c\mathbf{1}_{\mathcal{C}^r} - g_\sigma(T) \geq 0$, the map $c\Omega - \Psi_{g_\sigma(T)} = \Psi_{c\mathbf{1}_{\mathcal{C}^r} - g_\sigma(T)}$ is completely positive by Theorem 4.2.2, that is, $0 \leq \Phi \sigma \Psi \leq c\Omega$ in the completely positive order. For positive semidefinite X this gives $0 \leq (\Phi \sigma \Psi)(X) \leq c\Omega(X)$, hence $\|(\Phi \sigma \Psi)(X)\|_p \leq c\|\Omega(X)\|_p$ by the Weyl monotonicity principle [16, Section III.2]. Taking the supremum over positive semidefinite X with $\|X\|_p \leq 1$, which suffices for both completely positive maps [126], gives the first claim. The formula for $\|g_\sigma(T)\|_\infty$ is the spectral theorem, since $g_\sigma \geq 0$. For $p = 1$, the induced and completely bounded trace norms agree on completely positive maps [127, Section 3.3.2], and the formula is the last statement of Theorem 4.2.5, applied to Ω and the positive semidefinite operator $g_\sigma(T)$. $\square$

Remark 4.14.2. If Φ and Ψ are channels, then $\|\Omega\|_{1 \rightarrow 1} = \|\Omega^*(\mathbf{1}_{\mathcal{Y}})\|_\infty = 2$, so $\|\Phi \sigma \Psi\|_1 \leq 2\|g_\sigma(T)\|_\infty$. The bound depends on the spectrum of T only through the values of g_σ on it; since g_σ is not monotone in general (for instance $g_\#(t) = \sqrt{t(1-t)}$), it cannot be replaced by $g_\sigma(\|T\|_\infty)$.

4.14.2 A variational characterisation of the geometric mean

The following extremal property of the geometric mean is due to Ando [6]; see also [17, Chapter 4].

Lemma 4.14.3. *For positive semidefinite $A, B \in \mathcal{L}(\mathcal{V})$,*

$$A \# B = \max \left\{ X = X^* : \begin{pmatrix} A & X \\ X & B \end{pmatrix} \geq 0 \right\},$$

the maximum being taken in the Loewner order. In particular, the block matrix with off-diagonal entries $A \# B$ is positive semidefinite.

Proof. Let $A > 0$. By the Schur complement criterion [17, Chapter 1], the block matrix is positive if and only if $B - XA^{-1}X \geq 0$. With $M = A^{-1/2}BA^{-1/2}$ and $Y = A^{-1/2}XA^{-1/2}$ this reads $Y^2 \leq M$. Since $t \mapsto t^{1/2}$ is operator monotone, $Y \leq |Y| \leq M^{1/2}$, so

$$X = A^{1/2}YA^{1/2} \leq A^{1/2}M^{1/2}A^{1/2} = A \# B;$$

and $Y = M^{1/2}$ is admissible, so $A \# B$ is the maximum. For general $A, B \geq 0$, adding $\eta\mathbf{1}_{\mathcal{V}}$ to both diagonal blocks preserves positivity, so every admissible X satisfies $X \leq (A + \eta\mathbf{1}_{\mathcal{V}}) \# (B + \eta\mathbf{1}_{\mathcal{V}})$, which decreases to $A \# B$ as $\eta \downarrow 0$ [79]. Conversely, the block matrices with off-diagonal entry $(A + \eta\mathbf{1}_{\mathcal{V}}) \# (B + \eta\mathbf{1}_{\mathcal{V}})$ are positive, and letting $\eta \downarrow 0$ in the closed positive cone shows that $A \# B$ is admissible. $\square$

For an adjoint-preserving linear map $\Xi : \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{Y})$, define

$$\mathcal{B}_\Xi : \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathbb{C}^2 \otimes \mathcal{Y}), \quad \mathcal{B}_\Xi(X) = \begin{pmatrix} \Phi(X) & \Xi(X) \\ \Xi(X) & \Psi(X) \end{pmatrix}.$$

This block-positivity characterisation of the geometric mean of completely positive maps is also obtained by Okayasu [100], for normal completely positive maps between von Neumann algebras, and by Kian [76]; we include a short proof in the present notation.

Proposition 4.14.4. *The map $\Phi \# \Psi$ is the largest adjoint-preserving map Ξ , in the completely positive order, for which $\mathcal{B}_\Xi$ is completely positive. Equivalently,*

$$J(\Phi) \# J(\Psi) = \max \left\{ X = X^* : \begin{pmatrix} J(\Phi) & X \\ X & J(\Psi) \end{pmatrix} \geq 0 \right\}.$$

In terms of the ARN derivative, the corresponding maximisation over $L(\mathbb{C}^r)$ is solved by

$$\max \left\{ X = X^* : \begin{pmatrix} T & X \\ X & \mathbf{1}_{\mathbb{C}^r} - T \end{pmatrix} \geq 0 \right\} = (T(\mathbf{1}_{\mathbb{C}^r} - T))^{1/2} = g_{\#}(T).$$

Proof. Ordering the tensor factors as $\mathbb{C}^2 \otimes (\mathcal{Y} \otimes \mathcal{X})$, the Choi matrix of $\mathcal{B}_{\Xi}$ is

$$J(\mathcal{B}_{\Xi}) = \begin{pmatrix} J(\Phi) & J(\Xi) \\ J(\Xi) & J(\Psi) \end{pmatrix},$$

and $J(\Xi)$ is Hermitian because Ξ preserves adjoints. By Choi's theorem [28], $\mathcal{B}_{\Xi}$ is completely positive if and only if this block matrix is positive semidefinite, and $\Xi_1 \leq \Xi_2$ in the completely positive order if and only if $J(\Xi_1) \leq J(\Xi_2)$. The first two statements therefore follow from Lemma 4.14.3 with $A = J(\Phi)$ and $B = J(\Psi)$, since $J(\Phi \# \Psi) = J(\Phi) \# J(\Psi)$. For the last statement apply Lemma 4.14.3 with $A = T$ and $B = \mathbf{1}_{\mathbb{C}^r} - T$; since these commute,

$$T \# (\mathbf{1}_{\mathbb{C}^r} - T) = (T(\mathbf{1}_{\mathbb{C}^r} - T))^{1/2} = g_{\#}(T)$$

by the functional calculus, consistently with $J(\Phi) \# J(\Psi) = W g_{\#}(T) W^*$ from Theorem 4.2.15. $\square$

Remark 4.14.5. Since $X \mapsto \lambda_{\min}(\text{Tr}_{\mathcal{Y}} X)$ is monotone in the Loewner order, Propositions 4.3.1 and 4.14.4 give

$$\widehat{F}(\Phi, \Psi) = \max \left\{ \lambda_{\min}(\text{Tr}_{\mathcal{Y}} X) : X = X^*, \begin{pmatrix} J(\Phi) & X \\ X & J(\Psi) \end{pmatrix} \geq 0 \right\},$$

which is the semidefinite program in [98, eq. (20)] for the square root of their geometric channel fidelity (cf. [72] for the analogous program for the channel fidelity based on the Uhlmann fidelity); at the level of states it reduces to the semidefinite program for the Matsumoto fidelity [29, 90].

4.14.3 Cauchy–Schwarz bounds

Lemma 4.14.6. *Let $A, B \geq 0$ in $L(\mathcal{V})$ and let X be such that*

$$\begin{pmatrix} A & X \\ X^* & B \end{pmatrix} \geq 0.$$

Then

1. $X = A^{1/2} C B^{1/2}$ for some C with $\|C\|_{\infty} \leq 1$;
2. $\|X\|_{\text{ui}}^2 \leq \|A\|_{\text{ui}} \|B\|_{\text{ui}}$ for every unitarily invariant norm;
3. in particular $\|A \# B\|_{\text{ui}}^2 \leq \|A\|_{\text{ui}} \|B\|_{\text{ui}}$.

Proof. Part (1) is standard [17, Chapter 1]: for $A, B > 0$, positivity of the block matrix is equivalent to $B - X^* A^{-1} X \geq 0$, that is, to $\|A^{-1/2} X B^{-1/2}\|_{\infty} \leq 1$; the general case follows by replacing A, B with $A + \eta \mathbf{1}_{\mathcal{V}}, B + \eta \mathbf{1}_{\mathcal{V}}$ and passing to a limit point of the contractions C_{η} . Part (2) is due to Horn and Mathias [64]: writing $X = Y^* Z$ with $Y = A^{1/2}$ and $Z = C B^{1/2}$, the Cauchy–Schwarz inequality $\|Y^* Z\|_{\text{ui}}^2 \leq \|Y^* Y\|_{\text{ui}} \|Z^* Z\|_{\text{ui}}$ for unitarily invariant norms [18, 64] gives $\|X\|_{\text{ui}}^2 \leq \|A\|_{\text{ui}} \|B^{1/2} C^* C B^{1/2}\|_{\text{ui}}$, and $0 \leq B^{1/2} C^* C B^{1/2} \leq B$ gives $\|B^{1/2} C^* C B^{1/2}\|_{\text{ui}} \leq \|B\|_{\text{ui}}$ by the Weyl monotonicity principle [16, Section III.2]. Part (3) follows from part (2) and Lemma 4.14.3. $\square$

Theorem 4.14.7. *For all $\Phi, \Psi \in \text{CP}(\mathcal{X}, \mathcal{Y})$, every unitarily invariant norm and every $p \in [1, \infty]$,*

$$\|J(\Phi \# \Psi)\|_{\text{ui}}^2 \leq \|J(\Phi)\|_{\text{ui}} \|J(\Psi)\|_{\text{ui}}, \quad \|\Phi \# \Psi\|_{p \rightarrow p}^2 \leq \|\Phi\|_{p \rightarrow p} \|\Psi\|_{p \rightarrow p}.$$

The second inequality also holds for the norms $\|\cdot\|_{\text{cb}, p \rightarrow p}$, and in particular $\|\Phi \# \Psi\|_1^2 \leq \|\Phi\|_1 \|\Psi\|_1$.

Proof. The first inequality is Lemma 4.14.6(3) with $A = J(\Phi)$ and $B = J(\Psi)$. For the second, $\mathcal{B}_{\Phi \# \Psi}$ is completely positive by Proposition 4.14.4, so for positive semidefinite X the block matrix $\mathcal{B}_{\Phi \# \Psi}(X)$ is positive semidefinite, and Lemma 4.14.6(2) with the Schatten p -norm gives

$$\|(\Phi \# \Psi)(X)\|_p^2 \leq \|\Phi(X)\|_p \|\Psi(X)\|_p \leq \|\Phi\|_{p \rightarrow p} \|\Psi\|_{p \rightarrow p} \|X\|_p^2.$$

Since $\Phi \# \Psi$ is completely positive, it suffices to take the supremum over positive semidefinite X [126]. For the norms $\|\cdot\|_{\text{cb}, p \rightarrow p}$, apply the same argument, for an arbitrary complex Euclidean space $\mathcal{V}$, to $\Phi \otimes \mathbf{1}_{L(\mathcal{V})}$ and $\Psi \otimes \mathbf{1}_{L(\mathcal{V})}$, using

$$(\Phi \otimes \mathbf{1}_{L(\mathcal{V})}) \# (\Psi \otimes \mathbf{1}_{L(\mathcal{V})}) = (\Phi \# \Psi) \otimes \mathbf{1}_{L(\mathcal{V})},$$

which holds because $(A \otimes C) \# (B \otimes C) = (A \# B) \otimes C$ for $A, B, C \geq 0$ [79]. □

Chapter 5

Shadows of quantum divergences: testing regions, modular lifts and continuity

Summary. Divergences built from Kubo–Ando means are functionals of one positive contraction, the Arveson–Radon–Nikodym derivative, which also governs their continuity (Chapter 4). We study those that are not. To a pair of states we attach three classical shadows: the reverse test, the testing region encoded by the hockey-stick profile, and the Nussbaum–Szkola pair of the relative modular operator. The Umegaki relative entropy, the Hirche–Tomamichel f -divergences and the hypothesis-testing quantities are functionals of the testing region, whose shadow consists of atoms at the relative spectrum, dominated by those of the reverse test, and an absolutely continuous part given by the coupling of the second state across the Helstrom test. Consequently the measured, testing-region and maximal divergences of a strictly convex function coincide only for commuting pairs. Testing-region divergences with finite boundary values are uniformly continuous, with a dimension-free modulus, and the Umegaki relative entropy is continuous exactly when the second state is invertible or the entropy is infinite. The Petz divergences arise from the same construction applied to the relative modular operator and have the same boundary masses. Replacing the trace norm by the Thompson metric and the relative entropy by a conditional entropy, we show that the Stein exponent of adaptive channel discrimination is continuous exactly when the Choi matrix of the second channel is invertible or the exponent is infinite. We prove similar results for the strong-converse exponent and the regularised Petz divergences, and give a one-shot dual formula for testing-region channel divergences. Tensor products are treated in Chapter 6.

5.1 Introduction

The geometric fidelity, the geometric Rényi divergences and the maximal f -divergences of two channels are all of the form $\lambda_{\max}(\operatorname{Tr}_Y Wh(T)W^*)$, where T is the Arveson–Radon–Nikodym (ARN) derivative of Φ with respect to $\Phi + \Psi$ (Theorems 4.2.15 and 4.7.5 in Chapter 4). One positive contraction and one scalar function carry the whole theory: finiteness is a statement about the boundary values of h , continuity is the functional calculus of T , and the continuity points are characterised by a dichotomy (Theorem 4.12.4). That program stops at the Kubo–Ando means. The Umegaki relative entropy is not a function of T , nor are the trace distance, the hypothesis-testing relative entropy, the Petz and sandwiched Rényi divergences, or the regularised channel divergences that govern adaptive discrimination.

The guiding observation of this chapter is that T is one of several *classical shadows* of a quantum pair, each of which plays the role of T for a different family of divergences. For states ρ, σ with $\omega = \rho + \sigma$:

- (I) the *reverse test* (p_T, q_T) , the classical pair carried by the spectral measure of $T = \omega^{-1/2} \rho \omega^{-1/2}$ in the state ω . Its divergences are the maximal f -divergences and the Kubo–Ando means [91] (Chapter 4);
- (II) the *testing region*, encoded by the hockey-stick profile $e(\gamma) = \text{Tr}(\rho - \gamma\sigma)_+$ and its second derivative $\mu = e''$. Its divergences are the hypothesis-testing quantities, the Hirche–Tomamichel f -divergences [63] and, by Frenkel’s formula [46], the Umegaki relative entropy;
- (III) the *modular shadow*, the Nussbaum–Szkoła pair of the relative modular operator [99]. Its divergences are the Petz f -divergences.

A fourth family, the measured divergences, is a supremum over many classical pairs, and a fifth, the sandwiched and regularised divergences, appears only after tensor powers. We carry the program of Chapter 4 through families (II)–(V): a representation by a shadow, a finiteness criterion through boundary values, continuity moduli, the continuity points, and, where possible, channels. Three tools replace the functional calculus of T : the T-lift (Section 5.3), the modular lift (Section 5.5), and, where tensor products break the shadows, pinching, the Thompson metric and a conditional-entropy decomposition (Sections 5.6 and 5.7).

Main results.

- *Structure of the testing region* (Theorem 5.3.1). The shadow μ has atoms exactly at the relative spectrum, of weight $\text{Tr} \Pi_{\ker(\rho - \gamma\sigma)} \sigma$, and an absolutely continuous part with density

$$2 \sum_{\lambda > 0 > \lambda'} \frac{\text{Tr}(Q_\lambda \sigma Q_{\lambda'} \sigma)}{\lambda - \lambda'},$$

the coupling of σ across the Helstrom test. The atoms are dominated by those of the reverse test (Theorem 5.3.2), the hockey-sticks of the reverse test dominate those of the pair, and

$$D_{BS}(\rho \| \sigma) - D(\rho \| \sigma) = \int_1^\infty \left(\frac{E_\gamma^T - E_\gamma}{\gamma} (\rho \| \sigma) + \frac{E_\gamma^T - E_\gamma}{\gamma^2} (\sigma \| \rho) \right) d\gamma,$$

with a pointwise nonnegative integrand that vanishes identically exactly for commuting pairs. The measured, testing-region and maximal divergences of a strictly convex f coincide only for commuting pairs (Corollary 5.3.3).

- *Continuity without a dichotomy* (Theorems 5.4.1 and 5.4.3). Every testing-region f -divergence with $f(0^+), f'(\infty) < \infty$ satisfies

$$|D_f(\rho \| \sigma) - D_f(\rho' \| \sigma')| \leq \int_0^\infty \min\{\gamma, 1, \varepsilon_1 + \gamma\varepsilon_2\} f''(d\gamma)$$

in every dimension. The perturbations of size $O(\sqrt{t})$ that make the geometric fidelity jump from 1 to 0 move these divergences by $O(\sqrt{t})$. The same modulus holds for the measured f -divergences. The Umegaki relative entropy is continuous at (ρ, σ) if and only if σ is invertible or $D(\rho \| \sigma) = \infty$.

- *Tensor products* (Section 5.6). Explicit single-shot sandwiches

$$D_\alpha^P \leq D_\alpha^{\text{HT}} \leq D_\alpha^P + \frac{\log 2}{1-\alpha} \quad \text{for } \alpha < 1 \quad \text{and} \\ |D_\alpha^{\text{HT}} - \tilde{D}_\alpha| \leq \frac{\alpha}{\alpha-1} \log v(\sigma) \quad \text{for } \alpha > 1,$$

which give the regularisation theorem of [63] with rates. Their behaviour under tensor products beyond these sandwiches, including the question of super-multiplicativity, is studied in Chapter 6.

- *The adaptive Stein exponent* (Theorems 5.7.2, 5.7.4 and Corollary 5.7.5). For every divergence that is antitone and logarithmically homogeneous in the second argument, every n -use adaptive quantity is $n d_T$ -Lipschitz in Ψ , where d_T is the Thompson metric of the Choi matrices. For the Umegaki entropy we also control the first argument uniformly over all references, which gives

$$|\zeta(\Phi, \Psi) - \zeta(\Phi', \Psi')| \leq 2\varepsilon \log d_Y + (1 + \varepsilon)h\left(\frac{\varepsilon}{1+\varepsilon}\right) + \varepsilon \log \kappa(\Psi) + d_T(J(\Psi), J(\Psi'))$$

for the Stein exponent ζ of adaptive discrimination, $\varepsilon = \frac{1}{2}\|\Phi - \Phi'\|_\diamond$ and $\kappa(\Psi)$ the condition number of $J(\Psi)$. The exponent is continuous at (Φ, Ψ) if and only if $J(\Psi)$ is invertible or $\zeta(\Phi, \Psi) = \infty$.

- *Petz divergences* (Section 5.5). Through the modular derivative $\tilde{T} = \Delta(1 + \Delta)^{-1}$ every Petz f -divergence is a functional of a commuting pair of the kind treated in Chapter 4. Its shadow has the boundary masses of the testing region, not those of the reverse test. Bounded Petz divergences are continuous everywhere,

$$|Q_\alpha^P(\rho\|\sigma) - Q_\alpha^P(\rho'\|\sigma')| \leq (2\varepsilon_1)^\alpha + (2\varepsilon_2)^{1-\alpha}$$

in every dimension, and the testing-region and Petz Rényi divergences are ordered for every $\alpha \neq 1$.

- *Sandwiched and Petz exponents, and the one-shot formula* (Theorems 5.7.8, 5.7.11, 5.7.13 and 5.7.16, Corollary 5.7.9). For $\alpha > 1$ and invertible $J(\Psi)$,

$$|\tilde{D}_\alpha^{\text{reg}}(\Phi'\|\Psi) - \tilde{D}_\alpha^{\text{reg}}(\Phi\|\Psi)| \leq \frac{\alpha}{\alpha-1} d_X d_Y (m/\lambda_{\min}(J(\Psi)))^{\alpha-1} \|\Phi' - \Phi\|_\diamond,$$

so the strong-converse exponent of adaptive discrimination is continuous there. For the regularised Petz divergences, which lack a chain rule, Theorem 5.7.11 gives Thompson-Lipschitz dependence on Φ , stability under local perturbations uniformly in the number of uses, and continuity at isometries and at invertible $J(\Phi)$. Testing-region channel divergences have the exact one-shot formula of Theorem 5.7.13,

$$D_f^{\text{HT}}(\Phi\|\Psi) = \inf_{Y, Z} \lambda_{\max} \left(\text{Tr}_Y \left[\int Y_\gamma f''(d\gamma) + \int Z_\gamma \tilde{f}''(d\gamma) \right] \right),$$

the infimum over families with $Y_\gamma \geq 0$, $Y_\gamma \geq J(\Phi) - \gamma J(\Psi)$, $Z_\gamma \geq 0$ and $Z_\gamma \geq J(\Psi) - \gamma J(\Phi)$, in which the reverse test of Chapter 4 is one feasible family. An invertible input is optimal exactly when it solves $\Gamma_f(\theta) = D\theta$. Their regularisations are the Petz ($\alpha < 1$), Umegaki ($\alpha = 1$) and sandwiched ($\alpha > 1$) channel divergences.

Related work. The divergences of family (II) were introduced by Hirche and Tomamichel [63] and are also called hockey-stick f -divergences. Beigi, Hirche and Tomamichel [13] and Liu, Hirche and Cheng [84] compared their Rényi members with the Petz and sandwiched families. Hiai, Mosonyi and Tomamichel [58] have recently developed them in von Neumann algebras: a Neyman–Pearson representation, the detection of reversibility, the regularisation of the Rényi members to the Petz divergences below order one and to the sandwiched divergences above it, and partial results on when different quantum f -divergences coincide. Our setting is finite-dimensional, and our results are of a different kind: the structure of the shadow μ , explicit regularisation rates and continuity moduli, the rigidity statement of Corollary 5.3.3, and the channel results; super-multiplicativity is treated in Chapter 6. Continuity bounds for entropies through hockey-stick integral representations were developed by Berta, Lami and Tomamichel [15], and Theorem 5.4.1 is in the spirit of that method.

The analogy with Chapter 4 is close but the answers differ in an instructive way. For Kubo–Ando means the obstruction to continuity is the invertibility of the *sum* $J(\Phi + \Psi)$, because the reverse test can be destroyed by an $O(\sqrt{t})$ coupling of its kernel to its range. For the testing region that mechanism is invisible, and the only obstruction is a support leak, which needs the invertibility of the *second argument*.

The quantitative claims were also checked numerically; the checks are summarised in Section 5.9.

The chapter is organised as follows. Section 5.2 introduces the three shadows and the testing-region f -divergences. Section 5.3 determines the structure of the testing-region shadow and compares it with the reverse test, and Section 5.4 proves the continuity results on states. Section 5.5 treats the modular shadow and the Petz divergences, and Section 5.6 compares the testing region with the two multiplicative shadows on tensor powers. Section 5.7 contains the results on channels: Thompson stability, the adaptive Stein exponent, the sandwiched and regularised Petz divergences, and the one-shot formula. Section 5.8 summarises the classification, Section 5.9 lists the numerical checks, and Section 5.10 collects open problems.

5.2 Three shadows of a pair of states

Throughout, ρ, σ are density operators on a finite-dimensional Hilbert space $\mathcal{H}$, $\omega = \rho + \sigma$, and Π_X denotes the support projection of a positive operator X , with $\Pi_X^\perp = \mathbf{1} - \Pi_X$. All logarithms are natural.

Definition 5.2.1. For $A, B \geq 0$ and $\gamma \geq 0$ the *hockey-stick divergence* is

$$E_\gamma(A\|B) = \text{Tr}(A - \gamma B)_+ = \max_{0 \leq M \leq \mathbf{1}} \text{Tr } M(A - \gamma B).$$

The *profile* of (ρ, σ) is $e(\gamma) = E_\gamma(\rho\|\sigma)$, the *error profile* is

$$\eta(\gamma) = 1 - e(\gamma) = \text{Tr } \rho - \text{Tr}(\rho - \gamma\sigma)_+,$$

the minimal Bayesian error with weights $(1, \gamma)$, and the *boundary masses* are $m_0 = \text{Tr } \Pi_\rho^\perp \sigma$ and $m_\infty = \text{Tr } \Pi_\sigma^\perp \rho$.

Lemma 5.2.2. (a) e is convex and nonincreasing on $[0, \infty)$ with $e'(\gamma) \rightarrow 0$ as $\gamma \rightarrow \infty$, $e(0) = 1$ and $(1 - \gamma)_+ \leq e(\gamma) \leq 1$. Equivalently η is concave and nondecreasing, $\eta(0) = 0$, and $\gamma \mapsto \eta(\gamma)/\gamma$ is nonincreasing.

(b) For $A, B \geq 0$, $\lim_{L \rightarrow \infty} \text{Tr}(A - LB)_+ = \text{Tr} \Pi_B^\perp A$. Hence $e(\infty) = m_\infty$ and $e'(0^+) = -(1 - m_0)$.

(c) $E_\gamma(\sigma \parallel \rho) = \gamma e(1/\gamma) - \gamma + 1$ for $\gamma > 0$.

Proof. (a) e is a supremum of affine functions of γ with slopes $-\text{Tr} M\sigma \leq 0$, and a convex nonincreasing function that is bounded below has $e'(\infty) = 0$; the bounds follow from $M = \mathbf{1}$, $M = 0$ and $\text{Tr} M(\rho - \gamma\sigma) \leq \text{Tr} M\rho \leq 1$. (b) The test $M = \Pi_B^\perp$ gives $\geq$. For $\delta > 0$,

$$A \leq (1 + \delta)\Pi_B^\perp A \Pi_B^\perp + (1 + \delta^{-1})\Pi_B A \Pi_B,$$

and $B \geq b\Pi_B$ with b the smallest nonzero eigenvalue of B . For $L \geq (1 + \delta^{-1})\|A\|/b$ we get $A - LB \leq (1 + \delta)\Pi_B^\perp A \Pi_B^\perp$, so $\text{Tr}(A - LB)_+ \leq (1 + \delta)\text{Tr} \Pi_B^\perp A$. For the slope at 0 write, using $(X)_+ - (-X)_+ = X$,

$$\eta(\gamma) = \gamma - \text{Tr}(\gamma\sigma - \rho)_+ = \gamma(1 - E_{1/\gamma}(\sigma \parallel \rho)),$$

and apply the first statement to (σ, ρ) . (c) is the same identity. $\square$

Definition 5.2.3 ([63]). Let $f : (0, \infty) \rightarrow \mathbb{R}$ be convex with $f(1) = 0$, and let f'' denote its second distributional derivative, a positive Radon measure. The *testing-region* (Hirche–Tomamichel) f -divergence is

$$D_f^{\text{HT}}(\rho \parallel \sigma) = \int_1^\infty \left(E_\gamma(\rho \parallel \sigma) f''(d\gamma) + E_\gamma(\sigma \parallel \rho) \tilde{f}''(d\gamma) \right), \quad \tilde{f}(x) = xf(1/x).$$

For $\alpha \in (0, 1) \cup (1, \infty)$ and $f_\alpha(x) = \frac{x^\alpha - 1}{\alpha - 1}$ we write $Q_\alpha^{\text{HT}} = 1 + (\alpha - 1)D_{f_\alpha}^{\text{HT}}$ and $D_\alpha^{\text{HT}} = \frac{1}{\alpha - 1} \log Q_\alpha^{\text{HT}}$.

For smooth f , $\tilde{f}''(d\gamma) = \gamma^{-3} f''(1/\gamma) d\gamma$, which is the form used in [63]. Since D_f^{HT} only depends on f'' we may and do assume $f'(1) = 0$ whenever convenient, so that $f \geq 0$. By [46], $D_{x \log x}^{\text{HT}}$ is the Umegaki relative entropy $D(\rho \parallel \sigma) = \text{Tr} \rho(\log \rho - \log \sigma)$.

Definition 5.2.4. The *testing-region shadow* of (ρ, σ) is the positive measure $\mu = e''$ on $(0, \infty)$, completed by the atoms m_0 at 0 and m_∞ at ∞ .

The profile is the Legendre dual of the quantum Neyman–Pearson curve, so it carries the same information as the testing region and the quantum relative Lorenz curve [26, 113]. For commuting $\rho = \sum_x p_x |x\rangle\langle x|$, $\sigma = \sum_x q_x |x\rangle\langle x|$ one has $e(\gamma) = \sum_x q_x (p_x/q_x - \gamma)_+$, so μ is the law of the likelihood ratio p/q under q on $\{q > 0\}$, and $m_\infty = p(\{q = 0\})$.

Proposition 5.2.5 (Representation). *For every convex f with $f(1) = 0$,*

$$D_f^{\text{HT}}(\rho \parallel \sigma) = \int_0^\infty (e(\gamma) - (1 - \gamma)_+) f''(d\gamma) = \int_{(0, \infty)} f d\mu + f(0^+) m_0 + f'(\infty) m_\infty,$$

with $f'(\infty) = \lim_{x \rightarrow \infty} f(x)/x$ and the convention $0 \cdot \infty = 0$. Moreover $\mu((0, \infty)) = 1 - m_0$ and $\int \gamma d\mu = 1 - m_\infty$.

Proof. Write $g(\gamma) = e(\gamma) - (1 - \gamma)_+ \geq 0$. Lemma 5.2.2(c) and the substitution $\gamma \mapsto 1/\gamma$ turn the second term of Definition 5.2.3 into $\int_0^1 g f''$, which gives the first formula. For the second, both sides are linear in f'' with nonnegative integrands, so by Tonelli it suffices to treat the extreme kernels $f = k_s$, $k_s(x) = (x - s)_+$ for $s \geq 1$ and $k_s(x) = (s - x)_+$ for $s < 1$,

which have $f(1) = f'(1) = 0$ and $f'' = \delta_s$. Since e is convex with $e'(\infty) = 0$, $e(\infty) = m_\infty$ and $e'(0^+) = -(1 - m_0)$ (Lemma 5.2.2),

$$e(s) = m_\infty + \int_{(s,\infty)} (y - s) d\mu(y), \quad e(s) = 1 - s(1 - m_0) + \int_{(0,s]} (s - y) d\mu(y).$$

For $s \geq 1$ the first identity is $g(s) = \int k_s d\mu + k'_s(\infty)m_\infty$, and for $s < 1$ the second is $g(s) = \int k_s d\mu + k_s(0)m_0$. A general f with $f(1) = f'(1) = 0$ is $\int k_s f''(ds)$ by Taylor's formula, and $f(0^+)$, $f'(\infty)$ are the corresponding integrals, by monotone convergence. The masses follow from the first identity: its derivative is $e'(s) = -\mu((s, \infty))$, so $\mu((0, \infty)) = -e'(0^+) = 1 - m_0$, and at $s \rightarrow 0$ it gives $1 = m_\infty + \int y d\mu$. Adding $c(x - 1)$ to f changes the right-hand side by $c(\int (\gamma - 1) d\mu - m_0 + m_\infty) = 0$, so the formula holds without the normalisation $f'(1) = 0$ whenever no $\infty - \infty$ occurs. $\square$

Proposition 5.2.6 (Mellin formulas). *For $\alpha \in (0, 1)$,*

$$Q_\alpha^{\text{HT}}(\rho \parallel \sigma) = \int_{(0,\infty)} \gamma^\alpha d\mu(\gamma) = \alpha(1 - \alpha) \int_0^\infty \gamma^{\alpha-2} \eta(\gamma) d\gamma.$$

For $\alpha > 1$, $Q_\alpha^{\text{HT}} = +\infty$ if $m_\infty > 0$, and otherwise

$$Q_\alpha^{\text{HT}}(\rho \parallel \sigma) = \int_{(0,\infty)} \gamma^\alpha d\mu(\gamma) = \alpha(\alpha - 1) \int_0^\infty \gamma^{\alpha-2} e(\gamma) d\gamma.$$

Proof. Apply Proposition 5.2.5 to f_α , for which $f_\alpha(0^+) = \frac{-1}{\alpha-1}$, $f'_\alpha(\infty) = 0$ for $\alpha < 1$ and $= \infty$ for $\alpha > 1$, and $f''_\alpha(d\gamma) = \alpha\gamma^{\alpha-2}d\gamma$. With $\mu((0, \infty)) = 1 - m_0$ this gives $Q_\alpha^{\text{HT}} = \int \gamma^\alpha d\mu$ (plus $\infty \cdot m_\infty$ when $\alpha > 1$). For the integrals use the first formula of Proposition 5.2.5: $g = \min\{\gamma, 1\} - \eta$, and

$$\alpha(1 - \alpha) \int_0^\infty \gamma^{\alpha-2} \min\{\gamma, 1\} d\gamma = 1$$

for $\alpha < 1$, while for $\alpha > 1$ we use

$$\alpha(\alpha - 1) \int_0^1 \gamma^{\alpha-2} (1 - \gamma) d\gamma = 1.$$

$\square$

The reverse test and the modular shadow. Let $T = \omega^{-1/2} \rho \omega^{-1/2}$ on $\Pi_\omega \mathcal{H}$ with spectral projections E_t . The *reverse test* is the classical pair $p_T(t) = t \text{Tr } E_t \omega$, $q_T(t) = (1 - t) \text{Tr } E_t \omega$ on $\text{spec } T$. The preparation $t \mapsto \omega^{1/2} E_t \omega^{1/2} / \text{Tr } E_t \omega$ maps (p_T, q_T) to (ρ, σ) , and $D_f^T := D_f(p_T \parallel q_T)$ is the maximal f -divergence of (ρ, σ) for operator convex f [91], equal to $\text{Tr } \omega^{1/2} h_f(T) \omega^{1/2}$ with $h_f(t) = (1 - t)f(t/(1 - t))$ (Theorem 4.2.15 and Lemma 4.7.4 in Chapter 4). Its profile is

$$E_\gamma^T(\rho \parallel \sigma) := E_\gamma(p_T \parallel q_T) = \text{Tr } \omega^{1/2} (T - \gamma(\mathbf{1} - T))_+ \omega^{1/2}.$$

The *modular shadow* is the Nussbaum–Szkola pair $P_{ij} = a_i |\langle x_i | y_j \rangle|^2$, $Q_{ij} = b_j |\langle x_i | y_j \rangle|^2$ built from $\rho = \sum a_i |x_i\rangle \langle x_i|$ and $\sigma = \sum b_j |y_j\rangle \langle y_j|$; its f -divergences are the Petz f -divergences D_f^P [61, 99]. The measured divergence D_f^M is the supremum of D_f over the outcome distributions of all measurements.

Proposition 5.2.7 (Blackwell chain). *For every convex f with $f(1) = 0$: $D_f^M \leq D_f^{\text{HT}} \leq D_f^T$.*

Proof. Both terms of Definition 5.2.3 are nondecreasing in the hockey-stick divergences, and hockey-stick divergences satisfy data processing. Apply this to a measurement and to the preparation channel of the reverse test, and use that D_f^{HT} is the classical f -divergence on commuting pairs. $\square$

5.3 Structure of the testing-region shadow

Both ρ and σ vanish on $\ker \omega$, so we work on $\Pi_\omega \mathcal{H}$. Let

$$\Gamma = \{\gamma > 0 : \rho - \gamma\sigma \text{ is singular on } \Pi_\omega \mathcal{H}\},$$

the *relative spectrum*; it is the set of $t/(1-t)$ with $t \in \text{spec } T \cap (0, 1)$, since $Tu = tu$ with $v = \omega^{-1/2}u$ is $(1-t)\rho v = t\sigma v$. For $\gamma \notin \Gamma$ write $\rho - \gamma\sigma = \sum_\lambda \lambda Q_\lambda$ on $\Pi_\omega \mathcal{H}$ and $P_\gamma = \sum_{\lambda>0} Q_\lambda$, the Helstrom projection.

Theorem 5.3.1 (Structure). (a) On each interval of $(0, \infty) \setminus \Gamma$, e is real-analytic and $e'(\gamma) = -\text{Tr } P_\gamma \sigma$.

(b) On $(0, \infty) \setminus \Gamma$ the shadow has the density

$$\frac{d\mu}{d\gamma} = 2 \sum_{\lambda>0>\lambda'} \frac{\text{Tr}(Q_\lambda \sigma Q_{\lambda'} \sigma)}{\lambda - \lambda'} \geq 0.$$

(c) At $\gamma_0 \in \Gamma$ the shadow has the atom $\mu(\{\gamma_0\}) = \text{Tr } \Pi_V \sigma > 0$, where $V = \ker(\rho - \gamma_0 \sigma) \cap \Pi_\omega \mathcal{H}$.

(d) μ has no other part, and its absolutely continuous part vanishes identically if and only if $[\rho, \sigma] = 0$.

Proof. (a) For $\gamma \notin \Gamma$, $0 \notin \text{spec}(\rho - \gamma\sigma)$, so $e(\gamma) = \text{Tr } \phi(\rho - \gamma\sigma)$ with $\phi(x) = x_+$ smooth near the spectrum, and

$$\frac{d}{d\gamma} \text{Tr } \phi(H) = \text{Tr } \phi'(H) \dot{H} = -\text{Tr } P_\gamma \sigma.$$

(b) By the Daleckiĭ–Krein formula,

$$\dot{P}_\gamma = \sum_{\lambda, \lambda'} \frac{1_{\lambda>0} - 1_{\lambda'>0}}{\lambda - \lambda'} Q_\lambda (-\sigma) Q_{\lambda'},$$

and $-\text{Tr } \dot{P}_\gamma \sigma$ is the stated sum; only pairs with $\lambda > 0 > \lambda'$ survive. (c) By Rellich's theorem the eigenvalues of the analytic family $\rho - \gamma\sigma$ are analytic branches. The branches through 0 at γ_0 have derivatives equal to the eigenvalues of $-\Pi_V \sigma \Pi_V|_V$, which are negative because $\sigma v = 0$ and $\rho v = \gamma_0 \sigma v$ would force $v \in \ker \omega$. So exactly $\dim V$ eigenvalues cross from positive to negative, $P_{\gamma_0-\delta} - P_{\gamma_0+\delta} \rightarrow \Pi_V$, and e' jumps by $\text{Tr } \Pi_V \sigma$. (d) Γ is finite, so (a)–(c) describe μ on $(0, \infty)$. If ρ, σ commute, σ is block diagonal for every $\rho - \gamma\sigma$ and the density vanishes. Conversely, if the density vanishes then $Q_\lambda \sigma Q_{\lambda'} = 0$ for $\lambda > 0 > \lambda'$, so $\dot{P}_\gamma = 0$ and $[\sigma, P_\gamma] = 0$; then $P_\gamma = P_I$ is constant on each interval I of $(0, \infty) \setminus \Gamma$ and commutes with σ and with $\rho = (\rho - \gamma\sigma) + \gamma\sigma$, hence the P_I commute with each other. On a minimal common invariant subspace B of all P_I , the operator $\rho - \gamma\sigma$ is positive for small γ and negative for large γ or has constant sign; as it is nonincreasing in γ , $\rho|_B = \gamma_B \sigma|_B$ for some $\gamma_B \in [0, \infty]$. So ρ and σ commute. $\square$

The density in (b) is the coupling of σ between the positive and negative parts of the optimal test at level γ . This is the information that the reverse test discards, as the next theorem makes precise. Write μ_T for the shadow of the reverse test: the law of $t/(1-t)$ under q_T , an atomic measure at $\Gamma \cup \{0\}$ with $\mu_T(\{\gamma_0\}) = (1-t_0) \text{Tr } E_{t_0} \omega$, $t_0 = \gamma_0/(1+\gamma_0)$, together with the atom $m_\infty^T = \text{Tr } E_1 \omega$ at ∞ .

Theorem 5.3.2 (The T-lift). (a) $E_\gamma(\rho\|\sigma) \leq E_\gamma^T(\rho\|\sigma)$ for every $\gamma \geq 0$, with equality if and only if $P_{>c}^T \omega P_{<c}^T = 0$, where $c = \gamma/(1 + \gamma)$ and P^T are the spectral projections of T . The same holds with ρ and σ exchanged and $c = 1/(1 + \gamma)$. Equality for all γ holds if and only if $[\rho, \sigma] = 0$.

(b) The atoms are dominated: $\mu(\{\gamma_0\}) \leq \mu_T(\{\gamma_0\})$ for $\gamma_0 \in \Gamma$, $m_0 \leq \mu_T(\{0\})$ and $m_\infty \leq m_\infty^T$, with equality at γ_0 if and only if V is invariant under ω .

(c) The absolutely continuous mass of μ equals

$$\sum_{\gamma_0 \in \Gamma \cup \{0\}} (\mu_T(\{\gamma_0\}) - \mu(\{\gamma_0\})).$$

(d) For every convex f with $f(1) = 0$,

$$D_f^T(\rho\|\sigma) - D_f^{\text{HT}}(\rho\|\sigma) = \int_1^\infty \left((E_\gamma^T - E_\gamma)(\rho\|\sigma) f''(d\gamma) + (E_\gamma^T - E_\gamma)(\sigma\|\rho) \tilde{f}''(d\gamma) \right) \geq 0.$$

For $f = x \log x$ this is $D_{BS} - D$ with $f''(d\gamma) = d\gamma/\gamma$ and $\tilde{f}''(d\gamma) = d\gamma/\gamma^2$, which refines the inequality $D \leq D_{BS}$ of [57] level by level.

Proof. (a) With $X = T - \gamma(\mathbf{1} - T)$ we have $\rho - \gamma\sigma = \omega^{1/2} X \omega^{1/2}$. Put $A = \omega^{1/2} X_+ \omega^{1/2}$ and $B = \omega^{1/2} X_- \omega^{1/2}$. Since $\text{Tr } Y_+ = \min\{\text{Tr } Z : Z \geq 0, Z \geq Y\}$ and $A \geq A - B$, we get $E_\gamma \leq \text{Tr } A = E_\gamma^T$. If equality holds, an optimal test M satisfies $\text{Tr}(\mathbf{1} - M)A + \text{Tr } MB = 0$, so $A = MA$, $MB = 0$ and $AB = AMB = 0$; conversely $AB = 0$ gives $(A - B)_+ = A$. Since $\omega^{1/2}$ is injective on $\Pi_\omega \mathcal{H}$ and $X_\pm$ are invertible on the ranges of $P_{>c}^T$ and $P_{<c}^T$, $AB = 0$ is equivalent to $P_{>c}^T \omega P_{<c}^T = 0$. As γ runs over $(0, \infty)$ in both directions, c runs over $(0, 1)$, so equality everywhere means that ω commutes with every spectral projection of T ; then $\rho = \omega^{1/2} T \omega^{1/2}$ commutes with ω and with $\sigma = \omega - \rho$. The converse is clear.

(b) Let $\gamma_0 \in \Gamma$ and let B be an isometry onto V . On V , $\rho = \gamma_0 \sigma$ as quadratic forms, so $\text{Tr } \Pi_V \sigma = (1 - t_0) \text{Tr } B^* \omega B$. The spectral subspace of T at t_0 is $\omega^{1/2} V$, with projection $E_{t_0} = \omega^{1/2} B (B^* \omega B)^{-1} B^* \omega^{1/2}$, so $\text{Tr } E_{t_0} \omega = \text{Tr} (B^* \omega B)^{-1} B^* \omega^2 B$. Kadison's inequality for the unital map $Z \mapsto B^* Z B$ gives $B^* \omega^2 B \geq (B^* \omega B)^2$, hence $\text{Tr } E_{t_0} \omega \geq \text{Tr } B^* \omega B$, with equality if and only if $(\mathbf{1} - \Pi_V) \omega \Pi_V = 0$. The boundary atoms are the cases $V = \ker \rho \cap \Pi_\omega \mathcal{H}$ (where $\omega = \sigma$) and $V = \ker \sigma \cap \Pi_\omega \mathcal{H}$ (where $\omega = \rho$, using ρ -masses).

(c) Both μ_T (on $[0, \infty)$) and $\mu + m_0 \delta_0$ have total mass $\text{Tr } \sigma = 1$ by Proposition 5.2.5, and μ_T is atomic.

(d) D_f^T is the classical f -divergence of (p_T, q_T) , which is Definition 5.2.3 applied to the profile E_γ^T . Subtract and use (a). $\square$

Corollary 5.3.3 (Rigidity). Let f be convex with $f(1) = 0$ and f'' positive almost everywhere on $(0, \infty)$.

(i) If $D_f^{\text{HT}}(\rho\|\sigma) = D_f^T(\rho\|\sigma) < \infty$, then $[\rho, \sigma] = 0$.

(ii) If $f(0^+), f'(\infty) < \infty$ and $D_f^{\text{M}}(\rho\|\sigma) = D_f^{\text{HT}}(\rho\|\sigma)$, then $[\rho, \sigma] = 0$.

Proof. (i) The integrand of Theorem 5.3.2(d) is nonnegative and continuous in γ , so it vanishes for all γ ; apply Theorem 5.3.2(a). (ii) The supremum defining D_f^{M} is attained by some measurement M with at most $(\dim \mathcal{H})^2$ outcomes. Its outcome pair (p, q) has $E_\gamma(p\|q) \leq E_\gamma(\rho\|\sigma)$ for all γ , and equality of the integrals forces equality for every γ . For $\gamma \notin \Gamma$ the optimal test on $\Pi_\omega \mathcal{H}$ is unique, so $P_\gamma = \sum_{x: p_x > \gamma q_x} M_x$ there. These projections are nested in γ and continuous on each interval of $(0, \infty) \setminus \Gamma$, so they are locally constant, and then $\dot{P}_\gamma = 0$ forces the density of Theorem 5.3.1(b) to vanish. By Theorem 5.3.1(d), ρ and σ commute. $\square$

Example 5.3.4 (Two pure states). Let $\rho = |\psi\rangle\langle\psi|$, $\sigma = |\phi\rangle\langle\phi|$ with $c = |\langle\psi|\phi\rangle|^2 \in (0, 1)$. The 2×2 eigenvalue problem gives

$$\begin{aligned} e(\gamma) &= \frac{1}{2} \left(1 - \gamma + \sqrt{(1 + \gamma)^2 - 4c\gamma} \right), \\ \mu(d\gamma) &= \frac{2c(1 - c) d\gamma}{(\gamma^2 + 2(1 - 2c)\gamma + 1)^{3/2}}, \\ m_0 &= m_\infty = 1 - c. \end{aligned}$$

Here $\Gamma = \emptyset$ and the shadow is absolutely continuous, while T is a projection and $\mu_T = \delta_0$, $m_\infty^T = 1$: the reverse test sees two orthogonal states. Consequently the geometric fidelity is 0 for every $c < 1$, whereas

$$Q_{1/2}^{\text{HT}}(\rho\|\sigma) = 2c(1 - c) \int_0^\infty \frac{\gamma^{1/2} d\gamma}{(\gamma^2 + 2(1 - 2c)\gamma + 1)^{3/2}} \longrightarrow 1 \quad (c \rightarrow 1),$$

with $Q_{1/2}^{\text{HT}} = \frac{1}{4}B(\frac{3}{4}, \frac{3}{4}) \approx 0.4236$ at $c = \frac{1}{2}$.

5.4 Continuity on states

Normalise $f(1) = f'(1) = 0$. Then $f(0^+) = \int_{(0,1)} \gamma f''(d\gamma)$ and $f'(\infty) = \int_{[1,\infty)} f''(d\gamma)$, so

$$\int_0^\infty \min\{\gamma, 1\} f''(d\gamma) = f(0^+) + f'(\infty). \quad (5.4.1)$$

We call f *bounded* when this quantity is finite.

Theorem 5.4.1 (Uniform continuity). *Let f be bounded. For states $\rho, \rho', \sigma, \sigma'$ in any dimension, with $\varepsilon_1 = \frac{1}{2}\|\rho - \rho'\|_1$ and $\varepsilon_2 = \frac{1}{2}\|\sigma - \sigma'\|_1$,*

$$|D_f^{\text{HT}}(\rho\|\sigma) - D_f^{\text{HT}}(\rho'\|\sigma')| \leq \omega_f(\varepsilon_1, \varepsilon_2) := \int_0^\infty \min\{\gamma, 1, \varepsilon_1 + \gamma\varepsilon_2\} f''(d\gamma),$$

and $\omega_f(\varepsilon_1, \varepsilon_2) \rightarrow 0$ as $\varepsilon_1, \varepsilon_2 \rightarrow 0$. The same bound holds for the stabilised channel divergences

$$D_f^{\text{HT}}(\Phi\|\Psi) = \sup_{\psi} D_f^{\text{HT}}((\text{id} \otimes \Phi)(\psi) \| (\text{id} \otimes \Psi)(\psi))$$

with $\varepsilon_1 = \frac{1}{2}\|\Phi - \Phi'\|_\diamond$ and $\varepsilon_2 = \frac{1}{2}\|\Psi - \Psi'\|_\diamond$.

Proof. By Proposition 5.2.5, $D_f^{\text{HT}} = \int g f''$ with $g = e - (1 - \gamma)_+$, and $0 \leq g \leq \min\{\gamma, 1\}$ by Lemma 5.2.2. For a traceless Hermitian X ,

$$\max_{0 \leq M \leq 1} |\text{Tr} MX| = \frac{1}{2}\|X\|_1,$$

so $|e(\gamma) - e'(\gamma)| \leq \varepsilon_1 + \gamma\varepsilon_2$ for the two profiles. Hence $|g - g'| \leq \min\{\gamma, 1, \varepsilon_1 + \gamma\varepsilon_2\}$. The limit follows from (5.4.1) and dominated convergence. For channels, apply the state bound to each input and take suprema. $\square$

For the trace distance ($f'' = \delta_1$) the modulus is $\min\{1, \varepsilon_1 + \varepsilon_2\}$. For the squared Hellinger distance $f = (\sqrt{x} - 1)^2$, $f''(d\gamma) = \frac{1}{2}\gamma^{-3/2}d\gamma$, splitting at $\gamma = 2\varepsilon$, 1 and $1/(2\varepsilon)$ gives $\omega_f(\varepsilon, \varepsilon) \leq 4\sqrt{2\varepsilon}$, so the testing-region fidelity $Q_{1/2}^{\text{HT}} = 1 - \frac{1}{2}D_{(\sqrt{x}-1)^2}^{\text{HT}}$ has modulus at most $2\sqrt{2\varepsilon}$. Compare Corollary 4.12.5 in Chapter 4: the geometric fidelity, the reverse-test

value of the same f , is discontinuous at every pair whose sum is singular and whose fidelity is positive.

The same modulus holds for the measured divergences $D_f^M(\rho\|\sigma) = \sup_N D_f(N\rho\|N\sigma)$, the supremum over measurements N , and for their stabilised channel versions. Indeed, a measurement does not increase trace distances and ω_f is nondecreasing in both arguments, so Theorem 5.4.1 applied to each pair of outcome distributions, on which D_f^{HT} is the classical D_f by Proposition 5.2.5, gives

$$|D_f(N\rho\|N\sigma) - D_f(N\rho'\|N\sigma')| \leq \omega_f(\varepsilon_1, \varepsilon_2)$$

for every N , and the difference of two suprema, which are finite because f is bounded, is at most the supremum of the differences.

Corollary 5.4.2 (No dichotomy). *Every bounded testing-region divergence is continuous at every pair of states and channels. The perturbations of size $O(\sqrt{t})$ in Theorem 4.12.4 of Chapter 4, which move every Kubo–Ando mean of a pair with singular sum to its affine part, move D_f^{HT} by at most $\omega_f(O(\sqrt{t}), O(\sqrt{t}))$.*

For instance, with $\sigma = |0\rangle\langle 0|$ and $\rho_t = |v_t\rangle\langle v_t|$, $v_t = \sqrt{1-t}|0\rangle + \sqrt{t}|1\rangle$, the geometric fidelity is 0 for $t > 0$ and 1 at $t = 0$, while Example 5.3.4 gives

$$Q_{1/2}^{\text{HT}}(\rho_t\|\sigma) = 0.97904, 0.99968, 0.999996 \quad \text{at } t = 10^{-2}, 10^{-4}, 10^{-6}.$$

The discontinuities of mean-based divergences are therefore a property of the reverse test, not of quantum divergences as such.

Theorem 5.4.3 (Unbounded generators). *Let f be convex with $f(1) = 0$. Then D_f^{HT} is lower semicontinuous on pairs of states with values in $[0, \infty]$, and:*

- (a) *if $f'(\infty) = \infty$ and $f(0^+) < \infty$, then D_f^{HT} is continuous at (ρ, σ) if and only if σ is invertible or $D_f^{\text{HT}}(\rho\|\sigma) = \infty$;*
- (b) *if $f(0^+) = \infty$ and $f'(\infty) < \infty$, then the same holds with ρ in place of σ ;*
- (c) *if both are infinite, D_f^{HT} is continuous at (ρ, σ) if and only if both are invertible or $D_f^{\text{HT}}(\rho\|\sigma) = \infty$.*

In particular the Umegaki relative entropy is continuous at (ρ, σ) if and only if σ is invertible or $D(\rho\|\sigma) = \infty$.

Proof. Lower semicontinuity follows from Fatou’s lemma, because $g \geq 0$ is continuous in (ρ, σ) . (a) If $\sigma \geq \lambda \mathbf{1}$ with $\lambda > 0$, then $\sigma' \geq \frac{\lambda}{2} \mathbf{1}$ for nearby σ' , and $e(\gamma) = 0$ for $\gamma \geq 2/\lambda$ and all nearby ρ' , because $\rho' \leq \mathbf{1} \leq \gamma \sigma'$. So the integrand g vanishes beyond $2/\lambda$ uniformly. On $(0, 2/\lambda]$ it is dominated by $\min\{\gamma, 1\}$, which is f'' -integrable there because $f(0^+) < \infty$ and f'' is locally finite. If σ is singular and $D_f^{\text{HT}}(\rho\|\sigma) < \infty$, take a unit vector $v \in \ker \sigma$ and $\rho_t = (1-t)\rho + t|v\rangle\langle v|$. Then $m_\infty(\rho_t, \sigma) \geq t > 0$ and $D_f^{\text{HT}}(\rho_t\|\sigma) \geq f'(\infty)m_\infty = \infty$ by Proposition 5.2.5 (with $f \geq 0$), so continuity fails. At a pair with $D_f^{\text{HT}} = \infty$, lower semicontinuity gives continuity into $[0, \infty]$. (b) is (a) for $\tilde{f}$ and the exchanged pair, since $D_f^{\text{HT}}(\rho\|\sigma) = D_{\tilde{f}}^{\text{HT}}(\sigma\|\rho)$. (c) combines both. $\square$

Quantitative bounds at invertible σ , in terms of $D_{\max}$, are given in [15] through the same integral representation. Theorem 5.4.3 is the testing-region analogue of the dichotomy of Theorem 4.12.4, but the obstruction differs. There the obstruction is the singularity of the *sum*, through a coupling of size $O(\sqrt{t})$. Here it is the singularity of the *second argument*, through a leak of size $O(t)$.

The Petz divergences, whose shadow is the Nussbaum–Szkola pair, are treated in Section 5.5.

5.5 The modular shadow: Petz divergences

Write $\rho = \sum_i a_i E_i$ and $\sigma = \sum_j b_j F_j$ with distinct eigenvalues and spectral projections. Let L_ρ and R_σ be left and right multiplication on the Hilbert–Schmidt space; they commute, and $\Delta = L_\rho R_\sigma^{-1}$ is the relative modular operator. For convex f with $f(1) = 0$ the Petz f -divergence is

$$D_f^P(\rho\|\sigma) = \langle \mathbf{1}, P_f(L_\rho, R_\sigma) \mathbf{1} \rangle = \sum_{ij} \text{Tr}(E_i F_j) P_f(a_i, b_j),$$

with the perspective $P_f(a, b) = bf(a/b)$ extended by $P_f(a, 0) = af'(\infty)$ and $P_f(0, b) = bf(0^+)$ [61]. Its classical pair is the Nussbaum–Szkola pair $P_{ij} = a_i \text{Tr } E_i F_j$, $Q_{ij} = b_j \text{Tr } E_i F_j$.

Theorem 5.5.1 (The modular lift). *(a) With $\Omega = L_\rho + R_\sigma$ and the modular derivative $\tilde{T} = L_\rho \Omega^{-1} = \Delta(\mathbf{1} + \Delta)^{-1}$ on the support of Ω ,*

$$D_f^P(\rho\|\sigma) = \langle \mathbf{1}, \Omega^{1/2} h_f(\tilde{T}) \Omega^{1/2} \mathbf{1} \rangle, \quad h_f(t) = (1-t)f(t/(1-t)),$$

the formula of Theorem 4.2.15 and Lemma 4.7.4 for the commuting pair (L_ρ, R_σ) in the vector $\mathbf{1}$.

(b) The profile of the modular shadow is the Petz hockey-stick

$$E_\gamma^P(\rho\|\sigma) := \sum_{ij} (P_{ij} - \gamma Q_{ij})_+ = \sum_j \text{Tr } F_j (\rho - \gamma b_j \mathbf{1})_+,$$

and D_f^P is Definition 5.2.3 applied to E_γ^P in place of E_γ .

(c) The boundary masses of the modular shadow are $m_0 = \text{Tr } \Pi_\rho^\perp \sigma$ and $m_\infty = \text{Tr } \Pi_\sigma^\perp \rho$, the same as for the testing region, and in general strictly smaller than those of the reverse test. Consequently D_f^P and D_f^{HT} are finite under exactly the same conditions: no positive boundary mass may meet an infinite boundary value of f .

Proof. (a) $P_f(a, b) = (a+b)h_f(a/(a+b))$, and functional calculus of the commuting pair gives $P_f(L_\rho, R_\sigma) = \Omega^{1/2} h_f(\tilde{T}) \Omega^{1/2}$. (b) We have

$$\sum_i (a_i - \gamma b_j)_+ \text{Tr } E_i F_j = \text{Tr } F_j (\rho - \gamma b_j)_+.$$

The second statement is Proposition 5.2.5 for the commuting (classical) pair (P, Q) . (c) We have

$$m_0^{\text{NS}} = \sum_{i:a_i=0} \sum_j b_j \text{Tr } E_i F_j = \text{Tr } \Pi_\rho^\perp \sigma,$$

and symmetrically for m_∞ . The comparison with the reverse test is Theorem 5.3.2(b). $\square$

Theorem 5.5.1(c) separates two Lebesgue decompositions. The reverse test of Chapter 4 splits a pair along the spectral projections of T , whose extreme eigenspaces can be strictly larger than the support defects. The modular and testing-region shadows split it along the supports themselves.

Theorem 5.5.2 (Continuity of Petz divergences). *(a) If $f(0^+), f'(\infty) < \infty$, then D_f^P is jointly continuous on pairs of states.*

(b) For $\alpha \in (0, 1)$ and states in any dimension,

$$|Q_\alpha^P(\rho\|\sigma) - Q_\alpha^P(\rho'\|\sigma')| \leq (2\varepsilon_1)^\alpha + (2\varepsilon_2)^{1-\alpha}, \quad \varepsilon_1 = \frac{1}{2}\|\rho - \rho'\|_1, \quad \varepsilon_2 = \frac{1}{2}\|\sigma - \sigma'\|_1.$$

The same modulus holds for the stabilised channel quantity $\inf_\psi Q_\alpha^P$ with diamond distances.

(c) For unbounded f , D_f^P is lower semicontinuous, and its continuity points are exactly those of Theorem 5.4.3. In particular Q_α^P with $\alpha \in (1, 2]$ is continuous at (ρ, σ) if and only if σ is invertible or $\Pi_\rho \not\leq \Pi_\sigma$.

(d) Every n -use and regularised Petz quantity with $\alpha \in [0, 2]$ is d_T -Lipschitz in the second channel (Theorem 5.7.2).

Proof. (a) P_f is continuous on $[0, 1]^2$, with $|P_f(a, b)| \leq C(a + b)$ because a convex f with finite boundary values satisfies $|f(x)| \leq C(1 + x)$. Uniform approximation of P_f by polynomials in two variables makes $(\rho, \sigma) \mapsto \langle \mathbf{1}, P_f(L_\rho, R_\sigma)\mathbf{1} \rangle$ a uniform limit of continuous functions. (b) By Hölder's inequality,

$$|Q_\alpha^P(\rho\|\sigma) - Q_\alpha^P(\rho'\|\sigma')| \leq \|\rho^\alpha - \rho'^\alpha\|_{1/\alpha} \|\sigma^{1-\alpha} - \sigma'^{1-\alpha}\|_{1/(1-\alpha)} + \|\rho'^\alpha\|_{1/\alpha} \|\sigma^{1-\alpha} - \sigma'^{1-\alpha}\|_{1/(1-\alpha)}.$$

The two fixed norms equal 1. Ando's inequality for the operator monotone functions x^α and $x^{1-\alpha}$ [7] gives

$$\|\rho^\alpha - \rho'^\alpha\|_{1/\alpha} \leq \| |\rho - \rho'|^\alpha \|_{1/\alpha} = \|\rho - \rho'\|_1^\alpha,$$

and likewise for σ . For channels, apply this to each input. (c) D_f^P is the increasing limit of the bounded divergences with f'' truncated to $[1/k, k]$, hence lower semicontinuous. At invertible σ the perspective is continuous near the spectrum; at singular σ with finite value, the support leak of Theorem 5.4.3 gives $m_\infty > 0$ and an infinite value. (d) D_α^P is antitone in σ for $\alpha \in [0, 2]$ and logarithmically homogeneous. $\square$

The theory of Chapter 4 predicts discontinuity of a mean of the lifted pair wherever $\Omega = L_\rho + R_\sigma$ is singular, that is, wherever ρ and σ are both singular. Theorem 5.5.2(a) shows that this never happens for bounded f . The reason is that every perturbation of (ρ, σ) lifts to a perturbation of the commuting pair (L_ρ, R_σ) inside the commuting pairs. The $O(\sqrt{t})$ coupling between the kernel and the range of Ω that drives the dichotomy is never available.

Proposition 5.5.3 (Orderings between the shadows). (a) For operator convex f , $D_f^P \leq D_f^T$, the maximal f -divergence [61, 91].

(b) For $\alpha \in (0, 1)$, $D_\alpha^P \leq D_\alpha^{\text{HT}} \leq D_\alpha^P + \frac{\log 2}{1-\alpha}$ (Proposition 5.6.1).

(c) For $\alpha > 1$, $D_\alpha^M \leq D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha \leq D_\alpha^P$, by [84]. Theorem 6.3.2 in Chapter 6 gives a short proof of the middle inequality for $\alpha \geq 2$.

So the testing-region and modular shadows are ordered in Rényi divergence for every $\alpha \neq 1$, with the order reversing at $\alpha = 1$, where both give the Umegaki entropy.

5.6 Tensor products: what replaces multiplicativity

The reverse test and the modular shadow are multiplicative: the shadow of $(\rho \otimes \rho', \sigma \otimes \sigma')$ is the multiplicative convolution of the two shadows, since T and the Nussbaum–Szkoła pair factorise. The testing region is not. The Hirche–Tomamichel Rényi divergences are

not additive [63], so tensor powers must be handled by comparison with shadows that do factorise. We use the modular shadow below order one and the pinched pair above it. Write η_n for the error profile of $(\rho^{\otimes n}, \sigma^{\otimes n})$, $v(\sigma)$ for the number of distinct eigenvalues of σ , and $\mathcal{P}_\sigma$ for the pinching in the eigenbasis of σ .

Proposition 5.6.1 (Order below one). *For $\alpha \in (0, 1)$, $\frac{1}{2}Q_\alpha^P \leq Q_\alpha^{\text{HT}} \leq Q_\alpha^P$. Hence*

$$D_\alpha^P(\rho\|\sigma) \leq D_\alpha^{\text{HT}}(\rho\|\sigma) \leq D_\alpha^P(\rho\|\sigma) + \frac{\log 2}{1-\alpha}, \quad \left| \frac{1}{n}D_\alpha^{\text{HT}}(\rho^{\otimes n}\|\sigma^{\otimes n}) - D_\alpha^P(\rho\|\sigma) \right| \leq \frac{\log 2}{n(1-\alpha)}.$$

Proof. The upper bound is [13, (1.10)]. For the lower bound, the inequality of Nussbaum and Szkoła [99] applied to $(\rho, \gamma\sigma)$ gives $\eta(\gamma) \geq \frac{1}{2} \sum_{ij} \min\{P_{ij}, \gamma Q_{ij}\}$ for every γ . The right-hand side is half the error profile of the modular shadow, whose Mellin transform is Q_α^P by Proposition 5.2.6 applied to a commuting pair. The modular shadow is multiplicative, so the same holds for tensor powers with the same constant. $\square$

Proposition 5.6.2 (Order above one). *For $\alpha > 1$ and $\Pi_\rho \leq \Pi_\sigma$,*

$$\begin{aligned} Q_\alpha(\mathcal{P}_\sigma \rho\|\sigma) &\leq Q_\alpha^{\text{HT}}(\rho\|\sigma) \leq v(\sigma)^\alpha Q_\alpha(\mathcal{P}_\sigma \rho\|\sigma), \\ Q_\alpha(\mathcal{P}_\sigma \rho\|\sigma) &\leq \tilde{Q}_\alpha(\rho\|\sigma) \leq v(\sigma)^\alpha Q_\alpha(\mathcal{P}_\sigma \rho\|\sigma), \end{aligned}$$

where $Q_\alpha(\mathcal{P}_\sigma \rho\|\sigma) = \text{Tr}(\mathcal{P}_\sigma \rho)^\alpha \sigma^{1-\alpha}$ is classical and $\tilde{Q}_\alpha$ is the sandwiched quasi-entropy. Consequently

$$\begin{aligned} |D_\alpha^{\text{HT}}(\rho\|\sigma) - \tilde{D}_\alpha(\rho\|\sigma)| &\leq \frac{\alpha}{\alpha-1} \log v(\sigma), \\ \left| \frac{1}{n}D_\alpha^{\text{HT}}(\rho^{\otimes n}\|\sigma^{\otimes n}) - \tilde{D}_\alpha(\rho\|\sigma) \right| &\leq \frac{\alpha(d-1)}{\alpha-1} \frac{\log(n+1)}{n}. \end{aligned}$$

Proof. The pinching is a channel fixing σ , so data processing for D_α^{HT} (Proposition 5.2.7) and for $\tilde{D}_\alpha$ gives the lower bounds, and the pinched pair commutes. The pinching inequality $\rho \leq v(\sigma)\mathcal{P}_\sigma(\rho)$ [54] and the monotonicity of $A \mapsto \text{Tr}(A - \gamma\sigma)_+$ give $E_\gamma(\rho\|\sigma) \leq vE_{\gamma/v}(\mathcal{P}_\sigma \rho\|\sigma)$. The Mellin formula with $\alpha > 1$ has positive weights, and the substitution $\gamma = vu$ produces the factor v^α . The same inequality with the monotonicity of

$$\rho \mapsto \text{Tr}(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}})^\alpha$$

gives the bound for $\tilde{Q}_\alpha$. For tensor powers use $v(\sigma^{\otimes n}) \leq (n+1)^{d-1}$ and the additivity of $\tilde{D}_\alpha$. $\square$

Propositions 5.6.1 and 5.6.2 give the regularisation theorem of [63] with explicit rates: the testing-region Rényi divergence regularises to the Petz divergence below order one and to the sandwiched divergence above it. In shadow language, $\frac{1}{n} \log \int \gamma^\alpha d\mu_n$ converges to $\log Q_\alpha^P$ for $\alpha < 1$ and to $\log \tilde{Q}_\alpha$ for $\alpha > 1$, where μ_n is the shadow of n copies. These are exactly the cumulant functions whose Legendre transforms give the Hoeffding and strong-converse exponents [55, 94, 96]. The two halves of the asymptotic testing region are thus governed by the two multiplicative shadows that bracket the testing region. At $\alpha = 1$ both halves meet at the Umegaki entropy, the only additive functional consistent with optimal binary tests [48].

5.7 Channels: the Thompson metric and the adaptive Stein exponent

Let $\Phi, \Psi : \mathcal{B}(\mathcal{X}) \rightarrow \mathcal{B}(\mathcal{Y})$ be channels with Choi matrices $J(\Phi) = \sum_{ij} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|)$, $d_{\mathcal{X}} = \dim \mathcal{X}$, $d_{\mathcal{Y}} = \dim \mathcal{Y}$. (This Choi matrix has the input factor first, unlike the other chapters; see Table 3.1.) Write $\Phi \leq_{\text{cp}} \Phi'$ when $\Phi' - \Phi$ is completely positive, that is, $J(\Phi) \leq J(\Phi')$. The *Thompson metric* on channels is

$$d_{\text{T}}(\Psi, \Psi') = \inf\{t \geq 0 : \Psi \leq_{\text{cp}} e^t \Psi', \Psi' \leq_{\text{cp}} e^t \Psi\} = d_{\text{T}}(J(\Psi), J(\Psi')),$$

which is $+\infty$ unless the Choi matrices have the same support. An n -use adaptive strategy produces final states $\rho_n^{\Phi}, \rho_n^{\Psi}$ by interleaving the n uses with fixed channels on a growing memory. For a divergence $\mathbb{D}$ on positive operators let $\mathbb{D}^{(n)}(\Phi\|\Psi) = \sup \mathbb{D}(\rho_n^{\Phi}\|\rho_n^{\Psi})$ over adaptive strategies, let $\mathbb{D}^{\text{reg}} = \lim_n \frac{1}{n} \mathbb{D}(\Phi^{\otimes n}\|\Psi^{\otimes n})$ when it exists, and let

$$\mathbb{D}^A(\Phi\|\Psi) = \sup_{\mathcal{R}, \rho_{\mathcal{R}\mathcal{X}}, \sigma_{\mathcal{R}\mathcal{X}}} \left[\mathbb{D}((\text{id} \otimes \Phi)\rho \parallel (\text{id} \otimes \Psi)\sigma) - \mathbb{D}(\rho\|\sigma) \right]$$

be the amortised divergence [128], the supremum running over pairs with $\mathbb{D}(\rho\|\sigma) < \infty$.

Lemma 5.7.1. *If $\Psi \leq_{\text{cp}} c\Psi'$, then $\rho_n^{\Psi} \leq c^n \rho_n^{\Psi'}$ for every n -use adaptive strategy, and $(\text{id} \otimes \Psi)(\sigma) \leq c(\text{id} \otimes \Psi')(\sigma)$ for every $\sigma \geq 0$.*

Proof. The second statement is the definition. For the first, let $X_k \geq 0$ and $X'_k \geq 0$ be the states before the k -th use in the two runs of the same strategy, and suppose $X_k \leq c^{k-1} X'_k$. Then

$$(\text{id} \otimes \Psi)(X_k) \leq c(\text{id} \otimes \Psi')(X_k) \leq c^k(\text{id} \otimes \Psi')(X'_k),$$

because $\text{id} \otimes \Psi'$ is positive. The interleaved channels are positive and preserve the order. $\square$

Theorem 5.7.2 (Thompson stability in the second argument). *Let $\mathbb{D}$ be antitone in its second argument and logarithmically homogeneous, $\mathbb{D}(\rho\|c\sigma) = \mathbb{D}(\rho\|\sigma) - \log c$. Then for all channels and all n ,*

$$|\mathbb{D}^{(n)}(\Phi\|\Psi) - \mathbb{D}^{(n)}(\Phi\|\Psi')| \leq n d_{\text{T}}(\Psi, \Psi'),$$

and the same bound holds for parallel strategies. Consequently $\mathbb{D}^{\text{reg}}$ and $\mathbb{D}^A$ are 1-Lipschitz in Ψ for d_{T} . If $J(\Psi)$ is invertible and

$$\eta := d_{\mathcal{X}} \|\Psi - \Psi'\|_{\diamond} / \lambda_{\min}(J(\Psi)) < 1,$$

then $d_{\text{T}}(\Psi, \Psi') \leq -\log(1 - \eta)$.

Proof. If $t > d_{\text{T}}(\Psi, \Psi')$, Lemma 5.7.1 gives $e^{-nt} \rho_n^{\Psi} \leq \rho_n^{\Psi'} \leq e^{nt} \rho_n^{\Psi}$ for the same strategy, and the two hypotheses on $\mathbb{D}$ give $|\mathbb{D}(\rho_n^{\Phi}\|\rho_n^{\Psi'}) - \mathbb{D}(\rho_n^{\Phi}\|\rho_n^{\Psi})| \leq nt$. Take suprema. The amortised case is the second statement of Lemma 5.7.1 applied to each pair. For the last claim,

$$\|J(\Psi') - J(\Psi)\|_{\infty} \leq \|J(\Psi') - J(\Psi)\|_1 \leq d_{\mathcal{X}} \|\Psi' - \Psi\|_{\diamond},$$

so

$$(1 - \eta)J(\Psi) \leq J(\Psi') \leq (1 + \eta)J(\Psi).$$

$\square$

The hypotheses hold for the Umegaki, Petz ($\alpha \in [0, 2]$), sandwiched ($\alpha \geq \frac{1}{2}$), geometric ($\alpha \in (0, 2]$) and measured Rényi divergences, for $D_{\max}$, $D_{\min}$ and for the hypothesis-testing relative entropy. Theorem 5.7.2 rests on the fact that the Thompson metric is additive under tensor products, $d_T(\Psi^{\otimes n}, \Psi'^{\otimes n}) = n d_T(\Psi, \Psi')$, whereas the diamond norm of $\Psi^{\otimes n} - \Psi'^{\otimes n}$ saturates. This is how tensor products are handled in the second argument. The first argument needs a different idea, because no divergence is monotone there.

Lemma 5.7.3. *For every $\mathcal{R}$ and $\sigma_{\mathcal{R}\mathcal{X}} \geq 0$,*

$$\lambda_{\min}(J(\Psi)) \sigma_{\mathcal{R}} \otimes \mathbf{1}_{\mathcal{Y}} \leq (\text{id} \otimes \Psi)(\sigma) \leq \lambda_{\max}(J(\Psi)) \sigma_{\mathcal{R}} \otimes \mathbf{1}_{\mathcal{Y}}.$$

Proof. For pure $\sigma = (M \otimes \mathbf{1})|\Omega\rangle\langle\Omega|(M \otimes \mathbf{1})^*$, with $|\Omega\rangle = \sum_i |i\rangle|i\rangle$ and $M : \mathcal{X} \rightarrow \mathcal{R}$, we have $\sigma_{\mathcal{R}} = MM^*$ and $(\text{id} \otimes \Psi)(\sigma) = (M \otimes \mathbf{1})J(\Psi)(M \otimes \mathbf{1})^*$. Mixed states follow by linearity. $\square$

Theorem 5.7.4 (The first argument, uniformly over references). *Let $J(\Psi)$ be invertible with condition number $\kappa(\Psi)$, and let h be the binary entropy. For all channels Φ, Φ' , all $\mathcal{R}$ and all $\rho_{\mathcal{R}\mathcal{X}}, \sigma_{\mathcal{R}\mathcal{X}}$ with $D(\rho\|\sigma) < \infty$,*

$$\begin{aligned} & \left| D((\text{id} \otimes \Phi')\rho \parallel (\text{id} \otimes \Psi)\sigma) - D((\text{id} \otimes \Phi)\rho \parallel (\text{id} \otimes \Psi)\sigma) \right| \\ & \leq 2\varepsilon \log d_{\mathcal{Y}} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right) + \varepsilon \log \kappa(\Psi), \end{aligned}$$

where

$$\varepsilon = \frac{1}{2} \|(\text{id} \otimes (\Phi - \Phi'))\rho\|_1 \leq \frac{1}{2} \|\Phi - \Phi'\|_{\diamond}.$$

Hence the same bound, with $\varepsilon = \frac{1}{2} \|\Phi - \Phi'\|_{\diamond}$, holds for $|D^A(\Phi\|\Psi) - D^A(\Phi'\|\Psi)|$ and for the stabilised channel divergence.

Proof. Put $\omega = (\text{id} \otimes \Phi)\rho$, $\omega' = (\text{id} \otimes \Phi')\rho$ and $\xi = (\text{id} \otimes \Psi)\sigma$. Since $D(\rho\|\sigma) < \infty$, $\Pi_{\rho_{\mathcal{R}}} \leq \Pi_{\sigma_{\mathcal{R}}}$, and Lemma 5.7.3 gives $\Pi_{\omega}, \Pi_{\omega'} \leq \Pi_{\sigma_{\mathcal{R}}} \otimes \mathbf{1} = \Pi_{\xi}$. On this support, operator monotonicity of the logarithm and Lemma 5.7.3 give $\log \xi = \log \sigma_{\mathcal{R}} \otimes \mathbf{1} + K$ with $\log \lambda_{\min} \leq K \leq \log \lambda_{\max}$. Both ω and ω' have $\mathcal{R}$ -marginal $\rho_{\mathcal{R}}$, so the term $\text{Tr}(\omega' - \omega)(\log \sigma_{\mathcal{R}} \otimes \mathbf{1})$ vanishes and

$$\begin{aligned} D(\omega'\|\xi) - D(\omega\|\xi) &= [H(\mathcal{Y}|\mathcal{R})_{\omega} - H(\mathcal{Y}|\mathcal{R})_{\omega'}] - \text{Tr}(\omega' - \omega)(K - c_0 \mathbf{1}), \\ c_0 &= \frac{1}{2} \log(\lambda_{\min} \lambda_{\max}), \end{aligned}$$

using $\text{Tr}(\omega' - \omega) = 0$. Winter's continuity bound for the conditional entropy [129] controls the first bracket by $2\varepsilon \log d_{\mathcal{Y}} + (1 + \varepsilon)h(\varepsilon/(1 + \varepsilon))$, independently of $\mathcal{R}$. The second term is at most

$$\|\omega' - \omega\|_1 \|K - c_0\|_{\infty} \leq 2\varepsilon \cdot \frac{1}{2} \log \kappa(\Psi).$$

The amortised statement follows by taking suprema, since the subtracted term $D(\rho\|\sigma)$ does not depend on Φ . $\square$

The cancellation of $\log \sigma_{\mathcal{R}}$ is what makes the bound independent of the reference system, and hence of the number of uses hidden in an amortising pair. This is the conditional-entropy replacement for the tensor-power structure that hockey-stick divergences lack.

Corollary 5.7.5 (Continuity of the adaptive Stein exponent). *Let $\zeta(\Phi, \Psi)$ be the optimal type-II error exponent for adaptive discrimination of Φ against Ψ with vanishing type-I error. Then $\zeta = D^{\text{reg}} = D^A$ [41, 125], and:*

(a) ζ is lower semicontinuous on pairs of channels;

(b) if $J(\Psi)$ is invertible, then $\zeta(\Phi, \Psi) \leq \log(d_{\mathcal{X}}/\lambda_{\min}(J(\Psi)))$, and for all Φ' and all Ψ' with the same Choi support,

$$|\zeta(\Phi, \Psi) - \zeta(\Phi', \Psi')| \leq 2\varepsilon \log d_{\mathcal{Y}} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right) + \varepsilon \log \kappa(\Psi) + d_{\text{T}}(\Psi, \Psi'),$$

$$\varepsilon = \frac{1}{2}\|\Phi - \Phi'\|_{\diamond};$$

(c) ζ is continuous at (Φ, Ψ) if and only if $J(\Psi)$ is invertible or $\zeta(\Phi, \Psi) = \infty$.

Proof. (a) By superadditivity and Fekete's lemma, $D^{\text{reg}} = \sup_n \frac{1}{n} D(\Phi^{\otimes n} \| \Psi^{\otimes n})$, a supremum of lower semicontinuous functions. (b) Since $J(\Phi) \leq \text{Tr } J(\Phi) \mathbf{1} = d_{\mathcal{X}} \mathbf{1}$, we have $\Phi \leq_{\text{cp}} c\Psi$ with $c = d_{\mathcal{X}}/\lambda_{\min}(J(\Psi))$, and Lemma 5.7.1 gives $D(\rho_n^{\Phi} \| \rho_n^{\Psi}) \leq n \log c$. The estimate is Theorem 5.7.4 at Ψ followed by Theorem 5.7.2 at Φ' . (c) If $J(\Psi)$ is invertible, (b) gives continuity, because $d_{\text{T}}(\Psi, \Psi') \rightarrow 0$ as $\Psi' \rightarrow \Psi$ by Theorem 5.7.2. If $\zeta(\Phi, \Psi) = \infty$, (a) gives continuity into $[0, \infty]$. Otherwise $J(\Psi)$ is singular and $\zeta(\Phi, \Psi) < \infty$. Put $\Phi_t = (1 - t)\Phi + t\Theta$ with Θ the completely depolarising channel. Then $J(\Phi_t)$ is invertible, so its support is not contained in that of $J(\Psi)$, and the maximally entangled input gives $\zeta(\Phi_t, \Psi) \geq D(\Phi_t \| \Psi) = \infty$, while $\|\Phi_t - \Phi\|_{\diamond} \leq 2t$. $\square$

Remark 5.7.6. Corollary 5.7.5 is the regularised counterpart of Theorem 4.12.4 in Chapter 4, with “ $J(\Phi + \Psi)$ invertible or the fidelity vanishes” replaced by “ $J(\Psi)$ invertible or the exponent is infinite”. The same proof gives the dichotomy for the one-shot channel relative entropy. By Theorem 5.7.2, the strong-converse exponent and every other regularised quantity of an admissible $\mathbb{D}$ is Thompson-Lipschitz in Ψ . Continuity in Φ is Theorem 5.7.8 for the regularised sandwiched divergences, which have a chain rule, and Theorem 5.7.11 for the regularised Petz divergences, which do not. The continuity of $\tilde{D}_{\alpha}^{\text{reg}}$ in the order α at $\alpha = 1$ was recently established in [124].

Proposition 5.7.7 (The T-lift for channels). *Let $J(\Phi + \Psi) = WW^*$, $J(\Phi) = WTW^*$ as in Theorem 4.2.2 of Chapter 4, and $H_{\gamma} = \text{Tr}_{\mathcal{Y}} W(T - \gamma(\mathbf{1} - T))_+ W^*$. For every input $\psi_{\mathcal{R}\mathcal{X}}$ with $\mathcal{X}$ -marginal τ ,*

$$E_{\gamma}((\text{id} \otimes \Phi)\psi \| (\text{id} \otimes \Psi)\psi) \leq \text{Tr } \tau^{\text{T}} H_{\gamma},$$

so the channel hockey-stick divergence satisfies $E_{\gamma}^{\diamond}(\Phi \| \Psi) \leq \lambda_{\max}(H_{\gamma})$. With $f(1) = f'(1) = 0$ one has the operator identity

$$h_f(T) = \int_1^{\infty} ((T - \gamma(\mathbf{1} - T))_+ f''(d\gamma) + ((\mathbf{1} - T) - \gamma T)_+ \tilde{f}''(d\gamma)),$$

so integrating recovers

$$D_f^{\text{HT}}(\Phi \| \Psi) \leq D_f^{\text{T}}(\Phi \| \Psi) = \lambda_{\max}(\text{Tr}_{\mathcal{Y}} W h_f(T) W^*)$$

(Theorem 4.7.5 in Chapter 4) level by level, with the gap of Theorem 5.3.2(d) at each input.

Proof. With $\psi = (M \otimes \mathbf{1})|\Omega\rangle\langle\Omega|(M \otimes \mathbf{1})^*$ the two outputs are $V(T)V^*$ and $V(\mathbf{1} - T)V^*$ with $V = (M \otimes \mathbf{1})W$. The argument of Theorem 5.3.2(a) gives

$$E_{\gamma} \leq \text{Tr } V X_+ V^* = \text{Tr}(M^* M \otimes \mathbf{1}) W X_+ W^*,$$

and $M^* M = \tau^{\text{T}}$. $\square$

The sandwiched divergences and the strong-converse exponent

For $\alpha > 1$ the regularised sandwiched channel divergence equals its amortised version, $\tilde{D}_\alpha^{\text{reg}} = \tilde{D}_\alpha^A$, by the chain rule of Fawzi and Fawzi [43, Corollary 5.2]. The same work characterises the strong-converse exponent of adaptive channel discrimination through the family $\{\tilde{D}_\alpha^{\text{reg}}\}_{\alpha>1}$. The second argument is covered by Theorem 5.7.2. For the first argument the conditional-entropy decomposition is unavailable, and we replace it by a ratio estimate on the quasi-entropy.

Theorem 5.7.8 (First argument, sandwiched). *Let $\alpha > 1$, let $J(\Psi)$ be invertible with $\lambda = \lambda_{\min}(J(\Psi))$, and put $m = \min\{d_{\mathcal{X}}, d_{\mathcal{Y}}\}$. For all channels Φ, Φ' , all $\mathcal{R}$ and all $\rho_{\mathcal{R}\mathcal{X}}, \sigma_{\mathcal{R}\mathcal{X}}$ with $\tilde{D}_\alpha(\rho\|\sigma) < \infty$,*

$$\frac{|\tilde{Q}_\alpha((\text{id} \otimes \Phi')\rho\|(\text{id} \otimes \Psi)\sigma) - \tilde{Q}_\alpha((\text{id} \otimes \Phi)\rho\|(\text{id} \otimes \Psi)\sigma)|}{\tilde{Q}_\alpha(\rho\|\sigma)} \leq \alpha d_{\mathcal{Y}} \left(\frac{m}{\lambda}\right)^{\alpha-1} \|J(\Phi') - J(\Phi)\|_1.$$

Consequently

$$|\tilde{D}_\alpha^{\text{reg}}(\Phi'\|\Psi) - \tilde{D}_\alpha^{\text{reg}}(\Phi\|\Psi)| \leq \frac{\alpha}{\alpha-1} d_{\mathcal{X}} d_{\mathcal{Y}} \left(\frac{m}{\lambda}\right)^{\alpha-1} \|\Phi' - \Phi\|_\diamond,$$

and the same bound holds for the stabilised one-shot channel divergence.

Proof. Write $\omega = (\text{id} \otimes \Phi)\rho$, $\omega' = (\text{id} \otimes \Phi')\rho$ and $\xi = (\text{id} \otimes \Psi)\sigma$, and put $\beta = \frac{\alpha-1}{2\alpha}$, so that $\tilde{Q}_\alpha(\cdot\|\xi) = \text{Tr}(\xi^{-\beta} \cdot \xi^{-\beta})^\alpha$. Let $J(\Phi') - J(\Phi) = J_+ - J_-$ be the Jordan decomposition and $\Gamma_\pm$ the corresponding completely positive maps, so that $\omega' - \omega = \Delta_+ - \Delta_-$ with $\Delta_\pm = (\text{id} \otimes \Gamma_\pm)\rho \geq 0$.

- (i) By Lemma 5.7.3, $\xi \geq \lambda \sigma_{\mathcal{R}} \otimes \mathbf{1}$ on $\Pi_{\sigma_{\mathcal{R}}} \otimes \mathbf{1} = \Pi_\xi$, and since $x \mapsto x^{-2\beta}$ is operator monotone decreasing with $2\beta \in (0, 1)$, we get $\xi^{-2\beta} \leq \lambda^{-2\beta} \sigma_{\mathcal{R}}^{-2\beta} \otimes \mathbf{1}$ there. For $\Delta \geq 0$ supported in $\Pi_{\rho_{\mathcal{R}}} \otimes \mathbf{1} \leq \Pi_\xi$ this gives

$$\|\xi^{-\beta} \Delta \xi^{-\beta}\|_\alpha \leq \lambda^{-2\beta} \|(\sigma_{\mathcal{R}}^{-\beta} \otimes \mathbf{1}) \Delta (\sigma_{\mathcal{R}}^{-\beta} \otimes \mathbf{1})\|_\alpha.$$

- (ii) Lemma 5.7.3 for $\Gamma_\pm$ gives $\Delta_\pm \leq \|J_\pm\|_\infty \rho_{\mathcal{R}} \otimes \mathbf{1}$, hence, by data processing,

$$\|(\sigma_{\mathcal{R}}^{-\beta} \otimes \mathbf{1}) \Delta_\pm (\sigma_{\mathcal{R}}^{-\beta} \otimes \mathbf{1})\|_\alpha \leq \|J_\pm\|_\infty d_{\mathcal{Y}}^{1/\alpha} \tilde{Q}_\alpha(\rho_{\mathcal{R}}\|\sigma_{\mathcal{R}})^{1/\alpha} \leq \|J_\pm\|_\infty d_{\mathcal{Y}}^{1/\alpha} \tilde{Q}_\alpha(\rho\|\sigma)^{1/\alpha}.$$

- (iii) Since $\omega \leq m \rho_{\mathcal{R}} \otimes \mathbf{1}$ (either from $J(\Phi) \leq d_{\mathcal{X}} \mathbf{1}$ or from the general inequality $\omega \leq d_{\mathcal{Y}} \omega_{\mathcal{R}} \otimes \mathbf{1}$), monotonicity of $\tilde{Q}_\alpha$ gives

$$\tilde{Q}_\alpha(\omega\|\xi) \leq \lambda^{1-\alpha} \tilde{Q}_\alpha(\omega\|\sigma_{\mathcal{R}} \otimes \mathbf{1}) \leq \lambda^{1-\alpha} m^\alpha d_{\mathcal{Y}} \tilde{Q}_\alpha(\rho\|\sigma),$$

and the same for ω' .

- (iv) For $A, B \geq 0$,

$$|\text{Tr } A^\alpha - \text{Tr } B^\alpha| \leq \alpha \max\{\|A\|_\alpha, \|B\|_\alpha\}^{\alpha-1} \|A - B\|_\alpha,$$

by differentiating along the segment and using Hölder's inequality.

Apply (iv) to $A = \xi^{-\beta} \omega' \xi^{-\beta}$ and $B = \xi^{-\beta} \omega \xi^{-\beta}$, bound $\|A - B\|_\alpha$ by (i) and (ii) and the maxima by (iii). The powers of λ combine to $\lambda^{-(\alpha-1)^2/\alpha - (\alpha-1)/\alpha} = \lambda^{1-\alpha}$, and

$$\|J_+\|_\infty + \|J_-\|_\infty \leq \|J(\Phi') - J(\Phi)\|_1 \leq d_{\mathcal{X}} \|\Phi' - \Phi\|_\diamond.$$

This gives the first estimate. For the second, $R_\Phi(\rho, \sigma) = \tilde{Q}_\alpha(\omega\|\xi)/\tilde{Q}_\alpha(\rho\|\sigma)$ satisfies $\tilde{D}_\alpha^A(\Phi\|\Psi) = \frac{1}{\alpha-1} \log \sup R_\Phi$, and $\sup R_\Phi \geq 1$ because $\tilde{D}_\alpha^A \geq 0$. The supremum inherits the uniform bound, and log is 1-Lipschitz on $[1, \infty)$. $\square$

Corollary 5.7.9 (Continuity of the sandwiched exponents). *Let $\alpha > 1$.*

- (a) $\tilde{D}_\alpha^{\text{reg}}$ is lower semicontinuous, and it is continuous at (Φ, Ψ) if and only if $J(\Psi)$ is invertible or $\tilde{D}_\alpha^{\text{reg}}(\Phi \| \Psi) = \infty$. At invertible $J(\Psi)$ it obeys the bound of Theorem 5.7.8 in Φ and is d_T -Lipschitz in Ψ .
- (b) On every compact range $\alpha \in (1, A]$ the functions $\frac{\alpha-1}{\alpha} \tilde{D}_\alpha^{\text{reg}}(\cdot \| \Psi)$ are equi-Lipschitz, with constant $d_{\mathcal{X}d_{\mathcal{Y}}}(m/\lambda)^{A-1}$ in the diamond norm. Consequently the strong-converse exponent of adaptive discrimination [43, Theorem 5.5],

$$\overline{H}(r, \Phi, \Psi) = \sup_{\alpha > 1} \frac{\alpha - 1}{\alpha} (r - \tilde{D}_\alpha^{\text{reg}}(\Phi \| \Psi)),$$

is continuous at every pair with invertible $J(\Psi)$, for every rate $r < D_{\max}(\Phi \| \Psi)$.

Proof. (a) Lower semicontinuity and the discontinuity at singular $J(\Psi)$ are as in Corollary 5.7.5: a support leak gives $\tilde{D}_\alpha = \infty$. (b) The factor $\frac{\alpha-1}{\alpha}$ cancels the factor $\frac{\alpha}{\alpha-1}$ of Theorem 5.7.8, and $m/\lambda \geq 1$. For

$$r < D_{\max}(\Phi \| \Psi) = \lim_{\alpha \rightarrow \infty} \tilde{D}_\alpha(\Phi \| \Psi),$$

choose A with $\tilde{D}_A(\Phi \| \Psi) > r$. By the one-shot case of Theorem 5.7.8 and Theorem 5.7.2, $\tilde{D}_A(\Phi' \| \Psi') > r$ for nearby pairs. Monotonicity in α then makes every term with $\alpha > A$ negative, while the supremum is nonnegative (let $\alpha \downarrow 1$). So near (Φ, Ψ) the supremum runs over $(1, A]$, where the terms are equicontinuous in both arguments. $\square$

The first argument without a chain rule: regularised Petz divergences

For $\alpha \in (0, 1)$, $Q_\alpha^{\text{P}}(\rho \| \sigma) = \text{Tr } \rho^\alpha \sigma^{1-\alpha}$ is nondecreasing and homogeneous of degree α in ρ , and nondecreasing and homogeneous of degree $1 - \alpha$ in σ . No chain rule is known for these divergences, so the route of Theorem 5.7.8 is closed: we cannot control one use against an arbitrary reference and then pass from the amortised to the regularised divergence. Example 5.7.12 shows that a one-use estimate uniform over references cannot hold in general. We work instead with parallel uses and replace them one at a time. This works for perturbations that act locally on the output and the input. Write $D_\alpha^{(n)}(\Phi \| \Psi) = D_\alpha^{\text{P}}(\Phi^{\otimes n} \| \Psi^{\otimes n})$ and

$$D_\alpha^{\text{P,reg}} = \lim_n \frac{1}{n} D_\alpha^{(n)} = \sup_n \frac{1}{n} D_\alpha^{(n)},$$

the limit existing by superadditivity.

Lemma 5.7.10 (Local stability, uniformly over references). *Let $\alpha \in (0, 1)$ and $\lambda \mathbf{1} \leq J(\Psi) \leq \Lambda \mathbf{1}$ with $\lambda > 0$. Let W be an operator on $\mathcal{Y}$ with polar decomposition $W = U|W|$, and put $s = \|W^*W - \mathbf{1}\|_\infty$ and $u = \|U - \mathbf{1}\|_\infty$, with $2s(\Lambda/\lambda)^{1-\alpha} < 1$ and $2u\Lambda/\lambda < 1$. Then for every $\mathcal{R}$, every $\omega \geq 0$ on $\mathcal{R} \otimes \mathcal{Y}$ and every $\sigma \geq 0$ on $\mathcal{R} \otimes \mathcal{X}$, with $\xi = (\text{id} \otimes \Psi)\sigma$,*

$$\begin{aligned} Q_\alpha^{\text{P}}((\mathbf{1} \otimes W)\omega(\mathbf{1} \otimes W)^* \| \xi) &\geq c_W Q_\alpha^{\text{P}}(\omega \| \xi), \\ c_W &= (1 + s)^{\alpha-1} (1 - 2s(\Lambda/\lambda)^{1-\alpha}) (1 - 2u\Lambda/\lambda)^{1-\alpha}. \end{aligned}$$

Proof. Put $A = (\mathbf{1} \otimes W)\omega^{1/2}$ and $G = W^*W$. Then $(AA^*)^\alpha = A(A^*A)^{\alpha-1}A^*$, with the power taken on the support of A^*A , which is that of ω . Hence

$$Q_\alpha^{\text{P}}(AA^* \| \xi) = \text{Tr} \left[(\omega^{1/2}(\mathbf{1} \otimes G)\omega^{1/2})^{\alpha-1} \omega^{1/2}(\mathbf{1} \otimes W^*) \xi^{1-\alpha} (\mathbf{1} \otimes W)\omega^{1/2} \right].$$

(i) $\omega^{1/2}(\mathbf{1} \otimes G)\omega^{1/2} \leq (1+s)\omega$, and $x \mapsto x^{\alpha-1}$ is operator monotone decreasing, so the first factor is at least $(1+s)^{\alpha-1}\omega^{\alpha-1}$ on the support of ω . (ii) We have

$$(\mathbf{1} \otimes W^*)\xi^{1-\alpha}(\mathbf{1} \otimes W) = (\mathbf{1} \otimes |W|)\xi_U^{1-\alpha}(\mathbf{1} \otimes |W|)$$

with $\xi_U = (\text{id} \otimes \Psi_U)\sigma$, where $\Psi_U = U^*\Psi(\cdot)U$. The Choi matrix $J(\Psi_U) = (\mathbf{1} \otimes U^*)J(\Psi)(\mathbf{1} \otimes U)$ also lies between λ and Λ . By Lemma 5.7.3 and operator monotonicity of $x^{1-\alpha}$, $\xi_U^{1-\alpha} = N^{1/2}MN^{1/2}$ with $N = \sigma_{\mathcal{R}}^{1-\alpha} \otimes \mathbf{1}$ and $\lambda^{1-\alpha} \leq M \leq \Lambda^{1-\alpha}$ on the support of N . Now $E = \mathbf{1} \otimes (|W| - \mathbf{1})$ commutes with N and has $\|E\| \leq s$, and

$$(\mathbf{1} + E)M(\mathbf{1} + E) \geq M - 2s\Lambda^{1-\alpha} \geq (1 - 2s(\Lambda/\lambda)^{1-\alpha})M.$$

Conjugating by $N^{1/2}$ gives

$$(\mathbf{1} \otimes |W|)\xi_U^{1-\alpha}(\mathbf{1} \otimes |W|) \geq (1 - 2s(\Lambda/\lambda)^{1-\alpha})\xi_U^{1-\alpha}.$$

Finally, $\|J(\Psi_U) - J(\Psi)\| \leq 2u\Lambda$ gives $J(\Psi_U) \geq (1 - 2u\Lambda/\lambda)J(\Psi)$, hence $\xi_U^{1-\alpha} \geq (1 - 2u\Lambda/\lambda)^{1-\alpha}\xi^{1-\alpha}$. (iii) If $A_1 \geq A_0 \geq 0$ and $B_1 \geq B_0 \geq 0$, then $\text{Tr } A_1 B_1 \geq \text{Tr } A_0 B_0$. Apply this to the bounds of (i) and (ii). The result is

$$c_W \text{Tr } \omega^{\alpha-1} \omega^{1/2} \xi^{1-\alpha} \omega^{1/2} = c_W Q_{\alpha}^{\text{P}}(\omega \| \xi).$$

□

The only place where the reference enters is $N = \sigma_{\mathcal{R}}^{1-\alpha} \otimes \mathbf{1}$, and the perturbation $\mathbf{1} \otimes |W|$ commutes with it. This is the Petz counterpart of the cancellation of $\log \sigma_{\mathcal{R}}$ in Theorem 5.7.4.

Theorem 5.7.11 (First argument for regularised Petz divergences). *Let $\alpha \in (0, 1)$.*

(a) *For all channels and all n ,*

$$|D_{\alpha}^{(n)}(\Phi' \| \Psi) - D_{\alpha}^{(n)}(\Phi \| \Psi)| \leq n \frac{\alpha}{1-\alpha} d_{\text{T}}(\Phi, \Phi').$$

Hence $D_{\alpha}^{\text{P,reg}}$ is $\frac{\alpha}{1-\alpha}$ -Lipschitz in the first channel for the Thompson metric.

(b) *If $J(\Psi)$ is invertible, $D_{\alpha}^{\text{P,reg}}(\cdot \| \Psi)$ is lower semicontinuous.*

(c) *Let $\lambda \mathbf{1} \leq J(\Psi) \leq \Lambda \mathbf{1}$ with $\lambda > 0$. Let W be as in Lemma 5.7.10, let $\mathcal{W} = W \cdot W^*$, and let $\mathcal{V} = V \cdot V^*$ be a unitary channel on $\mathcal{X}$ with $v = \|V - \mathbf{1}\|_{\infty} < \lambda/(2\Lambda)$. If $\Phi' \geq_{\text{cp}} (1 - \delta) \mathcal{W} \circ \Phi \circ \mathcal{V}$, then for every n*

$$\frac{1}{n} D_{\alpha}^{(n)}(\Phi' \| \Psi) \leq \frac{1}{n} D_{\alpha}^{(n)}(\Phi \| \Psi) + \frac{\alpha}{1-\alpha} \log \frac{1}{1-\delta} + \frac{1}{1-\alpha} \log \frac{1}{c_W} + \log \frac{1}{1-2v\Lambda/\lambda}.$$

(d) *Consequently, if $J(\Psi)$ is invertible, $D_{\alpha}^{\text{P,reg}}(\cdot \| \Psi)$ is continuous at every Φ with invertible Choi matrix and at every isometric channel $\Phi = K \cdot K^*$. At a general Φ it is continuous along every sequence $\Phi_k \rightarrow \Phi$ with $\Phi_k \geq_{\text{cp}} (1 - \delta_k) \mathcal{W}_k \circ \Phi \circ \mathcal{V}_k$, where $\delta_k \rightarrow 0$, $\mathcal{W}_k \rightarrow \mathbf{1}$ and $\mathcal{V}_k \rightarrow \mathbf{1}$.*

Proof. (a) If $\Phi \leq_{\text{cp}} e^t \Phi'$, then $\Phi^{\otimes n}(\psi) \leq e^{nt} \Phi'^{\otimes n}(\psi)$ for every input (Lemma 5.7.1). Monotonicity and homogeneity in the first argument give

$$Q_{\alpha}^{\text{P}}(\Phi^{\otimes n} \psi \| \Psi^{\otimes n} \psi) \leq e^{nt\alpha} Q_{\alpha}^{\text{P}}(\Phi'^{\otimes n} \psi \| \Psi^{\otimes n} \psi).$$

Exchange the roles, take the infimum over ψ , and apply $-\frac{1}{1-\alpha} \log$. (b) For fixed n , $\Phi \mapsto \inf_{\psi} Q_{\alpha}^P(\Phi^{\otimes n} \psi \| \Psi^{\otimes n} \psi)$ is continuous, by Theorem 5.5.2(b) and compactness of the pure inputs with reference $\mathcal{X}^{\otimes n}$. It is positive, because $\Psi^{\otimes n} \psi \geq \lambda_{\min}(J(\Psi))^n \psi_{\mathcal{R}} \otimes \mathbf{1}$ has support containing that of $\Phi^{\otimes n} \psi$. So each $\frac{1}{n} D_{\alpha}^{(n)}$ is continuous, and their supremum is lower semicontinuous. (c) First, $\Phi'^{\otimes n} \geq_{\text{cp}} (1-\delta)^n (\mathcal{W} \Phi \mathcal{V})^{\otimes n}$, so

$$Q_{\alpha}^P(\Phi'^{\otimes n} \psi \| \Psi^{\otimes n} \psi) \geq (1-\delta)^{n\alpha} Q_{\alpha}^P((\mathcal{W} \Phi \mathcal{V})^{\otimes n} \psi \| \Psi^{\otimes n} \psi).$$

Next remove the n copies of $\mathcal{W}$ one at a time. When $\mathcal{W}$ acts on the k -th output, the remaining outputs form part of the reference, and the second argument is $\Psi^{\otimes n} \psi = (\text{id} \otimes \Psi_k)(\sigma_k)$ for a positive σ_k on the other outputs and the k -th input. Lemma 5.7.10 therefore applies at each step and gives a factor c_W^n . Finally, $(\Phi \mathcal{V})^{\otimes n} \psi = \Phi^{\otimes n} \psi_V$ with $\psi_V = (\mathbf{1} \otimes V^{\otimes n}) \psi (\mathbf{1} \otimes V^{\otimes n})^*$, and $\Psi^{\otimes n} \psi = (\Psi \circ \mathcal{V}^*)^{\otimes n} \psi_V$. Since

$$J(\Psi \circ \mathcal{V}^*) = (\overline{V} \otimes \mathbf{1}) J(\Psi) (\overline{V} \otimes \mathbf{1})^* \geq (1 - 2v\Lambda/\lambda) J(\Psi)$$

and Q_{α}^P is nondecreasing in its second argument, we get

$$Q_{\alpha}^P(\Phi^{\otimes n} \psi_V \| \Psi^{\otimes n} \psi) \geq (1 - 2v\Lambda/\lambda)^{n(1-\alpha)} \inf_{\psi'} Q_{\alpha}^P(\Phi^{\otimes n} \psi' \| \Psi^{\otimes n} \psi').$$

Collect the factors, take the infimum over ψ , and apply $-\frac{1}{n(1-\alpha)} \log$. (d) For invertible $J(\Phi)$, use (a) and the last statement of Theorem 5.7.2. Let $\Phi = K \cdot K^*$ with K an isometry, so $J(\Phi) = |\kappa\rangle\langle\kappa|$ with $\kappa = (\mathbf{1} \otimes K)\Omega$ and $\|\kappa\|^2 = d_{\mathcal{X}}$. If $\|J(\Phi') - J(\Phi)\|_{\infty} = \eta$, the top eigenvalue μ of $J(\Phi')$ satisfies $\mu \geq d_{\mathcal{X}} - \eta$. Its eigenvector, normalised as $\kappa' = (\mathbf{1} \otimes B)\Omega$ with $\|\kappa'\|^2 = d_{\mathcal{X}}$ and $\langle\kappa|\kappa'\rangle \geq 0$, tends to κ as $\eta \rightarrow 0$, so $B \rightarrow K$. Now

$$J(\Phi') \geq \frac{\mu}{d_{\mathcal{X}}} |\kappa'\rangle\langle\kappa'| = \frac{\mu}{d_{\mathcal{X}}} J(B \cdot B^*),$$

and $B \cdot B^* = \mathcal{W} \circ \Phi$ with $W = BK^* + \mathbf{1} - KK^* \rightarrow \mathbf{1}$. So $\Phi' \geq_{\text{cp}} (1-\delta)\mathcal{W} \circ \Phi$ with $\delta \leq \eta/d_{\mathcal{X}}$ and $s, u \rightarrow 0$. Then (c) gives $\limsup D_{\alpha}^{\text{P,reg}}(\Phi' \| \Psi) \leq D_{\alpha}^{\text{P,reg}}(\Phi \| \Psi)$, and (b) gives the matching lower bound. The last statement follows from (b) and (c) in the same way. $\square$

The hypothesis of (d) covers post-processing by any channels $\mathcal{N}_k \rightarrow \text{id}$ followed by mixing in noise, and pre-processing by unitaries $\mathcal{V}_k \rightarrow \text{id}$. A channel close to the identity has a Choi matrix close to $|\Omega\rangle\langle\Omega|$, and therefore dominates $(1-\delta)W \cdot W^*$ for its top Kraus operator $W \rightarrow \mathbf{1}$, by the eigenvector argument in the proof of (d).

Example 5.7.12 (No uniform one-use bound). On a qubit let $\Phi_{\varepsilon} = (1-\varepsilon)\text{id} + \varepsilon\mathcal{D}$, with $\mathcal{D}$ completely depolarising, and $\Phi' = \text{id}$, so $\|\Phi_{\varepsilon} - \Phi'\|_{\diamond} \leq 2\varepsilon$. Let $J(\Psi)$ be invertible. Take $\mathcal{R} = \mathbb{C}^2$, $\rho = |v\rangle\langle v|$ with $v = \sqrt{1-\delta}|00\rangle + \sqrt{\delta}|11\rangle$, and $\sigma = |1\rangle\langle 1| \otimes \frac{1}{2}$, so $\xi = (\text{id} \otimes \Psi)\sigma = |1\rangle\langle 1| \otimes \Psi(\frac{1}{2})$. Then

$$Q_{\alpha}^P(\Phi' \rho \| \xi) = \delta \langle 1 | \Psi(\frac{1}{2})^{1-\alpha} | 1 \rangle.$$

On the other hand $\Phi_{\varepsilon} \rho \geq \varepsilon \delta |1\rangle\langle 1| \otimes \frac{1}{2}$, so

$$Q_{\alpha}^P(\Phi_{\varepsilon} \rho \| \xi) \geq (\varepsilon \delta / 2)^{\alpha} \text{Tr} \Psi(\frac{1}{2})^{1-\alpha}.$$

Hence

$$\frac{Q_{\alpha}^P(\Phi' \rho \| (\text{id} \otimes \Psi)\sigma)}{Q_{\alpha}^P(\Phi_{\varepsilon} \rho \| (\text{id} \otimes \Psi)\sigma)} \leq \left(\frac{2}{\varepsilon}\right)^{\alpha} \delta^{1-\alpha} \xrightarrow{\delta \rightarrow 0} 0$$

for every fixed ε . No estimate $Q_\alpha^P(\Phi' \rho \| \Psi \sigma) \geq c(\varepsilon) Q_\alpha^P(\Phi_\varepsilon \rho \| \Psi \sigma)$ with $c(\varepsilon) > 0$ can hold uniformly over references, let alone one with $c(\varepsilon) \rightarrow 1$. Removing noise is the bad direction, and it is not local. The divergence itself is continuous at both points: at Φ_ε by Theorem 5.7.11(a), since $J(\Phi_\varepsilon)$ is invertible, and at id by (d).

What remains open is continuity at channels with singular Choi matrix along perturbations that rotate the support of $J(\Phi)$ by a unitary of $\mathcal{X} \otimes \mathcal{Y}$ that is not a product. Such perturbations are not of the form in (c), and Example 5.7.12 shows that slot-by-slot replacement cannot handle them in general.

Testing-region channel divergences: the one-shot formula

The supremum over inputs does not commute with the integral over γ , so the $\lambda_{\max}$ formula of Theorem 4.7.5 in Chapter 4 cannot hold verbatim. It holds in dual form. For an input with $\mathcal{X}$ -marginal τ put $\theta = \tau^\top$. By the proof of Proposition 5.7.7 the two outputs are unitarily equivalent to

$$\omega_\theta = (\sqrt{\theta} \otimes \mathbf{1}) J(\Phi) (\sqrt{\theta} \otimes \mathbf{1}), \quad \xi_\theta = (\sqrt{\theta} \otimes \mathbf{1}) J(\Psi) (\sqrt{\theta} \otimes \mathbf{1}),$$

operators on $\mathcal{R} \otimes \mathcal{Y}$ with $\mathcal{R} \cong \mathcal{X}$, and these formulas define $\omega_\theta, \xi_\theta$ for every $\theta \geq 0$. Let f be convex with $f(1) = f'(1) = 0$. Write $m = f''$ and $\tilde{m} = \tilde{f}''$, restricted to $[1, \infty)$, and define the *testing-region operator* of the input,

$$\Gamma_f(\theta) = \text{Tr}_{\mathcal{Y}} \left[\int (\omega_\theta - \gamma \xi_\theta)_+ m(d\gamma) + \int (\xi_\theta - \gamma \omega_\theta)_+ \tilde{m}(d\gamma) \right] \geq 0, \\ \text{Tr } \Gamma_f(\theta) = D_f^{\text{HT}}(\omega_\theta \| \xi_\theta).$$

We assume that the integrals converge, for instance when $m, \tilde{m}$ are finite or when $J(\Phi)$ and $J(\Psi)$ dominate each other, since then the integrands vanish for large γ .

Theorem 5.7.13 (One-shot formula). *(a) The function $F(\theta) = D_f^{\text{HT}}(\omega_\theta \| \xi_\theta)$ is concave and positively homogeneous on positive operators θ . At every invertible θ it is differentiable, and its gradient is $\theta^{-1/2} \Gamma_f(\theta) \theta^{-1/2}$.*

(b) (Dual formula)

$$D_f^{\text{HT}}(\Phi \| \Psi) = \max_{\theta} F(\theta) = \inf \left\{ \lambda_{\max} \left(\text{Tr}_{\mathcal{Y}} \left[\int Y_\gamma m(d\gamma) + \int Z_\gamma \tilde{m}(d\gamma) \right] \right) \right\},$$

where the maximum runs over states and the infimum over measurable families with $Y_\gamma \geq 0$, $Y_\gamma \geq J(\Phi) - \gamma J(\Psi)$, $Z_\gamma \geq 0$ and $Z_\gamma \geq J(\Psi) - \gamma J(\Phi)$. For f'' with finitely many atoms both sides are semidefinite programs.

(c) (Optimal inputs) An invertible θ is optimal if and only if $\Gamma_f(\theta) = c\theta$ for a scalar c , and then $c = D_f^{\text{HT}}(\Phi \| \Psi)$. In particular the maximally entangled input is optimal if and only if

$$\text{Tr}_{\mathcal{Y}} \left[\int (J(\Phi) - \gamma J(\Psi))_+ m(d\gamma) + \int (J(\Psi) - \gamma J(\Phi))_+ \tilde{m}(d\gamma) \right] \text{ is a multiple of } \mathbf{1}_{\mathcal{X}}.$$

(d) (The reverse test is one dual point) With $J(\Phi + \Psi) = WW^$ and $J(\Phi) = WTW^*$, the family $Y_\gamma = W(T - \gamma(\mathbf{1} - T))_+ W^*$, $Z_\gamma = W(\mathbf{1} - T - \gamma T)_+ W^*$ is feasible in (b). Its value is $\lambda_{\max}(\text{Tr}_{\mathcal{Y}} W h_f(T) W^*) = D_f^T(\Phi \| \Psi)$, the closed form of Theorem 4.7.5.*

Proof. Step 1. For Hermitian X on $\mathcal{R} \otimes \mathcal{Y}$ and $\theta \geq 0$,

$$\mathrm{Tr} [(\sqrt{\theta} \otimes \mathbf{1})X(\sqrt{\theta} \otimes \mathbf{1})]_+ = \inf \{ \mathrm{Tr}(\theta \otimes \mathbf{1})Y : Y \geq 0, Y \geq X \},$$

and for invertible θ the infimum is attained exactly at

$$Y = (\theta^{-1/2} \otimes \mathbf{1})[(\sqrt{\theta} \otimes \mathbf{1})X(\sqrt{\theta} \otimes \mathbf{1})]_+(\theta^{-1/2} \otimes \mathbf{1}).$$

Indeed, write $\tilde{X} = (\sqrt{\theta} \otimes \mathbf{1})X(\sqrt{\theta} \otimes \mathbf{1})$, put $\tilde{Y} = (\sqrt{\theta} \otimes \mathbf{1})Y(\sqrt{\theta} \otimes \mathbf{1})$, which satisfies $\tilde{Y} \geq 0$ and $\tilde{Y} \geq \tilde{X}$, and let P be the projection onto the positive eigenspace of $\tilde{X}$. Then

$$\mathrm{Tr}(\theta \otimes \mathbf{1})Y = \mathrm{Tr} \tilde{Y} \geq \mathrm{Tr} P \tilde{Y} P \geq \mathrm{Tr} P \tilde{X} P = \mathrm{Tr} \tilde{X}_+.$$

Equality forces $(\mathbf{1} - P)\tilde{Y}(\mathbf{1} - P) = 0$, hence $\tilde{Y} = P\tilde{Y}P$ and then $\tilde{Y} = \tilde{X}_+$. For singular θ use the optimiser at $\theta + \epsilon \mathbf{1}$ and let $\epsilon \rightarrow 0$.

(a). Apply Step 1 to $X = J(\Phi) - \gamma J(\Psi)$ and to $X = J(\Psi) - \gamma J(\Phi)$, noting

$$\omega_\theta - \gamma \xi_\theta = (\sqrt{\theta} \otimes \mathbf{1})(J(\Phi) - \gamma J(\Psi))(\sqrt{\theta} \otimes \mathbf{1}),$$

and integrate. This gives $F(\theta) = \inf_{\mathcal{C}} \mathrm{Tr} \theta G(Y, Z)$, with $G(Y, Z) = \mathrm{Tr}_{\mathcal{Y}}[\int Y dm + \int Z d\tilde{m}]$ and $\mathcal{C}$ the convex set of feasible families. So F is an infimum of linear functions: it is concave, homogeneous and upper semicontinuous. At invertible θ_0 the optimal family of Step 1 gives the supergradient $G_0 = \theta_0^{-1/2} \Gamma_f(\theta_0) \theta_0^{-1/2}$. Conversely, write G_θ for the value of G at the optimal family of θ . Since $\mathrm{Tr} \theta_0 G_\theta \geq F(\theta_0)$, we get

$$F(\theta) = \mathrm{Tr} \theta G_\theta \geq F(\theta_0) + \mathrm{Tr}(\theta - \theta_0)G_\theta.$$

As $\theta \rightarrow \theta_0$, $G_\theta \rightarrow G_0$ by continuity of $\theta \mapsto (\omega_\theta - \gamma \xi_\theta)_+$ and dominated convergence, so F is differentiable at θ_0 with gradient G_0 .

(b). The set of states is convex and compact, $\mathcal{C}$ is convex, and $(\theta, (Y, Z)) \mapsto \mathrm{Tr} \theta G(Y, Z)$ is affine and continuous in each variable. Sion's minimax theorem therefore gives

$$\max_{\theta} \inf_{\mathcal{C}} \mathrm{Tr} \theta G = \inf_{\mathcal{C}} \max_{\theta} \mathrm{Tr} \theta G = \inf_{\mathcal{C}} \lambda_{\max}(G).$$

The maximum is attained because F is upper semicontinuous. For finitely many atoms, $N = (\sqrt{\theta} \otimes \mathbf{1})M(\sqrt{\theta} \otimes \mathbf{1})$ turns $\max_{\theta} F$ into

$$\max \left\{ \sum_k c_k \mathrm{Tr} N_k X_k : 0 \leq N_k \leq \theta \otimes \mathbf{1}, \theta \geq 0, \mathrm{Tr} \theta = 1 \right\},$$

and (b) is its Lagrange dual. Both sides are strictly feasible.

(c). If θ is invertible and $\Gamma_f(\theta) = c\theta$, the gradient is $c\mathbf{1}$, and concavity gives

$$F(\theta') \leq F(\theta) + c \mathrm{Tr}(\theta' - \theta) = F(\theta)$$

for every state θ' ; moreover $c = \mathrm{Tr} \Gamma_f(\theta) = F(\theta)$. Conversely, an optimal invertible θ lies in the relative interior of the state space, so its gradient G satisfies $\mathrm{Tr}(\theta' - \theta)G \leq 0$ for all states θ' . That is, $G \leq (\mathrm{Tr} \theta G)\mathbf{1} = F(\theta)\mathbf{1}$, or $\Gamma_f(\theta) \leq F(\theta)\theta$, and equality of traces forces $\Gamma_f(\theta) = F(\theta)\theta$. At $\theta = \mathbf{1}/d_{\mathcal{X}}$ we have $\Gamma_f(\mathbf{1}/d_{\mathcal{X}}) = d_{\mathcal{X}}^{-1} \mathrm{Tr}_{\mathcal{Y}}[\cdots]$, with the bracket of (c).

(d). On Π_W ,

$$W(T - \gamma(\mathbf{1} - T))W^* = J(\Phi) - \gamma J(\Psi),$$

so the family is feasible. By the operator identity of Proposition 5.7.7,

$$\int Y_\gamma m(d\gamma) + \int Z_\gamma \tilde{m}(d\gamma) = Wh_f(T)W^*,$$

and the value is that of Theorem 4.7.5. □

Part (d) places the closed form of the reverse test inside the dual program of the testing region. The two divergences are the same infimum, once over all dominating families and once over the single family built from T . The inequality $D_f^{\text{HT}} \leq D_f^T$ for channels is thus weak duality. Part (c) is a nonlinear eigenvalue problem $\Gamma_f(\theta) = D\theta$. In the linear case $\Gamma(\theta) = \sqrt{\theta} A \sqrt{\theta}$ of the reverse test it reduces to the top eigenvector of $A = \text{Tr}_Y W h_f(T) W^*$. In our tests the iteration $\theta \mapsto \Gamma_f(\theta) / \text{Tr} \Gamma_f(\theta)$, started at $\mathbf{1}/d_{\mathcal{X}}$, converged to the optimum (Section 5.9), but we have no proof of convergence in general. The symmetric case and two cheap bounds follow.

Proposition 5.7.14 (Covariant pairs). *Suppose that Φ and Ψ are jointly covariant, $\Phi(U_g \cdot U_g^*) = V_g \Phi(\cdot) V_g^*$ and $\Psi(U_g \cdot U_g^*) = V_g \Psi(\cdot) V_g^*$, for a group $g \mapsto U_g$ acting irreducibly on $\mathcal{X}$. Then for every convex f the maximally entangled input is optimal:*

$$\begin{aligned} D_f^{\text{HT}}(\Phi \parallel \Psi) &= D_f^{\text{HT}}(J(\Phi)/d_{\mathcal{X}} \parallel J(\Psi)/d_{\mathcal{X}}) \\ &= \frac{1}{d_{\mathcal{X}}} \int_1^\infty \left(E_\gamma(J(\Phi) \parallel J(\Psi)) f''(d\gamma) + E_\gamma(J(\Psi) \parallel J(\Phi)) \tilde{f}''(d\gamma) \right), \end{aligned}$$

a function of the eigenvalues of $J(\Phi) - \gamma J(\Psi)$ only. The same holds for D_f^P and every other divergence that satisfies data processing and is additive on block-diagonal pairs.

Proof. For pure inputs the output divergence depends only on the input marginal τ on $\mathcal{X}$, since two purifications differ by an isometry on the reference, and mixed inputs do no better, by data processing. Given ψ , attach a classical flag and twirl:

$$\psi' = \int dg |g\rangle\langle g| \otimes (\mathbf{1} \otimes U_g) \psi (\mathbf{1} \otimes U_g)^*.$$

By covariance the outputs are block diagonal in g with unitarily rotated blocks, so the divergence of ψ' equals that of ψ . The marginal of ψ' on $\mathcal{X}$ is $\mathbf{1}/d_{\mathcal{X}}$ by irreducibility, so a purification of ψ' does at least as well as ψ and has the same value as the maximally entangled input. $\square$

For D_f^{HT} the statement is also a case of Theorem 5.7.13(c). Covariance makes $J(\Phi)$ and $J(\Psi)$ invariant under $\bar{U}_g \otimes V_g$, so the bracket in (c) commutes with the irreducible representation $\bar{U}_g$ and is a multiple of $\mathbf{1}$ by Schur's lemma. Criterion (c) is more general than covariance, since it is one matrix equation in $J(\Phi)$ and $J(\Psi)$.

Proposition 5.7.15 (Two-sided bounds). *For all channels and convex f ,*

$$\begin{aligned} D_f^{\text{HT}}(J(\Phi)/d_{\mathcal{X}} \parallel J(\Psi)/d_{\mathcal{X}}) &\leq D_f^{\text{HT}}(\Phi \parallel \Psi) \leq \int_1^\infty \left(E_\gamma^\diamond(\Phi \parallel \Psi) f''(d\gamma) + E_\gamma^\diamond(\Psi \parallel \Phi) \tilde{f}''(d\gamma) \right) \\ &\leq D_f^T(\Phi \parallel \Psi), \end{aligned}$$

where

$$E_\gamma^\diamond(\Phi \parallel \Psi) = \max\{\text{Tr} N(J(\Phi) - \gamma J(\Psi)) : 0 \leq N \leq \tau \otimes \mathbf{1}, \tau \geq 0, \text{Tr} \tau = 1\}$$

is a semidefinite program for each γ , and the last quantity is $\lambda_{\max}(\text{Tr}_Y W h_f(T) W^*)$. The first two inequalities are equalities for jointly covariant pairs. The middle inequality is an equality if and only if one input is optimal for m -almost every level of $E_\gamma(\Phi \parallel \Psi)$ and $\tilde{m}$ -almost every level of $E_\gamma(\Psi \parallel \Phi)$. The exact value is the program of Theorem 5.7.13(b).

Proof. The lower bound is the maximally entangled input. The middle inequality exchanges supremum and integral. The SDP form follows by substituting $N = (\sqrt{\tau^\top} \otimes \mathbf{1}) M (\sqrt{\tau^\top} \otimes \mathbf{1})$ for a test M , and the upper bound is Proposition 5.7.7. $\square$

Theorem 5.7.16 (Regularised values). *For parallel strategies let $D_\alpha^{\text{HT,reg}}(\Phi\|\Psi)$ be the limit of $\frac{1}{n}D_\alpha^{\text{HT}}(\Phi^{\otimes n}\|\Psi^{\otimes n})$. The limit exists and*

$$D_\alpha^{\text{HT,reg}} = \begin{cases} D_\alpha^{\text{P,reg}}, & \alpha \in (0, 1), \\ D^{\text{reg}}, & \alpha = 1, \\ \tilde{D}_\alpha^{\text{reg}}, & \alpha > 1. \end{cases}$$

In particular, for $\alpha > 1$ the regularised testing-region divergence is the quantity computed by the semidefinite hierarchy of [43].

Proof. For $\alpha < 1$ apply Proposition 5.6.1 to the outputs of every input. This gives $D_\alpha^{\text{P}} \leq D_\alpha^{\text{HT}} \leq D_\alpha^{\text{P}} + \frac{\log 2}{1-\alpha}$ for the n -fold channel divergences, and the Petz channel divergence is superadditive. For $\alpha = 1$ the divergences coincide input by input [46]. For $\alpha > 1$, the inequality $D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$ ([84]; Theorem 6.3.2 for $\alpha \geq 2$) applied input by input gives

$$\frac{1}{n}D_\alpha^{\text{HT}}(\Phi^{\otimes n}\|\Psi^{\otimes n}) \leq \frac{1}{n}\tilde{D}_\alpha(\Phi^{\otimes n}\|\Psi^{\otimes n}) \leq \tilde{D}_\alpha^{\text{reg}}.$$

For the lower bound, valid for all $\alpha > 1$, fix m and an input τ_m for m uses, and feed $\tau_m^{\otimes k}$. Proposition 5.6.2 applied to the k -fold output states gives

$$\frac{1}{km}D_\alpha^{\text{HT}} \rightarrow \frac{1}{m}\tilde{D}_\alpha(\Phi^{\otimes m}(\tau_m)\|\Psi^{\otimes m}(\tau_m)).$$

Data processing handles n not divisible by m , and we optimise over τ_m and m . □

5.8 The classification after this chapter

Family	Shadow	States	Channels
I. Means, maximal	reverse test T	complete (Chapter 4)	closed form; dichotomy at $J(\Phi + \Psi)$

Family	Shadow	States	Channels
II. Testing region (Umegaki, E_γ , D_f^{HT})	$\mu = e''$: atoms at the relative spectrum, continuous part from the Helstrom coupling	Theorems 5.3.1, 5.3.2; uniform moduli; dichotomy at σ ; super-multiplicative (see Chapter 6) at $\alpha = 2$, at every integer order against pure partners (Theorem 6.7.1), at $\alpha = 3$ for sign-free pairs, whenever one reference state has at most two distinct eigenvalues (Lieb–Thirring identity), against three-level partners with a nonnegative 3-cycle and for weakly correlated negative cycles on both sides (Corollaries 6.7.7 and 6.7.9), at $\alpha = 3$ to second order around trivial partners $\sigma' \propto 1$, to first order at $\alpha = \infty$, for $\alpha > 1$ with one factor measured, and for enough copies (Theorems 6.4.1–6.4.3, 6.5.3, 6.6.2, Corollary 6.6.3, Propositions 6.6.5, 6.8.3); the open case at $\alpha = 3$ reduces to singular states whose reference states have at least three distinct eigenvalues (Corollary 6.6.6), for three-level references to an inequality for K_3 alone (Proposition 6.7.11), and in general to a polynomial criterion (Theorem 6.7.12); at integer orders the conjecture is a correlation inequality on a simplex (Proposition 6.5.1); the conjecture is a convex-order statement for the Frenkel measure (Theorem 6.8.2)	exact one-shot formula: dual $\lambda_{\max}$ program, optimal inputs solve $\Gamma_f(\theta) = D\theta$, reverse test is one dual point (Theorem 5.7.13); covariant pairs; regularised values (Theorem 5.7.16)
III. Modular (Petz)	Nussbaum–Szkola; modular derivative $\tilde{T} = \Delta(\mathbf{1} + \Delta)^{-1}$	Theorems 5.5.1, 5.5.2: same boundary masses as II; continuous for bounded f ; dimension-free Hölder modulus for Q_α^{P} ; orderings with II	dimension-free modulus for the stabilised Q_α^{P} , $\alpha < 1$; Thompson-Lipschitz in Ψ ($\alpha \in [0, 2]$) and in Φ ($\alpha < 1$); first argument of $D_\alpha^{\text{P}, \text{reg}}$: local stability, continuity at isometries and at invertible $J(\Phi)$ (Theorem 5.7.11); $D_\alpha^{\text{HT}, \text{reg}} = D_\alpha^{\text{P}, \text{reg}}$ for $\alpha < 1$
IV. Measured	supremum of shadows	$D_f^{\text{M}} \leq D_f^{\text{HT}}$, equality only if commuting; the modulus of Theorem 5.4.1 for bounded f	the same modulus for the stabilised divergences; Thompson-Lipschitz in Ψ for the measured Rényi divergences (Theorem 5.7.2)
V. Regularised, sandwiched	limits of II, III and pinching	$D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$ for $\alpha > 1$ [84], with a short proof for $\alpha \geq 2$ (Theorem 6.3.2); rates in Propositions 5.6.1, 5.6.2	Stein exponent (Corollary 5.7.5); sandwiched and strong-converse exponents (Corollary 5.7.9); $D_\alpha^{\text{HT}, \text{reg}} = \tilde{D}_\alpha^{\text{reg}}$ for $\alpha > 1$

5.9 Numerical checks

Random states are Wishart-normalised and random channels come from random isometries.

Claim	Result
Proposition 5.2.6 (Mellin formulas)	agreement with Definition 5.2.3 to 7–8 digits for $\alpha \in \{0.3, 0.5, 0.8, 1.5, 2\}$
Theorem 5.3.1(b) (density)	finite differences of e match the formula to 6 digits at 13 points, $d = 3, 4$
Theorem 5.3.1(c), 5.3.2(b) (atoms)	jump of e' equals $\text{Tr } \Pi_V \sigma$ to 6 digits, and the testing-region atom is at most the reverse-test atom, at all 12 atoms of 4 random qutrit pairs
Theorem 5.3.2(a) (T-lift)	$\min(E_\gamma^T - E_\gamma) \geq -2.4 \cdot 10^{-15}$ over 300 pairs ($d \leq 4$, random ranks) and 100 values of γ ; gap $\leq 2 \cdot 10^{-16}$ for commuting pairs
Example 5.3.4	profile to $5 \cdot 10^{-13}$; Q_α^{HT} to 8 digits for $\alpha \in \{0.3, 0.5, 0.9\}$
Theorem 5.4.1 (squared Hellinger)	$ \Delta D_f^{\text{HT}} /\omega_f \in [0.004, 0.275]$ over 30 perturbed pairs
Proposition 5.6.1 (Nussbaum–Szkoła side)	$\min(2\eta - \eta_{\text{NS}}) \geq 1.3 \cdot 10^{-4}$ over 300 pairs; the reverse comparison $\eta \leq \eta_{\text{NS}}$ fails (by 0.23), so the profile comparison is one-sided, while $Q_\alpha^{\text{HT}}/Q_\alpha^{\text{P}} \in [0.942, 1]$
Proposition 5.6.2	$Q_\alpha^{\text{HT}}/Q_\alpha(\mathcal{P}_\sigma \rho \ \sigma) \in [1.002, 2.105]$ and $Q_\alpha^{\text{HT}}/(v^\alpha Q_\alpha(\mathcal{P}_\sigma \rho \ \sigma)) \in [0.111, 0.483]$ over 60 pairs, $\alpha \in \{1.5, 2\}$
Theorem 5.7.4	$\max(\text{lhs} - \text{rhs}) = -4.1 \cdot 10^{-3}$ over 300 random $(\Phi, \Phi', \Psi, \rho, \sigma)$ including rank-deficient inputs and Kraus rank 1–3
Theorem 5.7.2 ($n = 1$)	$\max(\Delta D - d_{\text{T}}) = -7.6 \cdot 10^{-3}$ over 300 trials
Theorem 5.5.2(b)	$\max(\Delta Q_\alpha^{\text{P}} - \text{bound}) = -5.0 \cdot 10^{-3}$ over 900 cases, $\alpha \in \{0.2, 0.5, 0.8\}$, random ranks
Theorem 5.7.8	$\max \Delta R /\text{bound} = 0.016$ over 300 random $(\Phi, \Phi', \Psi, \rho, \sigma)$, $\alpha \in \{1.3, 2, 3\}$
Proposition 5.7.14	for 20 random Pauli pairs the maximally entangled input minimises $Q_{1/2}^{\text{HT}}$ against 200 random pure inputs (smallest gap $1.7 \cdot 10^{-3}$)
Theorem 5.7.13	12 random qubit channel pairs, f'' with 5 atoms: the common-input SDP and the dual $\lambda_{\max}$ program agree to 7 digits; at the optimum $\ \Gamma_f(\theta) - D\theta\ \leq 1.6 \cdot 10^{-10}$; the iteration $\theta \mapsto \Gamma_f(\theta)/\text{Tr } \Gamma_f(\theta)$ from $\mathbf{1}/2$ reaches D within $3 \cdot 10^{-5}$ in 3,000 steps; in every case maximally entangled $< D < \text{middle bound} < \text{reverse test}$, e.g. $0.604 < 1.038 < 1.100 < 1.270$
Lemma 5.7.10	ratio/ $c_W \geq 1.09$ over 401 random cases with $\dim \mathcal{R} = 3$ and 18 adversarial searches, $\alpha \in \{0.2, 0.5, 0.8\}$; two-fold telescoping ratio/ $c_W^2 \geq 1.25$
Example 5.7.12	one-use ratio 0.47, 0.20, 0.074, 0.024, 0.0025 for $\delta = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-6}$ ($\alpha = \frac{1}{2}$, $\varepsilon = 0.05$), below the bound $(2/\varepsilon)^\alpha \delta^{1-\alpha}$

5.10 Open problems

1. Find the garbling that maps the reverse test to the testing-region pair, which exists by Theorem 5.3.2. Is it the rotation of the Helstrom projections in Theorem 5.3.1(b)?
2. Is $D_{\alpha}^{\text{P,reg}}(\cdot\|\Psi)$, $\alpha < 1$, continuous at every channel when $J(\Psi)$ is invertible? Theorem 5.7.11 reduces this to perturbations that rotate the support of a singular $J(\Phi)$ by a non-product unitary of $\mathcal{X} \otimes \mathcal{Y}$. Example 5.7.12 shows that any proof must use the n -fold structure, not a one-use bound. A related question is to remove the condition number from the constants of Theorems 5.7.4 and 5.7.8.
3. Does the iteration $\theta \mapsto \Gamma_f(\theta)/\text{Tr } \Gamma_f(\theta)$ of Theorem 5.7.13 always converge to an optimal input? In the linear case of the reverse test, the Cauchy–Schwarz inequality shows that the objective increases along the iteration. In all our tests the iteration converged to the optimum. For which pairs beyond the covariant ones is the maximally entangled input optimal, that is, when is the bracket of Theorem 5.7.13(c) scalar?

Chapter 6

Super-multiplicativity of hockey-stick Rényi quasi-entropies

Summary. The testing-region (Hirche–Tomamichel) Rényi quasi-entropies Q_α^{HT} of a pair of quantum states are the Mellin moments of its hockey-stick profile $e(\gamma) = \text{Tr}(\rho - \gamma\sigma)_+$. They are not multiplicative under tensor products, and we study the conjectured inequality $Q_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq Q_\alpha^{\text{HT}}(\rho \| \sigma) Q_\alpha^{\text{HT}}(\rho' \| \sigma')$. We prove it at $\alpha = 2$, at every integer order against pure partners, for every $\alpha > 1$ when one factor is measured, to first order as $\alpha \rightarrow \infty$, and for tensor powers with enough copies. At $\alpha = 3$ we prove it for sign-free pairs and whenever one reference state has at most two distinct eigenvalues, by writing the defect as an integral of Lieb–Thirring gaps, and a cycle inequality $|\text{Tr} ABC|^3 \leq \text{Tr}(A^2 B) \text{Tr}(B^2 C) \text{Tr}(C^2 A)$ for positive matrices gives further three-eigenvalue cases. At integer orders the conjecture is a correlation inequality for a uniform point of a simplex, which a Chebyshev inequality for Schur-convex functions settles for pure and sign-free pairs. In general it is the family of moment inequalities of a convex-order domination between two Frenkel measures. We also give a layer-cake proof of $D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$ for $\alpha \geq 2$, and show that one-way LOCC tests never beat the classical product of the shadows.

6.1 Introduction

For states ρ, σ on a finite-dimensional Hilbert space the hockey-stick divergences $E_\gamma(\rho \| \sigma) = \text{Tr}(\rho - \gamma\sigma)_+$ describe the testing region of the pair, and integrating them against f'' gives the testing-region f -divergences D_f^{HT} [63]. Their Rényi members are given by the quasi-entropies Q_α^{HT} , the Mellin moments of the hockey-stick profile (Section 6.2). Unlike the Petz and sandwiched quasi-entropies, Q_α^{HT} is not multiplicative under tensor products. This chapter is about the inequality

$$Q_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq Q_\alpha^{\text{HT}}(\rho \| \sigma) Q_\alpha^{\text{HT}}(\rho' \| \sigma'), \quad \alpha \in (0, 1) \cup (1, \infty),$$

which Conjecture 6.3.5 asserts for all pairs of states. For $\alpha > 1$ it says that D_α^{HT} is superadditive, and for $\alpha < 1$ that it is subadditive; equality holds when one of the two pairs commutes.

Main results We prove Conjecture 6.3.5 in the following cases. The testing-region quasi-entropy is super-multiplicative at $\alpha = 2$, where it is $1 + \chi_{\text{BKM}}^2$ and the proof reduces to a Gaussian-mixture inequality for the kernel $\frac{u/2}{\sinh(u/2)}$ (Theorem 6.4.1). At integer orders the trace formula turns the conjecture into a correlation inequality for a uniformly distributed

point of a simplex (Proposition 6.5.1). A Chebyshev-type inequality for Schur-convex functions (Lemma 6.5.2) then proves the conjecture at every integer order for pure states, at $\alpha = 3$ for every two sign-free pairs, and again at $\alpha = 2$ for all pairs (Theorem 6.5.3). Sign-free pairs include all qubit pairs, all pairs whose σ has two distinct eigenvalues, and all pure states. Integrating a pure partner out of the trace formula and applying the Lieb–Thirring inequality to mixtures of resolvents proves the conjecture against every pure partner at every integer order (Theorem 6.7.1), with no condition on the other pair. For pure states the general case is a classical convolution inequality, a Rényi-entropy submodularity for sums (Remark 6.4.4). The testing region of a product dominates that of the shadow tensored with any measurement of the second factor, so $D_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq D_\alpha^{\text{HT}}(\rho \| \sigma) + D_\alpha^{\text{M}}(\rho' \| \sigma')$ for every α (Theorem 6.4.2). The quasi-entropy is also super-multiplicative to first order as $\alpha \rightarrow \infty$ (Theorem 6.4.3).

At $\alpha = 3$ a second mechanism removes the sign condition. A two-time integral identity (Proposition 6.6.1) writes the defect against a partner whose σ' has two eigenvalues as an integral of Lieb–Thirring gaps $\text{Tr}[A^3 X^3] - \text{Tr}[(AX)^3]$ (Theorem 6.6.2), so the conjecture holds at $\alpha = 3$ whenever σ or σ' has at most two distinct eigenvalues, for all states and with the expected equality cases (Corollary 6.6.3). The same identity proves the conjecture at $\alpha = 3$ to second order in σ' around $\sigma' \propto \mathbf{1}$, for every ρ' (Proposition 6.6.5), gives $Q_3^{\text{HT}} \leq \text{Tr}[\rho \sigma^{-\gamma} \rho \sigma^{-\gamma} \rho \sigma^{2\gamma-2}]$ for every real γ , with the sandwiched quasi-entropy at $\gamma = \frac{2}{3}$ (Corollary 6.6.7), and reduces the open case at $\alpha = 3$ to singular ρ, ρ' whose reference states both have at least three distinct eigenvalues (Corollary 6.6.6). The Lieb–Thirring gap at exponent three is a rooted sum of Schur-product squares with the kernel $\frac{(u-z)(v-z)(z+2u)(z+2v)}{z+u+v}$ (Proposition 6.7.3), which proves inequality (6.5.1) (Corollary 6.7.4). The cycle inequality $|\text{Tr} ABC|^3 \leq \text{Tr}(A^2 B) \text{Tr}(B^2 C) \text{Tr}(C^2 A)$ for positive matrices (Theorem 6.7.5) lets a pure profile be fitted under the two-letter weights of any state, which gives the conjecture at $\alpha = 3$ against every partner with three eigenvalues and a nonnegative block 3-cycle (Corollary 6.7.7). A splitting of the defect into two-letter and three-letter parts proves it when both block 3-cycles are negative and not too strongly correlated (Theorem 6.7.8, Corollary 6.7.9). What remains for three-eigenvalue reference states is one inequality for the kernel K_3 on twisted pure data (Conjecture 6.7.10, Proposition 6.7.11). For general pairs at $\alpha = 3$ the conjecture follows from the nonnegativity of the covariances D_k of polynomial traces $\text{Tr}[A^i \rho A^j \rho A^l \rho]$ over compositions $i+j+l = k$ (Theorem 6.7.12), which holds at $k = 2$; at $k = 3$, $D_3 \geq 0$ follows from a single-system chirality inequality (Proposition 6.7.13).

The conjecture itself is the family of moment inequalities of one statement: the *Frenkel measure* $\nu(dv) = (e(v) - (1-v)_+)v^{-2}dv$, whose mass and first moment are the two relative entropies and whose moments are the Rényi quasi-entropies, should dominate, in the convex order, the Frenkel measure of the classical product of the two shadows (Theorem 6.8.2). For qubit pairs this reduces to a two-sign-change property of the profile difference, which holds in every test, and for tensor powers the conjecture holds once there are enough copies (Proposition 6.8.3). For every two non-commuting pairs the profile of the product crosses that of the classical product of the shadows at least twice, so a pair and its shadow cannot be compared through positive maps in either direction (Proposition 6.9.1). One-way LOCC tests never beat the shadow product, and numerically neither do separable or PPT tests (Proposition 6.9.2, Conjecture 6.9.3). The conjecture is therefore a statement about entangled tests. As a by-product, $D_2^{\text{HT}} < \tilde{D}_2$ for every non-commuting pair (Corollary 6.9.4).

On the Fourier side, Q_2^{HT} is the average of a modular correlation function over a logistic clock, and Theorem 6.4.1 says that synchronising the clocks of two factors increases it (Proposition 6.10.1). Dephasing each factor by a modular heat flow and letting the flows converge to the pinching reduces the conjecture to a local supermodularity condition at

product points (Theorem 6.10.2). For two qubit pairs at $\alpha = 3$ the conjecture becomes a single explicit inequality for the logarithmic mean (Proposition 6.10.3). This is the case of the simplex theorem in which both pairs have two eigenvalues (Corollary 6.10.4), and a second proof uses the variation-diminishing property of the Cauchy kernel.

We give an elementary proof of $D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$ for all real $\alpha \geq 2$ (Theorem 6.3.2). The full range $\alpha > 1$ is due to [84], after the integer case of [13]. With linear-programming duality it yields an operator layer-cake inequality (Corollary 6.3.4). Section 6.3 recalls from Section 5.6 the explicit single-shot sandwiches $D_\alpha^{\text{P}} \leq D_\alpha^{\text{HT}} \leq D_\alpha^{\text{P}} + \frac{\log 2}{1-\alpha}$ for $\alpha < 1$ and $|D_\alpha^{\text{HT}} - \tilde{D}_\alpha| \leq \frac{\alpha}{\alpha-1} \log v(\sigma)$ for $\alpha > 1$, which give the regularisation theorem of [63] with rates.

Related work The testing-region divergences were introduced by Hirche and Tomamichel [63] and are also called hockey-stick f -divergences. Beigi, Hirche and Tomamichel [13] and Liu, Hirche and Cheng [84] compared their Rényi members with the Petz and sandwiched families. Hiai, Mosonyi and Tomamichel [58] have recently developed them in von Neumann algebras: a Neyman–Pearson representation, the detection of reversibility, the regularisation of the Rényi members to the Petz divergences below order one and to the sandwiched divergences above it, and partial results on when different quantum f -divergences coincide.

Chapter 5 studies the testing region of a single pair: the structure of its shadow, continuity, the modular (Petz) shadow, and channel divergences. This chapter turns to tensor products. The reverse test and the modular shadow factorise under tensor products, but the testing region does not, and the question is what replaces multiplicativity.

Section 6.2 recalls the definitions and results of Chapter 5 that we use. Section 6.3 states the conjecture, together with the comparison inequalities for tensor powers and the inequality $D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$. Sections 6.4–6.7 prove the cases of the conjecture listed above. Section 6.8 identifies the conjecture with a convex-order statement for Frenkel measures, Section 6.9 shows why positive maps and restricted tests cannot prove it, and Section 6.10 treats modular clocks, dephasing flows and two qubit pairs. Section 6.11 summarises the numerical checks, and Section 6.12 lists open problems.

The quantitative claims were checked numerically, and the conjectures were tested; the checks are summarised in Section 6.11.

6.2 Preliminaries from Chapter 5

We use the notation and results of Section 5.2, recalled here without proofs. States ρ, σ are density operators on a finite-dimensional Hilbert space $\mathcal{H}$, $\omega = \rho + \sigma$, Π_X is the support projection of a positive operator X , and $\Pi_X^\perp = \mathbf{1} - \Pi_X$. Logarithms are natural.

For $A, B \geq 0$ and $\gamma \geq 0$, $E_\gamma(A\|B) = \text{Tr}(A - \gamma B)_+ = \max_{0 \leq M \leq \mathbf{1}} \text{Tr } M(A - \gamma B)$ is the hockey-stick divergence. The profile $e(\gamma) = E_\gamma(\rho\|\sigma)$, the error profile $\eta(\gamma) = 1 - e(\gamma)$ and the boundary masses $m_0 = \text{Tr } \Pi_\rho^\perp \sigma$ and $m_\infty = \text{Tr } \Pi_\sigma^\perp \rho$ are as in Definition 5.2.1, and Lemma 5.2.2 collects their properties: e is convex and nonincreasing with $(1-\gamma)_+ \leq e(\gamma) \leq 1$, η is concave and nondecreasing with $\eta(\gamma)/\gamma$ nonincreasing (part (a)), $e(\infty) = m_\infty$ and $e'(0^+) = -(1-m_0)$ (part (b)), and $E_\gamma(\sigma\|\rho) = \gamma e(1/\gamma) - \gamma + 1$ (part (c)). The testing-region (Hirche–Tomamichel) f -divergence D_f^{HT} of a convex f with $f(1) = 0$ is Definition 5.2.3 [63]; for $\alpha \in (0, 1) \cup (1, \infty)$ and $f_\alpha(x) = \frac{x^\alpha - 1}{\alpha - 1}$ we write $Q_\alpha^{\text{HT}} = 1 + (\alpha - 1)D_{f_\alpha}^{\text{HT}}$ and $D_\alpha^{\text{HT}} = \frac{1}{\alpha - 1} \log Q_\alpha^{\text{HT}}$. As D_f^{HT} depends only on f'' , we normalise $f'(1) = 0$ when convenient, and by Frenkel’s formula [46], $D_{x \log x}^{\text{HT}}$ is the Umegaki relative entropy $D(\rho\|\sigma) = \text{Tr } \rho(\log \rho - \log \sigma)$. The testing-region shadow $\mu = e''$, completed by the atoms m_0 at 0 and m_∞ at ∞ , is

Definition 5.2.4; for commuting ρ, σ with eigenvalues p_x, q_x it is the law of the likelihood ratio p/q under q on $\{q > 0\}$.

We use two formulas from Section 5.2 directly. The representation formula of Proposition 5.2.5 states that, for every convex f with $f(1) = 0$,

$$D_f^{\text{HT}}(\rho\|\sigma) = \int_0^\infty (e(\gamma) - (1 - \gamma)_+) f''(d\gamma) = \int_{(0, \infty)} f d\mu + f(0^+) m_0 + f'(\infty) m_\infty,$$

with $f'(\infty) = \lim_{x \rightarrow \infty} f(x)/x$ and the convention $0 \cdot \infty = 0$, and that $\mu((0, \infty)) = 1 - m_0$ and $\int \gamma d\mu = 1 - m_\infty$. The Mellin formulas of Proposition 5.2.6 state that, for $\alpha \in (0, 1)$,

$$Q_\alpha^{\text{HT}}(\rho\|\sigma) = \int_{(0, \infty)} \gamma^\alpha d\mu(\gamma) = \alpha(1 - \alpha) \int_0^\infty \gamma^{\alpha-2} \eta(\gamma) d\gamma,$$

and that, for $\alpha > 1$, $Q_\alpha^{\text{HT}} = +\infty$ if $m_\infty > 0$, and otherwise

$$Q_\alpha^{\text{HT}}(\rho\|\sigma) = \int_{(0, \infty)} \gamma^\alpha d\mu(\gamma) = \alpha(\alpha - 1) \int_0^\infty \gamma^{\alpha-2} e(\gamma) d\gamma.$$

The reverse test and the modular shadow Two further shadows of the pair factorise under tensor products; both are defined in Section 5.2. Let $T = \omega^{-1/2} \rho \omega^{-1/2}$ on $\Pi_\omega \mathcal{H}$. The *reverse test* is a classical pair (p_T, q_T) on $\text{spec } T$ that is mapped to (ρ, σ) by a preparation. For operator convex f , $D_f^T := D_f(p_T\|q_T)$ is the maximal f -divergence of (ρ, σ) [91], equal to $\text{Tr } \omega^{1/2} h_f(T) \omega^{1/2}$ with $h_f(t) = (1 - t)f(t/(1 - t))$ (Theorem 4.2.15 and Lemma 4.7.4 in Chapter 4), and the profile of the reverse test is $E_\gamma^T(\rho\|\sigma) := E_\gamma(p_T\|q_T) = \text{Tr } \omega^{1/2} (T - \gamma(\mathbf{1} - T))_+ \omega^{1/2}$. The *modular shadow* is the Nussbaum–Szkoła pair, whose f -divergences are the Petz f -divergences D_f^{P} [61, 99], and D_f^{M} is the measured divergence. These quantities satisfy the Blackwell chain $D_f^{\text{M}} \leq D_f^{\text{HT}} \leq D_f^{\text{T}}$ (Proposition 5.2.7).

6.3 Tensor products and the conjecture

Since T and the Nussbaum–Szkoła pair factorise, the reverse test and the modular shadow are multiplicative: the shadow of $(\rho \otimes \rho', \sigma \otimes \sigma')$ is the multiplicative convolution of the two shadows. The testing region does not factorise, and the Hirsch–Tomamichel Rényi divergences are not additive [63]. Section 5.6 therefore compares tensor powers with the factorising shadows, the modular shadow below order one and the pinched pair above it. We write η_n for the error profile of $(\rho^{\otimes n}, \sigma^{\otimes n})$, $v(\sigma)$ for the number of distinct eigenvalues of σ , d for the dimension, and $\mathcal{P}_\sigma$ for the pinching in the eigenbasis of σ . For $\alpha \in (0, 1)$, Proposition 5.6.1 (order below one) gives $\frac{1}{2} Q_\alpha^{\text{P}} \leq Q_\alpha^{\text{HT}} \leq Q_\alpha^{\text{P}}$. For $\alpha > 1$ and $\Pi_\rho \leq \Pi_\sigma$, Proposition 5.6.2 (order above one) places both $Q_\alpha^{\text{HT}}(\rho\|\sigma)$ and the sandwiched quasi-entropy $\tilde{Q}_\alpha(\rho\|\sigma)$ between the classical quantity $Q_\alpha(\mathcal{P}_\sigma \rho\|\sigma) = \text{Tr}(\mathcal{P}_\sigma \rho)^\alpha \sigma^{1-\alpha}$ and $v(\sigma)^\alpha Q_\alpha(\mathcal{P}_\sigma \rho\|\sigma)$. We use the resulting estimates: for $\alpha \in (0, 1)$,

$$\begin{aligned} D_\alpha^{\text{P}}(\rho\|\sigma) &\leq D_\alpha^{\text{HT}}(\rho\|\sigma) \leq D_\alpha^{\text{P}}(\rho\|\sigma) + \frac{\log 2}{1 - \alpha}, \\ \left| \frac{1}{n} D_\alpha^{\text{HT}}(\rho^{\otimes n}\|\sigma^{\otimes n}) - D_\alpha^{\text{P}}(\rho\|\sigma) \right| &\leq \frac{\log 2}{n(1 - \alpha)}, \end{aligned}$$

and for $\alpha > 1$ and $\Pi_\rho \leq \Pi_\sigma$,

$$\begin{aligned} |D_\alpha^{\text{HT}}(\rho\|\sigma) - \tilde{D}_\alpha(\rho\|\sigma)| &\leq \frac{\alpha}{\alpha - 1} \log v(\sigma), \\ \left| \frac{1}{n} D_\alpha^{\text{HT}}(\rho^{\otimes n}\|\sigma^{\otimes n}) - \tilde{D}_\alpha(\rho\|\sigma) \right| &\leq \frac{\alpha(d - 1)}{\alpha - 1} \frac{\log(n + 1)}{n}. \end{aligned}$$

Proposition 6.3.1 (Chernoff-type sandwich). *For $\alpha \in (0, 1)$,*

$$1 - E_1^T(\rho\|\sigma) \leq \eta(1) \leq Q_\alpha^{\text{HT}}(\rho\|\sigma) \leq \eta(1)^\alpha + \eta(1)^{1-\alpha} - \eta(1).$$

Proof. The first inequality is Theorem 5.3.2(a) at $\gamma = 1$. By Lemma 5.2.2(a), $\eta(\gamma) \geq \eta(1) \min\{\gamma, 1\}$ and $\eta(\gamma) \leq \min\{\min\{\gamma, 1\}, \eta(1) \max\{\gamma, 1\}\}$. Insert these into the Mellin formula of Proposition 5.2.6: $\alpha(1-\alpha) \int_0^\infty \gamma^{\alpha-2} \min\{\gamma, 1\} d\gamma = 1$ gives the lower bound, and splitting at $\eta(1)$ and $1/\eta(1)$ gives the upper bound. $\square$

Since $Q_\alpha^{\text{HT}} \leq Q_\alpha^{\text{P}}$ [13], the middle inequality sharpens the bound $\eta(1) \leq \text{Tr} \rho^\alpha \sigma^{1-\alpha}$ of Audenaert et al. [11]. The first inequality says that the reverse test always gives a lower bound on the Bayesian error. This is the mechanism behind every converse built from Kubo–Ando means, including the geometric Chernoff converse of Theorem 7.4.1 in Chapter 7.

The inequality $D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$ It was proved in [13] for integers $\alpha \geq 2$ and conjectured there for all $\alpha > 1$. Liu, Hirche and Cheng [84] proved it for all $\alpha > 1$ through the trace formula $Q_\alpha^{\text{HT}}(\rho\|\sigma) = (\alpha-1) \int_0^\infty \text{Tr}((\sigma+t)^{-1/2} \rho (\sigma+t)^{-1/2})^\alpha dt$. Proposition 5.6.2 bounds the gap by $\frac{\alpha}{\alpha-1} \log v(\sigma)$. We give a short independent proof for every real $\alpha \geq 2$. It uses only the Helstrom projections, and it shows where the exponent $\alpha = 2$ is critical.

Theorem 6.3.2. *For every $\alpha \geq 2$ and all states, $D_\alpha^{\text{HT}}(\rho\|\sigma) \leq \tilde{D}_\alpha(\rho\|\sigma)$.*

Proof. We may assume $\Pi_\rho \leq \Pi_\sigma$ and work on $\Pi_\sigma \mathcal{H}$, where σ is invertible. Put $\kappa = \frac{\alpha}{\alpha-1} \in (1, 2]$, $c = \frac{1}{2\kappa}$ and $X = \sigma^{-c} \rho \sigma^{-c}$, so that $\tilde{Q}_\alpha = \text{Tr} X^\alpha$.

Step 1 (a variational lower bound). For every $H \geq 0$, Hölder's and Young's inequalities give $\text{Tr} \rho H = \text{Tr} X(\sigma^c H \sigma^c) \leq \|X\|_\alpha \|\sigma^c H \sigma^c\|_\kappa \leq \frac{1}{\alpha} \text{Tr} X^\alpha + \frac{1}{\kappa} \text{Tr}(\sigma^c H \sigma^c)^\kappa$, that is,

$$\tilde{Q}_\alpha(\rho\|\sigma) \geq \alpha \text{Tr} \rho H - (\alpha-1) \text{Tr}(\sigma^c H \sigma^c)^\kappa.$$

Step 2 (the Helstrom layer cake). Let P_γ be the Helstrom projections and reparametrise by $u = \gamma^{\alpha-1}$, writing $M(u) = P_{u^{\kappa-1}}$. Choose $H = \int_0^\infty M(u) du = (\alpha-1) \int_0^\infty \gamma^{\alpha-2} P_\gamma d\gamma$, which is bounded because $P_\gamma = 0$ beyond the largest generalised eigenvalue. Then $\alpha \text{Tr} \rho H = \alpha(\alpha-1) \int \gamma^{\alpha-2} \text{Tr} \rho P_\gamma d\gamma$.

Step 3 (the derivative bound). Put $H_U = \int_0^U M(u) du$, $K_U = \sigma^c H_U \sigma^c$, $\Phi(U) = \text{Tr} K_U^\kappa$ and $\Psi(U) = \kappa \int_0^U u^{\kappa-1} \text{Tr} \sigma M(u) du$. Both vanish at $U = 0$ and are absolutely continuous, with $\Phi'(U) = \kappa \text{Tr} [\sigma^c K_U^{\kappa-1} \sigma^c M(U)]$. Since $0 \leq M \leq \mathbf{1}$, $H_U \leq U \mathbf{1}$, so $K_U \leq U \sigma^{1/\kappa}$. As $0 < \kappa-1 \leq 1$, the function $x^{\kappa-1}$ is operator monotone, so $K_U^{\kappa-1} \leq U^{\kappa-1} \sigma^{(\kappa-1)/\kappa}$ and

$$\sigma^c K_U^{\kappa-1} \sigma^c \leq U^{\kappa-1} \sigma^{2c+(\kappa-1)/\kappa} = U^{\kappa-1} \sigma.$$

Hence $\Phi'(U) \leq \kappa U^{\kappa-1} \text{Tr} \sigma M(U) = \Psi'(U)$, and

$$\text{Tr}(\sigma^c H \sigma^c)^\kappa = \Phi(\infty) \leq \Psi(\infty) = \alpha \int_0^\infty \gamma^{\alpha-1} \text{Tr} \sigma P_\gamma d\gamma.$$

Step 4. Combining the steps with the Mellin formula of Proposition 5.2.6,

$$\begin{aligned} \tilde{Q}_\alpha &\geq \alpha(\alpha-1) \int_0^\infty \gamma^{\alpha-2} \text{Tr} P_\gamma(\rho - \gamma\sigma) d\gamma \\ &= \alpha(\alpha-1) \int_0^\infty \gamma^{\alpha-2} E_\gamma(\rho\|\sigma) d\gamma = Q_\alpha^{\text{HT}}(\rho\|\sigma). \end{aligned}$$

$\square$

Remark 6.3.3 (The greedy argument and the range $(1, 2)$). Step 3 holds for every family $0 \leq M(u) \leq \mathbf{1}$, not only for Helstrom projections. In fact, minimising $\int \text{Tr } \sigma M(u) d(u^\kappa)$ over families with $\int M = H$ is a linear program whose dual objective is $\text{Tr } \Lambda H - C_\alpha Q_\alpha^{\text{HT}}(\Lambda \| \sigma)$. So the “operator bathtub inequality”

$$\text{Tr} \left(\sigma^{\frac{1}{2\kappa}} \left(\int_0^\infty M(u) du \right) \sigma^{\frac{1}{2\kappa}} \right)^\kappa \leq \int_0^\infty \text{Tr } \sigma M(u) d(u^\kappa) \quad (0 \leq M(u) \leq \mathbf{1})$$

for a fixed κ is equivalent to the inequality $Q_\alpha^{\text{HT}} \leq \tilde{Q}_\alpha$ at $\alpha = \kappa/(\kappa - 1)$ for all pairs $(\cdot, \sigma)$ (Corollary 6.3.4). For $\kappa > 2$ the pointwise inequality of Step 3 would require $K \leq A \Rightarrow \text{Tr } K^{\kappa-1} N \leq \text{Tr } A^{\kappa-1} N$ for all $0 \leq N \leq A$. That is operator monotonicity of $x^{\kappa-1}$, which fails, and we found violations numerically. The integrated inequality nevertheless held in all our tests, including 1,200 random families with $\alpha \in \{1.2, 1.5, 1.8\}$. For $\alpha \in (1, 2)$ the inequality holds by [84], but a direct operator proof of the bathtub inequality for $\kappa > 2$ needs a non-greedy argument.

Corollary 6.3.4 (An operator layer-cake inequality). *For every $\sigma \geq 0$, every $\kappa \geq 1$ and every measurable family $0 \leq M(u) \leq \mathbf{1}$ with $\int_0^\infty M(u) du$ finite,*

$$\text{Tr} \left(\sigma^{\frac{1}{2\kappa}} \left(\int_0^\infty M(u) du \right) \sigma^{\frac{1}{2\kappa}} \right)^\kappa \leq \int_0^\infty \text{Tr } \sigma M(u) d(u^\kappa).$$

For $\kappa \leq 2$ this is Step 3 of Theorem 6.3.2. For $\kappa > 2$ it follows from [84].

Proof. Put $\alpha = \kappa/(\kappa - 1)$, $H = \int M$, and $C = \kappa^{1-\alpha}/\alpha$. For every Hermitian Λ , the inequality $\text{Tr } XM \geq -\text{Tr } X_-$ for $0 \leq M \leq \mathbf{1}$ gives

$$\int_0^\infty \text{Tr } \sigma M(u) d(u^\kappa) \geq \text{Tr } \Lambda H - \int_0^\infty \text{Tr} (\Lambda - \kappa u^{\kappa-1} \sigma)_+ du = \text{Tr } \Lambda H - C Q_\alpha^{\text{HT}}(\Lambda \| \sigma),$$

by the substitution $\gamma = \kappa u^{\kappa-1}$ and the Mellin formula (both sides are homogeneous, so the formula extends to unnormalised Λ). By [84], $Q_\alpha^{\text{HT}} \leq \tilde{Q}_\alpha$. Writing $Y = \sigma^{-1/(2\kappa)} \Lambda \sigma^{-1/(2\kappa)}$ and $K = \sigma^{1/(2\kappa)} H \sigma^{1/(2\kappa)}$, we get $\text{Tr } \Lambda H - C \tilde{Q}_\alpha(\Lambda \| \sigma) = \text{Tr } Y K - C \text{Tr } Y^\alpha$. Its supremum over $Y \geq 0$ is $\text{Tr } K^\kappa$, because $\sup_y (yk - Cy^\alpha) = k^\kappa$ for this C . $\square$

Super-multiplicativity would give a third route to these inequalities, and it is a structural statement about how testing regions tensorize.

Conjecture 6.3.5 (Super-multiplicativity). *For all $\alpha \in (0, 1) \cup (1, \infty)$ and all pairs, $Q_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq Q_\alpha^{\text{HT}}(\rho \| \sigma) Q_\alpha^{\text{HT}}(\rho' \| \sigma')$.*

Below we prove the conjecture at $\alpha = 2$, at every integer order $\alpha = n$ for pure states, at every integer order against pure partners (Theorem 6.7.1), at $\alpha = 3$ for all sign-free pairs (Theorem 6.5.3), for all pairs in which one of the two reference states has at most two distinct eigenvalues (Corollary 6.6.3), against partners with three eigenvalues and a nonnegative 3-cycle (Corollary 6.7.7) and for weakly correlated negative 3-cycles on both sides (Corollary 6.7.9), for one measured factor, to first order at $\alpha = \infty$, and for tensor powers with enough copies, and we show that it is the family of moment inequalities of a convex-order statement about one measure (Theorem 6.8.2).

Equality holds when ρ' and σ' commute. In that case $E_\gamma(\rho \otimes \rho' \| \sigma \otimes \sigma') = \sum_y a_y e(\gamma b_y / a_y)$ and the Mellin transform factorises. For $\alpha > 1$ the conjecture says that D_α^{HT} is superadditive, and for $\alpha < 1$ that it is subadditive.

Theorem 6.3.6 (Reduction). *Fix $\alpha > 1$. If $Q_\alpha^{\text{HT}}(\rho^{\otimes n} \|\sigma^{\otimes n}) \geq Q_\alpha^{\text{HT}}(\rho \|\sigma)^n$ for infinitely many n , then $D_\alpha^{\text{HT}}(\rho \|\sigma) \leq \tilde{D}_\alpha(\rho \|\sigma)$. For $\alpha < 1$ the same hypothesis gives $D_\alpha^{\text{HT}} \geq D_\alpha^{\text{P}}$. So Conjecture 6.3.5 would give independent proofs of [84] and of [13, (1.10)], using only pinching and the inequalities of [11].*

Proof. For $\alpha > 1$, Proposition 5.6.2 applied to n copies gives $Q_\alpha^{\text{HT}}(\rho \|\sigma)^n \leq (n+1)^{\alpha(d-1)} \tilde{Q}_\alpha(\rho \|\sigma)^n$. Take n -th roots and let $n \rightarrow \infty$. For $\alpha < 1$, choose $0 < s_1 < \alpha < s_2 < 1$. The inequality of [11] gives $\eta_n(\gamma) \leq \gamma^{1-s}(Q_s^{\text{P}})^n$ for $s = s_1, s_2$. Using the bound with s_1 below $\gamma_* = (Q_{s_2}^{\text{P}}/Q_{s_1}^{\text{P}})^{n/(s_2-s_1)}$ and the bound with s_2 above it, the Mellin integral gives

$$Q_\alpha^{\text{HT}}(\rho^{\otimes n} \|\sigma^{\otimes n}) \leq \alpha(1-\alpha) \left(\frac{1}{\alpha-s_1} + \frac{1}{s_2-\alpha} \right) ((Q_{s_1}^{\text{P}})^\theta (Q_{s_2}^{\text{P}})^{1-\theta})^n, \quad \theta = \frac{s_2-\alpha}{s_2-s_1}.$$

Combine this with the hypothesis $Q_\alpha^{\text{HT}}(\rho \|\sigma)^n \leq Q_\alpha^{\text{HT}}(\rho^{\otimes n} \|\sigma^{\otimes n})$, take n -th roots, and let $s_1, s_2 \rightarrow \alpha$, using the continuity of $s \mapsto Q_s^{\text{P}}$. This gives $Q_\alpha^{\text{HT}} \leq Q_\alpha^{\text{P}}$. $\square$

6.4 Super-multiplicativity: three cases that can be proved

We prove Conjecture 6.3.5 at $\alpha = 2$, prove a one-sided version for every f in which one factor is replaced by a measurement, and prove it to first order as $\alpha \rightarrow \infty$. Throughout, $L(a, b) = \frac{a-b}{\log a - \log b}$ is the logarithmic mean, with $L(a, a) = a$.

Theorem 6.4.1 (Super-multiplicativity at $\alpha = 2$). *For all pairs of states,*

$$Q_2^{\text{HT}}(\rho \otimes \rho' \|\sigma \otimes \sigma') \geq Q_2^{\text{HT}}(\rho \|\sigma) Q_2^{\text{HT}}(\rho' \|\sigma'),$$

so D_2^{HT} is superadditive. When the right-hand side is finite, equality holds if and only if $[\rho, \sigma] = 0$ or $[\rho', \sigma'] = 0$.

Proof. If $\Pi_\rho \not\leq \Pi_\sigma$ or $\Pi_{\rho'} \not\leq \Pi_{\sigma'}$, both sides are infinite, since $Q_2^{\text{HT}} \geq 1$. Otherwise we work on the supports. Write $\sigma = \sum_j e^{x_j} |j\rangle\langle j|$ and $\rho_{jk} = \langle j|\rho|k\rangle$. The trace formula at $\alpha = 2$ [13, (2.11)] gives

$$Q_2^{\text{HT}}(\rho \|\sigma) = \int_0^\infty \text{Tr} [\rho(\sigma+t)^{-1} \rho(\sigma+t)^{-1}] dt = \sum_{jk} \frac{|\rho_{jk}|^2}{L(e^{x_j}, e^{x_k})} = 1 + \chi_{\text{BKM}}^2(\rho, \sigma),$$

where χ_{BKM}^2 is the Bogoliubov–Kubo–Mori χ^2 -divergence. Since

$$L(e^x, e^y) = e^{(x+y)/2} \frac{\sinh((x-y)/2)}{(x-y)/2},$$

$$Q_2^{\text{HT}}(\rho \|\sigma) = \sum_{jk} w_{jk} k(x_j - x_k),$$

$$w_{jk} = |\rho_{jk}|^2 e^{-(x_j+x_k)/2} = w_{kj} \geq 0, \quad k(u) = \frac{u/2}{\sinh(u/2)}.$$

The product pair has eigenvalues $e^{x_j+x'_l}$ and matrix entries $\rho_{jk}\rho'_{lm}$. With $u = x_j - x_k$ and $v = x'_l - x'_m$, and using $w'_{lm} = w'_{ml}$,

$$Q_2^{\text{HT}}(\rho \otimes \rho' \|\sigma \otimes \sigma') = \sum_{jklm} w_{jk} w'_{lm} k(u+v) = \sum_{jklm} w_{jk} w'_{lm} \frac{1}{2} (k(u+v) + k(u-v)).$$

Euler's product $\frac{\sinh z}{z} = \prod_{n \geq 1} (1 + \frac{z^2}{\pi^2 n^2})$ gives $k(u) = \prod_n (1 + \frac{u^2}{4\pi^2 n^2})^{-1} = \mathbb{E} e^{-u^2 V}$, where $V = \sum_n \frac{Y_n}{4\pi^2 n^2}$ and the Y_n are independent standard exponential variables ($\mathbb{E} V = \frac{1}{24}$). Hence

$$\begin{aligned} \frac{1}{2}(k(u+v) + k(u-v)) &= \mathbb{E}[e^{-(u^2+v^2)V} \cosh(2uvV)] \\ &\geq \mathbb{E}[e^{-u^2 V} e^{-v^2 V}] \geq \mathbb{E} e^{-u^2 V} \mathbb{E} e^{-v^2 V} = k(u)k(v). \end{aligned}$$

The last step is Chebyshev's association inequality for the two nonincreasing functions $V \mapsto e^{-u^2 V}$ and $V \mapsto e^{-v^2 V}$. Summing over j, k, l, m proves the inequality. If neither pair commutes, there is a term with $w_{jk}w'_{lm} > 0$, $u \neq 0$ and $v \neq 0$. For that term both inequalities are strict, because V is not almost surely constant. If one pair commutes, the product shadow factorises (the remark after Conjecture 6.3.5). $\square$

The proof uses the product structure in one place: the symmetrisation $v \mapsto -v$, which pairs the index (l, m) with (m, l) . The pointwise inequality $k(u+v) \geq k(u)k(v)$ is false, since $k(2u)/k(u)^2 = 2 \tanh(u/2)/u < 1$ for $u > 0$.

Theorem 6.4.2 (One factor measured). *Let (P, Q) be the classical pair carried by the shadow of (ρ, σ) , namely $Q = \mu + m_0 \delta_0$ and $P(d\gamma) = \gamma \mu(d\gamma) + m_\infty \delta_\infty$ on $[0, \infty]$. Then $E_\gamma(P\|Q) = e(\gamma)$, and for every measurement $N = \{N_y\}$ on the second factor, with $a_y = \text{Tr } \rho' N_y$, $b_y = \text{Tr } \sigma' N_y$, and every $\gamma \geq 0$,*

$$E_\gamma(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq \sum_y a_y e(\gamma b_y / a_y) = E_\gamma(P \otimes a \| Q \otimes b),$$

where $a e(\gamma b/a)$ is read as its limit when $a = 0$ or $b = 0$. The same holds with both pairs reversed. Consequently, for every convex f ,

$$D_f^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq \sup_N D_f(P \otimes a \| Q \otimes b),$$

and for every $\alpha \in (0, 1) \cup (1, \infty)$,

$$D_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq D_\alpha^{\text{HT}}(\rho \| \sigma) + D_\alpha^{\text{M}}(\rho' \| \sigma').$$

Proof. $E_\gamma(P\|Q) = \int_{(\gamma, \infty)} (\lambda - \gamma) d\mu(\lambda) + m_\infty = e(\gamma)$ by the first identity in the proof of Proposition 5.2.5. Let P_c be the Helstrom projection onto the positive part of $\rho - c\sigma$, with $P_\infty = 0$, and put $M_\gamma = \sum_y P_{\gamma b_y / a_y} \otimes N_y$. Then $0 \leq M_\gamma \leq 1$ and

$$\text{Tr } M_\gamma(\rho \otimes \rho' - \gamma \sigma \otimes \sigma') = \sum_y a_y \text{Tr } P_{c_y}(\rho - c_y \sigma) = \sum_y a_y e(c_y), \quad c_y = \gamma b_y / a_y.$$

On the classical side, $E_\gamma(P \otimes a \| Q \otimes b) = \sum_y a_y E_{c_y}(P \| Q) = \sum_y a_y e(c_y)$. Reversing both pairs gives the second family of hockey sticks. Definition 5.2.3 is monotone in both families, and D_f^{HT} agrees with D_f on classical pairs, so the f -divergence bound follows. For $f = f_\alpha$, classical Rényi divergences are additive and $D_\alpha(P\|Q) = D_\alpha^{\text{HT}}(\rho\|\sigma)$, since by Proposition 5.2.5 the two pairs have the same f -divergences. Taking the supremum over N gives $D_\alpha^{\text{M}}(\rho' \| \sigma')$. $\square$

For $\alpha > 1$ this is the inequality $Q_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq Q_\alpha^{\text{HT}}(\rho \| \sigma) Q_\alpha^{\text{M}}(\rho' \| \sigma')$, the conjecture with the second factor measured. For $\alpha < 1$ it gives $Q_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \leq Q_\alpha^{\text{HT}}(\rho \| \sigma) Q_\alpha^{\text{M}'}(\rho' \| \sigma')$. So, if the conjecture holds, $Q_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma')$ lies between $Q_\alpha^{\text{HT}} Q_\alpha^{\text{HT}'}$ and $Q_\alpha^{\text{HT}} Q_\alpha^{\text{M}'}$ for $\alpha < 1$. Pointwise, the bound of Theorem 6.4.2 combines Helstrom tests of the first factor with a

fixed measurement of the second. To combine Helstrom tests of both factors one would need Helstrom projections that are nested in γ , and they are not.

Near $\alpha = \infty$ the shadow concentrates at its top atom, where a second-order perturbation computation decides the question. Let σ be invertible, and let $\gamma_{\max} = e^{\mathcal{D}_{\max}(\rho\|\sigma)}$ be a simple eigenvalue of $\sigma^{-1/2}\rho\sigma^{-1/2}$. Let v be a unit vector with $\rho v = \gamma_{\max}\sigma v$, and put $H = \rho - \gamma_{\max}\sigma \leq 0$, $m = \langle v|\sigma|v \rangle$, $a = \sigma v - mv \perp v$ and

$$\kappa(\rho\|\sigma) = \frac{2\gamma_{\max}}{m} \langle a|(-H)^+|a \rangle \geq 0,$$

with $^+$ the Moore–Penrose inverse. Then $\kappa = 0$ for commuting pairs.

Theorem 6.4.3 (First order at $\alpha = \infty$). *In this setting, as $\alpha \rightarrow \infty$,*

$$\log Q_{\alpha}^{\text{HT}}(\rho\|\sigma) = \alpha \log \gamma_{\max} + \log m + \frac{\kappa(\rho\|\sigma)}{\alpha + 1} + O(\alpha^{-2}),$$

and for two such pairs $\kappa(\rho \otimes \rho'\|\sigma \otimes \sigma') \geq \kappa(\rho\|\sigma) + \kappa(\rho'\|\sigma')$. Hence

$$\log \frac{Q_{\alpha}^{\text{HT}}(\rho \otimes \rho'\|\sigma \otimes \sigma')}{Q_{\alpha}^{\text{HT}}(\rho\|\sigma) Q_{\alpha}^{\text{HT}}(\rho'\|\sigma')} = \frac{\kappa(\rho \otimes \rho'\|\sigma \otimes \sigma') - \kappa(\rho\|\sigma) - \kappa(\rho'\|\sigma')}{\alpha + 1} + O(\alpha^{-2}).$$

The leading coefficient is nonnegative, and when it is positive the conjecture holds for all sufficiently large α .

Proof. Asymptotics. Let $\gamma_2 < \gamma_{\max}$ be the next generalised eigenvalue. For $u \in (\gamma_2, \gamma_{\max}]$ exactly one eigenvalue $\lambda_1(u)$ of $\rho - u\sigma$ is positive, so $e(u) = \lambda_1(u)$ there. The eigenvalue is simple, so λ_1 is real-analytic near $\gamma_{\max}$, with $\lambda_1(\gamma_{\max}) = 0$, $\lambda_1'(\gamma_{\max}) = -m$ and, by second-order perturbation theory, $\lambda_1''(\gamma_{\max}) = 2\langle a|(-H)^+|a \rangle$, because the other eigenvalues of H are negative. In the Mellin formula $Q_{\alpha}^{\text{HT}} = \alpha(\alpha - 1) \int_0^{\gamma_{\max}} u^{\alpha-2} e(u) du$, the integral over $[0, \gamma_{\max} - \delta]$ is $O(\alpha^2(\gamma_{\max} - \delta)^{\alpha})$. On the rest insert $e(u) = m(\gamma_{\max} - u) + \frac{1}{2}\lambda_1''(\gamma_{\max})(\gamma_{\max} - u)^2 + O((\gamma_{\max} - u)^3)$ and use $\alpha(\alpha - 1) \int_0^{\gamma} u^{\alpha-2}(\gamma - u)^k du = \gamma^{\alpha+k-1} k! \Gamma(\alpha + 1) / \Gamma(\alpha + k)$. This gives $Q_{\alpha}^{\text{HT}} = m\gamma_{\max}^{\alpha} (1 + \kappa/(\alpha + 1) + O(\alpha^{-2}))$.

Superadditivity of κ . Put primes on the quantities of the second pair. The product has top generalised eigenvalue $\gamma_{\max}\gamma'_{\max}$, which is simple, with eigenvector $v \otimes v'$, and $\langle v \otimes v'|\sigma \otimes \sigma'|v \otimes v' \rangle = mm'$. Expanding $\rho = \gamma_{\max}\sigma + H$ and $\rho' = \gamma'_{\max}\sigma' + H'$,

$$-H_{12} := \gamma_{\max}\gamma'_{\max} \sigma \otimes \sigma' - \rho \otimes \rho' = A - H \otimes H', \quad A = \gamma_{\max} \sigma \otimes (-H') + \gamma'_{\max} (-H) \otimes \sigma'.$$

Since $H \otimes H' = (-H) \otimes (-H') \geq 0$, we have $0 \leq -H_{12} \leq A$. Both operators have kernel $\mathbb{C} v \otimes v'$, because $\langle x|A|x \rangle = 0$ forces $x \in (\mathcal{H} \otimes v') \cap (v \otimes \mathcal{H}')$ when $\sigma, \sigma' > 0$. Hence $(-H_{12})^+ \geq A^+$. The vector for the product is

$$a_{12} = (\sigma \otimes \sigma')(v \otimes v') - mm' v \otimes v' = m v \otimes a' + m' a \otimes v' + a \otimes a' \perp v \otimes v'.$$

For z in the range of A , $\langle z|A^+|z \rangle = \sup_x (2 \operatorname{Re} \langle x|z \rangle - \langle x|A|x \rangle)$. Take $x = \gamma_{\max}^{-1} v \otimes b' + \gamma_{\max}'^{-1} b \otimes v'$, with $b = (-H)^+ a$ and $b' = (-H')^+ a'$. Using $Hv = 0$, $(-H)b = a$ and $a \perp v$, we get $A(v \otimes b') = \gamma_{\max} \sigma v \otimes a'$ and $A(b \otimes v') = \gamma_{\max}' a \otimes \sigma' v'$. Therefore $\langle v \otimes b'|a_{12} \rangle = mc'$, $\langle b \otimes v'|a_{12} \rangle = m'c$, $\langle v \otimes b'|A|v \otimes b' \rangle = \gamma_{\max} mc'$, $\langle b \otimes v'|A|b \otimes v' \rangle = \gamma_{\max}' m'c$, and the cross term vanishes, where $c = \langle a|(-H)^+|a \rangle$ and $c' = \langle a'|(-H')^+|a' \rangle$. Hence

$$\langle a_{12}|(-H_{12})^+|a_{12} \rangle \geq \langle a_{12}|A^+|a_{12} \rangle \geq \frac{mc'}{\gamma_{\max}} + \frac{m'c}{\gamma_{\max}'},$$

and multiplying by $2\gamma_{\max}\gamma'_{\max}/(mm')$ gives $\kappa_{12} \geq \kappa + \kappa'$. $\square$

Remark 6.4.4 (Pure states). If $\rho = \psi$ is pure, then $(\sigma + t)^{-1/2}\psi(\sigma + t)^{-1/2}$ has rank one, and the trace formula gives $Q_\alpha^{\text{HT}}(\psi\|\sigma) = (\alpha - 1) \int_0^\infty (\int \frac{\nu(dq)}{q+t})^\alpha dt$ for $\alpha > 1$, where ν is the spectral measure of σ in ψ . For $\psi \otimes \psi'$ this measure is the multiplicative convolution of the two. With $t = e^x$ and $\mu(dy) = e^{(1/\alpha-1)y} (\log_* \nu)(dy)$, one finds $Q_\alpha^{\text{HT}}(\psi\|\sigma) = \|\mu * K_\alpha\|_\alpha^\alpha / \|K_\alpha\|_\alpha^\alpha$ with $K_\alpha(x) = e^{x/\alpha} / (1 + e^x)$. This is, up to normalisation, the density of $\log B$ for a beta-prime variable B with parameters $(\frac{1}{\alpha}, 1 - \frac{1}{\alpha})$, and it is the hyperbolic secant density at $\alpha = 2$. So for pure states Conjecture 6.3.5 is the classical inequality

$$\|\mu * \mu' * K_\alpha\|_\alpha \|K_\alpha\|_\alpha \geq \|\mu * K_\alpha\|_\alpha \|\mu' * K_\alpha\|_\alpha \quad \text{for finite positive measures } \mu, \mu' \text{ on } \mathbb{R}.$$

Minkowski's inequality gives the companion upper bound $\|\mu * K_\alpha\|_\alpha \leq \mu(\mathbb{R}) \|K_\alpha\|_\alpha$, and since $\mu(\mathbb{R}) = \langle \psi | \sigma^{(1-\alpha)/\alpha} | \psi \rangle$ this is exactly $Q_\alpha^{\text{HT}} \leq \tilde{Q}_\alpha$ for pure states. The lower bound is not a property of log-concave kernels in general: for $K = \mathbf{1}_{[0,1]}$ and $\mu = \mu' = \delta_0 + \delta_{1/2}$ it reads $(1 + 3^\alpha) \geq (1 + 2^{\alpha-1})^2$, which fails for every $\alpha \geq 5$. Any proof must therefore use the shape of K_α , as the Gaussian-mixture argument does at $\alpha = 2$.

In the language of Rényi entropies $h_\alpha(f) = \frac{1}{1-\alpha} \log \int f^\alpha$, with $X \sim \mu$, $X' \sim \mu'$ and $Z \sim K_\alpha$ independent, the pure-state conjecture reads

$$h_\alpha(X + X' + Z) + h_\alpha(Z) \leq h_\alpha(X + Z) + h_\alpha(X' + Z) \quad (\alpha > 1),$$

a submodularity of the Rényi entropy of sums in which the common summand Z has the law of $\log B$, $B \sim \text{Beta}'(\frac{1}{\alpha}, 1 - \frac{1}{\alpha})$. For the Shannon entropy and an arbitrary common summand this inequality is data processing: $h(X + X' + Z) - h(X' + Z) = I(X; X + X' + Z) \leq I(X; X + Z) = h(X + Z) - h(Z)$, since $X \rightarrow X + Z \rightarrow X + Z + X'$ is a Markov chain; this is the pure-state shadow of the additivity of the Umegaki entropy. For $\alpha \neq 1$ the inequality depends on the law of Z : it fails for Z uniform on $[0, 1)$ and fair coins $X, X' \in \{0, \frac{1}{2}\}$ when $\alpha \geq 5$ (the computation above), while for Gaussian Z it holds at every integer $\alpha = n$, because $\int \prod_{i=1}^n G(x - a_i) dx \propto \exp(-\frac{1}{2} \sum_i (a_i - \bar{a})^2)$ depends on $a \in \mathbb{R}^n$ only through the S_n -invariant form $Q_0(a) = \sum_i (a_i - \bar{a})^2$, and for such a kernel the symmetrisation of the proof of Theorem 6.4.1 works in every dimension: with a, b the i.i.d. samples, $\frac{1}{n!} \sum_{\pi \in S_n} e^{-\langle a - \bar{a}, \pi b - \bar{b} \rangle} \geq 1$ by Jensen's inequality, because $\frac{1}{n!} \sum_{\pi} \pi b = \bar{b} \mathbf{1}$. At $n = 2$ the kernel of K_2 is $e^{-(a_1 + a_2)/2} k(a_1 - a_2)$, a Gaussian scale mixture, and the same argument with Chebyshev's association inequality is Theorem 6.4.1. For $n \geq 3$ the kernel $\Psi_n(a) = \int \prod_i K_n(x - a_i) dx \propto e^{(1-n)\bar{a}} \int_0^\infty s^{n-2} \prod_i (1 + s e^{a_i - \bar{a}})^{-1} ds$ is not a function of $Q_0(a)$: for $n = 3$ it decays like $e^{-|a - \bar{a}|/\sqrt{6}}$ along $(u, 0, 0)$ and like $|a - \bar{a}| e^{-|a - \bar{a}|/\sqrt{2}}$ along $(u, -u, 0)$, and at $|a - \bar{a}| = 4$ the two values are 0.379 and 0.336. A mixture of Gaussians in non-invariant quadratic forms does not support the symmetrisation step. So the integer case $n \geq 3$, even for pure states, needs an argument that does not pass through Gaussian mixtures. Theorem 6.5.3 gives one, by majorization on a simplex, and proves the inequality above at every integer $\alpha \geq 2$.

At integer orders the pure-state case reduces to an inequality for one explicit function under the symmetric group, and the proof of Theorem 6.4.1 is the case $n = 2$ of that inequality.

Proposition 6.4.5 (Integer orders, pure states). *Let $n \geq 2$ be an integer and, for $w \in \mathbb{R}^n$ with $\sum_i w_i = 0$,*

$$k_n(w) = (n-1) \int_0^\infty \frac{dt}{\prod_{i=1}^n (t + e^{w_i})}, \quad k_n(0) = 1.$$

- (a) *For a pure state ψ and faithful σ with spectral measure ν of σ in ψ , let $a_1, \dots, a_n$ be i.i.d. with the law of $\log q$ under ν , and $\bar{a}$ their mean. Then $Q_n^{\text{HT}}(\psi\|\sigma) =$*

$\mathbb{E}[e^{(1-n)\bar{a}}k_n(a-\bar{a})]$, and $Q_n^{\text{HT}}(\psi \otimes \psi' || \sigma \otimes \sigma') = \mathbb{E}[e^{(1-n)(\bar{a}+\bar{b})}k_n(a-\bar{a}+b-\bar{b})]$ with b an independent sample for the second pair.

(b) If

$$\frac{1}{n!} \sum_{\pi \in S_n} k_n(u + \pi v) \geq k_n(u) k_n(v) \quad \text{for all } u, v \text{ with } \sum u_i = \sum v_i = 0, \quad (6.4.1)$$

then Conjecture 6.3.5 holds at $\alpha = n$ for all pure states and all faithful σ, σ' . Inequality (6.4.1) is implied by its logarithmic form $\frac{1}{n!} \sum_{\pi} \log k_n(u + \pi v) \geq \log k_n(u) + \log k_n(v)$.

(c) (6.4.1) holds at $n = 2$, where $k_2(u, -u) = \frac{u}{\sinh u}$ and the logarithmic form holds factor by factor in Euler's product. For $n = 3$, $k_3(w) = -2 \sum_{i=1}^3 w_i / \prod_{j \neq i} (e^{w_i} - e^{w_j})$.

(d) $\log k_n$ is concave on the hyperplane, so (6.4.1) and its logarithmic form are reverse Jensen inequalities: for a concave L with $L(0) = 0$ one has $L(u + \pi v) \leq L(u) + L(\pi v)$ for each π , and the orbit average must make up the difference.

Proof. (a) By the trace formula,

$$Q_n^{\text{HT}}(\psi || \sigma) = (n-1) \int_0^\infty \left(\int \frac{\nu(dq)}{q+t} \right)^n dt = \mathbb{E} \left[(n-1) \int_0^\infty \prod_i (q_i + t)^{-1} dt \right]$$

with $q_i = e^{a_i}$ i.i.d., and the substitution $t = e^{\bar{a}} \tau$ gives the formula. The spectral measure of $\sigma \otimes \sigma'$ in $\psi \otimes \psi'$ is the multiplicative convolution $\nu \star \nu'$, so the samples for the product are $a_i + b_i$. (b) Since b is exchangeable and $\bar{b}$ is permutation invariant, $\mathbb{E} k_n(a - \bar{a} + b - \bar{b}) = \mathbb{E} \frac{1}{n!} \sum_{\pi} k_n((a - \bar{a}) + \pi(b - \bar{b}))$. Apply (6.4.1) with $u = a - \bar{a}$, $v = b - \bar{b}$ and take expectations; the factors $e^{(1-n)\bar{a}}$ and $e^{(1-n)\bar{b}}$ are positive and independent. The logarithmic form implies (6.4.1) by the arithmetic-geometric mean inequality. (c) For $n = 2$, $k_2(u, -u) = \int_0^\infty \frac{dt}{(t+e^u)(t+e^{-u})} = \frac{2u}{e^u - e^{-u}} = \frac{u}{\sinh u}$, which is $k(2u)$ in the notation of Theorem 6.4.1, and (6.4.1) is the inequality $k(u+v) + k(u-v) \geq 2k(u)k(v)$ proved there. For the logarithmic form, Euler's product gives $-\log k_2 = \sum_{m \geq 1} h_m(u^2)$ with $h_m(s) = \log(1 + s/(\pi^2 m^2))$ concave and vanishing at 0, so $\frac{1}{2}[h_m((u+v)^2) + h_m((u-v)^2)] \leq h_m(u^2 + v^2) \leq h_m(u^2) + h_m(v^2)$ by concavity and subadditivity. For $n = 3$, the partial fraction expansion of $t \mapsto \prod_i (t+x_i)^{-1}$ gives $\int_0^\infty \prod_i (t+x_i)^{-1} dt = -\sum_i \log x_i / \prod_{j \neq i} (x_i - x_j)$, the second divided difference of $-\log$, since $\sum_i \prod_{j \neq i} (x_i - x_j)^{-1} = 0$. (d) With $x_i = e^{w_i}$, $\prod_i (t+x_i)^{-1} = \int_{\mathbb{R}_+^n} e^{-\langle r, x \rangle - t|r|_1} dr$, so $K(x) := \int_0^\infty \prod_i (t+x_i)^{-1} dt = \int_{\mathbb{R}_+^n} e^{-\langle r, x \rangle} dr / |r|_1$. In the coordinates $r_i = e^{\theta_i}$ this is $K(e^w) = \int F(\theta + w) G(\theta) d\theta$ with $F(z) = \exp(-\sum_i e^{z_i})$ and $G(\theta) = \exp(\sum_i \theta_i - \log \sum_i e^{\theta_i})$, both log-concave on $\mathbb{R}^n$. A convolution of log-concave functions is log-concave by the Prékopa-Leindler inequality, and so is the restriction $w \mapsto k_n(w)$ to the hyperplane. The last statement is the subadditivity of concave functions vanishing at 0. $\square$

Inequality (6.4.1) is a statement about a single function of $2(n-1)$ real variables. For $n \geq 3$, k_n is neither even nor a function of the invariant quadratic form, so the two ingredients of the $n = 2$ proof, the Gaussian mixture and the parallelogram law, are both unavailable. The divided-difference structure is what works: Theorem 6.5.3(b) below proves (6.4.1) for every n , with equality only at $u = 0$ or $v = 0$. The logarithmic form remains open for $n \geq 3$; it held in all our tests (Section 6.11).

6.5 Integer orders as a correlation inequality on a simplex

At integer orders the trace formula turns Conjecture 6.3.5 into a statement about one random point of a simplex, and a majorization argument proves it whenever the functions involved are real, convex and permutation invariant, possibly after symmetrisation. This is the multivariate form of the measure pictures used so far. The product quasi-entropy becomes the pairing of a measure on the simplex built from the first pair with a function built from the second.

Fix an integer $n \geq 2$, let $\Delta_{n-1} = \{\mu \in [0, 1]^n : \sum_i \mu_i = 1\}$, and let U be uniformly distributed on Δ_{n-1} . For faithful $\sigma = \sum_j q_j |j\rangle\langle j|$ and a word $a = (a_1, \dots, a_n)$ of indices, write $q_a = (q_{a_1}, \dots, q_{a_n})$ and $W_a = \rho_{a_1 a_2} \rho_{a_2 a_3} \cdots \rho_{a_n a_1}$ for the cycle weights of the trace formula, and put

$$\mathcal{F}_{\rho, \sigma}(\mu) = \sum_a W_a (\mu \cdot q_a)^{1-n}, \quad \mathcal{G}_{\rho, \sigma}(\mu) = \sum_a W_a k_{q_a}(\mu),$$

where $k_x(\mu) = (x_1 \cdots x_n \sum_{i=1}^n \mu_i / x_i)^{-1}$ for $x \in (0, \infty)^n$. Grouping the words by eigenvalues shows that $\mathcal{F}$ and $\mathcal{G}$ depend only on the block weights $\text{Tr}(P_{\lambda_1} \rho P_{\lambda_2} \rho \cdots P_{\lambda_n} \rho)$, where the P_λ are the spectral projections of σ , so they do not depend on the choice of eigenbasis.

Proposition 6.5.1 (Simplex representation). *Let σ, σ' be faithful and write $x \circ y$ for the entrywise product.*

(a) *For $x, y \in (0, \infty)^n$,*

$$(n-1) \int_0^\infty \prod_{i=1}^n \frac{dt}{t+x_i} = \mathbb{E}(U \cdot x)^{1-n}, \quad \mathbb{E}(U \cdot (x \circ y))^{1-n} = \mathbb{E}[(U \cdot y)^{1-n} k_x(U)].$$

(b) $Q_n^{\text{HT}}(\rho \| \sigma) = \mathbb{E} \mathcal{F}_{\rho, \sigma}(U) = \mathbb{E} \mathcal{G}_{\rho, \sigma}(U)$, and

$$Q_n^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') = \mathbb{E}[\mathcal{F}_{\rho, \sigma}(U) \mathcal{G}_{\rho', \sigma'}(U)].$$

So Conjecture 6.3.5 at $\alpha = n$ is the correlation inequality $\mathbb{E}[\mathcal{F}_{\rho, \sigma} \mathcal{G}_{\rho', \sigma'}] \geq \mathbb{E} \mathcal{F}_{\rho, \sigma} \mathbb{E} \mathcal{G}_{\rho', \sigma'}$ for the uniform law on Δ_{n-1} .

(c) $\mathcal{F}_{\rho, \sigma}$ and $\mathcal{G}_{\rho, \sigma}$ are invariant under cyclic shifts of the coordinates of μ , and the reversal $(\mu_1, \mu_2, \dots, \mu_n) \mapsto (\mu_1, \mu_n, \dots, \mu_2)$ replaces them by their complex conjugates. They are real when all cycle weights are real: always at $n = 2$, where $W_{jk} = |\rho_{jk}|^2$; for qubit pairs; for pure ρ ; and for ρ with nonnegative entries in an eigenbasis of σ . For pure $\rho = |\psi\rangle\langle\psi|$, $W_a = \prod_i |\psi_{a_i}|^2$, and they are invariant under all permutations of the coordinates.

Proof. (a) Let $E_1, \dots, E_n$ be independent standard exponential variables and $S = \sum_i E_i$. Then E/S is uniform on Δ_{n-1} and independent of S , which has the Gamma law with shape n , so $\mathbb{E} S^{1-n} = 1/(n-1)!$. Since $z^{1-n} = \frac{1}{(n-2)!} \int_0^\infty r^{n-2} e^{-rz} dr$ and $\mathbb{E} e^{-rE \cdot x} = \prod_i (1+rx_i)^{-1}$,

$$\frac{\mathbb{E}(U \cdot x)^{1-n}}{(n-1)!} = \mathbb{E}(E \cdot x)^{1-n} = \frac{1}{(n-2)!} \int_0^\infty \frac{r^{n-2} dr}{\prod_i (1+rx_i)} = \frac{1}{(n-2)!} \int_0^\infty \frac{dt}{\prod_i (t+x_i)},$$

by the substitution $r = 1/t$. For the second identity, $X_i = x_i E_i$ are independent exponential variables with means x_i . Writing $X = RV$ with $R = \sum_i X_i$ and $V \in \Delta_{n-1}$, the density of (R, V) is $(\prod_i x_i)^{-1} R^{n-1} \exp(-R \sum_i V_i/x_i)$ with respect to $dR dV_1 \cdots dV_{n-1}$, and the coordinate measure $dV_1 \cdots dV_{n-1}$ gives Δ_{n-1} total mass $1/(n-1)!$. Integrating out R

gives $\mathbb{E}(X \cdot y)^{1-n} = \mathbb{E}[(U \cdot y)^{1-n} k_x(U)] / (n-1)!$, while $\mathbb{E}(X \cdot y)^{1-n} = \mathbb{E}(E \cdot (x \circ y))^{1-n} = \mathbb{E}(U \cdot (x \circ y))^{1-n} / (n-1)!$.

(b) In the eigenbasis of σ , $\text{Tr}[(\rho(\sigma+t)^{-1})^n] = \sum_a W_a \prod_i (t+q_{a_i})^{-1}$, so the trace formula $Q_n^{\text{HT}} = (n-1) \int_0^\infty \text{Tr}[(\rho(\sigma+t)^{-1})^n] dt$ and the first identity of (a) give $Q_n^{\text{HT}} = \mathbb{E}\mathcal{F}$. The product pair has eigenvectors $|j\rangle \otimes |l\rangle$ with eigenvalues $q_j q'_l$, and its cycle weights are $W_a W'_b$. The second identity of (a), with $x = q'_b$ and $y = q_a$, gives $Q_n^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') = \sum_{a,b} W_a W'_b \mathbb{E}[(U \cdot q_a)^{1-n} k_{q'_b}(U)] = \mathbb{E}[\mathcal{F}_{\rho,\sigma} \mathcal{G}_{\rho',\sigma'}]$. With $y = (1, \dots, 1)$ it gives $\mathbb{E}k_x(U) = \mathbb{E}(U \cdot x)^{1-n}$, hence $\mathbb{E}\mathcal{G} = Q_n^{\text{HT}}$.

(c) Permuting the coordinates of μ by π replaces W_a by $W_{a \circ \pi}$ after relabelling the sum. A cycle weight is invariant under cyclic shifts of its word, and reversing the word conjugates it, because ρ is Hermitian. The remaining statements are direct. $\square$

Recall that μ is *majorized* by ν , $\mu \prec \nu$, if μ is a convex combination of permutations of ν , and that ϕ is *Schur-convex* if $\mu \prec \nu$ implies $\phi(\mu) \leq \phi(\nu)$. Schur-convex functions are permutation invariant, and every permutation-invariant convex function is Schur-convex [88, Ch. 3].

Lemma 6.5.2 (Chebyshev's inequality on the simplex). *Let $\phi, \chi : \Delta_{n-1} \rightarrow \mathbb{R}$ be bounded Borel functions that are Schur-convex. Then*

$$\mathbb{E}[\phi(U)\chi(U)] \geq \mathbb{E}\phi(U) \mathbb{E}\chi(U).$$

If ϕ and χ are convex and permutation invariant, and neither is almost everywhere constant, the inequality is strict.

Proof. Induction on n . For $n = 2$ write $U = (V, 1 - V)$ with V uniform on $[0, 1]$. A Schur-convex function of $(v, 1 - v)$ is a nondecreasing function of $|v - \frac{1}{2}|$, and Chebyshev's association inequality for the single variable $|V - \frac{1}{2}|$ gives the claim.

Let $n \geq 3$, let $T = \min_i U_i$, and let I be the index at which the minimum is attained, which is unique almost surely. By the law of total covariance,

$$\text{Cov}(\phi, \chi) = \mathbb{E}[\text{Cov}(\phi, \chi \mid T, I)] + \text{Cov}(\mathbb{E}[\phi \mid T, I], \mathbb{E}[\chi \mid T, I]).$$

Put $\Delta(s) = \{\mu \in \Delta_{n-1} : \mu_j \geq s \ \forall j\} = c + (1 - ns)(\Delta_{n-1} - c)$, where $c = (\frac{1}{n}, \dots, \frac{1}{n})$. Given $T = s$ and $I = i$, U is uniform on the facet $\{\mu_i = s\}$ of $\Delta(s)$. The map $\nu \mapsto \mu$ with $\mu_i = s$ and $\mu_j = s + (1 - ns)\nu_j$ for $j \neq i$ carries Δ_{n-2} onto this facet and the uniform law to the uniform law. It preserves majorization, because a doubly stochastic matrix acting on ν extends to one acting on μ that fixes the i -th coordinate. So the restrictions of ϕ and χ to the facet are Schur-convex functions of ν , and the first term is nonnegative by the induction hypothesis.

Given $T = s$, U is uniform on the boundary of $\Delta(s)$: the facets are congruent, and the gradient of $\min_i \mu_i$ has the same length on each of them. By permutation invariance, $\mathbb{E}[\phi \mid T, I] = A_\phi(1 - nT)$, with $A_\phi(r) = \mathbb{E}\phi(c + r(X - c))$ and X uniform on $\partial\Delta_{n-1}$. For $0 \leq r \leq r' \leq 1$, the point $c + r(x - c)$ is a convex combination of $c + r'(x - c)$ and c , which is the average of the permutations of $c + r'(x - c)$; hence $c + r(x - c) \prec c + r'(x - c)$, and A_ϕ is nondecreasing. The same holds for χ , and Chebyshev's inequality for the single variable T makes the second term nonnegative.

For strictness, let ϕ be convex and permutation invariant. Then $\phi(c) \leq \phi(\mu)$ for all μ , and A_ϕ is convex and nondecreasing with $A_\phi(0) = \phi(c)$. If $A_\phi = \phi(c)$ on $[0, 1]$, then $\mathbb{E}\phi(U) = \mathbb{E}A_\phi(1 - nT) = \phi(c)$, and since $\phi \geq \phi(c)$, ϕ is almost everywhere constant. Otherwise there is $r_\phi < 1$ with $A_\phi(r_\phi) > \phi(c)$, and convexity gives $A_\phi(r') - A_\phi(r) \geq (r' - r)(A_\phi(r_\phi) - \phi(c))/r > 0$ for $r_\phi \leq r < r' \leq 1$, so A_ϕ is strictly increasing on $[r_\phi, 1]$. If

the same holds for χ , then both $s \mapsto A_\phi(1 - ns)$ and $s \mapsto A_\chi(1 - ns)$ are strictly decreasing on some $[0, s_0)$. The variable T has positive density there, so, with T' an independent copy, the second term $\frac{1}{2}\mathbb{E}[(A_\phi(1 - nT) - A_\phi(1 - nT'))(A_\chi(1 - nT) - A_\chi(1 - nT'))]$ is strictly positive. $\square$

Call a pair (ρ, σ) *sign-free* if $\text{Tr}(P\rho Q\rho R\rho) \geq 0$ for all pairwise distinct spectral projections P, Q, R of σ , and *weakly sign-free* if $\text{Re Tr}(P\rho Q\rho R\rho) \geq 0$ for all of them. The condition is empty when σ has at most two distinct eigenvalues, in particular for every qubit pair. It holds for pure ρ , for ρ with nonnegative entries in an eigenbasis of σ , and for commuting pairs.

Theorem 6.5.3 (Integer orders through the simplex). *(a) ($n = 2$.) $\mathcal{F}_{\rho, \sigma}(v, 1 - v) = \mathcal{G}_{\rho, \sigma}(v, 1 - v) = \sum_{jk} |\rho_{jk}|^2 / (vq_j + (1 - v)q_k)$, so $Q_2^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') = \int_0^1 \mathcal{F}_{\rho, \sigma} \mathcal{F}_{\rho', \sigma'} dv$. These functions are convex and symmetric under $v \mapsto 1 - v$, so Lemma 6.5.2 gives Theorem 6.4.1 again.*

(b) Inequality (6.4.1) holds for every $n \geq 2$, strictly unless $u = 0$ or $v = 0$. Hence Conjecture 6.3.5 holds at every integer order $\alpha = n \geq 2$ for all pure states ψ, ψ' , and when the right-hand side is finite, equality holds only if $[\psi, \sigma] = 0$ or $[\psi', \sigma'] = 0$. More generally it holds at $\alpha = n$ when one pair is pure and all block weights $\text{Tr}(P_{\lambda_1} \rho P_{\lambda_2} \rho \cdots P_{\lambda_n} \rho)$ of the other are nonnegative, for instance a qubit pair or a pair in which ρ has nonnegative entries in an eigenbasis of σ . Theorem 6.7.1 below removes the condition on the other pair.

(c) ($n = 3$.) If (ρ', σ') is sign-free and (ρ, σ) is weakly sign-free, then

$$Q_3^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq Q_3^{\text{HT}}(\rho \| \sigma) Q_3^{\text{HT}}(\rho' \| \sigma'),$$

and when the right-hand side is finite, equality holds if and only if $[\rho, \sigma] = 0$ or $[\rho', \sigma'] = 0$. In particular this holds for all pairs in which σ and σ' have at most two distinct eigenvalues each, in any dimension. Corollary 6.6.3 below removes the weak sign-freeness condition on (ρ, σ) when σ' has at most two distinct eigenvalues.

If σ or σ' is not faithful, either both sides are infinite or one restricts to the supports, so for the inequalities the faithfulness assumption of Proposition 6.5.1 costs nothing.

Proof. (a) At $n = 2$, $k_{(q_j, q_k)}(v, 1 - v) = 1/(vq_k + (1 - v)q_j)$, and the sum is symmetric in (j, k) . Each term is convex in v .

(b) Let $y = e^u$ and $x = e^v$, so that $k_n(u) = \mathbb{E}(U \cdot y)^{1-n}$ and $k_n(v) = \mathbb{E}(U \cdot x)^{1-n}$ by Proposition 6.5.1(a). The same identity with $x \circ \pi$ in place of x gives $k_n(u + \pi v) = \mathbb{E}[(U \cdot y)^{1-n} k_{x \circ \pi}(U)]$, and $k_{x \circ \pi}(\mu) = k_x(\mu \circ \pi^{-1})$. Averaging over π and using the exchangeability of U ,

$$\begin{aligned} \frac{1}{n!} \sum_{\pi} k_n(u + \pi v) &= \mathbb{E}[\phi_y(U) \chi_x(U)], \\ \phi_y(\mu) &= \frac{1}{n!} \sum_{\pi} (\mu \circ \pi \cdot y)^{1-n}, \quad \chi_x(\mu) = \frac{1}{n!} \sum_{\pi} k_x(\mu \circ \pi), \end{aligned}$$

with $\mathbb{E} \phi_y = k_n(u)$ and $\mathbb{E} \chi_x = k_n(v)$. Both functions are convex and permutation invariant, because $(\mu \cdot y)^{1-n}$ and k_x are convex functions of a positive linear form. Lemma 6.5.2 gives (6.4.1). The function ϕ_y is a sum of convex functions, so it is almost everywhere constant only if every summand is affine, that is, only if y is constant, i.e. $u = 0$; likewise χ_x only if $v = 0$. This gives strictness. For the states, let ψ' be pure. Then $\mathcal{G}_{\psi', \sigma'}$ is

a nonnegative combination of the convex functions k_x and is permutation invariant, by Proposition 6.5.1(c). If all block weights of (ρ, σ) are nonnegative, $\mathcal{F}_{\rho, \sigma}$ is real and convex, and by exchangeability $\mathbb{E}[\mathcal{F}_{\rho, \sigma} \mathcal{G}_{\psi', \sigma'}] = \mathbb{E}[\bar{\mathcal{F}} \mathcal{G}_{\psi', \sigma'}]$, where $\bar{\mathcal{F}}$ is the average of $\mathcal{F}_{\rho, \sigma}$ over permutations of the coordinates; it is convex, permutation invariant and has the same mean. Lemma 6.5.2 and Proposition 6.5.1(b) give the inequality. For pure ψ , $\mathcal{F}_{\psi, \sigma}$ is the expectation of $(\mu \cdot q_a)^{1-n}$ over i.i.d. letters with law $|\psi_j|^2$; it is constant only if ψ lies in one eigenspace of σ , that is, only if $[\psi, \sigma] = 0$, which gives the equality statement.

(c) Group the words by spectral projections: $\mathcal{F}_{\rho, \sigma}(\mu) = \sum w(\lambda_1, \lambda_2, \lambda_3)(\mu_1 \lambda_1 + \mu_2 \lambda_2 + \mu_3 \lambda_3)^{-2}$, with $w(\lambda_1, \lambda_2, \lambda_3) = \text{Tr}(P_{\lambda_1} \rho P_{\lambda_2} \rho P_{\lambda_3} \rho)$. If two of the λ 's coincide, w is real and nonnegative. Up to cyclic order it is $\text{Tr}[(P \rho P)^3]$, or $\text{Tr}(A X X^*)$ with $A = P \rho P \geq 0$ and $X = P \rho Q$. Hence $\text{Re } \mathcal{F}_{\rho, \sigma}$ is a constant plus a nonnegative combination of convex functions for weakly sign-free pairs. It is invariant under all permutations, because by Proposition 6.5.1(c) the reversal conjugates $\mathcal{F}$ and the cyclic shifts fix it. For a sign-free pair all block weights are real, so $\mathcal{G}_{\rho', \sigma'}$ is real, convex and permutation invariant. Since $\mathbb{E} \mathcal{F} = Q_3^{\text{HT}}$ is real and the product formula is real,

$$Q_3^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') - Q_3^{\text{HT}}(\rho \| \sigma) Q_3^{\text{HT}}(\rho' \| \sigma') = \text{Cov}(\text{Re } \mathcal{F}_{\rho, \sigma}(U), \mathcal{G}_{\rho', \sigma'}(U)) \geq 0$$

by Lemma 6.5.2. For equality, suppose $[\rho, \sigma] \neq 0$. Then $P \rho Q \neq 0$ for two distinct spectral projections P, Q . The range of $X = P \rho Q$ lies in the range of $P \rho^{1/2}$, which is the range of $A = P \rho P$. So $A^{1/2} X \neq 0$ and $w(\lambda, \lambda, \lambda') = \text{Tr}(A X X^*) = \|A^{1/2} X\|_2^2 > 0$. The term $(\mu_1 \lambda + \mu_2 \lambda + \mu_3 \lambda')^{-2}$ is not affine on Δ_2 , so $\text{Re } \mathcal{F}_{\rho, \sigma}$ is not almost everywhere constant. The same argument applies to $\mathcal{G}_{\rho', \sigma'}$ when $[\rho', \sigma'] \neq 0$, and Lemma 6.5.2 gives strict inequality. If one pair commutes, equality holds by the remark after Conjecture 6.3.5. $\square$

By Remark 6.4.4, Theorem 6.5.3(b) is equivalent to a classical statement. For every integer $n \geq 2$, all independent real random variables X, X' , and $Z = \log B$ with $B \sim \text{Beta}'(\frac{1}{n}, 1 - \frac{1}{n})$ independent of them,

$$h_n(X + X' + Z) + h_n(Z) \leq h_n(X + Z) + h_n(X' + Z).$$

This follows first for finitely supported laws, which come from pure states, and then for compactly supported laws by approximation, since the densities of $X + Z$ converge uniformly with uniform exponential tails. For general finite positive measures μ and their restrictions μ_M to $[-M, M]$, Minkowski's inequality gives $\|(\mu - \mu_M) * K_n\|_n \leq (\mu - \mu_M)(\mathbb{R}) \|K_n\|_n \rightarrow 0$, so the convolution inequality of Remark 6.4.4 holds for all finite positive measures at every integer order.

Remark 6.5.4 (What the simplex picture leaves open). In the language of measures, Proposition 6.5.1 says that $Q_n^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma')$ is the pairing of the complex measure $\mathcal{F}_{\rho, \sigma} dU$ on Δ_{n-1} with the continuous function $\mathcal{G}_{\rho', \sigma'}$. The conjecture asks that the centred measure $(\mathcal{F}_{\rho, \sigma} - Q_n^{\text{HT}}) dU$ lie in the dual cone of all functions $\mathcal{G}_{\rho', \sigma'}$. Lemma 6.5.2 exhibits one cone that is positive against itself: centred Schur-convex densities pair nonnegatively with Schur-convex functions. Theorem 6.5.3 is the observation that sign-free pairs land in that cone. Three things fall outside it.

(i) *Chirality*. For general pairs at $n = 3$,

$$Q_3^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') - Q_3^{\text{HT}} Q_3^{\text{HT}} = \text{Cov}(\text{Re } \mathcal{F}_{\rho, \sigma}, \text{Re } \mathcal{G}_{\rho', \sigma'}) - \mathbb{E}[\text{Im } \mathcal{F}_{\rho, \sigma} \text{Im } \mathcal{G}_{\rho', \sigma'}].$$

The imaginary parts come from the orientation of 3-cycles through three distinct eigenvalues, $\text{Im } \text{Tr}(P \rho Q \rho R \rho)$. They are odd under transpositions of the coordinates, and the last term has no sign. Numerically it is at most $1.4 \cdot 10^{-3}$ of the covariance for random 3×3 products.

- (ii) *Schur-convexity.* The covariance term is covered by Lemma 6.5.2 whenever $\text{Re } \mathcal{F}_{\rho, \sigma}$ and $\text{Re } \mathcal{G}_{\rho', \sigma'}$ are Schur-convex. This is weaker than sign-freeness, and it held for $\text{Re } \mathcal{F}_3$ in all 600 random pairs we drew in dimensions 3 and 4, but not always: $\text{Re } \mathcal{G}_3$ failed in 4 of them. Convexity of $\text{Re } \mathcal{F}_3$ fails as well for some pairs found by adversarial search (see below). Super-multiplicativity held for every such pair against random partners.
- (iii) *Order $n \geq 4$.* $\text{Re } \mathcal{F}_n$ is only dihedrally invariant. The induction in Lemma 6.5.2 needs the full symmetric group on each facet, so two qubit pairs at $\alpha = 4$, whose cycle weights are nonnegative but not permutation invariant, are already outside its reach.

Against partners whose σ' has at most two distinct eigenvalues, (i) and (ii) are bypassed by Theorem 6.6.2 below, which never separates real and imaginary parts.

At $n = 3$, for a qubit pair (ρ', σ') and an arbitrary (ρ, σ) , $\mathcal{G}_{\rho', \sigma'}$ is a constant plus a sum of convex functions of single coordinates. Conditioning on U_1 then reduces the conjecture to one variable: it holds for (ρ, σ) against every qubit pair if and only if (for the converse, let $\rho'_{22} \rightarrow 0$ with $|\rho'_{12}|^2 \leq \rho'_{11}\rho'_{22}$ to isolate one kernel)

$$\Delta(a) = 2 \int_0^1 \theta (R(\theta) - Q_3^{\text{HT}}(\rho \parallel \sigma)) \frac{d\theta}{1 + (a-1)\theta} \geq 0 \quad \text{for all } a > 0,$$

where $R(\theta) = \mathbb{E}[\text{Re } \mathcal{F}_{\rho, \sigma} \mid U_1 = 1 - \theta] = \sum_{ijk} \rho_{ij} \rho_{jk} \rho_{ki} / (((1 - \theta)q_i + \theta q_j)((1 - \theta)q_i + \theta q_k))$. Exchangeability and the cyclic invariance of $\mathcal{F}$ give $2 \int_0^1 \theta R = 3 \int_0^1 \theta^2 R = Q_3^{\text{HT}}$, so the measure $\theta(R - Q_3^{\text{HT}}) d\theta$ annihilates affine functions on $[0, 1]$. If $R - Q_3^{\text{HT}}$ changes sign only twice and is positive near the ends, then $\Delta \geq 0$, by comparing the convex kernel $\frac{1}{1+(a-1)\theta}$ with its linear interpolant at the two crossings. For weakly sign-free pairs R is convex, being an average of $\text{Re } \mathcal{F}$ over segments that move affinely with θ , and the sign pattern follows. In general R is not convex: adversarial searches in dimensions 3 and 4 found pairs with $R'' < 0$ near $\theta = 0$ or $\theta = 1$. In every case, including those, $R - Q_3^{\text{HT}}$ had the pattern $+-+$ and $\Delta > 0$. In fact Δ is the function $\mathcal{D}_{\rho \parallel \sigma}$ of Theorem 6.6.2 below: Proposition 6.5.1(a) with $x = (1, 1, a)$, for which $k_x(\mu) = (a - (a-1)\mu_3)^{-1}$, writes $\mathcal{D}_{\rho \parallel \sigma}(a) = \mathbb{E}[(\text{Re } \mathcal{F}_{\rho, \sigma}(U) - Q_3^{\text{HT}})(a - (a-1)U_3)^{-1}]$, and conditioning on $U_3 = 1 - \theta$, whose density is 2θ , gives Δ . So $\Delta(a) \geq 0$ for every pair (ρ, σ) and every $a > 0$, with no condition on R .

For a qutrit pair the matrix can be eliminated. Put $K_3(x) = 2 \int_0^\infty \prod_i (t + x_i)^{-1} dt$ and $J_x(a) = \frac{1}{3} \sum_i K_3(x^{(i)}) - K_3(1, 1, a) K_3(x)$, where $x^{(i)}$ is x with x_i replaced by ax_i . Up to a positive factor, the contribution of a partner word $(\lambda, \lambda, a\lambda)$ is

$$3 \sum_{a' \neq b'} \rho_{a'a'} |\rho_{a'b'}|^2 J_{(q_{a'}, q_{a'}, q_{b'})}(a) + 6 \text{Re}(\rho_{12} \rho_{23} \rho_{31}) J_{(q_1, q_2, q_3)}(a).$$

Here $J \geq 0$ by Theorem 6.5.3(b), and only the 3-cycle can be negative. For each labelling $\{a', b', c\} = \{1, 2, 3\}$ and each $s > 0$, $|\rho_{12} \rho_{23} \rho_{31}| \leq \frac{1}{2} (s \rho_{b'b'} |\rho_{b'c}|^2 + s^{-1} \rho_{a'a'} |\rho_{a'c}|^2)$. Splitting the 3-cycle in this way at all three vertices and optimising s shows that every qutrit pair is super-multiplicative at $\alpha = 3$ against every pair whose σ' has two distinct eigenvalues, provided

$$J_{(q_1, q_2, q_3)}(a) \leq \sum_{c=1}^3 \left(J_{(q_{a'}, q_{a'}, q_c)}(a) J_{(q_{b'}, q_{b'}, q_c)}(a) \right)^{1/2} \quad \text{for all } q \in (0, \infty)^3, a > 0. \quad (6.5.1)$$

This is a scalar inequality between seven values of one explicit function, with equality when two of the q_i coincide. It is proved in Corollary 6.7.4 below, through a rooted sum-of-squares

form of the Lieb–Thirring gap; its pointwise analogue for R'' , which would follow from a Peano-kernel form of Δ and the Cauchy–Schwarz inequality, fails near $\theta = 0$. The same splitting with shared budgets does not suffice in dimension 5, although the inequality it is meant to prove still held in our searches there. Corollary 6.6.3 proves that conclusion directly, in every dimension.

Splitting the partner side word by word is not possible. The pointwise condition of Section 6.11 would require the inequality for single partner words with three distinct entries, and it is false at extreme spectra. Super-multiplicativity against the actual partners still held there, because a partner always carries its own repeated words. This is the obstruction that survives Theorem 6.6.2: what remains open at $\alpha = 3$ comes from partner words with three distinct eigenvalues.

6.6 Order three: the defect as an integral of Lieb–Thirring gaps

At $\alpha = 3$ there is a second mechanism. It needs no sign condition on ρ and no Schur-convexity, and it settles the conjecture against every partner whose reference state has at most two distinct eigenvalues. Throughout, $\sigma = \sum_k q_k |k\rangle\langle k|$ is faithful, $W_x = \rho_{x_1 x_2} \rho_{x_2 x_3} \rho_{x_3 x_1}$ for words $x \in [d]^3$, $q_x = (q_{x_1}, q_{x_2}, q_{x_3})$, and

$$K_3(z) = 2 \int_0^\infty \frac{dt}{\prod_{i=1}^3 (t + z_i)} = 2 \int_0^\infty \frac{t dt}{\prod_{i=1}^3 (1 + tz_i)} \quad (z \in (0, \infty)^3),$$

the second form by the substitution $t \mapsto 1/t$. This is the function K_3 of Remark 6.5.4, and $K_3(e^w) = k_3(w)$ when $\sum_i w_i = 0$, in the notation of Proposition 6.4.5. Moreover $K_3(z) = \mathbb{E}(U \cdot z)^{-2}$ by Proposition 6.5.1(a), $K_3(1, 1, 1) = 1$, and K_3 is symmetric and homogeneous of degree -2 . Expanding traces in the eigenbasis of σ ,

$$\mathrm{Tr} [f(\sigma) \rho g(\sigma) \rho h(\sigma) \rho] = \sum_x W_x f(q_{x_1}) g(q_{x_2}) h(q_{x_3}) \quad (6.6.1)$$

for all functions f, g, h on the spectrum, so $Q_3^{\mathrm{HT}}(\rho \| \sigma) = \sum_x W_x K_3(q_x)$. For $a > 0$ put

$$\begin{aligned} \mathcal{D}_{\rho \| \sigma}(a) &= 2 \int_0^\infty \mathrm{Tr} [\rho(\sigma + t)^{-1} \rho(\sigma + t)^{-1} \rho(a\sigma + t)^{-1}] dt - K_3(1, 1, a) Q_3^{\mathrm{HT}}(\rho \| \sigma) \\ &= \sum_x W_x [K_3(q_{x_1}, q_{x_2}, a q_{x_3}) - K_3(1, 1, a) K_3(q_x)], \end{aligned}$$

where $K_3(1, 1, a) = 2(a - 1 - \log a)/(a - 1)^2$. Since W_x is invariant under cyclic shifts of x , the position of the dilated letter does not matter. The function $\mathcal{D}_{\rho \| \sigma}$ vanishes identically for commuting pairs, and $\mathcal{D}_{\rho \| \sigma}(1) = \mathcal{D}'_{\rho \| \sigma}(1) = 0$ by Euler's identity $\sum_i z_i \partial_i K_3 = -2K_3$.

Proposition 6.6.1 (A two-time representation). *Let $z \in (0, \infty)^3$ and $a > 0$, and for $s, t > 0$ put $m_i = 1 + tz_i$ and $n_i = (1 + s) + t(a + s)z_i$. Then*

$$\begin{aligned} (a) \quad & 4 \int_0^\infty \int_0^\infty ts \frac{m_3^2}{m_1 m_2 n_3^3} ds dt = K_3(z_1, z_2, a z_3), \\ (b) \quad & 4 \int_0^\infty \int_0^\infty \frac{ts ds dt}{n_1 n_2 n_3} = K_3(1, 1, a) K_3(z). \end{aligned}$$

Proof. All integrands are positive, so the order of integration is free. (a) For $A, m > 0$ the substitution $u = sm$ gives $\int_0^\infty s(A + sm)^{-3} ds = m^{-2} \int_0^\infty ((A + u)^{-2} - A(A + u)^{-3}) du = \frac{1}{2Am^2}$.

Since $n_3 = (1 + atz_3) + sm_3$, the s -integral of $sm_3^2n_3^{-3}$ is $\frac{1}{2(1+atz_3)}$, and the left side becomes $2 \int_0^\infty t[(1 + tz_1)(1 + tz_2)(1 + atz_3)]^{-1} dt = K_3(z_1, z_2, atz_3)$. (b) Fix s and put $u_s = 1 + s$ and $v_s = a + s$, so that $n_i = u_s + v_s tz_i$. The substitution $t = (u_s/v_s)u$ gives $\int_0^\infty t dt / \prod_i (u_s + v_s tz_i) = u_s^{-1} v_s^{-2} K_3(z)/2$. For $a \neq 1$, the partial fractions $\frac{s}{(1+s)(a+s)^2} = \frac{-1}{(a-1)^2} \left(\frac{1}{1+s} - \frac{1}{a+s} \right) + \frac{a}{(a-1)(a+s)^2}$ give $\int_0^\infty \frac{s ds}{(1+s)(a+s)^2} = \frac{a-1-\log a}{(a-1)^2} = \frac{1}{2} K_3(1, 1, a)$, and the case $a = 1$ follows by continuity. Hence the left side is $4 \cdot \frac{1}{2} K_3(z) \cdot \frac{1}{2} K_3(1, 1, a)$. $\square$

Write $\xi_i = m_i/n_i$. Then the integrand of part (a) is $ts(m_1m_2m_3)^{-1}\xi_3^3$, and that of part (b) is $ts(m_1m_2m_3)^{-1}\xi_1\xi_2\xi_3$. Averaging (a) over the position of the dilated letter,

$$J_z(a) = \frac{1}{3} \sum_{i=1}^3 K_3(z^{(i)}) - K_3(1, 1, a)K_3(z) = 4 \int_0^\infty \int_0^\infty \frac{ts}{m_1m_2m_3} \left(\frac{1}{3} \sum_i \xi_i^3 - \xi_1\xi_2\xi_3 \right) ds dt \geq 0,$$

with J and $z^{(i)}$ as in Remark 6.5.4. So the case of (6.4.1) at $n = 3$ in which one argument has two equal coordinates is the arithmetic–geometric mean inequality. The noncommutative form of that inequality is the Lieb–Thirring inequality $\text{Tr}[(AX)^3] \leq \text{Tr}[A^3X^3]$ for $A, X \geq 0$, and it controls the defect for every ρ .

Theorem 6.6.2 (The defect against a two-level partner). *Let $\rho \geq 0$, let σ be faithful and let $a > 0$. For $s, t > 0$ put $M_t = \mathbf{1} + t\sigma$, $N_{t,s} = (1 + s)\mathbf{1} + t(a + s)\sigma$, $\rho_t = M_t^{-1/2} \rho M_t^{-1/2}$ and $X_{t,s} = M_t N_{t,s}^{-1}$. Then*

$$\mathcal{D}_{\rho\|\sigma}(a) = 4 \int_0^\infty \int_0^\infty ts \left(\text{Tr}[\rho_t^3 X_{t,s}^3] - \text{Tr}[(\rho_t X_{t,s})^3] \right) ds dt \geq 0,$$

and the integrand is nonnegative at every (t, s) . If $[\rho, \sigma] \neq 0$ and $a \neq 1$, then $\mathcal{D}_{\rho\|\sigma}(a) > 0$.

Proof. $M = M_t$ and $N = N_{t,s}$ are positive functions of σ , so they commute. Since $\rho_t X = M^{-1/2}(\rho N^{-1})M^{1/2}$, we have $\text{Tr}[(\rho_t X)^3] = \text{Tr}[(\rho N^{-1})^3]$, and $\text{Tr}[\rho_t^3 X^3] = \text{Tr}[\rho M^{-1} \rho M^{-1} \rho M^2 N^{-3}]$. With $m(q) = 1 + tq$ and $n(q) = (1 + s) + t(a + s)q$, (6.6.1) gives

$$\text{Tr}[\rho_t^3 X^3] = \sum_x W_x \frac{m(q_{x_3})^2}{m(q_{x_1})m(q_{x_2})n(q_{x_3})^3}, \quad \text{Tr}[(\rho_t X)^3] = \sum_x W_x \prod_{i=1}^3 n(q_{x_i})^{-1}.$$

These are finite sums of positive kernels that are $ts ds dt$ -integrable by Proposition 6.6.1. Integrating term by term, parts (a) and (b) give the two terms of $\mathcal{D}_{\rho\|\sigma}(a)$. Since $\rho_t \geq 0$ and $X \geq 0$, the Lieb–Thirring inequality $\text{Tr}[(X^{1/2} A X^{1/2})^r] \leq \text{Tr}[X^{r/2} A^r X^{r/2}]$ for $r \geq 1$ [9, 81], with $r = 3$ and $A = \rho_t$, gives $\text{Tr}[(\rho_t X)^3] = \text{Tr}[(X^{1/2} \rho_t X^{1/2})^3] \leq \text{Tr}[\rho_t^3 X^3]$.

For strictness we use the identity, valid for $B \geq 0$ with eigenvalues b_i and Hermitian Y written in an eigenbasis of B ,

$$\text{Tr}[Y^2 B^3] - \text{Tr}[B Y B Y B] = \frac{1}{2} \sum_{ij} |Y_{ij}|^2 (b_i + b_j)(b_i - b_j)^2 \geq 0, \quad (6.6.2)$$

with equality if and only if $[B, Y] = 0$. Fix s and let $t \rightarrow 0$. Then $X = (1 + s)^{-1}(\mathbf{1} + t\beta\sigma + O(t^2))$ with $\beta = (1 - a)/(1 + s)$, and $\rho_t = \rho + O(t)$. Expanding $\text{Tr}[A^3(\mathbf{1} + tY)^3] - \text{Tr}[(A(\mathbf{1} + tY))^3]$ to second order in t , the terms of orders 0 and 1 cancel, and

$$\text{Tr}[\rho_t^3 X^3] - \text{Tr}[(\rho_t X)^3] = 3t^2(1 - a)^2(1 + s)^{-5} (\text{Tr}[\sigma^2 \rho^3] - \text{Tr}[\rho \sigma \rho \sigma \rho]) + O(t^3).$$

If $[\rho, \sigma] \neq 0$ and $a \neq 1$, the bracket is positive by (6.6.2). So the continuous, nonnegative integrand is positive on a nonempty open set, and $\mathcal{D}_{\rho\|\sigma}(a) > 0$. $\square$

Corollary 6.6.3 (Super-multiplicativity at $\alpha = 3$ against two-level partners). *Let $\sigma' = \lambda_1 P + \lambda_2 P^\perp$ with $\lambda_1, \lambda_2 > 0$, a projection P and $P^\perp = \mathbf{1} - P$, and let $\rho \geq 0$, faithful σ and $\rho' \geq 0$ be arbitrary, in any dimensions. With $a = \lambda_2/\lambda_1$,*

$$Q_3^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') - Q_3^{\text{HT}}(\rho \| \sigma) Q_3^{\text{HT}}(\rho' \| \sigma') = \lambda_1^{-2} \left[S_1 \mathcal{D}_{\rho \| \sigma}(a) + S_2 a^{-2} \mathcal{D}_{\rho \| \sigma}(1/a) \right] \geq 0,$$

where $S_1 = 3 \text{Tr}[P \rho' P \rho' P^\perp \rho'] \geq 0$ and $S_2 = 3 \text{Tr}[P^\perp \rho' P^\perp \rho' P \rho'] \geq 0$. Hence Conjecture 6.3.5 holds at $\alpha = 3$ whenever σ or σ' has at most two distinct eigenvalues, with no condition on ρ , ρ' or the other reference state. When the right-hand side of the conjecture is finite, equality holds if and only if $[\rho, \sigma] = 0$ or $[\rho', \sigma'] = 0$.

Proof. Replacing σ' by σ'/λ_1 multiplies both sides by λ_1^2 , so let $\lambda_1 = 1$ and $\lambda_2 = a$. In an eigenbasis of σ' adapted to P , with eigenvalues $y_l \in \{1, a\}$, the product pair has cycle weights $W_x W'_{x'}$ and eigenvalues $q_x \circ y_{x'}$, so the defect is $\sum_{x, x'} W_x W'_{x'} [K_3(q_x \circ y_{x'}) - K_3(q_x) K_3(y_{x'})]$. If all letters of x' lie in one eigenspace, $y_{x'}$ is constant and the bracket vanishes by homogeneity. If exactly one letter lies in $\text{ran } P^\perp$, $y_{x'}$ is a cyclic shift of $(1, 1, a)$, and by cyclic invariance of W_x the sum over x is $\mathcal{D}_{\rho \| \sigma}(a)$; by (6.6.1) for ρ' , with one term for each position of the odd letter, the total weight of these words is S_1 . If exactly two letters lie in $\text{ran } P^\perp$, $y_{x'}$ is a cyclic shift of $(a, a, 1) = a(1, 1, 1/a)$, and homogeneity gives $a^{-2} \mathcal{D}_{\rho \| \sigma}(1/a)$, with total weight S_2 . Next, $S_1 = 3 \text{Tr}[(P \rho' P)(P \rho' P^\perp)(P \rho' P^\perp)^*] = 3 \|(P \rho' P)^{1/2} P \rho' P^\perp\|_2^2 \geq 0$, and likewise $S_2 \geq 0$. Theorem 6.6.2 gives the inequality, and exchanging the two factors gives the case in which σ has at most two eigenvalues.

For equality, if $[\rho', \sigma'] = 0$ then $P \rho' P^\perp = 0$ and $S_1 = S_2 = 0$, and if $[\rho, \sigma] = 0$ then $\mathcal{D}_{\rho \| \sigma} \equiv 0$. Suppose neither pair commutes. Then $a \neq 1$, so $\mathcal{D}_{\rho \| \sigma}(a) > 0$ by Theorem 6.6.2, and $P \rho' P^\perp \neq 0$. The range of $P \rho' P^\perp$ lies in the range of $P \rho'^{1/2}$, which is the range of $P \rho' P$, so $(P \rho' P)^{1/2} P \rho' P^\perp \neq 0$ and $S_1 > 0$. If σ or σ' is not faithful, either both sides are infinite or one restricts to the supports, as in Theorem 6.5.3. $\square$

Remark 6.6.4 (What the identity does and does not reach). (i) The proof treats the 3-cycles of ρ , negative or complex, exactly like the other words. So the chirality and the failures of Schur-convexity in Remark 6.5.4(i),(ii) play no role against two-level partners. Corollary 6.6.3 contains the inequality and the equality cases of Corollary 6.10.4, and the case of Theorem 6.5.3(c) in which one reference state has at most two eigenvalues. The function $\Delta(a)$ of Remark 6.5.4 equals $\mathcal{D}_{\rho \| \sigma}(a)$, so $\Delta \geq 0$ for every pair. For qutrits, Corollary 6.6.3 makes inequality (6.5.1) unnecessary for the conjecture; (6.5.1) itself is proved in Corollary 6.7.4.

- (ii) The mechanism does not reach partner words with three distinct eigenvalues. A Lieb–Thirring gap $\text{Tr}[A^3 X^3] - \text{Tr}[(AX)^3]$ in which A is ρ conjugated by a function of σ produces kernels with two equal outer factors, as in Proposition 6.6.1(a), whereas a partner word with eigenvalues $(1, a, b)$ needs $K_3(z_1, az_2, bz_3)$. Such words are not positive one at a time (Section 6.11). So a proof of the remaining case must use the positivity of ρ' as well, through the repeated words that every partner carries.
- (iii) In one variable, let Z_s solve $\sigma Z_s + s Z_s \sigma = \rho$, so that $(Z_s)_{jk} = \rho_{jk}/(q_j + s q_k)$, and put $p(s) = \text{Tr}[Z_s^* \rho Z_s]$. Then $\mathbb{P}_{\rho \| \sigma}(ds) = \frac{3}{Q_3^{\text{HT}}} p(s) (1+s)^{-2} ds$ and $\mathbb{P}_0(ds) = 3(1+s)^{-4} ds$ are probability measures with mean $\frac{1}{2}$, the second being the value of the first for commuting pairs, and $\mathcal{D}_{\rho \| \sigma}(a) = \frac{2}{3} Q_3^{\text{HT}}(\rho \| \sigma) (1-a)^2 \int (\mathbb{P}_{\rho \| \sigma} - \mathbb{P}_0)(ds)/(a+s)$. So Theorem 6.6.2 says that the Stieltjes transform of $\mathbb{P}_{\rho \| \sigma}$ dominates that of $\mathbb{P}_0$. The stronger statement $\mathbb{P}_{\rho \| \sigma} \geq_{\text{cx}} \mathbb{P}_0$ holds for weakly sign-free pairs. Indeed, for an eigenvalue triple

$\lambda = (\lambda_1, \lambda_2, \lambda_3)$ of σ let $\mathbb{P}_\lambda(ds) = \frac{1}{3} \sum_k (1+s)^{-2} ((\lambda_j + s\lambda_k)(\lambda_l + s\lambda_k))^{-1} ds$, the sum over the three positions k with $\{j, l\}$ the other two. It has mass $\frac{1}{3}K_3(\lambda)$ and mean $\frac{1}{6}K_3(\lambda)$, and grouping the words by eigenvalues gives $\mathbb{P}_{\rho\|\sigma} = \frac{3}{Q_3^{\text{HT}}} \sum_\lambda \text{Re } w(\lambda) \mathbb{P}_\lambda$, with the block weights $w(\lambda) = \text{Tr}(P_{\lambda_1} \rho P_{\lambda_2} \rho P_{\lambda_3} \rho)$ of Theorem 6.5.3(c). These have nonnegative real parts for weakly sign-free pairs. In the variable $\theta = s/(1+s)$ the density of $\mathbb{P}_\lambda$ with respect to $\mathbb{P}_0$ is a sum of products of reciprocals of the positive affine functions $(1-\theta)\lambda_j + \theta\lambda_k$, hence log-convex. So the set where it lies below its mean $\frac{1}{3}K_3(\lambda)$ is an interval. Equal masses and means exclude a single sign change, so the difference has the sign pattern $+-+$ or vanishes, and comparing a convex test function with its chord through the two crossings gives $\mathbb{P}_\lambda \geq_{\text{cx}} \frac{1}{3}K_3(\lambda)\mathbb{P}_0$. Whether $\mathbb{P}_{\rho\|\sigma} \geq_{\text{cx}} \mathbb{P}_0$ for all pairs is open.

The same identity gives the conjecture to second order around trivial partners $\sigma' \propto \mathbf{1}$, for every ρ' , and it reduces the remaining case to singular states.

Proposition 6.6.5 (Second order around a trivial partner). *Let $\rho, \rho' \geq 0$, let σ be faithful, let H' be Hermitian and $\sigma'_\varepsilon = e^{\varepsilon H'}$. Then*

$$Q_3^{\text{HT}}(\rho \otimes \rho' \|\sigma \otimes \sigma'_\varepsilon) - Q_3^{\text{HT}}(\rho \|\sigma) Q_3^{\text{HT}}(\rho' \|\sigma'_\varepsilon) = \frac{3}{2} (\tau_1 - \tau_2) \ell_{\rho\|\sigma} \varepsilon^2 + O(\varepsilon^3),$$

where $\tau_1 = \text{Tr}[\rho'^3 H'^2]$, $\tau_2 = \text{Tr}[H' \rho' H' \rho'^2]$ and

$$\ell_{\rho\|\sigma} = \mathcal{D}_{\rho\|\sigma}''(1) = 2 \int_0^\infty t \left(\text{Tr}[\rho_t^3 S_t^2] - \text{Tr}[\rho_t S_t \rho_t S_t \rho_t] \right) dt,$$

with $\rho_t = (\mathbf{1} + t\sigma)^{-1/2} \rho (\mathbf{1} + t\sigma)^{-1/2}$ and $S_t = t\sigma(\mathbf{1} + t\sigma)^{-1}$. Both factors are nonnegative: $\ell_{\rho\|\sigma} > 0$ if and only if $[\rho, \sigma] \neq 0$, and $\tau_1 > \tau_2$ if and only if $[\rho', H'] \neq 0$. When both hold, there is $\varepsilon_0 > 0$, depending on ρ, σ, ρ', H' , such that the conjecture at $\alpha = 3$ holds strictly for $0 < |\varepsilon| < \varepsilon_0$.

Proof. By (6.6.2) with $B = \rho'$ and $Y = H'$, $\tau_1 - \tau_2 \geq 0$, with equality if and only if $[\rho', H'] = 0$.

The expansion. Diagonalise H' with eigenvalues h_l . The defect is $\sum_{x,x'} W_x W_{x'}' \varphi_{x,x'}(\varepsilon)$ with $\varphi_{x,x'}(\varepsilon) = K_3(q_x \circ e^{\varepsilon h_{x'}}) - K_3(q_x) K_3(e^{\varepsilon h_{x'}})$, a real-analytic function with $\varphi(0) = 0$. By (6.6.1) for ρ' , $\sum_{x'} W_{x'}' h_{x'_i} = \text{Tr}[\rho'^3 H']$ and $\sum_{x'} W_{x'}' h_{x'_i}^2 = \tau_1$ for each i , and $\sum_{x'} W_{x'}' h_{x'_i} h_{x'_j} = \tau_2$ for each $i \neq j$. At $(1, 1, 1)$, $\partial_i K_3 = -\frac{2}{3}$, $\partial_i^2 K_3 = 1$ and $\partial_i \partial_j K_3 = \frac{1}{2}$ for $i \neq j$, from $K_3(z) = \mathbb{E}(U \cdot z)^{-2}$, $\mathbb{E} U_i = \frac{1}{3}$, $\mathbb{E} U_i^2 = \frac{1}{6}$ and $\mathbb{E} U_i U_j = \frac{1}{12}$. Write $u = q_x$ and $K = K_3(u)$. Euler's identities $\sum_i u_i \partial_i K = -2K$ and $\sum_{i,j} u_i u_j \partial_i \partial_j K = 6K$ give

$$\begin{aligned} \sum_{x'} W_{x'}' \varphi'_{x,x'}(0) &= \text{Tr}[\rho'^3 H'] \left(\sum_i u_i \partial_i K + 2K \right) = 0, \\ \sum_{x'} W_{x'}' \varphi''_{x,x'}(0) &= \tau_1 \left(\sum_i u_i \partial_i K + \sum_i u_i^2 \partial_i^2 K - K \right) + \tau_2 \left(\sum_{i \neq j} u_i u_j \partial_i \partial_j K - 3K \right) \\ &= (\tau_1 - \tau_2) \left(\sum_i u_i^2 \partial_i^2 K - 3K \right). \end{aligned}$$

On the other hand $\mathcal{D}_{\rho\|\sigma}''(1) = \sum_x W_x (u_3^2 \partial_3^2 K - K) = \frac{1}{3} \sum_x W_x (\sum_i u_i^2 \partial_i^2 K - 3K)$, by cyclic invariance and $\partial_a^2 K_3(1, 1, a)|_{a=1} = 1$. So the second derivative of the defect at 0 is $3(\tau_1 - \tau_2) \ell_{\rho\|\sigma}$.

The formula for $\ell_{\rho\|\sigma}$. Put $m_i = 1 + tz_i$ and $s_i = tz_i/m_i \in (0, 1)$. From the second form of K_3 , $z_i^2 \partial_i^2 K_3(z) = 2 \int_0^\infty t \prod_j m_j^{-1} 2s_i^2 dt$, so $\lambda(z) := \frac{1}{3} \sum_i z_i^2 \partial_i^2 K_3 - K_3 =$

$2 \int_0^\infty t \prod_j m_j^{-1} (\frac{2}{3} \sum_i s_i^2 - 1) dt$. For a polynomial $G(s_1, s_2, s_3)$, using $\frac{d}{dt} m_i^{-1} = -z_i m_i^{-2}$ and $\frac{ds_i}{dt} = s_i(1 - s_i)/t$,

$$\frac{d}{dt} \left[t^2 \prod_j m_j^{-1} G \right] = t \prod_j m_j^{-1} \mathcal{T}G, \quad \mathcal{T}G = \left(2 - \sum_i s_i \right) G + \sum_i s_i(1 - s_i) \partial_i G,$$

and the boundary terms vanish at 0 and at ∞ , so $\int_0^\infty t \prod_j m_j^{-1} \mathcal{T}G dt = 0$. For $G = -\frac{1}{6} \sum_i s_i - \frac{1}{2}$ one finds $\mathcal{T}G = \frac{1}{3} \sum_i s_i^2 + \frac{1}{3} \sum_{i < j} s_i s_j - 1$, hence

$$\lambda(z) = \frac{2}{3} \int_0^\infty t \prod_j m_j^{-1} \left(\sum_i s_i^2 - \sum_{i < j} s_i s_j \right) dt.$$

Summing against W_x and using (6.6.1) with cyclic invariance, $\sum_x W_x \prod_j m(q_{x_j})^{-1} \cdot \frac{1}{3} \sum_i s(q_{x_i})^2 = \text{Tr}[\rho_t^3 S_t^2]$ and $\sum_x W_x \prod_j m(q_{x_j})^{-1} \cdot \frac{1}{3} \sum_{i < j} s(q_{x_i}) s(q_{x_j}) = \text{Tr}[\rho_t S_t \rho_t S_t \rho_t]$, which gives the formula for $\ell_{\rho \parallel \sigma} = \sum_x W_x \lambda(q_x)$. By (6.6.2) with $B = \rho_t$ and $Y = S_t$, the integrand is nonnegative and vanishes identically if and only if $[S_t, \rho_t] = 0$ for all t . Since S_t is an injective function of σ commuting with $(\mathbf{1} + t\sigma)^{\pm 1/2}$, this happens if and only if $[\rho, \sigma] = 0$. $\square$

Corollary 6.6.6 (Reduction to singular states). *Fix pairs (ρ, σ) and (ρ', σ') with σ, σ' faithful. Let $\text{diag}_\sigma(\rho)$ be the diagonal part of ρ in a fixed eigenbasis of σ (when σ is degenerate it depends on the basis, which does no harm), assume $O = \rho - \text{diag}_\sigma(\rho) \neq 0$, put $\rho_s = \text{diag}_\sigma(\rho) + sO$ and $s_+ = \max\{s : \rho_s \geq 0\}$. Then $s_+ \in [1, \infty)$, ρ_{s_+} is singular, and*

$$Q_3^{\text{HT}}(\rho_s \otimes \rho' \parallel \sigma \otimes \sigma') - Q_3^{\text{HT}}(\rho_s \parallel \sigma) Q_3^{\text{HT}}(\rho' \parallel \sigma') = s^2 A + s^3 B, \quad A \geq 0.$$

Consequently, if the conjecture at $\alpha = 3$ holds for (ρ_{s_+}, σ) against (ρ', σ') , it holds for (ρ, σ) against (ρ', σ') . Applying this to both factors, Conjecture 6.3.5 at $\alpha = 3$ holds for all pairs if and only if it holds whenever ρ and ρ' are both singular and σ and σ' both have at least three distinct eigenvalues.

Proof. A word x has $e(x) \in \{0, 2, 3\}$ edges $x_i x_{i+1}$ with distinct letters, and $W_x(\rho_s) = s^{e(x)} W_x(\rho)$. Words with $e(x) = 0$ are constant and contribute nothing, by homogeneity. A word with $e(x) = 2$ is a cyclic shift of (b, b, c) with $b \neq c$, and $W_x(\rho) = \rho_{bb} |\rho_{bc}|^2 \geq 0$. By homogeneity and the symmetry of K_3 , the sum over the partner words is $\sum_{x'} W_{x'}' [K_3(q_x \circ y_{x'}) - K_3(q_x) K_3(y_{x'})] = q_b^{-2} \mathcal{D}_{\rho' \parallel \sigma'}(q_c/q_b)$, which is nonnegative by Theorem 6.6.2 applied to the partner. This gives $A \geq 0$. The off-diagonal part O is traceless, Hermitian and nonzero, so it has a negative eigenvalue and $s_+ < \infty$. Moreover $s_+ \geq 1$ because $\rho_1 = \rho$, and ρ_{s_+} is singular by maximality. If the defect at s_+ is nonnegative, then $A + s_+ B \geq 0$, and since $A \geq 0$ and $s \mapsto A + sB$ is affine, $A + B \geq 0$, which is the defect at $s = 1$. For the last statement, the case in which one reference state has at most two eigenvalues is Corollary 6.6.3. $\square$

The Lieb–Thirring inequality also gives a family of upper bounds for Q_3^{HT} in one line. It contains the case $\alpha = 3$ of Theorem 6.3.2.

Corollary 6.6.7 (Upper bounds at order three). *For $\rho \geq 0$, faithful σ and every real γ ,*

$$Q_3^{\text{HT}}(\rho \parallel \sigma) \leq \text{Tr} [\rho \sigma^{-\gamma} \rho \sigma^{-\gamma} \rho \sigma^{2\gamma-2}].$$

At $\gamma = \frac{2}{3}$ the right-hand side is the sandwiched quasi-entropy $\tilde{Q}_3(\rho \parallel \sigma)$, and at $\gamma = 0$ it is the Petz quasi-entropy $\text{Tr} \rho^3 \sigma^{-2}$.

Proof. By the second form of K_3 , $Q_3^{\text{HT}}(\rho\|\sigma) = 2 \int_0^\infty t \operatorname{Tr}[(\rho Y_t)^3] dt$ with $Y_t = (\mathbf{1} + t\sigma)^{-1}$. Apply the Lieb–Thirring inequality to $A = \sigma^{-\gamma/2} \rho \sigma^{-\gamma/2}$ and $X = \sigma^\gamma Y_t$: since $AX = \sigma^{-\gamma/2}(\rho Y_t) \sigma^{\gamma/2}$, it gives $\operatorname{Tr}[(\rho Y_t)^3] \leq \operatorname{Tr}[\rho \sigma^{-\gamma} \rho \sigma^{-\gamma} \rho \sigma^{2\gamma} Y_t^3]$. Multiply by $2t$, integrate, and use $2 \int_0^\infty t(1+tq)^{-3} dt = q^{-2}$. $\square$

6.7 Pure partners, a cycle inequality and negative cycles

Three further cases of the conjecture can be reached: pure partners at every integer order; partners with three eigenvalues and a nonnegative 3-cycle; and pairs whose 3-cycles are negative on both sides and not too strongly correlated. The tools are the Lieb–Thirring inequality, the Hermite–Genocchi formula, positivity of the two-letter terms (Theorem 6.6.2), a cycle inequality for positive matrices and the symmetric-group inequality (6.4.1). Throughout this section σ and σ' are faithful, which costs nothing by the remark after Theorem 6.5.3. For $z, w \in (0, \infty)^3$ put

$$\begin{aligned} \Theta(z, w) &= K_3(z \circ w) - K_3(z)K_3(w), \\ \mathcal{D}_{\rho\|\sigma}[w] &= \sum_x W_x \Theta(q_x, w) \\ &= 2 \int_0^\infty \operatorname{Tr}[(w_1\sigma + t)^{-1} \rho (w_2\sigma + t)^{-1} \rho (w_3\sigma + t)^{-1} \rho] dt - K_3(w) Q_3^{\text{HT}}(\rho\|\sigma), \end{aligned}$$

the *three-point defect*, so that $\mathcal{D}_{\rho\|\sigma}(a) = \mathcal{D}_{\rho\|\sigma}[1, 1, a]$ and the defect of a product is $\sum_{x'} W_{x'} \mathcal{D}_{\rho\|\sigma}[y_{x'}]$. If $\sigma' = \sum_e y_e P'_e$ has spectral projections P'_e , grouping the partner words by eigenvalues gives

$$Q_3^{\text{HT}}(\rho \otimes \rho' \|\sigma \otimes \sigma') - Q_3^{\text{HT}}(\rho\|\sigma) Q_3^{\text{HT}}(\rho'\|\sigma') = \sum_{e,f,g} w'_{efg} \mathcal{D}_{\rho\|\sigma}[y_e, y_f, y_g], \quad (6.7.1)$$

with the block weights $w'_{efg} = \operatorname{Tr}(P'_e \rho' P'_f \rho' P'_g \rho')$, and the same formula with the roles exchanged. The block weights w' are cyclically invariant and conjugated by reversal. A two-letter block weight $w'_{eef} = \operatorname{Tr}[(P'_e \rho' P'_e)(P'_e \rho' P'_f)(P'_e \rho' P'_f)^*]$ is nonnegative, and $\mathcal{D}_{\rho\|\sigma}[y_e, y_e, y_f] = y_e^{-2} \mathcal{D}_{\rho\|\sigma}(y_f/y_e) \geq 0$ by Theorem 6.6.2.

Theorem 6.7.1 (Pure partners at every integer order). *Let $n \geq 2$ be an integer, let $\rho \geq 0$ and faithful σ act on $\mathbb{C}^d$, let σ' be faithful on $\mathbb{C}^{d'}$ and let $\psi \in \mathbb{C}^{d'}$. Then*

$$Q_n^{\text{HT}}(\rho \otimes \psi \psi^* \|\sigma \otimes \sigma') \geq Q_n^{\text{HT}}(\rho\|\sigma) Q_n^{\text{HT}}(\psi \psi^* \|\sigma').$$

The inequality is strict unless $[\rho, \sigma] = 0$ or $[\psi \psi^, \sigma'] = 0$. There is no condition on ρ : its cycle weights may be negative or complex.*

Proof. Both sides are homogeneous of degree $2n$ in ψ , so let $\|\psi\| = 1$. Let $\sigma' = \sum_i y_i |i\rangle\langle i|$, $p_i = |\langle i|\psi\rangle|^2$, let Y be a random variable with law $\sum_i p_i \delta_{y_i}$, and put $m_t(z) = \mathbb{E}(t + Yz)^{-1}$ and $M_t = m_t(\sigma) = \sum_i p_i (t + y_i \sigma)^{-1}$. We use the trace formula $Q_n^{\text{HT}}(\rho\|\sigma) = (n-1) \int_0^\infty \operatorname{Tr}[(\rho(\sigma + t)^{-1})^n] dt$.

Step 1: the partner integrates out. $(\mathbf{1} \otimes \psi^*)(\sigma \otimes \sigma' + t)^{-1}(\mathbf{1} \otimes \psi) = M_t$, so $\operatorname{Tr}[(\rho \otimes \psi \psi^*)(\sigma \otimes \sigma' + t)^{-1}]^n = \operatorname{Tr}[(\rho M_t)^n]$, and $Q'_n := Q_n^{\text{HT}}(\psi \psi^* \|\sigma') = (n-1) \int_0^\infty m_t(1)^n dt$.

Step 2: the n -th power of a mixture of resolvents. For $\zeta \in \mathbb{R}$ and $a \in \mathbb{R}^n$ with $\zeta + a_i > 0$ for all i , the Hermite–Genocchi formula

$$\prod_{i=1}^n \frac{1}{\zeta + a_i} = \int_{\mathbb{R}_+^n} e^{-\sum_i r_i(\zeta + a_i)} dr = \int_{\Delta_{n-1}} du \int_0^\infty R^{n-1} e^{-R(\zeta + u \cdot a)} dR = \mathbb{E}(\zeta + U \cdot a)^{-n}$$

holds, by the polar coordinates $r = Ru$ and because $(n-1)! du$ is the uniform law on Δ_{n-1} . Let $Y_1, \dots, Y_n$ be independent copies of Y , independent of U , put $\nu = \sum_c U_c/Y_c$, and define a finite positive measure κ_t on $(0, \infty)$ by $\int \varphi d\kappa_t = \mathbb{E}[\varphi(t\nu)/(Y_1 \cdots Y_n)]$. For $z > 0$, the formula with $\zeta = z$ and $a_c = t/Y_c$ gives

$$\int (z+s)^{-n} \kappa_t(ds) = \mathbb{E}\left[\frac{1}{Y_1 \cdots Y_n} \prod_c \frac{1}{z+t/Y_c}\right] = \mathbb{E} \prod_c \frac{1}{t+Y_c z} = m_t(z)^n,$$

so $\int (\sigma+s)^{-n} \kappa_t(ds) = M_t^n$ by the functional calculus.

Step 3: the t -marginal. For $\varphi \geq 0$, $\int_0^\infty \int \varphi d\kappa_t dt = \mathbb{E}[(Y_1 \cdots Y_n)^{-1} \int_0^\infty \varphi(t\nu) dt] = c_n \int_0^\infty \varphi(s) ds$ with $c_n = \mathbb{E}[(Y_1 \cdots Y_n)^{-1}]$. For $\varphi = (1+s)^{-n}$, Step 2 with $z = 1$ gives $Q'_n = (n-1) \int_0^\infty \int (1+s)^{-n} \kappa_t(ds) dt = c_n$. So $\int_0^\infty \kappa_t dt = Q'_n ds$.

Step 4: Lieb–Thirring. For fixed t and s put $A = M_t^{1/2} \rho M_t^{1/2} \geq 0$ and $X_s = (\sigma+s)^{-1} M_t^{-1} \geq 0$, which commutes with M_t . Then $AX_s = M_t^{1/2} \rho (\sigma+s)^{-1} M_t^{-1/2}$, and $\text{Tr}[A^n X_s^n] = \text{Tr}[(\rho M_t)^{n-1} \rho (\sigma+s)^{-n} M_t^{1-n}]$. The Lieb–Thirring inequality gives $\text{Tr}[(\rho(\sigma+s)^{-1})^n] = \text{Tr}[(AX_s)^n] \leq \text{Tr}[A^n X_s^n]$, and by Step 2 the right-hand side integrates against κ_t to $\text{Tr}[(\rho M_t)^n]$. By Step 3, Tonelli's theorem and Step 1,

$$\begin{aligned} Q_n^{\text{HT}}(\rho \parallel \sigma) Q'_n &= (n-1) \int_0^\infty dt \int \text{Tr}[(\rho(\sigma+s)^{-1})^n] \kappa_t(ds) \\ &\leq (n-1) \int_0^\infty \text{Tr}[(\rho M_t)^n] dt = Q_n^{\text{HT}}(\rho \otimes \psi \psi^* \parallel \sigma \otimes \sigma'). \end{aligned}$$

Strictness. The defect is $(n-1) \int_0^\infty dt \mathbb{E}[\Gamma(t, t\nu)/(Y_1 \cdots Y_n)]$ with the continuous, non-negative Lieb–Thirring gap $\Gamma(t, s) = \text{Tr}[A^n X_s^n] - \text{Tr}[(AX_s)^n]$. If $[\psi \psi^*, \sigma'] = 0$, then Y is constant, all partner words are constant and the defect vanishes; if $[\rho, \sigma] = 0$ it vanishes as well. Otherwise Y has at least two values. Let $\eta_k = \mathbb{E} Y^{-k}$. As $t \rightarrow 0$ with ν fixed, $M_t = \sigma^{-1}(\eta_1 - t\eta_2 \sigma^{-1} + O(t^2))$ and $(\sigma + t\nu)^{-1} = \sigma^{-1}(\mathbf{1} - t\nu \sigma^{-1} + O(t^2))$, so $X = \eta_1^{-1}(\mathbf{1} + tB_1 + O(t^2))$ with $B_1 = (\eta_2/\eta_1 - \nu)\sigma^{-1}$, and $A = A_t \rightarrow A_0 = \eta_1 \sigma^{-1/2} \rho \sigma^{-1/2}$. For a fixed $A \geq 0$ and $X = \theta(\mathbf{1} + tB_1 + t^2 B_2 + O(t^3))$ with X, B_1, B_2 commuting, the terms of order 1 and t of $\text{Tr}[A^n X^n] - \text{Tr}[(AX)^n]$ cancel, and so do the terms containing B_2 . In an eigenbasis of A , with eigenvalues a_i , the coefficient of t^2 is

$$\theta^n \frac{n}{4} \sum_{i,j} |(B_1)_{ij}|^2 \sum_{k=1}^{n-1} (a_i^k - a_j^k)(a_i^{n-k} - a_j^{n-k}) \geq 0,$$

because the t^2 -part of $\text{Tr}[(AX)^n]$ is $\theta^n(n \text{Tr}[A^n B_2] + \frac{n}{2} \sum_{k=1}^{n-1} \text{Tr}[B_1 A^k B_1 A^{n-k}])$. It is positive if and only if $[A, B_1] \neq 0$. Since the gap vanishes to second order in $X - \theta \mathbf{1}$ for every A , replacing A_t by A_0 changes $\Gamma(t, t\nu)$ only by $O(t^3)$. Since $[A_0, \sigma^{-1}] \neq 0$, $\Gamma(t, t\nu)/t^2$ therefore has a positive limit for every $\nu \neq \eta_2/\eta_1$, locally uniformly in ν . The variable ν has a continuous distribution on the event that $Y_1, \dots, Y_n$ are not all equal, which has positive probability. So the integrand is positive on a set of positive measure, and the defect is positive. $\square$

The argument is a noncommutative Hölder inequality, a special case of Epstein's concavity theorem [40]: the partner contributes only the classical law of Y , and the n -th power of the mixed resolvent M_t is an integral of n -th powers of single resolvents. For a mixed partner $\rho' = VV^*$, $\sigma' = \sum_e y_e P'_e$, the same route at $n = 3$ would need a positive matrix-valued measure with $(\sum_e (t + y_e \sigma)^{-1} \otimes F_e)^3 = \int (\sigma + s)^{-3} \otimes \kappa_t(ds)$, where $F_e = V^* P'_e V \geq 0$. At $d = 1$ such a measure is unique, by Stieltjes inversion, and already for two non-commuting rank-one weights it is indefinite: for $F_1 = e_1 e_1^*$ and $F_2 = gg^*$ with

$g = (\cos \phi, \sin \phi)$, its density near the left end of its support is a positive multiple of a 2×2 matrix with determinant $-\cos^2 \phi \sin^2 \phi < 0$. So Theorem 6.7.1 and Corollary 6.6.3 are genuinely different mechanisms.

Corollary 6.7.2 (Adding commuting noise). *At $\alpha = 3$, adding to ρ' a positive operator that commutes with σ' does not decrease the defect of (ρ, σ) against (ρ', σ') , and likewise for ρ . In particular Conjecture 6.3.5 holds at $\alpha = 3$ for $\rho' = \psi\psi^* + N$ with $N \geq 0$ and $[N, \sigma'] = 0$, against every (ρ, σ) .*

Proof. In (6.7.1), constant block words contribute nothing, and three-letter block weights depend only on the off-diagonal blocks $P'_e \rho' P'_f$, $e \neq f$. A two-letter block weight changes from $\text{Tr}[(P'_e \rho' P'_e) O_{ef} O_{ef}^*]$ to $\text{Tr}[(P'_e(\rho' + N) P'_e) O_{ef} O_{ef}^*]$, with $O_{ef} = P'_e \rho' P'_f$. This does not decrease it, and its coefficient $\mathcal{D}_{\rho \parallel \sigma}[y_e, y_e, y_f]$ is nonnegative. Combine with Theorem 6.7.1. $\square$

The Lieb–Thirring gap at exponent three also has a rooted Schur-product structure. It turns the three-vertex splitting before (6.5.1) in Remark 6.5.4 into an identity, and it proves inequality (6.5.1).

Proposition 6.7.3 (A rooted sum of squares for the Lieb–Thirring gap). *Let $g(\xi) = \frac{1}{3} \sum_i \xi_i^3 - \xi_1 \xi_2 \xi_3$ and, for $z, u, v > 0$,*

$$H(z; u, v) = \frac{(u - z)(v - z)(z + 2u)(z + 2v)}{z + u + v}.$$

(a) *For each $z > 0$, $(u, v) \mapsto H(z; u, v)$ is a positive semidefinite kernel, $H(z; z, v) = 0$ and $H(z; u, u) = 3g(u, u, z)$. Moreover, for all $\xi \in (0, \infty)^3$,*

$$H(\xi_2; \xi_1, \xi_3) + H(\xi_3; \xi_2, \xi_1) + H(\xi_1; \xi_3, \xi_2) = 3g(\xi).$$

(b) *For Hermitian A and $X = \text{diag}(\xi) > 0$,*

$$\text{Tr}[A^3 X^3] - \text{Tr}[(AX)^3] = \sum_c \sum_{a,b} H(\xi_c; \xi_a, \xi_b) A_{ac} A_{cb} A_{ba} = \int_0^\infty \text{Tr}[e^{-\lambda X} Y_\lambda^* A Y_\lambda] d\lambda,$$

where $Y_\lambda = e^{-\lambda X} (2X[X, A] + [X, A]X)$. For $A \geq 0$ the last integrand is nonnegative, which proves $\text{Tr}[(AX)^3] \leq \text{Tr}[A^3 X^3]$ by a sum of squares.

Proof. (a) Since $(z + u + v)^{-1} = \int_0^\infty e^{-\lambda(z+u+v)} d\lambda$, $H(z; u, v) = \int_0^\infty e^{-\lambda z} \phi_\lambda(u) \phi_\lambda(v) d\lambda$ with $\phi_\lambda(u) = (u - z)(z + 2u)e^{-\lambda u}$, a positive semidefinite kernel. The next two identities are direct. For the rotation identity put $S = \xi_1 + \xi_2 + \xi_3$ and $P_i = (\xi_j - \xi_i)(\xi_k - \xi_i)$ for $\{i, j, k\} = \{1, 2, 3\}$. The three terms are $P_i(\xi_i + 2\xi_j)(\xi_i + 2\xi_k)/S$, and $(\xi_i + 2\xi_j)(\xi_i + 2\xi_k) = S^2 - (\xi_j - \xi_k)^2$. With $D = (\xi_1 - \xi_2)(\xi_2 - \xi_3)(\xi_3 - \xi_1)$ one checks $P_1(\xi_2 - \xi_3) = P_2(\xi_3 - \xi_1) = P_3(\xi_1 - \xi_2) = -D$, so $\sum_i P_i(\xi_j - \xi_k)^2 = -D \sum_i (\xi_j - \xi_k) = 0$. Hence the sum is $S \sum_i P_i = S(\sum_i \xi_i^2 - \sum_{i < j} \xi_i \xi_j) = \sum_i \xi_i^3 - 3\xi_1 \xi_2 \xi_3 = 3g(\xi)$.

(b) By (6.6.1), $\text{Tr}[A^3 X^3] - \text{Tr}[(AX)^3] = \sum_x W_x(A)g(\xi_x)$ after averaging $\xi_{x_1}^3$ over the rotations of x . By the rotation identity in (a),

$$\sum_x W_x(A)g(\xi_x) = \frac{1}{3} \sum_x \sum_{\text{rot}} W_x(A)H(\xi_{x_2}; \xi_{x_1}, \xi_{x_3}) = \sum_x W_x(A)H(\xi_{x_2}; \xi_{x_1}, \xi_{x_3}),$$

which is the double sum. For the integral, $(Y_\lambda)_{ac} = \phi_\lambda(\xi_a)A_{ac}$ with $z = \xi_c$, and summing $e^{-\lambda \xi_c} (Y_\lambda e_c)^* A (Y_\lambda e_c)$ over c gives the integrand. $\square$

Call a function C on $(0, \infty)^3$ *AM–GM representable* if

$$C(z) = \int \omega(d\theta) \mu_\theta(z_1) \mu_\theta(z_2) \mu_\theta(z_3) g(\xi_\theta(z_1), \xi_\theta(z_2), \xi_\theta(z_3))$$

for a positive measure ω and positive measurable functions μ_θ, ξ_θ , with an absolutely convergent integral. Such functions are symmetric and nonnegative.

Corollary 6.7.4 (Rooted inequality; inequality (6.5.1)). *If C is AM–GM representable, then for all $z \in (0, \infty)^3$*

$$C(z_1, z_2, z_3) \leq \sum_{c=1}^3 \sqrt{C(z_a, z_a, z_c) C(z_b, z_b, z_c)} \quad (\{a, b\} = \{1, 2, 3\} \setminus \{c\}).$$

This applies to $z \mapsto J_z(a)$ for every $a > 0$, so inequality (6.5.1) holds. It also applies to the pure-partner kernel $G_Y(z) = \mathbb{E} K_3(z \circ (Y_1, Y_2, Y_3)) - \mathbb{E} K_3(Y_1, Y_2, Y_3) K_3(z)$, for independent copies Y_c of a positive random variable with finite support.

Proof. By Proposition 6.7.3(a), $3g(\xi) = \sum_c H(\xi_c; \xi_a, \xi_b)$ over the three rotations, and the 2×2 matrix of $H(\xi_c; \cdot, \cdot)$ at (ξ_a, ξ_b) is positive semidefinite, so $|H(\xi_c; \xi_a, \xi_b)| \leq 3\sqrt{g(\xi_a, \xi_a, \xi_c)g(\xi_b, \xi_b, \xi_c)}$. With $\mu_1\mu_2\mu_3 = \sqrt{\mu_a^2\mu_c \cdot \mu_b^2\mu_c}$ this gives

$$\mu_1\mu_2\mu_3g(\xi) \leq \sum_c \sqrt{\mu_a^2\mu_cg(\xi_a, \xi_a, \xi_c) \cdot \mu_b^2\mu_cg(\xi_b, \xi_b, \xi_c)}$$

pointwise in θ , and the Cauchy–Schwarz inequality in ω gives the claim. For $J_z(a)$, the identity after Proposition 6.6.1 is a representation with $\omega = 4ts ds dt$, $\mu(z) = (1 + tz)^{-1}$ and $\xi(z) = (1 + tz)/((1 + s) + t(a + s)z)$. For G_Y , take $\omega = 2 dt \kappa_t(ds)$ from the proof of Theorem 6.7.1 at $n = 3$, $\mu(z) = m_t(z)$ and $\xi(z) = (z + s)^{-1}/m_t(z)$. Then $\mu_1\mu_2\mu_3g(\xi) = \frac{1}{3} \sum_c (\prod_{c'} m_t(z_{c'})) m_t(z_c)^{-3} (z_c + s)^{-3} - \prod_c (z_c + s)^{-1}$. By Steps 2 and 3 of that proof, the first term integrates to $2 \int_0^\infty \prod_c m_t(z_c) dt = \mathbb{E} K_3(z \circ Y)$ and the second to $\mathbb{E} K_3(Y) \cdot K_3(z)$. $\square$

An elementary proof of the case $\omega = \delta$ (the rooted arithmetic–geometric mean inequality $g(\xi) \leq \sum_c \sqrt{g(\xi_a, \xi_a, \xi_c)g(\xi_b, \xi_b, \xi_c)}$) also follows from $\sqrt{u(u + 2v)} \geq u + v - v^2/(u + v)$ applied to the three square roots, which share the value $u + v = \xi_1 + \xi_2 + \xi_3$.

For partners with three eigenvalues the remaining terms are the 3-cycles, and a trace inequality decides how much of a 3-cycle the two-letter terms can absorb.

Theorem 6.7.5 (A cycle inequality). *For positive semidefinite A, B, C of the same size,*

$$|\mathrm{Tr}(ABC)|^3 \leq \mathrm{Tr}(A^2B) \mathrm{Tr}(B^2C) \mathrm{Tr}(C^2A).$$

Proof. Since $\mathrm{Tr}(ABC) = \mathrm{Tr}[(A^{2/3}B^{1/3})(B^{2/3}C^{1/3})(C^{2/3}A^{1/3})]$, Hölder’s inequality for Schatten norms gives

$$|\mathrm{Tr}(ABC)| \leq \|A^{2/3}B^{1/3}\|_3 \|B^{2/3}C^{1/3}\|_3 \|C^{2/3}A^{1/3}\|_3.$$

Now $\|A^{2/3}B^{1/3}\|_3^3 = \mathrm{Tr}[(B^{1/3}A^{4/3}B^{1/3})^{3/2}]$. The Araki–Lieb–Thirring inequality

$$\mathrm{Tr}[(R^{r/2}P^rR^{r/2})^q] \leq \mathrm{Tr}[(R^{1/2}PR^{1/2})^{rq}] \quad \text{for } P, R \geq 0,$$

$0 < r \leq 1$ and $q > 0$ [9], applied with $P = A^2$, $R = B$, $r = \frac{2}{3}$ and $q = \frac{3}{2}$, bounds it by $\mathrm{Tr}(B^{1/2}A^2B^{1/2}) = \mathrm{Tr}(A^2B)$. The other two factors are bounded in the same way. $\square$

Lemma 6.7.6 (Domination by a pure profile). *Let $\Pi_1, \Pi_2, \Pi_3 \geq 0$, $W = \text{Tr}(\Pi_1 \Pi_2 \Pi_3) \neq 0$ and $w_{bbc} = \text{Tr}(\Pi_b^2 \Pi_c)$ for $b \neq c$. Then there is $p \in (0, \infty)^3$ with $p_1 p_2 p_3 = |W|$ and $p_b^2 p_c \leq w_{bbc}$ for all six ordered pairs $b \neq c$.*

Proof. By the Cauchy–Schwarz inequality, $|W| = |\langle \Pi_3 \Pi_2^{1/2}, \Pi_1 \Pi_2^{1/2} \rangle_2| \leq \sqrt{w_{112} w_{332}}$, and cyclically $|W|^2 \leq w_{221} w_{331}$ and $|W|^2 \leq w_{223} w_{113}$. These three “V” inequalities show that all $w_{bbc} > 0$. In logarithmic coordinates $x_b = \log p_b$ we need $x \in \mathbb{R}^3$ with $\sum_b x_b = \log |W|$ and $2x_b + x_c \leq L_{bc} := \log w_{bbc}$. By Farkas’ lemma this system is feasible unless there are $\theta \in [0, \infty)^6$, indexed by ordered pairs, with $\sum \theta_{bc}(2e_b + e_c) = \ell(1, 1, 1)$ for some ℓ , necessarily $\ell = \sum \theta_{bc}$, and $\sum \theta_{bc} L_{bc} < \ell \log |W|$. The condition on θ says that $\sum \theta_{bc} v_{bc} = 0$, where $v_{bc} = 2e_b + e_c - (1, 1, 1) = e_b - e_a$, a being the third index. The six vectors v_{bc} are the vertices of a regular hexagon in the plane $\{\sum_i x_i = 0\}$, and the cone of such θ is generated by the minimal positive dependencies of the hexagon. These are the three opposite pairs $\{\theta_{12}, \theta_{32}\}$, $\{\theta_{21}, \theta_{31}\}$, $\{\theta_{23}, \theta_{13}\}$ and the two alternating triples $\{\theta_{12}, \theta_{23}, \theta_{31}\}$, $\{\theta_{21}, \theta_{32}, \theta_{13}\}$. On the pairs, $\sum \theta L \geq \ell \log |W|$ is a V inequality. On the triples it reads $w_{112} w_{223} w_{331} \geq |W|^3$ and $w_{221} w_{332} w_{113} \geq |W|^3$, which are Theorem 6.7.5 for (Π_1, Π_2, Π_3) and (Π_2, Π_1, Π_3) , since $|\text{Tr}(\Pi_2 \Pi_1 \Pi_3)| = |W|$. $\square$

Corollary 6.7.7 (Partners with three eigenvalues). *Let $\sigma' = \sum_{e=1}^3 y_e P'_e$ have three distinct eigenvalues, of arbitrary multiplicities, let $\rho' \geq 0$ and let (ρ, σ) be arbitrary. With $W' = \text{Tr}(P'_1 \rho' P'_2 \rho' P'_3 \rho')$ and $Z = \mathcal{D}_{\rho \parallel \sigma}[y_1, y_2, y_3]$,*

$$Q_3^{\text{HT}}(\rho \otimes \rho' \parallel \sigma \otimes \sigma') - Q_3^{\text{HT}}(\rho \parallel \sigma) Q_3^{\text{HT}}(\rho' \parallel \sigma') \geq 6(\text{Re}(W' Z) - |W'| \text{Re} Z).$$

In particular Conjecture 6.3.5 holds at $\alpha = 3$ whenever W' is real and nonnegative, for every (ρ, σ) .

Proof. By (6.7.1), the defect is $3 \sum_{b \neq c} w'_{bbc} E_{bc} + 6 \text{Re}(W' Z)$ with $E_{bc} = \mathcal{D}_{\rho \parallel \sigma}[y_b, y_b, y_c] \geq 0$. If $W' = 0$ we are done. Otherwise apply Lemma 6.7.6 to $\Pi_e = \rho'^{1/2} P'_e \rho'^{1/2}$, for which $\text{Tr}(\Pi_1 \Pi_2 \Pi_3) = W'$ and $\text{Tr}(\Pi_b^2 \Pi_c) = w'_{bbc}$, and let ψ be a vector with $\|P'_e \psi\|^2 = p_e$. By Theorem 6.7.1 the pure partner $(\psi \psi^*, \sigma')$, whose block weights are $p_e p_f p_g$, has nonnegative defect $3 \sum p_b^2 p_c E_{bc} + 6 p_1 p_2 p_3 \text{Re} Z$. Since $E_{bc} \geq 0$ and $p_b^2 p_c \leq w'_{bbc}$, the defect of (ρ', σ') is at least $3 \sum p_b^2 p_c E_{bc} + 6 \text{Re}(W' Z)$, which is at least $6(\text{Re}(W' Z) - |W'| \text{Re} Z)$. $\square$

For two qutrit pairs with nondegenerate reference states, Corollaries 6.6.3 and 6.7.7 and the symmetry of the conjecture leave the case in which neither 3-cycle lies in $[0, \infty)$. The next result covers the case in which both 3-cycles are negative and not too strongly correlated, in every dimension.

Theorem 6.7.8 (Negative cycles on both sides). *Let (ρ, σ) and (ρ', σ') be pairs with σ, σ' faithful, let Δ be the pinching of ρ by the spectral projections of σ and $O = \rho - \Delta$, and likewise Δ', O' . Suppose:*

- (i) *every block 3-cycle $\text{Tr}(P_a \rho P_b \rho P_c \rho)$ of ρ , over distinct spectral projections of σ , is real and ≤ 0 , and likewise for ρ' ;*
- (ii) *there are $\beta, \beta' > 0$ with $\frac{1}{\beta} + \frac{1}{\beta'} \leq 1$, $\Delta + \beta O \geq 0$ and $\Delta' + \beta' O' \geq 0$.*

Then $Q_3^{\text{HT}}(\rho \otimes \rho' \parallel \sigma \otimes \sigma') \geq Q_3^{\text{HT}}(\rho \parallel \sigma) Q_3^{\text{HT}}(\rho' \parallel \sigma')$.

Proof. Write the block expansion of the defect, $\sum_{x, x'} w_x w'_{x'} \Theta(q_x, y_{x'})$ over block words x of σ and x' of σ' , with $q_x, y_{x'}$ the corresponding eigenvalue words; constant words contribute nothing. Let $\mathcal{A}\mathcal{A}'$, $\mathcal{B}\mathcal{A}'$, $\mathcal{A}\mathcal{B}'$ and $\mathcal{B}\mathcal{B}'$ be the partial sums over the pairs (x, x') in which

x is a two-letter word ($\mathcal{A}$) or a three-letter word ($\mathcal{B}$), and likewise x' ($\mathcal{A}'$ or $\mathcal{B}'$). Then $\text{Def} = \mathcal{A}\mathcal{A}' + \mathcal{B}\mathcal{A}' + \mathcal{A}\mathcal{B}' + \mathcal{B}\mathcal{B}'$. A block weight scales as $w_x(\Delta + sO) = s^{e(x)}w_x(\rho)$, with $e = 2$ for two-letter and $e = 3$ for three-letter words. So for every $s > 0$,

$$\mathcal{A}\mathcal{A}' + s\mathcal{B}\mathcal{A}' = s^{-2} \sum_{x' \text{ two-letter}} w'_{x'} \mathcal{D}_{\Delta+sO\|\sigma}[y_{x'}],$$

where $\mathcal{D}_{\tau\|\sigma}$ is defined by the same block formula for a Hermitian τ . If $\Delta + sO \geq 0$, each term is nonnegative, by Theorem 6.6.2 and $w'_{x'} \geq 0$ for two-letter x' . Put $\lambda = 1/\beta$. The set of $s \geq 0$ with $\Delta' + sO' \geq 0$ is an interval containing β' , and $(1 - \lambda)^{-1} \leq \beta'$, so in

$$\text{Def} = \lambda[\mathcal{A}\mathcal{A}' + \lambda^{-1}\mathcal{B}\mathcal{A}'] + (1 - \lambda)[\mathcal{A}\mathcal{A}' + (1 - \lambda)^{-1}\mathcal{A}\mathcal{B}'] + \mathcal{B}\mathcal{B}'$$

both brackets are nonnegative, the second by the same argument with the factors exchanged, using $\Theta(z, w) = \Theta(w, z)$. Finally, by (i), the block weight of a three-letter word depends only on the set T of its three eigenvalues, because it is real and conjugated by reversal; call it $w_T \leq 0$. Hence $\mathcal{B}\mathcal{B}' = 36 \sum_{T, T'} w_T w'_{T'} \Psi(q_T; y_{T'})$, with $q_T, y_{T'}$ the eigenvalue triples in any order and $\Psi(z; w) = \frac{1}{6} \sum_{\pi \in S_3} K_3(z_\pi \circ w) - K_3(z)K_3(w) \geq 0$ by Theorem 6.5.3(b), and $w_T w'_{T'} \geq 0$. $\square$

Corollary 6.7.9 (Two families with negative cycles). *(a) Let $\rho = D - N$ and $\rho' = D' - N'$ in eigenbases of σ and σ' , with $D, D' \geq 0$ diagonal and N, N' entrywise non-negative with zero diagonal. If $D_{bb} \geq 2 \sum_c N_{bc}$ and $D'_{ee} \geq 2 \sum_f N'_{ef}$ for all b, e , the conjecture holds at $\alpha = 3$.*

(b) Let $d = d' = 3$, let σ, σ' have three distinct eigenvalues, and let the 3-cycles W, W' be real and negative. Put $r_{bc} = |\rho_{bc}|/\sqrt{\rho_{bb}\rho_{cc}}$ and $r_{\max}(\rho) = \max_{b \neq c} r_{bc}$, and likewise for ρ' . If $r_{\max}(\rho) + r_{\max}(\rho') \leq 1$, the conjecture holds at $\alpha = 3$.

Proof. (a) All block 3-cycles are $-\text{Tr}(P_a N P_b N P_c N) \leq 0$, and $D - 2N$ and $D' - 2N'$ are diagonally dominant, so Theorem 6.7.8 applies with $\beta = \beta' = 2$; entries of N inside an eigenspace of σ belong to Δ and only help. (b) In the proof of Theorem 6.7.8 the bracket $\mathcal{A}\mathcal{A}' + s\mathcal{B}\mathcal{A}'$ is nonnegative under the weaker condition that $\mathcal{D}_{\Delta+sO\|\sigma}[y_e, y_e, y_f] \geq 0$ for the two-letter words of σ' . Averaging over the position of the odd letter gives

$$\mathcal{D}_{\Delta+sO\|\sigma}[y_e, y_e, y_f] = 3s^2 y_e^{-2} \left[\sum_{b \neq c} w_{bbc} J_{bc} + 2sW J_{123} \right], \quad w_{bbc} = \rho_{bb} |\rho_{bc}|^2,$$

with $J_{bc} = J_{(q_b, q_b, q_c)}(y_f/y_e)$ and $J_{123} = J_{(q_1, q_2, q_3)}(y_f/y_e) \geq 0$. By Corollary 6.7.4, $J_{123} \leq \sum_c \sqrt{J_{ac} J_{bc}}$, and $2s|W|\sqrt{J_{ac} J_{bc}} \leq w_{aac} J_{ac} + w_{bbc} J_{bc}$ as soon as $s^2 W^2 \leq w_{aac} w_{bbc}$. Since $\sqrt{w_{aac} w_{bbc}}/|W| = 1/r_{ab}$, this holds for $s \leq 1/r_{\max}(\rho)$. Each ordered pair occurs at exactly one vertex, so the bracket is nonnegative. Take $\lambda = r_{\max}(\rho)$, so that $\lambda^{-1} = 1/r_{\max}(\rho)$ and $(1 - \lambda)^{-1} \leq 1/r_{\max}(\rho')$, and conclude as in Theorem 6.7.8. $\square$

Part (b) covers, for instance, the singular states $D_p^{1/2}(\frac{3}{2}\mathbf{1} - \frac{1}{2}ee^\top)D_p^{1/2}$, with $D_p = \text{diag}(p)$, $e = (1, 1, 1)$ and all $r_{bc} = \frac{1}{2}$, on both sides. For pairs with a nonnegative 3-cycle the conjecture is Corollary 6.7.7. What remains for qutrits is the chiral case and the strongly correlated negative case $r_{\max}(\rho) + r_{\max}(\rho') > 1$. Both reduce to one inequality for the kernel K_3 alone, on data that are not the cycle weights of any qutrit state.

For $p \in (0, \infty)^3$ and $\varphi \in \mathbb{R}$, the *twisted pure data* are $w_x^{p, \varphi} = p_{x_1} p_{x_2} p_{x_3} e^{i\varphi n(x)}$ on $x \in [3]^3$, where $n(x) = 1$ on the rotations of $(1, 2, 3)$, $n(x) = -1$ on the rotations of $(1, 3, 2)$ and $n(x) = 0$ otherwise. For $\varphi \notin 2\pi\mathbb{Z}$ they are the cycle weights of the indefinite Hermitian matrix $D_p^{1/2} T_\varphi D_p^{1/2}$, where T_φ has unit diagonal, $(T_\varphi)_{12} = (T_\varphi)_{23} = (T_\varphi)_{31} = e^{i\varphi/3}$ and

$(T_\varphi)_{21} = (T_\varphi)_{32} = (T_\varphi)_{13} = e^{-i\varphi/3}$. They are not the cycle weights of any qutrit state: for a state, $\rho_{bb}|\rho_{bc}|^2 = p_b^2 p_c$ for all $b \neq c$ and $|\rho_{12}\rho_{23}\rho_{31}| = p_1 p_2 p_3$ force $\rho_{bb} = p_b$ and $|\rho_{bc}|^2 = \rho_{bb}\rho_{cc}$, so ρ has rank one and its 3-cycle is positive.

Conjecture 6.7.10 (Twisted pure data). *For all $q, y \in (0, \infty)^3$ with distinct entries, all $p, p' \in (0, \infty)^3$ and all φ, φ' ,*

$$\sum_{x, x' \in [3]^3} w_x^{p, \varphi} w_{x'}^{p', \varphi'} \Theta(q_x, y_{x'}) \geq 0.$$

Proposition 6.7.11 (Reduction of the three-eigenvalue case). *(a) If Conjecture 6.7.10 holds, then Conjecture 6.3.5 holds at $\alpha = 3$ for all pairs in which σ and σ' each have at most three distinct eigenvalues, of any multiplicities. For given pairs only the eigenvalue triples of σ, σ' and the phases $\varphi = \arg W$, $\varphi' = \arg W'$ of their block 3-cycles are needed.*

(b) Conjecture 6.7.10 holds when $\varphi \in 2\pi\mathbb{Z}$ or $\varphi' \in 2\pi\mathbb{Z}$.

(c) At real phases, $\varphi, \varphi' \in \pi\mathbb{Z}$, Conjecture 6.7.10 reduces to the single case $\varphi = \varphi' = \pi$.

Proof. For data w, w' on $[3]^3$ write $\text{Def}(w, w') = \sum_{x, x'} w_x w_{x'}' \Theta(q_x, y_{x'})$. By (6.7.1) and its symmetric form this is the defect when w, w' are block weights, and it is bilinear. For a two-letter word x and genuine block weights w' , the coefficient $\sum_{x'} w_{x'}' \Theta(q_x, y_{x'}) = \mathcal{D}_{\rho' \| \sigma'}[q_x]$ is nonnegative.

For twisted data $w^{p, \varphi}$ and a two-letter word x' with eigenvalues (y_e, y_e, y_f) , the cyclic invariance of the data and of Θ gives

$$\sum_x w_x^{p, \varphi} \Theta(q_x, y_{x'}) = y_e^{-2} \left[3 \sum_{b \neq c} p_b^2 p_c J_{bc} + 6 p_1 p_2 p_3 \cos \varphi J_{123} \right] \geq 0,$$

where $J_{bc} = J_{(q_b, q_b, q_c)}(y_f/y_e)$ and $J_{123} = J_{(q_1, q_2, q_3)}(y_f/y_e)$. The inequality holds because $0 \leq J_{123} \leq \sum_c \sqrt{J_{ac} J_{bc}}$ (Corollary 6.7.4) and $2 p_1 p_2 p_3 \sqrt{J_{ac} J_{bc}} \leq p_a^2 p_c J_{ac} + p_b^2 p_c J_{bc}$.

(a) If σ or σ' has at most two eigenvalues, use Corollary 6.6.3. If $W = 0$, the defect is the sum of its two-letter terms in w , which are nonnegative; similarly if $W' = 0$. Otherwise Lemma 6.7.6 gives p with $p_b^2 p_c \leq w_{bbc}$ and $p_1 p_2 p_3 e^{i\varphi} = W$. Replacing the two-letter weights of w by $p_b^2 p_c$ does not increase Def , because their coefficients are nonnegative, and leaves the three-letter part unchanged. So $\text{Def}(w, w') \geq \text{Def}(w^{p, \varphi}, w')$. In the same way, using the nonnegative coefficients of the display, $\text{Def}(w^{p, \varphi}, w') \geq \text{Def}(w^{p, \varphi}, w^{p', \varphi'}) \geq 0$.

(b) If $\varphi' = 0$, then $w^{p', 0}$ are the cycle weights of the pure partner $\sqrt{p'} \sqrt{p'}^\top$ against $\text{diag}(y)$, and $\text{Def} = \sum_x w_x^{p, \varphi} G(q_x)$ with $G = (\sum_e p'_e)^3 G_Y$, where G_Y is the pure-partner kernel of Corollary 6.7.4 for the law $p'/\sum_e p'_e$ on $\{y_e\}$. That corollary and the argument of the display give $3 \sum p_b^2 p_c G_{bbc} + 6 p_1 p_2 p_3 \cos \varphi G_{123} \geq 0$. The case $\varphi = 0$ follows from $\Theta(z, w) = \Theta(w, z)$.

(c) The data depend on the phases only through $e^{i\varphi}$ and $e^{i\varphi'}$, so at real phases there are four cases, and three of them are covered by (b). $\square$

Numerically Conjecture 6.7.10 holds with a margin. Over 8,000 random instances, the minimum over a grid of phases, divided by the value at $\varphi = \varphi' = 0$, is +0.0079, always attained at $\varphi = \varphi' = \pi$, and it approaches 0 only at degenerate configurations (Section 6.11).

A polynomial criterion

The kernel K_3 can be expanded into polynomial moments, which gives a sufficient condition for the conjecture at $\alpha = 3$ in terms of finitely many traces at each order. For $\rho \geq 0$ and $A \geq 0$ put

$$\mathcal{Q}_k(\rho \| A) = \binom{k+2}{2}^{-1} \sum_{i+j+l=k} \text{Tr} [A^i \rho A^j \rho A^l \rho],$$

$$D_k = \mathcal{Q}_k(\rho \otimes \rho' \| A \otimes B) - \mathcal{Q}_k(\rho \| A) \mathcal{Q}_k(\rho' \| B),$$

with $D_k = D_k(\rho, A; \rho', B)$. With $T_\lambda(\rho, A) = \text{Tr}[A^{\lambda_1} \rho A^{\lambda_2} \rho A^{\lambda_3} \rho]$ and λ uniform on the $\binom{k+2}{2}$ compositions of k into three parts, $\mathcal{Q}_k(\rho \otimes \rho' \| A \otimes B)$ is the mean of $T_\lambda(\rho, A) T_\lambda(\rho', B)$. So D_k is the covariance of $T_\lambda(\rho, A)$ and $T_\lambda(\rho', B)$. It vanishes when $[\rho, A] = 0$ or $[\rho', B] = 0$, and $D_0 = D_1 = 0$.

Theorem 6.7.12 (Polynomial reduction). *Let σ, σ' be faithful. For $t > 0$ put $s_t = t \|\sigma\|_\infty$, $A_t = s_t \mathbf{1} - t\sigma \geq 0$, $B_t = \sigma'(\mathbf{1} + s_t \sigma')^{-1}$ and $\rho'_t = (\mathbf{1} + s_t \sigma')^{-1/2} \rho' (\mathbf{1} + s_t \sigma')^{-1/2}$. Then*

$$Q_3^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') - Q_3^{\text{HT}}(\rho \| \sigma) Q_3^{\text{HT}}(\rho' \| \sigma') = 2 \int_0^\infty t \sum_{k \geq 2} \binom{k+2}{2} D_k(\rho, A_t; \rho'_t, B_t) dt.$$

Consequently, if $D_k(\rho, A; \rho', B) \geq 0$ for all integers $k \geq 2$ and all positive semidefinite ρ, ρ', A, B , then Conjecture 6.3.5 holds at $\alpha = 3$.

Proof. By the second form of K_3 , $K_3(z \circ w) = 2 \int_0^\infty t \prod_i (1 + t z_i w_i)^{-1} dt$. Also $K_3(z) K_3(w) = \mathbb{E}_U[(U \cdot z)^{-2} K_3(w)] = \mathbb{E}_U K_3((U \cdot z)w) = 2 \int_0^\infty t \mathbb{E}_U \prod_j (1 + t(U \cdot z)w_j)^{-1} dt$. So the defect is $2 \int_0^\infty t \Delta_t dt$ with $\Delta_t = \sum_{x, x'} W_x W'_{x'} k(tq_x, y_{x'})$ and

$$k(s, y) = \prod_i \frac{1}{1 + s_i y_i} - \mathbb{E}_U \prod_j \frac{1}{1 + (U \cdot s) y_j}.$$

For $s_0 > 0$, $1 + s_i y_j = (1 + s_0 y_j)(1 + (s_i - s_0)y'_j)$ with $y'_j = y_j/(1 + s_0 y_j)$, and $U \cdot s - s_0 = U \cdot (s - s_0 \mathbf{1})$. So $k(s, y) = \prod_j (1 + s_0 y_j)^{-1} k(s - s_0 \mathbf{1}, y')$. The factor $\prod_j (1 + s_0 y_{x'_j})^{-1}$ turns $W'_{x'}(\rho')$ into $W'_{x'}(\rho'_t)$ when $s_0 = s_t$. Put $\tilde{s} = s_t \mathbf{1} - tq_x$, the eigenvalue word of A_t , and note that $0 \leq \tilde{s}_i y'_j < 1$. The Hermite–Genocchi formula of Theorem 6.7.1 gives $\prod_i (1 - x_i)^{-1} = \mathbb{E}_V(1 - V \cdot x)^{-3}$ for $x \in [0, 1]^3$. Expanding $(1 - u)^{-3} = \sum_k \binom{k+2}{2} u^k$, which converges uniformly on the relevant compact range, gives

$$k(-\tilde{s}, y') = \sum_{k \geq 0} \binom{k+2}{2} \left[\mathbb{E}_V(V \cdot (\tilde{s} \circ y'))^k - \mathbb{E}_U(U \cdot \tilde{s})^k \mathbb{E}_V(V \cdot y')^k \right].$$

Since $\mathbb{E}_U(U \cdot a)^k = \binom{k+2}{2}^{-1} h_k(a)$ with h_k the complete homogeneous symmetric polynomial, (6.6.1) turns $\sum_x W_x \mathbb{E}_U(U \cdot a_x)^k$ into $\mathcal{Q}_k(\rho \| A_t)$, and likewise on the product space. Summing over the finitely many words gives $\Delta_t = \sum_k \binom{k+2}{2} D_k(\rho, A_t; \rho'_t, B_t)$, and the terms $k = 0, 1$ vanish. $\square$

Proposition 6.7.13 (The first two orders). $D_2(\rho, A; \rho', B) = \frac{1}{4} E(\rho, A) E(\rho', B) \geq 0$, where $E(\rho, A) = \text{Tr}[\rho^3 A^2] - \text{Tr}[\rho^2 A \rho A] = \text{Tr}[[A, \rho]^* \rho [A, \rho]]$. For $k = 3$, with $T_{300} = \text{Tr}[\rho^3 A^3]$, $T_{210} = \text{Tr}[A^2 \rho A \rho^2]$, $T_{111} = \text{Tr}[(A \rho)^3]$, $u = T_{300} - \text{Re } T_{210}$, $v = \text{Re } T_{210} - T_{111}$ and primed quantities for (ρ', B) ,

$$D_3 = \frac{18}{100} u u' + \frac{3}{100} (u + v)(u' + v') + \frac{6}{100} v v' - \frac{6}{10} \text{Im } T_{210} \text{Im } T'_{210}.$$

Here $u + v$ is the Lieb–Thirring gap and $u - v = \text{Tr}[AK\rho K] \geq 0$ with $K = -i[A, \rho]$, so $u \geq |v|$. Consequently $D_3 \geq 0$ whenever

$$|\text{Im Tr}[A^2 \rho A \rho^2]| \leq \frac{1}{\sqrt{20}} (\text{Tr}[\rho^3 A^3] - \text{Tr}[(\rho A)^3]) \quad (6.7.2)$$

holds for (ρ, A) and for (ρ', B) .

Proof. On the compositions of 2, T_λ equals $\text{Tr}[\rho^3 A^2]$ on the three rotations of $(2, 0, 0)$ and the real number $\text{Tr}[\rho^2 A \rho A]$ on the three rotations of $(1, 1, 0)$. A two-point covariance with equal masses is a quarter of the product of the differences. The commutator form of E follows by expansion, and $E \geq 0$ by (6.6.2). On the ten compositions of 3, T_λ equals T_{300} on the three rotations of $(3, 0, 0)$, T_{210} on the three rotations of $(2, 1, 0)$, $\overline{T_{210}}$ on those of $(2, 0, 1)$, and T_{111} at $(1, 1, 1)$. The covariance splits into the covariance of the real parts, a three-point covariance with masses $(\frac{3}{10}, \frac{6}{10}, \frac{1}{10})$, minus that of the imaginary parts, which have mean zero; this gives the formula. Expanding $\text{Tr}[A(A\rho - \rho A)\rho(A\rho - \rho A)]$ gives $v - u$, so $u - v = \text{Tr}[AK\rho K] = \text{Tr}[A^{1/2}K\rho K A^{1/2}] \geq 0$. Under (6.7.2), $\frac{6}{10}|\text{Im}T_{210}\text{Im}T'_{210}| \leq \frac{3}{100}(u + v)(u' + v')$, so $D_3 \geq \frac{18}{100}uu' + \frac{6}{100}vv' \geq \frac{12}{100}uu' \geq 0$. $\square$

In our tests $D_k \geq 0$ for $k = 2, \dots, 12$, and the ratio in (6.7.2) stayed below 0.07, against the 0.224 allowed. Our attempts to certify (6.7.2) by universal rooted kernels of the type of Proposition 6.7.3 failed: at very spread spectra the constants such a kernel must satisfy can be met only with equality. A proof will probably have to use the positivity of ρ across different roots.

6.8 What the conjecture is a shadow of: the Frenkel measure and the convex order

We now identify the measure that controls Conjecture 6.3.5 for all α at once. In this section all states are faithful, so that $m_0 = m_\infty = 0$ and all quantities below are finite.

Definition 6.8.1 (Frenkel measure). The *Frenkel measure* of (ρ, σ) is $\nu_{\rho\|\sigma}(dv) = g(v)v^{-2}dv$ on $(0, \infty)$, where $g(v) = e(v) - (1 - v)_+$.

The measure is positive and supported in $[\gamma_{\min}, \gamma_{\max}]$, the convex hull of the relative spectrum, because $e(v) = 1 - v$ below $\gamma_{\min}$ and $e = 0$ above $\gamma_{\max}$. By the first formula of Proposition 5.2.5,

$$D_f^{\text{HT}}(\rho\|\sigma) = \int v^2 f''(v) \nu_{\rho\|\sigma}(dv) \quad \text{for every convex } f,$$

so ν has total mass $D(\sigma\|\rho)$ (take $f = -\log$), first moment $D(\rho\|\sigma)$ (take $f = v \log v$; this is Frenkel's formula [46]), and $Q_\alpha^{\text{HT}} = 1 + \alpha(\alpha - 1) \int v^\alpha d\nu$ for every real α . For a classical pair (P, Q) with $Q(dv) = \mu(dv)$, $P(dv) = v\mu(dv)$ the same formulas hold with the classical f -divergence, since both are $\int f d\mu$.

Let (ρ, σ) and (ρ', σ') be two pairs, (P, Q) and (P', Q') the classical pairs carried by their shadows (Theorem 6.4.2), and write

$$e_{12}(v) = E_v(\rho \otimes \rho' \|\sigma \otimes \sigma'), \quad e_{\text{cl}}(v) = E_v(P \otimes P' \|\! Q \otimes Q') = \int \text{Tr}(\lambda' \rho - v\sigma)_+ d\mu'(\lambda'),$$

and $\delta = e_{12} - e_{\text{cl}}$, and $\nu_{12}, \nu_{\text{cl}}$ for the Frenkel measures of the two product pairs, so that $\nu_{12} - \nu_{\text{cl}}$ has density $\delta(v)v^{-2}$. Recall that a finite measure ν_2 dominates ν_1 in the *convex order*, $\nu_2 \geq_{\text{cx}} \nu_1$, if $\int \psi d\nu_2 \geq \int \psi d\nu_1$ for every convex ψ ; this forces equal masses and equal first moments.

Theorem 6.8.2 (The conjecture as a moment inequality). *In this setting:*

- (a) ν_{12} and ν_{cl} have the same mass $D(\sigma\|\rho) + D(\sigma'\|\rho')$ and the same first moment $D(\rho\|\sigma) + D(\rho'\|\sigma')$, and for every real $\alpha \neq 0, 1$,

$$\begin{aligned} Q_\alpha^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') - Q_\alpha^{\text{HT}}(\rho \| \sigma) Q_\alpha^{\text{HT}}(\rho' \| \sigma') &= \alpha(\alpha - 1) \int v^\alpha d(\nu_{12} - \nu_{\text{cl}})(v) \\ &= \alpha(\alpha - 1) \int_0^\infty v^{\alpha-2} \delta(v) dv. \end{aligned}$$

So Conjecture 6.3.5 says that $M(s) = \int_0^\infty v^{s-1} \delta(v) dv$ has the sign of $s(s+1)$ for all s ; it vanishes at $s = 0$ and $s = -1$.

- (b) If $\nu_{12} \geq_{\text{cx}} \nu_{\text{cl}}$, then Conjecture 6.3.5 holds for the two pairs and all $\alpha \in \mathbb{R} \setminus \{0, 1\}$. More generally, $D_f^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq D_f(P \otimes P' \| Q \otimes Q')$ for every convex f such that $v \mapsto v^2 f''(v)$ is convex, and the reverse inequality holds when $v^2 f''$ is concave. Equality holds for $f = v \log v$ and $f = -\log v$.

- (c) The convex-order statement is equivalent to $U(c) \geq 0$ for all $c > 0$, where

$$U(c) = \int_c^\infty \frac{(v-c)\delta(v)}{v^2} dv = D_{f_c}(\rho \otimes \rho' \| \sigma \otimes \sigma') - D_{f_c}(P \otimes P' \| Q \otimes Q'),$$

with $f_c(v) = (v+c) \log v$ on $[c, \infty)$, continued affinely and C^1 below c . Equivalently, with the single function $j(u) = ((u+1) \log u - 2(u-1))_+$, the Jeffreys kernel cut off at the Helstrom point,

$$U(c) = D_j^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') - D_j(P \otimes P' \| Q \otimes Q') \quad (c > 0).$$

Here $U(0) = U(\infty) = U'(0) = U'(\infty) = 0$ and $U''(c) = \delta(c)/c^2$.

- (d) If δ changes sign at most twice on $(0, \infty)$ and both pairs are non-commuting, then $\nu_{12} \geq_{\text{cx}} \nu_{\text{cl}}$, δ changes sign exactly twice, and it is positive near both ends of its support and negative in between.

Proof. (a) Classical Rényi quasi-entropies are multiplicative and relative entropies are additive, so the moments of ν_{cl} are $\int v^\alpha d\nu_{\text{cl}} = (Q_\alpha^{\text{HT}} Q_\alpha'^{\text{HT}} - 1)/(\alpha(\alpha - 1))$, and the mass and first moment are the sums of the relative entropies. For ν_{12} use additivity of D under tensor products. Subtracting the two formulas $Q_\alpha = 1 + \alpha(\alpha - 1) \int v^\alpha d\nu$ gives the identity. The zeros of M at $s = 0$ and $s = -1$ are the equalities of first moments and of masses.

(b) Apply the definition of the convex order to $\psi = v^2 f''$, which is admissible because both measures have compact support in $(0, \infty)$. For f_α , $v^2 f_\alpha'' = \alpha v^\alpha$ is convex for $\alpha \notin (0, 1)$ and concave for $\alpha \in (0, 1)$; in both cases $\alpha(\alpha - 1) \int v^\alpha d(\nu_{12} - \nu_{\text{cl}}) \geq 0$. For $f = v \log v$ and $f = -\log v$ the function $v^2 f''$ is affine, and both inequalities hold.

(c) For measures with equal mass and mean, the convex order is equivalent to $\int (v - c)_+ d(\nu_{12} - \nu_{\text{cl}}) \geq 0$ for all c , since the functions $(v - c)_+$ and the affine functions generate the convex cone. This is $U(c) \geq 0$. The boundary values follow from (a): $U(0) = M(0)$, $U'(c) = -\int_c^\infty \delta v^{-2} dv$, so $U'(0) = -M(-1)$, and both vanish. The identification of f_c : $f_c''(v) = (v - c)_+/v^2$, and $(v + c) \log v$ has second derivative $1/v - c/v^2$. For the second form, j vanishes on $(0, 1]$, $j(1) = j'(1) = 0$ and $j''(u) = (u - 1)_+/u^2$, so only the first family of hockey sticks enters Definition 5.2.3, and $E_u(\rho \| c\sigma) = E_{cu}(\rho \| \sigma)$ with the substitution $v = cu$ gives $D_j^{\text{HT}}(\rho \| c\sigma) = \int_c^\infty (v - c)v^{-2} e(v) dv$.

(d) Suppose $\delta \not\equiv 0$ changes sign at most twice. U is C^1 with $U'' = \delta/c^2$ and vanishes with its derivative at 0 and ∞ . If δ had no sign change, U would be convex or concave

with zero boundary values, hence $U \equiv 0$ and $\delta \equiv 0$. If δ had one sign change, U' would be monotone on each side of it and vanish at both ends, hence of one sign, and U would be monotone with zero boundary values; again $\delta \equiv 0$. So δ changes sign exactly twice, at $a < b$. If $\delta > 0$ on $(0, a)$, then U is convex on $(0, a)$ and on (b, ∞) and concave on (a, b) . On $(0, a)$, $U(0) = U'(0) = 0$ gives $U \geq 0$; on (b, ∞) , convexity with $U(\infty) = U'(\infty) = 0$ gives $U \geq 0$; on (a, b) concavity gives $U \geq \min\{U(a), U(b)\} \geq 0$. If instead $\delta < 0$ on $(0, a)$, the same argument gives $U \leq 0$, hence $\int \psi d(\nu_{12} - \nu_{cl}) \leq 0$ for all convex ψ , in particular $Q_2^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \leq Q_2^{\text{HT}} Q_2^{\text{HT}}$, which contradicts the strict inequality of Theorem 6.4.1 for non-commuting pairs. $\square$

Theorem 6.8.2 identifies the structure behind Conjecture 6.3.5. The quantity that tensorizes badly is the Frenkel measure, and the conjecture says that its Mellin transform is super-multiplicative at every order. The natural stronger statement is domination in the convex order,

$$\nu_{\rho \otimes \rho' \| \sigma \otimes \sigma'} \geq_{\text{cx}} \nu_{P \otimes P' \| Q \otimes Q'}, \quad (6.8.1)$$

which says that a quantum product pair is a mean-preserving spread of the classical product of the two shadows: the same relative entropies in both directions, more mass at the ends of the testing region. The one-shot picture behind (6.8.1) is the sign structure of δ . Near both ends of the support $\delta \geq 0$ to second order when the extreme relative eigenvalues are simple, by Theorem 6.4.3 applied to the pairs and to the reversed pairs ($\delta(v) = v \delta_{\text{rev}}(1/v)$ by Lemma 5.2.2(c)); this is where entangled tests beat every test built from one factor at a time. In between, $\delta < 0$ on a middle interval: the classical product of the shadows, which tests the first factor by its Helstrom test at the level set by an ideal measurement of the second, is easier to discriminate there than the quantum product. The two sign changes are the one-shot counterparts of the three asymptotic comparisons: for n copies and $v = e^{nr}$, the quantum product loses in the Chernoff and Hoeffding regimes, where its exponents are governed by $D_\alpha^{\text{P}} \leq D_\alpha^{\text{HT}}$ ($\alpha < 1$), and wins in the strong-converse regime, where they are governed by $\tilde{D}_\alpha \geq D_\alpha^{\text{HT}}$ ($\alpha > 1$), by [11, 55, 94, 96] and the ordering of Proposition 5.5.3.

Numerically (Section 6.11), δ changes sign exactly twice for every pair of qubit pairs we generated, so (d) applies and (6.8.1) holds for them. In dimensions 2×3 and higher, δ can change sign four, six and eight times, so (d) does not apply, but $U \geq 0$ held in every case, including adversarial searches. We therefore record (6.8.1) as the conjecture behind Conjecture 6.3.5. Theorem 6.4.1 is the case $\psi(v) = v^2$ of (6.8.1), and the second-order statement at $\alpha = 1$ is the case $\psi = v \log v$: superadditivity of the testing-region variance $\int v \log^2 v d\mu - D^2$.

Finally, the conjecture does hold for tensor powers once there are enough copies, with an explicit threshold, because the regularised values of Theorem 5.7.16 lie on the right side of the one-shot value.

Proposition 6.8.3 (Tensor powers). *Let ρ, σ be states on $\mathbb{C}^d$ with $\Pi_\rho \leq \Pi_\sigma$.*

- (a) *If $\alpha > 1$ and $D_\alpha^{\text{HT}}(\rho \| \sigma) < \tilde{D}_\alpha(\rho \| \sigma)$, then $Q_\alpha^{\text{HT}}(\rho^{\otimes n} \| \sigma^{\otimes n}) > Q_\alpha^{\text{HT}}(\rho \| \sigma)^n$ for every n with $n(\tilde{D}_\alpha - D_\alpha^{\text{HT}}) > \frac{\alpha(d-1)}{\alpha-1} \log(n+1)$.*
- (b) *If $\alpha \in (0, 1)$ and $D_\alpha^{\text{HT}}(\rho \| \sigma) > D_\alpha^{\text{P}}(\rho \| \sigma)$, then the same holds for every n with $n(D_\alpha^{\text{HT}} - D_\alpha^{\text{P}}) > \frac{\log 2}{1-\alpha}$.*

The hypotheses hold for every non-commuting pair in our tests, and for pure ρ and $\alpha > 1$ they hold whenever $[\rho, \sigma] \neq 0$, by the strict Minkowski inequality in Remark 6.4.4.

Proof. (a) Proposition 5.6.2 for n copies gives $\frac{1}{n}D_\alpha^{\text{HT}}(\rho^{\otimes n}\|\sigma^{\otimes n}) \geq \tilde{D}_\alpha(\rho\|\sigma) - \frac{\alpha(d-1)}{\alpha-1} \frac{\log(n+1)}{n}$, which exceeds $D_\alpha^{\text{HT}}(\rho\|\sigma)$ under the stated condition; for $\alpha > 1$ this is the claim. (b) Proposition 5.6.1 gives $D_\alpha^{\text{HT}}(\rho^{\otimes n}\|\sigma^{\otimes n}) \leq nD_\alpha^{\text{P}}(\rho\|\sigma) + \frac{\log 2}{1-\alpha} < nD_\alpha^{\text{HT}}(\rho\|\sigma)$, and for $\alpha < 1$ super-multiplicativity of Q is subadditivity of D . $\square$

6.9 Why channels cannot close the gap: the shadow discrepancy and restricted tests

The standard tools for comparing statistical experiments act through positive maps. These include Blackwell's theorem and Le Cam deficiencies, the Alberti–Uhlmann theorem for qubits [1], Buscemi's comparison with quantum side information [25], Matsumoto's reverse test [91], and Gour's bracketing of quantum Lorenz curves by classical polygons [48]. In Gour's language, inclusion of testing regions is the convex order of the shadows μ themselves [48, Cor. 7], while (6.8.1) is the convex order of the Frenkel measures, which sits two integrations higher. The next two results show that tools of this kind cannot prove Conjecture 6.3.5 on their own, and they locate the missing ingredient in entangled tests.

Proposition 6.9.1 (The shadow discrepancy is two-sided). *Let (ρ, σ) and (ρ', σ') be pairs of faithful states, neither commuting, and let (P', Q') be the classical pair carried by the shadow of (ρ', σ') . Then $\delta = e_{12} - e_{\text{cl}}$ changes sign at least twice. In particular:*

- (a) *there are levels $v_\pm$ with $E_{v_+}(\rho \otimes \rho' \|\sigma \otimes \sigma') > E_{v_+}(\rho \otimes P' \|\sigma \otimes Q')$ and $E_{v_-}(\rho \otimes \rho' \|\sigma \otimes \sigma') < E_{v_-}(\rho \otimes P' \|\sigma \otimes Q')$;*
- (b) *there is no positive trace-preserving map Λ with $\Lambda(P') = \rho'$, $\Lambda(Q') = \sigma'$, and none with $\Lambda(\rho') = P'$, $\Lambda(\sigma') = Q'$, although (ρ', σ') and (P', Q') have the same testing region. The ancilla pair (ρ, σ) witnesses both failures at once.*

Proof. By Theorem 6.8.2(a), $\int_0^\infty \delta(v)v^{-2}dv = M(-1) = 0$ and $Q_2^{\text{HT}}(\rho \otimes \rho' \|\sigma \otimes \sigma') - Q_2^{\text{HT}}Q_2^{\text{HT}} = 2 \int_0^\infty \delta(v)dv$. The right-hand side is strictly positive by the equality case of Theorem 6.4.1. So $\delta > 0$ somewhere and, because of the vanishing mass, $\delta < 0$ somewhere. The argument of Theorem 6.8.2(d) shows that a nonzero δ with $M(0) = M(-1) = 0$ cannot change sign only once. This proves (a), since $e_{\text{cl}}(v) = E_v(\rho \otimes P' \|\sigma \otimes Q')$ by the testing-region equality of (ρ, σ) and (P, Q) . (b) A positive map whose domain is commutative, or whose range is commutative, is completely positive. If Λ maps (P', Q') to (ρ', σ') , then $\text{id} \otimes \Lambda$ maps $(\rho \otimes P', \sigma \otimes Q')$ to the product pair, and data processing gives $e_{12} \leq e_{\text{cl}}$, which contradicts v_+ . If Λ is a measurement producing (P', Q') , Theorem 6.4.2 with that measurement gives $e_{12} \geq e_{\text{cl}}$, which contradicts v_- . $\square$

Part (b) is known in other forms. The first half follows from the minimality of the reverse test [91] and Corollary 5.3.3, and the second half from the strict inequality between measured and testing-region divergences; [89] gives an example of the same phenomenon. The new point is (a): a single product with any non-commuting ancilla shows that the shadow is incomparable with the pair in both directions. So Conjecture 6.3.5 cannot be proved by sending either factor to or from its shadow through a positive map, nor by any pointwise comparison of profiles. The comparison has to be integrated, and the next result says which tests make it possible.

For a level $v \geq 0$ put $h_v(\lambda) = \text{Tr}(\lambda\rho - v\sigma)_+$ and $h'_v(\lambda) = \text{Tr}(\lambda\rho' - v\sigma')_+$. Both are convex and positively homogeneous in $\lambda \geq 0$. A *one-way LOCC test* measures one factor with a POVM $\{N_y\}$ and then applies a test M_y on the other factor that depends on the outcome, so that $M = \sum_y M_y \otimes N_y$ or $M = \sum_y N_y \otimes M_y$. A *separable test* is one with both M and $\mathbf{1} - M$ separable.

Proposition 6.9.2 (One-way tests never beat the shadow product). *In the setting of Theorem 6.8.2,*

$$e_{\text{cl}}(v) = D_{h_v}^{\text{HT}}(\rho' \parallel \sigma') = D_{h'_v}^{\text{HT}}(\rho \parallel \sigma) \quad (v \geq 0),$$

and every one-way LOCC test M , in either direction, satisfies $\text{Tr } M(\rho \otimes \rho' - v \sigma \otimes \sigma') \leq e_{\text{cl}}(v)$. Consequently, at every level v_+ of Proposition 6.9.1 the product pair has a test that beats every one-way LOCC test by at least $\delta(v_+) > 0$. Such levels exist for every two non-commuting pairs.

Proof. For the classical product, $E_v(P \otimes P' \parallel Q \otimes Q') = \int \text{Tr}(\lambda \rho - v \sigma)_+ Q'(d\lambda) = D_{h_v}(P' \parallel Q')$, where $P' = \lambda Q'$ on the finite part. By Proposition 5.2.5, this equals $D_{h_v}^{\text{HT}}(\rho' \parallel \sigma')$; the second identity is the same computation with the roles exchanged. For $M = \sum_y M_y \otimes N_y$ with $a_y = \text{Tr } N_y \rho'$ and $b_y = \text{Tr } N_y \sigma'$,

$$\begin{aligned} \text{Tr } M(\rho \otimes \rho' - v \sigma \otimes \sigma') &= \sum_y \text{Tr } M_y(a_y \rho - v b_y \sigma) \\ &\leq \sum_y \text{Tr}(a_y \rho - v b_y \sigma)_+ = D_{h_v}(a \parallel b) \leq D_{h_v}^{\text{HT}}(\rho' \parallel \sigma') \end{aligned}$$

by data processing for $D_{h_v}^{\text{HT}}$ under the measurement N . The other direction uses h'_v . $\square$

Proposition 6.9.2 is in the spirit of Piani's inequalities for relative entropies under one-way LOCC measurements [105], now for every hockey stick and against the shadow. Numerically the bound extends much further, as the following conjecture records.

Conjecture 6.9.3 (Separable tests). *Every separable test satisfies $\text{Tr } M(\rho \otimes \rho' - v \sigma \otimes \sigma') \leq e_{\text{cl}}(v)$ for all $v \geq 0$. Equivalently, the testing region of $(\rho \otimes \rho', \sigma \otimes \sigma')$ under separable tests is contained in the testing region of the classical product of the shadows $(P \otimes P', Q \otimes Q')$.*

We tested the stronger statement with PPT tests ($M, \mathbf{1} - M, M^\Gamma, (\mathbf{1} - M)^\Gamma \geq 0$), which coincide with separable tests in dimensions 2×2 and 2×3 [65]. It held for 97 random product pairs in dimensions 2×2 , 2×3 and 3×3 , and against adversarial search in 2×2 , to the accuracy of the semidefinite solver (Section 6.11). It is not a consequence of Proposition 6.9.2: computed exactly by a linear program over measurements on a great circle, the one-way LOCC value stays below the PPT value by up to two thirds of the gap $e_{\text{cl}} - e_{\text{PPT}}$.

Conjecture 6.9.3 has a Blackwell form. After refinement a separable test is a coarse-graining of a POVM $\{G_j \otimes H_j\}$ with product elements. The claim is that the outcome experiment $p_j = \text{Tr } G_j \rho$, $q_j = \text{Tr } H_j \sigma'$ is a garbling of $(P \otimes P', Q \otimes Q')$, that is, the law of $x_j x'_j$ under q is dominated in convex order by the law of $\lambda \lambda'$ under $Q \otimes Q'$, where $x_j = \text{Tr } G_j \rho / \text{Tr } G_j \sigma$ and $x'_j = \text{Tr } H_j \rho' / \text{Tr } H_j \sigma'$. For every state τ on the second factor, $\{\text{Tr}(H_j \tau) G_j\}_j$ is a POVM on the first. This controls the law of x conditionally on any grouping of the outcomes in which the H_j add up to a POVM element, which is the case for one-way protocols, and it then gives the claim by conditioning and Blackwell's theorem on each factor. For general separable POVMs no such grouping exists. A proof would need a “quasi-preparation” of the shadow that is positive only on the cone generated by the G_j , which is a Blackwell problem for a polyhedral operator system.

Taken together, Propositions 6.9.1–6.9.2 and Conjecture 6.9.3 split the profile difference as

$$\delta = \underbrace{(e_{12} - e_{\text{SEP}})}_{\geq 0: \text{ gain of entangled tests}} - \underbrace{(e_{\text{cl}} - e_{\text{SEP}})}_{\geq 0 \text{ (conjecturally): deficit of separable tests}},$$

so Conjecture 6.3.5 says that, in the Frenkel convex order, entangled tests on a product pair gain more than separable tests lose against the classical product of the shadows. This is the opposite of quantum data hiding [92], where entangled tests gain a lot against a restricted class. Here the gain is of second order, near the two ends of the testing region (Theorem 6.4.3), and the restricted class is benchmarked by the shadow and not by the full product. Any proof must therefore use the entangled optimal tests, as the proof of Theorem 6.4.1 does implicitly through the symmetrisation $v \mapsto -v$, which pairs the matrix element ρ'_{lm} with ρ'_{ml} across the tensor product.

Theorems 6.3.2 and 6.4.1 also settle the sign of the regularisation gap $\Delta_\alpha^+ = \tilde{D}_\alpha - D_\alpha^{\text{HT}}$ of [48, (45)] at $\alpha = 2$ for all non-commuting pairs, not only for qubits.

Corollary 6.9.4 (Strict regularisation gaps). *Let ρ, σ be faithful with $[\rho, \sigma] \neq 0$.*

- (a) $D_2^{\text{HT}}(\rho\|\sigma) < \tilde{D}_2(\rho\|\sigma)$.
- (b) *If the largest eigenvalue $\gamma_{\max}$ of $\sigma^{-1/2}\rho\sigma^{-1/2}$ is simple and its eigenvector v (with $\rho v = \gamma_{\max}\sigma v$) is not an eigenvector of σ , then $D_\alpha^{\text{HT}}(\rho\|\sigma) < \tilde{D}_\alpha(\rho\|\sigma)$ for all sufficiently large α .*

Proof. $\tilde{D}_\alpha$ is additive and dominates D_α^{HT} for $\alpha > 1$ ([84]; for $\alpha \geq 2$ also Theorem 6.3.2), so

$$\tilde{D}_\alpha(\rho\|\sigma) = \frac{1}{2}\tilde{D}_\alpha(\rho^{\otimes 2}\|\sigma^{\otimes 2}) \geq \frac{1}{2}D_\alpha^{\text{HT}}(\rho^{\otimes 2}\|\sigma^{\otimes 2}).$$

It therefore suffices that $Q_\alpha^{\text{HT}}(\rho^{\otimes 2}\|\sigma^{\otimes 2}) > Q_\alpha^{\text{HT}}(\rho\|\sigma)^2$. At $\alpha = 2$ this is the equality case of Theorem 6.4.1. For (b), by Theorem 6.4.3 it suffices that $\kappa_{12} > 2\kappa$ for the doubled pair. In the proof of Theorem 6.4.3, $\langle z|A^+|z\rangle - (2\text{Re}\langle x|z\rangle - \langle x|A|x\rangle) = \langle x - x_*|A|x - x_*\rangle$ with $Ax_* = z = a_{12}$, and the chosen x satisfies $Ax = a_{12} + a \otimes a'$. Hence the inequality $\langle a_{12}|A^+|a_{12}\rangle \geq \frac{mc'}{\gamma_{\max}} + \frac{m'c}{\gamma_{\max}}$ is strict as soon as $a \otimes a' \neq 0$. For the doubled pair $a' = a$, and $a \neq 0$ exactly when $\sigma v \notin \mathbb{C}v$. $\square$

6.10 Modular clocks, dephasing flows and two qubit pairs

The results so far treat Conjecture 6.3.5 through measures on the testing region. A second family of measures lives on the time axis of the modular group of σ . Write $\sigma = \sum_j e^{x_j} |j\rangle\langle j|$. In this basis $\sigma^{iT} \rho \sigma^{-iT}$ has entries $\rho_{jk} e^{iT\omega_{jk}}$ with frequencies $\omega_{jk} = x_j - x_k$. Averaging over a random time T with law π multiplies each entry by the characteristic function $\hat{\pi}(\omega_{jk})$. For product pairs, $(\sigma \otimes \sigma')^{iT} = \sigma^{iT} \otimes \sigma'^{iT}$, so both factors are driven by one clock.

Proposition 6.10.1 (The logistic clock). *Let σ be faithful, $Y = \sigma^{-1/4} \rho \sigma^{-1/4}$ and*

$$g_{\rho, \sigma}(t) = \text{Tr} [Y \sigma^{it} Y \sigma^{-it}] = \sum_{jk} |Y_{jk}|^2 \cos(\omega_{jk} t).$$

Let T have the logistic density $\frac{\pi}{2} \text{sech}^2(\pi t)$. Then

$$Q_2^{\text{HT}}(\rho\|\sigma) = \mathbb{E} g_{\rho, \sigma}(T), \quad Q_2^{\text{HT}}(\rho \otimes \rho' \|\sigma \otimes \sigma') = \mathbb{E} [g_{\rho, \sigma}(T) g_{\rho', \sigma'}(T)].$$

So Theorem 6.4.1 says that the modular autocorrelation functions of two pairs are positively correlated under a common logistic clock: synchronising the clocks of the two factors can only increase Q_2^{HT} .

Proof. $|Y_{jk}|^2 = w_{jk}$ in the notation of the proof of Theorem 6.4.1, and $\mathbb{E} \cos(\omega T) = \frac{\omega/2}{\sinh(\omega/2)} = k(\omega)$ is the characteristic function of the logistic law. For products, $Y_{12} = Y \otimes Y'$ and the frequencies add. In this language, the proof of Theorem 6.4.1 represents T as a Gaussian scale mixture, $T \mid V \sim N(0, 2V)$. For a centred Gaussian T one has $\mathbb{E}[\cos(aT) \cos(bT)] - \mathbb{E} \cos(aT) \mathbb{E} \cos(bT) = e^{-(a^2+b^2)V} (\cosh(2abV) - 1) \geq 0$, and Chebyshev's inequality handles the mixture over V . $\square$

At integer orders the same Fourier analysis applies to all n . With the Mellin–Barnes formula $\frac{1}{t+q} = \frac{1}{2\pi i} \int_{\text{Re } z=c} \frac{\pi}{\sin \pi z} q^{-z} t^{z-1} dz$ on the lines $c = 1 - \frac{1}{n}$, the t -integral of the trace formula produces $\delta(\sum_k y_k)$. This gives

$$Q_n^{\text{HT}}(\rho \parallel \sigma) = \int_{H_n} W_n(y) G_{\rho, \sigma}(y) dy, \quad G_{\rho, \sigma}(y) = \text{Tr} [\hat{\rho} \sigma^{-iy_1} \hat{\rho} \sigma^{-iy_2} \dots \hat{\rho} \sigma^{-iy_n}],$$

where $\hat{\rho} = \sigma^{-\frac{n-1}{2n}} \rho \sigma^{-\frac{n-1}{2n}}$, on the hyperplane $H_n = \{\sum_k y_k = 0\}$, with the weight $W_n(y) \propto \prod_k \pi / \sin(\frac{\pi}{n} - i\pi y_k)$ normalised so that $\int_{H_n} W_n = 1$. The multi-time correlation function is multiplicative, $G_{\rho \otimes \rho', \sigma \otimes \sigma'} = G_{\rho, \sigma} G_{\rho', \sigma'}$, so Conjecture 6.3.5 at order n says $\int W_n G G' \geq \int W_n G \int W_n G'$. At $n = 2$ the weight is the positive logistic density of Proposition 6.10.1. For $n \geq 3$ it is complex. One may move the n contours to lines $\text{Re } z_k = c_k$ with $\sum_k c_k = n - 1$. Each factor $\pi / \sin(\pi z_k)$ is real on its line only when $c_k = \frac{1}{2}$, and then $\sum_k c_k = n - 1$ forces $n = 2$. This is the Fourier-side form of the obstruction in Proposition 6.4.5. The simplex representation of Proposition 6.5.1 avoids it. There the reference law is the uniform law on Δ_{n-1} , which is positive for every n , and the two factors enter through different functions, $\mathcal{F}$ and $\mathcal{G}$, instead of through one multiplicative correlation function.

The clock picture suggests interpolating instead of comparing directly. Let $\mathcal{E}_s^\sigma(\rho) = \mathbb{E} \sigma^{iT} \rho \sigma^{-iT}$ with $T \sim N(0, 2s)$. This is the *modular heat flow*: the quantum Markov semigroup $e^{s\mathcal{L}_\sigma}$ with $\mathcal{L}_\sigma(X) = -[\log \sigma, [\log \sigma, X]]$, which multiplies the (j, k) entry by $e^{-s\omega_{jk}^2}$. It is unital and fixes σ . As $s \rightarrow \infty$ the Gaussian laws converge, on the frequency group, to the Haar average, so $\mathcal{E}_s^\sigma(\rho)$ tends to the pinching of ρ with respect to σ .

Theorem 6.10.2 (Reduction to a local supermodularity). *Let $\rho, \sigma, \rho', \sigma'$ be faithful, $\alpha \in (0, 1) \cup (1, \infty)$, and*

$$\Psi(s, t) = \log Q_\alpha^{\text{HT}}(\mathcal{E}_s^\sigma \rho \otimes \mathcal{E}_t^{\sigma'} \rho' \parallel \sigma \otimes \sigma') - \log Q_\alpha^{\text{HT}}(\mathcal{E}_s^\sigma \rho \parallel \sigma) - \log Q_\alpha^{\text{HT}}(\mathcal{E}_t^{\sigma'} \rho' \parallel \sigma').$$

Then $\Psi(s, \infty) = \Psi(\infty, t) = 0$, and

$$\log \frac{Q_\alpha^{\text{HT}}(\rho \otimes \rho' \parallel \sigma \otimes \sigma')}{Q_\alpha^{\text{HT}}(\rho \parallel \sigma) Q_\alpha^{\text{HT}}(\rho' \parallel \sigma')} = \Psi(0, 0) = \int_0^\infty \int_0^\infty \partial_s \partial_t \log Q_\alpha^{\text{HT}}(\mathcal{E}_s^\sigma \rho \otimes \mathcal{E}_t^{\sigma'} \rho' \parallel \sigma \otimes \sigma') ds dt.$$

Consequently Conjecture 6.3.5 holds at order α for all pairs if the following local condition holds for all faithful pairs:

$$(S_\alpha) \quad \partial_s \partial_t \big|_{s=t=0} \log Q_\alpha^{\text{HT}}(\mathcal{E}_s^\sigma \rho \otimes \mathcal{E}_t^{\sigma'} \rho' \parallel \sigma \otimes \sigma') \geq 0.$$

The same holds with the Gaussian law replaced by any symmetric infinitely divisible law whose exponent vanishes only at 0, for instance the Poisson flow $e^{-s|\omega_{jk}|}$.

Proof. $\mathcal{E}_s^\sigma$ is a mixture of unitaries that commute with σ , so $\mathcal{E}_s^\sigma \rho$ is faithful and converges exponentially fast, with rate $e^{-s \min \omega_{jk}^2}$ over $\omega_{jk} \neq 0$, to a state commuting with σ . If one pair commutes, the product shadow factorises (the remark after Conjecture 6.3.5), so $\Psi(\infty, t) = \Psi(s, \infty) = 0$. By the trace formula, Q_α^{HT} is real-analytic on faithful states. The single-factor terms of Ψ have zero mixed derivative, so the double integral of $\partial_s \partial_t \Psi$ over

$[0, \infty)^2$ equals $\Psi(0, 0) - \Psi(\infty, 0) - \Psi(0, \infty) + \Psi(\infty, \infty) = \Psi(0, 0)$. The semigroup property $\mathcal{E}_{s+h} = \mathcal{E}_h \mathcal{E}_s$ turns the mixed derivative at (s, t) into the mixed derivative at $(0, 0)$ for the pairs $(\mathcal{E}_s \rho, \sigma)$ and $(\mathcal{E}_t \rho', \sigma')$, so (S_α) for all pairs makes the integrand nonnegative. $\square$

Condition (S_α) is local and stronger than the conjecture: it asks that $\log Q_\alpha^{\text{HT}}$ of a product be supermodular along the two independent flows. Numerically it holds for every pair and every order we tested (Section 6.11). For two qubit pairs it takes an especially concrete form. In the eigenbasis of σ write $\rho = \begin{pmatrix} p_1 & c \\ \bar{c} & p_2 \end{pmatrix}$. The flow multiplies c by $e^{-s\omega^2}$, so along the flows only $X = |c|^2$ and $Y = |c'|^2$ change, and (S_α) says that $(X, Y) \mapsto \log Q_\alpha^{\text{HT}}$ is supermodular.

Proposition 6.10.3 (Two qubit pairs at integer orders). *Let (ρ, σ) and (ρ', σ') be faithful qubit pairs and $n \geq 2$ an integer. With the diagonals in the eigenbases of σ, σ' fixed, $Q_n^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma')$ is a polynomial of degree $\lfloor n/2 \rfloor$ in each of $X = |c|^2$ and $Y = |c'|^2$. For $n = 2, 3$ it is bilinear, $c_{00} + c_{10}X + c_{01}Y + c_{11}XY$. Then (S_n) on the qubit flows, Conjecture 6.3.5 and $c_{00}c_{11} \geq c_{10}c_{01}$ are equivalent.*

- (a) For $n = 2$ the condition is $\mathbb{E}[\cos(\omega T) \cos(\omega' T)] \geq \mathbb{E} \cos(\omega T) \mathbb{E} \cos(\omega' T)$, which holds by Theorem 6.4.1.
- (b) For $n = 3$ the conjecture for all pairs of qubit pairs is equivalent to (and proved in Corollary 6.10.4)

$$\Phi(a, b) := D(1, ab) + 2D(a, b) - 6D(1, a)D(1, b) \geq 0 \quad \text{for all } a, b > 0,$$

where $D(x, y) = \int_0^\infty \frac{dt}{(t+1)(t+x)(t+y)} = \frac{L(x)-L(y)}{y-x}$ and $L(\gamma) = \frac{\log \gamma}{\gamma-1}$, the reciprocal of the logarithmic mean of γ and 1. Here a and b range over the eigenvalue ratios of σ and σ' and their inverses.

Proof. In $Q_n = (n-1) \int_0^\infty \text{Tr}[(\rho_{12}(\sigma_{12} + t)^{-1})^n] dt$, every term is a closed walk of length n on the product basis. Its first coordinate crosses between the two levels an even number of times, each crossing contributing c or $\bar{c}$, and the same holds for the second coordinate. This gives the polynomial structure. For a bilinear f , $\partial_X \partial_Y \log f = (c_{00}c_{11} - c_{10}c_{01})/f^2$ and $f(1, 1)f(0, 0) - f(1, 0)f(0, 1) = c_{00}c_{11} - c_{10}c_{01}$, which gives the equivalences. (a) is the computation in Proposition 6.10.1 with one frequency per factor.

(b) The term c_{00} is the product of the diagonal parts, and $c_{10} = c'_{00}E$, $c_{01} = c_{00}E'$, where E is the coefficient of X in $Q_3(\rho \| \sigma)$. So the condition is $c_{11} \geq EE'$. Walks with two crossings in a factor visit one level twice and the other once, in all three cyclic positions. Hence c_{11} is a sum over the two choices of doubled level in each factor of $\frac{1}{3} \sum_{\pi \in S_3}$ of the kernel $\int_0^\infty \prod_k (t + q_{i_k} q'_{\pi(i_k)})^{-1} dt$ with positive weights $W_\epsilon W'_{\epsilon'}$, which are the same weights that appear in E and E' . After scaling out the homogeneity, the difference $c_{11} - EE'$ becomes $\frac{2}{3} \sum_{\epsilon, \epsilon'} W_\epsilon W'_{\epsilon'} \Phi(a^\epsilon, b^{\epsilon'})$ times positive factors, with $a = q_2/q_1$ and $b = q'_2/q'_1$, and the kernel is $D(1, ab)$ when the singly visited levels of the two factors occupy the same position of the walk and $D(a, b)$ otherwise. The ratio $W_+/W_- = (p_1/p_2)(q_2/q_1)^{2/3}$ can take any positive value, so each term must be nonnegative separately. $\square$

Inequality (b) is the case of the symmetric-group inequality (6.4.1) at $n = 3$ in which each of u and v has two equal coordinates, so it is covered by Theorem 6.5.3.

Corollary 6.10.4 (Super-multiplicativity at $\alpha = 3$ for qubits). *For all pairs of qubit states,*

$$Q_3^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq Q_3^{\text{HT}}(\rho \| \sigma) Q_3^{\text{HT}}(\rho' \| \sigma'),$$

with equality if and only if $[\rho, \sigma] = 0$ or $[\rho', \sigma'] = 0$. Equivalently, $\Phi(a, b) \geq 0$ for all $a, b > 0$, with equality exactly when $a = 1$ or $b = 1$.

Proof. Qubit pairs are sign-free, because σ has at most two distinct eigenvalues, so Theorem 6.5.3(c) gives the inequality and the equality cases. The statement about Φ is (6.4.1) at $n = 3$ for u and v proportional to centred versions of $(0, 0, \log a)$ and $(0, 0, \log b)$, with the strictness of Theorem 6.5.3(b); it is also equivalent to the first statement by Proposition 6.10.3(b). The inequality and its equality cases are also the case of Corollary 6.6.3 in which both pairs are qubit pairs. $\square$

We first proved Corollary 6.10.4 by a one-variable Stieltjes-transform argument of the same kind as the θ -reduction in Remark 6.5.4, and we record it because it is explicit. Substituting $t = 1/\lambda$ gives $D(1, x) = \frac{1}{2}\mathbb{E}_\mu f_x$ and $D(x, y) = \frac{1}{2}\mathbb{E}_\mu [f_x f_y]$, with $\mu(d\lambda) = \frac{2\lambda d\lambda}{(1+\lambda)^3}$ and $f_x(\lambda) = \frac{1+\lambda}{1+x\lambda}$. The identity $f_{ab}(\lambda) = f_a(\lambda)f_b(a\lambda)$ then writes $2\Phi(a, b) = (1 - \frac{1}{b}) \int_0^\infty \frac{\kappa_a(d\lambda)}{1+b\lambda}$, where κ_a is a signed measure of mass 0 whose density is $\frac{2\lambda}{(1+\lambda)^3}$ times a rational function with a cubic numerator N_a . Since $N_a(-a) = -(a-1)^3(a+1)$ and $N_a(-1) = 3c_a(a-1)^3$, with $c_a = 2D(1, a) > 0$, have opposite signs, κ_a changes sign at most twice on $(0, \infty)$. The variation-diminishing property of the Cauchy kernel [70, Ch. 5, Thm. 3.1] and the signs of Φ at the ends of each line then locate the only zero at $b = 1$. The ingredients were checked symbolically and numerically (Section 6.11). This route needs the bilinearity of Q_3^{HT} in the flow variables, so it does not extend beyond qubits. The simplex route does, because it needs only convexity and symmetry.

Section 6.11 summarises the numerical evidence; we mention four checks here. For Theorem 6.3.2, $Q_\alpha^{\text{HT}}/\tilde{Q}_\alpha \leq 0.9980$ for 160 random pairs with $\alpha \in \{2, 2.5, 4, 7\}$. In the range $(1, 2)$, which is covered by [84], $Q_\alpha^{\text{HT}}/\tilde{Q}_\alpha \leq 1.0000$ for 120 random pairs in dimensions 2–5 and $\alpha \in \{1.2, 1.5\}$. The Helstrom layer-cake inequality of Step 3 held for 30 pairs with $\alpha \in \{1.2, 1.5\}$. Super-multiplicativity was tested on 40 random pairs of qubit pairs with $\alpha \in \{0.3, 0.7, 1.5, 2, 4\}$, and all 200 ratios are at least 1 to six decimal places. An adversarial search (Nelder–Mead on the ratio, well-conditioned states) for α from 0.05 to 40 found no violation. In every run the minimiser drove one pair towards commuting, the equality case, and 30-digit arithmetic at the minimisers gives ratios above 1 by exactly the predicted second-order amount. Apparent violations at near-singular σ' were rounding artefacts. At $\alpha = 40$, 30-digit arithmetic with the tensor products formed exactly gives ratio $-1 = +1.4 \cdot 10^{-10}$, where double precision had given $-1.2 \cdot 10^{-6}$. Finally, for tensor powers, $\frac{1}{n}D_2^{\text{HT}}$ increases monotonically towards $\tilde{D}_2$ for qubit powers up to $n = 32$, and $\frac{1}{n}D_{1/2}^{\text{HT}}$ decreases towards $D_{1/2}^{\text{P}}$ up to $n = 16$.

6.11 Numerical checks

Random states are Wishart-normalised.

Claim	Result
Proposition 6.3.1	no violation over 40 pairs and $\alpha \in \{0.2, 0.5, 0.8\}$
$D_\alpha^{\text{HT}} \leq \tilde{D}_\alpha$ [84]	$Q_\alpha^{\text{HT}}/\tilde{Q}_\alpha \in [0.665, 1.0000]$ over 120 pairs, $d \in \{2, \dots, 5\}$, $\alpha \in \{1.2, 1.5, 2, 3\}$

Claim	Result
Conjecture 6.3.5	all 200 ratios $Q_\alpha^{\text{HT}}(\rho \otimes \rho' \ \sigma \otimes \sigma') / (Q_\alpha^{\text{HT}} Q_\alpha^{\text{HT}})$ lie in $[1.000000, 1.0169]$: 40 random pairs of qubit pairs, $\alpha \in \{0.3, 0.7, 1.5, 2, 4\}$. Adversarial minimisation over well-conditioned states, 2×2 for $\alpha \in \{0.05, 0.2, 0.5, 0.8, 0.95, 1.05, 1.5, 2, 3, 6, 15, 40\}$ and 2×3 for $\alpha \in \{0.5, 1.5, 3\}$: every minimiser approaches a commuting pair, with ratio -1 in $[-3 \cdot 10^{-11}, 10^{-11}]$, the rounding floor (an exactly commuting pair returns $-3 \cdot 10^{-12}$ at $\alpha = 15$). Along the minimising direction the ratio is $1 + Ct^2$ with $C = +1.2 \cdot 10^{-6}$ ($\alpha = 15$) and $+6.2 \cdot 10^{-5}$ ($\alpha = 40$). With exact inputs in 30-digit arithmetic the two minimisers have ratio $-1 + 2.49 \cdot 10^{-14}$ and $+1.62 \cdot 10^{-12}$, which equals Ct_0^2 at their distances $t_0 = 1.45 \cdot 10^{-4}$ and $1.62 \cdot 10^{-4}$ from commuting.
Theorem 6.8.2(d), sign pattern of δ	2×2 : 100 random pairs, all $+-$ (the signs near both ends verified with an exact routine at relative distances 10^{-3} and 10^{-2} ; one apparent extra change at relative distance $2 \cdot 10^{-6}$ from $\gamma_{\max}$ is below the accuracy); 2×3 : 25 pairs $+-$ and 15 pairs $+-+-$ of 40; 3×3 : 16, 11, 3 pairs with two, four, six sign changes of 30; 3×4 : 5, 4, 1 of 10
Convex order (6.8.1)	$U \geq 0$ in all 180 pairs above: computed from both ends by piecewise Gauss–Legendre quadrature (breakpoints at every kink of δ), $\min U \geq -2 \cdot 10^{-11}$ in absolute terms against peaks of 10^{-6} to 10^{-2} , with quadrature defects $ M(0) , M(-1) \leq 6 \cdot 10^{-11}$; an adversarial Nelder–Mead minimisation of the dip of U over the middle 80% of its support, relative to its peak, moved it from $1.3 \cdot 10^{-4}$ to $1.1 \cdot 10^{-4}$ (2×2) and from $1.9 \cdot 10^{-3}$ to $1.6 \cdot 10^{-3}$ (2×3) in 1,000 evaluations, never approaching zero
Theorem 6.8.2(a)	$M(s)$ has the sign of $s(s+1)$ at $s \in \{-3, -2, -1.5, -0.7, -0.3, 0.5, 1, 2, 4\}$ for every pair above
$\varphi(\alpha) = \log Q_\alpha^{\text{HT}}(\rho \otimes \rho' \ \sigma \otimes \sigma') - \log Q_\alpha^{\text{HT}} - \log Q_\alpha^{\text{HT}}$ on $\alpha \in [-3, 6]$	$\varphi \geq 0$ on the whole grid (step 0.25) for 85 pairs (2×2 , 2×3 , 3×3), with $\min \varphi = 3.4 \cdot 10^{-7}$ on $[-3, 0]$; φ has double zeros at $\alpha = 0, 1$ and is not convex in any of the 85 cases (a local maximum inside $(0, 1)$ and another beyond 2), so super-additivity of the tilted testing-region variance is not available as a route
Proposition 6.4.5, inequality (6.4.1) (proved in Theorem 6.5.3(b))	$n = 3$: $\min \frac{1}{n!} \sum_\pi k_n(u + \pi v) / (k_n(u)k_n(v)) = 1.0000079$ over 300 random pairs; $n = 4$: 1.0000296 over 150; $n = 5$: 1.00051 over 40; adversarial minimisation (4 starts each) returns 1.0000000 at $u = 0$ or $v = 0$ only. Logarithmic form: minimum gap $+6.3 \cdot 10^{-6}$ ($n = 3$, 200 pairs) and $+3.5 \cdot 10^{-5}$ ($n = 4$, 100 pairs), adversarial minima $\geq +4 \cdot 10^{-16}$ at the equality cases. The closed form of k_3 agrees with the integral to 10^{-15}

Claim	Result
Proposition 6.5.1 (simplex representation)	$\mathbb{E}\mathcal{F}$, $\mathbb{E}\mathcal{G}$ and $\mathbb{E}[\mathcal{F}\mathcal{G}']$, computed by a tensor Gauss–Legendre rule on Δ_{n-1} , agree with the trace formula to relative error $2 \cdot 10^{-12}$ for $n = 2$ (3×3), $n = 3$ (3×2 , 3×3) and $n = 4$ (2×2 , 3×2), 3 random pairs each; for generic 3-dimensional pairs $\mathcal{F}$ is complex ($ \text{Im } \mathcal{F} / \mathcal{F} $ up to $1.7 \cdot 10^{-3}$), cyclically invariant and conjugated by reversal to 10^{-15} , and $\text{Re } \mathcal{F}_4$ is not S_4 -invariant (relative deviation $4 \cdot 10^{-5}$)
Lemma 6.5.2	$\text{Cov} \geq 0$ for random pairs of Schur-convex functions (symmetrised exponentials of linear forms, symmetrised $(\mu \cdot y)^{-p}$, symmetrised maxima of affine functions, $-\prod_i \mu_i$, increasing functions of $\sum_i \mu_i^2$ and of $\max_i \mu_i$) on Δ_2 , Δ_3 , Δ_4 : minimum correlation $+3.8 \cdot 10^{-2}$ ($n = 3$, 369 pairs), $+6.0 \cdot 10^{-3}$ ($n = 4$, 383 pairs) and $+1.0 \cdot 10^{-1}$ ($n = 5$, 288 pairs), after discarding numerically constant functions; the quadrature reproduces $\mathbb{E}U_1$ and $\mathbb{E}U_1^2$ to 10^{-14}
Theorem 6.5.3(b)	pure $\otimes$ pure ratios minus 1: $\geq 9.7 \cdot 10^{-6}$, $1.7 \cdot 10^{-6}$, $1.5 \cdot 10^{-4}$, $1.4 \cdot 10^{-7}$ at $n = 3, 4, 5, 6$ (24 products each, dimensions 2×2 to 2×4); pairs with nonnegative entries (including qubits) tensored with pure pairs: $\geq 3.6 \cdot 10^{-5}$, $9.4 \cdot 10^{-6}$, $2.0 \cdot 10^{-5}$ at $n = 4, 5, 6$ (20 products each)
Theorem 6.5.3(c)	64 products of sign-free pairs drawn from four classes (σ with two distinct eigenvalues and random ρ in dimensions 3, 4; nonnegative ρ ; pure ρ ; qubits): ratio minus 1 at least $2.7 \cdot 10^{-7}$; smallest relative Hessian eigenvalue of $\text{Re } \mathcal{F}_3$ on a grid $+6.9 \cdot 10^{-3}$, $+6.1 \cdot 10^{-2}$, $+1.3 \cdot 10^{-1}$ for the first three classes

Claim	Result
Remark 6.5.4	<p> $\text{Re } \mathcal{F}_3$ Schur-convex in all 600 random pairs in dimensions 3, 4, $\text{Re } \mathcal{G}_3$ in 596; for the 4 exceptions the ratio minus 1 is at least $2.0 \cdot 10^{-4}$ against 4 random partners each; the chirality term $\mathbb{E}[\text{Im } \mathcal{F} \text{Im } \mathcal{G}']$ is at most $1.4 \cdot 10^{-3}$ of the covariance (median 10^{-4}) over 20 random 3×3 products; the pointwise condition $\frac{1}{6} \sum_{\pi \in S_3} I_\pi(x) \geq 2 \int_0^\infty \prod_i (t + x_i)^{-1} dt Q_3^{\text{HT}}(\rho \parallel \sigma)$, with $I_\pi(x) = 2 \int_0^\infty \text{Tr}[\rho(t + x_{\pi_1}\sigma)^{-1} \rho(t + x_{\pi_2}\sigma)^{-1} \rho(t + x_{\pi_3}\sigma)^{-1}] dt$, which would give super-multiplicativity against every sign-free partner, held at 360 random points (ρ, σ, x) in dimensions 3, 4 (minimum relative margin $4.5 \cdot 10^{-7}$) but is false at extreme spectra: for $q = (2.03 \cdot 10^{-5}, 1.04 \cdot 10^3, 1.08 \cdot 10^{-2})$, $x = (2.82, 2.08 \cdot 10^5, 1.48 \cdot 10^{-4})$ and a rank-2 qutrit ρ with a negative 3-cycle the difference is $-1.27 \cdot 10^{-9}$ in 60-digit arithmetic (-8.6% of the sum of the absolute values of its terms), while against pure partners with eigenvalues x the full defect stays positive; inequality (6.5.1): minimum of the ratio right side/left side 1.0000 over 60,000 samples ($\log q_i, \log a$ uniform in $[-S, S]$, $S = 2, 6, 12$), attained at coinciding q_i; the exact worst phase for qutrits against two-equal partner words gives a normalised defect $\geq +0.029$ (600 samples, $S = 6, 12$), and for $d = 4, 5$ an adversarial search over all ρ gives minima $\geq -4 \cdot 10^{-14}$ (rounding at degenerate spectra); an adversarial search of the full defect for 3×2 and 3×3 products with log-eigenvalues in $[-12, 12]$ (and in $[-6, 6]$, $[-3, 3]$) found no violation, its minimisers approaching commuting pairs (two apparent violations at -10^{-15} were rounding: $+10^{-25}$ in 60-digit arithmetic); $R(\theta)$ convex in all 600 random pairs in dimensions 2, 3, 4, 6, while three adversarial pairs have $R'' < 0$ (down to -243) yet $R - Q_3^{\text{HT}}$ has two sign changes, $2 \int \theta R = 3 \int \theta^2 R$ to 12 digits, and $\Delta(a)/Q_3^{\text{HT}} > 0$ for a from 10^{-4} to 10^4 </p>
<p> Theorem 6.6.2, Proposition 6.6.1, Corollaries 6.6.3 and 6.6.7, Proposition 6.6.5 </p>	<p> the Lieb–Thirring double integral reproduces $\mathcal{D}_{\rho \parallel \sigma}(a)$ to 12 digits for a random qutrit pair and for a pair with a strongly negative 3-cycle ($a = 0.15, 3.7$), with a pointwise nonnegative integrand; the defect of qutrit$\otimes$qutrit products with $\sigma' = \text{diag}(1, 1, a)$ equals $S_1 \mathcal{D}(a) + S_2 a^{-2} \mathcal{D}(1/a)$ to 10–12 digits; the defect divided by ε^2 tends linearly to $\frac{3}{2}(\tau_1 - \tau_2)\ell_{\rho \parallel \sigma}$; the bounds of Corollary 6.6.7 hold for 200 random pairs ($d \leq 4$, $\gamma \in \{-1, 0, \frac{1}{3}, \frac{2}{3}, 1, 2\}$), with minimal relative slack $6 \cdot 10^{-6}$ </p>

Claim	Result
Theorem 6.7.1, Proposition 6.7.3, Corollaries 6.7.4, 6.7.7, 6.7.9, Theorems 6.7.5, 6.7.8, 6.7.12, Proposition 6.7.13	pure partners against chiral 3×3 pairs: ratio minus 1 at least $2.3 \cdot 10^{-4}$, $9.3 \cdot 10^{-5}$, $5.0 \cdot 10^{-5}$ at $n = 3, 4, 5$ (25 products); the rotation identity of H to $5 \cdot 10^{-15}$, $[H(z; a_i, a_j)]$ positive semidefinite in 500 random 7×7 instances, the trace identity to 12 digits; (6.5.1): maximal ratio 0.9992 over 400 samples with eigenvalue ratios up to e^{12} ; Theorem 6.7.5: minimal ratio 1.012 over $2 \cdot 10^4$ random triples of sizes 2–4 and random ranks; Corollary 6.7.7 against partners with a degenerate three-level σ' ; negative-cycle pairs of both families of Corollary 6.7.9: all defects positive; Δ_t equals the D_k -series of Theorem 6.7.12 to 9 digits; the formulas for D_2 , D_3 and $u - v$ to 10^{-14} ; ratio in (6.7.2) at most 0.036 over 3000 random pairs and about 0.06 after adversarial search; $D_k \geq 0$ for $k = 2, \dots, 12$ in 3000 random trials; Conjecture 6.7.10: over 8000 random instances the minimum over a 73×73 phase grid, divided by the value at $(0, 0)$, is $+0.0079$, always at (π, π)
Proposition 6.9.1	δ takes both signs in all 180 random pairs of non-commuting pairs of the census above (it must, by the proposition)
Proposition 6.9.2	one-way LOCC value computed exactly (linear program over rank-one POVMs on a great circle of the Bloch sphere, 2400 directions, both directions of communication): $\max_v (e_{\text{LOCC}} - e_{\text{cl}}) = -1.1 \cdot 10^{-8}$ over 6 qubit product pairs and 40 levels (discretisation error of the circle)
Conjecture 6.9.3 (PPT tests)	$\max_v (e_{\text{PPT}} - e_{\text{cl}}) / \max \delta \leq -3.5 \cdot 10^{-9}$ over 60 random 2×2 pairs (80 levels each) and $\leq -5.8 \cdot 10^{-9}$ after adversarial Nelder–Mead search (4 starts, 250 evaluations, 25 levels); $\leq -7.9 \cdot 10^{-10}$ over 25 random 2×3 pairs; $\leq +2.9 \cdot 10^{-8}$ over 12 random 3×3 pairs (solver tolerance, where PPT tests include PPT-entangled ones). The PPT tests capture a fraction $\leq 3 \cdot 10^{-9}$ of the positive part of δ , while all tests exceed e_{cl} on about 70% of the levels. PPT tests beat one-way LOCC tests by up to $0.66 \max_v (e_{\text{cl}} - e_{\text{PPT}})$
Proposition 6.10.1	Q_2^{HT} by the trace formula and by the logistic modular-clock average agree to 10 digits (3×3 example)
Theorem 6.10.2, flow monotonicity	dephasing the second factor by the modular heat flow or the Poisson flow, $R(t) = Q_\alpha^{\text{HT}}(\rho \otimes \mathcal{E}_t \rho' \ \sigma \otimes \sigma') / Q_\alpha^{\text{HT}}(\mathcal{E}_t \rho' \ \sigma')$ is nonincreasing in t in all 240 cases (30 random qubit products, $\alpha \in \{0.5, 2, 3, 6\}$, both flows)
Condition (S_α)	discrete mixed differences of $\log Q_\alpha^{\text{HT}}$ on a grid in (s, t) , with s, t from 0 to ∞ : no negative cell beyond -10^{-11} (the accuracy of $\log Q$) in 119 random products: $\alpha \in \{0.3, 0.5, 0.7, 1.5, 2, 3, 6, 15\}$, dimensions 2×2 and 2×3 , Gaussian flows, and Poisson flows at $\alpha = 0.3$

Claim	Result
Proposition 6.10.3, Corollary 6.10.4	bilinearity of Q_3^{HT} in (X, Y) to $4 \cdot 10^{-15}$ (5 random pairs of qubit pairs); the closed form $2e^{-2(s+r)}\Phi$ matches the direct computation of F to 12 digits; $\Phi \geq 0$ in 40-digit arithmetic on a 121×121 grid of $\log a, \log b \in [-30, 30]$, with minimum of $\Phi \max(1, a) \max(1, b) / (\log a \log b)^2$ equal to $6.5 \cdot 10^{-5}$, attained at the corner $(30, 30)$, where the leading term is $(2 \log a - 7)/a^2$; adversarial minimisation (40 starts) moves only towards infinity along the diagonal. The ingredients of the Stieltjes proof were checked symbolically (sympy: $\partial_b \Phi(a, 1) = 0$, the four limits, $N_a(-a)$ and $N_a(-1)$) and numerically ($2\Phi = (1 - \frac{1}{b})S_a(b)$ to 12 digits; κ_a has mass 0 and two sign changes for $a \in \{0.2, 3, 40\}$). At $\alpha = 4$, where Q_4^{HT} is biquadratic, the coefficient matrix is not always totally positive (relative minor $-9 \cdot 10^{-3}$), so that sufficient condition fails
Local form of Conjecture 6.3.5	for $\rho'_t = \rho'_c + tX$ with $[\rho'_c, \sigma'] = 0$, the coefficient $C(\alpha) = \lim_t t^{-2}(Q(\rho \otimes \rho'_t \ \sigma \otimes \sigma') - QQ'_t)$ is positive for $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.5, 2, 3, 5, 10\}$ in 4 random qubit cases ($C/(QQ')$ between $1.5 \cdot 10^{-5}$ and $3 \cdot 10^{-2}$), and $\lim_t t^{-2}\delta$ has the pattern $+-+$
Theorem 6.4.1	$\min Q_2^{\text{HT}}(\rho \otimes \rho' \ \sigma \otimes \sigma') / (Q_2^{\text{HT}} Q_2^{\text{HT}}) = 1.000006$ over 150 random pairs ($d \in \{2, 3\}$); the logarithmic-mean formula matches the trace formula to $6 \cdot 10^{-14}$
Theorem 6.4.2	$\min_\gamma [E_\gamma(\rho \otimes \rho' \ \sigma \otimes \sigma') - \sum_y a_y e(\gamma b_y / a_y)] = -10^{-15}$ over 60 qubit pairs, random 3-outcome measurements and 25 levels;
Theorem 6.4.3	$D_\alpha^{\text{HT}}(\text{prod}) - D_\alpha^{\text{HT}} - D_\alpha(N\rho' \ N\sigma') \geq 0.039$ for $\alpha \in \{0.5, 2\}$ $(\log Q_\alpha^{\text{HT}} - \alpha \log \gamma_{\max} - \log m)(\alpha + 1) =$ 0.1669, 0.1576, 0.1526, 0.1501 at $\alpha = 10, 20, 40, 80$, against $\kappa = 0.1475$; $\kappa_{12} \geq \kappa + \kappa'$ in 2,000 random 2×2 and 500 random 2×3 cases, and at an adversarial point in 30- and 60-digit arithmetic ($+4.9 \cdot 10^{-13}$)
Tensor powers (qubits, Schur–Weyl blocks)	$\frac{1}{n} D_2^{\text{HT}} =$ 0.3058, 0.3070, 0.3090, 0.3120, 0.3159, 0.3203 for $n = 1, 2, 4, 8, 16, 32$ (limit $\tilde{D}_2 = 0.3361$); $\frac{1}{n} D_{1/2}^{\text{HT}} = 0.1356, 0.1355, 0.1352, 0.1348, 0.1340$ for $n = 1, 2, 4, 8, 16$ (limit $D_{1/2}^{\text{P}} = 0.1281$); both monotone, at the slow rate allowed by Propositions 5.6.1 and 5.6.2
Theorem 6.3.2	$Q_\alpha^{\text{HT}} / \tilde{Q}_\alpha \leq 0.9980$ over 160 pairs, $d \in \{2, \dots, 5\}$, $\alpha \in \{2, 2.5, 4, 7\}$; the Helstrom layer-cake inequality of Step 3 held for 15 pairs at each of $\alpha \in \{1.2, 1.5, 2, 3, 5\}$ (ratio ≤ 0.9998), and for 1,200 random rank-one families with $\alpha \in \{1.2, 1.5, 1.8\}$ (ratio ≤ 0.988)

6.12 Open problems

1. Prove the convex-order statement (6.8.1), or at least Conjecture 6.3.5. The conjecture holds at $\alpha = 2$ (Theorem 6.4.1), at every integer order for pure states and at $\alpha = 3$ for sign-free pairs (Theorem 6.5.3), at every integer order against pure partners (Theo-

rem 6.7.1), at $\alpha = 3$ whenever one of the two reference states has at most two distinct eigenvalues (Corollary 6.6.3), at $\alpha = 3$ in the further cases of Corollaries 6.7.7 and 6.7.9, at $\alpha = 3$ to second order around trivial partners $\sigma' \propto \mathbf{1}$ (Proposition 6.6.5), to first order at $\alpha = \infty$ (Theorem 6.4.3), for $\alpha > 1$ with one factor measured (Theorem 6.4.2), and for tensor powers with enough copies (Proposition 6.8.3). By Theorem 6.3.6 it would give independent proofs of [84] and of [13, (1.10)]; the full regularised identity has since been proved directly in [58], by Haagerup reduction. Theorem 6.8.2 reduces (6.8.1) to $U(c) \geq 0$: the truncated Jeffreys divergence D_j , $j(u) = ((u+1)\log u - 2(u-1))_+$, of the quantum product pair dominates that of the classical product of the shadows for every scaling c of the second argument; for qubit pairs it reduces further to the two-sign-change property of δ , which held in every test. Two special cases of the conjecture are of independent interest: the pure-state case at non-integer orders, which is the Rényi submodularity of Remark 6.4.4 (proved here at integer orders through (6.4.1), Theorem 6.5.3(b); the stronger logarithmic form of (6.4.1) is open for $n \geq 3$); and the second-order case at $\alpha = 1$, superadditivity of the testing-region variance $\int \gamma \log^2 \gamma d\mu - D^2$, which is the case $\psi = v \log v$ of (6.8.1). Pointwise comparison of profiles cannot prove the conjecture, and neither can positive maps between a pair and its shadow: δ changes sign for every two non-commuting pairs (Proposition 6.9.1). One-way LOCC tests, and numerically all separable and PPT tests, never beat the classical product of the shadows (Proposition 6.9.2, Conjecture 6.9.3), so a proof must use entangled tests. A related question is a direct, non-greedy operator proof of Corollary 6.3.4 for $\kappa > 2$.

2. Extend Lemma 6.5.2 beyond Schur-convex functions. Is $\text{Cov}(\phi(U), \chi(U)) \geq 0$ for convex functions on Δ_{n-1} that are only invariant under the dihedral group D_n of cyclic shifts and reversals? This would give the conjecture at every integer order for all pairs with nonnegative cycle weights, in particular for all pairs of qubit pairs, by Proposition 6.5.1. Our induction fails because the stabiliser of a facet in D_n is too small. Numerically the answer looks positive: the minimum correlation over 600 random pairs of D_4 -invariant convex functions was $+7.7 \cdot 10^{-6}$, and for 40 random pairs of qubit pairs at $n = 4$ the correlation of $\mathcal{F}_4$ and $\mathcal{G}'_4$ was at least 0.23. At integer orders $n \geq 4$ the conjecture is now known for pure partners against arbitrary pairs (Theorem 6.7.1); the dihedral question would add all pairs of pairs with nonnegative cycle weights. At $\alpha = 3$ the conjecture is proved whenever σ or σ' has at most two distinct eigenvalues, for pure partners and pure partners with commuting noise, against partners with three eigenvalues and a nonnegative block 3-cycle, and for pairs whose block 3-cycles are negative on both sides and not too strongly correlated (Corollaries 6.6.3, 6.7.2, 6.7.7, 6.7.9, Theorems 6.7.1, 6.7.8). What remains has both reference states with at least three distinct eigenvalues, and by Corollary 6.6.6 both states may be taken singular: pairs with chiral block 3-cycles, strongly correlated pairs with negative 3-cycles on both sides, and, when a reference state has four or more eigenvalues, pairs whose real block 3-cycles have mixed signs, or are nonnegative on one side and nonpositive on the other. Three concrete targets would settle parts of it. (a) Conjecture 6.7.10, an inequality for the kernel K_3 alone, which by Proposition 6.7.11 gives every pair whose reference states have at most three eigenvalues, including $d = d' = 3$; at real phases it is the single corner $\varphi = \varphi' = \pi$. (b) The chirality inequality (6.7.2), which gives $D_3 \geq 0$ in the polynomial criterion of Theorem 6.7.12; more ambitiously, $D_k \geq 0$ for all $k \geq 2$, which gives the conjecture at $\alpha = 3$. (c) For a qutrit ρ against an arbitrary partner, with the partner kernel $C_x = \mathcal{D}_{\rho' \| \sigma'}[q_x]$, the bound $|C_{123}| \leq \min_{p>0} \sum_{b \neq c} p_b^2 p_c C_{bbc} / (2p_1 p_2 p_3)$, which by Lemma 6.7.6 implies the conjecture for every qutrit ρ ; Theorem 6.7.1 gives its negative real direction, and Corollary 6.7.4 gives the stronger rooted bound $|C_{123}| \leq \sum_c \sqrt{C_{aac} C_{bbc}}$ for two-level

and pure partners. Partner words with three distinct eigenvalues are not positive one at a time (the pointwise condition of Section 6.11 is false at extreme spectra), the product-space Lieb–Thirring and matrix-measure versions of Theorem 6.7.1 fail for structural reasons (Remark 6.6.4(ii) and the remark after Theorem 6.7.1), and no universal rooted kernel certifies (6.7.2), so a proof must use the positivity of both pairs together and across different roots. Remark 6.6.4(iii) records a stronger one-variable convex-order statement, open beyond weakly sign-free pairs. The local supermodularity (S_α) of Theorem 6.10.2, which implies the conjecture and held in every test, is a different route with the same first open cases.

3. Prove Conjecture 6.9.3: separable tests on a product pair are dominated by the classical product of the shadows. It holds for one-way LOCC tests (Proposition 6.9.2). Does it hold for two-way LOCC tests? In its Blackwell form it asks whether every outcome experiment of a product-element POVM is a garbling of $(P \otimes P', Q \otimes Q')$.
4. Is the adaptive Chernoff exponent, $-\lim \frac{1}{n} \log \inf \eta_n(1)$, the minimum over α of a regularised testing-region cumulant function? Proposition 6.3.1 gives one inequality for every strategy.

Chapter 7

Adaptive quantum channel discrimination as a controlled random walk

Summary. We study converse bounds on the error exponent of adaptive discrimination between two quantum channels, with n channel uses and arbitrary quantum memory. Diagonalising the Arveson–Radon–Nikodym derivative of a channel Φ with respect to $\Phi + \Psi$ writes any pair of channels as $\Phi = \sum_{\ell} \tau_{\ell} \Omega_{\ell}$, $\Psi = \sum_{\ell} (1 - \tau_{\ell}) \Omega_{\ell}$ with common completely positive level maps Ω_{ℓ} . We show that this turns every adaptive discrimination protocol into a post-processing of a controlled random walk. Consequently every adaptive strategy errs with probability at least $\frac{1}{4} \exp(-n\hat{C} - 1.67\Lambda\sqrt{n})$, where $\hat{C}$ is the geometric Chernoff divergence of the channels, computable from one eigendecomposition, and Λ is a Thompson-type constant that is finite for every pair. No other decomposition of the pair into common pieces gives a better bound of this kind, and the bound is not comparable with the barycentric converse of Ji, Mishra, Mosonyi and Wilde. A dynamic program gives finite- n lower bounds on the error. If the levels are orthogonal at an optimal input, product strategies attain $\hat{C}$; if moreover every level map is a multiple of a channel, a product strategy is exactly optimal for every n and every error criterion. Finally, $e^{-\hat{C}}$ is Lipschitz under perturbations that are small relative to the pair, but under general perturbations $\hat{C}$ is only lower semicontinuous. Near replacement channels with rank-deficient outputs, and for amplitude-damping channels with nearby damping rates, the fidelity, barycentric and geometric Chernoff converses are all infinite while the adaptive exponent tends to zero.

7.1 Introduction

7.1.1 Background and motivation

In binary discrimination of two channels $\Phi, \Psi \in \mathcal{C}(\mathcal{X}, \mathcal{Y})$, a tester uses the unknown channel n times, possibly adaptively and with quantum memory, and then guesses which channel it was. With equal priors the optimal error decays as $e^{-n\xi}$, where ξ is the adaptive Chernoff exponent. For classical channels adaptivity does not help and ξ is the largest Chernoff information over input letters [56]; for classical–quantum channels the same holds with the quantum Chernoff divergence [116]. For general channels adaptivity can help, even turning a finite non-adaptive exponent into perfect discrimination after two uses [53, 116], and no formula for ξ is known.

Lower bounds on ξ come from product inputs and the quantum Chernoff bound [11, 99]. General upper bounds that hold for adaptive strategies have so far come from chain rules. Katariya and Wilde [72] proved a chain rule for the geometric Rényi divergences of order $\alpha \in (0, 1)$ and deduced that ξ is at most the negative logarithm of the geometric fidelity of the channels, $-2 \ln \hat{Q}_{1/2}(\Phi \| \Psi)$; Nuradha and Wilde [98] used such bounds for query complexity. Ji, Mishra, Mosonyi and Wilde [66, 67] proved a barycentric converse built from the Belavkin–Staszewski channel divergence, which is evaluated by a convex program. With a single Rényi order, as in the fidelity converse, turning a bound on the non-i.i.d. final states into an error exponent loses up to a factor of two.

7.1.2 Contributions

Our starting point is a structural observation (Lemma 7.2.1 and Proposition 7.3.1). Let $T = \sum_{\ell} \tau_{\ell} P_{\ell}$ be the spectral decomposition of the Arveson–Radon–Nikodym (ARN) derivative of Φ with respect to $\Omega = \Phi + \Psi$ [10, 111] (Section 4.2 and Theorem 4.2.2 in Chapter 4). Then $\Phi = \sum_{\ell} \tau_{\ell} \Omega_{\ell}$ and $\Psi = \sum_{\ell} (1 - \tau_{\ell}) \Omega_{\ell}$ with the same completely positive level maps Ω_{ℓ} . Each use of Φ (respectively Ψ) therefore selects a level ℓ with probability $\tau_{\ell} \text{tr} \Omega_{\ell}(\rho)$ (respectively $(1 - \tau_{\ell}) \text{tr} \Omega_{\ell}(\rho)$) and then outputs the hypothesis-independent state $\Omega_{\ell}(\rho) / \text{tr} \Omega_{\ell}(\rho)$. By multilinearity, the final states of any adaptive protocol are post-processings of a classical pair of laws on level records, and the log-likelihood ratio of a record is a random walk whose step values do not depend on the inputs.

Our main results are the following.

1. *Chernoff converse for all pairs* (Theorem 7.4.1). Every adaptive strategy has error at least $\frac{1}{4} \exp(-n\hat{C} - \sqrt{2 \ln 4} \Lambda \sqrt{n})$, so $\xi \leq \hat{C}(\Phi, \Psi) = -\ln \min_{0 \leq s \leq 1} \hat{Q}_s(\Phi \| \Psi)$, the geometric Chernoff divergence. The proof uses no chain rule.
2. *Canonicity* (Proposition 7.4.3). Every decomposition of the pair into common completely positive pieces gives a random-walk converse, and the level decomposition gives the best one. In particular Theorem 7.4.1 is never weaker than the converse obtained from a classical-program simulation.
3. *Comparison* (Corollary 7.4.5, Proposition 7.4.6). $\hat{C}$ lies between one half of the fidelity converse and the fidelity converse, and for nearby channels it is $\frac{\delta^2}{8} F_{\text{RLD}} + O(\delta^3)$, one half of the fidelity converse. It is not comparable with the barycentric converse C_{bar} of [66]: $C_{\text{bar}} \leq \hat{C}$ for level-balanced pairs and $\hat{C} \leq C_{\text{bar}}$ for flagged pairs, and both inequalities can be strict.
4. *Finite- n bounds* (Corollary 7.3.2). A dynamic program over $O(n^m)$ log-likelihood values, where m is the number of distinct eigenvalues of T , lower-bounds the optimal adaptive error.
5. *Operational meaning, strong converse* (Proposition 7.4.4, Corollary 7.4.7). $\hat{C}$ is the exact exponent when the tester is told the level after every use, and the walk gives strong-converse bounds for all Rényi orders $\alpha > 1$.
6. *Exactness* (Theorems 7.5.1 and 7.5.3). If the levels are orthogonal at an optimal input, equivalently if T commutes with the environment state there, then $\xi = \hat{C}$ and product strategies are optimal. If in addition each level map is a multiple of a channel, a product strategy is exactly optimal for every n and every error criterion.
7. *Continuity and fragility* (Theorems 7.6.2, 7.6.4 and 7.6.5, Corollary 7.6.3). If all levels of two pairs lie in $[\varepsilon, 1 - \varepsilon]$, then $|e^{-\hat{C}(\Phi_1, \Psi_1)} - e^{-\hat{C}(\Phi_2, \Psi_2)}| \leq 2M(\varepsilon)\tau + 4\sqrt{2} \tanh \frac{dT}{2}$,

with $M(\varepsilon) \leq \varepsilon^{-1}$, in terms of the in-pair derivative distance τ and the Thompson distance d_T between the sums; this yields a converse valid on a whole neighbourhood of a model pair. The converse is lower semicontinuous, continuous where $J(\Phi + \Psi)$ is invertible, and infinite exactly when some input avoids every ambiguous level. Near replacement channels with rank-deficient outputs, all three known single-letter converses are infinite while the adaptive exponent tends to zero.

8. *Amplitude damping and a pure-program converse* (Proposition 7.6.8, Theorem 7.6.9, Corollary 7.6.10). The same happens for two amplitude-damping channels with damping rates $\gamma' \rightarrow \gamma$, although their Choi ranges intersect nontrivially. A converse obtained by simulating each channel use with a pure program state shows that the adaptive exponent is of order $(\gamma - \gamma')^2$ for $0 < \gamma < 1$, with upper and lower bounds that differ by the factor $2/\gamma$ to leading order; Conjecture 7.6.13, which we verified numerically for $\gamma = \frac{1}{2}$ and $n \leq 4$, would reduce this factor to 2.

Section 7.6 applies the continuity theory of the derivative developed in Chapter 4 to the converse itself. It shows where $\hat{C}$ is stable and where it, together with the other single-letter converses, breaks down.

7.1.3 Related work

The converse half of Theorem 7.5.3 can also be obtained by simulating each channel use with a classical program, as for programmable and teleportation-covariant channels [107]; the level decomposition gives a canonical, checkable sufficient condition for such a simulation. Theorem 7.4.1 covers the general case, in which the law of the program depends on the input. Theorem 7.6.9 extends the simulation approach to programs that are pure quantum states. For amplitude damping, Remark 7.6.12 relates the exponent to the bound of Demkowicz-Dobrzański and Maccone on the quantum Fisher information of adaptive strategies [30].

7.1.4 Notation

For finite-dimensional Hilbert spaces $\mathcal{X}, \mathcal{Y}$ we write $L(\mathcal{X}, \mathcal{Y})$ for the linear maps from $\mathcal{X}$ to $\mathcal{Y}$, $L(\mathcal{X}) = L(\mathcal{X}, \mathcal{X})$, $D(\mathcal{X})$ for the density operators on $\mathcal{X}$, and $C(\mathcal{X}, \mathcal{Y})$ for the channels (completely positive trace-preserving maps) from $L(\mathcal{X})$ to $L(\mathcal{Y})$. We write A^* for the adjoint, $\mathbf{1}$ for the identity operator, id for the identity map, tr for the trace, $\text{Tr}_{\mathcal{Y}}$ for the partial trace over $\mathcal{Y}$ and (e_k) for standard bases. The vectorisation $\text{vec} : L(\mathcal{X}, \mathcal{Y}) \rightarrow \mathcal{Y} \otimes \mathcal{X}$ is the linear map with $\text{vec}(yx^*) = y \otimes \bar{x}$, that is, $\text{vec}(e_i e_j^*) = e_i \otimes e_j$, and $J(\Phi) = \sum_{i,j} \Phi(e_i e_j^*) \otimes e_i e_j^*$ is the Choi matrix of a map Φ on $L(\mathcal{X})$. For maps, $\Phi \leq \Psi$ means that $\Psi - \Phi$ is completely positive. The trace norm, operator norm and diamond norm are $\|\cdot\|_1$, $\|\cdot\|_\infty$ and $\|\cdot\|_\diamond$; $\lambda_{\min}$ and $\lambda_{\max}$ are the extreme eigenvalues of a Hermitian operator; $A \#_s B$ is the weighted geometric mean of positive operators [79]; and $F(\rho, \sigma) = \|\rho^{1/2} \sigma^{1/2}\|_1^2$ is the fidelity. Logarithms are natural. This chapter writes $\|\cdot\|_\diamond$ for the diamond norm, which Chapter 4 writes $|||\cdot|||_1$, and $\mathbf{1}$ without a subscript. The function $g_s(t) = t^s(1-t)^{1-s}$ of Section 7.2 coincides on $(0, 1)$ with the function h_s of (4.5.1), not the function g_σ attached there to a mean σ .

7.1.5 Organization

The remainder of this chapter is organized as follows. Section 7.2 introduces the level decomposition and the quantity $\hat{C}$. Section 7.3 shows that every adaptive protocol is a controlled random walk and derives the dynamic-programming bound. Section 7.4 proves

the geometric Chernoff converse, shows that the level decomposition gives the best converse of this kind, compares it with the fidelity and barycentric converses and derives strong-converse bounds. Section 7.5 gives conditions under which the bound is attained. Section 7.6 studies the continuity of the converse, and Section 7.6.1 treats amplitude damping. Section 7.7 contains numerical illustrations, and Section 7.8 concludes with open problems.

7.2 The level decomposition

Throughout, $\Phi, \Psi \in \mathcal{C}(\mathcal{X}, \mathcal{Y})$ are channels and $d = \dim \mathcal{X}$. Let $\Omega = \Phi + \Psi$ have the Kraus decomposition $\Omega(X) = \sum_{k=1}^r V_k X V_k^*$ with linearly independent V_k , and let $0 \leq T \leq \mathbf{1}$ be the ARN derivative of Φ with respect to Ω , so that $\Phi = \Psi_T$ and $\Psi = \Psi_{\mathbf{1}-T}$, where $\Psi_K(X) = \sum_{k,l} K_{kl} V_k X V_l^*$. With $W : \mathbb{C}^r \rightarrow \mathcal{Y} \otimes \mathcal{X}$, $W e_k = \text{vec}(V_k)$, the Choi matrix of Ψ_K is $J(\Psi_K) = W K W^*$.

Lemma 7.2.1 (Level decomposition). *Let $T = \sum_{\ell=1}^m \tau_\ell P_\ell$ with distinct $\tau_\ell \in [0, 1]$ and orthogonal projections P_ℓ , and put $\Omega_\ell = \Psi_{P_\ell}$. Then each Ω_ℓ is completely positive and nonzero, $\sum_\ell \Omega_\ell = \Omega$,*

$$\Phi = \sum_{\ell=1}^m \tau_\ell \Omega_\ell, \quad \Psi = \sum_{\ell=1}^m (1 - \tau_\ell) \Omega_\ell,$$

and the Ω_ℓ do not depend on the choice of Kraus decomposition of Ω . Choosing an orthonormal eigenbasis (f_i) of T with $T f_i = t_i f_i$ and putting $K_i = \sum_k (f_i)_k V_k$ gives the common diagonal Kraus form $\Phi = \sum_i t_i K_i \cdot K_i^$, $\Psi = \sum_i (1 - t_i) K_i \cdot K_i^*$. Finally, $\Phi \leq e^a \Psi$ in the completely positive order if and only if $\tau_\ell \leq e^a (1 - \tau_\ell)$ for all ℓ .*

Proof. Linearity and positivity of $K \mapsto \Psi_K$ give the first claims, and $\Psi_{P_\ell} \neq 0$ because $K \mapsto \Psi_K$ is injective. A change of Kraus decomposition conjugates T , and hence each P_ℓ , by a unitary, and leaves Ψ_{P_ℓ} unchanged. The Kraus form follows from $T_{kl} = \sum_i t_i (f_i)_k (f_i)_l$. The last claim is the order isomorphism of the ARN theorem applied to $e^a(\mathbf{1} - T) - T$. $\square$

We write $S_\ell = \Omega_\ell^*(\mathbf{1}_\mathcal{Y}) \geq 0$ and $a_\ell(\rho) = \text{tr} \Omega_\ell(\rho) = \text{tr}(S_\ell \rho)$. Since Φ and Ψ are trace preserving,

$$\sum_\ell \tau_\ell S_\ell = \mathbf{1}, \quad \sum_\ell (1 - \tau_\ell) S_\ell = \mathbf{1}. \quad (7.2.1)$$

Let $\lambda_\ell = \ln \frac{\tau_\ell}{1 - \tau_\ell} \in [-\infty, \infty]$. A level is *ambiguous* if $0 < \tau_\ell < 1$; we write $\mathcal{I}$ for the set of ambiguous levels and $\Lambda = \max_{\ell \in \mathcal{I}} |\lambda_\ell|$. When $\mathcal{I}$ contains every level, Λ equals the Thompson distance $d_T(\Phi, \Psi) = \max\{D_{\max}(\Phi\|\Psi), D_{\max}(\Psi\|\Phi)\}$ by Lemma 7.2.1. For $s \in [0, 1]$ let $g_s(t) = t^s(1 - t)^{1-s}$ for $0 < t < 1$ and $g_s(0) = g_s(1) = 0$, and put

$$B_s = \sum_\ell g_s(\tau_\ell) S_\ell = \Psi_{g_s(T)}^*(\mathbf{1}), \quad Q_s(\rho) = \text{tr}(\rho B_s),$$

and define

$$\hat{C}(\Phi, \Psi) = -\ln \min_{0 \leq s \leq 1} \lambda_{\min}(B_s) \in [0, \infty].$$

Only ambiguous levels contribute to B_s , and $s \mapsto B_s$ is continuous on $[0, 1]$.

Remark 7.2.2. By the fundamental principle (Theorem 4.2.15 and Remark 4.2.16 in Chapter 4), $J(\Psi) \#_s J(\Phi) = W g_s(T) W^*$ for $0 < s < 1$. The closed form of the geometric Rényi quasi-entropy of channels [72] (Lemma 4.5.3) then gives $\lambda_{\min}(B_s) = \hat{Q}_s(\Phi\|\Psi) = \inf_\rho \text{tr}((\Psi \otimes \mathbf{1})(\rho) \#_s (\Phi \otimes \mathbf{1})(\rho))$, so $\hat{C} = \sup_s (1 - s) \hat{D}_s(\Phi\|\Psi)$ is the geometric Chernoff divergence. Computing it needs one eigendecomposition of an $r \times r$ matrix and a one-dimensional minimisation.

Definition 7.2.3. The pair is *level balanced* if every Ω_ℓ is a multiple of a channel, that is, $S_\ell = c_\ell \mathbf{1}$ for all ℓ . For $\sigma \in \mathcal{D}(\mathcal{X})$ let G_σ be the $r \times r$ matrix $(G_\sigma)_{lk} = \text{tr}(V_l^* V_k \sigma)$; it is the transpose of $\Omega^c(\sigma)$, where Ω^c is the complementary map of Ω for the Stinespring operator $\sum_k V_k \otimes e_k$. The levels are *orthogonal at σ* if $\text{tr}(B^* A \sigma) = 0$ whenever $A \in \mathcal{K}_\ell$ and $B \in \mathcal{K}_{\ell'}$ with $\ell \neq \ell'$, where $\mathcal{K}_\ell = \{\sum_k c_k V_k : c \in \text{ran } P_\ell\}$.

For $u = \text{vec}(A)$ we write $\Phi_u = (\Phi \otimes \mathbf{1})(uu^*)$ and $\Psi_u = (\Psi \otimes \mathbf{1})(uu^*)$; the reduced state of uu^* on the input is AA^* .

Lemma 7.2.4 (Orthogonal levels and commuting outputs). *Let $\sigma \in \mathcal{D}(\mathcal{X})$ and $u = \text{vec}(A)$ with $AA^* = \sigma$.*

- (a) *The levels are orthogonal at σ if and only if $[T, G_\sigma] = 0$.*
- (b) *If $[T, G_\sigma] = 0$, then Φ_u and Ψ_u commute, and there are orthonormal vectors x_i with $\Phi_u = \sum_i t_i g_i x_i x_i^*$ and $\Psi_u = \sum_i (1 - t_i) g_i x_i x_i^*$, where (t_i, g_i) runs over the joint eigenvalues of (T, G_σ) with $g_i > 0$. In particular $\text{tr}(\Phi_u^s \Psi_u^{1-s}) = Q_s(\sigma)$ for $0 \leq s \leq 1$, where M^0 denotes the support projection of M .*
- (c) *If σ is invertible and Φ_u, Ψ_u commute, then $[T, G_\sigma] = 0$.*

Proof. Let $X : \mathbb{C}^r \rightarrow \mathcal{Y} \otimes \mathcal{X}$ be given by $X e_k = \text{vec}(V_k A)$. Then $X^* X = G_\sigma$ and $(\Psi_K \otimes \mathbf{1})(uu^*) = X K X^*$ for every K , so $\Phi_u = X T X^*$ and $\Psi_u = X(\mathbf{1} - T)X^*$; also $\text{tr} \Psi_K(\sigma) = \text{tr}(K G_\sigma)$. (a) For $c \in \text{ran } P_\ell$ and $d \in \text{ran } P_{\ell'}$ we have $\text{tr}((\sum_k d_k V_k)^* (\sum_k c_k V_k) \sigma) = d^* G_\sigma c$, so orthogonality means $P_{\ell'} G_\sigma P_\ell = 0$ for $\ell \neq \ell'$, that is, G_σ commutes with every P_ℓ . (b) $\Phi_u \Psi_u = X T G_\sigma (\mathbf{1} - T) X^*$ and $\Psi_u \Phi_u = X (\mathbf{1} - T) G_\sigma T X^*$ agree when $[T, G_\sigma] = 0$. For a joint orthonormal eigenbasis $T f_i = t_i f_i$, $G_\sigma f_i = g_i f_i$, the vectors $X f_i$ are orthogonal with $\|X f_i\|^2 = g_i$, and $x_i = X f_i / \sqrt{g_i}$ give the stated forms. Hence $\text{tr}(\Phi_u^s \Psi_u^{1-s}) = \sum_i g_s(t_i) g_i = \text{tr}(g_s(T) G_\sigma) = \text{tr} \Psi_{g_s(T)}(\sigma) = Q_s(\sigma)$, the convention $g_s(0) = g_s(1) = 0$ matching the support projections at $s \in \{0, 1\}$. (c) If σ is invertible, X is injective because the V_k are linearly independent, so $X(T G_\sigma (\mathbf{1} - T) - (\mathbf{1} - T) G_\sigma T) X^* = 0$ forces $T G_\sigma (\mathbf{1} - T) = (\mathbf{1} - T) G_\sigma T$, which is $T G_\sigma = G_\sigma T$. $\square$

7.3 Every adaptive protocol is a controlled random walk

An adaptive strategy with n uses consists of a state $\rho_0 \in \mathcal{D}(\mathcal{X} \otimes \mathcal{R}_0)$, channels $\Gamma_k : \mathcal{L}(\mathcal{Y} \otimes \mathcal{R}_{k-1}) \rightarrow \mathcal{L}(\mathcal{X} \otimes \mathcal{R}_k)$ for $1 \leq k \leq n-1$, and a final measurement on $\mathcal{Y} \otimes \mathcal{R}_{n-1}$. We write $\rho_\Phi^{(n)}, \rho_\Psi^{(n)}$ for the final states, $p_{\text{adapt}}(n)$ for the infimum over strategies of the equal-prior error $\frac{1}{2} - \frac{1}{4} \|\rho_\Phi^{(n)} - \rho_\Psi^{(n)}\|_1$, and $\xi_{\text{adapt}} = \limsup_n -\frac{1}{n} \ln p_{\text{adapt}}(n)$. Product strategies $u^{\otimes n}$ with $u \in \mathcal{X} \otimes \mathcal{X}$ achieve $\xi_{\text{prod}} = \sup_u \xi_{\text{QC}}(\Phi_u, \Psi_u)$, with $\xi_{\text{QC}}(A, B) = -\ln \min_{0 \leq s \leq 1} \text{tr}(A^s B^{1-s})$ [11, 99].

Proposition 7.3.1 (Controlled random walk). *Fix an adaptive strategy with n uses. There are states $\rho_k(\ell_1, \dots, \ell_{k-1}) \in \mathcal{D}(\mathcal{X})$ and states $\sigma_{\vec{\ell}}$ on the final system such that*

$$\rho_\Phi^{(n)} = \sum_{\vec{\ell}} P_\Phi(\vec{\ell}) \sigma_{\vec{\ell}}, \quad \rho_\Psi^{(n)} = \sum_{\vec{\ell}} P_\Psi(\vec{\ell}) \sigma_{\vec{\ell}},$$

where $\vec{\ell} = (\ell_1, \dots, \ell_n)$ and

$$P_\Phi(\vec{\ell}) = \prod_{k=1}^n a_{\ell_k}(\rho_k) \tau_{\ell_k}, \quad P_\Psi(\vec{\ell}) = \prod_{k=1}^n a_{\ell_k}(\rho_k) (1 - \tau_{\ell_k}), \quad (7.3.1)$$

$$\rho_k = \rho_k(\ell_1, \dots, \ell_{k-1}).$$

P_Φ and P_Ψ are probability distributions, and the log-likelihood ratio of a record is $L(\vec{\ell}) = \sum_k \lambda_{\ell_k}$.

Proof. The final state is multilinear in the n channel uses. Inserting $\Phi = \sum_\ell \tau_\ell \Omega_\ell$ and $\Psi = \sum_\ell (1 - \tau_\ell) \Omega_\ell$ gives $\rho_\Phi^{(n)} = \sum_{\vec{\ell}} (\prod_k \tau_{\ell_k}) \Theta_{\vec{\ell}}$ and $\rho_\Psi^{(n)} = \sum_{\vec{\ell}} (\prod_k (1 - \tau_{\ell_k})) \Theta_{\vec{\ell}}$, where $\Theta_{\vec{\ell}} \geq 0$ is the output of the strategy when round k uses the completely positive map Ω_{ℓ_k} . Let ζ_k be the unnormalised state entering round k in this run; it depends only on $\ell_1, \dots, \ell_{k-1}$. Since the Γ_k preserve the trace and $\text{tr}(\Omega_\ell \otimes \text{id})(\zeta) = a_\ell(\text{Tr}_{\mathcal{R}} \zeta)$, induction gives $\text{tr} \Theta_{\vec{\ell}} = \prod_k a_{\ell_k}(\rho_k)$ with $\rho_k = \text{Tr}_{\mathcal{R}} \zeta_k / \text{tr} \zeta_k$; when $\text{tr} \zeta_k = 0$ the choice of ρ_k is irrelevant. Put $\sigma_{\vec{\ell}} = \Theta_{\vec{\ell}} / \text{tr} \Theta_{\vec{\ell}}$. Normalisation follows from (7.2.1). $\square$

Hence the final pair of every strategy is obtained from the classical pair (P_Φ, P_Ψ) by a classical-to-quantum channel, and its performance under any error criterion is dominated by that of (P_Φ, P_Ψ) . The inputs enter only through the laws $a_\ell(\rho_k)$ of the levels, while the step values λ_ℓ are fixed.

Corollary 7.3.2 (Computable finite- n bound). *Define $v_n(x) = \frac{1}{2} \min\{e^x, 1\}$ for $x \in [-\infty, \infty)$ and, for $k = n - 1, \dots, 0$,*

$$v_k(x) = \lambda_{\min} \left(\sum_{\ell: \tau_\ell < 1} (1 - \tau_\ell) v_{k+1}(x + \lambda_\ell) S_\ell \right).$$

Then $p_{\text{adapt}}(n) \geq v_0(0)$. The recursion only visits the values $x = \sum_\ell n_\ell \lambda_\ell$ with $\sum_\ell n_\ell \leq n$, so it needs $O(n^m)$ eigenvalue computations of $d \times d$ matrices. The value $v_0(0)$ is the exact optimal error of the problem in which the tester also learns the level after each use.

Proof. By Proposition 7.3.1, the error is at least $\frac{1}{2} \sum_{\vec{\ell}} \min\{P_\Phi, P_\Psi\} = \sum_{\vec{\ell}} P_\Psi(\vec{\ell}) v_n(L(\vec{\ell}))$, where records with $P_\Psi(\vec{\ell}) = 0$ contribute nothing. For a prefix of length k with Ψ -weight b and log-likelihood ratio x , the remaining contribution is b times a function of x and of the future inputs, and the next step enters linearly through $a_\ell(\rho_{k+1}) = \text{tr}(S_\ell \rho_{k+1})$. Backward induction with $\text{tr}(\rho M) \geq \lambda_{\min}(M)$ shows that this contribution is at least $b v_k(x)$, and choosing at each step an eigenvector for $\lambda_{\min}$ attains the bound when the level is revealed. $\square$

7.4 The geometric Chernoff converse

Theorem 7.4.1. *For every pair of channels, every $n \geq 1$ and every adaptive strategy with n uses,*

$$\frac{1}{2} - \frac{1}{4} \|\rho_\Phi^{(n)} - \rho_\Psi^{(n)}\|_1 \geq \frac{1}{4} \exp\left(-n \hat{C}(\Phi, \Psi) - \sqrt{2 \ln 4} \Lambda \sqrt{n}\right).$$

In particular $\xi_{\text{adapt}}(\Phi, \Psi) \leq \hat{C}(\Phi, \Psi)$.

Proof. We may assume $\hat{C} < \infty$. By Proposition 7.3.1 the error is at least $\frac{1}{2} \sum_{\vec{\ell}} \min\{P_\Phi(\vec{\ell}), P_\Psi(\vec{\ell})\}$, and we may restrict the sum to records with all $\ell_k \in \mathcal{I}$.

Tilting. For $\ell \in \mathcal{I}$ put $w_\ell(s) = g_s(\tau_\ell) = (1 - \tau_\ell) e^{s \lambda_\ell}$. For each k let $s_k \in [0, 1]$ minimise $s \mapsto Q_s(\rho_k)$; it depends only on the past. Let R be the law of the adaptive process on $\mathcal{I}^n$ with $R(\ell_k = \ell \mid \text{past}) = a_\ell(\rho_k) w_\ell(s_k) / Q_{s_k}(\rho_k)$. With $L = \sum_k \lambda_{\ell_k}$ and $M = \sum_k s_k \lambda_{\ell_k}$, a case distinction on the sign of L gives the identity

$$\min \left\{ \prod_k \tau_{\ell_k}, \prod_k (1 - \tau_{\ell_k}) \right\} = \prod_k (1 - \tau_{\ell_k}) e^{\min\{L, 0\}} = \prod_k w_{\ell_k}(s_k) e^{\min\{X, Y\}},$$

where $X = \sum_k (1 - s_k) \lambda_{\ell_k}$ and $Y = -\sum_k s_k \lambda_{\ell_k}$. Moreover

$$\prod_k a_{\ell_k}(\rho_k) w_{\ell_k}(s_k) = R(\vec{\ell}) \prod_k Q_{s_k}(\rho_k) \geq R(\vec{\ell}) e^{-n\hat{C}},$$

because $Q_s(\rho) \geq \lambda_{\min}(B_s) \geq e^{-\hat{C}}$. Hence the error is at least $\frac{1}{2} e^{-n\hat{C}} \mathbb{E}_R[e^{\min\{X, Y\}}]$.

Drift. The function $s \mapsto \ln Q_s(\rho) = \ln \sum_{\ell \in \mathcal{I}} a_\ell(\rho)(1 - \tau_\ell) e^{s\lambda_\ell}$ is smooth and convex, and $\mathbb{E}_R[\lambda_{\ell_k} \mid \text{past}] = \delta_k := \frac{d}{ds} \ln Q_s(\rho_k)|_{s=s_k}$. Minimality of s_k gives $\delta_k = 0$ if $0 < s_k < 1$, $\delta_k \geq 0$ if $s_k = 0$, and $\delta_k \leq 0$ if $s_k = 1$. In all cases $(1 - s_k)\delta_k \geq 0$ and $-s_k\delta_k \geq 0$, so X and Y are R -submartingales whose increments lie in $[-\Lambda, \Lambda]$.

Concentration. By the Azuma–Hoeffding inequality [12] applied to the martingale parts, $R(X \leq -n\delta) + R(Y \leq -n\delta) \leq 2e^{-n\delta^2/(2\Lambda^2)}$. Choosing $\delta = \Lambda\sqrt{2\ln 4/n}$ makes this at most $\frac{1}{2}$, so $\mathbb{E}_R[e^{\min\{X, Y\}}] \geq \frac{1}{2} e^{-n\delta}$, which gives the bound. $\square$

Remark 7.4.2. The proof uses no chain rule. For unequal priors $p, 1 - p$ the same bound holds with $\frac{1}{4}$ replaced by $\frac{1}{2} \min\{p, 1 - p\}$. If all levels are ambiguous, $\Lambda = d_T(\Phi, \Psi)$.

The proof only used that Φ and Ψ are mixtures of common completely positive pieces with weights τ_ℓ and $1 - \tau_\ell$. Any other such decomposition gives a converse in the same way, and the level decomposition gives the best one.

Proposition 7.4.3 (Canonicity). *Let $\Phi = \sum_{j=1}^J p_j \Gamma_j$ and $\Psi = \sum_j q_j \Gamma_j$ with nonzero completely positive maps Γ_j and $p_j, q_j \geq 0$, $p_j + q_j > 0$. Proposition 7.3.1 and the proof of Theorem 7.4.1 apply verbatim to this decomposition and give a converse with exponent*

$$C_\Gamma = -\ln \min_{0 \leq s \leq 1} \lambda_{\min} \left(\sum_{j: p_j q_j > 0} p_j^s q_j^{1-s} \Gamma_j^*(\mathbf{1}) \right).$$

For every such decomposition $C_\Gamma \geq \hat{C}(\Phi, \Psi)$, with equality for the level decomposition. In particular, if $\Phi = \sum_j p_j \mathcal{E}_j$ and $\Psi = \sum_j q_j \mathcal{E}_j$ for channels $\mathcal{E}_j$, then the Chernoff information of (p, q) is at least $\hat{C}(\Phi, \Psi)$.

Proof. Since $\Omega - (p_j + q_j)\Gamma_j$ is completely positive, the ARN theorem gives $R_j \geq 0$ with $(p_j + q_j)\Gamma_j = \Psi_{R_j}$. Injectivity of $K \mapsto \Psi_K$ and $\sum_j \Psi_{R_j} = \Omega = \Psi_{\mathbf{1}}$ give $\sum_j R_j = \mathbf{1}$, and $\Phi = \sum_j t_j \Psi_{R_j}$ with $t_j = p_j/(p_j + q_j)$ gives $T = \sum_j t_j R_j$. For $0 < s < 1$ the function g_s is operator concave on $[0, 1]$: $g_s(T) = (\mathbf{1} - T) \#_s T$, and the weighted geometric mean is jointly concave [79]. Jensen's operator inequality [51] therefore gives $\sum_j g_s(t_j) R_j \leq g_s(\sum_j t_j R_j) = g_s(T)$. The map $K \mapsto \Psi_K^*(\mathbf{1})$ is positive and $g_s(t_j)(p_j + q_j) = p_j^s q_j^{1-s}$, so

$$\sum_j p_j^s q_j^{1-s} \Gamma_j^*(\mathbf{1}) = \sum_j g_s(t_j) \Psi_{R_j}^*(\mathbf{1}) \leq \Psi_{g_s(T)}^*(\mathbf{1}) = B_s.$$

Taking $\lambda_{\min}$ and using continuity in s at the endpoints gives $C_\Gamma \geq \hat{C}$. For channels $\mathcal{E}_j$ we have $\mathcal{E}_j^*(\mathbf{1}) = \mathbf{1}$, and C_Γ is the Chernoff information of (p, q) . $\square$

Proposition 7.4.4 (Operational meaning). *Let $\tilde{\Phi}(X) = \sum_\ell \tau_\ell \Omega_\ell(X) \otimes e_\ell e_\ell^*$ and $\tilde{\Psi}(X) = \sum_\ell (1 - \tau_\ell) \Omega_\ell(X) \otimes e_\ell e_\ell^*$, the channels that also announce the level. Then $\xi_{\text{adapt}}(\tilde{\Phi}, \tilde{\Psi}) = \xi_{\text{prod}}(\tilde{\Phi}, \tilde{\Psi}) = \hat{C}(\Phi, \Psi)$.*

Proof. The flagged pair has the same τ_ℓ and S_ℓ , so Theorem 7.4.1 applies to it with the same $\hat{C}$. With a fixed input ρ , a product strategy that keeps the flags sees i.i.d. records with laws $(a_\ell(\rho)\tau_\ell)_\ell$ and $(a_\ell(\rho)(1 - \tau_\ell))_\ell$, whose Chernoff information is $-\ln \min_s Q_s(\rho)$. Maximising over ρ gives $-\ln \min_s \min_\rho Q_s(\rho) = \hat{C}$. $\square$

Corollary 7.4.5. (a) $-\ln \widehat{Q}_{1/2}(\Phi\|\Psi) \leq \widehat{C}(\Phi, \Psi) \leq -2\ln \widehat{Q}_{1/2}(\Phi\|\Psi)$. The right-hand side is the geometric fidelity converse of [72, 98].

(b) Let $\theta \mapsto \Phi_\theta$ be a C^2 family of channels with $J(\Phi_\theta) > 0$. Then

$$\widehat{C}(\Phi_\theta, \Phi_{\theta+\delta}) = \frac{\delta^2}{8} F_{\text{RLD}}(\theta) + O(\delta^3), \quad -2\ln \widehat{Q}_{1/2}(\Phi_\theta\|\Phi_{\theta+\delta}) = \frac{\delta^2}{4} F_{\text{RLD}}(\theta) + O(\delta^3),$$

where $F_{\text{RLD}}(\theta) = \lambda_{\max}(\text{Tr}_{\mathcal{Y}}[(\partial_\theta J)J^{-1}(\partial_\theta J)])$, $J = J(\Phi_\theta)$, is the RLD Fisher information of the channel family [72].

Proof. (a) The first inequality is the choice $s = \frac{1}{2}$. For the second, the Cauchy–Schwarz inequality gives $Q_{1/2}(\rho)^2 \leq Q_s(\rho)Q_{1-s}(\rho)$, and weighted AM–GM with (7.2.1) gives $Q_{1-s}(\rho) \leq \sum_\ell a_\ell(\rho)((1-s)\tau_\ell + s(1-\tau_\ell)) \leq 1$. Hence $Q_s(\rho) \geq \widehat{Q}_{1/2}^2$ for all s and ρ .

(b) All levels are ambiguous. Put $x_\ell = 2\tau_\ell - 1$, so that $\sum_\ell a_\ell(\rho) = 2$ and $\sum_\ell a_\ell(\rho)x_\ell = 0$. The expansion $(1+x)^s(1-x)^{1-s} = 1 + (2s-1)x - 2s(1-s)x^2 + O(x^3)$, uniform in $s \in [0, 1]$, gives $Q_s(\rho) = 1 - s(1-s)\chi(\rho) + O(\|2T - \mathbf{1}\|^3)$ with $\chi(\rho) = \sum_\ell a_\ell(\rho)x_\ell^2 = \text{tr } \Psi_{(2T-\mathbf{1})^2}(\rho)$. Hence $\widehat{C} = \frac{1}{4} \max_\rho \chi(\rho) + O(\|2T - \mathbf{1}\|^3)$ and $-2\ln \widehat{Q}_{1/2} = \frac{1}{2} \max_\rho \chi(\rho) + O(\|2T - \mathbf{1}\|^3)$. Since $J(\Phi) - J(\Psi) = W(2T - \mathbf{1})W^*$, $J(\Phi) + J(\Psi) = WW^*$ is invertible here, and W is injective,

$$\max_\rho \chi(\rho) = \lambda_{\max}\left(\text{Tr}_{\mathcal{Y}}[(J(\Phi) - J(\Psi))(J(\Phi) + J(\Psi))^{-1}(J(\Phi) - J(\Psi))]\right),$$

which is $\frac{\delta^2}{2} F_{\text{RLD}}(\theta) + O(\delta^3)$ for $\Phi = \Phi_\theta$, $\Psi = \Phi_{\theta+\delta}$, and $\|2T - \mathbf{1}\| = O(\delta)$. $\square$

We compare with the barycentric converse. For states put $D_{\text{BS}}(\omega\|\rho) = \text{tr}(\omega \ln(\omega^{1/2}\rho^{-1}\omega^{1/2}))$ [14], and for channels $\widehat{D}(\mathcal{T}\|\Phi) = \sup_\rho D_{\text{BS}}((\mathcal{T} \otimes \mathbf{1})(\rho)\|(\Phi \otimes \mathbf{1})(\rho))$. For two channels the bound of [66, Theorem 24] reads $\xi_{\text{adapt}} \leq C_{\text{bar}}(\Phi, \Psi)$ with

$$C_{\text{bar}}(\Phi, \Psi) = \sup_{0 \leq s \leq 1} \inf_{\mathcal{T} \in \mathcal{C}(\mathcal{X}, \mathcal{Y})} [s \widehat{D}(\mathcal{T}\|\Phi) + (1-s) \widehat{D}(\mathcal{T}\|\Psi)].$$

Proposition 7.4.6 (Comparison with the barycentric bound). (a) If the pair is level balanced, then $C_{\text{bar}}(\Phi, \Psi) \leq \widehat{C}(\Phi, \Psi)$.

(b) For the flagged pair of Proposition 7.4.4, $\widehat{C} = \xi_{\text{adapt}} \leq C_{\text{bar}}$.

Proof. (a) Here $B_s = Z(s)\mathbf{1}$ with $Z(s) = \sum_\ell c_\ell g_s(\tau_\ell)$, so $\widehat{C} = -\ln \min_s Z(s)$ and $Z'(s) = \sum_{\ell \in \mathcal{I}} c_\ell g_s(\tau_\ell) \lambda_\ell$. For $s \in [0, 1]$ the map $\mathcal{T}_s = \Psi_{g_s(T)}/Z(s)$ is a channel, since $\Psi_{g_s(T)}^*(\mathbf{1}) = B_s$. Its Choi matrix $W g_s(T) W^*/Z(s)$ is supported in $W \text{ran } P_{\mathcal{I}}$, where $P_{\mathcal{I}} = \sum_{\ell \in \mathcal{I}} P_\ell$, and this subspace lies in the supports of $J(\Phi)$ and of $J(\Psi)$ because T and $\mathbf{1} - T$ are invertible on $\text{ran } P_{\mathcal{I}}$; in particular $D_{\max}(\mathcal{T}_s\|\Phi)$ and $D_{\max}(\mathcal{T}_s\|\Psi)$ are finite. By the closed form of the Belavkin–Staszewski channel divergence under such a bound [42, 66] (Lemma 4.6.5 and Theorem 4.6.7(a) in Chapter 4), $\widehat{D}(\mathcal{T}_s\|\Phi) = \lambda_{\max} \text{Tr}_{\mathcal{Y}} [J^{1/2} \ln(J^{1/2} J(\Phi)^{-1} J^{1/2}) J^{1/2}]$ with $J = J(\mathcal{T}_s)$. The operator perspective is covariant under congruences, and W is injective, so the operator inside the partial trace equals $W h(T) W^*$, where $h(t) = \frac{g_s(t)}{Z(s)} \ln \frac{g_s(t)}{Z(s)t}$ is evaluated on the commuting pair $g_s(T)/Z(s)$, T . Since $\text{Tr}_{\mathcal{Y}}(W K W^*) = \Psi_K^*(\mathbf{1})^\top$ and $\ln \frac{g_s(t)}{t} = -(1-s) \ln \frac{t}{1-t}$, level balance gives, since conclusive levels contribute nothing,

$$\widehat{D}(\mathcal{T}_s\|\Phi) = \sum_{\ell \in \mathcal{I}} c_\ell h(\tau_\ell) = -(1-s) \frac{Z'(s)}{Z(s)} - \ln Z(s), \quad \widehat{D}(\mathcal{T}_s\|\Psi) = s \frac{Z'(s)}{Z(s)} - \ln Z(s),$$

the second by the same computation. Let s_* minimise Z on $[0, 1]$. Then $(1 - s_*)Z'(s_*) \geq 0$ and $s_*Z'(s_*) \leq 0$, so both divergences of $\mathcal{T}_{s_*}$ are at most $-\ln Z(s_*) = \widehat{C}$, and $C_{\text{bar}} \leq \sup_s [s\widehat{D}(\mathcal{T}_{s_*} \|\Phi) + (1 - s)\widehat{D}(\mathcal{T}_{s_*} \|\Psi)] \leq \widehat{C}$.

(b) Proposition 7.4.4 and $\xi_{\text{adapt}} \leq C_{\text{bar}}$. $\square$

Part (a) applies to replacement channels $\Phi = \text{tr}(\cdot)\phi$, $\Psi = \text{tr}(\cdot)\psi$ with invertible ϕ, ψ ; there $\widehat{C} = -\ln \min_s \text{tr}(\psi \#_s \phi)$, while $\xi_{\text{adapt}} = \xi_{\text{QC}}(\phi, \psi)$ because the outputs do not depend on the inputs. Part (b) applies to measurement channels whose POVM elements are proportional, $\Phi(\rho) = \sum_\ell a_\ell \langle e_\ell, \rho e_\ell \rangle e'_\ell e'^*_{\ell}$ and $\Psi(\rho) = \sum_\ell b_\ell \langle e_\ell, \rho e_\ell \rangle e'_\ell e'^*_{\ell}$ with orthonormal e'_ℓ , which are equivalent to their flagged pairs because the output reveals the level. Section 7.7 gives examples with strict inequality in both cases, so neither converse dominates the other. For level-balanced pairs whose levels are orthogonal at some input, (a) and Theorem 7.5.3 below give $C_{\text{bar}} = \widehat{C} = \xi_{\text{adapt}}$. At pairs with trivially intersecting Choi ranges both converses are infinite (Theorem 7.6.4(d)).

Corollary 7.4.7 (Strong converse). *Suppose $\tau_\ell < 1$ for all ℓ , that is, $D_{\max}(\Phi \|\Psi) < \infty$, and for $\alpha > 1$ put $D_\alpha^{\text{lev}}(\Phi \|\Psi) = \frac{1}{\alpha-1} \ln \lambda_{\max}(\sum_\ell \tau_\ell^\alpha (1 - \tau_\ell)^{1-\alpha} S_\ell)$. If an adaptive strategy with n uses has type-II error $\beta_n \leq e^{-nr}$ (accepting Φ when Ψ is true), then its type-I success probability satisfies*

$$1 - \alpha_n \leq \exp \left(-n \sup_{\alpha > 1} \frac{\alpha-1}{\alpha} (r - D_\alpha^{\text{lev}}(\Phi \|\Psi)) \right).$$

For $1 < \alpha \leq 2$, D_α^{lev} is the geometric Rényi divergence of the channels [42] (Corollary 4.6.12 in Chapter 4), and for these orders the bound also follows from its chain rule [42].

Proof. By Proposition 7.3.1 the pair $(1 - \alpha_n, \beta_n)$ is achieved by a classical test on records, and classical data processing for $\alpha > 1$ gives $(1 - \alpha_n)^\alpha \beta_n^{1-\alpha} \leq \sum_\ell P_\Phi^\alpha P_\Psi^{1-\alpha}$. Summing over ℓ_n , then ℓ_{n-1} , and so on, and using $\sum_\ell a_\ell(\rho) \tau_\ell^\alpha (1 - \tau_\ell)^{1-\alpha} \leq \lambda_{\max}(\sum_\ell \tau_\ell^\alpha (1 - \tau_\ell)^{1-\alpha} S_\ell)$, bounds the right-hand side by $e^{n(\alpha-1)D_\alpha^{\text{lev}}}$. $\square$

Remark 7.4.8. As $\alpha \downarrow 1$, $D_\alpha^{\text{lev}} \rightarrow \lambda_{\max}(\sum_\ell \tau_\ell \lambda_\ell S_\ell) = \widehat{D}(\Phi \|\Psi)$, the Belavkin–Staszewski channel divergence (Theorem 4.7.5 in Chapter 4); if some $\tau_\ell = 1$, both are $+\infty$ by the same theorem. These rates are robust: for two splittings of one sum with derivatives $T_1, T_2 \leq (1 - \varepsilon)\mathbf{1}$, the quantities $e^{(\alpha-1)D_\alpha^{\text{lev}}}$, $1 < \alpha \leq 2$, differ by at most $2((4 - \alpha)\varepsilon^{1-\alpha} + (\alpha - 1)\varepsilon^{-\alpha})\|T_1 - T_2\|_\infty$ (Corollary 4.6.12). The optimal adaptive Stein exponent is the regularised Umegaki channel divergence $D^{\text{reg}}(\Phi \|\Psi) \leq \widehat{D}(\Phi \|\Psi)$ [41], so the walk does not improve Stein’s exponent; Corollary 7.4.7 is useful for rates r above $\widehat{D}(\Phi \|\Psi)$.

7.5 When the bound is attained

Theorem 7.5.1 (Exact exponent). *Suppose that $\sigma_* \in \mathcal{D}(\mathcal{X})$ and $s_* \in [0, 1]$ satisfy $Q_{s_*}(\sigma_*) = e^{-\widehat{C}}$ and that the levels are orthogonal at σ_* , equivalently $[T, G_{\sigma_*}] = 0$. Then $\xi_{\text{adapt}} = \xi_{\text{prod}} = \widehat{C}$, attained by the product inputs $u^{\otimes n}$ with $u = \text{vec}(\sigma_*^{1/2})$.*

Proof. By Lemma 7.2.4(b), $\min_{0 \leq s \leq 1} \text{tr}(\Phi_u^s \Psi_u^{1-s}) \leq Q_{s_*}(\sigma_*) = e^{-\widehat{C}}$, so $\xi_{\text{QC}}(\Phi_u, \Psi_u) \geq \widehat{C}$, which the product strategy $u^{\otimes n}$ achieves [11, 99]. Theorem 7.4.1 gives the reverse inequality. $\square$

Remark 7.5.2 (Classical channels). For classical channels $P(y|x)$, $Q(y|x)$ the Kraus operators of Ω can be taken to be multiples of $e_y e_x^*$. Then B_s is diagonal with entries

$\sum_y P(y|x)^s Q(y|x)^{1-s}$, so $\hat{C} = \max_x C(P(\cdot|x), Q(\cdot|x))$, and the levels are orthogonal at every basis state. Theorem 7.5.1 with $\sigma_* = e_{x_*} e_{x_*}^*$ for a maximising letter x_* is Hayashi's theorem [56]. Such pairs are in general not level balanced.

Theorem 7.5.3 (Exact optimality at every n). *Suppose the pair is level balanced, $S_\ell = c_\ell \mathbf{1}$, and the levels are orthogonal at some $\sigma \in D(\mathcal{X})$. Let $\nu_\Phi(\ell) = \tau_\ell c_\ell$ and $\nu_\Psi(\ell) = (1 - \tau_\ell) c_\ell$, and let $u = \text{vec}(A)$ with $AA^* = \sigma$. Then for every n :*

- (a) *the final pair of every adaptive strategy is a post-processing of $(\nu_\Phi^{\otimes n}, \nu_\Psi^{\otimes n})$;*
- (b) *the pairs $(\Phi_u^{\otimes n}, \Psi_u^{\otimes n})$ and $(\nu_\Phi^{\otimes n}, \nu_\Psi^{\otimes n})$ are post-processings of each other.*

Consequently the product strategy $u^{\otimes n}$ is optimal among all adaptive strategies with n uses, simultaneously for every binary error criterion: Bayesian error with arbitrary priors and the whole Neyman–Pearson trade-off. All error exponents (Stein, Chernoff, Hoeffding, strong converse) are those of the classical pair (ν_Φ, ν_Ψ) , and, with D the Umegaki channel divergence and D^{reg} its regularisation,

$$D(\Phi\|\Psi) = D^{\text{reg}}(\Phi\|\Psi) = \hat{D}(\Phi\|\Psi) = D(\nu_\Phi\|\nu_\Psi).$$

Proof. (a) Level balance gives $a_\ell(\rho) = c_\ell$ for every ρ , so in (7.3.1) $P_\Phi = \nu_\Phi^{\otimes n}$ and $P_\Psi = \nu_\Psi^{\otimes n}$ for every strategy; apply Proposition 7.3.1. (b) By Lemma 7.2.4(b), $\Phi_u = \sum_\ell \tau_\ell R_\ell$ and $\Psi_u = \sum_\ell (1 - \tau_\ell) R_\ell$ with $R_\ell = \sum_{i: t_i = \tau_\ell} g_i x_i x_i^*$; these have orthogonal supports and $\text{tr } R_\ell = \text{tr}(P_\ell G_\sigma) = a_\ell(\sigma) = c_\ell$. Measuring the projections onto the supports of the R_ℓ produces (ν_Φ, ν_Ψ) , and preparing R_ℓ/c_ℓ on outcome ℓ reproduces the outputs; tensor powers follow. For the last display, $D(\Phi\|\Psi) \geq D(\Phi_u\|\Psi_u) = D(\nu_\Phi\|\nu_\Psi)$, while $D \leq D^{\text{reg}} \leq \hat{D}$ by additivity of the Belavkin–Staszewski channel divergence [42], and $\hat{D}(\Phi\|\Psi) = \lambda_{\max}(\sum_\ell \tau_\ell \lambda_\ell S_\ell) = \sum_\ell c_\ell \tau_\ell \lambda_\ell = D(\nu_\Phi\|\nu_\Psi)$ by the closed form of Theorem 4.7.5(b) in Chapter 4. $\square$

Remark 7.5.4 (Relation to programmable simulation). Level balance says that $\Phi = \sum_\ell \nu_\Phi(\ell) \mathcal{E}_\ell$ and $\Psi = \sum_\ell \nu_\Psi(\ell) \mathcal{E}_\ell$ are mixtures of the same channels $\mathcal{E}_\ell = \Omega_\ell/c_\ell$, and orthogonality says that the $\mathcal{E}_\ell$ are perfectly distinguishable at u . Part (a) is then the familiar simulation of each use by a classical program [107]. The level decomposition makes the criterion canonical and checkable: compute T , its spectral projections and the S_ℓ . Theorem 7.4.1 is the extension to programs whose law depends on the input, and Proposition 7.4.3 shows that it is never weaker than a program simulation.

Corollary 7.5.5 (Random-unitary channels). *Let $U_1, \dots, U_N$ be unitaries on $\mathcal{X}$ with $\text{tr}(U_i^* U_j) = d \delta_{ij}$, and let $\Phi = \sum_i p_i U_i \cdot U_i^*$, $\Psi = \sum_i q_i U_i \cdot U_i^*$ with probability vectors p, q . Then the Choi strategy $u = \text{vec}(\mathbf{1})/\sqrt{d}$ is exactly optimal among adaptive strategies for every n , and every exponent is that of the classical pair (p, q) .*

Proof. Discard the i with $p_i = q_i = 0$. With $V_i = \sqrt{p_i + q_i} U_i$, T is diagonal with entries $p_i/(p_i + q_i)$, each level map is $\sum_{i \in I_\ell} (p_i + q_i) U_i \cdot U_i^*$ with $S_\ell = \sum_{i \in I_\ell} (p_i + q_i) \mathbf{1}$, and the levels are orthogonal at $\mathbf{1}/d$. The level laws are $\nu_\Phi(\ell) = \sum_{i \in I_\ell} p_i$, and (p, q) is a refinement of (ν_Φ, ν_Ψ) with the same likelihood ratios. $\square$

The U_i need not belong to a nice error basis: for example $\{\mathbf{1}, Z, XD_1, X^2 D_2\}$ on $\mathbb{C}^3$, with X, Z the Weyl operators and D_1, D_2 diagonal unitaries, is an orthogonal family, and $XD_1, X^2 D_2$ are not multiples of Weyl operators for generic D_1, D_2 .

Corollary 7.5.6 (Covariant channels). *Let G act by (possibly projective) unitary representations U on $\mathcal{X}$ and V on $\mathcal{Y}$ with $\Phi(U_g X U_g^*) = V_g \Phi(X) V_g^*$ and $\Psi(U_g X U_g^*) = V_g \Psi(X) V_g^*$.*

Suppose U is irreducible and the representation $K \mapsto V_g^* K U_g$ of G on $\text{span}\{V_k\}$ is multiplicity free. Then the hypotheses of Theorem 7.5.3 hold with $\sigma = \mathbf{1}/d$, so the Choi strategy is exactly optimal for every n and the exponents are those of the commuting Choi states $J(\Phi)/d$, $J(\Psi)/d$.

Proof. On Kraus coefficients the action is by unitaries $\pi(g)$ commuting with T (uniqueness of the ARN derivative), hence with every P_ℓ . Thus each Ω_ℓ is covariant, and S_ℓ commutes with U , so $S_\ell = c_\ell \mathbf{1}$ by irreducibility. The Gram matrix $G_{\mathbf{1}/d}$ also commutes with every $\pi(g)$, and multiplicity freeness makes the commutant of $\{\pi(g)\}$ abelian, so $G_{\mathbf{1}/d}$ commutes with T ; Lemma 7.2.4(a) gives orthogonality at $\mathbf{1}/d$. $\square$

Remark 7.5.7. Corollary 7.5.6 covers Weyl-covariant channels in every dimension, where it also follows from teleportation stretching [107], and all $\text{SU}(2)$ -covariant channels from spin j to spin j' , since $L(\mathbb{C}^{2j+1}, \mathbb{C}^{2j'+1}) \cong \bigoplus_{k=|j-j'|}^{j+j'} \mathcal{V}_k$ is multiplicity free. Such channels are mixtures of the channels whose Kraus operators are the components of an irreducible tensor operator of rank k ; for different k these have Choi matrices with orthogonal supports.

7.6 Continuity of the converse

The converse $\hat{C}$ depends on the pair only through the derivative T and the frame W , so the continuity theory of the Arveson–Radon–Nikodym derivative developed in Chapter 4 applies to it. Two facts emerge. Along perturbations that are small relative to the pair in the completely positive order, $e^{-\hat{C}}$ is Lipschitz, with a constant governed by the Thompson constant Λ (Theorem 7.6.2). Along additive perturbations $\hat{C}$ is only lower semicontinuous, and it can jump to $+\infty$. In fact, near replacement channels with rank-deficient outputs there are pairs for which the fidelity converse, the barycentric converse and $\hat{C}$ are all infinite, while the adaptive exponent tends to zero (Theorem 7.6.5). Section 7.6.1 shows that the same happens for amplitude-damping channels with nearby damping rates.

For two pairs (Φ_1, Ψ_1) and (Φ_2, Ψ_2) we use the notation of Section 4.10 in Chapter 4. We put $\Omega_i = \Phi_i + \Psi_i$, $\Omega' = \Omega_1 + \Omega_2$ and $\bar{\Omega} = \frac{1}{2}\Omega'$. We write T_1, T_2 for the derivatives of Φ_1, Φ_2 within their own pairs, in a common frame, and $\tau = \|T_1 - T_2\|_\infty$. We write $d_T = d_T(\Omega_1, \Omega_2)$ for the Thompson distance between the sums. The relative distance δ between the pairs is the smallest $d \geq 0$ with $-d\Omega' \leq \Phi_1 - \Phi_2 \leq d\Omega'$ and $-d\Omega' \leq \Psi_1 - \Psi_2 \leq d\Omega'$ in the completely positive order. By Lemma 4.10.2, T_i is unitarily equivalent to the derivative of Section 7.2 for the pair (Φ_i, Ψ_i) , so its eigenvalues are the levels τ_ℓ of that pair. For channels $\|\bar{\Omega}^*(\mathbf{1})\|_\infty = 2$ and $\|\Omega'^*(\mathbf{1})\|_\infty = 4$.

Lemma 7.6.1. *Let $0 < \varepsilon \leq \frac{1}{2}$ and let E, E' be Hermitian with spectra in $[\varepsilon, 1 - \varepsilon]$. For every $s \in [0, 1]$,*

$$\|g_s(E) - g_s(E')\|_\infty \leq M(\varepsilon) \|E - E'\|_\infty, \quad M(\varepsilon) = \max_{0 \leq s \leq 1} (s\varepsilon^{s-1} + (1-s)\varepsilon^{-s}) \leq \varepsilon^{-1}.$$

Moreover $M(\varepsilon) = \varepsilon^{-1/2}$ for $\varepsilon \geq \frac{1}{50}$.

Proof. On $[\varepsilon, 1 - \varepsilon]$ we have $g_s(t) = t^s(1-t)^{1-s}$, and E commutes with $\mathbf{1} - E$, so

$$g_s(E) - g_s(E') = (E^s - E'^s)(\mathbf{1} - E)^{1-s} + E'^s((\mathbf{1} - E)^{1-s} - (\mathbf{1} - E')^{1-s}),$$

where $(\mathbf{1} - E)^{1-s}$ and E'^s have norm at most 1. For $0 < p < 1$ and $X, Y \geq \varepsilon \mathbf{1}$, the representation $x^p = \frac{\sin p\pi}{\pi} \int_0^\infty \lambda^{p-1} (1 - \lambda(x + \lambda)^{-1}) d\lambda$ and $\|(X + \lambda)^{-1} - (Y + \lambda)^{-1}\|_\infty \leq (\varepsilon + \lambda)^{-2} \|X - Y\|_\infty$ give $\|X^p - Y^p\|_\infty \leq p\varepsilon^{p-1} \|X - Y\|_\infty$, since $\frac{\sin p\pi}{\pi} \int_0^\infty \lambda^p (\varepsilon + \lambda)^{-2} d\lambda =$

$p\varepsilon^{p-1}$; the cases $p \in \{0, 1\}$ are trivial. Applying this with $p = s$ to E, E' and with $p = 1 - s$ to $\mathbf{1} - E, \mathbf{1} - E'$ gives the bound with $s\varepsilon^{s-1} + (1 - s)\varepsilon^{-s} \leq s\varepsilon^{-1} + (1 - s)\varepsilon^{-1}$. The value $\varepsilon^{-1/2}$ is the value at $s = \frac{1}{2}$, and a direct computation shows that this is the maximum for $\varepsilon \geq \frac{1}{50}$. $\square$

For small ε one has $M(\varepsilon) \sim (e\varepsilon \ln \frac{1}{\varepsilon})^{-1}$. The spectrum of the derivative of a pair lies in $[\varepsilon, 1 - \varepsilon]$ exactly when all its levels are ambiguous and $\Lambda \leq \ln \frac{1-\varepsilon}{\varepsilon}$, that is, $\varepsilon = (1 + e^\Lambda)^{-1}$ (Lemma 7.2.1).

Theorem 7.6.2 (Robustness). *Let (Φ_1, Ψ_1) and (Φ_2, Ψ_2) be pairs of channels with $\text{ran } J(\Phi_1 + \Psi_1) = \text{ran } J(\Phi_2 + \Psi_2)$, and suppose that the levels of both pairs lie in $[\varepsilon, 1 - \varepsilon]$ for some $\varepsilon \in (0, \frac{1}{2}]$. Then*

$$|e^{-\widehat{C}(\Phi_1, \Psi_1)} - e^{-\widehat{C}(\Phi_2, \Psi_2)}| \leq 2M(\varepsilon)\tau + 4\sqrt{2}\tanh \frac{d_T}{2}.$$

If $\delta < \frac{1}{2}$, the right-hand side is at most $(4\sqrt{2}M(\varepsilon)(1 - 4\delta^2)^{-1/2} + 8\sqrt{2})\delta$. If $\Phi_1 + \Psi_1 = \Phi_2 + \Psi_2$, then $d_T = 0$ and τ is the distance between the derivatives with respect to the common sum.

Proof. For $0 < s < 1$, Remark 7.2.2 and the proof of Theorem 4.10.4(c) give

$$\lambda_{\min}(B_s^{(i)}) = \lambda_{\min}(\text{Tr}_Y(W a_i g_s(T_i) a_i W^*)),$$

where $B_s^{(i)}$ is the operator B_s of the pair (Φ_i, Ψ_i) , W is the frame of Ω' and a_1, a_2 are as in Section 4.10. The partial-trace estimate (Lemma 4.4.10) for Ω' and the sandwich estimate (Lemma 4.10.3), with $I = [\varepsilon, 1 - \varepsilon]$, $\varphi = g_s$ and $2\gamma = \tanh \frac{d_T}{2}$ (Lemma 4.10.1), give

$$\begin{aligned} |\lambda_{\min}(B_s^{(1)}) - \lambda_{\min}(B_s^{(2)})| &\leq 4\left(\frac{1}{2}\omega_{g_s}^I(\tau) + \sqrt{2}\max_I g_s \tanh \frac{d_T}{2}\right) \\ &\leq 2M(\varepsilon)\tau + 4\sqrt{2}\tanh \frac{d_T}{2}, \end{aligned}$$

by Lemma 7.6.1 and $g_s \leq 1$. Both sides are continuous in s on $[0, 1]$, because all levels are ambiguous, so the bound holds for all $s \in [0, 1]$. Since $|\min_s x_s - \min_s y_s| \leq \sup_s |x_s - y_s|$, the first claim follows. For the second, $\tau \leq 2\sqrt{2}\delta(1 - 4\delta^2)^{-1/2}$ and $\tanh \frac{d_T}{2} \leq 2\delta$ by Lemmas 4.10.1 and 4.10.5. $\square$

Since $\delta \leq \max\{\|\Phi_1 - \Phi_2\|_\diamond, \|\Psi_1 - \Psi_2\|_\diamond\}/\nu(\Phi_1 + \Psi_1)^2$ whenever $\text{ran } J(\Phi_2 + \Psi_2) \subseteq \text{ran } J(\Phi_1 + \Psi_1)$ (Corollary 4.10.8), Theorem 7.6.2 makes $e^{-\widehat{C}}$ locally Lipschitz in the diamond norm along such perturbations. Here ν is the smallest operator norm of a normalised combination of Kraus operators (Section 4.2 and Lemma 4.8.2). The theorem also gives a converse that is valid for a whole neighbourhood of a model pair.

Corollary 7.6.3 (Converse for an uncertain pair). *Let (Φ_0, Ψ_0) be a pair of channels whose levels lie in $[\varepsilon, 1 - \varepsilon]$, let $\delta < \frac{1}{2}$ and $\kappa = 2\sqrt{2}\delta(1 - 4\delta^2)^{-1/2} < \varepsilon$, and put $R = 2M(\varepsilon - \kappa)\kappa + 8\sqrt{2}\delta$. Every pair of channels (Φ, Ψ) within relative distance δ of (Φ_0, Ψ_0) has its levels in $[\varepsilon - \kappa, 1 - \varepsilon + \kappa]$, and if $R < e^{-\widehat{C}(\Phi_0, \Psi_0)}$, then*

$$\xi_{\text{adapt}}(\Phi, \Psi) \leq \widehat{C}(\Phi, \Psi) \leq -\ln(e^{-\widehat{C}(\Phi_0, \Psi_0)} - R).$$

The finite- n bound of Theorem 7.4.1 holds for (Φ, Ψ) with this exponent and $\Lambda \leq \ln \frac{1-\varepsilon+\kappa}{\varepsilon-\kappa}$.

Proof. By Lemma 4.10.5 the sums have the same Choi support and $\tau \leq \kappa$, so the levels of (Φ, Ψ) are within κ of those of (Φ_0, Ψ_0) by Weyl's inequality. Apply Theorem 7.6.2 with $\varepsilon - \kappa$ in place of ε , and Theorem 7.4.1. $\square$

Additive perturbations behave differently. Recall that $\hat{C}$ is defined for every pair of completely positive maps by the same formula.

Theorem 7.6.4 (Semicontinuity). *Let (Φ, Ψ) be a pair of channels, and let $\mathcal{I}$, S_ℓ and Ω_ℓ be as in Section 7.2.*

- (a) $e^{-\hat{C}}$ is upper semicontinuous: if $(\Phi_n, \Psi_n) \rightarrow (\Phi, \Psi)$ in any norm, then $\liminf_n \hat{C}(\Phi_n, \Psi_n) \geq \hat{C}(\Phi, \Psi)$.
- (b) Mixing with common noise never increases $\hat{C}$: for every channel Θ and $\varepsilon \in [0, 1]$, $\hat{C}((1-\varepsilon)\Phi + \varepsilon\Theta, (1-\varepsilon)\Psi + \varepsilon\Theta) \leq \hat{C}(\Phi, \Psi)$, and the left-hand side tends to $\hat{C}(\Phi, \Psi)$ as $\varepsilon \downarrow 0$.
- (c) If $J(\Phi + \Psi)$ is invertible, then $\hat{C}$ is continuous at (Φ, Ψ) .
- (d) $\hat{C}(\Phi, \Psi) = +\infty$ if and only if the operators S_ℓ , $\ell \in \mathcal{I}$, have a common nonzero kernel vector, that is, some input state σ has $\text{tr } \Omega_\ell(\sigma) = 0$ for every ambiguous level. This is the case whenever $\text{ran } J(\Phi) \cap \text{ran } J(\Psi) = \{0\}$, and then also $-2 \ln \hat{Q}_{1/2}(\Phi \| \Psi) = C_{\text{bar}}(\Phi, \Psi) = +\infty$.

Proof. Since $s \mapsto \lambda_{\min}(B_s)$ is continuous on $[0, 1]$ and equals $\hat{Q}_s(\Phi \| \Psi)$ for $0 < s < 1$ (Remark 7.2.2), we have $e^{-\hat{C}} = \inf_{0 < s < 1} \hat{Q}_s(\Phi \| \Psi)$.

(a) Each $\hat{Q}_s$ is upper semicontinuous (Lemma 4.12.1(c)), and an infimum of upper semicontinuous functions is upper semicontinuous.

(b) Write $\Phi_\varepsilon, \Psi_\varepsilon$ for the mixtures. Joint concavity of the weighted geometric mean [79] and $J(\Theta) \#_s J(\Theta) = J(\Theta)$ give $J(\Psi_\varepsilon) \#_s J(\Phi_\varepsilon) \geq (1-\varepsilon)J(\Psi) \#_s J(\Phi) + \varepsilon J(\Theta)$. Since $\text{Tr}_{\mathcal{Y}} J(\Theta) = \mathbf{1}$ and $\lambda_{\min}$ is superadditive, $\hat{Q}_s(\Phi_\varepsilon \| \Psi_\varepsilon) \geq (1-\varepsilon)\hat{Q}_s(\Phi \| \Psi) + \varepsilon \geq \hat{Q}_s(\Phi \| \Psi)$, because $B_s \leq \sum_\ell (s\tau_\ell + (1-s)(1-\tau_\ell))S_\ell = \mathbf{1}$ by (7.2.1). Taking the infimum over s gives the inequality. Conversely, $J(\Phi_\varepsilon) \leq J(\Phi) + \varepsilon J(\Theta)$ and $J(\Psi_\varepsilon) \leq J(\Psi) + \varepsilon J(\Theta)$, so monotonicity and continuity from above of the mean give $\limsup_{\varepsilon \downarrow 0} \hat{Q}_s(\Phi_\varepsilon \| \Psi_\varepsilon) \leq \hat{Q}_s(\Phi \| \Psi)$ for each $s \in (0, 1)$, and hence $\limsup_{\varepsilon \downarrow 0} e^{-\hat{C}(\Phi_\varepsilon, \Psi_\varepsilon)} \leq e^{-\hat{C}(\Phi, \Psi)}$.

(c) Let $(\Phi_n, \Psi_n) \rightarrow (\Phi, \Psi)$. Since $J(\Phi + \Psi)$ is invertible, the relative distance δ_n between the pairs tends to 0 (Corollary 4.10.8), and so do τ_n and $\tanh \frac{d_{T,n}}{2}$ (Lemmas 4.10.1 and 4.10.5). Choose $\eta \in (0, \frac{1}{4})$ such that T has no eigenvalue in $(0, 2\eta) \cup (1-2\eta, 1)$. For $0 < s < 1$ let k_s be the function that equals g_s on $[2\eta, 1-2\eta]$, vanishes on $[0, \eta] \cup [1-\eta, 1]$ and is affine in between. Since g_s is concave and nonnegative on $[0, 1]$, $0 \leq k_s \leq g_s$. Since $(s, t) \mapsto g_s(t)$ is uniformly continuous on $[0, 1] \times [\eta, 1-\eta]$, the functions k_s have a common modulus of continuity ω_η . In the frame of the proof of Theorem 7.6.2, $g_s(T_n) \geq k_s(T_n)$ and positivity of $X \mapsto \text{Tr}_{\mathcal{Y}}(WXW^*)$ give

$$\begin{aligned} \hat{Q}_s(\Phi_n \| \Psi_n) &\geq \lambda_{\min} \text{Tr}_{\mathcal{Y}} (W a_n k_s(T_n) a_n W^*) \\ &\geq \lambda_{\min} \text{Tr}_{\mathcal{Y}} (W a k_s(T) a W^*) - 2\omega_\eta(\tau_n) - 4\sqrt{2} \tanh \frac{d_{T,n}}{2}, \end{aligned}$$

by the sandwich estimate with $\varphi = k_s$. By the choice of η , $k_s(T) = g_s(T)$, so the first term is $\hat{Q}_s(\Phi \| \Psi)$. Taking the infimum over s and the limit inferior over n shows that $e^{-\hat{C}}$ is lower semicontinuous at (Φ, Ψ) . Together with (a) this gives continuity.

(d) For $0 < s < 1$, $B_s = \sum_{\ell \in \mathcal{I}} g_s(\tau_\ell) S_\ell$ with $g_s(\tau_\ell) > 0$, and B_0, B_1 are combinations of the same S_ℓ with positive coefficients. So $\min_s \lambda_{\min}(B_s) = 0$ if and only if the S_ℓ , $\ell \in \mathcal{I}$, have a common kernel vector. If the Choi ranges intersect trivially, T is a projection (see the beginning of Section 4.12), so $\mathcal{I} = \emptyset$, $B_s = 0$ and $\hat{Q}_{1/2}(\Phi \| \Psi) = 0$. For C_{bar} , let $0 < s < 1$ and let $\mathcal{T}$ be a channel. With the maximally entangled input,

$\widehat{D}(\mathcal{T} \parallel \Phi) \geq \frac{1}{d} D_{\text{BS}}(J(\mathcal{T}) \parallel J(\Phi))$, and the right-hand side is $+\infty$ unless $\text{ran } J(\mathcal{T}) \subseteq \text{ran } J(\Phi)$ (Theorem 4.7.5(a) and Proposition 4.7.1). The same holds for Ψ , and $J(\mathcal{T}) \neq 0$, so one of the two divergences is infinite for every $\mathcal{T}$, and $C_{\text{bar}} = +\infty$. $\square$

Part (b) gives a numerically stable way to compute $\widehat{C}$ from below when $J(\Phi + \Psi)$ is singular. The convergence can be slow. For a random Kraus-rank 2 versus 3 qubit pair mixed with the completely depolarising channel, the gap in $e^{-\widehat{C}}$ was $7.7 \cdot 10^{-2}$, $5.0 \cdot 10^{-2}$, $3.8 \cdot 10^{-2}$ and $3.1 \cdot 10^{-2}$ at $\varepsilon = 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}$, that is, about $0.7 / \ln \frac{1}{\varepsilon}$. In this example the minimum over s is attained at $s = 0$, where conclusive levels $\tau_\ell = 0$, perturbed to $\varepsilon' > 0$, contribute $g_s(\varepsilon') \approx \varepsilon'^s$, which is not small for $s \lesssim 1 / \ln \frac{1}{\varepsilon'}$; so the convergence is only logarithmic. Part (d) shows that $\widehat{C}$ is often infinite. If $k_\Phi + k_\Psi \leq d \dim \mathcal{Y}$ for the Kraus ranks, the Choi ranges are subspaces of complementary dimensions. In our experiments with random pairs (qubit pairs of Kraus ranks 2 and 2, and qubit-to-qutrit pairs of Kraus ranks 3 and 3, 300 pairs in total), they always intersected trivially, so all three single-letter converses were infinite. The finite values of $\widehat{C}$ for low Kraus ranks in Table 7.5 require an ambiguous level, which exists there because $2 + 3 > 4$.

The next result shows that the jumps allowed by Theorem 7.6.4(a) occur, and that they are not seen by the true exponent.

Theorem 7.6.5 (Fragility of single-letter converses). *Let ϕ, ψ be states on $\mathbb{C}^{d'}$ with $\text{rank}(\phi + \psi) \leq d'/2$, and let $\Phi = \text{tr}(\cdot) \phi$ and $\Psi = \text{tr}(\cdot) \psi$. There are states $\phi_t \rightarrow \phi$ and $\psi_t \rightarrow \psi$, at trace distance $O(\sqrt{t})$, such that the replacement channels $\Phi_t = \text{tr}(\cdot) \phi_t$ and $\Psi_t = \text{tr}(\cdot) \psi_t$ satisfy, for every $t > 0$,*

$$\widehat{C}(\Phi_t, \Psi_t) = -2 \ln \widehat{Q}_{1/2}(\Phi_t \parallel \Psi_t) = C_{\text{bar}}(\Phi_t, \Psi_t) = +\infty,$$

whereas

$$\xi_{\text{adapt}}(\Phi_t, \Psi_t) = \xi_{\text{QC}}(\phi_t, \psi_t) \leq -\ln \text{tr}(\phi_t^{1/2} \psi_t^{1/2}) \longrightarrow -\ln \text{tr}(\phi^{1/2} \psi^{1/2}).$$

In particular, if $\phi = \psi$, then (Φ_t, Ψ_t) converges in the diamond norm to a pair of identical channels and $\xi_{\text{adapt}}(\Phi_t, \Psi_t) \rightarrow 0$, while all three converses are infinite along the sequence.

Proof. By Corollary 4.12.5 in Chapter 4 there are states ϕ_t, ψ_t with $\text{ran } \phi_t \cap \text{ran } \psi_t = \{0\}$ at distance $O(\sqrt{t})$ from ϕ, ψ ; they are obtained by a Halmos dilation of the derivative of the pair into the kernel of $\phi + \psi$. Since $J(\Phi_t) = \phi_t \otimes \mathbf{1}$ and $J(\Psi_t) = \psi_t \otimes \mathbf{1}$, the Choi ranges intersect trivially, and Theorem 7.6.4(d) gives the first display. The outputs of a replacement channel do not depend on its input, so the final states of every adaptive strategy are images of $\phi_t^{\otimes n}$ and $\psi_t^{\otimes n}$ under one channel, and the product strategy is optimal. Hence $\xi_{\text{adapt}} = \xi_{\text{QC}}(\phi_t, \psi_t)$ [11, 99], which is at most the value at $s = \frac{1}{2}$. Finally $\|\Phi_t - \Phi\|_\diamond = \|\phi_t - \phi\|_1$, and the square root is continuous. $\square$

Example 7.6.6. Let $\phi = \psi = \frac{1}{2}(e_1 e_1^* + e_2 e_2^*)$ on $\mathbb{C}^4$, and couple e_1 to e_3 and e_2 to e_4 : put $v_{1,2} = (e_1 \pm \sqrt{t} e_3) / \sqrt{1+t}$ and $w_{1,2} = (e_2 \pm \sqrt{t} e_4) / \sqrt{1+t}$, so that $\phi_t = \frac{1}{2}(v_1 v_1^* + w_1 w_1^*)$ and $\psi_t = \frac{1}{2}(v_2 v_2^* + w_2 w_2^*)$. These states are at trace distance $\sqrt{t} + O(t)$ from ϕ and their ranges intersect trivially, while $\text{tr}(\phi_t^{1/2} \psi_t^{1/2}) = (\frac{1-t}{1+t})^2$. Numerically $\xi_{\text{adapt}}(\Phi_t, \Psi_t) = \xi_{\text{QC}}(\phi_t, \psi_t) = 2 \ln \frac{1+t}{1-t} \approx 4t$, attained at $s = \frac{1}{2}$, whereas $B_s = 0$ for all s . The same happens for any two distinct pure output states, the simplest case.

Theorem 7.6.5 explains the gap between $\widehat{C}$ and the true exponent at replacement channels noted after Proposition 7.4.6. By Proposition 7.4.4, $\widehat{C}(\Phi, \Psi)$ is the exponent of the

flagged pair, which announces the level, and at mutually singular pairs every level is conclusive, so the flagged pair is perfectly distinguishable after one use while the original pair need not be. By Theorem 7.6.4(c), this cannot happen near pairs with $J(\Phi + \Psi)$ invertible. By Theorem 4.12.4(c), at every pair with $\dim \ker J(\Phi + \Psi) \geq \text{rank } T(\mathbf{1} - T) \geq 1$ and $\widehat{C} < \infty$ there are completely positive pairs within $O(\sqrt{t})$ for which $\widehat{C} = +\infty$. Whether these can always be chosen to be channels is open (compare the question on singular pairs in Section 4.13); Theorem 7.6.5 settles it for replacement channels.

Remark 7.6.7 (Independent branches). If $\Phi = (1 - p)\Phi_A + p\Phi_B$ and $\Psi = (1 - p)\Psi_A + p\Psi_B$ with $\text{ran } J(\Phi_A + \Psi_A) \cap \text{ran } J(\Phi_B + \Psi_B) = \{0\}$, then every connection splits over the two branches (Corollary 4.11.4). Consequently $B_s(\Phi, \Psi) = (1 - p)B_s(\Phi_A, \Psi_A) + pB_s(\Phi_B, \Psi_B)$, the levels of the mixture are those of the two branches, and $e^{-\widehat{C}(\Phi, \Psi)} \geq (1 - p)e^{-\widehat{C}(\Phi_A, \Psi_A)} + pe^{-\widehat{C}(\Phi_B, \Psi_B)}$.

7.6.1 Amplitude damping

Theorem 7.6.5 concerns replacement channels, whose outputs do not depend on the input. We now show that the same behaviour occurs for amplitude damping, a basic model of physical noise, and give a converse that remains finite there. For $0 < \gamma \leq 1$ let $\mathcal{A}_\gamma$ be the qubit amplitude-damping channel with Kraus operators

$$K_0 = e_0 e_0^* + \sqrt{1 - \gamma} e_1 e_1^*, \quad K_1 = \sqrt{\gamma} e_0 e_1^*,$$

and write L_0, L_1 for the Kraus operators of $\mathcal{A}_{\gamma'}$. For two damping rates put

$$f(\gamma, \gamma') = \sqrt{\gamma\gamma'} + \sqrt{(1 - \gamma)(1 - \gamma')}, \quad \mu_*(\gamma, \gamma') = \frac{\sqrt{\gamma\gamma'}}{1 - \sqrt{(1 - \gamma)(1 - \gamma')}}.$$

Here f is the Bhattacharyya coefficient of the Bernoulli laws $(\gamma, 1 - \gamma)$ and $(\gamma', 1 - \gamma')$, and $\mu_* \leq f \leq 1$. The second inequality is the Cauchy–Schwarz inequality, with equality if and only if $\gamma = \gamma'$, and the first is equivalent to $\sqrt{(1 - \gamma)(1 - \gamma')}(1 - f) \geq 0$.

Proposition 7.6.8 (Levels of an amplitude-damping pair). *Let $0 < \gamma, \gamma' \leq 1$ with $\gamma \neq \gamma'$, and let $\Phi = \mathcal{A}_\gamma$, $\Psi = \mathcal{A}_{\gamma'}$.*

(a) *The pair has three levels: $\tau = 1$ with $\Omega_1 = K_0 \cdot K_0^*$, $\tau = 0$ with $\Omega_0 = L_0 \cdot L_0^*$, and the single ambiguous level $\tau_a = \gamma/(\gamma + \gamma')$ with $\Omega_a = (\gamma + \gamma') e_0 e_1^* \cdot e_1 e_0^*$ and $S_a = (\gamma + \gamma') e_1 e_1^*$.*

(b) *$\text{ran } J(\Phi) \cap \text{ran } J(\Psi) = \text{span}\{\text{vec}(e_0 e_1^*)\} \neq \{0\}$, and nevertheless*

$$\widehat{C}(\Phi, \Psi) = -2 \ln \widehat{Q}_{1/2}(\Phi \| \Psi) = C_{\text{bar}}(\Phi, \Psi) = +\infty.$$

(c) *The common kernel vector of Theorem 7.6.4(d) is e_0 , and $\Phi(e_0 e_0^*) = \Psi(e_0 e_0^*) = e_0 e_0^*$.*

Proof. (a) Since $K_1 = \sqrt{\gamma/\gamma'} L_1$, the maps Φ and Ψ are combinations of $V_1 = K_0$, $V_2 = L_0$ and $V_3 = \sqrt{\gamma + \gamma'} e_0 e_1^*$, namely $\Omega = \Phi + \Psi = \sum_k V_k \cdot V_k^*$, $\Phi = V_1 \cdot V_1^* + \tau_a V_3 \cdot V_3^*$ and $\Psi = V_2 \cdot V_2^* + (1 - \tau_a) V_3 \cdot V_3^*$. The V_k are linearly independent because $\gamma \neq \gamma'$, so by injectivity of $K \mapsto \Psi_K$ the derivative is $T = \text{diag}(1, 0, \tau_a)$ in this frame, and Lemma 7.2.1 gives the levels. Finally $S_a = V_3^* V_3$.

(b) The ranges are

$$\text{ran } J(\Phi) = \text{span}\{\text{vec } K_0, \text{vec } e_0 e_1^*\}, \quad \text{ran } J(\Psi) = \text{span}\{\text{vec } L_0, \text{vec } e_0 e_1^*\},$$

and $\text{vec } K_0, \text{vec } L_0, \text{vec } e_0 e_1^*$ are linearly independent. By (a), $B_s = g_s(\tau_a)(\gamma + \gamma')e_1 e_1^*$ is singular for every s , so $\hat{C} = +\infty$ and $\hat{Q}_{1/2}(\Phi\|\Psi) = \lambda_{\min}(B_{1/2}) = 0$ (Remark 7.2.2). For C_{bar} argue as in the proof of Theorem 7.6.4(d). If both $\hat{D}(\mathcal{T}\|\Phi)$ and $\hat{D}(\mathcal{T}\|\Psi)$ are finite, then $\text{ran } J(\mathcal{T}) \subseteq \text{span}\{\text{vec } e_0 e_1^*\}$, so $\mathcal{T} = c e_0 e_1^* \cdot e_1 e_0^*$. Then $\mathcal{T}^*(\mathbf{1}) = c e_1 e_1^* \neq \mathbf{1}$, and $\mathcal{T}$ is not a channel. Hence for $0 < s < 1$ the barycentric objective is $+\infty$ for every channel $\mathcal{T}$.

(c) $S_a e_0 = 0$, and $K_0 e_0 = L_0 e_0 = e_0$, $K_1 e_0 = L_1 e_0 = 0$. $\square$

Example 4.12.8 in Chapter 4 proves $\hat{Q}_{1/2}(\mathcal{A}_\gamma\|\mathcal{A}_{\gamma'}) = 0$, there written $\hat{F}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) = 0$, for all $\gamma \neq \gamma'$ in $[0, 1]$ by a different argument, directly from the intersection of the Choi ranges in (b).

Part (c) explains why $\hat{C}$ is infinite. The input e_0 , at which B_s is singular, is one on which the two channels act identically. By Proposition 7.4.4, $\hat{C}$ is the exponent of the flagged pair, and on e_0 the flag reveals which of the conclusive levels $K_0 \cdot K_0^*$ and $L_0 \cdot L_0^*$ occurred. These two maps agree on $e_0 e_0^*$, so all the information comes from the flag, which the pair (Φ, Ψ) itself does not provide. In contrast, the adaptive Stein exponent of the pair itself is infinite: with the maximally entangled input the outputs $J(\Phi)/2$ and $J(\Psi)/2$ have different supports, so $D(\Phi\|\Psi) = D(\Psi\|\Phi) = +\infty$.

To bound ξ_{adapt} from above we have to leave the class of common decompositions, since by Proposition 7.4.3 every such decomposition, and in particular every classical program, gives an exponent at least $\hat{C} = +\infty$. We use instead programs that are pure quantum states. For channels $\Phi, \Psi \in \mathcal{C}(\mathcal{X}, \mathcal{Y})$ with Kraus operators $(K_k)_{k \leq r}$ and $(L_l)_{l \leq r}$, padded with zeros to a common length r , put

$$\mu_*(\Phi, \Psi) = \max \left\{ |c| : \sum_{k,l} C_{lk} L_l^* K_k = c \mathbf{1}_{\mathcal{X}} \text{ for some } C \in \mathcal{L}(\mathbb{C}^r), \|C\|_\infty \leq 1 \right\}.$$

This is a semidefinite program: maximise $\text{Re } c$ subject to the linear constraint and $\begin{pmatrix} \mathbf{1} & C \\ C^* & \mathbf{1} \end{pmatrix} \geq 0$. It does not depend on the Kraus families, because two families of the same channel are related by a partial isometry, which can be absorbed into C .

Theorem 7.6.9 (Pure-program converse). *Let Φ, Ψ be channels.*

(a) $\mu_*(\Phi, \Psi)$ is the largest overlap $|\langle \psi_\Psi, \psi_\Phi \rangle|$ of pure states for which there is a channel Λ with $\Phi(\rho) = \Lambda(\rho \otimes \psi_\Phi \psi_\Phi^*)$ and $\Psi(\rho) = \Lambda(\rho \otimes \psi_\Psi \psi_\Psi^*)$ for all ρ .

(b) For every n and every adaptive strategy with n uses,

$$\frac{1}{2} - \frac{1}{4} \|\rho_\Phi^{(n)} - \rho_\Psi^{(n)}\|_1 \geq \frac{1}{2} \left(1 - \sqrt{1 - \mu_*(\Phi, \Psi)^{2n}} \right) \geq \frac{1}{4} \mu_*(\Phi, \Psi)^{2n},$$

so that $\xi_{\text{adapt}}(\Phi, \Psi) \leq -2 \ln \mu_*(\Phi, \Psi)$.

(c) If $\Phi(\rho) = \Lambda(\rho \otimes \pi_\Phi)$ and $\Psi(\rho) = \Lambda(\rho \otimes \pi_\Psi)$ for some channel Λ and mixed program states π_Φ, π_Ψ , then $\xi_{\text{QC}}(\pi_\Phi, \pi_\Psi) \geq -\ln \mu_*(\Phi, \Psi)$. Thus no program simulation gives a converse below $-\ln \mu_*$.

Proof. (a) Let $V_\Phi = \sum_k K_k \otimes e_k$ and $V_\Psi = \sum_l L_l \otimes e_l$ be the Stinespring isometries into $\mathcal{Y} \otimes \mathbb{C}^r$, so that $V_\Psi^*(\mathbf{1} \otimes C)V_\Phi = \sum_{k,l} C_{lk} L_l^* K_k$. Let C and c be feasible with $|c| < 1$; if $|c| = 1$, then $\Phi = \Psi$ and there is nothing to prove. The unitary

$$U = \begin{pmatrix} C & (\mathbf{1} - CC^*)^{1/2} \\ (\mathbf{1} - C^*C)^{1/2} & -C^* \end{pmatrix}$$

on $\mathbb{C}^r \otimes \mathbb{C}^2$ has $(\mathbf{1} \otimes e_0^*)U(\mathbf{1} \otimes e_0) = C$. Hence $V'_\Phi = (\mathbf{1} \otimes U)(V_\Phi \otimes e_0)$ and $V'_\Psi = V_\Psi \otimes e_0$ are Stinespring isometries of Φ and Ψ with $V'_\Psi V'^*_\Phi = c\mathbf{1}$. Choose unit vectors $\psi_\Phi, \psi_\Psi \in \mathbb{C}^2$ with $\langle \psi_\Psi, \psi_\Phi \rangle = c$. The prescription $W(x \otimes \psi_\Phi) = V'_\Phi x$, $W(x \otimes \psi_\Psi) = V'_\Psi x$ preserves all inner products of the spanning set $\{x \otimes \psi_\Phi, x \otimes \psi_\Psi\}$ of $\mathcal{X} \otimes \mathbb{C}^2$, so it extends to an isometry, and $\Lambda = \text{Tr}_{\mathbb{C}^r \otimes \mathbb{C}^2} W \cdot W^*$ simulates the pair. Conversely, let Λ simulate the pair with pure programs, and let W be a Stinespring isometry of Λ . Then $W(\cdot \otimes \psi_\Phi)$ and $W(\cdot \otimes \psi_\Psi)$ are Stinespring isometries of Φ and Ψ , and their product is $\langle \psi_\Psi, \psi_\Phi \rangle \mathbf{1}$. Any Stinespring isometry of Φ has the form $(\mathbf{1} \otimes A)V_\Phi$ with a partial isometry A , and similarly for Ψ with some B . Hence the product equals $V_\Psi^*(\mathbf{1} \otimes B^*A)V_\Phi$ with $\|B^*A\|_\infty \leq 1$.

(b) Replace every use of the unknown channel by $\Lambda(\cdot \otimes \psi_i \psi_i^*)$, each time with a fresh copy of the program. The final states of the strategy are then $\mathcal{S}((\psi_i \psi_i^*)^{\otimes n})$ for one channel $\mathcal{S}$ that does not depend on i , as in the stretching argument of [107]. By monotonicity of the trace norm, the error is at least that of discriminating the pure states $\psi_\Phi^{\otimes n}$ and $\psi_\Psi^{\otimes n}$, whose overlap has modulus μ_*^n . This error is $\frac{1}{2}(1 - \sqrt{1 - \mu_*^{2n}})$, and $\sqrt{1 - x} \leq 1 - \frac{x}{2}$.

(c) By Uhlmann's theorem [122] there are purifications ψ_Φ, ψ_Ψ of π_Φ, π_Ψ with $|\langle \psi_\Psi, \psi_\Phi \rangle|^2 = F(\pi_\Phi, \pi_\Psi)$. The channel $\Lambda \circ (\text{id} \otimes \text{Tr}_{\text{purif}})$ simulates the pair with these pure programs, so $F(\pi_\Phi, \pi_\Psi) \leq \mu_*^2$ by (a). Finally $\min_s \text{tr}(\pi_\Phi^s \pi_\Psi^{1-s}) \leq \text{tr}(\pi_\Phi^{1/2} \pi_\Psi^{1/2}) \leq \|\pi_\Phi^{1/2} \pi_\Psi^{1/2}\|_1 = F(\pi_\Phi, \pi_\Psi)^{1/2}$. $\square$

For replacement channels $\mu_* = F(\phi, \psi)^{1/2}$. The inequality $\leq$ holds because, by (a) and monotonicity of the fidelity, $F(\Phi(\rho), \Psi(\rho)) \geq \mu_*^2$ for every ρ . For $\geq$, take $C = X \otimes \mathbf{1}$ in the Kraus operators $\sqrt{\lambda_j} \phi_j e_k^*$ and $\sqrt{\kappa_i} \psi_i e_k^*$ and maximise over contractions X . The pure-program converse $-\ln F(\phi, \psi)$ is then within a factor of 2 of the true exponent $\xi_{\text{QC}}(\phi, \psi)$. It is finite on the pairs of Theorem 7.6.5 and tends to $-\ln F(\phi, \psi)$ along them. So $\min\{\widehat{C}, C_{\text{bar}}, -2\ln \mu_*\}$ is a single-letter converse that does not jump at those pairs, which bears on Question 7.8.2. Unlike the other converses, μ_* is defined through Stinespring isometries rather than through Choi matrices. It need not be close to the true exponent in general, and it can vanish: $\mu_* = 0$ for two distinct unitary channels, for which indeed $\xi_{\text{adapt}} = +\infty$.

Corollary 7.6.10 (Fragility for amplitude damping). *For $0 < \gamma, \gamma' \leq 1$, $\mu_*(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) = \mu_*(\gamma, \gamma')$, and*

$$C((\gamma, 1 - \gamma) \| (\gamma', 1 - \gamma')) \leq \xi_{\text{prod}} \leq \xi_{\text{adapt}}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) \leq 2 \ln \frac{1 - \sqrt{(1 - \gamma)(1 - \gamma')}}{\sqrt{\gamma\gamma'}},$$

where C is the classical Chernoff information. Moreover $p_{\text{adapt}}(n) \geq \frac{1}{2}(1 - \sqrt{1 - \mu_*(\gamma, \gamma')^{2n}})$ for every n . For fixed $0 < \gamma < 1$ and $\gamma' = \gamma - \Delta$,

$$\frac{\Delta^2}{8\gamma(1 - \gamma)} + O(\Delta^3) \leq \xi_{\text{adapt}}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) \leq \frac{\Delta^2}{4\gamma^2(1 - \gamma)} + O(\Delta^3).$$

Consequently, as $\gamma' \rightarrow \gamma$ the pair $(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'})$ converges in the diamond norm to a pair of identical channels and $\xi_{\text{adapt}} \rightarrow 0$, while $\widehat{C}$, $-2\ln \widehat{Q}_{1/2}$ and C_{bar} are $+\infty$ for every $\gamma' \neq \gamma$. The conclusion of Theorem 7.6.5 therefore holds for amplitude damping, along a physical one-parameter family. No Halmos dilation is needed, and the Choi ranges need not intersect trivially.

Proof. With $a = \sqrt{(1 - \gamma)(1 - \gamma')}$ we have $L_0^* K_0 = \text{diag}(1, a)$, $L_1^* K_1 = \sqrt{\gamma\gamma'} e_1 e_1^*$, $L_0^* K_1 = \sqrt{\gamma} e_0 e_1^*$ and $L_1^* K_0 = \sqrt{\gamma'} e_1 e_0^*$. Hence

$$\sum_{k,l} C_{lk} L_l^* K_k = \begin{pmatrix} C_{00} & \sqrt{\gamma} C_{01} \\ \sqrt{\gamma'} C_{10} & a C_{00} + \sqrt{\gamma\gamma'} C_{11} \end{pmatrix}.$$

This is $c1$ if and only if $C_{01} = C_{10} = 0$, $C_{00} = c$ and $C_{11} = c(1-a)/\sqrt{\gamma\gamma'}$. Since $\sqrt{\gamma\gamma'} \leq 1-a$, the contraction constraint is $|c|(1-a) \leq \sqrt{\gamma\gamma'}$, which gives $\mu_*(\gamma, \gamma')$. The upper bounds follow from Theorem 7.6.9(b). The lower bound is the product strategy with input e_1 , whose outputs $\text{diag}(\gamma, 1-\gamma)$ and $\text{diag}(\gamma', 1-\gamma')$ commute. The expansions are elementary: $\sqrt{(1-\gamma)(1-\gamma+\Delta)} = 1-\gamma + \frac{\Delta}{2} - \frac{\Delta^2}{8(1-\gamma)} + O(\Delta^3)$ gives $-\ln \mu_* = \frac{\Delta^2}{8\gamma^2} + \frac{\Delta^2}{8\gamma(1-\gamma)} + O(\Delta^3)$, and the Chernoff information of two Bernoulli laws is $\frac{\Delta^2}{8\gamma(1-\gamma)} + O(\Delta^3)$. The diamond distance is at most $2 \sum_k \|K_k - L_k\|_\infty \rightarrow 0$. Proposition 7.6.8 gives the infinite converses. $\square$

Remark 7.6.11. The upper and lower bounds of Corollary 7.6.10 differ by the factor $2/\gamma$ to leading order, and simulation arguments cannot close this gap. Classical programs give an exponent at least $\hat{C} = +\infty$ by Proposition 7.4.3. By Theorem 7.6.9(c), every quantum program gives at least $-\ln \mu_* = \frac{\Delta^2}{8\gamma^2(1-\gamma)} + O(\Delta^3)$, which exceeds ξ_{prod} by the factor $1/\gamma$ to leading order.

Remark 7.6.12 (Local regime). For estimating γ itself, adaptivity does not help. In the Kraus representation above, $\beta = \sum_k \dot{K}_k^* K_k = \text{diag}(0, -\frac{1}{2}) + \text{diag}(0, \frac{1}{2}) = 0$ and $\alpha = \sum_k \dot{K}_k^* \dot{K}_k = \frac{1}{4\gamma(1-\gamma)} e_1 e_1^*$. The bound of Demkowicz-Dobrzański and Maccone for adaptive strategies [30] therefore gives quantum Fisher information at most $4n\|\alpha\|_\infty = n/(\gamma(1-\gamma))$ after n adaptive uses. This equals the Fisher information of the product input $e_1^{\otimes n}$. It suggests, but does not prove, that $\xi_{\text{adapt}} = \frac{\Delta^2}{8\gamma(1-\gamma)} + o(\Delta^2)$, which is the lower bound of Corollary 7.6.10 (compare Question 7.8.3).

Numerical computations suggest a sharper bound. By Uhlmann's theorem and the minimax theorem, the smallest root fidelity of the final states over all adaptive n -use strategies equals

$$\max_{\|C\|_\infty \leq 1} \min_H \text{Re tr} \left(H^\top \sum_{\vec{k}, \vec{l}} C_{\vec{l}\vec{k}} \text{vec}(K_{\vec{k}}) \text{vec}(L_{\vec{l}})^* \right),$$

where H runs over deterministic n -combs and $K_{\vec{k}} = K_{k_1} \otimes \cdots \otimes K_{k_n}$. After dualising the inner minimisation, this is a semidefinite program.

Conjecture 7.6.13. *For every adaptive strategy with n uses, $F(\rho_{\mathcal{A}_\gamma}^{(n)}, \rho_{\mathcal{A}_{\gamma'}}^{(n)}) \geq f(\gamma, \gamma')^{2n}$, with equality for the product input $e_1^{\otimes n}$. Consequently $p_{\text{adapt}}(n) \geq \frac{1}{2}(1 - \sqrt{1 - f^{2n}})$ and*

$$\xi_{\text{adapt}}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) \leq -2 \ln f(\gamma, \gamma') = \frac{\Delta^2}{4\gamma(1-\gamma)} + O(\Delta^3) \leq 2 \xi_{\text{prod}}(1 + O(\Delta)).$$

We solved this program for $\gamma = \frac{1}{2}$, $n \leq 4$ and $\gamma' \in \{0.05, 0.2, 0.4, 0.47, 0.49\}$. In each case the optimal value agreed with f^n to within 10^{-8} , with optimal $C = \mathbf{1}$. Thus, up to the accuracy of the solver, Conjecture 7.6.13 holds for these values of n and γ' . The conjecture cannot be proved by amortisation, because the inequality

$$\sqrt{F}((\mathcal{A}_\gamma \otimes \text{id})\rho, (\mathcal{A}_{\gamma'} \otimes \text{id})\sigma) \geq f(\gamma, \gamma') \sqrt{F}(\rho, \sigma)$$

fails for some pairs of *different* states ρ, σ of a qubit and a two-dimensional reference. A numerical search found a pair with $\sqrt{F}((\mathcal{A}_{1/2} \otimes \text{id})\rho, (\mathcal{A}_{1/5} \otimes \text{id})\sigma) = 0.948315 \sqrt{F}(\rho, \sigma) < f(\frac{1}{2}, \frac{1}{5}) \sqrt{F}(\rho, \sigma) = 0.948683 \sqrt{F}(\rho, \sigma)$, and we verified this inequality in 40-digit arithmetic. A proof must therefore use that both branches of a strategy start from the same state. In the dilation $\mathcal{A}_\gamma(\rho) = \text{Tr}_E U_\theta(\rho \otimes e_0 e_0^*) U_\theta^*$ with $\cos \theta = \sqrt{1-\gamma}$ and the excitation-preserving rotation U_θ , the two branches differ by the rotations $U_{\theta'-\theta}$ applied to fresh environments,

γ'	ξ_{prod}	Choi product	$-2 \ln f$ (Conj. 7.6.13)	$-2 \ln \mu_*$ (Cor. 7.6.10)	$\widehat{C}, -2 \ln \widehat{Q}_{1/2}, C_{\text{bar}}$
0.40	$5.077 \cdot 10^{-3}$	$3.272 \cdot 10^{-3}$	$1.015 \cdot 10^{-2}$	$2.252 \cdot 10^{-2}$	∞
0.47	$4.506 \cdot 10^{-4}$	$2.974 \cdot 10^{-4}$	$9.012 \cdot 10^{-4}$	$1.858 \cdot 10^{-3}$	∞
0.49	$5.001 \cdot 10^{-5}$	$3.323 \cdot 10^{-5}$	$1.000 \cdot 10^{-4}$	$2.021 \cdot 10^{-4}$	∞

Table 7.1: Exponents for $\mathcal{A}_{1/2}$ versus $\mathcal{A}_{\gamma'}$. ξ_{prod} is attained at the input e_1 , and a local optimisation over all inputs with a reference system found no larger value. The ratios $-2 \ln \mu_*/\xi_{\text{prod}}$ are 4.44, 4.12, 4.04 (limit $2/\gamma = 4$) and $-2 \ln f/\xi_{\text{prod}} = 2.00$.

γ'	n	adaptive (SDP)	parallel (SDP)	$e_1^{\otimes n}$	$\frac{1}{2}(1 - \sqrt{1 - f^{2n}})$ (Conj. 7.6.13)	$\frac{1}{2}(1 - \sqrt{1 - \mu_*^{2n}})$ (Cor. 7.6.10)
0.40	1	0.4500	0.4500	0.4500	0.4497	0.4254
	2	0.4409	0.4415	0.4450	0.4291	0.3951
	3	0.4247	0.4260	0.4260	0.4134	0.3722
0.47	1	0.4850	0.4850	0.4850	0.4850	0.4785
	2	0.4826	0.4827	0.4845	0.4788	0.4696
	3	0.4771	0.4775	0.4775	0.4740	0.4627
0.49	1	0.4950	0.4950	0.4950	0.4950	0.4929
	2	0.4942	0.4942	0.4950	0.4929	0.4900
	3	0.4924	0.4925	0.4925	0.4913	0.4877
0.20	1	0.3500	0.3500	0.3500	0.3419	0.2452
	2	0.3026	0.3050	0.3050	0.2821	0.1638
	3	0.2815	0.2854	0.3020	0.2397	0.1145

Table 7.2: Optimal n -use error for $\mathcal{A}_{1/2}$ versus $\mathcal{A}_{\gamma'}$, computed by semidefinite programming over testers [27, 49]: adaptive, best parallel (entangled, non-adaptive) and the product input $e_1^{\otimes n}$. The last two columns are the lower bounds given by Conjecture 7.6.13 (verified numerically for $n \leq 4$) and by Corollary 7.6.10. The bounds of Theorem 7.4.1 and Corollary 7.3.2 are 0 for every n , because the flagged pair is perfectly distinguishable after one use with the input e_0 .

and $f = \cos(\theta' - \theta)$. The conjecture asserts that no adaptive strategy can make these rotations add up coherently.

Table 7.1 lists the exponents for $\gamma = \frac{1}{2}$, and Table 7.2 compares the finite- n bounds with the optimal errors for $n \leq 3$. Table 7.2 also shows that adaptivity helps at finite n . For $\gamma' = 0.2$ and $n = 3$ the optimal adaptive error is 0.2815, against 0.2854 for the best parallel strategy and 0.3020 for $e_1^{\otimes 3}$. Whether adaptivity also improves the exponent is open. Three uses are too few to decide this, but if Conjecture 7.6.13 holds, any gain is at most a factor of 2.

7.7 Numerical illustration

All computations in this section were done with the author's own code, which is not part of the `code/` directory described in Section 2.6. Exact optimal adaptive errors for $n = 1, 2$ were computed by semidefinite programming over two-round testers [27, 49]. The barycentric bound C_{bar} was computed with a semidefinite approximation of the operator relative entropy [44], after which the exact objective was re-evaluated at the computed channel; for the measurement pair the problem reduces to a one-dimensional convex minimisation, which we solved directly.

Bounds on the exponent Table 7.3 compares the achievable exponent ξ_{prod} (best product input found by local optimisation, so a lower bound on ξ_{adapt}), the barycentric converse C_{bar} , the new converse $\hat{C}$, and the fidelity converse $-2 \ln \hat{Q}_{1/2}$. In every case $\hat{C}$ is within 4% of half the fidelity converse. The first five pairs satisfy the hypotheses of Theorem 7.5.1, so there $\xi_{\text{prod}} = \xi_{\text{adapt}} = \hat{C}$; also $C_{\text{bar}} = \hat{C}$, for the first four rows by Proposition 7.4.6(a) and Theorem 7.5.3, and for the classical row by [66]. The measurement pair (Example 7.7.1) is its own flagged pair, so $\xi_{\text{adapt}} = \hat{C} < C_{\text{bar}}$. For the replacement pair (Example 7.7.2) we have $C_{\text{bar}} < \hat{C}$, while the true exponent is the quantum Chernoff divergence of the two states. In the remaining, generic cases C_{bar} is between 2% and 18% below $\hat{C}$.

Example 7.7.1. Let $e_\ell = (\cos \frac{\theta_\ell}{2}, \sin \frac{\theta_\ell}{2})$ with $\theta = (0, \frac{\pi}{4}, \frac{7\pi}{8}, \frac{9\pi}{8})$. The weights $w \geq 0$ with $\sum_\ell w_\ell e_\ell e_\ell^* = \mathbf{1}$ form a segment; let a and b be its points at 10% and 90% of its length, $a \approx (0.9164, 0.0520, 0.4678, 0.5638)$ and $b \approx (0.5638, 0.4678, 0.0520, 0.9164)$, and let Φ, Ψ be the measurement channels with POVM elements $a_\ell e_\ell e_\ell^*$ and $b_\ell e_\ell e_\ell^*$.

Example 7.7.2. The replacement channels $\Phi = \text{tr}(\cdot) \phi$, $\Psi = \text{tr}(\cdot) \psi$ with $\phi = \frac{1}{2}(\mathbf{1} + 0.9 Z)$ and $\psi = \frac{1}{2}(\mathbf{1} + 0.9 X)$.

Exact optimality at finite n Theorem 7.5.3 predicts that the exact adaptive optimum, the error of the Choi product strategy and the dynamic-programming value of Corollary 7.3.2 coincide. Table 7.4 confirms this for a depolarising pair, an $\text{SU}(2)$ -covariant spin-1 pair (two levels), and a random-unitary qutrit pair built from $\{\mathbf{1}, Z, XD_1, X^2 D_2\}$ with random diagonal D_1, D_2 (four levels). The dynamic program and the Choi strategy agree to ten digits also for $n = 3$.

The finite- n bound in general For pairs outside the scope of Theorem 7.5.3, Table 7.5 compares the dynamic-programming bound with the exact optimum (where computable), with the explicit bound of Theorem 7.4.1, and with the previous finite- n bound $\frac{1}{2}(1 - \sqrt{1 - \hat{Q}_{1/2}^{2n}})$, which follows from the Fuchs–van de Graaf inequality and the chain rule for the geometric fidelity. The second pair has Kraus ranks 2 and 3, hence infinite Thompson

pair	ξ_{prod}	C_{bar}	$\widehat{C}$ (new)	$-2 \ln \widehat{Q}_{1/2}$	exact
depolarising $p = 0.1$ vs 0.3	0.02383	0.02383	0.02383	0.04738	Thm. 7.5.3
Pauli $(.7, .1, .1, .1)$ vs $(.6, .2, .1, .1)$	0.01058	0.01058	0.01058	0.02112	Thm. 7.5.3
SU(2), spin 1, $(.8, .15, .05)$ vs $(.6, .3, .1)$	0.02467	0.02467	0.02467	0.04928	Thm. 7.5.3
SU(2), spin 1, $(.5, .3, .2)$ vs $(.2, .5, .3)$	0.05297	0.05297	0.05297	0.10580	Thm. 7.5.3
classical 2×2	0.03380	0.03380	0.03380	0.06726	Thm. 7.5.1
measurement pair, Ex. 7.7.1	0.19705	0.24697	0.19705	0.39410	Prop. 7.4.4
replacement pair, Ex. 7.7.2	0.33136	0.54272	0.57077	1.14154	$\xi = \xi_{\text{QC}}(\phi, \psi)$
classical–quantum, non-commuting	0.08410	0.13035	0.13280	0.26538	–
GAD $\gamma = 0.2$ vs 0.4 ($N = 0.1$)	0.02126	0.11388	0.13913	0.27482	–
GAD $N = 0.1$ vs 0.3 ($\gamma = 0.3$)	0.00862	0.03247	0.03380	0.06726	–
random full-rank #1	0.32226	0.55379	0.60249	1.19341	–
random full-rank #2	0.19256	0.34726	0.37964	0.73026	–

Table 7.3: Bounds on the adaptive Chernoff exponent: $\xi_{\text{prod}} \leq \xi_{\text{adapt}} \leq \min\{C_{\text{bar}}, \widehat{C}\}$ and $\widehat{C} \leq -2 \ln \widehat{Q}_{1/2}$. The column “exact” names the result that identifies ξ_{adapt} .

pair, n	exact adaptive (SDP)	Choi product	dynamic program
depolarising, $n = 1$ / $n = 2$	0.4250 / 0.3725	0.4250 / 0.3725	0.4250 / 0.3725
SU(2) spin 1, $n = 1$ / $n = 2$	0.4000 / 0.3600	0.4000 / 0.3600	0.4000 / 0.3600
random-unitary qutrit, $n = 1$ / $n = 2$	0.4000 / 0.3650	0.4000 / 0.3650	0.4000 / 0.3650

Table 7.4: Exact optimality of the Choi strategy for level-balanced pairs with orthogonal levels.

	$n = 1$	$n = 2$	$n = 5$	$n = 10$	$n = 20$
<i>random full-rank qubit pair</i> ($\widehat{C} = 0.341$, $\Lambda = 2.51$)					
exact adaptive (SDP)	0.2158	0.1496	–	–	–
dynamic program (Cor. 7.3.2)	0.1758	0.1084	$2.84 \cdot 10^{-2}$	$3.92 \cdot 10^{-3}$	$9.66 \cdot 10^{-5}$
Theorem 7.4.1	$2.74 \cdot 10^{-3}$	$3.46 \cdot 10^{-4}$	$4.04 \cdot 10^{-6}$	$1.54 \cdot 10^{-8}$	$2.15 \cdot 10^{-12}$
fidelity bound	0.1484	0.0686	$8.32 \cdot 10^{-3}$	$2.72 \cdot 10^{-4}$	$2.96 \cdot 10^{-7}$
<i>random Kraus-rank 2 vs 3 qubit pair</i> ($d_T = \infty$, $\widehat{C} = 2.40$, $\Lambda = 0.90$)					
exact adaptive (SDP)	0.1213	0.0422	–	–	–
dynamic program (Cor. 7.3.2)	$3.94 \cdot 10^{-2}$	$3.11 \cdot 10^{-3}$	$1.52 \cdot 10^{-6}$	$4.63 \cdot 10^{-12}$	$4.29 \cdot 10^{-23}$
Theorem 7.4.1	$5.03 \cdot 10^{-3}$	$2.45 \cdot 10^{-4}$	$5.28 \cdot 10^{-8}$	$8.00 \cdot 10^{-14}$	$4.15 \cdot 10^{-25}$
fidelity bound	$3.85 \cdot 10^{-3}$	$5.88 \cdot 10^{-5}$	$2.12 \cdot 10^{-10}$	$1.79 \cdot 10^{-19}$	$1.29 \cdot 10^{-37}$

Table 7.5: Lower bounds on the optimal adaptive error.

distance, two conclusive levels and $\widehat{C} = 2.40$; Theorem 7.4.1 still applies with $\Lambda = 0.90$. The explicit bound has the better exponent but a large second-order term when Λ is large, and the dynamic program is the practical finite- n bound. For random full-rank pair #1 of Table 7.3, the optimal two-use adaptive error (0.1539) is strictly smaller than the optimal two-use parallel error (0.1605), so adaptivity helps at finite n .

Local regime For a random full-rank qubit channel Φ_0 , a second one Φ_1 , and $\Phi_\delta = (1 - \delta)\Phi_0 + \delta\Phi_1$, the ratio $\widehat{C}/(-2 \ln \widehat{Q}_{1/2})$ was 0.5019, 0.5009, 0.5003, 0.5001, 0.5000 for $\delta = 0.2, 0.1, 0.05, 0.02, 0.01$, and $8\widehat{C}/(\delta^2 F_{\text{RLD}})$ was 0.64, 0.76, 0.85, 0.93, 0.96, consistent with Corollary 7.4.5(b) and an $O(\delta)$ relative correction. The ratio $C_{\text{bar}}/\widehat{C}$ was 0.963, 0.976, 0.993, 0.9997 for $\delta = 0.2, 0.1, 0.05, 0.02$, so in this example the two converses have the

same leading term.

Continuity For 60 random pairs of full-Kraus-rank qubit channels and perturbations $(1 - \eta)(\Phi, \Psi) + \eta(\Phi', \Psi')$ with $10^{-3} \leq \eta \leq 10^{-1}$, the change of $e^{-\hat{C}}$ was at most 0.18 times the bound of Theorem 7.6.2, and the same held for 59 pairs of splittings of one sum. For the same kind of pair, $\hat{C}$ along the perturbations with $\eta = 10^{-1}, \dots, 10^{-4}$ was 0.230, 0.466, 0.526, 0.534 against 0.535 at the pair, in line with Theorem 7.6.4(c). The values in Example 7.6.6 were computed from the level decomposition and the Petz quasi-entropies directly.

7.8 Concluding remarks

We have shown that the level decomposition of a pair of channels turns every adaptive discrimination protocol into a controlled random walk. This yields the converse $\hat{C}$ with an explicit second-order term (Theorem 7.4.1), and the converse is attained under the orthogonality condition of Theorem 7.5.1. Single-letter converses of this type can, however, be infinite at pairs whose adaptive exponent is small (Theorem 7.6.5 and Corollary 7.6.10). We conclude with some open problems.

Question 7.8.1. Is $\xi_{\text{adapt}} = \xi_{\text{prod}}$ whenever $d_T(\Phi, \Psi) < \infty$? This holds for classical channels, for classical–quantum channels [116] and in the cases of Theorems 7.5.1 and 7.5.3. The known separations [53, 116] have $d_T = \infty$.

Question 7.8.2. Find a single-letter converse that is at most $\min\{C_{\text{bar}}, \hat{C}\}$ for all pairs. For replacement channels both exceed the true exponent $\xi_{\text{QC}}(\phi, \psi)$, so such a converse should see the non-commutativity of the outputs, which every common decomposition hides (Proposition 7.4.3). By Theorem 7.6.5, a converse that is within a bounded factor of ξ_{adapt} must be finite on pairs whose Choi ranges intersect trivially, and so cannot be built only from Choi-level geometric means or from divergences that require support inclusion.

Question 7.8.3. For nearby channels, is the adaptive Chernoff exponent $\frac{\delta^2}{8}F(\theta) + o(\delta^2)$ for an explicitly computable Fisher information F between the one achieved by product inputs and F_{RLD} ? In this regime $T = \frac{1}{2}\mathbf{1} + O(\delta)$. By Lemma 7.2.4 the product input at σ attains $-\ln \min_s Q_s(\sigma)$ whenever $[T, G_\sigma] = 0$, so one expects the gap between ξ_{prod} and $\hat{C}$ to be controlled by this commutator.

Question 7.8.4. Prove a Hoeffding converse for general pairs. The walk structure applies, but $s \mapsto \min_\rho \ln Q_s(\rho)$ need not be convex, so the natural candidate involves its convex envelope.

Question 7.8.5. Does Theorem 7.6.5 hold near every pair of channels with $\dim \ker J(\Phi + \Psi) \geq 1$ and $\hat{C} < \infty$, that is, are there always nearby channel pairs with trivially intersecting Choi ranges and bounded adaptive exponent? Theorem 7.6.4(c) excludes such jumps when $J(\Phi + \Psi)$ is invertible, and Theorem 4.12.4(c) provides completely positive approximants. For the geometric fidelity $\hat{Q}_{1/2}$, which Chapter 4 writes $\hat{F}$, that chapter gives channel-valued approximants in two cases: at a diagonal pair (Φ, Φ) with $J(\Phi)$ singular, output rotations of Φ make the fidelity jump (Theorem 4.12.9 and Corollary 4.12.10), and within the amplitude-damping family it vanishes for $\gamma' \neq \gamma$ (Example 4.12.8; compare Proposition 7.6.8(b)), although these Choi ranges intersect nontrivially. The question is restated, together with the related questions of Chapter 4, in Section 9.2.

Question 7.8.6. Extend the level decomposition to normal channels between von Neumann algebras and to bosonic channels with energy constraints. Adaptive quantum illumination is listed as open in [97].

Chapter 8

Noisy gates: proofs and further results

Summary. This chapter contains the proofs of the results stated in Chapter 1 and the further results behind them: closed forms for Pauli channels, including biased noise and correlated miscalibrations, the amplitude-damping analysis, the local and finite-miscalibration comparisons of the converses with the adaptive Stein and Chernoff exponents, the certification bounds, the numerical methods, and the refinements of results of Chapters 4, 5 and 7 that these proofs need.

This chapter is the companion of Chapter 1, which states the results and their physical meaning; here we prove them. It contains the proofs of the statements made in Sections 1.2–1.7 of Chapter 1, closed-form expressions for Pauli channels, the amplitude-damping analysis, the certification bounds and the numerical methods. It keeps the notation of Chapter 1, which Section 8.1 recalls, and the theorems, equations, figures and tables of that chapter are cited by their numbers there, which begin with 1. Results proved in Chapters 4, 5 and 7 are used by reference; Section 8.15 collects the refinements of those results that are needed here, together with pointers to the rest. Statements labelled as conjectures are not claimed to be proved.

8.1 Notation and conventions

Channels Φ, Ψ map operators on $\mathcal{X}$ ($\dim \mathcal{X} = d_X$) to operators on $\mathcal{Y}$ ($\dim \mathcal{Y} = d_Y$). The unnormalized Choi matrix is $J(\Phi) = \sum_{ij} \Phi(|i\rangle\langle j|) \otimes |i\rangle\langle j|$, with output first, so that $\text{Tr } J(\Phi) = d_X$ and $\text{Tr}_{\mathcal{Y}} J(\Phi) = \mathbf{1}$ for a channel. The Choi floor is $\lambda(\Phi) = \lambda_{\min}(J(\Phi))$, and $\tilde{\lambda} = \lambda/d_X$ is the floor of the Choi state J/d_X . The diamond norm $\|\cdot\|_{\diamond}$ is unnormalized ($\|\Phi - \Psi\|_{\diamond} \leq 2$ for channels). We write $\text{vec}(A) = \sum_{ij} A_{ij} |i\rangle |j\rangle$, so that $J(\text{Ad}_A) = \text{vec}(A) \text{vec}(A)^{\dagger}$ with $\text{Ad}_A(\rho) = A\rho A^{\dagger}$.

For positive definite A, B and $s \in [0, 1]$, $A \#_s B = A^{1/2} (A^{-1/2} B A^{-1/2})^s A^{1/2}$ is the weighted Kubo–Ando geometric mean [79], extended to positive semidefinite arguments by continuity from above, and $\# = \#_{1/2}$. The channel quantities are

$$\begin{aligned} \hat{F}(\Phi, \Psi) &= \lambda_{\min} \text{Tr}_{\mathcal{Y}} [J(\Phi) \# J(\Psi)], & \hat{Q}_s(\Phi \| \Psi) &= \lambda_{\min} \text{Tr}_{\mathcal{Y}} [J(\Psi) \#_s J(\Phi)], \\ \hat{C}(\Phi, \Psi) &= -\ln \inf_{s \in (0,1)} \hat{Q}_s, \end{aligned} \quad (8.1.1)$$

$$D_{\text{BS}}(\Phi \| \Psi) = \lambda_{\max} \text{Tr}_{\mathcal{Y}} [J_{\Phi}^{1/2} \ln (J_{\Phi}^{1/2} J_{\Psi}^{-1} J_{\Phi}^{1/2}) J_{\Phi}^{1/2}], \quad (8.1.2)$$

with $J_\Phi = J(\Phi)$ [42, 72] (Chapters 4 and 7). The Umegaki channel divergence $D(\Phi\|\Psi)$ is the supremum over input states ψ_{RX} of $D((\text{id} \otimes \Phi)\psi\|(\text{id} \otimes \Psi)\psi)$, and the Uhlmann channel root fidelity $F_U(\Phi, \Psi)$ is the minimum over inputs of $\|\sqrt{\omega}\sqrt{\sigma}\|_1$ for the two output states. The Thompson metric is $d_T(\Psi, \Psi') = \inf\{t : e^{-t}J(\Psi) \leq J(\Psi') \leq e^t J(\Psi)\}$ [120]. Fisher informations are denoted F_{RLD} , F_{BKM} , F_{SLD} and F_∞ for the asymptotic adaptive rate; the Uhlmann root fidelity is F_U .

For n qubits, $\mathcal{D}_p(\rho) = (1-p)\rho + p \text{Tr}(\rho)\mathbf{1}/2$ acts on each qubit, and a Pauli channel is $\mathcal{P}_q(\rho) = \sum_Q q_Q Q \rho Q$ with Q ranging over the d^2 Pauli strings ($d = 2^n$). Local depolarizing noise is $\mathcal{D}_p^{\otimes n} = \mathcal{P}_q$ with $q_Q = (1 - \frac{3p}{4})^{n-\text{wt}Q} (\frac{p}{4})^{\text{wt}Q}$, and global depolarizing noise is $\mathcal{D}_p^{\text{glob}}(\rho) = (1-p)\rho + p \text{Tr}(\rho)\mathbf{1}/d$. The gate model is

$$\mathcal{N}_\delta = \mathcal{P}_q \circ \text{Ad}_{R_P(\delta)} \circ \text{Ad}_U, \quad R_P(\delta) = e^{-i\delta P/2}, \quad (8.1.3)$$

with $P \neq \mathbf{1}$ a Pauli string of weight w , and $\mathcal{N}_{p,\delta}$ denotes the case $\mathcal{P}_q = \mathcal{D}_p^{\otimes n}$. For the qubit examples we take $P = Z$; the choice $P = X$ gives identical values, because the two pairs differ by a common input and output unitary. For fixed P the Pauli strings split into $d^2/2$ unordered pairs $\{Q, QP\}$, and for each pair we put

$$x_Q(\delta) = \frac{(q_Q - q_{QP})^2 \sin^2(\delta/2)}{4q_Q q_{QP}}, \quad g(x) = 1 - (1+x)^{-1/2}, \quad \varphi(x) = \sqrt{\frac{x}{1+x}} \text{arsinh } \sqrt{x}. \quad (8.1.4)$$

8.2 Choi matrices and the diamond norm

Lemma 8.2.1 (cf. Lemma 4.8.7). *Let $\mathcal{M}$ be a Hermitian-preserving map. Then $\|J(\mathcal{M})\|_\infty \leq \frac{d_X}{2} \|\mathcal{M}\|_\diamond + \frac{1}{2} |\text{Tr } J(\mathcal{M})| \leq d_X \|\mathcal{M}\|_\diamond$. In particular $\|J(\Phi) - J(\Phi')\|_\infty \leq \frac{d_X}{2} \|\Phi - \Phi'\|_\diamond$ for channels. Both constants are attained in every dimension: by $\mathcal{M} = \text{id}$ for d_X , and by $\text{id} - \text{Ad}_V$ with $\text{Tr } V = 0$, or by $\text{id} - \mathcal{D}_1^{\text{glob}}$, for $d_X/2$.*

Proof. The inequalities and the cases $\mathcal{M} = \text{id}$ and $\text{id} - \text{Ad}_V$ are Lemma 4.8.7 in Chapter 4. For $\text{id} - \mathcal{D}_1^{\text{glob}}$, $J = |\Omega\rangle\langle\Omega| - \mathbf{1}/d$ has norm $d-1/d$, and writing $\mathcal{D}_1^{\text{glob}}$ as the uniform mixture of the d^2 Weyl unitaries gives $\|\text{id} - \mathcal{D}_1^{\text{glob}}\|_\diamond \leq 2(1-d^{-2})$, so equality holds. $\square$

Hence, if $\|\Psi' - \Psi\|_\diamond \leq \epsilon$ and $J(\Psi) \geq \lambda \mathbf{1}$, then $\|J(\Psi') - J(\Psi)\|_\infty \leq r\lambda$ with $r = d_X \epsilon / (2\lambda)$, and

$$(1-r)J(\Psi) \leq J(\Psi') \leq (1+r)J(\Psi), \quad d_T(\Psi, \Psi') \leq \ln \frac{1}{1-r} \quad (r < 1). \quad (8.2.1)$$

The same holds with a Choi-state trace-norm hypothesis $\|J(\Psi)/d_X - J(\Psi')/d_X\|_1 \leq \epsilon$, since $\|X\|_\infty \leq \frac{1}{2}\|X\|_1$ for traceless Hermitian X .

8.3 The Choi-floor law (Theorem 1.4.1)

Proof of Eq. (1.4.2). By Eq. (8.2.1) and joint monotonicity of $\#_s$ in each argument, $J(\Psi')\#_s J(\Phi)$ lies between $[(1 \mp r)J(\Psi)]\#_s J(\Phi) = (1 \mp r)^{1-s} J(\Psi)\#_s J(\Phi)$. The map $Z \mapsto \lambda_{\min} \text{Tr}_Y Z$ is monotone and positively homogeneous, so $(1-r)^{1-s} \hat{Q}_s(\Phi\|\Psi) \leq \hat{Q}_s(\Phi\|\Psi') \leq (1+r)^{1-s} \hat{Q}_s(\Phi\|\Psi)$. For $s = \frac{1}{2}$ this is the statement for $\hat{F}$, and taking the minimum over s gives $|\hat{C}(\Phi, \Psi') - \hat{C}(\Phi, \Psi)| \leq \ln \frac{1}{1-r}$. No property of $J(\Phi)$ is used. $\square$

Proof of the bound for D_{BS} in Theorem 1.4.1. Let Φ be a channel and let P be the projection onto the support of J_Φ . Put $X_\Psi = PJ_\Phi^{1/2}J_\Psi^{-1}J_\Phi^{1/2}P$, regarded as a positive definite operator on $\text{ran } P$. From $(1-r)J_\Psi \leq J_{\Psi'} \leq (1+r)J_\Psi$ and the operator monotonicity of inversion, $(1+r)^{-1}J_\Psi^{-1} \leq J_{\Psi'}^{-1} \leq (1-r)^{-1}J_\Psi^{-1}$, hence $(1+r)^{-1}X_\Psi \leq X_{\Psi'} \leq (1-r)^{-1}X_\Psi$. The logarithm is operator monotone, so $\ln X_\Psi - \ln(1+r)P \leq \ln X_{\Psi'} \leq \ln X_\Psi + \ln \frac{1}{1-r}P$. Conjugating with $J_\Phi^{1/2}$ preserves the order, and $J_\Phi^{1/2}PJ_\Phi^{1/2} = J_\Phi$. Taking Tr_Y , which is positive, and using $\text{Tr}_Y J_\Phi = \mathbf{1}$ for a channel gives

$$T_\Psi - \ln(1+r)\mathbf{1} \leq T_{\Psi'} \leq T_\Psi + \ln \frac{1}{1-r}\mathbf{1}, \quad T_\Psi = \text{Tr}_Y [J_\Phi^{1/2} \ln(X_\Psi) J_\Phi^{1/2}],$$

and the largest eigenvalues of T_Ψ and $T_{\Psi'}$ are the BS divergences. Since $\lambda_{\max}$ is monotone and $\lambda_{\max}(T + c\mathbf{1}) = \lambda_{\max}(T) + c$, $-\ln(1+r) \leq D_{\text{BS}}(\Phi\|\Psi') - D_{\text{BS}}(\Phi\|\Psi) \leq \ln \frac{1}{1-r}$. For a completely positive Φ the same argument gives the bound multiplied by $\|\Phi^*(\mathbf{1})\|_\infty$. $\square$

Proof of the sharpness statement, Eq. (1.4.3). The upper bound $d_X/(4\lambda)$ follows from Eq. (1.4.2) as $\epsilon \rightarrow 0$. For the lower bound let $v_W = \text{vec}(W)$, W unitary, be a λ -eigenvector of $J(\Psi)$, so $\|v_W\|^2 = d$, and let $\Phi = \text{Ad}_W$. For $B > 0$ the rank-one formula $v_W v_W^\dagger \# B = v_W v_W^\dagger / \langle v_W, B^{-1} v_W \rangle^{1/2}$ and $\text{Tr}_Y v_W v_W^\dagger = \mathbf{1}$ give $\hat{F}(\text{Ad}_W, \Psi') = \langle v_W, J(\Psi')^{-1} v_W \rangle^{-1/2}$. Take any unitary W' with $\text{Tr } W^\dagger W' = 0$ (it need not lie in the floor eigenspace), $\mathcal{M} = \frac{1}{2}(\text{Ad}_W - \text{Ad}_{W'})$, which has $\|\mathcal{M}\|_\diamond = 1$, and $\Psi_e = \Psi + e\mathcal{M}$, a channel for small $|e|$ because $J(\Psi) > 0$. Then $J(\mathcal{M}) = \frac{1}{2}(v_W v_W^\dagger - v_{W'} v_{W'}^\dagger)$, $\frac{d}{de} \langle v_W, J(\Psi_e)^{-1} v_W \rangle|_{e=0} = -\lambda^{-2} \langle v_W, J(\mathcal{M}) v_W \rangle = -d^2/(2\lambda^2)$ and $\langle v_W, J(\Psi)^{-1} v_W \rangle = d/\lambda$. Hence $\frac{d}{de} \ln \hat{F} = \frac{1}{2} \cdot \frac{d^2/(2\lambda^2)}{d/\lambda} = \frac{d}{4\lambda}$. For a Pauli channel $\mathcal{P}_q \circ \text{Ad}_U$ the vectors $\text{vec}(QU)$ are eigenvectors with eigenvalues dq_Q , so the floor eigenspace always contains such a v_W ; for local noise Q has full weight, which is the correlated error mentioned after Theorem 1.4.1. $\square$

A numerical check in 40-digit arithmetic gives the ratio of the slope to $d/(4\lambda)$ as 1.0000000 ($n = 1, p = 10^{-2}$), 1.0000034 ($n = 1, 10^{-3}$), 0.9999816 ($n = 2, 10^{-2}$) and 0.9999995 ($n = 2, 10^{-3}$); a 60-digit finite difference confirms $d^2/(4p^n)$ for $n = 1, 2, 3$ to 4×10^{-10} .

Two-sided version. If both channels move, with $\eta_J = \max\{\|J(\Phi') - J(\Phi)\|_\infty, \|J(\Psi') - J(\Psi)\|_\infty\}$ and $\lambda = \min\{\lambda(\Phi), \lambda(\Psi)\}$, then $(1 - \eta_J/\lambda)\hat{F}(\Phi, \Psi) \leq \hat{F}(\Phi', \Psi') \leq (1 + \eta_J/\lambda)\hat{F}(\Phi, \Psi)$, and similarly for every Kubo–Ando connection (Theorem 4.8.8 in Chapter 4; Proposition 8.15.2 also allows separate floors). Since $\hat{F}$ is symmetric, Eq. (1.4.3) shows that each of the two terms is sharp separately; whether the constant $d_X/(2\lambda)$ is attained when both channels move is open.

Proposition 8.3.1 (Absolute slope at a qubit gate). *Let $\Psi = \mathcal{D}_p \circ \text{Ad}_U$ on one qubit and $\Sigma(p) = \sup_\Phi \sup_{\mathcal{M}} |\frac{d}{de} \hat{F}(\Phi, \Psi + e\mathcal{M})|_{e=0}$, over completely positive Φ and trace-annihilating $\mathcal{M}$ with $\|\mathcal{M}\|_\diamond = 1$. Then $\frac{2}{3\sqrt{3}} \frac{1}{p} (1 + o(1)) \leq \Sigma(p) \leq \frac{1}{p}$. The lower bound is attained by unitary Φ ; the supremum over unitary Φ is*

$$\Sigma_{\text{unit}}(p) = \max_\theta \frac{\sqrt{2}}{4} \frac{\cos^2 \theta / \beta^2 + \sin^2 \theta / \lambda^2}{(\cos^2 \theta / \beta + \sin^2 \theta / \lambda)^{3/2}}, \quad \beta = 2 - \frac{3p}{2}, \quad \lambda = \frac{p}{2}, \quad (8.3.1)$$

with $p \Sigma_{\text{unit}} = 0.384904$ at $p = 10^{-2}$ and 0.384900 at $p = 10^{-3}, 10^{-4}$.

Proof. The upper bound is Eq. (1.4.2) with $\hat{F} \leq 1$. For unitary Φ with Choi vector $v = \sqrt{d}(\cos \theta \hat{e}_1 - i \sin \theta \hat{e}_P)$ in the eigenbasis of $J(\Psi)$, $\frac{d\hat{F}}{de} = \frac{1}{2} \hat{F}^3 \langle u, J(\mathcal{M}) u \rangle$ with $u = J(\Psi)^{-1} v$, which is proportional to the Choi vector of a unitary. Lemma 8.2.1 bounds

$\langle u, J(\mathcal{M})u \rangle \leq \frac{d}{2}\|u\|^2$, with equality for $\mathcal{M} = \frac{1}{2}(\text{Ad}_{\hat{u}} - \text{Ad}_{\hat{u}'})$. This gives Σ_{unit} ; with $\sin^2 \theta = y\lambda/d$ and $p \rightarrow 0$ it tends to $\frac{d}{4\lambda} \max_y y(1+y)^{-3/2} = \frac{d}{4\lambda} \frac{2}{3\sqrt{3}}$. $\square$

Conjecture 8.3.2. $\Sigma(p) = \Sigma_{\text{unit}}(p)$, i.e., unitary Φ are the worst case.

A Nelder–Mead search over Kraus-rank-four Φ with the optimal direction from the diamond-norm semidefinite program found no value above Σ_{unit} at $p = 10^{-2}$ and 10^{-3} ; random starts did not converge within the budget, so this evidence is weak. One apparent violation at $p = 10^{-3}$ ($p\Sigma = 5.34$, above the proved bound 1) was an artifact of double-precision finite differences near a nearly singular $J(\Phi)$; recomputed in 80-digit arithmetic it is 0.3846.

Stein exponent. By Corollary 5.7.5 in Chapter 5, if $J(\Psi)$ is invertible with condition number $\kappa(\Psi)$, then $\zeta(\Phi\|\Psi) \leq \ln(d_X/\lambda(\Psi))$ and, for all Φ' and all Ψ' with the same Choi support,

$$|\zeta(\Phi\|\Psi) - \zeta(\Phi'\|\Psi')| \leq 2\epsilon' \ln d_Y + (1+\epsilon')h\left(\frac{\epsilon'}{1+\epsilon'}\right) + \epsilon' \ln \kappa(\Psi) + d_T(\Psi, \Psi'), \quad \epsilon' = \frac{1}{2}\|\Phi - \Phi'\|_{\diamond}, \quad (8.3.2)$$

where h is the binary entropy (natural logarithms). The proof there combines Winter's continuity bound for the conditional entropy [129] with a cancellation of the reference marginal, and the Thompson-metric Lipschitz property of all regularized quantities in the second argument. With Eq. (8.2.1), $d_T(\Psi, \Psi') \leq \ln \frac{1}{1-r}$.

8.4 Closed forms for Pauli channels

Theorem 8.4.1 (Pauli-block closed forms). *Let $\mathcal{N}_{\delta}$ be as in Eq. (8.1.3) with all $q_Q > 0$, and let $\theta_Q = 2 \operatorname{arsinh} \sqrt{x_Q}$. Then $\operatorname{Tr}_{\mathcal{Y}}$ of $J(\mathcal{N}_{\delta})\#_s J(\mathcal{N}_0)$ and of the Belavkin–Staszewski operator is a multiple of the identity, and, with sums over the $d^2/2$ pairs $\{Q, QP\}$,*

$$\hat{F}(\mathcal{N}_{\delta}, \mathcal{N}_0) = \sum_{\text{pairs}} (q_Q + q_{QP})(1 + x_Q)^{-1/2},$$

$$\hat{Q}_s = \sum_{\text{pairs}} (q_Q + q_{QP}) \frac{\cosh((s - \frac{1}{2})\theta_Q)}{\cosh(\theta_Q/2)}, \quad \hat{C} = -\ln \hat{F}, \quad (8.4.1)$$

$$D_{\text{BS}}(\mathcal{N}_{\delta}\|\mathcal{N}_0) = D_{\text{BS}}(\mathcal{N}_0\|\mathcal{N}_{\delta}) = 2 \sum_{\text{pairs}} (q_Q + q_{QP}) \varphi(x_Q), \quad (8.4.2)$$

$$D(\mathcal{N}_{\delta}\|\mathcal{N}_0) = \sin^2 \frac{\delta}{2} \sum_{\text{pairs}} (q_Q - q_{QP}) \ln \frac{q_Q}{q_{QP}},$$

$$F_{\text{U}}(\mathcal{N}_{\delta}, \mathcal{N}_0) = \sum_{\text{pairs}} \sqrt{(q_Q + q_{QP})^2 - (q_Q - q_{QP})^2 \sin^2 \frac{\delta}{2}}. \quad (8.4.3)$$

For D and F_{U} the maximally entangled input is optimal.

Proof sketch. (i) *Blocks.* $\operatorname{vec}(QR_P(\delta)U) = (\mathbf{1} \otimes U^T)[\cos \frac{\delta}{2} \operatorname{vec} Q - i \sin \frac{\delta}{2} \operatorname{vec}(QP)]$, so in the orthonormal basis $\{\operatorname{vec}(Q)/\sqrt{d}\}$, after the unitary $\mathbf{1} \otimes U^T$, $J(\mathcal{N}_{\delta}) = \bigoplus_{\text{pairs}} d R \operatorname{diag}(q_Q, q_{QP}) R^{\dagger}$ with R the 2×2 rotation $\exp(-i \frac{\delta}{2} \sigma_x)$. Every two-variable functional calculus of the two Choi matrices is block diagonal in the same way. (ii) *Partial trace.* Conjugating by $\operatorname{diag}(1, i)$ makes both blocks real, so each block of any such function has a purely imaginary off-diagonal entry; since $\operatorname{Tr}_{\mathcal{Y}} |\operatorname{vec} A\rangle \langle \operatorname{vec} B| = (B^{\dagger} A)^T$, the off-diagonal contributions $2 \operatorname{Re}(m_{12}) P^T/d$ vanish and $\operatorname{Tr}_{\mathcal{Y}}$ is proportional to $\mathbf{1}$. (iii) *2×2 identities.* For $A = \operatorname{diag}(a, b)$ and $B = R A R^{\dagger}$, which

share trace T and determinant, $K = B^{-1/2}AB^{-1/2}$ has eigenvalues $e^{\pm\theta}$, and one finds $\text{Tr}(B\#_s A) = T \cosh((s - \frac{1}{2})\theta) / \cosh \frac{\theta}{2}$ and $D_{\text{BS}}(A\|B) = T\theta \tanh \frac{\theta}{2}$, with $\sinh \frac{\theta}{2} = \sqrt{x}$. The minimum over s is at $s = \frac{1}{2}$ in every block. The Umegaki and Uhlmann expressions follow from the qubit formulas for commuting spectra. (iv) *Input optimality*. Both channels are covariant under conjugation by the Paulis commuting with P and under one antiunitary symmetry (complex conjugation, possibly composed with a Pauli anticommuting with P). Averaging any input over the generated group, with a classical flag, does not change the value of a divergence that satisfies data processing, is invariant under unitaries and complex conjugation, and is additive over flagged blocks; the averaged input marginal is $\mathbf{1}/d$, whose purification is maximally entangled. $\square$

All closed forms were compared with direct matrix computations for $n = 1, 2$, random U and local and global noise, in 40-digit arithmetic; the largest deviation was 3×10^{-35} .

8.5 The multi-qubit window

Theorem 8.5.1 (Window law). *Let $\mathcal{N}_{p,\delta} = \mathcal{D}_p^{\otimes n} \circ \text{Ad}_{R_P(\delta)} \circ \text{Ad}_U$ with $\text{wt}P = w$.*

(a) $\lambda(\mathcal{D}_p^{\otimes n} \circ \text{Ad}_V) = (p/2)^n$ for every unitary V ; the eigenspace is spanned by $\text{vec}(QV)$ with $\text{wt}Q = n$, and $\lambda_{\max} = d(1 - \frac{3p}{4})^n$. For global noise, $\lambda = p/d$.

(b) With $\kappa_w = [(4 - 3p)/p]^w$ and $x_P = (\kappa_w - 1)^2 \sin^2(\delta/2)/(4\kappa_w)$,

$$\begin{aligned} (1 - \frac{3p}{4})^n g(x_P) &\leq 1 - \hat{F}(\mathcal{N}_{p,\delta}, \mathcal{N}_{p,0}) \leq g(x_P), \\ 2(1 - \frac{3p}{4})^n \varphi(x_P) &\leq D_{\text{BS}}(\mathcal{N}_{p,\delta} \| \mathcal{N}_{p,0}) \leq 2\varphi(x_P). \end{aligned} \quad (8.5.1)$$

(c) Let $\delta_\eta = \inf\{\delta : 1 - \hat{F} \geq \eta\}$, $x_\eta = (1 - \eta)^{-2} - 1$ and $x'_\eta = g^{-1}(\eta/(1 - \frac{3p}{4})^n)$. Then

$$2 \arcsin \frac{2\sqrt{\kappa_w x_\eta}}{\kappa_w - 1} \leq \delta_\eta \leq 2 \arcsin \frac{2\sqrt{\kappa_w x'_\eta}}{\kappa_w - 1}, \quad \delta_\eta = 4\sqrt{x_\eta} \left(\frac{p}{4}\right)^{w/2} [1 + O(np)]. \quad (8.5.2)$$

(d) At $\Psi = \mathcal{D}_p^{\otimes n} \circ \text{Ad}_U$ the worst-case log-modulus of Theorem 1.4.1, over the other argument, is $\frac{d}{4\lambda} = \frac{1}{4}(4/p)^n$.

(e) For global noise $\delta_\eta \simeq 4\sqrt{x_\eta p}/d$. For every n , $D(\mathcal{N}_{p,\delta} \| \mathcal{N}_{p,0}) \leq w \ln \frac{4-3p}{p} \sin^2 \frac{\delta}{2}$ and $1 - F_U(\mathcal{N}_{p,\delta}, \mathcal{N}_{p,0}) \leq \sin^2 \frac{\delta}{2}$.

Proof. (a) $J(\mathcal{D}_p) = (1-p)|\Omega\rangle\langle\Omega| + \frac{p}{2}\mathbf{1}_4$ has spectrum $\{2 - \frac{3p}{2}, \frac{p}{2}, \frac{p}{2}, \frac{p}{2}\}$; tensor products and output unitaries give the rest. (b) In Theorem 8.4.1, $q_{QP}/q_Q = \kappa_1^{-k}$ with $|k| \leq w$, and x is increasing in $|\ln(q_{QP}/q_Q)|$, so $x_Q \leq x_P$, attained by the pair $\{\mathbf{1}, P\}$. Since the weights $q_Q + q_{QP}$ sum to one and g, φ are increasing, $1 - \hat{F} = \sum(q_Q + q_{QP})g(x_Q) \leq g(x_P)$, and the lower bound keeps only the pair $\{\mathbf{1}, P\}$, whose weight is at least $(1 - \frac{3p}{4})^n$. (c) Invert g . (d) Theorem 1.4.1 with $W = PU$ for $\text{wt}P = n$. (e) Theorem 8.4.1 with $|\ln(q/q')| \leq w \ln \frac{4-3p}{p}$ and $\sum |q - q'| \leq 1$; for Uhlmann, $\sqrt{s^2 - \Delta^2 t} \geq s - \Delta^2 t/s$. $\square$

Table 8.1 compares Theorem 8.5.1(c) with exact evaluations. The window $\delta_{0.1}$ was found by bisection in 50-digit arithmetic, from the exact 2×2 block reduction, which was checked against brute-force 16×16 and 64×64 geometric means (including CNOT and

Table 8.1: Window $\delta_{0,1}$ at which $\hat{F}(\mathcal{N}_{p,\delta}, \mathcal{N}_{p,0}) = 0.9$, divided by $p^{w/2}$, for local depolarizing noise and a miscalibration of weight $w = n$. The predicted limit is $1.9373 \cdot 2^{-w}$. Slopes are fitted to $\ln \delta_{0,1}$ against $\ln p$ for $p \leq 10^{-3}$.

generator	$p = 10^{-1}$	10^{-2}	10^{-3}	10^{-5}	limit	slope
Z ($n = 1$)	1.0719	0.9780	0.9696	0.9687	0.96864	0.50019
$Z \otimes Z$ ($n = 2$)	0.5734	0.4923	0.4851	0.4843	0.48432	1.00031
$Z^{\otimes 3}$ ($n = 3$)	0.3120	0.2482	0.2428	0.2422	0.24216	1.50047

Toffoli as U) to 10^{-35} . All 32 local values lie inside the two-sided bounds of (c), and all 16 global values agree with the prediction to 7×10^{-17} . With $\eta = 0.1$, $4\sqrt{x_\eta} = 1.9373$.

A weight-one error on two qubits ($Z \otimes \mathbf{1}$) has exactly the window of Z on one qubit (the second factor sums out in Theorem 8.4.1), $\delta_{0,1} = 0.9696\sqrt{p}$ at $p = 10^{-3}$ and $0.9686\sqrt{p}$ as $p \rightarrow 0$; global depolarizing noise gives $\delta_{0,1}/\sqrt{p} \rightarrow 0.4843$ ($n = 2$) and 0.2422 ($n = 3$), as predicted by $1.9373/d$.

8.6 The monotone-metric hierarchy

Symmetric monotone metrics on full-rank states correspond to operator monotone $f : (0, \infty) \rightarrow (0, \infty)$ with $f(t) = tf(1/t)$ and $f(1) = 1$ [104]. For $\rho = \sum_i \mu_i |e_i\rangle\langle e_i|$ and a Hermitian tangent A , $F^f(\rho; A) = \sum_{ij} |A_{ij}|^2 c_f(\mu_i, \mu_j)$ with $c_f(\mu, \nu) = 1/(\nu f(\mu/\nu))$. The extremes are the RLD metric ($f(t) = 2t/(1+t)$) and the SLD metric ($f(t) = (1+t)/2$); BKM is $f(t) = (t-1)/\ln t$.

Proposition 8.6.1. *Let $\rho > 0$, $A = A^\dagger$, $\text{Tr } A = 0$ and $\rho_t = \rho + tA$.*

- (i) $F^{\text{SLD}} \leq F^f \leq F^{\text{RLD}} = \text{Tr } A\rho^{-1}A$.
- (ii) As $t \rightarrow 0$: $1 - \text{Tr}(\rho\#\rho_t) = \frac{t^2}{8}F^{\text{RLD}} + O(t^3)$, $D_{\text{BS}}(\rho_t\|\rho) = \frac{t^2}{2}F^{\text{RLD}} + O(t^3)$, $D(\rho_t\|\rho) = \frac{t^2}{2}F^{\text{BKM}} + O(t^3)$ and $1 - \|\sqrt{\rho}\sqrt{\rho_t}\|_1 = \frac{t^2}{8}F^{\text{SLD}} + O(t^3)$.
- (iii) For $0 < \nu < \mu$: $c_{\text{RLD}}(\mu, \nu) = \frac{\mu+\nu}{2\mu\nu}$, $c_{\text{BKM}}(\mu, \nu) = \frac{\ln(\mu/\nu)}{\mu-\nu}$, $c_{\text{SLD}}(\mu, \nu) = \frac{2}{\mu+\nu}$. In general $c_f(\mu, \nu) = 1/(\mu f(\nu/\mu))$ stays bounded as $\nu \rightarrow 0$ if and only if $f(0^+) > 0$.
- (iv) $c_f(\mu, \mu) = 1/\mu$ for every f : incoherent perturbations do not separate the metrics.
- (v) For a trace-preserving family $\mathcal{N}_t$ with $J(\mathcal{N}_0) > 0$, $1 - \hat{F}(\mathcal{N}_t, \mathcal{N}_0) = \frac{t^2}{8}\hat{F}^{\text{RLD}} + O(t^3)$ and $D_{\text{BS}}(\mathcal{N}_t\|\mathcal{N}_0) = \frac{t^2}{2}\hat{F}^{\text{RLD}} + O(t^3)$, with $\hat{F}^{\text{RLD}} = \lambda_{\max} \text{Tr}_{\mathcal{Y}}(J J_0^{-1} J)$ the channel RLD Fisher information [71].

Proof sketch. (i) is Ref. [104] with the Kubo–Ando bounds $\frac{2t}{1+t} \leq f(t) \leq \frac{1+t}{2}$. For (ii) put $K = \rho^{-1/2}A\rho^{-1/2}$ and expand $h(\mathbf{1} + tK)$ inside $\text{Tr} \rho^{1/2}(\cdot)\rho^{1/2}$, with $h = \sqrt{\cdot}$ for $\#$ and $h(x) = x \ln x$ for BS [61]; $\text{Tr} \rho^{1/2}K^2\rho^{1/2} = \text{Tr } A\rho^{-1}A$. For Umegaki the second derivative is $\int_0^\infty \text{Tr } A(\rho+s)^{-1}A(\rho+s)^{-1}ds$, and the Uhlmann case is the Bures–SLD relation [24]. (iii) and (iv) are algebra. (v) applies (ii) inside the partial trace; the first-order term vanishes by trace preservation, and Weyl’s inequality gives the extremal eigenvalues. $\square$

For the qubit family ($n = 1$, $P = Z$) Theorem 8.4.1 gives exact expressions for all δ , with $x = 4(1-p)^2 \sin^2(\delta/2)/[p(4-3p)]$:

$$\begin{aligned} \hat{F} &= \frac{p}{2} + (1 - \frac{p}{2})(1+x)^{-1/2}, & D_{\text{BS}} &= (2-p)\varphi(x), \\ D &= (1-p)\sin^2 \frac{\delta}{2} \ln \frac{4-3p}{p}, & F_{\text{U}} &= \frac{p}{2} + \sqrt{(1 - \frac{p}{2})^2 - (1-p)^2 \sin^2 \frac{\delta}{2}}. \end{aligned} \tag{8.6.1}$$

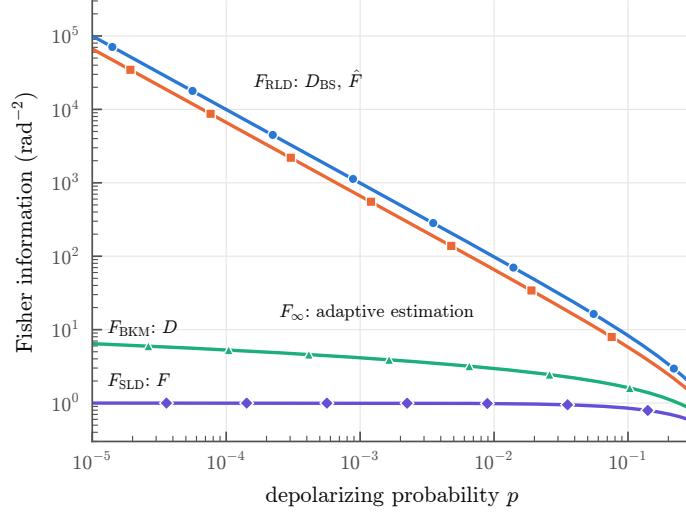


Figure 8.1: Exact Fisher informations per use at the depolarized qubit gate: F_{RLD} , which governs D_{BS} and $\hat{F}$ and grows as $1/p$; the asymptotic adaptive rate F_{∞} of Zhou and Jiang, which grows as $2/(3p)$; F_{BKM} , which governs the Umegaki divergence and grows as $\frac{1}{2} \ln(4/p)$; and F_{SLD} , which governs the Uhlmann fidelity and stays close to one.

Their second-order coefficients give the entries of Table 1.1, with $F_{\text{RLD}} = 2(1-p)^2(2-p)/[p(4-3p)]$, $F_{\text{BKM}} = \frac{1}{2}(1-p) \ln \frac{4-3p}{p}$ and $F_{\text{SLD}} = 2(1-p)^2/(2-p)$. For n qubits, $F_{\text{RLD}} = \frac{1}{4} \sum_{\text{pairs}} (q+q')(q-q')^2/(qq')$, $F_{\text{BKM}} = \frac{1}{2} \sum_{\text{pairs}} (q-q') \ln(q/q')$ and $F_{\text{SLD}} = \sum_{\text{pairs}} (q-q')^2/(q+q')$. These identities were checked symbolically and against 4×4 matrix evaluations at $\delta = 10^{-12}$ to relative accuracy 2×10^{-21} . Figure 8.1 shows the four Fisher informations as functions of p .

8.7 Amplitude damping

Let $\mathcal{A}_{\gamma}$ have Kraus operators $K_0 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|$ and $K_1 = \sqrt{\gamma}|0\rangle\langle 1|$, and write L_0, L_1 for those of $\mathcal{A}_{\gamma'}$. Proofs of the general statements used here are in Example 4.12.8, Section 7.6.1 and Section 8.15.

Infinite converses. For $\gamma \neq \gamma'$ in $(0, 1]$, $\text{ran} J(\mathcal{A}_{\gamma}) \cap \text{ran} J(\mathcal{A}_{\gamma'}) = \text{span}\{\text{vec } |0\rangle\langle 1|\}$ (the inclusion $\subseteq$, which is all that is used, holds for all $\gamma \neq \gamma'$ in $[0, 1]$; Example 4.12.8), and the level decomposition of the pair (Proposition 7.6.8) has conclusive levels $K_0 \cdot K_0^{\dagger}$ and $L_0 \cdot L_0^{\dagger}$ and a single ambiguous level with operator $S_a = (\gamma + \gamma')|1\rangle\langle 1|$. Every B_s entering $\hat{C}$ is a multiple of $|1\rangle\langle 1|$, hence singular, so $\hat{Q}_s = 0$ for all $s \in (0, 1)$; hence $\hat{C} = +\infty$, and $\hat{F} = \hat{Q}_{1/2} = 0$ gives $-2 \ln \hat{F} = +\infty$. For the barycentric converse [66], a channel $\mathcal{T}$ with both BS divergences finite would need $\text{ran} J(\mathcal{T}) \subseteq \text{span}\{\text{vec } |0\rangle\langle 1|\}$, forcing $\mathcal{T}^*(\mathbf{1}) \propto |1\rangle\langle 1| \neq \mathbf{1}$; so $C_{\text{bar}} = +\infty$. The common kernel vector is $|0\rangle$, on which the two channels act identically.

Program-simulation converse. For channels with Kraus operators (K_k) and (L_l) , padded to a common length m , let

$$\mu_*(\Phi, \Psi) = \max \left\{ |c| : \sum_{k,l} C_{lk} L_l^{\dagger} K_k = c \mathbf{1}, \|C\|_{\infty} \leq 1 \right\}, \quad (8.7.1)$$

a semidefinite program independent of the Kraus representations. By Theorem 7.6.9 in Chapter 7, (a) μ_* is the largest overlap of pure program states from which a single channel $\mathcal{L}$ simulates both channels, (b) every adaptive strategy with n uses has error probability

Table 8.2: Exponents for $\mathcal{A}_{1/2}$ against $\mathcal{A}_{\gamma'}$. The product-input value is a lower bound on ξ ; $-2 \ln \mu_*$ is a proved upper bound and $-2 \ln f$ a conjectured one. The single-letter converses $\hat{C}$, $-2 \ln \hat{F}$ and C_{bar} are infinite in every row.

γ'	product input $ 1\rangle$	$-2 \ln f$ (conjectured)	$-2 \ln \mu_*$ (proved)
0.40	5.077×10^{-3}	1.015×10^{-2}	2.252×10^{-2}
0.47	4.506×10^{-4}	9.012×10^{-4}	1.858×10^{-3}
0.49	5.001×10^{-5}	1.000×10^{-4}	2.020×10^{-4}

at least $\frac{1}{2}(1 - \sqrt{1 - \mu_*^{2n}}) \geq \frac{1}{4}\mu_*^{2n}$, so $\xi(\Phi, \Psi) \leq -2 \ln \mu_*$, and (c) no simulation by mixed programs gives a converse below $-\ln \mu_*$. The proof of (b) replaces each use of the unknown channel by $\mathcal{L}$ acting on a fresh program copy, as in the programmable simulations of Refs. [106, 107]. For amplitude damping the constraint forces C to be diagonal with $C_{00} = c$ and $C_{11} = c(1 - a)/\sqrt{\gamma\gamma'}$, $a = \sqrt{(1 - \gamma)(1 - \gamma')}$, which gives $\mu_* = \sqrt{\gamma\gamma'}/(1 - a)$ [Corollary 7.6.10]. The product input $|1\rangle$ produces the commuting outputs $\text{diag}(\gamma, 1 - \gamma)$ and $\text{diag}(\gamma', 1 - \gamma')$, whose classical Chernoff information is a lower bound on ξ . With $\Delta = \gamma - \gamma'$,

$$\frac{\Delta^2}{8\gamma(1 - \gamma)} + O(\Delta^3) \leq \xi(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) \leq \frac{\Delta^2}{4\gamma^2(1 - \gamma)} + O(\Delta^3). \quad (8.7.2)$$

The two bounds differ by the factor $2/\gamma$, and simulation arguments cannot close the gap: by Theorem 7.6.9(c), every program gives at least $-\ln \mu_*$, which exceeds the product-input value by $1/\gamma$ to leading order.

Conjecture (Conjecture 7.6.13). *For every adaptive n -use strategy, the fidelity of the two final states is at least f^{2n} with $f = \sqrt{\gamma\gamma'} + \sqrt{(1 - \gamma)(1 - \gamma')}$. Consequently $\xi(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) \leq -2 \ln f = \frac{\Delta^2}{4\gamma(1 - \gamma)} + O(\Delta^3)$.*

An exact semidefinite program (Uhlmann's theorem [122] plus a minimax over n -combs) returns f^n for the minimal root fidelity to within 10^{-8} for $n \leq 4$, $\gamma = \frac{1}{2}$ and $\gamma' \in \{0.05, 0.2, 0.4, 0.47, 0.49\}$. The conjecture cannot be proved by amortization: for one pair of different qubit–qubit input states the one-step fidelity ratio is $0.948315 < f(\frac{1}{2}, \frac{1}{5}) = 0.948683$ (40-digit check). The adaptive Fisher information for estimating γ is at most $n/(\gamma(1 - \gamma))$ [30], which the product input attains (Remark 7.6.12); this suggests, but does not prove, that ξ equals the product-input lower bound to leading order. Table 8.2 lists the values plotted in Fig. 1.2(a). Discrimination of amplitude-damping channels with finitely many uses is analyzed in Ref. [114].

For replacement channels $\Phi = \text{Tr}(\cdot)\phi$, $\Psi = \text{Tr}(\cdot)\psi$ one has $\mu_* = F_U(\phi, \psi)$, the root fidelity [see the discussion after Theorem 7.6.9], so $\min\{\hat{C}, C_{\text{bar}}, -2 \ln \mu_*\}$ is a single-letter converse that does not jump at the pairs of Fig. 1.2(b). That example is defined after Theorem 8.15.6; there $\|\Phi_t - \Phi\|_\diamond = 2\sqrt{t/(1 + t)}$, and $\xi = 2 \ln[(1 + t)/(1 - t)] = -2 \ln \mu_*$ exactly, so the program-simulation converse is tight. Numerically, this value agrees with an independently computed quantum Chernoff bound [11] to 1.5×10^{-8} , and $-2 \ln F_U(\phi_t, \psi_t)$ with $2 \ln[(1 + t)/(1 - t)]$ to 10^{-10} , while $\hat{F} \leq 6 \times 10^{-16}$ in double precision and $\leq 8 \times 10^{-11}$ from the semidefinite program.

8.8 The local Stein exponent

Proposition 8.8.1. *For the qubit pair $(\mathcal{N}_{p,\delta}, \mathcal{N}_{p,0})$, $\liminf_{\delta \rightarrow 0} \zeta/\delta^2 \geq F_\infty/2$ and $\limsup_{\delta \rightarrow 0} D_{\text{BS}}/\zeta \leq F_{\text{RLD}}/F_\infty = (2 - p)(3 - 2p)/(4 - 3p) < 3/2$ for every $p \in (0, 1)$.*

Table 8.3: Asymptotic adaptive Fisher information F_∞ and the RLD information F_{RLD} at the depolarized qubit gate, and the resulting bound on $\limsup_{\delta \rightarrow 0} D_{\text{BS}}/\zeta$.

p	pF_∞	pF_{RLD}	F_{RLD}/F_∞	$pF_\infty/2$ (bound on $p\zeta/\delta^2$)
10^{-1}	0.578571	0.831892	1.43784	0.289286
10^{-2}	0.657785	0.982569	1.49375	0.328893
10^{-3}	0.665778	0.998251	1.49938	0.332889
10^{-4}	0.666578	0.999825	1.49994	0.333289

Proof. Fix k and a parallel k -use strategy with an ancilla and input ρ . Its outputs ρ_k^δ have support $\text{supp}(\rho_R) \otimes \mathcal{Y}^{\otimes k}$ for all δ , because $J(\mathcal{N}_{p,\delta}) \geq \frac{p}{2}\mathbf{1}$ (by the Choi sandwich $(\text{id} \otimes \Psi)(\sigma) \geq \lambda_{\min}(J(\Psi)) \sigma_R \otimes \mathbf{1}$ of Lemma 5.7.3), and they are analytic in δ . Proposition 8.6.1(ii) gives $D(\rho_k^\delta \| \rho_k^0) = \frac{\delta^2}{2} F^{\text{BKM}}(\rho_k) + O(\delta^3)$, and $F^{\text{BKM}} \geq F^{\text{SLD}}$. Since ζ is the regularized (equivalently amortized) channel relative entropy [41, 125], $\zeta \geq \frac{1}{k} D(\rho_k^\delta \| \rho_k^0)$. Hence $\liminf_{\delta \rightarrow 0} \zeta/\delta^2 \geq \sup_k F_k^{\text{SLD}}/(2k) \geq F_\infty/2$, where F_k^{SLD} is the optimal k -use quantum Fisher information and $F_\infty = \lim_k F_k^{\text{SLD}}/k$ [131]. The upper side is $\zeta \leq D_{\text{BS}} = \frac{\delta^2}{2} F_{\text{RLD}} + O(\delta^4/p^2)$. Finally $(2-p)(3-2p) < \frac{3}{2}(4-3p)$ is equivalent to $p < \frac{5}{4}$. $\square$

Value of F_∞ . With Kraus operators $K_0 = \sqrt{1-3p/4}\mathbf{1}$ and $K_j = \sqrt{p/4}\sigma_j$, and generator $Z/2$, the span of $\{K_i^\dagger K_j\}$ is all Hermitian matrices, so the Hamiltonian-not-in-Kraus-span condition fails and the scaling is linear [31, 131]. The asymptotic rate is $F_\infty = 4 \min_{h:\beta(h)=0} \|\alpha(h)\|_\infty$ with $\alpha = (\dot{K} - ihK)^\dagger (\dot{K} - ihK)$ and $\beta = iK^\dagger (\dot{K} - ihK)$ [131]. We evaluated a feasible h (upper bound) and the minimax dual with $\rho_0 = \mathbf{1}/2$ (lower bound) in 50-digit arithmetic; they agree to a relative gap below 2×10^{-47} and with

$$F_\infty = \frac{2(1-p)^2}{p(3-2p)}, \quad pF_\infty = \frac{2}{3} - \frac{8p}{9} + O(p^2), \quad (8.8.1)$$

which is the closed form of Ref. [131, Sec. VII.A] for Pauli noise with $p_x = p_y = p_z = p/4$. A double-precision interior-point solution agrees to 3×10^{-9} . Table 8.3 lists values.

Conjecture (Conjecture 1.6.2, restated). $\lim_{\delta \rightarrow 0} D_{\text{BS}}/\zeta = F_{\text{RLD}}/F_\infty$, so that the ratio tends to $3/2$ as $p \rightarrow 0$.

The conjecture holds if the local coefficient of ζ is the SLD rate $F_\infty/2$. A priori it is governed by a relative-entropy (BKM-type) rate, which is at least the SLD rate. For the explicit strategies we evaluated, the BKM rate of a strategy strictly exceeds its own SLD rate, where by the rate of a k -use strategy we mean its Fisher information per use, F_k/k , which is twice the coefficient of δ^2 in the corresponding divergence per use. For example, the sequential strategy with maximally entangled input has $p \times (\text{SLD rate}) = 0.2334$ but $p \times (\text{BKM rate}) = 0.2992$, i.e., a relative-entropy coefficient $pD_k/(k\delta^2) \rightarrow 0.1496$. So the limit could lie strictly below F_{RLD}/F_∞ ; all that is established is the bracket $[1, F_{\text{RLD}}/F_\infty]$. Earlier sequential-strategy bounds, which gave $D_{\text{BS}}/\zeta \leq 3.34 + o(1)$, are superseded by Proposition 8.8.1.

8.9 First-order fragility: task versus bound

Theorem 1.4.1 concerns a *first-order* modulus: the change of a quantity per unit diamond-norm error in the reference channel. This section shows that the adaptive exponents obey the same law as the converses, and that they attain it at the worst case of Theorem 1.4.1.

The second-order curvature at the pair $(\mathcal{N}_{p,\delta}, \mathcal{N}_p)$ is a different question; the Chernoff side of it is treated in Sec. 8.10.

Conventions. In $\zeta(\Phi\|\Psi)$ the channel Φ is the null hypothesis and Ψ the alternative. The type-I error is that of rejecting Φ , and $\zeta \leq D_{\text{BS}}(\Phi\|\Psi)$. Theorem 1.4.1 perturbs the alternative Ψ . For a function q of the reference channel we define the first-order modulus

$$\text{Mod}_q(\Psi) = \limsup_{\epsilon \rightarrow 0} \sup_{\|\Psi' - \Psi\|_{\diamond} \leq \epsilon} \frac{|q(\Psi') - q(\Psi)|}{\epsilon}, \quad (8.9.1)$$

so that Eq. (1.4.3) reads $\sup_{\Phi} \text{Mod}_{-2 \ln \hat{F}(\Phi, \cdot)}(\Psi) = 1/(2\tilde{\lambda})$. We write $D_{\text{max}}(\Phi\|\Psi) = \ln \min\{t : J(\Phi) \leq tJ(\Psi)\}$. When $J(\Phi)$ is singular, $J(\Psi) \#_0 J(\Phi) = J(\Psi)$ by continuity from above, which is why $\hat{C}$ is defined with the infimum over $s \in (0, 1)$; for the rank-one pairs below the infimum is approached as $s \rightarrow 0$.

Lemma 8.9.1 (CP domination). *Suppose $\Psi - c\Phi$ is completely positive for some $c > 0$. Then the final states of every n -use adaptive strategy satisfy $\rho_{\Psi}^{(n)} \geq c^n \rho_{\Phi}^{(n)}$. For every test T with $\alpha_n = \text{Tr}(\mathbf{1} - T)\rho_{\Phi}^{(n)}$ and $\beta_n = \text{Tr} T \rho_{\Psi}^{(n)}$ it follows that $\beta_n \geq c^n(1 - \alpha_n)$ and $\frac{1}{2}(\alpha_n + \beta_n) \geq \frac{1}{2}c^n$. In particular $\zeta(\Phi\|\Psi) \leq D_{\text{max}}(\Phi\|\Psi)$ and $\xi(\Phi, \Psi) \leq D_{\text{max}}(\Phi\|\Psi)$.*

Proof. The final state is a composition of completely positive maps, each use of the channel contributing a factor $\Psi \otimes \text{id}$. Writing $\Psi = c\Phi + (\Psi - c\Phi)$ in every slot and expanding, all 2^n terms are positive semidefinite, and one of them is $c^n \rho_{\Phi}^{(n)}$. The best constant is $c = e^{-D_{\text{max}}(\Phi\|\Psi)}$. $\square$

Proposition 8.9.2 (Thompson–Lipschitz property). *Let Φ, Ψ, Ψ' be channels with $J(\Psi) > 0$. Each of $\zeta(\Phi\|\cdot)$, $\xi(\Phi, \cdot)$, $D_{\text{BS}}(\Phi\|\cdot)$, $\hat{C}(\Phi, \cdot)$, $-2 \ln \hat{F}(\Phi, \cdot)$, $D(\Phi\|\cdot)$ and $-2 \ln F_{\text{U}}(\Phi, \cdot)$ changes by at most $d_T(\Psi, \Psi')$ when Ψ is replaced by Ψ' . By Eq. (8.2.1), each has $\text{Mod} \leq 1/(2\tilde{\lambda}(\Psi))$.*

Proof. Let $t = d_T(\Psi, \Psi')$.

Operational exponents. For ζ this is Theorem 5.7.2 in Chapter 5; the following argument also covers ξ . Since $e^t \Psi - \Psi'$ is completely positive, the expansion in Lemma 8.9.1 gives $\rho_{\Psi'}^{(n)} \leq e^{nt} \rho_{\Psi}^{(n)}$ for the same strategy. Hence β_n and the Bayes error grow by at most e^{nt} , so $\zeta(\Phi\|\Psi') \geq \zeta(\Phi\|\Psi) - t$ and $\xi(\Phi, \Psi') \geq \xi(\Phi, \Psi) - t$. Exchanging Ψ and Ψ' gives the other direction.

The BS divergence. From $J(\Psi')^{-1} \geq e^{-t} J(\Psi)^{-1}$ and operator monotonicity of $\ln$ on the support of $J_{\Phi}^{1/2} J_{\Psi}^{-1} J_{\Phi}^{1/2}$, which equals the support of J_{Φ} , we get $J_{\Phi}^{1/2} \ln(\cdot) J_{\Phi}^{1/2} \geq$ (same at Ψ) $- t J_{\Phi}$. Moreover $\text{Tr}_Y J_{\Phi} = \mathbf{1}$.

The remaining quantities. $\hat{C}$ and $\hat{F}$ follow as in the proof of Theorem 1.4.1. D and F_{U} follow from monotonicity of $\ln$ and of the fidelity in each argument. $\square$

Thus every quantity considered, operational or not, is 1-Lipschitz in the relative (Thompson) error. The factor $1/\tilde{\lambda}$ arises only when a diamond-norm error is converted into a relative one, through Lemma 8.2.1.

Proposition 8.9.3 (The worst case of Theorem 1.4.1 is genuinely fragile). *Let $d_X = d_Y = d$ and let Ψ be a channel with $J(\Psi) > 0$. Let W be unitary with $J(\Psi)v_W = \mu v_W$ for $v_W = \text{vec } W$, and put $\tilde{\mu} = \mu/d$, $\hat{v}_W = v_W/\sqrt{d}$ and $\sigma = J(\Psi)/d$.*

(a) *The exponents, converses and comparison quantities all coincide:*

$$\begin{aligned} \zeta(\text{Ad}_W\|\Psi) &= \xi(\text{Ad}_W, \Psi) = D(\text{Ad}_W\|\Psi) = D_{\text{BS}}(\text{Ad}_W\|\Psi) \\ &= \hat{C}(\text{Ad}_W, \Psi) = -2 \ln \hat{F}(\text{Ad}_W, \Psi) = -2 \ln F_{\text{U}}(\text{Ad}_W, \Psi) = \ln(1/\tilde{\mu}). \end{aligned}$$

- (b) Let $\mathcal{M}$ be Hermitian preserving and trace annihilating. Along $\Psi + e\mathcal{M}$, each quantity in (a) is differentiable at $e = 0$ with the common derivative $-\langle \hat{v}_W, J(\mathcal{M})\hat{v}_W \rangle / \mu$. Each is in fact Fréchet differentiable at Ψ .
- (c) Each quantity in (a) has $\text{Mod} = 1/(2\tilde{\mu})$, attained by $\mathcal{M} = \frac{1}{2}(\text{Ad}_W - \text{Ad}_{W'})$ with $\text{Tr } W^\dagger W' = 0$. If $\mu = \lambda(\Psi)$, as at the worst case of Theorem 1.4.1, this equals $1/(2\tilde{\lambda})$, the bound of Proposition 8.9.2.
- (d) For $\Psi = \mathcal{P}_Q \circ \text{Ad}_U$ and $W = QU$, statement (a) holds at every Pauli channel $\mathcal{P}_{Q'} \circ \text{Ad}_U$ with $\tilde{\mu} = q'_Q$. Along the direction in (c), every quantity equals $\ln[1/(q_Q + e/2)]$ exactly.

Proof. (a) *Upper bound.* $J(\Psi) - c v_W v_W^\dagger \geq 0$ holds if and only if $c \leq 1/\langle v_W, J(\Psi)^{-1} v_W \rangle = \tilde{\mu}$. Lemma 8.9.1 with $c = \tilde{\mu}$ then gives $\zeta, \xi \leq \ln(1/\tilde{\mu})$.

Lower bound. Use the maximally entangled input and the test $\{\hat{v}_W \hat{v}_W^\dagger, \mathbf{1} - \hat{v}_W \hat{v}_W^\dagger\}$, accepting Ad_W only if all n outcomes are $\hat{v}_W$. This gives $\alpha_n = 0$ and $\beta_n = \langle \hat{v}_W, \sigma \hat{v}_W \rangle^n = \tilde{\mu}^n$. For a Pauli channel the test is the Bell measurement in the known frame.

Converses. For rank-one $J_\Phi = v_W v_W^\dagger$ and any $J > 0$ we have $J_\Phi^{1/2} \ln(J_\Phi^{1/2} J^{-1} J_\Phi^{1/2}) J_\Phi^{1/2} = \ln\langle v_W, J^{-1} v_W \rangle v_W v_W^\dagger$ and $J \#_s v_W v_W^\dagger = \langle v_W, J^{-1} v_W \rangle^{s-1} v_W v_W^\dagger$ for $s \in (0, 1]$. Moreover $\text{Tr}_Y v_W v_W^\dagger = \mathbf{1}$. Hence $D_{\text{BS}} = \hat{C} = -2 \ln \hat{F} = D_{\text{max}} = \ln\langle v_W, J^{-1} v_W \rangle$, which equals $\ln(1/\tilde{\mu})$.

Umegaki and Uhlmann. $\ln(1/\tilde{\mu}) = -\langle \hat{v}_W, \ln \sigma \hat{v}_W \rangle \leq D \leq D_{\text{max}}$. Also $\tilde{\mu} \leq F_U^2 \leq \langle \hat{v}_W, \sigma \hat{v}_W \rangle = \tilde{\mu}$, where the first inequality is domination and the second uses the maximally entangled input.

(b) Put $u(e) = \ln\langle \hat{v}_W, \sigma_e^{-1} \hat{v}_W \rangle$, $\ell_1(e) = -\langle \hat{v}_W, \ln \sigma_e \hat{v}_W \rangle$ and $\ell_2(e) = -\ln\langle \hat{v}_W, \sigma_e \hat{v}_W \rangle$. The arguments of (a) hold at every Ψ_e , and give $\ell_1 \leq D, \zeta \leq u$ and $\ell_2 \leq \xi, -2 \ln F_U \leq u$. The first four quantities of (a) are equal to u for all e . The three bounding functions are smooth and coincide at $e = 0$. Because $\hat{v}_W$ is an eigenvector of σ , their derivatives also coincide there and equal $-\langle \hat{v}_W, J(\mathcal{M})\hat{v}_W \rangle / (d\tilde{\mu})$.

(c) By Lemma 8.2.1, $|\langle \hat{v}_W, J(\mathcal{M})\hat{v}_W \rangle| \leq \frac{d}{2} \|\mathcal{M}\|_\diamond$. For the stated $\mathcal{M}$ the left side equals $d/2$ and $\|\mathcal{M}\|_\diamond = 1$.

(d) The vector $\text{vec}(QU)$ is an eigenvector of $J(\mathcal{P}_{Q'} \circ \text{Ad}_U)$ with eigenvalue dq'_Q . $\square$

Remark 8.9.4. At this worst case the Umegaki divergence and the Uhlmann fidelity also equal $\ln(1/\tilde{\lambda})$ and move at the rate $1/(2\tilde{\lambda})$. The worst direction changes a floor eigenvalue, and by Proposition 8.6.1(iv) such incoherent perturbations do not separate the metrics. The hierarchy RLD ($\sim 1/\tilde{\lambda}$), BKM ($\sim \ln(1/\tilde{\lambda})$) and SLD ($O(1)$) is therefore a statement about second-order curvature under coherent errors. It does not hold per unit of diamond distance.

For a general unitary $\Phi = \text{Ad}_W$ the same arguments give a sandwich:

$$-\ln\langle \hat{v}_W, \sigma \hat{v}_W \rangle \leq \xi \leq \zeta \leq D_{\text{BS}} = \hat{C} = -2 \ln \hat{F} = \ln\langle \hat{v}_W, \sigma^{-1} \hat{v}_W \rangle, \quad \zeta \geq -\langle \hat{v}_W, \ln \sigma \hat{v}_W \rangle. \quad (8.9.2)$$

By Jensen's inequality, its two ends coincide if and only if $\hat{v}_W$ is an eigenvector of σ .

Lemma 8.9.5 (Tight converses pass on their sensitivity). *Let $f \leq g$ near Ψ , with $J(\Psi) > 0$ and $f(\Psi) = g(\Psi)$, and let g be Fréchet differentiable at Ψ . Then $\text{Mod}_f(\Psi) \geq \text{Mod}_g(\Psi)$. If g is only directionally differentiable, then $\text{Mod}_f \geq \sup_{\|\mathcal{M}\|_\diamond=1} \max\{0, -g'(\Psi; \mathcal{M})\}$.*

Proof. Choose $\mathcal{M}$ with $\|\mathcal{M}\|_\diamond = 1$ along which g decreases, $g'(\Psi; \mathcal{M}) = -c \leq 0$. Then $f(\Psi + e\mathcal{M}) \leq g(\Psi + e\mathcal{M}) = f(\Psi) - ce + o(e)$, so $|f(\Psi + e\mathcal{M}) - f(\Psi)| \geq ce - o(e)$. $\square$

Applied with $f \in \{\zeta, \xi\}$ and g a converse, the lemma shows two things. First, wherever a converse is tight, the exponent is at least as sensitive at first order as the converse. Second, by Proposition 8.9.2, the exponent is never more sensitive than $1/(2\tilde{\lambda})$. A first-order spurious gap can therefore occur only at pairs where the converse already overestimates the exponent, that is, at pairs that are spurious at zeroth order.

Proposition 8.9.6 (Pauli-diagonal pairs). *Let $\Phi = \mathcal{P}_a \circ \text{Ad}_U$ and $\Psi = \mathcal{P}_b \circ \text{Ad}_U$ with all $b_Q > 0$. Then $\zeta(\Phi\|\Psi) = D(\Phi\|\Psi) = D_{\text{BS}}(\Phi\|\Psi) = \sum_Q a_Q \ln(a_Q/b_Q)$, $\xi(\Phi, \Psi) = \hat{C}(\Phi, \Psi) = -\ln \inf_{s \in (0,1)} \sum_Q a_Q^s b_Q^{1-s}$, and $\hat{F} = \sum_Q \sqrt{a_Q b_Q}$.*

Proof. Achievability. The Bell measurement in the known frame produces i.i.d. samples of a or of b , to which the classical Stein and Chernoff theorems apply.

Converse. As in the proof of Proposition 8.13.3(a), the final state of any strategy is $\Lambda(a^{\otimes n})$ or $\Lambda(b^{\otimes n})$ for a channel Λ independent of the rates. Any test therefore pulls back to a test between i.i.d. classical sources.

Values of the converses. The two Choi matrices commute, and Tr_Y of each projector onto $\text{vec}(QU)$ is $\mathbf{1}$. $\square$

On this class the single-letter converses are exact, so their sensitivity along Pauli directions belongs to the task at every order. The pair $(\mathcal{N}_{p,\delta}, \mathcal{N}_p)$ is not in this class, because the miscalibration is coherent.

Proposition 8.9.7 (A first-order spurious gap). *Take preparation channels ($d_X = 1$, so $\tilde{\lambda} = \lambda$ and the diamond norm is the trace norm). Let $\sigma = \text{diag}(1 - \lambda, \lambda)$ and $\phi = \sqrt{1 - y\lambda}|0\rangle + \sqrt{y\lambda}|1\rangle$ with $0 < y < 1/\lambda$.*

- (a) *The converses are $D_{\text{BS}} = \hat{C} = -2 \ln \hat{F} = \ln[(1 - y\lambda)/(1 - \lambda) + y]$, which tends to $\ln(1 + y)$.*
- (b) *The exponents are $\zeta = -(1 - y\lambda) \ln(1 - \lambda) - y\lambda \ln \lambda$ and $\xi = -\ln[(1 - y\lambda)(1 - \lambda) + y\lambda^2]$.*
- (c) *The first-order moduli are*

$$\begin{aligned} \text{Mod}_{D_{\text{BS}}} &= \frac{y}{2(1 + y)\lambda} [1 + O(\lambda)], \\ \text{Mod}_{\zeta} &= \frac{1}{2} |1 - y| + O(\lambda \ln^2 \frac{1}{\lambda}) \quad (y \neq 1), \quad \text{Mod}_{\xi} = \frac{1}{2} + O(\lambda), \end{aligned} \tag{8.9.3}$$

$$\text{and at } y = 1, \text{Mod}_{\zeta} = \sqrt{\lambda(1 - \lambda)} \ln[(1 - \lambda)/\lambda] / (1 - 2\lambda).$$

Proof. Exponents. The quantum Stein lemma gives $\zeta = D(\phi\|\sigma)$. For pure ϕ the quantity $\langle \phi, \sigma^t \phi \rangle$ is minimized over $t \in [0, 1]$ at $t = 1$, which gives ξ .

Moduli. For traceless $\mathcal{M}$ with $\|\mathcal{M}\|_1 \leq 1$, $\sup |\text{Tr } \mathcal{M} H| = \frac{1}{2} [\lambda_{\max}(H) - \lambda_{\min}(H)]$. The derivative operators are $H = vv^\dagger / \langle \phi, \sigma^{-1} \phi \rangle$ with $v = \sigma^{-1} \phi$ for the converse, the Hadamard product of $\phi \phi^\dagger$ with the divided differences of $\ln$ for D , and $\phi \phi^\dagger / \langle \phi, \sigma \phi \rangle$ for ξ . $\square$

The ratio $\text{Mod}_{D_{\text{BS}}} / \text{Mod}_{\zeta}$ is of order $1/\lambda$; at $y = 1$ the two moduli are $1/(4\lambda)$ and roughly $\sqrt{\lambda} \ln(1/\lambda)$. The values already differ at zeroth order, as Lemma 8.9.5 requires. For replacement channels on a qubit, every adaptive strategy outputs $\Lambda(\phi^{\otimes n})$ or $\Lambda(\sigma'^{\otimes n})$, so the same moduli hold along replacement directions.

The slope maximizer of Proposition 8.3.1. Let $\Phi = \text{Ad}_{R_Z(\theta)U}$ and $\Psi = \mathcal{D}_p \circ \text{Ad}_U$ with $\sin^2(\theta/2) = p/2$, the unitary that maximizes the absolute slope of $\hat{F}$. Here $\hat{F} \rightarrow 1/\sqrt{3}$, and all converses equal $\ln[(1 - p/2)/(1 - 3p/4) + 2] \rightarrow \ln 3$. Table 8.4 brackets ζ . The lower bound comes from sequential strategies with maximally entangled input. The upper bound

Table 8.4: Converse and Stein exponent at the slope maximizer of Proposition 8.3.1. The integer k is the number of sequential uses in the best lower-bound strategy. U_3 is Eq. (8.3.2) at the base pair (Ad_U, Ψ) , where $\zeta = -\ln(1 - 3p/4)$ by Proposition 8.9.3.

p	$D_{\text{BS}} = -2 \ln \hat{F}$	$\zeta \geq (k)$	$\zeta \leq D_\alpha^\# (1/\alpha)$	$\zeta \leq U_3$	ratio $\geq$	ratio $\leq$
10^{-2}	1.09945	0.17298 (14)	0.48014 (11/16)	0.78916	2.29	6.36
10^{-3}	1.09870	0.071789 (46)	0.22020 (3/4)	0.32478	4.99	15.3
10^{-4}	1.09862	0.028417 (151)	0.091412 (3/4)	0.12692	12.0	38.7

is the one-copy divergence $D_\alpha^\#$ of Ref. [43], re-checked and rescaled in 40-digit arithmetic. The converse overshoots ζ by a certified factor that grows as p decreases. The absolute slope $0.3849/p$ of Proposition 8.3.1 is therefore a slope of the converse at a pair where the converse is already spurious. The first-order modulus of ζ at this pair is not determined.

Numerical checks. All entries below were computed from direct 4×4 and 16×16 matrices in 40-digit arithmetic, without using the closed forms.

- *Proposition 8.9.3(a).* Cases: depolarizing noise with $p = 10^{-2}, 10^{-3}, 10^{-4}$; a biased Pauli channel ($q_Z = 10^{-3}, q_X = q_Y = 10^{-5}$) with W chosen both on and off the floor; and two qubits with $U = \text{CNOT}$ and $W = (Z \otimes X)U$. In every case $-2 \ln \hat{F}$, D_{BS} , D_{max} , the Umegaki divergence at maximally entangled input, the Bell-test exponent and $-2 \ln F_U$ equal $\ln(1/q_W)$ to 10^{-24} or better. $\hat{C}$ agrees to 10^{-12} , because its infimum is approached only as $s \rightarrow 0$.
- *Proposition 8.9.3(b).* In the direction of Theorem 1.4.1 and in three random directions per case, the derivatives of u , D_{BS} , ℓ_1 and ℓ_2 agree with $-\langle \hat{v}_W, J(\mathcal{M})\hat{v}_W \rangle / \mu$ to 3×10^{-16} . In that direction the slope times $2q_W$ is 1.
- *Proposition 8.9.2.* Over eight random pairs, the largest ratio $|\mathcal{M}q|/d_T$ is 0.74.
- *Proposition 8.9.6.* The deviations are at most 2×10^{-40} (D_{BS}) and 3×10^{-16} ($\hat{Q}_s$).
- *Proposition 8.9.7.* The moduli agree with finite-difference suprema to all printed digits.

8.10 The Chernoff exponent at a noisy gate

For the pair of Sec. 8.8, Theorem 8.4.1 gives the exact value

$$\hat{C}(\mathcal{N}_{p,\delta}, \mathcal{N}_p) = -\ln \left[\frac{p}{2} + \left(1 - \frac{p}{2}\right)(1+x)^{-1/2} \right] = \frac{F_{\text{RLD}}}{8} \delta^2 + O(\delta^4/p^2), \quad (8.10.1)$$

with $x = 4(1-p)^2 \sin^2(\delta/2)/[p(4-3p)]$. We compare it with ξ .

Lemma 8.10.1 (Midpoint bound). *Call B a strong-converse Stein bound if every n -use adaptive strategy and test T satisfy $\text{Tr } T \rho_\Psi^{(n)} \geq c(\text{Tr } T \rho_\Gamma^{(n)}) e^{-nB(\Gamma\|\Psi) - o(n)}$, for some function c with $c(t) > 0$ whenever $t > 0$. For such a B and any channels Φ, Ψ, Γ ,*

$$\xi(\Phi, \Psi) \leq \max\{B(\Gamma\|\Phi), B(\Gamma\|\Psi)\}. \quad (8.10.2)$$

The following are strong-converse Stein bounds:

- D_{max} , by Lemma 8.9.1;

- the geometric Rényi divergences $\hat{D}_\alpha$ with $\alpha \in (1, 2]$, by their chain rule [42] together with $\tilde{D}_\alpha \leq \hat{D}_\alpha$ and

$$\mathrm{Tr} T\sigma \geq (\mathrm{Tr} T\rho)^{\alpha/(\alpha-1)} e^{-\tilde{D}_\alpha(\rho\|\sigma)};$$

letting $\alpha \downarrow 1$ gives D_{BS} ;

- the one-copy $D_\alpha^\#$ [43].

At the gate, $\xi(\mathcal{N}_{p,\delta}, \mathcal{N}_p) \leq B(\mathcal{N}_{p,\delta/2} \|\mathcal{N}_p)$.

Proof. Let T be the decision “ Φ ”. Whether $\mathrm{Tr} T\rho_\Gamma \geq \frac{1}{2}$ or $\mathrm{Tr}(\mathbf{1} - T)\rho_\Gamma > \frac{1}{2}$, one of the two error terms is at least $\frac{1}{2}c(\frac{1}{2})e^{-nB-o(n)}$. At the gate take $\Gamma = \mathcal{N}_{p,\delta/2}$. Post-rotation by $R_Z(-\delta/2)$, and conjugation by X combined with a rotation, show that $B(\Gamma \|\mathcal{N}_{p,\delta}) = B(\Gamma \|\mathcal{N}_p)$. $\square$

The two-copy bound $\frac{1}{2}D_\alpha^\#(\mathcal{N}^{\otimes 2} \|\mathcal{N}_0^{\otimes 2})$ used for ζ_+ is not included in the list, because it bounds ζ only through $\zeta = D^{\mathrm{reg}}$, which is a weak converse.

Proposition 8.10.2 (Local Chernoff exponent). *For every $p \in (0, 1)$,*

$$\frac{F_\infty}{8} = \frac{(1-p)^2}{4p(3-2p)} \leq \liminf_{\delta \rightarrow 0} \frac{\xi}{\delta^2} \leq \limsup_{\delta \rightarrow 0} \frac{\xi}{\delta^2} \leq \frac{F_{\mathrm{RLD}}}{8} = \frac{(1-p)^2(2-p)}{4p(4-3p)}. \quad (8.10.3)$$

Hence $1 \leq \liminf_{\delta \rightarrow 0} \hat{C}/\xi \leq \limsup_{\delta \rightarrow 0} \hat{C}/\xi \leq F_{\mathrm{RLD}}/F_\infty = (2-p)(3-2p)/(4-3p) < 3/2$, the same bracket as for D_{BS}/ζ . As $p \rightarrow 0$, $p\xi/\delta^2 \in [\frac{1}{12}, \frac{1}{8}]$.

Proof. Upper bound. Use $\xi \leq \hat{C}$ (Theorem 7.4.1). Alternatively, Lemma 8.10.1 with $B = D_{\mathrm{BS}}$ gives $\xi \leq D_{\mathrm{BS}}(\mathcal{N}_{p,\delta/2} \|\mathcal{N}_p) = \frac{F_{\mathrm{RLD}}}{2}(\frac{\delta}{2})^2 + O(\delta^4/p^2)$, which has the same constant and uses only Ref. [42].

Lower bound. The argument follows Proposition 8.8.1. Fix k and a parallel k -use strategy with an ancilla. Its outputs ρ_k^δ have δ -independent support and are analytic in δ . The quantum Chernoff theorem [11, 99], applied to i.i.d. blocks, gives $\xi \geq \frac{1}{k}C(\rho_k^\delta, \rho_k^0)$. Moreover $C \geq -\ln \mathrm{Tr} \sqrt{\rho_k^\delta} \sqrt{\rho_k^0} \geq -\ln F_{\mathrm{U}} \geq 1 - F_{\mathrm{U}} = \frac{\delta^2}{8} F^{\mathrm{SLD}}(\rho_k) + O(\delta^3)$ by Proposition 8.6.1(ii). Taking the supremum over k and using Ref. [131] gives $F_\infty/8$. $\square$

The lower-bound argument discards a Wigner–Yanase factor: a fixed strategy has local coefficient $F^{\mathrm{WY}}(\rho_k)/(8k)$, with $F^{\mathrm{SLD}} \leq F^{\mathrm{WY}} \leq 2F^{\mathrm{SLD}}$. For the sequential strategy below, the Chernoff constant exceeds that strategy’s own SLD rate by a factor 1.186. The exact local constant within Eq. (8.10.3) is therefore open, as it is for ζ (Conjecture 1.6.2).

Explicit strategies. Each formula below was verified against explicit simulation for $k \leq 3$ to 20 digits. We consider three strategies.

- (P) *Bell measurement after each use.* This gives the classical Chernoff information of $(q_I, q_Z, \frac{p}{2})$ against $(q_I c^2 + q_Z s^2, q_Z c^2 + q_I s^2, \frac{p}{2})$ with $s = \sin \frac{\delta}{2}$, which is of order δ^4/p .
- (A) *k sequential uses on half of a Bell pair, with $U^\dagger$ between uses and a collective measurement on i.i.d. blocks:*

$$\xi \geq -\frac{1}{k} \ln \left[1 - \sin^2 \frac{k\delta}{2} \left(\sqrt{1 - \frac{3q_k}{4}} - \sqrt{\frac{q_k}{4}} \right)^2 \right], \quad q_k = 1 - (1-p)^k. \quad (8.10.4)$$

- (B) *The heralded code of Sec. 8.11, with abort on a flag, $r_k = (1 - 2a)^k$ and $a = (p/4)/(1 - p/2)$:*

$$\xi \geq -\frac{1}{k} \ln \left[1 - (1 - \frac{p}{2})^k (1 - \sqrt{1 - r_k^2}) \sin^2 \frac{k\delta}{2} \right]. \quad (8.10.5)$$

Table 8.5: Local constants $p\xi/\delta^2$ as $\delta \rightarrow 0$, maximized over the number k of sequential uses (in parentheses). The entry $pF_{\text{RLD}}/8$ equals $p\hat{C}/\delta^2$.

p	(A)	(B)	$pF_{\infty}/8$	$pF_{\text{RLD}}/8$	product ($k = 1$)
10^{-2}	0.034417 (51)	0.036303 (57)	0.082223	0.122821	0.00224
10^{-3}	0.034573 (514)	0.036504 (568)	0.083222	0.124781	0.000242
10^{-4}	0.034588 (5145)	0.036524 (5684)	0.083322	0.124978	0.0000247
$\rightarrow 0$	0.0345901	0.0365266	1/12	1/8	0

For (A) with $k = 1$, i.e. a product strategy with a collective measurement, $p\xi/\delta^2 \rightarrow p/4 \rightarrow 0$. Single-use strategies therefore have curvature of order one, and the $1/p$ curvature is produced by adaptive accumulation. As $p \rightarrow 0$, the local constants of (A) and (B) tend to

$$\begin{aligned} c_A &= \max_u \frac{u}{16} \left(\sqrt{4 - 3(1 - e^{-u})} - \sqrt{1 - e^{-u}} \right)^2 = 0.0345901, \\ c_B &= \max_u \frac{u}{4} e^{-u/2} \left(1 - \sqrt{1 - e^{-u}} \right) = 0.0365266. \end{aligned} \quad (8.10.6)$$

Table 8.5 lists their finite- p values.

Proposition 8.10.3 (Finite miscalibration). *(a) For $p \ll \delta \ll 1$, strategy (A) with $k = \lfloor \pi/\delta \rfloor$ gives $\xi \geq \frac{\delta}{2\pi} \ln \frac{\delta}{\pi p} [1 + o(1)]$. (b) By Eq. (8.3.2) (Corollary 5.7.5), $\xi \leq \zeta \leq U_3 = \frac{\delta}{2} [\ln \frac{4}{p} + \ln \frac{2}{\delta} + O(1)]$, and since $\hat{C}(c\sqrt{p}) \rightarrow \frac{1}{2} \ln(1 + c^2/4)$,*

$$\sup_{\delta} \frac{\hat{C}}{\xi} \geq \frac{0.2682 + o(1)}{\sqrt{p} \ln(1/p)}, \quad 0.2682 = \max_c \frac{\ln(1 + c^2/4)}{1.5c}. \quad (8.10.7)$$

Proof. (a) With $k\delta \simeq \pi$ the block Chernoff bound is $-\ln[q_k/2 + 2\sqrt{(1 - \frac{3q_k}{4})\frac{q_k}{4}}] = \frac{1}{2} \ln \frac{1}{q_k} [1 + o(1)]$, where $q_k \simeq \pi p/\delta$. Dividing by $k \simeq \pi/\delta$ gives the claim. (b) Expand U_3 with $\epsilon' = (1 - p) \sin(\delta/2)$ and use $h(\epsilon') = \epsilon' \ln(1/\epsilon') + O(\epsilon')$. At $\delta = c\sqrt{p}$ this gives $U_3 = \frac{\delta}{2} [\frac{3}{2} \ln \frac{1}{p} + O(1)]$; maximize over c . $\square$

The curvature of $\hat{C}$ is thus genuine for $\delta \lesssim p \ln(1/p)$ and spurious for $p \ln(1/p) \lesssim \delta \lesssim \sqrt{p}$. In the latter range ξ grows only linearly in δ (up to logarithms), while $\hat{C}$ keeps its $\delta^2/(8p)$ curvature. These are the same three scales as for the Stein exponent.

Table 8.6 gives certified numbers. The upper bound is $\xi_+ = \min\{\hat{C}(\delta), B(\delta/2)\}$, where B is the smallest of D_{BS} , the certified one-copy $D_{\alpha}^{\#}$ of Sec. 8.11, $D_{\text{max}} = 2 \operatorname{arsinh} \sqrt{x}$ and $\ln(4/p)$. The lower bound ξ_- is the best of (P), (A) and (B). Above the onset the binding upper bound is $D_{\alpha}^{\#}$ at $\delta/2$, so the certified overshoot rests on Refs. [42, 43] and Lemma 8.10.1 only. The bound (8.3.2) is never binding. The program-simulation converse of Theorem 7.6.9 gives $-2 \ln \mu_* \geq 1.34 \hat{C}$ at every tested point, so it does not help at this pair.

8.11 The Stein exponent at finite miscalibration

Lower bounds. Each lower bound is $\zeta \geq D_k/k$ for an explicit k -use strategy, maximized over integer $k \leq \max(10/p, 50/\delta)$. (A) Sequential, maximally entangled input, $U^{\dagger}$ between uses: $D_k = (1 - q) \sin^2(k\delta/2) \ln[1 + 4(1 - q)/q]$ with $q = 1 - (1 - p)^k$. (B) Sequential with a heralded two-qubit code $\{|00\rangle, |11\rangle\}$ (probe plus noiseless ancilla), in which bit-flip errors are flagged and phase errors act as logical dephasing: $D_k = (1 - \frac{p}{2})^k c_k \operatorname{artanh}(c_k) (1 - \cos k\delta)$ with $c_k =$

Table 8.6: Chernoff exponent at finite δ , with $x = \delta/(p \ln(1/p))$. Ratios divided by ξ_+ are certified lower bounds on $\hat{C}/\xi$; ratios divided by ξ_- are upper bounds.

	$p = 10^{-2}$	$p = 10^{-3}$	$p = 10^{-4}$
small- x value of $\hat{C}/\xi_-$ ($x = 0.02$)	3.38	3.42	3.42
smallest x with $\hat{C}/\xi_+ > 1.01$	3.17	1.59	1.26
maximum of $\hat{C}/\xi_+$	1.125 ($\delta = 0.29$)	2.483 ($\delta = 0.087$)	5.996 ($\delta = 0.029$)
peak of $\hat{C}/\xi_+$ times $\sqrt{p} \ln(1/p)$	0.52	0.54	0.55
maximum of $\hat{C}/\xi_-$	5.23	11.1	26.4
ξ_+/ξ_- over the grid	1.05–4.95	2.36–5.30	3.42–5.29

Table 8.7: Key numbers of Fig. 1.5 and Fig. 8.2, with $\bar{\delta} = \delta/(p \ln(1/p))$. Ratios divided by ζ_+ are certified lower bounds on the true ratio; ratios divided by ζ_- are upper bounds.

	$p = 10^{-2}$	$p = 10^{-3}$	$p = 10^{-4}$
D_{BS}/ζ_- at $\bar{\delta} = 0.01$	3.127	3.159	3.162
smallest $\bar{\delta}$ with $D_{\text{BS}}/\zeta_+ > 1.01$	1.26	0.79	0.63
maximum of D_{BS}/ζ_+	1.81 ($\delta = 0.23$)	3.74 ($\delta = 0.069$)	9.15 ($\delta = 0.023$)
position of the maximum, $\delta/\sqrt{p}$	2.31	2.18	2.31
$\hat{F}$ at the maximum	0.66	0.68	0.65
the same with U_3 and $\ln(4/p)$ only	1.29	3.06	7.70
maximum of D_{BS}/ζ_-	5.21	11.9	29.9
maximum of $(-2 \ln \hat{F})/\zeta_+$	1.07	2.17	5.36
ζ_+/ζ_- over the grid	1.01–3.89	2.28–4.05	3.01–4.08
peak of D_{BS}/ζ_+ times $\sqrt{p} \ln(1/p)$	0.83	0.82	0.84

$(1 - 2a)^k$, $a = (p/4)/(1 - p/2)$. Both formulas were verified by explicit simulation for $k \leq 3$ to 20 digits. As $\delta/p \rightarrow 0$ they give $p\zeta/\delta^2 \geq 0.149587$ (A) and 0.158119 (B). These finite-length strategies fall short of the asymptotic coefficient $pF_\infty/2 \rightarrow 1/3$ of Proposition 8.8.1, which is why the small- δ ratio $D_{\text{BS}}/\zeta_- \approx 0.5/0.158 = 3.16$ in Fig. 1.5(a) lies above the proved bracket $[1, F_{\text{RLD}}/F_\infty]$: the numerical bracket for $\bar{\delta} \lesssim 1$ is $[1, 3.16]$, and only in the limit $\delta \rightarrow 0$ does Proposition 8.8.1 narrow it to $[1, F_{\text{RLD}}/F_\infty] \subset [1, 3/2]$.

Upper bounds. $\zeta_+ = \min\{D_{\text{BS}}, U_3, \ln(4/p), D_\alpha^\#, \frac{1}{2}D_\alpha^\#(\mathcal{N}^{\otimes 2} \|\mathcal{N}_0^{\otimes 2})\}$. Here U_3 is Eq. (8.3.2) applied at the pair $(\mathcal{N}_{p,0}, \mathcal{N}_{p,0})$, where $\zeta = 0$ and $d_T = 0$, with $\epsilon' = (1 - p)|\sin(\delta/2)|$ (checked against the diamond-norm semidefinite program) and $\kappa = (4 - 3p)/p$. The Fawzi–Fawzi divergences satisfy $\zeta \leq \tilde{D}_\alpha^{\text{reg}} \leq D_\alpha^\#$ for $\alpha > 1$ [43]; we solved their semidefinite programs for $1/\alpha$ on a dyadic grid in $[\frac{1}{2}, \frac{31}{32}]$, re-checked each solution in 25-digit arithmetic and rescaled it to be exactly feasible, so each value is a valid upper bound. The two-copy program was solved on alternate grid points; it improved ζ_+ by up to 6.8% at $p = 10^{-2}$ and by at most 0.5% at $p = 10^{-3}$.

Which bound sets ζ_+ . For $\bar{\delta} \lesssim 1$ the minimum defining ζ_+ is D_{BS} itself. At every grid point from the onset of the overshoot through its maximum, the minimum is attained by a Fawzi–Fawzi divergence (one or two copies) [43]; the continuity bound U_3 of Corollary 5.7.5 in Chapter 5 attains the minimum only at $\bar{\delta} \geq 50$ ($p = 10^{-3}$) and $\bar{\delta} \geq 79$ ($p = 10^{-4}$), far beyond the peaks. The certified peak overshoots 1.81, 3.74 and 9.15 therefore depend on the Fawzi–Fawzi bound only. The row “the same with U_3 and $\ln(4/p)$ only” of Table 8.7 shows the weaker values obtained without the Fawzi–Fawzi bounds.

The remaining factor of at most 4.1 between ζ_- and ζ_+ is not attributed to either side: a

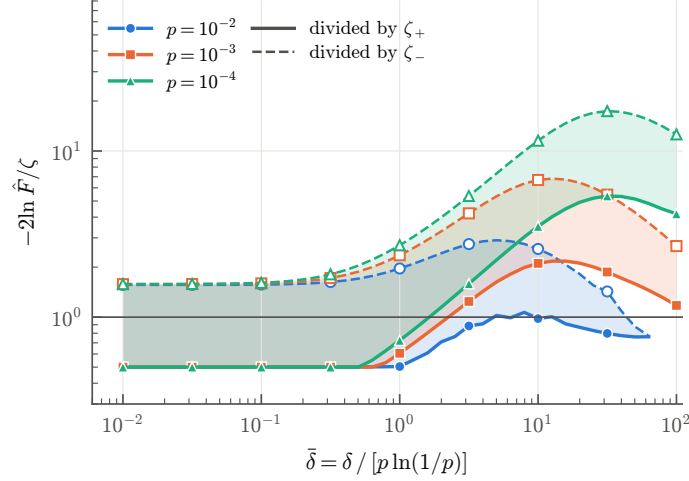


Figure 8.2: The same bracket as Fig. 1.5(a) for $-2 \ln \hat{F}$, which is a Chernoff-type quantity and not itself a Stein converse. Solid lines divide by ζ_+ , dashed lines by ζ_- .

strategy optimized for finite δ could raise ζ_- , and more copies in the Fawzi–Fawzi hierarchy could lower ζ_+ . A search over Schmidt inputs $\sqrt{q}|00\rangle + \sqrt{1-q}|11\rangle$ found the maximally entangled input optimal at three test points.

8.12 Biased Pauli noise and correlated miscalibrations

This section applies Theorem 8.4.1 beyond uniform depolarizing noise. We treat (i) a single qubit with dephasing-dominated (biased) Pauli noise, as in bias-preserving architectures [109, 121], for Z and X miscalibrations; (ii) the asymptotic adaptive Fisher information F_∞ for every Pauli channel and every Pauli-string miscalibration, in closed form; and (iii) the resulting local Stein bracket, including two qubits with a ZZ miscalibration.

Setting. A qubit Pauli channel has rates q_1, p_x, p_y, p_z with $q_1 = 1 - p_x - p_y - p_z$. Its bias ratio is $b = p_z/(p_x + p_y)$, so that depolarizing noise has $b = \frac{1}{2}$. In the examples $p_x = p_y = p_\perp = p_z/(2b)$. The gate is $\mathcal{N}_\delta = \mathcal{P}_q \circ \text{Ad}_{R_P(\delta)} \circ \text{Ad}_U$ as in Eq. (8.1.3). In the notation of Theorem 8.4.1, label the pairs $\{Q, QP\}$ by k , with rates (q_k, q'_k) , and put

$$T_k = q_k + q'_k, \quad \Gamma_k = \frac{4q_k q'_k}{q_k + q'_k}, \quad \Gamma = \sum_k \Gamma_k, \quad (8.12.1)$$

so that $\sum_k T_k = 1$, $0 \leq \Gamma_k \leq T_k$ and $(q_k - q'_k)^2 = T_k(T_k - \Gamma_k)$. Pairs with $T_k = 0$ are omitted. The pair containing the identity is $k = 1$, $\{\mathbf{1}, P\}$, with rates (q_1, q_P) ; we call q_P the *coupled rate* of the miscalibration. For a qubit the second pair consists of the two Paulis other than $\mathbf{1}$ and P . For $P = Z$ the pairs are $\{\mathbf{1}, Z\}$ and $\{X, Y\}$; for $P = X$ they are $\{\mathbf{1}, X\}$ and $\{Y, Z\}$.

Closed forms and the window

Proposition 8.12.1 (Biased qubit). *Let $s = \sin(\delta/2)$, $q_1 = 1 - p_z - 2p_\perp$, and let $x_k = (q_k - q'_k)^2 s^2 / (4q_k q'_k)$ as in Eq. (8.1.4).*

(a) For a Z miscalibration the pair $\{X, Y\}$ has $x = 0$, and for all δ

$$\begin{aligned}\hat{F} &= (1 - 2p_\perp)(1 + x_1)^{-1/2} + 2p_\perp, & D_{\text{BS}} &= 2(1 - 2p_\perp)\varphi(x_1), \\ x_1 &= \frac{(q_1 - p_z)^2 s^2}{4q_1 p_z},\end{aligned}\tag{8.12.2}$$

$$D = s^2(q_1 - p_z) \ln \frac{q_1}{p_z}, \quad F = 2p_\perp + \sqrt{(1 - 2p_\perp)^2 - (q_1 - p_z)^2 s^2}.\tag{8.12.3}$$

(b) For an X miscalibration both pairs contribute:

$$\begin{aligned}\hat{F} &= \frac{q_1 + p_\perp}{\sqrt{1 + x_1}} + \frac{p_z + p_\perp}{\sqrt{1 + x_2}}, \\ x_1 &= \frac{(q_1 - p_\perp)^2 s^2}{4q_1 p_\perp}, \quad x_2 = \frac{(p_z - p_\perp)^2 s^2}{4p_z p_\perp},\end{aligned}\tag{8.12.4}$$

$$\begin{aligned}D_{\text{BS}} &= 2(q_1 + p_\perp)\varphi(x_1) + 2(p_z + p_\perp)\varphi(x_2), \\ D &= s^2 \left[(q_1 - p_\perp) \ln \frac{q_1}{p_\perp} + (p_z - p_\perp) \ln \frac{p_z}{p_\perp} \right],\end{aligned}\tag{8.12.5}$$

$$F = \sqrt{(q_1 + p_\perp)^2 - (q_1 - p_\perp)^2 s^2} + \sqrt{(p_z + p_\perp)^2 - (p_z - p_\perp)^2 s^2}.\tag{8.12.6}$$

(c) In both cases, with the per-use Fisher informations

$$\begin{aligned}F_{\text{RLD}} &= \sum_k \frac{T_k(T_k - \Gamma_k)}{\Gamma_k}, & F_{\text{BKM}} &= \frac{1}{2} \sum_k (q_k - q'_k) \ln \frac{q_k}{q'_k}, \\ F_{\text{SLD}} &= \sum_k (T_k - \Gamma_k) = 1 - \Gamma,\end{aligned}\tag{8.12.7}$$

one has, as $\delta \rightarrow 0$,

$$-2 \ln \hat{F} = \frac{\delta^2}{4} F_{\text{RLD}} + O(\delta^4/p_\perp^2), \quad D_{\text{BS}} = \frac{\delta^2}{2} F_{\text{RLD}} + O(\delta^4/p_\perp^2),\tag{8.12.8}$$

$$D = \frac{\delta^2}{2} F_{\text{BKM}} + O(\delta^4 \ln \frac{1}{p_\perp}), \quad 1 - F = \frac{\delta^2}{8} F_{\text{SLD}} + O(\delta^4).\tag{8.12.9}$$

For a Z miscalibration $F_{\text{RLD}} = (1 - 2p_\perp)(q_1 - p_z)^2/(4q_1 p_z) = \frac{1}{4p_z}[1 + O(p_z)]$ and $F_{\text{BKM}} = \frac{1}{2}(q_1 - p_z) \ln(q_1/p_z)$. For an X miscalibration $F_{\text{RLD}} = \frac{(q_1 + p_\perp)(q_1 - p_\perp)^2}{4q_1 p_\perp} + \frac{(p_z + p_\perp)(p_z - p_\perp)^2}{4p_z p_\perp} = \frac{1}{4p_\perp}[1 + O(p_z)]$ and $F_{\text{BKM}} = \frac{1}{2}[\ln \frac{1}{p_\perp} + p_z \ln(2b)][1 + O(p_z)]$.

Proof. (a) and (b) are Theorem 8.4.1 with the pairs listed above; the maximally entangled input is optimal for D and F . (c) follows from $g(x) = \frac{x}{2} + O(x^2)$, $\varphi(x) = x + O(x^2)$, $s^2 = \frac{\delta^2}{4} + O(\delta^4)$ and $\sum_k T_k x_k = \frac{s^2}{4} \sum_k T_k (q_k - q'_k)^2/(q_k q'_k)$, together with $(q - q')^2 = T(T - \Gamma)$ and $4qq' = \Gamma T$. Every x_k is at most $s^2/(4 \min_Q qq)$, which gives the remainders. The coefficients agree with Proposition 8.6.1(v). $\square$

Proposition 8.12.2 (Window for any Pauli noise). *Let $\mathcal{N}_\delta$ be as in Eq. (8.1.3) on n qubits with $q_1 > q_P > 0$, let $T_1 = q_1 + q_P$, and let $\delta_\eta = \inf\{\delta : 1 - \hat{F}(\mathcal{N}_\delta, \mathcal{N}_0) \geq \eta\}$.*

(a) $T_1 g(x_1) \leq 1 - \hat{F} \leq T_1 g(x_1) + (1 - T_1)$, with equality on the left when every pair other than $\{1, P\}$ has $q_k = q'_k$.

(b) Put $\delta(y) = 2 \arcsin(2\sqrt{q_1 q_P y}/(q_1 - q_P))$ and $x_\eta = (1 - \eta)^{-2} - 1$. Then $\delta_\eta \leq \delta(g^{-1}(\eta/T_1))$, and, if $\eta > 1 - T_1$, $\delta_\eta \geq \delta(g^{-1}((\eta - 1 + T_1)/T_1))$. In particular

$$\delta_\eta = 4\sqrt{x_\eta q_P/q_1} [1 + O(1 - T_1) + O(q_P)].\tag{8.12.10}$$

The window is therefore set by the coupled rate q_P . For local depolarizing noise $q_P/q_1 = [p/(4-3p)]^w$, which recovers Theorem 8.5.1(c). For D_{BS} , the inequality $D_{BS} \geq 2T_1\varphi(x_1)$ gives the analogous upper bound on its window, and at fixed ratios of the rates the other pairs change the D_{BS} window only by a relative $O(1-T_1)$.

Proof. By Theorem 8.4.1, $1 - \hat{F} = \sum_k T_k g(x_k)$ with $0 \leq g < 1$; keep the first term, or bound the others by $\sum_{k \geq 2} T_k = 1 - T_1$. Since x_1 is increasing in $\delta \in [0, \pi]$, inverting g and solving $x_1 = y$ for δ gives (b). As $1 - T_1 \rightarrow 0$, both arguments of g^{-1} tend to η , and $2\sqrt{q_1 q_P}/(q_1 - q_P) = 2\sqrt{q_P/q_1} [1 + O(q_P)]$. $\square$

Remark 8.12.3 (Floor and window are set by different rates). For Z -biased noise with $p_\perp < p_z$, the Choi floor is $\tilde{\lambda} = p_\perp$ whatever the miscalibration, so the worst-case modulus (Theorem 1.4.1) is $d/(4\lambda) = 1/(4\tilde{\lambda}) = 1/(4p_\perp)$. The window of a specific miscalibration is instead set by its coupled rate. A Z miscalibration couples to p_z , the largest rate: $\delta_{0.1} = 1.937\sqrt{p_z}$, independent of the bias. An X miscalibration couples to $p_x = p_\perp$: $\delta_{0.1} = 1.937\sqrt{p_z/(2b)}$, which narrows as $b^{-1/2}$. At $p_z = 10^{-3}$ and $b = 100$ the two windows are 6.14×10^{-2} and 4.34×10^{-3} rad for $\hat{F}$, and 2.88×10^{-2} and 2.04×10^{-3} rad for D_{BS} (Fig. 1.4(a)). Pairs that do not contain the identity carry total weight $1 - T_1 \leq p_x + p_y + p_z$ and only shift the window at that order. If such a pair has exactly one vanishing rate, the two Choi matrices have different supports, and $D_{BS} = D = \zeta = \infty$ for every $\delta \neq 0$. If the coupled rate itself vanishes ($q_P = 0$), the window closes; for example, an X miscalibration under pure dephasing has $\hat{F}(\mathcal{N}_\delta, \mathcal{N}_0) = 0$ for every $\delta \neq 0$.

Asymptotic adaptive Fisher information for Pauli channels

Theorem 8.12.4 (Adaptive QFI of a Pauli-string rotation under Pauli noise). *Let $\mathcal{N}_\theta = \mathcal{P}_q \circ \text{Ad}_{R_P}(\theta) \circ \text{Ad}_U$ on n qubits, with Γ as in Eq. (8.12.1).*

- (a) *If $\Gamma = 0$, the Hamiltonian-not-in-Kraus-span condition holds and the optimal N -use Fisher information grows as N^2 . If $\Gamma > 0$ it fails, and*

$$F_\infty = \frac{1 - \Gamma}{\Gamma}, \quad F_1 = 1 - \Gamma = F_{\text{SLD}}, \quad (8.12.11)$$

where F_1 is the single-use entanglement-assisted Fisher information.

- (b) *If all $q_Q > 0$, then $F_{\text{RLD}} = \sum_k T_k^2/\Gamma_k - 1$ and*

$$\frac{F_{\text{RLD}}}{F_\infty} = \frac{\Gamma}{1 - \Gamma} \left(\sum_k \frac{T_k^2}{\Gamma_k} - 1 \right) \geq 1, \quad (8.12.12)$$

with equality if and only if Γ_k/T_k is the same for all pairs.

For a qubit with $P = Z$, Eq. (8.12.11) is Eq. (50) of Ref. [131], which reads $\Gamma = 4[p_x p_y/(p_x + p_y) + q_1 p_z/(q_1 + p_z)]$.

Proof. Since F_∞ and F_1 do not depend on the Kraus representation, take $K_Q = \sqrt{q_Q} Q U$ for each Pauli string Q with $q_Q > 0$, and $K_{QP} = \sqrt{q_Q p_P} Q P U$ for its partner, using the operator QP itself including its phase. Then $\dot{K}_Q = -\frac{i}{2}\sqrt{q_Q} Q P U$, and U drops out of every quantity below. The generator is $H = i \sum_Q K_Q^\dagger \dot{K}_Q = P/2$.

(a, scaling) $K_Q^\dagger K_{Q'}$ is proportional to the Pauli string QQ' , which is proportional to P only if $Q' \propto QP$. So $P/2$ lies in the Hermitian span of $\{K_Q^\dagger K_{Q'}\}$ if and only if some pair has both rates positive, that is, iff $\Gamma > 0$. The scaling statement is Thm. 1 of Ref. [131].

(a, value) By Thm. 2 of Ref. [131], $F_\infty = 4 \min_{h: \beta(h)=0} \|\alpha(h)\|_\infty$ with $\alpha = \sum_Q A_Q^\dagger A_Q$, $A_Q = \dot{K}_Q - i \sum_{Q'} h_{QQ'} K_{Q'}$ and $\beta = i \sum_Q K_Q^\dagger A_Q$.

Lower bound. Since $\|\alpha\|_\infty \geq \text{Tr } \alpha/d$, and the Pauli strings are orthogonal in the Hilbert–Schmidt product, $\text{Tr } \alpha/d$ is a sum of nonnegative terms. An entry $h_{QQ'}$ with $Q' \notin \{Q, QP\}$ enters $\text{Tr } \alpha/d$ only through the term $q_{Q'} |h_{QQ'}|^2$, and enters β only along Pauli strings other than $\mathbf{1}$ and P . The constant part $i \sum K^\dagger K = P/2$ has no such component. Setting these entries to zero therefore keeps $\beta = 0$ feasible and does not increase the objective, so it suffices to minimize over h supported on the pairs. Write the entries of pair k as $h_{QQ} = u$, $h_{QP, QP} = v$ and $h_{Q, QP} = (y + ie)/(2\sqrt{q_k q'_k})$. A direct computation gives that the pair contributes

$$a_k \mathbf{1} + a'_k P \text{ to } \alpha, \quad a_k = \frac{T_k}{4} + y + \frac{y^2 + e^2}{\Gamma_k} + q_k u^2 + q'_k v^2, \quad (8.12.13)$$

and $(q_k u + q'_k v) \mathbf{1} + (\frac{T_k}{2} + y)P$ to β , with $a'_k = 0$ when $u = v = 0$. For a pair with exactly one vanishing rate, y is absent and $a_k \geq T_k/4$. Hence

$$\min_{\beta=0} \frac{\text{Tr } \alpha}{d} = \min \left\{ \sum_k \left(\frac{T_k}{4} + y_k + \frac{y_k^2}{\Gamma_k} \right) : \sum_k y_k = -\frac{1}{2} \right\} = \frac{1}{4} - \frac{1}{2} + \frac{1}{4\Gamma} = \frac{1-\Gamma}{4\Gamma}, \quad (8.12.14)$$

attained at $y_k = -\Gamma_k/(2\Gamma)$ and $u = v = e = 0$.

Upper bound. At this minimizer every $a'_k = 0$, so $\alpha = \frac{1-\Gamma}{4\Gamma} \mathbf{1}$ and $\|\alpha\|_\infty = \text{Tr } \alpha/d$. This gives $F_\infty = (1-\Gamma)/\Gamma$.

Dropping the constraint $\sum_k y_k = -\frac{1}{2}$ gives $y_k = -\Gamma_k/2$, $\alpha = \frac{1-\Gamma}{4} \mathbf{1}$ and $F_1 = 4 \min_h \|\alpha\|_\infty = 1 - \Gamma$. This coincides with F_{SLD} at the maximally entangled input, by Theorem 8.4.1 and Eq. (8.12.7).

(b) The formula for F_{RLD} is Eq. (8.12.7) with $\sum_k T_k = 1$. The inequality is Cauchy–Schwarz, $\sum_k T_k^2/\Gamma_k \geq (\sum_k T_k)^2/\sum_k \Gamma_k = 1/\Gamma$, with equality iff T_k/Γ_k is constant. $\square$

Physically, F_{RLD} sees mainly the coupled rate, $F_{\text{RLD}} \simeq 1/\Gamma_1 \simeq 1/(4q_P)$, whereas $F_\infty \simeq 1/\Gamma$ is limited by every pair whose two Pauli errors both occur. As the total error $1 - q_1 \rightarrow 0$ at fixed ratios of the rates,

$$\frac{F_{\text{RLD}}}{F_\infty} = 1 + \sum_{k \geq 2} \frac{\Gamma_k}{\Gamma_1} + O(1 - q_1). \quad (8.12.15)$$

For a qubit this is $1 + \frac{q_a q_b}{(q_a + q_b) q_P}$, where q_a, q_b are the two spectator rates. The ratio is not bounded over Pauli channels: for an X miscalibration with $p_x \ll p_y = p_z$ it is $1 + p_y/(2p_x)$.

Local Stein bracket

Corollary 8.12.5 (Local Stein bracket for Pauli noise). *In the setting of Theorem 8.12.4 with all $q_Q > 0$,*

$$\liminf_{\delta \rightarrow 0} \frac{\zeta(\mathcal{N}_\delta \| \mathcal{N}_0)}{\delta^2} \geq \frac{F_\infty}{2} = \frac{1-\Gamma}{2\Gamma}, \quad 1 \leq \limsup_{\delta \rightarrow 0} \frac{D_{\text{BS}}(\mathcal{N}_\delta \| \mathcal{N}_0)}{\zeta(\mathcal{N}_\delta \| \mathcal{N}_0)} \leq \frac{F_{\text{RLD}}}{F_\infty}. \quad (8.12.16)$$

Proof. This is the argument of Proposition 8.8.1. The constant-support hypothesis holds because, for all δ , $J(\mathcal{N}_\delta)$ is unitarily equivalent to $J(\mathcal{N}_0)$, whose smallest eigenvalue is $2^n \min_Q q_Q > 0$. For every N -use strategy with outputs ρ_N^δ , one has $\zeta \geq D(\rho_N^\delta \| \rho_N^0)/N$ [41, 125], and Proposition 8.6.1(ii) gives $D = \frac{\delta^2}{2} F^{\text{BKM}} + O(\delta^3) \geq \frac{\delta^2}{2} F^{\text{SLD}} + O(\delta^3)$. Taking $N \rightarrow \infty$ over optimal strategies gives F_∞ (Theorem 8.12.4). The upper side is $\zeta \leq D_{\text{BS}} = \frac{\delta^2}{2} F_{\text{RLD}} + O(\delta^3)$ (Proposition 8.6.1(v)). $\square$

Table 8.8: Biased qubit noise with $p_z = 10^{-3}$ and $p_x = p_y = p_z/(2b)$. The first two columns after b give the windows of $\hat{F}$ ($1 - \hat{F} = 0.1$) and of D_{BS} ($D_{\text{BS}} = 0.1$). The local coefficients are $D_{\text{BS}}/\delta^2 \rightarrow F_{\text{RLD}}/2$ and $\zeta/\delta^2 \geq F_\infty/2$. All entries are exact evaluations of closed forms.

b	rotation	$\delta_{0.1}(\hat{F})$	$\delta_{0.1}(D_{\text{BS}})$	$F_{\text{RLD}}/2$	$F_\infty/2$	F_{RLD}/F_∞	F_{BKM}
$\frac{1}{2}$	Z or X	6.15×10^{-2}	2.89×10^{-2}	124.1	82.89	1.4975	3.44
10	Z	6.14×10^{-2}	2.88×10^{-2}	124.6	121.6	1.0249	3.45
10	X	1.37×10^{-2}	6.44×10^{-3}	2494	1280	1.9487	4.95
100	Z	6.14×10^{-2}	2.88×10^{-2}	124.6	124.3	1.0025	3.45
100	X	4.34×10^{-3}	2.04×10^{-3}	2.495×10^4	1.253×10^4	1.9911	6.10
1000	Z	6.14×10^{-2}	2.88×10^{-2}	124.6	124.6	1.0002	3.45
1000	X	1.37×10^{-3}	6.44×10^{-4}	2.495×10^5	1.250×10^5	1.9955	7.25

Biased qubit. For a Z miscalibration with $p_x = p_y = p_\perp$, Eq. (8.12.12) simplifies exactly to

$$\frac{F_{\text{RLD}}}{F_\infty} = T_1 + \frac{p_\perp T_1^2}{2q_1 p_z}, \quad T_1 = 1 - 2p_\perp, \quad (8.12.17)$$

which equals $(2-p)(3-2p)/(4-3p)$ (Proposition 8.8.1) at $p_\perp = p_z = p/4$. It satisfies $F_{\text{RLD}}/F_\infty < \frac{3}{2}$ whenever $p_\perp \leq p_z$ (i.e., $b \geq \frac{1}{2}$) and $q_1 > p_z$. To see this, set $t = p_\perp/p_z \in (0, 1]$ and $\sigma = 2(p_z + p_\perp) \in (0, 1)$. Then $\frac{3}{2} - F_{\text{RLD}}/F_\infty$ has the sign of $-[2t(t+1)\sigma^2 - (4t^2 + 2t - 1)\sigma + 2t^2 - 2]$, which is convex in σ and negative at $\sigma = 0$ (for $t < 1$) and at $\sigma = 1$; at $t = 1$ the bracket equals $\sigma(4\sigma - 5) < 0$. As $p_z \rightarrow 0$ at fixed bias, $F_{\text{RLD}}/F_\infty = 1 + \frac{1}{4b} + O(p_z)$. Bias toward the rotation axis thus closes the bracket.

For pure dephasing ($p_\perp = 0$) the range of $J(\mathcal{N}_\delta)$, $\text{span}\{\text{vec } \mathbf{1}, \text{vec } Z\}$, does not depend on δ . The argument of Corollary 8.12.5 applies on this support, and $D_{\text{BS}} = 2\varphi(x_1)$ exactly, so

$$\lim_{\delta \rightarrow 0} \frac{\zeta(\mathcal{N}_\delta \| \mathcal{N}_0)}{\delta^2} = \frac{F_\infty}{2} = \frac{F_{\text{RLD}}}{2} = \frac{(1 - 2p_z)^2}{8p_z(1 - p_z)}. \quad (8.12.18)$$

Here the local Stein coefficient is known exactly, and it coincides with the curvature of D_{BS} .

For an X miscalibration under the same noise the spectator pair $\{Y, Z\}$ has $\Gamma_2 \simeq 4p_\perp$, comparable to $\Gamma_1 \simeq 4p_\perp$. Equation (8.12.15) then gives $F_{\text{RLD}}/F_\infty \rightarrow (1 + 4b)/(1 + 2b) \in [\frac{3}{2}, 2)$, so the ratio $3/2$ is exceeded for every $b > \frac{1}{2}$. The exact ratio stays below 2 on a 30-digit grid of 22 952 points covering $p_\perp/p_z \in [10^{-6}, 1]$ and $2(p_z + p_\perp) \in [10^{-7}, 0.9999]$, with maximum 1.9999988; we have no proof of this bound. In this case $\zeta/\delta^2 \geq F_\infty/2 \simeq 1/(16p_\perp)$ for $b \gg 1$, so the narrow window of Remark 8.12.3 also reflects a genuine sensitivity of ζ , with D_{BS} larger by at most a factor 2. Table 8.8 lists values.

Correlated miscalibrations. For local depolarizing noise and P of weight w , the quantities of Theorem 8.12.4 depend on w but not on n . The reason is that qubits outside the support of P multiply both rates of a pair by the same factor. As $p \rightarrow 0$,

$$F_{\text{RLD}} = \frac{1}{4}(4/p)^w[1 + O(p)], \quad F_\infty = \Theta(p^{-\lceil w/2 \rceil}), \quad F_{\text{RLD}}/F_\infty = \Theta(p^{-\lfloor w/2 \rfloor}). \quad (8.12.19)$$

The leading term of Γ comes from the pairs $\{Q, QP\}$ in which Q and QP each have weight about $w/2$ on the support of P . For example, for $P = Z \otimes Z$ the weight-one errors $Z \otimes \mathbf{1}$ and $\mathbf{1} \otimes Z$ form a pair with $\Gamma_k \simeq p/2$, whereas the coupled pair $\{\mathbf{1}, ZZ\}$ has $\Gamma_1 \simeq p^2/4$. For two qubits this gives

$$F_\infty = \frac{2}{p}[1 + O(p)], \quad F_{\text{RLD}} = \frac{4}{p^2}[1 + O(p)], \quad \frac{F_{\text{RLD}}}{F_\infty} = \frac{2}{p}[1 + O(p)], \quad (8.12.20)$$

so Corollary 8.12.5 pins ζ/δ^2 only to $[1/p, 2/p^2]$ to leading order. Table 8.9 gives the limits for $w \leq 4$. Under Z -biased local noise ($p_z = 10^{-3}$, $b = 100$) a ZZ miscalibration gives

Table 8.9: Leading coefficients of F_∞ , F_{RLD} and their ratio for local depolarizing noise and a miscalibration $Z^{\otimes w}$ (any $n \geq w$). Values are exact closed forms at $p = 10^{-5}$, scaled as indicated; the limits are $2/3, 1, 3/2$ ($w = 1$), $2, 4, 2$ ($w = 2$), $2/3, 16, 24$ ($w = 3$) and $8/3, 64, 24$ ($w = 4$).

w	$p^{\lceil w/2 \rceil} F_\infty$	$p^w F_{\text{RLD}}$	$p^{\lfloor w/2 \rfloor} F_{\text{RLD}}/F_\infty$
1	0.666658	0.999983	1.499994
2	1.999945	3.999880	1.999995
3	0.666667	15.99928	23.99892
4	2.666529	63.99616	23.99980

$F_\infty = 498.5$ and $F_{\text{RLD}}/F_\infty = 499.5 \simeq 1/(2p_z)$. This is controlled by the weight-one Z errors and is no better than in the depolarizing case.

Whether the true local coefficient of ζ for a weight-two error follows $F_\infty/2 \sim 1/p$ or $F_{\text{RLD}}/2 \sim 2/p^2$ is open. Explicit sequential strategies give $\Theta(1/p)$, but their relative-entropy rate is not known to be optimal. Until this is settled, the $p^{w/2}$ window of D_{BS} for $w \geq 2$ (Theorem 8.5.1) is not known to be a property of ζ .

Numerical verification.

- The closed forms of Proposition 8.12.1 agree with direct 40-digit evaluations from 4×4 Choi matrices to within 1.2×10^{-33} relative. The checks covered 24 points with Z and X miscalibrations, symmetric and asymmetric transverse rates, and Choi floors down to 4×10^{-8} . Maximally entangled inputs beat 40 random inputs for D and F .
- All 60 exact windows (40-digit bisection) lie within the bounds of Proposition 8.12.2, and their ratio to Eq. (8.12.10) lies in $[1.006, 1.024]$, $[1.0006, 1.0024]$ and $[1.00006, 1.00024]$ at $p_z = 10^{-2}$, 10^{-3} and 10^{-4} .
- For the qubit we certified F_∞ at 29 biased points, including asymmetric $p_x \neq p_y$ and pure dephasing, with the method of Proposition 8.8.1: a 50-digit minimax lower bound over the full h , and $\|\alpha\|_\infty$ at the minimizer. The two agree with each other and with Eq. (8.12.11) to within 10^{-46} relative, and the double-precision interior-point solution agrees to within 2×10^{-7} .
- For two qubits (16 Kraus operators, input dimension 4) we used double-precision interior-point solutions; the within-pair minimizer evaluated in 40 digits, for which $\beta = 0$ to 10^{-36} and $\alpha \propto \mathbf{1}$; and a double-precision lower bound over the full 16×16 matrix h . These agree with Eq. (8.12.11) to within 5×10^{-6} , 4×10^{-36} and 2×10^{-12} relative, respectively, for ZZ at $p = 10^{-1}, \dots, 10^{-4}$, for $Z \otimes \mathbf{1}$, and for ZZ and XX under biased noise.
- In every case F_{RLD} computed as $\lambda_{\max} \text{Tr}_Y[J J^{-1} J]$ agrees with Eq. (8.12.7) to within 10^{-33} , and $\text{Tr}_Y[\cdot]$ is proportional to $\mathbf{1}$.

8.13 Certification

Proof of Theorem 1.7.1. The estimates are assumed to be channels, so $\Phi - \hat{\Phi}$ and $\Psi - \hat{\Psi}$ are trace annihilating and the constant $d_X/2$ of Lemma 8.2.1 applies. On the success event, Lemma 8.2.1 gives $\|J(\Phi) - J(\hat{\Phi})\|_\infty \leq r\hat{\lambda}$, so Weyl's inequality gives $\lambda(\Phi) \geq (1 - r)\hat{\lambda}$, and

$J(\hat{\Phi}) \geq \hat{\lambda} \mathbf{1}$ gives $(1-r)J(\hat{\Phi}) \leq J(\Phi) \leq (1+r)J(\hat{\Phi})$; likewise for Ψ . Joint monotonicity and degree-one homogeneity of $(A, B) \mapsto A \#_s B$ give $J(\Psi) \#_s J(\Phi) \geq (1-r)J(\hat{\Psi}) \#_s J(\hat{\Phi})$ for every s , hence $\hat{Q}_s(\Phi \|\Psi) \geq (1-r)\hat{Q}_s(\hat{\Phi} \|\hat{\Psi})$, and the minimum over s gives Eq. (1.7.1). The inequality $\xi \leq \hat{C}$ is Theorem 7.4.1 in Chapter 7, and Corollary 8.15.5 gives the finite- n bound with $\Lambda \leq d_T(\hat{\Phi}, \hat{\Psi}) + 2 \ln \frac{1}{1-r}$. $\square$

Stein version. On the same event, Eq. (8.3.2) at the base pair $(\hat{\Phi}, \hat{\Psi})$ with $\epsilon' \leq \epsilon/2$, together with $\zeta(\hat{\Phi} \|\hat{\Psi}) \leq D_{\text{BS}}(\hat{\Phi} \|\hat{\Psi})$, gives

$$\zeta(\Phi \|\Psi) \leq D_{\text{BS}}(\hat{\Phi} \|\hat{\Psi}) + \ln \frac{1}{1-r} + \epsilon \ln d_Y + \left(1 + \frac{\epsilon}{2}\right) h\left(\frac{\epsilon}{2+\epsilon}\right) + \frac{\epsilon}{2} \ln \kappa(\hat{\Psi}), \quad (8.13.1)$$

in which every term is computable from the estimates.

Sharpness of the radius. By Eq. (1.4.3), there are model pairs and members of the diamond ball of radius ϵ whose values of $-\ln \hat{F}$ differ by $\frac{d_X \epsilon}{4\tilde{\lambda}}(1 - o(1))$. A diamond-ball certificate with additive accuracy η therefore needs $\epsilon \leq 4\eta\tilde{\lambda}(1 + o(1))$.

Proposition 8.13.1 (Cost through a diamond ball). *(a) With a known lower bound $\tilde{\lambda}_0 \leq \tilde{\lambda}$ and diamond-norm tomography [93] at $\epsilon = \eta\tilde{\lambda}_0$, Theorem 1.7.1 certifies $\hat{C}$ and $-\ln \hat{F}$ to additive accuracy $\eta \leq 1$ with $N = O((d_X d_Y k + d_X \ln(1/\alpha))/(\eta\tilde{\lambda}_0)^2)$ uses per channel, where $1 - \alpha$ is the confidence level. This learner processes the channel outputs coherently. (b) Consider learners that prepare a fresh input (with ancilla) for each use and measure the output, possibly choosing later inputs from earlier outcomes. Any such learner that outputs $\hat{\Lambda}$ with $\|\hat{\Lambda} - \mathcal{N}_{p,\theta}\|_\diamond \leq \epsilon$ with probability at least $2/3$ for all θ needs $N \geq 0.0589(1-p)^2 \epsilon^{-2}(1 - o(1))$ uses.*

Proof of (b). Two hypotheses θ, θ' with $\|\mathcal{N}_{p,\theta} - \mathcal{N}_{p,\theta'}\|_\diamond = 2(1-p) \sin \frac{|\theta - \theta'|}{2} = 2\epsilon(1 + o(1))$ must be distinguished with errors at most $1/3$. For classically adaptive kernels, the Bhat-tacharyya coefficient of the outcome records is at least the product of the per-round minima, each of which is at least the channel root fidelity, equal to its Choi value by Theorem 8.4.1, i.e., at least $1 - \sin^2 \frac{\theta - \theta'}{2}$. Errors at most $1/3$ require a coefficient at most $\sqrt{8/9}$, which gives the bound with $0.0589 = \frac{1}{2} \ln \frac{9}{8}$. $\square$

Remark 8.13.2 (Learner classes). Part (a) is achieved by a coherent learner and part (b) bounds incoherent learners, so Proposition 8.13.1 does not establish a cost of order $1/(\eta\tilde{\lambda})^2$ within a single learner class. For coherent learners the two-point argument on the family $\mathcal{N}_{p,\theta}$ gives much less: with error-corrected probing the Fisher information per use for θ is at most $F_\infty = 2(1-p)^2/[p(3-2p)] \approx 2/(3p)$ (Sec. 8.8), so distinguishing θ, θ' with $|\theta - \theta'| \approx \epsilon/(1-p)$ requires only $N \gtrsim 1/(F_\infty \epsilon^2) \approx 3p/(2\epsilon^2)$, which at $\epsilon = 4\eta\tilde{\lambda} = \eta p$ is of order $1/(\eta^2 p)$, the order of the intrinsic cost of Proposition 8.13.3. Whether diamond-ball certification costs order $1/(\eta\tilde{\lambda})^2$ uses for coherent learners is open.

Proposition 8.13.3 (Intrinsic cost). *(a) Let $\Psi_0 = \mathcal{D}_p^{\otimes n} \circ \text{Ad}_U$, and let Ψ_1 be the Pauli channel obtained by multiplying q_P ($\text{wt} P = n$, so $q_P = \tilde{\lambda}$) by $[(1+\eta)/(1-\eta)]^2$ and decreasing q_1 accordingly. Any N -use strategy, adaptive or coherent and with arbitrary quantum memory, that estimates $\hat{F}(\text{Ad}_{PU}, \Psi)$ to relative accuracy better than η (i.e., with $|\hat{f} - \hat{F}| < \eta \hat{F}$ for its estimate $\hat{f}$) with probability at least $2/3$ under both hypotheses needs $N \geq c_0/(\eta^2 \tilde{\lambda})$, with $c_0 = 0.0246$ for $\eta \leq 0.1$, $p \leq 0.04$. (b) Preparing Bell pairs, applying the channel, undoing the known U and measuring in the Bell basis, $N = \lceil 3 \ln(2/\alpha)/(\eta^2 \tilde{\lambda}) \rceil$ uses estimate q_P , hence $\hat{F}(\text{Ad}_{PU}, \Psi) = \sqrt{q_P}$, to relative accuracy $(1 \pm \eta)$ with probability at least $1 - \alpha$.*

Table 8.10: Converse side: total channel uses per channel for a certified lower bound on $\hat{F}$ (and, for (b), an upper bound on D_{BS}) at accuracy $\eta = 0.05$, median run, with fitted exponents a in $N \propto p^{-a}$ and bootstrap 95% intervals. The last row is the reference scale $c_0/(\eta^2 \tilde{\lambda})$ of Proposition 8.13.3(a), which is proved for the estimand $\hat{F}(\text{Ad}_{PU}, \Psi)$ at success probability $2/3$, not for the simulated pair with the median-run criterion.

procedure	$p = 10^{-2}$	3×10^{-3}	10^{-3}	3×10^{-4}	10^{-4}	a
(a) diamond ball, Thm. 1.7.1, $\hat{F}$	3.6×10^{11}	4.0×10^{12}	3.6×10^{13}	4.0×10^{14}	3.6×10^{15}	2.000
(b) Pauli rates, $\hat{F}$	1.2×10^6	4.3×10^6	1.2×10^7	3.9×10^7	1.2×10^8	1.00 [0.99, 1.03]
(b) Pauli rates, D_{BS}	1.0×10^7	3.5×10^7	1.0×10^8	3.5×10^8	1.0×10^9	1.00 [0.97, 1.03]
reference $c_0/(\eta^2 \tilde{\lambda})$	3.9×10^3	1.3×10^4	3.9×10^4	1.3×10^5	3.9×10^5	1

Proof. (a) A Pauli channel is a mixture of unitaries, so the final state of any strategy is $\Lambda(q^{\otimes N})$ for a channel Λ independent of q . Under the two hypotheses $\hat{F} = \sqrt{qP}$ and $\hat{F} = \sqrt{qP}(1 + \eta)/(1 - \eta)$, so the two open acceptance intervals $\{|\hat{f} - \hat{F}| < \eta \hat{F}\}$ are disjoint, and a successful estimator yields a test with both errors at most $1/3$, and data processing gives $N \text{KL}(q^{(0)} \| q^{(1)}) \geq \frac{\ln 2}{3}$. The minimum of $N\eta^2 \tilde{\lambda}$ over $\eta \in [0.005, 0.1]$ and $p \in [10^{-6}, 0.04]$ is 0.02465 ; analytically $\text{KL} \leq \chi^2 \leq 16\eta^2 \tilde{\lambda}(1 + O(\eta + \tilde{\lambda}))$. The rank-one formula of Sec. 8.3 gives $\hat{F}(\text{Ad}_{PU}, \Psi_j) = \sqrt{q_P^{(j)}}$. (b) Multiplicative Chernoff bound for a binomial with mean NqP . $\square$

The perturbation in (a) changes the Pauli rates in the low-probability sector, where the information per use is $(\Delta q)^2/q \approx \eta^2 \tilde{\lambda}$, not $\eta^2 \tilde{\lambda}^2$. A two-point lower bound of order $1/(\eta \tilde{\lambda})^2$ for the certification task itself is therefore false at this instance; the quadratic cost belongs to certification through a diamond ball. Hypothesis-testing quantities are different: the hockey-stick channel divergence satisfies $|E_\gamma^\diamond(\Phi' \| \Psi') - E_\gamma^\diamond(\Phi \| \Psi)| \leq \frac{1}{2} \|\Phi - \Phi'\|_\diamond + \frac{\gamma}{2} \|\Psi - \Psi'\|_\diamond$, so additive accuracy η needs only $\epsilon \leq 2\eta/(1 + \gamma)$, independently of λ .

Conjecture 8.13.4. *For every channel pair with floor $\tilde{\lambda}$, the optimal number of uses for certifying $\hat{C}$ or $\hat{F}$ to relative accuracy η is $\Theta(\text{poly}(d)/(\eta^2 \tilde{\lambda}))$.*

Simulations. We simulated standard single-qubit Pauli process tomography (six input states, three measurement bases, 18 binary settings, exact binomial sampling) for the pair $\mathcal{N}_{p,\delta}, \mathcal{N}_{p,0}$ with $U = H$ and $\delta = 2\sqrt{p}$, where $\hat{F} \approx 0.71$. All confidence regions hold simultaneously with probability at least 0.95 (Hoeffding or Clopper–Pearson bounds with a union bound).

Converse side. A converse is certified by a lower confidence bound on $\hat{F}$ (an upper bound on $\hat{C} = -\ln \hat{F}$ for this pair) and an upper confidence bound on D_{BS} . We compared two procedures. (a) Linear inversion, which is exactly trace preserving, with the Hoeffding trace-norm ball $\|J_i - J_{\text{LI},i}\|_1 \leq \epsilon_{\text{tr}}$; Theorem 1.7.1 then gives $\hat{F} \geq (1 - r)\hat{F}(J_{\text{LI},1}, J_{\text{LI},0})$ with $r = \epsilon_{\text{tr}}/(2\tilde{\lambda})$, and the bound is vacuous if $r \geq 1$. (b) Direct estimation of the Pauli error rates in the known frame (three settings per channel: prepare $U^\dagger|P_+\rangle$ and measure P), with Clopper–Pearson intervals defining a polytope of Pauli channels. Because $\hat{F}$ is jointly concave and D_{BS} jointly convex while the Choi pair is affine in the rates, the minimum of $\hat{F}$ and the maximum of D_{BS} over the product polytope are attained at pairs of vertices and were computed exactly by vertex enumeration. Procedure (b) assumes that the noise is a Pauli channel in the known frame. The accuracy target was $\ln(\hat{F}_{\text{true}}/\hat{F}_{\text{LCB}}) \leq \eta$ in the median run, on a grid of $1/8$ decade with 100 runs (a) or 50 runs (b) per point.

The ratio of (a) to (b) is about $750/\tilde{\lambda}$ at every p . The p^{-2} of (a) holds by construction: the requirement $r = \epsilon_{\text{tr}}/(2\lambda_{\min}) \leq \eta$ with $\lambda_{\min} = p/2$, $\epsilon_{\text{tr}} = 36t$ and the Hoeffding width

$t = [\ln(1440)/(2N_{\text{set}})]^{1/2}$ alone gives $N = 18 \cdot 36^2 \ln(1440)/(2\eta^2 p^2) \approx 3.4 \times 10^7/p^2$ uses per channel, against $3.6 \times 10^7/p^2$ observed. The fitted exponents in $N \propto \eta^{-b}$ over $\eta \in \{0.1, 0.05, 0.025\}$ are $b = 1.94$ for (a) and 1.65–1.76 for (b), below the asymptotic value 2 because these widths are pre-asymptotic. Procedure (a) lies a constant factor of about 1.5×10^6 above the incoherent-learner bound of Proposition 8.13.1(b) at the sharp radius $\epsilon = 4\eta\tilde{\lambda}$, which reflects the constants of a non-optimal learner, not a different scaling. Procedure (b) assumes that both channels are Pauli channels in the known frame. On the converse side, the exact lower confidence bound over the model-free Clopper–Pearson box (a'') below is a concave minimization that we did not attempt, so whether model-free tomography attains the $1/\tilde{\lambda}$ law for certifying the converse is open. Procedure (b) lies a p -independent factor of about 300 above the reference scale $c_0/(\eta^2\tilde{\lambda})$; since that scale is proved for a different estimand and success criterion, this factor is not a gap to a lower bound for the simulated pair. Coverage, measured in 1000 runs per p at the budget of each procedure, was 1.000 (Clopper–Pearson interval $[0.996, 1]$) for every certified bound; the median certified $\hat{F}$ was 0.67–0.68 against a true value of 0.71. At $\delta = 1$, where $E_1 \approx 0.48$ and $F \approx 0.88$, the hockey-stick divergence (upper bound $\frac{1}{2}\|\hat{\Phi} - \hat{\Psi}\|_\diamond + \epsilon_{\text{tr}}$) and the Uhlmann fidelity (lower bound through the purified distance) were certified through the same ball with 3.4×10^7 and 5.9×10^{10} uses, independently of p (fitted slopes 0.000 and -0.008).

Opposite side. Upper bounds on $\hat{F}$ and lower bounds on D_{BS} , which limit distinguishability from below, are convex programs over the confidence region; we solved them with a safety margin of 10^{-5} above the solver tolerance, using the geometric Rényi divergence of order $7/8$ (monotone in the order) for D_{BS} . We used four regions: (a) the trace-norm ball, (a') a Hoeffding box on the 18 setting probabilities, (a'') a Clopper–Pearson box, i.e., exact binomial intervals per setting with no model assumption, and (b) the known-frame Pauli polytope. With the criterion that the certified divergence lies within a factor two of the truth (median of 10 runs), the total uses per channel at $p = 10^{-3}$ were 4.3×10^{10} , 1.3×10^9 , 9.4×10^6 and 2.4×10^5 , with fitted exponents 1.98, 1.93, 0.96 and 0.99 over $p \in [10^{-4}, 10^{-2}]$, the same range as Table 8.10. A variance-adaptive region thus removes the extra factor $1/\tilde{\lambda}$ without any model assumption, while the variance-blind regions (a) and (a') do not. For the box and ball regions, 1000 repetitions at $(p, N) = (10^{-2}, 10^6)$, $(10^{-3}, 10^8)$ and $(10^{-4}, 10^{10})$ shots per setting gave coverage 1.000 against a nominal 0.95. At the same sample sizes, and at $(p, N) = (10^{-3}, 10^{10})$, the plug-in $\hat{F}$ from linear inversion exceeded the truth, which makes the converse claim a stronger limit than is true, in 21.8%, 8.2%, 0.6% and 31.9% of runs. A hedged maximum-likelihood estimate [22] with $\beta = 1/2$ did so in up to 96% of runs at 10^2 – 10^3 shots per setting ($p \leq 10^{-3}$). A route that uses only the different-sums continuity bound of Theorem 4.10.4 in Chapter 4 with a Weyl-certified floor did not become non-vacuous below 10^{12} shots per setting.

Two qubits. For a CNOT gate with a $Z \otimes Z$ miscalibration under local noise, plug-in estimation of $\hat{F}$ to 10% accuracy in 90% of runs (product Pauli tomography, linear inversion with projection onto channels) needed 10^6 , 10^8 and 10^{10} shots per setting at $p = 0.1$, 0.03 and 0.01, i.e., $N \approx p^{-4}$ on a one-decade grid with three points, consistent with $1/\lambda^2$ and $\lambda = p^2/4$. These are plug-in estimates, not certified bounds.

8.14 Numerical methods

Geometric means were computed through the Arveson–Radon–Nikodym factorization of the pair, which is stable when both Choi matrices are invertible, and cross-checked against semidefinite programs. Closed forms were verified against direct matrix computations in 40–60-digit arithmetic (`mpmath`); semidefinite programs used `cvxpy` with Clarabel, with

SCS as a cross-check. Diamond norms used the Watrous semidefinite program [127]. Whenever a Choi eigenvalue fell below 10^{-8} , finite differences were evaluated in multiple-precision arithmetic, because square roots of round-off-level eigenvalues corrupt double-precision derivatives. The code, random seeds and raw data are in the `code/` directory of the repository that accompanies this book, <https://github.com/mjones1298/Continuity-of-divergences-python->; its README.md lists the script and output file behind each number of this chapter and of Chapter 1.

8.15 Refinements of results from Chapters 4, 5 and 7

The continuity of the geometric quantities, the Thompson-metric and first-argument continuity of the adaptive Stein exponent, the geometric Chernoff converse and the program-simulation converse, which Sections 1.3–1.7 of Chapter 1 and Sections 8.3, 8.7 and 8.13 of this chapter use, are proved in Chapters 4, 5 and 7. This section points to them and proves, for finite-dimensional channels, the refinements that are needed here and are not proved there. Maps act from $L(\mathcal{X})$ to $L(\mathcal{Y})$ and are completely positive (CP), and channels when stated. We use $\text{Tr}_Y J(\Lambda) = \Lambda^*(\mathbf{1})^T$, so that $\text{Tr}_Y J(\Phi) = \mathbf{1}$ for a channel, and $\text{Tr}_Y |\text{vec } A\rangle\langle\text{vec } B| = (B^\dagger A)^T$.

Kubo–Ando facts. For $s \in (0, 1)$ and positive semidefinite (PSD) A, B we use only the following properties of $\#_s$ [79]. (K1) $\#_s$ is monotone in each argument. (K2) $(aA)\#_s(bB) = a^{1-s}b^s A\#_s B$ for $a, b > 0$. (K3) $X(A\#_s B)X^\dagger = (XAX^\dagger)\#_s(XBX^\dagger)$ for invertible X , and $A\#_s B = A^{1-s}B^s$ for commuting $A, B > 0$. (K4) $\#_s$ is continuous on pairs of positive definite operators, and $A_n\#_s B_n \rightarrow A\#_s B$ whenever $A_n \downarrow A$ and $B_n \downarrow B$. (K5) $A\#_s B \leq (1-s)A + sB$ and $A\#_s A = A$. By the defining formula and (K4), the mean of two block-diagonal pairs is block diagonal. With the convention of Sec. 8.1, $\hat{Q}_0 = \hat{Q}_1 = 1$, and $\hat{Q}_s \leq 1$ for channels (proof of Theorem 7.6.4(b)), so $e^{-\hat{C}} = \inf_{0 < s < 1} \hat{Q}_s$; we use this form throughout.

Derivatives and levels. Let Φ, Ψ be CP maps and fix an injective W with $WW^\dagger = J(\Phi + \Psi)$, whose columns are linearly independent Kraus operators of $\Phi + \Psi$. The derivative T of Φ with respect to $\Phi + \Psi$ (Section 4.2) satisfies $0 \leq T \leq \mathbf{1}$, $J(\Phi) = WTW^\dagger$ and $J(\Psi) = W(\mathbf{1} - T)W^\dagger$. With $b_s(t) = t^s(1-t)^{1-s}$ and $b_s(0) = b_s(1) = 0$ for $0 < s < 1$, Theorem 4.2.15 gives $J(\Psi)\#_s J(\Phi) = W b_s(T) W^\dagger$, and Theorem 4.5.6 gives $\text{ran}(A\#_s B) = \text{ran } A \cap \text{ran } B$ for PSD A, B . The levels τ_ℓ of a pair of channels, the operators S_ℓ and B_s , the constant $\Lambda \leq d_T(\Phi, \Psi)$ and the identity $\hat{Q}_s(\Phi\|\Psi) = \lambda_{\min}(B_s)$ for $0 < s < 1$ are those of Section 7.2 (Lemma 7.2.1 and Remark 7.2.2, where b_s is written g_s).

Continuity of the geometric quantities and the scope of the dichotomy

$\hat{Q}_s$ and $\hat{F}$ are upper semicontinuous on pairs of CP maps (Lemma 4.12.1(c)), and hence so is $e^{-\hat{C}} = \inf_{0 < s < 1} \hat{Q}_s$; in particular $\hat{F}$ and $\hat{Q}_s$ are continuous at every pair at which they vanish. As functions on pairs of CP maps, $\hat{F}$ and $\hat{Q}_s$ are continuous at (Φ, Ψ) if and only if $J(\Phi + \Psi)$ is invertible or $\hat{F}(\Phi, \Psi) = 0$ (Theorem 4.12.4(a) and (d)(ii)). Restricted to pairs of channels, $\hat{F}$ is continuous at (Φ, Φ) if and only if $J(\Phi)$ is invertible, and if $J(\Phi)$ is singular the output rotations $\Phi_t = \text{Ad}_{e^{-itH}} \circ \Phi$ witness the discontinuity for suitable Hermitian H (Corollary 4.12.10 and Theorem 4.12.9(b)); whether the ‘only if’ direction holds at off-diagonal channel pairs with $J(\Phi + \Psi)$ singular and $\hat{F}(\Phi, \Psi) > 0$ is open (Section 4.13). For unitaries, $\hat{F}(\text{Ad}_U, \text{Ad}_{U'}) = 0$ whenever $U' \notin \mathbb{C}U$, while $\hat{F}(\text{Ad}_U, \text{Ad}_U) = 1$ (Corollary 4.12.6). For amplitude damping, $\text{ran } J(\mathcal{A}_\gamma) \cap \text{ran } J(\mathcal{A}_{\gamma'}) \subseteq \text{span}\{|\text{vec } 0\rangle\langle 1|\}$ and $\hat{F}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) = 0$ for $\gamma \neq \gamma'$ in $[0, 1]$ (Example 4.12.8), and $\hat{C} = -2 \ln \hat{F} = C_{\text{bar}} = +\infty$ for

$0 < \gamma, \gamma' \leq 1$ (Proposition 7.6.8). For pairs of channels, $\hat{C}$ is lower semicontinuous everywhere and continuous wherever $J(\Phi + \Psi)$ is invertible (Theorem 7.6.4(a,c)). The following proposition extends the last statement to completely positive maps.

Proposition 8.15.1. *Let Φ, Ψ be CP maps and $0 < s < 1$. If $J(\Phi + \Psi)$ is invertible, then $\hat{Q}_s, \hat{F}$ and $\hat{C}$ are continuous at (Φ, Ψ) .*

Proof. Near (Φ, Ψ) , $S' = J(\Phi' + \Psi')$ is invertible, and Theorem 4.2.15 with $W = S'^{1/2}$ gives $T' = S'^{-1/2}J(\Phi')S'^{-1/2}$, which is continuous in the pair, and $\hat{Q}_s(\Phi' \parallel \Psi') = \lambda_{\min} \text{Tr}_{\mathcal{Y}}(S'^{1/2}b_s(T')S'^{1/2})$. This is continuous for fixed s . For $\hat{C}$, since $e^{-\hat{C}}$ is upper semicontinuous, it remains to show $\liminf e^{-\hat{C}(\Phi', \Psi')} \geq e^{-\hat{C}(\Phi, \Psi)}$. Choose $\eta \in (0, \frac{1}{4})$ such that T has no eigenvalue in $(0, 2\eta) \cup (1 - 2\eta, 1)$, and let k_s equal b_s on $[2\eta, 1 - 2\eta]$, vanish on $[0, \eta] \cup [1 - \eta, 1]$ and be affine in between. Since b_s is concave and nonnegative, $0 \leq k_s \leq b_s$, and $k_s(T) = b_s(T)$. The family $\{k_s\}_{0 < s < 1}$ is uniformly bounded and equicontinuous, because $(s, t) \mapsto b_s(t)$ is uniformly continuous on $[0, 1] \times [\eta, 1 - \eta]$, so it is totally bounded in the supremum norm (Arzelà–Ascoli). Approximating by a finite net shows $\sup_s \|k_s(T') - k_s(T)\|_{\infty} \rightarrow 0$ as $T' \rightarrow T$. Hence $\hat{Q}_s(\Phi' \parallel \Psi') \geq \lambda_{\min} \text{Tr}_{\mathcal{Y}}(S'^{1/2}k_s(T')S'^{1/2}) \rightarrow \lambda_{\min} \text{Tr}_{\mathcal{Y}}(S^{1/2}b_s(T)S^{1/2}) = \hat{Q}_s(\Phi \parallel \Psi)$ uniformly in s , and taking the infimum over s gives the claim. $\square$

The two-sided Choi-floor law with separate floors

Proposition 8.15.2. *Let Φ, Φ', Ψ, Ψ' be CP maps. Put $r_{\Phi} = \|J(\Phi') - J(\Phi)\|_{\infty}/\lambda(\Phi)$ if $\lambda(\Phi) > 0$ and $r_{\Phi} = 0$ if $\Phi' = \Phi$, define r_{Ψ} likewise, and assume $r_{\Phi}, r_{\Psi} < 1$. Then for $0 < s < 1$*

$$(1 - r_{\Psi})^{1-s}(1 - r_{\Phi})^s \hat{Q}_s(\Phi \parallel \Psi) \leq \hat{Q}_s(\Phi' \parallel \Psi') \leq (1 + r_{\Psi})^{1-s}(1 + r_{\Phi})^s \hat{Q}_s(\Phi \parallel \Psi). \quad (8.15.1)$$

In particular

$$\sqrt{(1 - r_{\Phi})(1 - r_{\Psi})} \leq \frac{\hat{F}(\Phi', \Psi')}{\hat{F}(\Phi, \Psi)} \leq \sqrt{(1 + r_{\Phi})(1 + r_{\Psi})},$$

$$|\hat{C}(\Phi', \Psi') - \hat{C}(\Phi, \Psi)| \leq \ln \frac{1}{1 - \max\{r_{\Phi}, r_{\Psi}\}}.$$

With $\lambda = \min\{\lambda(\Phi), \lambda(\Psi)\}$ and η_J as in Sec. 8.3, the equal-floor bounds $1 \pm \eta_J/\lambda$, for $\hat{F}$ and for every Kubo–Ando connection, are Theorem 4.8.8(a,b). For channels, $\eta_J/\lambda \leq r = d_X \epsilon / (2\lambda)$ with $\epsilon = \max\{\|\Phi' - \Phi\|_{\diamond}, \|\Psi' - \Psi\|_{\diamond}\}$, by Lemma 8.2.1.

Proof. $J(\Phi) \geq \lambda(\Phi)\mathbf{1}$ gives $(1 - r_{\Phi})J(\Phi) \leq J(\Phi) - r_{\Phi}\lambda(\Phi)\mathbf{1} \leq J(\Phi') \leq J(\Phi) + r_{\Phi}\lambda(\Phi)\mathbf{1} \leq (1 + r_{\Phi})J(\Phi)$, and likewise for Ψ . Apply (K1) and (K2), and use that $Z \mapsto \lambda_{\min} \text{Tr}_{\mathcal{Y}} Z$ is monotone and positively homogeneous. $\square$

For $\Phi' = \Phi$ this is Eq. (1.4.2). Since $\hat{F}$ is symmetric, Eq. (1.4.3) shows that each of the two terms is sharp separately, when the other channel is unitary and therefore has floor 0. Whether the constant $d_X/(2\lambda)$ for $\ln \hat{F}$ is attained when both channels move and both floors are at least λ is open.

The Thompson metric and the adaptive Stein exponent

An n -use adaptive strategy consists of a state ρ_1 on $\mathcal{X} \otimes \mathcal{R}_1$, channels $\Gamma_k : \mathcal{L}(\mathcal{Y} \otimes \mathcal{R}_k) \rightarrow \mathcal{L}(\mathcal{X} \otimes \mathcal{R}_{k+1})$ for $k < n$, and a final measurement on $\mathcal{Y} \otimes \mathcal{R}_n$. Its final states are $\rho_n^{\Phi} = (\Phi \otimes$

$\text{id})\Gamma_{n-1} \cdots \Gamma_1(\Phi \otimes \text{id})(\rho_1)$ and ρ_n^Ψ . We use $\zeta(\Phi\|\Psi) = D^{\text{reg}}(\Phi\|\Psi) = \lim_n \frac{1}{n} D(\Phi^{\otimes n}\|\Psi^{\otimes n}) = D^A(\Phi\|\Psi)$ [41, 125]. Here D of channels is the Umegaki channel divergence, and $D^A(\Phi\|\Psi) = \sup[D((\text{id} \otimes \Phi)\rho\|(\text{id} \otimes \Psi)\sigma) - D(\rho\|\sigma)]$ is the amortized divergence, with the supremum over finite-dimensional $\mathcal{R}$ and states ρ, σ on $\mathcal{R} \otimes \mathcal{X}$ with $D(\rho\|\sigma) < \infty$. We write $\kappa(\Psi) = \lambda_{\max}(J(\Psi))/\lambda_{\min}(J(\Psi))$ and h for the binary entropy.

The Choi sandwich $\lambda_{\min}(J(\Psi))\sigma_{\mathcal{R}} \otimes \mathbf{1} \leq (\text{id} \otimes \Psi)(\sigma) \leq \lambda_{\max}(J(\Psi))\sigma_{\mathcal{R}} \otimes \mathbf{1}$ is Lemma 5.7.3. If $J(\Psi) \leq cJ(\Psi')$, then $\rho_n^\Psi \leq c^n \rho_n^{\Psi'}$ for every n -use adaptive strategy (Lemma 5.7.1). Consequently, for every divergence that is antitone in its second argument and satisfies $\mathbb{D}(\rho\|c\sigma) = \mathbb{D}(\rho\|\sigma) - \ln c$, such as the Umegaki, Belavkin–Staszewski or max-relative entropy, the suprema over adaptive or parallel n -use strategies are $n d_T$ -Lipschitz in Ψ , and $|\zeta(\Phi\|\Psi) - \zeta(\Phi\|\Psi')| \leq d_T(\Psi, \Psi')$ (Theorem 5.7.2). With Eq. (8.2.1), $d_T(\Psi, \Psi') \leq \ln \frac{1}{1-r}$ for $r = d_X\|\Psi - \Psi'\|_{\diamond}/(2\lambda(\Psi)) < 1$; the constant $d_X/2$ of Lemma 8.2.1 halves the radius used in Theorem 5.7.2. The first-argument bound for D and D^A , uniformly over references, is Theorem 5.7.4. Corollary 5.7.5 gives the properties of ζ used here: ζ is lower semicontinuous on pairs of channels; if $J(\Psi)$ is invertible, then $\zeta(\Phi\|\Psi) \leq \ln(d_X/\lambda(\Psi))$ and Eq. (8.3.2) holds; and ζ is continuous at (Φ, Ψ) if and only if $J(\Psi)$ is invertible or $\zeta(\Phi\|\Psi) = \infty$.

Remark 8.15.3. If $J(\Phi + \Psi)$ is invertible, then either $J(\Psi)$ is invertible or $\text{ran} J(\Phi) \not\subseteq \text{ran} J(\Psi)$, in which case $\zeta(\Phi\|\Psi) \geq D(\Phi\|\Psi) = \infty$. So ζ , too, is continuous at every pair with $J(\Phi + \Psi)$ invertible.

Remark 8.15.4. D_{BS} is continuous, with values in $(-\infty, +\infty]$, at every pair of CP maps with $J(\Phi + \Psi)$ invertible. Chapter 4 does not state this as a single result; it follows from that chapter as follows, since $D_{\text{BS}} = \hat{D}_f$ with $f(t) = t \ln t$, for which $f(0^+) = 0$ and $f'(\infty) = +\infty$. Let $(\Phi', \Psi') \rightarrow (\Phi, \Psi)$. Their relative distance δ to (Φ, Ψ) tends to 0 (Corollary 4.10.8), and hence so do τ and d_T (Lemma 4.10.5). If $J(\Psi)$ is invertible, then $J(\Psi') \geq \frac{1}{2}\lambda_{\min}(J(\Psi))\mathbf{1}$ eventually, so $D_{\max}(\Phi'\|\Psi')$ is bounded uniformly, and Theorem 4.10.4(e) gives $D_{\text{BS}}(\Phi'\|\Psi') \rightarrow D_{\text{BS}}(\Phi\|\Psi)$. If $J(\Psi)$ is singular, then $\text{ran} J(\Phi) \not\subseteq \text{ran} J(\Psi)$ and $D_{\text{BS}}(\Phi\|\Psi) = +\infty$ (Theorem 4.7.5(a)); a bound $J(\Phi') \leq cJ(\Psi')$ along a subsequence with bounded c would give $J(\Phi) \leq cJ(\Psi)$ in the limit, so $D_{\max}(\Phi'\|\Psi') \rightarrow +\infty$, while $\nu(\Phi' + \Psi')$ stays bounded below by Lemma 4.8.2(2) and $\|(\Phi' + \Psi')^*(\mathbf{1})\|_{\infty}$ stays bounded, and Corollary 4.7.7 gives $D_{\text{BS}}(\Phi'\|\Psi') \rightarrow +\infty$. Theorem 4.7.6 is the corresponding statement for perturbations with a fixed sum $\Phi + \Psi$, along which it holds at every pair.

The geometric Chernoff converse

For an n -use adaptive strategy as above, the optimal error with equal priors is $P_n^{\text{err}} = \frac{1}{2} - \frac{1}{4}\|\rho_n^\Phi - \rho_n^\Psi\|_1$, and $\xi(\Phi, \Psi) = \limsup_n -\frac{1}{n} \ln \inf P_n^{\text{err}}$, the infimum running over strategies.

Every adaptive protocol is a controlled random walk on the levels of the pair (Proposition 7.3.1). For every pair of channels, every $n \geq 1$ and every n -use adaptive strategy, $P_n^{\text{err}} \geq \frac{1}{4} \exp(-n\hat{C}(\Phi, \Psi) - \sqrt{2 \ln 4} \Lambda \sqrt{n})$ with $\Lambda \leq d_T(\Phi, \Psi)$, so that $\xi(\Phi, \Psi) \leq \hat{C}(\Phi, \Psi)$ (Theorem 7.4.1); moreover $-\ln \hat{F}(\Phi, \Psi) \leq \hat{C}(\Phi, \Psi) \leq -2 \ln \hat{F}(\Phi, \Psi)$ (Corollary 7.4.5(a)).

Corollary 8.15.5 (Finite- n converse for an uncertain pair). *On the success event of Theorem 1.7.1, every n -use adaptive strategy for the true pair has*

$$P_n^{\text{err}} \geq \frac{1}{4} \exp\left(-n\left[\hat{C}(\hat{\Phi}, \hat{\Psi}) + \ln \frac{1}{1-r}\right] - \sqrt{2 \ln 4} \left[d_T(\hat{\Phi}, \hat{\Psi}) + 2 \ln \frac{1}{1-r}\right] \sqrt{n}\right).$$

Proof. The proof of Theorem 1.7.1 (Sec. 8.13) gives $\hat{C}(\Phi, \Psi) \leq \hat{C}(\hat{\Phi}, \hat{\Psi}) + \ln \frac{1}{1-r}$ and $(1-r)J(\hat{\Phi}) \leq J(\Phi) \leq (1+r)J(\hat{\Phi})$, so $d_T(\Phi, \hat{\Phi}) \leq \ln \frac{1}{1-r}$, and likewise for Ψ . The triangle inequality for d_T and Theorem 7.4.1 with $\Lambda \leq d_T(\Phi, \Psi)$ give the claim. $\square$

Fragility of single-letter Chernoff converses

Write $\xi_{\text{QC}}(\phi, \psi) = -\ln \min_{0 \leq s \leq 1} \text{Tr} \phi^s \psi^{1-s}$ for the quantum Chernoff exponent of two states [11, 99], and $F_{\text{U}}(\phi, \psi) = \|\sqrt{\phi}\sqrt{\psi}\|_1$ for their root fidelity.

Theorem 7.6.5 constructs replacement channels at distance $O(\sqrt{t})$ from a given pair at which the three single-letter converses are infinite; the following version gives an explicit construction and distance.

Theorem 8.15.6 (Replacement channels). *Let ϕ, ψ be states on $\mathbb{C}^{d'}$ with $\text{rank}(\phi + \psi) \leq d'/2$, and let $\Phi = \text{Tr}(\cdot)\phi$, $\Psi = \text{Tr}(\cdot)\psi$. There are states ϕ_t, ψ_t with $\|\phi_t - \phi\|_1, \|\psi_t - \psi\|_1 \leq 2(\sqrt{t} + t)$ such that the replacement channels $\Phi_t = \text{Tr}(\cdot)\phi_t$ and $\Psi_t = \text{Tr}(\cdot)\psi_t$ satisfy, for every $t > 0$,*

$$\begin{aligned} \hat{C}(\Phi_t, \Psi_t) &= -2 \ln \hat{F}(\Phi_t, \Psi_t) = C_{\text{bar}}(\Phi_t, \Psi_t) = +\infty, \\ \xi(\Phi_t, \Psi_t) &= \xi_{\text{QC}}(\phi_t, \psi_t) \leq -\ln \text{Tr} \phi_t^{1/2} \psi_t^{1/2} \xrightarrow{t \rightarrow 0} -\ln \text{Tr} \phi^{1/2} \psi^{1/2}. \end{aligned}$$

If $\phi = \psi$, then $\|\Phi_t - \Phi\|_{\diamond}, \|\Psi_t - \Psi\|_{\diamond} \leq 2(\sqrt{t} + t)$ and $\xi(\Phi_t, \Psi_t) \rightarrow 0$, while the three converses are infinite for every $t > 0$.

Proof. Let $\mathcal{K} = \text{supp}(\phi + \psi)$. Since $\dim \mathcal{K} \leq d'/2$, there is a partial isometry V with $V^\dagger V$ the projection onto $\mathcal{K}$ and $\text{ran} V \subseteq \mathcal{K}^\perp$. Put $A_\pm = \mathbf{1} \pm \sqrt{t}V$, $\phi_t = A_+ \phi A_+^\dagger / (1+t)$ and $\psi_t = A_- \psi A_-^\dagger / (1+t)$. These are states, since $\text{Tr} V \phi = 0$ and $\text{Tr} V \phi V^\dagger = 1$. Also $\phi_t - \phi = [\sqrt{t}(V\phi + \phi V^\dagger) + t(V\phi V^\dagger - \phi)] / (1+t)$ has trace norm at most $2(\sqrt{t} + t)$, and likewise for ψ_t . If $A_+ x = A_- y$ with $x, y \in \mathcal{K}$, the components in $\mathcal{K}$ and $\mathcal{K}^\perp$ give $x = y$ and $Vx = -Vy$, so $x = y = 0$. Hence $\text{ran} \phi_t \cap \text{ran} \psi_t \subseteq A_+ \mathcal{K} \cap A_- \mathcal{K} = \{0\}$, and so $\text{ran} J(\Phi_t) \cap \text{ran} J(\Psi_t) = (\text{ran} \phi_t \cap \text{ran} \psi_t) \otimes \mathcal{X} = \{0\}$, because $J(\Phi_t) = \phi_t \otimes \mathbf{1}$. Theorem 4.5.6 gives $\hat{Q}_s(\Phi_t \| \Psi_t) = 0$ for $0 < s < 1$, hence $\hat{C} = -2 \ln \hat{F} = \infty$. For every channel $\mathcal{T}$, $J(\mathcal{T}) \neq 0$ cannot have range in both $\text{ran} J(\Phi_t)$ and $\text{ran} J(\Psi_t)$, so, as in the proof of Theorem 7.6.4(d), $C_{\text{bar}} = \infty$. A replacement channel ignores its input, so the final states of every adaptive strategy are $\mathcal{S}(\phi_t^{\otimes n})$ and $\mathcal{S}(\psi_t^{\otimes n})$ for one channel $\mathcal{S}$. The strategy that measures the n outputs is therefore optimal, and $\xi = \xi_{\text{QC}}(\phi_t, \psi_t)$ [11, 99]. This is at most the value at $s = \frac{1}{2}$, which is continuous in the states. Finally $\|\text{Tr}(\cdot)(\phi_t - \phi)\|_{\diamond} = \|\phi_t - \phi\|_1$. $\square$

The example of Fig. 1.2(b). For $\phi = \psi = \frac{1}{2}(|1\rangle\langle 1| + |2\rangle\langle 2|)$ on $\mathbb{C}^4$ and $V = |3\rangle\langle 1| + |4\rangle\langle 2|$, the states $\phi_t = \frac{1}{2}(v_+ v_+^\dagger + w_+ w_+^\dagger)$ and $\psi_t = \frac{1}{2}(v_- v_-^\dagger + w_- w_-^\dagger)$, with $v_\pm = (|1\rangle \pm \sqrt{t}|3\rangle) / \sqrt{1+t}$ and $w_\pm = (|2\rangle \pm \sqrt{t}|4\rangle) / \sqrt{1+t}$, are block diagonal with pure blocks. For every $s \in [0, 1]$, $\text{Tr} \phi_t^s \psi_t^{1-s} = |\langle v_+, v_- \rangle|^2 = (\frac{1-t}{1+t})^2$, so $\xi(\Phi_t, \Psi_t) = 2 \ln \frac{1+t}{1-t}$ exactly, and $\|\Phi_t - \Phi\|_{\diamond} = 2\sqrt{t}/(1+t)$. The root fidelity is $F_{\text{U}}(\phi_t, \psi_t) = \frac{1-t}{1+t}$, so the program-simulation converse of Theorem 7.6.9, for which $\mu_* = F_{\text{U}}(\phi_t, \psi_t)$ at replacement channels (see the discussion after that theorem), is tight here: $-2 \ln \mu_*(\Phi_t, \Psi_t) = \xi(\Phi_t, \Psi_t)$.

The program-simulation converse

Parts (a)–(c) of the program-simulation converse quoted in Sec. 8.7 are Theorem 7.6.9, whose fidelity is the square of the root fidelity F_{U} used here. For replacement channels $\Phi = \text{Tr}(\cdot)\phi$, $\Psi = \text{Tr}(\cdot)\psi$ one has $\mu_*(\Phi, \Psi) = F_{\text{U}}(\phi, \psi)$, and $-2 \ln \mu_*$ is finite along any approximating pairs of replacement channels and converges to its value at the limit pair, so $\min\{\hat{C}, C_{\text{bar}}, -2 \ln \mu_*\}$ does not jump along the pairs of Theorem 8.15.6 with $\phi = \psi$ (discussion after Theorem 7.6.9). For $0 < \gamma, \gamma' \leq 1$, $\mu_*(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) = \sqrt{\gamma\gamma'}/(1-a)$ with $a = \sqrt{(1-\gamma)(1-\gamma')}$ (Corollary 7.6.10).

Chapter 9

Open problems

This chapter collects the open problems and conjectures of Chapters 1 and 4–8, grouped by theme. Problems that were raised in more than one chapter are stated once, with references to every place where they arise. Conjectures are stated as conjectures; each is supported by the partial results and numerical evidence described in the chapter cited, and none is claimed here. The problems are formulated in the notation of Chapter 3; where a chapter uses a different convention, for instance the squared Uhlmann fidelity in Chapter 7, we say so.

9.1 Sharp constants for the Kubo–Ando family

The continuity estimates of Chapter 4 are dimension-free, and several are known to be sharp: the constant and the exponent of Theorem 4.4.6, the order λ^{-1} of Theorem 4.8.8 and the rate of Theorem 4.12.11. Others are not.

Problem 9.1.1 (Constants for pairs with different sums). Lemma 4.10.5 is sharp as $\delta \rightarrow 0$. Is its factor $(1 - 4\delta^2)^{-1/2}$ optimal for finite δ ? The constants of Corollaries 4.6.11 and 4.6.12, for the Belavkin–Staszewski relative entropy and the geometric Rényi divergences of order $\alpha \in (1, 2]$, are not optimised either. (Section 4.13.)

Question 9.1.2 (A Schur multiplier norm). The coefficient of the second term of Lemma 4.10.3 is sharp as $\gamma \rightarrow 0$. Can the second term be replaced by $2\gamma\|\varphi\|_\infty$ for all γ ? By the proof of Lemma 4.10.3, this would follow if the Schur multiplier norm of the matrix $[\sin(\psi_i + \psi_j)]_{i,j}$, for $\psi_i \in [-\frac{\pi}{4}, \frac{\pi}{4}]$, were $\max_i |\sin 2\psi_i|$. (Section 4.13.)

Problem 9.1.3 (The local Lipschitz constant at noisy gates). What is the exact local Lipschitz constant of $\log \hat{F}$ near a pair of channels whose Choi matrices are bounded below by λ ? When only one channel moves, the constant $d_X/(4\lambda)$ is attained, at every Pauli channel composed with a unitary (Theorem 1.4.1, Eq. (1.4.3)). When both channels move, Theorem 4.8.8 and Proposition 8.15.2 give the constant $d_X/(2\lambda)$, and Example 4.8.11 leaves a factor 2 between this constant and the best examples. Each of the two one-sided terms is sharp separately; whether $d_X/(2\lambda)$ is attained when both channels move and both floors are at least λ is open. (Section 4.13; Section 8.3.)

For the absolute rather than the logarithmic slope, Chapter 8 computes the worst case over unitary channels at a depolarized qubit gate, and Chapters 1 and 8 conjecture that it is the worst case overall.

Conjecture 9.1.4 (Unitary channels are the worst case; Conjecture 8.3.2). *Let $\Psi = \mathcal{D}_p \circ \text{Ad}_U$ be a depolarized qubit gate, and let $\Sigma(p)$ be the largest derivative $|\frac{d}{de} \hat{F}(\Phi, \Psi + e\mathcal{M})|_{e=0}$*

over completely positive Φ and trace-annihilating $\mathcal{M}$ with $\|\mathcal{M}\|_\diamond = 1$. Then $\Sigma(p)$ equals the same supremum restricted to unitary channels Φ , which is given in closed form by Proposition 8.3.1.

Chapter 8 describes the numerical evidence for this conjecture as weak. A related problem concerns the condition number in Chapter 5.

Problem 9.1.5 (Condition numbers in the first argument). Can the condition number $\kappa(\Psi)$ of $J(\Psi)$ be removed from the constants of Theorems 5.7.4 and 5.7.8? (Section 5.10.)

9.2 Singular pairs and the dichotomy within channels

The dichotomy of Theorem 4.12.4 describes the continuity points of the Kubo–Ando family on pairs of completely positive maps. Chapters 1, 4 and 7 all ask how much of it survives within channels, and what happens at the discontinuity points.

Problem 9.2.1 (Limit points at singular pairs). Let $c_\Omega = \dim \ker J(\Omega)$ with $\Omega = \Phi + \Psi$. When $1 \leq c_\Omega < \text{rank } T(\mathbf{1} - T)$, what is the exact set of limit points of $\hat{F}$ at (Φ, Ψ) ? (Section 4.13.)

Problem 9.2.2 (Rates at pairs of zero fidelity). At mutually singular pairs the fidelity is continuous at the sharp rate $d^{1/2}/\sin \theta$ (Theorem 4.12.11). At what rate is it continuous at pairs with $\hat{F} = 0$ that are not mutually singular, such as the amplitude-damping pairs $(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'})$ of Example 4.12.8? (Section 4.13.)

Problem 9.2.3 (The dichotomy within channels). Within channels, the dichotomy holds at the diagonal pairs (Φ, Φ) (Corollary 4.12.10). Does it hold at every pair of channels with $J(\Phi + \Psi)$ singular and $\hat{F}(\Phi, \Psi) > 0$? By Theorem 4.12.9(a), it would suffice to find a Hermitian H for which the short $[J(\Phi) \# J(\Psi)]_{\mathcal{S}_H}$ lowers $\lambda_{\min} \circ \text{Tr}_\gamma$. Is the bound of Theorem 4.12.9 an equality? (Section 4.13; the same question is stated in Section 1.3.)

The next problem asks for the same phenomenon at the level of converses, together with control of the operational exponent.

Problem 9.2.4 (Spurious fragility near every singular pair). Does Theorem 7.6.5 hold near every pair of channels with $\dim \ker J(\Phi + \Psi) \geq 1$ and $\hat{C} < \infty$? That is, are there always nearby pairs of channels with trivially intersecting Choi ranges and bounded adaptive Chernoff exponent? Theorem 7.6.4(c) excludes such jumps when $J(\Phi + \Psi)$ is invertible, and Theorem 4.12.4 provides completely positive approximants. Channel-valued approximants along which $\hat{F}$ jumps are known at the diagonal pairs (Theorem 4.12.9 and Corollary 4.12.10, as in Problem 9.2.3) and within the amplitude-damping family (Example 4.12.8 and Proposition 7.6.8(b)); there the Choi ranges intersect nontrivially, and Corollary 7.6.10 shows that the adaptive exponent tends to 0 as $\gamma' \rightarrow \gamma \in (0, 1)$. (Question 7.8.5.)

9.3 Beyond operator convex functions and finite dimensions

Question 9.3.1 (Merely convex functions). For a convex but not operator convex f , the maximal f -divergence is still defined through reverse tests [91], but the perspective formula need not hold. Is it still a function of the ARN derivative T , and of which function? (Section 4.13.)

Problem 9.3.2 (The garbling between two shadows). By Theorem 5.3.2, the testing-region pair of (ρ, σ) is a garbling of the reverse test. Find this garbling explicitly. Is it the rotation of the Helstrom projections in Theorem 5.3.1(b)? (Section 5.10.)

Problem 9.3.3 (Other norms, von Neumann algebras and bosonic channels). By Ando's inequality in unitarily invariant norms, Proposition 4.4.3(1) holds in every unitarily invariant norm, and Proposition 4.4.3(2) becomes $\|g_{\#}(E) - g_{\#}(E')\|_p \leq \|E - E'\|_{p/2}^{1/2}$ for $p \geq 2$, but statements for maps need norms adapted to the Schatten classes. Extend the continuity theory of Chapter 4 to normal completely positive maps between von Neumann algebras, with T in the commutant of a Stinespring representation, as suggested by the framework of Okayasu [100, 101]; the estimates there are already dimension-free. Extend the level decomposition of Chapter 7 to normal channels between von Neumann algebras and to bosonic channels with energy constraints; adaptive quantum illumination is listed as open in [97]. Extend the noise-floor laws of Chapter 1 to bosonic loss channels with energy constraints. (Section 4.13; Question 7.8.6; Section 1.8.)

9.4 Regularised channel divergences and one-shot formulas

Question 9.4.1 (Continuity of regularised Petz divergences). Is $D_{\alpha}^{\text{P,reg}}(\cdot\|\Psi)$, for $\alpha < 1$, continuous at every channel when $J(\Psi)$ is invertible? Theorem 5.7.11 reduces this to perturbations that rotate the support of a singular $J(\Phi)$ by a non-product unitary of $\mathcal{X} \otimes \mathcal{Y}$, and Example 5.7.12 shows that a proof must use the n -fold structure, not a one-use bound. (Section 5.10.)

Question 9.4.2 (Optimal inputs for testing-region channel divergences). Does the iteration $\theta \mapsto \Gamma_f(\theta)/\text{Tr} \Gamma_f(\theta)$ of Theorem 5.7.13 always converge to an optimal input? In the linear case of the reverse test the objective increases along the iteration, and in all the tests of Chapter 5 the iteration converged to the optimum. For which pairs beyond the covariant ones of Proposition 5.7.14 is the maximally entangled input optimal, that is, when is the bracket of Theorem 5.7.13(c) scalar? (Section 5.10.)

9.5 Super-multiplicativity of testing-region quasi-entropies

The central conjecture of Chapter 6 is the following.

Conjecture 9.5.1 (Super-multiplicativity; Conjecture 6.3.5). *For all $\alpha \in (0, 1) \cup (1, \infty)$ and all pairs of states,*

$$Q_{\alpha}^{\text{HT}}(\rho \otimes \rho' \| \sigma \otimes \sigma') \geq Q_{\alpha}^{\text{HT}}(\rho \| \sigma) Q_{\alpha}^{\text{HT}}(\rho' \| \sigma').$$

It holds at $\alpha = 2$ (Theorem 6.4.1), at every integer order for pure states and at $\alpha = 3$ for sign-free pairs (Theorem 6.5.3), at every integer order against pure partners (Theorem 6.7.1), at $\alpha = 3$ whenever one of the reference states has at most two distinct eigenvalues (Corollary 6.6.3) and in the further cases of Corollaries 6.7.7 and 6.7.9, at $\alpha = 3$ to second order around trivial partners (Proposition 6.6.5), to first order at $\alpha = \infty$ (Theorem 6.4.3), for $\alpha > 1$ with one factor measured (Theorem 6.4.2) and for tensor powers with enough copies (Proposition 6.8.3). By Theorem 6.3.6 it would give independent proofs of the inequalities $D_{\alpha}^{\text{HT}} \leq \tilde{D}_{\alpha}$ for $\alpha > 1$ [84] and $D_{\alpha}^{\text{HT}} \geq D_{\alpha}^{\text{P}}$ for $\alpha < 1$ [13]. The conjecture is the family of moment inequalities of the convex-order statement (6.8.1) for Frenkel measures (Theorem 6.8.2), and the general problem is to prove that statement. Section 6.12 discusses the following routes and special cases in detail.

Problem 9.5.2 (Special cases of independent interest). Prove Conjecture 9.5.1 for pure states at non-integer orders, which is a Rényi-entropy submodularity for sums (Remark 6.4.4); the stronger logarithmic form of the integer-order inequality is open for $n \geq 3$. Prove the second-order case at $\alpha = 1$, the superadditivity of the testing-region variance $\int \gamma \log^2 \gamma d\mu - D^2$. Give a direct, non-greedy operator proof of Corollary 6.3.4 for $\kappa > 2$. (Section 6.12, item 1.)

Question 9.5.3 (Dihedral Chebyshev inequality). Is $\text{Cov}(\phi(U), \chi(U)) \geq 0$ for a uniform point U of the simplex Δ_{n-1} and convex functions ϕ, χ on Δ_{n-1} that are invariant only under the dihedral group D_n ? This extension of Lemma 6.5.2 would prove Conjecture 9.5.1 at every integer order for all pairs with nonnegative cycle weights, in particular for all pairs of qubit pairs (Proposition 6.5.1). Numerical tests support a positive answer. (Section 6.12, item 2.)

Problem 9.5.4 (Order three). At $\alpha = 3$ the open case has both reference states with at least three distinct eigenvalues, and both states may be taken singular (Corollary 6.6.6). Three targets would settle parts of it: Conjecture 6.7.10, an inequality for the kernel K_3 alone on twisted pure data, which by Proposition 6.7.11 covers every pair whose reference states have at most three eigenvalues; the chirality inequality (6.7.2), which gives $D_3 \geq 0$ in the polynomial criterion of Theorem 6.7.12, and more ambitiously $D_k \geq 0$ for all $k \geq 2$, which gives the conjecture at $\alpha = 3$; and, for a qutrit ρ against an arbitrary partner, the bound on the partner kernel stated in Section 6.12, item 2(c). The local supermodularity of Theorem 6.10.2 is a different route with the same first open cases. (Section 6.12, item 2.)

Pointwise comparison of profiles cannot prove Conjecture 9.5.1, because the profile difference changes sign for every two non-commuting pairs (Proposition 6.9.1), and one-way LOCC tests never beat the classical product of the shadows (Proposition 6.9.2). Chapter 6 conjectures that the same holds for all separable tests, so that a proof must use entangled tests.

Conjecture 9.5.5 (Separable tests; Conjecture 6.9.3). *Let e_{cl} be the hockey-stick profile of the classical product $(P \otimes P', Q \otimes Q')$ of the testing-region shadows of (ρ, σ) and (ρ', σ') . Every separable test M satisfies $\text{Tr } M(\rho \otimes \rho' - v \sigma \otimes \sigma') \leq e_{\text{cl}}(v)$ for all $v \geq 0$.*

The conjecture holds for one-way LOCC tests (Proposition 6.9.2) and was tested numerically with PPT tests. Whether it holds for two-way LOCC tests is open; in Blackwell form it asks whether every outcome experiment of a POVM with product elements is a garbling of $(P \otimes P', Q \otimes Q')$. (Section 6.12, item 3.)

9.6 Adaptive discrimination of channels

Question 9.6.1 (When does adaptivity not help?). Is $\xi_{\text{adapt}} = \xi_{\text{prod}}$ whenever $d_T(\Phi, \Psi) < \infty$? This holds for classical channels, for classical-quantum channels [116] and in the cases of Theorems 7.5.1 and 7.5.3, and the known separations [53, 116] have $d_T = \infty$. (Question 7.8.1.)

Amplitude damping is a natural test case. For $0 < \gamma < 1$ and $\Delta = \gamma - \gamma'$, Corollary 7.6.10 brackets the adaptive Chernoff exponent between the Bernoulli Chernoff information of the product input, $\Delta^2/(8\gamma(1-\gamma)) + O(\Delta^3)$, and the program-simulation bound $-2 \ln \mu_* = \Delta^2/(4\gamma^2(1-\gamma)) + O(\Delta^3)$; the two differ by the factor $2/\gamma$ to leading order, and no program simulation can close the gap (Remark 7.6.11). With $f(\gamma, \gamma') = \sqrt{\gamma\gamma'} + \sqrt{(1-\gamma)(1-\gamma')}$, Chapter 7 conjectures the following; the fidelity F in it is the squared Uhlmann fidelity, as in that chapter.

Conjecture 9.6.2 (Fidelity of adaptive amplitude-damping strategies; Conjecture 7.6.13). *For every adaptive strategy with n uses, $F(\rho_{\mathcal{A}_\gamma}^{(n)}, \rho_{\mathcal{A}_{\gamma'}}^{(n)}) \geq f(\gamma, \gamma')^{2n}$, with equality for the product input $e_1^{\otimes n}$. Consequently*

$$\xi_{\text{adapt}}(\mathcal{A}_\gamma, \mathcal{A}_{\gamma'}) \leq -2 \ln f(\gamma, \gamma') = \frac{\Delta^2}{4\gamma(1-\gamma)} + O(\Delta^3) \leq 2\xi_{\text{prod}}(1 + O(\Delta)).$$

For $\gamma = \frac{1}{2}$, $n \leq 4$ and several values of γ' , exact semidefinite programs agree with the conjecture to the accuracy of the solver, and it cannot be proved by amortisation (Section 7.6.1; Section 8.7). A bound on the adaptive Fisher information suggests that ξ_{adapt} equals the product-input value to leading order (Remark 7.6.12); whether this is so is a further open question.

Problem 9.6.3 (A single-letter converse without spurious jumps). Find a single-letter converse for the adaptive Chernoff exponent that is at most $\min\{C_{\text{bar}}, \hat{C}\}$ for all pairs. For replacement channels both exceed the true exponent, so such a converse should see the non-commutativity of the outputs, which every common decomposition hides (Proposition 7.4.3). By Theorem 7.6.5, a converse within a bounded factor of ξ_{adapt} must be finite on pairs whose Choi ranges intersect trivially, so it cannot be built only from Choi-level geometric means or from divergences that require support inclusion. The program-simulation converse $-2 \ln \mu_*$ of Theorem 7.6.9 is finite on the pairs of Theorem 7.6.5 and on amplitude damping, which makes $\min\{\hat{C}, C_{\text{bar}}, -2 \ln \mu_*\}$ a natural candidate to study. (Question 7.8.2; Section 1.8.)

Problem 9.6.4 (The local Chernoff coefficient). For nearby channels, is the adaptive Chernoff exponent $\frac{\delta^2}{8} F(\theta) + o(\delta^2)$ for an explicitly computable Fisher information F between the one achieved by product inputs and F_{RLD} ? In this regime $T = \frac{1}{2}\mathbf{1} + O(\delta)$, and one expects the gap between ξ_{prod} and $\hat{C}$ to be controlled by the commutator in Lemma 7.2.4. At a depolarized qubit gate with a coherent miscalibration δ , Chapter 1 states, and Chapter 8 proves, that ξ/δ^2 lies asymptotically between $F_\infty/8$ and $F_{\text{RLD}}/8$ as $\delta \rightarrow 0$ (Proposition 8.10.2; the lower bound uses the error-corrected strategies of Zhou and Jiang [131]); where in this bracket the coefficient lies is open, and no Chernoff analogue of Conjecture 9.7.1 is stated. (Question 7.8.3; Section 1.8.)

Problem 9.6.5 (A Hoeffding converse). Prove a Hoeffding converse for general pairs of channels. The random-walk structure of Proposition 7.3.1 applies, but $s \mapsto \min_\rho \ln Q_s(\rho)$ need not be convex, so the natural candidate involves its convex envelope. (Question 7.8.4.)

Problem 9.6.6 (The Chernoff exponent from testing regions). Is the adaptive Chernoff exponent $-\lim_n \frac{1}{n} \log \inf \eta_n(1)$, where η_n is the error profile after n adaptive uses, the minimum over α of a regularised testing-region cumulant function? Proposition 6.3.1 gives one inequality for every strategy. (Section 6.12, item 4.)

Problem 9.6.7 (Operational consequences for other tasks). The geometric Rényi divergences with $\alpha \in (1, 2]$ give converse bounds for channel capacities [42], and geometric distinguishability measures limit channel discrimination and estimation [72]. Chapters 1 and 7 make the continuity and the discontinuity of the converses quantitative for discrimination: a converse valid on a neighbourhood of a model pair (Corollary 7.6.3), the Choi-floor law (Theorem 1.4.1) and the comparison with the Stein and Chernoff exponents (Section 1.6). The corresponding questions for capacity bounds and for estimation are open: how do these converses respond to model errors, and does the sensitivity belong to the task? (Section 4.13, item 4.)

9.7 Noisy gates and certification

The local comparison of Chapter 1 between the Belavkin–Staszewski converse and the adaptive Stein exponent at a depolarized qubit gate leaves the following conjecture.

Conjecture 9.7.1 (Local tightness of the Belavkin–Staszewski converse; Conjecture 1.6.2). *For the depolarized qubit gate,*

$$\lim_{\delta \rightarrow 0} \frac{D_{\text{BS}}(\mathcal{N}_{p,\delta} \| \mathcal{N}_p)}{\zeta(\mathcal{N}_{p,\delta} \| \mathcal{N}_p)} = \frac{F_{\text{RLD}}}{F_{\infty}},$$

which tends to 3/2 as $p \rightarrow 0$.

Chapter 1 proves that the ratio lies between 1 and $F_{\text{RLD}}/F_{\infty} < 3/2$ in the limit (Eq. (1.6.2)). It does not take a position on whether the conjecture holds: for the explicit strategies evaluated there, the relative-entropy rate exceeds the SLD rate of the same strategy, so the limit could lie strictly below $F_{\text{RLD}}/F_{\infty}$.

Problem 9.7.2 (Correlated errors). For an n -qubit gate the fragile window of the geometric converses scales as $\delta \sim p^{w/2}$ for a miscalibration of weight w . For $w \geq 2$ the local bracket between the converse and the Stein exponent widens, to $2/p$ for a two-qubit ZZ miscalibration under local depolarizing noise and as $p^{-\lfloor w/2 \rfloor}$ for weight w (Corollary 8.12.5). Does ζ inherit the narrow window of the converse for correlated errors, or is that window a spurious fragility? (Sections 1.6 and 8.12.)

Problem 9.7.3 (Certification with coherent learners). For learners that measure each channel output separately, certification of a geometric converse through a diamond-norm ball costs of order $1/(\eta\tilde{\lambda})^2$ channel uses (Proposition 8.13.1). Can coherent learners certify through a diamond ball with fewer uses? The two-point argument behind the incoherent bound does not extend to them (Remark 8.13.2). (Section 1.7.)

Problem 9.7.4 (Model-free certification on the converse side). Variance-adaptive estimation of Pauli error rates in a known frame certifies a lower bound on $\hat{F}$, hence an upper bound on the converse, with of order $1/(\eta^2\tilde{\lambda})$ uses. Does model-free tomography attain this $1/\tilde{\lambda}$ law on the converse side? (Section 1.7.)

Conjecture 9.7.5 (Intrinsic certification cost; Conjecture 8.13.4). *For every pair of channels with Choi-state floor $\tilde{\lambda}$, the optimal number of uses for certifying $\hat{C}$ or $\hat{F}$ to relative accuracy η is $\tilde{\Theta}(\text{poly}(d)/(\eta^2\tilde{\lambda}))$.*

For local depolarizing noise, Proposition 8.13.3 gives a lower bound of this order for one estimand and an estimator that attains it for Pauli channels in a known frame.

Bibliography

- [1] P. M. Alberti and A. Uhlmann. A problem relating to positive linear maps on matrix algebras. *Rep. Math. Phys.*, 18:163–176, 1980.
- [2] William N. Anderson, Jr. Shorted operators. *SIAM J. Appl. Math.*, 20(3):520–525, 1971. doi: 10.1137/0120053.
- [3] William N. Anderson, Jr. and Richard J. Duffin. Series and parallel addition of matrices. *J. Math. Anal. Appl.*, 26(3):576–594, 1969.
- [4] William N. Anderson, Jr. and George E. Trapp. Shorted operators. II. *SIAM J. Appl. Math.*, 28(1):60–71, 1975. doi: 10.1137/0128007.
- [5] Tsuyoshi Ando. Lebesgue-type decomposition of positive operators. *Acta Sci. Math. (Szeged)*, 38:253–260, 1976.
- [6] Tsuyoshi Ando. Concavity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra Appl.*, 26:203–241, 1979.
- [7] Tsuyoshi Ando. Comparison of norms $|||f(A) - f(B)|||$ and $|||f(|A - B|)|||$. *Math. Z.*, 197:403–409, 1988.
- [8] Tsuyoshi Ando, Chi-Kwong Li, and Roy Mathias. Geometric means. *Linear Algebra Appl.*, 385:305–334, 2004.
- [9] H. Araki. On an inequality of Lieb and Thirring. *Lett. Math. Phys.*, 19:167–170, 1990.
- [10] William B. Arveson. Subalgebras of C^* -algebras. *Acta Math.*, 123:141–224, 1969.
- [11] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, Ll. Masanes, A. Acín, and F. Verstraete. Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.*, 98:160501, 2007. doi: 10.1103/PhysRevLett.98.160501.
- [12] K. Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J.*, 19:357–367, 1967.
- [13] S. Beigi, C. Hirche, and M. Tomamichel. Some properties and applications of the new quantum f -divergences, 2025. arXiv:2501.03799.
- [14] V. P. Belavkin and P. Staszewski. C^* -algebraic generalization of relative entropy and entropy. *Ann. Inst. H. Poincaré Sect. A (N.S.)*, 37(1):51–58, 1982.
- [15] M. Berta, L. Lami, and M. Tomamichel. Continuity of entropies via integral representations, 2024. arXiv:2408.15226.

- [16] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, New York, 1997.
- [17] Rajendra Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.
- [18] Rajendra Bhatia and Chandler Davis. A Cauchy–Schwarz inequality for operators with applications. *Linear Algebra Appl.*, 223/224:119–129, 1995.
- [19] Rajendra Bhatia and Kalyan B. Sinha. Variation of real powers of positive operators. *Indiana Univ. Math. J.*, 43(3):913–925, 1994. doi: 10.1512/iumj.1994.43.43039.
- [20] Rajendra Bhatia, Stéphane Gaubert, and Tanvi Jain. Matrix versions of the Hellinger distance. *Lett. Math. Phys.*, 109:1777–1804, 2019.
- [21] Andreas Bluhm, Ángela Capel, Paul Gondolf, and Antonio Pérez-Hernández. General continuity bounds for quantum relative entropies. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 162–167, 2023. doi: 10.1109/ISIT54713.2023.10206734. arXiv:2305.10140.
- [22] R. Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Phys. Rev. Lett.*, 105:200504, 2010. doi: 10.1103/PhysRevLett.105.200504.
- [23] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.*, 8:14485, 2017. doi: 10.1038/ncomms14485.
- [24] S. L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439, 1994. doi: 10.1103/PhysRevLett.72.3439.
- [25] F. Buscemi. Comparison of quantum statistical models: equivalent conditions for sufficiency. *Commun. Math. Phys.*, 310:625–647, 2012.
- [26] F. Buscemi and G. Gour. Quantum relative Lorenz curves. *Phys. Rev. A*, 95:012110, 2017.
- [27] G. Chiribella, G. M. D’Ariano, and P. Perinotti. Theoretical framework for quantum networks. *Phys. Rev. A*, 80:022339, 2009.
- [28] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra Appl.*, 10(3):285–290, 1975.
- [29] Sam Cree and Jamie Sikora. A fidelity measure for quantum states based on the matrix geometric mean, 2020. arXiv:2006.06918.
- [30] R. Demkowicz-Dobrzański and L. Maccone. Using entanglement against noise in quantum metrology. *Phys. Rev. Lett.*, 113:250801, 2014. doi: 10.1103/PhysRevLett.113.250801.
- [31] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă. The elusive Heisenberg limit in quantum-enhanced metrology. *Nat. Commun.*, 3:1063, 2012. doi: 10.1038/ncomms2067.

- [32] R. Demkowicz-Dobrzański, J. Czejkowski, and P. Sekatski. Adaptive quantum metrology under general Markovian noise. *Phys. Rev. X*, 7:041009, 2017. doi: 10.1103/PhysRevX.7.041009.
- [33] Igor Devetak and Peter W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Comm. Math. Phys.*, 256:287–303, 2005.
- [34] R. G. Douglas. On majorization, factorization, and range inclusion of operators on Hilbert space. *Proc. Amer. Math. Soc.*, 17:413–415, 1966.
- [35] R. Duan, Y. Feng, and M. Ying. Entanglement is not necessary for perfect discrimination between unitary operations. *Phys. Rev. Lett.*, 98:100503, 2007. doi: 10.1103/PhysRevLett.98.100503.
- [36] W. Dür, M. Skotiniotis, F. Fröwis, and B. Kraus. Improved quantum metrology using quantum error correction. *Phys. Rev. Lett.*, 112:080801, 2014. doi: 10.1103/PhysRevLett.112.080801.
- [37] Ali Ebadian, Ismail Nikoufar, and Madjid Eshaghi Gordji. Perspectives of matrix convex functions. *Proc. Natl. Acad. Sci. USA*, 108(18):7313–7314, 2011.
- [38] Edward G. Effros. A matrix convexity approach to some celebrated quantum inequalities. *Proc. Natl. Acad. Sci. USA*, 106(4):1006–1008, 2009.
- [39] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi. Quantum certification and benchmarking. *Nat. Rev. Phys.*, 2:382, 2020. doi: 10.1038/s42254-020-0186-4.
- [40] H. Epstein. Remarks on two theorems of E. Lieb. *Commun. Math. Phys.*, 31:317–325, 1973.
- [41] K. Fang, O. Fawzi, R. Renner, and D. Sutter. Chain rule for the quantum relative entropy. *Phys. Rev. Lett.*, 124:100501, 2020.
- [42] Kun Fang and Hamza Fawzi. Geometric Rényi divergence and its applications in quantum channel capacities. *Comm. Math. Phys.*, 384(3):1615–1677, 2021. doi: 10.1007/s00220-021-04064-4.
- [43] H. Fawzi and O. Fawzi. Defining quantum divergences via convex optimization. *Quantum*, 5:387, 2021.
- [44] H. Fawzi, J. Saunders, and P. A. Parrilo. Semidefinite approximations of the matrix logarithm. *Found. Comput. Math.*, 19:259–296, 2019.
- [45] S. T. Flammia and J. J. Wallman. Efficient estimation of Pauli channels. *ACM Trans. Quantum Comput.*, 1(1):1, 2020. doi: 10.1145/3408039.
- [46] P. E. Frenkel. Integral formula for quantum relative entropy implies data processing inequality. *Quantum*, 7:1102, 2023.
- [47] Péter E. Frenkel, Milán Mosonyi, Péter Vrana, and Mihály Weiner. Error bounds for composite quantum hypothesis testing and a new characterization of the weighted Kubo–Ando geometric means, 2025. arXiv:2503.13379.

- [48] G. Gour. Quantum noncommutativity uniquely determines relative entropy, 2026. [arXiv:2607.01712](#).
- [49] G. Gutoski and J. Watrous. Toward a general theory of quantum games. In *Proc. 39th Annu. ACM Symp. Theory Comput. (STOC)*, pages 565–574, 2007.
- [50] Paul R. Halmos. Two subspaces. *Trans. Amer. Math. Soc.*, 144:381–389, 1969.
- [51] F. Hansen and G. K. Pedersen. Jensen’s operator inequality. *Bull. London Math. Soc.*, 35:553–564, 2003.
- [52] R. Harper, S. T. Flammia, and J. J. Wallman. Efficient learning of quantum noise. *Nat. Phys.*, 16:1184, 2020. doi: 10.1038/s41567-020-0992-8.
- [53] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous. Adaptive versus non-adaptive strategies for quantum channel discrimination. *Phys. Rev. A*, 81:032339, 2010.
- [54] M. Hayashi. Optimal sequence of quantum measurements in the sense of Stein’s lemma in quantum hypothesis testing. *J. Phys. A*, 35:10759–10773, 2002.
- [55] M. Hayashi. Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Phys. Rev. A*, 76:062301, 2007.
- [56] M. Hayashi. Discrimination of two channels by adaptive methods and its application to quantum system. *IEEE Trans. Inf. Theory*, 55:3807–3820, 2009.
- [57] F. Hiai and D. Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.*, 143:99–114, 1991.
- [58] F. Hiai, M. Mosonyi, and M. Tomamichel. Hockey stick f -divergences. [arXiv:2607.08760](#), 2026.
- [59] Fumio Hiai. Matrix analysis: matrix monotone functions, matrix means, and majorization. *Interdiscip. Inform. Sci.*, 16(2):139–248, 2010.
- [60] Fumio Hiai. Quantum f -divergences in von Neumann algebras. II. Maximal f -divergences. *J. Math. Phys.*, 60(1):012203, 2019. doi: 10.1063/1.5051427.
- [61] Fumio Hiai and Milán Mosonyi. Different quantum f -divergences and the reversibility of quantum operations. *Rev. Math. Phys.*, 29(7):1750023, 2017. doi: 10.1142/S0129055X17500234.
- [62] Fumio Hiai and Dénes Petz. *Introduction to Matrix Analysis and Applications*. Universitext. Springer, Cham; Hindustan Book Agency, New Delhi, 2014. doi: 10.1007/978-3-319-04150-6.
- [63] C. Hirche and M. Tomamichel. Quantum Rényi and f -divergences from integral representations. *Commun. Math. Phys.*, 2024. doi: 10.1007/s00220-024-05087-3.
- [64] Roger A. Horn and Roy Mathias. Cauchy–Schwarz inequalities associated with positive semidefinite matrices. *Linear Algebra Appl.*, 142:63–82, 1990.
- [65] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A*, 223:1–8, 1996.

- [66] K. Ji, H. K. Mishra, M. Mosonyi, and M. M. Wilde. Barycentric bounds on the error exponents of quantum hypothesis exclusion. *arXiv:2407.13728*, 2024.
- [67] K. Ji, H. K. Mishra, M. Mosonyi, and M. M. Wilde. Converse bounds for quantum hypothesis exclusion: A divergence-radius approach. *arXiv:2501.09712*, 2025.
- [68] Nathaniel Johnston, David W. Kribs, and Vern I. Paulsen. Computing stabilized norms for quantum operations via the theory of completely bounded maps. *Quantum Inf. Comput.*, 9(1–2):16–35, 2009.
- [69] M. Jones. Continuity of certain quantum channel divergences. *arXiv:2610.16270*, 2026.
- [70] S. Karlin. *Total Positivity*, volume I. Stanford University Press, 1968.
- [71] V. Katariya and M. M. Wilde. RLD Fisher information bound for multiparameter estimation of quantum channels. *New J. Phys.*, 23:073040, 2021.
- [72] Vishal Katariya and Mark M. Wilde. Geometric distinguishability measures limit quantum channel estimation and discrimination. *Quantum Inf. Process.*, 20(2):78, 2021. doi: 10.1007/s11128-021-02992-7.
- [73] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin. Quantum error correction for metrology. *Phys. Rev. Lett.*, 112:150802, 2014. doi: 10.1103/PhysRevLett.112.150802.
- [74] Sumeet Khatri, Kunal Sharma, and Mark M. Wilde. Information-theoretic aspects of the generalized amplitude-damping channel. *Phys. Rev. A*, 102(1):012401, 2020. doi: 10.1103/PhysRevA.102.012401. *arXiv:1903.07747*.
- [75] Mohsen Kian. Sharpness loss for Radon–Nikodym derivatives of completely positive maps, 2026. *arXiv:2609.09915*.
- [76] Mohsen Kian. Relative Kubo–Ando means of completely positive maps, 2026. *arXiv:2605.11701*.
- [77] A. Yu. Kitaev. Quantum computations: algorithms and error correction. *Russian Math. Surveys*, 52(6):1191–1249, 1997.
- [78] M. Kliesch and I. Roth. Theory of quantum system certification. *PRX Quantum*, 2:010201, 2021. doi: 10.1103/PRXQuantum.2.010201.
- [79] Fumio Kubo and Tsuyoshi Ando. Means of positive linear operators. *Math. Ann.*, 246(3):205–224, 1980.
- [80] Richard Kueng, David M. Long, Andrew C. Doherty, and Steven T. Flammia. Comparing experiments to the fault-tolerance threshold. *Phys. Rev. Lett.*, 117(17):170502, 2016. doi: 10.1103/PhysRevLett.117.170502. *arXiv:1510.05653*.
- [81] E. H. Lieb and W. Thirring. Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In *Studies in Mathematical Physics*, pages 269–303. Princeton University Press, 1976.
- [82] Friedrich Liese and Igor Vajda. On divergences and informations in statistics and information theory. *IEEE Trans. Inform. Theory*, 52(10):4394–4412, 2006.

- [83] Jianhua Lin. Divergence measures based on the Shannon entropy. *IEEE Trans. Inform. Theory*, 37(1):145–151, 1991.
- [84] P.-C. Liu, C. Hirche, and H.-C. Cheng. Layer cake representations for quantum divergences, 2025. arXiv:2507.07065.
- [85] Karl Löwner. Über monotone Matrixfunktionen. *Math. Z.*, 38:177–216, 1934.
- [86] Wei Luo, Chuanning Song, and Qingxiang Xu. Perturbation estimation for the parallel sum of Hermitian positive semi-definite matrices. *Linear Multilinear Algebra*, 67(10):1971–1984, 2019. doi: 10.1080/03081087.2018.1477915. arXiv:1806.07033.
- [87] Filip B. Maciejewski, Zoltán Zimborás, and Michał Oszmaniec. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum*, 4:257, 2020. doi: 10.22331/q-2020-04-24-257. arXiv:1907.08518.
- [88] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer, 2nd edition, 2011.
- [89] K. Matsumoto. An example of a quantum statistical model which cannot be mapped to a less informative one by any trace preserving positive map, 2014. arXiv:1409.5658.
- [90] Keiji Matsumoto. Reverse test and quantum analogue of classical fidelity and generalized fidelity, 2010. arXiv:1006.0302.
- [91] Keiji Matsumoto. A new quantum version of f -divergence. In *Reality and Measurement in Algebraic Quantum Theory*, volume 261 of *Springer Proc. Math. Stat.*, pages 229–273. Springer, Singapore, 2018.
- [92] W. Matthews, S. Wehner, and A. Winter. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Commun. Math. Phys.*, 291:813–843, 2009.
- [93] A. A. Mele and L. Bittel. Optimal learning of quantum channels in diamond distance. *arXiv:2512.10214*, 2025.
- [94] M. Mosonyi and T. Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. *Commun. Math. Phys.*, 334:1617–1648, 2015.
- [95] Milán Mosonyi and Fumio Hiai. Some continuity properties of quantum Rényi divergences. *IEEE Trans. Inform. Theory*, 70(4):2674–2700, 2024. doi: 10.1109/TIT.2023.3324758.
- [96] H. Nagaoka. The converse part of the theorem for quantum Hoeffding bound, 2006. arXiv:quant-ph/0611289.
- [97] R. Nair and M. Gu. Fundamental limits of quantum illumination. *Optica*, 7:771–774, 2020.
- [98] Theshani Nuradha and Mark M. Wilde. Query complexity of classical and quantum channel discrimination. *Quantum Sci. Technol.*, 10(4):045075, 2025. doi: 10.1088/2058-9565/ae0a79. arXiv:2504.12989.

- [99] M. Nussbaum and A. Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.*, 37:1040–1057, 2009.
- [100] Rui Okayasu. Geometric mean and Lebesgue-type decomposition of completely positive maps, 2026. arXiv:2605.06019.
- [101] Rui Okayasu. Operator-valued maximal f -divergences for completely positive maps, 2026. arXiv:2609.00554.
- [102] A. Oufkir and F. Girardi. Improved lower bounds for learning quantum channels in diamond distance. *arXiv:2601.04180*, 2026.
- [103] Vern I. Paulsen. *Completely Bounded Maps and Operator Algebras*, volume 78 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002.
- [104] D. Petz. Monotone metrics on matrix spaces. *Linear Algebra Appl.*, 244:81, 1996.
- [105] M. Piani. Relative entropy of entanglement and restricted measurements. *Phys. Rev. Lett.*, 103:160504, 2009.
- [106] S. Pirandola and C. Lupo. Ultimate precision of adaptive noise estimation. *Phys. Rev. Lett.*, 118:100502, 2017. doi: 10.1103/PhysRevLett.118.100502.
- [107] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira. Fundamental limits to quantum channel discrimination. *npj Quantum Inf.*, 5:50, 2019.
- [108] József Pitrik and Dániel Virostek. Quantum Hellinger distances revisited. *Lett. Math. Phys.*, 110:2039–2052, 2020.
- [109] Shruti Puri, Lucas St-Jean, Jonathan A. Gross, Alexander Grimm, Nicholas E. Fratini, Pavithran S. Iyer, Anirudh Krishna, Steven Touzard, Liang Jiang, Alexandre Blais, Steven T. Flammia, and S. M. Girvin. Bias-preserving gates with stabilized cat qubits. *Sci. Adv.*, 6:eaay5901, 2020. doi: 10.1126/sciadv.aay5901.
- [110] W. Pusz and S. L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. *Rep. Math. Phys.*, 8(2):159–170, 1975.
- [111] Maxim Raginsky. Radon–Nikodym derivatives of quantum operations. *J. Math. Phys.*, 44(11):5003–5020, 2003.
- [112] David Reeb, Michael J. Kastoryano, and Michael M. Wolf. Hilbert’s projective metric in quantum information theory. *J. Math. Phys.*, 52(8):082201, 2011. doi: 10.1063/1.3615729. arXiv:1102.5170.
- [113] J. M. Renes. Relative submajorization and its use in quantum resource theories. *J. Math. Phys.*, 57:122202, 2016.
- [114] M. Rexti and S. Mancini. Discriminating qubit amplitude damping channels. *J. Phys. A*, 54:165303, 2021. doi: 10.1088/1751-8121/abed68.
- [115] B. Russo and H. A. Dye. A note on unitary operators in C^* -algebras. *Duke Math. J.*, 33(2):413–416, 1966.
- [116] F. Salek, M. Hayashi, and A. Winter. Usefulness of adaptive strategies in asymptotic quantum channel discrimination. *Phys. Rev. A*, 105:022419, 2022.

- [117] L. Schmitt and D. Sutter. The regularized channel Rényi divergence is continuous. *arXiv:2609.29885*, 2026.
- [118] L. Seveso, F. Albarelli, M. G. Genoni, and M. G. A. Paris. On the discontinuity of the quantum Fisher information for quantum statistical models with parameter dependent rank. *J. Phys. A*, 53:02LT01, 2020. doi: 10.1088/1751-8121/ab599b.
- [119] W. Forrest Stinespring. Positive functions on C^* -algebras. *Proc. Amer. Math. Soc.*, 6:211–216, 1955.
- [120] A. C. Thompson. On certain contraction mappings in a partially ordered vector space. *Proc. Amer. Math. Soc.*, 14:438, 1963.
- [121] David K. Tuckett, Stephen D. Bartlett, and Steven T. Flammia. Ultrahigh error threshold for surface codes with biased noise. *Phys. Rev. Lett.*, 120:050505, 2018. doi: 10.1103/PhysRevLett.120.050505.
- [122] A. Uhlmann. The “transition probability” in the state space of a $*$ -algebra. *Rep. Math. Phys.*, 9(2):273–279, 1976.
- [123] D. Šafránek. Discontinuities of the quantum Fisher information and the Bures metric. *Phys. Rev. A*, 95:052320, 2017. doi: 10.1103/PhysRevA.95.052320.
- [124] J. Wang and Y. Yang. Continuity of regularized channel Rényi divergences, 2026. *arXiv:2609.28635*.
- [125] X. Wang and M. M. Wilde. Resource theory of asymmetric distinguishability for quantum channels. *Phys. Rev. Research*, 1:033169, 2019.
- [126] John Watrous. Notes on super-operator norms induced by Schatten norms. *Quantum Inf. Comput.*, 5(1):58–68, 2005.
- [127] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, Cambridge, 2018.
- [128] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur. Amortized channel divergence for asymptotic quantum channel discrimination. *Lett. Math. Phys.*, 110:2277, 2020. doi: 10.1007/s11005-020-01297-7.
- [129] A. Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. *Commun. Math. Phys.*, 347:291–313, 2016.
- [130] Christopher J. Wood and Jay M. Gambetta. Quantification and characterization of leakage errors. *Phys. Rev. A*, 97:032306, 2018. doi: 10.1103/PhysRevA.97.032306. *arXiv:1704.03081*.
- [131] S. Zhou and L. Jiang. Asymptotic theory of quantum channel estimation. *PRX Quantum*, 2:010343, 2021. doi: 10.1103/PRXQuantum.2.010343.
- [132] S. Zhou, M. Zhang, J. Preskill, and L. Jiang. Achieving the Heisenberg limit in quantum metrology using quantum error correction. *Nat. Commun.*, 9:78, 2018. doi: 10.1038/s41467-017-02510-3.