The Ultimate Matching Theorem

Michael Goldman[1] Francesco Grotto[2] Dario Trevisan[2]

[1]LAMA, CNRS, Université Gustave Eiffel, Université Paris-Est Créteil,
5 boulevard Descartes, 77454 Marne-la-Vallée, France
michael.goldman@cnrs.fr

[2]Università di Pisa, Dipartimento di Matematica,
Largo Bruno Pontecorvo 5, 56127 Pisa, Italia
francesco.grotto@unipi.it dario.trevisan@unipi.it

November 10, 2026

hexagon:2611.00294v1

Abstract

Let $X_1, \dots, X_N$ be independent uniform points in the unit square, matched either to an evenly spread deterministic configuration or to a second independent sample. For exponents $\alpha_1, \alpha_2 > 2$ with $1/\alpha_1 + 1/\alpha_2 = 1/2$, we prove that with probability tending to one a single permutation controls the jth coordinate displacements in the exponential Orlicz norm of exponent α_j, at the scale $\sqrt{\log N/N}$, for both $j = 1, 2$. This settles Talagrand's Ultimate Matching Conjecture for all pairs of finite critical exponents. The proof constructs, on the flat torus, a current whose divergence is Gaussian white noise projected on a dyadic grid and whose coordinates have the critical anisotropic exponential Orlicz bounds, and transfers it to matching through a primal optimal-transport reduction. At the hard endpoints $(2, \infty)$ and $(\infty, 2)$ the conjecture fails: a Brownian local-time construction shows that the normalized optimal matching cost diverges in probability. This file was mostly generated by LLMs as a result of interactions with us to try and understand the 'monolithic' versions of the proofs which may be found in the **associated GitHub repository** [28]. A digested version of the proofs will be made available soon.

Contents

1 Introduction **2**

I Critical Gaussian Orlicz flows on the torus **6**

2 The Gaussian flow problem and the strategy **6**

3 Gaussian increments along predictable partitions **12**

4 Local divergence inverses and Whitney partitions **14**

5 Anisotropic covariance selection **19**

6 The adaptive construction and the bootstrap **20**

7 Exponential moment estimates **27**

8 The terminal Gaussian correction **30**

II From matching to dyadic Gaussian Orlicz flows **36**

9 Overview of the reduction **36**

10 Notation and statement of the reduction **37**

11 Preliminary estimates **39**

12 From point configurations to histograms **41**

13 Transport and divergence currents **41**

14 Bounds on the histogram densities **46**

15 Coupling the histograms with dyadic Gaussian noise **46**

16 Proof of the reduction **48**

17 From the torus to the square **49**

18 Choice of dyadic mesh **51**

19 Consequences and the hard endpoint **52**

III Failure of the hard endpoint **53**

20 Statement, mechanism, and Gaussian reduction **53**

21 A pathwise dual inequality **54**

22 Brownian local time and arbitrarily large reward **55**

23 Realization by spatially adapted Gaussian bubbles **58**

24 Divergence of the Gaussian gauge and failure of matching **62**

Appendix **64**

A Geometric estimates for the adaptive construction **64**

Use of AI tools **68**

1 Introduction

1.1 The matching problem

Let $X_1, \dots, X_N$ be independent uniform points in the unit square $D = [0,1]^2$. Talagrand's *Ultimate Matching Conjecture*, posed as Research Problem 17.1.2 (equation (17.3)) of [25], asks for a permutation matching this random sample to *evenly spread* deterministic points $Y_1, \dots, Y_N$ with simultaneous exponential tail bounds in both coordinates. Here "evenly spread" has Talagrand's meaning: D admits a partition into N rectangles of area $1/N$, with one point Y_i in each rectangle R_i, such that $R_i \subset B(Y_i, 10/\sqrt{N})$. We also consider the matching of two independent uniform

samples, which satisfies the same conclusion. Our first result, which we call the Ultimate Matching Theorem, is the following.

Theorem 1.1. *Let α_1, α_2 be fixed finite exponents with*

$$\alpha_1, \alpha_2 > 2, \qquad \frac{1}{\alpha_1} + \frac{1}{\alpha_2} = \frac{1}{2}.$$

There exists a constant $L = L(\alpha_1, \alpha_2) > 0$, independent of N and of the evenly spread configuration, such that the following holds with probability tending to one as $N \to \infty$: for every fixed evenly spread deterministic $(Y_i)_{i=1}^N$*, or when $(Y_i)_{i=1}^N$ is a second independent uniform sample,* one and the same *permutation $\pi \in S_N$ satisfies, for both coordinates $j = 1, 2$,*

$$\sum_{i=1}^{N} \exp\left[\left(\frac{\sqrt{N}\,|X_i^j - Y_{\pi(i)}^j|}{L\sqrt{\log N}}\right)^{\alpha_j}\right] \le 2N, \qquad j = 1, 2. \tag{1.1}$$

The symmetric quartic case $(4, 4)$ is included. The proof is completed in Corollaries 18.1 and 16.1, which combine the Gaussian flow estimate of Theorem 2.1 with the reduction of Part II.

Finite subcritical exponents. The theorem also gives (1.1) for every fixed pair $0 < \alpha_1, \alpha_2 < \infty$ with $\alpha_1^{-1} + \alpha_2^{-1} > 1/2$, after changing the constant L. Indeed, set

$$c = 2(\alpha_1^{-1} + \alpha_2^{-1}) > 1, \qquad r_i = c\alpha_i.$$

Then $r_i > \alpha_i$, $r_i > 2$, and $r_1^{-1} + r_2^{-1} = 1/2$, so the critical theorem applies with exponents r_1, r_2. For the resulting permutation, write $t_{i,j} = \sqrt{N}\,|X_i^j - Y_{\pi(i)}^j|/(L\sqrt{\log N})$. Since $t^{\alpha_j} \le 1 + t^{r_j}$ for $t \ge 0$, Jensen's inequality with a fixed $A \ge 1$ gives

$$\frac{1}{N}\sum_{i=1}^{N} e^{(t_{i,j}/A)^{\alpha_j}} \le \left(\frac{1}{N}\sum_{i=1}^{N} e^{t_{i,j}^{\alpha_j}}\right)^{A^{-\alpha_j}} \le (2e)^{A^{-\alpha_j}} \le 2$$

when $A^{\alpha_j} \ge \log(2e)/\log 2$ for both $j = 1, 2$. Thus the same permutation works at the same scale, with L replaced by AL; the probability of success is unchanged.

The hard endpoints. The coordinate endpoints $(2, \infty)$ and $(\infty, 2)$ behave differently. Theorem 20.1 in Part III shows that there the optimal matching cost, divided by $\sqrt{\log N/N}$, diverges in probability, uniformly over deterministic reference configurations. The Gaussian construction of Part I applies only to finite exponents, and the endpoint obstruction is of a different nature. These coordinatewise endpoints should not be confused with the usual scalar L^∞ matching problem or with an endpoint of a power-cost problem. In the endpoint cases it is in fact possible to determine the exact first order asymptotic (in probability) behavior of the cost: for its (undigested, AI assisted) derivation we refer once again to the associated GitHub repository [28], since here we are mainly interested in the failure of Talagrand's conjecture.

1.2 Previous work

The dual route. The foundations of planar random matching go back to Ajtai, Komlós and Tusnády [1]. Talagrand's work [23, 24, 25] identified the multiscale nature of the problem and led to Research Problem 17.1.2 in the 2021 monograph. The guiding method is *dual*: one estimates transport or discrepancy functionals against classes of test functions, and controls the resulting random suprema through chaining, majorizing measures and geometric decompositions. This approach has produced sharp matching estimates in several regimes.

In dimension three, Chapter 18 of [25] proves a sharp matching theorem between an independent uniform sample and an evenly spread deterministic configuration, including coordinate-dependent

exponential costs and hard coordinate constraints (Theorem 18.1.1 and Exercise 18.1.4); Talagrand observes that the extension to dimensions $d > 3$ requires no new ideas. Above dimension two, irregularities at different scales cannot accumulate in the same way, whereas in the plane their cumulative effect produces the critical logarithmic factor. The remaining difficulty is therefore specifically two-dimensional.

The dual direction continues to develop: Cheng, Otto, Palmieri and Wachtel [12] study min–max matching through a geometrically linear approximation of a dual formulation, and Ouimet [21] proposes a solution to Problem 5.3 in Talagrand's *My favorite problems* [26], controlling simultaneously total distance and maximal displacement through electrical flows and planar Lipschitz truncation of lattice circulations.

The primal route. Our proof follows a complementary tradition. Rather than estimating the dual supremum, we construct an actual *primal* transport current carrying the discrepancy of the random measures. This approach was brought into focus by the statistical-physics prediction of Caracciolo, Lucibello, Parisi and Sicuro [11]: at leading order in quadratic bipartite matching, the displacement should behave like the gradient of a Poisson potential driven by empirical-density fluctuations, which is the linearization of the Monge–Ampère equation. Ambrosio, Stra and Trevisan [6] established the leading two-dimensional quadratic asymptotics rigorously, and finer results on the cost and the optimal map followed [3, 4]. In parallel, the variational regularity theory of optimal transport, through the harmonic approximation of Goldman and Otto [17] and its quantitative versions by Goldman, Huesmann and Otto [15, 16], quantified the approximation of displacements by Poisson fluxes; related methods treat nonuniform planar domains [5]. These results make it natural to study random matching through the Eulerian continuity equation and a compatible vector current.

For nonquadratic costs the corresponding equation is nonlinear. Ambrosio, Vitillaro and Trevisan [7] obtained sharp estimates for power-law costs, and Armegioiu, Goldman, Grotto and Trevisan [8] related finite power-cost asymptotics on the torus to a Gaussian variational-martingale problem and a parabolic Monge–Ampère Bellman equation. In this viewpoint successive logarithmic scales are accumulated through stochastic control; it directly motivates the construction of Part I. The power-cost results do not, by themselves, give the coordinatewise exponential Orlicz bounds proved here.

1.3 Strategy of the proof

Critical Gaussian currents. The main analytic task is to solve a divergence equation for mean-zero Gaussian white noise projected onto dyadic squares of mesh $h = 2^{-m}$ on the flat two-dimensional torus. With $\psi_r(t) = e^{|t|^r} - 1$ and $\|\cdot\|_{\psi_r}$ the associated Luxemburg norm, Part I constructs, with probability bounded below uniformly in m, a periodic current J with

$$\operatorname{div} J = W_h, \qquad \max_{i=1,2} \|J_i\|_{\psi_{r_i}} \le C\sqrt{m},$$

for every finite critical pair $(r_1, r_2) = (\alpha_1, \alpha_2)$ (Theorem 2.1). The difficulty is not the existence of an inverse of the divergence, but the selection of one whose two coordinate tails are simultaneously of the critical size.

At a single point, a Bellman equation suggests an inverse-Hessian covariance for a Gaussian martingale, and a supersolution evaluated backward in Bellman time formally gives a supermartingale. The pointwise choice, however, need not be compatible across space. Part I therefore reveals the Gaussian increments through predictable local divergence corrections on anisotropic Whitney cells. An adaptive metric follows the locally preferred covariance, and a discounted memory records spatial oscillations of the accumulated current. A simultaneous induction on the metric and the memory yields a contracting Lyapunov quantity, whose negative drift pays for the variance of the critical exponential tilt. A terminal Gaussian correction then transfers the current from the adaptive partition to the fixed dyadic grid.

From currents to matching. Part II supplies the reductions, independently of the adaptive construction. The minimum-flow gauge is uniformly Lipschitz in the dyadic Gaussian coefficients, so a fixed positive success probability upgrades to high probability by Gaussian concentration. An exact four-sheet folding transfers the torus estimate to the square. Dyadic regularization, a comparison between transport and static currents, and an interval-encoded KMT coupling relate the Gaussian source to one empirical histogram against Lebesgue measure, and then to two independent histograms. Talagrand's evenly spread points are a further $O(N^{-1/2})$ transport from Lebesgue measure, and a Birkhoff decomposition yields one permutation with both coordinate bounds. These steps give (1.1) in both formulations.

The endpoint obstruction. At the endpoint $(2, \infty)$ the reduction of Part II still applies, so it suffices to show that the square Gaussian minimum-flow gauge is not of order $\sqrt{m}$. Part III builds random dual test potentials of bounded cost whose expected pairing with the projected noise is too large. The input is a planar Brownian martingale with bounded terminal anisotropic Orlicz cost but arbitrarily large expected integrated square-root determinant, obtained from short Brownian local-time stages. An adaptive tree of piecewise affine Gaussian bubbles converts this determinant into a positive Gaussian pairing.

1.4 Organization

Part I (Sections 2–8) proves the Gaussian flow estimate on the torus; the geometric estimates for its adaptive construction are collected in Appendix A. Part II (Sections 9–19) reduces matching to the square Gaussian flow problem, upgrades fixed-probability bounds to high probability, transfers the torus estimate to the square, and completes the proof of Theorem 1.1. Part III (Sections 20–24) proves the failure at the hard endpoints.

Part I
Critical Gaussian Orlicz flows on the torus

In this part, W_h is the projection of mean-zero Gaussian white noise on the flat torus $\mathbb{T}^2$ onto the dyadic square grid of mesh $h = 2^{-m}$. For fixed $2 < r_1, r_2 < \infty$ with $1/r_1 + 1/r_2 = 1/2$, we bound the minimum coordinatewise Orlicz norm of a periodic current with divergence W_h by a multiple of $\sqrt{m}$, with probability bounded below independently of m (Theorem 2.1). The construction combines predictable Gaussian increments, local inverses of the divergence in anisotropic Whitney charts, and a contracting metric–memory functional. A terminal Gaussian correction transfers the adaptive current to the fixed dyadic source.

2 The Gaussian flow problem and the strategy

2.1 The Gaussian minimum-flow problem

Let W be mean-zero real white noise on the flat torus $\mathbb{T}^2$:

$$\mathbb{E}W(f)W(g) = \int_{\mathbb{T}^2} fg - \int_{\mathbb{T}^2} f \int_{\mathbb{T}^2} g. \tag{2.1}$$

For $h = 2^{-m}$, let P_h denote averaging on the fixed periodic h-square grid and write $W_h = P_h W$; thus $W_h|_Q = h^{-2}W(Q)$ and $\int_{\mathbb{T}^2} W_h = 0$. We use the Luxemburg norm $\|\cdot\|_{\psi_r}$ associated with $\psi_r(t) = e^{|t|^r} - 1$. Fix finite exponents $r_i > 2$ on the critical line

$$a_i = \frac{1}{r_i}, \qquad a_1 + a_2 = \frac{1}{2}. \tag{2.2}$$

We also set $a_* := \max_i a_i < 1/2$. For every mean-zero source f, define the minimum-flow gauge

$$\mathfrak{F}_{\boldsymbol{r}}(f) = \inf_{\operatorname{div} J = f} \max_{i=1,2} \|J_i\|_{\psi_{r_i}}, \tag{2.3}$$

where $\boldsymbol{r} = (r_1, r_2)$ and the divergence constraint is understood distributionally for periodic currents. The problem is to control this gauge for W_h.

Theorem 2.1. *For every fixed finite $r_1, r_2 > 2$ with $1/r_1 + 1/r_2 = 1/2$, there exist constants $C < \infty$, $\varepsilon_0 > 0$ and $m_0 < \infty$, depending only on the exponents, such that for $m \geq m_0$,*

$$\mathbb{P}\left(\mathfrak{F}_{\boldsymbol{r}}(W_h) \leq C\sqrt{m}\right) > \varepsilon_0. \tag{2.4}$$

Remark 2.2. The fixed-positive-probability estimate is the conclusion proved in this part. As shown in Part II (Subsection 11.1), the fixed-grid minimum-flow gauge is Lipschitz in the Gaussian coefficients with a constant independent of h. Gaussian concentration therefore upgrades (2.4) to $\mathbb{P}(\mathfrak{F}_{\boldsymbol{r}}(W_h) > C'\sqrt{m}) \leq 2e^{-cm}$ for suitable $C', c > 0$. This improvement requires no further adaptive-flow estimates and is not used below.

2.2 The quartic case

Consider first $r_1 = r_2 = 4$. The proof controls the exponential cost of the accumulated current through the phase

$$H(y) = (1 + y_1^2)^2 + (1 + y_2^2)^2.$$

Here $y \in \mathbb{R}^2$ is a current value: at a spatial point x and reveal generation k, it will be $y = U_k(x)$. It is not a spatial coordinate. The quadratic terms in H give nondegenerate curvature near the origin, while its growth at infinity is quartic.

The choice of the next covariance is suggested by the curvature of e^H. Since $D^2H \preceq CH^{1/2}I$, the matrix

$$B(y) = H(y)^{1/2}I + \nabla H(y)\nabla H(y)^T$$

controls both the Hessian term and the square of the linear variation of the phase. Among positive definite matrices of determinant one, the choice $C(y) = \sqrt{\det B(y)}B(y)^{-1}$ minimizes $\mathrm{tr}(B(y)G)$; the calculation is given in Section 5. For example, at $y = (R, 0)$ with $R \to \infty$,

$$B(y) \asymp \mathrm{diag}(R^6, R^2), \qquad C(y) \asymp \mathrm{diag}(R^{-2}, R^2).$$

The corresponding phase cost is $\mathrm{tr}(BC) \asymp R^4 \asymp H$, whereas an isotropic choice gives $\mathrm{tr}\, B \asymp R^6$. Thus the preferred covariance reduces the variance in the direction where the current is already large.

This pointwise choice cannot be imposed independently at every spatial point. The local divergence equation must be solved on a whole cell, in a chart chosen before the next noise is revealed. The construction therefore uses one covariance-control matrix G per cell and moves it towards C by bounded updates. A memory of old spatial oscillations controls the error between the state at the cell representative and the state at other points. The next subsection explains this simultaneous metric–memory induction.

For general fixed finite critical exponents, the same construction uses

$$F_i(t) = (1 + t^2)^{r_i/2}, \qquad H(y) = F_1(y_1) + F_2(y_2), \qquad v(y) = \nabla H(y). \tag{2.5}$$

2.3 Structure of the proof

The noise is revealed through a finite predictable sequence of Gaussian increments. At each stage a local divergence correction adds the newly revealed noise to the current, with a covariance chosen from the old history on each active cell. The aim is to control a spatial exponential modular of the accumulated current. The phase is H from (2.5), and the preferred covariance is selected by the regularized Hessian proxy in (5.1). No continuous-time Bellman equation is solved in the construction; its formal relation to the discrete argument is explained in Subsection 2.4.

The inverse-Hessian rule does not by itself solve the spatial problem. The preferred covariance $C(U_k(x))$ depends on x, whereas an inverse of the divergence must be constructed coherently on a whole region. Moreover, the covariance must be chosen *before* the next Gaussian variables are revealed. The construction therefore follows the pointwise Bellman covariance with a predictable metric and records the price of not following it exactly. This is done by a simultaneous induction on the routing metric and on a memory of spatial oscillations.

Section 3 reveals the source through a finite predictable tree. Section 4 provides a local divergence inverse in intrinsic Whitney coordinates and a normalized Gaussian increment with uniform conditional variance and Hölder envelopes E_{k+1}. These bounds do not depend on the eccentricity of the physical chart. We use the scales

$$s = hm^3, \qquad \delta = \frac{\varepsilon^2}{m}, \qquad \lambda_k = e^{-\Lambda k\delta}, \tag{2.6}$$

where $\varepsilon > 0$ is a small fixed step parameter and $\Lambda > 0$ is a sufficiently large fixed damping constant. The cumulative normalized current satisfies $\mathrm{div}\, U_k = \sqrt{\delta}\, P_{\mathcal{P}_k} W$; its unscaled counterpart is $U_k/\sqrt{\delta}$. The grid parameter m is fixed throughout the construction; k counts reveals and does not specify a common spatial size of the cells. The depth of any actual revealed chain is at most $J_0 \le C_{\mathrm{dep}} m$, for a fixed geometric constant C_{dep}.

Metric lag. On a parent Q, choose an old-measurable representative x_Q. Section 5 selects an updated determinant-one metric G_Q^+ by clipping the old metric G_k towards $C(U_k(x_Q))$. The clipping stays within a fixed factor of G_k and contracts, at x_Q, the lag

$$L_k(x)^2 = \mathrm{tr}\Big(G_k(x)C(U_k(x))^{-1}\Big).$$

If U_k were constant on Q, this would be a contraction throughout Q. But U_k varies in space while the metric is chosen once on Q, and the difference $U_k(x) - U_k(x_Q)$ is exactly what prevents applying the pointwise Bellman optimization independently at every location.

Memory. Each old increment is Hölder in the intrinsic metric in which it was created, and descendants shrink in these birth metrics, even after the bounded changes of metric between refinements. Hence its oscillation across a much later descendant is reduced by a factor r_{osc}^{k-j}. The old diagonal metric caps and relative comparisons of $H(U_j)$ then give, schematically for x, z in the current parent,

$$|U_{k,i}(x) - U_{k,i}(z)| \lesssim \sqrt{\delta M}\, H(U_k(x))^{a_i} \sum_{j \le k} r_{\mathrm{osc}}^{k-j} E_j(x).$$

Along the actual reveals of each branch we retain the discounted square function

$$T_k = \sum_{j \le k} \rho_1^{k-j} E_j^2, \qquad S_k = L_k^2 + T_k, \qquad r_{\mathrm{osc}}^2 < \rho_1 < 1.$$

Weighted Cauchy–Schwarz bounds the ancestor sum by $C\sqrt{T_k}$. Thus T_k pays for the oscillation of the old field between the representative, where the metric is selected, and the point where the estimate is used.

Simultaneous induction. Section 6 closes three estimates together: the coordinatewise caps $G_{k,ii} \le MH(U_k)^{2a_i}$, with a fixed large constant M; small relative changes of the phase $H(U_k)$; and the contraction of S_k. The *old* memory estimate controls $U_k(x) - U_k(x_Q)$, so the representative-based metric update preserves the diagonal caps with a strict margin. The *fresh* Gaussian envelope gives $|V_{k+1,i}| \lesssim \sqrt{\delta}\, H(U_k)^{a_i} E_{k+1}$, which preserves the caps and the phase comparison at the new state. Only then do we use the old stopped exponential modular, together with the lower area bound for active cells, to obtain $H(U_k) \lesssim m$ pointwise. Consequently $\sqrt{\delta H(U_k)} \lesssim \varepsilon$.

This small factor makes the metric update stable under spatial variation of the state. After the representative contraction and a comparison of the lag along the old-to-new state segment, which uses only old-state information, one obtains

$$L_{k+1}^2 \le q_L L_k^2 + C + C\varepsilon^2 T_k + C E_{k+1}^2, \qquad q_L < 1.$$

On an actual reveal, $T_{k+1} = \rho_1 T_k + E_{k+1}^2$. Choosing ρ_1 and ε with strict margins, the two inequalities combine into the renewal estimate

$$\boxed{S_{k+1} \le q_0 S_k + C(1 + E_{k+1}^2), \qquad q_0 < 1.} \tag{2.7}$$

The old memory justifies the new metric, and the contraction of the new metric pays for retaining the memory. This is why the two are propagated simultaneously.

Exponential estimate. The renewal bound (2.7) supplies the negative drift that turns the exponential Bellman trial into a spatial supermartingale. Write $H_k = H(U_k)$. The phase used in Section 7 is

$$\Phi_k = \lambda_k H_k (1 + \kappa\delta S_k),$$

where $\kappa > 0$ is fixed and the clock λ_k decreases slightly with each reveal. Under a fresh reveal, the linear variation of H_k is Gaussian. At criticality its conditional variance is bounded, up to constants, by $\lambda_k \delta H_k S_k$: the factor S_k measures how far the actual covariance is from the pointwise Bellman covariance. Exponentiating this linear term produces a positive drift of that size, and the term $\kappa\delta H_k S_k$ is added to the exponent to compensate it. On a good reveal, $H_{k+1} \le (1+\xi) H_k$ for a small fixed ξ, and (2.7) gives

$$H_{k+1} S_{k+1} - H_k S_k \le -d_0 H_k S_k + C H_k (1 + E_{k+1}^2), \qquad d_0 = 1 - q_0(1+\xi) > 0.$$

Thus the additional phase contributes $-\kappa d_0 \lambda_{k+1} \delta H_k S_k$, which dominates the conditional Gaussian variance cost once κ is large. The remaining errors have size $\lambda_{k+1}\delta H_k(1 + E_{k+1}^2)$. The square of the

envelope has a uniform conditional exponential moment, and its coefficient satisfies $\delta H_k \lesssim \varepsilon^2$ by the old modular bound $H_k \lesssim m$. The constant part is paid for by the decrease $\lambda_k - \lambda_{k+1} \asymp \lambda_{k+1}\delta$.

The order of these estimates matters. We bound the product $H_{k+1}S_{k+1}$ by relative phase control, rather than expanding the new lag inside the phase. Consequently the only linear Gaussian term is the variation of H_k, whose coefficient depends on the old state and on the metric selected *before* the noise is revealed. Its conditional moment-generating function can therefore be evaluated directly, and no independence between the Gaussian increment and its Hölder envelope is needed. This gives a one-step conditional estimate for the spatial exponential modular.

To iterate it, Section 7 uses a nonnegative killed-and-frozen numerical supermartingale. On histories that pass every envelope and modular test it is the spatial integral of e^{Φ_k}; on an envelope failure it is set to zero. Overshoot stopping and the separately controlled probability of an envelope failure then give a bounded spatial modular with a fixed positive probability. Terminal branches are frozen without artificial memory discounts. In this sense the construction plays the role of the formal Bellman supermartingale $u(T-t, X_t)$ from (2.9), with the metric–memory mechanism paying for spatial incompatibility.

Terminal correction. Section 8 transfers the divergence from the adaptive partition to the fixed-grid source W_h in physical coordinates. Combining that correction with the successful leaf current gives (2.4). The geometric estimates used in the construction are collected in Appendix A.

2.4 Bellman motivation

The starting analogy is the power-energy problem

$$\inf_{\operatorname{div} J = W_h} \int_{\mathbb{T}^2} |J(x)|^p \, \mathrm{d}x, \qquad 1 < p < \infty.$$

For fixed p, its logarithmic-scale asymptotics lead to a stochastic-control problem; see [8]. Fix a *spatial point* $x \in \mathbb{T}^2$ and reveal the Gaussian noise through successive logarithmic Fourier shells. The divergence constraint fixes the longitudinal contribution to the current at x, but transverse corrections can be optimized. In the continuous-scale limit, the accumulated two-component current at this fixed point is modeled by a controlled Gaussian martingale. Its state belongs to $\mathbb{R}^2$; we denote a generic state by y. The corresponding power-cost Bellman equation is

$$\partial_t u_p(t,y) = \frac{1}{2}\sqrt{\det D_y^2 u_p(t,y)}, \qquad u_p(0,y) = |y|^p. \tag{2.8}$$

Where its Hessian is positive definite, the optimal one-point covariance has the determinant-one inverse-Hessian form

$$C_p(t,y) = \sqrt{\det D_y^2 u_p(t,y)}\,(D_y^2 u_p(t,y))^{-1}.$$

This is the continuous-time antecedent of the selector $C(y)$ used later in (5.1).

Here is the supermartingale interpretation, which will guide the critical construction. For any smooth, strictly convex candidate u with positive Hessian, denote by C_u the same normalized inverse-Hessian matrix. Fix a terminal Bellman time $T > 0$ and an initial current state $X_0 = y$. Formally let the one-point current evolve as

$$\mathrm{d}X_t = \frac{1}{\sqrt{2}} C_u(T-t, X_t)^{1/2} \, \mathrm{d}\mathbf{B}_t, \qquad 0 \le t \le T,$$

where $\mathbf{B}_t$ is a standard planar Brownian motion modeling the newly revealed Gaussian innovations after the transverse control. Thus X_t is *the current accumulated at the fixed spatial point* x, not the position of a particle or a spatial current on the whole torus. Its conditional covariance per unit time is $\frac{1}{2}C_u(T-t, X_t)$. The factor $1/2$ is exactly the time normalization in (2.8). Since $\frac{1}{4}\operatorname{tr}(C_u D_y^2 u) = \frac{1}{2}\sqrt{\det D_y^2 u}$, Itô's formula gives

$$\mathrm{d}\, u(T-t, X_t) = \left[\frac{1}{4}\operatorname{tr}(C_u D_y^2 u) - \partial_t u\right](T-t, X_t)\,\mathrm{d}t + \mathrm{d}M_t \;\le\; \mathrm{d}M_t \tag{2.9}$$

whenever u is a Bellman supersolution. With equality in the Bellman equation the drift vanishes. In the power case, $u_p(0, X_T) = |X_T|^p$, so the supermartingale property yields the one-point terminal-cost bound $\mathbb{E}|X_T|^p \le u_p(T, y)$ (after localization and the necessary integrability justification). The reversal of Bellman time is essential: the PDE evolves in the remaining-time variable, while X_t accumulates newly revealed noise. This calculation is formal (it requires localization and a well-defined feedback martingale), but it identifies the goal: a supermartingale built from an exponential trial evaluated along the evolving current.

For the Orlicz problem, use the phase H from (2.5) and the critical weights $a_i = 1/r_i$. Try $u(t, y) = \exp\{\zeta_{\mathrm{B}}(t)H(y)\}$, with $\zeta_{\mathrm{B}} > 0$ an auxiliary increasing Bellman-time coefficient. To compare with the subcritical exponents, put $\sigma = a_1 + a_2 - \frac{1}{2} \ge 0$, so $\sigma = 0$ on the critical line. Differentiation and the 2×2 rank-one determinant identity yield

$$e^{-\zeta_{\mathrm{B}} H} \sqrt{\det D_y^2 u} \le C(\zeta_{\mathrm{B}} + \zeta_{\mathrm{B}}^{3/2}) H^{1-\sigma}. \tag{2.10}$$

As $\partial_t u = \zeta_{\mathrm{B}}' H u$, choosing $\zeta_{\mathrm{B}}' \ge C(\zeta_{\mathrm{B}} + \zeta_{\mathrm{B}}^{3/2})$ produces a supersolution for a prescribed finite interval, provided the initial coefficient is sufficiently small. In particular, the exponential Bellman supersolution persists at criticality, including the symmetric quartic example $r_1 = r_2 = 4$. The subcritical advantage is instead the large-amplitude power reserve $H^{-\sigma}$, which disappears when $\sigma = 0$.

The one-point limiting cost is heuristically finite, but this discards all the geometry. A pointwise covariance optimization does not construct a *spatially compatible* current $x \mapsto J(x)$ driven by the same white noise and satisfying $\operatorname{div} J = W_h$. At polynomial order the reconstruction errors can be controlled by Wiener-chaos estimates as in [8]. At critical exponential growth the missing power reserve means that the one-point supermartingale cannot by itself absorb the errors of spatial reconstruction. The solution sought in this part is therefore a *spatial* analogue of (2.9), coupled to an adaptive routing metric and a memory of spatial oscillations.

2.5 Notation

The notation used throughout Part I is collected in Table 1. All references point to the first definition of the corresponding object.

Table 1: Principal notation and location of first definitions.

Symbol	Meaning	Definition
W	Mean-zero white noise on $\mathbb{T}^2$.	(2.1)
P_h, W_h	Averaging on the fixed dyadic h-grid; $W_h = P_h W$.	p. 6
$\boldsymbol{r} = (r_1, r_2)$, a_i	Orlicz exponents and critical weights $a_i = 1/r_i$.	(2.2)
a_*	Larger critical weight $\max_i a_i < 1/2$.	p. 6
ψ_r, $\|\cdot\|_{\psi_r}$	Exponential Young function and Luxemburg norm.	p. 6
$\mathfrak{F}_{\boldsymbol{r}}$	Minimum coordinatewise Orlicz flow gauge.	(2.3)
h, s, δ, ε	Fixed-grid mesh, terminal intrinsic scale, time step $\delta = \varepsilon^2/m$, and fixed small ε.	(2.6)
J_0, C_{dep}	Deterministic bound $J_0 \le C_{\mathrm{dep}} m$ for actual reveal depth.	p. 7
λ_k, Λ	Deterministic damping clock and its fixed rate.	(2.6)
$\mathcal{P}_k$	Nested partition after the kth reveal.	p. 12
$\mathcal{F}_k$	Sigma-field of geometry and masses revealed through stage k.	p. 13
f_k, Δf_k	Adaptive projection of W and its next Gaussian increment.	p. 12
$Q_k(x)$, $M_k(x)$, $v_{k+1}(x)$	Cell containing x, its Gaussian average, and next conditional variance.	(3.4)

Symbol	Meaning	Definition
G_Q^+, A_Q, $\widehat{Q}$	Predictable routing metric, its square root, and the intrinsic parent chart.	(4.1)
K, c_0, D	Fixed adjacent-metric distortion, inball ratio, and current intrinsic parent diameter.	p. 14
θ, D_0	Whitney size cap and intrinsic terminal stopping threshold $D \le D_0 s$.	p. 15
γ, A, C_A	Whitney boundary-clearance, inball-displacement and receiver-size constants.	Lemma 4.1
ϑ	Child size contraction across an actual birth-chart refinement, distinct from θ.	p. 15
z, $\widehat{S}$	Intrinsic coordinates and transformed child.	(4.1)
$d_{G_Q^+}$	Distance in the predictable intrinsic metric.	p. 14
ℓ_S	Intrinsic scale of a Whitney child.	(4.7)
α, β	Hölder exponents: kernel α and field β.	p. 17
c_*	Fixed positive floor $\lvert S\rvert \ge c_* s^2$ for revealed-cell areas.	Lemma 3.2
$\widehat{K}_Q$, $\widehat{\mathcal{B}}_Q$	Ray kernel and local divergence inverse in intrinsic coordinates.	(4.2)
K_Q^{phys}, $\mathcal{B}_Q^{\mathrm{phys}}$	Corresponding physical Piola kernel and inverse.	(4.3)
$\widehat{P}$	Averaging onto the intrinsic Whitney child partition.	p. 17
$H(y)$, $v(y)$	Anisotropic exponential phase and its gradient.	(2.5)
$\zeta_{\mathrm{B}}(t)$, σ	Bellman time coefficient (distinct from λ_k) and subcritical reserve $\sigma = a_1 + a_2 - 1/2$.	pp. 10–10
$B(y)$, $C(y)$	Bellman matrix and determinant-one target metric.	(5.1)
$\mathcal{T}_C$	Spectral metric clipping toward C, with relative distortion bounded by K.	Lemma 5.1
$L(y, G)$	Lag between actual and target metrics.	(5.2)
G_k, G_k^+	Birth metric and predictable updated metric at a split.	p. 20
M	Fixed coefficient in the diagonal cap $G_{ii} \le MH(U)^{2a_i}$.	Lemma 5.1
η	Relative change allowed for $H(U_k)$ at an actual reveal.	p. 23
r_{osc}	Effective ancestor spatial-oscillation decay factor.	Lemma 6.3
q_d, q_L, q_0	Diagonal-cap, lag and combined-memory contraction factors.	pp. 26,26
ξ	Realized relative phase-change bound $\lvert H' - H\rvert \le \xi H$, controlled by ε.	p. 26
Z_{k+1}	Intrinsic normalized Gaussian increment field.	p. 20
V_{k+1}, U_k	Physical increment and cumulative divergence current.	(6.1)
$E_{k+1,S}$	Gaussian supremum–Hölder envelope on a child.	(6.4)
T_k, S_k	Discounted spatial memory and combined lag–memory energy.	(6.9)
ρ_1	Fixed memory discount; satisfies $r_{\mathrm{osc}}^2 < \rho_1 < q_0$.	(6.9), (6.17)
R	Global good-event cutoff $E_{k,S} \le R\sqrt{m}$ for Gaussian envelopes.	(6.6)
T_*	Fixed stopping threshold for the exponential modular.	p. 23
Φ_k, κ	Exponential Lyapunov phase and its positive memory penalty coefficient.	p. 8
$\mathcal{Z}_k$	Spatial exponential modular.	p. 23
$\mathcal{P}_*$	Final predictable stopping partition.	(3.2)
$g^{(1)}$, $g^{(2)}$	Revealed and unrevealed pieces of the terminal divergence correction.	(8.2)
$J^{(1)}$, $J^{(2)}$	Corresponding physical correction currents.	pp. 31–32
D_*	Uniform bound on physical terminal-leaf diameters.	(8.1)

Symbol	Meaning	Definition
$K_{\mathbb{T}}$, $d_{\mathbb{T}}$	Periodic Green vector kernel and fixed physical torus distance.	(8.4)
$K_{x,h}$, $\sigma_*^2(x)$	Fixed-grid averaged kernel and conditional variance of the terminal bridge current.	(8.5)

3 Gaussian increments along predictable partitions

This section describes how the noise is revealed. The partition is refined predictably, and Lemma 3.1 gives the conditional Gaussian law of the new masses at each split. Lemma 3.2 then bounds all revealed cell averages simultaneously, although the refinement tree depends on W.

Start with the full fundamental square, understood modulo opposite sides, as a single root Q_0; the mean-zero convention gives $W(Q_0) = 0$. The identified sides form a coordinate seam, not a physical boundary. A local ray current on Q_0 has zero trace across that seam and thus defines a periodic current. At generation k a finite partition $\mathcal{P}_k$ and the masses $\{W(Q) : Q \in \mathcal{P}_k\}$ have been revealed. The next subdivision of each active Q is selected using only this history, then its child masses are revealed. On inactive cells nothing is revealed.

Set

$$f_k = P_{\mathcal{P}_k} W = \sum_{Q \in \mathcal{P}_k} \frac{W(Q)}{|Q|} \mathbf{1}_Q, \qquad \Delta f_k = f_{k+1} - f_k.$$

For a child $S \subset Q$, define the new mass

$$\zeta_S = W(S) - \frac{|S|}{|Q|} W(Q).$$

Lemma 3.1. *Conditionally on the old history and chosen subdivision,*

$$\mathbb{E}_k \zeta_S = 0, \qquad \mathbb{E}_k(\zeta_S \zeta_T) = |S \cap T| - \frac{|S||T|}{|Q|}, \tag{3.1}$$

with conditional independence between distinct active parents. If $\mathcal{P}_*$ is a finite predictable stopping partition and $g \in L^2(\mathbb{T}^2)$ is deterministic, then

$$\operatorname{Var}\left((W - P_{\mathcal{P}_*} W)(g) \mid \mathcal{F}_*\right) = \sum_{Q \in \mathcal{P}_*} \int_Q |g - g_Q|^2, \qquad g_Q = \fint_Q g. \tag{3.2}$$

The conditional mean of this residual is zero.

Proof. Inside a parent Q, subtract from $\mathbf{1}_S$ its projection onto the constants on Q. The resulting test function is $\mathbf{1}_S - (|S|/|Q|)\mathbf{1}_Q$, which is orthogonal to the previously revealed parent mass. Isonormality gives (3.1). The new test functions are chosen predictably, hence their coordinates are Gaussian conditional on the old sigma-field. Apply this decomposition successively. At a stopping time pad the tree to a deterministic depth, freezing all inactive leaves; conditional Gaussian orthogonal projection on the resulting, random but predictably generated finite-dimensional subspace proves (3.2). Here the test function g is deterministic; for random tests measurable with respect to the final reveal history, the pairing with the isonormal residual would first have to be defined conditionally. The identity would fail if a future subdivision were selected *after* inspecting unrevealed bridge coefficients. □

If the new density is sent through any predictable linear kernel K_x, then $Z(x) = \Delta f_k(K_x)$ satisfies

$$\mathbb{E}_k e^{u \cdot Z(x)} = \exp\left(\tfrac{1}{2} u^T \operatorname{Cov}_k(Z(x)) u\right), \qquad u \in \mathbb{R}^2. \tag{3.3}$$

There is no Bernstein restriction on the tilt u. The adaptivity is handled through predictability, not by declaring the random tree independent of W.

3.1 Maxima of adaptive Gaussian cell averages

Let $\mathcal{P}_k$ be the partition after the kth reveal, with $\mathcal{P}_0 = \{\mathbb{T}^2\}$, and let $\mathcal{F}_k$ contain the partitions and masses revealed through that time. The refinement $\mathcal{P}_{k+1}$ is selected $\mathcal{F}_k$-measurably, before its new cell masses are observed. For a deterministic point x outside the cell boundaries, let $Q_k(x)$ be the cell of $\mathcal{P}_k$ containing x, and put

$$M_k(x) = \frac{W(Q_k(x))}{|Q_k(x)|}, \qquad v_{k+1}(x) = \frac{1}{|Q_{k+1}(x)|} - \frac{1}{|Q_k(x)|}. \tag{3.4}$$

An inactive cell is retained in subsequent partitions, so that $v_{k+1}(x) = 0$ on an inactive branch. The following estimate applies uniformly to this random refinement tree.

Lemma 3.2. *Let $\mathcal{P}_0, \dots, \mathcal{P}_J$ be a finite predictably refined partition sequence of $\mathbb{T}^2$, with inactive cells frozen and every nonempty cell of area at least $c_* s^2$. For $Q \in \mathcal{P}_k$ put $M(Q) = W(Q)/|Q|$. There are constants $C, c > 0$ depending only on c_* such that for every $b > 0$,*

$$\mathbb{P}\Big(\max_{0 \le k \le J} \max_{Q \in \mathcal{P}_k} |M(Q)| > \frac{b}{s} \Big) \le C s^{-2} \exp(-c b^2). \tag{3.5}$$

In particular, if $s = 2^{-m} m^3$, then

$$\max_{k,Q} \frac{|W(Q)|}{|Q|} \le C \frac{\sqrt{m}}{s} \quad \textit{with probability } 1 - o(1). \tag{3.6}$$

Proof. Fix a deterministic x for which (3.4) is well defined at every stage. Conditional on $\mathcal{F}_k$, the chosen subdivision of $Q_k(x)$ is fixed, and the centered child bridge is Gaussian. Hence $M_k(x)$ is a martingale with conditional increment variance $v_{k+1}(x) \ge 0$. The predictable variance telescopes *pathwise*:

$$\sum_{j=0}^{k-1} v_{j+1}(x) = |Q_k(x)|^{-1} - |Q_0(x)|^{-1} \le (c_* s^2)^{-1}.$$

For each fixed $u \in \mathbb{R}$, the exponential

$$\exp\Big\{ u(M_k(x) - M_0(x)) - \tfrac{1}{2} u^2 \sum_{j<k} v_{j+1}(x) \Big\}$$

is a nonnegative martingale. Apply the exponential maximal inequality, first with $u > 0$ and then with $-u$, and optimize at $u \asymp bs$ to obtain

$$\mathbb{P}\big(\max_{k \le J} |M_k(x)| > b/s \big) \le 2 \exp(-c b^2). \tag{3.7}$$

At the root $M_0(x) = W(\mathbb{T}^2) = 0$. If *any* revealed cell at any time has a large average, then the spatial set of x for which the running maximum is large contains the entire cell and hence has measure at least $c_* s^2$. Consequently, for the event B on the left of (3.5),

$$\mathbf{1}_B \le (c_* s^2)^{-1} \int_{\mathbb{T}^2} \mathbf{1}_{\{\max_k |M_k(x)| > b/s\}} \, \mathrm{d}x.$$

Take expectations and use (3.7). For $b = A\sqrt{m}$, $s^{-2} = 2^{2m} m^{-6}$, and A large enough that $cA^2 > 2 \log 2$, the upper bound is $o(1)$. □

No independence between the random partition and W is assumed.

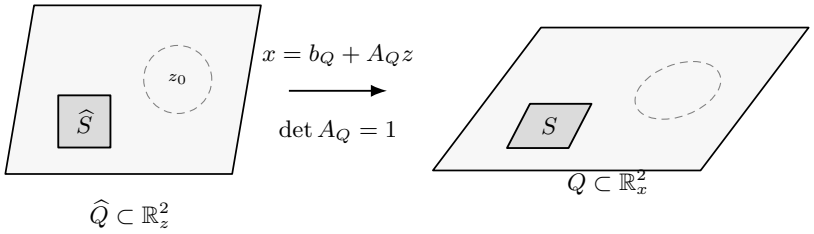


Figure 1: The predictable intrinsic chart. Whitney squares and kernel estimates are defined on $\widehat{Q}$; their physical images are generally parallelograms. The local affine map preserves area, but not Euclidean lengths. The dashed intrinsic ball becomes an ellipse in physical coordinates.

4 Local divergence inverses and Whitney partitions

At each refinement step the new current is obtained by inverting the divergence on an active parent cell and projecting onto the children of that cell. This section constructs the two geometric ingredients. The first is a ray inverse of the divergence, defined in an intrinsic chart attached to the predictable routing metric and transported to physical coordinates by a Piola transform. The second is a finite Whitney partition of the parent in the same chart, in which clipped boundary pieces are attached to full receiver squares. The section ends with the projected-kernel bounds of Lemma 4.2, which are the geometric input for the covariance estimates of Section 5. Proofs of some of the geometric statements are collected in Appendix A.

Physical and intrinsic coordinates. The current U_k, the two coordinate directions, and the phase $H(U_k)$ always live in the original physical coordinates on $\mathbb{T}^2$. The local inverse of divergence, however, is constructed in a chart chosen using the *old* Gaussian history. Fix an active physical parent Q, regarded as a convex set in one chosen lift of the torus, and let G_Q^+ be its new predictable routing metric. Set

$$A_Q = (G_Q^+)^{1/2} = A_Q^T, \quad \det A_Q = 1, \quad z = A_Q^{-1}(x - b_Q), \quad \widehat{Q} = A_Q^{-1}(Q - b_Q), \quad \widehat{S} = A_Q^{-1}(S - b_Q). \tag{4.1}$$

Here b_Q is an old-measurable choice of origin, and all displayed quantities are local to this parent. In particular, A_Q is generally not an automorphism of the torus. The intrinsic distance is

$$d_{G_Q^+}(x, y) = |A_Q^{-1}(x - y)|, \qquad d_{G_Q^+}(x, \partial Q) = \operatorname{dist}(A_Q^{-1}(x - b_Q), \partial\widehat{Q}).$$

The physical child $S = b_Q + A_Q\widehat{S}$ has $|S| = |\widehat{S}|$, but need not be a Euclidean square (Figure 1). All local radii, Whitney lengths, Hölder seminorms and point-source kernel estimates below are in the *intrinsic* z-chart. There is no uniform Euclidean physical kernel bound independent of the eccentricity of A_Q.

The transformed parent $\widehat{Q}$ is a uniformly shape-regular convex parallelogram: before updating the metric the active parent is a square in its birth chart, and successive metrics have a uniformly bounded relative Loewner distortion. Let $D = \operatorname{diam}(\widehat{Q})$ and choose an inball $B(z_0, c_0 D) \subset \widehat{Q}$. Here $K > 1$ denotes the fixed bound on distortion between a cell's birth metric G and the updated metric G^+: $K^{-1}G \preceq G^+ \preceq KG$ (see (5.6)). The inball ratio $c_0 = c_0(K) > 0$ is independent of the physical eccentricity of the chart.

Take a fixed nonnegative smooth probability bump ω supported in $B(0, c_0/4)$ and define

$$\widehat{\omega}_Q(w) = D^{-2}\omega((w - z_0)/D), \qquad \int_{\widehat{Q}} \widehat{\omega}_Q = 1, \quad \|D^j\widehat{\omega}_Q\|_\infty \le C_j D^{-2-j} \quad (j = 0, 1).$$

Define the *intrinsic* ray inverse by

$$\widehat{K}_Q(z, w) = (z - w)\int_1^\infty t\widehat{\omega}_Q(w + t(z - w))\, dt, \qquad \widehat{\mathcal{B}}_Q\widehat{f}(z) = \int_{\widehat{Q}} \widehat{K}_Q(z, w)\widehat{f}(w)\, dw. \tag{4.2}$$

For mean-zero $\widehat{f}$, integration along the ray proves $\operatorname{div}_z \widehat{\mathcal{B}}_Q \widehat{f} = \widehat{f}$ and zero normal trace on $\partial\widehat{Q}$. The *physical* Piola kernel and inverse are

$$K_Q^{\mathrm{phys}}(x,y) = A_Q \widehat{K}_Q(z,w), \qquad \mathcal{B}_Q^{\mathrm{phys}} f(x) = A_Q \widehat{\mathcal{B}}_Q \widehat{f}(z), \quad \widehat{f}(w) = f(b_Q + A_Q w), \tag{4.3}$$

where $z = A_Q^{-1}(x - b_Q)$ and $w = A_Q^{-1}(y - b_Q)$. Because $\det A_Q = 1$, no Jacobian factor is missing. A change of variables and the Piola identity give

$$\int_Q \mathcal{B}_Q^{\mathrm{phys}} f \cdot \nabla_x \phi \, dx = \int_{\widehat{Q}} \widehat{\mathcal{B}}_Q \widehat{f} \cdot \nabla_z \widehat{\phi} \, dz = -\int_Q f \phi \, dx, \qquad \int_Q f = 0. \tag{4.4}$$

The zero extension across the physical boundary has no flux. The inverse acts on the convex parent, not on its potentially disconnected terminal children.

Intrinsic kernel bounds. The scaled bump gives, for $z \neq w$,

$$|\widehat{K}_Q(z,w)| \le \frac{C}{|z-w|}, \qquad |D_z \widehat{K}_Q(z,w)| + |D_w \widehat{K}_Q(z,w)| \le \frac{C}{|z-w|^2}. \tag{4.5}$$

Furthermore,

$$\widehat{K}_Q(z,w) = 0 \quad \text{if} \quad |z-w| > C \operatorname{dist}(z, \partial\widehat{Q}). \tag{4.6}$$

Both properties are proved in Lemma A.3 (Appendix A.2): (4.5) follows by differentiating the normalized bump under the integral, and (4.6) from convexity, since a ray point $w + t(z-w)$ in the central bump forces $\operatorname{dist}(z, \partial\widehat{Q}) \ge cD/t$ while $|z-w| \le D/t$. In physical variables the uniform bounds read

$$|A_Q^{-1} K_Q^{\mathrm{phys}}(x,y)| \le \frac{C}{d_{G_Q^+}(x,y)}, \quad |A_Q^{-1} D_x K_Q^{\mathrm{phys}}(x,y) A_Q| + |A_Q^{-1} D_y K_Q^{\mathrm{phys}}(x,y) A_Q| \le \frac{C}{d_{G_Q^+}(x,y)^2}.$$

The support condition is likewise expressed in $d_{G_Q^+}$. For example $A_Q = \operatorname{diag}(L, L^{-1})$ makes the raw physical kernel as large as L times its intrinsic counterpart, and the bound $|K_Q^{\mathrm{phys}}(x,y)| \lesssim |x-y|^{-1}$ can fail by a factor of order L^2. Determinant one preserves area, not Euclidean lengths.

4.1 Whitney partitions in intrinsic coordinates

Truncating a Whitney decomposition and completing it by boundary sets of diameter $O(s)$ and area comparable to s^2 is the geometric construction of [5, Lemma 5.1 and Appendix A], also recalled in [18, Section 2.3, Lemma 2.1]. Here it is carried out inside each predictably chosen intrinsic parent. The explicit receiver construction below also gives the child-size cap and the constants, uniform under successive changes of chart, which the adaptive construction requires.

The constants $\theta \in (0,1)$ and $D_0 > 0$ denote, respectively, the upper fraction of a parent's intrinsic diameter allowed for a Whitney square and the terminal stopping threshold $D \le D_0 s$. The distinct contraction parameter $\vartheta \in (0,1)$ bounds the *birth-chart* size ratio at an actual refinement: $\ell_{j+1} \le \vartheta \ell_j$, with one corresponding diameter contraction for a newly terminal union. It is achieved by choosing θ small and D_0 large after K is fixed; Lemma A.1 gives the precise constants.

Lemma 4.1. *Fix a shape parameter $c_0 > 0$ and a sufficiently small fixed $\theta > 0$. There are γ, D_0, C depending only on c_0, θ so that, whenever $D > D_0 s$, the* intrinsic *parallelogram $\widehat{Q}$ admits a disjoint measurable partition into children $\widehat{S}$ with scales $\ell_S \ge s$:*

$$|\widehat{S}| \asymp \ell_S^2, \quad \operatorname{diam}(\widehat{S}) \le C\ell_S, \quad \ell_S \asymp \max\{s, \operatorname{dist}(z, \partial\widehat{Q})\} \quad (z \in \widehat{S}). \tag{4.7}$$

Nonterminal children are intrinsic *dyadic squares with $\ell_S \le \theta D$; terminal children have $\ell_S = s$, area at least s^2 and intrinsic diameter at most Cs. After choosing D_0 sufficiently large, the maximal intrinsic child diameter is a fixed small fraction of D. The physical partition consists of $S = b_Q + A_Q \widehat{S}$. If $D \le D_0 s$, the parent is left terminal.*

Since $\det A_Q = 1$, the physical children have the same areas as the intrinsic ones, but not in general the same Euclidean diameters.

Proof. Throughout this proof write Q, S, R, x for *intrinsic* subsets and points of $\widehat{Q}$; they are not physical coordinates. Use a dyadic lattice in this predictable intrinsic chart, with smallest squares of side s. Take *all inclusion-maximal* dyadic squares $R \subset Q$ with side $\ell_R \in [2s, \theta D]$ and $\operatorname{dist}(R, \partial Q) \geq \gamma \ell_R$. Since dyadic squares are nested or disjoint, these maximal squares have disjoint interiors. A maximal square's dyadic parent fails admissibility: it either is not contained in Q, violates the distance condition, or exceeds θD. These three alternatives imply $\operatorname{dist}(x, \partial Q) \lesssim_{\gamma,\theta} \ell_R$ for $x \in R$. The reverse inequality follows from admissibility. More explicitly, for a square of side ℓ, if its 2ℓ parent is not contained in Q, every point of the child is within $2\sqrt{2}\ell$ of the boundary; if the parent fails the distance requirement, the same bound is $(2\gamma + 2\sqrt{2})\ell$; if it exceeds the cap, then $D < 2\ell/\theta$. This proves the upper Whitney comparison, with constants fixed before the Gaussian experiment, so the nonterminal squares satisfy (4.7).

Every remaining point of Q belongs to an uncovered s-lattice square. If this s-square is wholly inside Q, retain it as a full uncovered square. Its $2s$-ancestor cannot be admissible: otherwise that ancestor would lie in some inclusion-maximal admissible square and would cover the s-square. The cap cannot fail since $\theta D > 4s$. If the ancestor fails containment or boundary clearance, every point of the uncovered s-square lies within $(2\gamma + 3\sqrt{2})s$ of ∂Q. In particular every full uncovered square is within Cs of the boundary of Q.

It remains to assign a *clipped* piece $P = Q \cap R$, where an s-square R meets ∂Q, to a full uncovered s-square. Choose $x \in P$. Then $\operatorname{dist}(x, \partial Q) \leq \sqrt{2}s$. Let z_0 be the center of an inball $B(z_0, c_0 D) \subset Q$ and set $x_t = (1-t)x + tz_0$ with $t = As/D$. Choose a constant $A = A(c_0)$ large enough that $c_0 As > 5s$, and choose $D_0 > 2A$ to ensure $t < 1/2$. Convexity and the supporting-line characterization of boundary distance give

$$5s < \operatorname{dist}(x_t, \partial Q) \leq (\sqrt{2} + A)s, \qquad |x_t - x| \leq As.$$

The s-lattice square R' containing x_t is therefore fully inside Q. Take γ larger than $(A + \sqrt{2})/2$ (and enlarge it by an absolute constant if necessary). Then R' cannot lie in an admissible dyadic square of side at least $2s$, because any such square would put x_t at boundary distance at least $2\gamma s$. Hence R' is an uncovered full receiver within $O(s)$ of P.

Choose one receiver by a deterministic tie-break rule for every clipped s-square and join its clipped piece to that receiver. A receiver can get only $O_A(1)$ pieces because their originating lattice squares lie within $O(As)$; the union has area at least s^2, diameter $O_A(s)$, and all points have boundary distance $O_A(s)$. Any full uncovered square not receiving a clipped piece is a terminal child by itself. The maximal squares and these terminal unions partition Q modulo a fixed measure-zero boundary convention; see Figure 2.

Choose D_0 only after A, γ, θ and the terminal diameter constant C_A have been fixed, so that $C_A s \leq \theta D$ whenever $D > D_0 s$. The exact old/new-chart side ratio and the metric transport factor are recorded in Lemmas A.1 and A.2. In particular, choose θ and then D_0 so that the birth-side shrink factor ϑ in those lemmas satisfies $\vartheta\sqrt{K} < 1$. All selections can be made by a deterministic lexicographic rule. The construction terminates in at most $C_K \log(1/s) = O(m)$ steps: every nonterminal square has intrinsic area at most a fixed fraction of its parent's area, and every nonempty cell has area at least cs^2. No lower bound on the area of an individual clipped fragment is needed. □

The half-open boundary convention and the finite receiver assignment are spelled out in Lemma A.4.

4.2 Boundary receivers and the choice of constants

We fix the constants in the following order. Fix an inball ratio c_0. First take $A > 5/c_0$ and $\gamma > (A + \sqrt{2})/2$. Next choose $\theta > 0$ small enough for subsequent metric contraction. Finally take

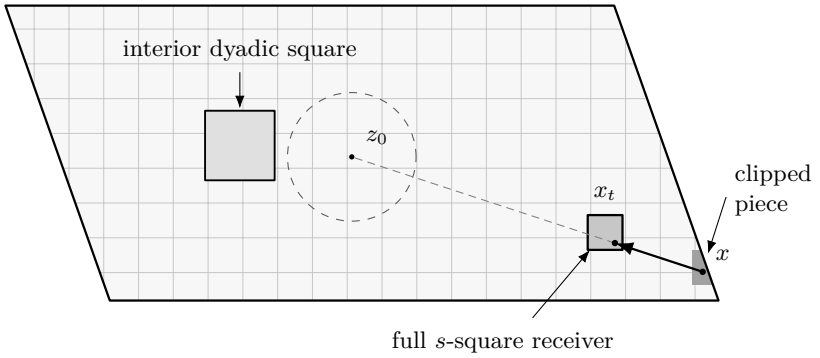


Figure 2: Clipped-square receivers in an intrinsic chart (schematic, not to scale). A boundary s-square may have an arbitrarily small intersection with $\widehat{Q}$. Moving x toward the inball center gives x_t, contained in an uncovered full s-square. The clipped fragment is assigned to that receiver; only the resulting union is a terminal child and has area at least s^2.

D_0 so large that $D_0 > 2A$, $\theta D_0 > 4$, and every $O_A(s)$ terminal union has diameter at most a prescribed small fraction of D in the current chart. The constants in the Whitney bounds may depend on c_0, A, γ, θ but *not* on D/s, the lattice translation, or the physical eccentricity of a determinant-one chart.

The two kinds of children behave differently. A maximal admissible square has its full area ℓ^2, whereas a clipped mesh piece may have arbitrarily small area. The receiving full s-square provides the area floor for the union, and only this union is a child of the partition. At an acute corner the receiving square may lie along a different face, but the inball displacement gives bounded distance and bounded multiplicity without any angle-dependent connectivity hypothesis. The operator $\mathcal{B}_Q$ is applied to the convex parent, not to a terminal union. Only the area, diameter and Whitney bounds of the children enter the projected kernel estimates, so neither connectedness nor a John constant for terminal unions is used.

Lattice conventions. All squares are half-open after applying one fixed lexicographic boundary convention. The s-lattice belongs to the predictable intrinsic chart, not to the fixed physical grid; it is translated according to a preassigned measurable rule. At exact lattice contacts we choose one closed representative for the support test and assign measure-zero interfaces arbitrarily. No Gaussian sample is consulted when making any of these choices. A parent with $D \le D_0 s$ is terminal, and no inverse is applied on its children. Lemma 4.1 is thus a finite, measurable construction rather than a limiting Whitney decomposition.

Notation for the projected kernels. We use $0 < \alpha < 1$ for the projected-kernel Hölder exponent. The Gaussian field and its spatial memory use a fixed smaller exponent $0 < \beta < \alpha$; constants may depend on both exponents. Here G_Q^+ is the updated physical covariance metric, $A_Q = (G_Q^+)^{1/2}$, $z = A_Q^{-1}(x - b_Q)$ and $\widehat{S} = A_Q^{-1}(S - b_Q)$. The length ℓ_S and every distance in Lemma 4.2 are measured in this updated intrinsic chart. The averaging operator $\widehat{P}$ is intrinsic; a physical vector is recovered only after multiplying by A_Q. The fixed physical distance of the torus occurs only in Section 8.

The cell-averaged singular-kernel estimates are related to the near- and far-cell bounds in [18, Section 2.4, Lemma 2.6] and to the covariance calculation for the histogram correction in [5, Section 4, Step 3]. In the present argument, the boundary-distance support (4.6) and the exact Gaussian covariance identity (6.2) give a pointwise conditional variance bound uniform in the intrinsic scale. The projected-kernel bounds needed for this are the following.

Lemma 4.2. *Let $\widehat{P}$ average over the children $\widehat{S}$ in the intrinsic chart $\widehat{Q}$, and let z, z' belong to*

one $\widehat{S}$ of intrinsic scale ℓ. For every fixed $0<\alpha<1$,

$$\|\widehat{P}\widehat{K}_Q(z,\cdot)\|_2 \le C, \qquad \|\widehat{P}\widehat{K}_Q(z,\cdot)\|_\infty \le C\ell^{-1}, \tag{4.8}$$
$$\|\widehat{P}(\widehat{K}_Q(z,\cdot)-\widehat{K}_Q(z',\cdot))\|_2 \le C(|z-z'|/\ell)^\alpha, \quad \|\widehat{P}(\widehat{K}_Q(z,\cdot)-\widehat{K}_Q(z',\cdot))\|_\infty \le C\ell^{-1}(|z-z'|/\ell)^\alpha.$$

The same inequalities hold after subtracting the parent-mean projection. They are uniform in the intrinsic chart.

Lemma 4.2 concerns the intrinsic kernel only; the passage to physical vectors is discussed after (4.12).

Proof. In this proof, for compactness, write Q,S,T,x,y and K_Q,P for the intrinsic objects $\widehat{Q},\widehat{S},\widehat{T},z,w$ and $\widehat{K}_Q,\widehat{P}$. Set $d_Q(x)=\operatorname{dist}(x,\partial Q)$. The Whitney bounds give $d_Q(x)\le C_W\ell$ for $x\in S$; by the support restriction, $K_Q(x,\cdot)$ is supported in $B(x,C\ell)$. If a child T meets this ball, choose $y\in T\cap B(x,C\ell)$. The distance-to-boundary function is 1-Lipschitz, so $\ell_T\le C_W\max\{s,d_Q(y)\}\le C\ell$. Thus $\operatorname{diam}T\le C\ell$ and the union of projected-support children lies in $B(x,C'\ell)$ and has area at most $C\ell^2$.

For the one-point L^∞ bound we separate near and far children. Choose $c_1>0$ smaller than the inverse of 10 times every fixed Whitney comparison constant. For any child T meeting $B(x,c_1\ell)$,

$$|T|\ge c\ell^2. \tag{4.9}$$

Indeed, if $\ell=s$ this is the global area floor $|T|\ge s^2$. If $\ell>s$, then S is a nonterminal square and admissibility implies $d_Q(x)\ge\gamma\ell$; the 1-Lipschitz property gives $d_Q(y)\ge\gamma\ell/2$ for $y\in B(x,c_1\ell)$. The pointwise Whitney comparison on T implies $\ell_T\ge c\ell$, and $|T|\ge c\ell_T^2$. For these near children,

$$\left|\fint_T K_Q(x,y)\,dy\right| \le \frac{C}{|T|}\int_{B(x,C'\ell)}\frac{dy}{|x-y|} \le \frac{C\ell}{\ell^2}=\frac{C}{\ell}.$$

On each other projected-support child T, all $y\in T$ satisfy $|x-y|\ge c_1\ell$, so the same bound follows pointwise from $|K_Q(x,y)|\le C/|x-y|$. Multiplication by the square root of the $O(\ell^2)$ projected-support area proves the L^2 estimate.

For the two-point estimates put $d=|x-x'|$. If $d\ge c_1\ell/4$, apply the two one-point estimates and enlarge the constant. Suppose $0<d<c_1\ell/4$. The combined supports lie in $B(x,C''\ell)$. In the two disks of radius $2d$ about the two poles, the two singular kernels have L^1 norm $O(d)$. Outside the disks, the mean-value estimate for $\nabla_x K_Q$ gives

$$\int_Q |K_Q(x,y)-K_Q(x',y)|\,dy \le Cd(1+\log(\ell/d)). \tag{4.10}$$

For a projected child meeting $B(x,c_1\ell/2)$, use (4.9) and divide (4.10) by $|T|$. For any other child the segment joining x,x' remains at distance at least $c_1\ell/4$ from that child; integrate the pointwise gradient bound to get Cd/ℓ^2. Hence

$$\|P(K_Q(x,\cdot)-K_Q(x',\cdot))\|_\infty \le C\ell^{-1}\frac{d}{\ell}(1+\log(\ell/d)) \le C_\alpha\ell^{-1}(d/\ell)^\alpha.$$

The projected support of the difference has area $O(\ell^2)$, which gives its L^2 bound. Subtracting the parent-constant projection changes the bounds by at most a fixed factor, since $P_QP=P_Q$ and a constant projection is an L^2 contraction; its L^∞ norm is bounded by that of the projected kernel. The same two-pole calculation on a whole parent gives the translation estimate

$$\int_Q |K_Q(x,y)-K_Q(x',y)|\,dy \le Cd(1+\log(D/d)), \qquad 0<d\le D/2. \tag{4.11}$$

□

Projected Piola evaluations. Let P^{phys} average over the physical children $S = b_Q + A_Q\widehat{S}$. Since $\det A_Q = 1$, projection commutes with pullback and multiplication by the old-measurable matrix:

$$A_Q^{-1}P^{\mathrm{phys}}K_Q^{\mathrm{phys}}(x,\cdot) = (\widehat{P}\widehat{K}_Q(z,\cdot)) \circ A_Q^{-1}(\cdot - b_Q), \qquad z = A_Q^{-1}(x - b_Q). \tag{4.12}$$

The L^2 and L^∞ norms of the normalized vector $A_Q^{-1}P^{\mathrm{phys}}K_Q^{\mathrm{phys}}$ are therefore controlled by (4.8). The corresponding raw physical vector can cost a factor $\|A_Q\|_{\mathrm{op}}$; this factor must remain in physical covariance and componentwise bounds. The parent mean subtraction enjoys the same commutation property.

5 Anisotropic covariance selection

This section defines the target metric $C(y)$ toward which the covariance of each fresh Gaussian increment is steered, and the lag $L(y, G)$ measuring how far a current metric G is from it. Lemma 5.1 collects the bounds on C, on L and on the spectral clipping $\mathcal{T}_C$ that the bootstrap of Section 6 uses.

Recall the phase H and its gradient v from (2.5). The continuous-time exponential Bellman trial has normalized Hessian $\zeta_{\mathrm{B}}D^2H + \zeta_{\mathrm{B}}^2 vv^T$, as recalled in (2.10). For the discrete construction we use instead a regularized Hessian proxy, whose diagonal part also dominates the curvature of H. On the critical line its determinant is comparable to H^2, so there is no large-amplitude factor $H^{-\sigma}$ to spare. The metric must therefore be adjusted dynamically.

Define

$$B(y) = \mathrm{diag}(H^{1-2a_1}, H^{1-2a_2}) + v(y)v(y)^T, \qquad C(y) = \sqrt{\det B(y)}B(y)^{-1}. \tag{5.1}$$

In two dimensions, $\det C = 1$. To interpret this choice, for any positive definite G with $\det G = 1$, the arithmetic–geometric mean inequality applied to $B^{1/2}GB^{1/2}$ gives

$$\mathrm{tr}(BG) \geq 2\sqrt{\det B} = \mathrm{tr}(BC),$$

with equality precisely when $G = C$. We measure the mismatch of an old metric G from the target by

$$L(y, G) = \sqrt{\mathrm{tr}(GC(y)^{-1})}. \tag{5.2}$$

Since $C^{-1} = B/\sqrt{\det B}$, the lag is exactly the relative covariance cost:

$$\frac{\mathrm{tr}(B(y)G)}{\mathrm{tr}(B(y)C(y))} = \frac{L(y,G)^2}{2}. \tag{5.3}$$

Thus $L \geq \sqrt{2}$, with equality precisely at $G = C(y)$. The matrix G will bound the fresh physical covariance from above (see (6.3)), whereas the spatial distance is defined by G^{-1}, as in (4.1).

Lemma 5.1. *On* (2.2)*, with constants depending on the fixed exponents,*

$$\det B \asymp H^2, \qquad C_{ii} \lesssim H^{2a_i}, \qquad \frac{vv^T}{H} \preceq C_0 C(y)^{-1}. \tag{5.4}$$

If $G_{ii} \leq MH(y)^{2a_i}$ *and* $\det G = 1$*, then*

$$L(y,G)^2 \leq C_M H(y), \qquad |\partial_{y_i} L(y,G)| \leq C_M H(y)^{1/2-a_i}. \tag{5.5}$$

For a fixed $K > 1$*, the determinant-one spectral clipping*

$$\mathcal{T}_C(G) = G^{1/2}\,\mathrm{clip}_{[K^{-1},K]}(G^{-1/2}CG^{-1/2})G^{1/2}$$

satisfies, with $G^+ = \mathcal{T}_C(G)$*,*

$$K^{-1}G \preceq G^+ \preceq KG, \quad G^+ \preceq K^{-1}G + C, \quad L(y,G^+) \leq K^{-1/2}L(y,G) + C_K. \tag{5.6}$$

Proof. Write $D_i = H^{1-2a_i}$. The rank-one determinant formula gives

$$\det B = D_1 D_2 \left(1 + \sum_i \frac{v_i^2}{D_i}\right) = H\left(1 + \sum_i v_i^2 H^{-1+2a_i}\right) \asymp H^2,$$

where the last comparison follows by examining the dominant summand $F_i(y_i) \geq H/2$. The formula for the inverse of a 2×2 positive matrix gives $C_{ii} = B_{jj}/\sqrt{\det B} \lesssim H^{2a_i}$. Also

$$C^{-1} = \frac{B}{\sqrt{\det B}} = \chi \operatorname{diag}(H^{-2a_i}) + ww^T, \quad \chi = \frac{H}{\sqrt{\det B}} \asymp 1, \quad w = \frac{v}{(\det B)^{1/4}}.$$

One has $|\partial_i \chi| \lesssim H^{-a_i}$ and $|\partial_i w_j| \lesssim H^{1/2-a_i-a_j}$. Represent L as the Euclidean norm of the direct sum of $\sqrt{\chi}\, G_{ii}^{1/2} H^{-a_i}$ and $G^{1/2}w$. Under the diagonal caps, differentiation of this vector proves (5.5) without dividing by a possibly small L.

The two eigenvalues of $G^{-1/2}CG^{-1/2}$ are reciprocal, and clipping both to $[K^{-1}, K]$ preserves their product. The inequality $\operatorname{clip}(t) \leq K^{-1} + t$ yields the Loewner bound. Along the common eigendirections the logarithmic mismatch is shortened by $\log K$. Separating the regimes of large and bounded log-mismatch, this gives $L(y, G^+) \leq K^{-1/2} L(y, G) + C_K$. □

6 The adaptive construction and the bootstrap

We now build the adaptive field U_k generation by generation and prove that it stays in the regime where the estimates of Sections 4 and 5 apply. At each active parent the metric of the next increment is obtained by clipping the current metric toward the target C of Section 5; the fresh Gaussian increment is then pushed forward by this metric. The main result, Proposition 6.5, shows by induction that, as long as no envelope is too large and the modular has not overshot, the diagonal caps, the relative phase comparisons and a contraction of the lag-plus-memory functional S_k hold simultaneously. These are the inputs of the exponential moment estimates of Section 7.

Throughout, every stage $k \to k+1$ has the same order. First the geometry is chosen from the old history: the representative, the metric G_Q^+, the intrinsic parent and its Whitney partition, and the stopping decisions are $\mathcal{F}_k$-measurable. Only then are the new Gaussian child masses revealed. Every quantity called old, predictable or old-measurable below is fixed before the noise of the current stage.

Initialize $U_0 = 0, G_0 = I, \mathcal{P}_0 = \{Q_0\}$. At an active physical parent Q in a chosen lift, choose a representative x_Q by a fixed old-measurable rule, and set $C_Q = C(U_k(x_Q))$, $G_Q^+ = \mathcal{T}_{C_Q}(G_k)$ and $A_Q = (G_Q^+)^{1/2}$. This matrix, the intrinsic parent $\widehat{Q}$, its bump and its Whitney partition are all fixed before the new child masses are revealed. After revelation let $\Delta f_k = (P_{\mathcal{P}_{k+1}} - P_{\mathcal{P}_k})W$ be the physical, parentwise mean-zero step density and put

$$\widehat{\Delta f}_k(w) = \Delta f_k(b_Q + A_Q w), \qquad Z_{k+1}(z) = \int_{\widehat{Q}} \widehat{K}_Q(z, w) \widehat{\Delta f}_k(w)\, dw.$$

Thus Z_{k+1} is an intrinsically normalized Gaussian vector field. The physical increment is its Piola push-forward:

$$V_{k+1}(x) = \sqrt{\delta}\, A_Q Z_{k+1}(A_Q^{-1}(x - b_Q)), \qquad U_{k+1} = U_k + V_{k+1}, \qquad \operatorname{div}_x U_{k+1} = \sqrt{\delta}\, f_{k+1}. \quad (6.1)$$

Every such increment has zero normal trace on its physical parent, so its zero extension is globally periodic. In particular the local matrix $A_Q \in \mathrm{SL}(2, \mathbb{R})$ is never used as a global coordinate change on $\mathbb{T}^2$.

Lemma 6.1. *Conditional on the old filtration and on the predictably selected chart, $Z_{k+1}(z)$ is a centered Gaussian vector with covariance matrix*

$$\operatorname{Cov}_k(Z_{k+1}(z))_{ab} = \int_{\widehat{Q}} [(\widehat{P} - \widehat{P}_{\widehat{Q}})\widehat{K}_{Q,a}(z, \cdot)](w)\, [(\widehat{P} - \widehat{P}_{\widehat{Q}})\widehat{K}_{Q,b}(z, \cdot)](w)\, dw. \quad (6.2)$$

In particular,

$$\operatorname{Cov}_k Z_{k+1}(z) \preceq CI, \qquad \operatorname{Cov}_k V_{k+1}(x) \preceq C\delta G_Q^+, \qquad \operatorname{div}_x V_{k+1} = \sqrt{\delta}\,\Delta f_k. \tag{6.3}$$

Proof. Fix the old history. The chart A_Q and the partition are then deterministic. Pulling back a white-noise test g by $x = b_Q + A_Q w$ preserves its L^2 inner product: the Jacobian is $\det A_Q = 1$. The conditional increment Δf_k is precisely the orthogonal Gaussian projection onto the span of child-constant functions with their parent constant removed. Pairing it with the two components of $\widehat{K}_Q(z,\cdot)$ gives the Gram matrix (6.2). Lemma 4.2 provides the intrinsic L^2 bound, so the Gram matrix is $\preceq CI$. The covariance identity $\operatorname{Cov}_k(V) = \delta A_Q \operatorname{Cov}_k(Z) A_Q^T$ then proves the physical bound, since $A_Q A_Q^T = G_Q^+$. For the divergence, $D_x V = \sqrt{\delta}\, A_Q (D_z Z) A_Q^{-1}$, so $\operatorname{tr} D_x V = \sqrt{\delta}\operatorname{tr} D_z Z = \sqrt{\delta}\widehat{\Delta f}_k$. Equivalently, the weak identity (4.4) proves the result without pointwise differentiation. The Piola normal-flux identity preserves the no-flux boundary condition. □

The physical field does not, in general, have covariance bounded by $C\delta I$; its covariance is controlled only through the anisotropic metric G_Q^+. The next lemma gives uniform sub-Gaussian control of the normalized field Z_{k+1} and of its Hölder seminorm on each child.

Lemma 6.2. *Fix $0 < \beta < \alpha < 1$. At generation k, condition on $\mathcal{F}_k$. Let Q be an active parent, let its intrinsic chart and Whitney children be selected $\mathcal{F}_k$-measurably as above, and let Z_{k+1} be the centered intrinsic field in* (6.1)*. For each newly revealed child $S = b_Q + A_Q \widehat{S}$ of intrinsic scale ℓ_S, choose the continuous version of Z_{k+1} on $\overline{\widehat{S}}$ and set*

$$E_{k+1,S} := \sup_{z \in \overline{\widehat{S}}} |Z_{k+1}(z)| + \ell_S^{\beta} [Z_{k+1}]_{C^{\beta}(\overline{\widehat{S}})}. \tag{6.4}$$

There are deterministic $c, C > 0$, depending only on α, β and the uniform intrinsic Whitney and kernel constants, such that, for every $t \geq 0$,

$$\mathbb{P}(E_{k+1,S} > t \mid \mathcal{F}_k) \leq Ce^{-ct^2}, \qquad \mathbb{E}_k e^{cE_{k+1,S}^2} \leq C, \tag{6.5}$$

possibly after reducing c. These estimates hold for each $\mathcal{F}_k$-measurably selected child and are uniform in k, m, its intrinsic scale, and the eccentricity of A_Q. The Hölder seminorm in (6.4) *is computed with the intrinsic z-distance; its physical pullback uses $d_{G_Q^+}(x,x') = |A_Q^{-1}(x - x')|$.*

If $J_0 \leq C_{\mathrm{dep}} m$ is the deterministic reveal-depth bound, $s = 2^{-m} m^3$, and each child has area at least $c_* s^2$, then for a sufficiently large fixed global envelope cutoff *$R > 0$,*

$$\mathbb{P}(\exists \text{ an actually attempted reveal } (k+1,S) : E_{k+1,S} > R\sqrt{m}) \leq CJ_0 s^{-2} e^{-cR^2 m} \leq e^{-c_R m} \tag{6.6}$$

for all sufficiently large m, with $c_R > 0$.

Proof. We work conditionally on $\mathcal{F}_k$ and write $\mathbb{P}_k(\cdot) = \mathbb{P}(\cdot \mid \mathcal{F}_k)$; in particular the chart, $\widehat{Q}$, the averaging projector and the child $\widehat{S}$ are deterministic. By the Gram formula (6.2) and the projected evaluations (4.8),

$$\sup_{z \in \widehat{S}} \mathbb{E}_k |Z_{k+1}(z)|^2 \leq C_0, \qquad \mathbb{E}_k |Z_{k+1}(z) - Z_{k+1}(z')|^2 \leq C_0 \left(\frac{|z - z'|}{\ell_S} \right)^{2\alpha} \quad (z, z' \in \widehat{S}). \tag{6.7}$$

The same bounds extend to the closure by continuity. Both components are centered Gaussian. The constants do not see the Piola matrix, because (6.7) concerns the normalized field Z_{k+1}.

We bound $E_{k+1,S}$ by a chaining argument. Set $\ell = \ell_S$ and fix an exponent γ with $\beta < \gamma < \alpha$. The intrinsic diameter bound $\operatorname{diam}(\widehat{S}) \leq C_W \ell$ allows us to construct, for every integer $j \geq 0$, a finite $2^{-j}\ell$-net $N_j \subset \overline{\widehat{S}}$ with $|N_j| \leq C 2^{2j}$: take a fixed countable dense set in the child, run a

deterministic greedy procedure with a fixed order, and retain the finitely many selected points. Since the intrinsic chart is predictable, the nets are measurable with respect to the old history. Each point $z \in \overline{\widehat{S}}$ has a representative $z_j \in N_j$ at distance at most $C2^{-j}\ell$. The nets need not be nested, nor need $\widehat{S}$ be connected.

At level zero test the two Gaussian coordinates at all points of N_0. At every level $j \geq 0$ test all coordinate differences between pairs of points in $N_j \cup N_{j+1}$ whose separation is at most $C_1 2^{-j}\ell$, where C_1 is chosen to cover both consecutive-level approximations and pairs of nearest representatives of nearby points. There are at most $C2^{4j}$ such tests. By (6.7), each tested difference has conditional standard deviation at most $C2^{-\alpha j}$. For $A \geq 1$ and $u \geq 0$, let $\mathcal{G}_u$ be the event on which every level-zero coordinate has absolute value at most $A(1+u)$ and every tested difference at level j has absolute value at most $A(1+u)2^{-\gamma j}$. The scalar Gaussian tail bound and a union bound give

$$\mathbb{P}_k(\mathcal{G}_u^c) \leq Ce^{-cA^2(1+u)^2} + C\sum_{j\geq 0} 2^{4j} \exp\Big[-cA^2(1+u)^2 2^{2(\alpha-\gamma)j}\Big] \leq C'e^{-c'u^2}. \tag{6.8}$$

Here A is fixed sufficiently large. The series converges uniformly: since $\alpha - \gamma > 0$, its exponential factor dominates 2^{4j}, and $(1+u)^2 \geq 1 + u^2$ separates the u-tail. The increment threshold is $2^{-\gamma j}$, without a factor growing with j; this is what yields a genuine C^β bound.

On $\mathcal{G}_u$, telescope the values at the representatives z_j of a point z from level zero to arbitrarily fine levels. Consecutive representatives are $O(2^{-j}\ell)$ apart, so every increment occurs among the tested pairs. The convergent geometric series $\sum_{j\geq 0} 2^{-\gamma j}$ gives

$$\sup_{\widehat{S}} |Z_{k+1}| \leq CA(1+u).$$

For z, z' with $d = |z - z'| \leq \ell$, choose j so that $2^{-j}\ell \asymp d$. Their level-j representatives are $O(2^{-j}\ell)$ apart, hence their difference is one of the tested pairs; the two finer telescopes cost at most $CA(1+u)2^{-\gamma j}$. Consequently,

$$|Z_{k+1}(z) - Z_{k+1}(z')| \leq CA(1+u)(d/\ell)^\gamma \leq CA(1+u)(d/\ell)^\beta.$$

For $d > \ell$ the supremum bound gives the same C^β estimate. Taking suprema over a fixed countable dense set and extending to the continuous version makes $E_{k+1,S}$ measurable. We conclude from (6.8) that $\mathbb{P}_k(E_{k+1,S} > CA(1+u)) \leq C'e^{-c'u^2}$, which, after adjusting constants, proves the first inequality in (6.5). Tail integration gives, for c smaller than the tail exponent,

$$\mathbb{E}_k e^{cE_{k+1,S}^2} = 1 + \int_0^\infty 2cte^{ct^2}\mathbb{P}_k(E_{k+1,S} > t)\, dt \leq C.$$

Finally, the area floor implies that at each generation there are at most $(c_* s^2)^{-1}$ children, including terminal unions. Because the next subdivision is $\mathcal{F}_k$-measurable, the conditional bound in (6.5) can be summed over all children selected for the next actual reveal, and then averaged over the past. A further union bound over the J_0 generations gives the first inequality of (6.6). The envelopes on distinct children need not be independent for this argument. Since $s = 2^{-m}m^3$ and $J_0 \leq C_{\mathrm{dep}}m$,

$$CJ_0 s^{-2}e^{-cR^2 m} \leq C'm^{-5}\exp\big[-(cR^2 - 2\log 2)m\big].$$

Choose fixed R with $cR^2 > 2\log 2$. This proves the second inequality in (6.6). □

Local estimates used in the induction. Conditionally on its old history, every active parent satisfies the following, by Lemmas 4.1, 4.2, 6.1 and 6.2. The chart and the entire child partition are predictable, and every child has area at least $c_* s^2$. Active children contract in birth charts, while terminal children have intrinsic diameter $O(s)$; the contraction is quantified in Lemma A.1,

and its use along an entire ancestry is recorded in Lemma 6.3. The increment $V_{k+1} = \sqrt{\delta} A_Q Z_{k+1}$ has zero normal trace on its parent and divergence $\sqrt{\delta}\,\Delta f_k$. For $x = b_Q + A_Q z$ in a child S,

$$\operatorname{Cov}_k V_{k+1}(x) \preceq C\delta G_Q^+, \qquad |V_{k+1,i}(x)| \le \sqrt{\delta}\sqrt{(G_Q^+)_{ii}}\, E_{k+1,S}, \qquad \mathbb{E}_k e^{cE_{k+1,S}^2} \le C,$$

and the same envelope controls the intrinsic C^β seminorm. The constants are independent of the physical eccentricity, whose effect is confined to the factors involving G_Q^+.

Lag and memory. For each x, let $E_j(x)$ be the envelope associated to its stage-j child, or zero on an inactive generation. Fix a deterministic memory discount $0 < \rho_1 < 1$; it is chosen in Section 6.3 so that the square of the geometric ancestor-decay factor (r_{osc} in Lemma 6.3) is strictly smaller than ρ_1. The subscript in ρ_1 is part of its name, not a generation index. Set

$$T_0 = 0, \quad T_{k+1} = \rho_1 T_k + E_{k+1}^2, \quad L_k(x) = L(U_k(x), G_k(x)), \quad S_k = L_k^2 + T_k. \tag{6.9}$$

Both terms of S_k are nonnegative. The memory T_k records old oscillations with an exponentially discounted weight, and L_k^2 measures how far the current routing metric lags behind the instantaneous Bellman metric. The recurrence for T_k applies only at actual reveals: on a frozen leaf, T_k, L_k and S_k remain unchanged, so padding the filtration does not discount the past.

Stopping rules. Fix the deterministic modular stop threshold $T_* > 0$. The construction uses exactly two stopping rules, both checked before the next subdivision is selected. If some previously attempted child envelope exceeded $R\sqrt{m}$, the history has been killed. If the old modular $\mathcal{Z}_k := \int_{\mathbb{T}^2} e^{\Phi_k}$ (see (7.1)) satisfies $\mathcal{Z}_k \ge T_*$, the history has overshot. Otherwise the next subdivision and metric update are chosen, the new child masses are revealed, and their envelopes are computed. If a new envelope exceeds $R\sqrt{m}$, we kill the scalar stopped supermartingale. If an overshoot occurs, we retain the attained modular value and freeze further geometric refinement. The diagonal caps and the relative phase comparisons are not tested by the construction. On every history that has been neither killed nor stopped by overshoot they are consequences of Proposition 6.5. This separates a random failure event from the field estimates valid on successful histories.

Birth and reveal convention. A stage-j square Q_j is born with metric G_j, which is also the physical covariance metric of the increment $V_j = \sqrt{\delta}\, G_j^{1/2} Z_j$. On an active parent Q_j the predictable next metric is G_j^+, and a newly created child is born with $G_{j+1} = G_j^+$. A child created as a clipped terminal union receives that increment once and then freezes. A square stopped directly at the threshold is not updated and reveals no further Gaussian increment. At padded times, the exponent of the ancestor contraction is the number of genuine reveals, not the calendar difference. We use essential suprema across cell boundaries and pointwise continuous versions on each cell interior.

6.1 Auxiliary estimates

Throughout the induction, $0 < \eta < 1$ is the fixed relative phase tolerance per actual Gaussian reveal: $(1-\eta)H(U_{j-1}) \le H(U_j) \le (1+\eta)H(U_{j-1})$. It is chosen small before the Gaussian step size ε. There is no new phase comparison on a frozen stage.

The two lemmas below are the deterministic inputs of the induction. Lemma 6.3 bounds the oscillation of U_b on a cell by the memory T_b, summing only over actual reveals and not over padded calendar times. Lemma 6.4 identifies short anisotropic segments along which the lag may be differentiated using only the old diagonal cap. Neither follows from the local ray-kernel estimate alone.

Lemma 6.3. *Let $Q_0 \supset \cdots \supset Q_b$ be the actual ancestral chain of a point, with exactly one new Gaussian increment V_j revealed when Q_j is born, $1 \le j \le b$. The birth metric G_j and intrinsic length ℓ_j belong to Q_j, and $E_j(Q_j)$ is the intrinsic C^β envelope of the normalized field Z_j, before the Piola factor $G_j^{1/2}$. Suppose the already certified caps and phase inequalities hold at each reveal:*

$$(G_j)_{ii} \le MH(U_j)^{2a_i}, \qquad (1-\eta)H(U_{j-1}) \le H(U_j) \le (1+\eta)H(U_{j-1}).$$

If the actual child contractions and metric comparisons of Lemma A.2 hold, define the effective ancestor-oscillation ratio r_{osc} by

$$r_{\mathrm{osc}} = (\vartheta\sqrt{K})^\beta (1-\eta)^{-a_*}, \qquad r_{\mathrm{osc}}^2 < \rho_1 < 1.$$

Then for x, z in the same stage-b cell (whether this cell is active, newly terminal, or directly stopped),

$$|U_{b,i}(x) - U_{b,i}(z)| \le \frac{C\sqrt{\delta M}}{\sqrt{1 - r_{\mathrm{osc}}^2/\rho_1}} H(U_b(x))^{a_i} \sqrt{T_b(x)}, \qquad T_b(x) = \sum_{j=1}^{b} \rho_1^{b-j} E_j(Q_j(x))^2. \quad (6.10)$$

The sum is over actual revelations only. After a cell becomes terminal, U, G and T are frozen there and no extra contraction factor is generated by padding the filtration to the deterministic horizon.

Proof. We bound each revealed increment separately and sum. For $1 \le j \le b$, the pair x, z lies in the actual stage-j child on which $V_j = \sqrt{\delta} G_j^{1/2} Z_j$ was constructed. The row bound for the Piola transform and the intrinsic envelope give

$$|V_{j,i}(x) - V_{j,i}(z)| \le \sqrt{\delta}\sqrt{(G_j)_{ii}}\, E_j(Q_j) \left(\frac{d_{G_j}(x,z)}{\ell_j}\right)^\beta.$$

The j-th cap and the inequalities along the actual chain imply

$$\sqrt{(G_j)_{ii}} \le \sqrt{M}(1-\eta)^{-a_i(b-j)} H(U_b(x))^{a_i}.$$

By Lemma A.2, with a final terminal transition counted only when it was actually revealed, $d_{G_j}(x,z)/\ell_j \le C(\vartheta\sqrt{K})^{b-j}$. Since U_0 is spatially constant, summing the revealed increments and using $a_i \le a_*$ yields

$$|U_{b,i}(x) - U_{b,i}(z)| \le C\sqrt{\delta M} H(U_b(x))^{a_i} \sum_{j=1}^{b} r_{\mathrm{osc}}^{b-j} E_j(Q_j(x)).$$

The weighted Cauchy–Schwarz inequality, applied over actual reveals only, gives

$$\sum_{j=1}^{b} r_{\mathrm{osc}}^{b-j} E_j \le \left(\sum_{n\ge 0} (r_{\mathrm{osc}}^2/\rho_1)^n\right)^{1/2} \left(\sum_{j=1}^{b} \rho_1^{b-j} E_j^2\right)^{1/2}.$$

For a directly stopped cell Q_b there is no subsequent increment; for a newly terminal union the b-th increment is included once. After termination the field does not change and the inequality is retained without further discounting. □

Lemma 6.4. *Fix the finite critical exponents and $K, M < \infty$. There are $b_0 > 0$ and $C = C(K, M, \boldsymbol{r})$ such that the following holds. Let $y \in \mathbb{R}^2$, let G be positive definite with $\det G = 1$ and $G_{ii} \le MH(y)^{2a_i}$, and let $\widetilde{G}$ satisfy $0 \prec \widetilde{G} \preceq KG$ and $\det \widetilde{G} = 1$. Suppose $y^{(1)}, y^{(2)}$ belong to the anisotropic rectangle*

$$\mathcal{R}_b(y) = \{w : |w_i - y_i| \le bH(y)^{a_i},\ i = 1, 2\}, \qquad 0 < b \le b_0.$$

At every w in the rectangle, and hence along every segment joining any two of its points,

$$\tfrac{1}{2}H(y) \le H(w) \le 2H(y), \quad G_{ii} \le C_M H(w)^{2a_i}, \quad \widetilde{G}_{ii} \le C_{K,M} H(w)^{2a_i}. \tag{6.11}$$

Consequently, for $G' \in \{G, \widetilde{G}\}$,

$$|L(y^{(1)}, G') - L(y^{(2)}, G')| \le C_{K,M}\sqrt{H(y)} \sum_{i=1}^{2} \frac{|y_i^{(1)} - y_i^{(2)}|}{H(y)^{a_i}}. \tag{6.12}$$

Proof. For $F_i(t) = (1+t^2)^{r_i/2}$ one has $|F_i'(t)| \le r_i(1+t^2)^{(r_i-1)/2}$. For $w \in \mathcal{R}_1(y)$, the bound $(1+y_i^2)^{1/2} \le H(y)^{a_i}$ gives $(1+w_i^2)^{1/2} \le 2H(y)^{a_i}$, hence $|\partial_i H(w)| \le CH(y)^{1-a_i}$. Integrating along coordinate segments, $|H(w) - H(y)| \le CbH(y)$ on $\mathcal{R}_b(y)$. Taking b_0 sufficiently small gives the first comparison in (6.11); the remaining two follow directly from the old diagonal cap and $\widetilde{G} \preceq KG$. Lemma 5.1 applies to G' at each point of the segment, with the enlarged fixed cap constant $C_{K,M}$, and gives $|\partial_i L(w, G')| \le C_{K,M} H(w)^{1/2-a_i}$. Integrating this derivative along the straight segment and using $H(w) \asymp H(y)$ proves (6.12). □

The proof uses only old-state information and the rectangle bound, not any contraction of the lag. Hence Lemma 6.4 applies to the segment from the old representative to a point and to the segment of a new increment, without assuming the new-state cap in advance.

6.2 The induction

The induction is carried out on the following events. For $k \ge 0$ let $\mathcal{H}_k \in \mathcal{F}_k$ be the old good event at stage k: the history has been neither killed nor stopped by overshoot before stage k, and

(H1) for every $j \le k$ and almost every x in an active or terminal cell, $(G_j)_{ii}(x) \le MH(U_j(x))^{2a_i}$ for $i = 1, 2$, and, if a Gaussian increment was actually revealed at x at stage $j \ge 1$, $(1-\eta)H(U_{j-1}(x)) \le H(U_j(x)) \le (1+\eta)H(U_{j-1}(x))$;

(H2) every envelope revealed up to stage k satisfies $E_{j,S} \le R\sqrt{m}$;

(H3) $\mathcal{Z}_k < T_*$.

Let $\mathcal{G}_{k+1} \in \mathcal{F}_{k+1}$ be the event that every child revealed at stage $k+1$ has $E_{k+1,S} \le R\sqrt{m}$. Part (i) of the next proposition concerns $\mathcal{F}_k$-measurable quantities only; part (ii) adds the fresh increment.

Proposition 6.5. *Fix $K > 1$ and choose the constants in the order of Section 6.3. Then the following holds for every $k \ge 0$.*

(i) On $\mathcal{H}_k$, before the stage-$(k+1)$ revelation, every active parent Q and almost every $x \in Q$ satisfy

$$H(U_k(x)) \le C_{T_*} m, \tag{6.13}$$

and the predictable metric $G^+ = G_Q^+$ satisfies (6.16) below.

(ii) On $\mathcal{H}_k \cap \mathcal{G}_{k+1}$, simultaneously for all new children and almost every point (and for the chosen continuous representatives inside each child),

$$\begin{aligned} (G_{k+1})_{ii}(x) &\le MH(U_{k+1}(x))^{2a_i}, \\ (1-\eta)H(U_k(x)) &\le H(U_{k+1}(x)) \le (1+\eta)H(U_k(x)), \\ S_{k+1}(x) &\le q_0 S_k(x) + C(1 + E_{k+1}(x)^2). \end{aligned} \tag{6.14}$$

On an inactive cell, U, G and S are unchanged, so the first two bounds persist and $S_{k+1} = S_k$.

In particular, $\mathcal{H}_k \cap \mathcal{G}_{k+1} \cap \{\mathcal{Z}_{k+1} < T_*\} \subset \mathcal{H}_{k+1}$.

Proof. Fix $k \geq 0$ and work on $\mathcal{H}_k$, simultaneously on the whole predictable tree. Steps 1, 2 and 4 involve only $\mathcal{F}_k$-measurable quantities and prove part (i); Steps 3 and 5 add the fresh increment and prove part (ii).

Step 1: the old oscillation estimate. For x in an active cell of stage k, the caps and phase comparisons in (H1) are exactly the hypotheses of Lemma 6.3 along the actual ancestral chain of x, with $b = k$; the child contractions of Lemma A.2 hold by construction. Since $\delta = \varepsilon^2/m$ and, by (H2), $T_k \leq R^2 m/(1-\rho_1)$, the estimate (6.10) gives, for x, z in the same current active cell,

$$|U_{k,i}(x) - U_{k,i}(z)| \leq C\sqrt{\delta M}\, H(U_k(x))^{a_i} \sqrt{T_k(x)} \leq C\varepsilon R\sqrt{M}\, H(U_k(x))^{a_i}. \tag{6.15}$$

The same estimate holds on a new terminal child, by the terminal case of Lemma A.2. Since terminal children are frozen, it is never invoked for a later refinement of such a child.

Step 2: the old representative and a strict cap margin. The derivative bounds $|\partial_i H(y)| \leq C H(y)^{1-a_i}$ imply that if $|z_i - y_i| \leq b H(y)^{a_i}$ with b small, then $H(z) = (1 + O(b))H(y)$. Thus (6.15) makes $H(U_k(x_Q))$ and $H(U_k(x))$ relatively close. The clipping inequality (5.6) gives

$$(G^+)_{ii} \leq K^{-1}(G_k)_{ii} + C(U_k(x_Q))_{ii} \leq \big(K^{-1}M + C(1 + O(\varepsilon R\sqrt{M}))\big) H(U_k(x))^{2a_i}.$$

Choose the strict diagonal-cap contraction factor $q_d \in (K^{-1}, 1)$, then take M large enough that $K^{-1}M + C \leq q_d M$ with slack, and finally take ε small enough to absorb the representative error. Thus

$$(G^+)_{ii} \leq q_d M H(U_k(x))^{2a_i}. \tag{6.16}$$

Step 3: the fresh increment and the relative phase. On the new intrinsic envelope cutoff, the physical row estimate $|e_i^T A_Q Z| \leq \sqrt{(G_Q^+)_{ii}}\, E_{k+1}$ and (6.16) yield

$$|V_{k+1,i}(x)| \leq C\sqrt{\delta M}\, H(U_k(x))^{a_i} E_{k+1}(x) \leq C\varepsilon R\sqrt{M}\, H(U_k(x))^{a_i}.$$

Consequently $|H(U_{k+1}) - H(U_k)| \leq \xi H(U_k)$, where $\xi = C\varepsilon R\sqrt{M}$ can be made smaller than any prescribed fixed $\eta > 0$. Moreover, with $q_d < (1-\eta)^{2a_*}$ (Section 6.3), (6.16) and this phase comparison preserve the diagonal cap $(G_{k+1})_{ii} \leq M H(U_{k+1})^{2a_i}$. This proves the first two lines of (6.14) for all new children simultaneously.

Step 4: the old modular gives a pointwise phase cap. On an active parent Q, the oscillation bound (6.15) of Step 1 makes $H(U_k)$ comparable throughout Q. Every active Q has area at least cs^2. By (H3), $\mathcal{Z}_k \leq T_*$, and $\lambda_k \geq \lambda_{J_0} \geq 1/2$ by the choice of ε after Λ and C_{dep} (Section 6.3). Integration over Q gives

$$T_* \geq \int_Q e^{\lambda_k H(U_k)} \geq c|Q| e^{cH(U_k(x))},$$

so $H(U_k(x)) \leq C(1 + \log(T_*/s^2)) \leq C_{T_*} m$. This proves (6.13), which is $\mathcal{F}_k$-measurable.

Step 5: contraction of the lag and memory. Set $y = U_k(x)$, $y_Q = U_k(x_Q)$ and $y' = U_{k+1}(x)$. Steps 1 and 3 establish $|y_{Q,i} - y_i| + |y'_i - y_i| \leq b H(y)^{a_i}$ with $b \leq C\varepsilon R\sqrt{M}$. Both metrics G_k and G^+ satisfy the old diagonal cap, with $G^+ \preceq K G_k$ and $\det G^+ = 1$. Once ε is chosen so that $b \leq b_0$, Lemma 6.4 applies with $y^{(1)} = y_Q$ and $y^{(2)} = y'$. It gives the cap throughout the segments $[y, y_Q]$ and $[y, y']$, which permits integration of (5.5) along them; no conclusion of the lag update is used. At the representative the clipping contracts the lag, $L(y_Q, G^+) \leq K^{-1/2} L(y_Q, G_k) + C_K$ by (5.6). Together with the segment estimates, this gives

$$L_{k+1}(x) \leq K^{-1/2} L_k(x) + C_K + C_M \sqrt{\delta H(U_k(x))}\big(\sqrt{T_k(x)} + E_{k+1}(x)\big).$$

By Step 4, $\sqrt{\delta H} \leq C_{T_*}\varepsilon$. Fix the lag-contraction factor $q_L > K^{-1}$ and a Young slack $\tau > 0$ with $(1+\tau)K^{-1} \leq q_L$. Squaring with this slack gives

$$L_{k+1}^2 \leq q_L L_k^2 + C_\tau + C_\tau \varepsilon^2 T_k + C_\tau \varepsilon^2 E_{k+1}^2.$$

Since $T_{k+1} = \rho_1 T_k + E_{k+1}^2$, choose $\max\{q_L, \rho_1\} < q_0 < 1$ (the combined energy contraction) and shrink ε further, obtaining

$$S_{k+1} \le q_0 S_k + C(1 + E_{k+1}^2).$$

This proves part (ii) simultaneously on all active children. On inactive branches nothing is renewed: the old bounds remain valid, and the clock is the only quantity that changes.

Finally, (H1) at stage $k+1$ is (6.14) together with the persistence on inactive cells, (H2) at stage $k+1$ is $\mathcal{G}_{k+1}$, and (H3) is the remaining condition $\mathcal{Z}_{k+1} < T_*$. All constants in Steps 1–5 are the fixed constants of Section 6.3; none depends on the parent. □

Since $U_0 = 0$, $G_0 = I$, $H(0) = 2$ and $M \ge 1$, the event $\mathcal{H}_0$ is the event $\{\mathcal{Z}_0 < T_*\}$, which is the whole space by the choice of T_* in Section 7. By induction on k, using the last inclusion of Proposition 6.5, $\mathcal{H}_k$ contains every history that has been neither killed nor stopped by overshoot before stage k.

6.3 Choice of parameters

All constants of Part I are fixed, with strictly positive margins, in the order of Table 2. Each constant depends only on the fixed exponents $\boldsymbol{r}$ and on the constants listed above it. None of them depends on the grid scale m or on the freshly revealed Gaussian data, and no constant is constrained by a constant chosen after it.

Constant	Requirement and source
$K > 1$, $0 < \beta < \alpha < 1$	Metric distortion and Hölder exponents; arbitrary.
c_0; A, γ	Inball ratio of intrinsic parents, depending on K; then the Whitney displacement and clearance constants, depending only on c_0 (Lemma 4.1).
q_d, $q_L \in (K^{-1}, 1)$; τ	Cap and lag contraction factors; Young slack $\tau > 0$ with $(1+\tau)K^{-1} \le q_L$ (Step 5 of Proposition 6.5).
η, q_0	$q_d < (1-\eta)^{2a_*}$, and q_0 with $q_L < q_0 < 1/(1+\eta)$.
ϑ; θ, D_0; C_{dep}	$\vartheta\sqrt{K} < 1$ and $r_{\mathrm{osc}} := (\vartheta\sqrt{K})^{\beta}(1-\eta)^{-a_*}$ with $r_{\mathrm{osc}}^2 < q_0$; implemented by the Whitney cap θ and the stopping threshold D_0 (Lemma A.1); this fixes the depth constant C_{dep}.
ρ_1	$r_{\mathrm{osc}}^2 < \rho_1 < q_0$.
M	$K^{-1}M + C \le q_d M$ with slack, where C comes from (5.4) and (5.6) (Step 2).
κ, Λ	Drift balance in the one-step contraction (Section 7).
T_*	$T_* > \mathcal{Z}_0$, large enough for a positive success probability (Proposition 7.4).
R	$cR^2 > 2\log 2$, with c the envelope tail constant (Lemma 6.2).
ε	Last: $b \le b_0$ in Lemma 6.4; $\xi < \eta$; absorption of the representative error in (6.16) and of $\varepsilon^2 T_k$ in Step 5; $\lambda_{J_0} \ge 1/2$, i.e. $\Lambda C_{\mathrm{dep}} \varepsilon^2 \le \log 2$; the smallness conditions of Section 7; the terminal corrections of Section 8.

Table 2: Order of the parameter choices in Part I.

In particular,

$$r_{\mathrm{osc}}^2 < \rho_1 < q_0 < 1, \qquad q_L < q_0, \qquad q_d < (1-\eta)^{2a_*}, \qquad q_0(1+\eta) < 1. \tag{6.17}$$

7 Exponential moment estimates

This section turns the deterministic bounds of Proposition 6.5 into a probabilistic estimate on the exponential modular of U_k. Proposition 7.3 shows that, at a single point and conditionally on the past, one refinement step does not increase e^{Φ_k} in expectation on the local envelope event. Proposition 7.4 integrates this in space, kills the process at envelope failures and stops it at

overshoot, and deduces from Ville's inequality that with probability bounded below all leaves terminate with the Orlicz bound (7.7).

Define the positive phase and its spatial modular by

$$\Phi_k(x) = \lambda_k H(U_k(x))\big(1 + \kappa\delta S_k(x)\big), \qquad \mathcal{Z}_k = \int_{\mathbb{T}^2} e^{\Phi_k(x)}\,\mathrm{d}x. \tag{7.1}$$

At the critical exponents, the variance associated with $\lambda_k H$ has the same order as the phase itself. The additional term $\kappa\delta HS$ has a negative drift under active refinement, which compensates for this covariance cost. The mechanism is that, after a Taylor expansion, the fresh Gaussian field enters the increment of Φ_k only through a linear term $\ell \cdot Z_{k+1}$ whose coefficient ℓ is old-measurable (Lemma 7.1). Its exponential moment is then computed exactly, and the envelope contributes only a quadratic correction (Lemma 7.2).

Lemma 7.1. *On an old good active cell, put $y = U_k(x)$, $H = H(y)$, $H' = H(U_{k+1}(x))$, $S = S_k(x)$, $S' = S_{k+1}(x)$, $E = E_{k+1}(x)$, and $\lambda = \lambda_{k+1}$. On the local envelope cutoff,*

$$\Phi_{k+1} - \Phi_k \le -\Lambda\lambda\delta H - \kappa d_0 \lambda\delta HS + \ell \cdot Z_{k+1} + A_\kappa \lambda\delta H(1 + E^2), \tag{7.2}$$

where

$$d_0 = 1 - q_0(1 + \xi) > 0, \qquad \ell = \lambda\sqrt{\delta}\, A_Q^T v(y) = \lambda\sqrt{\delta}\,(G_Q^+)^{1/2} v(y).$$

In particular, ℓ is measurable before *the fresh Gaussian variables are revealed.*

Proof. Taylor's formula and $F_i''(t) \lesssim H^{1-2a_i}$ along the short segment give

$$H' - H \le v(y) \cdot V_{k+1} + A\delta HE^2, \qquad |H' - H| \le \xi H. \tag{7.3}$$

By the memory inequality $S' \le q_0 S + A(1 + E^2)$ and positivity,

$$H'S' - HS \le (1 + \xi)H\big(q_0 S + A(1 + E^2)\big) - HS = -d_0 HS + A(1 + \xi)H(1 + E^2).$$

Choose ξ so that $q_0(1 + \xi) < 1$. Now use the exact algebraic identity

$$\Phi_{k+1} - \Phi_k = -(\lambda_k - \lambda)H(1 + \kappa\delta S) + \lambda(H' - H) + \kappa\lambda\delta(H'S' - HS).$$

The clock gives $\lambda_k - \lambda \ge \Lambda\lambda\delta$. Insert the two preceding bounds and $V_{k+1} = \sqrt{\delta} A_Q Z_{k+1}$. This proves (7.2). □

Lemma 7.2. *Suppose, conditionally on $\mathcal{F}_k$, Z is a centered Gaussian vector with covariance at most $\sigma^2 I$ and an accompanying (not necessarily independent) envelope satisfies $\mathbb{E}_k e^{cE^2} \le C$. For $0 \le u \le u_0 < c/4$ and any old-measurable $\ell \in \mathbb{R}^2$,*

$$\log \mathbb{E}_k e^{\ell \cdot Z + uE^2} \le \sigma^2 |\ell|^2 + C_* u. \tag{7.4}$$

Proof. Since E and Z may be dependent, we separate them by Cauchy–Schwarz:

$$\mathbb{E}_k e^{\ell \cdot Z + uE^2} \le (\mathbb{E}_k e^{2\ell \cdot Z})^{1/2} (\mathbb{E}_k e^{2uE^2})^{1/2}.$$

The first factor is evaluated exactly using Gaussianity. For the second, $\mathbb{E}_k(E^2 e^{2u_0 E^2})$ is uniformly finite by the assumed exponential-square moment. Integrating the u-derivative of its logarithm from 0 to u gives $C_* u$. □

Proposition 7.3. *For κ sufficiently large, Λ sufficiently large after κ, and ε sufficiently small after all fixed constants, on any old good active history,*

$$\mathbb{E}_k\left[\mathbf{1}_{\{E_{k+1}(x) \le R\sqrt{m}\}} e^{\Phi_{k+1}(x)}\right] \le e^{\Phi_k(x)}. \tag{7.5}$$

Proof. We exponentiate (7.2) and apply Lemma 7.2. First, from $vv^T \preceq CHC(y)^{-1}$ and $G^+ \preceq KG_k$,

$$|\ell|^2 = \lambda^2 \delta v^T A_Q A_Q^T v = \lambda^2 \delta v^T G_Q^+ v \le C_K \lambda \delta H \operatorname{tr}(G_k C(y)^{-1}) \le C_K \lambda \delta H S.$$

This is the only place where the lag enters. Set $u = A_\kappa \lambda \delta H$. By the old phase cap $H \le C_{T_*} m$, we have $u \le C\varepsilon^2$, so Lemma 7.2 applies. Exponentiating (7.2) on the local good event, then dropping its indicator in the upper bound, gives

$$\log \mathbb{E}_k\big[\mathbf{1}_{\{E \le R\sqrt{m}\}} e^{\Phi_{k+1}-\Phi_k}\big] \le -\Lambda \lambda \delta H - \kappa d_0 \lambda \delta H S + C_K \lambda \delta H S + C_\kappa \lambda \delta H.$$

Choosing $\kappa d_0 > C_K$ and then $\Lambda > C_\kappa$ makes the right-hand side nonpositive. □

Because the tilt ℓ is predictable and the fresh field enters linearly, the estimate involves no δH^2 term and no fourth cumulant.

7.1 Stopping rule

The one-step estimate is applied up to a stopping time. We define the numerical process on all outcomes of each attempted reveal.

Let J_0 be a fixed upper bound on active depth, and let $\mathcal{F}_k$ be the completed sigma-field generated by the geometry and all Gaussian cell masses revealed through stage k; as described in Section 6, the entire next subdivision is $\mathcal{F}_k$-measurable. We pad terminal branches to the deterministic horizon J_0 by doing nothing. The old good event is $A_k := \mathcal{H}_k \in \mathcal{F}_k$, defined before Proposition 6.5: previous envelopes were acceptable and the last attained modular is strictly below T_*; the caps and relative-phase bounds then hold by that proposition. At stage $k+1$ write

$$B_{k+1} = \mathcal{G}_{k+1} = \{\text{every newly revealed child envelope is } \le R\sqrt{m}\} \in \mathcal{F}_{k+1},$$

where no new child is assigned an envelope on a frozen branch.

Define a nonnegative adapted scalar process $\widehat{\mathcal{Z}}_k$, the killed-and-frozen modular, as follows:

(i) Initially $\widehat{\mathcal{Z}}_0 = \mathcal{Z}_0$.

(ii) Before failure or overshoot, attempt the next geometric refinement. If B_{k+1} fails, put $\widehat{\mathcal{Z}}_{k+1} = 0$ and thereafter keep the numerical process zero.

(iii) If B_{k+1} holds, use the actual new modular $\widehat{\mathcal{Z}}_{k+1} = \mathcal{Z}_{k+1}$. If it is at least T_*, freeze the geometric field and its memory for all later stages, allowing only λ to decrease.

(iv) On a branch that became terminal for geometric reasons, freeze (U, S) on that branch from then on, while the clock continues to decrease.

Proposition 7.4. *The process $\widehat{\mathcal{Z}}_k$ is a nonnegative supermartingale and*

$$\mathbb{P}(\exists j \le J_0 : \widehat{\mathcal{Z}}_j \ge T_*) \le \frac{\mathcal{Z}_0}{T_*}. \tag{7.6}$$

The global envelope-failure probability is $o(1)$. Consequently, for large fixed T_*, there is a constant $p_* > 0$ independent of m such that with probability at least p_* all leaves terminate and the leaf current satisfies

$$\operatorname{div} J_{\text{leaf}} = P_{\mathcal{P}_*} W, \qquad \max_i \|J_{\text{leaf},i}\|_{\psi_{r_i}} \le C\delta^{-1/2} \le C_{\boldsymbol{r}} \sqrt{m}. \tag{7.7}$$

Proof. We first prove the supermartingale property, case by case. Suppose the past has neither failed nor overshot. For each active x, Propositions 6.5 and 7.3 give

$$\mathbb{E}_k\left[\mathbf{1}_{B_{k+1}} e^{\Phi_{k+1}(x)}\right] \le \mathbb{E}_k\left[\mathbf{1}_{\{E_{k+1}(x) \le R\sqrt{m}\}} e^{\Phi_{k+1}(x)}\right] \le e^{\Phi_k(x)}.$$

The first inequality holds because the global envelope event is contained in the local one. On an inactive branch the field and memory are frozen and the clock decreases, so the integrand can only decrease and the same inequality holds. Integrating over x and using conditional Tonelli, we obtain

$$\mathbb{E}_k[\mathbf{1}_{B_{k+1}}\mathcal{Z}_{k+1}] \le \mathcal{Z}_k.$$

Since the definition sets $\widehat{\mathcal{Z}}_{k+1} = 0$ on B_{k+1}^c, the left side is exactly the conditional expectation of the numerical process at the new stage, until the first overshoot. After the overshoot there is no further geometric evolution, and $\lambda_{k+1} \le \lambda_k$ ensures $\widehat{\mathcal{Z}}_{k+1} \le \widehat{\mathcal{Z}}_k$ pathwise. After a failure the process is zero. No conditioning on the success of a future reveal is involved.

The initial values are deterministic: $U_0 = 0$, $H(U_0) = 2$, $G_0 = I$ and $T_0 = 0$; hence $\mathcal{Z}_0 = \exp(2(1 + \kappa\delta L(0, I)^2)) = O(1)$. For $\tau = \inf\{k : \widehat{\mathcal{Z}}_k \ge T_*\}$, Ville's inequality yields (7.6) for the killed process. Independently of this stopping argument, Lemma 6.2 bounds the probability of any actual attempted envelope failure by $o(1)$, using conditional tails and the deterministic bound on the number of attempted children. Write $\mathcal{E}_m$ for the event of any attempted envelope failure and $\mathcal{O}_m$ for the first overshoot along an all-envelope-good history. The successful completion event satisfies

$$\mathbb{P}(A_m) \ge 1 - \mathbb{P}(\mathcal{O}_m) - \mathbb{P}(\mathcal{E}_m) \ge 1 - \mathcal{Z}_0/T_* - o(1).$$

Thus T_* may be fixed sufficiently large so that this lower bound exceeds some $p_* > 0$.

On this good completion event the final positive modular is at most T_*. Since $\lambda_{J_0} \ge 1/2$ (after choosing ε small), $\Phi_{J_0} \ge \frac{1}{2}H(U_{J_0})$. The fixed exponential modular bound implies the coordinate Luxemburg bounds for U_{J_0} after an absolute dilation, and dividing by $\sqrt{\delta}$ proves (7.7). □

The threshold in (7.6) is tested on the killed process $\widehat{\mathcal{Z}}_k$, so Ville's bound concerns a first overshoot along a history on which all envelopes are good. On a killed history the process is not the modular of an admissible current. The actual-reveal geometry, including direct stops, terminal unions and padded times, is recorded in Lemma A.5. The next section corrects the adaptive current from $P_{\mathcal{P}_*}W$ to the fixed-grid source W_h.

8 The terminal Gaussian correction

The adaptive field from Proposition 7.4 solves the cell-averaged source PW, where P averages over the terminal leaves of a random partition. The target source is W_h, obtained by averaging over the fixed h-grid. In this section we construct a current with divergence $W_h - PW$ and show that it is negligible in the Orlicz norms under consideration; this completes the proof of Theorem 2.1. The difference $W_h - PW$ splits exactly into two pieces. The first is built from the revealed averages PW and has zero mean on each fixed h-square, so it is inverted square by square. The second is the fixed-grid average of the unrevealed residual $W - PW$; conditionally on the terminal history it is Gaussian, and we invert it by convolution with the periodic Green kernel. On the torus this is a periodic convolution, with no physical boundary data.

8.1 The exact splitting

On a successful history, every terminal leaf S has area at least cs^2 and diameter $O(s)$ in its birth metric G_S. This includes both the clipped unions born at a split and the active squares stopped without another split; by Lemma A.1, the latter have intrinsic diameter at most CD_0s. Write $A_S = G_S^{1/2}$ and $S = b_S + A_S\widehat{S}$ in its chosen lift. The diagonal cap holds in physical coordinate directions, so at the final good stage $H(U) \le C_{T_*}m$ gives

$$\|A_S\|_{\mathrm{op}}^2 \le \operatorname{tr} G_S \le M\sum_{i=1}^{2} H(U)^{2a_i} \le C_{M,T_*}m^{2a_*}.$$

We now pass from the local intrinsic geometry to the fixed physical flat-torus metric, in which

$$\operatorname{diam}_{\text{physical}}(S) \le D_* := Csm^{a_*}, \qquad a_* < \frac{1}{2}, \qquad D_*/h = Cm^{3+a_*}. \tag{8.1}$$

This bound uses the cap: with Cs in place of D_* it would fail to hold uniformly in the eccentricity.

Lemma 8.1. *On the successful first-failure event, every terminal leaf S, including a directly stopped square and a clipped union born at the last reveal, satisfies*

$$|S| \ge cs^2, \qquad \operatorname{diam}_{\text{physical}} S \le Csm^{a_*}.$$

The terminal partition $\mathcal{P}_*$ and the success event belong to the finite terminal revealed sigma field $\mathcal{F}_*$. For every fixed deterministic $g \in L^2(\mathbb{T}^2)$, conditional on $\mathcal{F}_*$,

$$\operatorname{Var}((W - P_{\mathcal{P}_*}W)(g) \mid \mathcal{F}_*) = \sum_{S\in\mathcal{P}_*} \int_S |g - g_S|^2.$$

Proof. There are two kinds of terminal leaves. A newly terminal union has G_S-intrinsic diameter at most $C_A s$, and a square stopped without another reveal has birth-metric diameter at most $\sqrt{K}D_0 s$. In both cases the receiver construction gives the area floor.

On successful completion the final modular is at most T_* and $\lambda_{J_0} \ge 1/2$. Lemma 6.3, applied at the last actual reveal of S, gives throughout S the same phase comparability that holds on an active cell. Since U is frozen afterwards, the comparison remains valid at time J_0. Integrating the final modular over S then gives $H(U|_S) \le C_{T_*}m$, also when S was created at the final reveal. The newly inherited diagonal cap, or the old cap for a directly stopped square, yields $\|G_S^{1/2}\|_{\text{op}} \le Cm^{a_*}$. Multiplying by the intrinsic diameter proves the physical bound.

The algorithm makes all geometric choices before the corresponding fresh Gaussian child contrasts are revealed, and it is padded to a finite deterministic horizon. Since stopping and the success event depend only on revealed contrasts, they are $\mathcal{F}_*$-measurable. The residual white noise in the orthogonal complement of the finally revealed cell indicators is conditionally Gaussian. Testing it against a deterministic g and using the conditional orthogonal projection gives the displayed variance. □

From now on all distances and derivatives are physical, not intrinsic; in particular the periodic Green kernel is never pulled back by an adaptive A_S. Write $P = P_{\mathcal{P}_*}$. The identity

$$W_h - PW = (P_h - I)PW + P_h(W - PW) =: g^{(1)} + g^{(2)} \tag{8.2}$$

holds exactly, as an identity of centered distributions.

8.2 Correction of the revealed part

The first source $g^{(1)}$ has zero integral on each fixed h-square. Applying the ray inverse on these squares gives a periodic current $J^{(1)}$ with

$$\operatorname{div} J^{(1)} = g^{(1)}, \qquad \left\|J^{(1)}\right\|_\infty \le Ch \left\|PW\right\|_\infty .$$

Lemma 3.2 controls the random final cells directly and gives, with probability $1 - o(1)$,

$$\|PW\|_\infty \le C\sqrt{m}/s, \qquad \left\|J^{(1)}\right\|_\infty \le C\sqrt{m}/m^3 = o(1). \tag{8.3}$$

8.3 Periodic kernel estimates for the unrevealed bridge

Let $G_{\mathbb{T}}$ be the mean-zero Green function of the torus Laplacian in the fixed physical coordinates, and choose the periodic vector kernel $K_{\mathbb{T}}(x,y) = \nabla_x G_{\mathbb{T}}(x-y)$ with the sign convention

$$\operatorname{div}_x K_{\mathbb{T}}(x,y) = \delta_y - 1.$$

The standard local expansion of the periodic Green function and smoothness away from the diagonal give

$$|K_{\mathbb{T}}(x,y)| \le C(1 + d_{\mathbb{T}}(x,y)^{-1}), \qquad |\nabla_y K_{\mathbb{T}}(x,y)| \le C(1 + d_{\mathbb{T}}(x,y)^{-2}). \tag{8.4}$$

In the source variable let $K_{x,h} = P_h^{(y)} K_{\mathbb{T}}(x,\cdot)$, a piecewise constant function with jumps along the fixed h-grid, and define

$$J^{(2)}(x) = (W - PW)(K_{x,h}).$$

Its divergence is $g^{(2)}$, with the vector notation understood coordinatewise. Lemma 8.1, applied to the deterministic function $K_{x,h}$, gives the exact conditional variance

$$\sigma_*^2(x) := \operatorname{Var}(J^{(2)}(x) \mid \mathcal{F}_*) = \sum_{Q \in \mathcal{P}_*} \int_Q |K_{x,h} - (K_{x,h})_Q|^2 \, \mathrm{d}y. \tag{8.5}$$

The next lemma collects the kernel bounds used to estimate this variance and the spatial modulus of $J^{(2)}$.

Lemma 8.2. *For h sufficiently small, uniformly in $x, y \in \mathbb{T}^2$,*

$$|K_{x,h}(y)| \le \frac{C}{h + d_{\mathbb{T}}(x,y)}. \tag{8.6}$$

If y, z lie in a common physical set of Euclidean torus diameter at most D_*, with $r = d_{\mathbb{T}}(x,y) \ge 4(D_* + 2h)$, then

$$|K_{x,h}(y) - K_{x,h}(z)| \le C \frac{D_* + h}{r^2}. \tag{8.7}$$

Finally, for $|u| \le 1/4$,

$$\|K_{\mathbb{T}}(\cdot + u, \cdot) - K_{\mathbb{T}}(\cdot,\cdot)\|_{L^1(\mathbb{T}^2)} \le C|u|(1 + |\log|u||). \tag{8.8}$$

Proof. Fix a periodic lift at distance less than $1/4$ from the pole. When the h-square containing y is at distance at least $2h$ from x, average (8.4) over that square and use $|x - z| \asymp d_{\mathbb{T}}(x,y)$. When the square is close to the pole, the integral of $1/|x-z|$ over an h-square is $O(h)$, and division by the area h^2 gives $O(h^{-1})$. The smooth periodic remainder contributes $O(1)$, which proves (8.6).

For the second claim, let R_y, R_z be the fixed h-squares containing y, z. If $\zeta \in R_y$ and $\zeta' \in R_z$, then $|\zeta - \zeta'| \le D_* + 2\sqrt{2}h$. Every point of the segment $[\zeta, \zeta']$ lies within $\sqrt{2}h + D_* + 2\sqrt{2}h \le \frac{3\sqrt{2}}{2}(D_* + 2h)$ of y, and $\frac{3\sqrt{2}}{2}(D_* + 2h) \le \frac{3\sqrt{2}}{8} r < \frac{3}{4} r$ because $r \ge 4(D_* + 2h)$. Hence the whole segment stays at distance at least $r/4$ from the pole. We apply the derivative bound for the original kernel $K_{\mathbb{T}}$ in (8.4) along this segment and average twice; this proves (8.7). The step function $K_{x,h}$ itself is never differentiated.

For (8.8), split the torus into $B(x, 2|u|)$ and its complement. In the ball, integrating the two $1/r$ singularities gives $C|u|$. Outside, the mean-value theorem gives the pointwise bound $C|u|/r^2$, whose integral in two dimensions is $C|u|(1 + |\log|u||)$. □

Proposition 8.3. *Under the leaf diameter bound* (8.1),

$$\sup_{x \in \mathbb{T}^2} \sigma_*^2(x) \le C \left[1 + \log_+ \left(\frac{D_*}{h} \right) \right] \le C \log m. \tag{8.9}$$

The estimate is deterministic once the successful terminal geometry is fixed, and it uses nothing about how the tree was selected.

Proof. Put $R_* = 8(D_* + 2h)$ and split the leaves into those meeting $B(x, R_*)$ and the remaining ones. Any near leaf lies inside $B(x, R_* + D_*) \subset B(x, CR_*)$. Subtracting a leaf average only decreases the L^2 norm, so by (8.6)

$$\sum_{Q\,\mathrm{near}} \int_Q |K_{x,h} - (K_{x,h})_Q|^2 \le C \int_{B(x,CR_*)} \frac{\mathrm{d}y}{(h + d_{\mathbb{T}}(x,y))^2} \le C\big[1 + \log_+(D_*/h)\big].$$

For a far leaf Q and $y, z \in Q$ we apply (8.7). Since the leaf is much smaller than its distance to the pole, this gives

$$\int_Q |K_{x,h} - (K_{x,h})_Q|^2 \le C|Q| \frac{(D_* + h)^2}{r_Q^4}, \quad r_Q = \mathrm{dist}_{\mathbb{T}}(x, Q).$$

Summing over far leaves and comparing with the integral in the two-dimensional radial variable bounds their contribution by

$$C(D_* + h)^2 \int_{R_*/2}^1 r^{-3}\,\mathrm{d}r \le C.$$

This proves the first inequality in (8.9), and the second follows from (8.1). Borderline cells are avoided by increasing the universal radius constants, which does not change the scaling. □

8.4 Orlicz and supremum estimates

We now turn the variance bound into estimates for $J^{(2)}$: a ψ_2 bound with fixed positive conditional probability, and an L^∞ bound with probability $1 - o(1)$. Interpolating between the two gives the ψ_{r_i} bounds.

Write $A_m \in \mathcal{F}_*$ for the event that the adaptive construction completes without numerical killing or modular overshoot, so that $\mathbb{P}(A_m) \ge p_* > 0$ by Proposition 7.4. On A_m, conditionally on $\mathcal{F}_*$, each coordinate $J_i^{(2)}(x)$ is centered Gaussian by Lemma 8.1, and its variance is at most $C_0 \log m$ by (8.9), with a fixed C_0.

We first bound the quadratic exponential modular. For a centered Gaussian variable X of variance σ^2 and $t^2 > 2\sigma^2$,

$$\mathbb{E}e^{X^2/t^2} = \Big(1 - \frac{2\sigma^2}{t^2}\Big)^{-1/2}.$$

With $A_1^2 = 8C_0/3$ and $t^2 = A_1^2 \log m$ one has $2\sigma^2/t^2 \le 3/4$, hence

$$\mathbb{E}\left[\exp\left(\frac{|J_i^{(2)}(x)|^2}{A_1^2 \log m}\right) \Big| \mathcal{F}_*\right] \le 2 \qquad (x \in \mathbb{T}^2,\ i = 1, 2).$$

By conditional Tonelli the spatial integral has conditional expectation at most 2. Markov's inequality at level 16, applied to each of the two coordinates, shows that with conditional probability at least $1 - 2 \cdot \frac{2}{16} = \frac{3}{4}$,

$$\int_{\mathbb{T}^2} \exp\left(\frac{|J_i^{(2)}|^2}{A_1^2 \log m}\right) \le 16, \qquad i = 1, 2.$$

On this event, Jensen's inequality for the concave map $t \mapsto t^{1/4}$ on the unit-area torus gives $\int_{\mathbb{T}^2} \exp(|J_i^{(2)}|^2/(4A_1^2 \log m)) \le 16^{1/4} = 2$, that is, $\int_{\mathbb{T}^2} \psi_2(J_i^{(2)}/(2A_1\sqrt{\log m})) \le 1$. Consequently the event

$$C_m = \Big\{\max_{i=1,2} \|J_i^{(2)}\|_{\psi_2} \le 2A_1\sqrt{\log m}\Big\}$$

satisfies

$$\mathbb{P}(C_m \mid \mathcal{F}_*) \geq \tfrac{3}{4} \qquad \text{on } A_m. \tag{8.10}$$

The L^∞ bound requires a separate high-probability argument, since quadratic exponential integrability in space alone does not control the supremum. Recall that $g^{(2)} = W_h - P_h PW$, and let D_m be the intersection of the fixed-h Gaussian maximum event and the predictable adaptive-cell maximum event. Since $s \gg h$, on D_m

$$\left\| g^{(2)} \right\|_\infty \leq C\sqrt{m}/h.$$

The fixed-grid event follows from the ordinary union bound over h^{-2} deterministic Gaussians, and the adaptive event is (3.6). Hence $\mathbb{P}(D_m^c) = o(1)$; this bound is unconditional. Since $J^{(2)} = K_{\mathbb{T}} * g^{(2)}$, the translation estimate (8.8) gives on D_m the deterministic modulus

$$|J^{(2)}(x) - J^{(2)}(z)| \leq C\frac{\sqrt{m}}{h}\, d_{\mathbb{T}}(x,z)(1 + |\log d_{\mathbb{T}}(x,z)|).$$

Fix a deterministic torus net $\mathcal{N}_m$ of mesh $q_m = 2^{-Lm}$, where $L > 3$ is fixed. It has at most $C2^{2Lm}$ points. Since $h = 2^{-m}$, the modulus above is $O(m^{3/2}2^{-(L-1)m}) = o(1)$ at distance q_m. Let N_m be the event that $|J^{(2)}(y)| \leq A_0\sqrt{m \log m}$ for every $y \in \mathcal{N}_m$. The conditional Gaussian tails at variance level $C_0 \log m$ and a union bound over the net give, on A_m,

$$\mathbb{P}(N_m^c \mid \mathcal{F}_*) \leq C2^{2Lm} e^{-A_0^2 m/(2C_0)} = o(1)$$

once $A_0^2 > 4C_0 L \log 2$. This estimate uses neither the fixed-grid nor the adaptive-cell maximum event. On $D_m \cap N_m$ the modulus extends the net bound to

$$\left\| J^{(2)} \right\|_\infty \leq C\sqrt{m \log m}. \tag{8.11}$$

For a scalar function f and fixed $r > 2$, the elementary inequality

$$\left(\frac{|f|}{A^{1-2/r}B^{2/r}}\right)^r \leq \left(\frac{|f|}{B}\right)^2 \quad \text{whenever} \quad \|f\|_\infty \leq A$$

shows, after a fixed dilation, that

$$\|f\|_{\psi_r} \leq C_r \|f\|_\infty^{1-2/r} \|f\|_{\psi_2}^{2/r}.$$

Consequently, on $C_m \cap D_m \cap N_m$,

$$\left\| J_i^{(2)} \right\|_{\psi_{r_i}} \leq C_{r_i} m^{1/2-1/r_i}\sqrt{\log m} = o(\sqrt{m}) \qquad (r_i < \infty). \tag{8.12}$$

The gain over $\sqrt{m}$ comes from the finiteness of r_i and is not uniform as $r_i \to \infty$.

8.5 Joint probability estimate

Since $A_m \in \mathcal{F}_*$, (8.10) gives $\mathbb{P}(A_m \cap C_m) = \mathbb{E}[\mathbf{1}_{A_m}\mathbb{P}(C_m \mid \mathcal{F}_*)] \geq \frac{3}{4}\mathbb{P}(A_m)$. The remaining two events are removed by subtracting their complements, which requires no conditional independence:

$$\begin{aligned}\mathbb{P}(A_m \cap C_m \cap D_m \cap N_m) &\geq \mathbb{P}(A_m \cap C_m) - \mathbb{P}(D_m^c) - \mathbb{P}(A_m \cap N_m^c) \\ &\geq \tfrac{3}{4}\,\mathbb{P}(A_m) - o(1) \geq \tfrac{3}{4}\,p_* - o(1) > 0.\end{aligned}$$

On this intersection, put $J = J_{\text{leaf}} + J^{(1)} + J^{(2)}$. The exact splitting (8.2) gives $\operatorname{div} J = W_h$, and the three estimates are

$$\max_i \|J_{\text{leaf},i}\|_{\psi_{r_i}} \leq C\sqrt{m}, \qquad \|J^{(1)}\|_\infty \leq Cm^{-5/2}, \qquad \|J_i^{(2)}\|_{\psi_{r_i}} \leq C_{r_i} m^{1/2-1/r_i}\sqrt{\log m}.$$

In the quartic case the last bound is $Cm^{1/4}\sqrt{\log m} = o(\sqrt{m})$, and for every fixed finite pair the same strict improvement follows from (8.12). The triangle inequality therefore gives, for some fixed $p_{**} > 0$,

$$\mathbb{P}(\mathfrak{F}_{\boldsymbol{r}}(W_h) \leq C\sqrt{m}) \geq p_{**} \tag{8.13}$$

for all sufficiently large m.

Completion of the proof of Theorem 2.1. The estimate (8.13) is the required fixed-positive-probability statement. Taking any $0 < \varepsilon_0 < p_{**}$ gives (2.4) and proves Theorem 2.1. The upgrade to a high-probability bound by Gaussian concentration, with constants independent of the mesh, is proved in Part II; see Remark 2.2.

Part II
From matching to dyadic Gaussian Orlicz flows

This part reduces the matching problem to a Gaussian minimum-flow problem on the square (Theorem 10.4), for every critical exponent pair including the hard endpoints, and transfers the torus estimate of Part I to the square (Section 17).

9 Overview of the reduction

Fix a critical exponent pair $\boldsymbol{r} = (r_1, r_2)$, with $1/r_1 + 1/r_2 = 1/2$. Let $\mathfrak{M}_N^{\boldsymbol{r}}$ denote the minimum, over permutations of two samples of N independent uniform points in $[0,1]^2$, of the maximum of their coordinatewise Orlicz matching norms. The natural scale is $s_N = \sqrt{(\log N)/N}$. We relate a bound at this scale to the minimum Orlicz norm of a current whose divergence is Gaussian white noise projected onto the dyadic partition of mesh $h_N \asymp s_N$. The two statements are equivalent at the level of high-probability bounds with deterministic constants; the equivalence does not identify the constants or the minimum costs. The one-cloud coupling also gives Talagrand's original matching problem with any evenly spread deterministic reference configuration (Corollary 16.1).

For finite critical exponents, Theorem 2.1 in Part I gives a fixed-positive-probability bound for the Gaussian current on the flat torus. The concentration principle of Section 11.1 upgrades it to high probability before the matching reduction is applied. The reduction also covers the hard endpoint, for which no Gaussian flow is constructed; see Part III. No Gaussian flow estimate is assumed in the proof of Theorem 10.4, and the concentration step does not use the flow construction of Part I; the proof of Theorem 10.4 does not use Part I.

The reduction uses the projection onto dyadic cells, rather than a Fourier or heat regularization. The random source produced by the reduction is then exactly the cellwise Gaussian source of the flow construction.

We now outline the argument. Choose h_N so that Nh_N^2 is comparable to $\log N$. With high probability, both empirical histograms then have densities bounded above and away from zero. Spreading the atoms uniformly within their cells changes each transport displacement by at most a multiple of h_N. Conversely, a coupling of the histograms gives a doubly stochastic matrix on the original labels, and a Birkhoff decomposition yields a single permutation controlling both coordinates. This reduces matching to transport between the two histograms (Proposition 12.1).

For bounded positive densities, transport with coordinatewise Orlicz costs is comparable to the minimum Orlicz norm of a static divergence current (Theorems 13.7 and 13.8). One direction follows by linearly interpolating densities along a given current. The reverse direction uses straight-segment interpolation associated with a strictly convex translation cost. The relevant density bound is obtained from noncrossing of transport rays and a Jacobian inequality; we give the details needed at a hard coordinate endpoint.

The remaining random step is a coupling of the histogram discrepancy with white noise on the same dyadic grid (Theorem 15.1). Encode the fixed binary rectangle tree by nested intervals and apply the classical one-dimensional KMT coupling to a Brownian bridge. Its increments give exactly the Gaussian cell masses. The rectangle discrepancies have an explicit Haar divergence inverse whose L^∞ norm is at most Cs_N with probability tending to one. This controls every coordinate Orlicz norm, including an L^∞ coordinate, and suffices to transfer bounds up to fixed constants.

Finally, the Gaussian statement on the square follows from the analogous torus statement (Section 17). Folding the unit torus onto the square pushes a periodic current to a no-flux current and sends white noise projected at mesh $h/2$ to square white noise projected at mesh h. The

equality of the Gaussian source laws is exact. We prove only the needed direction, from torus to square; folding does not give an inverse comparison.

10 Notation and statement of the reduction

Let $D = [0,1]^2$. Let $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ be independent sequences of independent uniform random variables in D, and set

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{X_i}, \qquad \nu_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{Y_i}, \qquad s_N = \sqrt{\frac{\log N}{N}}.$$

10.1 Matching and Orlicz norms

Throughout, the exponent pair $\boldsymbol{r} = (r_1, r_2)$ is fixed and satisfies the critical relation

$$2 \leq r_1, r_2 \leq \infty, \qquad \frac{1}{r_1} + \frac{1}{r_2} = \frac{1}{2}, \qquad \frac{1}{\infty} := 0. \tag{10.1}$$

Thus either $2 < r_1, r_2 < \infty$, or $\boldsymbol{r} = (2,\infty)$ or $(\infty, 2)$. We put $a_i = 1/r_i$ with $a_i = 0$ at the infinite endpoint. For $2 \leq r < \infty$, put

$$\psi_r(t) = e^{|t|^r} - 1.$$

For the endpoint, put

$$\psi_\infty(t) = \begin{cases} 0, & |t| \leq 1, \\ +\infty, & |t| > 1. \end{cases}$$

The Luxemburg norm associated with ψ_∞ is the L^∞ norm. For the exponent pair $\boldsymbol{r}$ we use the coordinate Young functions ψ_{r_1} and ψ_{r_2} throughout.

For a finite vector $b = (b_i)_{i=1}^N$, define

$$\|b\|_{\psi_{r_j},N} := \inf\left\{c > 0 : \frac{1}{N}\sum_{i=1}^{N}\psi_{r_j}(b_i/c) \leq 1\right\}.$$

For $r_j = \infty$, this is $\max_i |b_i|$.

For a permutation $\sigma \in S_N$, write $d_{i,j}^\sigma = \left|X_i^j - Y_{\sigma(i)}^j\right|$ and define

$$\mathfrak{M}_N^{\boldsymbol{r}} := \min_{\sigma\in S_N}\max_{j=1,2}\left\|(d_{i,j}^\sigma)_{i=1}^N\right\|_{\psi_{r_j},N}. \tag{10.2}$$

The two-sample form of the Ultimate Matching Conjecture is the following.

Definition 10.1. The conjecture $\mathrm{UMC}_{\boldsymbol{r}}$ is the assertion that there is a deterministic $C < \infty$ such that

$$\mathbb{P}\left(\mathfrak{M}_N^{\boldsymbol{r}} \leq C s_N\right) \longrightarrow 1. \tag{10.3}$$

Talagrand's Research Problem 17.1.2 [25] states the same coordinatewise criterion for a random sample and an evenly spread reference configuration. The symmetric quartic case is $\boldsymbol{r} = (4,4)$; the hard coordinate endpoint is $\boldsymbol{r} = (2,\infty)$.

Lemma 10.2. *For $\boldsymbol{r} = (4,4)$, the coordinate gauge in (10.2) is equivalent, up to a factor $\sqrt{2}$, to the isotropic gauge generated by $e^{|z|^4} - 1$.*

Proof. The lower bound follows from $|z_j| \leq |z|$. Conversely, if both coordinate Luxemburg norms are at most a, then $|z|^4 \leq 2(|z_1|^4 + |z_2|^4)$ and Cauchy–Schwarz yields

$$\frac{1}{N}\sum_i e^{|z_i|^4/(\sqrt{2}a)^4} \leq \prod_{j=1}^{2}\left(\frac{1}{N}\sum_i e^{|z_{i,j}|^4/a^4}\right)^{1/2} \leq 2.$$

□

10.2 Dyadic Gaussian flows

Write $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$. For $\Omega \in \{D, \mathbb{T}^2\}$, P_h^Ω denotes averaging onto the fixed h-square grid in Ω, acting on mean-zero functions when appropriate. With mean-zero white noise W^Ω (the isonormal process on $L_0^2(\Omega)$), set

$$W_h^\Omega = P_h^\Omega W^\Omega. \tag{10.4}$$

On the torus we also write $W_h = W_h^{\mathbb{T}}$ and $\mathfrak{F}_{\boldsymbol{r}} = \mathfrak{F}_{\boldsymbol{r}}^{\mathbb{T}}$ when there is no ambiguity.

Choose a sufficiently large constant A and a dyadic h_N satisfying

$$As_N \le h_N < 2As_N. \tag{10.5}$$

The Gaussian source on the square used in the reduction is

$$W_{h_N}^D = P_{h_N}^D W^D, \tag{10.6}$$

with the same white-noise normalization as W_h on the torus.

For a vector field $J = (J_1, J_2)$, put

$$\|J\|_{\boldsymbol{r}} := \max_{j=1,2} \|J_j\|_{\psi_{r_j}}. \tag{10.7}$$

For a mean-zero distribution f, we encode both the divergence equation and the no-flux boundary condition weakly. A vector field J is admissible for f when

$$-\int_D J \cdot \nabla\varphi \, dx = \langle f, \varphi \rangle \qquad \text{for every } \varphi \in C^\infty(\overline{D}). \tag{10.8}$$

For regular J, this is exactly $\operatorname{div} J = f$ in D together with the zero normal trace $J \cdot n = 0$ on ∂D. We then define

$$\mathfrak{F}_{\boldsymbol{r}}^D(f) := \inf \{ \|J\|_{\boldsymbol{r}} : J \text{ satisfies (10.8)} \}. \tag{10.9}$$

The corresponding Gaussian minimum-flow problem is the following.

Definition 10.3. The conjecture $\mathrm{GOMF}_{\boldsymbol{r}}$ asserts that there is a deterministic $C < \infty$ such that

$$\mathbb{P}\left(\mathfrak{F}_{\boldsymbol{r}}^D(W_{h_N}^D) \le C\sqrt{\log N} \right) \longrightarrow 1. \tag{10.10}$$

The main result of this part is the following equivalence.

Theorem 10.4. *For every fixed pair $\boldsymbol{r}$ satisfying* (10.1)*, and for A sufficiently large,* $\mathrm{UMC}_{\boldsymbol{r}}$ *and* $\mathrm{GOMF}_{\boldsymbol{r}}$ *are equivalent. More precisely, the following statements are equivalent:*

(i) $\mathbb{P}(\mathfrak{M}_N^{\boldsymbol{r}} \le Cs_N) \to 1$ *for some C;*

(ii)
$$\mathbb{P}\left(\mathfrak{F}_{\boldsymbol{r}}^D(P_{h_N}^D(\mu_N - \nu_N)) \le Cs_N \right) \to 1$$
for some C;

(iii)
$$\mathbb{P}\left(\mathfrak{F}_{\boldsymbol{r}}^D(W_{h_N}^D) \le C\sqrt{\log N} \right) \to 1$$
for some C.

All constants are deterministic and independent of N.

11 Preliminary estimates

We collect elementary facts on Orlicz norms and on the minimum-flow gauge. We then show that the gauge of projected white noise concentrates around its median, so that a bound with fixed positive probability implies a bound with high probability.

Lemma 11.1. *For every fixed $2 \le r < \infty$ on a probability space,*

$$\|Z\|_{\psi_r} \asymp_r \sup_{s \ge 1} s^{-1/r} \|Z\|_{L^s}. \tag{11.1}$$

Proof. The forward implication follows by integrating the tail furnished by $\mathbb{E}e^{|Z|^r/a^r} \le 2$. Conversely, expand the exponential in moments and use $k! \ge (k/e)^k$. □

Lemma 11.2. *Let Θ be any Young function and $K \ge 1$. If $\int \Theta(|f|/r) \le K$, then $\|f\|_{L^\Theta} \le Kr$.*

Proof. Convexity and $\Theta(0) = 0$ give $\Theta(t/K) \le K^{-1}\Theta(t)$. □

Lemma 11.3. *For mean-zero sources f, g,*

$$\left|\mathfrak{F}_{\boldsymbol{r}}^D(f) - \mathfrak{F}_{\boldsymbol{r}}^D(g)\right| \le \mathfrak{F}_{\boldsymbol{r}}^D(f-g), \qquad \mathfrak{F}_{\boldsymbol{r}}^D(af) = |a|\mathfrak{F}_{\boldsymbol{r}}^D(f).$$

Proof. Subtract admissible currents and use the triangle inequality in the product norm (10.7). □

11.1 Gaussian concentration for dyadic minimum-flow gauges

The argument of this subsection uses neither the matching reduction, nor the transport–flow comparison, nor any construction of the Gaussian current. It applies without change when one of the exponents equals ∞.

For $\Omega \in \{D, \mathbb{T}^2\}$, let $\mathfrak{F}_{\boldsymbol{r}}^\Omega$ be the minimum-flow gauge of (10.9), with the periodic divergence convention in the torus case. If $h = 2^{-m}$, write its projected mean-zero white noise in the form

$$W_h^\Omega(Z)|_Q = h^{-1}(Z_Q - \overline{Z}), \qquad \overline{Z} = h^2 \sum_Q Z_Q, \tag{11.2}$$

where the h-grid coefficients Z_Q are independent standard Gaussian variables.

Lemma 11.4. *Fix any $r_i \in [2, \infty]$. There is an absolute constant C, independent of h, the number of cells and the exponent pair, such that for $\Omega \in \{D, \mathbb{T}^2\}$ and coefficient arrays Z, Z',*

$$\left|\mathfrak{F}_{\boldsymbol{r}}^\Omega(W_h^\Omega(Z)) - \mathfrak{F}_{\boldsymbol{r}}^\Omega(W_h^\Omega(Z'))\right| \le C \|Z - Z'\|_{\ell^2}. \tag{11.3}$$

Proof. Write $v = Z - Z'$ and $g_Q = h(v_Q - \overline{v})$ for the mass of the source difference in a grid square Q. For any union A of grid squares, centering and Cauchy–Schwarz give

$$\left|\sum_{Q \in A} g_Q\right| \le 2h \|v\|_{\ell^2} \sqrt{\min\{\#A, \#A^c\}}.$$

The nearest-neighbor square grid satisfies the relative edge-isoperimetric inequality

$$|\partial A| \ge \sqrt{2}\sqrt{\min\{\#A, \#A^c\}},$$

where ∂A counts internal edges (and also periodic seam edges for $\Omega = \mathbb{T}^2$). Indeed, for the $k \times k$ grid with $k = h^{-1}$ and $\#A \le k^2/2$, the edge-isoperimetric inequality of Bollobás and Leader [9] gives $|\partial A| \ge \min\{2\sqrt{\#A}, k\} \ge \sqrt{2}\sqrt{\#A}$; the torus graph contains all edges of the square grid, so the same bound holds for $\Omega = \mathbb{T}^2$. Consequently every cut has enough capacity when each

internal edge is assigned capacity $Ch\,\|v\|_{\ell^2}$ in both directions. Gale's supply–demand theorem [14] (see also [19]), a consequence of the max-flow/min-cut theorem, then produces a consistent signed edge flux f_e with

$$\sum_{e\,\mathrm{out}(Q)} f_e - \sum_{e\,\mathrm{in}(Q)} f_e = g_Q, \qquad |f_e| \le Ch\,\|v\|_{\ell^2}\,.$$

On D, take the flux on exterior faces to be zero.

It remains to realize these integrated face fluxes as a continuum current. On $Q = [ih, (i+1)h] \times [jh, (j+1)h]$, put $t = (x-ih)/h$ and $u = (y-jh)/h$, and use globally consistent rightward/upward signed fluxes f_L, f_R, f_B, f_T. Define

$$J_1(x,y) = \frac{(1-t)f_L + tf_R}{h}, \qquad J_2(x,y) = \frac{(1-u)f_B + uf_T}{h}. \tag{11.4}$$

The normal components match across common faces (and periodically across torus seams). Their weak divergence on each cell is

$$\operatorname{div} J = \frac{f_R - f_L + f_T - f_B}{h^2} = h^{-1}(v_Q - \overline{v}).$$

This current satisfies the zero-normal-trace convention on D and $\max_i \|J_i\|_{L^\infty} \le C\,\|v\|_{\ell^2}$. Every Luxemburg ψ_r-norm with $r \ge 2$, including $r = \infty$, is bounded by a universal multiple of L^∞ on a space of unit area. The triangle inequality for minimum-flow gauges, Lemma 11.3, now gives (11.3). □

Proposition 11.5. *Let $h = 2^{-m}$, $\Omega \in \{D, \mathbb{T}^2\}$, and fix $r_1, r_2 \in [2, \infty]$. Suppose there are $B, p_0 > 0$ such that, for all sufficiently large m,*

$$\mathbb{P}\Big(\mathfrak{F}^{\Omega}_{\boldsymbol{r}}(W^{\Omega}_h) \le B\sqrt{m}\Big) \ge p_0.$$

Then there are $B', c > 0$ such that, for all sufficiently large m,

$$\mathbb{P}\Big(\mathfrak{F}^{\Omega}_{\boldsymbol{r}}(W^{\Omega}_h) > B'\sqrt{m}\Big) \le 2e^{-cm}. \tag{11.5}$$

Proof. By Lemma 11.4, the gauge is an L-Lipschitz function of a standard Gaussian vector, with L independent of the dimension. Gaussian concentration around a median M_m shows that $M_m \le B\sqrt{m} + C(L, p_0)$. Indeed, otherwise the lower tail below $B\sqrt{m}$ would have probability less than p_0. The upper concentration tail beyond a fixed larger multiple $B'\sqrt{m}$ is then at most $2e^{-cm}$. □

Corollary 11.6. *In the dyadic Gaussian condition (iii) of Theorem 10.4, it is equivalent to require the existence of $B, p_0 > 0$ such that*

$$\mathbb{P}\Big(\mathfrak{F}^{D}_{\boldsymbol{r}}(W^{D}_{h_N}) \le B\sqrt{\log N}\Big) \ge p_0$$

for all sufficiently large N. The same conclusion holds for a torus flow with its periodic divergence constraint.

Proof. The high-probability statement implies the positive-probability bound. Conversely, $h_N = 2^{-m(N)}$ and $m(N) \asymp \log N$, so Proposition 11.5 applies with a changed constant and yields the Gaussian condition (iii). □

Thus the Gaussian input to the matching equivalence need only be proved with a fixed positive probability.

12 From point configurations to histograms

We show that, up to an additive error of order h, matching the two point clouds is equivalent to transporting their dyadic histograms. This replaces the atomic problem by a transport problem between densities.

For a coupling π on $D \times D$, define

$$\mathfrak{T}_{\boldsymbol{r}}(\varrho_0, \varrho_1) := \inf_{\pi \in \Pi(\varrho_0\,\mathrm{d}x, \varrho_1\,\mathrm{d}x)} \max_{j=1,2} \|x_j - y_j\|_{L^{\psi_{r_j}}(\pi)}. \tag{12.1}$$

Let $P_h^D \mu_N$ be obtained by spreading each atomic mass uniformly on its dyadic cell.

Proposition 12.1. *There is a universal C such that, for every pair of point clouds and every dyadic h,*

$$\mathfrak{T}_{\boldsymbol{r}}(P_h^D \mu_N, P_h^D \nu_N) \le \mathfrak{M}_N^{\boldsymbol{r}} + Ch, \qquad \mathfrak{M}_N^{\boldsymbol{r}} \le C(\mathfrak{T}_{\boldsymbol{r}}(P_h^D \mu_N, P_h^D \nu_N) + h). \tag{12.2}$$

The constant is uniform over the critical exponents, including an L^∞ endpoint.

Proof. Fix a permutation. Couple the two associated cells using the same relative coordinate. Each coordinate displacement increases by at most $2h$, and the first inequality follows from the Luxemburg triangle inequality.

Conversely, let π couple the histograms. Attach posterior atomic labels inside every occupied cell and define

$$q_{ik} = N \int \lambda_i(x) \mu_k(y)\, \mathrm{d}\pi(x, y).$$

Then $Q = (q_{ik})$ is doubly stochastic. If the histogram coordinate norms are at most r, the fractional labelled coupling has both coordinate norms at most $r + 2h$. Decompose Q into permutation matrices. For finite coordinates, the average over this decomposition of the sum of the two coordinate modulars is at most 2, so one permutation has that sum at most 2; Lemma 11.2 absorbs this fixed factor. For an L^∞ coordinate, every positive entry of Q obeys the corresponding displacement bound, and a Birkhoff decomposition can be chosen inside the positive support. Thus the same permutation satisfies the hard coordinate bound. □

13 Transport and divergence currents

This section contains the deterministic core of the reduction. For densities bounded above and away from zero, we prove that the anisotropic transport gauge $\mathfrak{T}_{\boldsymbol{r}}$ is comparable to the minimum-flow gauge $\mathfrak{F}_{\boldsymbol{r}}^D$ of their difference, first for finite exponents and then at the endpoint $(2, \infty)$. The passage from transport to flow requires a bound on the densities along displacement interpolation, which we derive from noncrossing of transport rays and a Jacobian inequality. The domain D is convex, so straight segments remain in D.

13.1 Noncrossing of transport rays

Lemma 13.1. *Let $L : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ be strictly convex on its convex effective domain. Let $\Gamma \subset D \times D$ be c-cyclically monotone for $c(x, y) = L(y - x)$. For $0 < t < 1$, the map*

$$(x, y) \longmapsto (1 - t)x + ty$$

is injective on Γ.

Proof. Suppose $(1 - t)x_0 + ty_0 = (1 - t)x_1 + ty_1$ and put $v_i = y_i - x_i$. Then $x_1 - x_0 = t(v_0 - v_1)$, and the crossed displacements are

$$y_1 - x_0 = tv_0 + (1 - t)v_1, \qquad y_0 - x_1 = (1 - t)v_0 + tv_1.$$

They remain in the convex effective domain. If $v_0 \neq v_1$, strict convexity gives

$$L(y_1 - x_0) + L(y_0 - x_1) < L(v_0) + L(v_1),$$

contradicting two-point cyclical monotonicity. Hence $v_0 = v_1$, and then $x_0 = x_1$ and $y_0 = y_1$. □

13.2 Density bounds for displacement interpolation

Lemma 13.2. *Let $D \subset \mathbb{R}^2$ be bounded and convex, let $L \in C^2(\mathbb{R}^2)$, and assume that D^2L is uniformly positive definite on the bounded displacement set $D - D$. Let $c(x, y) = L(y - x)$ and let ϱ_0, ϱ_1 be smooth densities on D satisfying $0 < \underline{\varrho} \le \varrho_i \le \overline{\varrho}$. Then the c-optimal plan is induced by a unique map T. Moreover T is approximately differentiable almost everywhere, a Kantorovich potential is twice differentiable in the Alexandrov sense almost everywhere, and at almost every point where these differentials exist one has the measure-theoretic Jacobian identity*

$$\varrho_0(x) = \varrho_1(T(x)) \, |\det DT(x)|. \tag{13.1}$$

If, for some $t \in (0, 1)$, the interpolation map $T_t = (1-t)\mathrm{Id} + tT$ is injective on a full-$\varrho_0\,dx$-measure transport set, then the area formula applies without multiplicity and

$$\varrho_t(T_t(x)) \, |\det DT_t(x)| = \varrho_0(x) \tag{13.2}$$

almost everywhere on that set, where $\varrho_t\,dx = (T_t)_{\#}(\varrho_0\,dx)$.

Proof. For the translation cost considered here, strict convexity gives the twist condition and C^2 regularity on the bounded displacement set gives the local semiconcavity hypotheses. Existence and uniqueness of the optimal map are a special case of Villani's Monge theorem; see [27, Thm. 10.28 and Ex. 10.35]. Once injectivity and approximate differentiability are available, the Jacobian identity is precisely the Jacobian equation of [27, Thm. 11.1]. This theorem applies to $T_t = (1 - t)\mathrm{Id} + tT$ because approximate differentiability is preserved by affine combinations with the identity; the noncrossing lemma supplies injectivity for $0 < t < 1$. This yields (13.2) without a multiplicity factor. Only these almost-everywhere properties are used; no classical C^1 regularity of the optimal map is needed. □

Lemma 13.3. *Let $L \in C^2(\mathbb{R}^2)$ have Hessian uniformly positive definite on $D - D$, and let T be the optimal map for the cost $L(y - x)$ between smooth densities satisfying $0 < \underline{\varrho} \le \varrho_0, \varrho_1 \le \overline{\varrho}$. Put $T_t = (1 - t)\mathrm{Id} + tT$ and $\varrho_t = (T_t)_{\#}(\varrho_0\,\mathrm{d}x)$. Then*

$$\varrho_t \le \varrho_* := \frac{\overline{\varrho}^2}{\underline{\varrho}} \qquad (0 \le t \le 1). \tag{13.3}$$

Proof. Choose the Kantorovich-potential convention $\phi(x) + \chi(y) \le L(y - x)$ with equality at $y = T(x)$. Put $A = D^2L(T(x) - x)$. By Lemma 13.2, the required approximate differentials and the Jacobian identity exist almost everywhere. At Alexandrov points, touching by $x' \mapsto L(T(x) - x') - \chi(T(x))$ gives $D^2\phi \le A$, while differentiating $\nabla\phi(x) = -\nabla L(T(x) - x)$ gives

$$D^2\phi = A(I - DT).$$

Consequently

$$DT = I - A^{-1}D^2\phi = A^{-1/2}\big(I - A^{-1/2}D^2\phi A^{-1/2}\big)A^{1/2},$$

so DT is similar to the symmetric matrix

$$B = I - A^{-1/2}D^2\phi A^{-1/2} \ge 0.$$

Thus the eigenvalues λ_1, λ_2 of DT are real and nonnegative. In particular $\det DT \ge 0$, so the absolute value in (13.1) can now be removed. Fix $t \in (0, 1)$. By Lemma 13.1, T_t is injective on the

transport graph. Thus Lemma 13.2 applies without multiplicity. Moreover $\det DT_t > 0$ because $(1-t) + t\lambda_k > 0$, so the absolute value in (13.2) can also be removed. We have

$$\det DT_t = \prod_{k=1}^{2} ((1-t) + t\lambda_k) \geq \Big[(1-t) + t\sqrt{\det DT}\Big]^2 .$$

By (13.1), now without absolute values, $\det DT = \varrho_0/(\varrho_1 \circ T) \geq \underline{\varrho}/\overline{\varrho}$, hence $\det DT_t \geq \underline{\varrho}/\overline{\varrho}$. Using (13.2),

$$\varrho_t(T_t(x)) = \frac{\varrho_0(x)}{\det DT_t(x)} \leq \frac{\overline{\varrho}^2}{\underline{\varrho}}.$$

At $t = 0, 1$ the same bound follows directly from $\varrho_0, \varrho_1 \leq \overline{\varrho} \leq \overline{\varrho}^2/\underline{\varrho}$ (the unit-mass bounds imply $\underline{\varrho} \leq 1 \leq \overline{\varrho}$). □

Remark 13.4. With the opposite convention for the Kantorovich potential, all signs in the intermediate identities reverse simultaneously. The invariant statement is that DT is similar to a symmetric positive semidefinite matrix.

13.3 Boundary smoothing and approximation

The previous lemma requires smooth densities, whereas the histogram densities used later are piecewise constant. The passage between them requires a recovery approximation, not merely lower semicontinuity. We use the following local construction on the square.

Lemma 13.5. *Let $D = [0,1]^2$. There are Markov kernels $K_\tau(x, \mathrm{d}z)$, $0 < \tau < 1/4$, with smooth nonnegative densities $k_\tau(x, z)$ such that:*

(i) *K_τ preserves Lebesgue measure and is symmetric:*

$$\int_D k_\tau(x,z)\, \mathrm{d}z = 1, \qquad \int_D k_\tau(x,z)\, \mathrm{d}x = 1, \qquad k_\tau(x,z) = k_\tau(z,x).$$

(ii) *$k_\tau(x,z) = 0$ unless $|x - z|_\infty \leq \tau$.*

(iii) *If $S_\tau\varrho(z) = \int_D k_\tau(x,z)\varrho(x)\, \mathrm{d}x$, then $S_\tau\varrho$ is smooth, $S_\tau\varrho \to \varrho$ in $L^1(D)$, and*

$$\underline{\varrho} \leq \varrho \leq \overline{\varrho} \quad \Longrightarrow \quad \underline{\varrho} \leq S_\tau\varrho \leq \overline{\varrho}.$$

(iv) *If π couples $\varrho_0\, \mathrm{d}x$ and $\varrho_1\, \mathrm{d}x$, then*

$$\pi_\tau(\mathrm{d}x'\, \mathrm{d}y') := \int_{D\times D} K_\tau(x, \mathrm{d}x') K_\tau(y, \mathrm{d}y')\, \pi(\mathrm{d}x\, \mathrm{d}y)$$

couples $S_\tau\varrho_0\, \mathrm{d}x$ and $S_\tau\varrho_1\, \mathrm{d}x$. Moreover $|(y' - x') - (y - x)|_\infty \leq 2\tau$ on the support of the lifted coupling. Consequently, for every bounded continuous displacement cost $c(y - x)$,

$$\int c(y' - x')\, \mathrm{d}\pi_\tau \longrightarrow \int c(y - x)\, \mathrm{d}\pi.$$

If in addition $|y_2 - x_2| \leq r$ π-a.s., then $|y_2' - x_2'| \leq r + 2\tau$ π_τ-a.s.

Proof. Choose an even $C_c^\infty(\mathbb{R})$ probability density η_τ supported in $(-\tau, \tau)$. In one dimension define the reflected, two-periodic kernel on $[0,1]$ by the method of images

$$k_\tau^{(1)}(x,z) = \sum_{n\in\mathbb{Z}} \Big[\eta_\tau(z - x + 2n) + \eta_\tau(z + x + 2n)\Big], \qquad x, z \in [0,1].$$

It is nonnegative, symmetric, smooth, and its integral in either variable is one. Equivalently, it is convolution by η_τ after even 2-periodic reflection. Since $\tau < 1/4$, every nonzero image term has $|x - z| \le \tau$; for the reflected terms near 0 or 1, this follows from $x + z \le \tau$ or $2 - x - z \le \tau$. Take the product of the two one-dimensional kernels to obtain k_τ on D.

The Markov and bistochastic identities imply preservation of mass and the pointwise bounds $\underline{\varrho}, \overline{\varrho}$. Standard approximate-identity theory for the even periodic extension gives L^1 convergence and smoothness. The formula for π_τ gives the asserted marginals by Fubini. Local support gives $|x' - x|_\infty, |y' - y|_\infty \le \tau$, hence the displacement and hard-support bounds. Uniform continuity of a bounded continuous cost on the compact set $D - D$ gives the final convergence. □

Corollary 13.6. *Assume $0 < \underline{\varrho} \le \varrho_i \le \overline{\varrho}$ and let $\varrho_i^\tau = S_\tau \varrho_i$. If $r > \mathfrak{T}_{\boldsymbol{r}}(\varrho_0, \varrho_1)$ and $r' > r$, then for all sufficiently small τ there is a coupling of ϱ_0^τ and ϱ_1^τ whose finite-coordinate modulars at scale r' are bounded by a fixed constant independent of τ. At a hard coordinate, a bound $|y_j - x_j| \le r$ is replaced by $|y'_j - x'_j| \le r'$. The fixed modular constant may be absorbed into a fixed further enlargement of r' by Lemma 11.2.*

Proof. Start from a coupling witnessing the gauge at scale r and apply Lemma 13.5. For finite coordinates the costs at the larger scale r' are bounded continuous functions on $D - D$ and their integrals converge to those of the original coupling, which are at most one at scale r and hence at most one at scale r'. For a hard coordinate choose $\tau < (r' - r)/2$. □

13.4 Finite critical exponents

For finite r_1, r_2, set

$$L_r(z) = \sum_{j=1}^{2} \psi_{r_j}(|z_j|/r).$$

The generalized Benamou–Brenier formula for this convex Lagrangian is standard; see, for example, [22, 27]:

$$\inf_\pi \int L_r(y - x)\, \mathrm{d}\pi = \inf_{\partial_t \varrho_t + \operatorname{div} J_t = 0} \int_0^1 \int_D \varrho_t L_r(J_t/\varrho_t)\, \mathrm{d}x\, \mathrm{d}t. \tag{13.4}$$

Theorem 13.7. *Assume $2 < r_1, r_2 < \infty$ and $1/r_1 + 1/r_2 = 1/2$, and let $0 < \underline{\varrho} \le \varrho_0, \varrho_1 \le \overline{\varrho}$. Then*

$$C^{-1}_{\underline{\varrho}, \overline{\varrho}} \mathfrak{F}^D_{\boldsymbol{r}}(\varrho_0 - \varrho_1) \le \mathfrak{T}_{\boldsymbol{r}}(\varrho_0, \varrho_1) \le C_{\underline{\varrho}, \overline{\varrho}} \mathfrak{F}^D_{\boldsymbol{r}}(\varrho_0 - \varrho_1), \tag{13.5}$$

with a constant depending only on $\underline{\varrho}, \overline{\varrho}$ and the fixed exponents.

Proof. Flow to transport. Let $\operatorname{div} J = \varrho_0 - \varrho_1$, put $\varrho_t = (1 - t)\varrho_0 + t\varrho_1$ and $J_t = J$, and let $a > \|J\|_{\boldsymbol{r}}$. Since $\varrho_t \ge \underline{\varrho}$ and the perspective $s \mapsto s\psi_{r_j}(q/s)$ is decreasing,

$$\int_0^1 \int \varrho_t \psi_{r_j}\left(\frac{|J_j|}{\varrho_t(a/\underline{\varrho})}\right) \le \underline{\varrho} \int \psi_{r_j}(|J_j|/a) \le \underline{\varrho}.$$

Thus the total $L_{a/\underline{\varrho}}$ action is at most $2\underline{\varrho} \le 2$. By (13.4) and Lemma 11.2, the two displacement norms are at most $2a/\underline{\varrho}$.

Transport to flow. Take $r > \mathfrak{T}_{\boldsymbol{r}}(\varrho_0, \varrho_1)$ and then fix $r' > r$. Let $\varrho_i^\tau = S_\tau \varrho_i$. By Corollary 13.6, for small τ there is a coupling of the smooth densities $\varrho_0^\tau, \varrho_1^\tau$ whose two coordinate modulars at scale r' are bounded by a universal fixed constant. After a fixed enlargement of r' using Lemma 11.2, we may assume that their sum is at most $2 + o_\tau(1)$.

For $\varepsilon > 0$, minimize $L_{r'}(z) + \varepsilon|z|^2$ between the smoothed marginals. The minimizer is induced by a map and Lemma 13.3 applies. Its interpolation densities obey $\varrho_t^{\tau,\varepsilon} \le \varrho_* = \overline{\varrho}^2/\underline{\varrho}$, uniformly in

both τ and ε. If $v_t(T_t(x)) = T(x) - x$ and $J_t = \varrho_t v_t$, set $J^{\tau,\varepsilon} = \int_0^1 J_t \,\mathrm{d}t$ and $a(x) = \int_0^1 \varrho_t(x)\,\mathrm{d}t \le \varrho_*$. Coordinatewise Jensen gives

$$\varrho_* \psi_{r_j}\left(\frac{|J_j^{\tau,\varepsilon}|}{\varrho_* r'}\right) \le \int_0^1 \varrho_t \psi_{r_j}(|v_{t,j}|/r')\,\mathrm{d}t.$$

Hence the currents are bounded, uniformly in τ, ε, in the corresponding Orlicz spaces at scale $C_{\underline{\varrho},\overline{\varrho}} r'$.

First let $\varepsilon \downarrow 0$. Then let $\tau \downarrow 0$. The superlinear Young functions give uniform integrability of the current coordinates, hence weak compactness in L^1; the divergence constraint passes distributionally because $\varrho_i^\tau \to \varrho_i$ in L^1. Lower semicontinuity of convex integral functionals gives a limiting current J with $\operatorname{div} J = \varrho_0 - \varrho_1$ and $\|J_j\|_{L^{\psi_{r_j}}} \le C_{\underline{\varrho},\overline{\varrho}} r'$. Finally let $r' \downarrow r$ and $r \downarrow \mathfrak{T}_{\boldsymbol{r}}(\varrho_0, \varrho_1)$.

The smoothing step is where Corollary 13.6 enters: it gives an upper bound for the smoothed transport problem, whereas lower semicontinuity alone gives the inequality in the wrong direction. □

13.5 The critical endpoint $(2, \infty)$

Assume, after exchanging coordinates if necessary, that $\boldsymbol{r} = (2, \infty)$.

Theorem 13.8. *If $0 < \underline{\varrho} \le \varrho_0, \varrho_1 \le \overline{\varrho}$, then*

$$\mathfrak{T}_{(2,\infty)}(\varrho_0, \varrho_1) \asymp_{\underline{\varrho},\overline{\varrho}} \mathfrak{F}^D_{(2,\infty)}(\varrho_0 - \varrho_1). \tag{13.6}$$

Proof. Flow to transport. Let $a > \max(\|J_1\|_{\exp L^2}, \|J_2\|_\infty)$ and use the linear density path. Then $|J_2/\varrho_t| \le a/\underline{\varrho}$. By the weak admissibility identity (10.8), extending ϱ_t and J by zero outside D produces a distributional continuity equation on $\mathbb{R}^2$ with no boundary source. The standard superposition principle for continuity equations, see [2], then produces a dynamic plan concentrated on trajectories in $\overline{D}$ whose endpoints satisfy $|y_2 - x_2| \le a/\underline{\varrho}$. Jensen along trajectories and the perspective estimate control the first coordinate exponential-square modular at the same scale, up to a fixed factor.

Transport to flow. Let r exceed the endpoint transport gauge and fix $r' > r$. Choose a coupling π_0 with

$$|y_2 - x_2| \le r \quad \pi_0\text{-a.s.}, \qquad \int \psi_{r_1}(|y_1 - x_1|/r)\,\mathrm{d}\pi_0 \le 1.$$

Smooth the two marginals and the coupling by Lemma 13.5. For all sufficiently small τ the resulting coupling π_0^τ has smooth marginals $\varrho_i^\tau = S_\tau \varrho_i$, still satisfying $\underline{\varrho} \le \varrho_i^\tau \le \overline{\varrho}$, and

$$|y_2 - x_2| \le r' \quad \pi_0^\tau\text{-a.s.}, \qquad \int \psi_{r_1}(|y_1 - x_1|/r')\,\mathrm{d}\pi_0^\tau \le 1 + o_\tau(1).$$

Thus it is enough to obtain estimates uniform in τ for these smooth marginals.

Use the explicit smooth convex barriers

$$b_k(s) = k\,(|s| - 1)_+^4 + k^{-1} s^2.$$

They are C^2, strictly convex, satisfy $\sup_{|s| \le 1} b_k(s) \le k^{-1}$, and diverge uniformly on $\{|s| \ge 1 + \eta\}$ for every $\eta > 0$. For fixed τ, minimize the strictly convex smooth costs

$$L_{r',k}(z) = \psi_{r_1}(|z_1|/r') + b_k(z_2/r').$$

The smoothed benchmark gives a cost bounded uniformly in k and τ. For every k, Lemma 13.3 applies to the smooth marginals and gives $\varrho_t^{\tau,k} \le \varrho_* = \overline{\varrho}^2/\underline{\varrho}$. Passing first $k \to \infty$ yields a coupling π^τ satisfying

$$|y_2 - x_2| \le r' \quad \pi^\tau\text{-a.s.}, \qquad \int \psi_{r_1}(|y_1 - x_1|/r')\,\mathrm{d}\pi^\tau \le C,$$

and the interpolation density bound passes to the limit by weak convergence against nonnegative test functions.

Let $\varrho_t^\tau = (e_t)_\# \pi^\tau$ and $J_t^\tau = (e_t)_\#((y-x)\pi^\tau)$. Disintegrating over e_t shows

$$|J_{t,2}^\tau| \le r' \varrho_t^\tau \le r' \varrho_* \quad \text{a.e.}$$

Hence $J^\tau = \int_0^1 J_t^\tau \, \mathrm{d}t$ satisfies $\|J_2^\tau\|_\infty \le r' \varrho_*$, while the first component is uniformly bounded in $\exp L^2$ by the same perspective/Jensen estimate as in Theorem 13.7. Extracting a subsequence as $\tau \downarrow 0$, use weak compactness in L^1 for the first component and weak-* compactness in L^∞ for the second. The divergence equation passes to the limit and convex lower semicontinuity yields a current J with

$$\operatorname{div} J = \varrho_0 - \varrho_1, \qquad \|J_1\|_{\exp L^2} + \|J_2\|_\infty \le C_{\underline{\varrho},\overline{\varrho}} r'.$$

Finally let $r' \downarrow r$ and then $r \downarrow \mathfrak{T}_{(2,\infty)}$. □

The hard constraint enters this proof only in the limit $k \to \infty$; no uniform L^∞ bound is used for the approximating problems before the convex barriers become infinite.

14 Bounds on the histogram densities

The transport–flow comparison of the previous section requires densities bounded above and below. We check that the histograms satisfy this with probability tending to one.

Lemma 14.1. *For every fixed $\varepsilon \in (0,1)$, if A in (10.5) is sufficiently large, then*

$$\mathbb{P}\left(1-\varepsilon \le P_{h_N}^D \mu_N, P_{h_N}^D \nu_N \le 1+\varepsilon \text{ on } D\right) \longrightarrow 1.$$

Proof. A terminal cell count has mean $\lambda_N = N h_N^2 \asymp A^2 \log N$. Chernoff's inequality and a union bound over the $h_N^{-2} = N/\lambda_N$ cells give

$$\mathbb{P}\left(\max_Q |N\mu_N(Q) - \lambda_N| > \varepsilon \lambda_N\right) \le \frac{2N}{\lambda_N} e^{-c\varepsilon^2 \lambda_N},$$

and similarly for ν_N. □

Combining Proposition 12.1, Theorems 13.7 and 13.8, and Lemma 14.1 proves the equivalence of statements (i) and (ii) in Theorem 10.4.

15 Coupling the histograms with dyadic Gaussian noise

It remains to compare the histogram source with the projected Gaussian source. We use the one-dimensional empirical-process approximation of Komlós, Major and Tusnády [20], in the nonasymptotic form of Bretagnolle and Massart [10, Theorem 1]; see also [13, Theorem 1.8 and equation (1.1)]. For the empirical distribution function F_N of independent uniform variables on $[0,1]$, it provides a coupled standard Brownian bridge B such that, with

$$\Delta_N = \sup_{0 \le t \le 1} |F_N(t) - t - N^{-1/2} B(t)|,$$

one has $\mathbb{P}\{\Delta_N > (12 \log N + x)/N\} \le 2e^{-x/6}$ for all $x \ge 0$. In particular there are absolute positive constants C, c such that

$$\mathbb{P}\{\Delta_N > C(\log N + x)/N\} \le C e^{-cx}, \qquad N \ge 2, \quad x > 0. \tag{15.1}$$

We encode the fixed planar partition by intervals and apply this theorem directly. Since we only compare bounds with fixed constants, a correction of size $O(s_N)$ with probability tending to one suffices.

A deterministic divergence inverse. Resolve the fixed $h = 2^{-m}$ square grid by alternating vertical and horizontal equal binary splits, for $2m$ generations. Each level-k rectangle has area 2^{-k} and aspect ratio between $1/2$ and 2. For an internal rectangle Q with children Q^0, Q^1, put

$$\mathsf{h}_Q = |Q|^{-1/2}(\mathbf{1}_{Q^0} - \mathbf{1}_{Q^1}).$$

For a vertical split $Q = I \times K$, let v_I vanish at both endpoints of I and satisfy $v_I' = \mathbf{1}_{I^0} - \mathbf{1}_{I^1}$. Set $V_Q = |Q|^{-1/2} v_I(x)\mathbf{1}_K(y)e_1$, extended by zero; for horizontal splits use the analogous e_2 field. Integration by parts in the active coordinate gives

$$\int_D V_Q \cdot \nabla\phi = -\int_D \mathsf{h}_Q\phi, \qquad \|V_Q\|_\infty \le C, \qquad \phi \in C^1(\overline{D}). \tag{15.2}$$

There is no boundary contribution: on faces normal to the active component the primitive vanishes, and on the other faces it is tangential.

If g is constant on the terminal squares and $\int_D g = 0$, set $d_Q = \int_{Q^0} g - \int_{Q^1} g$. The tree Haar basis gives

$$g = \sum_Q \frac{d_Q}{\sqrt{|Q|}}\mathsf{h}_Q, \qquad R = \sum_Q \frac{d_Q}{\sqrt{|Q|}}V_Q.$$

Thus $\operatorname{div} R = g$ with zero normal flux in the weak sense of (15.2). At almost every point only one rectangle per generation contributes. Consequently

$$\|R\|_\infty \le C \max_Q |d_Q| \sum_{k=0}^{2m-1} 2^{k/2} \le \frac{C}{h}\max_Q |d_Q|. \tag{15.3}$$

All sums are finite. For $m = 0$ the mean-zero grid-constant source is zero.

The one-cloud coupling. Order the terminal squares recursively, placing all descendants of Q^0 before those of Q^1, and assign them consecutive intervals of length h^2 in $[0,1]$. Every tree rectangle Q then corresponds to a single interval I_Q of length $|Q|$. Take the uniform sample in (15.1). For each sample variable in a leaf interval I_S, independently choose a uniform point in S. These points are independent and uniform in D, and their counts on tree rectangles equal the interval counts. The construction does not condition on the event where the KMT bound holds. Fixed interval endpoints are avoided almost surely.

For $I_S = (a_S, b_S]$, define W_h^X on S by

$$\int_S W_h^X = B(b_S) - B(a_S).$$

These masses are centered jointly Gaussian and satisfy

$$\operatorname{Cov}\left(\int_S W_h^X, \int_T W_h^X\right) = |S|\mathbf{1}_{S=T} - |S||T|.$$

Hence W_h^X has exactly the law of $P_h^D W^D$, including the mean-zero constraint. For $g = P_h^D\mu_N - 1 - N^{-1/2}W_h^X$, interval increments of the error in (15.1) give $|\int_Q g| \le 2\Delta_N$ and $|d_Q| \le 4\Delta_N$. The deterministic inverse therefore supplies a current R_h^X with

$$\operatorname{div} R_h^X = P_h^D\mu_N - 1 - N^{-1/2}W_h^X, \qquad \mathbb{P}\left\{\left\|R_h^X\right\|_\infty > C\frac{\log N + x}{Nh}\right\} \le Ce^{-cx}. \tag{15.4}$$

Its normal flux is zero. The joint construction may depend on N and h; this is enough for the probability comparisons below.

Theorem 15.1. *If $Nh^2 \asymp \log N$, there is a coupling of two independent uniform point clouds and a field W_h^D having exactly the law of $P_h^D W^D$, together with a zero-normal-flux current Q_h, such that*

$$\operatorname{div} Q_h = P_h^D(\mu_N - \nu_N) - \sqrt{2/N}\, W_h^D. \tag{15.5}$$

For a deterministic $C < \infty$,

$$\mathbb{P}\Big\{ \|Q_h\|_\infty \le C\frac{\log N}{Nh} \Big\} \longrightarrow 1. \tag{15.6}$$

In particular $\|Q_h\|_{\boldsymbol{r}} \le C s_N$ with probability tending to one for every critical exponent pair, including the endpoints.

Proof. Use independent copies of (15.4) for the two clouds and put

$$W_h^D = (W_h^X - W_h^Y)/\sqrt{2}, \qquad Q_h = R_h^X - R_h^Y.$$

The Gaussian law and the divergence identity follow directly. Take x to be a sufficiently large fixed multiple of $\log N$ in (15.4) and use a union bound. Since $Nh^2 \asymp \log N$, we have $\log N/(Nh) \asymp s_N$. Finally, on the unit-area square, $\|f\|_{\psi_r} \le (\log 2)^{-1/r}\, \|f\|_\infty$ for $2 \le r < \infty$, and $\|f\|_{\psi_\infty} = \|f\|_\infty$. □

The correction Q_h and the Gaussian source W_h^D are in general dependent; the comparison below does not use independence.

16 Proof of the reduction

We now assemble the previous sections into the proof of Theorem 10.4, and then derive the one-cloud version of the matching estimate.

Proof of Theorem 10.4. By Proposition 12.1, atomic matching and histogram transport are equivalent at scale s_N, because $h_N \asymp s_N$. On the event of Lemma 14.1, the transport–flow comparison Theorems 13.7 and 13.8 identifies histogram transport with the corresponding static minimum-flow gauge, for finite and endpoint exponents respectively. This proves (i)⇔(ii).

By Lemma 11.3 and Theorem 15.1 and the homogeneity of the minimum-flow gauge,

$$\left| \mathfrak{F}_{\boldsymbol{r}}^D(P_{h_N}^D(\mu_N - \nu_N)) - \sqrt{\frac{2}{N}}\, \mathfrak{F}_{\boldsymbol{r}}^D(W_{h_N}^D) \right| \le \|Q_{h_N}\|_{\boldsymbol{r}}\,.$$

By (15.6), the right-hand side is at most $C s_N$ with probability tending to one for every critical pair. Intersecting this event with a high-probability bound for either flow gauge transfers that bound to the other after changing its constant; no independence is required. Since $\sqrt{2/N}\,\sqrt{\log N} = \sqrt{2}\, s_N$, this proves (ii)⇔(iii). □

16.1 One cloud and Talagrand's evenly spread references

The same argument gives the original one-cloud formulation of the Ultimate Matching Conjecture, and not only its two-sample analogue. It uses no additional Gaussian input.

Corollary 16.1. *Fix a critical exponent pair $\boldsymbol{r}$ as in* (10.1) *and let h_N satisfy* (10.5). *Suppose that for some deterministic C,*

$$\mathbb{P}\Big(\mathfrak{F}_{\boldsymbol{r}}^D(W_{h_N}^D) \le C\sqrt{\log N} \Big) \longrightarrow 1. \tag{16.1}$$

Let $Y_1, \dots, Y_N$ be any deterministic evenly spread family in D: there are rectangles R_i partitioning D up to null sets, with $|R_i| = 1/N$, $Y_i \in R_i$, and $R_i \subset B(Y_i, 10/\sqrt{N})$. Then for some deterministic $C' < \infty$, independent of the chosen family,

$$\mathbb{P}\left(\min_{\sigma \in S_N} \max_{j=1,2} \big\| (X_i^j - Y_{\sigma(i)}^j)_{i=1}^N \big\|_{\psi_{r_j}, N} \le C' \sqrt{\frac{\log N}{N}} \right) \longrightarrow 1. \tag{16.2}$$

The conclusion includes a hard coordinate endpoint whenever the corresponding Gaussian hypothesis (16.1) holds; the corollary does not supply that hypothesis.

Proof. Let $\varrho_h = P^D_{h_N}\mu_N$. The one-cloud coupling (15.4) gives

$$\varrho_h - 1 = N^{-1/2} W^X_{h_N} + \operatorname{div} R^X_{h_N},$$

where $W^X_{h_N}$ has the law of $W^D_{h_N}$ and $\left\|R^X_{h_N}\right\|_\infty \le C s_N$ with probability tending to one. The same bound holds in the coordinate Orlicz gauge. Therefore (16.1) and the flow-gauge triangle inequality give

$$\mathfrak{F}^D_{\boldsymbol{r}}(\varrho_h - 1) \le N^{-1/2}\mathfrak{F}^D_{\boldsymbol{r}}(W^X_{h_N}) + \left\|R^X_{h_N}\right\|_{\boldsymbol{r}} \le C_1 s_N$$

with probability tending to one. By Lemma 14.1 applied to the first cloud, $1/2 \le \varrho_h \le 3/2$ on another event of probability tending to one. Hence the transport–flow comparison, Theorems 13.7 and 13.8, yields a coupling from $\varrho_h\,\mathrm{d}x$ to $\mathrm{d}x$ whose two coordinate Luxemburg norms are bounded by $C_2 s_N$.

There is also a coupling from μ_N to $\varrho_h\,\mathrm{d}x$ which moves each atom only inside its h_N-cell, with coordinate displacement at most h_N. Finally, the deterministic map which sends $x \in R_i$ to Y_i couples Lebesgue measure to $N^{-1}\sum_i \delta_{Y_i}$, moving every point by at most $10/\sqrt{N}$. Gluing these three couplings and using the Luxemburg triangle inequality gives a coupling γ between the two equal-weight atomic measures for which

$$\max_j \|x_j - y_j\|_{L^{\psi_{r_j}}(\gamma)} \le C_2 s_N + h_N + 10/\sqrt{N} \le C_3 s_N.$$

It remains to replace γ by a permutation. The coupling γ determines a doubly stochastic matrix, so by Birkhoff's theorem it is a convex combination of matching permutations. At a scale slightly larger than $C_3 s_N$, the mean over these permutations of the sum of the two coordinate Young-function modulars is at most 2. Therefore some single permutation has this sum at most 2, and a fixed further enlargement of scale makes each coordinate modular at most 1. For $r_j = \infty$, all positive entries of the doubly stochastic matrix already obey the hard coordinate bound; we select a permutation in that support while controlling the remaining finite-coordinate modular. This yields (16.2) with a constant uniform over the evenly spread family. □

17 From the torus to the square

The reduction requires a Gaussian source on the square, whereas the flow construction of Part I yields a periodic current on $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$. In this section we show that folding the torus onto the square relates the two projected sources exactly, so that a torus bound implies the corresponding square bound (Proposition 17.3).

17.1 Folding of currents

On $\mathbb{T}^1 = [0,1)$ define

$$F_1(u) = 2\min\{u, 1-u\}, \qquad F = (F_1, F_1) : \mathbb{T}^2 \longrightarrow D = [0,1]^2.$$

The notation $F = (F_1, F_1)$ means $F(u_1, u_2) = (F_1(u_1), F_1(u_2))$. Outside the folds $u_i = 0, 1/2$, the derivative is $DF(u) = 2R_u$, where R_u is diagonal with entries ± 1. Each $y \in (0,1)^2$ has exactly four preimages. For integrable scalar functions and vector fields define

$$(\mathscr{S}_0 f)(y) = \frac{1}{4}\sum_{F(u)=y} f(u), \qquad (\mathscr{S}_1 J)(y) = \frac{1}{2}\sum_{F(u)=y} R_u J(u). \tag{17.1}$$

These formulas are the pushforward against normalized Lebesgue measure; in particular $F_{\#}\mathrm{d}u = \mathrm{d}y$. For mean-zero distributions the scalar pushforward is defined by $\langle \mathscr{S}_0 f, \varphi\rangle = \langle f, \varphi \circ F\rangle$.

Lemma 17.1. *Suppose $J \in L^1(\mathbb{T}^2;\mathbb{R}^2)$ and $\operatorname{div}_{\mathbb{T}} J = f$ in distributions. Then $J^D = \mathscr{S}_1 J$ satisfies*

$$-\int_D J^D \cdot \nabla\varphi \, \mathrm{d}y = \langle \mathscr{S}_0 f, \varphi\rangle \quad \text{for every } \varphi \in C^1(\overline{D}).$$

In particular, J^D has zero normal trace on ∂D. For every Young function Θ and $j = 1, 2$,

$$\int_D \Theta\left(\frac{|J_j^D|}{2a}\right) \mathrm{d}y \le \int_{\mathbb{T}^2} \Theta\left(\frac{|J_j|}{a}\right) \mathrm{d}u, \qquad \|J_j^D\|_{L^\Theta(D)} \le 2\|J_j\|_{L^\Theta(\mathbb{T}^2)}. \tag{17.2}$$

The same inequality holds when $L^\Theta = L^\infty$.

Proof. On every inverse sheet $\mathrm{d}u = \mathrm{d}y/4$, and the chain rule gives $\nabla(\varphi \circ F)(u) = 2R_u \nabla\varphi(F(u))$. Thus

$$-\int_{\mathbb{T}^2} J \cdot \nabla(\varphi \circ F) \, \mathrm{d}u = -\int_D \left(\frac{1}{2}\sum_{F(u)=y} R_u J(u)\right) \cdot \nabla\varphi(y) \, \mathrm{d}y.$$

For an L^1 current, testing with the periodic Lipschitz function $\varphi \circ F$ is justified by smooth approximation; in our application the source is a finite-dimensional Gaussian cell function. The distributional divergence identity on the torus gives the first claim. Since the test functions need not vanish on ∂D, this identity also shows that the fold creates no boundary flux. Finally $J_j^D = 2 \cdot (1/4) \sum \varepsilon_{u,j} J_j(u)$, and Jensen's inequality together with $F_{\#}\mathrm{d}u = \mathrm{d}y$ proves (17.2). □

17.2 Compatibility with dyadic projections

Recall that $P_h^{\mathbb{T}}$ and P_h^D are the fixed-grid conditional expectations. Restricted to L_0^2, their ranges are the mean-zero cellwise-constant subspaces. Accordingly the projected noises $W_h^{\mathbb{T}}, W_h^D$ have exactly the meaning specified in (10.4). If $h = 2^{-m}$ is a square mesh, take the torus mesh

$$\widetilde{h} = h/2 = 2^{-(m+1)}.$$

The four folding lines coincide with grid lines at this mesh. Thus every torus $\widetilde{h}$-square is mapped bijectively onto a square h-square, and the following deterministic relation holds up to cell-boundary null sets:

$$\mathscr{S}_0 P_{\widetilde{h}}^{\mathbb{T}} f = P_h^D \mathscr{S}_0 f. \tag{17.3}$$

Lemma 17.2. *Let $W^{\mathbb{T}}$ be centered mean-zero isonormal white noise over $L_0^2(\mathbb{T}^2)$. For $h = 2^{-m}$ and $\widetilde{h} = h/2$,*

$$\mathscr{S}_0 W_{\widetilde{h}}^{\mathbb{T}} \overset{\text{law}}{=} P_h^D W^D, \tag{17.4}$$

where W^D is standard mean-zero white noise on D.

Proof. For deterministic $\phi, \psi \in L_0^2(D)$, the isonormal covariance of the folded field is

$$\begin{aligned}\mathbb{E}[\langle \mathscr{S}_0 W_{\widetilde{h}}^{\mathbb{T}}, \phi\rangle \langle \mathscr{S}_0 W_{\widetilde{h}}^{\mathbb{T}}, \psi\rangle] &= \langle P_{\widetilde{h}}^{\mathbb{T}}(\phi \circ F), P_{\widetilde{h}}^{\mathbb{T}}(\psi \circ F)\rangle_{L^2(\mathbb{T}^2)} \\ &= \langle P_h^D \phi, P_h^D \psi\rangle_{L^2(D)}.\end{aligned}$$

The second equality holds because F is measure preserving and the reflected dyadic mesh is exactly aligned. Both fields are centered Gaussian on the same finite-dimensional cell space, so equality of covariances proves equality in law. □

Proposition 17.3. *For every pair $\boldsymbol{r}$ for which the coordinate Orlicz spaces are defined, and every $m \ge 1$,*

$$\mathbb{P}\left\{\mathfrak{F}_{\boldsymbol{r}}^D(P_{2^{-m}}^D W^D) \le 2t\right\} \ge \mathbb{P}\left\{\mathfrak{F}_{\boldsymbol{r}}^{\mathbb{T}}(P_{2^{-(m+1)}}^{\mathbb{T}} W^{\mathbb{T}}) \le t\right\}, \qquad t > 0.$$

Consequently, a bound $O(\sqrt{m})$ with high probability for the unit-torus Gaussian flow implies the identical exact-cell-source square Gaussian estimate, with at most a universal change of constant.

Proof. For any torus sample and any admissible periodic current, use Lemma 17.1 to obtain a no-flux square current for the folded source, at coordinate cost at most twice the torus cost. Take infima (or approximating sequences) and then use Lemma 17.2 to identify the source law. □

Remark 17.4. A general no-flux square current cannot be pulled back to the torus at unchanged Gaussian source law simply by reversing the four-sheeted averaging: the fold destroys information from three independent complementary sheet modes. The matching–square-flow equivalence is symmetric, but the folding result is used only in the torus-to-square direction.

18 Choice of dyadic mesh

The torus Gaussian bound is formulated along the meshes $h_m = 2^{-m}$ and expressed in $\sqrt{m}$, whereas the matching reduction uses the mesh h_N of (10.5). We check that these two parametrizations describe the same family of dyadic projections, and then combine the previous sections into a matching criterion stated on the torus.

In the square matching reduction one chooses a dyadic h_N with $A\sqrt{\log N/N} \le h_N < 2A\sqrt{\log N/N}$. Given N, write $h_N = 2^{-m(N)}$, so that

$$m(N) = \frac{\log N}{2\log 2} + O(\log\log N), \qquad \sqrt{m(N)} \asymp \sqrt{\log N}.$$

The square-source estimate follows from the torus theorem at the single adjacent depth $m(N)+1$. Conversely, a square-source high-probability bound formulated at every h_N gives a bound along every m by choosing integers N_m with

$$N_m \asymp A^2 m 4^m, \qquad A\sqrt{\frac{\log N_m}{N_m}} \le 2^{-m} < 2A\sqrt{\frac{\log N_m}{N_m}}.$$

For instance, one may take $N_m = \lceil 4A^2(\log 2)m4^m\rceil$: then $A\sqrt{(\log N_m)/N_m}/2^{-m} \to 1/\sqrt{2}$, which lies strictly between $1/2$ and 1. This reverse calibration concerns the square Gaussian gauges; it does not reverse the torus fold.

Corollary 18.1. *Fix $\boldsymbol{r}$ satisfying* (10.1)*. Suppose that there exist constants $B, p_0 > 0$ such that, for every sufficiently large m,*

$$\mathbb{P}\left\{\mathfrak{F}^{\mathbb{T}}_{\boldsymbol{r}}(P^{\mathbb{T}}_{2^{-m}}W^{\mathbb{T}}) \le B\sqrt{m}\right\} \ge p_0.$$

Then Talagrand's square matching estimate follows:

$$\mathbb{P}\left\{\mathfrak{M}^{\boldsymbol{r}}_N \le C'\sqrt{\log N/N}\right\} \longrightarrow 1$$

for a deterministic $C' < \infty$.

Proof. Proposition 11.5, applied on $\mathbb{T}^2$, upgrades the assumed lower probability bound to an exponentially small failure probability at a possibly larger constant. Choose the square mesh h_N as in Section 18 and apply Proposition 17.3 at the adjacent torus mesh $h_N/2$. This gives the square Gaussian gauge estimate with high probability. The three-way matching–histogram–square-Gaussian equivalence Theorem 10.4 finishes the proof, since $\sqrt{2/N}\,\sqrt{\log N} = \sqrt{2}s_N$. □

Thus the fixed-positive-probability estimate of Theorem 2.1 suffices for the finite critical matching estimates. Part III shows that this torus criterion cannot hold at the hard endpoint $(2,\infty)$, since its square consequence is false.

19 Consequences and the hard endpoint

We conclude Part II by recording what the reduction gives for the quartic pair and for the hard endpoint, which is the subject of Part III. The reduction of Theorem 10.4 applies to every critical exponent pair, including both hard-coordinate endpoints. The different outcomes for finite and infinite exponents concern the Gaussian minimum-flow estimate itself, not the validity of the matching–Gaussian comparison.

The quartic case. For $\boldsymbol{r} = (4, 4)$, the isotropic comparison Lemma 10.2 identifies the matching statement, up to universal changes of constants, with the square Gaussian estimate

$$\mathbb{P}\left\{ \inf_{\substack{\operatorname{div} J = W^D_{h_N} \\ J\cdot n = 0}} \|J\|_{\exp L^4(D)} \le C\sqrt{\log N} \right\} \longrightarrow 1.$$

Thus Theorem 2.1, combined with Corollary 18.1, resolves the finite critical pairs, including the symmetric quartic case.

The hard endpoint. For $\boldsymbol{r} = (2, \infty)$ the corresponding square gauge is

$$\mathfrak{F}^D_{(2,\infty)}(W^D_{h_N}) = \inf_{\substack{\operatorname{div} J = W^D_{h_N} \\ J\cdot n = 0}} \max\{\|J_1\|_{\exp L^2(D)}, \|J_2\|_\infty\}.$$

The comparison theorem remains applicable, including the L^∞ coordinate: the endpoint KMT correction is of the target order and can be absorbed by changing the fixed constant. However, Theorem 24.1 in Part III shows that, on the square,

$$\frac{\mathfrak{F}^D_{(2,\infty)}(W^D_{2^{-m}})}{\sqrt{m}} \xrightarrow{\mathbb{P}} +\infty.$$

Consequently the hard-endpoint matching assertion is false (Theorem 20.1). The one-way torus-to-square implication of Proposition 17.3 also rules out a fixed-positive-probability torus upper bound at this endpoint.

Dual tests and the cutoff. The complementary Young function to $\psi_r(t) = e^{|t|^r} - 1$ grows, for large arguments, as $q(\log q)^{1/r}$; near zero it grows as $q^{r/(r-1)}$. At $r = 2$, the admissible dual gradient cost is therefore comparable to $|q|\sqrt{\log(2 + |q|)}$, while the associate coordinate for L^∞ is L^1. The lower-bound argument of Part III only needs a weak Orlicz–Hölder inequality for smooth potentials (Lemma 21.1); it does not use an identification of the full Banach dual of an exponential Orlicz space. Such an identification would require care, because ψ_r does not satisfy the Δ_2 condition at infinity.

The dyadic projection is necessary: unregularized planar white noise is almost surely not in H^{-1}, so it has no static L^2 divergence current. The specific matching mesh $h_N \asymp \sqrt{\log N/N}$, however, is fixed by the occupancy and histogram-regularization requirements of this part, not by the ultraviolet divergence alone.

Part III
Failure of the hard endpoint

In this part we prove that at the hard coordinate endpoint $(2,\infty)$, and by symmetry at $(\infty,2)$, the optimal matching cost divided by $\sqrt{\log N/N}$ diverges in probability (Theorem 20.1). The obstruction we show here is qualitative, but a determination of the exact divergence rate of the cost can be found in the GitHub repository [28] associated to this file.

20 Statement, mechanism, and Gaussian reduction

This section states the result, explains the mechanism, and reduces the theorem to a statement about the projected Gaussian source on the square. This part uses its own notation; where an object coincides with one of Part II, we say so.

Let $D=[0,1]^2$, let $(X_i)_{i=1}^N$ and $(Y_i)_{i=1}^N$ be independent uniform samples in D, and write

$$s_N=\sqrt{\frac{\log N}{N}},\qquad \psi_2(t)=e^{t^2}-1.$$

For a permutation π, denote by Δ_j^π the N-vector of jth coordinate displacements $X_i^j-Y_{\pi(i)}^j$, with the normalized Luxemburg norm $\|\cdot\|_{\psi_2,N}$. Put

$$M_N=\min_{\pi\in S_N}\max\left\{\|\Delta_1^\pi\|_{\psi_2,N},\max_i|\Delta_2^\pi(i)|\right\}. \tag{20.1}$$

Theorem 20.1. *For the two independent point clouds,*

$$\frac{M_N}{s_N}\xrightarrow[N\to\infty]{\mathbb{P}}+\infty. \tag{20.2}$$

The assertion holds after exchanging the coordinate axes. In fact, the following one-cloud conclusion is uniform over all *deterministic reference configurations $z^{(N)}=(z_i^{(N)})_{i=1}^N\in D^N$, including the evenly spread configurations of Corollary 16.1. For every fixed $A<\infty$,*

$$\sup_{z^{(N)}\in D^N}\mathbb{P}\left\{\min_{\pi\in S_N}\max\left(\|(X_i^1-z_{\pi(i)}^{(N),1})_i\|_{\psi_2,N},\max_i|X_i^2-z_{\pi(i)}^{(N),2}|\right)\le As_N\right\}\longrightarrow 0. \tag{20.3}$$

Idea of the proof. Part I establishes *upper* Gaussian-flow bounds for finite exponents, which Part II transfers to matching. Here we work in the opposite direction: we build random test potentials of bounded dual cost whose expected pairing with the projected noise is too large for any fixed endpoint bound. Three probability spaces are involved. Brownian motion supplies a model martingale; an auxiliary uniform point in the square samples the gradients of a deterministic tree of triangles; finally, Gaussian white noise selects the signs of the bubbles on that tree. The spatial-branch law holds for *every* assignment of signs, whereas Gaussian freshness produces a positive expectation over those signs. The two conditionings must be kept separate.

Why the endpoint is different. The complementary Young function to $\exp(t^2)-1$ grows like $|v|\sqrt{\log(2+|v|)}$. Thus a dual gradient potential u is charged according to

$$\Phi(a)=\int_0^{|a|}\sqrt{\log(2+r)}\,\mathrm{d}r,\qquad \Psi(a,b)=\Phi(a)+|b|. \tag{20.4}$$

We seek an upper-triangular predictable control for the Brownian martingale $\int_0^1 L_t\,\mathrm{d}B_t$ with bounded terminal Ψ-cost but large

$$\mathbb{E}\int_0^1\sqrt{\det L_t}\,\mathrm{d}t. \tag{20.5}$$

In an upper-triangular parametrization, $Y_t > 0$ defines a change of probability measure, and the ratio $M_t = X_t/Y_t$ is a centered martingale under the Y_t-biased law. A sequence of short local-time stages generates a large accumulated reward while keeping $|M_t|$ small when the logarithmic weight on Y_t is large. The terminal cost therefore remains uniformly bounded even as the reward increases. The obstruction is thus structural; it is not the failure of a particular candidate flow.

Spatially, a determinant-one chart converts $\sqrt{\det L_t}$ into the L^2 size of a bubble. Choosing its sign from fresh Gaussian noise turns this size into a positive expected pairing. Unlike a deterministic test, the adaptive potentials exploit this reward without increasing the convex terminal gradient cost.

Plan of the proof. The argument has four steps: Brownian reward (Lemma 22.1), spatial realization (Lemma 23.3), a weak dual lower bound (Lemma 21.1), and concentration together with the matching transfer (Theorem 24.1). The Brownian and spatial steps are new. From Part II we use the square-flow Lipschitz estimate Lemma 11.4, the triangle inequality of Lemma 11.3, the mesh choice (10.5), and the probability-level comparisons of Proposition 12.1, Theorem 13.8, Lemma 14.1 and Theorem 15.1. The argument takes place directly on the square: the torus-to-square folding of Section 17 goes in one direction only, and it is not reversed here.

The square Gaussian source. Let $\mathcal{W}$ be isonormal white noise on $L^2(D)$, with mean-zero restriction $\mathcal{W}_0$. Let P_h^D be cellwise averaging on the fixed dyadic square grid of mesh $h = 2^{-m}$, and put $W_h = P_h^D \mathcal{W}_0$. Thus W_h is a random *function*, not a continuum white-noise distribution; its pairing with a test u equals $\mathcal{W}(P_h^D u)$ even for a random u (interpreted through the finite cell coefficients). For a mean-zero distribution f, define the square, no-flux minimum-flow gauge

$$F_h(f) = \inf_{\substack{J=(J_1,J_2) \\ -\int_D J\cdot\nabla\varphi = \langle f,\varphi\rangle \ \forall \varphi \in C^\infty(\overline{D})}} \max\{\|J_1\|_{L^{\psi_2}(D)}, \|J_2\|_{L^\infty(D)}\}, \qquad \mathcal{F}_m = F_{2^{-m}}(W_{2^{-m}}). \tag{20.6}$$

The subscript h merely indicates the mesh on which f is supported; admissible currents are continuum fields. Thus F_h is the gauge $\mathfrak{F}^D_{(2,\infty)}$ of (10.9), and W_h is the square source W_h^D of (10.4). The Gaussian coefficient representation is

$$W_h|_Q = h^{-1}(Z_Q - \overline{Z}), \qquad \overline{Z} = h^2 \sum_Q Z_Q, \tag{20.7}$$

with independent standard normal Z_Q, as in (11.2). By Lemma 11.4, the coefficient-space gauge $Z \mapsto \mathcal{F}_m(Z)$ is C_{Lip}-Lipschitz with C_{Lip} independent of m. We shall prove

$$\frac{\mathcal{F}_m}{\sqrt{m}} \xrightarrow[m\to\infty]{\mathbb{P}} +\infty \tag{20.8}$$

for this square source, and then deduce Theorem 20.1 in Section 24 from the eventwise comparisons of Part II.

21 A pathwise dual inequality

The test potentials of Section 23 are chosen from the noise itself. The following inequality bounds their pairing with the source by the flow gauge, pathwise, so that no independence between test and noise is needed.

Lemma 21.1. *There is an absolute constant C_D such that, for every dyadic h, for every mean-zero Lipschitz $u : D \to \mathbb{R}$, and for every realized finite-dimensional source W_h,*

$$|\mathcal{W}(P_h^D u)| \le C_D\, F_h(W_h) \left(1 + \int_D \Psi(\partial_1 u, \partial_2 u)\,\mathrm{d}x\right). \tag{21.1}$$

The assertion holds pathwise even when the test u is selected measurably from the same noise $\mathcal{W}$.

Proof. The complementary Young bound, with an absolute C, is

$$|ab| \le e^{a^2} - 1 + C\Phi(b) \qquad (a, b \in \mathbb{R}). \tag{21.2}$$

Indeed, if $|a| \le 2\sqrt{\log(2+|b|)}$, the term $C\Phi(b)$ pays for $|ab|$. Otherwise $|b| \le e^{a^2/4}$ and $|a| \ge 2\sqrt{\log 2}$, so $|a|e^{a^2/4} \le C(e^{a^2}-1)$; the lower bound on $|a|$ justifies this estimate for small arguments as well.

Fix $r > F_h(W_h)$ and an admissible current with $\|J_1\|_{\psi_2}, \|J_2\|_\infty \le r$. The weak no-flux identity, first for smooth tests and then by Lipschitz approximation, gives

$$|\mathcal{W}(P_h^D u)| = \left| \int_D J \cdot \nabla u \right| \le r \left(1 + C \int_D \Phi(\partial_1 u) + \int_D |\partial_2 u| \right).$$

Here $\int_D (e^{(J_1/r)^2} - 1) \le 1$ because D has unit area. Taking the infimum in r proves the claim. The inequality is deterministic for each current and each test, so it does not use any independence between u and the noise. Since $P_h^D u$ lies in a fixed finite cell space, its pairing with the realized W_h is unambiguous. □

22 Brownian local time and arbitrarily large reward

This section constructs the model martingale: a bounded, predictable, upper-triangular control whose terminal Ψ-cost is bounded by an absolute constant while its determinant reward (20.5) is as large as we wish. It is the only place where the endpoint structure of Ψ enters.

Lemma 22.1. *For every $D_0 > 0$ there are deterministic times $0 = t_0 < t_1 < \cdots < t_K = 1$ and bounded continuous functions*

$$L_j : (\mathbb{R}^2)^{j-1} \longrightarrow \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : a, d > 0 \right\}, \qquad j = 1, \dots, K,$$

such that $L_t = L_j(B_{t_1} - B_{t_0}, \dots, B_{t_{j-1}} - B_{t_{j-2}})$ on $[t_{j-1}, t_j)$ is predictable, takes values in a fixed compact set of upper-triangular matrices with positive diagonals, and

$$\mathbb{E}\,\Psi\left(\int_0^1 L_t \, \mathrm{d}B_t \right) \le C_*, \qquad \mathbb{E} \int_0^1 \sqrt{\det L_t}\, \mathrm{d}t \ge D_0, \tag{22.1}$$

where $C_* < \infty$ is absolute, independent of D_0.

Heuristics. The following formal computation explains why bounded cost and unbounded reward are compatible; it is not used in the proof. In the construction below, the control satisfies $\sigma_t \beta_t = R$ during an active stage. The reward rate $\sqrt{\det \mathcal{L}_t} = Y_t \sqrt{R}$ therefore has conditional mean $\sqrt{R}$, and the reward is $\sqrt{R}$ times the active real duration. Stage j lasts R_j/R in real time, so the reward is of order $\sqrt{R}$ as soon as $\sum_j R_j \ge R$. Each stage contributes a resource of order $q\sqrt{K_q}$, with $q = \gamma/J$ and $K_q \approx 2\log(J/\gamma)$. Summing over J stages gives $\sum_j R_j \approx \gamma\sqrt{2\log(J/\gamma)}$, hence a reward of order $(\log J)^{1/4}$, which is unbounded in J.

The cost stays bounded because the two coordinates are tied together. By stage j, the multiplicative coordinate Y has accumulated logarithmic variance $s \le 2S_j$, which inflates the logarithmic weight in Φ by a factor $\sqrt{1+s} \lesssim \sqrt{S_j}$. Meanwhile the local-time stopping confines the ratio $M = X/Y$ to $|M| \le b_j = S_j^{-1/2}$. The product $|M|\sqrt{1+s}$ is therefore bounded at every stage; this is the content of (22.11). In this way the $\sqrt{\log}$ growth of the complementary Young function is exactly compensated by the confinement $|M| \le S_j^{-1/2}$.

Proof. Step 1: independent operational stages. Each successful stage gains about $q\sqrt{\log(1/q)}$, at the price of a failure probability $O(q)$. With J stages and $q = \gamma/J$, the total resource $\sum_j R_j$ defined below grows like $\sqrt{\log J}$, while the probability of some failure stays small. Fix a small absolute $\gamma > 0$ and an integer $J \gg 1$. Put

$$q = \frac{\gamma}{J}, \qquad K_q = \log(1 + q^{-2}), \qquad S_j = 2^j, \qquad b_j = S_j^{-1/2}, \qquad \ell_j = q b_j,$$
$$w_j(v) = \frac{S_j}{K_q(v + \ell_j^2)}, \qquad 0 \le v \le b_j^2.$$

In stage j let U be a fresh standard real Brownian motion started at 0, and stop it at

$$T_j = \tau_{\ell_j} \wedge \tau_{\{\pm b_j\}} \wedge b_j^2,$$

where τ_ℓ is the first time that Tanaka-normalized local time at 0 reaches ℓ. A stage succeeds when τ_{ℓ_j} is the first stopping event; then $U_{T_j} = 0$. Put

$$R_j = \int_0^{T_j} \sqrt{w_j(v)}\, \mathrm{d}v.$$

Tanaka's formula at the bounded stop yields $\mathbb{E}|U_{T_j}| = \mathbb{E}L^0_{T_j} \le \ell_j$, so $\mathbb{P}(\text{exit at } \pm b_j) \le q$. Lévy's local-time identity and Brownian reflection imply

$$c \min\{1, \ell/\sqrt{v}\} \le \mathbb{P}(\tau_\ell > v) \le C \min\{1, \ell/\sqrt{v}\} \qquad (\ell, v > 0). \tag{22.2}$$

The probability of reaching the time cap without success is therefore at most Cq. For a sufficiently small absolute c_0 and $\ell_j^2 \le v \le c_0 b_j^2$, subtracting the $O(q)$ total failure probability from the left side of (22.2) gives

$$\mathbb{P}(T_j > v, \text{ stage } j \text{ succeeds}) \ge c\frac{\ell_j}{\sqrt{v}}. \tag{22.3}$$

Consequently, by the tail integral formula,

$$\begin{aligned} \mathbb{E}[R_j; \text{success}] &\ge \frac{c\sqrt{S_j}\,\ell_j}{\sqrt{K_q}} \int_{\ell_j^2}^{c_0 b_j^2} \frac{\mathrm{d}v}{v} \ge cq\sqrt{K_q}, \\ 0 \le R_j &\le \frac{\sqrt{S_j}}{\sqrt{K_q}} \int_0^{b_j^2} \frac{\mathrm{d}v}{\sqrt{v + \ell_j^2}} \le \frac{2}{\sqrt{K_q}}. \end{aligned} \tag{22.4}$$

Each operational stage succeeds with probability at least $1 - Cq$. Thus the probability that all J stages succeed is at least $1 - CJq \ge 1 - C\gamma$. Conditioned on this event, the resources remain independent. Their sum S_R satisfies

$$\mu_J := \mathbb{E}[S_R \mid \text{all succeed}] \ge cJq\sqrt{K_q} = c\gamma\sqrt{K_q} \longrightarrow \infty, \qquad \mathrm{Var}(S_R \mid \text{all succeed}) \le \frac{2}{\sqrt{K_q}}\mu_J.$$

Fix any prescribed real-time resource $R > 0$. Once γ is chosen sufficiently small and J is large enough that $\mu_J \ge 2R$, Chebyshev's inequality gives $\mathbb{P}(S_R < R \mid \text{all succeed}) = o_J(1)$. Hence

$$\mathbb{P}\left(\text{all stages succeed and } \sum_{j=1}^{J} R_j \ge R\right) \ge \frac{1}{2}. \tag{22.5}$$

Step 2: embedding in ordinary time. Run the stages successively using the first coordinate B^1 of a planar Brownian motion; each completed successful stage restarts at $M = 0$. Within stage j, transform operational time v into real time by

$$\begin{aligned} \theta_j(v) &= \frac{1}{R}\int_0^v \sqrt{w_j(w)}\, \mathrm{d}w, \qquad v = \theta_j^{-1}(t - t_j^{\mathrm{start}}), \\ \sigma_t &= (\theta_j'(v))^{-1/2}, \qquad \beta_t = \sqrt{w_j(v)}\,(\theta_j'(v))^{-1/2}. \end{aligned} \tag{22.6}$$

Within a stage, $\sigma_t^2\,\mathrm{d}t = \mathrm{d}v$. Consequently the increment of $M_t = \int_0^t \sigma_r\,\mathrm{d}B_r^1$, reparametrized by v, is a standard Brownian motion. The strong Markov property at the stage starting times makes successive successful stages independent. The stopping rules depend only on B^1, and σ, β are predictable in that filtration. Stop at the first failed stage, the end of stage J, or time 1, and put $\sigma = \beta = 0$ afterward. During an active stage,

$$\sigma_t\beta_t = R, \qquad \int_{\text{stage } j} \beta_t^2\,\mathrm{d}t \le \int_0^{b_j^2} w_j(v)\,\mathrm{d}v = \frac{S_j}{K_q}\log(1+q^{-2}) = S_j. \tag{22.7}$$

The real duration of stage j is R_j/R. On the event in (22.5), every stage reached before time 1 succeeds and the total available duration is at least 1. The process is therefore active for almost every real time $t \in [0,1]$ on that event, so

$$\mathbb{E}\int_0^1 \sqrt{\sigma_t\beta_t}\,\mathrm{d}t \ge \frac{1}{2}\sqrt{R}. \tag{22.8}$$

For fixed J, R the coefficients are bounded, though their bounds may depend on J, R.

Step 3: a triangular two-dimensional martingale. Set

$$s_t = \int_0^t \beta_r^2\,\mathrm{d}r, \qquad Y_t = \exp\!\left(\int_0^t \beta_r\,\mathrm{d}B_r^2 - \frac{1}{2}s_t\right), \qquad X_t = Y_t M_t.$$

The clock depends only on B^1, so conditionally on its entire path $\mathbb{E}(Y_t \mid B^1) = 1$. The product rule, with zero quadratic covariation between the B^1 and B^2 integrals, gives

$$\mathrm{d}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \underbrace{\begin{pmatrix} Y_t\sigma_t & M_tY_t\beta_t \\ 0 & Y_t\beta_t \end{pmatrix}}_{\mathcal{L}_t} \mathrm{d}\begin{pmatrix} B_t^1 \\ B_t^2 \end{pmatrix}, \qquad \sqrt{\det \mathcal{L}_t} = Y_t\sqrt{\sigma_t\beta_t}. \tag{22.9}$$

It follows from (22.8) that

$$\mathbb{E}\int_0^1 \sqrt{\det \mathcal{L}_t}\,\mathrm{d}t \ge \frac{1}{2}\sqrt{R}. \tag{22.10}$$

We now bound the terminal cost. If j is the last stage visited, then $|M_1| \le b_j$ and $s_1 \le \sum_{k\le j} S_k \le 2S_j$. Conditionally on B^1 and under weighting by Y_1, the random variable $\log Y_1$ is Gaussian with mean $s_1/2$ and variance s_1. Since $\Phi(x) \le C|x|\sqrt{\log(2+|x|)}$ and $|M_1| \le 1$,

$$\begin{aligned} \mathbb{E}[\Phi(X_1) \mid B^1] &\le C|M_1|\,\mathbb{E}\Big[Y_1\sqrt{\log(2+Y_1|M_1|)} \mid B^1\Big] \\ &\le C|M_1|\sqrt{1+s_1} \le C. \end{aligned} \tag{22.11}$$

Also $\mathbb{E}|Y_1 - 1| \le 2$. The convex cost concerns $(X_1, Y_1 - 1)$, not (X_1, Y_1): its second coordinate is centered. Since $(X_0, Y_0 - 1) = (0,0)$, we have proved

$$\mathbb{E}\Psi(X_1, Y_1 - 1) \le C_* \quad \text{with } C_* \text{ independent of } J, R.$$

Step 4: bounded continuous simple controls. First stop the pair $(X, Y-1)$ at $Y = H$, or at time 1. Conditional Jensen and optional sampling preserve the terminal Ψ bound. Before this stopping time, $Y_t \le H$, $|M_t| \le 1$, and the coefficients σ, β are bounded for fixed J, R; hence the stopped integrand is bounded. As $H \to \infty$, monotone convergence recovers the reward (22.10). Choose R so large that this reward exceeds $D_0 + 2$, and then choose H so that the stopped reward exceeds $D_0 + 1$.

It remains to pass to simple controls with continuous coefficients. Approximate the bounded predictable, upper-triangular nonnegative-diagonal integrand in $L^2(\mathbb{P}\otimes\mathrm{d}t)$ by simple predictable processes on deterministic time intervals. Approximate all finite-dimensional coefficient functions

by bounded continuous functions of finitely many earlier Brownian increments. Refine the partition to include observation times, clip each matrix to a common compact box of upper-triangular matrices with nonnegative diagonals, and add εI. These operations preserve L^2 convergence as $\varepsilon \downarrow 0$. The Itô isometry gives convergence of terminal integrals in L^2, hence convergence of their Ψ costs in L^1, since $\Psi(z) \le C(1+|z|^p)$ for some $p<2$ and the approximants are uniformly square integrable. Continuity of $\sqrt{\det L}$ on the common compact box gives convergence of the rewards in $L^1(\mathbb{P}\otimes \mathrm{d}t)$. Fix one sufficiently accurate approximation. The diagonal lower bound $\varepsilon > 0$, the deterministic partition and the terminal estimate are then as claimed, with an absolute C_* after increasing it once. □

23 Realization by spatially adapted Gaussian bubbles

This section transfers the Brownian model of Lemma 22.1 to test potentials on D (Lemma 23.3). The model is realized without sampling Brownian motion in physical space: the Brownian increments are replaced by bounded innovations, obtained by following the children of a uniform spatial point through a tree of triangles. The family of bubbles is chosen with the help of white noise, but the resulting spatial estimates hold for every fixed realization of the chosen signs.

Reference geometry. Let $T = \{(x_1,x_2) : x_1, x_2 \ge 0,\ x_1+x_2 \le 1\}$. Choose four disjoint small axis-parallel dyadic squares strictly inside T, whose centers form a square, and put a translated tent $(r - \|x-c\|_\infty)_+$ in each, with checkerboard signs. Their sum b is compactly supported in $\operatorname{int} T$, piecewise affine, and satisfies

$$\int_T b(x)\,\ell(x)\,\mathrm{d}x = 0 \quad \text{for every affine } \ell. \tag{23.1}$$

The checkerboard arrangement cancels both the total mass and the first moments; this is what will give Gaussian orthogonality between ancestors and descendants. The law of $\nabla b(x)$ for uniform $x \in T$ is supported on $0, \pm e_1, \pm e_2$ and symmetric under signed coordinate permutations. Choose $c_b > 0$ so that $|T|^{-1}\int_T \nabla b \otimes \nabla b = c_b^2 I$, and denote by ν the law of $c_b^{-1}\nabla b(U)$ for uniform $U \in T$. Then ν is bounded, centered, has covariance I, and is invariant under signed coordinate permutations. Fix a sufficiently fine finite dyadic triangulation into congruent right-isosceles triangles of leg 2^{-k_0} on which b is affine. The integer k_0 and template b are absolute.

Two conditional laws. The next two lemmas serve different purposes. In the first, we condition on earlier Gaussian observations to justify freshness of the next selected bubble. In the second, we freeze all bubble signs and condition on the spatial branch to obtain the i.i.d. innovation law and Jensen's stopping inequality.

Lemma 23.1. *Suppose that a finite sequence of nonzero Hilbert-space vectors is chosen predictably from an isonormal Gaussian process: at step k, the vector b_k is measurable with respect to previously revealed pairings $\mathcal{W}(b_1),\dots,\mathcal{W}(b_{k-1})$, takes one of finitely many values on each finite history, and is orthogonal to every previous vector on that history. Let $\mathcal{G}_{k-1}$ denote that history. Then, conditionally on $\mathcal{G}_{k-1}$, $\mathcal{W}(b_k)$ is centered Gaussian with variance $\|b_k\|_2^2$. For every finite-valued $\mathcal{G}_{k-1}$-measurable Hilbert-space vector v_k,*

$$\mathbb{E}\left[\operatorname{sgn}\mathcal{W}(b_k)\,\mathcal{W}(v_k) \mid \mathcal{G}_{k-1}\right] = \sqrt{\frac{2}{\pi}}\,\frac{\langle v_k, b_k\rangle}{\|b_k\|_2}. \tag{23.2}$$

Here v_k may itself depend on the previous observations and need not be orthogonal to previous directions. The finite-valued assumption is the one used below, and it makes all random isonormal pairings unambiguous.

Proof. We proceed inductively through the finite decision tree. On any branch determined by previously observed pairings, the selected directions are fixed vectors, and the event selecting that branch is measurable with respect to their Gaussian coordinates. Conditional on the actual observed values, the restriction of white noise to the orthogonal complement of their span remains a fresh isonormal Gaussian process. The next direction belongs to this complement, so its normalized pairing is a fresh standard Gaussian. Decompose $v_k = \langle v_k, b_k\rangle b_k/\|b_k\|_2^2 + r_k$. Although $\mathcal{W}(r_k)$ may have a past-measurable component, its remaining centered component is conditionally independent of $\mathcal{W}(b_k)$, and the past-measurable component has zero pairing with its centered sign. Since $\mathbb{E}|G| = \sqrt{2/\pi}$ for $G \sim N(0,1)$, the formula follows. An induction over the finite histories justifies the conditioning. □

Lemma 23.2. *Fix a bounded continuous simple control L as in Lemma 22.1 and a deterministic assignment $Q \mapsto s(Q) \in \{\pm 1\}$ of signs to all potential parents. Let U be uniform on D, independent of an auxiliary sequence of independent ν-distributed vectors, and let $(\mathcal{S}_k)_{k=0}^n$ be a filtration with the following properties.*

(a) Complete congruent subdivisions. *Before microstep k, the active triangle containing U is $\mathcal{S}_{k-1}$-measurable and has the form $Q = x_Q + d_Q A O_Q T$, where A is the current chart, O_Q is a signed coordinate permutation, and x_Q, d_Q, A, O_Q are $\mathcal{S}_{k-1}$-measurable. Its children are the images of the reference triangulation under the same affine map.*

(b) Signs fixed before revelation. *The increment recorded at microstep k is*

$$\xi_k = s(Q)\, O_Q\, c_b^{-1} \nabla b(\widehat{U}), \qquad \widehat{U} = d_Q^{-1} O_Q^T A^{-1}(U - x_Q) \in T,$$

and $\mathcal{S}_k$ is generated by $\mathcal{S}_{k-1}$ and the child of Q containing U.

(c) Predictable retiling. *At the end of a control block, each active triangle is retiled into complete triangles of the form in (a) for the new chart, plus a discarded remainder. Which retained triangle contains U, or whether U lies in the discarded remainder, is $\mathcal{S}$-measurable before the next microstep. After U is discarded, the increments ξ_k are the subsequent auxiliary ν-vectors, which are added to $\mathcal{S}_k$.*

(d) No further information. *Apart from (a)–(c), $\mathcal{S}_k$ carries no information on the position of U inside its current triangle.*

The bubble construction of Lemma 23.3, with all signs frozen after the white noise is fixed, satisfies (a)–(d). Then the ideal increments $\xi_1, \dots, \xi_n$ are independent with law ν. The actual gradient path is the ideal vector martingale stopped at the first discarded region. In particular, for the convex Ψ in (20.4),

$$\mathbb{E}_{U\sim \mathrm{Unif}(D)} \Psi(G_\tau(U)) \le \mathbb{E}_{U,\mathrm{aux}} \Psi(G_n), \tag{23.3}$$

where G_k is the corresponding ideal gradient martingale, τ is the geometric stopping index, and the expectation on the right also includes the independent innovations appended after loss.

Once all signs are frozen, Lemma 23.2 is a statement about the uniform point U and the auxiliary innovations only; the white noise plays no role in it.

Proof. Fix all bubble signs. With U uniform on D, the initial current triangle is one of two congruent halves of D. Conditional on its identity, U is uniform in that triangle. At a microstep, the complete affine child triangulation partitions the parent into congruent equal-area triangles. Thus the normalized reference increment has law ν, irrespective of the parent, its signed permutation O, and the already fixed sign. Conditional on any retained new-chart triangle after a retiling, U is again uniform in that triangle. The portion discarded at a retiling is identified *before* the next increment; thereafter fresh independent artificial ν-increments supply the ideal continuation.

By (b)–(d), the filtration records the active parent, the selected child and its innovation, each retiling/survival decision, and the appended innovations after loss; it does not disclose the exact location of U inside the current triangle. Hence, conditional on $\mathcal{S}_{k-1}$, the point $\widehat{U}$ is uniform in T on survival, and by the invariance of ν under signed coordinate permutations the next ideal increment always has law ν. Iterated conditioning proves the claimed i.i.d. law and the martingale property of G_k. Discarding is a stopping time for this filtration, and conditional Jensen for G_n at the stopped σ-algebra yields (23.3). The geometric decisions may depend on previous innovations through the chosen charts; predictability is sufficient. □

Lemma 23.3. *Fix a simple control L supplied by Lemma 22.1 and put $a_t = \sqrt{\det L_t}$. There are mean-zero random piecewise affine Lipschitz potentials u_m on D, built from $\mathcal{W}$ and $h = 2^{-m}$, such that*

$$\int_D \Psi(\partial_1 u_m, \partial_2 u_m)\,\mathrm{d}x \le \mathbb{E}\Psi\left(\int_0^1 L_t\,\mathrm{d}B_t\right) + o_L(1) \quad \textit{almost surely,} \tag{23.4}$$

$$\liminf_{m\to\infty} \frac{\mathbb{E}\,\mathcal{W}(P_h^D u_m)}{\sqrt{m}} \ge c_*\,\mathbb{E}\int_0^1 a_t\,\mathrm{d}t, \tag{23.5}$$

where $c_* > 0$ is absolute. The error and the minimum admissible m may depend on the fixed control, but c_* does not.

Proof. The potential u_m is a sum of signed bubbles on an adaptive tree of triangles. Steps 1 and 2 build the tree and prove the cost bound (23.4) for every choice of signs; Steps 3 and 4 average over the Gaussian signs and prove the reward bound (23.5).

Step 1: charts and depths. Let $0 = t_0 < \cdots < t_K = 1$ be the fixed control partition and put

$$n = \left\lfloor \frac{m}{4k_0} \right\rfloor, \qquad n_j = \lfloor nt_j \rfloor - \lfloor nt_{j-1} \rfloor.$$

The jth block has n_j microsteps. Along any active branch in that block, the matrix L_j depends only on completed previous block increments. Write $a_j = \sqrt{\det L_j}$ and choose the chart A_j by

$$A_j^{-T} = \frac{L_j}{a_j}, \qquad \det A_j = 1. \tag{23.6}$$

The chart and its inverse have operator norms bounded by a constant depending only on the fixed control. An active triangle can be written $Q = x_Q + d_Q A_j O_Q T$ with O_Q a signed coordinate permutation. On each microstep it is subdivided using the image of the fixed reference triangulation.

At the beginning of each new control block, freeze the old gradient (which is affine on each current triangle) and retile that triangle on a new-chart square lattice of spacing $2^{-\lfloor\sqrt{m}\rfloor} d_Q$. Retain only lattice squares wholly inside the triangle, and split each square into two. Since the old and new charts have uniformly bounded operator norms for the fixed control, the omitted boundary strip has relative area at most $C_L 2^{-\lfloor\sqrt{m}\rfloor}$, uniformly in the parent. The initial two triangles are treated similarly. Because there are only $K+1$ retilings, the total discarded area is at most

$$\varepsilon_m \le C_L(K+1)2^{-\lfloor\sqrt{m}\rfloor} = o_L(1). \tag{23.7}$$

For every active triangle,

$$d_Q \ge 2^{-k_0 n - (K+1)\lfloor\sqrt{m}\rfloor} \ge 2^{-m/3} \tag{23.8}$$

for all sufficiently large m depending on L.

Step 2: bubbles and spatial Brownian limit. For each active parent Q in block j, define, extended by zero off Q,

$$b_Q(x) = \frac{d_Q a_j}{c_b\sqrt{n}}\, b\Big(d_Q^{-1} O_Q^T A_j^{-1}(x - x_Q)\Big). \tag{23.9}$$

Reveal $\mathcal{W}(b_Q)$, select $s_Q = \operatorname{sgn} \mathcal{W}(b_Q)$, and add $s_Q b_Q$ to u_m. The total potential is continuous, piecewise affine, and Lipschitz, because each bubble is supported strictly inside its parent and the tree is finite. Moreover $\int_D b_Q = 0$ by (23.1), so $\int_D u_m = 0$.

On a child of Q, the new gradient increment is exactly

$$\nabla(s_Q b_Q) = \frac{L_j \xi_Q}{\sqrt{n}}, \qquad \xi_Q = s_Q c_b^{-1} O_Q \nabla b. \tag{23.10}$$

At each block end, store the sum of the normalized innovations $\xi_Q/\sqrt{n}$ along the branch. The next block's L_{j+1} is evaluated from the completed earlier block sums, before any new bubble is revealed. The spatial branch lemma Lemma 23.2 applies even though the signs were chosen from $\mathcal{W}$, because we first freeze all those signs. Thus the spatial random point is the only randomness used in the martingale law; the Gaussian noise is averaged only in Step 3.

The ideal block sums converge jointly to independent Brownian increments with covariances $(t_j - t_{j-1})I$ by the multivariate central limit theorem. Since all coefficient maps L_j are bounded and continuous, the ideal terminal gradient converges in law to $\int_0^1 L_t \, \mathrm{d}B_t$. Its increments have magnitude at most $C_L/\sqrt{n}$, so conditional Hoeffding–Azuma bounds give uniformly sub-Gaussian tails for the ideal gradient. The function Ψ has subquadratic growth. Therefore its terminal expectations converge, and the stopped-martingale Jensen bound (23.3) gives (23.4) uniformly over every possible assignment of bubble signs.

Step 3: freshness of Gaussian rewards. Reveal the bubbles breadth first, with a deterministic order within each level. When b_Q is chosen, all previous Gaussian pairings and their signs are already known. Every previous bubble is either disjoint from Q or affine on Q. Hence (23.1) gives

$$\langle b_Q, b_{Q'} \rangle_{L^2(D)} = 0 \quad \text{whenever } b_{Q'} \text{ was revealed earlier.} \tag{23.11}$$

For a fixed m, every past innovation belongs to a finite set, and the simple control sees only finitely many completed block sums. Hence only finitely many chart and bubble choices can arise, and Lemma 23.1 applies to $v_Q = P_h^D b_Q$. Since P_h^D is an orthogonal projection,

$$\mathbb{E}\left[s_Q \mathcal{W}(P_h^D b_Q) \mid \mathcal{G}_{Q^-} \right] = \sqrt{\frac{2}{\pi}} \, \frac{\|P_h^D b_Q\|_2^2}{\|b_Q\|_2}. \tag{23.12}$$

The square-cell Poincaré inequality and affine scaling give

$$\|P_h^D b_Q - b_Q\|_2 \le Ch \|\nabla b_Q\|_2 \le C_L \frac{h}{d_Q} \|b_Q\|_2 = o_L(1) \|b_Q\|_2 \tag{23.13}$$

uniformly in active Q, by (23.8). Writing $|T| = 1/2$, the determinant-one change of variables in (23.9) also shows

$$\|b_Q\|_2 = \frac{\|b\|_{L^2(T)}}{c_b |T|} \, \frac{a_j |Q|}{\sqrt{n}}. \tag{23.14}$$

Thus for all large m, with a universal positive c,

$$\mathbb{E}[s_Q \mathcal{W}(P_h^D b_Q) \mid \mathcal{G}_{Q^-}] \ge \frac{c \, a_j |Q|}{\sqrt{n}}. \tag{23.15}$$

Step 4: sum across the tree. At any microstep, the sum of $a_j|Q|$ over active parents equals the spatial expectation of the coefficient a_j along a surviving branch. By (23.7), replacing surviving by ideal branches changes this expectation by at most $C_L \varepsilon_m$. For each fixed sign realization the ideal law is the i.i.d. law of Lemma 23.2, so the ideal expectation does not depend on the signs. Taking Gaussian expectations, summing (23.15) over the n microsteps, and using the blockwise central limit theorem gives

$$\liminf_{m\to\infty} \frac{\mathbb{E}\mathcal{W}(P_h^D u_m)}{\sqrt{m}} \ge c \lim_{m\to\infty} \sqrt{\frac{n}{m}} \frac{1}{n} \sum_{\text{microsteps}} \mathbb{E}_{\text{ideal}} a_{j(\text{step})} = \frac{c}{\sqrt{4k_0}} \, \mathbb{E} \int_0^1 \sqrt{\det L_t} \, \mathrm{d}t.$$

The final constant depends only on the fixed reference template, not on the control or the prescribed reward. This is (23.5). □

24 Divergence of the Gaussian gauge and failure of matching

We combine the three previous sections to prove the Gaussian divergence (20.8), first in expectation and then in probability, and transfer it to matching with the comparisons of Part II.

Theorem 24.1. *For the projected* square *white noise in* (20.7)*,*

$$\frac{F_{2^{-m}}(W_{2^{-m}})}{\sqrt{m}} \xrightarrow[m\to\infty]{\mathbb{P}} +\infty. \tag{24.1}$$

Proof. Fix an arbitrary $D_0 > 0$ and choose once and for all the simple control of Lemma 22.1. By Lemmas 21.1 and 23.3, for all sufficiently large m,

$$|\mathcal{W}(P_h^D u_m)| \le C_D \mathcal{F}_m (C_* + 2) \quad \text{almost surely,}$$

while

$$\liminf_{m\to\infty} \frac{\mathbb{E}\mathcal{W}(P_h^D u_m)}{\sqrt{m}} \ge c_* D_0.$$

Consequently

$$\liminf_{m\to\infty} \frac{\mathbb{E}\mathcal{F}_m}{\sqrt{m}} \ge \frac{c_* D_0}{C_D(C_* + 2)}. \tag{24.2}$$

The denominator and c_* are independent of D_0. We first send $m \to \infty$ for this fixed control, and then let D_0 range over arbitrarily large fixed values. This proves $\mathbb{E}\mathcal{F}_m/\sqrt{m} \to \infty$.

The mesh-independent Lipschitz estimate Lemma 11.4 and the Gaussian concentration inequality imply, for every $t > 0$,

$$\mathbb{P}\{\mathcal{F}_m \le \mathbb{E}\mathcal{F}_m - t\} \le \exp\left(-\frac{t^2}{2C_{\mathrm{Lip}}^2}\right).$$

For any fixed A, eventually $\mathbb{E}\mathcal{F}_m \ge 2A\sqrt{m}$. Substituting $t = A\sqrt{m}$ gives (24.1). □

The equivalence of Theorem 10.4 alone would only show that no fixed constant bounds M_N/s_N with probability tending to one. To obtain divergence in probability, we transfer the estimates at the level of probabilities.

Proof of Theorem 20.1. Choose the dyadic mesh h_N of (10.5), with $A_0 s_N \le h_N < 2A_0 s_N$ and $m(N) = -\log_2 h_N \asymp \log N$. On the high-probability histogram ellipticity event of Lemma 14.1, the atomic-to-histogram comparison Proposition 12.1 and the endpoint transport–flow theorem Theorem 13.8 give, with constants independent of N,

$$F_{h_N}\big(P_{h_N}^D(\mu_N - \nu_N)\big) \le C(M_N + h_N). \tag{24.3}$$

In the coupling of Theorem 15.1 there is an exact divergence corrector Q_{h_N} with

$$P_{h_N}^D(\mu_N - \nu_N) = \sqrt{\frac{2}{N}}\, W_{h_N} + \operatorname{div} Q_{h_N}, \qquad \max\{\|Q_{h_N,1}\|_{\psi_2}, \|Q_{h_N,2}\|_\infty\} \le C s_N \tag{24.4}$$

on an event of probability $1 - o(1)$ (the finite horizontal coordinate has an even smaller $o_{\mathbb{P}}(s_N)$ correction). The triangle inequality for the minimum-flow gauge (Lemma 11.3) then yields, pathwise on this event,

$$\sqrt{\frac{2}{N}}\, F_{h_N}(W_{h_N}) \le C(M_N + h_N + s_N). \tag{24.5}$$

For each fixed $A < \infty$, the event $\{M_N \le A s_N\}$ therefore implies $\{\mathcal{F}_{m(N)} \le C_A\sqrt{m(N)}\}$ on the good coupling event. By Theorem 24.1,

$$\mathbb{P}(M_N \le A s_N) \le \mathbb{P}\{\mathcal{F}_{m(N)} \le C_A\sqrt{m(N)}\} + o(1) \longrightarrow 0.$$

This proves (20.2).

For the one-cloud statement, fix any deterministic reference $z^{(N)} \in D^N$ and let $p_N(A; z^{(N)})$ be the probability of matching one uniform sample to it with both coordinate costs at most As_N. For two independent samples, the corresponding events are independent and each has probability $p_N(A; z^{(N)})$. Composing the two permutations via their common reference, and using the triangle inequality for the discrete Luxemburg norm and for the supremum, produces a two-sample matching of cost at most $2As_N$. Thus

$$p_N(A; z^{(N)})^2 \leq \mathbb{P}(M_N \leq 2As_N) \longrightarrow 0.$$

The right side is independent of the reference; taking the supremum over $z^{(N)} \in D^N$ proves (20.3). Finally, the symmetry of the sample law under exchange of axes proves the other hard endpoint. □

A Geometric estimates for the adaptive construction

This appendix collects the deterministic geometric estimates behind the adaptive construction of Part I. Throughout, $K > 1$ is the distortion bound between consecutive metrics, $c_0 = c_0(K)$ the inball ratio, s the terminal scale, and θ, D_0, ϑ the Whitney parameters of Section 4. Appendix A.1 compares a cell's sizes in its birth metric and in later metrics; these contraction bounds feed the ancestor-decay factor $\vartheta\sqrt{K}$ used in Sections 6 and 7. Appendix A.2 proves the bounds (4.5) and (4.6) on the ray kernel, which are the input of Lemma 4.2. Appendix A.3 gives the finite Whitney partition with boundary receivers used in Lemma 4.1, fixes the order of the constants, and records the indexing of actual refinements along a revealed ancestral chain.

A.1 Chart contraction and terminal cells

We use the metric convention

$$d_G(x,z) = |G^{-1/2}(x-z)|, \qquad \operatorname{diam}_G E = \sup_{x,z\in E} d_G(x,z).$$

Lemma A.1. *Fix $K > 1$. At an active generation let Q be a square of side ℓ in its birth chart G, where $\det G = 1$. Let G^+ be the predictably selected next metric, with*

$$K^{-1}G \preceq G^+ \preceq KG, \qquad \det G^+ = 1.$$

Let $D = \operatorname{diam}_{G^+} Q$. Suppose that $D > D_0 s$ and that Q is split by the balanced Whitney construction of Lemma 4.1. Write C_A for a uniform bound such that every terminal union created by this split has G^+-diameter at most $C_A s$. Then the following claims hold simultaneously:

$$\sqrt{2/K}\,\ell \le D \le \sqrt{2K}\,\ell, \tag{A.1}$$

$$\ell(S) \le \theta\sqrt{2K}\,\ell \quad \text{for every nonterminal child } S, \tag{A.2}$$

$$\operatorname{diam}_{G^+} S \le (C_A\sqrt{2K}/D_0)\,\ell \quad \text{for every terminal union } S. \tag{A.3}$$

In particular, for any prescribed $\vartheta > 0$ one can choose $\theta > 0$ sufficiently small, and subsequently D_0 sufficiently large, so that (i) every new nonterminal side obeys $\ell(S) \le \vartheta\ell$ and (ii) every newly terminal union obeys $\operatorname{diam}_{G^+} S \le \vartheta\ell$. The choices depend only on the predetermined shape and clipping parameters, not on s, the depth, the parent location, or the eccentricity of G in fixed physical coordinates. A parent declared terminal without splitting has $\operatorname{diam}_G Q \le \sqrt{K} D_0 s$.

Proof. The transformation from birth coordinates to new coordinates is

$$A = (G^+)^{-1/2} G^{1/2}.$$

Since $K^{-1}G \preceq G^+ \preceq KG$, its singular values lie in $[K^{-1/2}, K^{1/2}]$. The diameter of a Euclidean square of side ℓ is $\sqrt{2}\ell$; this proves (A.1). The Whitney upper cap in the new chart is $\ell(S) \le \theta D$, which combined with (A.1) proves (A.2). For a terminal child $\operatorname{diam}_{G^+} S \le C_A s$. Since the parent was split, $D > D_0 s$, and by $D \le \sqrt{2K}\ell$ we have $s < \sqrt{2K}\ell/D_0$. This proves (A.3). Take $\theta < \vartheta/\sqrt{2K}$ and $D_0 > C_A\sqrt{2K}/\vartheta$, together with the other finite lower bounds of Lemma A.4. The constant C_A of the receiver construction is determined before D_0 is chosen and is independent of D_0. Finally, a no-split parent with $D \le D_0 s$ satisfies

$$\operatorname{diam}_G Q = \sqrt{2}\ell \le \sqrt{K} D \le \sqrt{K} D_0 s.$$

□

Lemma A.2. *Suppose the sequence of active ancestors of a cell Q_k is $Q_j \supset Q_{j+1} \supset \cdots \supset Q_k$, with each Q_p a nonterminal square of side ℓ_p in G_p for $j \le p \le k$, and*

$$\ell_{p+1} \le \vartheta \ell_p, \qquad K^{-1} G_p \preceq G_{p+1} \preceq K G_p.$$

Then

$$\frac{\operatorname{diam}_{G_j} Q_k}{\ell_j} \le \sqrt{2}(\vartheta\sqrt{K})^{k-j}. \tag{A.4}$$

If instead Q_k is a terminal union born from the last active parent Q_{k-1}, the same inequality holds with $\sqrt{2}$ replaced by an absolute fixed constant.

A terminal leaf is never refined again, so the lemma concerns only actual refinements.

Proof. By multiplication of the consecutive Loewner comparisons, $G_j^{-1} \preceq K^{k-j} G_k^{-1}$, whence $\operatorname{diam}_{G_j} Q_k \le K^{(k-j)/2} \operatorname{diam}_{G_k} Q_k$. For a square $\operatorname{diam}_{G_k} Q_k = \sqrt{2}\ell_k$ and $\ell_k \le \vartheta^{k-j}\ell_j$. For a terminal union use the terminal version of Lemma A.1 at the last step, giving the same product with a fixed prefactor (Figure 3). Once terminal the cell is frozen. □

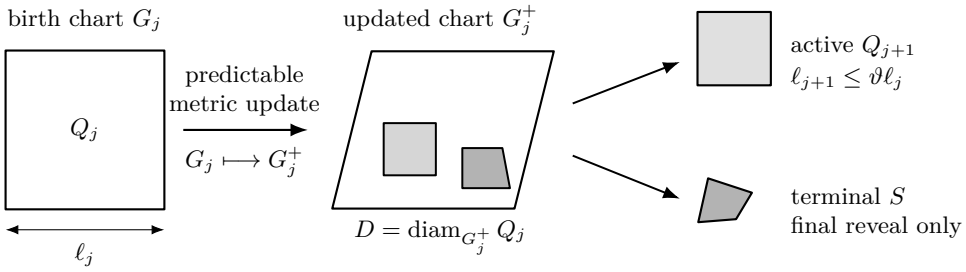


Figure 3: One actual refinement step. The parent is a square in its birth metric G_j but is generally oblique in the updated chart G_j^+. A newly active child receives birth metric $G_{j+1} = G_j^+$; a terminal union receives its last increment and is not refined further. Comparing diameters in earlier birth metrics also incurs the factor $\sqrt{K}$ per actual chart transition.

Finite depth and area floor. All the square sides and diameter contractions in Lemmas A.1 and A.2 are measured in the specified birth charts. A split-created terminal union enters only through the estimate for its final transition; it is not a square in a subsequent chart and is never split again. The factor $\sqrt{K}$ in (A.4) comes in addition to the factor $\sqrt{2K}$ in the newborn-side bound (A.2), and each is counted exactly once. Every active square is born with side at least $2s$, by construction; its physical area is the square of that side because $\det G = 1$. Since $\ell_p \le \vartheta^p \ell_0$ along a chain of active squares, the number of strict refinements is at most

$$J_0 \le 2 + \frac{\log(\ell_0/(2s))}{|\log \vartheta|} = O(\log(1/s)) = O(m).$$

Every terminal union contains an entire uncovered s-square; therefore all cells, at every revealed stage, have area at least s^2.

A.2 Ray kernel estimates

We work on a convex parent in the intrinsic chart of Section 4 and, as in the proof of Lemma 4.2, drop the hats: Q, x, y, K_Q and ω_Q stand for $\widehat{Q}$, z, w, $\widehat{K}_Q$ and $\widehat{\omega}_Q$. Thus $D = \operatorname{diam} Q$, $B(z_0, c_0 D) \subset Q$, and $\omega_Q(z) = D^{-2}\omega((z - z_0)/D)$ is the rescaled bump of (4.2), with $\omega \ge 0$ smooth, of integral one and supported in $B(0, c_0/4)$. Then $\int_Q \omega_Q = 1$ and $\|\nabla^j \omega_Q\|_\infty \le C_j D^{-2-j}$ for $j = 0, 1$. All constants below may depend on c_0 and the fixed bump, but not on D/s, the chart eccentricity, or the random realization.

Lemma A.3. *For $x, y \in Q$, $x \neq y$, the vector kernel*

$$K_Q(x,y) = (x-y)\int_1^\infty t\omega_Q(y+t(x-y))\,dt$$

satisfies

$$|K_Q(x,y)| \le C|x-y|^{-1}, \qquad |\nabla_x K_Q(x,y)| + |\nabla_y K_Q(x,y)| \le C|x-y|^{-2}.$$

It vanishes when $|x-y| > C\,\mathrm{dist}(x,\partial Q)$.

Proof. Write $r = |x-y|$. On the support of the integrand, $z = y + t(x-y) \in \operatorname{supp}\omega_Q \subset Q$, so $tr = |z-y| \le D$. The integral therefore ends by D/r. The bounds on the bump give

$$|K_Q(x,y)| \le CrD^{-2}\int_1^{D/r} t\,dt \le C/r.$$

Differentiating the prefactor contributes at most $CD^{-2}\int_1^{D/r} t\,dt \le C/r^2$. Differentiating the bump in x or y contributes at most $CrD^{-3}\int_1^{D/r} t^2\,dt \le C/r^2$. These are the asserted gradient bounds.

For the support assertion, choose a supporting line P for the convex domain which realizes $\mathrm{dist}(x,\partial Q)$, with nonnegative inward signed distance δ_P. A point z in the bump support satisfies $\delta_P(z) \ge cD$ for fixed $c > 0$, whereas $\delta_P(y) \ge 0$. As $x = (1-t^{-1})y + t^{-1}z$, one gets $\mathrm{dist}(x,\partial Q) = \delta_P(x) \ge cD/t$. Combining with $r \le D/t$ proves the support assertion. □

In the notation of Section 4, Lemma A.3 is exactly (4.5) and (4.6).

A.3 Finite Whitney geometry and actual-reveal indexing

We record the finite partition of a clipped parallelogram and the indexing of actual refinements. The stopped scalar process itself is treated in Proposition 7.4. All assertions about disjointness of cells, pointwise values, and Gaussian evaluations are understood up to Lebesgue-null cell boundaries, with a fixed half-open dyadic convention. Continuous versions of the local Gaussian fields are used on the closures of individual children. The constants below never depend on the physical eccentricity of the old metric.

Lemma A.4. *Let $Q \subset \mathbb{R}^2$ be convex with $B(z_0, c_0 D) \subset Q$, $D = \operatorname{diam} Q$, and let $\mathcal{D}_s$ be a translated, fixed, nested dyadic lattice of s-squares. Fix $A > 8/c_0$ and $\gamma > (A+\sqrt{2})/2 + 2$. For any $\theta \in (0, 1/8)$ there is a constant $C_A = C(c_0, A, \gamma, \theta)$ such that, if $D > D_0 s$ with $D_0 > \max\{2A, 4/\theta\}$, the following algorithm gives a finite partition, modulo null boundaries.*

- *(i) Select the inclusion-maximal dyadic squares $B \subset Q$ whose side $2^j s$ satisfies $2s \le 2^j s \le \theta D$ and $\mathrm{dist}(B,\partial Q) \ge \gamma 2^j s$. Keep them as individual children.*
- *(ii) Among the remaining s-squares, retain as* receivers *those which are wholly contained in Q. All such squares lie within $C_A s$ of ∂Q.*
- *(iii) For every remaining s-square R meeting both the interior and the complement of Q, take the measurable clipped piece $R \cap Q$ and attach it to one receiver within $C_A s$, using deterministic lexicographic tie breaking.*
- *(iv) Each receiver, together with all pieces assigned to it, is a single terminal child. A receiver with no assigned pieces is also a terminal child.*

The selected large squares and resulting terminal children are pairwise disjoint modulo null sets and cover Q modulo null sets. Every terminal child S satisfies

$$s^2 \le |S| \le C_A s^2, \qquad \operatorname{diam} S \le C_A s, \qquad \sup_{z\in S} \mathrm{dist}(z,\partial Q) \le C_A s.$$

If B is a selected large square, its side ℓ_B satisfies $\ell_B \le \theta D$ and $\gamma\ell_B \le \operatorname{dist}(z, \partial Q) \le C_A\ell_B$ for $z \in B$. In particular the partition has the pointwise Whitney comparisons needed for the projected kernel in Lemma 4.2.

Proof. Any s-square meeting Q is either contained in a selected large square, is a full uncovered square, or has nonempty intersection with Q and with its complement. Half-open boundaries make these three assignments disjoint modulo null sets. Inclusion-maximal squares in the same dyadic lattice are disjoint or nested, so distinct selected squares are disjoint. The dyadic hierarchy is finite because its sides are between s and θD.

If a full uncovered s-square R had an admissible $2s$ ancestor, that ancestor would lie in some selected square; this contradicts uncoveredness. Since $2s < \theta D$, the $2s$ ancestor fails either containment in Q or the boundary-distance inequality. In either case its points, and therefore all of R, are within $(2\gamma + 3\sqrt{2})s$ of ∂Q. Thus there are no uncovered squares far from the boundary, regardless of the orientation or translation of the lattice.

For a clipped R, take $z \in R \cap Q$, so $\operatorname{dist}(z, \partial Q) \le \sqrt{2}s$. Set $z_t = (1-t)z + tz_0$, $t = As/D < 1/2$. By the inball and convexity, and then by the Lipschitz property of boundary distance,

$$Ac_0 s \le \operatorname{dist}(z_t, \partial Q) \le (A + \sqrt{2})s, \qquad |z_t - z| \le As.$$

Since $Ac_0 > 8$, the unique half-open s-square R' with $z_t \in R'$ is contained in Q. Were R' covered by a selected dyadic square of side $L \ge 2s$, then $\operatorname{dist}(z_t, \partial Q) \ge \gamma L \ge 2\gamma s$, contrary to $2\gamma > A + \sqrt{2}$. Thus R' is uncovered and is an eligible receiver. Its distance from R is at most $(A + 2\sqrt{2})s$. This construction works at an acute corner even when R meets two boundary sides; no side is chosen, only the inball center.

Every piece assigned to a given receiver originates in one of the at most $C(A)$ lattice squares in a fixed-radius $O_A(s)$ neighborhood. The receiver supplies area at least s^2, and the union of at most $C(A)$ clipped pieces has area at most $C_A s^2$ and diameter at most $C_A s$. The boundary-distance assertion follows either from the uncovered-full-square estimate or the clipped-square estimate.

Finally, for a selected B of side ℓ, its parent $B^{\uparrow}$ of side 2ℓ is not admissible. If it fails containment, the distance to ∂Q is at most $2\sqrt{2}\ell$ on B; if it fails clearance, the upper bound is $(2\gamma + 2\sqrt{2})\ell$; if it fails the cap, then $D < 2\ell/\theta$ and every point of Q is within D of the boundary. The lower bound on B is its admissibility. These are the stated uniform comparisons. The proof is independent of the relative translation of the lattice and of all angles except through the fixed inball ratio c_0. □

Choice of constants. Fix the clipping constant $K > 1$. A square in its birth metric transformed to the updated chart has singular values between $K^{-1/2}$ and $K^{1/2}$; thus it has an inball ratio at least $c_0 = c_0(K) > 0$. Fix $A > 8/c_0$, then $\gamma > (A + \sqrt{2})/2 + 2$. These determine the terminal union constant C_A, independently of s and of D_0. Fix the contraction target ϑ; then choose $\theta < \vartheta/\sqrt{2K}$ and

$$D_0 > \max\{2A, 4/\theta, C_A\sqrt{2K}/\vartheta\}.$$

Increasing D_0 never changes C_A or the selected-square comparisons. Consequently Lemma A.1 applies uniformly to the exact finite partition, and by reducing ϑ we enforce $\vartheta\sqrt{K} < 1$ and subsequently $(\vartheta\sqrt{K})^{2\beta}(1-\eta)^{-2a_*} < \rho_1$.

Lemma A.5. *Let $Q_0 \supset Q_1 \supset \cdots \supset Q_b$ be the cells of the* actual *revealed ancestral chain of a point, where Q_j is born with metric G_j, and each revealed increment V_j is routed in G_j. Suppose every strict transition from Q_{p-1} to Q_p satisfies the comparison and side/diameter bounds of Lemma A.1. If Q_b is a square or a newly terminal union, then for every $1 \le j \le b$,*

$$\frac{\operatorname{diam}_{G_j}(Q_b)}{\ell_j} \le C(\vartheta\sqrt{K})^{b-j},$$

where ℓ_j is the side of Q_j in its birth metric when Q_j is a square; for a newly terminal Q_j its last intrinsic diameter serves as the length at j. If Q_b is stopped directly, without a further reveal, its G_b-diameter is at most $\sqrt{K}D_0 s$. At padded times $k > b$, $U_k = U_b$, $G_k = G_b$ and $T_k = T_b$ on this leaf. No new factor $(\vartheta\sqrt{K})^{k-b}$ is assigned to its geometry, and no new Gaussian envelope is charged.

Proof. For a chain of square children, the intrinsic birth-side contractions multiply to $\ell_b \le \vartheta^{b-j}\ell_j$. As in the proof of Lemma A.2, the Loewner comparisons imply $G_j^{-1} \preceq K^{b-j}G_b^{-1}$ without any commutativity assumption; this gives the asserted ancestor diameter bound. If the final child is a newly terminal union, apply the terminal-diameter estimate only on the final transition and use the ordinary side contractions earlier. If the parent stops without splitting, its new-chart diameter is $\le D_0 s$, while the old/new metric comparison gives $\operatorname{diam}_{G_b} Q_b \le \sqrt{K}D_0 s$. Frozen padded times have no strict transitions. □

Stopping convention. Lemma A.5 counts only actual refinements. Once a leaf is terminal, its field and memory are frozen; only the decreasing clock continues to act on its exponential integrand. At a failed fresh-envelope test, the numerical modular is killed. These are the conventions of the stopped-supermartingale argument in Proposition 7.4. On successful completion, the terminal partition and success event are measurable with respect to the revealed history, so the unrevealed source is the conditional Gaussian bridge of Lemma 3.1.

Use of AI tools

The starting point of this work was the previous project [8] on the finite p cost case. Intense interactions with LLMs spanning several weeks led to a first proof of the Ultimate matching conjecture in the subcritical case, leaving the critical case (as well a the endpoint cases) seemingly out of reach. Somewhat unexpectedly to us, a solution for the critical finite exponent case was finally found during another round of exploration, see the 'monolithic' file `finite-exponents.pdf` in the repository. After this GPT-6 quickly obtained the exact asymptotic for the endpoint cases, see the 'monolithic' file `endpoints.pdf` in the same repository. The content of this file was mostly generated and proof-checked by the LLMs during interactions with us aimed at understanding the 'monolithic' versions. We have an overall understanding of the proofs but plan to provide a fully digested version shortly.

References

[1] M. Ajtai, J. Komlós and G. Tusnády, *On optimal matchings*, Combinatorica **4** (1984), 259–264.

[2] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhäuser, 2008.

[3] L. Ambrosio and F. Glaudo, *Finer estimates on the 2-dimensional matching problem*, J. Éc. polytech. Math. **6** (2019), 737–765.

[4] L. Ambrosio, F. Glaudo and D. Trevisan, *On the optimal map in the 2-dimensional random matching problem*, Discrete Contin. Dyn. Syst. **39** (2019), 7291–7308.

[5] L. Ambrosio, M. Goldman and D. Trevisan, *On the quadratic random matching problem in two-dimensional domains*, Electron. J. Probab. **27** (2022), paper no. 54.

[6] L. Ambrosio, F. Stra and D. Trevisan, *A PDE approach to a 2-dimensional matching problem*, Probab. Theory Related Fields **173** (2019), 433–477.

[7] L. Ambrosio, F. Vitillaro and D. Trevisan, *Sharp PDE estimates for random two-dimensional bipartite matching with power cost function*, Rend. Lincei Mat. Appl. **35** (2024), 323–342.

[8] V. Armegioiu, M. Goldman, F. Grotto and D. Trevisan, *Exact asymptotics for the 2D Euclidean random matching problem*, arXiv:2609.08717 (2026).

[9] B. Bollobás and I. Leader, *Edge-isoperimetric inequalities in the grid*, Combinatorica **11** (1991), 299–314.

[10] J. Bretagnolle and P. Massart, *Hungarian constructions from the nonasymptotic viewpoint*, Ann. Probab. **17** (1989), 239–256.

[11] S. Caracciolo, C. Lucibello, G. Parisi and G. Sicuro, *Scaling hypothesis for the Euclidean bipartite matching problem*, Phys. Rev. E **90** (2014), 012118.

[12] X. Cheng, F. Otto, M. Palmieri and A. Wachtel, *A geometrically linear approximation of min–max optimal matching*, arXiv:2609.08680 (2026).

[13] R. M. Dudley, *Uniform Central Limit Theorems*, 2nd ed., Cambridge University Press, 2014.

[14] D. Gale, *A theorem on flows in networks*, Pacific J. Math. **7** (1957), 1073–1082.

[15] M. Goldman, M. Huesmann and F. Otto, *Quantitative linearization results for the Monge–Ampère equation*, Comm. Pure Appl. Math. **74** (2021), 2483–2560.

[16] M. Goldman, M. Huesmann and F. Otto, *Almost sharp rates of convergence for the average cost and displacement in the optimal matching problem*, preprint, 2023; Abel Symposium proceedings.

[17] M. Goldman and F. Otto, *A variational proof of partial regularity for optimal transportation maps*, Ann. Sci. Éc. Norm. Supér. **53** (2020), 1209–1233.

[18] M. Goldman and D. Trevisan, *Optimal transport methods for combinatorial optimization over two random point sets*, Probab. Theory Related Fields **188** (2024), 1315–1384.

[19] A. J. Hoffman, *Some recent applications of the theory of linear inequalities to extremal combinatorial analysis*, in: Combinatorial Analysis, Proc. Sympos. Appl. Math. **10**, Amer. Math. Soc., Providence, 1960, 113–127.

[20] J. Komlós, P. Major and G. Tusnády, *An approximation of partial sums of independent RV's and the sample DF. I*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **32** (1975), 111–131.

[21] F. Ouimet, *A solution to Problem 5.3 in Talagrand's My Favorite Problems: simultaneous sum and bottleneck bounds for random planar matchings*, preprint (2026), doi:10.13140/RG.2.2.26527.93606/2.

[22] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Birkhäuser, 2015.

[23] M. Talagrand, *Matching random samples in many dimensions*, Ann. Appl. Probab. **2** (1992), 846–856.

[24] M. Talagrand, *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*, Springer, 2005.

[25] M. Talagrand, *Upper and Lower Bounds for Stochastic Processes: Decomposition Theorems*, 2nd ed., Springer, 2021.

[26] M. Talagrand, *My favorite problems*, manuscript, 2026.

[27] C. Villani, *Optimal Transport: Old and New*, Springer, 2009.

[28] `https://github.com/grottofrancesco/UltimateMatchingTheorem`.