

Aging in Spherical Spin Glass Dynamics

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

With the exception of this note, this entire document is machine written. We have taken care to supervise this writing to ensure that the proofs meet a minimum standard of readability and that previous literature is appropriately cited. However, the resulting document remains lacking in exposition and overall coherence. We are nonetheless releasing this manuscript on Hexagon to disseminate the result as quickly as possible, because we believe it may be of interest to the mathematics and physics communities. We welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

We thank Luca Leuzzi for many helpful comments on the theory of spherical spin glasses and his encouragement to explore their dynamics rigorously.

P.L. was partially supported by NSF grant DMS-2450004.

ABSTRACT

We study Langevin dynamics for the pure spherical p -spin glass with $p \geq 3$, with covariance Nq^p and generator $\Delta_S + \beta \nabla_S H_N \cdot \nabla_S$, started from a uniform point independent of the disorder. We let the dimension tend to infinity on fixed time horizons and study the long-time behavior of the resulting hard-spherical Cugliandolo–Kurchan equations for the correlation C and the response R . At every positive temperature, their solution has a unique stationary limit at bounded time differences, which satisfies the fluctuation–dissipation theorem (FDT), and its energy converges. For $\beta \geq \beta_d$, where β_d is the dynamical transition, we show that the plateau q of the stationary correlation is the larger solution of the marginality relation $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$, and we compute the limiting energy. Correlations with every fixed earlier state tend to zero. The limits of large waiting time and large time difference therefore do not commute, which proves aging of the correlation, including at the transition point.

We also prove that the response at a late time, as a function of the time difference, converges to its stationary profile in L^2 and uniformly over all earlier times; that FDT in the later time holds uniformly over all earlier times; and that the integral of the squared violation of FDT in the earlier time tends to zero, while for $\beta > \beta_d$ the positive part $(\partial_s C - R)_+$ of this violation keeps a positive integrated mass. For $\beta \geq \beta_d$, the stationary correlation approaches its plateau as a power law, with the exponent predicted by mode-coupling theory and a finite positive amplitude. For the cubic model and $\beta \geq \beta_d$, including the transition point, we prove a lower bound on the total susceptibility and a product inequality for correlations at three times; for $\beta > \beta_d$, we prove, for each fixed level below the plateau, an explicit asymptotic lower bound on the first passage time to that level multiplied by the entropy production during the passage. The aging proofs for the cubic and quartic models represent the correlation by vectors in a Hilbert space indexed by trees, and general pure interactions are treated with averages over finitely many independent futures and established algorithmic threshold bounds. The full fluctuation–dissipation relation of the aging regime and the time scale of aging remain open.

Contents

1	Introduction	1
1.1	Spin glass dynamics and aging	1
1.2	The model and the Cugliandolo–Kurchan equations.	2
1.3	The physical picture	5
1.4	Main results	9
1.5	Relation to previous work	17
1.6	Scope and open problems.	24
1.7	Organization of the paper.	25
2	Overview of the proofs	27
2.1	The difficulty	27
2.2	From the diffusion to stationary clusters	28
2.3	The scalar equation and its constraints	30
2.4	The causal tree realization and reflection	31
2.5	Selection of the plateau from two threshold inputs	33
2.6	Why the threshold inputs hold	35
2.7	Response localization and persistent violation.	39
2.8	Three-time geometry, susceptibility, and the Fisher cost	41
2.9	The stationary power law	43
2.10	Dependencies between the sections	44
PART I	The Cugliandolo–Kurchan system and its stationary limits	46
3	The model and the hard-spherical Cugliandolo–Kurchan equations	47
3.1	The Langevin dynamics.	47
3.2	The hard-spherical CK system	50
3.3	Comparison of conventions	52
3.4	The bridge on fixed time horizons.	53
3.5	The response as a Brownian primitive.	58
3.6	Landscape bounds	60
4	Structural identities and uniform estimates	63
4.1	The energy identity.	64
4.2	The Green covariance identity.	68
4.3	The energy–response tradeoff and the multiplier bound	72

4.4	Regularity in time	73
4.5	The backward response equation	74
4.6	Response domination and compactness	76
5	Entropy production and stationary limits	79
5.1	Relative entropy and an exact fluctuation–dissipation identity	80
5.2	An integrable Fisher majorant.	84
5.3	Flatness in the waiting time.	90
5.4	Stationary translated clusters	92
PART II	Plateau selection, forgetting, and correlation aging	96
6	The scalar equation of a stationary cluster	97
7	The causal tree realization and reflection	102
7.1	The tree Hilbert space	103
7.2	The exact causal realization	105
7.3	The evolution of a projected state	108
7.4	Reflection and complete monotonicity.	111
8	Selection of the plateau and correlation aging	119
8.1	Prefix centers and the threshold inputs	120
8.2	Marginality of positive plateaus.	122
8.3	Convergence at every fixed temperature.	125
8.4	The dynamical transition and the selected energy	127
8.5	Forgetting of fixed waiting times	130
8.6	Correlation aging.	131
9	Tree ceilings and the threshold inputs for cubic and quartic models	133
9.1	The cubic tree ceiling	134
9.2	The quartic tree ceiling	137
9.3	The energy ceiling for prefix centers.	140
9.4	Threshold-saturating prefix centers	141
9.5	Forgetting at marginal saturation	142
10	The threshold inputs for general pure models	144
10.1	The algorithmic threshold and its overlap penalty	146
10.2	Finite forks and the exact energy subtraction	152
10.3	Adaptive disclosure and orthogonality to early queries	162
10.4	Transfer to a fixed waiting-time state	166
10.5	Proof of Proposition 8.1.	171

PART III	Response and the fluctuation–dissipation relation	172
11	Localization of the response and of correlation derivatives	173
11.1	The full response row	175
11.2	The later correlation derivative	178
11.3	The earlier derivative and the squared FDT defect	181
11.4	The direct thermal covariance	185
11.5	The remote response for general pure models	187
12	Persistent integrated violation of the fluctuation–dissipation relation	193
12.1	A weighted slow ratio	194
12.2	Transport of the response budget	196
12.3	The persistent defect	197
PART IV	Three-time geometry in the cubic model	200
13	Saturated tree vectors and the product inequality	201
13.1	The universal operator at a saturated root	202
13.2	Positive saturated triples	208
14	Three-time correlations and the susceptibility	213
14.1	From causal centers to actual correlations	214
14.2	The susceptibility lower bound	220
15	A Fisher cost for relaxation below the plateau	227
PART V	The stationary approach to the plateau	235
16	Algebraic relaxation in the stationary sector	236
16.1	Inputs and statement of the result.	237
16.2	Normalization and comparison	239
16.3	Comparison profiles at the critical exponent.	243
16.4	The quadratic balance and its logarithmic Dirichlet form	252
16.5	An integrable-source renewal equation	256
17	Further questions	264
17.1	The complete slow fluctuation–dissipation relation	264
17.2	The sharp susceptibility upper bound	265
17.3	A nondegenerate clock and physical relaxation rates.	266
17.4	Selection below the dynamical transition	266
17.5	Other orders of limits.	267
	References	268

Introduction

1.1 Spin glass dynamics and aging

The static theory of mean-field spin glasses describes the Gibbs measure of a random Hamiltonian in high dimension. For the spherical p -spin model studied in this paper, the limiting free energy is given by the Crisanti–Sommers formula, a variational problem over probability measures on $[0, 1]$ whose minimizer is the Parisi measure [46]. This formula was proved by Talagrand for even models [120] and by Chen in general [42]. For the pure model, the Parisi measure is a point mass at zero at high temperature, and it has one step of replica symmetry breaking at every temperature below the static transition [6, Section 5]; the two-point support property follows from [120, Proposition 2.2]. These results concern the equilibrium measure. In this paper, we study the dynamics.

The natural dynamics is the Langevin diffusion on the sphere, in which the configuration moves in the direction of increasing Hamiltonian, perturbed by Brownian noise; the inverse temperature sets the strength of the drift relative to the noise. This diffusion is reversible with respect to the Gibbs measure, and for each fixed dimension N , it converges to the Gibbs measure as time tends to infinity. At sufficiently low temperature, however, the spectral gap of its generator is exponentially small in N , with probability tending to one over the disorder [77], so the slowest relaxation modes have relaxation times that grow exponentially with N . On time scales that do not grow with N , the dynamics can stay far from equilibrium. In this paper, we take the limit $N \rightarrow \infty$ first, on every fixed time interval, and only then study the long-time behavior of the limiting dynamics. In this order of limits, the Gibbs measure and its Parisi measure need not describe the behavior at long times. The physical theory predicts instead that the system ages.

Aging is observed in disordered materials that do not reach equilibrium on any accessible time scale. After a rapid cooling to low temperature, called a quench, the response of such a material to a small perturbation depends on the time that has elapsed since the quench, and it becomes slower as the sample gets older. In spin glasses, Lundgren, Svedlindh, Nordblad, and Beckman found that magnetic relaxation in CuMn depends on the time spent at low temperature before the measurement [109]; the experimental picture is reviewed in [121]. In a trap model, the system jumps between traps and spends a random time in each of them. Bouchaud introduced weak ergodicity breaking through

a trap model whose mean trapping time diverges: the system keeps aging, although the phase space does not split into mutually inaccessible regions [33, p. 1707]. Here the correlation measures the overlap of the configurations at two times, and the response measures the linear effect of a small force applied at the earlier time; we define both precisely in Section 1.2. In the formulation used for mean-field aging, the correlation with every fixed earlier configuration eventually vanishes, while relaxation becomes slower as the system gets older [52, arXiv version 1, Section 3].

Mean-field spin glasses with p -spin interactions are a setting in which aging can be derived from a microscopic dynamics. Derrida introduced the random energy model, in which the energies of different configurations are independent, as a limit of a family of p -spin models [62], and the Ising p -spin models themselves were studied by Gross and Mézard [85] and by Gardner [74]. Sompolinsky and Zippelius developed the dynamical mean-field theory of spin glass relaxation [115, 116]. This theory replaces the interaction of one coordinate with all the others by a random force and a memory of its own past, both determined self-consistently, and it describes the dynamics through closed equations for the correlation and response functions. Kirkpatrick and Thirumalai connected the dynamics of p -spin models with the glass transition of supercooled liquids, which are liquids cooled below their freezing temperature without crystallizing [97, 98]. For $p \geq 3$, the dynamical equations resemble those of the mode-coupling theory of supercooled liquids, a self-consistent theory in which correlation functions satisfy nonlinear equations with memory and, at low temperature, develop a plateau, that is, a positive limit at large time differences (see [83] for a review). In these equations, the correlation of the equilibrium dynamics stops decaying to zero below a dynamical transition temperature. This temperature is higher than the static transition temperature, at which the Parisi measure changes. Kirkpatrick, Thirumalai, and Wolynes built on this correspondence a scaling theory of glasses in finite dimension, in which relaxation below the dynamical transition proceeds by thermally activated rearrangements, where the noise carries the system across barriers of the energy landscape [99]. The spherical p -spin model was solved at equilibrium by Crisanti and Sommers [46]. Crisanti, Horner, and Sommers studied its dynamics in [48], including the dynamical transition above the static one and the power-law relaxation near it. Cugliandolo and Kurchan constructed an asymptotic solution of the nonequilibrium equations at long times [51]. We describe their predictions in Section 1.3, after we introduce the model and the equations that they analyzed.

1.2 The model and the Cugliandolo–Kurchan equations

We fix an integer $p \geq 3$ and $\beta > 0$, and write $\xi(q) = q^p$. For $N \geq 2$, let $S_N = \{x \in \mathbb{R}^N : \|x\|^2 = N\}$. The Hamiltonian H_N is a centered Gaussian process on $\mathbb{R}^N$ with covariance

$$\mathbb{E}H_N(x)H_N(y) = N\xi\left(\frac{x \cdot y}{N}\right),$$

realized as in (3.1). This is the pure spherical p -spin model in the normalization of the rigorous static theory. Its Gibbs measure at inverse temperature β has density proportional to $e^{\beta H_N}$ with respect to the uniform probability measure on S_N . Let X_t be the Langevin dynamics on S_N with generator

$$\Delta_S + \beta \nabla_S H_N \cdot \nabla_S,$$

where Δ_S and ∇_S are the Laplace–Beltrami operator and gradient on S_N ; equivalently, X_t solves (3.3), whose noise has amplitude $\sqrt{2}$. The Gibbs measure is reversible for this diffusion. The initial point X_0 is uniform on S_N and independent of the disorder and of the Brownian motion. Since the uniform measure is the Gibbs measure at $\beta = 0$, this initial condition models a quench from infinite temperature. The drift points in the direction of increasing H_N , so the limiting energies and the threshold energy in this paper are positive; physics works usually minimize the Hamiltonian, which reverses these signs.

Since the uniform initial law is not invariant for the dynamics, its natural observables depend on two times. The correlation

$$C_N(t, s) = \frac{X_t \cdot X_s}{N}$$

is the overlap of the configurations at times s and t . The response describes how the configuration at a time t reacts to a small force applied at an earlier time s . Precisely, for a deterministic locally square-integrable force $h(u)$ projected onto the tangent space of S_N , the response is the zero-initial strong solution V_t of the forced linearized equation along X , with the drift and noise derivatives evaluated on X . It satisfies $V_t = \int_0^t U_N(t, u) h(u) du$ for the random response matrix $U_N(t, u)$ (Section 3.1). We write

$$\bar{r}_N(t, s) = \frac{1}{N} \mathbb{E} \operatorname{tr} U_N(t, s), \quad 0 \leq s \leq t,$$

for its normalized trace, averaged over the disorder, the initial point, and the noise. Thus $\bar{r}_N(t, s)$ is the mean linear response of a coordinate at time t to a unit impulse of force on the same coordinate at time s , that is, to a force concentrated near time s with unit time integral. We measure the response to a force added to the drift without a factor β . The finite-dimensional response has diagonal value $\bar{r}_N(t, t) = \kappa_N = (N - 1)/N$, because the force is projected onto the $(N - 1)$ -dimensional tangent space, and its limit below has diagonal value one (Remark 3.1).

As $N \rightarrow \infty$, these functions have deterministic limits on fixed time horizons. For every fixed $0 \leq T < \infty$, the correlation $C_N(t, s)$ converges in probability to a function $C(t, s)$, uniformly on $[0, T]^2$, and the averaged response $\bar{r}_N(t, s)$ converges to a function $R(t, s)$,

uniformly on $0 \leq s \leq t \leq T$. Together with a function $\lambda(t)$, which keeps $C(t, t) = 1$, the limits form the unique solution of the system of integro-differential equations

$$\begin{aligned}\partial_t R(t, s) &= -\lambda(t)R(t, s) + \beta^2 \int_s^t \zeta''(C(t, u))R(t, u)R(u, s) \, du, \\ \partial_t C(t, s) &= -\lambda(t)C(t, s) + \beta^2 \int_0^t \zeta''(C(t, u))R(t, u)C(u, s) \, du \\ &\quad + \beta^2 \int_0^s \zeta'(C(t, u))R(s, u) \, du, \\ \lambda(t) &= 1 + \beta^2 \int_0^t [\zeta'(C(t, u)) + C(t, u)\zeta''(C(t, u))]R(t, u) \, du,\end{aligned}$$

for $t \geq s \geq 0$, with $C(t, t) = R(t, t) = 1$, where C is extended to $s > t$ by symmetry. We call these the Cugliandolo–Kurchan (CK) equations; they are stated with their conventions in (3.9)–(3.12). Existence and uniqueness of the solution were proved by Dembo, Guionnet, and Mazza [60, Proposition 1.1], and the solution satisfies $0 \leq C \leq 1$ and $R \geq 0$ (see Proposition 3.2). Sellke proved fixed-horizon convergence of C_N and of the Brownian observable representing the integrated response [112, Theorem 4.3]. The force-response identification in Lemma 3.4 equates the annealed mean of its rescaled version $\Gamma_N(t, s) = X_t \cdot B_s / (\sqrt{2}N)$ with $\int_0^s \bar{r}_N(t, u) \, du$, where B drives the diffusion. He built on the theorem of Ben Arous, Dembo, and Guionnet for the soft-spherical dynamics [24, Theorem 1.2], in which the constraint $\|X_t\|^2 = N$ is replaced by a confining potential. We state these results in our normalization in Proposition 3.3 and Lemma 3.4, and we deduce the uniform convergence of $\bar{r}_N$ in Remark 3.5. Throughout the paper, the *CK solution* is this solution (C, R, λ) , which describes the Langevin dynamics started from the uniform point.

The CK equations involve only C , R , and λ ; the disorder no longer appears. The function λ is a Lagrange multiplier for the spherical constraint, and the integrals are memory terms. In the physical derivation through dynamical mean-field theory, each coordinate of X_t behaves in the limit like the solution x_t of the linear equation (see [49, arXiv version 2, Section 9.5])

$$dx_t = \left(-\lambda(t)x_t + \beta^2 \int_0^t \zeta''(C(t, u))R(t, u)x_u \, du + \eta_t \right) dt + \sqrt{2} \, db_t.$$

Here x_0 , b , and η are independent; x_0 is a standard Gaussian variable, b is a standard Brownian motion, and η is a centered Gaussian process with covariance $\mathbb{E}\eta_t\eta_u = \beta^2\zeta'(C(t, u))$. The interaction of one coordinate with all the others produces the memory kernel $\beta^2\zeta''(C(t, u))R(t, u)$, which weights the effect of the past values of x on its present drift, and the noise η , which is called colored because its values at different times are correlated. The CK equations express the self-consistency conditions that $C(t, s) = \mathbb{E}x_t x_s$, that $R(t, s)$ is the response of x_t to a force applied at time s , and that $\mathbb{E}x_t^2 = 1$, which determines λ . We use this effective process only as a guide. Its variation-of-constants formula writes

$C(t, s)$ as the sum of the contributions of the initial condition, of the Brownian noise, and of the colored noise, each propagated by the response. This is the Green covariance identity of Proposition 4.2, which we prove directly from the CK equations.

We write

$$e_\beta(t) = \beta p \int_0^t C(t, u)^{p-1} R(t, u) \, du, \quad \chi(t, s) = \int_s^t R(t, u) \, du.$$

The first quantity is the limit of the mean energy $N^{-1} \mathbb{E} H_N(X_t)$, and $\lambda(t) = 1 + p\beta e_\beta(t)$ (Proposition 4.1). The second is the integrated response, or susceptibility. It is the response at time t to a constant unit force switched on at time s . A field h added to the Hamiltonian adds the force βh to the drift, so the magnetization per unit field is $\beta \chi(t, s)$ in our conventions. It is the analog of the magnetization measured in aging experiments after a small field is switched on at time s , the zero-field-cooled protocol [121].

We also use the entropy production of the dynamics. Fix the disorder, and let ρ_t be the law of X_t . Since the dynamics is reversible, the relative entropy of ρ_t with respect to the Gibbs measure decreases in time. Its rate of decrease is the relative Fisher information, the integral against ρ_t of the squared norm of the gradient of the logarithm of the density of ρ_t with respect to the Gibbs measure. We call this rate the entropy production at time t . The initial relative entropy is at most $\beta^2 N/2$ on average over the disorder, so the total entropy production per coordinate is at most $\beta^2/2$ on average (Lemma 5.1). We write $a_N(t)$ for the entropy production at time t , divided by N and averaged over the disorder. Following the usual terminology for disorder averages, we also call $a_N(t)$ the normalized annealed Fisher information (Section 5).

Other works use other normalizations. Cugliandolo and Kurchan [51, Equations (1)–(2)] minimize a Hamiltonian whose couplings are indexed by distinct unordered p -tuples and have variance $p!/(2N^{p-1})$. Their limiting covariance convention is $N\xi(x \cdot y/N)/2$, with an inverse temperature β^{CK} in front of the drift and noise with variance rate 2. At the level of the limiting equations, their convention agrees with ours when $\beta^{\text{CK}} = \sqrt{2}\beta$, with no change of time. In the physical normalization of time, where the drift is the gradient itself and the noise has variance rate 2θ at temperature θ , a rescaling of time gives $\beta^{\text{CK}} = 1/\theta$, and hence $\beta = 1/(\sqrt{2}\theta)$. The mathematical works [24, 60, 112] use Brownian noise with variance rate one; their drift parameter is $\alpha = \beta/2$, and their time is $v = 2t$. Table 3.1 in Section 3.3 collects the dictionary. All formulas in this paper are written in our conventions, and the formulas that we quote from [47, 51, 60, 112] have been translated.

1.3 The physical picture

We first recall what happens in equilibrium. Suppose that the diffusion is started from the Gibbs measure instead of the uniform measure. Then the process is stationary, and

its correlation and response depend only on the time difference $\tau = t - s$, which we also call the lag: they have the form $c(\tau)$ and $r(\tau)$. They satisfy the fluctuation–dissipation theorem (FDT),

$$r(\tau) = -c'(\tau), \quad \tau > 0,$$

which expresses the response to a small force through the spontaneous fluctuations of the system. FDT is a consequence of reversibility (Remark 3.1); see Kubo [100] for the general linear-response theory. Agarwal [1] treats response around nonthermal steady states under detailed balance; nonequilibrium diffusion response identities are discussed in [8]. In our conventions it carries no factor β . When $C(t, s) = c(t - s)$, it reads $R(t, s) = \partial_s C(t, s)$. Reversibility also gives a spectral representation. By the spectral theorem for the generator, the equilibrium correlation is a mixture of decaying exponentials, $c(\tau) = \int_{[0, \infty)} e^{-x\tau} \nu(dx)$ for a probability measure ν . By Bernstein’s theorem [122], a continuous function c on $[0, \infty)$ with $c(0) = 1$ has this form if and only if it is *completely monotone* on $(0, \infty)$, that is, infinitely differentiable there with

$$(-1)^n c^{(n)}(\tau) \geq 0, \quad n \geq 0, \tau > 0.$$

In particular, a completely monotone function is nonnegative, nonincreasing, and convex. We use the terms FDT and complete monotonicity in this sense throughout.

At high temperature, the dynamics started from the uniform point approaches this equilibrium behavior. Dembo, Guionnet, and Mazza proved that for β small enough, the CK solution relaxes exponentially fast and converges to a stationary solution that satisfies FDT [60, Proposition 1.2 and Theorem 1.3].

Below the dynamical transition, the picture is different. Let s be the waiting time, which is the age of the system when a measurement starts, and let $\tau = t - s$ be the lag. Cugliandolo and Kurchan constructed an asymptotic solution of the CK equations at long times [51]. It has the following features; the equation numbers refer to the arXiv version of [51].

- *A stationary regime.* When $s \rightarrow \infty$ at fixed lag τ , we have $C(s + \tau, s) \rightarrow c(\tau)$ and $R(s + \tau, s) \rightarrow r(\tau)$, and the limits satisfy FDT. On bounded windows of late times, the correlation and the response behave as in equilibrium.
- *A positive plateau.* The stationary correlation does not decay to zero. It decreases to a positive limit $q = \lim_{\tau \rightarrow \infty} c(\tau)$, called the plateau; in the physics literature, q is also called the dynamical Edwards–Anderson parameter.
- *Weak ergodicity breaking and an aging regime.* Below the plateau, the correlation decays on time scales that grow with the waiting time s , and $C(t, s) \rightarrow 0$ as $t \rightarrow \infty$ for every fixed s . The aging regime consists of pairs of late times $s < t$ with $t - s \rightarrow \infty$ at which $C(t, s)$ stays below the plateau.

- *Marginality.* The plateau is fixed by a marginality condition [51, Equation (7)], which says that a stability eigenvalue vanishes; we explain its meaning below. In our conventions, it reads $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$.
- *A constant fluctuation–dissipation ratio.* In the aging regime, FDT fails, and response and correlation are related instead by $R(t, s) \approx x \partial_s C(t, s)$, with a constant $x = (p-2)(1-q)/q$ given in the paragraph after [51, Equation (12)]; the corresponding response moments are [51, Equation (13)]. Equilibrium FDT corresponds to $x = 1$, and below the dynamical transition $x < 1$. Cugliandolo, Kurchan, and Peliti later interpreted x as the ratio of the temperature of the noise to an effective temperature of the slow motion, which is therefore higher [55].
- *The energy.* The limit of the energy $e_\beta(t)$ as $t \rightarrow \infty$ is given by [51, Equation (9)], combined with the relation between response and correlation in the aging regime; in our conventions, it is $\beta[1 + (p-2)q^{p-1} - (p-1)q^p]$.

The first three features imply that the two iterated limits of the correlation differ:

$$\lim_{\tau \rightarrow \infty} \lim_{s \rightarrow \infty} C(s + \tau, s) = q > 0, \quad \lim_{s \rightarrow \infty} \lim_{\tau \rightarrow \infty} C(s + \tau, s) = 0.$$

Taking the limit of large waiting time first describes the relaxation of an arbitrarily old system over a bounded time window. Taking the limit of large lag first asks whether the configuration at a fixed time is eventually forgotten. In equilibrium, both limits equal $\lim_{\tau \rightarrow \infty} c(\tau)$. In this model, their difference is a mathematical signature of aging: at large waiting times, the correlation first approaches the plateau, as it would in equilibrium, and it then decays further on a time scale that diverges with the waiting time. Thus stationarity on bounded windows does not describe the relaxation over the whole history. The noncommutation of the limits detects this slower relaxation without determining its time scale.

The dynamical transition is the temperature at which x reaches one [51]. In our conventions, the marginality condition and $x = 1$ give the transition parameters

$$q_d = \frac{p-2}{p-1}, \quad \beta_d^2 = \frac{(p-1)^{p-1}}{p(p-2)^{p-2}}.$$

For the cubic model, $q_d = 1/2$ and $\beta_d = \sqrt{4/3}$. For $\beta < \beta_d$, the physical picture predicts a zero plateau [48, 51]. Equivalently, β_d is the smallest β for which the equation $\beta^2 p q^{p-2}(1-q) = 1$ has a solution $q \in (0, 1)$, and this solution is q_d (Section 3.3). In the physical picture, this is the equation for the plateau of the equilibrium dynamics in the replica symmetric phase, so the equilibrium dynamics stops decorrelating at the inverse temperature β_d [48].

The picture can be compared with the static theory. Let β_s be the static transition, so that the Parisi measure is the point mass at zero for $\beta \leq \beta_s$ and has two atoms for $\beta > \beta_s$

[6]. Then $\beta_d < \beta_s$; for $p = 3$, we have $\beta_d \approx 1.155$ and $\beta_s \approx 1.207$ (see [119, Theorem 1] for the equations that determine β_s in our normalization). Since $\xi''(q) = p(p-1)q^{p-2}$, the marginality condition reads $\beta^2 \xi''(q)(1-q)^2 = 1$. Consider a trial measure with one step of replica symmetry breaking, $\theta \delta_0 + (1-\theta)\delta_q$ with $\theta \in (0,1)$, in the Crisanti–Sommers functional. In the replica calculation, its stability against perturbations of the overlaps within a block is measured by the replicon eigenvalue, which is a positive multiple of $1 - \beta^2 \xi''(q)(1-q)^2$. Thus the marginality condition says that the replicon eigenvalue vanishes, so that the solution is marginally stable. If, in addition, the derivative of the functional in q vanishes, then $\theta = x$. For $\beta > \beta_s$, the Parisi measure is obtained instead by optimizing the functional over both θ and q , and it is strictly stable. Barrat, Franz, and Parisi identified the asymptotic dynamical condition with this static marginal stability [12, Section IV, Equation (25)]; see also [49, arXiv version 2, Section 15.1].

In this sense, the plateau and the energy of the dynamics are those of marginally stable states rather than those of the equilibrium states that the Parisi measure describes, and the fluctuation–dissipation ratio x is the weight θ of the corresponding marginal solution. Cugliandolo and Kurchan noted that, in this model, the overlap distribution defined by the dynamics differs from the static one [51, the paragraph after Equation (15)]. Franz, Mézard, Parisi, and Pelti related the dynamical response to the static Parisi measure under additional hypotheses, and they pointed out that the pure spherical p -spin model does not satisfy this relation [73]. For $\beta_d < \beta \leq \beta_s$, the Parisi measure carries no positive overlap, while the predicted dynamical plateau is positive.

The limiting energy also has a landscape interpretation, developed alongside the calculation of Cugliandolo and Kurchan. The Thouless–Anderson–Palmer (TAP) free energy is a function of the local magnetizations whose stable local minima represent candidate metastable states. Crisanti and Sommers counted the solutions of the TAP equations of the spherical model and computed the threshold energy at which these solutions become marginally stable [47, Equations (8)–(9)]. At this threshold, the lower edge of the limiting spectrum of the Hessian of the TAP free energy reaches zero [51, arXiv version 1, p. 4]; see also [49, arXiv version 2, Section 15.4]. The energy of these threshold states depends on the temperature, and it agrees with the limiting energy of Cugliandolo and Kurchan [47, Equation (14)]. As $\beta \rightarrow \infty$, it tends to the threshold energy $E_{\text{th}} = 2\sqrt{(p-1)/p}$ of the landscape, which is strictly smaller than the ground-state energy $\lim_{N \rightarrow \infty} N^{-1} \mathbb{E} \max_{S_N} H_N$; for $p = 3$, we have $E_{\text{th}} \approx 1.633$, while the ground-state energy is approximately 1.657 [43]. Kurchan and Laloux related slow dynamics to the geometry of the landscape near the threshold [103, Section V]. Related analyses of metastable states, barriers, and stationary points of the TAP free energy are in [38, 104].

Cugliandolo and Kurchan developed the structure of the aging regime further in [52], where they studied the Sherrington–Kirkpatrick model. They distinguished weak ergodicity breaking from weak long-term memory, which means that the integrated

response to a force applied during any fixed time interval $[0, S]$ vanishes as $t \rightarrow \infty$ [52, arXiv version 1, Section 3]. They also observed that the asymptotic equations of the aging regime do not change when time is replaced by an increasing function of time and the response is transformed accordingly, so that these equations do not determine the time scale of aging [52, arXiv version 1, Section 4]. Further, they studied the relations between correlations at three widely separated times [52, arXiv version 1, Section 5]. Two important examples of such relations are the dynamical ultrametric relation $C(t, u) = \min\{C(t, s), C(s, u)\}$ for $u < s < t$, analogous to the ultrametric organization of equilibrium overlaps, and a product relation; for spherical p -spin models, they discussed a product relation [52, arXiv version 1, Section 5.1, Equation (5.1.8)]. Reviews of the two regimes and their response are given in [35, 49]. The idealized mode-coupling scenario of the glass transition was developed by Bengtzelius, Götze, and Sjölander and by Leutheusser [27, 107]. Mode-coupling theory also describes how a critical or marginal stationary correlation $c(\tau)$ approaches its plateau: $c(\tau) - q$ decays like τ^{-a} , where $a \in (0, 1/2)$ solves $\Gamma(1 - a)^2 / \Gamma(1 - 2a) = \lambda_{\text{MCT}}$, Γ is Euler's gamma function, and the constant λ_{MCT} , called the exponent parameter, is not to be confused with the multiplier $\lambda(t)$. The correspondence between disordered models and mode-coupling theory was worked out by Bouchaud, Cugliandolo, Kurchan, and Mézard in [34]. They gave this exponent relation at equilibrium [34, arXiv version, Equation (IV.3)] and for the stationary profile of the aging solution [34, arXiv version, Equation (IV.6)], where they attribute the calculation to Cugliandolo and Le Doussal [53]. For the pure spherical model, the exponent parameter of the aging solution is $x/2 = (p - 2)(1 - q)/(2q)$. This relation is written explicitly by Kim and Latz [96, the paragraph after Equation (12)]; see also [49, arXiv version 2, Equation (12.36)].

The order of limits is essential in this picture. As explained in Section 1.1, the dynamics converges to its Gibbs measure for each finite N . In the CK description, the dimension tends to infinity first, on every fixed time interval, and only then do the observation and waiting times tend to infinity. Aging in this sense is a property of a deterministic pair of functions (C, R) , and it is not a statement about time scales that grow with N . On such time scales, the dynamics can be activated, and the relevant mechanisms are different; we discuss the rigorous results on activated aging in Section 1.5. Neither a theorem about finite-dimensional mixing nor a formal solution of the asymptotic equations determines the behavior of the CK solution in the order of limits used here.

1.4 Main results

The goal of this paper is to prove the main features of the picture of Section 1.3 for pure p -spin models with $p \geq 3$ started from a uniform point. The physical calculation imposes a long-time form on the solution and solves reduced asymptotic equations. We prove

its conclusions for the CK solution itself, which is the limit of the Langevin dynamics, without assuming that form. At every positive temperature, we prove that the CK solution becomes stationary on bounded windows of lags and satisfies FDT there, that its energy converges, and that its correlation and response with every fixed earlier time tend to zero. For $\beta \geq \beta_d$, that is, at and below the dynamical transition temperature, we prove that the plateau is marginal and that the limiting energy is the one predicted by Cugliandolo and Kurchan. The two iterated limits of the correlation then differ, which is the sense in which the solution ages. We also prove that the response at a late time to forces applied at all earlier times concentrates at short lags in the mean-square sense, that the violation of FDT vanishes in a mean-square sense, and that, for $\beta > \beta_d$, its positive part keeps a positive integrated mass. For $\beta \geq \beta_d$, we prove that the stationary correlation approaches its plateau as a power law with a positive and finite amplitude. For the cubic model, we prove a lower bound on the susceptibility and an inequality between correlations at three times for $\beta \geq \beta_d$, and, for $\beta > \beta_d$, a lower bound that relates the time needed to relax below the plateau to the entropy production. We do not identify the time scale of aging, and we do not prove the full fluctuation–dissipation relation of the aging regime.

Let (C, R, λ) be the CK solution of Section 1.2. All long-time limits in the following statements concern this deterministic solution. They describe the Langevin dynamics only in the order of limits described in Section 1.3: first $N \rightarrow \infty$ on a fixed time horizon, through Proposition 3.3, and then the long-time limits.

Theorem 1.1 (Stationary limit, marginal plateau, and aging). *Fix $p \geq 3$ and $\beta > 0$. There exist functions $c_\beta, r_\beta: [0, \infty) \rightarrow \mathbb{R}$ and numbers $e_\beta(\infty)$ and ℓ_β such that $e_\beta(T) \rightarrow e_\beta(\infty)$ as $T \rightarrow \infty$ and*

$$C(T+t, T+s) \longrightarrow c_\beta(t-s), \quad R(T+t, T+s) \longrightarrow r_\beta(t-s), \quad \lambda(T) \longrightarrow \ell_\beta, \quad (1.1)$$

where the first two limits are locally uniform in $t \geq s \geq 0$. The function c_β is completely monotone on $(0, \infty)$, $r_\beta = -c'_\beta$, and $c_\beta(0) = r_\beta(0) = 1$. For every finite $S \geq 0$,

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq S} \{C(t, s) + R(t, s)\} = 0. \quad (1.2)$$

If $\beta \geq \beta_d$, then the plateau $q_\beta = \lim_{\tau \rightarrow \infty} c_\beta(\tau)$ is the larger of the two solutions $q \in (0, 1)$ of

$$\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1, \quad (1.3)$$

and $q_\beta \geq q_d$. In this range,

$$e_\beta(\infty) = \beta[1 + (p-2)q_\beta^{p-1} - (p-1)q_\beta^p], \quad (1.4)$$

$$\ell_\beta = \frac{1}{1-q_\beta} + \beta^2 p(1-q_\beta^{p-1}), \quad (1.5)$$

and the correlation limits do not commute:

$$\lim_{\tau \rightarrow \infty} \lim_{s \rightarrow \infty} C(s + \tau, s) = q_\beta > 0, \quad \lim_{s \rightarrow \infty} \lim_{\tau \rightarrow \infty} C(s + \tau, s) = 0. \quad (1.6)$$

Theorem 1.1 has two parts. The first holds at every temperature. It says that the CK solution becomes stationary at bounded lags: near every late time T , the correlation and the response are close to functions of the lag alone, these functions satisfy FDT, and the correlation has the spectral representation of an equilibrium correlation. Thus, on bounded windows of late times, the correlation and the response have the stationary FDT structure of an equilibrium dynamics, even at low temperature, where the dynamics keeps aging. The limits in (1.1) are taken along the whole family $T \rightarrow \infty$, so the stationary profile does not depend on a subsequence. By (1.2), the solution also forgets every fixed earlier time: its correlation with the configuration at a fixed time s , and its response to a force applied at time s , tend to zero.

The second part concerns $\beta \geq \beta_d$, and it identifies the plateau and the energy with those of the asymptotic solution of Cugliandolo and Kurchan. The relation (1.3) is their marginality condition [51, Equation (7)], and (1.4) is their energy [51, Equation (9)], combined with the relation between response and correlation in the aging regime stated after [51, Equation (12)]; both are written here in the conventions of Section 3.3. Since the plateau is positive while the correlation with every fixed earlier time vanishes, the two iterated limits in (1.6) differ. We call this noncommutation *correlation aging*. It does not identify the time scale on which $C(s + \tau, s)$ decreases below q_β . In physical terms, Theorem 1.1 shows that, for $\beta \geq \beta_d$, the dynamics started from a uniform point does not reach equilibrium: its plateau is the marginal one, and its limiting energy is the energy predicted for the threshold states. For example, if $\beta_d < \beta < \beta_s$, the Parisi measure is the point mass at zero and the equilibrium energy per coordinate is β , while by (1.4), we have $e_\beta(\infty) = \beta - \beta q_\beta^{p-1}[(p-1)q_\beta - (p-2)] < \beta$, since $q_\beta > q_d$.

The theorem includes the transition point $\beta = \beta_d$, where $q_\beta = q_d$ and $e_\beta(\infty) = \beta_d$. The value $e_{\beta_d}(\infty) = \beta_d$ coincides with the value $e_\beta(\infty) = \beta$ obtained whenever the plateau is zero (Remark 8.7). Letting $\beta \rightarrow \infty$ in (1.4), we find that $e_\beta(\infty) \rightarrow E_{\text{th}} = 2\sqrt{(p-1)/p}$ (Remark 8.6). This limit is the threshold energy of the pure p -spin landscape [7, Equation (2.13)], and it agrees with the low-temperature results of Sellke for Langevin dynamics on bounded time horizons [112, Theorems 1.1 and 1.7]. The limit $\beta \rightarrow \infty$ is taken after all other limits, and it says nothing about gradient flow at zero temperature.

We prove Theorem 1.1 at the end of Section 8.6, by combining Proposition 8.4, Theorems 8.5 and 8.8, and Corollary 8.9 with Proposition 5.7 and Corollary 7.7. The main difficulty is that the correlation equation involves the whole history: on late windows of bounded length, its memory integrals involve the overlap of late states with the remote past, which a limit on such windows does not control. The selection of the plateau and the forgetting of earlier times rest on two facts, which we state as Proposition 8.1 in

Section 8.1. To state them, we write $C(t, s) = \langle v_t, v_s \rangle$ for unit vectors v_t in a deterministic Hilbert space whose components are indexed by trees, and we project v_t onto the components whose leaves record the initial condition or noise up to an earlier time, while their internal interaction times may be later (Section 7). The representation comes with an energy functional, the tree energy, which reproduces $e_\beta(t)$ on v_t . The first fact says that, for every stationary limit with positive plateau, the tree energy of these projections, after normalization, does not exceed the threshold energy E_{th} in the limit. The second fact says that if the whole translated family converges to one stationary profile whose plateau is positive and marginal, then the solution forgets every fixed earlier time. We prove that proposition for $p \in \{3, 4\}$ in Section 9, using a deterministic inequality in this Hilbert space, and for every $p \geq 3$ in Section 10, using the algorithmic threshold of Huang and Sellke [92, Theorem 1 and Corollary 1.8].

To describe what happens for $\beta < \beta_d$, we use the following elementary facts about the marginality equation. Let

$$q_m = \frac{p-2}{p}, \quad \beta_m^2 = \frac{p^{p-1}}{4(p-1)(p-2)^{p-2}}.$$

The function $q \mapsto p(p-1)q^{p-2}(1-q)^2$ attains its maximum β_m^{-2} on $[0, 1]$ at q_m , so the marginality equation $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$ has a solution in $(0, 1)$ if and only if $\beta \geq \beta_m$; the value β_m is not a physical transition. For $\beta \geq \beta_m$, we write $q_+(\beta)$ for its solution in $[q_m, 1)$. We have $\beta_m < \beta_d$, $q_m < q_d$, and $q_+(\beta_d) = q_d$, and for $\beta \geq \beta_d$ we write $q_\beta = q_+(\beta)$; see Sections 8.2 and 8.4.

For $\beta < \beta_d$, Theorem 1.1 does not determine the plateau on the whole interval. For $\beta < \beta_m$, the plateau is zero. For $\beta_m \leq \beta < \beta_d$, the proof shows that it is either zero or $q_+(\beta)$, and it proves convergence in both cases (Proposition 8.4 and Remark 8.7). The physical picture predicts the zero plateau there [48, 51]. We return to this question in Section 17.4.

In the remaining statements, when p and β are fixed, we abbreviate $c = c_\beta$, $r = r_\beta$, $q = \lim_{\tau \rightarrow \infty} c_\beta(\tau)$, and $\ell = \ell_\beta$. All these statements use the stationary profile selected in Theorem 1.1. The only exception in notation is Theorem 1.4, where c denotes a level in $(0, q)$, as in Sections 14 and 15.

The stationary limit in Theorem 1.1 describes the response $R(t, s)$ at bounded lags $t - s$, and the forgetting statement describes it at bounded earlier times s . The next theorem concerns the response at a late time t to forces applied at all earlier times $s \in [0, t]$ simultaneously. We write it as a function of the lag $v = t - s \in [0, t]$, so that $R(t, t - v)$ is the response at time t to an impulse at time $t - v$. We also write

$$E(t, s) = \partial_s C(t, s) - R(t, s), \quad t > s > 0,$$

for the violation of FDT in the earlier time s , and $z_+ = \max\{z, 0\}$ for $z \in \mathbb{R}$. In equilibrium, $E = 0$. For $\beta > \beta_d$, the physical picture predicts $E(t, s) \approx (1 - x) \partial_s C(t, s) \geq$

0 in the aging regime, and this quantity is expected to be small there, because the correlation changes on a diverging time scale.

Theorem 1.2 (Response at late times and the violation of FDT). *Fix $p \geq 3$ and $\beta > 0$. Then, as $t \rightarrow \infty$,*

$$\begin{aligned} \mathbf{1}_{[0,t]}(v)R(t, t-v) &\longrightarrow r(v), \\ \mathbf{1}_{[0,t]}(v) \partial_s C(t, s)|_{s=t-v} &\longrightarrow r(v) \end{aligned} \quad \text{strongly in } L^2(0, \infty), \quad (1.7)$$

where the functions on the left are set equal to zero for $v > t$. Moreover,

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |R(t, s) - r(t-s)| = 0, \quad \chi(t, 0) = o(\sqrt{t}). \quad (1.8)$$

Let $E(t, s) = \partial_s C(t, s) - R(t, s)$. There exists a nonnegative integrable function a_F on $[0, \infty)$, obtained as an L^1_{loc} limit of the normalized annealed Fisher informations along one subsequence of dimensions, such that $\int_0^\infty a_F(s) ds \leq \beta^2/2$, $|E(t, s)|^2 \leq a_F(s)$ for $t > s > 0$, and $\sup_{s \geq T} a_F(s) \rightarrow 0$ as $T \rightarrow \infty$. Further,

$$\lim_{t \rightarrow \infty} \int_0^t |E(t, s)|^2 ds = 0. \quad (1.9)$$

If $\beta > \beta_d$, let

$$q = q_\beta, \quad x = \frac{(p-2)(1-q)}{q} \in (0, 1), \quad m_\infty = q(1 - x^{1/p}) > 0.$$

Then

$$\lim_{A \rightarrow \infty} \lim_{t \rightarrow \infty} \frac{p \int_0^{t-A} C(t, u)^{p-1} R(t, u) du}{C(t, t-A)^p - C(t, 0)^p} = x, \quad (1.10)$$

while, for every fixed $S \geq 0$,

$$\lim_{L \rightarrow \infty} \liminf_{t \rightarrow \infty} \int_S^{t-L} E(t, u)_+ du \geq m_\infty, \quad \liminf_{t \rightarrow \infty} t \int_0^t E(t, u)_+^2 du \geq m_\infty^2. \quad (1.11)$$

The first half of Theorem 1.2 says that the response and the derivative of the correlation in the earlier time concentrate at short lags. The stationary limit in Theorem 1.1 controls the response only on bounded windows of lags, while (1.7) concerns all lags $v \in [0, t]$, an interval whose length tends to infinity. In particular, for large A , the squared response integrated over lags larger than A is small at late times, and by (1.8), the response converges uniformly over all earlier times. This does not imply that the response integrated over these lags vanishes: the total susceptibility $\chi(t, 0)$ grows more slowly than $\sqrt{t}$, but we do not prove that it stays bounded. The convergence of the second function in (1.7) is a strong form of FDT in the earlier time: the violation $E(t, \cdot)$ tends to zero in $L^2(0, t)$.

The bound $|E|^2 \leq a_F$ explains why FDT holds in this sense. The function a_F is a limit, along one subsequence of dimensions, of the normalized entropy production a_N of Section 1.2, and its integral is at most $\beta^2/2$. The bound $|E(t, s)|^2 \leq a_F(s)$ says that the violation of FDT at an earlier time s is controlled by the entropy produced at time s . Since $\sup_{s \geq T} a_F(s) \rightarrow 0$ as $T \rightarrow \infty$, the violation is uniformly small when the earlier time s is large. This is the mechanism of Cugliandolo, Dean, and Kurchan [54], which we make uniform in the dimension. We call a_F the Fisher majorant. It need not be unique, but we choose one dimension subsequence and one majorant and use them throughout, including in Theorem 1.4.

FDT also holds in the later time, uniformly over all earlier times, for every $p \geq 3$ and $\beta > 0$:

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |\partial_t C(t, s) + r(t - s)| = \lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |\partial_t C(t, s) + R(t, s)| = 0$$

(Theorem 11.4). On the diagonal, $\partial_t C(t, t) = -1$ denotes the derivative from the causal side with the earlier time held fixed; $c'(0) = -1$ is a right derivative.

The second half of Theorem 1.2 shows that the violation of FDT nevertheless carries a positive integrated mass in the aging regime. For every fixed $S \geq 0$, the integral of the positive part of $E(t, \cdot)$ over $[S, t - L]$ has lower limit at least m_∞ when first $t \rightarrow \infty$ and then $L \rightarrow \infty$, as in (1.11), while the L^2 norm of $E(t, \cdot)$ tends to zero. These statements hold together because the violation is integrated over the growing interval $[S, t - L]$, whose length is of order t ; the theorem does not say how the violation is distributed over this interval. By the second inequality in (1.11), the L^2 norm of E_+ cannot decay faster than $t^{-1/2}$. Cugliandolo, Dean, and Kurchan explained in [54] why entropy production allows small pointwise violations of FDT to coexist with a nonzero integrated violation over an expanding time interval. In Theorem 1.2, we prove such a coexistence for the CK solution. The ratio in (1.10) identifies one moment of the slow response with the fluctuation–dissipation ratio x of [51]. Indeed, if the response satisfied $R(t, u) = x \partial_u C(t, u)$ for $u \leq t - A$, as the constant ratio of [51] predicts for large A , then the ratio would equal x , since $p \int_0^{t-A} C(t, u)^{p-1} \partial_u C(t, u) \, du = C(t, t - A)^p - C(t, 0)^p$. We prove this weighted identity, but not the pointwise relation.

The Fisher majorant is constructed in Theorem 5.3. The convergence in (1.7) and the bound $\chi(t, 0) = o(\sqrt{t})$ are proved in Theorems 11.1 and 11.5 and Corollary 11.2, the identity (1.10) in Proposition 12.1, and the bounds (1.11) in Theorem 12.4.

The convergence over all earlier times also shows where the correlation at the plateau comes from. The Green covariance identity of Proposition 4.2 writes $C(t, s)$ as the sum of three terms: the contribution $R(t, 0)R(s, 0)$ of the initial condition, a thermal term, which is the contribution of the Brownian noise, and a memory term, which is the contribution of the colored noise of the effective process. For $t \geq s$, the thermal term and its stationary

counterpart are

$$B(t, s) = 2 \int_0^s R(t, u) R(s, u) \, du, \quad b_{\text{th}}(\tau) = 2 \int_0^\infty r(\tau + v) r(v) \, dv.$$

For every $p \geq 3$ and $\beta > 0$, we show in Corollary 11.7 that

$$\lim_{S \rightarrow \infty} \sup_{t \geq s \geq S} |B(t, s) - b_{\text{th}}(t - s)| = 0. \quad (1.12)$$

The function b_{th} is completely monotone and tends to zero. Hence the thermal term vanishes on every late pair of times whose lag diverges. The initial term also vanishes, by (1.2). Thus any correlation retained on such a pair is carried by the memory term, which comes from the interaction of each coordinate with the others.

For the cubic model, we prove two further predictions of the physical picture in one-sided form. Cugliandolo and Kurchan described the aging regime through the relations between correlations at three times [52, arXiv version 1, Section 5]. For spherical models, they discussed a relation in which correlations below the plateau compose multiplicatively: for three times that are far apart, $C(t, u)/q$ is the product of $C(t, s)/q$ and $C(s, u)/q$ [52, arXiv version 1, Section 5.1, Equation (5.1.8)].

Theorem 1.3 (Three-time inequality and susceptibility in the cubic model). *Let $p = 3$ and $\beta \geq \beta_d$. For every sequence of triples $u_n < s_n < t_n$ with $u_n \rightarrow \infty$, $s_n - u_n \rightarrow \infty$, and $t_n - s_n \rightarrow \infty$,*

$$\liminf_{n \rightarrow \infty} \{qC(t_n, u_n) - C(t_n, s_n)C(s_n, u_n)\} \geq 0. \quad (1.13)$$

The same holds for $qC(t_n, s_n) - C(t_n, u_n)C(s_n, u_n)$ and for $qC(s_n, u_n) - C(t_n, s_n)C(t_n, u_n)$. Further, under the same hypotheses,

$$\liminf_{t \rightarrow \infty} \chi(t, 0) \geq 2(1 - q). \quad (1.14)$$

For the cubic model, $\beta_d = \sqrt{4/3}$ and $q \geq 1/2$, with $q = 1/2$ at $\beta = \beta_d$; all three product inequalities and the susceptibility bound include this point. The inequality (1.13) is a one-sided version of the product relation of Cugliandolo and Kurchan for the CK solution, with no assumption on the time scale of aging. In words, after dividing correlations by q , the correlation across two successive long time intervals is asymptotically at least the product of the correlations across the two intervals. For a level $c \in (0, q)$, consider the last time after s at which the correlation with the configuration at time s is at least c . The inequality (1.13) implies that these last times compose: for fixed levels $c_1, c_2 \in (0, q)$ and a fixed level below $c_1 c_2 / q$, and for all sufficiently large s , the last time for the lower level is at least the last time for c_2 , started from the last time for c_1 (Corollary 14.4). These statements allow the correlation to return above a level before its last crossing, and (1.13) does not assert a product equality.

The bound (1.14) concerns the total response to a constant force switched on at time zero. If FDT held at all pairs of times, we would have $\chi(t, 0) = 1 - C(t, 0) \rightarrow 1$. Of this amount, the stationary regime contributes $\int_0^\infty r(v) dv = 1 - q$, so (1.14) forces at least another $1 - q$ of integrated response outside every fixed window of lags. In the full physical prediction, including the absence of response mass at correlation zero, the aging regime contributes xq instead of q , which gives $\chi(t, 0) \rightarrow (1 - q) + xq = (p - 1)(1 - q)$; for $p = 3$, this is $2(1 - q)$, which is smaller than one when $q > 1/2$. We prove only the lower bound. Both results come from the geometry of the tree vectors of Section 7 whose energies approach the threshold: the product inequality from Theorem 13.5, and the susceptibility bound from a limiting operator associated with such vectors (Theorem 13.2). They are proved in Theorems 14.3 and 14.7.

The next result relates relaxation below the plateau to entropy production. For a level $c \in (0, q)$ and a waiting time s , we consider the time $\tau_c(s)$ that elapses after s until the correlation with the configuration at time s first reaches the level c .

Theorem 1.4 (Entropy production and relaxation below the plateau in the cubic model). *Let $p = 3$ and $\beta > \beta_d$, and let a_F be the Fisher majorant of Theorem 1.2. If $s_n \rightarrow \infty$, $\tau_n \rightarrow \infty$, and*

$$\tau_n \int_{s_n}^{s_n + \tau_n} a_F(u) du \rightarrow 0,$$

then $C(s_n + \tau_n, s_n) \rightarrow q$ and $\chi(s_n + \tau_n, s_n) \rightarrow 1 - q$. For $0 < c < q$, the first passage time

$$\tau_c(s) = \inf\{\tau \geq 0 : C(s + \tau, s) \leq c\}$$

is finite, tends to infinity as $s \rightarrow \infty$, and satisfies

$$\liminf_{s \rightarrow \infty} \tau_c(s) \int_s^{s + \tau_c(s)} a_F(u) du \geq \min\left\{1, \left[\frac{3\beta^2(2q - 1)(q - c)^2}{\ell + 15\beta^2}\right]^2\right\} > 0. \quad (1.15)$$

The same lower bound holds after replacing the integral by $\int_s^\infty a_F(u) du$.

The first statement concerns windows $[s_n, s_n + \tau_n]$ whose starting times and lengths both tend to infinity. If the length of the window times the integral of a_F over it tends to zero, then at the end of the window, the correlation with its start tends to the plateau, and the susceptibility tends to its stationary value $1 - q = \int_0^\infty r(v) dv$, as if the solution were stationary. Relaxation below the plateau therefore requires entropy production: to reach a fixed level $c < q$, the elapsed time multiplied by the integral of a_F over the elapsed time must stay bounded below, and (1.15) gives an explicit lower bound. As in Section 15, we call the quantity bounded in (1.15) the Fisher cost of the passage. Denoting the right side of (1.15) by κ_c , the bound with $\int_s^\infty a_F$ shows that

$$\tau_c(s) \geq \frac{\kappa_c - o(1)}{\int_s^\infty a_F(u) du} \quad (s \rightarrow \infty).$$

Since a_F is integrable, the right side tends to infinity: the time needed to relax below the plateau grows at least as fast as the reciprocal of the remaining integral $\int_s^\infty a_F$ of the chosen Fisher majorant. This is a quantitative form of aging, but it does not give a growth law for $\tau_c(s)$, because we do not know how fast $\int_s^\infty a_F$ decays. The strict condition $\beta > \beta_d$ is essential to this positive bound: its factor $2q - 1$ vanishes at the transition. The proof, in Theorem 15.1, combines the backward response equation with the product inequality of Theorem 1.3.

Our last main result says that, for $\beta \geq \beta_d$, the stationary correlation approaches its plateau algebraically: $c(t) - q$ is asymptotic to a positive constant times t^{-a} , with the exponent predicted by mode-coupling theory.

Theorem 1.5 (Power-law approach to the plateau). *Fix $p \geq 3$ and $\beta \geq \beta_d$. There exists a unique $a \in (0, 1/2)$ such that*

$$\frac{\Gamma(1-a)^2}{\Gamma(1-2a)} = \frac{(p-2)(1-q)}{2q}, \quad (1.16)$$

and there exists $A_\beta \in (0, \infty)$ such that, as $t \rightarrow \infty$,

$$c(t) - q \sim A_\beta t^{-a}, \quad (-1)^n c^{(n)}(t) \sim A_\beta(a)_n t^{-a-n} \quad (n \geq 1). \quad (1.17)$$

Here $(a)_n = a(a+1) \cdots (a+n-1)$, and n is fixed.

The relation (1.16) is the exponent relation of mode-coupling theory for the stationary profile of the aging solution, with exponent parameter $x/2$ [34, arXiv version, Equation (IV.6)], [96]. At the transition, $x = 1$ and $a \approx 0.395$ for every p ; as $\beta \rightarrow \infty$, the right side of (1.16) tends to zero and a tends to $1/2$. The critical power law, with an amplitude that depends on the dynamics at finite time differences, is part of the classical asymptotic analysis of mode-coupling theory [81, Section 4, Equation (6), pp. 880–881]. In Theorem 16.1, we prove it for the stationary profile selected by the CK solution, together with a positive finite amplitude and the asymptotics of every fixed derivative. In particular, the response $r = -c'$ decays like $aA_\beta t^{-a-1}$, and the approach to the plateau has no single exponential relaxation time. These exponents describe how the stationary profile approaches its plateau after the limit of large waiting time; they are not exponents of the aging regime. None of the geometric or Fisher-cost arguments uses this power law.

1.5 Relation to previous work

We now compare our results and methods with previous work.

The physical predictions and their later refinements. The predictions tested by our results come from the asymptotic solution of Cugliandolo and Kurchan [51], described

in Section 1.3. Their derivation assumes that the stationary and aging regimes separate, and it imposes an asymptotic form on the correlation and response in the aging regime. The time scale of the aging regime is not determined by this calculation, because the asymptotic equations admit reparametrizations of time [52, arXiv version 1, Section 4]. In Theorem 1.1, we prove the stationary FDT, the marginal plateau, the energy, and the forgetting of fixed earlier times for the CK solution itself, without assuming that the regimes separate. Theorem 1.2 identifies one weighted moment of the slow response with the ratio x , and Theorem 1.3 gives one side of the product relation between three times. The full slow fluctuation–dissipation relation and the time scale of aging remain open. The effective-temperature interpretation of x was developed by Cugliandolo, Kurchan, and Peliti [55]. Franz, Mézard, Parisi, and Peliti related the dynamical response to the equilibrium overlap distribution under additional hypotheses [73], and they pointed out that the unperturbed spherical p -spin model does not satisfy this relation. We do not use equilibrium overlaps. Altieri, Biroli, and Cammarota developed a formulation of aging within dynamical mean-field theory in [2], and Chamon, Kennett, Castillo, and Cugliandolo developed the reparametrization invariance of the slow dynamics in [41]. Ghimentì, Berthier, Kurchan, and van Wijland related accelerated glass dynamics to time-reparametrization softness, through numerical comparisons and an asymptotic analysis of a nonreciprocally coupled spherical p -spin model [78, pp. 4–5 and SI Appendix, Section 7].

The picture has also been tested numerically and refined. Kim and Latz integrated the CK equations numerically over long times and compared the result with the predicted fluctuation–dissipation relation and with the crossover to the aging regime [96]. Andrianov and Lefèvre proposed a form of the time scale of aging by matching the stationary and aging regimes within a scaling expansion [3]. Folena, Franz, and Ricci-Tersenghi found that in zero-temperature quenches of mixed spherical models the long-time dynamics can depend on the initial condition and can reach energies below the threshold [70, Sections V–VII]. Folena and Zamponi tested weak ergodicity breaking numerically in mean-field spin glasses over long times, mainly in mixed models [69], and Lang, Sachdev, and Diehl developed a numerical renormalization method that reaches longer integration times [105]. Citro and Ricci-Tersenghi obtained numerical evidence for strong ergodicity breaking, in which the correlation with the initial configuration stays positive at long times, in the spherical $3 + 4$ model quenched from near the dynamical transition to zero temperature [45]. The studies of mixed models and of other initial conditions indicate that the selection of the threshold in Theorem 1.1 depends on our hypotheses.

Rigorous limits on fixed time horizons and the CK equations. The rigorous derivation of mean-field dynamical equations began with the large deviation methods on path space of Ben Arous and Guionnet for Langevin dynamics of spin glasses: a large deviation principle for asymmetric interactions [17] and a large deviation upper bound for symmet-

ric interactions [18], in regimes of short times or high temperature. Grunwald proved an analogous asymmetric large deviation principle and symmetric upper bound for Glauber dynamics, the spin-flip dynamics of Ising spin glasses, without restrictions on the fixed time horizon or temperature [86]. Guionnet obtained averaged propagation of chaos, that is, asymptotic independence of finitely many coordinates, and a quenched single-spin limit expressed as convergence in law over the disorder [87]. For spherical models, Ben Arous, Dembo, and Guionnet proved in [24, Theorem 1.2] that the empirical correlation and integrated response of the soft-spherical Langevin dynamics converge on fixed time horizons to the solution of the CK equations. Dembo, Guionnet, and Mazza took the limit of these deterministic equations as the soft constraint becomes hard [60, Proposition 1.1]. In the same paper, they proved exponential relaxation and convergence to a stationary FDT solution at sufficiently high temperature [60, Proposition 1.2 and Theorem 1.3], and they classified the solutions of the scalar memory equation by the largest admissible plateau and a stability inequality [60, Proposition 1.4]. They also showed that the CK solution has stationary FDT limits at every temperature, under the asymptotic conditions (6.5) and (6.6) of their paper [60, Proposition 6.1]. Sellke proved the convergence theorem for the hard-spherical diffusion itself [112, Theorem 4.3], by coupling it with soft-spherical diffusions [112, Lemmas 3.1 and 3.2]. Guionnet and Mazza studied the long-time behavior of nonlinear Kraichnan equations, which are integral equations for response kernels with a prescribed covariance, in [89], using positive noncrossing expansions: series with nonnegative terms indexed by pairings of ordered points whose arcs do not cross. Dembo and Subag derived dynamical equations on fixed time horizons for initial conditions in prescribed bands near critical points [58] and for Gibbs distributed initial conditions [59]. The latter result includes the initialization at inverse temperature zero, which is the uniform start, together with the limit of a hard spherical constraint [59, Theorem 1.8(a) and Proposition 1.10]. They also characterize stationary FDT dynamics and relaxation within equilibrium bands [59, Proposition 1.14 and Corollary 1.18]. The equilibrium relaxation conclusions concern Gibbs starts at the ambient temperature, whereas our long-time selection follows a quench from the uniform law to inverse temperature $\beta > 0$. Dembo, Lubetzky, and Zeitouni proved universality on fixed time horizons for Langevin-like dynamics with independent asymmetric interactions [61, Theorem 1.1]. For diffusions with linear random matrix interactions, Dembo and Gheissari proved universality of averaged observables for both symmetric and asymmetric couplings, using a stochastic Taylor expansion [57, Theorem 1 and Corollary 3]. Surveys of this program are [14] and [88, pp. 117–121].

Rigorous dynamical mean-field theory has recently been extended to first-order methods with random data [39] and to stochastic gradient descent [76], and its equations have been derived by a dynamical cavity method [56]. Long-time results are known for Langevin dynamics in spiked matrix models, which are random matrices with a low-rank

signal [108]. For Langevin diffusions with adaptive drift, Fan, Ko, Loureiro, Lu, and Shen proved convergence of the linear response on fixed horizons [65], and long-time results under an approximate time-translation invariance hypothesis, which they verify under a log-Sobolev inequality that is uniform in the dimension [66, Theorems 2.5 and 2.9]. For the orthogonally invariant Sherrington–Kirkpatrick model, Fan, Misiakiewicz, Wang, and Wen proved stationary FDT limits under rapid mixing on time scales that do not depend on the dimension and suitable confining curvature [67, Lemma 4.1]. The models and hypotheses of these works differ from ours.

We take the results of [24, 60, 112] as input, through Propositions 3.2 and 3.3, and prove long-time statements for the CK solution at every temperature. The main difference from [60, Proposition 6.1] is that our stationary limits (Proposition 5.7) require no asymptotic hypothesis on the solution: entropy production supplies the needed control.

Aging for the two-spin model and on other time scales. Ben Arous, Dembo, and Guionnet proved aging for the soft-spherical model with $p = 2$ in the order of limits used here [21]. Their analysis uses the spectral decomposition of the interaction matrix, which reduces the dynamics to coupled Ornstein–Uhlenbeck processes. This linear structure is absent for $p \geq 3$, where a positive stationary plateau must satisfy the nonlinear marginality condition (1.3).

A different rigorous theory of aging concerns time scales that grow with the dimension, on which the dynamics is activated and is trapped in deep states for long times. Ben Arous, Bovier, and Gayraud proved aging for the Glauber dynamics of the random energy model below the critical temperature [22, 23]. Ben Arous and Černý showed that the arcsine law of a stable subordinator is a universal aging scheme for trap models: under their hypotheses, the probability that the process occupies the same trap at two late times converges to a function of their ratio given by the generalized arcsine law [16]. For random-hopping-time dynamics, in which the jumps do not depend on the energy and the holding times do, Ben Arous, Bovier, and Černý proved that Ising p -spin models with $p \geq 3$ age like the random energy model on intermediate exponential time scales [25], and Ben Arous and Gün obtained extremal aging, which is governed by the longest holding times, on subexponential time scales [19]. Bovier and Gayraud proved convergence of clock processes in random environments, and aging for random-hopping-time dynamics of the p -spin SK model [36]. For the Metropolis dynamics of the random energy model, Černý and Wassmer proved convergence of the clock process and explained why aging of the two-time correlation does not follow directly from it [40, Theorem 1.1 and Remark 1.2(3)], and Gayraud proved two-time aging [75]. These results let the observation times grow with N and describe the dynamics through clock processes, which record the total time spent in the states visited by the dynamics; this is the trap mechanism of [33]. In the CK limit, the dimension tends to infinity first on fixed horizons, activation is invisible, and the aging of Theorem 1.1 is a property of the

deterministic CK solution. The differences between these orders of limits are discussed in the surveys [14, 15] and [88, pp. 120–121].

Landscape and algorithmic thresholds, shattering, and metastability. In physics, the threshold energy $E_{\text{th}} = 2\sqrt{(p-1)/p}$ is the energy of the marginal TAP states at zero temperature [47, Equation (9)]. Auffinger, Ben Arous, and Černý used random matrix theory to compute the complexity of the critical points of the pure spherical p -spin Hamiltonian, that is, the exponential growth rate of their number. Up to the sign convention, E_{th} is the energy E_∞ of their analysis [7, Equation (2.13)]. Auffinger and Ben Arous extended these computations to mixed models [5]. In pure models, Subag proved by a second moment computation that, for each fixed energy strictly between the threshold and the ground-state energy, the number of critical points with energy above it, divided by its expectation, converges to one in L^2 , in our sign convention [117, Theorem 1 and Corollary 2]. In optimization, E_{th} is the algorithmic threshold of pure models, that is, the largest energy that algorithms of a natural stable class can reach. Subag constructed an algorithm for spherical models, based on the Hessian, that reaches the energy $\int_0^1 \sqrt{\xi''(u)} \, du$, which equals E_{th} for $\xi(u) = u^p$ [118]. Montanari [110] and El Alaoui, Montanari, and Sellke [63] constructed message passing algorithms for Ising models. These algorithms iteratively use the gradient of the Hamiltonian, and with probability tending to one, they reach energies arbitrarily close to the ground-state energy under a no-overlap-gap assumption on the Parisi order parameter: in [110, Assumption 1], that the support of the Parisi measure is an interval $[0, q_*(\beta)]$ for all sufficiently large β , and in [63, arXiv version 1, Assumption 2], that the zero-temperature Parisi minimizer is strictly increasing on $[0, 1)$. Huang and Sellke proved a matching hardness result. Consider algorithms that map the Gaussian disorder coefficients to points of the ball of radius $\sqrt{N}$ in $\mathbb{R}^N$, with Lipschitz constants, in the Euclidean metrics, bounded independently of N . Then the probability that the energy per coordinate of the output exceeds the algorithmic threshold by a fixed positive amount tends to zero as $N \rightarrow \infty$. They proved this for even models in [94], and for multi-species spherical models, including odd pure models and external fields, in [92, Theorem 1 and Corollary 1.8]. Sellke proved that at low temperature, Langevin dynamics reaches energies close to E_{th} on bounded time horizons [112, Theorem 1.1]. He also proved that, from an initial point independent of the disorder, it stays below E_{th} on each fixed time horizon with probability tending to one as $N \rightarrow \infty$ [112, Theorem 1.7]. Our energy formula agrees with these results when the zero-temperature limit is taken after the long-time limits (Remark 8.6).

For every stationary limit with positive plateau, we use E_{th} as an upper bound for the limiting energy of normalized prefix centers, which are the projections of late tree vectors described in Section 1.4 (Proposition 8.1). For $p \in \{3, 4\}$, we prove this bound with a deterministic inequality for unit vectors in a Hilbert space graded by ordered trees

(Theorem 9.2 and Lemma 9.3). The inequality has a sharp constant, and the deficit from this constant controls the squared norm of the components of a vector at every fixed height, where the height of a tree is the largest distance from its root to a leaf. Branching constructions appear in the algorithms of Subag [118, Remark 6] and in the overlap geometry of Huang and Sellke [94, Remarks 1.2 and 1.3]; here we retain quantitative height defects inside the causal covariance realization. For general p , we use the theorem of Huang and Sellke [92] together with the coupling of Sellke [112, Lemmas 3.1 and 3.2].

Sellke's work on the marginal stability of approximate ground states [113] concerns a different part of the landscape, and we do not relate the stationary response to a Hessian. The finite-dimensional dynamics has also been studied directly. Gheissari and Jagannath estimated the spectral gap of spherical spin glass dynamics [77], and Ben Arous, Gheissari, and Jagannath bounded the evolution of the energy and the gradient by differential inclusions [26]. Ben Arous and Jagannath separated shattering, a decomposition into exponentially many well-separated and individually subdominant regions whose union has the full limiting free energy, from metastability, which concerns long residence times near a region before escape [20]. El Alaoui, Montanari, and Sellke proved shattering of the Gibbs measure of pure spherical models in a range of temperatures, and showed that Langevin dynamics started from the Gibbs measure keeps a large correlation with its initial point on exponentially long time scales [64, Corollary 2.9]. This plateau concerns equilibrium initial data and time scales that grow with N . The plateau in Theorem 1.1 concerns a uniform initial point and the limit in which the dimension tends to infinity first.

Methods. We describe the origin of the main tools of the proofs; Section 2 explains how they fit together.

Entropy production and FDT. Agarwal [1] treats response around nonthermal steady states under detailed balance; nonequilibrium diffusion response identities are given in [8]. Cugliandolo, Dean, and Kurchan bounded violations of FDT by the entropy production in [54, arXiv version 2, Equations (12)–(14) and (18)]. We use the Bochner formula for a Laplacian with a gradient drift [9, Proposition 3] and the calculus of reversible diffusions [10]. In Section 5, we establish dimension-uniform bounds for the projected hard-spherical response and pass them to a single Fisher majorant for the CK solution, which controls the violation of FDT at all pairs of times $t > s > 0$ and gives stationary FDT limits.

Effective processes and the Green identity. The covariance identity of Proposition 4.2 has the variation-of-constants interpretation of the linear effective single-spin equation with white and colored Gaussian noise [49, arXiv version 2, Section 9.5, Equation (9.56), p. 55]. We prove the identity directly from the CK equations in Section 4.

Expansions with nonnegative terms. Positive noncrossing expansions of the response equation appear in the work of Guionnet and Mazza [89, Section 2.1] and of Dembo,

Guionnet, and Mazza [60, Section 2]. In Section 7, we expand the covariance instead. We organize the expansion by ordered trees whose leaves record either the initial condition or a time of the thermal noise. This realizes $C(t, s)$ as the inner product of unit vectors in one deterministic Hilbert space, and the projection onto trees with early leaves plays the role of conditioning on the early noise.

Independent futures. Trajectories with a common past and independent future noise were used by Cugliandolo and Dean [50, arXiv version 1, Section II] and by Barrat, Burioni, and Mézard, who classified aging by the overlaps of such trajectories [11, Section 2]. The overlap of two independent continuations of lengths a and b equals the correlation at the reflected lag $a + b$ in equilibrium [11, Section 2, Equation (9)]. Franz and Virasoro proposed averaging cloned futures to estimate the magnetization of the region of configurations that the dynamics explores on intermediate time scales [72, arXiv version 3, Section III], and noise-averaged magnetizations and their relation to the threshold are central to the dynamical TAP approach of Biroli [31, arXiv version 2, Section III.B]. Time-averaged magnetizations whose normalized squared norm tends to q also appear in [52, arXiv version 1, Section 5.1, Equation (5.1.9)]. Our prefix centers are deterministic counterparts of these averages. In Proposition 7.5, we prove the reflected formula for the stationary limits of the CK solution: the older memory cancels exactly in a closed integral equation for the times after the earlier time, before any limit is taken. For general p , Section 10 uses finitely many branched future trajectories of the diffusion.

Complete monotonicity. The reflected kernel $(a, b) \mapsto c(a + b)$ is positive semidefinite, and together with the boundedness and continuity of c , this implies that c is completely monotone, by the Bernstein–Widder theorem and its semigroup form [28, 122]. Franosch and Voigtmann proved complete monotonicity for Brownian mode-coupling equations of mixtures by a different route, an iteration argument [71, arXiv version 1, Section III].

Gaussian conditioning. In Section 10, we prove forgetting for general p by revealing the disorder in stages, through the gradients of the Hamiltonian at finitely many points chosen from the history, and by applying the external-field threshold of [92, Theorem 1 and Corollary 1.8] at each stage. An Euler approximation of an earlier state lies in the span of the initial point, finitely many Brownian increments, and the computed gradients, and we condition on the Gaussian coefficients needed to determine the gradients on this span. Gaussian conditioning is classical [95]. Adaptive Gaussian conditioning is used by Bayati and Montanari for approximate message passing [13, arXiv version 4, Section 3.8] and by Bolthausen for the iterative construction of solutions of the TAP equations [32, Section 3]. El Alaoui, Montanari, and Sellke conditioned the disorder tensor on adaptive gradient observations in their analysis of message passing for spin glasses [63, arXiv version 1, Appendix A], and Huang and Sellke extended this conditioning to multi-species spherical models [93, Appendix A]. We adapt the method to a fixed finite history of the Langevin dynamics.

Mode-coupling and renewal tools. The scalar memory equation of a stationary profile belongs to the class studied in mode-coupling theory [83, Sections 3.1–3.2] and [82, Section 4.2.1, Equations (4.15)–(4.17), pp. 192–193]. The structural theory of these equations is developed in [84] and [71], including the characterization of the plateau [71, arXiv version 1, Section IV], and Sjögren interpreted them through regenerative processes, which restart at random times [114]. Our proof of Theorem 1.5 in Section 16 combines comparison principles, a weighted Schur inequality for the convolution in the scalar equation, a factorization of gamma variables, which turns a quadratic balance into a renewal equation, the key renewal theorem [68], and the monotone density theorem [30, Section 1.7.3].

The three-time inequality. The product relation for three times is part of the physical analysis of Cugliandolo and Kurchan [52, arXiv version 1, Section 5.1, Equation (5.1.8)]. We prove a uniform one-sided inequality for the CK solution in Theorem 14.3, from the constrained geometry of tree vectors whose energies approach the threshold (Theorem 13.5). We use it for the bound on the Fisher cost in Theorem 1.4.

Concurrent work. Several recent works study related questions with different models, initial conditions, or hypotheses. Dembo and Subag analyzed dynamics started from Gibbs measures, including limits on fixed time horizons and the stationary FDT dynamics and equilibrium band relaxation described above [59]. Their theorem includes the initialization at inverse temperature $\beta_0 = 0$, hence the uniform start [59, Theorem 1.8(a) and Proposition 1.10]; for initialization from a low-temperature Gibbs measure, [59, Theorem 1.8(b)] assumes additional conditions on the mixture and on the equilibrium Parisi measure (genericity and strict one-step replica symmetry breaking, in the sense of [59, Section 1]). The works [65–67] on dynamical mean-field equations, described above, concern other models; their long-time results use approximate time-translation invariance or mixing hypotheses. Lang, Citro, Leuzzi, and Ricci-Tersenghi computed the aging phase diagram and the asymptotic energies of mixed spherical models at zero temperature, within a formal asymptotic analysis [106]. Cipolloni and Cugliandolo studied coupled spherical p -spin systems, including their aging regimes and fluctuation–dissipation relations [44]. Our contribution relative to these works is the set of long-time conclusions for the uniformly initialized CK solution of pure models: the selection of the plateau and the energy, the forgetting of fixed earlier times, and the results on the response.

1.6 Scope and open problems

Our conclusions concern the CK solution started from a uniform point, not an arbitrary solution of a reduced asymptotic system, and they apply to Langevin dynamics only through the limit on fixed time horizons of Proposition 3.3. For every $p \geq 3$, the stationary limits, the convergence of the energy, the forgetting of fixed earlier times, and

the convergence of the response over all earlier times hold at every positive temperature. The identification of the plateau and the energy, correlation aging, and the stationary power law hold for $\beta \geq \beta_d$, and the weighted ratio and the persistent violation of FDT hold for $\beta > \beta_d$. For the cubic model, the three-time inequality and the susceptibility bound hold for $\beta \geq \beta_d$, while the positive Fisher cost requires $\beta > \beta_d$.

Several parts of the physical picture remain open, and we formulate them in Section 17:

- the full slow fluctuation–dissipation relation, of which (1.10) is one weighted moment (Section 17.1);
- an upper bound matching (1.14) for the susceptibility (Section 17.2);
- the time scale of aging, which the three-time inequality and the Fisher cost constrain but do not determine (Section 17.3);
- the selection of the plateau for $\beta_m \leq \beta < \beta_d$ (Section 17.4);
- the behavior on time scales that grow with the dimension, and at zero temperature (Section 17.5).

1.7 Organization of the paper

Section 2 gives an overview of the proofs. They have three main ingredients. From entropy production, we obtain stationary FDT limits on bounded late windows. The representation of the correlation by vectors in a Hilbert space indexed by trees separates the contributions of the earlier and the later noise. An energy bound for the earlier contributions selects the marginal plateau and shows that fixed earlier configurations are forgotten. The rest of the paper has five parts.

Part I contains the foundations. Section 3 states the model, the conventions, the CK equations, and the published inputs, including the limit theorem on fixed time horizons. In Section 4, we prove the energy identity, the Green covariance identity, and estimates that are uniform in time. In Section 5, we use entropy production to construct the Fisher majorant and to show that all limits of the CK solution on late time windows are stationary and satisfy FDT.

Part II proves Theorem 1.1. Section 6 derives the scalar equation of a stationary limit and its constraints. Section 7 realizes C as the Gram kernel of unit vectors in a Hilbert space graded by trees, and proves the reflection identity and complete monotonicity. Section 8 states the two threshold inputs (Proposition 8.1) and deduces from them the selection of the plateau, the forgetting of earlier times, and correlation aging. Section 9 proves the threshold inputs for $p \in \{3, 4\}$ with deterministic inequalities in the tree space, and Section 10 proves them for every $p \geq 3$.

Part III proves Theorem 1.2. In Section 11, we prove the convergence of the response and of the derivatives of the correlation over all earlier times, and the convergence of the

thermal term. In Section 12, we prove the weighted ratio and the persistent violation of FDT.

Part IV concerns the cubic model and proves Theorems 1.3 and 1.4. Section 13 studies tree vectors near the threshold, Section 14 transfers their geometry to the CK correlations and proves the susceptibility bound, and Section 15 proves the bound on the Fisher cost.

Part V contains Section 16, which proves Theorem 1.5, and Section 17, which discusses the open problems. Section 16 can be read once Part II is available.

Table 1.1 lists, for each main theorem, where it is proved and the main results that the proof uses.

Result	Proved in	Main inputs
Theorem 1.1	Section 8; the proof is assembled at the end of Section 8.6	Proposition 3.3; Proposition 5.7; Proposition 6.1; Proposition 7.5 and Corollary 7.7; Proposition 8.1, proved in Section 9 for $p = 3, 4$ and in Section 10 for all p
Theorem 1.2	Theorem 5.3; Theorems 11.1 and 11.5 and Corollary 11.2; Proposition 12.1 and Theorem 12.4	Theorem 1.1; Proposition 4.2; Proposition 7.5; Theorem 9.2 ($p = 3$) or Section 11.5 (all p); Lemma 12.3
(1.12)	Corollary 11.7	Corollary 7.7; Theorem 11.1
Theorem 1.3	Theorems 14.3 and 14.7	Theorem 9.2; Proposition 9.5; Theorems 13.2 and 13.5; Theorems 11.1 and 11.4 and Corollary 11.2
Theorem 1.4	Theorem 15.1	Proposition 4.5; Theorem 5.3; Corollary 11.2 and Theorem 11.4; Theorem 14.3
Theorem 1.5	Theorem 16.1	Proposition 6.1; Corollary 7.7; Theorem 8.5

Table 1.1. The main results, the sections and results in which they are proved, and the main inputs of their proofs.

Overview of the proofs

In this section, we explain the main steps of the proofs and the ideas behind them. We describe each step informally and point to the precise statements in later sections. In Section 2.1, we explain what makes the long-time behavior of the CK solution hard to analyze. Sections 2.2 to 2.6 describe the proof of Theorem 1.1, and Sections 2.7 to 2.9 describe the proofs of the other main theorems. Throughout, $p \geq 3$ and $\beta > 0$ are fixed, $m = p - 1$, and (C, R, λ) is the CK solution of Section 1.2. We illustrate several formulas in the cubic model, where $m = 2$, $\beta_d = \sqrt{4/3}$, and $q_d = 1/2$.

2.1 The difficulty

After the dimension limit, which we take first on every fixed time horizon (Proposition 3.3), we must study the deterministic solution (C, R, λ) of the integro-differential equations (3.9)–(3.11) as both of its times tend to infinity. The aging statement (1.6) combines two assertions of different kinds. The first concerns bounded windows of lags at late times, where we must show that the solution becomes stationary and that its stationary correlation decreases to a positive plateau q_β . The second concerns the whole history: we must show that $C(t, s) \rightarrow 0$ as $t \rightarrow \infty$ for every fixed s . Note that a positive stationary plateau implies neither the second assertion nor the noncommutation of the two limits in (1.6). The stationary profile records $C(T + t, T + s)$ only for bounded $t - s$, and it carries no information about $C(t, s)$ for fixed s .

Cugliandolo and Kurchan derived the plateau, the energy, and the loss of memory from an asymptotic solution of the CK equations [51]. Their derivation separates a stationary regime at bounded lags from an aging regime below the plateau, in which it imposes an asymptotic relation between response and correlation; within this ansatz, the plateau is fixed by the marginality condition. A proof cannot start from these assumptions. We do not know in advance that the two regimes separate, on which time scale the aging regime lives, or how response and correlation are related there, and our results determine neither the time scale nor the full relation.

The difficulty is visible in the equations. If we translate both observation times to late times and keep their difference bounded, the memory integral of the response equation (3.9) involves only the times between them. The memory integrals of the correlation equation (3.10), however, run over $[0, t]$ and $[0, s]$, and on a late window they involve

the overlap of late states with the remote past, which a limit on late windows does not control.

We make no assumption on this older memory. There are three main ingredients. The first is entropy production. We prove that the total entropy produced by the diffusion relative to its Gibbs measure is at most $\beta^2 N/2$ on average over the disorder, and deduce that every limit of the solution on late windows is stationary and satisfies FDT (Section 2.2). The closed response equation then constrains the plateau, but it does not select it (Section 2.3). The second ingredient is a realization of C as the Gram kernel of unit vectors v_t in a Hilbert space graded by trees, whose leaves record the times of the thermal noise. Projecting a late state onto the trees with early leaves plays the role of conditioning on the early noise, and the Gram kernel of these projections satisfies an equation from which the older memory cancels exactly (Section 2.4). The third ingredient is an energy ceiling for normalized projections of late states. Together with the scalar constraints, it forces every positive plateau to be marginal, and when it is saturated, it forces the projections to forget every fixed earlier state (Sections 2.5 and 2.6). For $p \in \{3, 4\}$, the ceiling is a deterministic inequality in the tree space; for general p , we use the algorithmic threshold of Huang and Sellke [92, Theorem 1 and Corollary 1.8].

2.2 From the diffusion to stationary clusters

Inputs. The bridge of Proposition 3.3 is Sellke’s convergence theorem for hard-spherical Langevin dynamics [112, Theorem 4.3]: on every fixed time horizon, empirical correlations and a Brownian observable converge, and by Lemma 3.4, this observable is the integrated response to an additive force. In finite dimension, the normalized trace response has diagonal $\kappa_N = (N - 1)/N$. Its annealed mean converges to R uniformly on fixed causal triangles, by the compact-time modulus of the tangent flow and uniqueness of the limit identified by the primitives (Remark 3.5); this is stronger than the convergence of primitives that the published bridge gives. Integrating by parts in the Gaussian disorder, we find that $e_\beta(t)$ is the limit of the mean energy and that $\lambda = 1 + p\beta e_\beta$ (Proposition 4.1). Our basic structural tool is the Green covariance identity of Proposition 4.2,

$$\begin{aligned} C(t, s) = & R(t, 0)R(s, 0) + 2 \int_0^{t \wedge s} R(t, u)R(s, u) \, du \\ & + \beta^2 p \int_0^t \int_0^s R(t, u)R(s, w)C(u, w)^{p-1} \, dw \, du, \end{aligned} \tag{2.1}$$

whose three terms come from the initial condition, the thermal noise, and the colored memory, each propagated by the response. It has the variation-of-constants interpretation of the linear effective single-spin equation with white and colored Gaussian noise [49, arXiv version 2, Section 9.5, equation (9.56), p. 55]. We prove the identity directly from the CK equations, whose derivation and rigorous fixed-horizon limits are discussed in

[24, 51, 58]. We also show that $1 \leq \lambda \leq \Lambda$ and $0 \leq R \leq C$, and that the translated kernels $C_T(a, b) = C(T + a, T + b)$, $R_T(a, b) = R(T + a, T + b)$, and $\lambda_T(a) = \lambda(T + a)$ are precompact as $T \rightarrow \infty$ (Section 4).

Entropy production. Fix the disorder, let ρ_t be the density of X_t with respect to the uniform measure ω_N on S_N , and recall that the diffusion is reversible with respect to the Gibbs density $g_N \propto e^{\beta H_N}$. The relative entropy of ρ_t with respect to g_N decreases at the rate $I_H(t)$, the relative Fisher information. Since the initial law is uniform, the initial relative entropy is $\log Z_{N,\beta} - \beta \int H_N d\omega_N$, whose mean is at most $\log \mathbb{E} Z_{N,\beta} = \beta^2 N/2$ by Jensen's inequality. With $a_N(t) = N^{-1} \mathbb{E} I_H(t)$, we arrive at the budget

$$\int_0^\infty a_N(t) dt \leq \frac{\beta^2}{2} \quad (2.2)$$

of Lemma 5.1. The link with FDT is the exact identity of Lemma 5.2. If c_H and r_H are the correlation and the response at fixed disorder, and $m_\tau(x)$ is the conditional mean of $X_{s+\tau}$ given $X_s = x$, then for $t > s$ we have

$$r_H(t, s) - \partial_s c_H(t, s) = \frac{1}{N} \int \rho_s \left\langle m_{t-s}, \nabla_s \log \frac{\rho_s}{g_N} \right\rangle d\omega_N.$$

At $s = 0$, the derivative in the earlier time is taken from the right. The right side vanishes in equilibrium, and by the Cauchy–Schwarz inequality its square is at most $I_H(s)/N$. After averaging over the disorder, we find that $|\bar{r}_N(t, s) - \partial_s \bar{c}_N(t, s)|^2 \leq a_N(s)$: the violation of FDT at the earlier time s is controlled by the entropy production at time s . This is the mechanism of Cugliandolo, Dean, and Kurchan [54, arXiv version 2, equations (12)–(13), p. 3, and (18), p. 4].

A single Fisher majorant. We need this bound for the CK solution at every pair of times. However, the budget (2.2) bounds a_N only in L^1 , and the integrability of a limit of a_N would not make it small at every late time. To overcome this, we turn to the Bochner formula [9, Proposition 3, equation (4b), p. 187], which shows that $e^{-Kt} I_H(t)$ is nonincreasing on an event of overwhelming probability for the disorder, with K depending only on p and β . We apply Helly's selection theorem to the Fisher informations restricted to this disorder event (compare the selection proof in [29, Theorem 25.9, p. 336]), and use the landscape moments and entropy budget to show that the disorder outside this event contributes an L^1 mass that vanishes as $N \rightarrow \infty$. This gives one subsequence (N_j) , which we keep for the rest of the paper, along which the full normalized annealed Fisher information a_{N_j} converges in L^1_{loc} to a function a_F , the Fisher majorant. We prove in Theorem 5.3 that

$$\begin{aligned} \int_0^\infty a_F(u) du &\leq \frac{\beta^2}{2}, \quad \lim_{T \rightarrow \infty} \sup_{s \geq T} a_F(s) = 0, \\ |\partial_s C(t, s) - R(t, s)|^2 &\leq a_F(s), \quad t > s > 0, \end{aligned} \quad (2.3)$$

where the decay of $\sup_{s \geq T} a_F(s)$ follows from integrability and the monotonicity of $e^{-Kt} a_F(t)$. We choose the right-continuous representative of this limit and use the same a_F below, including for the physical relaxation cost. One function thus dominates the backward defect $E(t, s) = \partial_s C(t, s) - R(t, s)$ on the whole triangle. In the same way, using a second current identity and the gradient bound for the semigroup, we show that a correlation at fixed lag changes little when both times are shifted. For $0 \leq \tau \leq L$ and $0 \leq w_1 \leq w_2$, the difference $|C(w_2 + \tau, w_2) - C(w_1 + \tau, w_1)|$ is at most $B_L \int_{w_1}^{w_2} \sqrt{a_F(u)} \, du$, where B_L depends only on p , β , and L (Lemma 5.5); compare [54, arXiv version 2, equation (15), p. 3].

Stationary clusters. As a_F is integrable, the right sides of these bounds vanish on late windows of bounded length. Combining this with the precompactness of the translated kernels, we prove in Proposition 5.7 that every limit of $(C_{T_n}, R_{T_n}, \lambda_{T_n})$ along a sequence $T_n \rightarrow \infty$ has the form $c(|a - b|)$, $r(a - b)$, and ℓ , where $c(0) = r(0) = 1$, $0 \leq c \leq 1$, $r = -c' \geq 0$, and the constant ℓ equals $1 + p\beta e_\infty$ for the limit e_∞ of the energies. Stationarity comes from the flatness bound, and FDT from integrating the Fisher bound over a bounded window. As the memory integral of the response equation runs only between the two observation times, we also obtain the closed stationary response equation

$$r'(\tau) = -\ell r(\tau) + \beta^2 p(p-1) \int_0^\tau c(\tau-u)^{p-2} r(\tau-u) r(u) \, du, \quad (2.4)$$

in which the coefficient $\beta^2 p(p-1)$ is $6\beta^2$ for $p = 3$. We call (c, r, ℓ) a stationary cluster and $q = \lim_{\tau \rightarrow \infty} c(\tau)$ its plateau. Note that Proposition 5.7 determines neither q nor ℓ , that different sequences T_n may a priori lead to different clusters, and that it says nothing about $C(t, s)$ for fixed s .

2.3 The scalar equation and its constraints

Integrating (2.4) once, we obtain the scalar memory equation of Proposition 6.1,

$$c'(t) = -1 + \int_0^t \phi(c(t-u)) r(u) \, du, \quad \phi(y) = \ell - \beta^2 p + \beta^2 p y^{p-1}, \quad (2.5)$$

which is the equation classified by Dembo, Guionnet, and Mazza [60, arXiv version 1, Proposition 1.4, p. 5, and Section 5, pp. 17–20]. We derive its consequences directly, because Section 8 needs the constants and a strict inequality in exactly the following form. As $t \rightarrow \infty$, the integral tends to $\phi(q)(1-q)$, and the limit of $c'(t)$ must vanish because c is bounded. This gives the plateau identity

$$\phi(q)(1-q) = 1, \quad \text{that is,} \quad \ell = L(q) = \frac{1}{1-q} + \beta^2 p(1-q^{p-1}).$$

Next, ϕ is nondecreasing and c is nonincreasing. Replacing $c(t - u)$ by $c(t)$ in (2.5), we find that $c'(t) \geq F(c(t))$, where $F(y) = \phi(y)(1 - y) - 1$. The memory term of the stationary response equation is nonnegative, so $r(t) \geq e^{-\ell t} > 0$. Then c decreases strictly from 1 to q , every $y \in (q, 1)$ equals $c(t_y)$ for some t_y , and $F(y) \leq c'(t_y) < 0$. We have obtained the strict sign condition

$$F(q) = 0, \quad F(y) < 0 \quad (q < y < 1).$$

In particular, $F'(q) \leq 0$, which is the submarginality inequality $\beta^2 p(p - 1)q^{p-2}(1 - q)^2 \leq 1$. In the cubic model, these read $\ell = (1 - q)^{-1} + 3\beta^2(1 - q^2)$, $F(y) = (\ell - 3\beta^2 + 3\beta^2 y^2)(1 - y) - 1$, and $6\beta^2 q(1 - q)^2 \leq 1$. Finally, Lemma 6.2 states that, for a prescribed multiplier, the stationary system has at most one solution.

These are constraints on a possible cluster, not a selection of its plateau. If $q = 0$, then $\ell = 1 + \beta^2 p$ and the inequality says nothing. If $q > 0$, then ℓ is determined by q , and the plateau is submarginal. To prove equality in the marginality relation, we need an inequality in the opposite direction. It will come from an upper bound on the energy of the part of a late state that was already determined at an earlier time.

2.4 The causal tree realization and reflection

The tree space. In the Green identity (2.1), the first two terms form the inner product of $R(t, 0)e_0 + \sqrt{2}\mathbf{1}_{[0,t]}R(t, \cdot)$ and $R(s, 0)e_0 + \sqrt{2}\mathbf{1}_{[0,s]}R(s, \cdot)$ in $\mathbb{R}e_0 \oplus L^2(\mathbb{R}_+)$, where a unit vector e_0 represents the initial condition. If $C(u, w) = \langle v_u, v_w \rangle$, then $C(u, w)^m = \langle v_u^{\otimes m}, v_w^{\otimes m} \rangle$, and the third term is an inner product as well. We therefore look for unit vectors such that

$$v_t = R(t, 0)e_0 + \sqrt{2}\mathbf{1}_{[0,t]}R(t, \cdot) + \beta\sqrt{p} \int_0^t R(t, u) \mathcal{I}(v_u^{\otimes m}) \, du, \quad (2.6)$$

where $\mathcal{I}$ is an isometry from m -fold tensor powers into a space orthogonal to the first two terms. Iterating (2.6), we expand v_t into terms indexed by finite rooted trees in which every vertex has either no children or m ordered children (Section 7.1). Each leaf carries the initial vector or the time of a noise increment, and each internal vertex records one use of the memory integral. For example, when $p = 3$, the component of v_t indexed by the tree with two leaves, evaluated at two noise times $y_1, y_2 \leq t$, is

$$w_t(y_1, y_2) = 2\sqrt{3}\beta \int_{y_1 \vee y_2}^t R(t, u)R(u, y_1)R(u, y_2) \, du.$$

Noise received at times y_1 and y_2 is transported to a time u by the response, combined there by the cubic interaction, and transported to time t ; at a leaf that carries the initial vector, a factor $\sqrt{2}R(u, y_i)$ becomes $R(u, 0)$ (Example 7.2). Accordingly, $\mathcal{H}$ is an orthogonal sum, over trees, of tensor powers of $\mathbb{R}e_0 \oplus L^2(\mathbb{R}_+)$, and $\mathcal{I}$ grafts m trees onto a new root.

We write Π_j for the projection onto trees of height j , and P_T for the prefix projection, which keeps each component only where all of its leaf times are at most T . As the leaves of a graft are the leaves of its subtrees, P_T commutes with grafting.

The realization. In Proposition 7.1, we construct a norm-continuous curve of unit vectors $v_t \in \mathcal{H}$ that solves (2.6) and satisfies

$$\langle v_t, v_s \rangle = C(t, s), \quad P_t v_t = v_t, \quad h(v_t) = e_\beta(t), \quad \text{where } h(v) = \sqrt{p} \langle v, \mathcal{I}(v^{\otimes m}) \rangle.$$

We call h the tree energy. Every component of v_t is a nonnegative function of the leaf labels. The construction is a positive Picard iteration: the Gram kernels of the truncations at height n increase with n , because R and C are nonnegative, and their limit solves the Green equation, whose solution with values in $[-1, 1]$ is unique. Positive noncrossing expansions of the response equation appear in [89, Section 2.1] and [60, Section 2]; here we expand the covariance.

Prefix centers. For $T, a \geq 0$, we call $M_T(a) = P_T v_{T+a}$ a prefix center. It retains the components whose initial and thermal-noise leaf labels lie at or before T ; their internal memory integrals may still extend beyond T . For $s \leq T$, we have $P_T v_s = v_s$, and therefore $\langle v_{T+a}, v_s \rangle = \langle M_T(a), v_s \rangle$; that is, the correlation of a late state with any state up to time T is carried by its prefix center. At first sight, the Gram kernel of the prefix centers is as hard to compute as C , since the centers are built from the whole older history. However, the older memory cancels exactly. Let $K_T(t, s) = \langle P_T v_t, P_T v_s \rangle$. Applying P_T to (2.6), we see that K_T satisfies a Green identity in which the noise integral stops at T . If $u \leq T$, then $P_T v_u = v_u$, and therefore $K_T(u, w) = C(u, w)$. When we subtract the identity for K_T from (2.1), the initial terms, the noise integrals over $[0, T]$, and every block of the memory integral with $u \leq T$ or $w \leq T$ cancel. For $t, s \geq T$, only the future rectangle remains:

$$\begin{aligned} C(t, s) - K_T(t, s) &= 2 \int_T^{t \wedge s} R(t, u) R(s, u) \, du \\ &\quad + \beta^2 p \int_T^t \int_T^s R(t, u) R(s, w) \{C(u, w)^m - K_T(u, w)^m\} \, dw \, du. \end{aligned}$$

In the variables $a = t - T$ and $b = s - T$, this is the closed equation (7.20) for $Q_T(a, b) = \langle M_T(a), M_T(b) \rangle$ on a fixed rectangle. It holds for every T , before any limit is taken, and it uses no decay of the older memory.

Reflection. Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to a stationary cluster (c, r, ℓ) . The coefficients of the closed equation converge, and by a Gronwall argument, every solution of the limiting equation with values in $[0, 1]$ is the limit of Q_{T_n} . Comparing double Laplace transforms, and using only the stationary

response equation, we check that $c(a + b)$ is such a solution. This proves the reflection identity of Proposition 7.5: for every $A < \infty$,

$$\lim_{n \rightarrow \infty} \sup_{0 \leq a, b \leq A} |\langle M_{T_n}(a), M_{T_n}(b) \rangle - c(a + b)| = 0. \quad (2.7)$$

In particular, $\|M_{T_n}(L)\|^2 \rightarrow d_L = c(2L)$. We compute the energy of a center by the same kind of exact subtraction. When we subtract $h(v_t) = e_\beta(t)$ from $h(P_T v_t)$, only an integral over the last L units of time remains (7.27). Passing to the limit, we find that

$$h(M_{T_n}(L)) \longrightarrow h_L = e_\infty + \beta p \int_0^L r(\tau) \{c(2L - \tau)^{p-1} - c(\tau)^{p-1}\} d\tau. \quad (2.8)$$

These limits hold for every stationary cluster, whatever its plateau.

We may think of $M_T(a)$ as the average of the state at time $T + a$ over independent continuations of the thermal noise after time T . For a stationary reversible Markov process, the conditional means of two such continuations of lengths a and b have overlap $c(a + b)$ (Remark 7.6); this reflected formula appears in the work of Barrat, Burioni, and Mézard [11, Section 2, Equation (9)]. Averages over independent futures and time-averaged magnetizations [31, 50, 52, 72] are discussed in Section 1.5. In our proof, the reflected overlap comes from the exact cancellation above.

Complete monotonicity. By (2.7), the Hankel kernel $(a, b) \mapsto c(a + b)$ is a limit of Gram kernels, and it is positive semidefinite. Using a short semigroup argument in the spirit of the Bernstein–Widder theorem [28, 122], we deduce the representation of Corollary 7.7,

$$c(t) = \int_{[0, \infty)} e^{-xt} \nu(dx), \quad \nu(\{0\}) = q,$$

for a probability measure ν . In particular, c and r are completely monotone. Franosch and Voigtmann proved complete monotonicity for Brownian mode-coupling equations of mixtures by a different route, an iteration argument [71, arXiv version 1, Section III]. Finally, after time T no new noise leaf survives the projection P_T . By Lemma 7.3, the curve $t \mapsto P_T v_t$ is then differentiable on (T, ∞) , and its derivative is the image under P_T of an explicit polynomial function of v_t . We use this in Section 11 to bound $\partial_t C$ far from the diagonal.

2.5 Selection of the plateau from two threshold inputs

The threshold inputs. Besides the fact that the correlation of a late state with an earlier state is carried exactly by the center, Section 8 takes as input two facts about prefix centers, which we state as Proposition 8.1. Recall that h is homogeneous of degree p , so that $h_L/d_L^{p/2}$ is the limiting energy of the normalized center $M_{T_n}(L)/\|M_{T_n}(L)\|$. Let $E_{\text{th}} =$

$2\sqrt{(p-1)/p}$ be the threshold energy of the pure p -spin landscape [7, Equation (2.13)]. Informally, Proposition 8.1 states the following.

- (a) For every stationary cluster with positive plateau, the limiting center energies satisfy $h_L \leq E_{\text{th}} d_L^{p/2}$ for every $L \geq 0$.
- (b) If the translated kernels converge along every sequence $T \rightarrow \infty$ to one stationary profile, and if its plateau is positive and marginal, then the solution forgets every fixed earlier state: $\sup_{0 \leq s \leq S} C(t, s) \rightarrow 0$ as $t \rightarrow \infty$, for every $S < \infty$.

We prove Theorem 1.1 from these inputs in Section 8, and we prove the inputs in Sections 9 and 10.

An algebraic identity. Since c decreases to q , we have $d_L \rightarrow q$. Using dominated convergence in (2.8) and $p \int_0^\infty c^{p-1} r = 1 - q^p$, we find that

$$h_L \longrightarrow h = e_\infty - \beta \{1 - q^p - pq^{p-1}(1 - q)\}.$$

If $q > 0$, inserting the plateau identity $\ell = L(q)$ and $e_\infty = (\ell - 1)/(p\beta)$ into this formula, we obtain the identity of Lemma 8.2,

$$\frac{h}{q^{p/2}} = \sqrt{\frac{p-1}{p}} \left(z + \frac{1}{z} \right) \geq E_{\text{th}}, \quad z = \beta \sqrt{p(p-1)} q^{(p-2)/2} (1 - q). \quad (2.9)$$

Equality holds if and only if $z = 1$, that is, if and only if $\beta^2 p(p-1) q^{p-2} (1 - q)^2 = 1$. For $p = 3$, we have $h = e_\infty - \beta \{1 - q^3 - 3q^2(1 - q)\}$, $z = \beta \sqrt{6q} (1 - q)$, and $E_{\text{th}} = 2\sqrt{2/3}$. Note that this identity is algebra about a stationary cluster, and that it uses no threshold input.

Marginality. Let (c, r, ℓ) be a stationary cluster with $q > 0$. Letting $L \rightarrow \infty$ in part (a) of Proposition 8.1, we find that $h \leq E_{\text{th}} q^{p/2}$, while (2.9) gives $h \geq E_{\text{th}} q^{p/2}$. Equality must hold, and the plateau is marginal. This is the inequality in the opposite direction that the scalar constraints were missing. For $\beta > \beta_m$, the marginality equation has two roots in $(0, 1)$, and the strict sign condition excludes the smaller one. Indeed, at a marginal plateau we have $F(q) = F'(q) = 0$, and if $q < q_m = (p-2)/p$, then $F''(q) > 0$, so that F would be positive just above q . Hence $q = q_+(\beta)$, the root in $[q_m, 1)$ (Lemma 8.3).

Convergence at every temperature. It remains to show that the cluster does not depend on the sequence T_n . By marginality and the plateau identity, the multiplier of every cluster is either $1 + \beta^2 p$, when $q = 0$, or $L(q_+(\beta))$; it lies in a set with at most two elements. The function λ is continuous and bounded. By the intermediate value theorem, every point between $\liminf_{t \rightarrow \infty} \lambda(t)$ and $\limsup_{t \rightarrow \infty} \lambda(t)$ is the limit of $\lambda(T_n)$ along some sequence $T_n \rightarrow \infty$, and by compactness, each such limit is the multiplier of a cluster. An interval contained in a set with at most two elements is a single point, so λ converges. Once

the multiplier is fixed, Lemma 6.2 identifies the profile, and precompactness upgrades subsequential convergence to convergence of the whole translated family (Proposition 8.4). This holds for every $\beta > 0$, and the energy converges as well.

The selected plateau. For $\beta \geq \beta_d$, we exclude the zero plateau with the same sign condition. If $q = 0$, then $\ell = 1 + \beta^2 p$ and $F(y) = y\{\beta^2 p y^{p-2}(1-y) - 1\}$, and we find that

$$F(q_d) = q_d \left(\frac{\beta^2}{\beta_d^2} - 1 \right) \geq 0,$$

which contradicts the sign condition on $(0, 1)$. For $p = 3$, this reads $F(1/2) = (3\beta^2/4 - 1)/2$. At $\beta = \beta_d$, we have $F(q_d) = 0$, and the contradiction comes from the strictness of the sign condition; the endpoint needs no continuity argument in β . We conclude that $q = q_\beta = q_+(\beta) \geq q_d$, and the energy formula (1.4) follows from $\ell = L(q_\beta)$ and marginality (Theorem 8.5). For $\beta < \beta_m$, only the zero plateau is possible. For $\beta_m \leq \beta < \beta_d$, we prove convergence, but we do not decide between the two alternatives (Remark 8.7).

Forgetting and aging. If $q > 0$, the convergence of the whole translated family and marginality are the hypotheses of part (b) of Proposition 8.1, which implies that $\sup_{s \leq t} C(t, s) \rightarrow 0$. If $q = 0$, we need no threshold input. For $s \leq t - L$, we have $C(t, s) = \langle M_{t-L}(L), v_s \rangle \leq \|M_{t-L}(L)\|$ by the Cauchy–Schwarz inequality. The right side tends to $c(2L)^{1/2}$ as $t \rightarrow \infty$, and $c(2L) \rightarrow 0$ as $L \rightarrow \infty$. Using the domination $R \leq C$, we transfer forgetting to the response (Theorem 8.8). For $\beta \geq \beta_d$, the ordered limits (1.6) follow: the first inner limit is $c(\tau)$, which tends to $q_\beta > 0$, and the second vanishes by forgetting (Corollary 8.9).

The order of the argument. But there is an apparent circularity here, since the threshold inputs are proved in later sections, and part (b) assumes a marginal plateau, which is what part (a) is used to prove. The way out is the order of deduction. Part (a) concerns an arbitrary stationary cluster with positive plateau, and its proofs in Sections 9 and 10 use only Sections 3 to 7. We deduce marginality from part (a) and the identity (2.9), which uses no threshold input. Convergence at every temperature and marginality are then the hypotheses of part (b), from which we deduce forgetting.

2.6 Why the threshold inputs hold

The cubic tree ceiling

For $p = 3$, grafting is binary, the tree energy is $h = \sqrt{3}f$ with $f(v) = \langle v, \mathcal{I}(v \otimes v) \rangle$, and $E_{\text{th}} = \sqrt{3}F_*$ with $F_* = 2\sqrt{2}/3$. We prove in Theorem 9.2 that $|f(v)| < F_*$ for every unit

vector v , that F_* is the supremum, and that, with $b_j = \|\Pi_j v\|^2$,

$$\sum_{j \geq 0} b_j^2 \leq 4\sqrt{2} (F_* - f(v)). \quad (2.10)$$

The proof is a heightwise Cauchy–Schwarz inequality followed by a Riemann sum. Let $B_j = b_0 + \dots + b_j$, with $B_{-1} = B_{-2} = 0$. Since a graft is one level taller than its taller subtree, the component of $\mathcal{I}(v \otimes v)$ at height j is built from the components of v below height j , and its squared norm is $B_{j-1}^2 - B_{j-2}^2 = b_{j-1}s_j^2$, where $s_j = (B_{j-1} + B_{j-2})^{1/2}$ (Lemma 9.1). Using the Cauchy–Schwarz inequality at each height, and then $2\sqrt{b_j b_{j-1}} \leq b_j + b_{j-1}$, we find that

$$f(v) = \sum_{j \geq 1} \langle \Pi_j v, \Pi_j \mathcal{I}(v \otimes v) \rangle \leq \sum_{j \geq 1} \sqrt{b_j b_{j-1}} s_j \leq \frac{1}{2} \sum_{j \geq 1} (b_j + b_{j-1}) s_j.$$

We have $s_j \leq \sqrt{2B_{j-1}}$ and $s_{j+1} \leq \sqrt{2B_j}$, and the last sum is therefore at most the trapezoid sum of the concave function $\sqrt{2y}$ over the partition $0 \leq B_0 \leq B_1 \leq \dots$ of $[0, 1]$. This trapezoid sum is at most

$$\int_0^1 \sqrt{2y} \, dy = \frac{2\sqrt{2}}{3} = F_*,$$

and keeping track of the gaps in these steps gives (2.10). The supremum is approached by unit vectors with mass $1/(n+1)$ at each of the heights $0, \dots, n$, whose component at each height $j \geq 1$ is a positive multiple of the graft of the lower components. For these vectors, the first inequality above is an equality, and $f(v)$ is the midpoint Riemann sum of the same integral over the first n of the $n+1$ intervals of the partition; the omitted interval contributes at most $\sqrt{2}/(n+1)$.

The defect bound says more than $f(v) < F_*$. Together with the Cauchy–Schwarz inequality, it implies that $\|\Pi_{\leq H} v\|^2 \leq (4\sqrt{2}(H+1)(F_* - f(v)))^{1/2}$ for every $H \geq 0$, where $\Pi_{\leq H} = \Pi_0 + \dots + \Pi_H$: a unit vector close to the ceiling has little mass at every fixed set of heights. The proof also retains the gaps of the two inequalities above, an angular defect and a Young defect (9.3), which we use in Section 13. For $p = 4$, a difference of cubes replaces the difference of squares, and the ceiling is $f(v) = \langle v, \mathcal{I}(v^{\otimes 3}) \rangle < \sqrt{3}/2$ for unit vectors v , with $h = 2f$ and $E_{\text{th}} = \sqrt{3}$ (Lemma 9.3). To prove part (a) for $p \in \{3, 4\}$, we apply the ceiling to the normalized centers $M_{T_n}(L)/\|M_{T_n}(L)\|$ and let $n \rightarrow \infty$ using Proposition 7.5. The ceilings are elementary inequalities in the tree space, and their proofs use neither the landscape results of [7] nor the algorithmic results of [92, 94].

Height escape and forgetting

Under the hypotheses of part (b), the limits of Proposition 7.5 hold as $T \rightarrow \infty$ without passing to a sequence, and the normalized centers saturate the threshold. Indeed, the

deficit $\delta_L = E_{\text{th}} - h_L/d_L^{p/2}$ is nonnegative by part (a), and $h_L/d_L^{p/2}$ tends to $h/q^{p/2}$, which equals E_{th} by the equality case of (2.9). Then $\delta_L \rightarrow 0$, and by the defect bound, the mass of the centers leaves every fixed set of heights (Proposition 9.5):

$$\limsup_{t \rightarrow \infty} \|\Pi_{\leq H} P_{t-L} v_t\|^2 \leq d_L \{\kappa_p (H+1) \delta_L\}^{1/2},$$

with $\kappa_3 = 4\sqrt{2/3}$ and $\kappa_4 = \sqrt{3}$. An earlier state behaves in the opposite way. The curve $s \mapsto v_s$ is norm-continuous, so the set $\{v_s : s \leq S\}$ is compact, and $\|(\text{Id} - \Pi_{\leq H})v_s\| \rightarrow 0$ as $H \rightarrow \infty$, uniformly in $s \leq S$. For $s \leq S \leq t-L$, we have $C(t, s) = \langle P_{t-L} v_t, v_s \rangle$, and splitting v_s at height H , we find that

$$|C(t, s)| \leq \|\Pi_{\leq H} P_{t-L} v_t\| + \|(\text{Id} - \Pi_{\leq H})v_s\|.$$

We first let $t \rightarrow \infty$ with L and H fixed, then let $L \rightarrow \infty$, and finally let $H \rightarrow \infty$. This proves part (b) for $p \in \{3, 4\}$.

General pure models

For general p , we have no deterministic ceiling in the tree space. In Section 10, we prove both inputs for every $p \geq 3$ by an argument in finite dimension, which does not use the tree ceilings.

Forks. In finite dimension, the analog of a prefix center is a mean over independent futures. At a late time T , we continue the trajectory by K independent Brownian futures of length L , which we call a fork. We let $m_{N,T}$ be the average of their endpoints, the center of the fork. With a current identity of the kind used in Section 5, we show that the overlap of two branches of lengths a and d concentrates near $c(a+d)$ along T_n ; this is the reflection identity in finite dimension (Lemma 10.3). We compute the energy of the center exactly by a Gaussian integration by parts, and subtracting the energy of a single branch removes the common past in the dimension limit. Unlike the cancellation of Section 7, this one holds only in the limit, because on a fixed time horizon the overlaps of all branches with a state of the common past concentrate around the same value. As $N = N_j \rightarrow \infty$ and then $T = T_n \rightarrow \infty$, we find that (Proposition 10.4)

$$\frac{\|m_{N,T}\|^2}{N} \longrightarrow Q_{K,L} = \frac{1}{K} + \left(1 - \frac{1}{K}\right)c(2L), \quad \frac{\mathbb{E}H_N(m_{N,T})}{N} \longrightarrow h_{K,L},$$

where $h_{K,L}$ is explicit. Using the hard-soft approximation of Sellke [112, Lemmas 3.1 and 3.2], we approximate the normalized center in mean square by Lipschitz functions of the disorder, with constants that do not depend on N . The algorithmic threshold of Huang and Sellke [92, Theorem 1 and Corollary 1.8] applies to these approximations, and it implies that $h_{K,L} \leq E_{\text{th}} Q_{K,L}^{p/2}$; this follows the argument by which Sellke showed that the

energy of the dynamics itself stays below the threshold [112, Section 3]. Letting $K \rightarrow \infty$, we have $Q_{K,L} \rightarrow d_L$ and $h_{K,L} \rightarrow h_L$, which proves part (a). Every horizon, number of branches, and approximation parameter is fixed while the dimension tends to infinity, and the limits are taken in the order

$$N_j \rightarrow \infty, \quad n \rightarrow \infty, \quad \varepsilon \downarrow 0, \quad K \rightarrow \infty, \quad L \rightarrow \infty, \quad M \rightarrow \infty$$

of (10.1), where ε is the accuracy of the Lipschitz approximation and M is the number of steps of the Euler–Maruyama scheme below.

Adaptive disclosure. Part (b) needs more. Under marginality, the normalized fork centers saturate the threshold as $K \rightarrow \infty$ and then $L \rightarrow \infty$ (Lemma 10.5). If a fork ends at time t , the mean overlap of its center with the earlier state X_s is the mean of $X_t \cdot X_s / N$, which tends to $C(t, s)$. We must therefore show that a Lipschitz algorithm close to the threshold is almost orthogonal to X_s . For directions independent of the disorder, this follows from the external-field version of the threshold [92, Corollary 1.8], which by convex duality implies the overlap penalty of Lemma 10.2. Let ζ_N be an auxiliary seed independent of the disorder tensor J , let $w = w(\zeta_N)$ satisfy $\|w\| = \sqrt{N}$, and let $\mathcal{A}_N(J, \zeta_N)$ take values in the ball of radius $\sqrt{N}$. If, for almost every value of the seed, $J \mapsto \mathcal{A}_N(J, \zeta_N)$ is L_{Lip} -Lipschitz, with a deterministic constant L_{Lip} that does not depend on N , then

$$\limsup_{N \rightarrow \infty} \left\{ \mathbb{E} \frac{H_N(\mathcal{A}_N)}{N} + \beta_d \mathbb{E} \left| \frac{\langle w, \mathcal{A}_N \rangle}{N} \right|^p \right\} \leq E_{\text{th}}.$$

However, X_s depends on the disorder. To overcome this, we disclose the disorder in stages (Lemma 10.6). At each stage, we reveal the part of the tensor that determines the next gradient query. Conditionally on what has been revealed, the rest of the tensor is standard Gaussian on a smaller space, and the overlap penalty applies to the new query direction. By induction over finitely many stages, we show that an algorithm that saturates the threshold is asymptotically orthogonal to every span obtained in this way from a seed-measurable initial span of fixed dimension by a fixed number of gradient queries (Proposition 10.7). In the proof of Proposition 10.8, we apply this only on the fixed fork product law, whose seed consists of X_0 , the full trunk Wiener path, and the K full future Wiener paths, jointly independent of J . For its Euler–Maruyama scheme with M steps, the initial span is generated by X_0 and the M trunk Brownian increments, and the scheme adds at most M gradient queries. The resulting span contains a vector approximating this fork’s trunk state X_s , with dimension-limit normalized mean-square error at most Cs/M , where C depends only on p , β , and s . We take the limits in (10.1), using $\varepsilon = 1/(m+1)$, and let $M \rightarrow \infty$ last. This proves that $C(t, s) \rightarrow 0$ for every fixed s (Proposition 10.8), uniformly in $s \leq S$ by the modulus of continuity of C . Adaptive Gaussian conditioning of this kind goes back to Bolthausen [32, Section 3] and to Bayati

and Montanari [13, arXiv version 4, Section 3.8]. Conditioning of the disorder tensor on adaptive gradient observations was developed for message passing in spin glasses in [63, arXiv version 1, Appendix A] and extended to multi-species spherical models in [93, Appendix A]; we combine it with the external-field threshold.

2.7 Response localization and persistent violation

The remote response. Theorem 1.1 controls the response only on bounded windows of lags, while a row $s \mapsto R(t, s)$ has length t . Convergence on every bounded window does not exclude a small contribution spread over the remote part $[0, t - L]$ of the row. In Section 11, we first prove that the remote part carries no squared mass:

$$\lim_{L \rightarrow \infty} \limsup_{t \rightarrow \infty} \int_0^{t-L} R(t, u)^2 \, du = 0. \quad (2.11)$$

Again we use the prefix centers. The component of v_{T+L} at height zero is the leaf vector $R(T+L, 0)e_0 + \sqrt{2}\mathbf{1}_{[0, T+L]}R(T+L, \cdot)$, and P_T restricts its noise part to $[0, T]$. This yields

$$2 \int_0^T R(T+L, u)^2 \, du \leq \|\Pi_0 M_T(L)\|^2 \leq \|M_T(L)\|^2 \longrightarrow c(2L) \quad (T \rightarrow \infty),$$

which proves (2.11) when $q = 0$. If $q > 0$, then $c(2L) \rightarrow q$, and we need to show that the normalized centers have vanishing mass at height zero. For $p = 3$, this follows from the term $j = 0$ of (2.10), because the normalized centers saturate the threshold. For general p , we test the center of a fork against the Gaussian vector $G_N = \int_0^T k(u) \, dB_u$, built from the Brownian path with a weight k proportional to $R(T+L, \cdot)$. A weighted form of the Brownian response identity shows that the mean overlap of $m_{N,T}$ with G_N tends to $(2 \int_0^T R(T+L, u)^2 \, du)^{1/2}$. Since G_N is independent of the disorder, the overlap penalty forces this overlap to be small, with no adaptive disclosure (Proposition 11.11).

Whole rows. Together with convergence on recent strips, (2.11) implies that the lag rows $v \mapsto \mathbf{1}_{[0,t]}(v)R(t, t-v)$ converge to r in $L^2(0, \infty)$, and by the Cauchy–Schwarz inequality, we have $\chi(t, 0) = o(\sqrt{t})$ (Theorem 11.1). By the backward response equation, we have $\partial_s R \leq \Lambda R$, so that a large remote value $R(t, s)$ would force squared mass on $[s-1, s]$; this upgrades the convergence to the uniform norm (Corollary 11.2). For the earlier derivative, we differentiate the Green identity in its earlier time. On a bounded interval of earlier times, forgetting makes every term vanish as $t \rightarrow \infty$, except a colored term $\int_0^t R(t, u)C(u, s)^{p-1} \, du$, which we control by positivity and a modulus of continuity uniform in t : the Green identity bounds $C(t, s)$ from below by a weighted integral of this term over earlier times, with positive weights; forgetting makes that integral vanish, and the uniform modulus in (11.17) then forces the term to vanish pointwise. This shows that

$E(t, s) \rightarrow 0$ for every fixed $s > 0$. Since $|E(t, s)|^2 \leq a_F(s)$ with an integrable a_F , dominated convergence over the whole row then shows that $\int_0^t |E(t, s)|^2 ds \rightarrow 0$ (Theorem 11.5). This is where we need one majorant for the whole triangle. For every $p \geq 3$, the projected drift and the reflection identity also bound $\partial_t C$ far from the diagonal (Theorem 11.4). More precisely, for every $L > 0$,

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq t-L} |\partial_t C(t, s)| \leq \sqrt{c''(2L)} \leq \frac{\sqrt{1-q}}{eL},$$

and, using convergence on bounded lag windows, we obtain

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |\partial_t C(t, s) + r(t-s)| = \lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |\partial_t C(t, s) + R(t, s)| = 0.$$

The diagonal derivative is taken from the causal side and equals -1 , while $c'(0) = -1$ is a right derivative. These conclusions about forward FDT hold at every positive temperature. For every $p \geq 3$, we also deduce (1.12) from complete-row convergence (Corollary 11.7).

Persistent integrated violation. These statements about squared norms are compatible with a persistent violation of FDT. Let $\beta > \beta_d$, and recall that $x = (p-2)(1-q)/q \in (0, 1)$. The energy identity fixes one weighted integral of the response over the whole row,

$$p \int_0^t C(t, u)^{p-1} R(t, u) du = \frac{e_\beta(t)}{\beta} \rightarrow 1 - d_*,$$

where $d_* = 1 - e_\beta(\infty)/\beta = (1-x)q^p$ by the energy formula (1.4). A recent strip of length A contributes $p \int_0^A c^{p-1} r = 1 - c(A)^p$, as it would for an equilibrium profile. The rest of the row carries the weighted response $c(A)^p - d_*$, while $C(t, \cdot)^p$ rises across it from $C(t, 0)^p \rightarrow 0$ to $C(t, t-A)^p \rightarrow c(A)^p$. It follows that the ratio in (1.10) tends to $1 - d_*/c(A)^p$, and then to x as $A \rightarrow \infty$ (Proposition 12.1). In the same way, we find the exact identity

$$p \int_0^t C(t, u)^{p-1} E(t, u) du = 1 - C(t, 0)^p - \frac{e_\beta(t)}{\beta} \rightarrow d_* > 0.$$

To remove the weight C^{p-1} , we prove the transport inequality of Lemma 12.3: if $z \geq 0$ is absolutely continuous on $[u, v]$, $r_0 \geq 0$, and $J = \int_u^v z^{p-1} r_0$, then $\int_u^v (z' - r_0)_+ \geq [z(v) - (z(u)^p + pJ)^{1/p}]_+$. Applying this on $[S, t-L]$ with $z = C(t, \cdot)$ and $r_0 = R(t, \cdot)$, we find that, in the limit $t \rightarrow \infty$, the positive part of the defect has integral at least $m_L = c(L) - [c(L)^p - d_*]^{1/p}$ there. As $L \rightarrow \infty$, m_L increases to $m_\infty = q(1-x^{1/p})$ (Theorem 12.4). This lower bound holds on an interval whose length is of order t , and by the Cauchy-Schwarz inequality, we have $\liminf_{t \rightarrow \infty} t \int_0^t E(t, u)_+^2 du \geq m_\infty^2$. This is the coexistence of small pointwise violation and integrated violation over an expanding interval described by Cugliandolo, Dean, and Kurchan [54].

2.8 Three-time geometry, susceptibility, and the Fisher cost

In the cubic model, we use more of the tree geometry than the energy ceiling. Throughout this section, $p = 3$.

Saturated tree vectors. In Section 13, which is deterministic, we study unit vectors $v \in \mathcal{H}$ with a small deficit $\delta(v) = F_* - f(v)$. Using the retained defects, we bound weighted sums of the variations of the height amplitudes $\|\Pi_j v\|$ and of the angular errors between each component $\Pi_j v$ and the graft $\Pi_j \mathcal{I}(v \otimes v)$ of the lower components. We deduce that the root operator $\mathcal{I}^* v$, viewed as a Hilbert–Schmidt operator on $\mathcal{H}$, is then close to an operator whose coefficients in the normalized height directions depend only on the height masses (Lemma 13.1). In the coordinate $x \in (0, 1)$ given by the cumulative mass, these operators converge to one universal operator. More precisely, if $\|v_n\| = 1$ and $f(v_n) \rightarrow F_*$, there exist isometric embeddings V_n of $\mathcal{H}$ into $L^2(0, 1) \oplus \mathcal{K}$, where $\mathcal{K}$ is a fixed auxiliary Hilbert space, such that $V_n v_n = \mathbf{1}$ and

$$V_n \mathcal{I}^* v_n V_n^* \longrightarrow \mathcal{T} \oplus 0 \quad \text{in Hilbert–Schmidt norm,}$$

where $\mathcal{T}$ is the integral operator with kernel $T(x, y) = (2 \max(x, y))^{-1/2}$ (Theorem 13.2). The operator $\mathcal{T}$ is positive and trace class, with $\text{tr } \mathcal{T} = \sqrt{2}$ and $\langle \mathbf{1}, \mathcal{T} \mathbf{1} \rangle = F_*$. For two vectors with nonnegative inner products at every height, the same defects show that little of their overlap is carried by heights at which a smoothed cosine between their prefixes is far from one (Lemma 13.4). For three such vectors, the prefixes that carry the overlaps of two of the pairs are nested, which leads to the product inequality of Theorem 13.5,

$$\langle v_1, v_2 \rangle \geq \langle v_1, v_3 \rangle \langle v_2, v_3 \rangle - \omega\left(\max_i \delta(v_i)\right), \quad \lim_{\delta \downarrow 0} \omega(\delta) = 0.$$

Three times. Let $\beta \geq \beta_d$, including $\beta = \beta_d$, where $q = 1/2$. To apply these results to the CK solution, we use the normalized centers $U_{s,D} = M_{s,D} / \|M_{s,D}\|$ of a fixed smoothing age D , where $M_{s,D} = P_{s-D} v_s$. Their components are nonnegative, and by Proposition 9.5 their deficits are small when D is large. The obstacle is to relate their overlaps to the correlations. A center is close to the average $D^{-1} \int_{s-D}^s v_u \, du$ of the trajectory over the preceding window (Lemma 14.1), and the overlap of two such averages is close to $C(t, s)$ when the derivatives of C are small far from the diagonal, which is what Section 11 provides. We find that $\langle U_{t,D}, U_{s,D} \rangle \approx C(t, s)/q$ for late pairs whose distance is large compared with D (Proposition 14.2), and the product inequality becomes Theorem 14.3: for late triples $u < s < t$ with diverging gaps,

$$qC(t, u) \geq C(t, s)C(s, u) - o(1).$$

This is a one-sided version of the spherical product relation of Cugliandolo and Kurchan [52, arXiv version 1, Section 5.1, Equation (5.1.8)], which we prove for the actual CK solution without any assumption on the time scale of aging. It implies composition bounds for the last times at which the correlation crosses given levels (Corollary 14.4).

The susceptibility. For the bound on $\chi(t, 0)$, we use the root operator of a center. For $p = 3$, the root frame (7.6) reads

$$\mathcal{I}^* P_T v_t = \beta \sqrt{3} \int_0^t R(t, u) |P_T v_u\rangle \langle P_T v_u| \, du,$$

so that its trace is a weighted response mass. We take $t_n \rightarrow \infty$ along which $\chi(t_n, 0)$ tends to its lower limit, ages $L_n \rightarrow \infty$, and the normalized centers U_n of $M_{T_n}(L_n)$, where $T_n = t_n - L_n$; these centers approach the ceiling. We split their root operators into an old part A_n , the integral over $u \leq T_n$, and a recent part B_n . On recent windows of fixed length, the projected states $P_{T_n} v_u$ are close to $\sqrt{q} U_n$ by reflection, and in the coordinates of the universal operator, B_n converges to $\beta \sqrt{3} \sqrt{q} (1 - q) |\mathbf{1}\rangle \langle \mathbf{1}|$, which equals $2^{-1/2} |\mathbf{1}\rangle \langle \mathbf{1}|$ by marginality. By Theorem 13.2, the old part then converges in Hilbert–Schmidt norm to $\mathcal{T} - 2^{-1/2} |\mathbf{1}\rangle \langle \mathbf{1}|$, a positive operator with trace $2^{-1/2}$.

However, convergence in Hilbert–Schmidt norm does not give convergence of the trace. To overcome this, we rely on the lower semicontinuity of the trace on positive operators, after replacing each old term $|v_u\rangle \langle v_u|$ by $|M_{u,D}\rangle \langle M_{u,D}|$, whose norm converges. By causality, the errors on time strips at distance at least D from each other are orthogonal (Lemma 14.6), and their total is controlled by the squared remote response, which vanishes by (2.11). The trace of the smoothed old part is asymptotically $\beta \sqrt{3} q^{-1/2} c(2D)$ times the old response mass $\chi(t_n, 0) - (1 - q)$, and comparing it with $2^{-1/2}$, we find that

$$\liminf_{t \rightarrow \infty} \chi(t, 0) \geq (1 - q) + \frac{q(1 - q)}{c(2D)} \rightarrow 2(1 - q) \quad (D \rightarrow \infty),$$

which is Theorem 14.7. This holds for every $\beta \geq \beta_d$, including the endpoint, but we have no matching upper bound.

A Fisher cost. In Section 15, we combine the same Fisher majorant a_F from (2.3) with the product inequality. Let $\beta > \beta_d$, so that $q > 1/2$, and consider windows $[s_n, t_n]$ with $s_n \rightarrow \infty$, $t_n - s_n \rightarrow \infty$, and $C(t_n, s_n) \rightarrow z \in (0, q)$, such that

$$\varepsilon_n^2 = (t_n - s_n) \int_{s_n}^{t_n} a_F(u) \, du \rightarrow \varepsilon^2 \leq 1.$$

By the Cauchy–Schwarz inequality and $|E(w, u)|^2 \leq a_F(u)$, the integrated backward defect on every subinterval of the window is at most ε_n ; in particular, the response mass of the window is $1 - C(t_n, s_n)$ up to ε_n . Integrating the backward response equation (4.21) over the window, we find that $1 - \ell(1 - C(t_n, s_n)) + 3\beta^2 I_n$ is at most of order ε_n in the limit, where

$$I_n = \int_{s_n}^{t_n} [1 - C(u, s_n)^2] R(t_n, u) \, du.$$

In the bulk of the window, we have $C(t_n, u)C(u, s_n) \leq qC(t_n, s_n) + o(1)$ by the product inequality. As the row $u \mapsto C(t_n, u)$ has derivative $R(t_n, u) + E(t_n, u)$, a clipped primitive

turns this into an explicit lower bound for I_n , and we compute the recent strip from stationary convergence. With the calibration $6\beta^2 q(1-q)^2 = 1$ and $\ell = (1-q)^{-1} + 3\beta^2(1-q^2)$, the two estimates are compatible only if

$$3\beta^2(2q-1)(q-z)^2 \leq (\ell + 9\beta^2)\varepsilon + 6\beta^2\varepsilon^2$$

(Lemma 15.3), and the left side is positive because $q > 1/2$. This strict margin vanishes at $\beta = \beta_d$, which is why the positive Fisher cost is restricted to $\beta > \beta_d$, even though the product inequality includes the endpoint. For $\varepsilon = 0$, this is impossible, which excludes every limit $z \in (0, q)$ of $C(t_n, s_n)$ along windows on which $\tau \int a_F$ tends to zero. Limits above q are excluded by the remote envelope (14.6). We also exclude a zero limit: the first crossing of a fixed level $z' \in (0, q)$ occurs inside the window, at a lag that tends to infinity, and the crossing windows have no larger Fisher cost, so that the obstruction applies to them. Every such window therefore ends at the plateau. At the first passage below a level $c \in (0, q)$, the inequality bounds ε from below, which gives (1.15) (Theorem 15.1).

2.9 The stationary power law

For every $p \geq 3$, let $\beta \geq \beta_d$. In Section 16, we study the scalar equation (2.5) for the selected profile, using only this equation with its plateau identity, marginality, and complete monotonicity. The function $u(z) = [c((1-q)z) - q]/(1-q)$ satisfies (Lemma 16.3)

$$u' + u + M(u) * u' = 0, \quad u(0) = 1,$$

where $*$ is the convolution on $[0, \infty)$ and M is a polynomial with nonnegative coefficients such that $M(y) = y + \vartheta y^2 + O(y^3)$, with $\vartheta = (p-2)(1-q)/(2q)$. For $p = 3$, we have $M(y) = y + \vartheta y^2$. The linear coefficient of M is exactly one, by marginality. For a pure power $w(t) = t^{-a}$, the beta integral gives

$$(w * w)' = \frac{\Gamma(1-a)^2}{\Gamma(1-2a)} w^2,$$

so that the exponent relation (1.16) says that the two quadratic terms of the equation balance for w . This is the classical exponent relation of mode-coupling theory [81, Section 4, p. 881], in its version for the aging state [34, Equation (IV.6)] and [96]. As the balance is quadratic, it fixes the exponent but not the amplitude. The proof of an amplitude has three further steps.

Comparison. Using a comparison principle (Lemma 16.4), we turn the sign of the residual of an explicit profile into an inequality between that profile and u . For the shifted power $g(t) = (1+t)^{-a}$, the choice of a makes the quadratic part of the residual positive but smaller than g^3 . The sign of the residual is then decided by cubic terms. We fix it by

correcting g with a small multiple of g^2 (Lemmas 16.5 and 16.6). Our comparison profiles keep u itself up to a fixed time and continue it by a rescaled corrected power, which absorbs the memory of the initial segment. We obtain $A_-(1+t)^{-a} \leq u(t) \leq A_+(1+t)^{-a}$ with constants $0 < A_- \leq A_+ < \infty$ (Proposition 16.7). These bounds determine the exponent, but they still allow $t^a u(t)$ to oscillate between A_- and A_+ on a logarithmic time scale.

A logarithmic Dirichlet form. The convolution square $u * u$ satisfies the quadratic balance $(u * u)' = \vartheta u^2 + \mathcal{E}$, where the error $\mathcal{E}$ is integrable against the weight $t^{-(1-2a)}$ (Lemma 16.8). We integrate the balance against this weight and compare the result with the Schur test for a homogeneous kernel, keeping its exact remainder. We find that the function $D(x) = e^{ax} u(e^x)$ has finite energy for an explicit Dirichlet form on the line (Lemma 16.9), which implies that $D(x+y) - D(x) \rightarrow 0$ as $x \rightarrow \infty$, for every fixed y . This still allows D to vary slowly without converging.

A renewal equation. We take the Laplace transform of the balance at the point e^{-x} . Since the product of independent gamma and beta variables with suitable parameters has a gamma distribution (Lemma 16.13), the transformed balance becomes a linear renewal equation for a smoothed version of D^2 . Its source is integrable: the variance part of the source is controlled by the Dirichlet form, and the rest by $\mathcal{E}$. Using the key renewal theorem [68, Chapter XI, Section 1], with a renewal density that we control near zero and at infinity (Lemma 16.12), we show that the smoothed function converges. Its difference from D^2 tends to zero by the second step, so that D converges, and the bounds of the first step make the limit positive and finite. This proves that $c(t) - q \sim A_\beta t^{-a}$. Finally, complete monotonicity and the monotone density theorem [30, Section 1.7.3] give the asymptotics of every fixed derivative, without differentiating an asymptotic relation (Theorem 16.1). The tools are classical, and the structural theory of these memory equations is developed in [71, 84, 114]; we apply them to the stationary profile selected by the CK solution.

2.10 Dependencies between the sections

Table 1.1 lists, for each main theorem, the section in which it is proved and the main results that its proof uses. We add three remarks on the order of reading. First, for Theorem 1.1, a reader interested in $p \in \{3, 4\}$ may read Section 9 and skip Section 10, and a reader interested in general p may skip Section 9, since Section 10 does not use the tree ceilings. For Theorem 1.2 with $p \neq 3$ and a positive plateau, Section 11.5 uses the forks and the overlap penalty of Section 10. Second, Section 13 uses only the tree space and the cubic ceiling with its retained defects. Third, Part V uses only Propositions 5.7 and 6.1, Theorem 8.5, and Corollary 7.7, and the power law is not used in the proofs of

the other main theorems; it serves only to verify the power-law hypotheses in the last assertions of Corollaries 11.6 and 11.7.

PART I

The Cugliandolo–Kurchan system and its
stationary limits

The model and the hard-spherical Cugliandolo–Kurchan equations

In this section, we fix the model, the dynamics, and our conventions, and state the hard-spherical Cugliandolo–Kurchan (CK) equations together with the published results that connect them with the Langevin dynamics. We prove three statements. Proposition 3.3 transfers Sellke’s convergence theorem for the hard-spherical diffusion on fixed time horizons [112, arXiv version 5, Theorem 4.3, p. 23] to our normalization and to a uniform initial point. Lemma 3.4 identifies the Brownian observable in that theorem with the integrated response to an additive force. Lemma 3.6 collects bounds on the Hamiltonian and its derivatives that are uniform in the dimension. This section also uses the variational-response estimates from Steps 1 and 2 of the proof of Proposition 4.1, which depend only on Lemma 3.6: Remark 3.1 uses the actual L^2 derivative and the quenched moment estimate (4.3), and the proof of Lemma 3.4 uses (4.4).

These results are the finite-dimensional input of the paper, and we use them throughout Section 4. The landscape bounds enter again in the entropy estimates of Section 5 and in the branched trajectories of Section 10, and we use the bridge wherever we obtain a statement about the CK solution as a limit of the diffusion, in Sections 5, 10, and 11.

Throughout the paper, $p \geq 3$ and $\beta > 0$ are fixed, and $\xi(q) = q^p$.

3.1 The Langevin dynamics

For $N \geq 2$, let $S_N = \{x \in \mathbb{R}^N : \|x\|^2 = N\}$, and let ω_N be the uniform probability measure on S_N . The disorder is an ordered coefficient tensor $J = (J_{i_1 \dots i_p})_{i_1, \dots, i_p=1}^N$ whose N^p entries are independent standard Gaussian random variables. The Hamiltonian is

$$H_N(x) = N^{-(p-1)/2} \sum_{i_1, \dots, i_p=1}^N J_{i_1 \dots i_p} x_{i_1} \cdots x_{i_p}, \quad x \in \mathbb{R}^N. \quad (3.1)$$

Since the entries of J are independent and the tensor is not symmetrized, we have

$$\mathbb{E} H_N(x) H_N(y) = N^{-(p-1)} \sum_{i_1, \dots, i_p=1}^N \prod_{j=1}^p x_{i_j} y_{i_j} = N^{-(p-1)} (x \cdot y)^p = N \xi\left(\frac{x \cdot y}{N}\right) \quad (3.2)$$

for all $x, y \in \mathbb{R}^N$. No factor $p!$ appears, because the coefficients are ordered.

We write ∇H_N and $\nabla^2 H_N$ for the Euclidean gradient and Hessian. For $x \in \mathbb{R}^N$, let $P_x = I - xx^\top / N$. For $x \in S_N$, the matrix P_x is the orthogonal projection onto the tangent space $T_x S_N = \{v \in \mathbb{R}^N : v \cdot x = 0\}$. We write ∇_S , ∇_S^2 , and Δ_S for the gradient, the Riemannian Hessian, and the Laplace–Beltrami operator of S_N , with the metric induced from $\mathbb{R}^N$. Note that $\nabla_S H_N(x) = P_x \nabla H_N(x)$. We also write

$$\kappa_N = \frac{N-1}{N}.$$

We consider the diffusion

$$dX_t = \beta P_{X_t} \nabla H_N(X_t) dt + \sqrt{2} P_{X_t} dB_t - \kappa_N X_t dt, \quad (3.3)$$

where B is a standard Brownian motion in $\mathbb{R}^N$. The initial point X_0 has law ω_N , and X_0 , J , and B are independent. The equation has a unique strong solution up to its explosion time, because its coefficients are polynomials. The solution stays on S_N . By Itô's formula, $Z_t = \|X_t\|^2 - N$ satisfies a linear equation $dZ_t = Z_t dA_t$ for a continuous semimartingale A , because $x \cdot P_x v = -(\|x\|^2 - N)(x \cdot v)/N$ and $2 \operatorname{tr}(P_x^2) - 2\kappa_N \|x\|^2$ vanishes on S_N ; since $Z_0 = 0$, we have $Z \equiv 0$, and there is no explosion.

For a smooth function F on $\mathbb{R}^N$ and $x \in S_N$, Itô's formula shows that the generator of (3.3) is

$$\mathcal{L}_N F(x) = \operatorname{tr}(P_x \nabla^2 F(x)) - \kappa_N x \cdot \nabla F(x) + \beta \nabla_S H_N(x) \cdot \nabla F(x). \quad (3.4)$$

The first two terms are $\Delta_S F(x)$, by the classical formula for the Laplacian of a sphere of radius $\sqrt{N}$, and the last one is $\beta \nabla_S H_N \cdot \nabla_S F$. This shows that $\mathcal{L}_N = \Delta_S + \beta \nabla_S H_N \cdot \nabla_S$. Equivalently, $\mathcal{L}_N F = g_N^{-1} \operatorname{div}_S(g_N \nabla_S F)$ with

$$g_N = Z_{N,\beta}^{-1} e^{\beta H_N}, \quad Z_{N,\beta} = \int e^{\beta H_N} d\omega_N,$$

so that the Gibbs measure $g_N d\omega_N$ is reversible. The sign convention is gradient ascent, in the sense that the drift pushes X_t toward large values of H_N , and the limiting energies of this paper are positive. Because the noise amplitude is $\sqrt{2}$, the diffusion part of $\mathcal{L}_N$ is the Laplacian itself.

We write

$$C_N(t, s) = \frac{X_t \cdot X_s}{N}, \quad s, t \geq 0,$$

for the empirical correlation.

Response. We measure the response to an additive tangent force in the drift, without a factor β . Write

$$b(x) = \beta P_x \nabla H_N(x) - \kappa_N x, \quad \sigma_j(x) = \sqrt{2} P_x e_j, \quad j = 1, \dots, N.$$

On each finite horizon, using a measurable representative, the response to a deterministic $h \in L^2_{\text{loc}}([0, \infty); \mathbb{R}^N)$ is the zero-initial-state strong solution of the forced linearized equation

$$dV_t = Db(X_t)[V_t] dt + P_{X_t}h(t) dt + \sum_{j=1}^N D\sigma_j(X_t)[V_t] dB_t^j, \quad V_0 = 0. \quad (3.5)$$

It is constructed on each finite horizon in the space of predictable square-integrable processes. For each fixed waiting time $u \geq 0$ and finite horizon $H \geq u$, the response matrix $U_N(t, u)$, $u \leq t \leq H$, is the continuous adapted matrix whose columns solve the homogeneous variational equation along the actual trajectory X :

$$\begin{aligned} d(U_N(t, u)v) &= Db(X_t)[U_N(t, u)v] dt + \sum_{j=1}^N D\sigma_j(X_t)[U_N(t, u)v] dB_t^j, \\ U_N(u, u)v &= P_{X_u}v, \end{aligned} \quad (3.6)$$

for $v \in \mathbb{R}^N$. The equation is constructed from the N coordinate columns; linearity and pathwise uniqueness identify every direction with their linear combination. The choices on different finite horizons agree on their common time interval, on one event for all vectors. The matrix acts as zero in the normal direction X_u and maps every vector into $T_{X_t}S_N$. For each fixed $H > u$, its terminal values are also the Fréchet derivatives into $L^2(\mathbb{P}_B; \mathbb{R}^N)$ of the nonlinear solution restarted at $X_u + P_{X_u}v$, evaluated at $v = 0$. For this ambient restart, choose smooth compactly supported extensions agreeing with the displayed coefficients on a neighborhood of the sphere; the variational equation along X uses those physical coefficients. This definition needs no simultaneous samplewise smoothness in all starting points. For classical background on smooth spatial stochastic flows, see [101, Section 1, pp. 1022–1023] and [102, Section 4.6, Theorem 4.6.5, p. 173].

For every such deterministic h , the forced variational equation has the Bochner integral representation in $L^2(\mathbb{P}_B; \mathbb{R}^N)$,

$$V_t = \int_0^t U_N(t, u)h(u) du. \quad (3.7)$$

To justify the representation, first take a bounded continuous force on a fixed horizon. Integrating the causal homogeneous response in its forcing time and applying stochastic Fubini gives the forced variational equation; pathwise uniqueness identifies the result with V . The fixed-horizon L^2 estimates make both the forced terminal solution and the kernel integral continuous linear maps of the deterministic force in $L^2([0, T]; \mathbb{R}^N)$. Density of bounded continuous functions extends the identity to every deterministic L^2 force. The same causal-source argument applies to bounded continuous adapted random directions, including the disorder directions used in Proposition 4.1. In words, $U_N(t, u)$

is the homogeneous linearized response to a tangent-projected impulse at time u . We define its annealed normalized trace by

$$\bar{r}_N(t, u) = \frac{1}{N} \mathbb{E} \operatorname{tr} U_N(t, u), \quad 0 \leq u \leq t. \quad (3.8)$$

This expectation is finite by the fourth-moment bound (4.4) below. The finite- N diagonal is $\bar{r}_N(t, t) = \kappa_N$, because $U_N(t, t) = P_{X_t}$ and $\operatorname{tr} P_x = N - 1$ on S_N ; the limiting response will have diagonal value one.

Remark 3.1 (The equilibrium fluctuation–dissipation relation in this convention). In this convention, stationary equilibrium FDT reads $r = -c'$, with no factor of β ; at lag zero, the derivative is taken from the right. To see this, fix J and let $P_\tau^H = e^{\tau \mathcal{L}_N}$. For the Gibbs initial law, the correlation and response at lag τ are

$$c_{N,G}(\tau) = \frac{1}{N} \sum_{i=1}^N \int x_i P_\tau^H x_i g_N \, d\omega_N, \quad r_{N,G}(\tau) = \frac{1}{N} \sum_{i=1}^N \int \nabla_S x_i \cdot \nabla_S P_\tau^H x_i g_N \, d\omega_N.$$

The second formula is the trace of the mean of U_N for the Gibbs initial law. The actual tangent-projected restart is Fréchet differentiable into L^2 , and the bounded linear expectation operator commutes with that derivative. The Markov restart property identifies the expected response with the expected derivative of $P_{t-u}^H \operatorname{id}$ at the waiting state. Gibbs stationarity then gives the displayed integral, whose integrand is $\sum_i \nabla_S x_i \cdot \nabla_S P_{t-u}^H x_i$ because $\nabla_S x_i = P_x \ell_i$. The fixed-disorder moment estimate (4.3) ensures integrability. Since $\mathcal{L}_N$ is symmetric with respect to $g_N \, d\omega_N$, we have

$$c'_{N,G}(\tau) = \frac{1}{N} \sum_{i=1}^N \int x_i \mathcal{L}_N P_\tau^H x_i g_N \, d\omega_N = -r_{N,G}(\tau).$$

A force written as $\beta P_x h$ would have response $\beta r_{N,G}$, and the relation would read $r = -\beta c'$ in that convention.

3.2 The hard-spherical CK system

By a solution of the hard-spherical CK system, we mean a triple (C, R, λ) of continuous functions, with C defined and symmetric on $[0, \infty)^2$, R defined on the causal triangle $\{(t, s) : 0 \leq s \leq t\}$, and λ defined on $[0, \infty)$, such that, for every $s \geq 0$, the functions

$t \mapsto C(t, s)$ and $t \mapsto R(t, s)$ are continuously differentiable on $[s, \infty)$ and satisfy, for $t \geq s$,

$$\partial_t R(t, s) = -\lambda(t)R(t, s) + \beta^2 \int_s^t \zeta''(C(t, u))R(t, u)R(u, s) \, du, \quad (3.9)$$

$$\begin{aligned} \partial_t C(t, s) = & -\lambda(t)C(t, s) + \beta^2 \int_0^t \zeta''(C(t, u))R(t, u)C(u, s) \, du \\ & + \beta^2 \int_0^s \zeta'(C(t, u))R(s, u) \, du, \end{aligned} \quad (3.10)$$

$$\begin{aligned} \lambda(t) = & 1 + \beta^2 \int_0^t [\zeta'(C(t, u)) + C(t, u)\zeta''(C(t, u))]R(t, u) \, du \\ = & 1 + \beta^2 p^2 \int_0^t C(t, u)^{p-1}R(t, u) \, du, \end{aligned} \quad (3.11)$$

with the boundary data

$$C(t, t) = R(t, t) = 1, \quad t \geq 0. \quad (3.12)$$

At $t = s$, the derivatives are one-sided. The second line of (3.11) holds because $\zeta'(q) + q\zeta''(q) = pq^{p-1} + p(p-1)q^{p-1} = p^2q^{p-1}$.

The correlation equation (3.10) describes C only on the causal side $t \geq s$, and the diagonal requires a word. At $s = t$, the symmetry $C(u, t) = C(t, u)$ and (3.11) turn (3.10) into

$$\partial_t C(t, s)|_{s=t} = -\lambda(t) + (\lambda(t) - 1) = -1. \quad (3.13)$$

The derivative in (3.13) holds the earlier argument fixed and approaches the diagonal from the causal side. The function $t \mapsto C(t, t)$ is constant; its derivative is zero. For the diffusion, Itô's formula gives $d(X_t \cdot X_t)/N = 2X_t \cdot dX_t/N + 2\kappa_N dt$, where the drift part of $2X_t \cdot dX_t/N$ is $-2\kappa_N dt$. Thus the inward drift and the quadratic variation balance exactly at finite N , and their coefficients tend to -2 and $+2$ as $N \rightarrow \infty$. The multiplier λ enforces the limiting balance and keeps $C(t, t) = 1$; this is the meaning of the term 1 in (3.11).

We write

$$e_\beta(t) = \beta p \int_0^t C(t, u)^{p-1}R(t, u) \, du, \quad \chi(t, s) = \int_s^t R(t, u) \, du. \quad (3.14)$$

By (3.11), we have $\lambda(t) = 1 + p\beta e_\beta(t)$, and Proposition 4.1 shows that $e_\beta(t)$ is the limit of the mean energy $N^{-1}\mathbb{E}H_N(X_t)$. The quantity $\chi(t, s)$ is the integrated response, or susceptibility.

We use the following published facts throughout the paper.

Proposition 3.2 (Well-posedness and positivity). *The hard-spherical CK system (3.9)–(3.12) has a unique solution (C, R, λ) . It satisfies*

$$0 \leq C(t, s) \leq 1 \quad (s, t \geq 0), \quad R(t, s) \geq 0 \quad (t \geq s \geq 0), \quad (3.15)$$

and C is a positive-semidefinite kernel: for all $n \geq 1$, times $t_1, \dots, t_n \geq 0$, and $a \in \mathbb{R}^n$, we have $\sum_{i,j=1}^n a_i a_j C(t_i, t_j) \geq 0$.

Existence and uniqueness, the bounds (3.15), and the positive semidefiniteness of C are proved by Dembo, Guionnet, and Mazza [60, arXiv version 1, Proposition 1.1, p. 4, and Lemma 2.1, p. 6] for the hard-spherical equations in the noise-one convention recalled in Section 3.3. Their hypotheses are met by the polynomial $\zeta(q) = q^p$, the initial value $C(0,0) = 1$, and any fixed temperature; their Proposition 1.1 has no restriction to high temperature. Existence and uniqueness in the noise-one convention are also recorded in [112, arXiv version 5, Proposition 4.2, p. 23]. The time change (3.19) below, which is a bijection between solutions of the two systems, transfers all of these properties to (3.9)–(3.12). The positive semidefiniteness of C can also be read off from Proposition 3.3. For fixed times t_i and coefficients a_i , we have

$$\sum_{i,j=1}^n a_i a_j C_N(t_i, t_j) = \frac{1}{N} \left\| \sum_{i=1}^n a_i X_{t_i} \right\|^2 \geq 0,$$

and the left side converges in probability to $\sum_{i,j} a_i a_j C(t_i, t_j)$.

We record two consequences that are used constantly. Because C and R are continuous, each of them is bounded on every compact set of its domain. The right sides of (3.9) and (3.10) are continuous on the causal triangle $\{0 \leq s \leq t\}$, and so are the partial derivatives $\partial_t R$ and $\partial_t C$.

3.3 Comparison of conventions

The factors of two in the CK equations depend on the noise amplitude, on the normalization of time, temperature, and covariance, and on how an external force enters the drift. We compare our convention with the two conventions of the works on which we rely.

The noise-one convention. The works of Ben Arous, Dembo, and Guionnet [24], of Dembo, Guionnet, and Mazza [60], and of Sellke [112] use the noise coefficient one. On S_N , the generator is

$$\frac{1}{2} \Delta_S + \alpha \nabla_S H_N \cdot \nabla_S,$$

with the same covariance $N\zeta(x \cdot y/N)$. Here $\alpha > 0$ is the drift coefficient, which Sellke denotes by β ; his Gibbs measure is proportional to $e^{2\alpha H_N}$ [112, arXiv version 5, Equation (1.2) and the following paragraph, p. 2]. In [60], the dynamics descends a Hamiltonian, and reversing its sign gives the law of H_N . In [24, Equations (1.2), (1.6), and (1.7)], specialized to $\nu(q) = q^p$, the displayed Hamiltonian H^{BAG} has the law of $-2H_N$, and the force in the drift is $-\nabla H^{\text{BAG}}/2$, which has the law of ∇H_N . If x_v is the noise-one diffusion with $\alpha = \beta/2$, then $X_t = x_{2t}$ is our diffusion; this is Step 1 of the proof of Proposition 3.3.

Under this time change, the multiplier of the noise-one system is multiplied by two, its memory coefficient α^2 is multiplied by four, and its noise diagonal one becomes two. This produces exactly (3.9)–(3.11). We do not rescale the covariance $N\zeta$.

The convention of Cugliandolo and Kurchan. In [51, Equations (1)–(4)], the couplings $J_{i_1 \dots i_p}$, $i_1 < \dots < i_p$, have variance $p!/(2N^{p-1})$, so that the Hamiltonian H^{CK} has covariance $N(x \cdot y/N)^p/2$ up to terms that do not affect the limiting equations, and the energy is minimized. The Langevin equation is $\partial_t s = -\beta^{\text{CK}} \nabla H^{\text{CK}}(s) - z(t)s + \eta$, with white noise of variance 2, and the response has diagonal value one. At the level of the limiting CK equations, the covariance of $-\sqrt{2}H^{\text{CK}}$ has the normalization of the covariance of H_N , and the identification is

$$H_N \leftrightarrow -\sqrt{2}H^{\text{CK}}, \quad \beta^{\text{CK}} = \sqrt{2}\beta,$$

with no change of time. With these substitutions, the causal forms of equations (3)–(4) of [51], for $t > s$, coincide with (3.9)–(3.10),¹ their multiplier $z(t) = 1 - p\beta^{\text{CK}}E(t)$ is $\lambda(t)$, and their energy per spin is $E(t) = -e_\beta(t)/\sqrt{2}$. Their equation (7) for the threshold plateau is the marginality relation $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$.

Table 3.1 collects the dictionary.

In our convention, the predicted dynamical transition (β_d, q_d) is the point of the marginality curve $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$ at which $x = (p-2)(1-q)/q$ equals one; see the text after equation (12) and the concluding discussion in [51]. The condition $x = 1$ gives $q_d = (p-2)/(p-1)$. Substituting this value into the marginality relation, we obtain $\beta_d^2 = (p-1)^{p-1}/(p(p-2)^{p-2})$. Equivalently, $\beta_d^2 = 1/(p g(q_d))$ with $g(q) = q^{p-2}(1-q)$. Indeed, we have $g'(q) = q^{p-3}[(p-2) - (p-1)q]$, and g is maximal on $[0, 1]$ at q_d , with $g(q_d) = (p-2)^{p-2}/(p-1)^{p-1}$.

3.4 The bridge on fixed time horizons

The CK equations describe the diffusion (3.3) only after we send the dimension to infinity on a fixed time horizon. Three published results combine to give this limit. Ben Arous, Dembo, and Guionnet proved the limit for the soft-spherical diffusion, including the empirical correlation and the Brownian observable for the integrated response [24, arXiv version 1, Equation (1.9), p. 4, and Theorem 1.2, p. 5]. Dembo, Guionnet, and Mazza proved that the deterministic soft-spherical equations converge to the hard-spherical ones [60, Proposition 1.1 and its proof]. The convergence of the deterministic equations is not by itself a statement about the hard-spherical diffusion. Sellke proved the convergence theorem for the hard-spherical diffusion itself [112, arXiv version 5, Theorem 4.3, p. 23],

¹In arXiv:cond-mat/9303036v1, the factor $C^{p-1}(t, t')$ in the first memory integral of equation (3) should read $C^{p-1}(t, t'')$.

	this paper	noise one [60, 112]	CK [51]
time	t	$v = 2t$	t
covariance of H	Nq^p	Nq^p	$Nq^p/2$
temperature parameter	β	$\alpha = \beta/2$	$\beta^{\text{CK}} = \sqrt{2}\beta$
diffusion part of generator	Δ_S	$\Delta_S/2$	Δ_S
multiplier	$\lambda(t)$	$\mu(v) = \lambda(v/2)/2$	$z(t) = \lambda(t)$
constant term of multiplier	1	1/2	1
memory coefficients	$\beta^2 \xi'', \beta^2 \xi'$	$\alpha^2 \xi'', \alpha^2 \xi'$	$(\beta^{\text{CK}})^2 \xi''/2, (\beta^{\text{CK}})^2 \xi'/2$
noise diagonal	2	1	2
mean energy	$e_\beta(t)$	$e_\beta(v/2)$	$E(t) = -e_\beta(t)/\sqrt{2}$
equilibrium FDT	$r = -c'$	$r = -2c'$	$r = -c'$
transition, $p = 3$	$\beta_d^2 = 4/3$	$\alpha_d^2 = 1/3$	$(\beta_d^{\text{CK}})^2 = 8/3$
transition, general p	β_d	$\alpha_d = \beta_d/2$	$\beta_d^{\text{CK}} = \sqrt{2}\beta_d$
plateau at transition	q_d	q_d	q_d

Table 3.1. The dictionary between the conventions. The CK column refers to the normalization of the limiting CK equations. In each column, r and c denote the stationary response and correlation as functions of the lag measured in that column's time, and $\beta_d^2 = (p-1)^{p-1}/(p(p-2)^{p-2})$ and $q_d = (p-2)/(p-1)$. In the convention of [51], the transition temperature for $p = 3$ is $(3/8)^{1/2}$.

by coupling it with soft-spherical diffusions [112, arXiv version 5, Lemma 3.1, pp. 16–18]. We use Sellke's theorem for the correlation and for the Brownian observable. The next proposition states it in our normalization.

Proposition 3.3 (The bridge on fixed time horizons). *Let (C, R, λ) be the solution of the hard-spherical CK system of Proposition 3.2, and let*

$$C_N(t, s) = \frac{X_t \cdot X_s}{N}, \quad \Gamma_N(t, s) = \frac{X_t \cdot B_s}{\sqrt{2}N}.$$

For every fixed $0 \leq T < \infty$, we have, uniformly on $0 \leq s \leq t \leq T$,

$$C_N(t, s) \longrightarrow C(t, s), \quad \Gamma_N(t, s) \longrightarrow \int_0^s R(t, u) \, du \quad \text{in probability.} \quad (3.16)$$

That is, for every $\varepsilon > 0$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\sup_{0 \leq s \leq t \leq T} \left\{ |C_N(t, s) - C(t, s)| + \left| \Gamma_N(t, s) - \int_0^s R(t, u) \, du \right| \right\} > \varepsilon \right) = 0. \quad (3.17)$$

Both convergences also hold after taking expectations, uniformly on the same triangle. The limiting response satisfies $R(t, t) = 1$.

Proof. The proof has five steps. In Step 1, we write our diffusion as a time change of Sellke’s diffusion. In Steps 2 and 3, we check that the same time change maps Sellke’s CK kernels to the solution of (3.9)–(3.12) and his Brownian observable to Γ_N ; this proves (3.16) for a deterministic initial point. In Step 4, we use a rotation to extend the convergence to the uniform initial point. In Step 5, we prove the convergence of expectations by a uniform integrability bound.

Step 1: the time change. Sellke’s diffusion [112, arXiv version 5, Equation (1.2), p. 2] has noise coefficient one and drift coefficient α (which Sellke denotes by β):

$$dx_v = \alpha P_{x_v} \nabla H_N(x_v) dv + P_{x_v} dW_v - \frac{\kappa_N}{2} x_v dv, \quad (3.18)$$

where W is a standard Brownian motion in $\mathbb{R}^N$ and H_N is the Hamiltonian (3.1). We take $\alpha = \beta/2$. Given our X_0 , J , and B , let

$$W_v = \sqrt{2} B_{v/2}, \quad x_v = X_{v/2}, \quad v \geq 0.$$

Then W is a standard Brownian motion with respect to the filtration $(\mathcal{F}_{v/2})_{v \geq 0}$, where $(\mathcal{F}_t)$ is the filtration of X_0 , J , and B . For a continuous adapted integrand G , we have $\int_0^{v/2} G_r dr = \frac{1}{2} \int_0^v G_{w/2} dw$ and $\sqrt{2} \int_0^{v/2} G_r dB_r = \int_0^v G_{w/2} dW_w$. Inserting these two identities into the integral form of (3.3) at time $v/2$, we obtain

$$x_v = x_0 + \int_0^v \left[\frac{\beta}{2} P_{x_w} \nabla H_N(x_w) - \frac{\kappa_N}{2} x_w \right] dw + \int_0^v P_{x_w} dW_w,$$

which is (3.18) with $\alpha = \beta/2$. Conversely, $X_t = x_{2t}$ and $B_t = W_{2t}/\sqrt{2}$. The disorder and the initial point are unchanged, and they remain independent of W .

Step 2: the CK kernels. Let C^S , R^S , and μ^S be the hard-spherical CK kernels and multiplier of [112, arXiv version 5, Definition 4, pp. 22–23] with parameter α ; by [112, arXiv version 5, Proposition 4.2, p. 23], which rests on [24, arXiv version 1, Theorem 1.2, p. 5] and the proof of [60, Proposition 1.1], they are the unique continuous, locally bounded solution of the following system. In our notation for ξ , and with the later time first as in [112], they solve, for $v \geq w \geq 0$,

$$\begin{aligned} \partial_v R^S(v, w) &= -\mu^S(v) R^S(v, w) + \alpha^2 \int_w^v \xi''(C^S(v, y)) R^S(v, y) R^S(y, w) dy, \\ \partial_v C^S(v, w) &= -\mu^S(v) C^S(v, w) + \alpha^2 \int_0^v \xi''(C^S(v, y)) R^S(v, y) C^S(y, w) dy \\ &\quad + \alpha^2 \int_0^w \xi'(C^S(v, y)) R^S(w, y) dy, \\ \mu^S(v) &= \frac{1}{2} + \alpha^2 p^2 \int_0^v C^S(v, y)^{p-1} R^S(v, y) dy, \end{aligned}$$

with $C^S(v, v) = R^S(v, v) = 1$. We claim that

$$C(t, s) = C^S(2t, 2s), \quad R(t, s) = R^S(2t, 2s), \quad \lambda(t) = 2\mu^S(2t). \quad (3.19)$$

Let $\tilde{C}$, $\tilde{R}$, and $\tilde{\lambda}$ denote the right sides of (3.19). By the uniqueness in Proposition 3.2, it suffices to show that $(\tilde{C}, \tilde{R}, \tilde{\lambda})$ solves (3.9)–(3.12). The boundary data are immediate. For the multiplier, the substitution $y = 2u$ gives

$$\tilde{\lambda}(t) = 1 + 2\alpha^2 p^2 \int_0^{2t} C^S(2t, y)^{p-1} R^S(2t, y) \, dy = 1 + 4\alpha^2 p^2 \int_0^t \tilde{C}(t, u)^{p-1} \tilde{R}(t, u) \, du,$$

which is (3.11) since $4\alpha^2 = \beta^2$. For the response, we have $\partial_t \tilde{R}(t, s) = 2(\partial_v R^S)(2t, 2s)$ by the chain rule, and the substitution $y = 2u$ in the memory integral yields

$$\partial_t \tilde{R}(t, s) = -2\mu^S(2t) \tilde{R}(t, s) + 4\alpha^2 \int_s^t \zeta''(\tilde{C}(t, u)) \tilde{R}(t, u) \tilde{R}(u, s) \, du.$$

This is (3.9). Each memory term acquires the coefficient $4\alpha^2 = \beta^2$, with one factor of two from differentiating in the new time and the other from changing the integration variable. The same computation, applied to the two memory integrals of the correlation equation, yields (3.10). This proves (3.19).

Step 3: the observables and a deterministic initial point. Fix $x_0 \in S_N$, and consider for the moment the diffusion started from the deterministic point $X_0 = x_0$. Sellke’s theorem [112, arXiv version 5, Theorem 4.3, p. 23] concerns the diffusion (3.18) started from a deterministic point of S_N , with the Hamiltonian (3.1). Applied on the horizon $[0, 2T]$, it states that, almost surely,

$$\frac{x_v \cdot x_w}{N} \longrightarrow C^S(v, w), \quad \frac{x_v \cdot W_w}{N} \longrightarrow \int_0^w R^S(v, y) \, dy,$$

uniformly on $0 \leq w \leq v \leq 2T$, for any coupling of the diffusions in different dimensions. In particular, these convergences hold in probability, which is all we use. The hypotheses of the theorem are satisfied, since the Hamiltonian is the pure p -spin Hamiltonian with ordered independent standard Gaussian coefficients, the initial point is deterministic, and the horizon $2T$ and the parameter $\alpha = \beta/2$ are fixed as $N \rightarrow \infty$. By Step 1, we have $C_N(t, s) = x_{2t} \cdot x_{2s}/N$ and

$$\Gamma_N(t, s) = \frac{X_t \cdot B_s}{\sqrt{2}N} = \frac{1}{2} \cdot \frac{x_{2t} \cdot W_{2s}}{N}.$$

By (3.19) and the substitution $y = 2u$, the limit of the latter is

$$\frac{1}{2} \int_0^{2s} R^S(2t, y) \, dy = \int_0^s R^S(2t, 2u) \, du = \int_0^s R(t, u) \, du.$$

The triangle $0 \leq w \leq v \leq 2T$ corresponds to $0 \leq s \leq t \leq T$. This proves (3.17) for the diffusion started from x_0 .

Step 4: the uniform initial point. For $x_0 \in S_N$, let $\delta_N(x_0, \varepsilon)$ be the probability in (3.17) for the diffusion started from x_0 . We claim that $\delta_N(x_0, \varepsilon)$ does not depend on x_0 . Let e_1 be the first coordinate vector, let $b = x_0 / \sqrt{N}$, and let $O = O_{x_0}$ be the Householder reflection

$$O = I - \frac{2(e_1 - b)(e_1 - b)^\top}{\|e_1 - b\|^2} \quad (b \neq e_1), \quad O = I \quad (b = e_1).$$

It is orthogonal, depends measurably on x_0 , and satisfies $O(\sqrt{N}e_1) = x_0$. Let X be the diffusion started from x_0 , and set $Y_t = O^\top X_t$, $B'_t = O^\top B_t$, and $J' = (O^\top)^{\otimes p} J$, that is,

$$J'_{j_1 \dots j_p} = \sum_{i_1, \dots, i_p=1}^N J_{i_1 \dots i_p} O_{i_1 j_1} \cdots O_{i_p j_p}.$$

The matrix $O^{\otimes p}$ is orthogonal on $\mathbb{R}^{N^p}$, so J' is again a tensor of independent standard Gaussian entries. Also, B' is a standard Brownian motion, independent of J' . Let H'_N be the Hamiltonian with coefficients J' ; then $H'_N(y) = H_N(Oy)$. Using $P_{Oy} = OP_y O^\top$ and $\nabla H'_N(y) = O^\top \nabla H_N(Oy)$, and applying $O^\top$ to (3.3), we obtain

$$dY_t = \beta P_{Y_t} \nabla H'_N(Y_t) dt + \sqrt{2} P_{Y_t} dB'_t - \kappa_N Y_t dt, \quad Y_0 = \sqrt{N}e_1.$$

We conclude that (Y, J', B') has the law of the diffusion started from $\sqrt{N}e_1$, together with its disorder and its noise. The matrix O is orthogonal, so that $Y_t \cdot Y_s = X_t \cdot X_s$ and $Y_t \cdot B'_s = X_t \cdot B_s$; in other words, C_N and Γ_N take the same values for both triples. This implies that $\delta_N(x_0, \varepsilon) = \delta_N(\sqrt{N}e_1, \varepsilon)$.

Now let X_0 be uniform and independent of (J, B) . The strong solution is a measurable function of (X_0, J, B) . By independence and Fubini's theorem, the probability in (3.17) is

$$\int_{S_N} \delta_N(x_0, \varepsilon) \omega_N(dx_0) = \delta_N(\sqrt{N}e_1, \varepsilon),$$

which tends to zero by Step 3. This proves (3.17), that is, (3.16).

Step 5: expectations. Since $|X_t \cdot B_s| \leq \sqrt{N} \|B_s\|$, we have

$$\sup_{0 \leq s \leq t \leq T} |\Gamma_N(t, s)| \leq \frac{\sup_{s \leq T} \|B_s\|}{\sqrt{2N}}.$$

By Doob's maximal inequality, $\mathbb{E} \sup_{s \leq T} \|B_s\|^2 \leq 4\mathbb{E} \|B_T\|^2 = 4NT$, and therefore

$$\mathbb{E} \sup_{0 \leq s \leq t \leq T} |\Gamma_N(t, s)|^2 \leq 2T$$

for every N . The function $(t, s) \mapsto \int_0^s R(t, u) du$ is continuous on the triangle, and we let ρ_T be its maximum there. Then the random variables $\sup_{0 \leq s \leq t \leq T} |\Gamma_N(t, s) - \int_0^s R(t, u) du|$

have second moments at most $2(2T + \rho_T^2)$, and in particular they are uniformly integrable. They tend to zero in probability, and uniform integrability implies that their expectations tend to zero. The same holds for $\sup_{0 \leq s \leq t \leq T} |C_N(t, s) - C(t, s)|$, which is bounded by 2. We have shown that

$$\lim_{N \rightarrow \infty} \mathbb{E} \sup_{0 \leq s \leq t \leq T} \left(|C_N(t, s) - C(t, s)|^2 + \left| \Gamma_N(t, s) - \int_0^s R(t, u) \, du \right| \right) = 0, \quad (3.20)$$

where for the square we used $|C_N - C|^2 \leq 2|C_N - C|$. Because $|\mathbb{E}Z| \leq \mathbb{E}|Z|$, this implies the convergence of the expectations, uniformly on the triangle. Finally, $R(t, t) = 1$ is part of the boundary data (3.12). $\square$

3.5 The response as a Brownian primitive

Proposition 3.3 concerns a Brownian observable. The next lemma identifies its mean with the integrated response to an additive force, which is how the Brownian observable is used as a response in [24, arXiv version 1, Equation (1.9), p. 4, and Theorem 1.2, p. 5] and in [112, arXiv version 5, Theorem 4.3, p. 23]. The identification is the classical connection between linear response and Gaussian integration by parts. The original Wiener translation formulas of Cameron and Martin [37, Theorems 1–2, pp. 386–387] assume a shift whose derivative has bounded variation and use Wiener paths with variance $t/2$; for a modern deterministic-shift formulation, see [90, Exercise 4.29, p. 27, and Theorem 4.44, p. 31]. We use the backward Itô equation for the actual conditional mean, which fixes the tangent projection and the normalization of the noise directly.

Lemma 3.4 (Brownian primitive of the averaged force response). *Let $U_N(t, u)$ be the pathwise response matrix (3.6), and let $\bar{r}_N(t, u) = N^{-1} \mathbb{E} \operatorname{tr} U_N(t, u)$ as in (3.8). Then, for $0 \leq s \leq t$,*

$$\int_0^s \bar{r}_N(t, u) \, du = \mathbb{E} \Gamma_N(t, s) = \frac{\mathbb{E}[X_t \cdot B_s]}{\sqrt{2} N}. \quad (3.21)$$

The same identity holds with the disorder and the initial point fixed, and with the expectation taken over the Brownian motion only. At equal times, $\bar{r}_N(t, t) = (N - 1)/N$.

Proof. Fix the disorder J and the initial point x_0 , and write $\mathbb{E}_B$ for expectation over B . Put $m_\tau = P_\tau^H \operatorname{id}$, the mean of the actual solution started from a sphere point, and

$$F_u = m_{t-u}(X_u), \quad 0 \leq u \leq t.$$

The mean has the spatial and time regularity needed for the backward Kolmogorov equation. Applying Itô's formula along the actual sphere trajectory, the drift vanishes and

$$dF_u = \sqrt{2} D m_{t-u}(X_u) P_{X_u} dB_u.$$

The predictable integrand is square-integrable on this fixed horizon, $F_0 = m_t(x_0)$ is deterministic, and $F_t = X_t$. The actual L^2 derivative of the projected restart, followed by the bounded linear expectation operator, and the Markov restart identity give

$$\mathbb{E}_B U_N(t, u) = \mathbb{E}_B [Dm_{t-u}(X_u) P_{X_u}]$$

at every fixed deterministic u . These are statements about the constructed solution and its variational equation, rather than assumed response identities.

For $0 \leq s \leq t$, the Itô product formula, or the covariance identity for square-integrable stochastic integrals, yields

$$\mathbb{E}_B [X_t^i B_s^i] = \sqrt{2} \int_0^s \mathbb{E}_B [Dm_{t-u}(X_u) P_{X_u}]_{ii} \, du.$$

The deterministic initial mean contributes zero because $\mathbb{E}_B B_s^i = 0$. Equivalently, the coordinate form is

$$\int_0^s \mathbb{E}_B [U_N(t, u)]_{ii} \, du = \frac{1}{\sqrt{2}} \mathbb{E}_B [X_t^i B_s^i]. \quad (3.22)$$

Summing over i and dividing by N , we obtain the quenched form of (3.21):

$$\int_0^s \frac{1}{N} \mathbb{E}_B \operatorname{tr} U_N(t, u) \, du = \frac{\mathbb{E}_B [X_t \cdot B_s]}{\sqrt{2} N}. \quad (3.23)$$

The right side is integrable over the disorder and initial point, since its absolute value is bounded by $\sqrt{s/2}$. The normalized trace on the left has a fourth moment bounded uniformly on each fixed causal window by (4.4). Fubini's theorem therefore permits averaging over the independent disorder and initial point, giving (3.21). Finally, $\bar{r}_N(t, t) = (N-1)/N$ because $U_N(t, t) = P_{X_t}$ and $\operatorname{tr} P_{X_t} = N-1$. $\square$

Combining Proposition 3.3 and Lemma 3.4, we find that, for every $0 \leq T < \infty$,

$$\lim_{N \rightarrow \infty} \sup_{0 \leq s \leq t \leq T} \left| \int_0^s \bar{r}_N(t, u) \, du - \int_0^s R(t, u) \, du \right| = 0. \quad (3.24)$$

In particular, the susceptibility $\chi(t, s)$ of (3.14) is the limit of the integrated response $\int_s^t \bar{r}_N(t, u) \, du = \mathbb{E}[X_t \cdot (B_t - B_s)] / (\sqrt{2} N)$, uniformly on $0 \leq s \leq t \leq T$.

Remark 3.5. Proposition 3.3 and Lemma 3.4 give the convergence (3.24) of the primitives of the annealed response. Primitive convergence alone does not imply pointwise convergence of $\bar{r}_N$. The published convergence theorem supplies this statement about primitives, which suffices for the integrated uses in this paper. Where a response is integrated against a continuous weight, as in the proof of Proposition 4.1, we may pass to the limit through the primitives, using the bound on the total variation of $\bar{r}_N(t, u) \, du$ supplied by (4.4).

The annealed means satisfy the stronger conclusion

$$\sup_{0 \leq s \leq t \leq T} |\bar{r}_N(t, s) - R(t, s)| \longrightarrow 0 \quad (N \rightarrow \infty), \quad 0 \leq T < \infty.$$

Indeed, the projected variational equation, its time-increment estimates, and the uniform exponential moments of the landscape give a common modulus of continuity for the annealed traces on each fixed causal time window. Together with (4.4), this gives precompactness by the Arzelà–Ascoli theorem. Every uniformly convergent subsequence has the same primitives as R by (3.24); continuity and uniqueness of a function from its primitives identify its limit with R . Precompactness then gives convergence along the whole sequence. This conclusion concerns the deterministic annealed normalized traces; we assert no convergence or concentration of the random response matrices U_N .

3.6 Landscape bounds

For later estimates, we define the normalized multilinear norm of the coefficient tensor,

$$Q_N = \frac{1}{\sqrt{N}} \sup_{\|v_1\|=\dots=\|v_p\|=1} \left| \sum_{i_1, \dots, i_p=1}^N J_{i_1 \dots i_p} v_1^{i_1} \cdots v_p^{i_p} \right|, \quad (3.25)$$

where the supremum is over unit vectors $v_1, \dots, v_p \in \mathbb{R}^N$. Closely related uniform tensor estimates are in [24, arXiv version 1, Appendix B, Lemma B.1 and Equations (B.5)–(B.7), pp. 27–29]. The next lemma shows that Q_N is of order one, with uniform moments and an exponentially small probability of being large, and that it controls the Hamiltonian and its first two derivatives on S_N .

Lemma 3.6 (Uniform landscape moments). *For every fixed $0 \leq k < \infty$, we have $\sup_N \mathbb{E} Q_N^k < \infty$, and for every fixed $b < \infty$, we have $\sup_N \mathbb{E} e^{bQ_N} < \infty$. There exist constants $A_p, c_p > 0$, depending only on p , such that $\mathbb{P}(Q_N > A_p) \leq e^{-c_p N}$ for all $N \geq 2$. Moreover, on S_N ,*

$$\begin{aligned} \frac{|H_N|}{N} &\leq Q_N, & \frac{\|\nabla H_N\|}{\sqrt{N}} &\leq pQ_N, & \|\nabla^2 H_N\|_{\text{op}} &\leq p(p-1)Q_N, \\ \frac{|\Delta_S H_N|}{N} &\leq p^2 Q_N, & \|\nabla_S^2 H_N\|_{\text{op}} &\leq p^2 Q_N. \end{aligned} \quad (3.26)$$

Because $\nabla_S H_N = P_x \nabla H_N$ and $\|P_x\|_{\text{op}} \leq 1$, the gradient bound implies $\|\nabla_S H_N\| \leq pQ_N \sqrt{N}$ on S_N .

Proof. The proof has three steps. We first use a net argument to reduce Q_N to finitely many Gaussian variables, then deduce a tail bound, and finally prove the derivative bounds. We write $J(v_1, \dots, v_p) = \sum_{i_1, \dots, i_p} J_{i_1 \dots i_p} v_1^{i_1} \cdots v_p^{i_p}$ for the multilinear form of J , and

$$M_N = \sup_{\|v_1\|=\dots=\|v_p\|=1} |J(v_1, \dots, v_p)| = \sqrt{N} Q_N.$$

Step 1: a net and the telescoping bound. Let $\varepsilon = 1/(2p)$, and let $\mathcal{N}$ be a maximal ε -separated subset of the unit sphere of $\mathbb{R}^N$. By maximality, every unit vector is within distance ε of $\mathcal{N}$. The balls of radius $\varepsilon/2$ around the points of $\mathcal{N}$ are disjoint and contained in the ball of radius $1 + \varepsilon/2$. Comparing volumes gives

$$|\mathcal{N}| \leq \left(\frac{1 + \varepsilon/2}{\varepsilon/2} \right)^N = \left(1 + \frac{2}{\varepsilon} \right)^N = (1 + 4p)^N.$$

Let M_N^{net} be the maximum of $|J(w_1, \dots, w_p)|$ over $w_1, \dots, w_p \in \mathcal{N}$. Given unit vectors $v_1, \dots, v_p$, choose $w_k \in \mathcal{N}$ such that $\|v_k - w_k\| \leq \varepsilon$. By multilinearity, we have the telescoping identity

$$J(v_1, \dots, v_p) - J(w_1, \dots, w_p) = \sum_{k=1}^p J(w_1, \dots, w_{k-1}, v_k - w_k, v_{k+1}, \dots, v_p).$$

In the k th term, all arguments except $v_k - w_k$ are unit vectors, and by homogeneity its absolute value is at most $\|v_k - w_k\| M_N \leq \varepsilon M_N$. Summing over k , we have $|J(v_1, \dots, v_p)| \leq M_N^{\text{net}} + p\varepsilon M_N$. Taking the supremum over $v_1, \dots, v_p$, and using that M_N is finite, we conclude that

$$M_N \leq (1 - p\varepsilon)^{-1} M_N^{\text{net}} = 2M_N^{\text{net}}.$$

Step 2: the tail bound and the moments. For fixed unit vectors $w_1, \dots, w_p$, the variable $J(w_1, \dots, w_p)$ is a centered Gaussian with variance $\sum_{i_1, \dots, i_p} (w_1^{i_1} \cdots w_p^{i_p})^2 = \prod_k \|w_k\|^2 = 1$, because the entries of J are independent. By the Gaussian tail bound, we have $\mathbb{P}(|J(w_1, \dots, w_p)| > y) \leq 2e^{-y^2/2}$ for $y \geq 0$. The product net $\mathcal{N}^p$ has at most $(1 + 4p)^{pN}$ elements. By Step 1 and a union bound, for $x \geq 0$,

$$\mathbb{P}(Q_N > x) \leq \mathbb{P}\left(M_N^{\text{net}} > \frac{x\sqrt{N}}{2}\right) \leq 2 \exp\left(N\kappa_p - \frac{Nx^2}{8}\right), \quad \kappa_p = p \log(1 + 4p).$$

Let $A_p = 4\sqrt{\kappa_p}$. For $x \geq A_p$, the inequality $x^2/16 \geq \kappa_p$ gives

$$\mathbb{P}(Q_N > x) \leq 2e^{-Nx^2/16} \leq 2e^{-x^2/8}, \quad x \geq A_p, \quad N \geq 2. \quad (3.27)$$

At $x = A_p$, the first bound in (3.27) is $2e^{-N\kappa_p} \leq e^{-c_p N}$ with $c_p = \kappa_p - \frac{1}{2} \log 2 > 0$, because $N \geq 2$. For $k > 0$ and $b \geq 0$, we write the moments as integrals of the tail and use (3.27) beyond A_p :

$$\mathbb{E}Q_N^k \leq A_p^k + \int_{A_p}^{\infty} kx^{k-1} 2e^{-x^2/8} dx, \quad \mathbb{E}e^{bQ_N} \leq e^{bA_p} + \int_{A_p}^{\infty} be^{bx} 2e^{-x^2/8} dx.$$

Both right sides are finite and do not depend on N . The case $k = 0$ is immediate, and for $b < 0$, we have $e^{bQ_N} \leq 1$. This proves the assertions about the moments and the tail.

Step 3: the derivative bounds. We write a point of S_N as $x = \sqrt{N}v$ with $\|v\| = 1$. Then $H_N(x) = N^{-(p-1)/2}J(x, \dots, x) = \sqrt{N}J(v, \dots, v)$ and $|H_N(x)| \leq \sqrt{N}M_N = NQ_N$. Each derivative of H_N replaces one copy of x by a test direction. For a unit vector w , we have

$$\nabla H_N(x) \cdot w = N^{-(p-1)/2} \sum_{k=1}^p J(x, \dots, x, \underbrace{w}_k, x, \dots, x),$$

a sum of p terms, each of absolute value at most $N^{-(p-1)/2}N^{(p-1)/2}M_N = M_N$. Summing these p bounds gives $\|\nabla H_N(x)\| \leq pM_N = p\sqrt{N}Q_N$. Similarly, $\nabla^2 H_N(x)(w, w')$ is $N^{-(p-1)/2}$ times a sum of $p(p-1)$ terms in which two of the slots carry the unit vectors w, w' and the others carry x . Each term is at most $N^{(p-2)/2}M_N$ in absolute value, and the bound $|\nabla^2 H_N(x)(w, w')| \leq p(p-1)N^{-1/2}M_N = p(p-1)Q_N$ follows. This proves the first three inequalities in (3.26).

For the intrinsic Hessian, let $w \in T_x S_N$ be a unit vector, and let $\gamma(r) = x \cos(r/\sqrt{N}) + \sqrt{N}w \sin(r/\sqrt{N})$ be the great circle through x in the direction w . It is a unit-speed geodesic of S_N with $\gamma''(0) = -x/N$, and we have

$$\nabla_S^2 H_N(x)(w, w) = \frac{d^2}{dr^2} H_N(\gamma(r)) \Big|_{r=0} = \nabla^2 H_N(x)(w, w) - \frac{x \cdot \nabla H_N(x)}{N}.$$

The polynomial H_N is homogeneous of degree p , and Euler's identity reads $x \cdot \nabla H_N(x) = pH_N(x)$. By polarization, we obtain, as quadratic forms on $T_x S_N$,

$$\nabla_S^2 H_N(x) = P_x \nabla^2 H_N(x) P_x - \frac{pH_N(x)}{N} P_x. \quad (3.28)$$

By the bounds already proved, $\|\nabla_S^2 H_N\|_{\text{op}} \leq p(p-1)Q_N + pQ_N = p^2Q_N$. Finally, $\Delta_S H_N(x)$ is the trace of $\nabla_S^2 H_N(x)$ on the $(N-1)$ -dimensional space $T_x S_N$, which gives $|\Delta_S H_N(x)| \leq (N-1)p^2Q_N \leq Np^2Q_N$. This proves the last two inequalities. $\square$

Structural identities and uniform estimates

In this section, we prove the structural identities of the CK solution and the estimates on it that hold uniformly in time. Proposition 4.1 shows that the CK integral $e_\beta(t)$ is the limit of the mean energy of the diffusion, and that $\lambda(t) = 1 + p\beta e_\beta(t)$. Proposition 4.2, the Green covariance identity, expresses the correlation as an explicit quadratic functional of the response. We realize it by unit vectors Z_t in a Hilbert space. From these, we deduce an energy–response tradeoff and the bound $1 \leq \lambda \leq \Lambda$ (Proposition 4.3), Lipschitz and Hölder bounds in time (Proposition 4.4), an equation for the response in its earlier time (Proposition 4.5), and the domination $R \leq C$ together with the compactness of the CK solution along late time windows (Proposition 4.6). All constants are uniform over the half-line $[0, \infty)$.

We use the results of Section 3. Proposition 4.1 uses Proposition 3.3 and Lemmas 3.4 and 3.6, and Proposition 4.4 uses Propositions 3.3 and 4.1 and Lemma 3.6. All other proofs in this section are deterministic, and they use only the CK equations (3.9)–(3.12) and the published facts of Proposition 3.2. The results of this section are used throughout the rest of the paper. The bounds and the compactness enter the construction of stationary limits in Section 5. The Green covariance identity is the starting point of the tree realization in Section 7 and of the response estimates in Section 11. The energy identity identifies the energy of a stationary cluster in Sections 5 and 8. We use the backward response equation in Sections 11 and 15, and the domination $R \leq C$ transfers the forgetting of correlations to the response in Section 8.

The Green covariance identity is the variation-of-constants formula for the linear effective process of dynamical mean-field theory, in which a single coordinate is driven by its own response kernel, by white noise, and by a Gaussian noise with covariance $\beta^2 \xi'(C)$. The single-coordinate retarded equation and its colored Gaussian noise are described in [49, arXiv version 2, Section 9.5, Equation (9.56), p. 55; see also Equations (9.13)–(9.14), p. 45]. The dynamical equations behind this picture are those of Cugliandolo and Kurchan [51, arXiv version 1, Equations (2)–(4), pp. 2–3], established for the soft-spherical model in [24, arXiv version 1, Theorem 1.2, p. 5]; for a conditional Gaussian-field description in a related setting, see [58, arXiv version 2, Remark 1.4, Equations (1.23)–(1.24), p. 5]. These works provide the background. We prove the identity in the exact form (4.10) directly from the CK equations, so that the later uses of the complete memory are transparent. The other estimates of this section are elementary, and we include their

proofs for completeness.

Throughout this section, (C, R, λ) is the solution of the hard-spherical CK system given by Proposition 3.2, and e_β is defined by (3.14).

4.1 The energy identity

Proposition 4.1 (Energy identity on fixed horizons). *For every $T < \infty$, the annealed normalized energy converges uniformly on $[0, T]$, and its limit is*

$$e_\beta(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} H_N(X_t) = \beta p \int_0^t C(t, u)^{p-1} R(t, u) \, du, \quad \lambda(t) = 1 + p\beta e_\beta(t). \quad (4.1)$$

In the physics literature, the energy is expressed in the same way through the correlation–response integral of [51, Equation (5)]. The sentence after that equation prints the energy with the sign opposite to their equation (9); the dictionary of Section 3.3, with $E(t) = -e_\beta(t)/\sqrt{2}$, agrees with the sign in equation (9). The proof below justifies the hard-sphere form through the primitives of the response and a rank-one correction in finite dimension. It is a Gaussian integration by parts in the disorder. Differentiating the trajectory with respect to one Gaussian coefficient produces the pathwise response matrix $U_N(t, u)$ of (3.6), and differentiating the covariance (3.2) splits the result into a trace term and a rank-one correction. The trace term carries the correlation weight $C_N(t, u)^{p-1}$ against the response; in the limit, it becomes the CK integral. Two estimates make this rigorous. The first is a fourth-moment bound for the response columns and normalized trace that is uniform in the dimension and in the waiting time u , and the second is the convergence (3.24) of the primitives of the annealed response. The rank-one correction is controlled by the operator norm, whose second moment is at most a constant times N .

Proof. The proof has four steps. In Step 1, we bound the drift derivative and the noise derivative trace along the sphere. In Step 2, a stopped fourth-moment estimate gives uniform bounds on the response columns and normalized trace. In Step 3, we integrate by parts in the disorder and obtain the exact identity (4.6). In Step 4, we pass to the limit $N \rightarrow \infty$.

Recall the variational definition (3.6) and the normalized trace (3.8). The physical coefficients are

$$b(x) = \beta P_x \nabla H_N(x) - \kappa_N x, \quad \sigma_j(x) = \sqrt{2} P_x e_j, \quad j = 1, \dots, N.$$

Step 1: bounds for the variational coefficients. The landscape derivative bounds give $\|Db(x)[v]\| \leq A_N \|v\|$ on S_N , with $A_N = 1 + \beta p(p+1)Q_N$. For the noise, $D\sigma_j(x)[v] = -\sqrt{2}N^{-1}(vx_j + xv_j)$. Since $\|x\|^2 = N$, the inequality $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ implies

$$\sum_{j=1}^N \|D\sigma_j(x)[v]\|^2 \leq B_N \|v\|^2, \quad B_N = \frac{8}{N}. \quad (4.2)$$

These estimates apply to all ambient vectors, so no samplewise diffeomorphism or invertibility assertion is needed.

Step 2: column, trace and operator moments. Fix $u \geq 0$ and a deterministic $v \in \mathbb{R}^N$. The true variational solution starts from $P_{X_u}v$, which is $\mathcal{F}_u$ -measurable and has norm at most $\|v\|$. Apply Itô's formula to the fourth power of its norm, stopped at the first exit from a ball. The bounds in Step 1, Gronwall's inequality, and Fatou's lemma when removing the stopping give, conditionally on the disorder,

$$\mathbb{E}_B \|U_N(t, u)v\|^4 \leq \|v\|^4 \exp\{(4A_N + 2 + 6B_N)(t - u)\}, \quad 0 \leq u \leq t. \quad (4.3)$$

For a $\mathcal{F}_u$ -measurable initial direction of norm at most a deterministic a , the same estimate holds with a^4 in place of $\|v\|^4$. For $N \geq 1$, the rate in (4.3) is at most $54 + 4\beta p(p + 1)Q_N$. Consequently, for $T < \infty$, the constant

$$K_T = e^{54T} \sup_N \mathbb{E} e^{4\beta p(p+1)TQ_N} < \infty, \quad \Theta_T = \sqrt{1 + K_T},$$

is independent of the dimension and waiting time. Its finiteness follows from Lemma 3.6. Put $\tau_N(t, u) = N^{-1} \text{tr } U_N(t, u)$. Jensen's inequality for the average of the diagonal entries, $|e_i \cdot U_N e_i| \leq \|U_N e_i\|$, and the elementary bound $\|U_N\|_{\text{op}}^2 \leq \sum_{i=1}^N \|U_N e_i\|^2$ show that

$$\begin{aligned} \sup_N \sup_{u \geq 0, 0 \leq t-u \leq T} \sup_{1 \leq i \leq N} \mathbb{E} \|U_N(t, u)e_i\|^4 &\leq K_T, \\ \sup_N \sup_{u \geq 0, 0 \leq t-u \leq T} \mathbb{E} |\tau_N(t, u)|^4 &\leq K_T, \\ \mathbb{E} \|U_N(t, u)\|_{\text{op}}^2 &\leq N\Theta_T^2, \quad 0 \leq t - u \leq T. \end{aligned} \quad (4.4)$$

In particular, $\mathbb{E} |\tau_N(t, u)|^2 \leq \Theta_T^2$, $\mathbb{E} |\tau_N(t, u)| \leq \Theta_T$, and $\mathbb{E} \|U_N(t, u)\|_{\text{op}} \leq \sqrt{N}\Theta_T$. The bounds depend only on the lag window T , even though the waiting state depends on the disorder and the past. Jointly measurable response representatives and their product-time integrability follow from the actual projected-restart derivative and these column moments.

Step 3: Gaussian integration by parts. For a multi-index $\nu = (i_1, \dots, i_p) \in \{1, \dots, N\}^p$, let $f_\nu(x) = N^{-(p-1)/2} x_{i_1} \cdots x_{i_p}$, so that $H_N = \sum_\nu J_\nu f_\nu$. On S_N , we have $|x_i| \leq \sqrt{N}$, so $|f_\nu| \leq \sqrt{N}$ and $\|\nabla f_\nu\| \leq p$. Changing the coefficient J_ν by ε adds $\varepsilon \beta P_x \nabla f_\nu(x)$ to the drift of (3.3), which is a tangent force $\varepsilon \beta P_{X_u} \nabla f_\nu(X_u)$ at time u . As in (3.7), the variation-of-constants formula yields

$$\partial_{J_\nu} X_t = \beta \int_0^t U_N(t, u) \nabla f_\nu(X_u) \, du, \quad (4.5)$$

where the derivative is taken in L^2 for the actual terminal solution. The first-order expansion of f_ν on the compact sphere gives

$$\begin{aligned}\partial_{J_\nu} \mathbb{E}_{X_0, B} f_\nu(X_t) &= \mathbb{E}_{X_0, B} [\nabla f_\nu(X_t) \cdot \partial_{J_\nu} X_t], \\ |\partial_{J_\nu} \mathbb{E}_{X_0, B} f_\nu(X_t)| &\leq \beta p^2 \int_0^t \mathbb{E}_{X_0, B} \|U_N(t, u)\|_{\text{op}} \, du.\end{aligned}$$

Step 2 makes the right side integrable under the disorder law at every fixed N . As X_0 and B are independent of J , we apply Gaussian integration by parts to $J \mapsto \mathbb{E}_{X_0, B} f_\nu(X_t)$, conditionally on the other disorder coefficients. Summing over the N^p multi-indices and using Fubini's theorem, we obtain

$$\mathbb{E} H_N(X_t) = \sum_\nu \mathbb{E} [J_\nu f_\nu(X_t)] = \beta \int_0^t \mathbb{E} \operatorname{tr} \left[U_N(t, u) \sum_\nu \nabla f_\nu(X_u) \nabla f_\nu(X_t)^\top \right] \, du.$$

To compute the sum over ν , we differentiate the identity $\sum_\nu f_\nu(x) f_\nu(y) = N^{-(p-1)} (x \cdot y)^p$, the middle expression in (3.2), once in x_j and once in y_i . With $c = x \cdot y / N$, this gives

$$\sum_\nu \partial_i f_\nu(y) \partial_j f_\nu(x) = pc^{p-1} \delta_{ij} + \frac{p(p-1)}{N} c^{p-2} x_i y_j.$$

Taking $x = X_t$ and $y = X_u$, so that $c = C_N(t, u)$, we have

$$\sum_\nu \nabla f_\nu(X_u) \nabla f_\nu(X_t)^\top = p C_N(t, u)^{p-1} I + \frac{p(p-1)}{N} C_N(t, u)^{p-2} X_t X_u^\top.$$

Inserting this into the previous display, using $\operatorname{tr}(U_N X_t X_u^\top) = X_u^\top U_N X_t$, and dividing by N , we obtain

$$\begin{aligned}\frac{1}{N} \mathbb{E} H_N(X_t) &= \beta p \int_0^t \mathbb{E} \left[C_N(t, u)^{p-1} \frac{\operatorname{tr} U_N(t, u)}{N} \right] \, du \\ &\quad + \frac{\beta p(p-1)}{N^2} \int_0^t \mathbb{E} [C_N(t, u)^{p-2} X_u^\top U_N(t, u) X_t] \, du.\end{aligned}\tag{4.6}$$

The second line is a rank-one correction. Using $|C_N| \leq 1$, $|X_u^\top U_N(t, u) X_t| \leq N \|U_N(t, u)\|_{\text{op}}$, and Step 2, its absolute value is at most $\beta p(p-1) T \Theta_T / \sqrt{N}$ for $0 \leq t \leq T$. This tends to zero uniformly in $t \leq T$.

Step 4: the limit $N \rightarrow \infty$. Fix $T < \infty$. In this step, only the dimension tends to infinity, with the horizon T fixed. We first replace the random weight $C_N(t, u)^{p-1}$ in the first line of (4.6) by $C(t, u)^{p-1}$. Using that $x \mapsto x^{p-1}$ is $(p-1)$ -Lipschitz on $[-1, 1]$, the bound $\mathbb{E} |\tau_N(t, u)|^2 \leq \Theta_T^2$, and the Cauchy-Schwarz inequality, we find that the error is at most

$$\begin{aligned}&\beta p(p-1) \int_0^t \mathbb{E} [|C_N(t, u) - C(t, u)| |\tau_N(t, u)|] \, du \\ &\leq \beta p(p-1) T \Theta_T \left(\mathbb{E} \sup_{0 \leq u \leq t \leq T} |C_N(t, u) - C(t, u)|^2 \right)^{1/2},\end{aligned}$$

which tends to zero uniformly in $t \leq T$ by (3.20). After this replacement, and by Fubini's theorem, the first line of (4.6) becomes $\beta p \int_0^t w(t, u) \bar{r}_N(t, u) \, du$, where

$$w(t, u) = C(t, u)^{p-1}, \quad 0 \leq u \leq t \leq T.$$

It remains to show that

$$\lim_{N \rightarrow \infty} \sup_{0 \leq t \leq T} \left| \int_0^t w(t, u) (\bar{r}_N(t, u) - R(t, u)) \, du \right| = 0. \quad (4.7)$$

The convergence of primitives (3.24) does not imply this by itself, because $\bar{r}_N(t, \cdot)$ may change sign; we also need a bound on the total variation of the signed measure $\bar{r}_N(t, u) \, du$. Because $|\bar{r}_N(t, u)| \leq \mathbb{E}|\tau_N(t, u)| \leq \Theta_T$, we have

$$\int_0^t |\bar{r}_N(t, u)| \, du \leq T\Theta_T, \quad 0 \leq t \leq T,$$

for every N . Let r_T be the maximum of the continuous function R on the triangle. Let

$$\eta_N = \sup_{0 \leq s \leq t \leq T} \left| \int_0^s (\bar{r}_N(t, u) - R(t, u)) \, du \right|,$$

which tends to zero by (3.24). Fix $\delta > 0$ and a partition $0 = \tau_0 < \tau_1 < \dots < \tau_m = T$ of mesh at most δ . Let $w_\delta(t, u) = w(t, \tau_{k-1})$ for $u \in [\tau_{k-1}, \tau_k)$, and let $\omega_T(\delta)$ be the modulus of continuity of w on the compact triangle $\{0 \leq u \leq t \leq T\}$, so that $\sup |w - w_\delta| \leq \omega_T(\delta)$ and $\omega_T(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. For fixed $t \leq T$, the integral $\int_0^t w_\delta(t, u) (\bar{r}_N - R)(t, u) \, du$ is a sum of at most m terms $w(t, \tau_{k-1}) \int_{\tau_{k-1} \wedge t}^{\tau_k \wedge t} (\bar{r}_N - R)(t, u) \, du$. Each integral, including the one over a possibly truncated last interval, is a difference of two primitives evaluated at points of $[0, t]$, and $|w| \leq 1$. This gives

$$\sup_{0 \leq t \leq T} \left| \int_0^t w(t, u) (\bar{r}_N - R)(t, u) \, du \right| \leq \omega_T(\delta) T (\Theta_T + r_T) + 2m\eta_N,$$

where the first term bounds the integral of $(w - w_\delta)(\bar{r}_N - R)$ by the total variation bounds above. We first let $N \rightarrow \infty$ with the partition fixed, which removes the term $2m\eta_N$, and then let $\delta \rightarrow 0$. This proves (4.7).

Combining Step 3, the replacement of the weight, and (4.7), we conclude that

$$\lim_{N \rightarrow \infty} \sup_{0 \leq t \leq T} \left| \frac{1}{N} \mathbb{E} H_N(X_t) - \beta p \int_0^t C(t, u)^{p-1} R(t, u) \, du \right| = 0,$$

which is the first assertion of (4.1). The identity $\lambda(t) = 1 + p\beta e_\beta(t)$ follows from (3.11). $\square$

4.2 The Green covariance identity

We now regard the CK solution as fixed and use its coefficients to define a linear problem. Let

$$\Sigma(t, u) = \beta^2 p(p-1)C(t, u)^{p-2}R(t, u), \quad 0 \leq u \leq t, \quad (4.8)$$

which is continuous and, by (3.15), nonnegative. Let $\mathcal{D}$ be the linear Volterra operator

$$\mathcal{D}f(t) = f'(t) + \lambda(t)f(t) - \int_0^t \Sigma(t, u)f(u) \, du.$$

Note that $\beta^2 \zeta''(C)R = \Sigma$ and $\beta^2 \zeta'(C) = \beta^2 p C^{p-1}$. The CK equations take the following form. We extend $R(\cdot, s)$ by zero on $[0, s)$. Then, for fixed s , (3.9) says that $\mathcal{D}R(\cdot, s)(t) = 0$ for $t > s$, with $R(s, s) = 1$. That is, the response is the causal Green kernel of $\mathcal{D}$, with unit diagonal. Also, (3.10) says that

$$\mathcal{D}C(\cdot, s)(t) = \beta^2 p \int_0^s C(t, u)^{p-1} R(s, u) \, du, \quad t > s. \quad (4.9)$$

Although the coefficients λ and Σ come from the nonlinear CK system, they are fixed continuous functions in what follows.

Proposition 4.2 (Green covariance identity). *For every $t, s \geq 0$,*

$$\begin{aligned} C(t, s) &= R(t, 0)R(s, 0) + 2 \int_0^{t \wedge s} R(t, u)R(s, u) \, du \\ &\quad + \beta^2 p \int_0^t \int_0^s R(t, u)R(s, v)C(u, v)^{p-1} \, dv \, du. \end{aligned} \quad (4.10)$$

In particular,

$$C(t, 0) = R(t, 0), \quad 2 \int_0^t R(t, u)^2 \, du \leq 1. \quad (4.11)$$

The three terms of (4.10) are the contributions of the initial condition, of the white noise, whose covariance is 2 in our convention, and of the colored noise with covariance $\beta^2 p C^{p-1} = \beta^2 \zeta'(C)$, each propagated by the response.

Proof. Let $Q(t, s)$ be the right side of (4.10). It is continuous and symmetric. The proof has four steps. We verify that Q satisfies the same linear equation as C off the diagonal, and the same equation as C along the diagonal; a Gronwall argument then shows that $Q = C$. Finally, we deduce (4.11). Throughout, all integrands are continuous and bounded on bounded sets, and whenever we exchange the order of integration below, we apply Fubini's theorem on a bounded region.

Step 1: the off-diagonal equation. We write $Q = Q_1 + Q_2 + Q_3$ for the three terms:

$$Q_1(t, s) = R(t, 0)R(s, 0), \quad Q_2(t, s) = 2 \int_0^{t \wedge s} R(t, u)R(s, u) \, du,$$

$$Q_3(t, s) = \beta^2 p \int_0^t \int_0^s R(t, u)R(s, v)K(u, v) \, dv \, du, \quad K(u, v) = C(u, v)^{p-1}.$$

Fix s and let $t > s$. By (3.9), we have $\partial_t R(t, 0) = -\lambda(t)R(t, 0) + \int_0^t \Sigma(t, z)R(z, 0) \, dz$, and hence

$$\partial_t Q_1(t, s) = -\lambda(t)Q_1(t, s) + \int_0^t \Sigma(t, z)Q_1(z, s) \, dz.$$

For $t > s$, we have $Q_2(t, s) = 2 \int_0^s R(t, u)R(s, u) \, du$. Differentiating under the integral and using (3.9) gives

$$\partial_t Q_2(t, s) = -\lambda(t)Q_2(t, s) + 2 \int_0^s R(s, u) \int_u^t \Sigma(t, z)R(z, u) \, dz \, du.$$

The region $\{0 \leq u \leq s, u \leq z \leq t\}$ is $\{0 \leq z \leq t, 0 \leq u \leq z \wedge s\}$, and exchanging the integrals yields

$$\partial_t Q_2(t, s) = -\lambda(t)Q_2(t, s) + \int_0^t \Sigma(t, z)Q_2(z, s) \, dz.$$

Note that the inner integral produces $Q_2(z, s)$ with its upper limit $z \wedge s$, which accounts correctly for the memory times $z < s$. For the colored term, differentiating in the upper limit t produces a boundary term with $R(t, t) = 1$, and differentiating $R(t, u)$ inside the integral produces the memory term after the same exchange of integrals. We obtain

$$\partial_t Q_3(t, s) = -\lambda(t)Q_3(t, s) + \int_0^t \Sigma(t, z)Q_3(z, s) \, dz + \beta^2 p \int_0^s K(t, v)R(s, v) \, dv.$$

Adding the three equations, we conclude that, for $t > s$,

$$\partial_t Q(t, s) = -\lambda(t)Q(t, s) + \int_0^t \Sigma(t, z)Q(z, s) \, dz + \beta^2 p \int_0^s C(t, v)^{p-1}R(s, v) \, dv. \quad (4.12)$$

By (4.9), the correlation C satisfies the same equation. Note that the white-noise integral has no boundary term when $t > s$, because its upper limit $t \wedge s = s$ does not move.

Step 2: the diagonal equation. Along the diagonal, the upper limit of the white-noise integral moves. We differentiate $Q(t, t) = Q_1(t, t) + Q_2(t, t) + Q_3(t, t)$, using (3.9), $R(t, t) = 1$, and the symmetry of K . Both arguments now move, and the computations of Step 1 give

$$\frac{d}{dt} Q_i(t, t) = -2\lambda(t)Q_i(t, t) + 2 \int_0^t \Sigma(t, u)Q_i(u, t) \, du + b_i(t), \quad i = 1, 2, 3,$$

where $b_1(t) = 0$, $b_2(t) = 2$, and $b_3(t) = 2\beta^2 p \int_0^t K(t, u) R(t, u) \, du$. Adding these identities, we obtain

$$\frac{d}{dt} Q(t, t) = -2\lambda(t) Q(t, t) + 2 \int_0^t \Sigma(t, u) Q(u, t) \, du + 2\beta^2 p \int_0^t C(t, u)^{p-1} R(t, u) \, du + 2. \quad (4.13)$$

The last term 2 is the boundary term $2R(t, t)^2$ from the moving upper limit of the white-noise integral. The constant function $C(t, t) = 1$ satisfies the same equation. Indeed, if we substitute $Q = C$ and $C(t, t) = 1$, the right side of (4.13) becomes

$$-2\lambda(t) + 2 \int_0^t \Sigma(t, u) C(t, u) \, du + 2\beta^2 p \int_0^t C(t, u)^{p-1} R(t, u) \, du + 2,$$

and since $\Sigma(t, u) C(t, u) = \beta^2 p(p-1) C(t, u)^{p-1} R(t, u)$, this vanishes by the multiplier equation (3.11). Both Q and C equal one at $(0, 0)$.

Step 3: uniqueness. Let $D = Q - C$. It is continuous and symmetric, $D(0, 0) = 0$, and by Steps 1 and 2 it satisfies the homogeneous equations

$$\begin{aligned} \partial_t D(t, s) &= -\lambda(t) D(t, s) + \int_0^t \Sigma(t, z) D(z, s) \, dz, \quad t > s, \\ \frac{d}{dt} D(t, t) &= -2\lambda(t) D(t, t) + 2 \int_0^t \Sigma(t, z) D(z, t) \, dz. \end{aligned}$$

Fix $T < \infty$, and let $\bar{\lambda}_T = \sup_{[0, T]} |\lambda|$, $\bar{\Sigma}_T = \sup_{0 \leq u \leq t \leq T} |\Sigma(t, u)|$, and

$$d(t) = \sup_{0 \leq u, v \leq t} |D(u, v)|, \quad 0 \leq t \leq T,$$

which is nondecreasing and bounded. Integrating the diagonal equation from 0 to s and bounding $|D(z, v)| \leq d(v)$ for $z \leq v$, we find

$$|D(s, s)| \leq (2\bar{\lambda}_T + 2\bar{\Sigma}_T T) \int_0^s d(v) \, dv.$$

For $t > s$, we integrate the off-diagonal equation from s to t and find

$$|D(t, s)| \leq |D(s, s)| + (\bar{\lambda}_T + \bar{\Sigma}_T T) \int_s^t d(v) \, dv,$$

where we used that for $v \in [s, t]$ and $z \leq v$, both arguments of $D(v, s)$ and $D(z, s)$ lie in $[0, v]$. These two bounds and the symmetry of D imply that $|D(t', s')| \leq (2\bar{\lambda}_T + 2\bar{\Sigma}_T T) \int_0^t d(v) \, dv$ whenever $t', s' \leq t$. Taking the supremum over t' and s' , we have

$$d(t) \leq (2\bar{\lambda}_T + 2\bar{\Sigma}_T T) \int_0^t d(v) \, dv, \quad 0 \leq t \leq T.$$

By Gronwall's inequality, $d \equiv 0$ on $[0, T]$. Since the horizon T was arbitrary, we conclude that $Q = C$, which is (4.10).

Step 4: consequences. At $s = 0$, the white-noise and colored terms of (4.10) vanish and $R(0, 0) = 1$, so that $C(t, 0) = R(t, 0)$. At $s = t$, the left side of (4.10) is one, and the three terms on the right are nonnegative, because $R \geq 0$ and $C \geq 0$ by (3.15). In particular, $2 \int_0^t R(t, u)^2 \, du \leq 1$. $\square$

Feature vectors. We now realize (4.10) as a Gram identity. By Proposition 3.2, C is a positive-semidefinite kernel, and the Schur product theorem [111, Section 4, Satz VII, p. 14] implies that its entrywise power $K = C^{p-1}$ is positive semidefinite as well. Let $\mathcal{K}$ be the Hilbert space generated by vectors V_u , $u \geq 0$, with

$$\langle V_u, V_v \rangle = C(u, v)^{p-1}.$$

Explicitly, we take finite formal linear combinations of the symbols V_u , define the inner product by bilinearity, take the quotient by the vectors of norm zero, and complete. This is the classical positive-kernel Hilbert-space construction [4, Part I, Section 2(4), p. 344]. We call $\mathcal{K}$ the colored feature space. The vectors V_u have norm one, and

$$\|V_u - V_v\|^2 = 2\{1 - C(u, v)^{p-1}\}, \quad (4.14)$$

so $u \mapsto V_u$ is norm-continuous, because C is continuous. We may therefore define the Bochner integral

$$A_t = \int_0^t R(t, u) V_u \, du \in \mathcal{K}$$

of a continuous function on a compact interval. Since bounded linear functionals commute with the integral, we have

$$\langle A_t, A_s \rangle = \int_0^t \int_0^s R(t, u) R(s, v) C(u, v)^{p-1} \, dv \, du, \quad \langle V_t, A_t \rangle = \int_0^t C(t, u)^{p-1} R(t, u) \, du.$$

In the direct sum $\mathbb{R} \oplus L^2(\mathbb{R}_+) \oplus \mathcal{K}$, we define

$$Z_t = R(t, 0) \oplus \sqrt{2} R(t, \cdot) \mathbf{1}_{[0, t]} \oplus \beta \sqrt{p} A_t. \quad (4.15)$$

Adding the inner products of the three orthogonal components, we see that (4.10) says exactly

$$\langle Z_t, Z_s \rangle = C(t, s), \quad \|Z_t\| = 1.$$

The initial, white-noise, and colored contributions to the correlation are orthogonal components of Z_t , and the colored component retains the complete history. Section 7 refines the colored component into a causal realization graded by trees.

4.3 The energy–response tradeoff and the multiplier bound

Proposition 4.3 (Energy–response tradeoff and uniform multiplier bound). *For every $t \geq 0$,*

$$\frac{e_\beta(t)^2}{p} + C(t, 0)^2 + 2 \int_0^t R(t, u)^2 \, du \leq 1, \quad (4.16)$$

and

$$1 \leq \lambda(t) \leq \Lambda, \quad \Lambda = \frac{1 + \sqrt{1 + 4\beta^2 p^3}}{2}. \quad (4.17)$$

Both estimates are deterministic consequences of the CK equations and of Proposition 3.2. Proposition 4.1 is needed only to interpret $e_\beta(t)$ as an energy.

Proof. The proof has three steps. We first prove the tradeoff, then deduce a first bound on the multiplier, and finally improve it to (4.17) using a lower bound on the response.

Step 1: the tradeoff. Let $M_t = \int_0^t C(t, u)^{p-1} R(t, u) \, du$. With the notation of the feature vectors above, we have $M_t = \langle V_t, A_t \rangle$, and since $\|V_t\| = 1$, the Cauchy–Schwarz inequality yields $M_t^2 \leq \|A_t\|^2$. On the diagonal, (4.10) reads $1 = R(t, 0)^2 + 2 \int_0^t R(t, u)^2 \, du + \beta^2 p \|A_t\|^2$. Combining these two facts, we find

$$\beta^2 p M_t^2 + R(t, 0)^2 + 2 \int_0^t R(t, u)^2 \, du \leq 1.$$

Since $e_\beta(t) = \beta p M_t$ and $\lambda(t) - 1 = \beta^2 p^2 M_t$, we have $e_\beta(t)^2/p = \beta^2 p M_t^2 = (\lambda(t) - 1)^2/(\beta^2 p^3)$. Using $C(t, 0) = R(t, 0)$ from (4.11), we obtain (4.16), and equivalently

$$\frac{(\lambda(t) - 1)^2}{\beta^2 p^3} + R(t, 0)^2 + 2 \int_0^t R(t, u)^2 \, du \leq 1. \quad (4.18)$$

Step 2: a first bound. Note that $M_t \geq 0$ and $\lambda(t) \geq 1$, because $C \geq 0$ and $R \geq 0$. By (4.18), we also have $\lambda(t) \leq 1 + \beta p^{3/2}$. In particular, $\Lambda_0 = \sup_{t \geq 0} \lambda(t)$ satisfies $1 \leq \Lambda_0 < \infty$.

Step 3: the improvement. Since $\Sigma \geq 0$ and $R \geq 0$, (3.9) implies that $\partial_t R(t, u) \geq -\lambda(t) R(t, u)$ for $t \geq u$. Then $t \mapsto R(t, u) \exp \int_u^t \lambda(v) \, dv$ is nondecreasing, and since $R(u, u) = 1$, we have

$$R(t, u) \geq \exp\left(-\int_u^t \lambda(v) \, dv\right) \geq e^{-\Lambda_0(t-u)}.$$

Squaring this bound and integrating gives

$$R(t, 0)^2 + 2 \int_0^t R(t, u)^2 \, du \geq e^{-2\Lambda_0 t} + \frac{1 - e^{-2\Lambda_0 t}}{\Lambda_0} \geq \frac{1}{\Lambda_0},$$

where in the last inequality we used that the middle expression is a convex combination of 1 and $1/\Lambda_0$, and that $\Lambda_0 \geq 1$. Inserting this into (4.18), we find $(\lambda(t) - 1)^2 \leq \beta^2 p^3 (1 - \Lambda_0^{-1})$ for every t . Taking times t with $\lambda(t) \rightarrow \Lambda_0$ (the supremum need not be attained), we obtain

$$(\Lambda_0 - 1)^2 \leq \beta^2 p^3 (1 - \Lambda_0^{-1}) = \beta^2 p^3 \frac{\Lambda_0 - 1}{\Lambda_0}.$$

If $\Lambda_0 > 1$, we divide by $\Lambda_0 - 1$ and find $\Lambda_0^2 - \Lambda_0 - \beta^2 p^3 \leq 0$, which implies that Λ_0 is at most the positive root Λ of $x^2 - x - \beta^2 p^3$. If $\Lambda_0 = 1$, then $\Lambda_0 \leq \Lambda$ because $\Lambda \geq 1$. This proves (4.17). $\square$

4.4 Regularity in time

Proposition 4.4 (Uniform regularity in time). *There exist finite constants $D_{p,\beta}$ and $L_{p,\beta}$, depending only on p and β , such that e_β and λ are $L_{p,\beta}$ -Lipschitz on $[0, \infty)$, and, for all $u, v \geq 0$ and $h \geq 0$,*

$$|C(u+h, v) - C(u, v)| \leq D_{p,\beta} h + \sqrt{2h}, \quad |C(u, v+h) - C(u, v)| \leq D_{p,\beta} h + \sqrt{2h}. \quad (4.19)$$

Moreover, with Λ as in (4.17), for all $t \geq s \geq 0$,

$$0 \leq 1 - C(t, s) \leq \Lambda(t - s). \quad (4.20)$$

We obtain the estimates (4.19) from the diffusion in finite dimension and pass them to the limit with u, v , and h fixed; their constants do not depend on the times involved. The bound (4.20) is stronger near the diagonal. It implies that any unit vectors v_t in a Hilbert space with $\langle v_t, v_s \rangle = C(t, s)$, such as Z_t in (4.15), satisfy

$$\|v_t - v_s\|^2 = 2\{1 - C(t, s)\} \leq 2\Lambda|t - s|,$$

so that $t \mapsto v_t$ is $1/2$ -Hölder. The modulus (4.19) alone would only yield the exponent $1/4$.

Proof. Let $\gamma_j = \sup_N \mathbb{E} Q_N^j$ for $j = 1, 2$, which are finite by Lemma 3.6 and depend only on p .

Step 1: the energy and the multiplier. Let $e_N(t) = N^{-1} \mathbb{E} H_N(X_t)$. By Itô's formula on S_N and (3.4),

$$H_N(X_t) - H_N(X_s) = \int_s^t [\Delta_S H_N(X_r) + \beta \|\nabla_S H_N(X_r)\|^2] dr + \sqrt{2} \int_s^t \nabla_S H_N(X_r) \cdot dB_r.$$

For fixed J , the integrand of the stochastic integral is bounded, and the stochastic integral has mean zero; by (3.26), the integrands are integrable over the disorder as well. By (3.26),

we have $|\Delta_S H_N|/N \leq p^2 Q_N$ and $\|\nabla_S H_N\|^2/N \leq \|\nabla H_N\|^2/N \leq p^2 Q_N^2$ on S_N . Taking expectations gives

$$|e_N(t) - e_N(s)| \leq p^2(\gamma_1 + \beta\gamma_2)|t - s|,$$

with a constant independent of N , s , and t . By Proposition 4.1, we have $e_N(t) \rightarrow e_\beta(t)$ for every t , and the Lipschitz bound passes to the limit, showing that e_β is Lipschitz with constant $L_e = p^2(\gamma_1 + \beta\gamma_2)$. By (4.1), $\lambda = 1 + p\beta e_\beta$ is Lipschitz with constant $p\beta L_e$. We take $L_{p,\beta} = \max\{1, p\beta\} L_e$.

Step 2: the correlation modulus. We decompose the increment of the path into its drift and martingale parts:

$$X_{u+h} - X_u = \int_u^{u+h} [\beta P_{X_r} \nabla H_N(X_r) - \kappa_N X_r] dr + \sqrt{2} \int_u^{u+h} P_{X_r} dB_r.$$

By (3.26), the drift has norm at most $(1 + \beta p Q_N) \sqrt{N}$. By the Itô isometry and $\text{tr}(P_x^2) = \text{tr} P_x = N - 1$, the martingale part has mean square norm $2(N - 1)h \leq 2Nh$. Using the Cauchy–Schwarz inequality for the martingale part, we find

$$\mathbb{E} \frac{\|X_{u+h} - X_u\|}{\sqrt{N}} \leq (1 + \beta p \gamma_1)h + \sqrt{2h}.$$

By the Cauchy–Schwarz inequality and $\|X_v\| = \sqrt{N}$, we also have, pathwise,

$$|C_N(u + h, v) - C_N(u, v)| = \frac{|(X_{u+h} - X_u) \cdot X_v|}{N} \leq \frac{\|X_{u+h} - X_u\|}{\sqrt{N}},$$

and no relation between the times u and v is needed. Taking expectations, we deduce that $|\mathbb{E} C_N(u + h, v) - \mathbb{E} C_N(u, v)| \leq D_{p,\beta} h + \sqrt{2h}$ with $D_{p,\beta} = 1 + \beta p \gamma_1$. We now let $N \rightarrow \infty$ with u , v , and h fixed. By Proposition 3.3, applied on the horizon $\max\{u + h, v\}$ and using the symmetry of C_N and C , we have $\mathbb{E} C_N \rightarrow C$ at both points, which proves the first bound in (4.19). The second bound follows from the first and the symmetry of C .

Step 3: the diagonal increment. Let $t > s$. All terms on the right side of (3.10) other than the first are nonnegative, by (3.15). Since $0 \leq C \leq 1$ and $\lambda \leq \Lambda$ by Proposition 4.3, we find

$$\partial_t C(t, s) \geq -\lambda(t) C(t, s) \geq -\Lambda.$$

Integrating from s to t and using $C(s, s) = 1$, we obtain $C(t, s) \geq 1 - \Lambda(t - s)$, which is the upper bound in (4.20). The lower bound is $C \leq 1$. $\square$

4.5 The backward response equation

The response equation (3.9) is an equation in the later time t . The next proposition gives the corresponding equation in the earlier time s .

Proposition 4.5 (Backward response equation). *For $t > s \geq 0$, the response satisfies*

$$\partial_s R(t, s) = \lambda(s)R(t, s) - \int_s^t R(t, u)\Sigma(u, s) \, du. \quad (4.21)$$

Both partial derivatives $\partial_t R$ and $\partial_s R$ extend continuously from $\{0 < s < t\}$ to each compact causal triangle $\{0 \leq s \leq t \leq T\}$. At $s = 0 < t$, $\partial_s R$ is the right derivative; on the positive diagonal, $\partial_t R$ is the right derivative and $\partial_s R$ the left derivative along their causal rows. At $(0, 0)$ these symbols denote the continuous extensions of the interior slopes.

Proof. The idea is to write the response as a series of Volterra compositions, to reverse the order of the factors, and to differentiate the resulting formula in its lower endpoint; this needs only the continuity of Σ , not its derivative. Guionnet and Mazza give a related positive noncrossing expansion for the normalized response in a nonlinear Kraichnan equation [89, arXiv version 1, Section 2.1 and Lemma 2.1, pp. 7–8]. Here we keep Σ fixed and use the Neumann series of the resulting linear Volterra equation. For continuous kernels k_1 and k_2 on the causal triangle, we define the Volterra composition

$$(k_1 * k_2)(t, s) = \int_s^t k_1(t, u)k_2(u, s) \, du, \quad 0 \leq s \leq t,$$

which is associative by Fubini's theorem on bounded regions. Let

$$G(t, s) = \exp\left(-\int_s^t \lambda(v) \, dv\right), \quad 0 \leq s \leq t.$$

Note that $0 < G \leq 1$, because $\lambda \geq 1$ by Proposition 4.3.

Step 1: the Neumann series. By (3.9), $\partial_t[R(t, s)/G(t, s)] = (\Sigma * R)(t, s)/G(t, s)$ for $t \geq s$. Integrating from s to t and using $R(s, s) = 1$ gives the variation-of-constants formula $R = G + G * \Sigma * R$. Fix T , and let $\bar{\Sigma}_T = \sup_{0 \leq u \leq t \leq T} \Sigma(t, u)$. The kernel $(G * \Sigma)^n * G$ is an integral over $2n$ ordered intermediate times in $[s, t]$ of a product of n factors Σ and $n + 1$ factors G . As $0 < G \leq 1$, it is at most $\bar{\Sigma}_T^n$ times the volume of the simplex of ordered times:

$$|(G * \Sigma)^n * G|(t, s) \leq \bar{\Sigma}_T^n \frac{(t - s)^{2n}}{(2n)!}, \quad 0 \leq s \leq t \leq T.$$

The Neumann series $\sum_{n \geq 0} (G * \Sigma)^n * G$ therefore converges uniformly on the triangle, and its sum $\tilde{R}$ satisfies $\tilde{R} = G + G * \Sigma * \tilde{R}$. The difference $\tilde{R} - R$ satisfies the homogeneous equation $f = G * \Sigma * f$, and iterating this equation n times shows that $|f| \leq \bar{\Sigma}_T^n T^{2n} \sup |f| / (2n)!$ for every n , which forces $f = 0$. We have shown that

$$R = \sum_{n \geq 0} (G * \Sigma)^n * G.$$

Step 2: reversing the composition. By associativity, $(G * \Sigma)^n * G = G * (\Sigma * G)^n$, and both are the composition of the word $G\Sigma G \cdots \Sigma G$. Separating the last two factors $\Sigma * G$ of each term with $n \geq 1$, we obtain

$$R = G + \sum_{n \geq 0} (G * \Sigma)^n * G * \Sigma * G = G + R * \Sigma * G.$$

Explicitly, for $0 \leq s \leq t$,

$$R(t, s) = G(t, s) + \int_s^t R(t, u) (\Sigma * G)(u, s) \, du, \quad (\Sigma * G)(u, s) = \int_s^u \Sigma(u, v) G(v, s) \, dv.$$

Step 3: differentiation in the earlier time. We have $\partial_s G(v, s) = \lambda(s)G(v, s)$ and $G(s, s) = 1$. Differentiating the inner integral in its lower endpoint yields

$$\partial_s (\Sigma * G)(u, s) = -\Sigma(u, s) + \lambda(s)(\Sigma * G)(u, s),$$

which is continuous in (u, s) . The outer integral has lower endpoint s , and its boundary term vanishes because $(\Sigma * G)(s, s) = 0$. By the Leibniz rule, we have

$$\begin{aligned} \partial_s R(t, s) &= \lambda(s)G(t, s) + \int_s^t R(t, u) [-\Sigma(u, s) + \lambda(s)(\Sigma * G)(u, s)] \, du \\ &= \lambda(s) [G + R * \Sigma * G](t, s) - (R * \Sigma)(t, s). \end{aligned}$$

Since $G + R * \Sigma * G = R$, this is (4.21). We differentiated only continuous integrands through their moving endpoints.

The right side of (4.21) is continuous on the closed triangle $\{0 \leq s \leq t \leq T\}$, with value $\lambda(t)$ at $s = t$. The forward derivative given by (3.9) is also continuous there, with value $-\lambda(s)$ at $t = s$. This proves the assertions about continuity. $\square$

4.6 Response domination and compactness

For $T \geq 0$, we consider the translated kernels

$$C_T(a, b) = C(T + a, T + b), \quad R_T(a, b) = R(T + a, T + b) \quad (a \geq b), \quad \lambda_T(a) = \lambda(T + a), \quad (4.22)$$

defined for offsets $a, b \geq -T$. On every compact set of offsets, they are defined once T is large enough. We consider correlations on squares of offsets, and responses only on closed causal triangles $\{a \geq b\}$, where their diagonal value is one.

Proposition 4.6 (Response domination and bounded-lag compactness). *Let $a_{p,\beta} = \beta^2 p(p-1)$. For every $t \geq s \geq 0$,*

$$0 \leq R(t, s) \leq C(t, s) \leq 1, \quad |\partial_t R(t, s)| + |\partial_s R(t, s)| \leq 2\{\Lambda + a_{p,\beta}(t-s)\}, \quad (4.23)$$

with the boundary interpretations of Proposition 4.5. The translated triples (C_T, R_T, λ_T) are precompact for locally uniform convergence as $T \rightarrow \infty$, in the sense that every sequence $T_n \rightarrow \infty$ has a subsequence along which C_{T_n} , R_{T_n} , and λ_{T_n} converge uniformly on every compact subset of $\mathbb{R}^2$, of $\{(a, b) \in \mathbb{R}^2 : a \geq b\}$, and of $\mathbb{R}$, respectively.

Proof. The proof has three steps. We first prove the domination $R \leq C$, then the derivative bounds, and finally the compactness.

Step 1: the domination. Fix $s \geq 0$, and let $D_s(t) = C(t, s) - R(t, s)$ for $t \geq s$. We split the first memory integral of (3.10) at s . With the notation (4.8), the correlation satisfies

$$\partial_t C(t, s) = -\lambda(t)C(t, s) + \int_s^t \Sigma(t, u)C(u, s) \, du + F_s(t), \quad t \geq s,$$

where

$$F_s(t) = \int_0^s \Sigma(t, u)C(u, s) \, du + \beta^2 p \int_0^s C(t, u)^{p-1} R(s, u) \, du \geq 0.$$

In words, for fixed s , the correlation solves the same linear Volterra equation as the response, with the same initial value one at $t = s$ and the additional nonnegative forcing F_s , which collects the memory from times before s . Subtracting (3.9), we find

$$D'_s(t) = -\lambda(t)D_s(t) + \int_s^t \Sigma(t, u)D_s(u) \, du + F_s(t), \quad D_s(s) = 0.$$

Let $\tilde{D}_s(t) = \int_s^t R(t, v)F_s(v) \, dv$. Differentiating, using $R(t, t) = 1$ and (3.9), and exchanging the order of integration over $\{s \leq v \leq u \leq t\}$, we find that $\tilde{D}_s$ satisfies the same equation with $\tilde{D}_s(s) = 0$. The difference $f = D_s - \tilde{D}_s$ satisfies $f' = -\lambda f + \int_s^t \Sigma(t, u)f(u) \, du$ and $f(s) = 0$. On $[s, T]$, with $\bar{f}(t) = \sup_{[s, t]} |f|$, we obtain $\bar{f}(t) \leq (\bar{\lambda}_T + \bar{\Sigma}_T T) \int_s^t \bar{f}(v) \, dv$, with $\bar{\lambda}_T$ and $\bar{\Sigma}_T$ as in the proof of Proposition 4.2, and Gronwall's inequality shows that $f = 0$. We have shown that

$$C(t, s) - R(t, s) = \int_s^t R(t, v)F_s(v) \, dv \geq 0.$$

Together with (3.15), this implies that $0 \leq R \leq C \leq 1$.

Step 2: the derivative bounds. By Step 1 and (3.15), we have $0 \leq \Sigma(t, u) \leq a_{p, \beta}$. In the forward equation (3.9), the memory term is therefore at most $a_{p, \beta}(t - s)$ in absolute value, and the multiplier term is at most Λ by (4.17). This gives $|\partial_t R(t, s)| \leq \Lambda + a_{p, \beta}(t - s)$. The same argument applied to the backward equation (4.21) gives $|\partial_s R(t, s)| \leq \Lambda + a_{p, \beta}(t - s)$. This proves (4.23); at $t = s$, we use the continuous extensions of Proposition 4.5.

Step 3: compactness. For $m \in \mathbb{N}$, let $W_m = [-m, m]^2$ and $W_m^{\text{caus}} = \{(a, b) \in W_m : a \geq b\}$. For $T \geq m$, the translated kernels are defined on these sets. On them, we have the following bounds, uniformly in $T \geq m$:

- $|C_T| \leq 1$, and C_T has the modulus of continuity (4.19) in each coordinate, by Proposition 4.4;
- $0 \leq R_T \leq 1$, and R_T is Lipschitz on the convex set W_m^{caus} with constant $2\{\Lambda + 2a_{p,\beta}m\}$, by (4.23), since the lag is at most $2m$ there;
- $1 \leq \lambda_T \leq \Lambda$, and λ_T is $L_{p,\beta}$ -Lipschitz, by Propositions 4.3 and 4.4.

Let $T_n \rightarrow \infty$. By the Arzelà–Ascoli theorem, there exists a subsequence along which the three components converge uniformly on W_1 , W_1^{caus} , and $[-1, 1]$. From it, we extract a further subsequence that converges uniformly on W_2 , W_2^{caus} , and $[-2, 2]$, and so on. The diagonal subsequence converges uniformly on W_m , W_m^{caus} , and $[-m, m]$ for every m . Every compact set of offsets in the respective domains lies in one of these sets for m large, and this proves the precompactness. Offsets may be negative; each compact set lies in the domain once T_n exceeds its distance from the origin. We use the response only inside the causal triangles, and we need no continuity across the diagonal. $\square$

Entropy production and stationary limits

In this section, we prove that the CK solution becomes stationary on late time windows and satisfies the fluctuation–dissipation theorem (FDT) there. Our tool is the relative entropy of the diffusion with respect to its Gibbs measure. It decreases at a rate equal to the relative Fisher information. Since the initial law is uniform, the entropy produced over all times is at most $\beta^2 N/2$ on average over the disorder (Lemma 5.1). We then show in Lemma 5.2 that the Fisher information at an earlier time s controls the backward FDT defect $\partial_s C - R$ at that time. In Theorem 5.3, we pass these estimates to the limit along one subsequence of dimensions. We obtain an integrable function a_F , the Fisher majorant, whose square root dominates the defect for all $t > s > 0$. In Lemma 5.5, we establish that the Fisher information also controls the change of a correlation at fixed lag when both observation times are shifted. Combining these estimates with compactness, we prove Proposition 5.7. It states that every limit point of the translated kernels is a stationary FDT profile with a constant multiplier, and that its response satisfies a closed equation.

We use the following results of Sections 3 and 4: the fixed-horizon bridge (Proposition 3.3), the Brownian primitive of the response (Lemma 3.4), the landscape bounds (Lemma 3.6), the energy identity (Proposition 4.1), the Green covariance identity (Proposition 4.2), and the uniform estimates of Propositions 4.3 to 4.6. The stationary clusters of Proposition 5.7 are the starting point of Sections 6 to 8. We use the Fisher majorant again in Sections 10, 11, 14, and 15, and Section 10 reuses the subsequence of dimensions chosen in Theorem 5.3.

The entropy dissipation identity, the Bochner formula, and the gradient bound for the semigroup are standard calculus for reversible diffusions. The Bochner formula for a Laplacian with a gradient drift is recorded in [9, Proposition 3, equation (4b), p. 187], and the diffusion and entropy calculus is presented in [10, Chapter 3 and Proposition 5.2.2, p. 245]. We derive the spherical response identity of Lemma 5.2 directly; for related response identities around nonthermal steady states under detailed balance, see [1, Sections 2–3, pp. 27–30]; nonequilibrium diffusion response identities are given in [8, equations (6) and (11), pp. 010602-2–010602-3]. Cugliandolo, Dean, and Kurchan described the physical mechanism that links entropy production to violations of FDT, and they derived bounds on these violations from the entropy production [54, arXiv version 2, equations (12)–(13), p. 3, and (18), p. 4]; their integrated bounds are discussed in the unnumbered passage “Integrated bounds for purely relaxational systems” on p. 3 of that

arXiv version. For the projected hard-spherical response, we establish dimension-uniform bounds and pass them to a single Fisher majorant for the CK solution. We give the calculations in full because the sign, the spherical projection, and the normalization of the response all matter later.

Standing notation. Throughout this section, $p \geq 3$ and $\beta > 0$ are fixed. For a fixed realization of the disorder J , the Hamiltonian H_N is fixed. Recall from (3.4) that the generator of (3.3) is

$$\mathcal{L}_N = \Delta_S + \beta \langle \nabla_S H_N, \nabla_S \cdot \rangle.$$

We write $P_\tau = P_\tau^H = e^{\tau \mathcal{L}_N}$ for its semigroup. All differential operators act on functions on S_N and are spherical: ∇_S , div_S , Δ_S , and ∇_S^2 are the gradient, divergence, Laplacian, and Hessian of S_N , as in Section 3.1, and Ric is its Ricci tensor. We write $\Gamma(f, g) = \langle \nabla_S f, \nabla_S g \rangle$ and $\Gamma(f) = \Gamma(f, f)$. The subscript H marks quantities computed for the fixed Hamiltonian H_N . The expectation $\mathbb{E}_J$ averages over the Gaussian tensor, while $\mathbb{E}$ also averages over the initial point and the Brownian motion. Let ρ_t be the density of the law of X_t given J with respect to ω_N . The actual diffusion realizes a strictly positive density on S_N , with $\rho_0 = 1$ because the initial point is uniform and independent of J . We use its C^3 spatial regularity and the joint continuity, up to time zero, of ρ_t , $\nabla_S \rho_t$, its forward-equation slope, and the spatial gradient of that slope. The forward equation holds on S_N at positive times and gives its right derivative at zero. All density, score, and current quantities used below are jointly measurable in the disorder and their time parameters; we establish their integrability over the disorder in the estimates below. We require no positivity or forward equation for an extension off S_N . Finally, Q_N is the tensor norm (3.25). Derivatives at time zero are always right derivatives; derivatives at positive times are ordinary derivatives.

5.1 Relative entropy and an exact fluctuation–dissipation identity

Recall from Section 3.1 that the Gibbs density $g_N = Z_{N,\beta}^{-1} e^{\beta H_N}$, with $Z_{N,\beta} = \int e^{\beta H_N} d\omega_N$, is reversible for $\mathcal{L}_N$. We write

$$v_t^{\text{cur}} = \nabla_S \log(\rho_t / g_N)$$

for the relative score of ρ_t ; the superscript distinguishes it from the tree vectors v_t of Section 7. The probability current of the diffusion is $j_t = \beta \rho_t \nabla_S H_N - \nabla_S \rho_t = -\rho_t v_t^{\text{cur}}$, which satisfies $\partial_t \rho_t + \text{div}_S j_t = 0$; we write all formulas below in terms of v_t^{cur} . We define the relative entropy, the relative Fisher information, and its normalized annealed average by

$$D_H(t) = \int \rho_t \log(\rho_t / g_N) d\omega_N, \quad I_H(t) = \int \|v_t^{\text{cur}}\|^2 \rho_t d\omega_N, \quad a_N(t) = \frac{1}{N} \mathbb{E}_J I_H(t).$$

Lemma 5.1 (Finite normalized entropy budget). *For every realization of the disorder and every $t \geq 0$,*

$$\partial_t \rho_t = \operatorname{div}_S(\rho_t v_t^{\text{cur}}), \quad D'_H(t) = -I_H(t), \quad \int_0^\infty I_H(t) \, dt \leq D_H(0). \quad (5.1)$$

The first identity is an equation on S_N , and both derivatives at $t = 0$ are taken from the right. Moreover,

$$\int_0^\infty a_N(t) \, dt \leq \frac{\beta^2}{2}, \quad \frac{D_H(0)}{N} \leq 2\beta Q_N. \quad (5.2)$$

Proof. Note that $\rho_t v_t^{\text{cur}} = \nabla_S \rho_t - \beta \rho_t \nabla_S H_N$. Then the forward equation $\partial_t \rho_t = \Delta_S \rho_t - \beta \operatorname{div}_S(\rho_t \nabla_S H_N)$ of the diffusion is the first identity in (5.1). Next, we differentiate the relative entropy. Using $\int \partial_t \rho_t \, d\omega_N = 0$ and integrating by parts, we find

$$D'_H(t) = \int \log(\rho_t / g_N) \operatorname{div}_S(\rho_t v_t^{\text{cur}}) \, d\omega_N = - \int \rho_t \|v_t^{\text{cur}}\|^2 \, d\omega_N = -I_H(t).$$

At positive times, we may differentiate under the integral because the forward slope is continuous and the density is strictly positive on the compact sphere. Since the entropy and Fisher information are continuous up to zero, the positive-time identity gives $D'_H(0+) = -I_H(0)$. The same argument, applied to the continuous forward slope, gives the right derivative of the density at zero. Integrating from 0 to T , we have $\int_0^T I_H(t) \, dt = D_H(0) - D_H(T) \leq D_H(0)$, where we used the fact that the relative entropy of two probability densities is nonnegative. Letting $T \rightarrow \infty$ proves the third assertion of (5.1).

We now bound the initial entropy. Since $\rho_0 = 1$, we have

$$D_H(0) = \log Z_{N,\beta} - \beta \int H_N \, d\omega_N.$$

By (3.26), we have $|H_N| \leq NQ_N$ on S_N . As ω_N is a probability measure, this implies that

$$-\beta NQ_N \leq \log Z_{N,\beta} \leq \beta NQ_N, \quad \left| \int H_N \, d\omega_N \right| \leq NQ_N.$$

This proves the second inequality in (5.2), and by Lemma 3.6, both $\log Z_{N,\beta}$ and $\int H_N \, d\omega_N$ are integrable over the disorder. For each $x \in S_N$, the variable $H_N(x)$ is a centered Gaussian of variance $N\zeta(1) = N$. By Fubini's theorem, we have $\mathbb{E}_J \int H_N \, d\omega_N = 0$ and

$$\mathbb{E}_J Z_{N,\beta} = \int \mathbb{E}_J e^{\beta H_N(x)} \, d\omega_N(x) = e^{\beta^2 N/2}.$$

Using Jensen's inequality, we obtain

$$\mathbb{E}_J D_H(0) = \mathbb{E}_J \log Z_{N,\beta} \leq \log \mathbb{E}_J Z_{N,\beta} = \frac{\beta^2 N}{2}.$$

Finally, by Tonelli's theorem and the third assertion of (5.1),

$$\int_0^\infty a_N(t) \, dt = \frac{1}{N} \mathbb{E}_J \int_0^\infty I_H(t) \, dt \leq \frac{1}{N} \mathbb{E}_J D_H(0) \leq \frac{\beta^2}{2},$$

which is the first inequality in (5.2). $\square$

We next express correlations and responses through the semigroup. Let $m_\tau = P_\tau \text{id}$ be the vector of functions $(P_\tau x_1, \dots, P_\tau x_N)$. By the Markov property, $m_\tau(x)$ is the conditional mean of $X_{s+\tau}$ given $X_s = x$ and J . Jensen's inequality and $\|X_{s+\tau}\| = \sqrt{N}$ imply that $\|m_\tau(x)\| \leq \sqrt{N}$. For $t \geq s$, we define the quenched correlation and the normalized impulse response by

$$c_H(t, s) = \frac{1}{N} \int \rho_s(x) \langle x, m_{t-s}(x) \rangle \, d\omega_N(x), \quad (5.3)$$

$$r_H(t, s) = \frac{1}{N} \int \rho_s \sum_{i=1}^N \Gamma(x_i, P_{t-s} x_i) \, d\omega_N. \quad (5.4)$$

By the Markov property again, $c_H(t, s) = N^{-1} \mathbb{E}[X_t \cdot X_s \mid J]$, and $\mathbb{E}_J c_H(t, s) = \mathbb{E} c_N(t, s)$. The function r_H is the response to an additive tangent force. Recall that a deterministic force $h(s) \in \mathbb{R}^N$ enters the linearized equation (3.5) as $P_{X_s} h(s)$, without a factor of β . Because $\nabla_s x_i = P_x e_i$, the actual mean response is expressed through the derivatives of the transition mean. For bounded continuous h , (3.7) and the expected projected-restart derivative identify the i th coordinate of the mean linearized response with

$$\int_0^t \int \rho_s \sum_{j=1}^N h_j(s) \Gamma(x_j, P_{t-s} x_i) \, d\omega_N \, ds.$$

In other words, $N r_H(t, s)$ is the trace of the impulse response matrix of the quenched mean. As in Remark 3.1, the actual tangent-projected restart is Fréchet differentiable into L^2 , and the bounded linear expectation operator commutes with that derivative. It follows that the expected trace of $U_N(t, s)$ is the mean of $\sum_i \Gamma(x_i, P_{t-s} x_i)(X_s)$. The actual waiting-state law is $\rho_s \, d\omega_N$, so this is exactly the integral in (5.4). Integrability follows from the fixed-disorder fourth-moment estimate (4.3). We set

$$\bar{c}_N = \mathbb{E}_J c_H, \quad \bar{r}_N = \mathbb{E}_J r_H.$$

By the discussion above, $\bar{r}_N$ is the annealed normalized response (3.8), and $r_H(s, s) = N^{-1} \int \rho_s \text{tr } P_x \, d\omega_N = \kappa_N$, where $\kappa_N = (N-1)/N$ as in Section 3.1.

Lemma 5.2 (Exact nonequilibrium fluctuation–dissipation identity). *For $0 \leq s < t$,*

$$r_H(t, s) - \partial_s c_H(t, s) = \frac{1}{N} \int \rho_s \langle m_{t-s}, v_s^{\text{cur}} \rangle \, d\omega_N, \quad (5.5)$$

where the derivative is taken with the terminal time t fixed, from the right when $s = 0$. In particular,

$$|\bar{r}_N(t, s) - \partial_s \bar{c}_N(t, s)|^2 \leq a_N(s), \quad (5.6)$$

and, for $0 \leq s \leq t$,

$$\left| 1 - \bar{c}_N(t, s) - \int_s^t \bar{r}_N(t, u) \, du \right|^2 \leq (t - s) \int_s^t a_N(u) \, du. \quad (5.7)$$

Proof. We begin with an integration by parts formula for the law at time s . For smooth functions f, g on S_N , we have

$$\int \rho_s f \mathcal{L}_N g \, d\omega_N = - \int \rho_s \Gamma(f, g) \, d\omega_N - \int \rho_s f \langle v_s^{\text{cur}}, \nabla_S g \rangle \, d\omega_N. \quad (5.8)$$

Indeed, integrating $\Delta_S g$ by parts against $\rho_s f$ gives $-\int \rho_s \Gamma(f, g) - \int f \langle \nabla_S \rho_s, \nabla_S g \rangle$, and the remaining term combines with the drift term of $\mathcal{L}_N$ because $\nabla_S \rho_s - \beta \rho_s \nabla_S H_N = \rho_s v_s^{\text{cur}}$.

Next, we differentiate (5.3) in s , using the forward equation $\partial_s \rho_s = \mathcal{L}_N^* \rho_s$, where $\mathcal{L}_N^*$ is the adjoint with respect to ω_N , and the backward equation $\partial_s P_{t-s} x_i = -\mathcal{L}_N P_{t-s} x_i$. This gives

$$\partial_s c_H(t, s) = \frac{1}{N} \int \rho_s \sum_{i=1}^N [\mathcal{L}_N(x_i P_{t-s} x_i) - x_i \mathcal{L}_N P_{t-s} x_i] \, d\omega_N.$$

Using the product rule $\mathcal{L}_N(fg) = f \mathcal{L}_N g + g \mathcal{L}_N f + 2\Gamma(f, g)$, we can write the bracket as $(P_{t-s} x_i) \mathcal{L}_N x_i + 2\Gamma(x_i, P_{t-s} x_i)$. Applying (5.8) with $f = P_{t-s} x_i$ and $g = x_i$ to the first of these two terms, we obtain

$$\partial_s c_H(t, s) = \frac{1}{N} \int \rho_s \sum_{i=1}^N [\Gamma(x_i, P_{t-s} x_i) - (P_{t-s} x_i) \langle v_s^{\text{cur}}, \nabla_S x_i \rangle] \, d\omega_N.$$

The first term is $r_H(t, s)$. In the second, we have $\sum_i (P_{t-s} x_i) \langle v_s^{\text{cur}}, P_x e_i \rangle = \langle m_{t-s}, v_s^{\text{cur}} \rangle$, because v_s^{cur} is tangent and $\nabla_S x_i = P_x e_i$. This proves (5.5), with its sign and coefficient. As a check of the normalization, consider $\beta = 0$. Then the uniform law is invariant, $\rho_s = 1$, and $v_s^{\text{cur}} = 0$. The coordinates satisfy $\Delta_S x_i = -\kappa_N x_i$, so that $c_H(t, s) = e^{-\kappa_N(t-s)}$ and $r_H(t, s) = \kappa_N e^{-\kappa_N(t-s)} = \partial_s c_H(t, s)$.

By the Cauchy–Schwarz inequality and $\|m_{t-s}\| \leq \sqrt{N}$, the square of the right side of (5.5) is at most $I_H(s)/N$. To average over the disorder, we need a bound on $\partial_s c_H$ that is integrable over J . In the formula for $\partial_s c_H$ before the integration by parts, the first term is $N^{-1} \int \rho_s \langle m_{t-s}, \mathcal{L}_N \text{id} \rangle \, d\omega_N$. Using $\mathcal{L}_N \text{id} = \beta \nabla_S H_N - \kappa_N \text{id}$ and the bound $\|\nabla_S H_N\| \leq p Q_N \sqrt{N}$ from (3.26), we see that its absolute value is at most $\kappa_N + \beta p Q_N$. The second term is $2N^{-1} \int \rho_s \text{tr}(D_S m_{t-s} P_x) \, d\omega_N$, and its absolute value is at most $2\|D_S m_{t-s}\|_{\text{op}} \leq$

$2e^{\beta p^2 Q_N(t-s)}$ by the gradient bound (5.17) below, whose proof uses only the Bochner formula. We conclude that

$$|\partial_s c_H(t, s)| \leq 1 + \beta p Q_N + 2e^{\beta p^2 Q_N(t-s)}, \quad |r_H(t, s)| \leq e^{\beta p^2 Q_N(t-s)}.$$

These bounds are integrable over the disorder by Lemma 3.6, uniformly for $t - s$ in compact sets, and dominated convergence justifies $\partial_s \bar{c}_N = \mathbb{E}_J \partial_s c_H$ for $0 < s < t$. The quenched and annealed correlations, responses, and currents are continuous on each closed causal row, so the identities for the earlier-time derivatives of the correlations at positive times extend to the right derivative at $s = 0$. By Jensen's inequality, we have

$$|\bar{r}_N(t, s) - \partial_s \bar{c}_N(t, s)|^2 \leq \mathbb{E}_J |r_H(t, s) - \partial_s c_H(t, s)|^2 \leq \frac{1}{N} \mathbb{E}_J I_H(s) = a_N(s),$$

which is (5.6).

For (5.7), note that $\bar{c}_N(t, t) = 1$ because $\|X_t\|^2 = N$. The function $u \mapsto \bar{c}_N(t, u)$ is continuous on $[s, t]$, with a bounded derivative on $[s, t]$; in particular, it is Lipschitz and is the integral of its derivative. Integrating (5.6) over $u \in [s, t]$, we obtain

$$\left| 1 - \bar{c}_N(t, s) - \int_s^t \bar{r}_N(t, u) \, du \right| \leq \int_s^t \sqrt{a_N(u)} \, du,$$

and one more application of the Cauchy–Schwarz inequality completes the proof. $\square$

5.2 An integrable Fisher majorant

Recall that the landscape cutoff A_p of Lemma 3.6 satisfies $\mathbb{P}(Q_N > A_p) \leq e^{-c_p N}$. We write $G_N = \{Q_N \leq A_p\}$ for the corresponding event on the disorder and set

$$K = 2\beta p^2 A_p.$$

The constant K depends only on p and β . We extract the subsequence in the following theorem once, before any later waiting-time sequence, stationary cluster, or finite-fork profile is chosen, and we keep the same strictly increasing dimension subsequence (N_j) for the rest of the paper. We take all limits in N at fixed times; the long-time statements concern only the limiting function a_F and the deterministic CK solution.

Theorem 5.3 (Fisher control of the backward FDT defect). *There exist a dimension subsequence (N_j) and a nonnegative, locally bounded, right-continuous function a_F on $[0, \infty)$ such that*

$$a_{N_j} \longrightarrow a_F \quad \text{in } L^1_{\text{loc}}([0, \infty)), \quad \int_0^\infty a_F(u) \, du \leq \frac{\beta^2}{2}. \quad (5.9)$$

The function $e^{-Kt}a_F(t)$ is nonincreasing. In particular, for $s \geq \varepsilon > 0$,

$$a_F(s) \leq \frac{K}{1 - e^{-K\varepsilon}} \int_{s-\varepsilon}^s a_F(u) \, du, \quad \sup_{s \geq T} a_F(s) \longrightarrow 0 \quad (T \rightarrow \infty). \quad (5.10)$$

For every $t > s > 0$, the earlier-time derivative $\partial_2 C(t, s) = \partial_s C(t, s)$ exists, is continuous, and satisfies

$$|\partial_2 C(t, s) - R(t, s)|^2 \leq a_F(s). \quad (5.11)$$

Every row $C(t, \cdot)$ is absolutely continuous on $[0, t]$, and for $t \geq v \geq u \geq 0$,

$$\left| C(t, v) - C(t, u) - \int_u^v R(t, s) \, ds \right| \leq \int_u^v \sqrt{a_F(s)} \, ds. \quad (5.12)$$

Finally, e_β is locally absolutely continuous, and

$$|e'_\beta(s)|^2 \leq p^2 A_p^2 a_F(s) \quad \text{for almost every } s \geq 0. \quad (5.13)$$

We write

$$\mu(ds) = a_F(s) \, ds, \quad \mu([0, \infty)) \leq \beta^2/2. \quad (5.14)$$

Proof. The proof has six steps. The first three construct a_F . In Step 1, we use the Bochner formula to show that, on the event G_N , the Fisher information grows at most like e^{Kt} . Then the average of the Fisher information over G_N , multiplied by e^{-Kt} , is a bounded nonincreasing function of t , and we use Helly's selection theorem to extract a subsequence along which these functions converge (Step 2). In Step 3, we show that the Fisher information on the complement of G_N has total mass tending to zero, which leads to (5.9) and (5.10). In the last three steps, we transfer the finite-dimensional estimates to the CK solution. In Step 4, we pass the integrated form of Lemma 5.2 to the limit, which gives (5.12), together with (5.11) for almost every s . In Step 5, we deduce from the Green covariance identity that $\partial_2 C$ is continuous, which upgrades the bound to every s . In Step 6, we treat the energy in the same way.

Step 1: the Bochner identity and growth of the Fisher information. Fix the disorder. Let $f_t = \rho_t/g_N$ and $\psi_t = \log f_t$, and let $\nu_N(dx) = g_N(x) \omega_N(dx)$ be the Gibbs measure. Then $v_t^{\text{cur}} = \nabla_S \psi_t$ and $I_H(t) = \int f_t \Gamma(\psi_t) \, d\nu_N$. Recall that the generator $\mathcal{L}_N$ is symmetric with respect to ν_N , with $\int f \mathcal{L}_N g \, d\nu_N = - \int \Gamma(f, g) \, d\nu_N$. In terms of f_t , the forward equation reads $\partial_t f_t = \mathcal{L}_N f_t$. As $\mathcal{L}_N(e^\psi) = e^\psi(\mathcal{L}_N \psi + \Gamma(\psi))$, we also have $\partial_t \psi_t = \mathcal{L}_N \psi_t + \Gamma(\psi_t)$. Differentiating I_H , we find

$$I'_H(t) = \int (\mathcal{L}_N f_t) \Gamma(\psi_t) \, d\nu_N + 2 \int f_t \Gamma(\psi_t, \mathcal{L}_N \psi_t) \, d\nu_N + 2 \int f_t \Gamma(\psi_t, \Gamma(\psi_t)) \, d\nu_N.$$

By symmetry, the first integral equals $\int f_t \mathcal{L}_N \Gamma(\psi_t) \, d\nu_N$. It also equals $-\int \Gamma(f_t, \Gamma(\psi_t)) \, d\nu_N = -\int f_t \Gamma(\psi_t, \Gamma(\psi_t)) \, d\nu_N$, because $\nabla_S f_t = f_t \nabla_S \psi_t$. The third integral is therefore $-2 \int f_t \mathcal{L}_N \Gamma(\psi_t) \, d\nu_N$, and we obtain

$$I'_H(t) = - \int f_t \mathcal{L}_N \Gamma(\psi_t) \, d\nu_N + 2 \int f_t \Gamma(\psi_t, \mathcal{L}_N \psi_t) \, d\nu_N = -2 \int f_t \Gamma_2(\psi_t) \, d\nu_N,$$

where $\Gamma_2(\psi) = \frac{1}{2}\mathcal{L}_N\Gamma(\psi) - \Gamma(\psi, \mathcal{L}_N\psi)$. For $\mathcal{L}_N = \Delta_S + \beta \langle \nabla_S H_N, \nabla_S \cdot \rangle$, the Bochner formula reads $\Gamma_2(\psi) = \|\nabla_S^2 \psi\|_{\text{HS}}^2 + (\text{Ric} - \beta \nabla_S^2 H_N)(\nabla_S \psi, \nabla_S \psi)$. Using $f_t \, d\nu_N = \rho_t \, d\omega_N$, we conclude that

$$I'_H(t) = -2 \int \rho_t \left\{ \|\nabla_S^2 \psi_t\|_{\text{HS}}^2 + (\text{Ric} - \beta \nabla_S^2 H_N)(\nabla_S \psi_t, \nabla_S \psi_t) \right\} d\omega_N. \quad (5.15)$$

This identity is asserted for $t > 0$. Its differentiations and integrations by parts use the spatial C^3 density, positivity on the compact sphere, and joint continuity of the forward slope and its spatial gradient. In particular, we compute the time derivative of the relative score by differentiating the intrinsic forward equation along the sphere. Continuity of I_H at zero will suffice to include the initial time in the resulting comparison.

For $N \geq 2$, the sphere S_N of radius $\sqrt{N}$ has $\text{Ric} = \frac{N-2}{N}$ times its metric, which is nonnegative, and $\|\nabla_S^2 H_N\|_{\text{op}} \leq p^2 Q_N$ by (3.26). Then (5.15) implies that $I'_H \leq 2\beta p^2 Q_N I_H$ at positive times. On the event G_N , we have $I'_H \leq K I_H$, so $e^{-Kt} I_H(t)$ is nonincreasing for $t > 0$. Its continuity at zero extends this monotonicity to $[0, \infty)$. We define

$$b_N(t) = \frac{1}{N} \mathbb{E}_J[\mathbf{1}_{G_N} I_H(t)], \quad \tilde{b}_N(t) = e^{-Kt} b_N(t).$$

As an average of nonincreasing functions, $\tilde{b}_N$ is nonincreasing; no differentiation under $\mathbb{E}_J$ is needed. At time zero, $\nabla_S \psi_0 = -\beta \nabla_S H_N$. Differentiating the covariance $N(x \cdot y/N)^p$ in both variables and setting $y = x \in S_N$, we see that $\mathbb{E}_J[\nabla H_N(x) \nabla H_N(x)^\top] = pI + \frac{p(p-1)}{N} x x^\top$. The projection P_x removes the second term, because $P_x x = 0$. We obtain $\mathbb{E}_J \|\nabla_S H_N(x)\|^2 = p \, \text{tr} P_x = p(N-1)$, and by Tonelli's theorem,

$$a_N(0) = \frac{\beta^2}{N} \int \mathbb{E}_J \|\nabla_S H_N\|^2 \, d\omega_N = \beta^2 p \frac{N-1}{N}.$$

We have shown that $\tilde{b}_N$ is nonincreasing and $0 \leq \tilde{b}_N(t) \leq \tilde{b}_N(0) \leq a_N(0) \leq \beta^2 p$ for all $t \geq 0$. We also note that the bound $I'_H \leq 2\beta p^2 Q_N I_H$ holds for every disorder. It implies $I_H(t) \leq I_H(0) e^{2\beta p^2 Q_N t} \leq \beta^2 p^2 N Q_N^2 e^{2\beta p^2 Q_N t}$, and by Lemma 3.6, a_N is finite and bounded on compact intervals, uniformly in N .

Step 2: Helly selection. We use the Helly diagonal-selection argument; compare the proof of [29, Theorem 25.9, p. 336]. The sequences $(\tilde{b}_N(\theta))_N$ are bounded, and a diagonal extraction over the positive rationals gives a subsequence (N_j) along which $\tilde{b}_{N_j}(\theta)$ converges for every rational $\theta > 0$. The limits form a nonincreasing function on the positive rationals with values in $[0, \beta^2 p]$. Let $\tilde{b}(t)$ be the right-continuous function on $[0, \infty)$ defined by the limit of these values as $\theta \downarrow t$ through the rationals. Let $t > 0$ be a continuity point of $\tilde{b}$. For rationals $\theta < t < \theta'$, we have $\tilde{b}_{N_j}(\theta') \leq \tilde{b}_{N_j}(t) \leq \tilde{b}_{N_j}(\theta)$ by monotonicity. Letting $j \rightarrow \infty$ and then $\theta \uparrow t$, $\theta' \downarrow t$, we find that $\tilde{b}_{N_j}(t) \rightarrow \tilde{b}(t)$. Since a

monotone function has at most countably many discontinuities, we conclude that $\tilde{b}_{N_j} \rightarrow \tilde{b}$ almost everywhere. We define

$$a_F(t) = e^{Kt} \tilde{b}(t).$$

Then a_F is nonnegative and right-continuous, $a_F(t) \leq \beta^2 p e^{Kt}$, and $e^{-Kt} a_F(t) = \tilde{b}(t)$ is nonincreasing. For every $M < \infty$, we have $0 \leq b_{N_j}(t) \leq \beta^2 p e^{KM}$ for $t \in [0, M]$, and by dominated convergence, $b_{N_j} \rightarrow a_F$ in $L^1([0, M])$. The value of a_F at zero is the right limit; no convergence of $b_N(0)$ to $a_F(0)$ is used. We have shown that

$$b_{N_j} \longrightarrow a_F \quad \text{in } L^1_{\text{loc}}([0, \infty)).$$

Step 3: the Fisher mass on the bad event. The difference $a_N - b_N = N^{-1} \mathbb{E}_J[\mathbf{1}_{G_N^c} I_H]$ is nonnegative. By the quenched budget in (5.1) and the bound $D_H(0) \leq 2\beta N Q_N$ of Lemma 5.1, we have

$$\begin{aligned} \int_0^\infty (a_N - b_N)(t) \, dt &= \frac{1}{N} \mathbb{E}_J \left[\mathbf{1}_{G_N^c} \int_0^\infty I_H(t) \, dt \right] \\ &\leq 2\beta \mathbb{E}_J[Q_N \mathbf{1}_{G_N^c}] \leq 2\beta (\mathbb{E}_J Q_N^2)^{1/2} \mathbb{P}(G_N^c)^{1/2}. \end{aligned}$$

The right side tends to zero by Lemma 3.6. Together with Step 2, this implies that $a_{N_j} \rightarrow a_F$ in $L^1_{\text{loc}}([0, \infty))$. For every M , we have $\int_0^M a_F = \lim_j \int_0^M a_{N_j} \leq \beta^2/2$ by Lemma 5.1, and monotone convergence in M proves (5.9).

The monotonicity of $e^{-Kt} a_F(t)$ means that $a_F(u) \geq e^{-K(s-u)} a_F(s)$ for $u \leq s$. Integrating this inequality over $u \in [s - \varepsilon, s]$, we have

$$\int_{s-\varepsilon}^s a_F(u) \, du \geq a_F(s) \int_0^\varepsilon e^{-Kw} \, dw = a_F(s) \frac{1 - e^{-K\varepsilon}}{K},$$

which is the first part of (5.10); here $K > 0$ because $\beta > 0$ and $A_p > 0$. Taking $\varepsilon = 1$, we have $\sup_{s \geq T} a_F(s) \leq K(1 - e^{-K})^{-1} \mu([T - 1, \infty))$ for $T \geq 1$, and the right side tends to zero because μ is a finite measure. This proves (5.10). We note that integrability of a_F alone would not give this uniform decay; it comes from the monotonicity of $e^{-Kt} a_F(t)$.

Step 4: passage of the row inequality to the limit. Fix $t \geq v \geq u \geq 0$. Recall from the proof of Lemma 5.2 that $\bar{c}_N(t, \cdot)$ is the integral of its derivative on $[u, t]$. Integrating (5.6) over $[u, v]$, we have

$$\left| \bar{c}_N(t, v) - \bar{c}_N(t, u) - \int_u^v \bar{r}_N(t, s) \, ds \right| \leq \int_u^v \sqrt{a_N(s)} \, ds.$$

We let $N = N_j \rightarrow \infty$ in each term. By Proposition 3.3, $\bar{c}_N(t, v) = \mathbb{E} C_N(t, v) \rightarrow C(t, v)$, and similarly at (t, u) . The response integral equals $\mathbb{E}[X_t \cdot (B_v - B_u)] / (\sqrt{2}N)$ by Lemma 3.4; it is a difference of the Brownian observables of Proposition 3.3, whose expectations

converge to $\int_u^v R(t, s) \, ds$. For the right side, we use the elementary inequality $|\sqrt{x} - \sqrt{y}|^2 \leq |x - y|$ for $x, y \geq 0$. On every compact interval I , it yields

$$\int_I |\sqrt{a_{N_j}} - \sqrt{a_F}|^2 \leq \int_I |a_{N_j} - a_F| \longrightarrow 0,$$

and by the Cauchy–Schwarz inequality, $\sqrt{a_{N_j}} \rightarrow \sqrt{a_F}$ in $L^1(I)$. This proves (5.12) for every $t \geq v \geq u \geq 0$. We fixed the subsequence in Step 2, before choosing the endpoints, and the correlations and the primitives of the response converge along the full sequence of dimensions. No further extraction is involved.

Next, fix t and let $F_t(s) = C(t, s) - \int_0^s R(t, w) \, dw$ for $0 \leq s \leq t$. By (5.12), we have, for every finite family of disjoint intervals $(u_i, v_i) \subset [0, t]$,

$$\sum_i |F_t(v_i) - F_t(u_i)| \leq \int_{\cup_i (u_i, v_i)} \sqrt{a_F(s)} \, ds.$$

The right side is small when the total length of the intervals is small, because $\sqrt{a_F}$ is locally bounded. Hence F_t is absolutely continuous. As $s \mapsto \int_0^s R(t, w) \, dw$ is continuously differentiable, the row $C(t, \cdot)$ is absolutely continuous on $[0, t]$. At every point $s \in (0, t)$ where F_t is differentiable and which is a Lebesgue point of $\sqrt{a_F}$, dividing (5.12) by $v - u$ and letting the interval shrink to s , we have $|F'_t(s)| \leq \sqrt{a_F(s)}$. We have shown that

$$|\partial_2 C(t, s) - R(t, s)|^2 \leq a_F(s) \quad \text{for almost every } s \in (0, t).$$

The exceptional set may depend on t ; this causes no difficulty, since we carry out Step 5 on each fixed row with the fixed function a_F .

Step 5: the continuous representative of the earlier derivative. Let $\dot{R}(s, u) = \partial_1 R(s, u)$. Recall from (3.9) and Proposition 4.5 that $\dot{R}$ is continuous on every compact causal triangle, with one-sided values on the diagonal. For $t > s$, the Green identity (4.10) reads

$$\begin{aligned} C(t, s) &= R(t, 0)R(s, 0) + 2 \int_0^s R(t, u)R(s, u) \, du \\ &\quad + \beta^2 p \int_0^t R(t, u) \int_0^s R(s, v)C(u, v)^{p-1} \, dv \, du. \end{aligned}$$

The variable s enters only through $R(s, 0)$, $R(s, u)$, $R(s, v)$, and the upper limits of the integrals over u and v . Differentiating in s and using $R(s, s) = 1$, we obtain, for $t > s > 0$,

$$\begin{aligned} \partial_2 C(t, s) &= R(t, 0)\dot{R}(s, 0) + 2R(t, s) + 2 \int_0^s R(t, u)\dot{R}(s, u) \, du \\ &\quad + \beta^2 p \int_0^t R(t, u) \left\{ C(u, s)^{p-1} + \int_0^s \dot{R}(s, v)C(u, v)^{p-1} \, dv \right\} \, du. \end{aligned} \quad (5.16)$$

All integrands are continuous on the fixed compact sets of integration, which justifies differentiation under the integral sign. No derivative of $C(u, v)^{p-1}$ is needed, because

u and v are integration variables that do not depend on s . The right side of (5.16) is jointly continuous on the open triangle $t > s > 0$. Hence, for each fixed t , the row $C(t, \cdot)$ is continuously differentiable on $(0, t)$, and its classical derivative agrees with the almost-everywhere derivative of Step 4.

Fix $t > s > 0$, and choose points $s_n \downarrow s$ with $s_n < t$ at which the bound of Step 4 holds. Then $|\partial_2 C(t, s_n) - R(t, s_n)|^2 \leq a_F(s_n)$. As $n \rightarrow \infty$, the left side converges to $|\partial_2 C(t, s) - R(t, s)|^2$ by the continuity of (5.16) and of R , and $a_F(s_n) \rightarrow a_F(s)$ because a_F is right-continuous. This proves (5.11).

Step 6: the energy. Let $e_H(s) = N^{-1} \int \rho_s H_N d\omega_N$, so that $e_N(s) = \mathbb{E}_J e_H(s) = N^{-1} \mathbb{E} H_N(X_s)$. By the first identity in (5.1) and an integration by parts, we have

$$e'_H(s) = -\frac{1}{N} \int \rho_s \langle v_s^{\text{cur}}, \nabla_s H_N \rangle d\omega_N.$$

Using $\|\nabla_s H_N\| \leq pQ_N \sqrt{N}$ and the Cauchy–Schwarz inequality, we find that $|e'_H(s)| \leq pQ_N \sqrt{I_H(s)/N}$. We split the average over the disorder according to G_N . On G_N , we have $Q_N \leq A_p$, and Jensen’s inequality yields $\mathbb{E}_J[\mathbf{1}_{G_N} pQ_N \sqrt{I_H(s)/N}] \leq pA_p \sqrt{b_N(s)}$. On the complement, by the Cauchy–Schwarz inequality, $\mathbb{E}_J[\mathbf{1}_{G_N^c} pQ_N \sqrt{I_H(s)/N}] \leq \sqrt{\eta_N(s)}$, where

$$\eta_N(s) = \frac{p^2}{N} \mathbb{E}_J[Q_N^2 I_H(s) \mathbf{1}_{G_N^c}].$$

By (5.1) and $D_H(0) \leq 2\beta NQ_N$, the total mass of η_N satisfies

$$\int_0^\infty \eta_N(s) ds \leq 2\beta p^2 \mathbb{E}_J[Q_N^3 \mathbf{1}_{G_N^c}] \leq 2\beta p^2 (\mathbb{E}_J Q_N^6)^{1/2} \mathbb{P}(G_N^c)^{1/2} \rightarrow 0.$$

For fixed $u \leq v$, Tonelli’s theorem and the two bounds above give

$$|e_N(v) - e_N(u)| \leq \int_u^v \mathbb{E}_J |e'_H(s)| ds \leq pA_p \int_u^v \sqrt{b_N(s)} ds + \left((v-u) \int_0^\infty \eta_N(s) ds \right)^{1/2}.$$

We let $N = N_j \rightarrow \infty$. By Proposition 4.1, $e_N(v) \rightarrow e_\beta(v)$ and $e_N(u) \rightarrow e_\beta(u)$; by Step 2 and the argument of Step 4, $\int_u^v \sqrt{b_{N_j}} \rightarrow \int_u^v \sqrt{a_F}$; and the last term tends to zero. We conclude that

$$|e_\beta(v) - e_\beta(u)| \leq pA_p \int_u^v \sqrt{a_F(s)} ds, \quad 0 \leq u \leq v.$$

As in Step 4, we see that e_β is locally absolutely continuous and that $|e'_\beta(s)| \leq pA_p \sqrt{a_F(s)}$ at almost every s , which is (5.13). Here we use only the convergence of energies at fixed times, not the convergence of their derivatives. $\square$

We note two features of Theorem 5.3. First, the majorant a_F is produced by one subsequence of dimensions and need not be unique; the inequalities it satisfies concern

the unique deterministic CK solution; the bounds on the defect hold at every pair of times, and the bound on the derivative of the energy holds for almost every time. Second, (5.11) and (5.10) imply that the defect $\partial_2 C(t, s) - R(t, s)$ tends to zero uniformly in t as $s \rightarrow \infty$. This does not identify the slow fluctuation–dissipation ratio: both R and $\partial_2 C$ may tend to zero while $R/\partial_2 C$ has a nontrivial limit, and the defect may accumulate over a diverging interval of earlier times. For $\beta > \beta_d$, Section 12 shows that a positive amount of defect does accumulate in this way.

The bound (5.11) also holds at the two ends of each row, with one-sided derivatives. We record this for later use.

Remark 5.4 (The ends of a row). Let $t > 0$. The one-sided derivatives $\partial_2 C(t, 0+)$ and $\partial_2 C(t, t-)$ of $s \mapsto C(t, s)$ exist, and

$$|\partial_2 C(t, 0+) - R(t, 0)|^2 \leq a_F(0), \quad \partial_2 C(t, t-) = R(t, t) = 1.$$

In other words, (5.11) holds for $0 \leq s \leq t$, with derivatives at $s = 0$ and $s = t$ taken from the interior of the row. Indeed, let $\Delta(t, s)$ denote the right side of (5.16). It is continuous for $0 \leq s \leq t$, and by the mean value theorem, the one-sided derivatives at the ends of the row exist and equal $\Delta(t, 0)$ and $\Delta(t, t)$. The first bound follows by choosing $s_n \downarrow 0$ in (5.11) and using the right continuity of a_F at zero. For the second, we differentiate the diagonal Green identity

$$1 = C(t, t) = R(t, 0)^2 + 2 \int_0^t R(t, u)^2 \, du + \beta^2 p \int_0^t \int_0^t R(t, u) R(t, v) C(u, v)^{p-1} \, dv \, du$$

in t . Using $R(t, t) = 1$ and the symmetry of C , we find that the initial, white-noise, and colored-noise terms have derivatives $2R(t, 0)\dot{R}(t, 0)$, $2 + 4 \int_0^t R(t, u)\dot{R}(t, u) \, du$, and twice the colored-noise term of $\Delta(t, t)$, respectively. Comparing with (5.16), we see that the derivative of the right side equals $2\Delta(t, t) - 2$; the white-noise boundary term is counted once on the diagonal, while $2\Delta(t, t)$ counts it twice. Since the left side is constant, $\Delta(t, t) = 1$.

5.3 Flatness in the waiting time

The bound (5.11) controls the earlier observation time with the terminal time fixed. To compare two correlations with the same lag at different waiting times, we also need to control the effect of the probability current on a conditional-mean observable. A bound of this kind for one-time observables appears in [54, arXiv version 2, equation (15), p. 3].

Lemma 5.5 (Semigroup gradient bound and the waiting-time derivative). *For every $\tau \geq 0$ and $x \in S_N$,*

$$\|D_S m_\tau(x)\|_{\text{op}} \leq e^{\beta p^2 Q_N \tau}. \quad (5.17)$$

Let $\Psi_\tau(x) = N^{-1} \langle x, m_\tau(x) \rangle$. Then

$$\|\nabla_S \Psi_\tau(x)\| \leq N^{-1/2} (1 + e^{\beta p^2 Q_N \tau}), \quad (5.18)$$

$$\frac{d}{dw} c_H(w + \tau, w) = - \int \rho_w \langle v_w^{\text{cur}}, \nabla_S \Psi_\tau \rangle d\omega_N. \quad (5.19)$$

The waiting-time identity holds for all $w \geq 0$, with a right derivative at $w = 0$. For $0 \leq \tau \leq L$ and $0 \leq w_1 \leq w_2$,

$$|C(w_2 + \tau, w_2) - C(w_1 + \tau, w_1)| \leq B_L \int_{w_1}^{w_2} \sqrt{a_F(u)} du, \quad B_L = 1 + e^{\beta p^2 A_p L}. \quad (5.20)$$

Proof. We begin with the gradient bound. Let $K_N = \beta p^2 Q_N$. Recall from Step 1 of the proof of Theorem 5.3 that $\text{Ric} \geq 0$ and $\|\nabla_S^2 H_N\|_{\text{op}} \leq p^2 Q_N$. By the Bochner formula, we have the curvature bound $\Gamma_2(f) \geq -K_N \Gamma(f)$ for smooth f . Fix τ , x , and a smooth function f , and let $\Phi(u) = P_u \Gamma(P_{\tau-u} f)(x)$ for $0 \leq u \leq \tau$. Using $\partial_u \Gamma(P_{\tau-u} f) = -2\Gamma(P_{\tau-u} f, \mathcal{L}_N P_{\tau-u} f)$, we find

$$\Phi'(u) = P_u [\mathcal{L}_N \Gamma(P_{\tau-u} f) - 2\Gamma(P_{\tau-u} f, \mathcal{L}_N P_{\tau-u} f)](x) = 2P_u \Gamma_2(P_{\tau-u} f)(x) \geq -2K_N \Phi(u),$$

where we used that P_u preserves order. By Gronwall's inequality, $\Phi(\tau) \geq e^{-2K_N \tau} \Phi(0)$, that is,

$$\|\nabla_S P_\tau f\|^2 \leq e^{2K_N \tau} P_\tau \|\nabla_S f\|^2.$$

This is the classical gradient bound under a lower bound on the curvature; see [10, Theorem 3.2.3, p. 144]. We apply it to $f_z(x) = \langle z, x \rangle$ for a unit vector $z \in \mathbb{R}^N$. Then $\|\nabla_S f_z\| = \|P_x z\| \leq 1$ and $\nabla_S P_\tau f_z = \nabla_S \langle z, m_\tau \rangle = (D_S m_\tau)^* z$. By the gradient bound, we now have $\|(D_S m_\tau(x))^* z\| \leq e^{K_N \tau}$. Taking the supremum over unit vectors z proves (5.17). This bound has no factor depending on the dimension; applying the scalar estimate only to the N coordinate functions and summing would lose such a factor.

Next, we have $\nabla_S \Psi_\tau = N^{-1} [P_x m_\tau + (D_S m_\tau)^* x]$. As $\|x\| = \sqrt{N}$ and $\|m_\tau\| \leq \sqrt{N}$, (5.17) implies (5.18). At fixed lag τ , we have $c_H(w + \tau, w) = \int \rho_w \Psi_\tau d\omega_N$ by (5.3). Differentiating at $w > 0$ and using the first identity in (5.1) and an integration by parts, we obtain (5.19). The current is continuous up to zero, and the exact current integral over a waiting-time interval therefore gives the same identity with the right derivative at $w = 0$.

It remains to prove (5.20). Let $0 \leq \tau \leq L$. On the event G_N , we have $Q_N \leq A_p$, and by (5.18), (5.19), and the Cauchy-Schwarz inequality,

$$\left| \frac{d}{dw} c_H(w + \tau, w) \right| \leq B_L \sqrt{I_H(w)/N}.$$

Averaging over G_N and using Jensen's inequality, we obtain $\mathbb{E}_J[1_{G_N} \sqrt{I_H(w)/N}] \leq \sqrt{b_N(w)}$, with b_N as in Step 1 of the proof of Theorem 5.3. On the complement of

G_N , we use only $|c_H| \leq 1$ at the two endpoints; no control of the gradient is needed there. Combining these bounds, we find

$$|\bar{c}_N(w_2 + \tau, w_2) - \bar{c}_N(w_1 + \tau, w_1)| \leq B_L \int_{w_1}^{w_2} \sqrt{b_N(w)} \, dw + 2\mathbb{P}(G_N^c).$$

We now let $N = N_j \rightarrow \infty$. The correlations converge by Proposition 3.3, $\int_{w_1}^{w_2} \sqrt{b_{N_j}} \rightarrow \int_{w_1}^{w_2} \sqrt{a_F}$ by Steps 2 and 4 of the proof of Theorem 5.3, and $\mathbb{P}(G_N^c) \leq e^{-c_p N} \rightarrow 0$. This proves (5.20). $\square$

We now collect the uniform late estimates that follow from Theorem 5.3 and Lemma 5.5. Since μ is a finite measure, they show that, on late windows of bounded length, the integrated FDT relation holds, and the correlation at fixed lag and the energy do not depend on the waiting time, up to errors that vanish as the window moves to infinity.

Corollary 5.6 (Uniform late estimates). *Fix $L, \Theta \geq 0$. For every $T \geq 0$,*

$$\sup_{s \geq T, 0 \leq \tau \leq L} \left| 1 - C(s + \tau, s) - \int_s^{s+\tau} R(s + \tau, u) \, du \right| \leq \sqrt{L\mu([T, \infty))}, \quad (5.21)$$

and for every $T \geq \Theta$,

$$\sup_{s \geq T, |\theta| \leq \Theta, 0 \leq \tau \leq L} |C(s + \theta + \tau, s + \theta) - C(s + \tau, s)| \leq B_L \sqrt{\Theta\mu([T - \Theta, \infty))}, \quad (5.22)$$

$$\sup_{s \geq T, |\theta| \leq \Theta} |e_\beta(s + \theta) - e_\beta(s)| \leq pA_p \sqrt{\Theta\mu([T - \Theta, \infty))}. \quad (5.23)$$

The three right sides tend to zero as $T \rightarrow \infty$ with L and Θ fixed.

Proof. For (5.21), we use (5.12) on the row $t = s + \tau$ between $u = s$ and $v = t$, together with $C(t, t) = 1$. Its right side is at most $\sqrt{\tau\mu([s, s + \tau])} \leq \sqrt{L\mu([T, \infty))}$ by the Cauchy–Schwarz inequality. For (5.22), we apply (5.20) with w_1 and w_2 the smaller and the larger of s and $s + \theta$. The interval $[w_1, w_2]$ has length $|\theta| \leq \Theta$ and lies in $[T - \Theta, \infty)$, and the Cauchy–Schwarz inequality gives the bound as before. The proof of (5.23) is the same, using the increment bound $|e_\beta(v) - e_\beta(u)| \leq pA_p \int_u^v \sqrt{a_F}$ from Step 6 of the proof of Theorem 5.3. The last assertion holds because $\mu([T, \infty)) \rightarrow 0$ as $T \rightarrow \infty$. $\square$

5.4 Stationary translated clusters

Recall the translated kernels C_T , R_T , and λ_T of (4.22), with correlations considered on squares of offsets and responses only on closed causal triangles $\{a \geq b\}$. We say that

a triple $(C_\infty, R_\infty, \lambda_\infty)$ is a *translated cluster* if there exists a sequence $T_n \rightarrow \infty$ such that $C_{T_n} \rightarrow C_\infty$, $R_{T_n} \rightarrow R_\infty$, and $\lambda_{T_n} \rightarrow \lambda_\infty$ uniformly on compact sets of offsets in their respective domains.

Proposition 5.7 (Stationary FDT translated clusters). *The translated family (C_T, R_T, λ_T) is precompact for locally uniform convergence on compact offset sets. Every translated cluster has the form*

$$C_\infty(a, b) = c(|a - b|), \quad R_\infty(a, b) = r(a - b) \quad (a \geq b), \quad \lambda_\infty(a) = \ell,$$

where ℓ is a constant and

$$c(0) = 1, \quad r(0) = 1, \quad c' = -r, \quad 0 \leq c \leq 1, \quad r \geq 0. \quad (5.24)$$

In particular, c is nonincreasing, its limit $q = \lim_{\tau \rightarrow \infty} c(\tau)$ belongs to $[0, 1)$, and $\int_0^\infty r(u) \, du = 1 - q$. Along the same sequence T_n , the energy translates $e_\beta(T_n + a)$ converge locally uniformly to a constant e_∞ , and $\ell = 1 + p\beta e_\infty$. Moreover, r is continuously differentiable on $[0, \infty)$ and satisfies the closed stationary response equation

$$r'(\tau) = -\ell r(\tau) + \beta^2 p(p-1) \int_0^\tau c(\tau-u)^{p-2} r(\tau-u) r(u) \, du. \quad (5.25)$$

We call the triple (c, r, ℓ) of Proposition 5.7 a *stationary cluster*, and q its *plateau*. Different sequences T_n may a priori lead to different stationary clusters.

Proof. The proof has four steps. We first recall compactness. We then show that the correlation of every translated cluster is stationary, and that its response is stationary and satisfies FDT. Finally, we identify the multiplier and pass to the limit in the response equation. The order of limits is the same throughout: we have already taken the dimension limit in Theorem 5.3, the window of offsets is fixed, and only $T_n \rightarrow \infty$ varies.

Step 1: compactness. The precompactness of the translated family is proved in Proposition 4.6. The bounds $0 \leq C_T \leq 1$ and $R_T \geq 0$, the symmetry of C_T , and the diagonal values $C_T(a, a) = R_T(a, a) = 1$ pass to every translated cluster.

Step 2: stationarity of the correlation. Fix a sequence $T_n \rightarrow \infty$ along which the translated family converges to a translated cluster $(C_\infty, R_\infty, \lambda_\infty)$. Let $a \geq b$ and θ be fixed offsets. We have

$$C_{T_n}(a + \theta, b + \theta) - C_{T_n}(a, b) = C(s_n + \theta + \tau, s_n + \theta) - C(s_n + \tau, s_n),$$

where $s_n = T_n + b$ and $\tau = a - b$. Once $s_n \geq |\theta|$, we may apply (5.22) with $L = \tau$, $\Theta = |\theta|$, and $T = s_n$, and its right side tends to zero because $s_n \rightarrow \infty$. We conclude that $C_\infty(a + \theta, b + \theta) = C_\infty(a, b)$. Let $c(\tau) = C_\infty(\tau, 0)$ for $\tau \geq 0$. Taking $\theta = -b$, we have $C_\infty(a, b) = c(a - b)$ for $a \geq b$, and by symmetry $C_\infty(a, b) = c(|a - b|)$. The function c is continuous, $0 \leq c \leq 1$, and $c(0) = 1$.

Step 3: the response and FDT. Let $a \geq b$. By (5.21) with $s = T_n + b$, $\tau = a - b$, and $L = a - b$, we have

$$\left| 1 - C_{T_n}(a, b) - \int_b^a R_{T_n}(a, u) \, du \right| \leq \sqrt{(a - b) \mu([T_n + b, \infty))} \longrightarrow 0.$$

Since the response converges uniformly on the causal triangle containing the segment $\{(a, u) : b \leq u \leq a\}$, we may pass to the limit inside the integral. This gives

$$\int_b^a R_\infty(a, u) \, du = 1 - c(a - b).$$

For $a = 0$ and $b = -\tau$, this reads $1 - c(\tau) = \int_{-\tau}^0 R_\infty(0, u) \, du$. As $R_\infty(0, \cdot)$ is continuous on $(-\infty, 0]$, the function c is continuously differentiable on $[0, \infty)$, with $c'(\tau) = -R_\infty(0, -\tau)$ and a one-sided derivative at $\tau = 0$. Differentiating the general identity in b with a fixed, we have $R_\infty(a, b) = -c'(a - b)$ for $a \geq b$. We set $r = -c'$. Then r is continuous and nonnegative, $r(0) = R_\infty(0, 0) = 1$, and $R_\infty(a, b) = r(a - b)$. This proves that the response is stationary and satisfies FDT; neither property was assumed.

As $r \geq 0$ and $c \geq 0$, the function c is nonincreasing and bounded below, and the limit q exists. As r is continuous with $r(0) = 1$, there exists $\delta > 0$ such that $r \geq 1/2$ on $[0, \delta]$. Then $q \leq c(\delta) \leq 1 - \delta/2 < 1$. Integrating $c' = -r$ gives $\int_0^\infty r(u) \, du = \lim_{\tau \rightarrow \infty} (1 - c(\tau)) = 1 - q$.

Step 4: the multiplier and the response equation. By (4.1), we have $e_\beta(t) = (\lambda(t) - 1)/(p\beta)$ at every time t . Then $e_\beta(T_n + a) \rightarrow (\lambda_\infty(a) - 1)/(p\beta)$ locally uniformly in a . For fixed offsets a and a' , (5.23) with $s = T_n + a'$, $\theta = a - a'$, and $\Theta = |a - a'|$ implies that $e_\beta(T_n + a) - e_\beta(T_n + a') \rightarrow 0$. The limit of the energy translates is therefore a constant e_∞ , and λ_∞ is the constant $\ell = 1 + p\beta e_\infty$.

Finally, we pass to the limit in the response equation. Fix $\tau \geq 0$ and let $s = T_n$. Integrating (3.9) in its first variable from s to $s + \tau$, using $R(s, s) = 1$, and substituting $u = s + u'$ in the memory integral, we obtain

$$\begin{aligned} R(s + \tau, s) - 1 &= \int_0^\tau \left[-\lambda(s + v)R(s + v, s) + \beta^2 p(p - 1) \int_0^v C(s + v, s + u')^{p-2} \right. \\ &\quad \left. \times R(s + v, s + u')R(s + u', s) \, du' \right] dv. \end{aligned}$$

Here we used $\xi''(x) = p(p - 1)x^{p-2}$. All arguments of C , R , and λ are translated times with offsets in $[0, \tau]$, in a fixed causal triangle. The memory integral of (3.9) runs only over times between the two observation times, and no other part of the history enters. By the locally uniform convergence along T_n , together with $C_\infty(a, b) = c(|a - b|)$, $R_\infty(a, b) = r(a - b)$, and $\lambda_\infty = \ell$, we find

$$r(\tau) = 1 + \int_0^\tau \left[-\ell r(v) + \beta^2 p(p - 1) \int_0^v c(v - u)^{p-2} r(v - u)r(u) \, du \right] dv.$$

The integrand is continuous in v , which implies that r is continuously differentiable on $[0, \infty)$ and satisfies (5.25), with $r'(0) = -\ell$. This completes the proof. $\square$

Proposition 5.7 describes every translated cluster, but it does not determine the plateau q or the multiplier ℓ , it does not show that the whole translated family converges, and it says nothing about the correlation with a fixed early state. We answer these questions in Section 8, using the scalar equation of Section 6 and the causal covariance representation of Section 7.

PART II

Plateau selection, forgetting, and correlation aging

The scalar equation of a stationary cluster

The CK response equation (3.9) is causal, in the sense that its memory integral runs only over times between the earlier and the later observation time. When both times are translated to late times and their lag stays bounded, the equation involves only the translated kernels on a bounded window, and it passes to every translated cluster. This is how we obtained the closed equation (5.25) in Proposition 5.7. The correlation equation (3.10) is different. Its memory integrals run over the whole past, and the stationary limit does not control their older part. In this section, we extract from the closed response equation the information that it carries about a stationary cluster (c, r, ℓ) .

The main result, Proposition 6.1, is a scalar memory equation for c . It determines the multiplier ℓ in terms of the plateau q , and it gives a strict sign condition on the interval $(q, 1)$, which implies the inequality $\beta^2 p(p-1)q^{p-2}(1-q)^2 \leq 1$. We also show that the response equation with a prescribed multiplier has at most one solution (Lemma 6.2). We use only Proposition 5.7. In Section 8, the multiplier formula and the sign condition enter the proofs of marginality (Lemma 8.3) and of the selection of the plateau (Theorem 8.5), and the uniqueness lemma identifies the stationary profile in Proposition 8.4. The scalar equation (6.1) is also the starting point of Section 16.

The scalar memory equation (6.1) is the equation classified by Dembo, Guionnet, and Mazza [60, arXiv version 1, Proposition 1.4, p. 5, and Section 5, pp. 17–20]. For $b > 0$ and a nondecreasing function $\phi: [0, 1] \rightarrow \mathbb{R}$ with Lipschitz continuous derivative and $\sup_{0 \leq y \leq 1} \phi(y)(1-y) \geq b$, they characterized the plateau of its solution through the largest admissible root and proved the corresponding stability inequality. The sign condition (6.3) and the inequality (6.4) already follow from their scalar classification; here we obtain them by a direct comparison of the memory term. These hypotheses hold for the polynomial ϕ of (6.1), with $b = 1$ by (6.2), but we do not use their result. We derive the equation and its consequences directly from Proposition 5.7, for two reasons. Our response is normalized differently, and Section 8 needs the constants and the strict inequality in the form given here. We also note that their reduction of the CK solution to the scalar equation at all temperatures [60, arXiv version 1, Proposition 6.1, pp. 21–22] assumes the asymptotic conditions (6.5) and (6.6) of that paper. Here, Proposition 5.7 applies to every stationary cluster without such assumptions.

Throughout this section, $p \geq 3$ and $\beta > 0$ are fixed. Derivatives of stationary profiles at $t = 0$ are right derivatives on $[0, \infty)$.

Proposition 6.1 (Scalar constraints). *Let (c, r, ℓ) be a stationary cluster with plateau q , as in Proposition 5.7, and let $B = \beta^2 p$ and $\gamma = \ell - B$. Then $r = -c'$ is strictly positive on $[0, \infty)$, and*

$$c'(t) = -1 + \int_0^t \phi(c(t-u))r(u) \, du, \quad \phi(y) = \gamma + By^{p-1}. \quad (6.1)$$

The plateau $q = \lim_{t \rightarrow \infty} c(t)$ belongs to $[0, 1)$ and satisfies

$$\phi(q)(1-q) = 1, \quad \ell = L(q), \quad \text{where } L(q) = \frac{1}{1-q} + \beta^2 p(1-q^{p-1}). \quad (6.2)$$

Moreover, if $F(y) = \phi(y)(1-y) - 1$, then

$$F(q) = 0, \quad F(y) < 0 \quad (q < y < 1). \quad (6.3)$$

In particular,

$$\beta^2 p(p-1)q^{p-2}(1-q)^2 \leq 1. \quad (6.4)$$

Proof. We first show that r is positive and integrate the response equation once, which gives (6.1). We then let $t \rightarrow \infty$ in (6.1) to identify the multiplier. Finally, we compare the memory integral in (6.1) with its value at the current correlation, which yields the sign condition.

Step 1: positivity and the scalar equation. With $\beta^2 p(p-1) = B(p-1)$, the stationary response equation (5.25) reads

$$r'(t) = -\ell r(t) + B(p-1) \int_0^t c(t-u)^{p-2} r(t-u)r(u) \, du.$$

By variation of constants and $r(0) = 1$, we have

$$r(t) = e^{-\ell t} + B(p-1) \int_0^t e^{-\ell(t-v)} \int_0^v c(v-u)^{p-2} r(v-u)r(u) \, du \, dv.$$

Every factor in the double integral is nonnegative, because $c \geq 0$ and $r \geq 0$. This implies that $r(t) \geq e^{-\ell t} > 0$. Then c is strictly decreasing, and $c(t) > q$ for every $t \geq 0$.

Next, we integrate the response equation from 0 to t . Using $r(0) = 1$ and $\int_0^t r(u) \, du = 1 - c(t)$, we have

$$r(t) - 1 = -\ell(1 - c(t)) + B(p-1) \int_0^t \int_0^v c(v-u)^{p-2} r(v-u)r(u) \, du \, dv.$$

The double integral is over the bounded triangle $\{0 \leq u \leq v \leq t\}$, where the integrand is continuous; Fubini's theorem applies there without any assumption on the behavior of c and r at infinity. Exchanging the order of integration and substituting $w = v - u$, we find

$$\begin{aligned} & (p-1) \int_0^t \int_0^v c(v-u)^{p-2} r(v-u)r(u) \, du \, dv \\ &= \int_0^t r(u) \left[(p-1) \int_0^{t-u} c(w)^{p-2} r(w) \, dw \right] du. \end{aligned}$$

As $c' = -r$ and $c(0) = 1$, the derivative of $c(w)^{p-1}$ is $-(p-1)c(w)^{p-2}r(w)$, and the bracket equals $1 - c(t-u)^{p-1}$. Inserting this into the previous identity, we obtain

$$r(t) = 1 - (\ell - B)(1 - c(t)) - B \int_0^t c(t-u)^{p-1} r(u) \, du.$$

Finally, we substitute $r(t) = -c'(t)$ on the left and $1 - c(t) = \int_0^t r(u) \, du$ on the right. Recalling that $\gamma = \ell - B$, we arrive at

$$c'(t) = -1 + \int_0^t (\gamma + Bc(t-u)^{p-1}) r(u) \, du,$$

which is (6.1).

Step 2: the multiplier. By Proposition 5.7, the limit q exists, $q < 1$, and $\int_0^\infty r(u) \, du = 1 - q$. We claim that the integral in (6.1) converges to $\phi(q)(1 - q)$ as $t \rightarrow \infty$. Let $M_\phi = \sup_{[0,1]} |\phi|$, and fix $U > 0$. For $t > U$, we split the integral at U and find

$$\begin{aligned} & \left| \int_0^t \phi(c(t-u)) r(u) \, du - \phi(q)(1 - q) \right| \\ & \leq \sup_{0 \leq u \leq U} |\phi(c(t-u)) - \phi(q)| \int_0^U r(u) \, du + 2M_\phi \int_U^\infty r(u) \, du. \end{aligned}$$

Here we used $0 \leq c \leq 1$ and $\int_0^\infty r = 1 - q$. For $0 \leq u \leq U$, we have $q \leq c(t-u) \leq c(t-U)$. Then $c(t-u) \rightarrow q$ uniformly in $u \in [0, U]$ as $t \rightarrow \infty$, and the first term tends to zero because ϕ is continuous. Letting $t \rightarrow \infty$ and then $U \rightarrow \infty$, we obtain the claim, since r is integrable. By (6.1), we have $c'(t) \rightarrow \phi(q)(1 - q) - 1$. If this limit were nonzero, the bounded function c would grow or decrease linearly at infinity, which is impossible. We conclude that $\phi(q)(1 - q) = 1$. Since $\phi(q) = \ell - B + Bq^{p-1}$, this identity is equivalent to $\ell = L(q)$, which proves (6.2).

Step 3: the sign condition. The function ϕ is nondecreasing on $[0, 1]$, because $B > 0$. As c is nonincreasing, we have $c(t-u) \geq c(t)$ for $0 \leq u \leq t$. Using this and $r \geq 0$ in (6.1), we find

$$c'(t) \geq -1 + \phi(c(t)) \int_0^t r(u) \, du = -1 + \phi(c(t))(1 - c(t)) = F(c(t)).$$

By Step 1, c decreases continuously and strictly from $c(0) = 1$ to its limit q . Every $y \in (q, 1)$ is then equal to $c(t_y)$ for some $t_y \in (0, \infty)$, and

$$F(y) \leq c'(t_y) = -r(t_y) < 0.$$

Together with $F(q) = \phi(q)(1 - q) - 1 = 0$, which is (6.2), this proves (6.3).

Since the polynomial F vanishes at q and is negative on $(q, 1)$, we have $F'(q) \leq 0$. Using $\phi(q) = 1/(1 - q)$, we compute

$$F'(q) = B(p - 1)q^{p-2}(1 - q) - \phi(q) = \frac{\beta^2 p(p - 1)q^{p-2}(1 - q)^2 - 1}{1 - q}.$$

As $1 - q > 0$, the inequality $F'(q) \leq 0$ is (6.4). $\square$

Proposition 6.1 gives constraints that every stationary cluster satisfies; it does not determine the plateau. If $q = 0$, then (6.2) gives $\ell = L(0) = 1 + \beta^2 p$, and (6.4) carries no information. If $q > 0$, the multiplier is determined by q , and (6.4) is a one-sided version of the marginality relation $\beta^2 p(p - 1)q^{p-2}(1 - q)^2 = 1$ of Theorem 1.1. To select q , Section 8 combines these constraints with the upper bound on the energy of prefix centers from Proposition 8.1.

We now turn to uniqueness. We write $f * g$ for the convolution on $[0, \infty)$, given by $(f * g)(\tau) = \int_0^\tau f(\tau - u)g(u) du$. A solution of the system in the next lemma is understood in integrated form: a pair (c, r) of locally bounded Borel measurable functions on $[0, \infty)$ such that, for every $t \geq 0$,

$$c(t) = 1 - \int_0^t r(u) du, \quad r(t) = 1 + \int_0^t \left[-\ell r(v) + \beta^2 p(p - 1)((c^{p-2}r) * r)(v) \right] dv.$$

The integrated equations are pointwise identities, not merely identities almost everywhere. As the integrands on their right sides are locally integrable, both functions are continuous. Every pair of continuously differentiable functions satisfying the system, with right derivatives at zero, is a solution in this sense; by Proposition 5.7, this applies to every stationary cluster with multiplier ℓ .

Lemma 6.2 (Uniqueness with prescribed multiplier). *For a fixed finite ℓ , the system*

$$c' = -r, \quad r' = -\ell r + \beta^2 p(p - 1)(c^{p-2}r) * r, \quad c(0) = r(0) = 1,$$

has at most one solution in this locally bounded integrated sense on $[0, \infty)$.

Proof. Let (c_1, r_1) and (c_2, r_2) be two solutions, and fix $T > 0$. Let $M \geq 1$ be a bound for $|c_1|$, $|c_2|$, $|r_1|$, and $|r_2|$ on $[0, T]$, and define, for $0 \leq t \leq T$,

$$D(t) = \sup_{0 \leq v \leq t} (|c_1(v) - c_2(v)| + |r_1(v) - r_2(v)|).$$

We first bound the difference of the memory integrands. Let $0 \leq u \leq v \leq T$, and write $a = v - u$ and $b = u$. Telescoping the three factors, we have

$$\begin{aligned} & |c_1(a)^{p-2}r_1(a)r_1(b) - c_2(a)^{p-2}r_2(a)r_2(b)| \\ & \leq |c_1(a)^{p-2} - c_2(a)^{p-2}| |r_1(a)r_1(b)| + |c_2(a)^{p-2}| |r_1(a) - r_2(a)| |r_1(b)| \\ & \quad + |c_2(a)^{p-2}r_2(a)| |r_1(b) - r_2(b)| \\ & \leq pM^{p-1}\{D(a) + D(b)\}. \end{aligned}$$

In the last line, we used $|x^{p-2} - y^{p-2}| \leq (p-2)M^{p-3}|x-y|$ for $|x|, |y| \leq M$, and $|c_1 - c_2|, |r_1 - r_2| \leq D$. Since $D \geq 0$, we have $\int_0^v \{D(v-u) + D(u)\} du = 2 \int_0^v D(u) du \leq 2 \int_0^t D(u) du$ for $v \leq t$. Subtracting the integrated equations of the two solutions, we obtain, for $0 \leq t \leq T$,

$$|c_1(t) - c_2(t)| \leq \int_0^t D(u) du,$$

$$|r_1(t) - r_2(t)| \leq (|\ell| + 2\beta^2 p^2 (p-1)M^{p-1}T) \int_0^t D(u) du.$$

The right sides are nondecreasing in t . Taking the supremum over $[0, t]$, we find

$$D(t) \leq K_T \int_0^t D(u) du, \quad K_T = 1 + |\ell| + 2\beta^2 p^2 (p-1)M^{p-1}T,$$

where K_T depends on ℓ, p, β, T , and the bound M . The function D is bounded and nondecreasing on $[0, T]$, and Gronwall's inequality implies that $D = 0$ on $[0, T]$. As T was arbitrary, the two solutions coincide. $\square$

The causal tree realization and reflection

In this section, we realize the CK correlation as the Gram kernel of a curve v_t of unit vectors in a deterministic Hilbert space. We use this realization to study the part of a late state that was already determined at an earlier time. The main results are the exact causal realization, Proposition 7.1; the evolution of a projected state, Lemma 7.3; the reflection identity for stationary clusters, Proposition 7.5; and the spectral representation of every stationary correlation profile, Corollary 7.7.

We use the CK equations of Section 3, the physical bounds (3.15), the Green covariance identity of Proposition 4.2, the energy identity (4.1), the uniform multiplier bound (4.17), the diagonal increment bound (4.20), the regularity of the response from Propositions 4.5 and 4.6, and the stationary clusters of Proposition 5.7. The results of this section are used in most of the later sections. The limits d_L and h_L of Proposition 7.5 are the objects of Proposition 8.1, and they enter the proofs of marginality and forgetting in Section 8. The ceilings of Section 9 and the geometry of Sections 13 and 14 rest on the tree space and its height grading. In Section 11, we use the leaf component of a projected state and Lemma 7.3. The complete monotonicity asserted in Theorem 1.1 is part of Corollary 7.7, which is also used in Sections 11, 14, and 16.

We start from the Green identity (4.10), which we derived in Proposition 4.2 by variation of constants from the CK equations of [24, 60]. Positive noncrossing expansions of the associated response equation occur in the work of Guionnet and Mazza [89, Section 2.1 and Lemma 2.1] and of Dembo, Guionnet, and Mazza [60, Section 2]. Guionnet and Mazza also describe their full Fock-space interpretation. We build on these positive expansions. The tree realization below is a positive expansion of the covariance, graded by ordered trees so that every term records its height and the times of the thermal noise that it uses; we do not identify this grading with a Wiener chaos decomposition. All constructions take place in one deterministic Hilbert space. The prefix projection P_T defined below plays the role of conditioning on the noise up to time T , but we never condition on the finite-dimensional probability space.

Standing assumptions and notation. Throughout this section, $p \geq 3$ and $\beta > 0$ are fixed, $m = p - 1$, and (C, R, λ) is the uniformly initialized hard CK solution of Section 3. All Hilbert spaces and tensor products are real. We write

$$\Sigma(t, u) = \beta^2 p(p-1) C(t, u)^{p-2} R(t, u), \quad 0 \leq u \leq t,$$

for the kernel (4.8), so that the response equation (3.9) reads

$$\partial_t R(t, s) = -\lambda(t)R(t, s) + \int_s^t \Sigma(t, u)R(u, s) \, du, \quad t > s. \quad (7.1)$$

By (3.15) and Proposition 4.6, we have $0 \leq R(t, s) \leq C(t, s) \leq 1$ for $t \geq s$, and in particular $\Sigma \geq 0$.

7.1 The tree Hilbert space

We first explain the purpose of the construction. Recall from Proposition 4.2 that the correlation is the sum of three terms,

$$\begin{aligned} C(t, s) &= R(t, 0)R(s, 0) + 2 \int_0^{t \wedge s} R(t, u)R(s, u) \, du \\ &\quad + \beta^2 p \int_0^t \int_0^s R(t, u)R(s, w)C(u, w)^m \, dw \, du. \end{aligned}$$

The first term comes from the initial condition, the second from the thermal noise, and the third from the colored memory. The first two terms are the inner product of the vectors $R(t, 0)e_0 + \sqrt{2}\mathbf{1}_{[0, t]}R(t, \cdot)$ and $R(s, 0)e_0 + \sqrt{2}\mathbf{1}_{[0, s]}R(s, \cdot)$ in $\mathbb{R}e_0 \oplus L^2(\mathbb{R}_+)$. The second component of such a vector records the noise received at each time $u \leq t$, transported to time t by the response. If $C(u, w) = \langle v_u, v_w \rangle$ for some vectors v_u , then $C(u, w)^m = \langle v_u^{\otimes m}, v_w^{\otimes m} \rangle$, and the third term is also an inner product. This suggests looking for vectors v_t with

$$v_t = R(t, 0)e_0 + \sqrt{2}\mathbf{1}_{[0, t]}R(t, \cdot) + \beta\sqrt{p} \int_0^t R(t, u)\mathcal{I}(v_u^{\otimes m}) \, du,$$

where $\mathcal{I}$ is an isometry from the m -fold tensor power into a space orthogonal to the first two terms. Iterating this equation, we expand v_t as a sum of terms indexed by trees. Each leaf of a tree carries either the initial vector e_0 or the time of a noise increment, and each internal vertex records one use of the memory integral. The point of this bookkeeping is that every term of v_t comes with the times of the noise it uses. The terms whose noise times are at most T form the part of the state at time t built from the initial vector and noise received by time T ; their memory integrals may still extend beyond T . We now make this precise.

Trees and grafting. Let $\mathcal{E} = \mathbb{R}e_0 \oplus L^2(\mathbb{R}_+)$, where e_0 is a unit vector. By a *tree* we mean a finite rooted tree in which every vertex has either no children or exactly m children, and the children of each vertex are ordered. We call vertices without children leaves. We write $|\tau|$ for the number of leaves of a tree τ , whose leaves we list from left to right, and $\bullet$ for the tree with one vertex. For trees $\tau_1, \dots, \tau_m$, we write $[\tau_1, \dots, \tau_m]$ for the tree whose

root has the subtrees $\tau_1, \dots, \tau_m$, in this order. We call this operation grafting. Every tree other than $\bullet$ is the graft of a unique ordered m -tuple of trees.

We define $\mathcal{H}$ as the orthogonal direct sum, over all trees τ , of one copy of $\mathcal{E}^{\otimes|\tau|}$:

$$\mathcal{H} = \bigoplus_{\tau} \mathcal{E}^{\otimes|\tau|}.$$

The summand indexed by $\bullet$ is $\mathcal{E}$. We call it the one-leaf summand. The space $\mathcal{H}$ is separable, because there are countably many trees and each summand is separable. Since tensor products distribute over orthogonal direct sums, we have

$$\mathcal{H}^{\otimes m} \cong \bigoplus_{(\tau_1, \dots, \tau_m)} \mathcal{E}^{\otimes(|\tau_1| + \dots + |\tau_m|)}.$$

We define the grafting map $\mathcal{I}: \mathcal{H}^{\otimes m} \rightarrow \mathcal{H}$ by sending the summand indexed by $(\tau_1, \dots, \tau_m)$ identically onto the summand indexed by $[\tau_1, \dots, \tau_m]$; the leaves of τ_1 come first, then those of τ_2 , and so on. Since every tree other than $\bullet$ is the graft of a unique m -tuple, $\mathcal{I}$ is an isometry, and its range is the orthogonal complement of the one-leaf summand. Its adjoint satisfies $\mathcal{I}^* \mathcal{I} = \text{Id}$, and $\mathcal{I}^*$ vanishes on the one-leaf summand. We use ordered trees because they make grafting an isometry, with no symmetry factors. When $p = 3$, we identify a tensor $x \otimes y \in \mathcal{H} \otimes \mathcal{H}$ with the Hilbert–Schmidt operator $z \mapsto \langle y, z \rangle x$ on $\mathcal{H}$; this is how we use $\mathcal{I}^*$ in Sections 13 and 14.

Leaf labels and positivity. We now realize $\mathcal{E}$ as an L^2 space. Let $\mathbb{L} = \{\star\} \sqcup [0, \infty)$ carry the measure $\delta_{\star} + du$. Then $\mathcal{E} = L^2(\mathbb{L})$, with $e_0 = \mathbf{1}_{\{\star\}}$, and $\mathcal{E}^{\otimes k} = L^2(\mathbb{L}^k)$. An element of the summand indexed by τ is then a function of the leaf labels of τ . The label $\star$ marks the initial vector, and a label $u \geq 0$ marks a noise time. We say that a vector of $\mathcal{H}$ is nonnegative if each of its components is a nonnegative function almost everywhere. Tensor products and grafts of nonnegative vectors are nonnegative.

Heights. We define the height of $\bullet$ to be zero, and the height of $[\tau_1, \dots, \tau_m]$ to be one plus the largest height of $\tau_1, \dots, \tau_m$. We write Π_j for the orthogonal projection of $\mathcal{H}$ onto the sum of the summands indexed by trees of height j , and we set $\Pi_{\leq H} = \sum_{j=0}^H \Pi_j$. Note that $\sum_{j \geq 0} \Pi_j = \text{Id}$ in the strong sense, because every tree has a finite height.

Prefix projections. For $T \geq 0$, let $\mathcal{E}_T = \mathbb{R}e_0 \oplus L^2([0, T])$, and let ω_T be the orthogonal projection of $\mathcal{E}$ onto $\mathcal{E}_T$, that is, $\omega_T(ae_0 + g) = ae_0 + \mathbf{1}_{[0, T]}g$. We define the prefix projection

$$P_T = \bigoplus_{\tau} \omega_T^{\otimes|\tau|},$$

and we write $\mathcal{H}_T$ for its range, the closed subspace generated by trees whose leaf factors lie in $\mathcal{E}_T$. In terms of leaf labels, P_T multiplies the component indexed by τ by the

indicator function that all leaf labels lie in $\{\star\} \cup [0, T]$. Since P_T , Π_j , and $\mathcal{I}$ act summand by summand, and the leaves of a graft are the leaves of its root subtrees, we have

$$P_T P_S = P_{T \wedge S}, \quad P_T \Pi_j = \Pi_j P_T, \quad P_T \mathcal{I} = \mathcal{I} P_T^{\otimes m}. \quad (7.2)$$

In words, P_T is a causal projection. A vector built from leaves with noise times at most T belongs to $\mathcal{H}_T$, and P_T commutes with grafting. The projections P_T , $\text{Id} - P_T$, and Π_j preserve nonnegativity, since each of them multiplies every component by an indicator function.

7.2 The exact causal realization

Proposition 7.1 (Exact causal realization). *There exists a norm-continuous curve $t \mapsto v_t \in \mathcal{H}$, $t \geq 0$, with $v_t \in \mathcal{H}_t$ for every t , such that*

$$v_t = R(t, 0)e_0 + \sqrt{2} \mathbf{1}_{[0, t]} R(t, \cdot) + \beta \sqrt{p} \int_0^t R(t, u) \mathcal{I}(v_u^{\otimes m}) \, du, \quad (7.3)$$

$$\langle v_t, v_s \rangle = C(t, s), \quad \|v_t\| = 1, \quad (7.4)$$

where the integral is a Bochner integral in $\mathcal{H}$. For every tree τ , the component of v_t indexed by τ is a nonnegative function of the leaf labels; in particular, $\langle \Pi_j v_t, \Pi_j v_s \rangle \geq 0$ for all $j \geq 0$ and $t, s \geq 0$. Let h be the tree energy, defined by

$$h(v) = \sqrt{p} \langle v, \mathcal{I}(v^{\otimes m}) \rangle, \quad v \in \mathcal{H}. \quad (7.5)$$

Then $h(v_t) = e_\beta(t)$ and $\lambda(t) = 1 + p\beta h(v_t)$ for every $t \geq 0$. Moreover, for all $T \geq 0$ and $t \geq 0$,

$$\mathcal{I}^* P_T v_t = \beta \sqrt{p} \int_0^t R(t, u) (P_T v_u)^{\otimes m} \, du. \quad (7.6)$$

Before the proof, we write out the first terms of the expansion in the cubic case.

Example 7.2 (The first two heights for $p = 3$). Let $p = 3$, so that grafting is binary. We write η_t for the component of v_t at height zero, which is the leaf vector

$$\eta_t = R(t, 0)e_0 + \sqrt{2} \mathbf{1}_{[0, t]} R(t, \cdot).$$

The only tree of height one is $[\bullet, \bullet]$, with two leaves. Iterating (7.3) once, we see that the component of v_t at height one is

$$w_t = \beta \sqrt{3} \int_0^t R(t, u) \mathcal{I}(\eta_u \otimes \eta_u) \, du.$$

As a function of two noise times $y_1, y_2 \in [0, t]$, it is

$$w_t(y_1, y_2) = 2\sqrt{3} \beta \int_{y_1 \vee y_2}^t R(t, u) R(u, y_1) R(u, y_2) \, du.$$

Noise received at the times y_1 and y_2 is transported to a time u by the response, combined there by the cubic interaction, and transported to time t . At labels involving $\star$, a factor $\sqrt{2}R(u, y_i)$ is replaced by $R(u, 0)$. The trees of height two are $[[\bullet, \bullet], \bullet]$, $[\bullet, [\bullet, \bullet]]$, and $[[\bullet, \bullet], [\bullet, \bullet]]$, and the corresponding components of v_t are

$$\begin{aligned} \beta\sqrt{3} \int_0^t R(t, u) \mathcal{I}(w_u \otimes \eta_u) \, du, \quad & \beta\sqrt{3} \int_0^t R(t, u) \mathcal{I}(\eta_u \otimes w_u) \, du, \\ & \beta\sqrt{3} \int_0^t R(t, u) \mathcal{I}(w_u \otimes w_u) \, du. \end{aligned}$$

The prefix projection acts only on the labels. For instance, we see that $P_T w_t$ keeps the values of w_t at noise times $y_1, y_2 \leq T$, while the time integral defining it still runs up to t . Finally, the Gram kernel of $\eta_t + w_t$ is

$$A_0(t, s) + 3\beta^2 \int_0^t \int_0^s R(t, u) R(s, w) A_0(u, w)^2 \, dw \, du, \quad A_0(t, s) = \langle \eta_t, \eta_s \rangle,$$

which is the first step of the Picard iteration of the Green identity (4.10). In the proof below, we show that these iterations increase to C .

Proof of Proposition 7.1. The proof has five steps. We first construct the component of v_t indexed by each tree through a finite recursion. We then show that the Gram kernels of the truncations at bounded height increase to C , using a uniqueness property of the Green equation. This allows us to sum the components. Finally, we compute the projected root frame and the energy.

Step 1: the tree components. Let $\eta_t = R(t, 0)e_0 + \sqrt{2}\mathbf{1}_{[0, t]}R(t, \cdot) \in \mathcal{E}$, and set $w_\bullet(t) = \eta_t$. If $\tau = [\tau_1, \dots, \tau_m]$, we define, in the summand indexed by τ ,

$$w_\tau(t) = \beta\sqrt{p} \int_0^t R(t, u) w_{\tau_1}(u) \otimes \cdots \otimes w_{\tau_m}(u) \, du. \quad (7.7)$$

This is a recursion on the number of vertices. The curve $t \mapsto \eta_t$ is norm-continuous, because R is uniformly continuous on compact causal triangles and bounded by one. By induction on the number of vertices, each w_τ is a norm-continuous curve, and the integral in (7.7) is a Bochner integral of a continuous integrand. Also by induction, $w_\tau(t)$ is nonnegative, because $R \geq 0$ and tensor products and integrals of nonnegative vectors are nonnegative, and $w_\tau(t)$ lies in $\mathcal{E}_t^{\otimes |\tau|}$, because $w_{\tau_i}(u) \in \mathcal{E}_u^{\otimes |\tau_i|} \subset \mathcal{E}_t^{\otimes |\tau_i|}$ for $u \leq t$.

Step 2: the truncated Gram kernels. For $n \geq 0$, let $v_t^{(n)}$ be the sum of $w_\tau(t)$ over the finitely many trees τ of height at most n , and let $K_n(t, s) = \langle v_t^{(n)}, v_s^{(n)} \rangle$. A tree of height at most $n \geq 1$ is either $\bullet$ or the graft of m trees of height at most $n - 1$. Summing (7.7) over these trees, we have

$$v_t^{(n)} = \eta_t + \beta\sqrt{p} \int_0^t R(t, u) \mathcal{I}((v_u^{(n-1)})^{\otimes m}) \, du.$$

Recall that η_t is orthogonal to the range of $\mathcal{I}$ and that $\mathcal{I}$ is an isometry. Using also $\langle x^{\otimes m}, y^{\otimes m} \rangle = \langle x, y \rangle^m$, we find that $K_0 = A_0$ and

$$K_n(t, s) = A_0(t, s) + \beta^2 p \int_0^t \int_0^s R(t, u) R(s, w) K_{n-1}(u, w)^m dw du, \quad n \geq 1, \quad (7.8)$$

where

$$A_0(t, s) = \langle \eta_t, \eta_s \rangle = R(t, 0) R(s, 0) + 2 \int_0^{t \wedge s} R(t, u) R(s, u) du.$$

The squared coefficients are $(\sqrt{2})^2 = 2$ and $(\beta\sqrt{p})^2 = \beta^2 p$, which are the coefficients of the Green identity (4.10). Note that $K_n \geq 0$, because the components are nonnegative. We claim that

$$0 \leq K_{n-1} \leq K_n \leq C, \quad n \geq 1. \quad (7.9)$$

Indeed, the Green identity reads $C = A_0 + \beta^2 p \iint RR C^m$, and its double integral is nonnegative by (3.15). This implies that $K_0 = A_0 \leq C$. If $0 \leq K_{n-1} \leq C$, then (7.8) and $R \geq 0$ imply that $K_n \leq C$. In the same way, $K_1 \geq K_0$, and if $K_{n-1} \geq K_{n-2} \geq 0$, then $K_n \geq K_{n-1}$. This proves (7.9). Then K_n increases pointwise to a limit K with $0 \leq K \leq C \leq 1$, and by monotone convergence in (7.8), K solves the Green equation

$$K(t, s) = A_0(t, s) + \beta^2 p \int_0^t \int_0^s R(t, u) R(s, w) K(u, w)^m dw du. \quad (7.10)$$

Step 3: uniqueness for the Green equation. We prove that, for every $A < \infty$, (7.10) has at most one solution on $[0, A]^2$ with values in $[-1, 1]$. The kernel C is such a solution by Proposition 4.2, so uniqueness implies that $K = C$, that is, K_n increases to C . Let K and K' be two such solutions and let $\Delta = K - K'$. Using $|x^m - y^m| \leq m|x - y|$ for $x, y \in [-1, 1]$ and $0 \leq R \leq 1$, we have, for $t, s \leq A$,

$$|\Delta(t, s)| \leq \beta^2 p m \int_0^t \int_0^s |\Delta(u, w)| dw du.$$

Let $\rho(a) = \sup_{t, s \leq a} |\Delta(t, s)|$, a bounded nondecreasing function. We bound $|\Delta(u, w)|$ by $\rho(\max\{u, w\})$ and divide the square $[0, a]^2$ according to which coordinate is larger. For $a \leq A$, this yields

$$\rho(a) \leq \beta^2 p m \int_0^a \int_0^a \rho(\max\{u, w\}) dw du = 2\beta^2 p m \int_0^a z \rho(z) dz \leq 2\beta^2 p m A \int_0^a \rho(z) dz.$$

By Gronwall's inequality, we have $\rho = 0$ on $[0, A]$. As A is arbitrary, we conclude that $K = C$.

Step 4: the curve v_t . Using Step 3, we find that $\sum_\tau \|w_\tau(t)\|^2 = \lim_n K_n(t, t) = C(t, t) = 1$. The components lie in orthogonal summands, and the series $v_t = \sum_\tau w_\tau(t)$ therefore converges in $\mathcal{H}$, with $v_t^{(n)} \rightarrow v_t$. Then $\langle v_t, v_s \rangle = \lim_n K_n(t, s) = C(t, s)$, which is (7.4). By the identity $\|v_t - v_s\|^2 = 2\{1 - C(t, s)\}$ and the continuity of C , the curve $t \mapsto v_t$ is norm-continuous. Each component of v_t is nonnegative and lies in $\mathcal{E}_t^{\otimes |\tau|}$, which shows that $v_t \in \mathcal{H}_t$. The inner product $\langle \Pi_j v_t, \Pi_j v_s \rangle$ is a sum of inner products of nonnegative functions, and it is nonnegative.

It remains to prove (7.3). The integrand $u \mapsto R(t, u) \mathcal{I}(v_u^{\otimes m})$ is continuous on $[0, t]$, and the integral is a Bochner integral. The orthogonal projection onto the summand indexed by $\tau = [\tau_1, \dots, \tau_m]$ is bounded and commutes with the integral, and it maps $\mathcal{I}(v_u^{\otimes m})$ to $w_{\tau_1}(u) \otimes \dots \otimes w_{\tau_m}(u)$. Comparing with (7.7), we see that the component of the right side of (7.3) indexed by τ is $w_\tau(t)$. Its one-leaf component is $\eta_t = w_\bullet(t)$. Both sides of (7.3) therefore have the same components.

Step 5: the root frame and the energy. We apply P_T to (7.3). The projection P_T is bounded and commutes with grafting by (7.2), which gives

$$P_T v_t = P_T \eta_t + \beta \sqrt{p} \int_0^t R(t, u) \mathcal{I}((P_T v_u)^{\otimes m}) \, du.$$

The vector $P_T \eta_t$ lies in the one-leaf summand, which $\mathcal{I}^*$ annihilates, and $\mathcal{I}^* \mathcal{I} = \text{Id}$. Applying $\mathcal{I}^*$ to the last display yields (7.6). We now let $T = t$. For $u \leq t$, we have $v_u \in \mathcal{H}_u \subset \mathcal{H}_t$, so $P_t v_u = v_u$. Taking the inner product of (7.6) with $v_t^{\otimes m}$, we obtain

$$\begin{aligned} h(v_t) &= \sqrt{p} \langle \mathcal{I}^* v_t, v_t^{\otimes m} \rangle = p\beta \int_0^t R(t, u) \langle v_u, v_t \rangle^m \, du \\ &= p\beta \int_0^t C(t, u)^{p-1} R(t, u) \, du = e_\beta(t), \end{aligned}$$

where the last equality is (4.1). By the same identity, we also have $\lambda(t) = 1 + p\beta e_\beta(t) = 1 + p\beta h(v_t)$. This completes the proof. $\square$

7.3 The evolution of a projected state

After time T , no new noise leaf survives the projection P_T , and the projected curve $P_T v_t$ becomes differentiable. The next lemma computes its derivative. In Section 11, we use it to turn reflection identities into bounds on derivatives at remote times (Theorem 11.4). The polynomial h of (7.5) is Fréchet differentiable on $\mathcal{H}$, and we write $\nabla h(v)$ for its gradient.

Lemma 7.3 (Local polynomial drift after a causal projection). *Define $\mathcal{B}: \mathcal{H} \rightarrow \mathcal{H}$ by*

$$\mathcal{B}(v) = -v + \beta \nabla h(v) - p\beta h(v)v.$$

For every fixed $T \geq 0$, the curve $t \mapsto P_T v_t$ is continuously differentiable on (T, ∞) , and

$$\frac{d}{dt} P_T v_t = P_T \mathcal{B}(v_t), \quad t > T. \quad (7.11)$$

The map $\mathcal{B}$ is Lipschitz on the closed unit ball of $\mathcal{H}$. If there exists $K < \infty$ such that $1 - C(t, s) \leq K|t - s|$ for all $t, s \geq 0$, then there exists $K' < \infty$, depending only on p, β , and K , such that

$$\left\| \frac{d}{dt} P_T v_t - \frac{d}{ds} P_T v_s \right\| \leq K' |t - s|^{1/2} \quad \text{for all } T \geq 0 \text{ and } t, s > T.$$

Proof. We differentiate the projected realization, identify the result with the gradient of h along the curve, and prove the Lipschitz bounds. Throughout, we write

$$\Phi(u) = \beta \sqrt{p} \mathcal{I}(v_u^{\otimes m}), \quad \Psi(t) = \int_0^t R(t, u) \Phi(u) \, du,$$

so that $v_t = \eta_t + \Psi(t)$ by (7.3), with η_t the leaf vector of the proof of Proposition 7.1. The curve Φ is continuous, with $\|\Phi(u)\| \leq \beta \sqrt{p}$.

Step 1: differentiation. By Proposition 4.5, the partial derivative $\partial_t R$ extends continuously to every compact causal triangle. For $t > T$, we have

$$P_T \eta_t = R(t, 0) e_0 + \sqrt{2} \mathbf{1}_{[0, T]} R(t, \cdot),$$

and the cutoff of the noise component no longer moves with t . On every compact interval of (T, ∞) , the difference quotients of $R(\cdot, u)$ converge to $\partial_t R(\cdot, u)$ uniformly in $u \in [0, T]$. It follows that $P_T \eta_t$ is continuously differentiable on (T, ∞) , with derivative $\partial_t R(t, 0) e_0 + \sqrt{2} \mathbf{1}_{[0, T]} \partial_t R(t, \cdot)$. Inserting the response equation (7.1), and noting that the noise component of $P_T \eta_z$ is $\sqrt{2} R(z, y)$ for $y \leq z \wedge T$ and zero for other y , we have

$$\frac{d}{dt} P_T \eta_t = -\lambda(t) P_T \eta_t + \int_0^t \Sigma(t, z) P_T \eta_z \, dz, \quad t > T. \quad (7.12)$$

Next, using that R and $\partial_t R$ are continuous on compact causal triangles and that $R(t, t) = 1$, we see that the curve Ψ is continuously differentiable on $(0, \infty)$, with

$$\Psi'(t) = \Phi(t) + \int_0^t \partial_t R(t, u) \Phi(u) \, du.$$

We insert (7.1) and exchange the order of integration on the triangle $\{0 \leq u \leq z \leq t\}$, which is allowed because the integrand is continuous and bounded in norm. We obtain

$$\Psi'(t) = \Phi(t) - \lambda(t) \Psi(t) + \int_0^t \Sigma(t, z) \Psi(z) \, dz. \quad (7.13)$$

Adding (7.12) and the image of (7.13) under P_T , we have, for $t > T$,

$$\frac{d}{dt} P_T v_t = P_T \left[-\lambda(t) v_t + \Phi(t) + \int_0^t \Sigma(t, z) v_z \, dz \right]. \quad (7.14)$$

Step 2: the gradient along the curve. For $v, w \in \mathcal{H}$, the derivative of h is

$$Dh(v)[w] = \sqrt{p} \langle w, \mathcal{I}(v^{\otimes m}) \rangle + \sqrt{p} \sum_{i=1}^m \left\langle \mathcal{I}^* v, v^{\otimes(i-1)} \otimes w \otimes v^{\otimes(m-i)} \right\rangle.$$

Recall from (7.6) with $T = t$ that $\mathcal{I}^* v_t = \beta \sqrt{p} \int_0^t R(t, u) v_u^{\otimes m} du$. Each of the m terms of the sum then equals $\beta \sqrt{p} \int_0^t R(t, u) C(t, u)^{m-1} \langle v_u, w \rangle du$, and we find that

$$\beta \nabla h(v_t) = \Phi(t) + \beta^2 p(p-1) \int_0^t C(t, u)^{p-2} R(t, u) v_u du = \Phi(t) + \int_0^t \Sigma(t, u) v_u du. \quad (7.15)$$

Although ordered grafting is not symmetric, the m contractions agree here because the root tensor of v_t is an integral of symmetric tensor powers; this produces the factor $p-1$ in Σ . We also have $\lambda(t) = 1 + p\beta h(v_t)$ by Proposition 7.1. Inserting this and (7.15) into (7.14), we arrive at (7.11). Its right side is continuous in t , because $t \mapsto v_t$ and $\mathcal{B}$ are continuous, and this shows that $P_T v_t$ is continuously differentiable on (T, ∞) .

Step 3: the Lipschitz bounds. The multilinear form $(y_0, \dots, y_m) \mapsto \langle y_0, \mathcal{I}(y_1 \otimes \dots \otimes y_m) \rangle$ has norm at most one, and we obtain h and Dh by inserting v or w into its $p = m+1$ slots. Changing one slot at a time, we find, for v, v' in the closed unit ball,

$$\begin{aligned} |h(v)| &\leq \sqrt{p}, & |h(v) - h(v')| &\leq p\sqrt{p} \|v - v'\|, \\ \|\nabla h(v) - \nabla h(v')\| &\leq p(p-1)\sqrt{p} \|v - v'\|. \end{aligned}$$

Writing $h(v)v - h(v')v' = h(v)(v - v') + \{h(v) - h(v')\}v'$, we have

$$\begin{aligned} \|\mathcal{B}(v) - \mathcal{B}(v')\| &\leq \{1 + \beta p(p-1)\sqrt{p} + p\beta(\sqrt{p} + p\sqrt{p})\} \|v - v'\| \\ &= (1 + 2\beta p^2 \sqrt{p}) \|v - v'\|. \end{aligned}$$

Now suppose that $1 - C(t, s) \leq K|t - s|$ for all t, s . Then $\|v_t - v_s\|^2 = 2\{1 - C(t, s)\} \leq 2K|t - s|$. Since P_T is a contraction, (7.11) implies that, for $t, s > T$,

$$\left\| \frac{d}{dt} P_T v_t - \frac{d}{ds} P_T v_s \right\| \leq (1 + 2\beta p^2 \sqrt{p}) \sqrt{2K} |t - s|^{1/2}.$$

This is the last assertion, with $K' = (1 + 2\beta p^2 \sqrt{p}) \sqrt{2K}$. □

Remark 7.4 (The diagonal increment bound). The hypothesis of the last assertion of Lemma 7.3 holds with $K = \Lambda$, the constant of (4.17). Indeed, by (4.20) in Proposition 4.4 and the symmetry of C , we have $1 - C(t, s) \leq \Lambda|t - s|$ for all $t, s \geq 0$. The derivatives of $P_T v_t$ are then $1/2$ -Hölder on (T, ∞) with the constant $(1 + 2\beta p^2 \sqrt{p}) \sqrt{2\Lambda}$, which does not depend on T .

7.4 Reflection and complete monotonicity

We now study the part of a late state that was already determined at an earlier time. By Proposition 5.7, entropy production identifies every bounded-lag cluster of the translated kernels as a stationary profile that satisfies the FDT relation. It neither selects the plateau of this profile nor controls the older part of the correlation memory. To study the plateau, we project late states onto the information carried by earlier noise leaves. For $T \geq 0$ and $a \geq 0$, we define the *prefix center*

$$M_T(a) = P_T v_{T+a}.$$

It is the part of the state at time $T + a$ built from the initial vector and noise received by time T , including memory interactions after T . The main observation of this section is that the Gram kernel of the prefix centers satisfies a closed equation on a fixed future rectangle. Every memory term that involves a time before T cancels exactly, before any limit is taken. The stationary FDT relation then identifies the limit of this kernel. This is the reflection identity, which says that the overlap of the centers at offsets a and b converges to $c(a + b)$.

Recall that a stationary cluster (c, r, ℓ) is a locally uniform limit of the translated kernels $(C_{T_n}, R_{T_n}, \lambda_{T_n})$ along a sequence $T_n \rightarrow \infty$, as in Proposition 5.7, and that its energy is $e_\infty = (\ell - 1)/(p\beta)$.

Proposition 7.5 (Reflection of a stationary cluster). *Fix $p \geq 3$ and $\beta > 0$. Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to a stationary cluster (c, r, ℓ) , as in Proposition 5.7, and let $e_\infty = (\ell - 1)/(p\beta)$ be its energy. Then, for every finite $A \geq 0$,*

$$\lim_{n \rightarrow \infty} \sup_{0 \leq a, b \leq A} |\langle M_{T_n}(a), M_{T_n}(b) \rangle - c(a + b)| = 0. \quad (7.16)$$

Moreover, for every fixed $L \geq 0$,

$$\lim_{n \rightarrow \infty} \|M_{T_n}(L)\|^2 = d_L, \quad d_L = c(2L), \quad (7.17)$$

$$\lim_{n \rightarrow \infty} h(M_{T_n}(L)) = h_L, \quad h_L = e_\infty + \beta p \int_0^L r(\tau) \{c(2L - \tau)^{p-1} - c(\tau)^{p-1}\} d\tau. \quad (7.18)$$

The plateau q of the cluster may be zero; the proposition covers every $p \geq 3$, every $\beta > 0$, and every stationary cluster.

Remark 7.6 (Independent futures). The reflected lag $a + b$ has a simple heuristic explanation. Let (x_t) be a stationary reversible Markov process in $\mathbb{R}^N$ with correlation $c(t) = \mathbb{E}[x_0 \cdot x_t]/N$. Two continuations from a common state x_T , with independent noise after time T , have the conditional means $\mathbb{E}[x_{T+a} \mid x_T]$ and $\mathbb{E}[x_{T+b} \mid x_T]$ at times $T + a$

and $T + b$. The Markov semigroup of the process is self-adjoint in the stationary law, so stationarity and the semigroup property give

$$\frac{1}{N} \mathbb{E} [\mathbb{E}[x_{T+a} \mid x_T] \cdot \mathbb{E}[x_{T+b} \mid x_T]] = \frac{1}{N} \mathbb{E}[x_0 \cdot x_{a+b}] = c(a+b).$$

Trajectories with a common past and independent future noise were studied by Barrat, Burioni, and Mézard, and the formula $c(a+b)$ for the overlap of two such continuations appears in [11, equation (9) and Section 4.1]. Franz and Virasoro proposed weighted averages over cloned futures to improve estimates of quasi-state magnetizations [72, arXiv version 3, Section III]; see also Biroli [31, arXiv version 2, Section III.B]. The prefix center plays the role of the conditional mean in the tree space. The CK dynamics is not stationary, and Proposition 7.5 justifies the reflected overlap for the stationary clusters of the CK solution, by an exact cancellation in the CK equations.

Proof of Proposition 7.5. The proof has four steps. First, for every fixed T , we derive a closed equation for the projected Gram kernel on the future rectangle. Second, we pass to the limit in its coefficients and show that the limiting equation is stable, in the sense that any solution with values in $[0, 1]$ is the limit of the projected kernels. Third, we verify with Laplace transforms that $c(a+b)$ is such a solution, which proves (7.16) and (7.17). Fourth, we compute the energy of a prefix center by an exact subtraction. Throughout, we write $B = \beta^2 p$.

Step 1: the closed equation on the future rectangle. Fix $T \geq 0$, and let $K_T(t, s) = \langle P_T v_t, P_T v_s \rangle$ for $t, s \geq 0$. We apply P_T to (7.3). The one-leaf component $P_T \eta_t$ is orthogonal to the grafted part, P_T commutes with grafting by (7.2), and $\mathcal{I}$ is an isometry. As in Step 2 of the proof of Proposition 7.1, we obtain, for $t, s \geq T$,

$$\begin{aligned} K_T(t, s) &= R(t, 0)R(s, 0) + 2 \int_0^T R(t, u)R(s, u) \, du \\ &\quad + B \int_0^t \int_0^s R(t, u)R(s, w)K_T(u, w)^m \, dw \, du. \end{aligned} \tag{7.19}$$

Here the noise integral stops at T , since the noise components of $P_T \eta_t$ and $P_T \eta_s$ vanish beyond T . The projection has no effect as soon as one argument lies in the past. Indeed, if $u \leq T$, then $v_u \in \mathcal{H}_T$, which implies that $P_T v_u = v_u$ and

$$K_T(u, w) = \langle v_u, P_T v_w \rangle = \langle P_T v_u, v_w \rangle = \langle v_u, v_w \rangle = C(u, w).$$

The same holds if $w \leq T$. We subtract (7.19) from the Green identity (4.10). The initial terms cancel, the noise integrals over $[0, T]$ cancel, and in the memory integrals every

block with $u \leq T$ or $w \leq T$ cancels. Only the integrals over $(T, t] \times (T, s]$ remain, and we have

$$\begin{aligned} C(t, s) - K_T(t, s) &= 2 \int_T^{t \wedge s} R(t, u) R(s, u) \, du \\ &\quad + B \int_T^t \int_T^s R(t, u) R(s, w) \{C(u, w)^m - K_T(u, w)^m\} \, dw \, du. \end{aligned}$$

For $a, b \geq 0$, let $Q_T(a, b) = K_T(T + a, T + b) = \langle M_T(a), M_T(b) \rangle$. Shifting all times by T , we arrive at

$$\begin{aligned} Q_T(a, b) &= C(T + a, T + b) - 2 \int_0^{a \wedge b} R(T + a, T + u) R(T + b, T + u) \, du \\ &\quad - B \int_0^a \int_0^b R(T + a, T + u) R(T + b, T + w) \\ &\quad \cdot \{C(T + u, T + w)^m - Q_T(u, w)^m\} \, dw \, du. \end{aligned} \quad (7.20)$$

This equation holds for every T ; no limit has been taken, and no decay of the older memory has been used. We also note that

$$0 \leq Q_T(a, b) \leq C(T + a, T + b) \leq 1.$$

Indeed, the components of $P_T v_t$ and $(\text{Id} - P_T)v_t$ are nonnegative, which implies $C(t, s) = K_T(t, s) + \langle (\text{Id} - P_T)v_t, (\text{Id} - P_T)v_s \rangle \geq K_T(t, s) \geq 0$.

Step 2: the limiting equation, uniqueness, and stability. We fix $0 \leq A < \infty$ and write $Q_n = Q_{T_n}$, $C_n(a, b) = C(T_n + a, T_n + b)$, and $R_n(a, u) = R(T_n + a, T_n + u)$. We also abbreviate $\Theta_n(u, w) = R_n(a, u)R_n(b, w)$ and $\theta(u, w) = r(a - u)r(b - w)$, for fixed a and b . By Proposition 5.7, we have $C_n(a, b) \rightarrow c(|a - b|)$ uniformly for $a, b \in [0, A]$, and $R_n(a, u) \rightarrow r(a - u)$ uniformly for $0 \leq u \leq a \leq A$. The limiting form of (7.20) is the equation

$$\begin{aligned} Q(a, b) &= c(|a - b|) - 2 \int_0^{a \wedge b} r(a - u)r(b - u) \, du \\ &\quad - B \int_0^a \int_0^b r(a - u)r(b - w) \{c(|u - w|)^m - Q(u, w)^m\} \, dw \, du \end{aligned} \quad (7.21)$$

for a function Q on $[0, A]^2$. We first show that it has at most one bounded solution. Let Q and Q' be two solutions with $|Q|, |Q'| \leq \Gamma$, and let $\Delta = Q - Q'$. Using $0 \leq r \leq 1$ and $|x^m - y^m| \leq m\Gamma^{m-1}|x - y|$ for $|x|, |y| \leq \Gamma$, we have

$$|\Delta(a, b)| \leq Bm\Gamma^{m-1} \int_0^a \int_0^b |\Delta(u, w)| \, dw \, du,$$

and the Gronwall argument of Step 3 in the proof of Proposition 7.1 shows that $\Delta = 0$.

Next, we prove stability. Let Q be a solution of (7.21) with values in $[0, 1]$. Subtracting (7.21) from (7.20) at $T = T_n$, we find

$$\begin{aligned} Q_n(a, b) - Q(a, b) &= \{C_n(a, b) - c(|a - b|)\} - 2 \int_0^{a \wedge b} \{\Theta_n(u, u) - \theta(u, u)\} du \\ &\quad - B \int_0^a \int_0^b \{\Theta_n(u, w)C_n(u, w)^m - \theta(u, w)c(|u - w|)^m\} dw du \\ &\quad + B \int_0^a \int_0^b \{\Theta_n(u, w) - \theta(u, w)\} Q(u, w)^m dw du \\ &\quad + B \int_0^a \int_0^b \Theta_n(u, w) \{Q_n(u, w)^m - Q(u, w)^m\} dw du. \end{aligned}$$

Let ε_n be the supremum over $a, b \in [0, A]$ of the sum of the absolute values of the first four terms on the right. All kernels take values in $[0, 1]$ and converge uniformly, and the intervals and rectangles of integration have measure bounded in terms of A only. Then $\varepsilon_n \rightarrow 0$. In the last term, we use $0 \leq \Theta_n \leq 1$ and $|x^m - y^m| \leq m|x - y|$ for $x, y \in [0, 1]$, which applies by Step 1. This gives

$$|Q_n(a, b) - Q(a, b)| \leq \varepsilon_n + Bm \int_0^a \int_0^b |Q_n(u, w) - Q(u, w)| dw du, \quad a, b \in [0, A].$$

Let $\rho_n(z) = \sup_{a, b \leq z} |Q_n(a, b) - Q(a, b)|$. As in Step 3 of the proof of Proposition 7.1, we have $\rho_n(z) \leq \varepsilon_n + 2BmA \int_0^z \rho_n(y) dy$ for $z \leq A$, and Gronwall's inequality implies that

$$\sup_{0 \leq a, b \leq A} |Q_n(a, b) - Q(a, b)| \leq \varepsilon_n e^{2BmA^2} \rightarrow 0. \quad (7.22)$$

It remains to find a solution of (7.21) with values in $[0, 1]$.

Step 3: identification of the solution through Laplace transforms. We check that $Q(a, b) = c(a + b)$ solves (7.21). Note that $0 \leq c \leq 1$. Step 2 then proves (7.16), and the case $a = b = L$ proves (7.17).

For a bounded function f on $[0, \infty)$, let $D_f(a, b) = f(|a - b|) - f(a + b)$. Note that $c(|u - w|)^m - c(u + w)^m = D_{c^m}(u, w)$. Then the function $c(a + b)$ solves (7.21) if and only if, for all $a, b \geq 0$,

$$D_c(a, b) = 2 \int_0^{a \wedge b} r(a - u)r(b - u) du + B \int_0^a \int_0^b r(a - u)r(b - w)D_{c^m}(u, w) dw du. \quad (7.23)$$

We prove (7.23) by comparing double Laplace transforms. For a bounded measurable f on $[0, \infty)$ and $z > 0$, we write $\hat{f}(z) = \int_0^\infty e^{-zt} f(t) dt$, and for a bounded measurable F on $[0, \infty)^2$ and $z_1, z_2 > 0$, we write $\hat{F}(z_1, z_2) = \int_0^\infty \int_0^\infty e^{-z_1 a - z_2 b} F(a, b) db da$.

We begin with the transform of D_f . Splitting $[0, \infty)^2$ into the regions $a \geq b$ and $a < b$, and substituting $a = b + x$ in the first and $b = a + x$ in the second, we find

$$\int_0^\infty \int_0^\infty e^{-z_1 a - z_2 b} f(|a - b|) db da = \frac{\hat{f}(z_1) + \hat{f}(z_2)}{z_1 + z_2}.$$

Substituting $y = a + b$ and integrating over $a \in [0, y]$, we have, for $z_1 \neq z_2$,

$$\int_0^\infty \int_0^\infty e^{-z_1 a - z_2 b} f(a + b) \, db \, da = \int_0^\infty f(y) \frac{e^{-z_2 y} - e^{-z_1 y}}{z_1 - z_2} \, dy = \frac{\widehat{f}(z_2) - \widehat{f}(z_1)}{z_1 - z_2}.$$

Subtracting the two identities and putting them over the common denominator $z_1^2 - z_2^2$, we find that

$$\widehat{D}_f(z_1, z_2) = \frac{2\{z_1 \widehat{f}(z_1) - z_2 \widehat{f}(z_2)\}}{z_1^2 - z_2^2}, \quad z_1 \neq z_2. \quad (7.24)$$

Next, we compute the transforms of the stationary profiles. Let $\sigma(u) = Bm c(u)^{m-1} r(u) = -B(c^m)'(u)$. The stationary response equation (5.25) reads $r' = -\ell r + \sigma * r$, where $*$ denotes convolution on $[0, \infty)$. The functions r and σ are nonnegative and integrable, with $\int_0^\infty r = 1 - q$ and $\int_0^\infty \sigma = B(1 - q^m)$. Then r' is integrable, and an integration by parts gives $z\widehat{r}(z) - 1 = -\ell\widehat{r}(z) + \widehat{\sigma}(z)\widehat{r}(z)$. Recall that r is continuous and nonnegative with $r(0) = 1$, so that $\widehat{r}(z) > 0$. This yields

$$\widehat{r}(z)^{-1} = z + \ell - \widehat{\sigma}(z). \quad (7.25)$$

In the same way, $c' = -r$ and $(c^m)' = -\sigma/B$, together with $c(0) = 1$, give

$$z\widehat{c}(z) = 1 - \widehat{r}(z), \quad z\widehat{c^m}(z) = 1 - \widehat{\sigma}(z)/B. \quad (7.26)$$

We now compute the transform of the right side of (7.23). Integrating first over $a, b \geq u$, we find that the transform of the noise term is

$$2 \int_0^\infty e^{-(z_1 + z_2)u} \, du \, \widehat{r}(z_1)\widehat{r}(z_2) = \frac{2\widehat{r}(z_1)\widehat{r}(z_2)}{z_1 + z_2}.$$

The memory term is a two-dimensional convolution of $r(a)r(b)$ with the bounded function D_{c^m} , and its transform is $B\widehat{r}(z_1)\widehat{r}(z_2)\widehat{D}_{c^m}(z_1, z_2)$. All these integrals converge absolutely, because r is integrable and D_{c^m} is bounded, which justifies the use of Fubini's theorem. Using (7.24) with $f = c^m$ and (7.26), we can write

$$B\widehat{D}_{c^m}(z_1, z_2) = \frac{2\{\widehat{\sigma}(z_2) - \widehat{\sigma}(z_1)\}}{z_1^2 - z_2^2}.$$

Adding the transforms of the two terms, we have, for $z_1 \neq z_2$,

$$\begin{aligned} \frac{2\widehat{r}(z_1)\widehat{r}(z_2)}{z_1 + z_2} + B\widehat{r}(z_1)\widehat{r}(z_2)\widehat{D}_{c^m}(z_1, z_2) &= \frac{2\widehat{r}(z_1)\widehat{r}(z_2)}{z_1^2 - z_2^2} \{z_1 - z_2 + \widehat{\sigma}(z_2) - \widehat{\sigma}(z_1)\} \\ &= \frac{2\widehat{r}(z_1)\widehat{r}(z_2)}{z_1^2 - z_2^2} \{\widehat{r}(z_1)^{-1} - \widehat{r}(z_2)^{-1}\} \\ &= \frac{2\{\widehat{r}(z_2) - \widehat{r}(z_1)\}}{z_1^2 - z_2^2} = \widehat{D}_c(z_1, z_2). \end{aligned}$$

In the second line, we used (7.25) at z_1 and at z_2 . In the last equality, we used (7.24) with $f = c$ and the first identity in (7.26).

Both sides of (7.23) are bounded continuous functions on $[0, \infty)^2$. Indeed, the left side is bounded by one, the noise term by $2 \int_0^\infty r = 2(1 - q)$, and the memory term by $B(1 - q)^2$. We have shown that their double Laplace transforms agree for $z_1 \neq z_2$, and by continuity of the transforms in (z_1, z_2) they agree everywhere on $(0, \infty)^2$. Let F denote either side. For fixed $z_2 > 0$, the function $a \mapsto \int_0^\infty e^{-z_2 b} F(a, b) db$ is bounded and continuous, by dominated convergence, and its Laplace transform is $\hat{F}(\cdot, z_2)$. By the uniqueness theorem for Laplace transforms of continuous functions, the two such functions coincide for every $a \geq 0$. For fixed a , the functions $b \mapsto F(a, b)$ are bounded and continuous and, by what we just proved, have the same Laplace transform. They therefore coincide. This proves (7.23), and hence (7.16) and (7.17).

Step 4: the energy subtraction. By (7.6), we have

$$h(P_T v_t) = \sqrt{p} \langle \mathcal{I}^* P_T v_t, (P_T v_t)^{\otimes m} \rangle = \beta p \int_0^t R(t, u) K_T(t, u)^m du.$$

For $u \leq T$, we have $K_T(t, u) = C(t, u)$ by Step 1. We subtract the identity $h(v_t) = e_\beta(t) = \beta p \int_0^t R(t, u) C(t, u)^m du$ of Proposition 7.1 from the last display. The integrals over $[0, T]$ cancel. At $t = T + L$, using $K_T(T + L, u) = Q_T(L, u - T) = Q_T(u - T, L)$, we obtain

$$h(P_T v_{T+L}) = e_\beta(T + L) + \beta p \int_T^{T+L} R(T + L, u) \{Q_T(u - T, L)^m - C(T + L, u)^m\} du. \quad (7.27)$$

Only an integral over a future interval of fixed length L remains. Now let $T = T_n$. By Proposition 5.7, we have $e_\beta(T_n + L) \rightarrow e_\infty$. With the change of variables $\tau = T_n + L - u \in [0, L]$, we have

$$\begin{aligned} R(T_n + L, T_n + L - \tau) &\rightarrow r(\tau), & C(T_n + L, T_n + L - \tau) &\rightarrow c(\tau), \\ Q_n(L - \tau, L) &\rightarrow c(2L - \tau), \end{aligned}$$

uniformly in $\tau \in [0, L]$, by Proposition 5.7 and (7.16). All these functions take values in $[0, 1]$, and the integral in (7.27) converges to the integral in (7.18). This completes the proof. $\square$

By the reflection identity, the Hankel kernel $(a, b) \mapsto c(a + b)$ is a limit of Gram kernels. This leads to the following representation of every stationary profile.

Corollary 7.7 (Spectral representation of every stationary cluster). *Let (c, r, ℓ) be a stationary cluster, as in Proposition 5.7, with plateau q . Then there exists a probability measure ν on $[0, \infty)$ such that*

$$c(t) = \int_{[0, \infty)} e^{-xt} \nu(dx) \quad (t \geq 0), \quad \nu(\{0\}) = q. \quad (7.28)$$

In particular, c and $r = -c'$ are completely monotone on $(0, \infty)$.

Proof. The representation of a bounded continuous function with positive semidefinite Hankel matrices as a Laplace transform is classical; see Widder [122, Chapter IV, Section 12, Theorems 12a–12b, pp. 160–161] for Bernstein’s theorem on completely monotone functions, and Berg, Christensen, and Ressel [28, Chapter 4, Corollary 4.5, p. 114] for positive definite functions on semigroups. Complete monotonicity for Brownian mode-coupling equations is also classical; Fransch and Voigtmann proved it for their mixture model by an iteration argument [71, arXiv version 1, Section III]. Here it follows from the reflection identity, and we include the short semigroup argument to keep the proof self-contained. The argument has three steps. We show that the Hankel kernel is positive semidefinite, construct a semigroup of shifts, and apply the spectral theorem.

Step 1: positivity. Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to (c, r, ℓ) . For $a_1, \dots, a_k \geq 0$ and $\alpha_1, \dots, \alpha_k \in \mathbb{R}$, we have

$$\sum_{i,j=1}^k \alpha_i \alpha_j \langle M_{T_n}(a_i), M_{T_n}(a_j) \rangle = \left\| \sum_{i=1}^k \alpha_i M_{T_n}(a_i) \right\|^2 \geq 0.$$

Letting $n \rightarrow \infty$ and using (7.16), we obtain

$$\sum_{i,j=1}^k \alpha_i \alpha_j c(a_i + a_j) \geq 0. \quad (7.29)$$

Step 2: the shift semigroup. Let $\mathcal{V}$ be the real vector space of finite formal linear combinations of symbols φ_a , $a \geq 0$, with the bilinear form determined by $\langle \varphi_a, \varphi_b \rangle = c(a + b)$. By (7.29), this form is positive semidefinite. Let $\mathcal{G}$ be the Hilbert space obtained by taking the quotient of $\mathcal{V}$ by its null vectors and completing. For $w = \sum_i \alpha_i \varphi_{a_i} \in \mathcal{V}$ and $\sigma \geq 0$, let $w_\sigma = \sum_i \alpha_i \varphi_{a_i + \sigma}$ and

$$\psi_w(\sigma) = \langle w_\sigma, w_\sigma \rangle = \sum_{i,j} \alpha_i \alpha_j c(a_i + a_j + 2\sigma).$$

Since $|c| \leq 1$, we see that $\psi_w(\sigma) \leq (\sum_i |\alpha_i|)^2$. For $n \geq 0$, we have $\psi_w((n+1)\sigma) = \langle w_{n\sigma}, w_{(n+2)\sigma} \rangle$, and the Cauchy–Schwarz inequality gives

$$\psi_w((n+1)\sigma)^2 \leq \psi_w(n\sigma) \psi_w((n+2)\sigma). \quad (7.30)$$

We claim that $\psi_w(\sigma) \leq \psi_w(0)$. If $\psi_w(0) = 0$, then (7.30) with $n = 0$ implies that $\psi_w(\sigma) = 0$. Suppose that $\psi_w(0) > 0$ and $\psi_w(\sigma) > \psi_w(0)$, and let $\kappa = \psi_w(\sigma) / \psi_w(0) > 1$. Using (7.30) and induction on n , we find that $\psi_w((n+1)\sigma) \geq \kappa \psi_w(n\sigma) > 0$ for every $n \geq 0$. Then $\psi_w(n\sigma) \geq \kappa^n \psi_w(0) \rightarrow \infty$, which contradicts the bound $\psi_w \leq (\sum_i |\alpha_i|)^2$. This proves the claim.

The claim says that $w \mapsto w_\sigma$ does not increase the norm. In particular, it maps null vectors to null vectors, and it defines a contraction S_σ of $\mathcal{G}$ with $S_\sigma \varphi_a = \varphi_{a+\sigma}$. Each S_σ is self-adjoint, because $\langle S_\sigma \varphi_a, \varphi_b \rangle = c(a+b+\sigma) = \langle \varphi_a, S_\sigma \varphi_b \rangle$. Further, $S_0 = \text{Id}$ and $S_\sigma S_{\sigma'} = S_{\sigma+\sigma'}$, so that $S_\sigma = S_{\sigma/2}^* S_{\sigma/2}$ is a positive operator.

Step 3: the spectral representation. For an integer $j \geq 1$, the operator $S_{1/j}$ is positive and $(S_{1/j})^j = S_1$. By uniqueness of positive j -th roots, we have $S_{1/j} = S_1^{1/j}$, and therefore $S_\sigma = S_1^\sigma$ for every rational $\sigma \geq 0$. Let μ be the spectral measure of the positive contraction S_1 at the vector φ_0 . It is a measure on $[0, 1]$ with total mass $\|\varphi_0\|^2 = c(0) = 1$, and for rational $\sigma > 0$ we have

$$c(\sigma) = \langle \varphi_0, S_\sigma \varphi_0 \rangle = \int_{(0,1]} y^\sigma \mu(dy).$$

Both sides are continuous in $\sigma > 0$, the right side by dominated convergence, and the identity extends to every $\sigma > 0$. Letting $\sigma \downarrow 0$, we see that $\mu((0, 1]) = c(0) = 1$, that is, $\mu(\{0\}) = 0$. Let ν be the image of μ under the map $y \mapsto -\log y$ from $(0, 1]$ to $[0, \infty)$. Then ν is a probability measure, and the identity above becomes the first relation in (7.28), also at $t = 0$. By dominated convergence, $c(t) \rightarrow \nu(\{0\})$ as $t \rightarrow \infty$, which gives $\nu(\{0\}) = q$. Finally, for $n \geq 0$ and $t > 0$, the function $x \mapsto x^n e^{-xt}$ is bounded on $[0, \infty)$, uniformly for t in compact subsets of $(0, \infty)$. Differentiating under the integral, we find

$$(-1)^n c^{(n)}(t) = \int_{[0,\infty)} x^n e^{-xt} \nu(dx) \geq 0, \quad (-1)^n r^{(n)}(t) = \int_{[0,\infty)} x^{n+1} e^{-xt} \nu(dx) \geq 0,$$

and c and r are completely monotone on $(0, \infty)$. □

Selection of the plateau and correlation aging

In this section, we prove the parts of Theorem 1.1 that concern the plateau, the energy, and the loss of memory. We begin with Lemma 8.3, which states that every positive plateau of a stationary cluster is marginal. In Proposition 8.4, we prove that the translated kernels, the multiplier, and the energy converge at every temperature. Theorem 8.5 identifies the plateau and the energy for $\beta \geq \beta_d$, and Theorem 8.8 shows that the solution forgets every fixed earlier state. Corollary 8.9 states the two ordered limits of the correlation. At the end of Section 8.6, we combine these results into a proof of Theorem 1.1.

We take as input two facts about prefix centers, which we state as Proposition 8.1 in Section 8.1 and prove in Sections 9 and 10. Besides these two facts, we use the stationary clusters of Proposition 5.7, the scalar constraints of Proposition 6.1, the uniqueness result Lemma 6.2, the causal tree realization of Proposition 7.1, the reflection identity of Proposition 7.5, and the uniform bounds of Section 4. The convergence proved in Proposition 8.4 and the marginality proved in Lemma 8.3 are exactly the hypotheses of part (b) of Proposition 8.1. We use the selected profile, the marginality relation, and the forgetting theorem throughout Parts III to V.

Standing assumptions and notation. Throughout this section, we fix $p \geq 3$ and $\beta > 0$, and (C, R, λ) denotes the uniformly initialized hard CK solution of Section 3. All long-time limits below concern this deterministic solution, for which the dimension limit has already been taken on each fixed time horizon. Recall from (4.1) that the energy $e_\beta(t)$ satisfies $\lambda(t) = 1 + p\beta e_\beta(t)$. For $T \geq 0$, the translated kernels are $C_T(a, b) = C(T + a, T + b)$ for $a, b \geq -T$, $R_T(a, b) = R(T + a, T + b)$ for $-T \leq b \leq a$, and $\lambda_T(a) = \lambda(T + a)$ for $a \geq -T$.

We fix the Fisher majorant a_F and the dimension subsequence (N_j) of Theorem 5.3 once, before choosing any waiting-time sequence or stationary cluster. All comparisons with finite-dimensional forks use this same dimension subsequence. We may pass to subsequences of waiting times to obtain stationary clusters, without changing the dimension extraction. The majorant is an auxiliary choice; the CK solution and the stationary profile identified below are unique.

As in Proposition 5.7, a *stationary cluster* is a triple (c, r, ℓ) obtained as the locally uniform limit of $(C_{T_n}, R_{T_n}, \lambda_{T_n})$ along some sequence $T_n \rightarrow \infty$. Proposition 5.7 implies that $c(0) = r(0) = 1$, $c' = -r$, $0 \leq c \leq 1$, and $r \geq 0$, that the plateau $q = \lim_{\tau \rightarrow \infty} c(\tau)$

lies in $[0, 1)$, and that $\int_0^\infty r(u) du = 1 - q$. By (4.23), we have $0 \leq R \leq C$, and passing to the limit gives $r \leq c$. As in Propositions 5.7 and 7.5, we write $e_\infty = (\ell - 1)/(p\beta)$ for the energy of the cluster. With $B = \beta^2 p$ and $\gamma = \ell - B$, we use the functions of Proposition 6.1,

$$\phi(y) = \gamma + By^{p-1}, \quad F(y) = \phi(y)(1 - y) - 1.$$

Recall from that proposition that $F(q) = 0$, that $F(y) < 0$ for $q < y < 1$, and that

$$\ell = \frac{1}{1 - q} + \beta^2 p(1 - q^{p-1}).$$

We also use the tree realization of Section 7. The unit vectors v_t satisfy $\langle v_t, v_s \rangle = C(t, s)$, and v_t lies in the closed subspace $\mathcal{H}_t$ generated by trees whose noise leaves carry times at most t . We write P_T for the orthogonal projection onto $\mathcal{H}_T$. The tree energy is $h(v) = \sqrt{p} \langle v, \mathcal{I}(v^{\otimes m}) \rangle$ with $m = p - 1$, and $h(v_t) = e_\beta(t)$.

8.1 Prefix centers and the threshold inputs

For $T \geq 0$ and $a \geq 0$, we define the prefix center

$$M_T(a) = P_T v_{T+a}.$$

Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to a stationary cluster (c, r, ℓ) with energy e_∞ . Recall from Proposition 7.5 that, for every fixed $L \geq 0$,

$$\begin{aligned} \|M_{T_n}(L)\|^2 &\longrightarrow d_L = c(2L), \\ h(M_{T_n}(L)) &\longrightarrow h_L = e_\infty + \beta p \int_0^L r(\tau) \{c(2L - \tau)^{p-1} - c(\tau)^{p-1}\} d\tau. \end{aligned}$$

Note that h is homogeneous of degree p . If $d_L > 0$, the energy of the normalized center $M_{T_n}(L) / \|M_{T_n}(L)\|$ converges to $h_L / d_L^{p/2}$.

We now explain why the prefix centers are the right objects for selecting the plateau. The state v_{T+L} is built from the initial vector and from the noise leaves at times up to $T + L$. The projection P_T keeps the part built from the initial vector and noise received by time T , including memory interactions after T . In physical terms, we may think of $M_T(L)$ as the average of the state at time $T + L$ over independent continuations of the thermal noise after time T . This interpretation goes back to the trajectories with a common past and independent futures studied by Cugliandolo and Dean [50, arXiv version 1, Section II] and by Barrat, Burioni, and Mézard [11, Section 4], and to the averages over cloned futures proposed by Franz and Virasoro [72, arXiv version 3, Section III]. In agreement with this interpretation, the limiting squared norm $d_L = c(2L)$ is the stationary correlation at the reflected lag $2L$, which is the overlap of two continuations of length

L from a common past; this reflected formula appears in [11, Section 2, Equation (9)]. Time-averaged magnetizations of a similar kind appear in the analysis of Cugliandolo and Kurchan [52, arXiv version 1, Section 5.1], and noise-averaged magnetizations and their relation to the threshold appear in the dynamical TAP approach of Biroli [31, arXiv version 2, Section III.B]. We use three properties of the centers.

- Every earlier state v_s with $s \leq T$ lies in $\mathcal{H}_T$, so that $\langle v_{T+L}, v_s \rangle = \langle M_T(L), v_s \rangle$. The correlation of a late state with a fixed earlier state is carried entirely by the center.
- For a stationary cluster with $q > 0$, the normalized limiting energies of the centers satisfy $h_L/d_L^{p/2} \leq E_{\text{th}}$ for every $L \geq 0$. Here $E_{\text{th}} = 2\sqrt{(p-1)/p}$ is the threshold energy of the pure p -spin landscape [7, Equation (2.13)].
- Suppose that the whole translated family converges to one stationary profile whose plateau is positive and marginal. Then the normalized energies of the centers approach E_{th} as $L \rightarrow \infty$, and this forces the solution to forget every fixed earlier state.

The first property is exact. The ceiling in the second and the forgetting in the third are the content of the following proposition, which is the only input that we take from Sections 9 and 10.

Proposition 8.1 (Threshold inputs). *Fix $p \geq 3$ and $\beta > 0$.*

- (a) *Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to a stationary cluster (c, r, ℓ) with plateau $q > 0$, as in Proposition 5.7, and let d_L and h_L be the limits in (7.17) and (7.18). Then, for every $L \geq 0$,*

$$h_L \leq E_{\text{th}} d_L^{p/2}. \quad (8.1)$$

- (b) *Suppose that the translated kernels converge along every sequence $T \rightarrow \infty$ to one stationary profile (c, r) , and that its plateau q is positive and satisfies $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$. Then, for every $0 \leq S < \infty$,*

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq S} |C(t, s)| = 0. \quad (8.2)$$

For $p \in \{3, 4\}$, both parts are proved in Section 9. For every $p \geq 3$, they are proved in Section 10.

For $p \in \{3, 4\}$, we prove both parts in Section 9 using a deterministic inequality on the tree Hilbert space (Theorem 9.2 and Lemma 9.3). It states that a unit tree vector has energy strictly below E_{th} , and that near equality forces its mass to leave every fixed set of tree heights. Applying this inequality to the normalized centers, we obtain part (a). For part (b), we show that the mass of the centers escapes from every fixed height as they approach the threshold, while an earlier state v_s with $s \leq S$ is approximated uniformly by finitely many heights. From the same argument, we also obtain the quantitative estimate Proposition 9.5, which we use again in Sections 11, 13, and 14.

For every $p \geq 3$, we prove both parts in Section 10. There we replace the projection by the mean of finitely many independent futures of the finite-dimensional dynamics from a common past. Using the fixed-horizon approximation of Sellke [112, Lemmas 3.1–3.2], we approximate the normalized mean by globally Lipschitz functions of the disorder coefficients. The algorithmic threshold of Huang and Sellke [92, Theorem 1 and Corollary 1.8] bounds the energy of these approximations by E_{th} in the dimension limit, and we transfer this bound to the mean by sending the approximation error to zero. Part (a) follows from Proposition 10.4 by letting the number of futures tend to infinity at fixed L . Part (b) follows from Proposition 10.8 and the finite-net argument in Section 10.5. Its proof combines the external-field form of the same threshold with adaptive Gaussian conditioning and an Euler approximation of the earlier state.

The references to later sections do not create a circular argument. The proofs of part (a) in Sections 9 and 10 apply to an arbitrary stationary cluster with positive plateau, and they use only results of Sections 3 to 7. Part (b) contains in its hypotheses the two facts that we deduce from part (a) in this section: convergence of the whole translated family, and marginality of the plateau. Besides these hypotheses, its proofs use the algebraic identity (8.5), which we prove in Lemma 8.2 without using Proposition 8.1. In this section, the order of deduction is as follows:

- part (a) and Lemma 8.2 imply marginality, Lemma 8.3;
- marginality implies convergence at every temperature, Proposition 8.4, and the selection of the plateau, Theorem 8.5;
- these results verify the hypotheses of part (b), which then gives forgetting, Theorem 8.8, and correlation aging, Corollary 8.9.

8.2 Marginality of positive plateaus

We set

$$q_m = \frac{p-2}{p}, \quad \beta_m^2 = \frac{p^{p-1}}{4(p-1)(p-2)^{p-2}}, \quad E_{\text{th}} = 2\sqrt{\frac{p-1}{p}}. \quad (8.3)$$

To explain the role of q_m and β_m , let $G(q) = p(p-1)q^{p-2}(1-q)^2$ for $q \in [0, 1]$, so that the marginality relation $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$ reads $\beta^2 G(q) = 1$. We have

$$G'(q) = p(p-1)q^{p-3}(1-q)\{(p-2) - pq\}.$$

The function G vanishes at 0 and 1, increases strictly on $[0, q_m]$, and decreases strictly on $[q_m, 1]$. Its maximum is

$$G(q_m) = \frac{4(p-1)(p-2)^{p-2}}{p^{p-1}} = \beta_m^{-2}.$$

In other words, β_m is the inverse temperature at which $\beta^2 G$ first reaches one on $[0, 1]$. For $\beta \geq \beta_m$, we denote by $q_+(\beta)$ the unique root of $\beta^2 G(q) = 1$ in $[q_m, 1]$; at $\beta = \beta_m$, it is the double root q_m . For $\beta > \beta_m$, the equation has a second root, which lies in $(0, q_m)$, and for $\beta < \beta_m$ it has no root in $[0, 1]$.

The next lemma computes the limit of the center energies of Proposition 7.5 as L tends to infinity, and compares it with E_{th} . It is a computation with the scalar constraints of Proposition 6.1 and uses no threshold input. We use it again in Sections 9 and 10.

Lemma 8.2 (Limiting energy of prefix centers). *Let (c, r, ℓ) be a stationary cluster with plateau q , as in Proposition 5.7, let $e_\infty = (\ell - 1)/(p\beta)$, and let d_L and h_L be defined by (7.17) and (7.18). Then $d_L \rightarrow q$ and $h_L \rightarrow h$ as $L \rightarrow \infty$, where*

$$h = e_\infty - \beta \{1 - q^p - pq^{p-1}(1 - q)\}. \quad (8.4)$$

If $q > 0$, define $z = \beta \sqrt{p(p-1)} q^{(p-2)/2} (1 - q) > 0$. Then

$$\frac{h}{q^{p/2}} = \sqrt{\frac{p-1}{p}} \left(z + \frac{1}{z} \right) \geq E_{\text{th}}, \quad (8.5)$$

and equality holds in (8.5) if and only if $\beta^2 p(p-1) q^{p-2} (1 - q)^2 = 1$.

Proof. As c is nonincreasing with limit q , we have $d_L = c(2L) \rightarrow q$. For h_L , we extend the integrand in (7.18) by zero outside $[0, L]$ and write the integral as an integral over $[0, \infty)$ of

$$g_L(\tau) = \begin{cases} r(\tau) \{c(2L - \tau)^{p-1} - c(\tau)^{p-1}\}, & 0 \leq \tau \leq L, \\ 0, & \tau > L. \end{cases}$$

For fixed τ , we have $2L - \tau \rightarrow \infty$ as $L \rightarrow \infty$, and $g_L(\tau)$ converges to $r(\tau) \{q^{p-1} - c(\tau)^{p-1}\}$. Since $0 \leq c \leq 1$, we have $|g_L(\tau)| \leq r(\tau)$, and r is integrable with $\int_0^\infty r = 1 - q$. By dominated convergence, we obtain

$$\lim_{L \rightarrow \infty} h_L = e_\infty + \beta p q^{p-1} (1 - q) - \beta p \int_0^\infty r(\tau) c(\tau)^{p-1} d\tau.$$

Using $r = -c'$, $c(0) = 1$, and $c(\tau) \rightarrow q$, we find

$$\int_0^\infty r(\tau) c(\tau)^{p-1} d\tau = -\frac{1}{p} \int_0^\infty (c^p)'(\tau) d\tau = \frac{1 - q^p}{p}.$$

Inserting this into the previous display, we obtain (8.4).

Now suppose that $q > 0$. Using the formula for ℓ in (6.2), we write the energy of the cluster as

$$e_\infty = \frac{\ell - 1}{p\beta} = \frac{q}{p\beta(1 - q)} + \beta(1 - q^{p-1}).$$

We insert this into (8.4) and note that $(1 - q^{p-1}) - (1 - q^p) = -q^{p-1}(1 - q)$. This gives

$$h = \frac{q}{p\beta(1-q)} + \beta(p-1)q^{p-1}(1-q).$$

By the definition of z , we have

$$\sqrt{\frac{p-1}{p}} z = \beta(p-1)q^{(p-2)/2}(1-q), \quad \sqrt{\frac{p-1}{p}} \frac{1}{z} = \frac{q^{-(p-2)/2}}{p\beta(1-q)},$$

and the sum of the two right sides is $h/q^{p/2}$. This proves the identity in (8.5). For $z > 0$, we have $z + z^{-1} \geq 2$, with equality if and only if $z = 1$. Recalling that $2\sqrt{(p-1)/p} = E_{\text{th}}$, we deduce the inequality in (8.5), and equality holds if and only if $z = 1$, that is, if and only if $z^2 = \beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$. $\square$

Lemma 8.3 (Marginality of every positive cluster plateau). *Let (c, r, ℓ) be a stationary cluster, as in Proposition 5.7, with plateau $q > 0$. Then*

$$\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1, \quad q \geq q_m. \quad (8.6)$$

Proof. First, part (a) of Proposition 8.1 gives an upper bound on the limiting center energy that matches the lower bound in (8.5), and this forces marginality. Second, the strict sign condition of Proposition 6.1 excludes the smaller root.

Step 1: marginality. Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to (c, r, ℓ) . Since $q > 0$, part (a) of Proposition 8.1 applies along this sequence, and (8.1) states that $h_L \leq E_{\text{th}} d_L^{p/2}$ for every $L \geq 0$. Letting $L \rightarrow \infty$ and using $d_L \rightarrow q$ and $h_L \rightarrow h$ from Lemma 8.2, we obtain

$$h \leq E_{\text{th}} q^{p/2}.$$

Together with the lower bound in (8.5), this forces $h/q^{p/2} = E_{\text{th}}$. The equality case of Lemma 8.2 then implies the first relation in (8.6).

Step 2: the larger root. We use the functions ϕ and F from the beginning of this section; recall that $F(q) = 0$ and $F(y) < 0$ for $q < y < 1$. Differentiating, we have

$$F'(y) = B(p-1)y^{p-2}(1-y) - \phi(y), \quad F''(y) = B(p-1)y^{p-3}\{(p-2) - py\}.$$

Since $\phi(q) = 1/(1-q)$ by (6.2), we find

$$F'(q) = \frac{\beta^2 p(p-1)q^{p-2}(1-q)^2 - 1}{1-q} = 0,$$

where we used Step 1 in the last equality. Suppose that $0 < q < q_m$. Then $F''(q) > 0$. Since F is a polynomial with $F(q) = F'(q) = 0$, Taylor's theorem implies that

$$F(q + \varepsilon) = \frac{1}{2}F''(q)\varepsilon^2 + O(\varepsilon^3) > 0$$

for all sufficiently small $\varepsilon > 0$. This contradicts $F(q + \varepsilon) < 0$. We conclude that $q \geq q_m$. $\square$

The first relation in (8.6) is the marginality condition of the asymptotic solution of Cugliandolo and Kurchan [51, Equation (7)], written in the conventions of Section 3. In view of the discussion after (8.3), the relations (8.6) say that $\beta \geq \beta_m$ and $q = q_+(\beta)$. A stationary cluster with positive plateau can occur only for $\beta \geq \beta_m$, and its plateau is then determined by β .

8.3 Convergence at every fixed temperature

Proposition 8.4 (Convergence at every fixed temperature). *For every $p \geq 3$ and $\beta > 0$, the multiplier $\lambda(t)$ and the energy $e_\beta(t)$ have limits ℓ_β and $e_\beta(\infty)$ as $t \rightarrow \infty$. All bounded-lag translates of C and R converge locally uniformly to a unique stationary FDT profile (c, r) , that is, for every $0 \leq A < \infty$,*

$$\sup_{0 \leq s \leq t \leq A} \{ |C(T+t, T+s) - c(t-s)| + |R(T+t, T+s) - r(t-s)| \} \longrightarrow 0 \quad (T \rightarrow \infty). \quad (8.7)$$

Its plateau q belongs to the following set:

$$q \in \begin{cases} \{0\}, & 0 < \beta < \beta_m, \\ \{0, q_+(\beta)\}, & \beta \geq \beta_m. \end{cases} \quad (8.8)$$

The positive branch is used only on its domain $\beta \geq \beta_m$. The limiting multiplier is

$$\ell_\beta = \frac{1}{1-q} + \beta^2 p (1 - q^{p-1}),$$

as in (6.2), and $e_\beta(\infty) = (\ell_\beta - 1)/(p\beta)$.

Proof. First, the multiplier of every stationary cluster lies in a set with at most two elements. Second, the late limit points of λ form an interval, each point of which is the multiplier of a stationary cluster; combined with the first step, this forces λ to converge. Third, we identify all stationary clusters using Lemma 6.2 and upgrade subsequential convergence to convergence of the whole translated family by precompactness.

Step 1: possible multipliers. Let (c, r, ℓ) be a stationary cluster with plateau q . If $q = 0$, then (6.2) gives $\ell = 1 + \beta^2 p$. If $q > 0$, then Lemma 8.3 and the discussion after it imply that $\beta \geq \beta_m$ and $q = q_+(\beta)$, and by (6.2) we have $\ell = (1 - q_+(\beta))^{-1} + \beta^2 p (1 - q_+(\beta)^{p-1})$. In both cases, ℓ belongs to the set

$$\mathcal{L}_\beta = \begin{cases} \{1 + \beta^2 p\}, & \beta < \beta_m, \\ \{1 + \beta^2 p, (1 - q_+(\beta))^{-1} + \beta^2 p (1 - q_+(\beta)^{p-1})\}, & \beta \geq \beta_m, \end{cases}$$

which has at most two elements.

Step 2: convergence of the multiplier. Recall from Propositions 4.3 and 4.4 that λ is continuous on $[0, \infty)$ and satisfies $1 \leq \lambda \leq \Lambda$. Let

$$a = \liminf_{t \rightarrow \infty} \lambda(t), \quad b = \limsup_{t \rightarrow \infty} \lambda(t),$$

which are finite. We say that y is a late limit point of λ if there exists a sequence $T_n \rightarrow \infty$ such that $\lambda(T_n) \rightarrow y$. We claim that every $y \in [a, b]$ is a late limit point. For $y \in \{a, b\}$, this holds by the definition of the lower and upper limits. If $a < y < b$, then for every n there exist times $t_1, t_2 \geq n$ such that $\lambda(t_1) < y < \lambda(t_2)$. By the intermediate value theorem, there exists T_n between t_1 and t_2 such that $\lambda(T_n) = y$, and $T_n \geq n$. This proves the claim.

Now let $y \in [a, b]$, and let $T_n \rightarrow \infty$ be such that $\lambda(T_n) \rightarrow y$. By Proposition 5.7, a subsequence of the translated family $(C_{T_n}, R_{T_n}, \lambda_{T_n})$ converges locally uniformly to a stationary cluster (c, r, ℓ) , and along it λ_{T_n} converges to the constant ℓ . Evaluating at offset zero, we find $\ell = y$, and Step 1 gives $y \in \mathcal{L}_\beta$. We have shown that

$$[a, b] \subseteq \mathcal{L}_\beta.$$

An interval contained in a finite set is a single point. Hence $a = b$, and $\lambda(t)$ converges to a limit $\ell_\beta \in \mathcal{L}_\beta$. By (4.1), we have $e_\beta(t) = (\lambda(t) - 1)/(p\beta)$, and the energy converges to $e_\beta(\infty) = (\ell_\beta - 1)/(p\beta)$.

Step 3: identification of the profile. Let (c, r, ℓ) be any stationary cluster, obtained along a sequence $T_n \rightarrow \infty$. Since $\lambda(T_n + a) \rightarrow \ell_\beta$ for every offset a , we have $\ell = \ell_\beta$. By (5.24) and (5.25), the pair (c, r) solves

$$c' = -r, \quad r' = -\ell_\beta r + \beta^2 p(p-1)(c^{p-2}r) * r, \quad c(0) = r(0) = 1,$$

and it is bounded, since $0 \leq r \leq c \leq 1$. Lemma 6.2 states that this system has at most one locally bounded solution. All stationary clusters therefore coincide; we denote their common profile by (c, r) . The argument uses only the value ℓ_β of the multiplier. In particular, it applies when the two elements of $\mathcal{L}_\beta$ coincide, which happens at $\beta = \beta_d$. Every stationary cluster also has the same energy $e_\infty = (\ell_\beta - 1)/(p\beta) = e_\beta(\infty)$.

Suppose that (8.7) fails for some $0 \leq A < \infty$. Then there exist $\varepsilon > 0$ and a sequence $T_n \rightarrow \infty$ such that, for every n , the supremum in (8.7) at $T = T_n$ is at least ε . Using the precompactness in Proposition 5.7, we extract a subsequence of the translates that converges locally uniformly to a stationary cluster, which is (c, r, ℓ_β) by the preceding paragraph. Along this subsequence, the supremum over the compact set $\{0 \leq s \leq t \leq A\}$ tends to zero, which is a contradiction. This proves (8.7).

Finally, if the plateau q of c is positive, then Lemma 8.3 implies that $\beta \geq \beta_m$ and $q = q_+(\beta)$. This proves (8.8), and it shows that the second alternative is absent for $\beta < \beta_m$. The formula for ℓ_β is (6.2) for the cluster (c, r, ℓ_β) . $\square$

8.4 The dynamical transition and the selected energy

Recall the dynamical transition parameters

$$q_d = \frac{p-2}{p-1}, \quad \beta_d^2 = \frac{(p-1)^{p-1}}{p(p-2)^{p-2}}.$$

We compare them with the constants in (8.3). First, $q_d - q_m = (p-2)/(p(p-1)) > 0$. Second, a direct computation shows that $G(q_d) = p(p-2)^{p-2}/(p-1)^{p-1} = \beta_d^{-2}$. As G decreases strictly on $[q_m, 1]$, we find that $\beta_d^{-2} < G(q_m) = \beta_m^{-2}$, that is, $\beta_d > \beta_m$, and that $q_+(\beta_d) = q_d$. For $\beta > \beta_d$, we have $G(q_+(\beta)) = \beta^{-2} < G(q_d)$, and then $q_+(\beta) > q_d$. For $\beta \geq \beta_d$, we write

$$q_\beta = q_+(\beta),$$

so that $q_\beta \geq q_d$, with equality exactly at $\beta = \beta_d$.

Theorem 8.5 (The selected stationary sector). *For every $p \geq 3$ and $\beta \geq \beta_d$, the unique stationary plateau is q_β . In particular, the limit ℓ_β of $\lambda(t)$ and the limiting energy satisfy*

$$\ell_\beta = \frac{1}{1-q_\beta} + \beta^2 p (1 - q_\beta^{p-1}), \quad (8.9)$$

$$e_\beta(\infty) = \beta \{1 + (p-2)q_\beta^{p-1} - (p-1)q_\beta^p\}. \quad (8.10)$$

For every $0 \leq A < \infty$,

$$\sup_{0 \leq s \leq t \leq A} \{ |C(T+t, T+s) - c(t-s)| + |R(T+t, T+s) - r(t-s)| \} \longrightarrow 0 \quad (T \rightarrow \infty), \quad (8.11)$$

where $r = -c'$ and $c(\tau) \rightarrow q_\beta$. At the endpoint, $e_{\beta_d}(\infty) = \beta_d$ and $\ell_{\beta_d} = 1 + p\beta_d^2$.

Proof. In view of Proposition 8.4, it suffices to exclude the alternative $q = 0$. Suppose that $q = 0$. Then (6.2) gives $\ell_\beta = 1 + \beta^2 p$, which means that $\gamma = \ell_\beta - \beta^2 p = 1$ and

$$F(y) = (1 + \beta^2 p y^{p-1})(1-y) - 1 = y \{ \beta^2 p y^{p-2}(1-y) - 1 \}.$$

Since $q_d^{p-2}(1-q_d) = (p-2)^{p-2}/(p-1)^{p-1}$, we have $\beta^2 p q_d^{p-2}(1-q_d) = \beta^2/\beta_d^2$ and

$$F(q_d) = q_d \left(\frac{\beta^2}{\beta_d^2} - 1 \right).$$

For $\beta \geq \beta_d$, this is nonnegative. On the other hand, q_d lies in $(0, 1) = (q, 1)$, so $F(q_d) < 0$ by (6.3). This contradiction shows that $q > 0$.

At $\beta = \beta_d$, we have $F(q_d) = 0$, and the contradiction comes from the strict inequality in (6.3). In Proposition 6.1, that strict inequality follows from the bound $c'(t) \geq F(c(t))$

and from $c' = -r < 0$ at every finite time. In this way, the endpoint $\beta = \beta_d$ is included without any continuity argument in β .

We have shown that $q > 0$, and Proposition 8.4 then implies that $q = q_+(\beta) = q_\beta$. The formula (8.9) is the formula for ℓ_β in Proposition 8.4, (8.11) is (8.7), and the relation $r = -c'$ is part of (5.24). For the energy, we use the marginality relation $\beta^2 G(q_\beta) = 1$. Multiplying it by $q_\beta / (p\beta(1 - q_\beta))$, we obtain

$$\frac{q_\beta}{p\beta(1 - q_\beta)} = \beta(p - 1)q_\beta^{p-1}(1 - q_\beta).$$

Using $e_\beta(\infty) = (\ell_\beta - 1) / (p\beta)$ and (8.9), we find

$$e_\beta(\infty) = \frac{q_\beta}{p\beta(1 - q_\beta)} + \beta(1 - q_\beta^{p-1}) = \beta(p - 1)q_\beta^{p-1}(1 - q_\beta) + \beta(1 - q_\beta^{p-1}),$$

and expanding the right side yields (8.10). At $\beta = \beta_d$, we have $q_\beta = q_d$ and $(p - 1)q_d = p - 2$, so that

$$(p - 2)q_d^{p-1} - (p - 1)q_d^p = q_d^{p-1}\{(p - 2) - (p - 1)q_d\} = 0.$$

This shows that $e_{\beta_d}(\infty) = \beta_d$, and (4.1) gives $\ell_{\beta_d} = 1 + p\beta_d e_{\beta_d}(\infty) = 1 + p\beta_d^2$. $\square$

The plateau and the energy in Theorem 8.5 are those predicted by Cugliandolo and Kurchan. In the normalization of [51], the inverse temperature is $\sqrt{2}\beta$ and the energy is $-e/\sqrt{2}$; with this translation, (8.10) is their Equation (9), evaluated with the fluctuation-dissipation ratio $x = (p - 2)(1 - q)/q$ given after their Equation (12).

Remark 8.6 (The subsequent low-temperature limit). As $\beta \rightarrow \infty$, we have $q_\beta \rightarrow 1$, $\beta(1 - q_\beta) \rightarrow [p(p - 1)]^{-1/2}$, and

$$\lim_{\beta \rightarrow \infty} e_\beta(\infty) = 2\sqrt{\frac{p - 1}{p}} = E_{\text{th}}.$$

This agrees with the threshold energy of the pure p -spin landscape, computed by Crisanti and Sommers [47, Equation (9)] in the normalization of [51] and by Auffinger, Ben Arous, and Černý [7, Equation (2.13)] in the present one. It also agrees with the results of Sellke [112, Theorems 1.1 and 1.7], which show that, at low temperature, the energy of Langevin dynamics on a bounded time horizon comes close to E_{th} . Here $\beta \rightarrow \infty$ is taken after the dimension limit and after both long-time limits, and the calculation says nothing about the gradient flow at zero temperature.

To verify these limits, recall that $q_\beta \geq q_d > 0$. The marginality relation gives

$$(1 - q_\beta)^2 = \frac{1}{\beta^2 p(p - 1)q_\beta^{p-2}} \leq \frac{1}{\beta^2 p(p - 1)q_d^{p-2}} \rightarrow 0,$$

so $q_\beta \rightarrow 1$. Taking square roots in the same relation, we find

$$\beta(1 - q_\beta) = \frac{q_\beta^{-(p-2)/2}}{\sqrt{p(p-1)}} \rightarrow \frac{1}{\sqrt{p(p-1)}}.$$

Let $g(y) = 1 + (p-2)y^{p-1} - (p-1)y^p$, so that $e_\beta(\infty) = \beta g(q_\beta)$ by (8.10). We have $g(1) = 0$ and

$$g'(y) = (p-1)y^{p-2}\{(p-2) - py\}, \quad g'(1) = -2(p-1).$$

Let $\varepsilon = 1 - q_\beta$. By Taylor's theorem, there exists $\eta \in (q_\beta, 1)$ such that

$$g(q_\beta) = g(1) - g'(1)\varepsilon + \frac{1}{2}g''(\eta)\varepsilon^2 = 2(p-1)\varepsilon + \frac{1}{2}g''(\eta)\varepsilon^2.$$

On $[0, 1]$, we have $|g''| \leq K_p$ with $K_p = (p-1)\{(p-2)^2 + p(p-1)\}$, and therefore

$$|e_\beta(\infty) - 2(p-1)\beta(1 - q_\beta)| \leq \frac{1}{2}K_p \beta(1 - q_\beta)^2.$$

As $\beta \rightarrow \infty$, the right side equals $\frac{1}{2}K_p \cdot \beta(1 - q_\beta) \cdot (1 - q_\beta)$ and tends to zero, while $2(p-1)\beta(1 - q_\beta) \rightarrow 2(p-1)/\sqrt{p(p-1)} = E_{\text{th}}$.

Remark 8.7 (The unselected intermediate interval). For $0 < \beta < \beta_m$, Proposition 8.4 gives $q = 0$; then $\ell_\beta = 1 + \beta^2 p$ and $e_\beta(\infty) = \beta$. For $\beta_m \leq \beta < \beta_d$, it proves convergence but allows either the zero branch or the positive branch $q_+(\beta)$. In the latter case, writing $q_+ = q_+(\beta)$, we have

$$e_\beta(\infty) = \beta + \frac{q_+\{(p-2) - (p-1)q_+\}}{p\beta(p-1)(1 - q_+)^2} > \beta.$$

Indeed, the marginality relation can be written as $\beta^2 p q_+^{p-1} = q_+ / ((p-1)(1 - q_+)^2)$. Combining it with the formula for ℓ_β in Proposition 8.4, we have

$$e_\beta(\infty) - \beta = \frac{\ell_\beta - 1 - \beta^2 p}{p\beta} = \frac{1}{p\beta} \left\{ \frac{q_+}{1 - q_+} - \frac{q_+}{(p-1)(1 - q_+)^2} \right\},$$

and bringing the bracket to a common denominator gives the stated formula. For the sign, note that $\beta < \beta_d$ implies $G(q_+) = \beta^{-2} > G(q_d)$. As G decreases strictly on $[q_m, 1]$, we obtain $q_m \leq q_+ < q_d$. In particular, $(p-2) - (p-1)q_+ > 0$, and the correction is positive. We do not decide which alternative occurs in this interval, and the transition is not classified from below there; see Section 17.

8.5 Forgetting of fixed waiting times

Theorem 8.8 (Forgetting of fixed waiting times). *Let $p \geq 3$ and $\beta > 0$, and let (c, r) be the stationary profile of Proposition 8.4. Then*

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq S} \{C(t, s) + R(t, s)\} = 0 \quad (0 \leq S < \infty). \quad (8.12)$$

If the stationary plateau is zero, the stronger whole-row conclusion holds:

$$\sup_{0 \leq s \leq t} \{|C(t, s) - c(t - s)| + |R(t, s) - r(t - s)|\} \rightarrow 0 \quad (t \rightarrow \infty). \quad (8.13)$$

Proof. The proof has four steps. We first reduce (8.12) to a statement about C . When the plateau q is positive, we verify the hypotheses of part (b) of Proposition 8.1. When $q = 0$, we argue directly: the norm of the prefix center tends to zero when first $t \rightarrow \infty$ and then $L \rightarrow \infty$. Finally, we prove (8.13) by splitting each row into a recent strip and a remote part.

Step 1: reduction to the correlation. Recall from (4.23) that $0 \leq R(t, s) \leq C(t, s)$ for $t \geq s \geq 0$. For $t \geq S$, we have

$$\sup_{0 \leq s \leq S} \{C(t, s) + R(t, s)\} \leq 2 \sup_{0 \leq s \leq S} C(t, s),$$

and it suffices to show that $\sup_{0 \leq s \leq S} C(t, s) \rightarrow 0$ as $t \rightarrow \infty$.

Step 2: positive plateau. Suppose that $q > 0$. Then the translated kernels converge along every sequence $T \rightarrow \infty$ to the single profile (c, r) by Proposition 8.4, and the plateau satisfies $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$ by Lemma 8.3. These are the hypotheses of part (b) of Proposition 8.1. Then (8.2) implies that $\sup_{0 \leq s \leq S} |C(t, s)| \rightarrow 0$.

Step 3: zero plateau. Suppose that $q = 0$. Fix $L \geq 0$ and let $t \geq L$. For $0 \leq s \leq t - L$, the vector v_s lies in $\mathcal{H}_s \subseteq \mathcal{H}_{t-L}$ by Proposition 7.1, so $P_{t-L}v_s = v_s$. As P_{t-L} is an orthogonal projection, we have

$$C(t, s) = \langle v_t, P_{t-L}v_s \rangle = \langle P_{t-L}v_t, v_s \rangle = \langle M_{t-L}(L), v_s \rangle.$$

By the Cauchy–Schwarz inequality and $\|v_s\| = 1$, we obtain

$$\sup_{0 \leq s \leq t-L} C(t, s)^2 \leq \|M_{t-L}(L)\|^2.$$

We now let $t \rightarrow \infty$ with L fixed. By Proposition 8.4, the translated kernels converge to the stationary cluster (c, r, ℓ_β) along every sequence $T_n \rightarrow \infty$. Then Proposition 7.5 applies

along every such sequence, and (7.17) gives $\|M_{T_n}(L)\|^2 \rightarrow c(2L)$. As the sequence is arbitrary, we have $\|M_T(L)\|^2 \rightarrow c(2L)$ as $T \rightarrow \infty$. Setting $T = t - L$, we obtain

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq t-L} C(t, s)^2 \leq c(2L). \quad (8.14)$$

This bound holds for every plateau, but it is useful only when $q = 0$. Fix $0 \leq S < \infty$. For $t \geq S + L$, we have $[0, S] \subseteq [0, t - L]$, and (8.14) implies $\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq S} C(t, s)^2 \leq c(2L)$ for every $L \geq 0$. Letting $L \rightarrow \infty$ and using $c(2L) \rightarrow q = 0$, we find that $\sup_{0 \leq s \leq S} C(t, s) \rightarrow 0$. Together with Steps 1 and 2, this proves (8.12).

Step 4: the whole row when $q = 0$. Fix $L > 0$, and let $t \geq L$. We split $[0, t]$ into the recent strip $[t - L, t]$ and the remote part $[0, t - L]$. On the recent strip, we use (8.7) with $T = t - L$ and $A = L$. Writing $s = T + u$ with $u \in [0, L]$, we have $C(t, s) = C(T + L, T + u)$ and $c(t - s) = c(L - u)$, and similarly for R and r . We obtain

$$\sup_{t-L \leq s \leq t} \{|C(t, s) - c(t - s)| + |R(t, s) - r(t - s)|\} \rightarrow 0 \quad (t \rightarrow \infty).$$

On the remote part, we have $t - s \geq L$. Recall that $0 \leq R \leq C$, that $0 \leq r \leq c$, and that c is nonincreasing. For $0 \leq s \leq t - L$, we find

$$\begin{aligned} |C(t, s) - c(t - s)| + |R(t, s) - r(t - s)| &\leq C(t, s) + c(t - s) + R(t, s) + r(t - s) \\ &\leq 2C(t, s) + 2c(L). \end{aligned}$$

By (8.14), the upper limit as $t \rightarrow \infty$ of the supremum of the right side over $0 \leq s \leq t - L$ is at most $2c(2L)^{1/2} + 2c(L)$. Combining the two parts, we have

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq t} \{|C(t, s) - c(t - s)| + |R(t, s) - r(t - s)|\} \leq 2c(2L)^{1/2} + 2c(L),$$

where $t \rightarrow \infty$ is taken with L fixed. Since $q = 0$, the right side tends to zero as $L \rightarrow \infty$. This proves (8.13). $\square$

8.6 Correlation aging

Corollary 8.9 (Correlation aging in the ordered limits). *For every $p \geq 3$ and $\beta \geq \beta_d$,*

$$\lim_{\tau \rightarrow \infty} \lim_{s \rightarrow \infty} C(s + \tau, s) = q_\beta > 0, \quad \lim_{s \rightarrow \infty} \lim_{\tau \rightarrow \infty} C(s + \tau, s) = 0. \quad (8.15)$$

Proof. Fix $\tau \geq 0$. Applying (8.11) with $A = \tau$ and $T = s$ to the pair of offsets $(\tau, 0)$, we find that $\lim_{s \rightarrow \infty} C(s + \tau, s) = c(\tau)$. As $\tau \rightarrow \infty$, we have $c(\tau) \rightarrow q_\beta$ by Theorem 8.5, and $q_\beta \geq q_d > 0$. This gives the first limit. Next, fix $s \geq 0$. Applying (8.12) with $S = s$ and $t = s + \tau$, we find that $C(s + \tau, s) \rightarrow 0$ as $\tau \rightarrow \infty$. The inner limit is zero for every s , and so is the outer limit. $\square$

The first inner limit in (8.15) is the stationary correlation $c(\tau)$, and the second inner limit vanishes at every fixed waiting time. Corollary 8.9 concerns only these two ordered limits. It does not identify the time scale on which $C(s + \tau, s)$ decreases below q_β , and it does not give a fluctuation–dissipation relation in the slow regime. We discuss these questions in Section 17.

Proof of Theorem 1.1. We collect the results of this section, which rely on Proposition 8.1, proved in Sections 9 and 10. Fix $p \geq 3$ and $\beta > 0$. By Proposition 8.4, the multiplier and the energy converge to ℓ_β and $e_\beta(\infty)$, and the translated kernels converge locally uniformly to the stationary profile $(c_\beta, r_\beta) = (c, r)$. This is (1.1). In Step 3 of the proof of Proposition 8.4, we showed that every stationary cluster equals (c, r, ℓ_β) , so the profile is unique. By Proposition 5.7, we have $c(0) = r(0) = 1$ and $r = -c'$, and c is completely monotone by Corollary 7.7. The forgetting statement (1.2) is (8.12). For $\beta \geq \beta_d$, Theorem 8.5 shows that the plateau is $q_\beta \geq q_d$, which is the larger solution of (1.3), and it gives (1.4) and (1.5). Finally, (1.6) is (8.15). This completes the proof of Theorem 1.1.

Tree ceilings and the threshold inputs for cubic and quartic models

In this section, we prove Proposition 8.1 for $p \in \{3, 4\}$. That proposition is stated in Section 8.1, and Section 8 deduces from it the selection of the plateau, forgetting, and correlation aging. We prove the proposition for every $p \geq 3$ in Section 10.

Both parts rest on a deterministic inequality in the tree space of Section 7.1. For $p = 3$, the cubic tree ceiling, Theorem 9.2, states that every unit tree vector has energy strictly below E_{th} , and that a unit vector with energy close to E_{th} has little mass at every fixed finite set of heights. Lemma 9.3 is the analogous statement for $p = 4$, and Corollary 9.4 restates both in terms of the tree energy. In Section 9.3, we prove part (a) of Proposition 8.1 by applying a ceiling to normalized prefix centers and passing to the limit with Proposition 7.5. Under the hypotheses of part (b), the normalized centers approach the threshold; this is Proposition 9.5. Their mass then leaves every fixed set of heights, while an earlier state is approximated by finitely many heights, uniformly on a bounded time interval. Combining these two facts, we prove part (b) in Section 9.5.

Besides the tree space and the realization of Proposition 7.1, we use the reflection identity of Proposition 7.5, the precompactness in Proposition 5.7, the plateau identity (6.2), and the limiting center energy of Lemma 8.2. The last lemma is an algebraic statement about a stationary cluster and does not use Proposition 8.1. Hence the arguments of this section and of Section 8 are not circular. The ceilings are used again later: the participation bound enters the proof of Theorem 11.1, and the retained angular and Young defects (9.3) are used in Section 13. Proposition 9.5 is used in Sections 11 and 14.

The value $E_{\text{th}} = 2\sqrt{(p-1)/p}$ is the threshold energy of the pure p -spin model. It appears in the complexity of critical points computed by Auffinger, Ben Arous, and Černý [7, equation (2.13)], where the Hamiltonian is minimized and the threshold is $-E_{\text{th}}$. It is the threshold for Lipschitz algorithms identified by Huang and Sellke, for even models in [94] and including odd pure models in [92, Theorem 1 and Corollary 1.8], and Sellke proved that low-temperature Langevin dynamics approaches it when the temperature is also sent to zero [112, Theorems 1.1 and 1.7]. The ceilings below are elementary inequalities in the tree space, and we do not use the cited results in the proofs of this section. They also differ from the branching constructions of the algorithmic literature, such as those of Subag [118, Remark 6] and of Huang and Sellke [94, Remarks 1.2–1.3].

and Sections 3 and 7], which concern configurations in finite dimension. Here we retain quantitative height defects inside the causal covariance realization and use their saturation to prove escape from every fixed set of heights.

Standing notation. Throughout this section, $p \in \{3, 4\}$ and $m = p - 1$, and $\mathcal{H}$, $\mathcal{I}$, Π_j , and P_T are the tree space, the grafting isometry, and the height and prefix projections of Section 7.1 for this arity. For $v \in \mathcal{H}$ and $j \geq 0$, we write

$$b_j = \|\Pi_j v\|^2, \quad a_j = \sqrt{b_j}, \quad B_j = \sum_{i \leq j} b_i,$$

with $b_j = B_j = 0$ for $j < 0$. Recall that $\sum_j \Pi_j = \text{Id}$, so that the masses b_j sum to $\|v\|^2$ and B_j increases to $\|v\|^2$. The tree energy is $h(v) = \sqrt{p} \langle v, \mathcal{I}(v^{\otimes m}) \rangle$, as in (7.5). It is a homogeneous polynomial of degree p , that is, $h(\alpha v) = \alpha^p h(v)$ for $\alpha \geq 0$.

9.1 The cubic tree ceiling

We begin with a computation that holds for both arities. It expresses the height profile of a graft through the cumulative masses B_j .

Lemma 9.1 (Heights of a graft). *Let $v \in \mathcal{H}$. Then $\Pi_0 \mathcal{I}(v^{\otimes m}) = 0$, and for every $j \geq 1$,*

$$\Pi_j \mathcal{I}(v^{\otimes m}) = \Pi_j \mathcal{I}((\Pi_{\leq j-1} v)^{\otimes m}), \quad \|\Pi_j \mathcal{I}(v^{\otimes m})\|^2 = B_{j-1}^m - B_{j-2}^m. \quad (9.1)$$

Proof. The ranges of the projections Π_i , $i \geq 0$, are mutually orthogonal and span $\mathcal{H}$. It follows that $\mathcal{H}^{\otimes m}$ is the orthogonal sum of the subspaces $\Pi_{i_1} \mathcal{H} \otimes \cdots \otimes \Pi_{i_m} \mathcal{H}$ over $(i_1, \dots, i_m)$. Because a graft has height one plus the largest height of its root subtrees, the map $\mathcal{I}$ sends the subspace indexed by $(i_1, \dots, i_m)$ into the range of Π_j with $j = 1 + \max_k i_k$. In particular, $\Pi_0 \mathcal{I} = 0$, and for $j \geq 1$,

$$\Pi_j \mathcal{I}(v^{\otimes m}) = \sum_{\max_k i_k = j-1} \mathcal{I}(\Pi_{i_1} v \otimes \cdots \otimes \Pi_{i_m} v).$$

Only the components $\Pi_i v$ with $i \leq j-1$ appear, which gives the first identity in (9.1). The terms of the sum are orthogonal, since $\mathcal{I}$ is an isometry and the subspaces are orthogonal, and the term indexed by $(i_1, \dots, i_m)$ has squared norm $b_{i_1} \cdots b_{i_m}$. Summing over the m -tuples with largest entry $j-1$, we obtain

$$\|\Pi_j \mathcal{I}(v^{\otimes m})\|^2 = \left(\sum_{i \leq j-1} b_i \right)^m - \left(\sum_{i \leq j-2} b_i \right)^m = B_{j-1}^m - B_{j-2}^m,$$

as desired. □

We now treat the cubic case. The next estimate is the reason that a normalized prefix center selects marginality. It is an inequality in the tree space, and it involves no finite-dimensional critical points.

Theorem 9.2 (The cubic tree ceiling and its retained defects). *Let $p = 3$, so that grafting is binary, and let $F_* = 2\sqrt{2}/3$. For a unit vector $v \in \mathcal{H}$, define*

$$f(v) = \langle v, \mathcal{I}(v \otimes v) \rangle, \quad \delta = F_* - f(v),$$

and let $b_j = \|\Pi_j v\|^2$, $a_j = \sqrt{b_j}$, and $B_j = \sum_{i \leq j} b_i$ for $j \geq 0$, with $B_{-1} = B_{-2} = 0$. Then $|f(v)| < F_$, and*

$$\sum_{j \geq 0} b_j^2 \leq 4\sqrt{2} \delta, \quad \|\Pi_{\leq H} v\|^2 \leq \sqrt{4\sqrt{2}(H+1)} \delta \quad (H \geq 0). \quad (9.2)$$

The supremum of f over unit vectors equals F_ , and it is not attained. Further, let $s_j = \sqrt{B_{j-1} + B_{j-2}}$, and for $j \geq 1$ let $\theta_j \in [0, \pi]$ be the angle between $\Pi_j v$ and $\Pi_j \mathcal{I}(v \otimes v)$, with $\theta_j = 0$ when one of these two vectors is zero. Then*

$$\sum_{j \geq 1} a_j a_{j-1} s_j (1 - \cos \theta_j) \leq \delta, \quad \frac{1}{2} \sum_{j \geq 1} s_j (a_j - a_{j-1})^2 \leq \delta. \quad (9.3)$$

Proof. The proof has three steps. First, we write $f(v)$ as a sum over heights and bound it by two sums S and Y , keeping track of the gaps. Second, we compare Y with the integral $\int_0^1 \sqrt{2y} \, dy = F_*$, again keeping the gap. Third, we construct unit vectors whose values of f approach F_* .

Step 1: heightwise Cauchy–Schwarz and Young inequalities. By Lemma 9.1 with $m = 2$, we have, for $j \geq 1$,

$$\|\Pi_j \mathcal{I}(v \otimes v)\|^2 = B_{j-1}^2 - B_{j-2}^2 = b_{j-1}(B_{j-1} + B_{j-2}) = a_{j-1}^2 s_j^2.$$

Using that the height projections are orthogonal and that $\Pi_0 \mathcal{I} = 0$, we find that

$$f(v) = \sum_{j \geq 1} \langle \Pi_j v, \Pi_j \mathcal{I}(v \otimes v) \rangle = \sum_{j \geq 1} a_j a_{j-1} s_j \cos \theta_j. \quad (9.4)$$

When one of the two vectors in the j -th inner product is zero, the coefficient $a_j a_{j-1} s_j$, which is the product of their norms, vanishes. The convention on θ_j therefore does not matter. We define

$$S = \sum_{j \geq 1} a_j a_{j-1} s_j, \quad Y = \frac{1}{2} \sum_{j \geq 1} (b_j + b_{j-1}) s_j.$$

All these series converge absolutely, because $s_j \leq \sqrt{2}$ and $\sum_j b_j = 1$. Using (9.4) and the inequality $a_j a_{j-1} \leq \frac{1}{2}(b_j + b_{j-1})$, we see that $|f(v)| \leq S \leq Y$, with the exact gaps

$$S - f(v) = \sum_{j \geq 1} a_j a_{j-1} s_j (1 - \cos \theta_j), \quad Y - S = \frac{1}{2} \sum_{j \geq 1} s_j (a_j - a_{j-1})^2. \quad (9.5)$$

Step 2: comparison with an integral. Let $g(y) = \sqrt{2y}$, a concave function on $[0, 1]$ with $\int_0^1 g(y) dy = F_*$. Reindexing the second half of Y and using $s_1 = \sqrt{b_0}$, we have

$$Y = \frac{1}{2}b_0^{3/2} + \frac{1}{2} \sum_{j \geq 1} b_j(s_j + s_{j+1}).$$

Since B_j increases to one, we may split $\int_0^1 g$ at the points $B_0 \leq B_1 \leq \dots$. Using $\int_0^{B_0} g = F_* b_0^{3/2}$, we find that

$$F_* - Y = \left(F_* - \frac{1}{2}\right)b_0^{3/2} + \sum_{j \geq 1} \left[\int_{B_{j-1}}^{B_j} g(y) dy - \frac{b_j}{2}(s_j + s_{j+1}) \right]. \quad (9.6)$$

We bound each bracket from below. Because g is concave, the trapezoid rule underestimates its integral, that is,

$$\int_{B_{j-1}}^{B_j} g(y) dy \geq \frac{b_j}{2} \{g(B_{j-1}) + g(B_j)\}.$$

We have $s_j \leq g(B_{j-1})$, because $B_{j-2} \leq B_{j-1}$. If $b_j > 0$, then

$$g(B_j) - s_{j+1} = \sqrt{2B_j} - \sqrt{B_j + B_{j-1}} = \frac{b_j}{\sqrt{2B_j} + \sqrt{B_j + B_{j-1}}} \geq \frac{b_j}{2\sqrt{2}},$$

because the denominator is at most $2\sqrt{2}$. Inserting these bounds, we see that the bracket with index j is at least $\frac{b_j}{2} \cdot \frac{b_j}{2\sqrt{2}} = b_j^2 / (4\sqrt{2})$, and this also holds when $b_j = 0$. For the first term, we use $b_0^{3/2} \geq b_0^2$ and

$$F_* - \frac{1}{2} - \frac{1}{4\sqrt{2}} = \frac{13\sqrt{2} - 12}{24} > 0.$$

Combining these bounds with (9.6), we obtain

$$F_* - Y \geq \frac{1}{4\sqrt{2}} \sum_{j \geq 0} b_j^2. \quad (9.7)$$

We now collect the gaps. We have the exact decomposition

$$\delta = F_* - f(v) = (F_* - Y) + (Y - S) + (S - f(v)),$$

in which all three terms are nonnegative. By (9.7), the first term is at least $(4\sqrt{2})^{-1} \sum_j b_j^2$, which yields the first inequality in (9.2). The second and third terms are the two sums in (9.3) by (9.5), which proves (9.3). Using the Cauchy–Schwarz inequality, we also have

$$\|\Pi_{\leq H} v\|^2 = \sum_{j=0}^H b_j \leq \left((H+1) \sum_{j \geq 0} b_j^2 \right)^{1/2} \leq \sqrt{4\sqrt{2}(H+1)} \delta,$$

which is the second inequality in (9.2). Finally, some b_j is positive, because $\sum_j b_j = 1$, and we have $|f(v)| \leq Y \leq F_* - (4\sqrt{2})^{-1} \sum_j b_j^2 < F_*$. In particular, no unit vector attains the value F_* .

Step 3: sharpness. We construct unit vectors $v^{(n)}$ with $f(v^{(n)}) \rightarrow F_*$. We give the construction for a general arity $m \in \{2, 3\}$, since the proof of Lemma 9.3 uses it with $m = 3$. Fix an integer $n \geq 1$ and let $\rho = (n + 1)^{-1}$. Let $w_0 = \sqrt{\rho} e_0$, a vector in the one-leaf summand. For $j = 1, \dots, n$, we define inductively

$$\gamma_j = \Pi_j \mathcal{I}(w_{j-1}^{\otimes m}), \quad u_j = \sqrt{\rho} \frac{\gamma_j}{\|\gamma_j\|}, \quad w_j = w_{j-1} + u_j.$$

We check by induction that this is well defined, and that w_{j-1} has mass ρ at each height $0, \dots, j-1$ and no mass at larger heights. This holds for w_0 . If it holds for w_{j-1} , then Lemma 9.1 applied to w_{j-1} gives

$$\|\gamma_j\|^2 = (j\rho)^m - ((j-1)\rho)^m > 0,$$

and u_j is well defined. The vector u_j lies in the range of Π_j and has squared norm ρ , and therefore w_j has mass ρ at each height $0, \dots, j$ and no mass at larger heights. Let $v^{(n)} = w_n$. It is a unit vector, with $b_j = \rho$ for $0 \leq j \leq n$ and $b_j = 0$ for $j > n$.

We now show that all the angles vanish. Let $1 \leq j \leq n$. Since $\Pi_{\leq j-1} v^{(n)} = w_{j-1}$, Lemma 9.1 implies that $\Pi_j \mathcal{I}((v^{(n)})^{\otimes m}) = \gamma_j$. The component $\Pi_j v^{(n)} = u_j$ is a positive multiple of γ_j , which means that $\theta_j = 0$. For $j > n$, the component $\Pi_j v^{(n)}$ vanishes. Only the heights $1, \dots, n$ contribute to f , and we find that

$$f(v^{(n)}) = \sum_{j=1}^n \langle u_j, \gamma_j \rangle = \sum_{j=1}^n \sqrt{\rho} \|\gamma_j\| = \sum_{j=1}^n \sqrt{\rho} \{(j\rho)^m - ((j-1)\rho)^m\}^{1/2}. \quad (9.8)$$

For $m = 2$, each summand equals $\rho \sqrt{(2j-1)\rho} = \rho g((j-\frac{1}{2})\rho)$. In other words, $f(v^{(n)})$ is the midpoint Riemann sum for $\int g$ over the first n of the $n+1$ intervals $[(j-1)\rho, j\rho]$ that partition $[0, 1]$. Since g is continuous on $[0, 1]$ and bounded by $\sqrt{2}$, the omitted last interval contributes at most $\sqrt{2}\rho$, and we obtain

$$\lim_{n \rightarrow \infty} f(v^{(n)}) = \int_0^1 \sqrt{2y} \, dy = F_*.$$

Together with Step 2, this shows that the supremum of f equals F_* and is not attained. $\square$

9.2 The quartic tree ceiling

For $p = 4$, grafting is ternary. The proof follows the cubic one, with a difference of cubes of the cumulative masses in place of a difference of squares. The comparison function becomes linear, and the ceiling calculation telescopes.

Lemma 9.3 (The quartic tree ceiling). *Let $p = 4$, so that grafting has arity three. For a unit vector $v \in \mathcal{H}$, define $f(v) = \langle v, \mathcal{I}(v^{\otimes 3}) \rangle$. Then*

$$|f(v)| < \frac{\sqrt{3}}{2}, \quad \sum_{j \geq 0} \|\Pi_j v\|^4 \leq 2\sqrt{3} \left(\frac{\sqrt{3}}{2} - f(v) \right). \quad (9.9)$$

The supremum of f over unit vectors equals $\sqrt{3}/2$, and it is not attained. In particular, let $h(v) = 2f(v)$ be the tree energy and $E_{\text{th}} = \sqrt{3}$. If (v_n) are unit vectors with $h(v_n) \rightarrow E_{\text{th}}$, then $\|\Pi_{\leq H} v_n\| \rightarrow 0$ for every $H \geq 0$.

Proof. The proof has three steps, parallel to those of Theorem 9.2: a heightwise bound, a comparison that keeps the gap, and the sharpness construction.

Step 1: heightwise bound. In this proof, we let

$$s_j = \sqrt{B_{j-1}^2 + B_{j-1}B_{j-2} + B_{j-2}^2}, \quad j \geq 1.$$

By Lemma 9.1 with $m = 3$, we have

$$\|\Pi_j \mathcal{I}(v^{\otimes 3})\|^2 = B_{j-1}^3 - B_{j-2}^3 = b_{j-1}s_j^2 = a_{j-1}^2 s_j^2.$$

As in Step 1 of the proof of Theorem 9.2, using the Cauchy–Schwarz inequality at each height and the inequality $a_j a_{j-1} \leq \frac{1}{2}(b_j + b_{j-1})$, we have

$$|f(v)| \leq S \leq Y, \quad S = \sum_{j \geq 1} a_j a_{j-1} s_j, \quad Y = \frac{1}{2} \sum_{j \geq 1} (b_j + b_{j-1}) s_j.$$

Reindexing the second half of Y and using $s_1 = B_0 = b_0$, we have

$$Y = \frac{1}{2} b_0^2 + \frac{1}{2} \sum_{j \geq 1} b_j (s_j + s_{j+1}).$$

Step 2: the comparison. Note that $b_j(B_j + B_{j-1}) = B_j^2 - B_{j-1}^2$. As B_j increases to one, the sum $\sum_{j \geq 0} b_j(B_j + B_{j-1})$ telescopes to one, and we can write

$$\frac{\sqrt{3}}{2} = \frac{\sqrt{3}}{2} b_0^2 + \frac{1}{2} \sum_{j \geq 1} b_j (\sqrt{3} B_{j-1} + \sqrt{3} B_j),$$

and subtracting the expression for Y , we arrive at

$$\frac{\sqrt{3}}{2} - Y = \frac{\sqrt{3}-1}{2} b_0^2 + \frac{1}{2} \sum_{j \geq 1} b_j \left[(\sqrt{3} B_{j-1} - s_j) + (\sqrt{3} B_j - s_{j+1}) \right]. \quad (9.10)$$

We have $s_j \leq \sqrt{3}B_{j-1}$, because $B_{j-2} \leq B_{j-1}$. To bound the second difference, let $X \geq X' \geq 0$. Then

$$\frac{(2X + X')^2}{3} - (X^2 + XX' + X'^2) = \frac{(X - X')(X + 2X')}{3} \geq 0.$$

This implies that $\sqrt{X^2 + XX' + X'^2} \leq (2X + X')/\sqrt{3}$, and therefore

$$\sqrt{3}X - \sqrt{X^2 + XX' + X'^2} \geq \sqrt{3}X - \frac{2X + X'}{\sqrt{3}} = \frac{X - X'}{\sqrt{3}}.$$

Taking $X = B_j$ and $X' = B_{j-1}$, we find that $\sqrt{3}B_j - s_{j+1} \geq b_j/\sqrt{3}$, and the summand with index $j \geq 1$ in (9.10) is at least $b_j^2/(2\sqrt{3})$. The first term is also at least $b_0^2/(2\sqrt{3})$, because

$$\frac{\sqrt{3}-1}{2} - \frac{1}{2\sqrt{3}} = \frac{2-\sqrt{3}}{2\sqrt{3}} > 0.$$

We conclude that

$$\frac{\sqrt{3}}{2} - Y \geq \frac{1}{2\sqrt{3}} \sum_{j \geq 0} b_j^2.$$

Since $f(v) \leq Y$, this yields the second inequality in (9.9). Further, $|f(v)| \leq Y$ and $\sum_j b_j^2 > 0$ for a unit vector, which gives $|f(v)| < \sqrt{3}/2$. This also shows that the value $\sqrt{3}/2$ is not attained.

Step 3: sharpness and finite heights. Let $v^{(n)}$ be the unit vectors constructed in Step 3 of the proof of Theorem 9.2, now with $m = 3$, and let $\rho = (n+1)^{-1}$. Inserting $m = 3$ into (9.8), we obtain

$$f(v^{(n)}) = \sum_{j=1}^n \sqrt{\rho} \{ \rho^3 (j^3 - (j-1)^3) \}^{1/2} = \rho^2 \sum_{j=1}^n \sqrt{3j^2 - 3j + 1}.$$

For $j \geq 1$, we have $3(j-1)^2 \leq 3j^2 - 3j + 1 \leq 3j^2$, so that

$$\frac{\sqrt{3}}{2} \cdot \frac{n(n-1)}{(n+1)^2} \leq f(v^{(n)}) \leq \frac{\sqrt{3}}{2} \cdot \frac{n}{n+1}.$$

Both bounds tend to $\sqrt{3}/2$, which proves that the supremum of f equals $\sqrt{3}/2$.

Finally, since $h = 2f$, the second inequality in (9.9) reads $\sum_j b_j^2 \leq \sqrt{3}\{E_{\text{th}} - h(v)\}$. For $H \geq 0$, the Cauchy-Schwarz inequality gives

$$\|\Pi_{\leq H} v\|^2 = \sum_{j=0}^H b_j^2 \leq \left((H+1) \sum_{j \geq 0} b_j^2 \right)^{1/2} \leq \{ \sqrt{3} (H+1) (E_{\text{th}} - h(v)) \}^{1/2},$$

and the right side tends to zero along any sequence of unit vectors with $h(v_n) \rightarrow E_{\text{th}}$. $\square$

For later use, we restate the two ceilings in terms of the tree energy h and the threshold E_{th} .

Corollary 9.4 (The ceilings in terms of the tree energy). *Let $p \in \{3, 4\}$, and let $\kappa_3 = 4\sqrt{2/3}$ and $\kappa_4 = \sqrt{3}$. Every unit vector $v \in \mathcal{H}$ satisfies $|h(v)| < E_{\text{th}}$ and*

$$\sum_{j \geq 0} \|\Pi_j v\|^4 \leq \kappa_p (E_{\text{th}} - h(v)), \quad \|\Pi_{\leq H} v\|^2 \leq \{\kappa_p (H + 1) (E_{\text{th}} - h(v))\}^{1/2} \quad (H \geq 0). \quad (9.11)$$

The supremum of h over unit vectors equals E_{th} .

Proof. For $p = 3$, we have $h = \sqrt{3}f$ and $E_{\text{th}} = 2\sqrt{2/3} = \sqrt{3}F_*$, so that $F_* - f(v) = \{E_{\text{th}} - h(v)\}/\sqrt{3}$. Then Theorem 9.2 implies that $|h(v)| < E_{\text{th}}$ and that $\sum_j b_j^2 \leq (4\sqrt{2}/\sqrt{3})\{E_{\text{th}} - h(v)\} = \kappa_3\{E_{\text{th}} - h(v)\}$. For $p = 4$, we have $h = 2f$ and $E_{\text{th}} = \sqrt{3}$, so that $\sqrt{3}/2 - f(v) = \{E_{\text{th}} - h(v)\}/2$. Then Lemma 9.3 implies that $|h(v)| < E_{\text{th}}$ and that $\sum_j b_j^2 \leq \sqrt{3}\{E_{\text{th}} - h(v)\} = \kappa_4\{E_{\text{th}} - h(v)\}$. In both cases, the second inequality in (9.11) follows from the first by the Cauchy–Schwarz inequality, as in the two proofs above, and the supremum statement follows from those of Theorem 9.2 and Lemma 9.3. $\square$

9.3 The energy ceiling for prefix centers

We now prove the first threshold input. Recall the prefix centers $M_T(a) = P_T v_{T+a}$ of Section 7.4.

Proof of part (a) of Proposition 8.1 for $p \in \{3, 4\}$. Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to a stationary cluster (c, r, ℓ) with plateau $q > 0$, and fix $L \geq 0$. We abbreviate $M_n = M_{T_n}(L)$. By (7.17), we have $\|M_n\|^2 \rightarrow d_L = c(2L)$, and $d_L \geq q > 0$ because c is nonincreasing with limit q . Then $M_n \neq 0$ for all large n , and $U_n = M_n/\|M_n\|$ is a unit vector. Corollary 9.4 implies that $h(U_n) < E_{\text{th}}$. Since h is homogeneous of degree p , we have

$$h(M_n) = \|M_n\|^p h(U_n) < E_{\text{th}} \|M_n\|^p.$$

Letting $n \rightarrow \infty$ and using (7.18) on the left side and (7.17) on the right side, we obtain $h_L \leq E_{\text{th}} d_L^{p/2}$, which is (8.1). $\square$

The proof uses only the limits of Proposition 7.5 along the given sequence. In particular, it uses neither marginality nor convergence of the whole translated family, which Section 8 deduces from part (a).

9.4 Threshold-saturating prefix centers

Under the hypotheses of part (b) of Proposition 8.1, the upper bound of part (a) becomes an equality in the limit of large offsets. The next proposition states this and converts it into a bound on the mass of the centers at low heights. It is the input to the proof of part (b) below, and we use it again in Sections 11 and 14.

Suppose that the translated kernels converge along every sequence $T \rightarrow \infty$ to one stationary profile (c, r) with plateau q . Every stationary cluster is then of the form (c, r, ℓ) , and (6.2) determines its multiplier as $\ell = (1 - q)^{-1} + \beta^2 p(1 - q^{p-1})$. The quantities d_L and h_L in (7.17) and (7.18) then do not depend on the sequence, and the energy $e_\infty = (\ell - 1)/(p\beta)$ in (7.18) is the limit of the energy $e_\beta(t)$; see Step 1 of the proof below.

Proposition 9.5 (Threshold-saturating prefix centers). *Let $p \in \{3, 4\}$ and $\beta > 0$. Suppose that the translated kernels converge along every sequence $T \rightarrow \infty$ to one stationary profile (c, r) , and that its plateau q is positive and satisfies $\beta^2 p(p - 1)q^{p-2}(1 - q)^2 = 1$. Let d_L and h_L be defined by (7.17) and (7.18) for this profile. Then*

$$\delta_L = E_{\text{th}} - \frac{h_L}{d_L^{p/2}} \geq 0 \quad (L \geq 0), \quad \lim_{L \rightarrow \infty} \delta_L = 0. \quad (9.12)$$

Moreover, for every $L \geq 0$ and every integer $H \geq 0$,

$$\limsup_{t \rightarrow \infty} \|\Pi_{\leq H} P_{t-L} v_t\|^2 \leq d_L \{\kappa_p(H + 1)\delta_L\}^{1/2}, \quad (9.13)$$

where $\kappa_3 = 4\sqrt{2/3}$ and $\kappa_4 = \sqrt{3}$ are the constants of Corollary 9.4.

The normalized prefix center approaches the threshold; the proposition makes no threshold assertion about the state v_t itself.

Proof. The proof has three steps. We first show that the limits of Proposition 7.5 hold as $T \rightarrow \infty$ without passing to a sequence. Then we prove (9.12) from part (a) of Proposition 8.1 and the identity (8.5). Finally, we apply Corollary 9.4 to normalized centers.

Step 1: reflection along every sequence. Let $T_n \rightarrow \infty$ be arbitrary. Using the precompactness in Proposition 5.7, we see that every subsequence of (T_n) has a further subsequence along which $(C_{T_n}, R_{T_n}, \lambda_{T_n})$ converges to a stationary cluster. By hypothesis, this cluster is (c, r, ℓ) , with the multiplier ℓ given above. Along the further subsequence, Proposition 7.5 gives the limits (7.17) and (7.18), with values d_L and h_L that do not depend on the subsequence, and $\lambda(T_n) \rightarrow \ell$. Since every subsequence has a further subsequence with the same limits, the whole sequence has them. As (T_n) was arbitrary, we conclude that, for every $L \geq 0$,

$$\lim_{T \rightarrow \infty} \|M_T(L)\|^2 = d_L, \quad \lim_{T \rightarrow \infty} h(M_T(L)) = h_L, \quad (9.14)$$

and that $\lambda(T) \rightarrow \ell$, so that $e_\beta(T) \rightarrow e_\infty$ by (4.1).

Step 2: the deficits. Recall that c is nonincreasing with limit $q > 0$. Then $d_L \geq q > 0$, and δ_L is well defined. By part (a) of Proposition 8.1, proved in Section 9.3, we have $h_L \leq E_{\text{th}} d_L^{p/2}$, that is, $\delta_L \geq 0$. To prove that $\delta_L \rightarrow 0$, recall from Lemma 8.2 that $d_L \rightarrow q$ and $h_L \rightarrow h$ as $L \rightarrow \infty$, where h is given by (8.4), and that

$$\frac{h}{q^{p/2}} = \sqrt{\frac{p-1}{p}} \left(z + \frac{1}{z} \right), \quad z = \beta \sqrt{p(p-1)} q^{(p-2)/2} (1-q),$$

by (8.5). The hypothesis $\beta^2 p(p-1) q^{p-2} (1-q)^2 = 1$ says that $z^2 = 1$. Since $z > 0$, we obtain $z = 1$ and $h/q^{p/2} = 2\sqrt{(p-1)/p} = E_{\text{th}}$. We conclude that

$$\lim_{L \rightarrow \infty} \frac{h_L}{d_L^{p/2}} = \frac{h}{q^{p/2}} = E_{\text{th}},$$

which proves (9.12).

Step 3: the mass at low heights. Fix $L \geq 0$ and $H \geq 0$. By (9.14), we have $\|M_T(L)\|^2 \rightarrow d_L > 0$, and in particular $M_T(L) \neq 0$ for all large T . We let $U_T = M_T(L)/\|M_T(L)\|$. Using the homogeneity of h and (9.14), we find that

$$h(U_T) = \frac{h(M_T(L))}{\|M_T(L)\|^p} \longrightarrow \frac{h_L}{d_L^{p/2}} = E_{\text{th}} - \delta_L \quad (T \rightarrow \infty).$$

By Corollary 9.4, we have $\|\Pi_{\leq H} U_T\|^2 \leq \{\kappa_p(H+1)(E_{\text{th}} - h(U_T))\}^{1/2}$, and therefore

$$\limsup_{T \rightarrow \infty} \|\Pi_{\leq H} U_T\|^2 \leq \{\kappa_p(H+1)\delta_L\}^{1/2}.$$

Since $\|\Pi_{\leq H} M_T(L)\|^2 = \|M_T(L)\|^2 \|\Pi_{\leq H} U_T\|^2$ and $\|M_T(L)\|^2 \rightarrow d_L$, we see that

$$\limsup_{T \rightarrow \infty} \|\Pi_{\leq H} M_T(L)\|^2 \leq d_L \{\kappa_p(H+1)\delta_L\}^{1/2}.$$

Setting $T = t - L$, we have $M_{t-L}(L) = P_{t-L} v_t$, and $t \rightarrow \infty$ if and only if $T \rightarrow \infty$. This proves (9.13). $\square$

9.5 Forgetting at marginal saturation

We now prove the second threshold input. When $s \leq t - L$, a fixed earlier state v_s lies in the range of P_{t-L} , and its overlap with the late state v_t equals its overlap with the prefix center $P_{t-L} v_t$. Proposition 9.5 shows that the center has almost no mass at low heights once L is large, while the earlier states v_s , $s \leq S$, have almost all their mass at bounded heights. The overlap is therefore small.

Proof of part (b) of Proposition 8.1 for $p \in \{3, 4\}$. Suppose that the hypotheses of part (b) hold, and fix $0 \leq S < \infty$. The proof uses three parameters: the offset $L \geq 0$, the height cutoff $H \geq 0$, and the time t . We first let $t \rightarrow \infty$ with L and H fixed, then let $L \rightarrow \infty$, and finally let $H \rightarrow \infty$.

Let $L, H \geq 0$, let $t \geq S + L$, and let $0 \leq s \leq S$. Then $s \leq t - L$, and $v_s \in \mathcal{H}_s \subset \mathcal{H}_{t-L}$ by Proposition 7.1, that is, $P_{t-L}v_s = v_s$. As P_{t-L} is an orthogonal projection, we have

$$C(t, s) = \langle v_t, v_s \rangle = \langle v_t, P_{t-L}v_s \rangle = \langle P_{t-L}v_t, v_s \rangle.$$

We split $v_s = \Pi_{\leq H}v_s + (\text{Id} - \Pi_{\leq H})v_s$ and move the first projection to the other side of the inner product. Using $\|v_s\| = 1$ and $\|P_{t-L}v_t\| \leq 1$, we find

$$|C(t, s)| \leq \|\Pi_{\leq H}P_{t-L}v_t\| + \|(\text{Id} - \Pi_{\leq H})v_s\|. \quad (9.15)$$

We first show that the second term is small uniformly in $s \leq S$. Let

$$\varepsilon_H = \sup_{0 \leq s \leq S} \|(\text{Id} - \Pi_{\leq H})v_s\|.$$

The map $s \mapsto v_s$ is norm-continuous by Proposition 7.1, and the set $\{v_s : 0 \leq s \leq S\}$ is compact. Recall that $\sum_j \Pi_j = \text{Id}$ strongly, which means that $(\text{Id} - \Pi_{\leq H})x \rightarrow 0$ as $H \rightarrow \infty$ for every $x \in \mathcal{H}$. Let $\varepsilon > 0$. There exist $s_1, \dots, s_k \in [0, S]$ such that every v_s with $s \leq S$ lies within ε of some v_{s_i} , and there exists H_0 such that $\|(\text{Id} - \Pi_{\leq H})v_{s_i}\| < \varepsilon$ for all i and $H \geq H_0$. Since $\text{Id} - \Pi_{\leq H}$ is a contraction, this implies $\|(\text{Id} - \Pi_{\leq H})v_s\| < 2\varepsilon$ for all $s \leq S$ and $H \geq H_0$. As ε is arbitrary, we have shown that $\varepsilon_H \rightarrow 0$ as $H \rightarrow \infty$.

We now take the limits in the stated order. Taking the supremum over $s \leq S$ in (9.15) and letting $t \rightarrow \infty$, we obtain from (9.13) that

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq S} |C(t, s)| \leq d_L^{1/2} \{\kappa_p(H+1)\delta_L\}^{1/4} + \varepsilon_H.$$

The left side depends on neither L nor H . We let $L \rightarrow \infty$ with H fixed. Since $d_L \leq 1$ and $\delta_L \rightarrow 0$ by (9.12), the first term on the right tends to zero. Then we let $H \rightarrow \infty$, and $\varepsilon_H \rightarrow 0$. We conclude that $\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq S} |C(t, s)| = 0$, which is (8.2). $\square$

This completes the proof of Proposition 8.1 for $p \in \{3, 4\}$.

The threshold inputs for general pure models

In this section, we prove Proposition 8.1 for every $p \geq 3$. For $p \in \{3, 4\}$, Section 9 proves both parts of that proposition by a deterministic inequality for tree vectors. The argument given here does not use the tree ceilings and works for every pure model. Part (a) of Proposition 8.1 is the input to the proof of marginality in Lemma 8.3, and part (b) is the input to the proof of forgetting in Theorem 8.8. The finite forks of Section 10.2, the overlap penalty of Lemma 10.2, and the saturation identity (10.28) are used again in Section 11.5, where we extend the localization of the response to every $p \geq 3$.

We use the landscape bounds of Lemma 3.6, the fixed-horizon bridge of Proposition 3.3, and the Brownian form of the averaged response in Lemma 3.4. From Section 4, we use the energy identity of Proposition 4.1 and the bound (4.4) on the moments of the response, proved there. From Section 5, we use the entropy budget of Lemma 5.1, the integration by parts formula (5.8), the Fisher majorant of Theorem 5.3, and the stationary clusters of Proposition 5.7. We also use the formulas (7.17) and (7.18) for d_L and h_L . The proof of part (b) uses, in addition, the identities (8.4) and (8.5) and the correlation modulus (4.19).

Ideas of the proof. We obtain the energy ceiling in part (a) from the algorithmic threshold of Huang and Sellke. At a late time T , we continue the trajectory by K independent Brownian futures of length L , which we call a fork, and average their endpoints. The current identity of Section 5 shows that two branches of the fork have overlap close to $c(a + d)$, where a and d are their lengths. We compute the energy of the average exactly by Gaussian integration by parts. Subtracting the energy of a single branch removes the common past in the dimension limit. Using the hard–soft approximation of Sellke, we approximate the average by a function of the disorder that is Lipschitz with a constant independent of the dimension. The algorithmic threshold then bounds its energy. Letting $K \rightarrow \infty$, we obtain part (a).

For part (b), we add a deterministic field to the threshold. By convex duality, the field version of the threshold penalizes the overlap of a near-threshold algorithm with any direction that is independent of the disorder. However, at positive times, an earlier state X_s depends on the disorder. We therefore disclose the disorder in stages, along a finite sequence of gradient queries, and apply the penalty to one new direction at each stage. An Euler–Maruyama approximation places X_s close to the span of finitely many such

queries. Under marginality, the normalized average of a fork saturates the threshold, and the penalty forces its overlap with X_s to vanish. Up to normalization, this overlap is $C(t, s)$.

Attribution. The algorithmic threshold is [92, Theorem 1 and Corollary 1.8]; for even p it was proved earlier in [94]. In [112, Section 3], Sellke combined it with a Lipschitz approximation of Langevin dynamics on a fixed time interval to show that the energy of the dynamics stays below the threshold. We follow that argument in Section 10.2. Trajectories with a common past and independent future noise are a standard construction in the physics of aging. Cugliandolo and Dean considered replicas that share their noise up to a waiting time and receive independent noise afterward in [50, arXiv version 1, Section II]. Barrat, Burioni, and Mézard classified aging through the overlaps of such copies in [11, Sections 2 and 4]. Franz and Virasoro proposed averaging cloned futures to estimate a quasi-state magnetization in [72, arXiv version 3, Section III], and noise-averaged magnetizations are central to Biroli's dynamical TAP approach in [31, arXiv version 2, Section III.B]. Here we prove the overlap and energy identities of the forks from the current identity of Section 5 and from Gaussian integration by parts. The disclosure argument of Section 10.3 is a Gaussian conditioning argument; we credit its antecedents there.

Order of limits. Every horizon, number of branches, field strength, approximation accuracy, and Lipschitz constant below is fixed while the dimension tends to infinity, and we take the long-time and outer limits only afterward. We take dimension limits along the subsequence (N_j) of Theorem 5.3, and we often write $N \rightarrow \infty$ for $N = N_j$, $j \rightarrow \infty$. The parameters and their order are as follows:

$$N_j \rightarrow \infty, \quad n \rightarrow \infty, \quad \varepsilon \downarrow 0, \quad K \rightarrow \infty, \quad L \rightarrow \infty, \quad M \rightarrow \infty. \quad (10.1)$$

Here n indexes a sequence of late times (T_n in Section 10.2 and t_n in Section 10.4), ε is the accuracy of the Lipschitz approximation in Proposition 10.4, K is the number of branches of a fork, L is their length, and M is the number of steps of the Euler–Maruyama scheme of the proof of Proposition 10.8. Each parameter is fixed while the limits to its left are taken. Part (a) uses only the first four limits, at a fixed L ; part (b) uses all six. For the forgetting proof, it suffices to take the approximation accuracies $\varepsilon = 1/(m+1)$, $m \in \mathbb{N}$, so $\varepsilon \downarrow 0$ there means $m \rightarrow \infty$. We remove the tolerance and the field strengths of the threshold theorem inside the proof of Lemma 10.2, after the dimension limit at fixed values of all other parameters. Constants of the Lipschitz maps may depend on every fixed parameter, and they are used only before the limits in which that parameter varies.

10.1 The algorithmic threshold and its overlap penalty

Throughout this section, $\|\cdot\|$ is the Euclidean norm,

$$B_N = \{x \in \mathbb{R}^N : \|x\| \leq \sqrt{N}\},$$

and Π_{B_N} is the metric projection onto B_N , which is 1-Lipschitz. The coefficient tensor J of (3.1) lies in the N^p -dimensional space $(\mathbb{R}^N)^{\otimes p}$, with its Euclidean norm, and has the standard Gaussian law there. More explicitly, we keep the ordered coefficients and use the Hilbert norm

$$\|J\|_{\text{coef}}^2 = \sum_{i_1, \dots, i_p=1}^N J_{i_1 \dots i_p}^2.$$

This norm is not rescaled with the dimension. When the dependence on J matters, we write $H_{N,J}$ for H_N . An algorithm is a measurable map $\mathcal{A}_N = \mathcal{A}_N(J, \zeta)$ with values in B_N , where the seed ζ is a random variable independent of J (possibly absent). For $0 \leq L_{\text{Lip}} < \infty$, we say that $\mathcal{A}_N$ is L_{Lip} -Lipschitz if, for almost every value of ζ ,

$$\|\mathcal{A}_N(J, \zeta) - \mathcal{A}_N(J', \zeta)\| \leq L_{\text{Lip}} \|J - J'\|_{\text{coef}} \quad \text{for all } J, J' \in (\mathbb{R}^N)^{\otimes p}.$$

The constant L_{Lip} is deterministic and does not depend on N or on the seed.

Recall the tensor norm Q_N of (3.25) and the constants $A_p, c_p > 0$ of Lemma 3.6, for which $\mathbb{P}(Q_N > A_p) \leq e^{-c_p N}$. We write

$$Q_* = \sup_N (\mathbb{E} Q_N^2)^{1/2},$$

which is finite by Lemma 3.6. By the definition of Q_N , a contraction of J against vectors $u_1, \dots, u_p$ is at most $\sqrt{N} Q_N \|u_1\| \cdots \|u_p\|$ in absolute value. A derivative of order $k \leq 2$ of H_N at x , applied to k vectors, is $N^{-(p-1)/2}$ times a sum of $p(p-1) \cdots (p-k+1)$ such contractions in which $p-k$ slots equal x . For $R \geq 1$ and $\|x\| \leq R\sqrt{N}$, this implies that

$$\begin{aligned} \frac{|H_N(x)|}{N} &\leq R^p Q_N, & \|\nabla H_N(x)\| &\leq p R^{p-1} Q_N \sqrt{N}, \\ \|\nabla^2 H_N(x)\|_{\text{op}} &\leq p(p-1) R^{p-2} Q_N. \end{aligned} \tag{10.2}$$

For $R = 1$, the first two bounds read

$$\sup_{x \in B_N} \frac{|H_N(x)|}{N} \leq Q_N, \quad \sup_{x \in B_N} \|\nabla H_N(x)\| \leq p Q_N \sqrt{N}. \tag{10.3}$$

In particular, the normalized energy of any algorithm is bounded by Q_N , whose moments are bounded uniformly in N . Now let x and y be random vectors of norm at most $R\sqrt{N}$.

The segment between them lies in the same ball. Using the mean value theorem, (10.2), and the Cauchy–Schwarz inequality, we find

$$\frac{\mathbb{E}|H_N(x) - H_N(y)|}{N} \leq pR^{p-1}Q_* \left(\mathbb{E} \frac{\|x - y\|^2}{N} \right)^{1/2}. \quad (10.4)$$

For $\kappa \geq 0$, let $q_1(\kappa) = (\kappa^2 / (p(p-2)))^{1/(p-1)}$, and define

$$\text{ALG}_p(\kappa) = \begin{cases} E_{\text{th}} + \frac{p-2}{p} \sqrt{p(p-1)} q_1(\kappa)^{p/2}, & \kappa^2 \leq p(p-2), \\ \sqrt{p + \kappa^2}, & \kappa^2 \geq p(p-2). \end{cases} \quad (10.5)$$

In terms of κ , the first line equals $E_{\text{th}} + \gamma_p \kappa^{p/(p-1)}$ with $\gamma_p = ((p-2)/p) \sqrt{p(p-1)} [p(p-2)]^{-p/(2(p-1))}$.

Theorem 10.1 (External-field Lipschitz threshold of Huang and Sellke). *Fix an integer $p \geq 3$, $0 \leq L_{\text{Lip}} < \infty$, $\kappa \geq 0$, and $\eta > 0$. There exist $c > 0$ and N_0 , depending only on p , L_{Lip} , κ , and η , such that the following holds for every $N \geq N_0$. Let $w \in \mathbb{R}^N$ be deterministic with $\|w\| = \sqrt{N}$, and let $\mathcal{A}_N : (\mathbb{R}^N)^{\otimes p} \rightarrow B_N$ be L_{Lip} -Lipschitz. Then*

$$\mathbb{P} \left(\frac{H_N(\mathcal{A}_N(J)) + \kappa \langle w, \mathcal{A}_N(J) \rangle}{N} \geq \text{ALG}_p(\kappa) + \eta \right) \leq e^{-cN}. \quad (10.6)$$

Theorem 10.1 is the case of one species and $\zeta(q) = q^p$ of [92, Theorem 1 and Corollary 1.8]. We now check that our setting satisfies the hypotheses there. In [92], the Hamiltonian is $\langle h, \sigma \rangle$ plus a Gaussian part on the ball of radius $\sqrt{N}$, the field is h times the all-ones vector $\mathbf{1}$ with $h \geq 0$, and an algorithm is a function of the disorder coefficients, concatenated into one vector with the Euclidean distance. Theorem 1 there states that, for fixed $\tau, \eta > 0$ and N large, every τ -Lipschitz algorithm reaches energy at least $\text{ALG} + \eta$ with probability at most e^{-cN} , with c and the threshold for N depending on τ , η , the covariance, and the field only. For one species with $\zeta(q) = q^p$, the Gaussian part is exactly H_N of (3.1), whose ordered coefficients are independent standard Gaussians; we take $h = \kappa$. Our algorithms are functions of J alone, the field being deterministic. Corollary 1.8 of [92], which covers odd and even p , states that $\text{ALG} = (\zeta'(1) + \kappa^2)^{1/2}$ if $\zeta'(1) + \kappa^2 \geq \zeta''(1)$, and otherwise

$$\text{ALG} = q_1 \zeta''(q_1)^{1/2} + \int_{q_1}^1 \zeta''(q)^{1/2} dq,$$

where $q_1 \in [0, 1)$ solves $\zeta'(q_1) + \kappa^2 = q_1 \zeta''(q_1)$. For $\zeta(q) = q^p$, the first condition reads $\kappa^2 \geq p(p-2)$, the equation for q_1 reads $\kappa^2 = p(p-2)q_1^{p-1}$, and

$$\begin{aligned} q_1 \zeta''(q_1)^{1/2} + \int_{q_1}^1 \zeta''(q)^{1/2} dq &= \sqrt{p(p-1)} \left(q_1^{p/2} + \frac{2}{p} (1 - q_1^{p/2}) \right) \\ &= E_{\text{th}} + \frac{p-2}{p} \sqrt{p(p-1)} q_1^{p/2}, \end{aligned}$$

where we used $2\sqrt{p(p-1)}/p = E_{\text{th}}$. This is (10.5). At $\kappa^2 = p(p-2)$, we have $q_1 = 1$ and both lines of (10.5) equal $\sqrt{p(p-1)}$.

It remains to replace $\mathbf{1}$ by an arbitrary direction w . Let O be an orthogonal matrix such that $O\mathbf{1} = w$. The map $J \mapsto O^{\otimes p}J$ is a linear isometry of $(\mathbb{R}^N)^{\otimes p}$ that preserves the standard Gaussian law, and $H_{N,O^{\otimes p}J}(Ox) = H_{N,J}(x)$ for all x . Given $\mathcal{A}_N$, the map $\mathcal{A}'_N(J) = O^\top \mathcal{A}_N(O^{\otimes p}J)$ is L_{Lip} -Lipschitz with values in B_N , and, with $J' = O^{\otimes p}J$,

$$H_{N,J}(\mathcal{A}'_N(J)) + \kappa \langle \mathbf{1}, \mathcal{A}'_N(J) \rangle = H_{N,J'}(\mathcal{A}_N(J')) + \kappa \langle w, \mathcal{A}_N(J') \rangle.$$

The tensor J' has the law of J , and [92, Theorem 1] applied to $\mathcal{A}'_N$ gives (10.6), with constants that do not depend on w or on $\mathcal{A}_N$. For $\kappa = 0$, we have $\text{ALG}_p(0) = E_{\text{th}}$, and Theorem 10.1 is the pure-model threshold, recorded in this form in [112, Proposition 1.9].

In this argument, we may always use the larger positive bound $L_{\text{Lip}} + 1$. This includes $L_{\text{Lip}} = 0$ and gives the same uniform choice of c and N_0 for every algorithm with the prescribed original bound. We need no separate measurability assumption on a deterministic algorithm, since Lipschitz maps are continuous.

Constructive finite-tree qualifications. The branching proof of the threshold uses finite trees whose depth and branching degrees are fixed before the dimension tends to infinity. Two auxiliary statements in [92, Appendix A] require qualifications in a constructive version. In Lemma A.8, if the absolute values of the leaf energies are bounded by $C > 0$ and the depth is $D \geq 1$, the retained branching degree $\lfloor K\varepsilon/(3CD) \rfloor$ is obtained in the range $0 < \varepsilon \leq CD$, and K is then taken large enough to make this degree positive. This is the range used when $\varepsilon \downarrow 0$; an unrestricted error can make the displayed degree exceed K .

For the base-preserving fork correction of Lemma A.9, we impose $k < N$ when k orthogonal increments are retained. After dividing the vertices by $\sqrt{N}$, suppose that the base x has squared norm $q \geq q_0 > 0$, the target increment has squared norm $q' \geq 0$, and the approximate increments $y_1, \dots, y_k$ satisfy, for $\varepsilon \geq 0$,

$$|||y_i||^2 - q'| \leq \varepsilon, \quad |\langle x, y_i \rangle| \leq \varepsilon, \quad |\langle y_i, y_j \rangle| \leq \varepsilon \quad (i \neq j).$$

Under the sufficient small-error conditions $\sqrt{\varepsilon} \leq q_0$ and $4(k+1)\sqrt{\varepsilon} \leq 1$, the base can be kept exactly and the increments corrected to mutually orthogonal vectors $z_i \perp x$ with $\|z_i\|^2 = q'$ and $\|z_i - y_i\| \leq e_k(\varepsilon)$, where

$$e_k(\varepsilon) = (4\sqrt{k+1} + 1)\sqrt{\varepsilon} + 2\sqrt{\sqrt{\varepsilon}}.$$

This bound permits $q' = 0$ and tends to zero for fixed k as $\varepsilon \downarrow 0$. It supplies a vanishing modulus, independent of the dimension, in place of the piecewise numerical modulus printed in that lemma.

We handle a schedule with zero initial overlap before this correction. For fixed $\delta \in (0, 1)$, we scale the normalized vertices, which have norm at most one, by $\sqrt{1 - \delta}$ and add a common orthogonal vector of norm $\sqrt{\delta}$. The new schedule of target overlaps is $q_i^{(\delta)} = (1 - \delta)q_i + \delta$, so its root is at least δ ; the terminal vectors change by at most $2\sqrt{\delta}$. We choose the dimension larger than the number of vertices of the retained finite tree, as well as its branching degree, so that the common orthogonal direction and all fork corrections exist. For fixed depth and retained branching, the accumulated correction tends to zero, and δ can be removed afterward. Repeated overlap levels are allowed: their target increments have norm zero.

In the algorithmic application, we can retain a fixed deterministic set of children, because the success event imposes the desired energy bound at every original terminal leaf. Internal vertices are first obtained from averages of the original terminal outputs. Restricting the independent Gaussian increments of the tree to this fixed set preserves their joint law. We therefore need no distributional assertion about an adaptively selected subtree. The depths, branching degrees, error tolerances, and dimension cutoffs are chosen before the algorithm and the deterministic field direction. The calibrated overlap schedule may depend on the algorithm, since the finite-tree ceiling is uniform over all admissible schedules. This is the uniformity asserted in Theorem 10.1.

The field version of the threshold penalizes the overlap of an algorithm with a direction independent of J . To state the penalty, recall $q_d = (p - 2)/(p - 1)$ and $\beta_d^2 = (p - 1)^{p-1}/(p(p - 2)^{p-2})$. The constant β_d is the dynamical transition point, and it enters here only as the coefficient of the penalty. The inverse temperature β of the dynamics plays no role in Theorem 10.1 and Lemma 10.2.

Lemma 10.2 (Absolute-overlap energy cost). *Fix an integer $p \geq 3$ and $0 \leq L_{\text{Lip}} < \infty$. There exist finite real numbers $\phi_N = \phi_N(p, L_{\text{Lip}}) \geq 0$, for every $N \geq 1$, with $\phi_N \rightarrow 0$ as $N \rightarrow \infty$ such that the following holds. Let ζ be a random variable independent of J , let $w = w(\zeta) \in \mathbb{R}^N$ satisfy $\|w\| = \sqrt{N}$, let $\mathcal{A}_N$ be an L_{Lip} -Lipschitz algorithm with seed ζ , and set $a_N = \langle w, \mathcal{A}_N \rangle / N$. Then, almost surely,*

$$\mathbb{E} \left[\frac{H_N(\mathcal{A}_N)}{N} + \beta_d |a_N|^p \mid \zeta \right] \leq E_{\text{th}} + \phi_N. \quad (10.7)$$

In particular,

$$\limsup_{N \rightarrow \infty} \left\{ \mathbb{E} \frac{H_N(\mathcal{A}_N)}{N} + \beta_d \mathbb{E} |a_N|^p \right\} \leq E_{\text{th}}. \quad (10.8)$$

The lemma makes no assumption on the concentration of a_N , and the seed may be absent, in which case w is deterministic.

Proof. The proof has three steps. We first compute the best ceiling that the field thresholds give for a given overlap, then we use finitely many fields at once, and finally we refine the set of fields after the dimension limit.

Step 1: the conjugate ceiling. For $0 \leq a \leq 1$, we define

$$\mathcal{E}_p(a) = \inf_{\kappa \geq 0} \{ \text{ALG}_p(\kappa) - \kappa a \}.$$

We claim that

$$\mathcal{E}_p(a) = \begin{cases} E_{\text{th}} - \beta_d a^p, & 0 \leq a \leq \sqrt{q_d}, \\ \sqrt{p(1-a^2)}, & \sqrt{q_d} \leq a \leq 1, \end{cases} \quad \text{and} \quad E_{\text{th}} - \mathcal{E}_p(a) \geq \beta_d a^p \quad (0 \leq a \leq 1). \quad (10.9)$$

The function ALG_p is convex on $[0, \infty)$, since $\kappa^{p/(p-1)}$ and $\sqrt{p + \kappa^2}$ are convex. It is also continuously differentiable. Indeed, let $\kappa_* = \sqrt{p(p-2)}$. On $[0, \kappa_*]$, its derivative is $p(p-1)^{-1} \gamma_p \kappa^{1/(p-1)}$, which increases from 0 to

$$\frac{p}{p-1} \gamma_p \kappa_*^{1/(p-1)} = \frac{p-2}{p-1} \sqrt{\frac{p(p-1)}{p(p-2)}} = \sqrt{q_d}.$$

On $[\kappa_*, \infty)$, its derivative is $\kappa / \sqrt{p + \kappa^2}$, which increases from $\sqrt{q_d}$ to 1. By convexity, for $0 \leq a < 1$ the infimum defining $\mathcal{E}_p(a)$ is attained where $\text{ALG}'_p(\kappa) = a$.

If $a \leq \sqrt{q_d}$, this point is $\kappa = ((p-1)a / (p\gamma_p))^{p-1}$. At this point, we have $\gamma_p \kappa^{1/(p-1)} = (p-1)a/p$, and therefore

$$\mathcal{E}_p(a) = E_{\text{th}} + \kappa(\gamma_p \kappa^{1/(p-1)} - a) = E_{\text{th}} - \frac{\kappa a}{p} = E_{\text{th}} - \frac{1}{p} \left(\frac{p-1}{p\gamma_p} \right)^{p-1} a^p.$$

Inserting the value of γ_p , we have

$$\begin{aligned} \left(\frac{p-1}{p\gamma_p} \right)^{p-1} &= \left(\frac{p-1}{p-2} \right)^{p-1} (p(p-1))^{-(p-1)/2} (p(p-2))^{p/2} \\ &= p^{1/2} (p-1)^{(p-1)/2} (p-2)^{-(p-2)/2}, \end{aligned}$$

and dividing by p , we find that the coefficient of a^p is $((p-1)^{p-1} / (p(p-2)^{p-2}))^{1/2} = \beta_d$.

If $\sqrt{q_d} \leq a < 1$, the point is $\kappa = \sqrt{p} a / \sqrt{1-a^2}$, and

$$\mathcal{E}_p(a) = \sqrt{\frac{p}{1-a^2}} - \frac{\sqrt{p} a^2}{\sqrt{1-a^2}} = \sqrt{p(1-a^2)}.$$

At $a = 1$, we have $\mathcal{E}_p(1) = 0$, because the function $\sqrt{p + \kappa^2} - \kappa$ decreases to 0 as $\kappa \rightarrow \infty$. The two formulas agree at $a = \sqrt{q_d}$, where both equal $\sqrt{p/(p-1)}$. This proves the first part of (10.9), and shows that $\mathcal{E}_p$ is continuous on $[0, 1]$.

Let $D_p(a) = E_{\text{th}} - \mathcal{E}_p(a)$. For $a \leq \sqrt{q_d}$, we have $D_p(a) = \beta_d a^p$. For $\sqrt{q_d} < a < 1$, we have $D'_p(a) = \sqrt{p} a / \sqrt{1-a^2}$. The function $a^{p-2} \sqrt{1-a^2}$ on $[0, 1]$ is maximal at $a^2 = q_d$, where it equals

$$q_d^{(p-2)/2} (1-q_d)^{1/2} = (p-2)^{(p-2)/2} (p-1)^{-(p-1)/2} = \frac{1}{\sqrt{p} \beta_d}.$$

Since $D'_p(a)/(p\beta_d a^{p-1}) = (\sqrt{p}\beta_d a^{p-2}\sqrt{1-a^2})^{-1}$, we conclude that $D'_p(a) \geq p\beta_d a^{p-1}$ on $(\sqrt{q_d}, 1)$. Integrating this inequality from $\sqrt{q_d}$, where $D_p(\sqrt{q_d}) = \beta_d q_d^{p/2}$, and using the continuity of D_p at $a = 1$, we find that $D_p(a) \geq \beta_d a^p$ on $[\sqrt{q_d}, 1]$. This proves (10.9).

Step 2: finitely many fields. Fix a finite set $\mathcal{G} \subset [0, \infty)$ with $0 \in \mathcal{G}$, and $\eta > 0$. We define

$$V_{\mathcal{G}}(a) = \min_{\kappa \in \mathcal{G}} \{ \text{ALG}_p(\kappa) - \kappa a \}, \quad 0 \leq a \leq 1.$$

Note that $|V_{\mathcal{G}}| \leq M_{\mathcal{G}}$, where $M_{\mathcal{G}} = E_{\text{th}} + \max \mathcal{G}$, because $0 \in \mathcal{G}$ and $\text{ALG}_p \geq 0$. We condition on ζ . Then w is fixed, J is standard Gaussian, and $J \mapsto \mathcal{A}_N(J, \zeta)$ is L_{Lip} -Lipschitz. For each $\kappa \in \mathcal{G}$, we use Theorem 10.1 with the field strength κ and the directions w and $-w$. Let $c(\mathcal{G}, \eta) > 0$ and $N_0(\mathcal{G}, \eta)$ be the smallest constant and the largest threshold for N among these $|\mathcal{G}|$ applications. For $N \geq N_0(\mathcal{G}, \eta)$, the event

$$\mathfrak{B} = \bigcup_{\kappa \in \mathcal{G}} \left\{ \frac{H_N(\mathcal{A}_N)}{N} + \kappa |a_N| \geq \text{ALG}_p(\kappa) + \eta \right\}$$

has conditional probability at most $2|\mathcal{G}|e^{-c(\mathcal{G}, \eta)N}$, uniformly in ζ . Off $\mathfrak{B}$, we have $H_N(\mathcal{A}_N)/N < V_{\mathcal{G}}(|a_N|) + \eta$; note that $|a_N| \leq 1$, since $\|w\| = \sqrt{N}$ and $\mathcal{A}_N \in B_N$. On $\mathfrak{B}$, we use $|H_N(\mathcal{A}_N)|/N \leq Q_N$ from (10.3). Recall that J is independent of ζ , which gives $\mathbb{E}[Q_N^2 | \zeta] = \mathbb{E}Q_N^2 \leq Q_*^2$. Let

$$r_N(\mathcal{G}, \eta) = (Q_* + M_{\mathcal{G}})(2|\mathcal{G}|e^{-c(\mathcal{G}, \eta)N})^{1/2}.$$

Using the Cauchy–Schwarz inequality, we obtain

$$\mathbb{E} \left[\frac{H_N(\mathcal{A}_N)}{N} - V_{\mathcal{G}}(|a_N|) \mid \zeta \right] \leq \eta + r_N(\mathcal{G}, \eta) \quad (10.10)$$

for $N \geq N_0(\mathcal{G}, \eta)$. For $1 \leq N < N_0(\mathcal{G}, \eta)$, we set $r_N(\mathcal{G}, \eta) = Q_* + M_{\mathcal{G}}$. Then (10.10) still holds, directly from $|H_N(\mathcal{A}_N)|/N \leq Q_N$ and $|V_{\mathcal{G}}| \leq M_{\mathcal{G}}$. The finite quantity $r_N(\mathcal{G}, \eta)$ depends only on p , L_{Lip} , $\mathcal{G}$, and η , and tends to zero as $N \rightarrow \infty$.

Step 3: refining the fields. Let $\mathcal{G}_k = \{i2^{-k} : 0 \leq i \leq 4^k\}$. These sets increase with k , and their union is dense in $[0, \infty)$. The functions $V_{\mathcal{G}_k}$ are continuous on $[0, 1]$ and decrease in k . For each $a \in [0, 1]$, the map $\kappa \mapsto \text{ALG}_p(\kappa) - \kappa a$ is continuous, and its infimum over the dense set $\bigcup_k \mathcal{G}_k$ equals $\mathcal{E}_p(a)$. The limit $\mathcal{E}_p$ is continuous on $[0, 1]$. By Dini's theorem, the quantity $s_k = \sup_{0 \leq a \leq 1} \{V_{\mathcal{G}_k}(a) - \mathcal{E}_p(a)\}$ tends to zero. By (10.9) and (10.10) with $\mathcal{G} = \mathcal{G}_k$ and $\eta = 1/k$, we have

$$\begin{aligned} \mathbb{E} \left[\frac{H_N(\mathcal{A}_N)}{N} + \beta_d |a_N|^p \mid \zeta \right] &\leq E_{\text{th}} + \mathbb{E} \left[\frac{H_N(\mathcal{A}_N)}{N} - \mathcal{E}_p(|a_N|) \mid \zeta \right] \\ &\leq E_{\text{th}} + s_k + \frac{1}{k} + r_N(\mathcal{G}_k, 1/k) \end{aligned}$$

for every $k \geq 1$. We define $\phi_N = \inf_{k \geq 1} \{s_k + 1/k + r_N(\mathcal{G}_k, 1/k)\}$, which depends only on p and L_{Lip} . This infimum is nonnegative and finite for every $N \geq 1$, since already the term with $k = 1$ is finite. Then (10.7) holds. Moreover, $\limsup_N \phi_N \leq s_k + 1/k$ for every k , and in particular $\phi_N \rightarrow 0$. Taking expectations in (10.7) and letting $N \rightarrow \infty$ gives (10.8). $\square$

10.2 Finite forks and the exact energy subtraction

In this section, we fix $\beta > 0$ and a stationary cluster (c, r, ℓ) of the translated CK kernels, as in Proposition 5.7, realized along a sequence $T_n \rightarrow \infty$. Its plateau is q and its energy is $e_\infty = (\ell - 1)/(p\beta)$. Recall from (5.24) that $0 \leq c \leq 1$, that $r = -c' \geq 0$, so that c is nonincreasing, and that $\int_0^\infty r(\tau) d\tau = 1 - q$. Recall also that $e_\beta(T_n + a) \rightarrow e_\infty$ for every $a \geq 0$, by (4.1) and the convergence of the translated multipliers.

Entropy notation. We use the notation of Section 5. For fixed disorder, ρ_t is the density of X_t with respect to ω_N and g_N is the Gibbs density. We write

$$v_t^{\text{cur}} = \nabla_S \log(\rho_t / g_N), \quad I_H(t) = \int \|v_t^{\text{cur}}\|^2 \rho_t d\omega_N, \quad a_N(t) = \frac{1}{N} \mathbb{E} I_H(t)$$

for the relative score, the Fisher information, and its normalized mean. This is the notation of Section 5; the superscript distinguishes the score from the tree vectors v_t . Let $\mu_N(dt) = a_N(t) dt$. Lemma 5.1 implies that $\mu_N([0, \infty)) \leq \beta^2/2$. By (5.9), $\mu_{N_j}(I) \rightarrow \mu(I)$ for every bounded interval I , where $\mu(dt) = a_F(t) dt$ is the finite measure of (5.14). We write P_t^H for the diffusion semigroup at fixed disorder and $\mathcal{L}_N$ for its generator.

Forks. For each fixed integer $K \geq 1$, we use the product probability space carrying J , the uniform initial point X_0 , a full Brownian path $(B_u)_{u \geq 0}$, and K full Brownian paths $B^{(1)}, \dots, B^{(K)}$ on $[0, \infty)$. These coordinate blocks are mutually independent. We call

$$\zeta_N = (X_0, (B^{(b)})_{1 \leq b \leq K}, B)$$

the fork seed, with its uniform-initial and Wiener product law; the disorder has its independent standard Gaussian law. The seed space and law depend on K , which we suppress in the notation. Fix $T \geq 0$ and $L \geq 0$. The trajectory X solves (3.3) driven by B . For $b \leq K$, let Z_b be the solution of (3.3) on $[0, T + L]$ with $Z_b(0) = X_0$, driven by the Brownian motion that equals B_u for $u \leq T$ and $B_T + B_{u-T}^{(b)}$ for $u \geq T$. Then $Z_b(u) = X_u$ for $u \leq T$, and after time T the paths $Z_1, \dots, Z_K$ are continuations of X from X_T by independent noises. We write

$$Y_b = Z_b(T + L), \quad m_{N,T} = \frac{1}{K} \sum_{b=1}^K Y_b,$$

and call $m_{N,T}$ the center of the fork. When needed, we indicate the dependence on K and L in the notation. The Brownian motion that drives Z_b is a standard Brownian motion independent of (J, X_0) . Hence each pair (J, Z_b) has the same law as $(J, (X_u)_{u \leq T+L})$. The branches share their past before T and are not independent. Note that $\|m_{N,T}\| \leq \sqrt{N}$, because $\|Y_b\| = \sqrt{N}$.

Response moments. Let $U_b(t, u)$ be the variational response matrix of the path Z_b , defined as in (3.6). Each driving process is Brownian for its own full filtration and is independent of (J, X_0) , even though the branches share a common past. Therefore the proof of (4.4) applies to every branch. With the constants K_S and Θ_S defined there, its coordinate columns and normalized trace have fourth moment at most K_S whenever $0 \leq t - u \leq S$. In particular,

$$\mathbb{E} \left| \frac{\text{tr } U_b(t, u)}{N} \right|^2 \leq \Theta_S^2, \quad \mathbb{E} \|U_b(t, u)\|_{\text{op}}^2 \leq N \Theta_S^2.$$

We use these estimates on the whole interval $[0, T + L]$, with $S = T + L$, and on the final interval $[T, T + L]$, with $S = L$. The latter bounds are uniform in T .

Lemma 10.3 (Entropy comparison for two branches). *Let $T, a, d \geq 0$, and let Y and Z be the endpoints of two branches at time T of lengths a and d , that is, of two continuations of X from X_T driven by independent Brownian motions. Set $Q_{N,T}(a, d) = \langle Y, Z \rangle / N$. Then, for $k \in \{1, 2\}$,*

$$\left| \mathbb{E} Q_{N,T}(a, d)^k - \mathbb{E} C_N(T + 2a + d, T + a)^k \right| \leq 2k e^{\beta p^2 A_p(a+d)} \sqrt{a \mu_N([T, T + a])} + 2e^{-c_p N}. \quad (10.11)$$

Moreover, if $T_n \rightarrow \infty$ is a sequence along which the translated kernels converge to a stationary cluster with correlation profile c , then, for every $L < \infty$,

$$\lim_{n \rightarrow \infty} \limsup_{j \rightarrow \infty} \sup_{0 \leq a, d \leq L} \mathbb{E} \left| Q_{N_j, T_n}(a, d) - c(a + d) \right|^2 = 0. \quad (10.12)$$

Proof. For $k \in \{1, 2\}$, let $\Phi_k(x) = (x/\sqrt{N})^{\otimes k}$, with values in $(\mathbb{R}^N)^{\otimes k}$ and its Euclidean norm, so that $\langle \Phi_k(x), \Phi_k(y) \rangle = (\langle x, y \rangle / N)^k$. On S_N , we have $\|\Phi_k\| = 1$ and $\|D_S \Phi_k\|_{\text{op}} \leq k/\sqrt{N}$. Fix the disorder. For $0 \leq z \leq a$, we define

$$f_z = P_{a-z}^H \Phi_k, \quad g_z = P_{d+z}^H \Phi_k, \quad F_J(z) = \int_{S_N} \rho_{T+z} \langle f_z, g_z \rangle d\omega_N,$$

where the semigroup acts on each coordinate of Φ_k .

We first identify the endpoint values. By the Markov property and the conditional independence of the two branches given the past up to time T , we have $\mathbb{E} F_J(0) = \mathbb{E} Q_{N,T}(a, d)^k$. By the semigroup property, $F_J(a)$ is the conditional expectation of $\langle \Phi_k(X_{T+a}), \Phi_k(X_{T+2a+d}) \rangle$ given J , so that $\mathbb{E} F_J(a) = \mathbb{E} C_N(T + 2a + d, T + a)^k$.

Next we compute F'_j . We have $\partial_z f_z = -\mathcal{L}_N f_z$, $\partial_z g_z = \mathcal{L}_N g_z$, and $\partial_z \rho_{T+z} = \mathcal{L}_N^* \rho_{T+z}$. By the product rule for the generator, applied to each pair of coordinates, $\mathcal{L}_N \langle f, g \rangle = \langle \mathcal{L}_N f, g \rangle + \langle f, \mathcal{L}_N g \rangle + 2 \sum_i \langle D_S f[e_i], D_S g[e_i] \rangle$, where (e_i) is an orthonormal frame of the tangent space. Using these identities, we find

$$\begin{aligned} F'_j(z) &= \int \rho_{T+z} \{ \mathcal{L}_N \langle f_z, g_z \rangle - \langle \mathcal{L}_N f_z, g_z \rangle + \langle f_z, \mathcal{L}_N g_z \rangle \} d\omega_N \\ &= 2 \int \rho_{T+z} \left\{ \langle f_z, \mathcal{L}_N g_z \rangle + \sum_i \langle D_S f_z[e_i], D_S g_z[e_i] \rangle \right\} d\omega_N. \end{aligned}$$

When we apply (5.8) to each coordinate of f_z and g_z , the gradient terms cancel, and we obtain the exact current identity

$$F'_j(z) = -2 \int \rho_{T+z} \left\langle P_{a-z}^H \Phi_k, D_S(P_{d+z}^H \Phi_k) [v_{T+z}^{\text{cur}}] \right\rangle d\omega_N. \quad (10.13)$$

Note that this identity does not use stationarity of ρ_{T+z} : the score v_{T+z}^{cur} records how far ρ_{T+z} is from the Gibbs density.

We now bound the right side of (10.13) on the event $G_N = \{Q_N \leq A_p\}$. By Jensen's inequality, $\|P_t^H \Phi_k\| \leq 1$. In the proof of Lemma 5.5, the Bochner formula, the nonnegative Ricci curvature of S_N , and the Hessian bound of (3.26) give the gradient bound $\|\nabla_S P_t^H \varphi\|^2 \leq e^{2\beta p^2 Q_N t} P_t^H \|\nabla_S \varphi\|^2$ for smooth scalar φ ; this is the classical gradient bound under a lower bound on the curvature. Applied to $\varphi = \langle w, \Phi_k \rangle$ for a unit vector w in $(\mathbb{R}^N)^{\otimes k}$, and followed by the supremum over w , it gives

$$\|D_S P_t^H \Phi_k\|_{\text{op}} \leq \frac{k}{\sqrt{N}} e^{\beta p^2 A_p t} \quad \text{on } G_N,$$

with no factor depending on the dimension of $(\mathbb{R}^N)^{\otimes k}$. Using the Cauchy–Schwarz inequality, we find that, on G_N and for $z \leq a$,

$$|F'_j(z)| \leq \frac{2k}{\sqrt{N}} e^{\beta p^2 A_p(a+d)} \int \rho_{T+z} \|v_{T+z}^{\text{cur}}\| d\omega_N \leq 2k e^{\beta p^2 A_p(a+d)} \left(\frac{I_H(T+z)}{N} \right)^{1/2}.$$

Taking expectations and using Jensen's inequality, then integrating in z and using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \int_0^a \mathbb{E}[\mathbf{1}_{G_N} |F'_j(z)|] dz &\leq 2k e^{\beta p^2 A_p(a+d)} \int_0^a \sqrt{a_N(T+z)} dz \\ &\leq 2k e^{\beta p^2 A_p(a+d)} \sqrt{a \mu_N([T, T+a])}. \end{aligned}$$

On G_N^c , we have $|F_j(0)|, |F_j(a)| \leq 1$, and $\mathbb{P}(G_N^c) \leq e^{-c_p N}$. Writing $\mathbb{E}[F_j(0) - F_j(a)] = -\mathbb{E}[\mathbf{1}_{G_N} \int_0^a F'_j(z) dz] + \mathbb{E}[\mathbf{1}_{G_N^c} (F_j(0) - F_j(a))]$ and combining the last two bounds with the identification of the endpoint values, we obtain (10.11).

For (10.12), fix L and let $a, d \leq L$. In (10.11), we may replace $a \mu_N([T, T + a])$ by $L \mu_N([T, T + L])$ and the exponential factor by $e^{2\beta p^2 A_p L}$. Then (5.9) implies that $\mu_{N_j}([T, T + L]) \rightarrow \mu([T, T + L])$ for each fixed T . By Proposition 3.3 on the horizon $T + 3L$ and $|C_N| \leq 1$, the moments $\mathbb{E}C_N(T + 2a + d, T + a)^k$ converge to $C(T + 2a + d, T + a)^k$ uniformly in $a, d \leq L$. Combining these facts with (10.11), we find

$$\limsup_{j \rightarrow \infty} \sup_{0 \leq a, d \leq L} \left| \mathbb{E}Q_{N_j, T}(a, d)^k - C(T + 2a + d, T + a)^k \right| \leq 4e^{2\beta p^2 A_p L} \sqrt{L \mu([T, T + L])}.$$

Along $T = T_n$, the right side tends to zero because μ is a finite measure, and $C(T_n + 2a + d, T_n + a) \rightarrow c(a + d)$ uniformly in $a, d \leq L$, by the locally uniform convergence of the translated correlations on $[0, 3L]^2$. Finally, writing $Q = Q_{N, T}(a, d)$ and $c_0 = c(a + d) \in [0, 1]$, we have

$$\mathbb{E}|Q - c_0|^2 = (\mathbb{E}Q^2 - c_0^2) - 2c_0(\mathbb{E}Q - c_0),$$

and the convergence of the first two moments, uniformly in $a, d \leq L$, proves (10.12). $\square$

We now state the main result of this section. For $K \geq 1$ and $L \geq 0$, we define

$$Q_{K, L} = \frac{1}{K} + \left(1 - \frac{1}{K}\right) c(2L), \quad (10.14)$$

$$k_{K, L}(\tau) = \frac{c(\tau) + (K - 1)c(2L - \tau)}{K}, \quad 0 \leq \tau \leq L, \quad (10.15)$$

$$h_{K, L} = e_\infty - \beta p \int_0^L r(\tau) \left\{ c(\tau)^{p-1} - k_{K, L}(\tau)^{p-1} \right\} d\tau. \quad (10.16)$$

Note that $Q_{K, L}$ is a convex combination of 1 and $c(2L)$, so that $q \leq c(2L) \leq Q_{K, L} \leq 1$, and that $0 \leq k_{K, L} \leq 1$.

Proposition 10.4 (Finite forks and their center). *Fix $\beta > 0$ and let (c, r, ℓ) be a stationary cluster with plateau $q > 0$ and energy e_∞ , realized along $T_n \rightarrow \infty$. Fix $K \geq 1$ and $L \geq 0$, and let $m_{N, T}$ be the center of the fork at time T with K branches of length L .*

(a) *The squared norm and the energy of the center satisfy*

$$\lim_{n \rightarrow \infty} \limsup_{j \rightarrow \infty} \mathbb{E} \left| \frac{\|m_{N_j, T_n}\|^2}{N_j} - Q_{K, L} \right|^2 = 0, \quad (10.17)$$

$$\lim_{n \rightarrow \infty} \limsup_{j \rightarrow \infty} \left| \frac{\mathbb{E}H_{N_j}(m_{N_j, T_n})}{N_j} - h_{K, L} \right| = 0. \quad (10.18)$$

(b) *The limiting energy is below the threshold:*

$$h_{K, L} \leq E_{\text{th}} Q_{K, L}^{p/2}. \quad (10.19)$$

(c) For every $T \geq 0$ and $\varepsilon > 0$, there exist algorithms $\mathcal{A}_{N,T}^\varepsilon = \mathcal{A}_{N,T}^\varepsilon(J, \zeta_N)$ with values in B_N and seed ζ_N , which are L_{Lip} -Lipschitz for a constant L_{Lip} depending only on p, β, T, K, L , and ε , such that

$$\limsup_{j \rightarrow \infty} \frac{1}{N_j} \mathbb{E} \|\mathcal{A}_{N_j,T}^\varepsilon - \Pi_{B_{N_j}}(m_{N_j,T} / \sqrt{Q_{K,L}})\|^2 \leq \varepsilon^2. \quad (10.20)$$

Moreover, along $T = T_n$, the vector $\Pi_{B_N}(m_{N,T} / \sqrt{Q_{K,L}})$ may be replaced by $m_{N,T} / \sqrt{Q_{K,L}}$, or by $\hat{m}_{N,T} = \sqrt{N} m_{N,T} / \|m_{N,T}\|$ (with $\hat{m}_{N,T} = 0$ if $m_{N,T} = 0$): the scaled mean-square distance between $\Pi_{B_N}(m_{N,T} / \sqrt{Q_{K,L}})$ and either vector tends to zero as $j \rightarrow \infty$ and then $n \rightarrow \infty$.

By (10.3), the normalized energies of the algorithms in part (c) are bounded by Q_N . Part (c) also transfers overlaps. If V_N is a random vector on the same probability space with $\|V_N\| \leq \sqrt{N}$ (for example $V_N = X_s$ with $s \leq T$), then, by the Cauchy–Schwarz inequality, we have

$$\left| \mathbb{E} \frac{\langle \mathcal{A}_{N,T}^\varepsilon, V_N \rangle}{N} - \mathbb{E} \frac{\langle m_{N,T}, V_N \rangle}{N \sqrt{Q_{K,L}}} \right| \leq \left(\frac{1}{N} \mathbb{E} \|\mathcal{A}_{N,T}^\varepsilon - m_{N,T} / \sqrt{Q_{K,L}}\|^2 \right)^{1/2}. \quad (10.21)$$

Proof. The proof has five steps. In the first two, we compute the energy of the center exactly by Gaussian integration by parts and subtract the energy of a single branch. In the third, we show that the common past before T contributes nothing in the dimension limit, and in the fourth we evaluate the contribution of the last L units of time; this proves part (a). In the fifth step, we construct the Lipschitz algorithms of part (c) and deduce part (b) from the threshold. We write $t = T + L$.

Step 1: Gaussian integration by parts. For an ordered index $I = (i_1, \dots, i_p)$, let $G_I(x) = N^{-(p-1)/2} x_{i_1} \cdots x_{i_p}$, so that $H_N = \sum_I J_I G_I$. Since $\sum_I G_I(x) G_I(z) = N(\langle x, z \rangle / N)^p$, differentiating first in x_i and then in z_l gives

$$\sum_I \partial_i G_I(x) \partial_l G_I(z) = p \left(\frac{\langle x, z \rangle}{N} \right)^{p-1} \delta_{il} + \frac{p(p-1)}{N} \left(\frac{\langle x, z \rangle}{N} \right)^{p-2} z_i x_l. \quad (10.22)$$

Differentiating the actual terminal solution into L^2 with respect to the disorder, as in the proof of Proposition 4.1, we have

$$\partial_{J_I} m_{N,T} = \frac{\beta}{K} \sum_{b=1}^K \int_0^t U_b(t, u) \nabla G_I(Z_b(u)) \, du.$$

The seed is independent of J , and the response has the column and operator moments above. As in the proof of Proposition 4.1, apply Gaussian integration by parts to the expected coordinate observable in each disorder coefficient, using its genuine L^2 derivative:

$$\mathbb{E} [J_I G_I(m_{N,T})] = \mathbb{E} [\langle \nabla G_I(m_{N,T}), \partial_{J_I} m_{N,T} \rangle],$$

and $\mathbb{E}H_N(m_{N,T})$ is the sum of these terms over I . For $u \leq t$, we write

$$\gamma_b(u) = \frac{\langle Y_b, Z_b(u) \rangle}{N}, \quad \bar{\gamma}_b(u) = \frac{\langle m_{N,T}, Z_b(u) \rangle}{N} = \frac{1}{K} \sum_{b'=1}^K \frac{\langle Y_{b'}, Z_b(u) \rangle}{N}.$$

Contracting (10.22) with $U_b(t, u)$ gives the exact identity

$$\begin{aligned} \frac{\mathbb{E}H_N(m_{N,T})}{N} &= \frac{\beta p}{K} \sum_{b=1}^K \int_0^t \mathbb{E} \left[\frac{\text{tr } U_b(t, u)}{N} \bar{\gamma}_b(u)^{p-1} \right] du + \mathcal{R}_{N,T}, \\ \mathcal{R}_{N,T} &= \frac{\beta p(p-1)}{K} \sum_{b=1}^K \int_0^t \mathbb{E} \left[\bar{\gamma}_b(u)^{p-2} \frac{\langle Z_b(u), U_b(t, u) m_{N,T} \rangle}{N^2} \right] du. \end{aligned} \quad (10.23)$$

Using $\|m_{N,T}\|, \|Z_b(u)\| \leq \sqrt{N}$ and $|\bar{\gamma}_b| \leq 1$, we find that $|\mathcal{R}_{N,T}| \leq \beta p(p-1)t \Theta_t / \sqrt{N}$.

Step 2: the exact subtraction. The same computation with $m_{N,T}$ replaced by Y_b gives $\mathbb{E}H_N(Y_b)/N = \beta p \int_0^t \mathbb{E}[N^{-1} \text{tr } U_b(t, u) \gamma_b(u)^{p-1}] du + O(t \Theta_t / \sqrt{N})$. The pair (J, Y_b) has the law of (J, X_t) , and therefore $\mathbb{E}H_N(Y_b)/N = e_N(t)$, where $e_N(t) = \mathbb{E}H_N(X_t)/N$. Averaging over b and subtracting (10.23), before taking any limit, we obtain

$$\begin{aligned} e_N(t) - \frac{\mathbb{E}H_N(m_{N,T})}{N} &= \frac{\beta p}{K} \sum_{b=1}^K \int_0^t \mathbb{E} \left[\frac{\text{tr } U_b(t, u)}{N} \{ \gamma_b(u)^{p-1} - \bar{\gamma}_b(u)^{p-1} \} \right] du \\ &\quad + O(t \Theta_t N^{-1/2}), \end{aligned} \quad (10.24)$$

where the implicit constant depends only on p and β . Note that this identity retains the whole history of the trajectory.

Step 3: the common past. Let $u \leq T$. Then $Z_b(u) = X_u$ for every b , and the pair (Y_b, X_u) has the law of (X_t, X_u) . By Proposition 3.3 and $|C_N| \leq 1$, we have $\mathbb{E}|\gamma_b(u) - C(t, u)|^2 \rightarrow 0$ as $N \rightarrow \infty$, uniformly in $u \leq T$. The same holds for $\bar{\gamma}_b(u)$, which is the average of $\gamma_{b'}(u)$ over b' . The function $x \mapsto x^{p-1}$ is $(p-1)$ -Lipschitz on $[-1, 1]$, and the normalized trace has second moment at most Θ_t^2 . Using the Cauchy–Schwarz inequality, we find that the part of the integral in (10.24) over $u \leq T$ is at most

$$\beta p(p-1) T \Theta_t \sup_{b \leq K, u \leq T} (\mathbb{E}|\gamma_b(u) - \bar{\gamma}_b(u)|^2)^{1/2}$$

in absolute value. This tends to zero as $N \rightarrow \infty$ with T fixed. We use no estimate on the decay of the memory here. The contribution of the past is small because, at a fixed time horizon, both overlaps concentrate around the same deterministic value $C(t, u)$.

Step 4: the last L units of time. Let $u = T + v$ with $0 \leq v \leq L$, and write $\tau = L - v$. The pair $(Y_b, Z_b(T + v))$ has the law of (X_t, X_{T+v}) . By Proposition 3.3, the overlap $\gamma_b(T + v)$

converges in L^2 to $C(T+L, T+v)$ as $N \rightarrow \infty$, uniformly in v . Along $T = T_n$, this tends to $c(L-v) = c(\tau)$, uniformly in $v \leq L$. For $b' \neq b$, the vectors $Y_{b'}$ and $Z_b(T+v)$ are the endpoints of two branches at T of lengths L and v . By (10.12), the overlap $\langle Y_{b'}, Z_b(T+v) \rangle / N$ converges in L^2 to $c(L+v) = c(2L-\tau)$, in the double limit $j \rightarrow \infty$ and then $n \rightarrow \infty$, uniformly in $v \leq L$. Averaging over b' , we find

$$\lim_{n \rightarrow \infty} \limsup_{j \rightarrow \infty} \sup_{0 \leq v \leq L} \left(\mathbb{E} |\gamma_b(T_n + v) - c(L-v)|^2 + \mathbb{E} |\bar{\gamma}_b(T_n + v) - k_{K,L}(L-v)|^2 \right) = 0.$$

On this interval, we have $t-u \leq L$, and the normalized response trace has second moment at most Θ_L^2 , uniformly in T . As in Step 3, using the Cauchy–Schwarz inequality, we may replace $\gamma_b^{p-1} - \bar{\gamma}_b^{p-1}$ in (10.24) by the deterministic weight

$$\omega(\tau) = c(\tau)^{p-1} - k_{K,L}(\tau)^{p-1}, \quad \tau = t-u,$$

with an error whose $\limsup_j$ tends to zero as $n \rightarrow \infty$.

It remains to pass to the limit in $\beta p \int_T^t \bar{r}_N(t, u) \omega(t-u) du$, where $\bar{r}_N(t, u) = \mathbb{E} \operatorname{tr} U_b(t, u) / N$ is the averaged response of Lemma 3.4, which does not depend on b . Lemma 3.4 and Proposition 3.3 imply that $\int_{s_1}^{s_2} \bar{r}_N(t, u) du \rightarrow \int_{s_1}^{s_2} R(t, u) du$ for all $T \leq s_1 \leq s_2 \leq t$. The signed measure $\bar{r}_N(t, u) du$ on $[T, t]$ has total variation

$$\int_T^t |\bar{r}_N(t, u)| du \leq \int_T^t \mathbb{E} \left| \frac{\operatorname{tr} U_b(t, u)}{N} \right| du \leq L \Theta_L.$$

As $R \geq 0$, the integral of $R(t, \cdot)$ over $[T, t]$ is the limit of the masses of these signed measures and obeys the same bound. The weight ω is continuous on $[0, L]$ and is the uniform limit of step functions. For a step function, the integral converges as $N \rightarrow \infty$, and the total variation bound controls the error of the approximation, uniformly in N . We conclude that, for fixed T ,

$$\lim_{N \rightarrow \infty} \int_T^t \bar{r}_N(t, u) \omega(t-u) du = \int_T^t R(t, u) \omega(t-u) du.$$

Along $T = T_n$, we have $R(T_n + L, T_n + v) \rightarrow r(L-v)$ uniformly in $v \leq L$, by the translated convergence on the closed causal triangle. Moreover, $e_N(t) \rightarrow e_\beta(T+L)$ by Proposition 4.1, and $e_\beta(T_n + L) \rightarrow e_\infty$. Inserting these limits into (10.24), we obtain

$$\lim_{n \rightarrow \infty} \limsup_{j \rightarrow \infty} \left| \frac{\mathbb{E} H_{N_j}(m_{N_j, T_n})}{N_j} - e_\infty + \beta p \int_0^L r(\tau) \omega(\tau) d\tau \right| = 0,$$

which is (10.18) by the definition (10.16).

For the norm, we have

$$\frac{\|m_{N,T}\|^2}{N} = \frac{1}{K} + \frac{1}{K^2} \sum_{b \neq b'} \frac{\langle Y_b, Y_{b'} \rangle}{N}.$$

Each overlap $\langle Y_b, Y_{b'} \rangle / N$ with $b \neq b'$ has the law of $Q_{N,T}(L, L)$, which tends to $c(2L)$ in L^2 in the double limit, by (10.12). There are $K(K-1)$ such terms, and $K(K-1)/K^2 = 1 - 1/K$. This proves (10.17) and completes the proof of part (a).

Step 5: Lipschitz algorithms and the threshold. Fix T and $\varepsilon > 0$. We may assume that $\varepsilon < 1$. Let $\varepsilon_1 = \varepsilon/(2\sqrt{K})$. We use the hard-soft approximation of [112, Lemmas 3.1–3.2]. We first check its hypotheses. In [112], the Hamiltonian is our H_N , with an ordered tensor of independent standard Gaussian coefficients and the Euclidean distance between coefficient tensors, and the dynamics is $dx_v = (\beta_S \nabla_{\text{sp}} H_N(x_v) - (N-1)(2N)^{-1} x_v) dv + P_{x_v} dW_v$, with noise coefficient one. With $\beta_S = \beta/2$, $X_u = x_{2u}$, and $B_u = W_{2u}/\sqrt{2}$, this is (3.3), as in the proof of Proposition 3.3, and our horizon t becomes $2t$. Lemma 3.1 there states that, for a fixed initial point on S_N and a Hamiltonian satisfying their derivative bounds (which hold for H_N with probability at least $1 - e^{-c_1 N}$, by [112, Proposition 1.11]), the hard dynamics stays within $\varepsilon_1 \sqrt{N}$ of a soft spherical dynamics driven by the same Brownian motion, on the whole horizon, with probability at least $1 - e^{-c_1 N}$. Here the confinement parameter of the soft dynamics is chosen large, depending on p, β, t , and ε_1 , and $c_1 > 0$ and the threshold for N depend on the same parameters. Lemma 3.2 there states that, for a fixed initial point and almost every Brownian path, there exists a map ℓ from coefficient tensors to B_N , Lipschitz with a constant depending only on p, β, t , and ε_1 , such that $(1 + \varepsilon_1)\ell(J)$ equals the soft endpoint with probability at least $1 - e^{-c_1 N}$, where the probability is over both the disorder and the Brownian motion. The coefficient-Lipschitz extension uses Kirszbraun's theorem in the proof of Lemma 3.2 of [112]; footnote 2 there gives a constructive choice that depends measurably on the Brownian path.

The construction of [112] is for a fixed initial point. To include the uniform initial point, we choose measurably an orthogonal matrix O_{x_0} such that $O_{x_0}(\sqrt{N}e_1) = x_0$. Rotating the tensor by $(O_{x_0}^T)^{\otimes p}$ and the Brownian path by $O_{x_0}^T$, solving from $\sqrt{N}e_1$, and rotating the output back by O_{x_0} gives the solution from x_0 , because H_N and (3.3) are equivariant under rotations. Conditionally on $X_0 = x_0$, the rotated tensor and Brownian path have the same joint law as the original ones, and the tensor rotation is an isometry of $(\mathbb{R}^N)^{\otimes p}$, which does not change Lipschitz constants.

We apply this to each concatenated path Z_b , whose driving Brownian motion is a function of the seed ζ_N . We obtain maps $\ell_b(J) = \ell_b(J, \zeta_N) \in B_N$, $b \leq K$, which are Lipschitz in J with a constant $L_{\text{Lip},1}$ depending only on p, β, t , and ε_1 . Let $\mathcal{G}_N$ be the event on which H_N satisfies the derivative bounds of [112] and, for every $b \leq K$, the conclusions of both lemmas hold for the branch b . On $\mathcal{G}_N$, we have $\|(1 + \varepsilon_1)\ell_b(J) - Y_b\| \leq \varepsilon_1 \sqrt{N}$ for every $b \leq K$, and a union bound gives $\mathbb{P}(\mathcal{G}_N^c) \leq 3Ke^{-c_1 N}$. The union bound does not

require the branches to be independent. We define

$$\mathcal{A}_{N,T}^\varepsilon(J) = \Pi_{B_N} \left(\frac{1 + \varepsilon_1}{\sqrt{Q_{K,L}}} \frac{1}{K} \sum_{b=1}^K \ell_b(J) \right). \quad (10.25)$$

Since Π_{B_N} is 1-Lipschitz and $Q_{K,L} \geq 1/K$, this is an L_{Lip} -Lipschitz algorithm with seed ζ_N and $L_{\text{Lip}} = 2L_{\text{Lip},1}\sqrt{K}$. On $\mathcal{G}_N$, the distance between $\mathcal{A}_{N,T}^\varepsilon$ and $\Pi_{B_N}(m_{N,T}/\sqrt{Q_{K,L}})$ is at most $\varepsilon_1\sqrt{N}/\sqrt{Q_{K,L}} \leq \varepsilon_1\sqrt{KN} = \varepsilon\sqrt{N}/2$; on $\mathcal{G}_N^c$, it is at most $2\sqrt{N}$. Using $\mathbb{P}(\mathcal{G}_N^c) \leq 3Ke^{-c_1N}$, we find

$$\frac{1}{N} \mathbb{E} \|\mathcal{A}_{N,T}^\varepsilon - \Pi_{B_N}(m_{N,T}/\sqrt{Q_{K,L}})\|^2 \leq \frac{\varepsilon^2}{4} + 12Ke^{-c_1N},$$

which proves (10.20).

Next we compare the three normalizations. Let $\nu_N = \|m_{N,T}\|^2/N$. For $x \geq 0$, we have

$$\left| \sqrt{x/Q_{K,L}} - 1 \right| = \frac{|x - Q_{K,L}|}{\sqrt{Q_{K,L}}(\sqrt{x} + \sqrt{Q_{K,L}})} \leq \frac{|x - Q_{K,L}|}{Q_{K,L}}. \quad (10.26)$$

The scaled distance between $m_{N,T}/\sqrt{Q_{K,L}}$ and its projection onto B_N is $(\sqrt{\nu_N/Q_{K,L}} - 1)_+$. If $m_{N,T} \neq 0$, the scaled distance between the projection and $\hat{m}_{N,T}$ is $(1 - \sqrt{\nu_N/Q_{K,L}})_+$, and both vectors vanish if $m_{N,T} = 0$. By (10.26) and $Q_{K,L} \geq q$, both distances are at most $|\nu_N - Q_{K,L}|/q$. Then (10.17) implies that their mean squares tend to zero as $j \rightarrow \infty$ and then $n \rightarrow \infty$. This completes the proof of part (c).

Finally, we prove part (b). Since $\mathcal{A}_{N,T}^\varepsilon$ is an L_{Lip} -Lipschitz algorithm with seed ζ_N , Lemma 10.2, applied with any ζ_N -measurable direction and with the nonnegative overlap term dropped, gives

$$\limsup_{N \rightarrow \infty} \frac{\mathbb{E} H_N(\mathcal{A}_{N,T}^\varepsilon)}{N} \leq E_{\text{th}}.$$

This is the pure-model threshold, the case $\kappa = 0$ of Theorem 10.1, averaged over the seed. Next we compare energies. Note that $m_{N,T}/\sqrt{Q_{K,L}}$ has norm at most $\sqrt{N}/q$. Using (10.4) with $R = q^{-1/2}$ and the homogeneity $H_N(m_{N,T}/\sqrt{Q_{K,L}}) = Q_{K,L}^{-p/2} H_N(m_{N,T})$, we find

$$Q_{K,L}^{-p/2} \frac{\mathbb{E} H_N(m_{N,T})}{N} \leq \frac{\mathbb{E} H_N(\mathcal{A}_{N,T}^\varepsilon)}{N} + p q^{-(p-1)/2} Q_* \Delta_{N,T},$$

where

$$\Delta_{N,T} = \left(\frac{1}{N} \mathbb{E} \|\mathcal{A}_{N,T}^\varepsilon - m_{N,T}/\sqrt{Q_{K,L}}\|^2 \right)^{1/2}.$$

By (10.20) and the comparison of normalizations, we have

$$\limsup_{n \rightarrow \infty} \limsup_{j \rightarrow \infty} \Delta_{N_j, T_n} \leq \varepsilon.$$

We take $j \rightarrow \infty$ with $T = T_n$ fixed, then $n \rightarrow \infty$, and finally $\varepsilon \downarrow 0$, in the order (10.1). By (10.18), we obtain $Q_{K,L}^{-p/2} h_{K,L} \leq E_{\text{th}}$, which is (10.19). The constants of the algorithms depend on T , but we use them only at a fixed T , before the limit $n \rightarrow \infty$. $\square$

We next record the limits of the fork parameters as $K \rightarrow \infty$ and then $L \rightarrow \infty$. Recall that $d_L = c(2L)$ and h_L are given by (7.17) and (7.18), with the cluster energy e_∞ , and that h is the quantity in (8.4).

Lemma 10.5 (The fork parameters as $K \rightarrow \infty$ and $L \rightarrow \infty$). *Let (c, r, ℓ) be a stationary cluster with plateau $q \geq 0$ and energy e_∞ .*

(i) *For every $L \geq 0$ and $K \geq 1$,*

$$0 \leq Q_{K,L} - d_L \leq \frac{1}{K}, \quad |h_{K,L} - h_L| \leq \frac{\beta p(p-1)}{K}. \quad (10.27)$$

(ii) *As $L \rightarrow \infty$, we have $d_L \rightarrow q$ and $h_L \rightarrow h = e_\infty - \beta\{1 - q^p - pq^{p-1}(1-q)\}$.*

(iii) *If $q > 0$ and $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$, then*

$$\lim_{L \rightarrow \infty} \lim_{K \rightarrow \infty} \frac{h_{K,L}}{Q_{K,L}^{p/2}} = E_{\text{th}}. \quad (10.28)$$

Proof. For part (i), we have $Q_{K,L} - d_L = (1 - c(2L))/K \in [0, 1/K]$. By (7.18) and (10.16),

$$h_{K,L} - h_L = \beta p \int_0^L r(\tau) \left\{ k_{K,L}(\tau)^{p-1} - c(2L - \tau)^{p-1} \right\} d\tau.$$

The integrand is dominated by $r(\tau)$ times a quantity that tends to zero uniformly. Indeed, $k_{K,L}(\tau) - c(2L - \tau) = (c(\tau) - c(2L - \tau))/K$, and both $k_{K,L}(\tau)$ and $c(2L - \tau)$ lie in $[0, 1]$. By the mean value theorem,

$$\left| k_{K,L}(\tau)^{p-1} - c(2L - \tau)^{p-1} \right| \leq \frac{p-1}{K}.$$

Together with $\int_0^L r \leq 1$, this proves (10.27).

Part (ii) is part of Lemma 8.2. For completeness, we recall the computation. Using $r = -c'$, we find that $\int_0^L r(\tau)c(\tau)^{p-1} d\tau = (1 - c(L)^p)/p$. As c is nonincreasing, we have $c(2L - \tau) \in [q, c(L)]$ for $0 \leq \tau \leq L$, and $\int_0^L r = 1 - c(L)$. Thus the integral of $r(\tau)c(2L - \tau)^{p-1}$ over $[0, L]$ lies between $q^{p-1}(1 - c(L))$ and $c(L)^{p-1}(1 - c(L))$. Both bounds tend to $q^{p-1}(1 - q)$. Inserting these two integrals into (7.18), we find that h_L tends to $e_\infty + \beta p q^{p-1}(1 - q) - \beta(1 - q^p) = h$.

For part (iii), part (i) and $d_L \geq q > 0$ imply that $h_{K,L}/Q_{K,L}^{p/2} \rightarrow h_L/d_L^{p/2}$ as $K \rightarrow \infty$, and part (ii) gives $h_L/d_L^{p/2} \rightarrow h/q^{p/2}$ as $L \rightarrow \infty$. By the equality case of (8.5) in Lemma 8.2, the marginality hypothesis implies that $h/q^{p/2} = E_{\text{th}}$. $\square$

Part (iii) uses marginality, which is proved in Lemma 8.3 from part (a) of Proposition 8.1. The proof of that part in Section 10.5 uses only Proposition 10.4 and part (i), and there is no circularity.

10.3 Adaptive disclosure and orthogonality to early queries

The overlap penalty of Lemma 10.2 applies to directions independent of J . At positive times, however, an earlier state X_s depends on J . The goal of this section is to prove that an algorithm that saturates the threshold is almost orthogonal to every subspace explored by a fixed finite number of gradient queries. We condition J on the information revealed by the queries. This is adaptive Gaussian conditioning, in the line of Bolthausen's iterative construction of solutions of the TAP equations [32, Section 3]. Bayati and Montanari used the same method in their analysis of approximate message passing [13, arXiv version 4, Section 3.8, Lemma 11], and they credit it to Bolthausen. The underlying facts on Gaussian vectors and their orthogonal decompositions are classical; see [95, Chapter 9, Theorem 9.1, pp. 127–128, and Theorems 9.8–9.9, pp. 131–132]. We use projection onto the span of the deterministic observations, which also permits linearly dependent observation families. El Alaoui, Montanari, and Sellke conditioned the disorder tensor on adaptive gradient observations in their analysis of message passing for spin glasses [63, arXiv version 1, Appendix A, Lemma A.2, and Theorem 7(a)], and Huang and Sellke extended this conditioning to multi-species spherical models [93, arXiv version 1, Appendix A, Lemma A.2, and Theorem 3(a)]. The displayed inverse-Gram formulas in these AMP analyses use rank and conditioning hypotheses. Our projection-based disclosure below also covers dependent queries. Here we disclose all tensor blocks needed to determine gradients throughout the explored span. We combine this disclosure with the external-field threshold to prove forgetting of a fixed finite history.

Outer parameters. Let θ be a finite list of outer parameters with a prescribed order of limits; in Section 10.4, $\theta = (n, \varepsilon, K, L)$, in the order (10.1). For nonnegative quantities $F_{N,\theta}$, we write $\lim_{\theta} \limsup_N F_{N,\theta} = 0$ if the upper limit in N , followed by the upper limits in the outer parameters in their prescribed order, vanishes. We do not assume that any intermediate limit exists.

Standing hypotheses. For each N and θ , let $\zeta_{N,\theta}$ be a random variable independent of J , which contains all auxiliary randomness used both by the algorithms and by the disclosure procedure below. Its measurable space and law may depend on θ ; below we write ζ_N with θ suppressed. For each N and θ , let $\mathcal{A}_{N,\theta} = \mathcal{A}_{N,\theta}(J, \zeta_N)$ be an $L_{\text{Lip},\theta}$ -Lipschitz algorithm with seed ζ_N . The constant $L_{\text{Lip},\theta}$ is deterministic and independent

of N , but it may be unbounded in the outer limits. We define the energy deficit

$$\delta_\theta = \max \left\{ 0, E_{\text{th}} - \liminf_{N \rightarrow \infty} \mathbb{E} \frac{H_N(\mathcal{A}_{N,\theta})}{N} \right\}. \quad (10.29)$$

Lemma 10.2 implies that $\limsup_N \mathbb{E} H_N(\mathcal{A}_{N,\theta})/N \leq E_{\text{th}}$. The hypothesis $\lim_\theta \delta_\theta = 0$ says exactly that the algorithms saturate the threshold in mean.

For a subspace $W \subset \mathbb{R}^N$, we write

$$\mathcal{T}(W) = (W^\perp)^{\otimes p} \subset (\mathbb{R}^N)^{\otimes p},$$

and P_V for the orthogonal projection onto a subspace V of $\mathbb{R}^N$ or of $(\mathbb{R}^N)^{\otimes p}$. If $W \subset W'$, then $\mathcal{T}(W') \subset \mathcal{T}(W)$, and we write $\mathcal{T}(W) \ominus \mathcal{T}(W')$ for the orthogonal complement of $\mathcal{T}(W')$ in $\mathcal{T}(W)$. We choose orthonormal bases of measurable subspaces measurably, by Gram–Schmidt in a fixed order with deterministic fallbacks.

Lemma 10.6 (Adaptive tensor disclosure). *Let $W_0 \subset \mathbb{R}^N$ be a ζ_N -measurable subspace, and let $\mathcal{F}_0$ be the σ -algebra generated by ζ_N and $P_{\mathcal{T}(W_0)^\perp} J$. For $j \geq 0$, given W_j and $\mathcal{F}_j$, let u_j be an $\mathcal{F}_j$ -measurable vector that is either zero or a unit vector in $W_j^\perp$, and set*

$$W_{j+1} = W_j + \text{span}(u_j), \quad \mathcal{F}_{j+1} = \sigma\left(\mathcal{F}_j, P_{\mathcal{T}(W_j) \ominus \mathcal{T}(W_{j+1})} J\right).$$

Then, for every j :

- (i) $P_{\mathcal{T}(W_j)^\perp} J$ is $\mathcal{F}_j$ -measurable, and conditionally on $\mathcal{F}_j$, the tensor $P_{\mathcal{T}(W_j)} J$ is standard Gaussian on $\mathcal{T}(W_j)$;
- (ii) if $x_j \in W_j$ is $\mathcal{F}_j$ -measurable, then $\nabla H_N(x_j)$ is $\mathcal{F}_j$ -measurable. In particular, u_j may be chosen as the normalized projection of $\nabla H_N(x_j)$ onto $W_j^\perp$, and as zero if that projection vanishes.

Proof. The conditional Gaussian assertions are understood almost surely with respect to the disclosed history. A zero query discloses the zero tensor block and leaves the residual law unchanged. We prove part (i) by induction on j . For $j = 0$, conditionally on ζ_N , the tensor J is standard Gaussian, because it is independent of ζ_N , and the subspaces $\mathcal{T}(W_0)$ and $\mathcal{T}(W_0)^\perp$ are fixed. The projections of a standard Gaussian vector onto two complementary orthogonal subspaces are independent standard Gaussian vectors on them. Revealing $P_{\mathcal{T}(W_0)^\perp} J$ leaves $P_{\mathcal{T}(W_0)} J$ standard Gaussian on $\mathcal{T}(W_0)$.

Suppose that part (i) holds for j . Conditionally on $\mathcal{F}_j$, the vector u_j and the subspaces in the orthogonal decomposition

$$\mathcal{T}(W_j) = \mathcal{T}(W_{j+1}) \oplus (\mathcal{T}(W_j) \ominus \mathcal{T}(W_{j+1}))$$

are fixed, and $P_{\mathcal{T}(W_j)} J$ is standard Gaussian on $\mathcal{T}(W_j)$. Its projections onto the two summands are independent standard Gaussian vectors. Revealing the second leaves the

first, which is $P_{\mathcal{T}(W_{j+1})}J$, standard Gaussian on $\mathcal{T}(W_{j+1})$. Since $\mathcal{T}(W_{j+1})^\perp = \mathcal{T}(W_j)^\perp \oplus (\mathcal{T}(W_j) \ominus \mathcal{T}(W_{j+1}))$, the projection $P_{\mathcal{T}(W_{j+1})^\perp}J$ is $\mathcal{F}_{j+1}$ -measurable. This completes the induction. At each step, we choose the new direction before we reveal the corresponding part of J ; this is what makes the induction work, and we never condition on a subspace that depends on undisclosed randomness.

For part (ii), let $x \in W_j$ and $y \in \mathbb{R}^N$. We have

$$\langle \nabla H_N(x), y \rangle = N^{-(p-1)/2} \left\langle J, \sum_{i=1}^p x^{\otimes(i-1)} \otimes y \otimes x^{\otimes(p-i)} \right\rangle.$$

Since $p \geq 2$, each tensor in the sum has at least one factor $x \in W_j$. It is orthogonal to every product tensor with all factors in $W_j^\perp$, and these product tensors span $\mathcal{T}(W_j)$. The right side is equal to $N^{-(p-1)/2}$ times the inner product of $P_{\mathcal{T}(W_j)^\perp}J$ with that tensor. By part (i), it is $\mathcal{F}_j$ -measurable when $x = x_j$ is. $\square$

Proposition 10.7 (Orthogonality to every fixed finite query span). *Assume the standing hypotheses and $\lim_\theta \delta_\theta = 0$. Fix integers $k_0 \geq 0$ and $M \geq 0$. For each N and θ , let $W_0 \subset W_1 \subset \dots \subset W_M$ be subspaces constructed as in Lemma 10.6 from the seed ζ_N , with $\dim W_0 \leq k_0$. Then*

$$\lim_\theta \limsup_{N \rightarrow \infty} \mathbb{E} \frac{\|P_{W_M} \mathcal{A}_{N,\theta}\|^2}{N} = 0. \quad (10.30)$$

In particular, if $Y_N \in W_M \cap B_N$ is a random vector, then $\lim_\theta \limsup_N \mathbb{E} |\langle \mathcal{A}_{N,\theta}, Y_N \rangle / N|^2 = 0$.

Proof. For $0 \leq j \leq M$, we define

$$\psi_{j,\theta} = \limsup_{N \rightarrow \infty} \mathbb{E} \frac{\|P_{W_j} \mathcal{A}_{N,\theta}\|^2}{N}.$$

We bound $\psi_{0,\theta}$ directly and then show by induction that $\psi_{j+1,\theta}$ is controlled by $\psi_{j,\theta}$ and δ_θ . We use the constants of the threshold only at fixed θ , before any outer limit, and the argument does not require $L_{\text{Lip},\theta}$ to stay bounded.

Step 1: the initial span. Let $e_1, \dots, e_{k_0}$ be a ζ_N -measurable orthonormal basis of W_0 , completed by zero vectors if $\dim W_0 < k_0$. We obtain such a frame by Gram–Schmidt on the projected coordinate vectors, keeping nonzero columns in a fixed order and then padding by zero columns. For $N \geq 1$, fix a deterministic unit vector $e_{\text{ref}} \in \mathbb{R}^N$, and for every i set $w_i = \sqrt{N}e_i$ on $\{e_i \neq 0\}$ and $w_i = \sqrt{N}e_{\text{ref}}$ on $\{e_i = 0\}$. This gives a ζ_N -measurable direction of norm $\sqrt{N}$ on every seed event. We apply Lemma 10.2 with $L_{\text{Lip}} = L_{\text{Lip},\theta}$ and direction w_i , and we drop its nonnegative overlap penalty on $\{e_i = 0\}$. Since the pairing with e_i vanishes there, this yields

$$\beta_d \limsup_{N \rightarrow \infty} \mathbb{E} \left| \frac{\langle \mathcal{A}_{N,\theta}, e_i \rangle}{\sqrt{N}} \right|^p \leq E_{\text{th}} - \liminf_{N \rightarrow \infty} \mathbb{E} \frac{H_N(\mathcal{A}_{N,\theta})}{N} \leq \delta_\theta.$$

For each i , by Jensen's inequality and $p \geq 2$, we have $\mathbb{E}|\langle \mathcal{A}_{N,\theta}, e_i \rangle|^2 / N \leq (\mathbb{E}|\langle \mathcal{A}_{N,\theta}, e_i \rangle| / \sqrt{N})^p$. Summing over i , we obtain

$$\psi_{0,\theta} \leq k_0 \left(\frac{\delta_\theta}{\beta_d} \right)^{2/p}. \quad (10.31)$$

Step 2: the energy of the residual. Fix $j < M$, write $W = W_j$, and let $\mathcal{A}^\perp = P_{W^\perp} \mathcal{A}_{N,\theta}$. Both $\mathcal{A}_{N,\theta}$ and $\mathcal{A}^\perp$ lie in B_N . Using (10.3) and the mean value theorem, we have

$$\frac{|H_N(\mathcal{A}_{N,\theta}) - H_N(\mathcal{A}^\perp)|}{N} \leq pQ_N \frac{\|P_W \mathcal{A}_{N,\theta}\|}{\sqrt{N}}.$$

By the Cauchy–Schwarz inequality, the expectation of the right side has upper limit at most $pQ_* \sqrt{\psi_{j,\theta}}$. Combining this with the definition (10.29) of δ_θ , we find

$$\liminf_{N \rightarrow \infty} \mathbb{E} \frac{H_N(\mathcal{A}^\perp)}{N} \geq E_{\text{th}} - \delta_\theta - pQ_* \sqrt{\psi_{j,\theta}}. \quad (10.32)$$

Step 3: the residual model. Since k_0 and M are fixed before the dimension limit, we may restrict to $N > k_0 + M$. Every residual dimension below is then positive. We now show that $\mathcal{A}^\perp$ is the output of a Lipschitz algorithm for a pure p -spin model in dimension $N' = N - \dim W$, and that its energy has the same normalization up to a factor tending to one. We condition on $\mathcal{F}_j$. Then W and u_j are fixed, and we choose an $\mathcal{F}_j$ -measurable orthonormal basis $f_1, \dots, f_{N'}$ of $W^\perp$. Let $J' \in (\mathbb{R}^{N'})^{\otimes p}$ be the coordinates of $P_{\mathcal{T}(W)} J$ in the basis $(f_{i_1} \otimes \dots \otimes f_{i_p})$. By Lemma 10.6(i), J' is a standard Gaussian vector in the N'^p -dimensional space $(\mathbb{R}^{N'})^{\otimes p}$, conditionally on $\mathcal{F}_j$. Let $y' \in \mathbb{R}^{N'}$ be the coordinates of $\mathcal{A}^\perp$ in the basis (f_i) , and set $x = \sqrt{N'/N} y'$. Then $\|x\| \leq \sqrt{N'}$, that is, $x \in B_{N'}$. Since $(\mathcal{A}^\perp)^{\otimes p} \in \mathcal{T}(W)$, only $P_{\mathcal{T}(W)} J$ contributes to $H_N(\mathcal{A}^\perp)$, and a change of normalization yields

$$\frac{H_N(\mathcal{A}^\perp)}{N} = N^{-(p+1)/2} \langle J', y'^{\otimes p} \rangle = \sqrt{\frac{N'}{N}} \frac{H'_{N'}(x)}{N'}, \quad (10.33)$$

where $H'_{N'}(x) = N'^{-(p-1)/2} \langle J', x^{\otimes p} \rangle$ is the pure p -spin Hamiltonian in dimension N' with coefficient tensor J' . Indeed, $\langle J', y'^{\otimes p} \rangle = (N/N')^{p/2} \langle J', x^{\otimes p} \rangle$, and $N^{-(p+1)/2} (N/N')^{p/2} N'^{(p-1)/2} = (N'/N)^{1/2} N'^{-1}$. Moreover, J is an affine isometric function of J' given $\mathcal{F}_j$, and $\mathcal{A}^\perp \mapsto x$ is a contraction. Therefore, conditionally on $\mathcal{F}_j$, the map $J' \mapsto x$ is $L_{\text{Lip},\theta}$ -Lipschitz with values in $B_{N'}$. Let $u' \in \mathbb{R}^{N'}$ be the coordinates of u_j when $u_j \neq 0$, and the first coordinate vector otherwise. The direction $w' = \sqrt{N'} u'$ has norm $\sqrt{N'}$, and when $u_j \neq 0$,

$$\frac{\langle w', x \rangle}{N'} = \frac{\langle u', y' \rangle}{\sqrt{N}} = \frac{\langle u_j, \mathcal{A}_{N,\theta} \rangle}{\sqrt{N}},$$

where we used $u_j \in W^\perp$ in the last equality.

Step 4: the penalty in the residual model. Let $\alpha_j = \langle u_j, \mathcal{A}_{N,\theta} \rangle / \sqrt{N}$, which is zero if $u_j = 0$. The conditional law of J' given $\mathcal{F}_j$ is standard Gaussian, and on each event $\{N' = n'\}$ the tensor J' is independent of $\mathcal{F}_j$. We may therefore apply (10.7) in dimension N' , with $\mathcal{F}_j$ in the role of the seed, to the algorithm $J' \mapsto x$ and the direction w' . If $u_j \neq 0$, its overlap term is $\beta_d |\alpha_j|^p$; if $u_j = 0$, we drop its nonnegative overlap term. Since $N' \geq N - k_0 - j$ and $\phi_N \geq 0$, we obtain

$$\mathbb{E} \left[\frac{H'_{N'}(x)}{N'} + \beta_d |\alpha_j|^p \middle| \mathcal{F}_j \right] \leq E_{\text{th}} + \sup_{N'' \geq N - k_0 - j} \phi_{N''}(p, L_{\text{Lip},\theta}).$$

We use (10.33) to write

$$\frac{H_N(\mathcal{A}^\perp)}{N} + \beta_d |\alpha_j|^p = \sqrt{\frac{N'}{N}} \left(\frac{H'_{N'}(x)}{N'} + \beta_d |\alpha_j|^p \right) + \left(1 - \sqrt{\frac{N'}{N}} \right) \beta_d |\alpha_j|^p.$$

Here $\sqrt{N'/N} \leq 1$, $0 \leq 1 - \sqrt{N'/N} \leq (k_0 + j)/N$, and $|\alpha_j| \leq 1$. Taking expectations and letting $N \rightarrow \infty$, we find

$$\limsup_{N \rightarrow \infty} \left\{ \mathbb{E} \frac{H_N(\mathcal{A}^\perp)}{N} + \beta_d \mathbb{E} |\alpha_j|^p \right\} \leq E_{\text{th}}.$$

Combining this with (10.32), we obtain $\beta_d \limsup_N \mathbb{E} |\alpha_j|^p \leq \delta_\theta + pQ_* \sqrt{\psi_{j,\theta}}$. By Jensen's inequality as in Step 1, and since $\|P_{W_{j+1}} \mathcal{A}_{N,\theta}\|^2 = \|P_{W_j} \mathcal{A}_{N,\theta}\|^2 + N|\alpha_j|^2$, we conclude that

$$\psi_{j+1,\theta} \leq \psi_{j,\theta} + \left(\frac{\delta_\theta + pQ_* \sqrt{\psi_{j,\theta}}}{\beta_d} \right)^{2/p}. \quad (10.34)$$

Step 5: the outer limits. The right sides of (10.31) and (10.34) are continuous nondecreasing functions of $(\delta_\theta, \psi_{j,\theta})$ that vanish at $(0,0)$. Upper limits pass through such functions. Taking the outer limits in their prescribed order and inducting on $j \leq M$, we find $\lim_\theta \psi_{j,\theta} = 0$ for every $j \leq M$. This is (10.30). Finally, if $Y_N \in W_M \cap B_N$, then

$$\left| \frac{\langle \mathcal{A}_{N,\theta}, Y_N \rangle}{N} \right|^2 = \left| \frac{\langle P_{W_M} \mathcal{A}_{N,\theta}, Y_N \rangle}{N} \right|^2 \leq \frac{\|P_{W_M} \mathcal{A}_{N,\theta}\|^2}{N},$$

which proves the last assertion. $\square$

10.4 Transfer to a fixed waiting-time state

We now combine the forks with finite-query orthogonality on their fixed product probability spaces. The hypotheses of the next proposition are those of Proposition 8.1(b).

Under them, the translated family (C_T, R_T, λ_T) converges along every sequence $T_n \rightarrow \infty$ to one stationary cluster (c, r, ℓ) . Indeed, by the precompactness in Proposition 5.7, every subsequence of (T_n) has a further subsequence along which the family converges, and the limit is a stationary cluster whose kernels are (c, r) . Its multiplier is determined by r : setting $\tau = 0$ in (5.25) and using $r(0) = 1$, we find $\ell = -r'(0)$. All subsequential limits are therefore equal. In particular, Proposition 10.4 applies along every sequence $T_n \rightarrow \infty$.

Proposition 10.8 (Fixed-waiting-time decorrelation). *Fix $p \geq 3$ and $\beta > 0$. Suppose that the translated kernels converge along every sequence $T \rightarrow \infty$ to one stationary profile (c, r) , and that its plateau q is positive and satisfies $\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1$. Then*

$$\lim_{t \rightarrow \infty} C(t, s) = 0 \quad \text{for every fixed } s \geq 0.$$

Proof. The proof is by contradiction. Recall from (3.15) that $C \geq 0$. Suppose that, for some $s \geq 0$, there exist $\gamma > 0$ and $t_n \rightarrow \infty$ such that $C(t_n, s) \geq \gamma$ for all n . The proof has three steps. We first express $C(t_n, s)$ through the center of a fork that ends at t_n . Then we show that the Lipschitz algorithms approximating the normalized center saturate the threshold in the order (10.1). Finally, on the fixed fork product law, we approximate X_s by a finite-query Euler vector and use Proposition 10.7 to force their overlap with X_s to vanish, which contradicts the first step.

Step 1: the overlap of the center. Fix $K \geq 1$ and $L \geq 0$, and set $T_n = \max\{s+1, t_n - L\}$ for every n . Then $T_n \rightarrow \infty$, $T_n \geq s$, and for all sufficiently large n we have $T_n + L = t_n$. Let $m_{N,n}$ be the center of the fork at time T_n with K branches of length L . For these sufficiently large n , the branches end at time t_n . The time s lies in the common past, and each pair (Y_b, X_s) has the law of (X_{t_n}, X_s) . This implies that

$$\mathbb{E} \frac{\langle m_{N,n}, X_s \rangle}{N} = \mathbb{E} C_N(t_n, s), \quad (10.35)$$

and by Proposition 3.3, the right side tends to $C(t_n, s)$ as $N \rightarrow \infty$. In the limit in n , the endpoint of the fork is the time t_n itself, so we do not need any regularity of $C(\cdot, s)$ in its first variable.

Step 2: saturation of the threshold. Let $\theta = (n, \varepsilon, K, L)$, where $\varepsilon = 1/(m+1)$ and $K = k+1$ for $m, k \in \mathbb{N}$, with the order of (10.1), and let $\mathcal{A}_{N,\theta} = \mathcal{A}_{N,T_n}^\varepsilon$ be the algorithms of Proposition 10.4(c) for this fork. They satisfy the standing hypotheses of Section 10.3 on the fixed fork product law of Section 10.2, with seed ζ_N consisting of X_0 , the K full future Wiener paths, and the full trunk Wiener path B , jointly independent of J , and with a constant $L_{\text{Lip},\theta}$ depending only on p, β , the fixed waiting time s , the horizon $T_n + L$, K , and ε . We write $Q = Q_{K,L}$, and recall that $q \leq Q \leq 1$. Let

$$\Delta_{N,\theta} = \left(\frac{1}{N} \mathbb{E} \|\mathcal{A}_{N,\theta} - m_{N,n} / \sqrt{Q}\|^2 \right)^{1/2}.$$

By (10.20) and the comparison of normalizations in Proposition 10.4(c), we have

$$\limsup_{n \rightarrow \infty} \limsup_{j \rightarrow \infty} \Delta_{N_j, \theta} \leq \varepsilon. \quad (10.36)$$

The vector $m_{N,n}/\sqrt{Q}$ has norm at most $\sqrt{N/q}$. Using (10.4) with $R = q^{-1/2}$ and the homogeneity of H_N , we find

$$\mathbb{E} \frac{H_N(\mathcal{A}_{N,\theta})}{N} \geq Q^{-p/2} \mathbb{E} \frac{H_N(m_{N,n})}{N} - C_q \Delta_{N,\theta}, \quad C_q = p q^{-(p-1)/2} Q_*.$$

The constant C_q does not depend on n , ε , K , or L ; this is where we use $q > 0$. Then (10.18) implies that $\lim_n \liminf_j Q^{-p/2} \mathbb{E} H_{N_j}(m_{N_j,n})/N_j = Q^{-p/2} h_{K,L}$. Together with (10.36), this gives

$$\limsup_{n \rightarrow \infty} \delta_\theta \leq \max \left\{ 0, E_{\text{th}} - \frac{h_{K,L}}{Q_{K,L}^{p/2}} + C_q \varepsilon \right\}.$$

We let $\varepsilon \downarrow 0$, then $K \rightarrow \infty$, then $L \rightarrow \infty$. By (10.28), which applies because q is marginal, we conclude that $\lim_\theta \delta_\theta = 0$ in the order (10.1).

Step 3: orthogonality on the fork product law and the contradiction. We use the algorithms of Step 2 on precisely the fork seed and disorder product space just described, and take the limits in precisely the order (10.1), with $\varepsilon = 1/(m+1)$. All disclosure histories below include the full fork seed. If $s = 0$, then X_0 lies in the seed-measurable subspace $W_0 = \text{span}(X_0)$, and Proposition 10.7 with $k_0 = 1$ and no disclosure steps gives zero limiting squared overlap. For $s > 0$, fix $M \geq 1$ before any limit. We construct the Euler–Maruyama approximation [91, Sections 4–5] of this fork’s trunk state X_s and prove the required normalized error directly.

The coefficients. Let $\Psi_N = \Pi_{B_N}$ be the metric projection onto the ball of radius $\sqrt{N}$. We use $P_y = I - yy^\top/N$ for every $y \in \mathbb{R}^N$, as in Section 3. For $x \in \mathbb{R}^N$, we define

$$b_N(x) = \beta P_{\Psi_N(x)} \nabla H_N(\Psi_N(x)) - \frac{N-1}{N} \Psi_N(x), \quad \sigma_N(x) = \sqrt{2} P_{\Psi_N(x)}.$$

Since $\Psi_N(x) = x$ on S_N , these are the coefficients of (3.3) on S_N , and

$$X_t = X_0 + \int_0^t b_N(X_r) \, dr + \int_0^t \sigma_N(X_r) \, dB_r.$$

Let $G_N = \{Q_N \leq A_p\}$, and let $y = \Psi_N(x)$ and $z = \Psi_N(x')$, so that $\|y\|, \|z\| \leq \sqrt{N}$ and $\|y - z\| \leq \|x - x'\|$. The matrix P_y has eigenvalues 1 and $1 - \|y\|^2/N \in [0, 1]$, which gives $\|P_y\|_{\text{op}} \leq 1$ and $\|P_y\|_{\text{HS}}^2 \leq N$. Moreover,

$$\|P_y - P_z\|_{\text{HS}} = \frac{\|yy^\top - zz^\top\|_{\text{HS}}}{N} \leq \frac{(\|y\| + \|z\|)\|y - z\|}{N} \leq \frac{2\|x - x'\|}{\sqrt{N}}.$$

On G_N , (10.2) with $R = 2$ gives $\|\nabla H_N(y)\| \leq p2^{p-1}A_p\sqrt{N}$ and $\|\nabla H_N(y) - \nabla H_N(z)\| \leq p(p-1)2^{p-2}A_p\|y - z\|$. Writing $P_y\nabla H_N(y) - P_z\nabla H_N(z) = P_y(\nabla H_N(y) - \nabla H_N(z)) + (P_y - P_z)\nabla H_N(z)$, we obtain, on G_N and for all $x, x' \in \mathbb{R}^N$,

$$\begin{aligned} \|b_N(x) - b_N(x')\| &\leq K_b\|x - x'\|, & \|\sigma_N(x) - \sigma_N(x')\|_{\text{HS}} &\leq \frac{2\sqrt{2}}{\sqrt{N}}\|x - x'\|, \\ \|b_N(x)\|^2 &\leq K'_b N, & \|\sigma_N(x)\|_{\text{HS}}^2 &\leq 2N, \end{aligned} \quad (10.37)$$

with $K_b = 1 + \beta p A_p 2^{p-2}(3p + 5)$ and $K'_b = (3\beta p 2^{p-1}A_p + 2)^2$. The Hilbert–Schmidt bound on $\sigma_N(x) - \sigma_N(x')$ is the reason that the error below does not grow with N : a bound of order one for each matrix entry would lose a factor of N .

The Euler–Maruyama scheme. Let $t_i = is/M$ for $0 \leq i \leq M$. We define $Z_0 = X_0$ and

$$Z_{i+1} = Z_i + \frac{s}{M} b_N(Z_i) + \sigma_N(Z_i)(B_{t_{i+1}} - B_{t_i}), \quad 0 \leq i < M.$$

For $r \in [t_i, t_{i+1})$, let $r_- = t_i$, and let $Z_r = Z_i + (r - t_i)b_N(Z_i) + \sigma_N(Z_i)(B_r - B_{t_i})$, so that

$$Z_t = X_0 + \int_0^t b_N(Z_{r_-}) \, dr + \int_0^t \sigma_N(Z_{r_-}) \, dB_r$$

and $Z_s = Z_M$. We set $Y_{N,M} = \Pi_{B_N}(Z_M)$.

The error on G_N . Fix $J \in G_N$ and X_0 , and let $\mathbb{E}_B$ denote expectation over the Brownian motion. Let $u(t) = \mathbb{E}_B\|X_t - Z_t\|^2/N$ and $v(r) = \mathbb{E}_B\|Z_r - Z_{r_-}\|^2/N$. Since Z_{r_-} is independent of $B_r - B_{r_-}$, the Itô isometry and (10.37) give

$$v(r) \leq 2\left(\frac{s}{M}\right)^2 K'_b + 2\frac{s}{M}\frac{2N}{N} \leq C_1\frac{s}{M}, \quad C_1 = 2sK'_b + 4.$$

Subtracting the integral equations for X_t and Z_t , and using the Cauchy–Schwarz inequality for the drift and the Itô isometry for the noise, we find, for $t \leq s$,

$$u(t) \leq 2sK_b^2 \int_0^t \mathbb{E}_B \frac{\|X_r - Z_{r_-}\|^2}{N} \, dr + \frac{16}{N} \int_0^t \mathbb{E}_B \frac{\|X_r - Z_{r_-}\|^2}{N} \, dr.$$

Inserting the bound $\mathbb{E}_B\|X_r - Z_{r_-}\|^2/N \leq 2u(r) + 2v(r)$, we obtain $u(t) \leq C_2 \int_0^t u(r) \, dr + C_2 C_1 s^2/M$ with $C_2 = 4sK_b^2 + 32$. By Gronwall's inequality,

$$u(s) \leq C_1 C_2 s e^{C_2 s} \frac{s}{M}.$$

The projection Π_{B_N} is 1-Lipschitz and fixes $X_s \in S_N$, which gives $\mathbb{E}_B\|X_s - Y_{N,M}\|^2/N \leq u(s)$ on G_N . On G_N^c , we have $\|X_s - Y_{N,M}\|^2/N \leq 4$, and $\mathbb{P}(G_N^c) \leq e^{-c_p N}$. Thus, with $C = C_1 C_2 s e^{C_2 s}$,

$$\limsup_{N \rightarrow \infty} \mathbb{E} \frac{\|X_s - Y_{N,M}\|^2}{N} \leq C \frac{s}{M}. \quad (10.38)$$

The expectation here is on the fixed fork product law. Its projection onto (J, X_0, B) has exactly the original disorder, uniform initial and Wiener product law, so the error bound is independent of n, ε, K , and L .

The query span. Let W_0 be the span of X_0 and the increments $B_{t_{i+1}} - B_{t_i}$, $0 \leq i < M$. It is measurable with respect to the full fork seed ζ_N and has dimension at most $M + 1$. Suppose that $Z_i \in W_i$ and that Z_i is $\mathcal{F}_i$ -measurable, where $\mathcal{F}_i$ is the disclosure history generated using this full fork seed, which holds for $i = 0$. The point $\Psi_N(Z_i)$ is a multiple of Z_i and lies in W_i . By Lemma 10.6(ii), the gradient $g_i = \nabla H_N(\Psi_N(Z_i))$ is $\mathcal{F}_i$ -measurable. We choose u_i as the normalized projection of g_i onto $W_i^\perp$ (zero if it vanishes), so that $g_i \in W_{i+1}$. Since $P_y w = w - y \langle y, w \rangle / N$, the step from Z_i to Z_{i+1} is a linear combination of Z_i , $\Psi_N(Z_i)$, g_i , and $B_{t_{i+1}} - B_{t_i}$, with coefficients that are $\mathcal{F}_i$ -measurable. This shows that $Z_{i+1} \in W_{i+1}$ and that Z_{i+1} is $\mathcal{F}_{i+1}$ -measurable. After M steps, $Z_M \in W_M$, and $Y_{N,M}$, a multiple of Z_M , lies in W_M . Each step adds at most one dimension, which gives $\dim W_M \leq 2M + 1$.

Passing from the Euler vector to the waiting state. By Step 2, Proposition 10.7 applies to these particular full-fork disclosure histories, with $k_0 = M + 1$ and M queries. Hence the squared overlap of $\mathcal{A}_{N,\theta}$ with $Y_{N,M}$ vanishes after the dimension limit and then the limits in $n, \varepsilon = 1/(m + 1)$, K , and L . Since $\|\mathcal{A}_{N,\theta}\| \leq \sqrt{N}$, we have

$$\mathbb{E} \left| \frac{\langle \mathcal{A}_{N,\theta}, X_s \rangle}{N} \right|^2 \leq 2\mathbb{E} \left| \frac{\langle \mathcal{A}_{N,\theta}, Y_{N,M} \rangle}{N} \right|^2 + 2\mathbb{E} \frac{\|X_s - Y_{N,M}\|^2}{N}.$$

By (10.38), the nested upper limit of the left side is at most $2Cs/M$. It does not depend on M . Sending $M \rightarrow \infty$ last (and using the $s = 0$ argument above when needed), we obtain

$$\lim_{\theta} \limsup_{N \rightarrow \infty} \mathbb{E} \left| \frac{\langle \mathcal{A}_{N,\theta}, X_s \rangle}{N} \right|^2 = 0. \quad (10.39)$$

On the other hand, by (10.21) with $V_N = X_s$ and by (10.35), we have

$$\frac{\mathbb{E} C_N(t_n, s)}{\sqrt{Q}} \leq \left(\mathbb{E} \left| \frac{\langle \mathcal{A}_{N,\theta}, X_s \rangle}{N} \right|^2 \right)^{1/2} + \Delta_{N,\theta}.$$

As $N = N_j \rightarrow \infty$, the left side tends to $C(t_n, s)/\sqrt{Q}$, which is at least γ because $Q \leq 1$. We conclude that, for every fixed ε, K , and L and all sufficiently large n ,

$$\gamma \leq \limsup_{j \rightarrow \infty} \left(\mathbb{E} \left| \frac{\langle \mathcal{A}_{N_j,\theta}, X_s \rangle}{N_j} \right|^2 \right)^{1/2} + \limsup_{j \rightarrow \infty} \Delta_{N_j,\theta}.$$

We take the upper limit as $n \rightarrow \infty$, then let $\varepsilon \downarrow 0$, $K \rightarrow \infty$, and $L \rightarrow \infty$, in the order (10.1). By (10.39), the first term tends to zero, and by (10.36), the second is at most ε after the limit in n . Hence $\gamma \leq 0$, which contradicts $\gamma > 0$. This completes the proof. $\square$

10.5 Proof of Proposition 8.1

We now prove both parts of Proposition 8.1 for every $p \geq 3$.

Proof of Proposition 8.1(a). Let $T_n \rightarrow \infty$ be a sequence along which the translated kernels converge to a stationary cluster (c, r, ℓ) with plateau $q > 0$, and fix $L \geq 0$. For every $K \geq 1$, Proposition 10.4 applies along T_n , and (10.19) gives

$$h_{K,L} \leq E_{\text{th}} Q_{K,L}^{p/2}.$$

By Lemma 10.5(i), we have $|Q_{K,L} - d_L| \leq 1/K$ and $|h_{K,L} - h_L| \leq \beta p(p-1)/K$. The second bound is the dominated convergence argument made explicit: the difference $h_{K,L} - h_L$ is βp times the integral of $r(\tau)$ times $k_{K,L}(\tau)^{p-1} - c(2L - \tau)^{p-1}$, which is at most $(p-1)/K$ in absolute value, and $\int_0^L r \leq 1$. Letting $K \rightarrow \infty$ and using the continuity of $x \mapsto x^{p/2}$, we obtain

$$h_L \leq E_{\text{th}} d_L^{p/2},$$

which is (8.1). $\square$

Letting $L \rightarrow \infty$ in (8.1) and using Lemma 10.5(ii), we obtain

$$h \leq E_{\text{th}} q^{p/2}, \tag{10.40}$$

where h is given by (8.4). This is the form in which part (a) enters the proof of Lemma 8.3. It holds for every stationary cluster with positive plateau, and its proof does not use marginality.

Proof of Proposition 8.1(b). The hypotheses of part (b) are those of Proposition 10.8, which gives $C(t, s) \rightarrow 0$ as $t \rightarrow \infty$ for each fixed s . It remains to make the convergence uniform in $s \in [0, S]$. Let $k \geq 1$ be an integer and let $s_i = i/k$ for $0 \leq i \leq \lceil kS \rceil$. Every $s \in [0, S]$ is within $1/k$ of some s_i , and by the correlation modulus (4.19) in the second variable,

$$|C(t, s) - C(t, s_i)| \leq \frac{D_{p,\beta}}{k} + \sqrt{\frac{2}{k}}$$

for all t . Since $C \geq 0$, we obtain

$$\sup_{0 \leq s \leq S} |C(t, s)| \leq \max_i C(t, s_i) + \frac{D_{p,\beta}}{k} + \sqrt{\frac{2}{k}}.$$

The maximum is over finitely many points, and it tends to zero as $t \rightarrow \infty$ by Proposition 10.8. Letting $t \rightarrow \infty$ and then $k \rightarrow \infty$ proves (8.2). $\square$

PART III

Response and the fluctuation–dissipation
relation

Localization of the response and of correlation derivatives

In this section, we study complete rows of the response and of both partial derivatives of the correlation and show that they converge to their stationary limits. Our main result, Theorem 11.1, states that for every $p \geq 3$ and $\beta > 0$, the lag row $v \mapsto R(t, t - v)$, extended by zero, converges to r in $L^2(0, \infty)$. From it we deduce uniform convergence of response rows (Corollary 11.2). We also show that the earlier derivative $\partial_s C(t, s)$ converges in the same L^2 sense, so that the squared backward FDT defect vanishes on whole rows (Theorem 11.5). Further, we bound the later derivative $\partial_t C(t, s)$ far from the diagonal for every $p \geq 3$ (Theorem 11.4), prove a smoothing estimate in the earlier time (Corollary 11.6), and prove that the thermal part of the Green covariance identity converges uniformly on late triangles (Corollary 11.7).

We use structural identities from Section 4 (Propositions 4.2, 4.3, 4.5, and 4.6), the Fisher bound of Theorem 5.3, convergence and forgetting from Section 8 (Proposition 8.4 and Theorem 8.8), and the causal tree realization (Propositions 7.1 and 7.5, Lemma 7.3, and Corollary 7.7). When the plateau is positive, the proof of Theorem 11.1 also uses the cubic tree ceiling of Section 9 if $p = 3$, and the finite forks of Section 10 for general p . We give the second argument at the end of this section, in Section 11.5. The results of this section are used in Section 12, which relies on Theorem 11.5, and in Sections 14 and 15, which use Theorems 11.1 and 11.4 and Corollary 11.2, together with the Fisher bound (11.14). Combined with Theorem 5.3, they prove (1.7), (1.8), and (1.9) in Theorem 1.2.

All convergence statements in this section are in the L^2 norm or in the uniform norm on a row. They do not control the integral of a row over its remote part. In Section 12, we prove that the backward FDT defect, although small in L^2 , keeps a positive integrated mass there.

Standing assumptions and notation. Throughout this section, an integer $p \geq 3$ and $\beta > 0$ are fixed, except where a statement restricts p . Let (c, r) be the stationary profile of Proposition 8.4, and let $q = \lim_{v \rightarrow \infty} c(v) \in [0, 1)$ be its plateau. Recall from Proposition 5.7 and Corollary 7.7 that

$$c(0) = r(0) = 1, \quad r = -c', \quad \int_0^\infty r(v) \, dv = 1 - q,$$

and that c and r are completely monotone on $(0, \infty)$. In particular, r is nonnegative and nonincreasing, and $r(v) \rightarrow 0$ as $v \rightarrow \infty$ because r is integrable.

By Proposition 8.4, the translated kernels converge along every sequence of times tending to infinity. Taking the translation time equal to $t - A$, we find that, for every fixed $0 \leq A < \infty$,

$$\lim_{t \rightarrow \infty} \sup_{0 \leq v \leq A} \{|C(t, t - v) - c(v)| + |R(t, t - v) - r(v)|\} = 0. \quad (11.1)$$

We call $\{(t, s) : t - A \leq s \leq t\}$ the recent strip of length A . Recall also the forgetting statement (8.12): for every fixed $S \geq 0$, we have $\sup_{0 \leq s \leq S} \{C(t, s) + R(t, s)\} \rightarrow 0$ as $t \rightarrow \infty$.

We use the bounds $0 \leq R(t, s) \leq C(t, s) \leq 1$ for $t \geq s$ from Proposition 4.6, and $1 \leq \lambda \leq \Lambda$ from (4.17). Since the memory integral in the response equation (3.9) is nonnegative, we have $\partial_t R(t, s) \geq -\lambda(t)R(t, s)$ for $t > s$. Integrating this inequality from the diagonal, where $R(s, s) = 1$, we find

$$R(t, s) \geq \exp\left(-\int_s^t \lambda(u) \, du\right) \geq e^{-\Lambda(t-s)}, \quad t \geq s \geq 0. \quad (11.2)$$

For $t \geq 0$, we extend the lag row of the response by zero:

$$f_t(v) = \mathbf{1}_{[0, t]}(v) R(t, t - v), \quad v \geq 0. \quad (11.3)$$

We write $\|\cdot\|_2$ for the norm of $L^2(0, \infty)$. By (4.11), we have $\|f_t\|_2^2 = \int_0^t R(t, u)^2 \, du \leq 1/2$. For $t \geq L \geq 0$, we write

$$S_2(t, L) = \int_0^{t-L} R(t, u)^2 \, du = \int_L^\infty f_t(v)^2 \, dv$$

for the squared response in the remote part of the row. Here and below, an integral with upper endpoint $t - L$ is considered only for $t \geq L$.

Let v_t be the tree vectors of Proposition 7.1, so that $\langle v_t, v_s \rangle = C(t, s)$ and $\|v_t\| = 1$. Let P_T be the prefix projections, let Π_j be the projections onto trees of height j , and let $h(v) = \sqrt{p} \langle v, \mathcal{I}(v^{\otimes(p-1)}) \rangle$ be the tree energy. Recall that v_s belongs to the range of P_T when $s \leq T$. For $T, a \geq 0$, we set $M_T(a) = P_T v_{T+a}$. As the whole translated family converges, Proposition 7.5 applies along every sequence $T \rightarrow \infty$. For every fixed $L \geq 0$, it implies that

$$\|M_T(L)\|^2 \longrightarrow d_L = c(2L), \quad h(M_T(L)) \longrightarrow h_L \quad (T \rightarrow \infty), \quad (11.4)$$

where h_L is given by (7.18) with $e = e_\beta(\infty)$. If $q > 0$, then $d_L \geq q > 0$, and we define

$$\delta_L = E_{\text{th}} - \frac{h_L}{d_L^{p/2}}.$$

By Proposition 8.1(a), we have $\delta_L \geq 0$ for every L . Recall also that the plateau is then marginal, by Lemma 8.3. We prove in (11.26) below that $\delta_L \rightarrow 0$ as $L \rightarrow \infty$ for every $p \geq 3$. For $p \in \{3, 4\}$, both properties are part of Proposition 9.5, whose hypotheses hold by Proposition 8.4 and Lemma 8.3. In words, the tree energy of the normalized prefix centers $M_T(L)/\|M_T(L)\|$ approaches the threshold energy E_{th} when $T \rightarrow \infty$ first and then $L \rightarrow \infty$.

11.1 The full response row

Theorem 11.1 (Strong localization of the response). *Fix $p \geq 3$ and $\beta > 0$. Then*

$$\lim_{t \rightarrow \infty} \|f_t - r\|_{L^2(0, \infty)} = 0. \quad (11.5)$$

Further, for each fixed $L \geq 0$,

$$\lim_{t \rightarrow \infty} \int_0^{t-L} R(t, u)^2 \, du = \int_L^\infty r(v)^2 \, dv, \quad \lim_{t \rightarrow \infty} \frac{\chi(t, 0)}{\sqrt{t}} = 0. \quad (11.6)$$

The main step is to show that the response keeps no squared mass in the remote part of the row:

$$\lim_{L \rightarrow \infty} \limsup_{t \rightarrow \infty} S_2(t, L) = 0. \quad (11.7)$$

Once (11.7) is known, the theorem follows from stationary convergence on the recent strip. We prove (11.7) in three cases, which use different inputs.

- $q = 0$ and every $p \geq 3$. The white-leaf component of the prefix center $M_T(L)$ carries the squared response $2S_2(T + L, L)$, and its norm is at most $\|M_T(L)\|$. Here we use only the tree realization (Proposition 7.1) and reflection (Proposition 7.5).
- $q > 0$ and $p = 3$. We normalize the prefix center and control its white-leaf component by the height inequality of the cubic tree ceiling (Theorem 9.2). Besides the inputs of the first case, we need that $\delta_L \rightarrow 0$, which is part of Proposition 9.5 in Section 9.
- $q > 0$ and every $p \geq 3$. We test the mean of finitely many independent continuations against a Gaussian direction built from the Brownian path, and use the overlap penalty of Lemma 10.2 for an independent direction. This case relies on the finite forks of Section 10 and is carried out in Section 11.5 (Proposition 11.11).

The third case contains the second. We include the second because it is short and uses only the deterministic tree geometry.

Proof of Theorem 11.1. We prove (11.7) in Steps 1–3 and deduce the theorem from it in Step 4.

Step 1: the white-leaf bound. Fix $T, L \geq 0$. By (7.3), the component of v_{T+L} in the one-leaf summand $\mathbb{R}e_0 \oplus L^2(\mathbb{R}_+)$ is $R(T+L, 0)e_0 + \sqrt{2}\mathbf{1}_{[0, T+L]}R(T+L, \cdot)$. The remaining term of (7.3) lies in the range of the grafting map, which is orthogonal to that summand. Recall that P_T commutes with Π_0 , and that on the one-leaf summand it restricts the noise-time function to $[0, T]$. We conclude that

$$\|\Pi_0 M_T(L)\|^2 = R(T+L, 0)^2 + 2 \int_0^T R(T+L, u)^2 du. \quad (11.8)$$

As Π_0 is an orthogonal projection, (11.8) implies that $2S_2(T+L, L) \leq \|\Pi_0 M_T(L)\|^2 \leq \|M_T(L)\|^2$. Letting $T \rightarrow \infty$ and using (11.4), we find that, with $t = T+L$,

$$\limsup_{t \rightarrow \infty} S_2(t, L) \leq \frac{d_L}{2} = \frac{c(2L)}{2}. \quad (11.9)$$

This bound holds for every $p \geq 3$ and every plateau. If $q = 0$, then $c(2L) \rightarrow 0$ as $L \rightarrow \infty$, and (11.9) proves (11.7).

Step 2: a positive plateau in the cubic model. Suppose that $q > 0$ and $p = 3$, and fix $L \geq 0$. For all large T , we have $\|M_T(L)\|^2 \geq q/2$ by (11.4); for such T , let $z_T = M_T(L)/\|M_T(L)\|$. For $p = 3$, the tree energy is $h = \sqrt{3}f$, where $f(v) = \langle v, \mathcal{I}(v \otimes v) \rangle$ is the functional of Theorem 9.2, and $E_{\text{th}} = 2\sqrt{2/3} = \sqrt{3}F_*$ with $F_* = 2\sqrt{2/3}$. Keeping only the term $j = 0$ in the first inequality of (9.2), we have

$$\|\Pi_0 z_T\|^4 \leq 4\sqrt{2} [F_* - f(z_T)] = 4\sqrt{2/3} [E_{\text{th}} - h(z_T)].$$

On the other hand, (11.8) implies that $\|\Pi_0 z_T\|^2 \geq 2S_2(T+L, L)/\|M_T(L)\|^2$. Combining the two bounds, we find

$$S_2(T+L, L) \leq \frac{\|M_T(L)\|^2}{2} \left(4\sqrt{2/3} [E_{\text{th}} - h(z_T)] \right)^{1/2}.$$

The tree energy is homogeneous of degree three, which gives $h(z_T) = h(M_T(L))/\|M_T(L)\|^3$, and (11.4) implies that $h(z_T) \rightarrow h_L/d_L^{3/2} = E_{\text{th}} - \delta_L$. Letting $T \rightarrow \infty$ with L fixed, we obtain

$$\limsup_{t \rightarrow \infty} S_2(t, L) \leq \frac{d_L}{2} \sqrt{4\sqrt{2/3} \delta_L}. \quad (11.10)$$

Proposition 9.5 states that $\delta_L \rightarrow 0$ as $L \rightarrow \infty$. As $d_L \leq 1$, the right side of (11.10) tends to zero, and (11.7) follows.

Step 3: a positive plateau for general p . Suppose that $q > 0$ and $p \geq 3$. Proposition 11.11 states that, for every $L > 0$,

$$\limsup_{t \rightarrow \infty} S_2(t, L) \leq \frac{d_L}{2} \beta_d^{-2/p} \delta_L^{2/p},$$

and the right side tends to zero as $L \rightarrow \infty$ by (11.26). The proof of Proposition 11.11, in Section 11.5, is independent of Theorem 11.1 and its consequences. This completes the proof of (11.7) in all cases.

Step 4: from the remote bound to strong convergence. We first check that $r \in L^2(0, \infty)$. For fixed V , we have $f_t \rightarrow r$ uniformly on $[0, V]$ by (11.1), so $\int_0^V r^2 = \lim_t \int_0^V f_t^2 \leq 1/2$. Letting $V \rightarrow \infty$, we find $\|r\|_2^2 \leq 1/2$.

Next, fix $L > 0$ and $t \geq L$. Splitting $\|f_t - r\|_2^2$ at L , and using $(x - y)^2 \leq 2x^2 + 2y^2$ on $[L, \infty)$ together with $\int_L^\infty f_t^2 = S_2(t, L)$, we have

$$\|f_t - r\|_2^2 \leq L \sup_{0 \leq v \leq L} |R(t, t - v) - r(v)|^2 + 2S_2(t, L) + 2 \int_L^\infty r(v)^2 dv.$$

The first term tends to zero by (11.1), and taking the upper limit in t , we find

$$\limsup_{t \rightarrow \infty} \|f_t - r\|_2^2 \leq 2 \limsup_{t \rightarrow \infty} S_2(t, L) + 2 \int_L^\infty r(v)^2 dv.$$

Letting $L \rightarrow \infty$, and using (11.7) and $r \in L^2$, we obtain (11.5).

For the first limit in (11.6), fix $L \geq 0$. Using the triangle inequality, and the fact that restriction to $[L, \infty)$ is a contraction of $L^2(0, \infty)$, we have

$$\left| S_2(t, L)^{1/2} - \left(\int_L^\infty r(v)^2 dv \right)^{1/2} \right| \leq \|\mathbf{1}_{[L, \infty)}(f_t - r)\|_2 \leq \|f_t - r\|_2,$$

which tends to zero by (11.5). Finally, note that $\chi(t, 0) = \int_0^t R(t, u) du = \int_0^t f_t(v) dv$. Using the Cauchy–Schwarz inequality on $[0, t]$, we obtain

$$0 \leq \chi(t, 0) \leq \int_0^t r(v) dv + \int_0^t |f_t(v) - r(v)| dv \leq 1 - q + \sqrt{t} \|f_t - r\|_2.$$

After division by $\sqrt{t}$, the second limit in (11.6) follows from (11.5). $\square$

We next upgrade Theorem 11.1 to uniform convergence. The backward response equation prevents $s \mapsto R(t, s)$ from increasing quickly. A large remote value $R(t, s)$ would then force squared mass on the interval $[s - 1, s]$, which (11.6) excludes.

Corollary 11.2 (Uniform absolute localization). *Fix $p \geq 3$ and $\beta > 0$. Then*

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |R(t, s) - r(t - s)| = 0. \quad (11.11)$$

In particular, if $s_n \rightarrow \infty$ and $t_n - s_n \rightarrow \infty$, then $R(t_n, s_n) \rightarrow 0$.

Proof. Recall that the kernel Σ in (4.8) is nonnegative. Since $R \geq 0$, the backward response equation (4.21) implies that $\partial_s R(t, s) \leq \lambda(s)R(t, s) \leq \Lambda R(t, s)$ for $t > s$. Then $w \mapsto e^{-\Lambda w} R(t, w)$ is nonincreasing on $[0, t]$, that is, $R(t, u) \geq e^{-\Lambda(s-u)} R(t, s)$ for $u \leq s \leq t$. Squaring this inequality and integrating it over $u \in [s-1, s]$, we have, for $1 \leq s \leq t$,

$$R(t, s)^2 \frac{1 - e^{-2\Lambda}}{2\Lambda} \leq \int_{s-1}^s R(t, u)^2 du.$$

If moreover $s \leq t - L$, the integral on the right is at most $S_2(t, L)$. Using (11.6), we find that, for every fixed $L \geq 0$,

$$\limsup_{t \rightarrow \infty} \sup_{1 \leq s \leq t-L} R(t, s) \leq \left(\frac{2\Lambda}{1 - e^{-2\Lambda}} \int_L^\infty r(v)^2 dv \right)^{1/2}.$$

Recall from (8.12) that $\sup_{0 \leq s \leq 1} R(t, s) \rightarrow 0$. On the recent strip $t - L \leq s \leq t$, the difference $|R(t, s) - r(t - s)|$ tends to zero uniformly by (11.1). On the remaining part $0 \leq s \leq t - L$, we have $|R(t, s) - r(t - s)| \leq R(t, s) + r(L)$, since r is nonincreasing. Combining these three estimates, we have

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |R(t, s) - r(t - s)| \leq \left(\frac{2\Lambda}{1 - e^{-2\Lambda}} \int_L^\infty r(v)^2 dv \right)^{1/2} + r(L).$$

The right side tends to zero as $L \rightarrow \infty$, which proves (11.11). For the last assertion, note that $R(t_n, s_n) \leq |R(t_n, s_n) - r(t_n - s_n)| + r(t_n - s_n)$, and both terms tend to zero. $\square$

11.2 The later correlation derivative

We now bound $\partial_t C(t, s)$ when $t - s$ is large. By Lemma 7.3, the projected curve $a \mapsto M_T(a)$ is continuously differentiable for $a > 0$, with

$$\frac{d}{da} M_T(a) = P_T \mathcal{B}(v_{T+a}), \quad \mathcal{B}(v) = -v + \beta \nabla h(v) - p\beta h(v)v,$$

and $\mathcal{B}$ is Lipschitz on the closed unit ball of $\mathcal{H}$. Reflection identifies limits of inner products of prefix centers, not of their derivatives. For this reason, we pass through difference quotients. This requires a modulus of continuity for the projected derivative that is uniform in T , which we now record.

Lemma 11.3 (Hölder continuity of the projected drift). *Fix $p \geq 3$ and $\beta > 0$. Let $L_{\mathcal{B}}$ be a Lipschitz constant of $\mathcal{B}$ on the closed unit ball of $\mathcal{H}$, and set $\kappa_{\mathcal{B}} = L_{\mathcal{B}} \sqrt{2\Lambda}$. Then, for all $t \geq s \geq 0$,*

$$0 \leq 1 - C(t, s) \leq \Lambda(t - s), \quad \|v_t - v_s\| \leq \sqrt{2\Lambda(t - s)},$$

and, for all $t, t' \geq 0$ and $T \geq 0$,

$$\|\mathcal{B}(v_t) - \mathcal{B}(v_{t'})\| \leq \kappa_{\mathcal{B}} |t - t'|^{1/2}, \quad \|P_T \mathcal{B}(v_t) - P_T \mathcal{B}(v_{t'})\| \leq \kappa_{\mathcal{B}} |t - t'|^{1/2}.$$

Proof. The first bound is (4.20) in Proposition 4.4. The second bound follows from the identity $\|v_t - v_s\|^2 = 2 - 2C(t, s)$. The vectors v_t lie in the unit ball, and the third bound follows from the second and the Lipschitz property of $\mathcal{B}$. The last bound holds because P_T is a contraction. $\square$

Theorem 11.4 (Remote derivative bound and forward FDT). *Fix an integer $p \geq 3$ and $\beta > 0$. For every fixed $L > 0$,*

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq t-L} |\partial_t C(t, s)| \leq \sqrt{c''(2L)} \leq \frac{\sqrt{1-q}}{eL}. \quad (11.12)$$

Further,

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |\partial_t C(t, s) - c'(t-s)| = 0, \quad \lim_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |\partial_t C(t, s) + R(t, s)| = 0. \quad (11.13)$$

Here $\partial_t C(t, t)$ is the right derivative of $u \mapsto C(u, t)$ at $u = t$; it equals -1 for every $t \geq 0$. The value $c'(0) = -1$ is also the right derivative.

Proof. All inputs to this proof hold for every integer $p \geq 3$: the projected drift of Lemma 7.3, its uniform modulus in Lemma 11.3, prefix reflection in Proposition 7.5, complete monotonicity in Corollary 7.7, and uniform response localization in Corollary 11.2.

The proof has three steps. First, we bound the norm of the projected derivative $M'_T(L)$ through difference quotients and reflection. Since $\partial_t C(T+L, s) = \langle M'_T(L), v_s \rangle$ for every $s \leq T$, we deduce (11.12) from this bound. Second, we show that $\partial_t C(t, t-v) \rightarrow c'(v)$ uniformly on each recent strip, by differentiating stationary convergence through forward difference quotients. Third, we combine the two estimates with Corollary 11.2.

Step 1: the remote bound. Fix $L > 0$ and $\eta \in (0, 1]$. For every $T \geq 0$, we have $M_T(L + \eta) - M_T(L) = \int_0^\eta M'_T(L + w) dw$. Lemma 11.3 then implies that

$$\left\| M'_T(L) - \frac{M_T(L + \eta) - M_T(L)}{\eta} \right\| \leq \frac{1}{\eta} \int_0^\eta \|M'_T(L + w) - M'_T(L)\| dw \leq \frac{2}{3} \kappa_B \sqrt{\eta}.$$

Applying (7.16) to the three inner products in the expansion of the squared norm, we have

$$\lim_{T \rightarrow \infty} \left\| \frac{M_T(L + \eta) - M_T(L)}{\eta} \right\|^2 = \frac{c(2L + 2\eta) - 2c(2L + \eta) + c(2L)}{\eta^2}.$$

Combining the last two displays, we find

$$\limsup_{T \rightarrow \infty} \|M'_T(L)\| \leq \frac{2}{3} \kappa_B \sqrt{\eta} + \left(\frac{c(2L + 2\eta) - 2c(2L + \eta) + c(2L)}{\eta^2} \right)^{1/2}.$$

The left side does not depend on η . The function c is smooth on $(0, \infty)$, and the second difference quotient tends to $c''(2L)$ as $\eta \downarrow 0$. Taking $T \rightarrow \infty$ first and then $\eta \downarrow 0$, we have

$$\limsup_{T \rightarrow \infty} \|M'_T(L)\| \leq \sqrt{c''(2L)}.$$

If $s \leq T$, then v_s lies in the range of P_T , so $C(T + a, s) = \langle M_T(a), v_s \rangle$ for all $a \geq 0$, and $\partial_t C(T + L, s) = \langle M'_T(L), v_s \rangle$. As $\|v_s\| = 1$, we have $|\partial_t C(T + L, s)| \leq \|M'_T(L)\|$ for every $s \leq T$. Setting $T = t - L$ proves the first inequality in (11.12).

For the second inequality, we write the representation (7.28) as $c(v) = q + \int_{(0, \infty)} e^{-xv} \nu(dx)$, where $\nu((0, \infty)) = 1 - q$. Differentiating under the integral for $v > 0$, we have

$$c''(2L) = \int_{(0, \infty)} x^2 e^{-2Lx} \nu(dx) \leq \frac{1 - q}{e^2 L^2},$$

where we used that $\sup_{x>0} x^2 e^{-2Lx} = (eL)^{-2}$, attained at $x = 1/L$.

Step 2: the recent strip. For $t > s$, we apply the projected drift with $T = s$ to $C(t, s) = \langle P_s v_t, v_s \rangle$. This gives

$$\partial_t C(t, s) = \langle P_s \mathcal{B}(v_t), v_s \rangle = \langle \mathcal{B}(v_t), v_s \rangle.$$

The right side is continuous in $t \geq s$. At $t = s$ it equals -1 , because $\langle \nabla h(v), v \rangle = ph(v)$ by homogeneity and $\|v_s\| = 1$; this is the derivative from the causal side at the diagonal. Lemma 11.3 also implies that $|\partial_t C(t, s) - \partial_t C(t', s)| \leq \kappa_B |t - t'|^{1/2}$ for all $t, t' \geq s$.

Fix $L > 0$ and $\eta \in (0, 1]$. For $0 \leq v \leq L$ and $t \geq L$, we hold the earlier time $t - v$ fixed in a forward difference quotient. The Hölder bound just stated gives

$$\left| \partial_t C(t, t - v) - \frac{C(t + \eta, t - v) - C(t, t - v)}{\eta} \right| \leq \frac{1}{\eta} \int_0^\eta \kappa_B \sqrt{w} \, dw = \frac{2}{3} \kappa_B \sqrt{\eta}.$$

By (11.1) with $A = L + \eta$, the difference quotient converges to $[c(v + \eta) - c(v)]/\eta$ uniformly in $v \in [0, L]$. We obtain

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq v \leq L} |\partial_t C(t, t - v) - c'(v)| \leq \frac{2}{3} \kappa_B \sqrt{\eta} + \sup_{0 \leq v \leq L} \left| \frac{c(v + \eta) - c(v)}{\eta} - c'(v) \right|.$$

The function $c' = -r$ is uniformly continuous on $[0, L + 1]$, including the right derivative $c'(0) = -1$, and the last supremum tends to zero as $\eta \downarrow 0$. We conclude that $\sup_{t-L \leq s \leq t} |\partial_t C(t, s) - c'(t - s)| \rightarrow 0$ for every fixed L .

Step 3: the whole row. On the remote part $0 \leq s \leq t - L$, we have $|\partial_t C(t, s) - c'(t - s)| \leq |\partial_t C(t, s)| + r(L)$. Using Step 2 and (11.12), we find

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq s \leq t} |\partial_t C(t, s) - c'(t - s)| \leq \sqrt{c''(2L)} + r(L).$$

Letting $L \rightarrow \infty$ proves the first limit in (11.13). As $c' = -r$, the second limit follows from the first and Corollary 11.2. $\square$

11.3 The earlier derivative and the squared FDT defect

For $t > s > 0$, let $E(t, s) = \partial_s C(t, s) - R(t, s)$ be the backward FDT defect. By Theorem 5.3, the derivative $\partial_s C$ is continuous on the open causal triangle, and there exists a nonnegative integrable function a_F , obtained on one dimension subsequence, such that

$$|E(t, s)|^2 \leq a_F(s) \quad (t > s > 0), \quad \int_0^\infty a_F(s) \, ds < \infty, \quad \lim_{S \rightarrow \infty} \sup_{s \geq S} a_F(s) = 0. \quad (11.14)$$

We do not need a_F to be unique. The bound (11.14) is the spherical form used here of the entropy-production bounds on FDT violation of Cugliandolo, Dean, and Kurchan [54, arXiv version 2, equations (12)–(13), p. 3, and (18), p. 4]. The common majorant in (11.14) is proved in Theorem 5.3, and the whole-row conclusions below are proved here. It shows that the defect is uniformly small when the earlier time is late. The row $[0, t]$ grows with t , however, and this alone does not control the defect on the whole row. The next theorem shows that the integral of the squared defect over the row nevertheless tends to zero.

Theorem 11.5 (Strong backward FDT). *Fix $p \geq 3$ and $\beta > 0$. Then*

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq S} |\partial_s C(t, s)| = 0 \quad \text{for each fixed } 0 \leq S < \infty, \quad \lim_{t \rightarrow \infty} \int_0^t |E(t, s)|^2 \, ds = 0. \quad (11.15)$$

Further, if $g_t(v) = \mathbf{1}_{[0, t]}(v) \partial_s C(t, t - v)$, then

$$\lim_{t \rightarrow \infty} \|g_t - r\|_{L^2(0, \infty)} = 0. \quad (11.16)$$

For $t > 0$, the earlier derivative at $s = 0$ is the right derivative. At $s = t > 0$, it is the left derivative of $u \mapsto C(t, u)$ and equals 1; we set $E(t, t) = 0$.

Proof. We proceed in three steps. First, we represent the colored memory term of the Green identity through feature vectors. A positivity argument, together with forgetting, shows that this term is negligible on every bounded interval of earlier times. Second, we differentiate the Green identity in its earlier time. This expresses $\partial_s C(t, s)$ through C , R , and the colored term, and it proves the first claim in (11.15). Third, the Fisher majorant allows dominated convergence over the whole row, and Theorem 11.1 implies (11.16).

Step 1: the colored term forgets bounded earlier intervals. For $s \geq 0$, let $\phi_s = v_s^{\otimes(p-1)} \in \mathcal{H}^{\otimes(p-1)}$. These are unit vectors with $\langle \phi_u, \phi_s \rangle = C(u, s)^{p-1}$ and $\|\phi_u - \phi_s\|^2 = 2[1 - C(u, s)^{p-1}]$. In particular, $s \mapsto \phi_s$ is norm continuous. We define

$$F_t = \int_0^t R(t, u) \phi_u \, du, \quad \mathcal{K}(t, s) = \langle F_t, \phi_s \rangle = \int_0^t R(t, u) C(u, s)^{p-1} \, du,$$

where the first integral is a Bochner integral of a continuous function on $[0, t]$. Note that $\mathcal{K}(t, s) \geq 0$. When $s = t$, the colored memory term in the Green identity (4.10) equals $\beta^2 p \|F_t\|^2$. Since all three terms of (4.10) are nonnegative, we obtain $\beta^2 p \|F_t\|^2 \leq C(t, t) = 1$. By the Cauchy–Schwarz inequality, we have, for all $t, s, v \geq 0$,

$$|\mathcal{K}(t, s) - \mathcal{K}(t, v)| \leq (\beta^2 p)^{-1/2} \sqrt{2[1 - C(s, v)^{p-1}]}. \quad (11.17)$$

This modulus does not depend on t . For $s \leq t$, Fubini's theorem expresses the colored memory term in (4.10) as $\beta^2 p \int_0^s R(s, v) \mathcal{K}(t, v) dv$. The other two terms are nonnegative, and therefore

$$C(t, s) \geq \beta^2 p \int_0^s R(s, v) \mathcal{K}(t, v) dv. \quad (11.18)$$

We claim that $\mathcal{K}(t, s) \rightarrow 0$ as $t \rightarrow \infty$, for every fixed $s > 0$. Suppose instead that $\mathcal{K}(t_n, s) \geq \varepsilon$ along a sequence $t_n \rightarrow \infty$, for some $\varepsilon > 0$. Since C is continuous and $C(s, s) = 1$, there exists $h \in (0, s]$ such that the right side of (11.17) is at most $\varepsilon/2$ for $v \in [s - h, s]$. Then $\mathcal{K}(t_n, v) \geq \varepsilon/2$ on $[s - h, s]$. Inserting this and the lower bound (11.2) into (11.18), we obtain

$$C(t_n, s) \geq \frac{\beta^2 p \varepsilon}{2} \int_{s-h}^s e^{-\Lambda(s-v)} dv = \frac{\beta^2 p \varepsilon}{2\Lambda} (1 - e^{-\Lambda h}) > 0,$$

which contradicts $C(t_n, s) \rightarrow 0$ from (8.12). For $s = 0$ and any $s' > 0$, (11.17) implies that $\mathcal{K}(t, 0) \leq \mathcal{K}(t, s') + (\beta^2 p)^{-1/2} \sqrt{2[1 - C(s', 0)^{p-1}]}$. Letting $t \rightarrow \infty$ and then $s' \downarrow 0$, we find $\mathcal{K}(t, 0) \rightarrow 0$.

Finally, fix S and $\varepsilon > 0$. The correlation is uniformly continuous on $[0, S]^2$ with $C(s, s) = 1$. Hence there exists a finite set $\{s_1, \dots, s_k\} \subset [0, S]$ such that every $s \in [0, S]$ has some s_i with $(\beta^2 p)^{-1/2} \sqrt{2[1 - C(s, s_i)^{p-1}]} \leq \varepsilon$. Then (11.17) implies that $\sup_{0 \leq s \leq S} \mathcal{K}(t, s) \leq \max_i \mathcal{K}(t, s_i) + \varepsilon$. Letting $t \rightarrow \infty$ and then $\varepsilon \downarrow 0$, we conclude that

$$\lim_{t \rightarrow \infty} \sup_{0 \leq s \leq S} \mathcal{K}(t, s) = 0. \quad (11.19)$$

Step 2: the backward correlation equation. Fix $t > 0$. By the Green identity (4.10) and Fubini's theorem, we have, for $0 \leq s \leq t$,

$$C(t, s) = R(t, 0)R(s, 0) + 2 \int_0^s R(t, u)R(s, u) du + \beta^2 p \int_0^s R(s, v) \mathcal{K}(t, v) dv.$$

By Proposition 4.5, the derivative $\partial_1 R$ is continuous on each compact causal triangle, and $\mathcal{K}(t, \cdot)$ is continuous. We may therefore differentiate in $s \in (0, t)$ under the integral signs. Using $R(s, s) = 1$, we obtain

$$\begin{aligned} \partial_s C(t, s) &= 2R(t, s) + \beta^2 p \mathcal{K}(t, s) + R(t, 0) \partial_1 R(s, 0) \\ &\quad + 2 \int_0^s R(t, u) \partial_1 R(s, u) du + \beta^2 p \int_0^s \partial_1 R(s, v) \mathcal{K}(t, v) dv. \end{aligned}$$

The first two terms are the boundary terms of the thermal and colored integrals. We now insert the response equation (3.9), which reads $\partial_1 R(s, v) = -\lambda(s)R(s, v) + \int_v^s \Sigma(s, w)R(w, v) dw$ with $\Sigma(s, w) = \beta^2 p(p-1)C(s, w)^{p-2}R(s, w)$. The terms with $-\lambda(s)$ reproduce $-\lambda(s)$ times the Green representation of $C(t, s)$. In the memory terms, we exchange the order of integration on the triangle $\{0 \leq u \leq w \leq s\}$, and they become

$$\int_0^s \Sigma(s, w) \left[R(t, 0)R(w, 0) + 2 \int_0^w R(t, u)R(w, u) du + \beta^2 p \int_0^w R(w, v)\mathcal{K}(t, v) dv \right] dw.$$

By the Green identity at (t, w) , the bracket equals $C(t, w)$. We have shown that, for $0 < s < t$,

$$\begin{aligned} \partial_s C(t, s) = & -\lambda(s)C(t, s) + 2R(t, s) + \beta^2 p(p-1) \int_0^s C(s, u)^{p-2}R(s, u)C(t, u) du \\ & + \beta^2 p \mathcal{K}(t, s). \end{aligned} \quad (11.20)$$

All integrations and differentiations here take place on finite rectangles, where the integrands are continuous and bounded. The right side of (11.20) is continuous on $[0, t]$, so it also gives the right derivative at $s = 0$ and the left derivative at $s = t$.

As a check, we let $s \uparrow t$. Using the definition of $\mathcal{K}$ and (3.11), we find

$$\begin{aligned} \beta^2 p \mathcal{K}(t, t) + \beta^2 p(p-1) \int_0^t C(t, u)^{p-1}R(t, u) du &= \beta^2 p^2 \int_0^t C(t, u)^{p-1}R(t, u) du \\ &= \lambda(t) - 1. \end{aligned}$$

Inserting this into (11.20), we obtain $\partial_s C(t, t-) = -\lambda(t) + 2 + \lambda(t) - 1 = 1$, and $E(t, t-) = 0$, as the unit diagonal response requires.

Now fix S and let $t > S$. Bounding $C(s, u)^{p-2} \leq 1$ and $R(s, u) \leq 1$ in (11.20), we have, for $0 \leq s \leq S$,

$$|\partial_s C(t, s)| \leq (\Lambda + \beta^2 p(p-1)S) \sup_{0 \leq u \leq S} C(t, u) + 2 \sup_{0 \leq u \leq S} R(t, u) + \beta^2 p \sup_{0 \leq u \leq S} \mathcal{K}(t, u).$$

Each supremum tends to zero, by (8.12) and (11.19). This proves the first claim in (11.15).

Step 3: the whole row. Let $\tilde{E}_t(s) = \mathbf{1}_{(0, t)}(s)E(t, s)$ for $s > 0$. For fixed $s > 0$, the first claim in (11.15) and (8.12) imply that $\tilde{E}_t(s) \rightarrow 0$ as $t \rightarrow \infty$. Further, $|\tilde{E}_t(s)|^2 \leq a_F(s)$ for all t by (11.14), and a_F is integrable on $(0, \infty)$. By the dominated convergence theorem, we conclude that

$$\lim_{t \rightarrow \infty} \int_0^t |E(t, s)|^2 ds = \lim_{t \rightarrow \infty} \int_0^\infty |\tilde{E}_t(s)|^2 ds = 0.$$

This is the second claim in (11.15). After the change of variables $s = t - v$, we have $\|g_t - f_t\|_2^2 = \int_0^t |E(t, s)|^2 ds$. Then Theorem 11.1 implies that $\|g_t - r\|_2 \leq \|g_t - f_t\|_2 + \|f_t - r\|_2 \rightarrow 0$, which is (11.16). $\square$

The proof does not require a unique limit of the Fisher information, nor a majorant chosen anew for each terminal time: one function a_F dominates every row. The positivity of C and R enters through (11.18).

The next corollary states that remote values of the response and of the earlier derivative are small, uniformly on the late triangle, and that remote correlations change little across a window of fixed length in the earlier time. It is used in Section 14.

Corollary 11.6 (Uniform remote decay and earlier smoothing). *Fix $p \geq 3$ and $\beta > 0$. Then*

$$\lim_{D \rightarrow \infty} \limsup_{S \rightarrow \infty} \sup_{s \geq S, t \geq s+D} (R(t, s) + |\partial_s C(t, s)|) = 0. \quad (11.21)$$

More precisely, for each fixed $D > 0$,

$$\limsup_{S \rightarrow \infty} \sup_{s \geq S+D, t \geq s+D} \int_{s-D}^s |\partial_u C(t, u)| \, du \leq \sqrt{D} \left(\int_D^\infty r(v)^2 \, dv \right)^{1/2}. \quad (11.22)$$

If $r(v) = O(v^{-1-a})$ as $v \rightarrow \infty$ for some $a > 0$, the right side of (11.22) is $O(D^{-a})$ as $D \rightarrow \infty$.

When $\beta \geq \beta_d$, Theorem 16.1 gives $r(v) \sim a A_\beta v^{-1-a}$, with the exponent $a \in (0, 1/2)$ defined there, and the last hypothesis holds.

Proof. Let

$$\epsilon(U) = \sup_{t \geq U} \|f_t - r\|_2, \quad \epsilon_\infty(U) = \sup_{t \geq U} \sup_{0 \leq s \leq t} |R(t, s) - r(t-s)|.$$

Both tend to zero as $U \rightarrow \infty$, by Theorem 11.1 and Corollary 11.2. For $t > s > 0$, we have $|\partial_s C(t, s)| \leq R(t, s) + \sqrt{a_F(s)}$ by (11.14).

For the pointwise statement, let $s \geq S$ and $t \geq s + D$. As r is nonincreasing, $R(t, s) \leq r(D) + \epsilon_\infty(S + D)$, which implies

$$R(t, s) + |\partial_s C(t, s)| \leq 2r(D) + 2\epsilon_\infty(S + D) + \sup_{u \geq S} \sqrt{a_F(u)}.$$

We take $S \rightarrow \infty$, using (11.14), and then $D \rightarrow \infty$. This proves (11.21).

For the integrated statement, fix $D > 0$, and let $s \geq S + D$ and $t \geq s + D$. Then $[s - D, s] \subset [S, t - D]$ and $t \geq S + 2D$. Recall that $\mu([S, \infty)) = \int_S^\infty a_F(u) \, du$, where μ is the measure of (5.14). By the Cauchy–Schwarz inequality,

$$\begin{aligned} \int_{s-D}^s |\partial_u C(t, u)| \, du &\leq \sqrt{D} \left(\int_{s-D}^s R(t, u)^2 \, du \right)^{1/2} + \sqrt{D \mu([S, \infty))} \\ &\leq \sqrt{D} S_2(t, D)^{1/2} + \sqrt{D \mu([S, \infty))}. \end{aligned}$$

As in Step 4 of the proof of Theorem 11.1, we have $S_2(t, D)^{1/2} \leq (\int_D^\infty r^2)^{1/2} + \|f_t - r\|_2$. The supremum in (11.22) is therefore at most

$$\sqrt{D} \left[\left(\int_D^\infty r(v)^2 dv \right)^{1/2} + \epsilon(S + 2D) \right] + \sqrt{D \mu([S, \infty))}.$$

Since a_F is integrable, $\mu([S, \infty)) \rightarrow 0$, and letting $S \rightarrow \infty$ proves (11.22). Finally, if $r(v) \leq A_0 v^{-1-a}$ for $v \geq v_0$, then for $D \geq v_0$ we have

$$\sqrt{D} \left(\int_D^\infty r(v)^2 dv \right)^{1/2} \leq \sqrt{D} \left(\frac{A_0^2 D^{-1-2a}}{1+2a} \right)^{1/2} = \frac{A_0}{\sqrt{1+2a}} D^{-a}.$$

This completes the proof. $\square$

11.4 The direct thermal covariance

The Green identity (4.10) splits $C(t, s)$ into an initial term, a thermal term, which comes directly from the white noise, and a colored memory term. We define the thermal term and its stationary counterpart by

$$B(t, s) = 2 \int_0^{t \wedge s} R(t, u) R(s, u) du, \quad b_{\text{th}}(\tau) = 2 \int_0^\infty r(\tau + v) r(v) dv.$$

We say that a sequence of pairs (t_n, s_n) with $t_n \geq s_n$ is a late pair with diverging lag if $s_n \rightarrow \infty$ and $t_n - s_n \rightarrow \infty$.

Corollary 11.7 (Thermal covariance on the full late triangle). *Fix $p \geq 3$ and $\beta > 0$. Then*

$$\lim_{S \rightarrow \infty} \sup_{t \geq s \geq S} |B(t, s) - b_{\text{th}}(t - s)| = 0. \quad (11.23)$$

The function b_{th} is completely monotone on $(0, \infty)$ and tends to zero at infinity. In particular, $B(t_n, s_n) \rightarrow 0$ along every late pair with diverging lag, and $B(t, t) \rightarrow 2 \int_0^\infty r(v)^2 dv$. If $C(t_n, s_n) \rightarrow c_0$ along a late pair with diverging lag, then

$$\lim_{n \rightarrow \infty} \beta^2 p \int_0^{t_n} \int_0^{s_n} R(t_n, u) R(s_n, v) C(u, v)^{p-1} dv du = c_0. \quad (11.24)$$

If $r(\tau) \sim a A \tau^{-1-a}$ as $\tau \rightarrow \infty$ for some constants $a, A > 0$, then

$$b_{\text{th}}(\tau) \sim 2(1-q) r(\tau) \sim 2(1-q) a A \tau^{-1-a}. \quad (11.25)$$

By Theorem 16.1, the hypothesis of the last assertion holds when $\beta \geq \beta_d$, with $A = A_\beta$. The identity (11.24) says that any correlation retained along a late pair with diverging lag comes from the colored memory term, with its whole integration rectangle.

Proof. Let $\tau = t - s \geq 0$. After the substitution $u = s - v$, the zero extension (11.3) gives

$$B(t, s) = 2 \int_0^\infty f_t(\tau + v) f_s(v) \, dv,$$

because $f_s(v) = 0$ for $v > s$. Let $(\mathcal{T}_\tau g)(v) = g(\tau + v)$; this translation is a contraction of $L^2(0, \infty)$. Then $B(t, s) = 2 \langle \mathcal{T}_\tau f_t, f_s \rangle$ and $b_{\text{th}}(\tau) = 2 \langle \mathcal{T}_\tau r, r \rangle$. Using $\|f_s\|_2, \|r\|_2 \leq 1/\sqrt{2}$, we have

$$|B(t, s) - b_{\text{th}}(\tau)| \leq 2 |\langle \mathcal{T}_\tau (f_t - r), f_s \rangle| + 2 |\langle \mathcal{T}_\tau r, f_s - r \rangle| \leq \sqrt{2} (\|f_t - r\|_2 + \|f_s - r\|_2),$$

uniformly in $\tau \geq 0$. With $\epsilon(S) = \sup_{t \geq S} \|f_t - r\|_2$, the left side of (11.23) is at most $2\sqrt{2}\epsilon(S)$, which tends to zero by Theorem 11.1.

Next, we write $c(v) = q + \int_{(0, \infty)} e^{-xv} \nu(dx)$ as in (7.28). Let $\sigma(dx) = x \nu(dx)$ on $(0, \infty)$. Then $r(v) = \int_{(0, \infty)} e^{-xv} \sigma(dx)$ for $v > 0$, and $\sigma((0, \infty)) = \lim_{v \downarrow 0} r(v) = 1$ by monotone convergence. By Tonelli's theorem,

$$b_{\text{th}}(\tau) = 2 \int_{(0, \infty)^2} \frac{e^{-x\tau}}{x+y} \sigma(dx) \sigma(dy) = \int_{(0, \infty)} e^{-x\tau} \rho_{\text{th}}(dx),$$

where $\rho_{\text{th}}(dx) = 2\sigma(dx) \int_{(0, \infty)} (x+y)^{-1} \sigma(dy)$. The measure ρ_{th} is positive with total mass $b_{\text{th}}(0) = 2\|r\|_2^2 \leq 1$. Hence b_{th} is the Laplace transform of a finite positive measure on $(0, \infty)$, and it is completely monotone on $(0, \infty)$. Also $b_{\text{th}}(\tau) \leq 2\|r\|_2 \|\mathbf{1}_{[\tau, \infty)} r\|_2 \rightarrow 0$.

Along a late pair with diverging lag, we have $|B(t_n, s_n)| \leq |B(t_n, s_n) - b_{\text{th}}(t_n - s_n)| + b_{\text{th}}(t_n - s_n) \rightarrow 0$. Taking $\tau = 0$ in (11.23), we find $B(t, t) \rightarrow b_{\text{th}}(0) = 2\|r\|_2^2$. The initial term of the Green identity satisfies $0 \leq R(t_n, 0)R(s_n, 0) \leq R(t_n, 0) \rightarrow 0$, by (8.12). Subtracting the initial and thermal terms from (4.10) at (t_n, s_n) proves (11.24).

For (11.25), note that $r > 0$ by Proposition 6.1, and that

$$\frac{b_{\text{th}}(\tau)}{r(\tau)} = 2 \int_0^\infty \frac{r(\tau + v)}{r(\tau)} r(v) \, dv.$$

As r is nonincreasing, the ratio $r(\tau + v)/r(\tau)$ lies in $[0, 1]$. It tends to one for each fixed v , by the assumed asymptotics. The measure $r(v) \, dv$ has finite mass $1 - q$. Using dominated convergence, we find $b_{\text{th}}(\tau)/r(\tau) \rightarrow 2(1 - q)$. $\square$

Remark 11.8 (Square integrable versus integrated response). The convergence statements of this section concern L^2 norms and uniform norms of rows. A slowly varying contribution spread over a growing part of the row can have vanishing squared norm and a positive integral. This happens for the response in the cubic model. For $p = 3$ and $\beta \geq \sqrt{4/3}$, Theorem 14.7 gives $\liminf_t \chi(t, 0) \geq 2(1 - q)$, while $\int_0^L f_t(v) \, dv \rightarrow \int_0^L r(v) \, dv \leq 1 - q$ for each fixed L . We conclude that, for every fixed L ,

$$\liminf_{t \rightarrow \infty} \int_0^{t-L} R(t, u) \, du \geq 1 - q > 0,$$

although $\int_0^{t-L} R(t, u)^2 du \rightarrow \int_L^\infty r^2$, which tends to zero as $L \rightarrow \infty$. In Section 12, we prove a statement of this kind for the backward FDT defect.

11.5 The remote response for general pure models

In this section, we prove Proposition 11.11. It is the case $q > 0$ of the remote estimate (11.7) for general $p \geq 3$, which we used in Step 3 of the proof of Theorem 11.1. For $p = 3$, we controlled the white-leaf component of a normalized prefix center directly, by the height inequality of the cubic tree ceiling. For general p , we have no such inequality. Instead, we test the mean of finitely many independent continuations of the dynamics against one Gaussian direction built from the Brownian path. The overlap penalty for a direction independent of the disorder (Lemma 10.2) then plays the role of the height inequality. That lemma rests on the external-field algorithmic threshold of Huang and Sellke [92, arXiv version 2, Theorem 1, p. 4, and Corollary 1.8, p. 8]; see Theorem 10.1. As the test direction is independent of the disorder, we do not need the adaptive disclosure argument that Section 10 uses for forgetting.

Setting. We work in the setting of Section 10. We take dimension limits along the subsequence (N_j) of Theorem 5.3 and write $N \rightarrow \infty$ for $N = N_j \rightarrow \infty$. We keep every time horizon, branch number, and approximation parameter fixed while $N \rightarrow \infty$, and take long-time limits afterward.

Fix an integer $K \geq 1$ and times $T > 0$ and $L \geq 0$. We run the diffusion (3.3) up to time T , driven by the Brownian motion $(B_u)_{u \leq T}$, which we call the trunk. From X_T , we run K continuations of duration L , driven by Brownian motions independent of each other and of $(J, X_0, (B_u)_{u \leq T})$. Let $Y_1, \dots, Y_K$ be their endpoints at time $T + L$, and let $m_{N,T} = K^{-1} \sum_{b=1}^K Y_b$ be their mean, which also depends on K and L . For each b , the Brownian path driving the b -th continuation, extended by the trunk on $[0, T]$, is a standard Brownian motion on $[0, T + L]$ independent of (J, X_0) . Since (3.3) has a unique strong solution, the vector $(J, X_0, (B_u)_{u \leq T}, Y_b)$ has the same law as $(J, X_0, (B_u)_{u \leq T}, X_{T+L})$.

Let $Q_{K,L}$ and $h_{K,L}$ be given by (10.14) and (10.16), with $e = e_\beta(\infty)$. When $q > 0$, every sequence $T \rightarrow \infty$ realizes the stationary profile (c, r) by Proposition 8.4, and we may apply Proposition 10.4 along every such sequence. By Lemma 10.5(i), we have, for every fixed $L \geq 0$,

$$\lim_{K \rightarrow \infty} Q_{K,L} = c(2L) = d_L, \quad \lim_{K \rightarrow \infty} h_{K,L} = h_L.$$

We now show that the quantity δ_L defined at the start of Section 11 satisfies

$$\delta_L = E_{\text{th}} - \frac{h_L}{d_L^{p/2}} \geq 0, \quad \lim_{L \rightarrow \infty} \delta_L = 0. \quad (11.26)$$

The inequality is (8.1); it also follows by letting $K \rightarrow \infty$ in (10.19). The plateau is marginal by Lemma 8.3, so (10.28) applies, and it gives the limit, because $h_{K,L}/Q_{K,L}^{p/2} \rightarrow h_L/d_L^{p/2}$ as $K \rightarrow \infty$. For $p \in \{3, 4\}$, the quantity δ_L is the same as in (9.12).

Two lemmas. The first lemma converts the absolute-overlap cost of Lemma 10.2 into a bound on the overlap with a Gaussian vector from an independent seed. We write $B_N = \{x \in \mathbb{R}^N : \|x\| \leq \sqrt{N}\}$. Recall that $\beta_d = \beta_d(p)$ is a constant of the model, defined in Section 1.3; it does not depend on the inverse temperature β .

Lemma 11.9 (Gaussian seed overlap has a threshold cost). *Fix $p \geq 3$ and $0 \leq L_0 < \infty$. For each N , let ζ_N be a random element independent of the coefficient tensor J , let G_N be a standard Gaussian vector in $\mathbb{R}^N$ that is a measurable function of ζ_N , and let $A_N = A_N(J, \zeta_N) \in B_N$ be measurable. Suppose that, for almost every value of ζ_N , the map $J \mapsto A_N(J, \zeta_N)$ is globally L_0 -Lipschitz, for the Euclidean distances on the ordered coefficient tensors and on $\mathbb{R}^N$. Let $\delta \geq 0$, and suppose that*

$$\liminf_{N \rightarrow \infty} \frac{\mathbb{E} H_N(A_N)}{N} \geq E_{\text{th}} - \delta.$$

Then

$$\limsup_{N \rightarrow \infty} \mathbb{E} \left| \frac{\langle A_N, G_N \rangle}{N} \right| \leq \beta_d^{-1/p} \delta^{1/p}. \quad (11.27)$$

The constant $\beta_d^{-1/p}$ depends neither on L_0 nor on the law of ζ_N . The same holds when N runs through a subsequence.

The algorithm A_N may depend on G_N ; only J must be independent of the seed. If A_N also depends on parameters held fixed while $N \rightarrow \infty$, the lemma applies at each of their values, and its conclusion passes to limits in these parameters taken afterward.

Proof. Let $w_N = \sqrt{N} G_N / \|G_N\|$ if $G_N \neq 0$, and $w_N = \sqrt{N} e_1$ otherwise. Then w_N is a measurable function of ζ_N with $\|w_N\| = \sqrt{N}$. Since $\|A_N\| \leq \sqrt{N}$, we have

$$\left| \frac{\langle A_N, G_N \rangle}{N} - \frac{\langle A_N, w_N \rangle}{N} \right| \leq \frac{\|G_N - w_N\|}{\sqrt{N}} = \left| \frac{\|G_N\|}{\sqrt{N}} - 1 \right|.$$

For $x \geq 0$, we have $|x - 1| \leq |x^2 - 1|$. The variable $\|G_N\|^2/N$ has mean one and variance $2/N$, and the expectation of the right side is at most $\mathbb{E} |\|G_N\|^2/N - 1| \leq \sqrt{2/N}$, which tends to zero.

We use Lemma 10.2 conditionally on ζ_N , with direction w_N , which is ζ_N -measurable, and Lipschitz constant L_0 . Writing $a_N = \langle A_N, w_N \rangle / N$, we have

$$\limsup_{N \rightarrow \infty} \left\{ \frac{\mathbb{E} H_N(A_N)}{N} + \beta_d \mathbb{E} |a_N|^p \right\} \leq E_{\text{th}}.$$

Since $\limsup_N (x_N + y_N) \geq \liminf_N x_N + \limsup_N y_N$, the hypothesis on the energy implies that $\beta_d \limsup_N \mathbb{E}|a_N|^p \leq \delta$. Hölder's inequality gives $\mathbb{E}|a_N| \leq (\mathbb{E}|a_N|^p)^{1/p}$, and hence $\limsup_N \mathbb{E}|a_N| \leq (\delta/\beta_d)^{1/p}$. Combining this with the first paragraph proves (11.27). The constants in Lemma 10.2 depend neither on the seed nor on the direction. $\square$

The second lemma identifies the overlap of the mean $m_{N,T}$ with a Gaussian vector built from the trunk. It is a weighted form of the Brownian response identity of Lemma 3.4 and Proposition 3.3.

Lemma 11.10 (A weighted Brownian response identity). *Fix $K \geq 1$, $T > 0$, and $L \geq 0$, and let $k : [0, T] \rightarrow \mathbb{R}$ be a continuous deterministic function with $\int_0^T k(u)^2 du = 1$. Let $G_N = \int_0^T k(u) dB_u$, where B is the trunk Brownian motion. Then G_N is a standard Gaussian vector in $\mathbb{R}^N$, independent of the disorder, and*

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\frac{\langle m_{N,T}, G_N \rangle}{N} \right] = \sqrt{2} \int_0^T k(u) R(T+L, u) du. \quad (11.28)$$

Proof. The coordinates of G_N are Wiener integrals of the independent coordinates of B . They are independent centered Gaussian variables with variance $\int_0^T k^2 = 1$, and they are independent of (J, X_0) because B is.

Let $t = T + L$. Since G_N is a function of the trunk, the equality of laws noted above gives $\mathbb{E} \langle Y_b, G_N \rangle = \mathbb{E} \langle X_t, G_N \rangle$ for each b . Averaging over b , we have $\mathbb{E} \langle m_{N,T}, G_N \rangle = \mathbb{E} \langle X_t, G_N \rangle$. Suppose first that $k = \sum_{i=1}^n k_i \mathbf{1}_{(u_{i-1}, u_i]}$ is a step function, with $0 = u_0 < \dots < u_n = T$. Then

$$\frac{\mathbb{E} \langle X_t, G_N \rangle}{N} = \sum_{i=1}^n k_i \frac{\mathbb{E} \langle X_t, B_{u_i} - B_{u_{i-1}} \rangle}{N} = \sqrt{2} \sum_{i=1}^n k_i [\mathbb{E} \Gamma_N(t, u_i) - \mathbb{E} \Gamma_N(t, u_{i-1})],$$

with Γ_N as in Proposition 3.3. That proposition states that $\mathbb{E} \Gamma_N(t, u) \rightarrow \int_0^u R(t, w) dw$, and the right side converges to $\sqrt{2} \int_0^T k(u) R(t, u) du$.

For a general continuous k and $\varepsilon > 0$, we choose a step function k_ε with $\sup_{[0, T]} |k - k_\varepsilon| \leq \varepsilon$. We have $\|X_t\|^2 = N$ and, by the Itô isometry, $\mathbb{E} \|\int_0^T (k - k_\varepsilon) dB\|^2 = N \int_0^T (k - k_\varepsilon)^2$. Using the Cauchy–Schwarz inequality, we find

$$\frac{1}{N} \left| \mathbb{E} \left\langle X_t, \int_0^T (k - k_\varepsilon) dB \right\rangle \right| \leq \left(\int_0^T (k - k_\varepsilon)^2 \right)^{1/2} \leq \sqrt{T} \varepsilon,$$

uniformly in N . Also $|\sqrt{2} \int_0^T (k - k_\varepsilon) R(t, \cdot)| \leq \sqrt{2} T \varepsilon$, since $0 \leq R \leq 1$. Using the step-function case for k_ε , we obtain

$$\limsup_{N \rightarrow \infty} \left| \frac{\mathbb{E} \langle X_t, G_N \rangle}{N} - \sqrt{2} \int_0^T k(u) R(t, u) du \right| \leq (\sqrt{T} + \sqrt{2} T) \varepsilon.$$

Letting $\varepsilon \downarrow 0$ proves (11.28). $\square$

The Brownian observable goes back to Ben Arous, Dembo, and Guionnet [24, arXiv version 1, equation (1.9), p. 4, and Theorem 1.2, p. 5], and its hard-sphere limit is [112, arXiv version 5, Definition 3, p. 22, and Theorem 4.3, p. 23], which underlies Proposition 3.3. By Lemma 3.4, the left side of (11.28) at finite N equals $\sqrt{2} \int_0^T k(u) \bar{r}_N(T+L, u) du$ when k is a step function; this is the Cameron–Martin identity for a deterministic force with profile k . We need only its limit, which Proposition 3.3 provides without a pointwise limit of the finite-dimensional response.

The remote response. We now prove the remaining case of (11.7).

Proposition 11.11 (Remote squared response for general pure models). *Fix $p \geq 3$ and $\beta > 0$, and suppose that the plateau q is positive. Then, for every $L > 0$,*

$$\limsup_{t \rightarrow \infty} \int_0^{t-L} R(t, u)^2 du \leq \frac{d_L}{2} \beta_d^{-2/p} \delta_L^{2/p}. \quad (11.29)$$

In particular, (11.7) holds.

The idea is as follows. We choose the weight k in Lemma 11.10 proportional to the remote response row $u \mapsto R(T+L, u)$ on $[0, T]$. Then the overlap of $m_{N,T}$ with G_N is $\sqrt{2S_2(T+L, L)}$ in the limit. After normalization by $\sqrt{Q_{K,L}}$, the mean $m_{N,T}$ has energy close to E_{th} , and Lemma 11.9 forces this overlap to be small. The coefficient-Lipschitz approximations of Proposition 10.4 make Lemma 11.9 applicable.

Proof. Fix $L > 0$ and $K \geq 1$, and abbreviate $Q = Q_{K,L}$ and $\delta_{K,L} = E_{\text{th}} - h_{K,L}/Q^{p/2}$. Recall from (10.19) that $\delta_{K,L} \geq 0$, and note that $Q \geq c(2L) \geq q > 0$. The proof has four steps. We take the limits in the following order:

- (i) $N = N_j \rightarrow \infty$, with T, K, L , and the approximation parameter ε fixed;
- (ii) $T \rightarrow \infty$, along an arbitrary sequence;
- (iii) $\varepsilon \downarrow 0$;
- (iv) $K \rightarrow \infty$.

Step 1: the test direction. For $T > 0$, the response is positive by (11.2), and $S_2(T+L, L) = \int_0^T R(T+L, u)^2 du > 0$. We define

$$k_{T,L}(u) = \frac{R(T+L, u)}{S_2(T+L, L)^{1/2}}, \quad 0 \leq u \leq T, \quad G_N = \int_0^T k_{T,L}(u) dB_u.$$

The weight $k_{T,L}$ is continuous and deterministic: it depends on the CK solution and on the fixed times T and L , not on the disorder or the dimension. It satisfies $\int_0^T k_{T,L}^2 = 1$. Lemma 11.10 implies that

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\frac{\langle m_{N,T}, G_N \rangle}{N} \right] = \sqrt{2} \frac{\int_0^T R(T+L, u)^2 du}{S_2(T+L, L)^{1/2}} = \sqrt{2S_2(T+L, L)}. \quad (11.30)$$

Step 2: the normalized mean and its Lipschitz approximation. Let Π_{B_N} be the metric projection onto B_N , and let $\bar{m}_{N,T} = \Pi_{B_N}(m_{N,T}/\sqrt{Q})$. For $\varepsilon > 0$, let $\mathcal{A}_{N,T}^\varepsilon$ be the maps of Proposition 10.4, so that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \|\mathcal{A}_{N,T}^\varepsilon - \bar{m}_{N,T}\|^2 \leq \varepsilon^2,$$

as in (10.20). These maps take values in B_N . Conditionally on the initial point and on all Brownian paths, they are globally Lipschitz in J , with a constant that depends neither on N nor on the seed. The vector G_N is a function of the trunk Brownian path. It is therefore measurable with respect to this seed, which is independent of J .

We next compare energies. The vectors $m_{N,T}/\sqrt{Q}$, $\bar{m}_{N,T}$, and $\mathcal{A}_{N,T}^\varepsilon$ have norm at most $\sqrt{N/q}$. By homogeneity and (10.3), we have $\|\nabla H_N(x)\| \leq pQ_N q^{-(p-1)/2} \sqrt{N}$ when $\|x\| \leq \sqrt{N/q}$, where Q_N is the tensor norm (3.25). Integrating along segments and applying the Cauchy–Schwarz inequality, we find that, for random vectors x, y of norm at most $\sqrt{N/q}$,

$$\left| \frac{\mathbb{E} H_N(x) - \mathbb{E} H_N(y)}{N} \right| \leq D_q \left(\frac{\mathbb{E} \|x - y\|^2}{N} \right)^{1/2}, \quad D_q = p q^{-(p-1)/2} \sup_N (\mathbb{E} Q_N^2)^{1/2},$$

and $D_q < \infty$ by Lemma 3.6. We set

$$\gamma_T = \limsup_{N \rightarrow \infty} \left| \frac{\mathbb{E} H_N(m_{N,T})}{N} - h_{K,L} \right|, \quad \kappa_T = \limsup_{N \rightarrow \infty} \left(\frac{\mathbb{E} \|\bar{m}_{N,T} - m_{N,T}/\sqrt{Q}\|^2}{N} \right)^{1/2}.$$

Note that $H_N(m_{N,T}/\sqrt{Q}) = Q^{-p/2} H_N(m_{N,T})$ by homogeneity. Comparing energies through the intermediate vector $\bar{m}_{N,T}$, we obtain

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{\mathbb{E} H_N(\mathcal{A}_{N,T}^\varepsilon)}{N} &\geq Q^{-p/2} (h_{K,L} - \gamma_T) - D_q (\varepsilon + \kappa_T) \\ &= E_{\text{th}} - \delta_{K,L} - Q^{-p/2} \gamma_T - D_q (\varepsilon + \kappa_T). \end{aligned} \quad (11.31)$$

Step 3: the seed cost. By Step 2, the maps $\mathcal{A}_{N,T}^\varepsilon$ satisfy the hypotheses of Lemma 11.9. The seed is formed by the initial point and all Brownian paths, the Gaussian vector is the vector G_N of Step 1, and $\delta = \delta_{K,L} + Q^{-p/2} \gamma_T + D_q (\varepsilon + \kappa_T)$ by (11.31). Lemma 11.9 implies that

$$\limsup_{N \rightarrow \infty} \mathbb{E} \left| \frac{\langle \mathcal{A}_{N,T}^\varepsilon, G_N \rangle}{N} \right| \leq \beta_d^{-1/p} (\delta_{K,L} + Q^{-p/2} \gamma_T + D_q (\varepsilon + \kappa_T))^{1/p}.$$

To return to $m_{N,T}$, we use $\mathbb{E} \|G_N\|^2 = N$ and the Cauchy–Schwarz inequality; the fact that G_N and the approximation share the trunk seed plays no role here. We have

$$\left| \mathbb{E} \frac{\langle m_{N,T}/\sqrt{Q} - \mathcal{A}_{N,T}^\varepsilon, G_N \rangle}{N} \right| \leq \left(\frac{\mathbb{E} \|m_{N,T}/\sqrt{Q} - \bar{m}_{N,T}\|^2}{N} \right)^{1/2} + \left(\frac{\mathbb{E} \|\bar{m}_{N,T} - \mathcal{A}_{N,T}^\varepsilon\|^2}{N} \right)^{1/2}.$$

Combining the last two displays with (11.30), we find that, for every $T > 0$,

$$\sqrt{\frac{2S_2(T+L, L)}{Q}} \leq \beta_d^{-1/p} (\delta_{K,L} + Q^{-p/2} \gamma_T + D_q(\varepsilon + \kappa_T))^{1/p} + \varepsilon + \kappa_T. \quad (11.32)$$

Step 4: the long-time limits. Let $T_n \rightarrow \infty$ be an arbitrary sequence. Using (10.18), we have $\gamma_{T_n} \rightarrow 0$. For κ_T , note that $\|\Pi_{B_N}(x) - x\| = (\|x\| - \sqrt{N})_+$, and that $|\sqrt{y/Q} - 1| \leq |y - Q|/Q$ for $y \geq 0$. With $y = \|m_{N,T}\|^2/N$, we find $\kappa_T^2 \leq \limsup_N \mathbb{E}[\|m_{N,T}\|^2/N - Q]^2/Q^2$, and (10.17) implies that $\kappa_{T_n} \rightarrow 0$. The left side of (11.32) does not depend on ε . Taking $T = T_n \rightarrow \infty$ in (11.32) and then $\varepsilon \downarrow 0$, we obtain

$$\limsup_{n \rightarrow \infty} S_2(T_n + L, L) \leq \frac{Q}{2} \beta_d^{-2/p} \delta_{K,L}^{2/p}.$$

Since the sequence T_n was arbitrary, the same bound holds for $\limsup_{t \rightarrow \infty} S_2(t, L)$. This upper limit does not depend on K . As $K \rightarrow \infty$, we have $Q_{K,L} \rightarrow d_L$ and $\delta_{K,L} \rightarrow \delta_L$, by Lemma 10.5(i). This proves (11.29). As $d_L \leq 1$ and $\delta_L \rightarrow 0$ by (11.26), the right side of (11.29) tends to zero as $L \rightarrow \infty$, which is (11.7). $\square$

Together with Steps 1, 2, and 4 of the proof of Theorem 11.1, Proposition 11.11 proves strong L^2 convergence of the response rows, the exact limits of the squared tails, and $\chi(t, 0) = o(\sqrt{t})$ for every pure model with $p \geq 3$.

Persistent integrated violation of the fluctuation–dissipation relation

By Theorem 11.5, the backward FDT defect $E(t, s) = \partial_s C(t, s) - R(t, s)$ tends to zero in $L^2(0, t)$ as $t \rightarrow \infty$. In this section, we prove that, above the dynamical transition, the defect nevertheless does not disappear. Its positive part keeps an integrated mass bounded below by a positive constant on the part of the row away from any fixed initial interval $[0, S]$ and any fixed recent strip $[t - L, t]$. This is what we mean by a persistent integrated violation: the defect is small in L^2 , but it has positive integrated mass, spread over a history whose length grows with t . The two statements are compatible. By the Cauchy–Schwarz inequality, a mass m spread over an interval of length ℓ forces a squared norm of at least m^2/ℓ , and here ℓ grows like t .

The main result is Theorem 12.4, and its proof has two ingredients. The first is Proposition 12.1. The energy identity determines the response integrated against the weight $pC(t, u)^{p-1}$ over the whole row, and stationary convergence determines the contribution of a recent strip. The remainder identifies one weighted ratio of the slow response with the number $x = (p - 2)(1 - q)/q$. The second ingredient is Lemma 12.3, an elementary inequality for one absolutely continuous function, which converts a deficit in this weighted response budget into a lower bound on the unweighted positive defect. This inequality is sharp (Remark 12.5).

The proofs use the energy identity (4.1), the selected plateau and energy of Theorem 8.5 together with stationary convergence (8.11), forgetting (8.12), the absolute continuity of correlation rows from Theorem 5.3, and Theorem 11.5. The results of this section prove (1.10) and (1.11) in Theorem 1.2. We return to them in Section 17.

The physical mechanism is the one described by Cugliandolo, Dean, and Kurchan in [54, arXiv version 2, equations (12)–(14), p. 3, equation (18), p. 4, and the integrated-bounds discussion on p. 3]. Entropy production bounds the violation of the fluctuation–dissipation relation, and the violation is therefore small at late times; its integral over a time interval that grows with the age of the system may nevertheless remain of order one. In our setting, the smallness at late times is Theorem 11.5, which rests on the entropy bound of Theorem 5.3. The lower bound on the integrated violation for the CK solution is proved here, from the energy identity.

Standing assumptions and notation. Throughout this section, we fix an integer $p \geq 3$ and $\beta > \beta_d$. By Theorem 8.5, the plateau $q = q_\beta$ is the larger root of

$$\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1, \quad (12.1)$$

and the limiting energy is

$$e_\infty = e_\beta(\infty) = \lim_{t \rightarrow \infty} e_\beta(t) = \beta[1 + (p-2)q^{p-1} - (p-1)q^p]. \quad (12.2)$$

Since $\beta > \beta_d$, we have $q > q_d = (p-2)/(p-1)$, as noted before Theorem 8.5. Indeed, the function $g(y) = y^{p-2}(1-y)^2$ is decreasing on $[q_m, 1]$, where $q_m = (p-2)/p < q_d$, and $\beta^2 p(p-1)g(q_d) = \beta^2/\beta_d^2 > 1$. The larger root of (12.1) therefore lies in $(q_d, 1)$. We set

$$x = \frac{(p-2)(1-q)}{q}, \quad d_* = 1 - \frac{e_\infty}{\beta}. \quad (12.3)$$

The inequality $q > q_d$ is equivalent to $(p-2)(1-q) < q$, which means that $x \in (0, 1)$. Using (12.2), we find

$$d_* = (p-1)q^p - (p-2)q^{p-1} = q^p \left[(p-1) - \frac{p-2}{q} \right] = (1-x)q^p > 0,$$

where we used $(p-1) - (p-2)/q = 1 - (p-2)(1-q)/q$. Equivalently,

$$\frac{e_\infty}{\beta} - (1 - q^p) = q^p - d_* = xq^p. \quad (12.4)$$

We write $E(t, s) = \partial_s C(t, s) - R(t, s)$ for $t > s > 0$, as in Section 11.3, and $E_+ = \max\{E, 0\}$. For $t > 0$, we use the right earlier derivative at $s = 0$ and set $E(t, t) = 0$, consistently with the left earlier derivative at the diagonal. Recall from Theorem 5.3 that, for each t , the row $u \mapsto C(t, u)$ is absolutely continuous on $[0, t]$, and that its derivative on $(0, t)$ is the continuous function $\partial_u C(t, u)$. In particular, $E(t, \cdot)$ is integrable on $(0, t)$. Let (c, r) be the stationary profile, with $r = -c'$, $c(0) = 1$, and c decreasing to q . Stationary convergence (8.11) implies that, for every fixed $A \geq 0$,

$$\lim_{t \rightarrow \infty} \sup_{0 \leq v \leq A} \{|C(t, t-v) - c(v)| + |R(t, t-v) - r(v)|\} = 0. \quad (12.5)$$

12.1 A weighted slow ratio

Proposition 12.1 (An exact weighted slow ratio). *For each fixed $A > 0$, the denominator below is positive for all large t , the ratio has a limit as $t \rightarrow \infty$, and the subsequent limit as $A \rightarrow \infty$ is*

$$\lim_{A \rightarrow \infty} \lim_{t \rightarrow \infty} \frac{p \int_0^{t-A} C(t, u)^{p-1} R(t, u) \, du}{C(t, t-A)^p - C(t, 0)^p} = x. \quad (12.6)$$

Further, the weighted signed defect has the full-row limit

$$\lim_{t \rightarrow \infty} p \int_0^t C(t, u)^{p-1} E(t, u) \, du = d_*. \quad (12.7)$$

The idea is as follows. By the energy identity, the total weighted response $p \int_0^t C^{p-1} R$ tends to e_∞/β . A recent strip of length A contributes $1 - c(A)^p$, as for an equilibrium profile, and its contribution tends to $1 - q^p$ as $A \rightarrow \infty$. By (12.4), the older part of the row carries the remaining weighted response xq^p , while the correlation raised to the power p falls from about q^p to zero across it.

Proof. By the energy identity (4.1), we have $p \int_0^t C(t, u)^{p-1} R(t, u) \, du = e_\beta(t)/\beta$, which tends to $e_\infty/\beta = 1 - d_*$. Fix $A > 0$. Using the change of variables $u = t - v$ and (12.5), we find

$$\begin{aligned} p \int_{t-A}^t C(t, u)^{p-1} R(t, u) \, du &= p \int_0^A C(t, t-v)^{p-1} R(t, t-v) \, dv \\ &\longrightarrow p \int_0^A c(v)^{p-1} r(v) \, dv = 1 - c(A)^p, \end{aligned}$$

where we used $r = -c'$ and $c(0) = 1$ in the last step. Subtracting, we see that the numerator in (12.6) converges to $c(A)^p - d_*$. We have $C(t, 0) \rightarrow 0$ by (8.12) and $C(t, t-A) \rightarrow c(A)$ by (12.5). The denominator therefore converges to $c(A)^p \geq q^p > 0$; in particular, it is positive for all large t . The ratio converges to

$$\frac{c(A)^p - d_*}{c(A)^p} = 1 - \frac{d_*}{c(A)^p}.$$

As $A \rightarrow \infty$, we have $c(A) \rightarrow q$, and this tends to $1 - d_*/q^p = x$. The two limits are taken in this order, and no interchange of limits occurs.

For (12.7), fix $t > 0$. The row $u \mapsto C(t, u)$ is absolutely continuous on $[0, t]$ with values in $[0, 1]$. It follows that $u \mapsto C(t, u)^p$ is absolutely continuous, with derivative $pC(t, u)^{p-1} \partial_u C(t, u)$ almost everywhere. Integrating over $[0, t]$ and using $C(t, t) = 1$, we have

$$p \int_0^t C(t, u)^{p-1} \partial_u C(t, u) \, du = 1 - C(t, 0)^p.$$

Subtracting the energy identity, we arrive at the exact identity

$$p \int_0^t C(t, u)^{p-1} E(t, u) \, du = 1 - C(t, 0)^p - \frac{e_\beta(t)}{\beta},$$

whose right side tends to $1 - (1 - d_*) = d_*$. This argument does not require the row to be monotone. $\square$

Remark 12.2 (Interpretation of x). In the asymptotic solution of Cugliandolo and Kurchan, the number $x = (p - 2)(1 - q)/q$ is the value of the fluctuation–dissipation ratio in the aging regime; see the paragraph after [51, arXiv version 1, Equation (12), p. 6]. There, for late pairs of times with correlation below the plateau, the response is predicted to be x times the earlier derivative of the correlation, in the normalization of this paper. Proposition 12.1 identifies x with one ratio of weighted integrals over the older part of the row. It does not show that the ratio $R(t, s)/\partial_s C(t, s)$ converges to x along late pairs; we discuss that statement, and its integrated form, in Section 17.

12.2 Transport of the response budget

We now remove the weight C^{p-1} . On the interval $[S, t - L]$, the correlation $u \mapsto C(t, u)$ rises from about zero to about $c(L)$. The weighted response budget $p \int_S^{t-L} C^{p-1} R$ on this interval tends to $c(L)^p - d_*$, which is smaller than $c(L)^p$. If the defect were nonpositive on the interval, the correlation could not rise faster than the response allows, and this budget would control the increase of $C(t, \cdot)^p$. The following lemma quantifies the failure of this control in terms of the positive part of the defect. No monotonicity is required.

To see the idea, let $z \geq 0$ be absolutely continuous and $r_0 \geq 0$. If $z' \leq r_0$ almost everywhere, then $(z^p)' = pz^{p-1}z' \leq pz^{p-1}r_0$, which implies that $z(v)^p \leq z(u)^p + p \int_u^v z^{p-1}r_0$. The lemma bounds the positive part of $z' - r_0$ from below when this inequality fails.

Lemma 12.3 (Response-budget transport). *Let $u < v$, let $z : [u, v] \rightarrow [0, \infty)$ be absolutely continuous, and let $r_0 \geq 0$ be integrable on $[u, v]$. Set*

$$J = \int_u^v z(s)^{p-1} r_0(s) \, ds, \quad e_0 = z' - r_0, \quad D = \left[z(v) - (z(u)^p + pJ)^{1/p} \right]_+.$$

Then

$$\int_u^v (e_0)_+ \, ds \geq D, \quad (v - u) \int_u^v (e_0)_+^2 \, ds \geq D^2. \quad (12.8)$$

Here J is finite because z is bounded on $[u, v]$, and e_0 is integrable because z' and r_0 are. If $(e_0)_+$ is not square integrable, the left side of the second inequality is infinite.

Proof. Let $\bar{z} = (z(u)^p + pJ)^{1/p}$. If $\bar{z} \geq z(v)$, then $D = 0$, and both claims hold because $(e_0)_+ \geq 0$. Suppose that $\bar{z} < z(v)$, so that $D = z(v) - \bar{z} > 0$.

We first treat the case $\bar{z} > 0$. For $a > 0$, we define the clipped power weight and its primitive by

$$w_a(y) = \min \left\{ \left(\frac{y}{a} \right)^{p-1}, 1 \right\}, \quad V_a(y) = \begin{cases} \frac{y^p}{pa^{p-1}}, & 0 \leq y \leq a, \\ y - \frac{p-1}{p}a, & y \geq a. \end{cases}$$

The function V_a is continuously differentiable on $[0, \infty)$, and its derivative $V'_a = w_a$ takes values in $[0, 1]$. Then $V_a \circ z$ is absolutely continuous, with derivative $w_a(z)z'$ almost everywhere. We use two pointwise bounds. As $0 \leq w_a \leq 1$, we have $(e_0)_+ \geq w_a(z)e_0$. As $r_0 \geq 0$ and $w_a(y) \leq (y/a)^{p-1}$, we have $w_a(z)r_0 \leq z^{p-1}r_0/a^{p-1}$. Integrating, we obtain

$$\int_u^v (e_0)_+ \, ds \geq \int_u^v w_a(z)z' \, ds - \int_u^v w_a(z)r_0 \, ds \geq V_a(z(v)) - V_a(z(u)) - \frac{J}{a^{p-1}}.$$

We now take $a = \bar{z}$ and compute the right side. From $z(u)^p \leq \bar{z}^p$, we have $z(u) \leq \bar{z}$ and $V_{\bar{z}}(z(u)) = z(u)^p / (p\bar{z}^{p-1})$. From $z(v) > \bar{z}$, we have $V_{\bar{z}}(z(v)) = z(v) - (p-1)\bar{z}/p$. Using $z(u)^p + pJ = \bar{z}^p$, we find

$$\begin{aligned} V_{\bar{z}}(z(v)) - V_{\bar{z}}(z(u)) - \frac{J}{\bar{z}^{p-1}} &= z(v) - \frac{p-1}{p}\bar{z} - \frac{z(u)^p + pJ}{p\bar{z}^{p-1}} \\ &= z(v) - \frac{p-1}{p}\bar{z} - \frac{\bar{z}}{p} = z(v) - \bar{z} = D. \end{aligned}$$

This is the first inequality in (12.8). The primitive $V_{\bar{z}}$ is evaluated only at the endpoints, and excursions and returns of z inside $[u, v]$ do not affect the bound.

If $\bar{z} = 0$, then $z(u) = 0$ and $J = 0$. For $0 < \varepsilon < z(v)$, the computation above with $a = \varepsilon$ gives

$$\int_u^v (e_0)_+ \, ds \geq V_\varepsilon(z(v)) - V_\varepsilon(0) - 0 = z(v) - \frac{p-1}{p}\varepsilon.$$

Letting $\varepsilon \downarrow 0$, we find $\int_u^v (e_0)_+ \geq z(v) = D$. Finally, the Cauchy–Schwarz inequality gives $D^2 \leq (\int_u^v (e_0)_+)^2 \leq (v-u) \int_u^v (e_0)_+^2$, which is the second inequality. $\square$

12.3 The persistent defect

Theorem 12.4 (A positive slow defect and a physical rate constraint). *Set*

$$m_L = c(L) - [c(L)^p - d_*]^{1/p}, \quad m_\infty = q(1 - x^{1/p}) > 0.$$

For every fixed $S \geq 0$ and $L > 0$,

$$\liminf_{t \rightarrow \infty} \int_S^{t-L} E_+(t, u) \, du \geq m_L, \quad (12.9)$$

$$\liminf_{t \rightarrow \infty} (t - S - L) \int_S^{t-L} E_+(t, u)^2 \, du \geq m_L^2. \quad (12.10)$$

Further, $m_L \uparrow m_\infty$ as $L \rightarrow \infty$. In particular,

$$\lim_{t \rightarrow \infty} \int_0^t E_+(t, u)^2 \, du = 0, \quad \liminf_{t \rightarrow \infty} t \int_0^t E_+(t, u)^2 \, du \geq q^2(1 - x^{1/p})^2 > 0. \quad (12.11)$$

The backward FDT defect tends to zero in $L^2(0, t)$, but not at a rate $o(t^{-1/2})$: by (12.11), we have $\liminf_t \sqrt{t} \|E(t, \cdot)\|_{L^2(0, t)} \geq m_\infty$. The constant m_L is the limit of the transport distance D of Lemma 12.3 on the interval $[S, t - L]$; we show in the proof that its radicand $c(L)^p - d_*$ is positive.

Proof. We first compute the limits of the endpoint values and of the weighted response budget on $[S, t - L]$. We then apply Lemma 12.3 on that interval. Finally, we study m_L and deduce (12.11).

Step 1: the budget on the remote part of the row. Fix $S \geq 0$ and $L > 0$, and let $t > S + L$. We set

$$J_t = \int_S^{t-L} C(t, u)^{p-1} R(t, u) \, du.$$

As in the proof of Proposition 12.1, with $A = L$, we have $p \int_0^{t-L} C(t, u)^{p-1} R(t, u) \, du \rightarrow c(L)^p - d_*$. The omitted initial interval is negligible. Indeed, since $0 \leq C \leq 1$, we have

$$0 \leq p \int_0^S C(t, u)^{p-1} R(t, u) \, du \leq pS \sup_{0 \leq u \leq S} R(t, u) \rightarrow 0,$$

by (8.12). We conclude that $pJ_t \rightarrow c(L)^p - d_*$. Further, $C(t, S) \rightarrow 0$ by (8.12), and $C(t, t - L) \rightarrow c(L)$ by (12.5).

Step 2: transport. We use Lemma 12.3 on $[S, t - L]$ with $z(u) = C(t, u)$ and $r_0(u) = R(t, u)$. The function z is absolutely continuous with values in $[0, 1]$, r_0 is nonnegative and bounded, and $e_0 = z' - r_0 = E(t, \cdot)$ almost everywhere on $(S, t - L)$. The transport distance is

$$D_t = \left[C(t, t - L) - (C(t, S)^p + pJ_t)^{1/p} \right]_+.$$

By Step 1 and the continuity of $y \mapsto y^{1/p}$ and of $y \mapsto y_+$, we have

$$\lim_{t \rightarrow \infty} D_t = \left[c(L) - (c(L)^p - d_*)^{1/p} \right]_+ = m_L.$$

The radicand is positive: $c(L) \geq q$ and (12.4) give $c(L)^p - d_* \geq q^p - d_* = xq^p > 0$. It is also smaller than $c(L)^p$, because $d_* > 0$. Hence $m_L > 0$. Using Lemma 12.3, we obtain

$$\int_S^{t-L} E_+(t, u) \, du \geq D_t, \quad (t - L - S) \int_S^{t-L} E_+(t, u)^2 \, du \geq D_t^2.$$

Taking the lower limit as $t \rightarrow \infty$ proves (12.9) and (12.10).

Step 3: the constants and the rate. Let $\varphi(y) = y - (y^p - d_*)^{1/p}$ for $y > d_*^{1/p}$. Then

$$\varphi'(y) = 1 - y^{p-1}(y^p - d_*)^{(1-p)/p} = 1 - \left(\frac{y^p}{y^p - d_*}\right)^{(p-1)/p} < 0,$$

and φ is strictly decreasing. We have $q^p > d_*$, and $c(L)$ decreases to q as $L \rightarrow \infty$. It follows that $m_L = \varphi(c(L))$ increases to

$$\varphi(q) = q - (q^p - d_*)^{1/p} = q - (xq^p)^{1/p} = q(1 - x^{1/p}) = m_\infty,$$

where we used (12.4). As $x < 1$, we have $m_\infty > 0$.

We have $0 \leq E_+ \leq |E|$, and Theorem 11.5 implies that $\int_0^t E_+(t, u)^2 du \rightarrow 0$. For the lower bound in (12.11), fix $L > 0$ and take $S = 0$. For $t > L$, we have

$$t \int_0^t E_+(t, u)^2 du \geq (t - L) \int_0^{t-L} E_+(t, u)^2 du.$$

The lower limit of the right side is at least m_L^2 , by (12.10). Letting $L \rightarrow \infty$, we find $\liminf_t t \int_0^t E_+^2 \geq m_\infty^2 = q^2(1 - x^{1/p})^2$. This completes the proof. $\square$

The lower limit in (12.9) is nonincreasing in L , because the integrand is nonnegative and the interval shrinks as L grows. It therefore has a limit in $[0, \infty]$ as $L \rightarrow \infty$. Since $m_{L'} \geq m_L$ for $L' \geq L$, we obtain, for every fixed $S \geq 0$,

$$\lim_{L \rightarrow \infty} \liminf_{t \rightarrow \infty} \int_S^{t-L} E_+(t, u) du \geq m_\infty. \quad (12.12)$$

Together with (12.11), this is (1.11).

Remark 12.5 (Sharpness of the transport inequality). The constant in Lemma 12.3 is sharp for the problem with prescribed endpoint values, interval length, and weighted response budget. Let nonnegative endpoint values z_0, z_1 and a budget $J \geq 0$ be given, with $\bar{z} = (z_0^p + pJ)^{1/p} < z_1$, and let $0 < \varepsilon < v - u$. On $[u, u + \varepsilon]$, we let z increase linearly from z_0 to $\bar{z}$ and set $r_0 = z'$. Then $e_0 = 0$ there, and

$$\int_u^{u+\varepsilon} z^{p-1} r_0 ds = \int_u^{u+\varepsilon} z^{p-1} z' ds = \frac{\bar{z}^p - z_0^p}{p} = J.$$

On $[u + \varepsilon, v]$, we let z increase linearly from $\bar{z}$ to z_1 and set $r_0 = 0$. Then $e_0 = (z_1 - \bar{z})/(v - u - \varepsilon)$ there, and

$$\int_u^v (e_0)_+ ds = z_1 - \bar{z} = D, \quad (v - u) \int_u^v (e_0)_+^2 ds = \frac{v - u}{v - u - \varepsilon} D^2.$$

The first inequality in (12.8) is an equality for each ε , and the second becomes an equality as $\varepsilon \downarrow 0$. This optimality concerns the inequality of Lemma 12.3 for the stated class of functions. We do not know whether the constants m_L and m_∞ are optimal for the CK solution.

PART IV

Three-time geometry in the cubic model

Saturated tree vectors and the product inequality

This section and the next concern the cubic model. In this section, we study unit vectors of the tree Hilbert space whose cubic tree energy is close to the ceiling E_{th} of Theorem 9.2. We prove two results about such vectors. The first, Theorem 13.2, states that the root operator $\mathcal{I}^*v$ of a vector close to the ceiling is close, in suitable coordinates, to one universal integral operator on $L^2(0,1)$. The second, Theorem 13.5, is a product inequality for three such vectors whose overlaps are nonnegative at every height. Lemmas 13.1 and 13.4 are the intermediate steps.

The section is deterministic. It uses the tree Hilbert space of Section 7 and the cubic ceiling of Theorem 9.2, with the retained defects (9.2) and (9.3) proved in Section 9. Nothing about the CK equations is used. In Section 14, we apply both theorems to the normalized prefix centers of the cubic CK solution: Theorem 13.5 gives the three-time product inequality of Theorem 14.3, and Theorem 13.2 is the main input of the susceptibility bound in Theorem 14.7.

Notation. Throughout this section, $p = 3$, so that the grafting map $\mathcal{I}: \mathcal{H} \otimes \mathcal{H} \rightarrow \mathcal{H}$ is binary. Recall from Section 7 that $\mathcal{H}$ is a real separable Hilbert space, that $\mathcal{I}$ is an isometry, and that its range is the orthogonal complement of the one-leaf summand. The one-leaf summand is the subspace of height zero, and we have

$$\mathcal{I}\mathcal{I}^* = \text{Id} - \Pi_0. \quad (13.1)$$

We write Π_j for the projection onto trees of height exactly j and $\Pi_{\leq H} = \sum_{j=0}^H \Pi_j$; the projections Π_j are orthogonal and sum to the identity. As in Section 7, we identify a tensor $x \otimes y$ with the Hilbert–Schmidt operator $|x\rangle\langle y|$, which maps z to $\langle y, z \rangle x$. Then $\mathcal{I}^*v$ is a Hilbert–Schmidt operator on $\mathcal{H}$, and $\|\mathcal{I}^*v\|_{\text{HS}} \leq \|v\|$. We write $\langle A, B \rangle_{\text{HS}} = \text{tr}(A^*B)$, so that

$$\langle |a\rangle\langle b|, |c\rangle\langle d| \rangle_{\text{HS}} = \langle a, c \rangle \langle b, d \rangle.$$

For $v \in \mathcal{H}$, let

$$f(v) = \langle v, \mathcal{I}(v \otimes v) \rangle = \langle \mathcal{I}^*v, |v\rangle\langle v| \rangle_{\text{HS}}, \quad F_* = \frac{2\sqrt{2}}{3},$$

and, for a unit vector v , let $\delta(v) = F_* - f(v)$ be its *deficit*. By Theorem 9.2, we have $\delta(v) > 0$ for every unit vector v . The cubic tree energy of (7.5) is $h(v) = \sqrt{3}f(v)$, and

$E_{\text{th}} = \sqrt{3}F_*$. Hence the deficit $\delta(v) = (E_{\text{th}} - h(v))/\sqrt{3}$ is the energy deficit divided by $\sqrt{3}$.

For a unit vector v , we use the height masses of Theorem 9.2:

$$v_j = \Pi_j v, \quad b_j = \|v_j\|^2, \quad B_j = \sum_{i \leq j} b_i, \quad a_j = \sqrt{b_j}, \quad \bar{b} = \sup_{j \geq 0} b_j,$$

with $B_{-1} = B_{-2} = 0$. Note that $B_j \uparrow 1$ as $j \rightarrow \infty$.

The height rule. A graft has height one plus the larger of the heights of its two children. This means that $\mathcal{I}$ maps $\Pi_i \mathcal{H} \otimes \Pi_k \mathcal{H}$ into $\Pi_{\max(i,k)+1} \mathcal{H}$, and

$$\Pi_j \mathcal{I}(x \otimes y) = \sum_{\max(i,k)=j-1} \mathcal{I}(\Pi_i x \otimes \Pi_k y) \quad (j \geq 1), \quad \Pi_0 \mathcal{I} = 0. \quad (13.2)$$

Taking adjoints, we find that $\mathcal{I}^* \Pi_0 = 0$ and, for $j \geq 1$,

$$\mathcal{I}^* \Pi_j v = \sum_{\max(i,k)=j-1} \Pi_i (\mathcal{I}^* v) \Pi_k. \quad (13.3)$$

In words, the height- j component of v is carried by the blocks $\Pi_i (\mathcal{I}^* v) \Pi_k$ of the root operator with $\max(i,k) = j-1$. We call j the *root height* of these blocks. Blocks with different root heights are orthogonal in the Hilbert-Schmidt inner product, and the restriction of $\mathcal{I}^*$ to $\Pi_j \mathcal{H}$ is an isometry for $j \geq 1$, by (13.1).

The retained defects. Let v be a unit vector with deficit δ . By (13.2) and the isometry of $\mathcal{I}$ (see also Lemma 9.1), the projected graft $g_j = \Pi_j \mathcal{I}(v \otimes v)$ satisfies

$$\|g_j\|^2 = \sum_{\max(i,k)=j-1} b_i b_k = B_{j-1}^2 - B_{j-2}^2 = b_{j-1} s_j^2, \quad s_j = \sqrt{B_{j-1} + B_{j-2}}. \quad (13.4)$$

Let θ_j be the angle between v_j and g_j , with the convention $\cos \theta_j = 1$ when $v_j = 0$ or $g_j = 0$. We use three consequences of Theorem 9.2. The first inequality in (9.2) implies that

$$\bar{b} \leq \left(\sum_{j \geq 0} b_j^2 \right)^{1/2} \leq (4\sqrt{2}\delta)^{1/2} = 2^{5/4}\delta^{1/2}, \quad (13.5)$$

and (9.3) states that

$$\sum_{j \geq 1} a_j a_{j-1} s_j (1 - \cos \theta_j) \leq \delta, \quad \frac{1}{2} \sum_{j \geq 1} s_j (a_j - a_{j-1})^2 \leq \delta. \quad (13.6)$$

13.1 The universal operator at a saturated root

By (13.5), a vector close to the ceiling has small mass at every height. The retained defects (13.6) give more information: they bound weighted sums of the variations $a_j - a_{j-1}$ of

the height amplitudes and of the angular errors $1 - \cos \theta_j$ between v_j and the projected graft g_j of the lower components. The next lemma shows that, as a consequence, the root operator of v is close to an operator whose coefficients in the normalized height directions $v_j / \sqrt{b_j}$, for $b_j > 0$, depend only on the height masses.

Lemma 13.1 (Coherent approximation of a root). *Let v be a unit tree vector with $\delta = \delta(v) \leq 1$. There exists a symmetric Hilbert–Schmidt operator $\tilde{A}_v$ with $\|\tilde{A}_v\|_{\text{HS}} = 1$ and height blocks*

$$\Pi_i \tilde{A}_v \Pi_k = \frac{|v_i\rangle\langle v_k|}{\sqrt{B_h + B_{h-1}}}, \quad h = \max(i, k), \quad (13.7)$$

where a zero denominator means a zero block, such that

$$\|\mathcal{I}^* v - \tilde{A}_v\|_{\text{HS}} \leq K \delta^{1/4}. \quad (13.8)$$

Here K is an absolute constant; one can take $K = 5$.

Proof. The proof has three steps. First, we define $\tilde{A}_v$ one root height at a time and compute its norm. Second, we compute the error at each root height. Third, we sum the errors: at root heights above a small cumulative mass ε , we use the retained defects; the remaining root heights form an initial segment of small mass.

Step 1: definition and normalization. For $j \geq 1$, let g_j and s_j be as in (13.4). We define the component of $\tilde{A}_v$ at root height j by

$$\tilde{A}^{(j)} = \frac{\mathcal{I}^* g_j}{s_j} \quad \text{if } s_j > 0, \quad \tilde{A}^{(j)} = 0 \quad \text{if } s_j = 0,$$

and we set $\tilde{A}_v = \sum_{j \geq 1} \tilde{A}^{(j)}$. Using (13.2) and $\mathcal{I}^* \mathcal{I} = \text{Id}$, we have

$$\mathcal{I}^* g_j = \sum_{\max(i, k)=j-1} |v_i\rangle\langle v_k|.$$

With $h = j - 1$, the denominator s_j equals $\sqrt{B_h + B_{h-1}}$, and the blocks of $\tilde{A}_v$ are those in (13.7). The block with indices (k, i) is the adjoint of the block with indices (i, k) , and therefore $\tilde{A}_v$ is symmetric. By (13.4) and the isometry of $\mathcal{I}$, we have $\|\tilde{A}^{(j)}\|_{\text{HS}}^2 = \|g_j\|^2 / s_j^2 = b_{j-1}$ when $s_j > 0$. When $s_j = 0$, we have $B_{j-1} = 0$, and in particular $b_{j-1} = 0$. The components with different root heights are orthogonal. Summing over j , we find

$$\|\tilde{A}_v\|_{\text{HS}}^2 = \sum_{j \geq 1} b_{j-1} = \sum_{h \geq 0} b_h = 1.$$

Step 2: the error at one root height. By (13.3), we have $\mathcal{I}^*v = \sum_{j \geq 1} \mathcal{I}^*v_j$, and $\mathcal{I}^*v_j$ lies in the same root-height component as $\tilde{A}^{(j)}$. By orthogonality,

$$\|\mathcal{I}^*v - \tilde{A}_v\|_{\text{HS}}^2 = \sum_{j \geq 1} \|\mathcal{I}^*v_j - \tilde{A}^{(j)}\|_{\text{HS}}^2.$$

For $j \geq 1$, the vector v_j lies in the range of $\mathcal{I}$, and (13.1) gives $\|\mathcal{I}^*v_j\|_{\text{HS}}^2 = b_j$. If $s_j > 0$, then, since g_j also lies in the range of $\mathcal{I}$,

$$\langle \mathcal{I}^*v_j, \tilde{A}^{(j)} \rangle_{\text{HS}} = \frac{\langle v_j, \mathcal{I}\mathcal{I}^*g_j \rangle}{s_j} = \frac{\langle v_j, g_j \rangle}{s_j} = a_j a_{j-1} \cos \theta_j,$$

where we used $\|g_j\| = a_{j-1}s_j$ from (13.4). The error at root height j is

$$\|\mathcal{I}^*v_j - \tilde{A}^{(j)}\|_{\text{HS}}^2 = b_j + b_{j-1} - 2a_j a_{j-1} \cos \theta_j = (a_j - a_{j-1})^2 + 2a_j a_{j-1}(1 - \cos \theta_j). \quad (13.9)$$

If $s_j = 0$, then $b_{j-1} = 0$ and the error is b_j , in agreement with (13.9).

Step 3: summation. If $\delta = 1$, then (13.8) holds with $K = 2$, because both operators have Hilbert–Schmidt norm at most one. We assume from now on that $\delta < 1$. Fix $\varepsilon \in (0, 1)$.

We first consider the root heights $j \geq 1$ with $B_{j-2} \geq \varepsilon$. For these, $s_j^2 = B_{j-1} + B_{j-2} \geq 2\varepsilon$. Multiplying each term of (13.9) by $s_j/\sqrt{2\varepsilon} \geq 1$ and using (13.6), we obtain

$$\sum_{j: B_{j-2} \geq \varepsilon} \|\mathcal{I}^*v_j - \tilde{A}^{(j)}\|_{\text{HS}}^2 \leq \frac{1}{\sqrt{2\varepsilon}} \sum_{j \geq 1} s_j \left[(a_j - a_{j-1})^2 + 2a_j a_{j-1}(1 - \cos \theta_j) \right] \leq \frac{4\delta}{\sqrt{2\varepsilon}}.$$

The remaining root heights form an initial segment. Let $m_\varepsilon = \min\{k \geq 0 : B_k \geq \varepsilon\}$, which is finite because $B_k \uparrow 1$. The condition $B_{j-2} < \varepsilon$ holds exactly when $1 \leq j \leq m_\varepsilon + 1$, because B is nondecreasing. On this segment,

$$\sum_{j=1}^{m_\varepsilon+1} b_j \leq B_{m_\varepsilon-1} + b_{m_\varepsilon} + b_{m_\varepsilon+1} \leq \varepsilon + 2\bar{b}, \quad \sum_{j=1}^{m_\varepsilon+1} b_{j-1} = B_{m_\varepsilon} \leq \varepsilon + \bar{b}.$$

Using $\|x - y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$ together with $\|\mathcal{I}^*v_j\|_{\text{HS}}^2 = b_j$ and $\|\tilde{A}^{(j)}\|_{\text{HS}}^2 = b_{j-1}$, we find that the errors on this segment sum to at most $4\varepsilon + 6\bar{b}$.

Adding the two bounds and using (13.5), we obtain

$$\|\mathcal{I}^*v - \tilde{A}_v\|_{\text{HS}}^2 \leq 4\varepsilon + 6 \cdot 2^{5/4} \delta^{1/2} + \frac{4\delta}{\sqrt{2\varepsilon}}.$$

We choose $\varepsilon = \delta^{1/2}$. Then $\delta^{3/4} \leq \delta^{1/2}$, because $\delta < 1$, and the right side is at most $(4 + 6 \cdot 2^{5/4} + 2\sqrt{2})\delta^{1/2} \leq 25\delta^{1/2}$. This proves (13.8) with $K = 5$. $\square$

When the height mesh $\bar{b}$ is small, the denominator in (13.7) is close to $\sqrt{2\bar{B}_h}$, and B_h varies almost continuously with the height. In the coordinate $x = B_h$ on $(0, 1)$, the blocks then become the values of one function of two variables. This is the content of the next theorem.

Theorem 13.2 (Universal threshold root). *Let $\mathcal{K}$ be a fixed separable infinite-dimensional Hilbert space, and let $\mathbf{1}$ denote the constant function one in $L^2(0, 1)$. If $\|v_n\| = 1$ and $f(v_n) \rightarrow F_*$, then there exist isometric embeddings $V_n: \mathcal{H} \rightarrow L^2(0, 1) \oplus \mathcal{K}$ satisfying $V_n v_n = \mathbf{1}$, for which*

$$V_n \mathcal{I}^* v_n V_n^* \longrightarrow \mathcal{T} \oplus 0 \quad \text{in Hilbert–Schmidt norm,} \quad T(x, y) = \frac{1}{\sqrt{2 \max(x, y)}}. \quad (13.10)$$

Here $\mathcal{T}$ is the integral operator on $L^2(0, 1)$ with kernel T . The operator $\mathcal{T}$ is positive and trace class, with the positive frame

$$\mathcal{T} = \frac{1}{\sqrt{2}} |\mathbf{1}\rangle \langle \mathbf{1}| + \frac{1}{2\sqrt{2}} \int_0^1 z^{-3/2} |\mathbf{1}_{[0,z]}\rangle \langle \mathbf{1}_{[0,z]}| \, dz. \quad (13.11)$$

In particular, $\text{tr } \mathcal{T} = \sqrt{2}$ and $\|\mathcal{T}\|_{\text{HS}} = 1$.

Remark 13.3 (Interpretation of the universal kernel). In the coordinates of Theorem 13.2, a point $x \in (0, 1)$ stands for the height at which the cumulative squared mass of v_n reaches x . The kernel says that, at the threshold, the root of v_n couples two children at cumulative masses x and y with a weight that depends only on the larger of the two, that is, on the taller child. In this continuum description, the weight $(2 \max(x, y))^{-1/2}$ is, up to null sets, the unique nonnegative function of $\max(x, y)$ that gives squared root mass dx to each band of cumulative mass dx : the pairs (x', y') with $\max(x', y') \in [x, x + dx]$ have area $2x \, dx$, and $T^2 \cdot 2x \, dx = dx$. (At a finite root height j , the coherent approximation of Lemma 13.1 carries the squared mass b_{j-1} of the height below.) The threshold value itself is the quadratic form of $\mathcal{T}$ on the constant function:

$$\langle \mathbf{1}, \mathcal{T} \mathbf{1} \rangle = 2 \int_0^1 \frac{x}{\sqrt{2x}} \, dx = \frac{2\sqrt{2}}{3} = F_*.$$

This is consistent with $f(v_n) = \langle \mathcal{I}^* v_n, |v_n\rangle \langle v_n| \rangle_{\text{HS}} \rightarrow F_*$. The convergence (13.10) holds in Hilbert–Schmidt norm only, and we do not claim that the traces of the operators $\mathcal{I}^* v_n$ converge to $\text{tr } \mathcal{T}$. For this reason, Section 14.2 uses the frame (13.11) only through the lower semicontinuity of the trace.

Proof of Theorem 13.2. The proof has six steps. After a bound on the height mesh, we embed each v_n so that its heights become consecutive intervals of $(0, 1)$. In these coordinates, the coherent approximation of Lemma 13.1 becomes an integral operator with a step

kernel. We then replace the step kernel by a simpler one and show that it converges to T . In the last step, we prove the properties of $\mathcal{T}$.

We write $\delta_n = \delta(v_n)$, and $b_{n,j}$, $B_{n,j}$, and $\bar{b}_n$ for the height masses of v_n . When no confusion can arise, we drop the index n .

Step 1: the height mesh. By (13.5), we have $\bar{b}_n \leq 2^{5/4} \delta_n^{1/2}$, which tends to zero because $\delta_n \rightarrow 0$.

Step 2: the embedding. For each j with $b_j > 0$, let $\hat{v}_j = v_j / \sqrt{b_j}$ and $J_j = (B_{j-1}, B_j)$. The vectors $\hat{v}_j$ are orthonormal, and so are the functions $b_j^{-1/2} \mathbf{1}_{J_j}$, because the intervals J_j are disjoint and have lengths b_j . Let E_n be the closed span of the vectors $\hat{v}_j$, and let $W_n: E_n \rightarrow L^2(0, 1)$ be the isometry with $W_n \hat{v}_j = b_j^{-1/2} \mathbf{1}_{J_j}$. The orthogonal complement $E_n^\perp$ is separable, because $\mathcal{H}$ is separable (it is a countable direct sum of tensor powers of a separable space). It is also infinite-dimensional, because it contains every vector of the one-leaf summand that is orthogonal to $\hat{v}_0$ (or the whole summand if $b_0 = 0$), and the one-leaf summand is infinite-dimensional by Section 7.1. We may therefore choose a unitary map $U_n: E_n^\perp \rightarrow \mathcal{K}$. We define $V_n(x + y) = W_n x \oplus U_n y$ for $x \in E_n$ and $y \in E_n^\perp$. Then V_n is an isometry, and

$$V_n v_n = \sum_{j: b_j > 0} \sqrt{b_j} W_n \hat{v}_j = \sum_{j: b_j > 0} \mathbf{1}_{J_j} = \mathbf{1},$$

since the intervals J_j cover $(0, 1)$ up to a countable set. Conjugation by V_n preserves Hilbert–Schmidt norms, positivity, and traces, because $V_n^* V_n$ is the identity of $\mathcal{H}$. Further, $V_n |x\rangle \langle y| V_n^* = |V_n x\rangle \langle V_n y|$.

Step 3: the coherent approximation as a step kernel. Let $\tilde{A}_n = \tilde{A}_{v_n}$ be the operator of Lemma 13.1, which applies once $\delta_n \leq 1$. Since $V_n v_{n,i} = \mathbf{1}_{J_i}$, the blocks (13.7) are transported to

$$V_n \frac{|v_{n,i}\rangle \langle v_{n,k}|}{\sqrt{B_h + B_{h-1}}} V_n^* = \frac{|\mathbf{1}_{J_i}\rangle \langle \mathbf{1}_{J_k}|}{\sqrt{B_h + B_{h-1}}}, \quad h = \max(i, k).$$

In other words, $V_n \tilde{A}_n V_n^*$ is the integral operator on $L^2(0, 1)$, extended by zero on $\mathcal{K}$, with the step kernel

$$\tilde{T}_n(x, y) = \frac{1}{\sqrt{B_h + B_{h-1}}}, \quad x \in J_i, y \in J_k, h = \max(i, k).$$

Recall that the Hilbert–Schmidt norm of an integral operator is the L^2 norm of its kernel. Then Lemma 13.1 implies that

$$\|V_n \mathcal{I}^* v_n V_n^* - \tilde{T}_n\|_{\text{HS}} = \|\mathcal{I}^* v_n - \tilde{A}_n\|_{\text{HS}} \leq 5\delta_n^{1/4} \rightarrow 0.$$

Here and below, we identify a kernel with its integral operator.

Step 4: replacement of the denominator. Let T_n be the step kernel equal to $(2B_h)^{-1/2}$ on $J_i \times J_k$, with $h = \max(i, k)$. The set of pairs (x, y) with $\max(i, k) = h$ has area $\sum_{\max(i, k)=h} b_i b_k = B_h^2 - B_{h-1}^2 = b_h(B_h + B_{h-1})$. Using $B_h + B_{h-1} = 2B_h - b_h$, we find

$$\|\tilde{T}_n - T_n\|_{L^2}^2 = \sum_{h: b_h > 0} b_h \left(1 - \sqrt{1 - \frac{b_h}{2B_h}}\right)^2 \leq \sum_{h: b_h > 0} \min\left\{b_h, \frac{b_h^3}{4B_h^2}\right\},$$

where we used $0 \leq 1 - \sqrt{1 - x} \leq x$ for $x \in [0, 1]$. Fix $\varepsilon \in (0, 1)$. The heights with $B_h < \varepsilon$ form an initial segment, on which $\sum b_h < \varepsilon$. On the heights with $B_h \geq \varepsilon$, we use the second term in the minimum and $\sum_h b_h = 1$. We obtain

$$\|\tilde{T}_n - T_n\|_{L^2}^2 \leq \varepsilon + \frac{\bar{b}_n^2}{4\varepsilon^2}.$$

By Step 1, the upper limit of the right side as $n \rightarrow \infty$ is ε , and letting $\varepsilon \downarrow 0$ gives $\|\tilde{T}_n - T_n\|_{L^2} \rightarrow 0$.

Step 5: convergence to the universal kernel. Let $x \in J_i$ and $y \in J_k$, with $h = \max(i, k)$. Then $\max(x, y) < \max(B_i, B_k) = B_h$, because $x < B_i$, $y < B_k$, and B is nondecreasing. If $h = i$, then $x > B_{h-1} \geq B_h - \bar{b}_n$, and if $h = k$, the same holds for y . This gives

$$\max(x, y) < B_h < \max(x, y) + \bar{b}_n.$$

Recalling that $T_n(x, y) = (2B_h)^{-1/2}$, we find

$$T(x, y) \left(\frac{\max(x, y)}{\max(x, y) + \bar{b}_n} \right)^{1/2} \leq T_n(x, y) \leq T(x, y).$$

The union of the rectangles $J_i \times J_k$ has full measure, and $\bar{b}_n \rightarrow 0$ by Step 1. Hence $T_n \rightarrow T$ almost everywhere, with $0 \leq T_n \leq T$. The kernel T is square integrable:

$$\int_0^1 \int_0^1 T(x, y)^2 \, dy \, dx = 2 \int_0^1 \int_0^x \frac{1}{2x} \, dy \, dx = 1.$$

By dominated convergence with the dominating function T^2 , we have $\|T_n - T\|_{L^2} \rightarrow 0$. Combining Steps 3–5, we obtain

$$\|V_n \mathcal{I}^* v_n V_n^* - \mathcal{T} \oplus 0\|_{\text{HS}} \leq \|\mathcal{I}^* v_n - \tilde{A}_n\|_{\text{HS}} + \|\tilde{T}_n - T_n\|_{L^2} + \|T_n - T\|_{L^2} \rightarrow 0,$$

which is (13.10). The computation above also shows that $\|\mathcal{T}\|_{\text{HS}} = \|T\|_{L^2} = 1$.

Step 6: the frame. For $z \in (0, 1]$, let $P_z = |\mathbf{1}_{[0,z]}\rangle\langle\mathbf{1}_{[0,z]}|$. It is positive, of rank one, and has trace z . The map $z \mapsto P_z$ is continuous in trace norm, and

$$\int_0^1 z^{-3/2} \|P_z\|_1 \, dz = \int_0^1 z^{-1/2} \, dz = 2.$$

Thus the integral in (13.11) converges in trace norm, and the right side of (13.11) defines a positive trace-class operator $\mathcal{T}'$ with

$$\mathrm{tr} \, \mathcal{T}' = \frac{1}{\sqrt{2}} + \frac{1}{2\sqrt{2}} \int_0^1 z^{-3/2} \cdot z \, dz = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = \sqrt{2}.$$

The kernel of P_z is $\mathbf{1}\{\max(x, y) \leq z\}$. For $f, g \in L^2(0, 1)$, Fubini's theorem gives

$$\begin{aligned} \langle f, \mathcal{T}' g \rangle &= \int_0^1 \int_0^1 f(x) g(y) \left[\frac{1}{\sqrt{2}} + \frac{1}{2\sqrt{2}} \int_{\max(x, y)}^1 z^{-3/2} \, dz \right] dx \, dy \\ &= \int_0^1 \int_0^1 f(x) T(x, y) g(y) \, dx \, dy. \end{aligned}$$

Fubini's theorem applies because $|f(x)g(y)|$ times the bracket is integrable: the bracket equals $T(x, y)$, since $\int_m^1 z^{-3/2} \, dz = 2(m^{-1/2} - 1)$, and T is square integrable. We have shown that $\mathcal{T}' = \mathcal{T}$, which proves (13.11), the positivity of $\mathcal{T}$, and the value of its trace. $\square$

13.2 Positive saturated triples

We next compare different vectors close to the ceiling. When two vectors have nonnegative inner products at every height, the overlaps at different heights cannot cancel. The pair-deficit lemma below shows that, for two vectors close to the ceiling, little cross mass is carried by heights at which a smoothed cosine between their prefixes is far from one. For three vectors, we choose for each of two pairs a prefix that carries almost all of the overlap of the pair and on which the two vectors are nearly parallel. These two prefixes are nested, and this gives the product inequality.

For two unit vectors v, w , we write B_j^v and B_j^w for their cumulative height masses.

Lemma 13.4 (Prefix alignment measured by a pair deficit). *Let v, w be unit tree vectors with $d_j = \langle \Pi_j v, \Pi_j w \rangle \geq 0$ for all $j \geq 0$. Let $D_j = \sum_{i \leq j} d_i$, with $D_{-1} = 0$, and set*

$$\rho_j = \frac{D_j + D_{j-1}}{\sqrt{(B_j^v + B_{j-1}^v)(B_j^w + B_{j-1}^w)}}, \quad \Delta(v, w) = \sum_{j \geq 0} d_j (1 - \rho_j),$$

with $\rho_j = 0$ when the denominator vanishes. Then $0 \leq \rho_j \leq 1$ for all j . If moreover $\delta(v) \leq 1$ and $\delta(w) \leq 1$, then

$$0 \leq \Delta(v, w) \leq K(\delta(v)^{1/4} + \delta(w)^{1/4}), \quad (13.12)$$

where K is an absolute constant.

The quantity $\Delta(v, w)$ is an average of $1 - \rho_j$ with respect to the cross masses d_j , whose total is $\langle v, w \rangle$. The ratio ρ_j is a smoothed version of $D_j / \sqrt{B_j^v B_j^w}$, the cosine of the angle between the prefixes $\Pi_{\leq j} v$ and $\Pi_{\leq j} w$. When the overlap of v and w is bounded away from zero and both deficits are small, the bound (13.12) says that most of the overlap of v and w , measured by the cross masses d_j , is collected at heights where ρ_j is close to one.

Proof. The ratio ρ_j is nonnegative, because all d_j are nonnegative. By the Cauchy–Schwarz inequality on prefixes, $D_j = \langle \Pi_{\leq j} v, \Pi_{\leq j} w \rangle \leq (B_j^v B_j^w)^{1/2}$, and the same holds with $j-1$ in place of j . The two-term Cauchy–Schwarz inequality $\sqrt{\alpha_1 \beta_1} + \sqrt{\alpha_2 \beta_2} \leq \sqrt{(\alpha_1 + \alpha_2)(\beta_1 + \beta_2)}$ then gives

$$D_j + D_{j-1} \leq \sqrt{(B_j^v + B_{j-1}^v)(B_j^w + B_{j-1}^w)}.$$

This shows that $\rho_j \leq 1$, and $\Delta(v, w) \geq 0$ follows.

Next, we compute the Hilbert–Schmidt inner product of the coherent approximations of Lemma 13.1. Blocks with different index pairs (i, k) are orthogonal, and the block (i, k) contributes $\langle v_i, w_i \rangle \langle v_k, w_k \rangle = d_i d_k$ divided by the two denominators in (13.7). For each h , we have $\sum_{\max(i, k)=h} d_i d_k = D_h^2 - D_{h-1}^2 = d_h(D_h + D_{h-1})$. Therefore

$$\langle \tilde{A}_v, \tilde{A}_w \rangle_{\text{HS}} = \sum_{h \geq 0} \frac{D_h^2 - D_{h-1}^2}{\sqrt{(B_h^v + B_{h-1}^v)(B_h^w + B_{h-1}^w)}} = \sum_{h \geq 0} d_h \rho_h = \langle v, w \rangle - \Delta(v, w). \quad (13.13)$$

If a denominator vanishes, then the corresponding components of v or w vanish at all heights up to h , and the terms on both sides are zero. The series converge absolutely, since the sum of the absolute values of the block inner products is at most $\|\tilde{A}_v\|_{\text{HS}} \|\tilde{A}_w\|_{\text{HS}} = 1$ by the Cauchy–Schwarz inequality. In the last equality, we used $\sum_h d_h = \langle v, w \rangle$.

On the other hand, by (13.1), we have

$$\langle \mathcal{I}^* v, \mathcal{I}^* w \rangle_{\text{HS}} = \langle v, \mathcal{I} \mathcal{I}^* w \rangle = \langle v, w \rangle - d_0.$$

Subtracting this from (13.13), we find

$$\Delta(v, w) = d_0 + \langle \mathcal{I}^* v, \mathcal{I}^* w \rangle_{\text{HS}} - \langle \tilde{A}_v, \tilde{A}_w \rangle_{\text{HS}}.$$

Using $\|\mathcal{I}^* w\|_{\text{HS}} \leq 1$, $\|\tilde{A}_v\|_{\text{HS}} = 1$, and Lemma 13.1, we bound the difference of the two inner products by

$$\|\mathcal{I}^* v - \tilde{A}_v\|_{\text{HS}} \|\mathcal{I}^* w\|_{\text{HS}} + \|\tilde{A}_v\|_{\text{HS}} \|\mathcal{I}^* w - \tilde{A}_w\|_{\text{HS}} \leq 5\delta(v)^{1/4} + 5\delta(w)^{1/4}.$$

The leaf overlap satisfies $d_0 \leq (b_0^v + b_0^w)/2$, and (13.5) implies that $b_0^v \leq 2^{5/4} \delta(v)^{1/2} \leq 2^{5/4} \delta(v)^{1/4}$, and similarly for w . Combining these bounds, we obtain (13.12) with $K = 7$. $\square$

We now state the product inequality. Note that if $\langle \Pi_h v_i, \Pi_h v_j \rangle \geq 0$ for all h , then $\langle v_i, v_j \rangle = \sum_h \langle \Pi_h v_i, \Pi_h v_j \rangle$ lies in $[0, 1]$.

Theorem 13.5 (Product inequality for positive threshold states). *For $\delta \in (0, 1]$, let $\omega(\delta)$ be the supremum of*

$$[\langle v_1, v_3 \rangle \langle v_2, v_3 \rangle - \langle v_1, v_2 \rangle]_+$$

over all triples of unit tree vectors v_1, v_2, v_3 such that $\langle \Pi_h v_i, \Pi_h v_j \rangle \geq 0$ for all $h \geq 0$ and $i, j \in \{1, 2, 3\}$, and such that $\max_i \delta(v_i) \leq \delta$. Then $\omega(\delta) \rightarrow 0$ as $\delta \downarrow 0$. In other words, every such triple satisfies

$$\langle v_1, v_2 \rangle \geq \langle v_1, v_3 \rangle \langle v_2, v_3 \rangle - \omega\left(\max_i \delta(v_i)\right), \quad (13.14)$$

with an error that tends to zero uniformly over such triples as their maximal deficit tends to zero.

Proof. The proof has four steps. We fix $\kappa \in (0, 1]$ and first treat triples with $\langle v_1, v_3 \rangle \geq \kappa$ and $\langle v_2, v_3 \rangle \geq \kappa$. For each $i \in \{1, 2\}$, we choose a height H_i such that the prefix $\Pi_{\leq H_i}$ contains almost all of the overlap between v_i and v_3 , and such that the prefixes of v_i and v_3 up to H_i are nearly parallel. The prefixes up to H_1 and H_2 are nested, which lets us compare v_1 and v_2 on the shorter one. The last step treats triples with a small overlap.

Throughout, we write

$$c_{ij} = \langle v_i, v_j \rangle, \quad d_{ij,h} = \langle \Pi_h v_i, \Pi_h v_j \rangle, \quad D_{ij,h} = \sum_{k \leq h} d_{ij,k}, \quad B_{i,h} = \sum_{k \leq h} \|\Pi_k v_i\|^2.$$

We let $\rho_{i3,h}$ be the ratio of Lemma 13.4 for the pair (v_i, v_3) , and set

$$\bar{b} = \sup_{i,h} \|\Pi_h v_i\|^2, \quad \varepsilon = \bar{b} + \Delta(v_1, v_3) + \Delta(v_2, v_3), \quad \eta = \sqrt{\varepsilon}.$$

Let $\delta = \max_i \delta(v_i) \leq 1$. By (13.5) and Lemma 13.4, we have

$$\varepsilon \leq 2^{5/4} \delta^{1/2} + 28 \delta^{1/4} \leq K_0 \delta^{1/4} \quad (13.15)$$

with the absolute constant $K_0 = 31$. In Steps 1–3, we assume that $c_{13}, c_{23} \geq \kappa$ and that $\eta \leq \kappa/4$; in particular, $\varepsilon \leq 1$, so that $\varepsilon \leq \eta \leq \varepsilon^{1/4}$.

Step 1: the choice of the grades H_i . Fix $i \in \{1, 2\}$. As $h \rightarrow \infty$, the partial sums $D_{i3,h}$ increase to c_{i3} . Let h_0 be the first height with $D_{i3,h_0} \geq c_{i3} - \eta$. The heights $h \geq h_0$ carry the cross mass

$$\sum_{h \geq h_0} d_{i3,h} = c_{i3} - D_{i3,h_0-1} > \eta,$$

since either $h_0 = 0$ and $c_{i3} \geq \kappa > \eta$, or $D_{i3,h_0-1} < c_{i3} - \eta$. We claim that there exists $H_i \geq h_0$ with

$$d_{i3,H_i} > 0 \quad \text{and} \quad 1 - \rho_{i3,H_i} \leq \frac{2\Delta(v_i, v_3)}{\eta} \leq 2\eta. \quad (13.16)$$

Indeed, if $1 - \rho_{i3,h} > 2\Delta(v_i, v_3)/\eta$ for every $h \geq h_0$ with $d_{i3,h} > 0$, then

$$\Delta(v_i, v_3) \geq \sum_{h \geq h_0} d_{i3,h}(1 - \rho_{i3,h}) > \frac{2\Delta(v_i, v_3)}{\eta} \cdot \eta = 2\Delta(v_i, v_3),$$

which is impossible, since $\Delta(v_i, v_3) \geq 0$. (The strict inequality holds because some $d_{i3,h}$ with $h \geq h_0$ is positive.) The second inequality in (13.16) follows from $\Delta(v_i, v_3) \leq \varepsilon = \eta^2$. This is the reason for the choice of H_i : the pair deficit is an average of $1 - \rho_{i3,h}$ with weights $d_{i3,h}$, the alignment can fail only at heights where the cross mass is small, and the heights beyond h_0 carry cross mass at least η . Because $H_i \geq h_0$, we also have

$$D_{i3,H_i} \geq c_{i3} - \eta. \quad (13.17)$$

Step 2: alignment on the prefix up to H_i . Let $x_i = \Pi_{\leq H_i} v_i$ and $z_i = \Pi_{\leq H_i} v_3$, and write

$$\Gamma = D_{i3,H_i} = \langle x_i, z_i \rangle, \quad X = B_{i,H_i} = \|x_i\|^2, \quad Z = B_{3,H_i} = \|z_i\|^2.$$

By (13.17), we have $\Gamma \geq \kappa - \kappa/4 = 3\kappa/4$. Using $\Gamma \leq \sqrt{XZ}$ and $X, Z \leq 1$, we find $X, Z \geq \Gamma^2 \geq 9\kappa^2/16$. The last increments $\xi = \|\Pi_{H_i} v_i\|^2$, $\zeta = \|\Pi_{H_i} v_3\|^2$, and $\gamma = d_{i3,H_i}$ satisfy $\xi, \zeta \leq \bar{b} \leq \varepsilon$ and $\gamma \geq 0$, and

$$\rho_{i3,H_i} = \frac{\Gamma - \gamma/2}{\sqrt{(X - \xi/2)(Z - \zeta/2)}} \leq \frac{\Gamma}{\sqrt{XZ}} \cdot \frac{1}{1 - \sigma}, \quad \sigma = \frac{8\varepsilon}{9\kappa^2} \leq \frac{1}{18},$$

where we used $(X - \xi/2)/X \geq 1 - \sigma$, $(Z - \zeta/2)/Z \geq 1 - \sigma$, and $\varepsilon \leq \kappa^2/16$. Using $\Gamma/\sqrt{XZ} \leq 1$ and $1/(1 - \sigma) \leq 1 + 2\sigma$, we have $\rho_{i3,H_i} \leq \Gamma/\sqrt{XZ} + 2\sigma$. Combining this with (13.16), we obtain

$$1 - \frac{\Gamma}{\sqrt{XZ}} \leq (1 - \rho_{i3,H_i}) + 2\sigma \leq 2\eta + \frac{16\varepsilon}{9\kappa^2} \leq \frac{4\eta}{\kappa^2}.$$

Let $\alpha_i = \Gamma/Z$. Then

$$\|x_i - \alpha_i z_i\|^2 = X - \frac{\Gamma^2}{Z} = X \left(1 - \frac{\Gamma^2}{XZ}\right) \leq 2 \left(1 - \frac{\Gamma}{\sqrt{XZ}}\right) \leq \frac{8\eta}{\kappa^2},$$

and $\|\alpha_i z_i\| = \Gamma/\sqrt{Z} \leq \sqrt{X} \leq 1$. We have established that

$$x_i = \alpha_i z_i + e_i, \quad \|e_i\| \leq \frac{2\sqrt{2}}{\kappa} \varepsilon^{1/4}, \quad \|\alpha_i z_i\| \leq 1 \quad (i = 1, 2). \quad (13.18)$$

Step 3: comparison on the shorter prefix. By symmetry in the indices 1 and 2, we may assume that $H_1 \leq H_2$. Applying $\Pi_{\leq H_1}$ to (13.18) with $i = 2$, and using $\Pi_{\leq H_1} z_2 =$

$\Pi_{\leq H_1} v_3 = z_1$, we find $\Pi_{\leq H_1} v_2 = \alpha_2 z_1 + e'_2$ with $\|e'_2\| \leq \|e_2\|$. Because the overlaps $d_{12,h}$ are all nonnegative, we have

$$c_{12} = \sum_h d_{12,h} \geq \sum_{h \leq H_1} d_{12,h} = \langle x_1, \Pi_{\leq H_1} v_2 \rangle = \langle \alpha_1 z_1 + e_1, \alpha_2 z_1 + e'_2 \rangle.$$

We expand the last inner product. We have $\|\alpha_1 z_1\| \leq 1$, $\|\alpha_2 z_1\| \leq \|\alpha_2 z_2\| \leq 1$, and $\|e_1\| \leq \|x_1\| + \|\alpha_1 z_1\| \leq 2$. The three cross terms are therefore bounded in absolute value by $\|e'_2\| + \|e_1\| + 2\|e'_2\| \leq (8\sqrt{2}/\kappa)\varepsilon^{1/4}$, and

$$c_{12} \geq \alpha_1 \alpha_2 B_{3,H_1} - \frac{8\sqrt{2}}{\kappa} \varepsilon^{1/4}.$$

For the main term, we use (13.17), $B_{3,H_2} \leq 1$, and $c_{13}, c_{23} \leq 1$:

$$\alpha_1 \alpha_2 B_{3,H_1} = D_{13,H_1} \frac{D_{23,H_2}}{B_{3,H_2}} \geq (c_{13} - \eta)(c_{23} - \eta) \geq c_{13}c_{23} - 2\eta \geq c_{13}c_{23} - 2\varepsilon^{1/4}.$$

Combining the last two displays, we find

$$c_{12} \geq c_{13}c_{23} - \frac{14}{\kappa} \varepsilon^{1/4} \quad \text{whenever } c_{13}, c_{23} \geq \kappa \text{ and } \varepsilon \leq \frac{\kappa^2}{16}. \quad (13.19)$$

Step 4: uniformity. Let $\kappa \in (0, 1]$, and let $\delta \leq 1$ be so small that $K_0 \delta^{1/4} \leq \kappa^2/16$. Let v_1, v_2, v_3 be an admissible triple with $\max_i \delta(v_i) \leq \delta$. If $c_{13} < \kappa$ or $c_{23} < \kappa$, then $c_{13}c_{23} - c_{12} \leq c_{13}c_{23} < \kappa$, because all overlaps lie in $[0, 1]$. Otherwise, (13.15) and (13.19) imply that $c_{13}c_{23} - c_{12} \leq 14\kappa^{-1}K_0^{1/4}\delta^{1/16}$. This shows that

$$\omega(\delta) \leq \max\{\kappa, 14\kappa^{-1}K_0^{1/4}\delta^{1/16}\},$$

and $\limsup_{\delta \downarrow 0} \omega(\delta) \leq \kappa$. Since $\kappa \in (0, 1]$ is arbitrary, we conclude that $\omega(\delta) \rightarrow 0$, as desired. $\square$

Three-time correlations and the susceptibility

This section applies the deterministic results of Section 13 to the cubic CK solution. In Section 14.1, we establish that the normalized prefix centers have asymptotically the same overlaps as the correlations divided by the plateau (Proposition 14.2). Combined with Theorem 13.5, this gives the three-time product inequality of Theorem 14.3 and composition bounds for last-crossing times (Corollary 14.4). In Section 14.2, we prove the lower bound $\liminf_{t \rightarrow \infty} \chi(t, 0) \geq 2(1 - q)$ (Theorem 14.7). These results prove Theorem 1.3. The product inequality is used again in Section 15, where it bounds the memory term of the response equation, and the last-crossing bounds are discussed in Section 17.

Standing assumptions and notation. Throughout this section, $p = 3$ and $\beta \geq \beta_d = \sqrt{4/3}$, including the critical endpoint. We write q for the selected plateau and (c, r) for the stationary profile of Theorem 8.5, with $r = -c'$. Recall that $q \geq 1/2$, and that the bounded-lag translates of C and R converge locally uniformly, as in (8.11). We use the tree vectors v_t and the prefix projections P_T of Proposition 7.1, and the notation f , F_* , and $\delta(\cdot)$ of Section 13.

It is convenient to index a prefix center by its terminal time and its age. For $s \geq D \geq 0$, we define

$$M_{s,D} = P_{s-D} v_s, \quad U_{s,D} = \frac{M_{s,D}}{\|M_{s,D}\|},$$

the latter whenever $M_{s,D} \neq 0$. In the notation of Section 7.4, $M_{s,D} = M_{s-D}(D)$: it is the part of the state at time s that is determined by the noise up to time $s - D$. We call D the smoothing age of the center.

Fix $D \geq 0$. Recall from Theorem 8.5 that the translated kernels converge along every sequence of times tending to infinity. Hence Proposition 7.5 applies along every such sequence, and its conclusions hold as $T \rightarrow \infty$. In particular, (7.17) gives $\|M_{s,D}\|^2 \rightarrow c(2D) \geq q$ as $s \rightarrow \infty$, and $U_{s,D}$ is defined for all large s . The plateau is positive and marginal (Theorem 8.5 and Lemma 8.3), which means that Proposition 9.5 applies. Let h_D and δ_D be as in (7.18) and (9.12). The cubic tree energy (7.5) is $h = \sqrt{3}f$, the function f is homogeneous of degree three, and $E_{\text{th}} = \sqrt{3}F_*$. By (7.17) and (7.18), we have

$$f(U_{s,D}) = \frac{f(M_{s,D})}{\|M_{s,D}\|^3} = \frac{h(M_{s,D})}{\sqrt{3}\|M_{s,D}\|^3} \rightarrow \frac{h_D}{\sqrt{3}c(2D)^{3/2}} = \frac{E_{\text{th}} - \delta_D}{\sqrt{3}} = F_* - \frac{\delta_D}{\sqrt{3}}$$

as $s \rightarrow \infty$. Together with (9.12), this gives

$$\lim_{s \rightarrow \infty} [F_* - f(U_{s,D})] = \frac{\delta_D}{\sqrt{3}}, \quad \lim_{D \rightarrow \infty} \delta_D = 0. \quad (14.1)$$

In all limits in this section, we let the waiting time tend to infinity with the smoothing age fixed, and increase the smoothing age only afterward.

14.1 From causal centers to actual correlations

In this section, we assume that $\beta \geq \beta_d$. We use the following inputs:

- from Section 11, the remote bound (11.12) on the later derivative (Theorem 11.4), and the bound (11.22) on the variation of a correlation row across a window (Corollary 11.6), which rests on the strong localization of the response (Theorem 11.1) and on the Fisher control of the backward defect (Theorem 5.3);
- from Section 9, Proposition 9.5, through (14.1);
- from Section 13, Theorem 13.5;
- from Section 7.4, the complete monotonicity of c (Corollary 7.7).

At $\beta = \beta_d$, we have $q = 1/2$ by Theorem 8.5, and $6\beta^2 q(1-q)^2 = 1$ by Lemma 8.3. Then the plateau is positive and marginal, so Proposition 9.5 applies at the endpoint. The response and Fisher inputs above hold for every $\beta > 0$. The argument below uses $q > 0$ and these inputs, and therefore includes $\beta = \beta_d$ directly. It does not use a strictly positive Fisher margin; the positive relaxation cost of Section 15 retains the strict hypothesis $\beta > \beta_d$.

These inputs imply that late, remote correlations change little across a window of length D , uniformly after the waiting-time limit. For the later coordinate, let $s \geq S$ and $t \geq s + D$. Then $s \leq t - D$ and $t \geq S + D$, so (11.12) with $L = D$ gives

$$\limsup_{S \rightarrow \infty} \sup_{t \geq s+D, s \geq S} |\partial_t C(t, s)| \leq \sqrt{c''(2D)}. \quad (14.2)$$

The earlier coordinate is controlled by (11.22). Both bounds become small after multiplication by the window length:

$$D^2 c''(2D) \rightarrow 0, \quad D \int_D^\infty r(u)^2 du \rightarrow 0 \quad (D \rightarrow \infty). \quad (14.3)$$

To prove the first limit, we use the spectral representation (7.28). Differentiating under the integral for $t > 0$, we have

$$D^2 c''(2D) = \int_{[0, \infty)} (Dx)^2 e^{-2Dx} \nu(dx).$$

The integrand is bounded by $\sup_{y \geq 0} y^2 e^{-2y} = e^{-2}$ and tends to zero as $D \rightarrow \infty$ for every $x \geq 0$, including $x = 0$. Since ν is a probability measure, the first limit follows from dominated convergence. For the second, recall that r is nonincreasing and $\int_0^\infty r = 1 - q$. This implies that $Dr(D) \leq \int_0^D r \leq 1 - q$, and

$$D \int_D^\infty r(u)^2 du \leq Dr(D) \int_D^\infty r(u) du \leq (1 - q) \int_D^\infty r(u) du \rightarrow 0.$$

We use only the complete monotonicity of c here, and not the power law of Section 16.

We place the causal centers and the three-time inequality below in the line of the normalized time-averaged magnetizations and the spherical product relation discussed by Cugliandolo and Kurchan [52, arXiv version 1, Section 5.1, Equations (5.1.8)–(5.1.9)]. In that work, the product relation is an asymptotic relation of the physical theory. Here we prove a uniform one-sided inequality for the CK solution, without assuming a slow clock.

The next lemma compares a center with the average of the trajectory over the window that precedes it. For $s \geq D > 0$, we define the Bochner integral

$$Z_{s,D} = \frac{1}{D} \int_{s-D}^s v_u du.$$

It is well defined because $u \mapsto v_u$ is continuous, and $\|Z_{s,D}\| \leq 1$.

Lemma 14.1 (Causal centers are approximated by local averages). *For every fixed $D > 0$,*

$$\lim_{s \rightarrow \infty} \|M_{s,D} - Z_{s,D}\|^2 = \eta_D, \quad \eta_D = c(2D) + \frac{2}{D^2} \int_0^D (D - \tau)c(\tau) d\tau - \frac{2}{D} \int_D^{2D} c(\tau) d\tau. \quad (14.4)$$

Moreover, $\eta_D \geq 0$ and $\eta_D \rightarrow 0$ as $D \rightarrow \infty$.

Proof. We expand $\|M_{s,D} - Z_{s,D}\|^2 = \|M_{s,D}\|^2 + \|Z_{s,D}\|^2 - 2 \langle M_{s,D}, Z_{s,D} \rangle$ and compute the limit of each term. By (7.17), the first term tends to $c(2D)$. Next,

$$\begin{aligned} \|Z_{s,D}\|^2 &= \frac{1}{D^2} \int_{s-D}^s \int_{s-D}^s C(u, w) du dw \\ &\rightarrow \frac{1}{D^2} \int_0^D \int_0^D c(|u - w|) du dw = \frac{2}{D^2} \int_0^D (D - \tau)c(\tau) d\tau, \end{aligned}$$

where we used the uniform convergence (8.11) on the window $[s - D, s]$. For the cross term, let $T = s - D$, and write $Q_T(a, b) = \langle M_T(a), M_T(b) \rangle$ for the Gram kernel of the prefix centers of Section 7.4. For $u = T + a$ with $a \in [0, D]$, we have $\langle M_{s,D}, v_u \rangle = \langle P_T v_{T+D}, P_T v_{T+a} \rangle = Q_T(D, a)$. As noted at the start of this section, (7.16) holds as $T \rightarrow \infty$. It implies that $Q_T(D, a)$ tends to $c(D + a)$ uniformly in $a \in [0, D]$, and

$$\langle M_{s,D}, Z_{s,D} \rangle = \frac{1}{D} \int_0^D Q_T(D, a) da \rightarrow \frac{1}{D} \int_0^D c(D + a) da = \frac{1}{D} \int_D^{2D} c(\tau) d\tau.$$

This proves (14.4). The limit η_D is nonnegative, as a limit of squared norms.

To show that $\eta_D \rightarrow 0$, let $g = c - q$, which is nonnegative, nonincreasing, and tends to zero. The constant q contributes $q + q - 2q = 0$ to η_D , so

$$\eta_D = g(2D) + \frac{2}{D^2} \int_0^D (D - \tau)g(\tau) \, d\tau - \frac{2}{D} \int_D^{2D} g(\tau) \, d\tau.$$

The first term tends to zero. The second is at most $2D^{-1} \int_0^D g$, which tends to zero because $g(\tau) \rightarrow 0$. The absolute value of the third is at most $2g(D)$, which also tends to zero. $\square$

Proposition 14.2 (Uniform aging Gram of normalized centers). *We have*

$$\lim_{D \rightarrow \infty} \limsup_{S \rightarrow \infty} \sup_{s \geq S+D, t \geq s+2D} \left| \langle U_{t,D}, U_{s,D} \rangle - \frac{C(t,s)}{q} \right| = 0. \quad (14.5)$$

In particular, the correlation of two late times with a diverging separation does not exceed the plateau in the limit: if $s_n \rightarrow \infty$ and $t_n - s_n \rightarrow \infty$, then

$$\limsup_{n \rightarrow \infty} C(t_n, s_n) \leq q. \quad (14.6)$$

In (14.5), the supremum is over pairs for which both centers are defined. For fixed D , this includes all pairs in the range once S is large.

Proof. The proof has four steps. We first replace each center by the local average of Lemma 14.1. We then compare the overlap of two local averages with the correlation of their endpoints. Combining the two steps gives the claim for the unnormalized centers, and in the last step we divide by their norms.

Step 1: centers and local averages. Since $\|M_{s,D}\| \leq \|v_s\| = 1$ and $\|Z_{t,D}\| \leq 1$, we have

$$|\langle M_{t,D}, M_{s,D} \rangle - \langle Z_{t,D}, Z_{s,D} \rangle| \leq \|M_{t,D} - Z_{t,D}\| + \|M_{s,D} - Z_{s,D}\|.$$

The limit in (14.4) is a limit as the single variable s tends to infinity, so that $\limsup_{S \rightarrow \infty} \sup_{s \geq S} \|M_{s,D} - Z_{s,D}\| = \eta_D^{1/2}$.

Step 2: local averages and correlations. Let $s \geq S + D$ and $t \geq s + 2D$. Then

$$\langle Z_{t,D}, Z_{s,D} \rangle = \frac{1}{D^2} \int_{t-D}^t \int_{s-D}^s C(u, w) \, dw \, du.$$

For $u \in [t-D, t]$ and $w \in [s-D, s]$, we compare $C(u, w)$ with $C(t, s)$ by changing the two coordinates in succession:

$$|C(u, w) - C(t, s)| \leq \int_u^t |\partial_x C(x, w)| \, dx + \int_{s-D}^s |\partial_y C(t, y)| \, dy.$$

In the first integral, $x \geq u \geq t - D \geq s + D \geq w + D$ and $w \geq s - D \geq S$, and the integrand is bounded by the supremum in (14.2). The second integral is the quantity bounded in (11.22), since $s \geq S + D$ and $t \geq s + D$. Averaging over u and w and using these two bounds, we find

$$\limsup_{S \rightarrow \infty} \sup_{s \geq S+D, t \geq s+2D} |\langle Z_{t,D}, Z_{s,D} \rangle - C(t, s)| \leq D\sqrt{c''(2D)} + \left(D \int_D^\infty r(u)^2 du\right)^{1/2}.$$

Step 3: unnormalized centers. Combining Steps 1 and 2, we obtain

$$\limsup_{S \rightarrow \infty} \sup_{s \geq S+D, t \geq s+2D} |\langle M_{t,D}, M_{s,D} \rangle - C(t, s)| \leq \mathcal{E}_D, \quad (14.7)$$

where

$$\mathcal{E}_D = 2\eta_D^{1/2} + D\sqrt{c''(2D)} + \left(D \int_D^\infty r(u)^2 du\right)^{1/2}.$$

Lemma 14.1 and (14.3) imply that $\mathcal{E}_D \rightarrow 0$ as $D \rightarrow \infty$.

Step 4: normalization. For centers that are defined,

$$\left| \langle U_{t,D}, U_{s,D} \rangle - \frac{C(t, s)}{q} \right| \leq \frac{|\langle M_{t,D}, M_{s,D} \rangle - C(t, s)|}{\|M_{t,D}\| \|M_{s,D}\|} + C(t, s) \left| \frac{1}{\|M_{t,D}\| \|M_{s,D}\|} - \frac{1}{q} \right|.$$

As $s \rightarrow \infty$, the squared norm $\|M_{s,D}\|^2$ tends to $c(2D) \geq q$, and this limit is again uniform over $s \geq S$ as $S \rightarrow \infty$. Using (14.7) and $C \leq 1$, we find that the upper limit as $S \rightarrow \infty$ of the supremum in (14.5) is at most

$$\frac{\mathcal{E}_D}{c(2D)} + \left| \frac{1}{c(2D)} - \frac{1}{q} \right| \leq \frac{\mathcal{E}_D}{q} + \frac{c(2D) - q}{q^2}.$$

This tends to zero as $D \rightarrow \infty$, because $\mathcal{E}_D \rightarrow 0$ and $c(2D) \rightarrow q$. This proves (14.5).

For (14.6), let $e(S, D)$ denote the supremum in (14.5). By the Cauchy–Schwarz inequality, $\langle U_{t,D}, U_{s,D} \rangle \leq 1$, and we have $C(t, s) \leq q + qe(S, D)$ for all pairs in its range. For fixed S and D , the pair (s_n, t_n) lies in this range for all large n , and therefore $\limsup_n C(t_n, s_n) \leq q + qe(S, D)$. Letting $S \rightarrow \infty$ and then $D \rightarrow \infty$ proves (14.6). $\square$

We now transfer the product inequality of Theorem 13.5 to the correlations. For $S \geq 0$ and $D > 0$, let

$$\mathcal{A}_{S,D} = \{(u, s, t) : u \geq S + D, s \geq u + 2D, t \geq s + 2D\}.$$

Theorem 14.3 (Actual three-time product inequality). *For the cubic CK solution at $\beta \geq \beta_d$, including the critical endpoint,*

$$\lim_{D \rightarrow \infty} \limsup_{S \rightarrow \infty} \sup_{(u,s,t) \in \mathcal{A}_{S,D}} [C(t, s)C(s, u) - qC(t, u)]_+ = 0. \quad (14.8)$$

The same holds with the bracket replaced by $[C(t, s)C(t, u) - qC(s, u)]_+$ or by $[C(t, u)C(s, u) - qC(t, s)]_+$. In particular, for every sequence of triples $u_n < s_n < t_n$ with $u_n \rightarrow \infty$, $s_n - u_n \rightarrow \infty$, and $t_n - s_n \rightarrow \infty$,

$$\liminf_{n \rightarrow \infty} \{qC(t_n, u_n) - C(t_n, s_n)C(s_n, u_n)\} \geq 0, \quad (14.9)$$

and the same holds for the other two inequalities.

Proof. We first check that the centers satisfy the positivity hypothesis of Theorem 13.5. Recall from Proposition 7.1 that every component of v_t is a nonnegative function of the leaf labels. The projections P_T and Π_h preserve this nonnegativity, since they multiply every component by an indicator function (Section 7.1). The inner product of two nonnegative vectors is nonnegative, and therefore $\langle \Pi_h M_{t,D}, \Pi_h M_{s,D} \rangle \geq 0$ for all h, s, t , and D , and the same holds for the normalized centers.

Let ω be the function of Theorem 13.5. Fix D so large that $\delta_D/\sqrt{3} + 1/D \leq 1$, and write $\delta'_D = \delta_D/\sqrt{3} + 1/D$. Using (14.1), we choose S_D such that $U_{x,D}$ is defined and $F_* - f(U_{x,D}) \leq \delta'_D$ for all $x \geq S_D$. Let $S \geq S_D$ and $(u, s, t) \in \mathcal{A}_{S,D}$. All three times are at least S , so the triple

$$v_1 = U_{t,D}, \quad v_2 = U_{u,D}, \quad v_3 = U_{s,D}$$

satisfies the hypotheses of Theorem 13.5 with maximal deficit at most δ'_D . Writing $g_{xy} = \langle U_{x,D}, U_{y,D} \rangle$, we deduce from (13.14) that

$$g_{tu} \geq g_{ts}g_{su} - \omega(\delta'_D).$$

We now replace the overlaps by correlations. Let $e(S, D)$ be the supremum in (14.5). The three pairs (s, t) , (u, s) , and (u, t) all lie in the range of that supremum, since $s \geq u + 2D \geq S + D$, $t \geq s + 2D$, and $t \geq u + 2D$. We obtain $|C(x, y) - qg_{xy}| \leq qe$ for each pair, with $e = e(S, D)$. All correlations and normalized overlaps lie in $[0, 1]$. Using $C(t, s) \leq q(g_{ts} + e)$, $C(s, u) \leq q(g_{su} + e)$, and $C(t, u) \geq q(g_{tu} - e)$, we find

$$C(t, s)C(s, u) - qC(t, u) \leq q^2(g_{ts}g_{su} - g_{tu}) + q^2(3e + e^2) \leq q^2\omega(\delta'_D) + q^2(3e + e^2).$$

We take the supremum over $\mathcal{A}_{S,D}$ and let $S \rightarrow \infty$. Proposition 14.2 implies that $\bar{e}_D = \limsup_{S \rightarrow \infty} e(S, D)$ tends to zero as $D \rightarrow \infty$, and $\omega(\delta'_D) \rightarrow 0$ by Theorem 13.5, since $\delta'_D \rightarrow 0$. We conclude that

$$\limsup_{S \rightarrow \infty} \sup_{\mathcal{A}_{S,D}} [C(t, s)C(s, u) - qC(t, u)]_+ \leq q^2\omega(\delta'_D) + q^2(3\bar{e}_D + \bar{e}_D^2) \rightarrow 0$$

as $D \rightarrow \infty$, which is (14.8). The other two inequalities follow in the same way, by choosing $U_{t,D}$ or $U_{u,D}$ as the vector v_3 in Theorem 13.5.

Finally, let (u_n, s_n, t_n) be a sequence as in the statement. For fixed S and D , the triple (u_n, s_n, t_n) lies in $\mathcal{A}_{S,D}$ for all large n . Then the upper limit of $[C(t_n, s_n)C(s_n, u_n) -$

$qC(t_n, u_n)]_+$ as $n \rightarrow \infty$ is at most the supremum over $\mathcal{A}_{S,D}$, for every S and D . Letting $S \rightarrow \infty$ and then $D \rightarrow \infty$, we find that this upper limit is zero, which implies (14.9). The other two inequalities are handled in the same way. $\square$

The product inequality implies composition bounds for relaxation times. For a level $c \in (0, q)$, we consider the last time at which the correlation with a given earlier time is at least c . Here and in Section 15, the letter c without an argument, possibly with an index, denotes such a level; the stationary profile is always written with its argument, as in $c(W)$.

Corollary 14.4 (Composition of last-crossing relaxation times). *For a level $c \in (0, q)$ and $s \geq 0$, define*

$$\Theta_c(s) = \sup\{t \geq s : C(t, s) \geq c\}.$$

For levels $c_1, c_2 \in (0, q)$ and $0 < c_0 < c_1 c_2 / q$, we have, for every sufficiently large s ,

$$\Theta_{c_0}(s) \geq \Theta_{c_2}(\Theta_{c_1}(s)). \quad (14.10)$$

More generally, let $c_1, \dots, c_k \in (0, q)$ be a fixed finite list of levels, and let $s_0 = s$ and $s_j = \Theta_{c_j}(s_{j-1})$ for $1 \leq j \leq k$. Then

$$\liminf_{s \rightarrow \infty} C(s_k, s) \geq q^{1-k} \prod_{j=1}^k c_j.$$

Proof. We first show that $\Theta_c(s)$ is finite and that $C(\Theta_c(s), s) = c$. By Theorem 8.8, $C(t, s) \rightarrow 0$ as $t \rightarrow \infty$, and the set $\{t \geq s : C(t, s) \geq c\}$ is bounded. It contains s , because $C(s, s) = 1$, and it is closed, because C is continuous. Its supremum $\Theta_c(s)$ therefore belongs to it, that is, $C(\Theta_c(s), s) \geq c$. If this inequality were strict, then by continuity some $t > \Theta_c(s)$ would also satisfy $C(t, s) > c$, which is impossible.

Next, we show that $\Theta_c(s) - s \rightarrow \infty$ as $s \rightarrow \infty$. Fix $W > 0$. By the stationary window (8.11), we have $C(s + W, s) \rightarrow c(W)$, and the profile satisfies $c(W) \geq q > c$. Then $C(s + W, s) > c$ for all large s , which gives $\Theta_c(s) \geq s + W$.

Let $s_1 = \Theta_{c_1}(s)$ and $s_2 = \Theta_{c_2}(s_1)$. As $s \rightarrow \infty$, we have $s \rightarrow \infty$, $s_1 - s \rightarrow \infty$, and, since $s_1 \geq s$, also $s_2 - s_1 \rightarrow \infty$. Applying (14.9) to the triple $s < s_1 < s_2$ along any sequence of values of s tending to infinity, and using $C(s_2, s_1) = c_2$ and $C(s_1, s) = c_1$, we obtain

$$\liminf_{s \rightarrow \infty} C(s_2, s) \geq \frac{c_1 c_2}{q} > c_0.$$

Then $C(s_2, s) \geq c_0$ for all large s , and s_2 belongs to the set that defines $\Theta_{c_0}(s)$. This proves (14.10).

For the finite list, we prove by induction on j that $\liminf_{s \rightarrow \infty} C(s_j, s) \geq q^{1-j} \prod_{i \leq j} c_i$. For $j = 1$, we have $C(s_1, s) = c_1$. Let $j \geq 2$. The times s_j are nondecreasing in j , which

gives $s_{j-1} - s \geq s_1 - s \rightarrow \infty$, and $s_j - s_{j-1} \rightarrow \infty$ because $s_{j-1} \rightarrow \infty$. Applying (14.9) to the triple $s < s_{j-1} < s_j$ and using $C(s_j, s_{j-1}) = c_j$, we find

$$q \liminf_{s \rightarrow \infty} C(s_j, s) \geq c_j \liminf_{s \rightarrow \infty} C(s_{j-1}, s) \geq c_j q^{2^{-j}} \prod_{i < j} c_i,$$

where the last inequality is the induction hypothesis. This completes the induction. $\square$

Remark 14.5. The definition of $\Theta_c(s)$ allows returns: the correlation $C(t, s)$ may fall below c and rise above it again before the last crossing. Theorem 14.3 and Corollary 14.4 are inequalities. They do not assert a multiplicative composition law for correlations, and they do not identify an aging clock; see Section 17.

14.2 The susceptibility lower bound

Throughout this section, $\beta \geq \beta_d$; the endpoint $\beta = \beta_d$, at which $q = 1/2$, is included. We have $\int_0^\infty r = 1 - q$, because $r = -c'$, $c(0) = 1$, and $c(\tau) \rightarrow q$. For $p = 3$, the marginality equation (8.6) reads $6\beta^2 q(1 - q)^2 = 1$, that is,

$$\beta\sqrt{3}\sqrt{q}(1 - q) = \frac{1}{\sqrt{2}}. \quad (14.11)$$

In this section, we use Theorem 13.2 from Section 13, Proposition 9.5 from Section 9 (through (14.1)), and the strong localization of the response row, Theorem 11.1, from Section 11. We also use the root frame (7.6) and Proposition 7.5. For $p = 3$, the root frame reads

$$\mathcal{I}^* P_T v_t = \beta\sqrt{3} \int_0^t R(t, u) |P_T v_u\rangle \langle P_T v_u| \, du. \quad (14.12)$$

The root operator of a center is a positive integral of the rank-one operators $|P_T v_u\rangle \langle P_T v_u|$, weighted by the response. Theorem 13.2 determines the limit of the normalized root in Hilbert–Schmidt norm, but this alone does not determine the limit of its trace, which is where the total response mass appears. We obtain a lower bound on the trace by first replacing each old rank-one term $|v_u\rangle \langle v_u|$ by the smoothed term $|M_{u,D}\rangle \langle M_{u,D}|$, whose norm converges. Because of causality, the errors made on time strips at distance at least D from each other are orthogonal. Their total is therefore controlled by the squared response row on the old times, which tends to zero by Theorem 11.1. We need no bound on the integral $\chi(t, 0)$ of the response row; by (11.6), it is only known to be $o(\sqrt{t})$.

Lemma 14.6 (Orthogonality of disjoint causal strips). *For $D > 0$ and $u \geq D$, let*

$$F_{u,D} = |v_u\rangle \langle v_u| - |M_{u,D}\rangle \langle M_{u,D}|.$$

If $u, w \geq D$ and $|u - w| \geq D$, then $\langle F_{u,D}, F_{w,D} \rangle_{\text{HS}} = 0$. For $D \leq T \leq t$, we have

$$\left\| \int_D^T R(t, u) F_{u,D} \, du \right\|_{\text{HS}}^2 \leq 8D \int_D^T R(t, u)^2 \, du. \quad (14.13)$$

If $t_n \rightarrow \infty$, $L_n \rightarrow \infty$, and $t_n - L_n \geq D$, then the Hilbert–Schmidt distance between

$$\int_0^{t_n - L_n} R(t_n, u) |v_u\rangle \langle v_u| \, du \quad \text{and} \quad \int_D^{t_n - L_n} R(t_n, u) |M_{u,D}\rangle \langle M_{u,D}| \, du$$

tends to zero.

Proof. We first note that the integrals are Bochner integrals in trace norm. For each fixed $x \in \mathcal{H}$, the map $T \mapsto P_T x$ is continuous: on each tree summand, P_T multiplies the leaf kernel by indicator functions of leaf times at most T , and the continuity follows from dominated convergence, first on each summand and then on their direct sum. Together with the continuity of $u \mapsto v_u$, this shows that $u \mapsto M_{u,D}$ is continuous on $[D, \infty)$. Then the rank-one operators above are continuous in trace norm, with trace norms at most one.

Suppose that $u \geq w + D$; the opposite ordering follows by symmetry. Then $w \leq u - D$, and we have $v_w \in \mathcal{H}_w \subset \mathcal{H}_{u-D}$ and $M_{w,D} \in \mathcal{H}_{w-D} \subset \mathcal{H}_{u-D}$. For both vectors $x \in \{v_w, M_{w,D}\}$, this gives $\langle v_u, x \rangle = \langle P_{u-D} v_u, x \rangle = \langle M_{u,D}, x \rangle$. Expanding the four rank-one pairings and using $\langle |a\rangle \langle a|, |b\rangle \langle b| \rangle_{\text{HS}} = \langle a, b \rangle^2$, we find

$$\begin{aligned} \langle F_{u,D}, F_{w,D} \rangle_{\text{HS}} &= \langle v_u, v_w \rangle^2 - \langle v_u, M_{w,D} \rangle^2 - \langle M_{u,D}, v_w \rangle^2 + \langle M_{u,D}, M_{w,D} \rangle^2 \\ &= \langle M_{u,D}, v_w \rangle^2 - \langle M_{u,D}, M_{w,D} \rangle^2 - \langle M_{u,D}, v_w \rangle^2 + \langle M_{u,D}, M_{w,D} \rangle^2 = 0. \end{aligned}$$

For (14.13), note that $\|F_{u,D}\|_{\text{HS}} \leq \|v_u\|^2 + \|M_{u,D}\|^2 \leq 2$, and that $R \geq 0$. Let $g = \mathbf{1}_{[D,T]} R(t, \cdot)$, extended by zero to $\mathbb{R}$, and $k = \mathbf{1}_{[-D,D]}$. By the orthogonality just proved,

$$\begin{aligned} \left\| \int_D^T R(t, u) F_{u,D} \, du \right\|_{\text{HS}}^2 &= \int_D^T \int_D^T R(t, u) R(t, w) \langle F_{u,D}, F_{w,D} \rangle_{\text{HS}} \, du \, dw \\ &\leq 4 \int_{\mathbb{R}} g(u) (k * g)(u) \, du. \end{aligned}$$

By the Cauchy–Schwarz inequality and Young’s convolution inequality, the right side is at most $4\|g\|_2 \|k * g\|_2 \leq 4\|k\|_1 \|g\|_2^2 = 8D \int_D^T R(t, u)^2 \, du$.

For the last assertion, let $T_n = t_n - L_n$, and let f_t be the lag row (11.3). Substituting $\tau = t_n - u$, we have

$$\int_0^{T_n} R(t_n, u)^2 \, du = \int_{L_n}^{t_n} f_{t_n}(\tau)^2 \, d\tau \leq 2\|f_{t_n} - r\|_{L^2(0,\infty)}^2 + 2 \int_{L_n}^{\infty} r(\tau)^2 \, d\tau \longrightarrow 0, \quad (14.14)$$

by (11.5) and $r \in L^2(0, \infty)$. The difference of the two operators in the statement is

$$\int_0^D R(t_n, u) |v_u\rangle \langle v_u| \, du + \int_D^{T_n} R(t_n, u) F_{u,D} \, du.$$

The first term has Hilbert–Schmidt norm at most $\int_0^D R(t_n, u) \, du \leq \sqrt{D} (\int_0^{T_n} R(t_n, u)^2 \, du)^{1/2}$, by the Cauchy–Schwarz inequality. The second term is bounded by (14.13) with $T = T_n$. Both bounds tend to zero by (14.14). $\square$

Theorem 14.7 (Susceptibility lower bound). *For the uniform-start cubic CK solution at $\beta \geq \beta_d$,*

$$\liminf_{t \rightarrow \infty} \chi(t, 0) \geq 2(1 - q). \quad (14.15)$$

More precisely, for every fixed $D > 0$,

$$\liminf_{t \rightarrow \infty} \chi(t, 0) \geq (1 - q) + \frac{q(1 - q)}{c(2D)}. \quad (14.16)$$

Proof. Let $\gamma = \liminf_{t \rightarrow \infty} \chi(t, 0) \in [0, \infty]$. If $\gamma = \infty$, there is nothing to prove. Otherwise, we choose $t_n \rightarrow \infty$ with $\chi(t_n, 0) \rightarrow \gamma$. All later bounds on the total response mass are along this sequence. Because $c(2D) \downarrow q$ as $D \rightarrow \infty$, the bound (14.16) implies (14.15). We therefore fix $D > 0$ and prove (14.16).

The proof has six steps.

1. We choose ages $L_n \rightarrow \infty$ such that the centers $U_n = U_{t_n, L_n}$ approach the ceiling and have the limits (14.17) below.
2. By (14.12), the root operator $\mathcal{I}^* U_n$ splits into an old part A_n , which is an integral over the times before $T_n = t_n - L_n$, and a recent part B_n .
3. In the coordinates of Theorem 13.2, the recent part converges to the rank-one operator $2^{-1/2} |\mathbf{1}\rangle \langle \mathbf{1}|$.
4. It follows that the old part converges in Hilbert–Schmidt norm to a positive operator $\mathcal{S}$ with trace $2^{-1/2}$.
5. We smooth the old part with Lemma 14.6. The smoothed operator is positive, and the lower semicontinuity of the trace shows that its trace is at least $2^{-1/2}$ in the limit.
6. The trace of the smoothed operator is explicit in the limit, and it is proportional to the old response mass $\gamma - (1 - q)$. Comparing with Step 5 gives (14.16).

Step 1: the diagonal choice. Recall the notation $M_T(a) = P_T v_{T+a}$ and $Q_T(a, b) = \langle M_T(a), M_T(b) \rangle$ from the proof of Lemma 14.1. Fix an integer $k \geq 1$. By (7.16), by (14.1) with $D = k$ and $s = T + k$, and by (8.11), the following three statements hold for all large T :

- (i) $\sup_{0 \leq a, b \leq k} |Q_T(a, b) - c(a + b)| \leq 1/k$;
- (ii) $M_T(k) \neq 0$, and $|f(M_T(k) / \|M_T(k)\|) - F_* + \delta_k / \sqrt{3}| \leq 1/k$;
- (iii) $|\int_0^k R(T + k, T + k - \tau) \, d\tau - (1 - c(k))| \leq 1/k$.

For (iii), we used that $\int_0^k R(T + k, T + k - \tau) \, d\tau \rightarrow \int_0^k r = 1 - c(k)$. Let θ_k be a threshold such that (i)–(iii) hold for all $T \geq \theta_k$, chosen nondecreasing in k and satisfying $\theta_k \geq k$.

Discarding finitely many n , we may assume that $t_n \geq 1 + \theta_1$. Let L_n be the largest integer $k \geq 1$ with $t_n \geq k + \theta_k$. For each fixed K , we have $t_n \geq K + \theta_K$ for all large n , because $t_n \rightarrow \infty$; therefore $L_n \rightarrow \infty$. Let $T_n = t_n - L_n$. Then $T_n \geq \theta_{L_n} \geq L_n \rightarrow \infty$, and (i)–(iii) hold with $k = L_n$ and $T = T_n$. We define

$$d_n = \|M_{t_n, L_n}\|^2 = Q_{T_n}(L_n, L_n), \quad U_n = U_{t_n, L_n} = \frac{M_{T_n}(L_n)}{\sqrt{d_n}}.$$

Note that $\chi(t_n, T_n) = \int_{T_n}^{t_n} R(t_n, u) \, du = \int_0^{L_n} R(t_n, t_n - \tau) \, d\tau$. Since $c(L_n) \rightarrow q$, $c(2L_n) \rightarrow q$, and $\delta_{L_n} \rightarrow 0$, the statements (i)–(iii) give

$$d_n \rightarrow q, \quad \chi(t_n, T_n) \rightarrow 1 - q, \quad f(U_n) \rightarrow F_*. \quad (14.17)$$

We also record a consequence of (i) for the recent states. Fix $W > 0$, and let n be so large that $L_n \geq W$. For $0 \leq \tau \leq W$, we have $P_{T_n} v_{t_n - \tau} = M_{T_n}(L_n - \tau)$, and (i) gives

$$\begin{aligned} \|P_{T_n} v_{t_n - \tau} - \sqrt{q} U_n\|^2 &= Q_{T_n}(L_n - \tau, L_n - \tau) + q - \frac{2\sqrt{q} Q_{T_n}(L_n - \tau, L_n)}{\sqrt{d_n}} \\ &= c(2L_n - 2\tau) + q - \frac{2\sqrt{q} c(2L_n - \tau)}{\sqrt{d_n}} + o(1), \end{aligned}$$

uniformly in $\tau \in [0, W]$. The function c is nonincreasing with limit q , which gives $c(2L_n - 2\tau) \rightarrow q$ and $c(2L_n - \tau) \rightarrow q$ uniformly in $\tau \in [0, W]$. Together with $d_n \rightarrow q$, this shows that

$$\sup_{0 \leq \tau \leq W} \|P_{T_n} v_{t_n - \tau} - \sqrt{q} U_n\| \rightarrow 0. \quad (14.18)$$

Step 2: the old and recent parts of the root. For $u \leq T_n$, we have $v_u \in \mathcal{H}_u \subset \mathcal{H}_{T_n}$ and $P_{T_n} v_u = v_u$. Applying (14.12) with $T = T_n$ and $t = t_n$, and dividing by $\sqrt{d_n}$, we obtain $\mathcal{I}^* U_n = A_n + B_n$, where

$$\begin{aligned} A_n &= \frac{\beta\sqrt{3}}{\sqrt{d_n}} \int_0^{T_n} R(t_n, u) |v_u\rangle \langle v_u| \, du, \\ B_n &= \frac{\beta\sqrt{3}}{\sqrt{d_n}} \int_{T_n}^{t_n} R(t_n, u) |P_{T_n} v_u\rangle \langle P_{T_n} v_u| \, du. \end{aligned}$$

Both operators are positive and trace class, as Bochner integrals in trace norm of positive rank-one operators with trace norms at most one.

Step 3: the recent part. By (14.17), Theorem 13.2 applies to the unit vectors U_n . Let V_n be the embeddings it provides, so that $V_n U_n = \mathbf{1}$ and $V_n \mathcal{I}^* U_n V_n^* \rightarrow \mathcal{T} \oplus 0$ in Hilbert–Schmidt norm. We transport every operator by V_n and write $\hat{X} = V_n X V_n^*$. This preserves

Hilbert–Schmidt distances, positivity, and traces. We write $|\mathbf{1}\rangle\langle\mathbf{1}|$ also for the operator $|\mathbf{1}\rangle\langle\mathbf{1}| \oplus 0$. We claim that

$$\widehat{B}_n \longrightarrow \frac{1}{\sqrt{2}}|\mathbf{1}\rangle\langle\mathbf{1}| \quad \text{in Hilbert–Schmidt norm.} \quad (14.19)$$

To prove the claim, fix $W > 0$ and let n be so large that $L_n \geq W$. We split $B_n = B'_n + B''_n$ at the time $t_n - W$, where B'_n is the integral over $[t_n - W, t_n]$. We write $w_{n,\tau} = P_{T_n} v_{t_n - \tau}$ and $\chi_n(W) = \int_0^W R(t_n, t_n - \tau) d\tau$. Using the rank-one estimate

$$\| |x\rangle\langle x| - |y\rangle\langle y| \|_{\text{HS}} \leq (\|x\| + \|y\|)\|x - y\|$$

with $x = w_{n,\tau}$ and $y = \sqrt{q} U_n$, both of norm at most one, we find

$$\left\| B'_n - \frac{\beta\sqrt{3}q\chi_n(W)}{\sqrt{d_n}} |U_n\rangle\langle U_n| \right\|_{\text{HS}} \leq \frac{2\beta\sqrt{3}}{\sqrt{d_n}} \chi_n(W) \sup_{0 \leq \tau \leq W} \|w_{n,\tau} - \sqrt{q} U_n\| \longrightarrow 0,$$

by (14.18). Note that $\chi_n(W) \rightarrow \int_0^W r$, by (8.11). Since $V_n |U_n\rangle\langle U_n| V_n^* = |\mathbf{1}\rangle\langle\mathbf{1}|$ and $d_n \rightarrow q$, it follows that

$$\widehat{B}'_n \longrightarrow \beta\sqrt{3}\sqrt{q} \left(\int_0^W r(\tau) d\tau \right) |\mathbf{1}\rangle\langle\mathbf{1}|.$$

For the remaining part, since $\|P_{T_n} v_u\| \leq 1$, we have

$$\begin{aligned} \|\widehat{B}''_n\|_{\text{HS}} &\leq \frac{\beta\sqrt{3}}{\sqrt{d_n}} \int_{T_n}^{t_n - W} R(t_n, u) du = \frac{\beta\sqrt{3}}{\sqrt{d_n}} [\chi(t_n, T_n) - \chi_n(W)] \\ &\longrightarrow \frac{\beta\sqrt{3}}{\sqrt{q}} \left[(1 - q) - \int_0^W r(\tau) d\tau \right], \end{aligned}$$

by (14.17). Combining the last two displays, we obtain

$$\limsup_{n \rightarrow \infty} \left\| \widehat{B}_n - \frac{1}{\sqrt{2}} |\mathbf{1}\rangle\langle\mathbf{1}| \right\|_{\text{HS}} \leq \left| \beta\sqrt{3}\sqrt{q} \int_0^W r - \frac{1}{\sqrt{2}} \right| + \frac{\beta\sqrt{3}}{\sqrt{q}} \left[(1 - q) - \int_0^W r \right].$$

The left side does not depend on W . As $W \rightarrow \infty$, we have $\int_0^W r \rightarrow 1 - q$, and the right side tends to zero by (14.11). This proves (14.19).

Step 4: the old part. By Theorem 13.2 and (14.19), $\widehat{A}_n = V_n \mathcal{I}^* U_n V_n^* - \widehat{B}_n$ converges in Hilbert–Schmidt norm to $\mathcal{S} \oplus 0$, where

$$\mathcal{S} = \mathcal{T} - \frac{1}{\sqrt{2}} |\mathbf{1}\rangle\langle\mathbf{1}| = \frac{1}{2\sqrt{2}} \int_0^1 z^{-3/2} |\mathbf{1}_{[0,z]}\rangle\langle\mathbf{1}_{[0,z]}| dz.$$

The second expression comes from the frame (13.11). It shows that $\mathcal{S}$ is positive and that $\text{tr } \mathcal{S} = \text{tr } \mathcal{T} - 2^{-1/2} = 2^{-1/2}$.

Step 5: smoothing and lower semicontinuity of the trace. Let n be so large that $T_n \geq D$, and define the smoothed old part

$$A_{n,D} = \frac{\beta\sqrt{3}}{\sqrt{d_n}} \int_D^{T_n} R(t_n, u) |M_{u,D}\rangle \langle M_{u,D}| \, du.$$

It is a positive trace-class operator. By the last assertion of Lemma 14.6, applied with the sequences t_n and L_n , and since $d_n \rightarrow q$, we have $\|A_{n,D} - A_n\|_{\text{HS}} \rightarrow 0$. By Step 4, $\hat{A}_{n,D} \rightarrow \mathcal{S} \oplus 0$ in Hilbert–Schmidt norm. We claim that

$$\liminf_{n \rightarrow \infty} \text{tr } A_{n,D} \geq \text{tr } \mathcal{S} = \frac{1}{\sqrt{2}}. \quad (14.20)$$

This is the classical lower semicontinuity of the trace on positive operators, and not continuity, since the convergence holds only in Hilbert–Schmidt norm. We include the short proof. Let Q be an orthogonal projection of finite rank on $L^2(0,1) \oplus \mathcal{K}$. For Hilbert–Schmidt operators X and Y , we have $\text{tr}(QXQ) = \text{tr}(QX) = \langle Q, X \rangle_{\text{HS}}$, so

$$|\text{tr}(QXQ) - \text{tr}(QYQ)| \leq \|Q\|_{\text{HS}} \|X - Y\|_{\text{HS}} = (\text{rank } Q)^{1/2} \|X - Y\|_{\text{HS}}.$$

In particular, $\text{tr}(Q\hat{A}_{n,D}Q) \rightarrow \text{tr}(Q(\mathcal{S} \oplus 0)Q)$. For a positive trace-class operator X , we have $\text{tr}(QXQ) \leq \text{tr } X$, as one sees by computing the trace in an orthonormal basis that contains a basis of the range of Q . Since $\text{tr } \hat{A}_{n,D} = \text{tr } A_{n,D}$, we obtain

$$\liminf_{n \rightarrow \infty} \text{tr } A_{n,D} \geq \text{tr}(Q(\mathcal{S} \oplus 0)Q).$$

Let (e_i) be an orthonormal basis of $L^2(0,1)$, and let Q_k be the projection onto the span of $e_1, \dots, e_k$. Then $\text{tr}(Q_k(\mathcal{S} \oplus 0)Q_k) = \sum_{i \leq k} \langle e_i, \mathcal{S}e_i \rangle$ increases to $\text{tr } \mathcal{S}$ as $k \rightarrow \infty$. This proves (14.20).

Step 6: the trace of the smoothed old part. The trace of a rank-one operator $|x\rangle\langle x|$ is $\|x\|^2$, so

$$\text{tr } A_{n,D} = \frac{\beta\sqrt{3}}{\sqrt{d_n}} \int_D^{T_n} R(t_n, u) z_D(u) \, du, \quad z_D(u) = \|M_{u,D}\|^2.$$

Recall from (7.17) that $z_D(u) \rightarrow c(2D)$ as $u \rightarrow \infty$. By (14.17), the old response mass satisfies

$$\int_0^{T_n} R(t_n, u) \, du = \chi(t_n, 0) - \chi(t_n, T_n) \longrightarrow \gamma - (1 - q).$$

Fix $u_0 \geq D$. For n so large that $T_n \geq u_0$, the Cauchy–Schwarz inequality and (14.14) give $\int_0^{u_0} R(t_n, u) \, du \leq \sqrt{u_0} (\int_0^{T_n} R(t_n, u)^2 \, du)^{1/2} \rightarrow 0$. Splitting the integrals at u_0 and using $0 \leq z_D \leq 1$ and $c(2D) \leq 1$, we find

$$\begin{aligned} & \left| \int_D^{T_n} R(t_n, u) z_D(u) \, du - c(2D) \int_0^{T_n} R(t_n, u) \, du \right| \\ & \leq 2 \int_0^{u_0} R(t_n, u) \, du + \sup_{u \geq u_0} |z_D(u) - c(2D)| \int_0^{T_n} R(t_n, u) \, du. \end{aligned}$$

We recall that $\gamma < \infty$. The upper limit of the right side as $n \rightarrow \infty$ is at most $\sup_{u \geq u_0} |z_D(u) - c(2D)|(\gamma - (1 - q))$, which tends to zero as $u_0 \rightarrow \infty$. Together with $d_n \rightarrow q$, this gives

$$\mathrm{tr} A_{n,D} \longrightarrow \frac{\beta\sqrt{3}}{\sqrt{q}} c(2D)(\gamma - (1 - q)).$$

Comparing with (14.20) and using (14.11), we obtain

$$\frac{\beta\sqrt{3}}{\sqrt{q}} c(2D)(\gamma - (1 - q)) \geq \frac{1}{\sqrt{2}} = \beta\sqrt{3}\sqrt{q}(1 - q).$$

Multiplying by $\sqrt{q}/(\beta\sqrt{3})$, we find $c(2D)(\gamma - (1 - q)) \geq q(1 - q)$, which is (14.16). This completes the proof. $\square$

Remark 14.8. In the asymptotic solution of Cugliandolo and Kurchan [51, arXiv version 1, paragraph after Equation (12) and Equation (13), p. 6], the slow sector has the constant fluctuation–dissipation ratio $x = (p - 2)(1 - q)/q$. Under this prediction, the total susceptibility tends to $(1 - q) + xq = (p - 1)(1 - q)$, which is $2(1 - q)$ for $p = 3$; formally, this is the relation (17.1) with $c = 0$. Theorem 14.7 proves the corresponding lower bound. The matching upper bound (17.2) is open; see Section 17.

A Fisher cost for relaxation below the plateau

In this section, we prove Theorem 15.1, which is Theorem 1.4. It concerns the relaxation of the cubic CK solution below the plateau, measured in physical time. We show that a growing time window $[s, s + \tau]$ on which $\tau \int_s^{s+\tau} a_F$ tends to zero ends at the plateau, and that every first passage of the correlation below a level $c \in (0, q)$ costs a positive amount of this quantity.

In the proof, we compare the backward response equation of Proposition 4.5, integrated over the window, with a lower bound on its memory term. The Fisher control of the backward defect (Theorem 5.3), which is the form taken here by the entropy-production mechanism of Cugliandolo, Dean, and Kurchan [54, arXiv version 2, Equations (12)–(13), p. 3, and Equation (18), p. 4], makes the integrated defect on the window small. The additional input is the uniform three-time inequality of Theorem 14.3: for late times $v < u < w$ with large gaps, it gives $C(w, u)C(u, v) \leq qC(w, v) + o(1)$, and we use it to bound the memory term from below. Its product form has a physical predecessor in [52, arXiv version 1, Section 5.1, Equation (5.1.8)]. After the calibration of the selected sector, the two bounds differ by $3\beta^2(2q - 1)(q - c)^2$, which is positive because $q > 1/2$. We also use the remote envelope (14.6) of Proposition 14.2, the response localization of Corollary 11.2 and Theorem 11.4, the stationary window (8.11), and the forgetting of fixed waiting times (Theorem 8.8). These results prove Theorem 1.4, and their implications for relaxation rates are discussed in Section 17.

Standing assumptions and notation. Throughout this section, $p = 3$ and $\beta > \beta_d = \sqrt{4/3}$. By Theorem 8.5 and the discussion preceding it, the selected plateau satisfies $q > q_d = 1/2$. We write $\ell = \ell_\beta = \lim_{t \rightarrow \infty} \lambda(t)$. For $p = 3$, the marginality equation (8.6) and the multiplier formula (8.9) read

$$6\beta^2 q(1 - q)^2 = 1, \quad \ell = \frac{1}{1 - q} + 3\beta^2(1 - q^2). \quad (15.1)$$

Let a_F be the Fisher majorant fixed in Theorem 5.3, from the same dimension extraction used in the physical-fork comparisons. Let $E(t, s) = \partial_s C(t, s) - R(t, s)$ be the backward FDT defect. Recall from (11.14) that $|E(t, s)|^2 \leq a_F(s)$ for $t > s > 0$, and that $\int_0^\infty a_F < \infty$. Every row $C(t, \cdot)$ is absolutely continuous on $[0, t]$, with derivative $R(t, \cdot) + E(t, \cdot)$ almost everywhere (Theorem 5.3). We also use that $0 \leq R \leq C \leq 1$ (Proposition 4.6), that $\lambda \geq 1$ by (3.11), which gives $\ell \geq 1$, and that for $p = 3$ the kernel (4.8) is $\Sigma(t, u) =$

$6\beta^2 C(t, u)R(t, u)$. As in Section 14.1, the letter c without an argument denotes a level in $(0, q)$, and the stationary profile is always written with its argument, as in $c(2D)$.

Theorem 15.1 (Fisher-small windows and all-level passage costs). *Suppose that $s_n, \tau_n \geq 0$, $s_n \rightarrow \infty$, $\tau_n \rightarrow \infty$, and*

$$\tau_n \int_{s_n}^{s_n + \tau_n} a_F(u) \, du \longrightarrow 0. \quad (15.2)$$

Then

$$C(s_n + \tau_n, s_n) \rightarrow q, \quad \chi(s_n + \tau_n, s_n) \rightarrow 1 - q. \quad (15.3)$$

For a level $c \in (0, q)$ and $s \geq 0$, define the first passage lag

$$\tau_c(s) = \inf\{\tau \geq 0 : C(s + \tau, s) \leq c\}.$$

It is positive and finite, satisfies $C(s + \tau_c(s), s) = c$, tends to infinity as $s \rightarrow \infty$, and satisfies

$$\liminf_{s \rightarrow \infty} \tau_c(s) \int_s^{s + \tau_c(s)} a_F(u) \, du \geq \mathfrak{c}_\beta(c), \quad (15.4)$$

$$\liminf_{s \rightarrow \infty} \tau_c(s) \int_s^\infty a_F(u) \, du \geq \mathfrak{c}_\beta(c), \quad (15.5)$$

where the explicit positive constant is

$$\mathfrak{c}_\beta(c) = \min \left\{ 1, \left[\frac{3\beta^2(2q-1)(q-c)^2}{\ell + 15\beta^2} \right]^2 \right\}. \quad (15.6)$$

Remark 15.2 (The constant and the endpoint). By (15.1), we have $3\beta^2 = (2q(1-q))^2$ and $\ell = (1+3q)/(2q(1-q))$, so that

$$\ell + 15\beta^2 = \frac{6 + 2q - 3q^2}{2q(1-q)^2}, \quad \frac{3\beta^2(2q-1)(q-c)^2}{\ell + 15\beta^2} = \frac{(2q-1)(q-c)^2}{6 + 2q - 3q^2}.$$

For $1/2 < q < 1$ and a level $c \in (0, q)$, we have $6 + 2q - 3q^2 = 5 + (1-q)(1+3q) > 5$ and $0 < (2q-1)(q-c)^2 < 1$. The ratio is therefore less than $1/5$, and the minimum in (15.6) is always attained by its second term. At the endpoint $\beta = \beta_d$, the plateau is $q = 1/2$ and the formula for the constant vanishes. Although the cubic product inequality of Theorem 14.3 includes this endpoint, the positive Fisher cost and the Fisher-small-window conclusion above are asserted only for $\beta > \beta_d$. The strict inequality is essential to the argument below, in which the positive margin $3\beta^2(2q-1)(q-c)^2$ is the source of the cost.

The theorem follows from a quantitative statement about a single sequence of windows, which we prove first.

Lemma 15.3 (A sequential physical-window obstruction). *Let $0 \leq s_n \leq t_n$, $s_n \rightarrow \infty$, and $t_n - s_n \rightarrow \infty$, and suppose that $C(t_n, s_n) \rightarrow z \in (0, q)$. Suppose further that*

$$\varepsilon_n = \left[(t_n - s_n) \int_{s_n}^{t_n} a_F(u) \, du \right]^{1/2} \longrightarrow \varepsilon \leq 1.$$

Then

$$3\beta^2(2q-1)(q-z)^2 \leq (\ell + 9\beta^2)\varepsilon + 6\beta^2\varepsilon^2. \quad (15.7)$$

Proof. The proof has six steps.

1. The Fisher control bounds the integrated defect on every subinterval of the window by ε_n .
2. We integrate the backward response equation over the window. The result relates the response mass of the window to the memory integral I_n defined in (15.11) below, with an error of order ε_n .
3. We split I_n into an early strip, a bulk, and a recent strip of length $2D$, and compute the recent strip and the values of the correlation row at the ends of the bulk.
4. In the bulk, the product inequality bounds $1 - C(u, s)^2$ from below by a function of the correlation row $C(t, u)$, which we integrate exactly.
5. We take the limits $n \rightarrow \infty$ and then $D \rightarrow \infty$, and obtain a lower bound for I_n .
6. The calibration (15.1) turns the two estimates into (15.7).

We temporarily suppress the index n and write $s = s_n$, $t = t_n$, $c_n = C(t, s)$, and $Y(u) = C(t, u)$ for $u \in [s, t]$. We assume throughout that n is large; in particular, $s > 0$.

Step 1: Fisher control on the window. For all $s \leq v \leq w \leq t$, the bound $|E(w, u)| \leq \sqrt{a_F(u)}$, the Cauchy–Schwarz inequality, and $a_F \geq 0$ give

$$\int_v^w |E(w, u)| \, du \leq \left((w - v) \int_v^w a_F(u) \, du \right)^{1/2} \leq \varepsilon_n. \quad (15.8)$$

The function Y is absolutely continuous with derivative $R(t, \cdot) + E(t, \cdot)$, and $Y(t) = 1$. Integrating over $[s, t]$, we find $1 - c_n = \chi(t, s) + \int_s^t E(t, u) \, du$. By (15.8) with $v = s$ and $w = t$, and since $c_n \geq 0$, we have

$$|\chi(t, s) - (1 - c_n)| \leq \varepsilon_n, \quad \chi(t, s) \leq 1 + \varepsilon_n. \quad (15.9)$$

This bound on the response mass of the window comes from the Fisher control alone. We have no bound on $\chi(t, 0)$, and none is needed.

Step 2: the integrated response equation. Recall from Proposition 4.5 that the derivative $\partial_v R(t, v)$ is continuous on the closed triangle $s \leq v \leq t$, and note that $R(t, t) = 1$. Integrating (4.21) in its second variable over $[s, t]$, and exchanging the order of integration

on the finite triangle $\{s \leq v \leq u \leq t\}$, where the integrand is continuous and bounded, we obtain

$$1 - R(t, s) = \int_s^t \lambda(u) R(t, u) \, du - 6\beta^2 \int_s^t R(t, u) \int_s^u C(u, v) R(u, v) \, dv \, du. \quad (15.10)$$

We rewrite the inner integral with the chain rule. For $s \leq u \leq t$, the function $v \mapsto C(u, v)^2$ is absolutely continuous on $[s, u]$, with derivative $2C(u, v)(R(u, v) + E(u, v))$, and $C(u, u) = 1$. Hence

$$\int_s^u C(u, v) R(u, v) \, dv = \frac{1 - C(u, s)^2}{2} - J_n(u), \quad J_n(u) = \int_s^u C(u, v) E(u, v) \, dv.$$

Using $0 \leq C \leq 1$ and (15.8) with $v = s$ and $w = u$, we find $|J_n(u)| \leq \varepsilon_n$. We define the memory integral

$$I_n = \int_s^t [1 - C(u, s)^2] R(t, u) \, du. \quad (15.11)$$

Inserting the expression for the inner integral into (15.10), writing $\lambda = \ell + (\lambda - \ell)$, and rearranging, we find the exact identity

$$\begin{aligned} 1 - \ell(1 - c_n) + 3\beta^2 I_n &= R(t, s) + \int_s^t (\lambda(u) - \ell) R(t, u) \, du \\ &\quad + \ell [\chi(t, s) - (1 - c_n)] + 6\beta^2 \int_s^t R(t, u) J_n(u) \, du. \end{aligned} \quad (15.12)$$

We bound the absolute values of the last three terms. The first is at most $\sup_{u \geq s_n} |\lambda(u) - \ell| (1 + \varepsilon_n)$, which tends to zero because $\lambda(u) \rightarrow \ell$ (Theorem 8.5) and ε_n is bounded. By (15.9), the second is at most $\ell \varepsilon_n$, and the third is at most $6\beta^2 \varepsilon_n \chi(t, s) \leq 6\beta^2 (1 + \varepsilon_n) \varepsilon_n$. We conclude that

$$|1 - \ell(1 - c_n) + 3\beta^2 I_n| \leq R(t, s) + o(1) + [\ell + 6\beta^2 (1 + \varepsilon_n)] \varepsilon_n, \quad (15.13)$$

where $o(1)$ tends to zero as $n \rightarrow \infty$.

Step 3: the recent strip and the ends of the bulk. Fix $D > 0$, and let n be so large that $t - s > 4D$. Let $s' = s + 2D$ and $t' = t - 2D$, and split I_n into the integrals over $[s, s']$, $[s', t']$, and $[t', t]$. The first is nonnegative, since $R \geq 0$ and $C \leq 1$.

On the recent strip, we first show that $C(u, s)$ is close to c_n . For $u \in [t', t]$,

$$|C(u, s) - c_n| \leq \int_u^t |\partial_x C(x, s)| \, dx \leq 2D \sup_{x \in [t', t]} |\partial_x C(x, s)|.$$

Fix $L > 0$. For large n , every $x \in [t', t]$ satisfies $x - s \geq t - s - 2D \geq L$, and $x \geq t' \rightarrow \infty$. Then (11.12) implies that $\limsup_n \sup_{x \in [t', t]} |\partial_x C(x, s)| \leq \sqrt{c''(2L)}$. The right side tends

to zero as $L \rightarrow \infty$, by the second inequality in (11.12). Together with $c_n \rightarrow z$, this shows that $\sup_{u \in [t', t]} |C(u, s) - z| \rightarrow 0$. Using (8.11), we find that the response mass of the strip satisfies $\int_{t'}^t R(t, u) \, du = \int_0^{2D} R(t, t - \tau) \, d\tau \rightarrow \int_0^{2D} r = 1 - c(2D)$. Therefore

$$\int_{t'}^t [1 - C(u, s)^2] R(t, u) \, du \longrightarrow [1 - c(2D)](1 - z^2). \quad (15.14)$$

We next consider the values of Y at the ends of the bulk. The stationary window (8.11) also gives $Y(t') = C(t, t - 2D) \rightarrow c(2D)$. At the other end,

$$\begin{aligned} |Y(s') - c_n| &\leq \int_s^{s+2D} R(t, u) \, du + \int_s^{s+2D} |E(t, u)| \, du \\ &\leq 2D \sup_{s \leq u \leq s+2D} R(t, u) + \left(2D \int_{s_n}^{\infty} a_F(u) \, du \right)^{1/2}. \end{aligned}$$

The first term tends to zero by Corollary 11.2, because $t - u \geq t - s - 2D \rightarrow \infty$ uniformly for $u \leq s + 2D$ and $r(\tau) \rightarrow 0$. The second tends to zero because a_F is integrable and $s_n \rightarrow \infty$. Therefore $Y(s') \rightarrow z$.

Step 4: the bulk. Let

$$\pi_{n,D} = \sup_{s' \leq u \leq t'} [Y(u)C(u, s) - qc_n]_+.$$

For $u \in [s', t']$, the times $s < u < t$ have gaps $u - s \geq 2D$ and $t - u \geq 2D$. For fixed S , we have $s_n \geq S + D$ for large n , and then the triple (s, u, t) belongs to the set $\mathcal{A}_{S,D}$ of Theorem 14.3, and $[Y(u)C(u, s) - qc_n]_+ = [C(t, u)C(u, s) - qC(t, s)]_+$ is the bracket in (14.8). Letting $n \rightarrow \infty$ and then $S \rightarrow \infty$, we find

$$\lim_{D \rightarrow \infty} \limsup_{n \rightarrow \infty} \pi_{n,D} = 0. \quad (15.15)$$

To use the product bound, we introduce a clipped function and its primitive. For $\kappa > 0$, let

$$\varphi_\kappa(y) = \left[1 - \frac{\kappa^2}{y^2} \right]_+ \quad (y > 0), \quad \varphi_\kappa(0) = 0,$$

and

$$\Phi_\kappa(y) = \int_0^y \varphi_\kappa(x) \, dx = \begin{cases} 0, & 0 \leq y \leq \kappa, \\ y + \kappa^2/y - 2\kappa, & y > \kappa. \end{cases}$$

The function Φ_κ is continuously differentiable on $[0, \infty)$, with $\Phi'_\kappa = \varphi_\kappa \in [0, 1]$, and $(\kappa, y) \mapsto \Phi_\kappa(y)$ is continuous on $(0, \infty) \times [0, \infty)$.

Let $\kappa_n = qc_n$ and $\kappa'_n = \kappa_n + \pi_{n,D}$. For large n , we have $c_n \geq z/2 > 0$ and $\kappa_n > 0$. We claim that

$$1 - C(u, s)^2 \geq \varphi_{\kappa'_n}(Y(u)), \quad s' \leq u \leq t'.$$

If $Y(u) \leq \kappa'_n$, the right side is zero. Otherwise, by the definition of $\pi_{n,D}$, we have $Y(u)C(u, s) \leq \kappa'_n$, that is, $0 \leq C(u, s) \leq \kappa'_n/Y(u)$, and the claim follows. Next, for $0 < \kappa \leq \kappa'$ and $y \geq 0$, we have

$$0 \leq \varphi_\kappa(y) - \varphi_{\kappa'}(y) \leq 1 - \frac{\kappa^2}{\kappa'^2} \leq \frac{2(\kappa' - \kappa)}{\kappa},$$

as one checks separately for $y \leq \kappa$, $\kappa < y \leq \kappa'$, and $y > \kappa'$. With $\kappa = \kappa_n$ and $\kappa' = \kappa'_n$, and using (15.9), we obtain

$$\int_{s'}^{t'} [1 - C(u, s)^2] R(t, u) \, du \geq \int_{s'}^{t'} \varphi_{\kappa_n}(Y(u)) R(t, u) \, du - \frac{2(1 + \varepsilon_n)}{qc_n} \pi_{n,D}.$$

The function Y is absolutely continuous on $[s', t']$ with $Y' = R(t, \cdot) + E(t, \cdot)$ almost everywhere, and Φ_{κ_n} is continuously differentiable with a bounded derivative. Hence $\Phi_{\kappa_n} \circ Y$ is absolutely continuous, with derivative $\varphi_{\kappa_n}(Y)Y'$, and

$$\int_{s'}^{t'} \varphi_{\kappa_n}(Y(u)) R(t, u) \, du = \Phi_{\kappa_n}(Y(t')) - \Phi_{\kappa_n}(Y(s')) - \int_{s'}^{t'} \varphi_{\kappa_n}(Y(u)) E(t, u) \, du.$$

This identity is oriented: it does not require Y to be monotone, and it holds when Y returns to earlier values or crosses the level κ_n . By (15.8) and $0 \leq \varphi_{\kappa_n} \leq 1$, the last integral is at most ε_n in absolute value. We conclude that

$$\int_{s'}^{t'} [1 - C(u, s)^2] R(t, u) \, du \geq \Phi_{qc_n}(Y(t')) - \Phi_{qc_n}(Y(s')) - \varepsilon_n - \frac{2(1 + \varepsilon_n)}{qc_n} \pi_{n,D}. \quad (15.16)$$

Step 5: the limits. We first take the lower limit as $n \rightarrow \infty$ with D fixed. We have $qc_n \rightarrow qz$, and by Step 3, $Y(t') \rightarrow c(2D)$ and $Y(s') \rightarrow z$. By the continuity of $(\kappa, y) \mapsto \Phi_\kappa(y)$,

$$\Phi_{qc_n}(Y(t')) - \Phi_{qc_n}(Y(s')) \longrightarrow \Phi_{qz}(c(2D)) - \Phi_{qz}(z) = \int_z^{c(2D)} \left(1 - \frac{q^2 z^2}{y^2}\right) dy,$$

where we used that $y \geq z > qz$ on $[z, c(2D)]$, so that no clipping occurs. Combining the nonnegativity of the early strip, (15.14), and (15.16), we obtain

$$\liminf_{n \rightarrow \infty} I_n \geq [1 - c(2D)](1 - z^2) + \int_z^{c(2D)} \left(1 - \frac{q^2 z^2}{y^2}\right) dy - \varepsilon - \frac{2(1 + \varepsilon)}{qz} \limsup_{n \rightarrow \infty} \pi_{n,D}.$$

The left side does not depend on D . We let $D \rightarrow \infty$, using $c(2D) \downarrow q$ and (15.15). Since $\int_z^q (1 - q^2 z^2 / y^2) dy = q - z - q^2 z + qz^2$, this gives

$$\liminf_{n \rightarrow \infty} I_n \geq (1 - q)(1 - z^2) + q - z - q^2 z + qz^2 - \varepsilon. \quad (15.17)$$

Step 6: the calibrated margin. Let $\Psi(z) = (1-q)(1-z^2) + q - z - q^2z + qz^2$ be the main term in (15.17). Expanding, we find $\Psi(z) = 1 - (1+q^2)z + (2q-1)z^2$, and hence

$$\Psi(z) - (1-q^2)(1-z) = q^2 - 2q^2z + (2q-1)z^2.$$

By the second identity in (15.1), we have $1 - \ell(1-z) = (z-q)/(1-q) - 3\beta^2(1-q^2)(1-z)$, and by the first, $1/(1-q) = 6\beta^2q(1-q)$. Therefore

$$\begin{aligned} 1 - \ell(1-z) + 3\beta^2\Psi(z) &= 3\beta^2[2q(1-q)(z-q) + q^2 - 2q^2z + (2q-1)z^2] \\ &= 3\beta^2(2q-1)(q-z)^2. \end{aligned} \quad (15.18)$$

In the second line, we expanded both sides: each equals $3\beta^2[(2q-1)z^2 - 4q^2z + 2qz + 2q^3 - q^2]$.

We now combine the estimates. Since $s_n \rightarrow \infty$ and $t_n - s_n \rightarrow \infty$, Corollary 11.2 implies that $R(t_n, s_n) \rightarrow 0$. Taking the lower limit of the left side of (15.13) and using $c_n \rightarrow z$, (15.17), and (15.18), we obtain

$$3\beta^2(2q-1)(q-z)^2 - 3\beta^2\varepsilon \leq \liminf_{n \rightarrow \infty} [1 - \ell(1-c_n) + 3\beta^2I_n] \leq [\ell + 6\beta^2(1+\varepsilon)]\varepsilon.$$

The memory bound (15.17) loses $3\beta^2\varepsilon$, and the response equation (15.13) loses $\ell\varepsilon + 6\beta^2\varepsilon + 6\beta^2\varepsilon^2$. Adding these losses, we arrive at

$$3\beta^2(2q-1)(q-z)^2 \leq (\ell + 9\beta^2)\varepsilon + 6\beta^2\varepsilon^2,$$

which is (15.7). □

Proof of Theorem 15.1. We first prove (15.3). Let $t_n = s_n + \tau_n$. Then $\varepsilon_n = (\tau_n \int_{s_n}^{t_n} a_F)^{1/2} \rightarrow 0$ by (15.2), and (14.6) implies that $\limsup_n C(t_n, s_n) \leq q$. Since C takes values in $[0, 1]$, it suffices to show that every subsequential limit z of $C(t_n, s_n)$ equals q ; we already know that $z \leq q$. Passing to a subsequence, we may assume that $C(t_n, s_n) \rightarrow z$.

If $z \in (0, q)$, then Lemma 15.3 with $\varepsilon = 0$ implies that $3\beta^2(2q-1)(q-z)^2 \leq 0$. This is impossible, because $q > 1/2$ and $z < q$.

If $z = 0$, we fix $z' \in (0, q)$. For large n , we have $C(t_n, s_n) < z'$, while $C(s_n, s_n) = 1 > z'$. The correlation is continuous, and the first crossing lag $\sigma_n = \inf\{\sigma \geq 0 : C(s_n + \sigma, s_n) \leq z'\}$ satisfies $\sigma_n \leq \tau_n$ and $C(s_n + \sigma_n, s_n) = z'$. The lags σ_n tend to infinity. Indeed, for fixed $W > 0$, (8.11) shows that $C(s_n + \sigma, s_n) \rightarrow c(\sigma) \geq q > z'$ uniformly in $\sigma \in [0, W]$, and this forces $\sigma_n > W$ for large n . Since $a_F \geq 0$ and $\sigma_n \leq \tau_n$,

$$\sigma_n \int_{s_n}^{s_n + \sigma_n} a_F(u) \, du \leq \tau_n \int_{s_n}^{s_n + \tau_n} a_F(u) \, du \rightarrow 0.$$

Lemma 15.3, applied to the windows $[s_n, s_n + \sigma_n]$ with z' in place of z and with $\varepsilon = 0$, again gives a contradiction. We conclude that $z = q$, and that $C(t_n, s_n) \rightarrow q$. Finally, by (15.9), we have $|\chi(t_n, s_n) - (1 - C(t_n, s_n))| \leq \varepsilon_n \rightarrow 0$, and therefore $\chi(t_n, s_n) \rightarrow 1 - q$.

We now turn to the passage lag. Fix a level $c \in (0, q)$ and $s \geq 0$. By Theorem 8.8, $C(s + \tau, s) \rightarrow 0$ as $\tau \rightarrow \infty$, and the set $\{\tau \geq 0 : C(s + \tau, s) \leq c\}$ is nonempty. It is closed, because C is continuous. Then $\tau_c(s)$ is finite and $C(s + \tau_c(s), s) \leq c$. We have $C(s, s) = 1$ and $C(s + \tau, s) > c$ for $0 \leq \tau < \tau_c(s)$, and by continuity, $C(s + \tau_c(s), s) = c$. As in the case $z = 0$ above, (8.11) and the bound $c(\tau) \geq q > c$ on the profile show that, for each fixed W , we have $C(s + \tau, s) > c$ for all $\tau \in [0, W]$ once s is large. We conclude that $\tau_c(s) \rightarrow \infty$.

Let $A = \liminf_{s \rightarrow \infty} \tau_c(s) \int_s^{s+\tau_c(s)} a_F(u) \, du \in [0, \infty]$. If $A \geq 1$, then (15.4) holds, since $\mathfrak{c}_\beta(c) \leq 1$. Suppose that $A < 1$, and choose $s_n \rightarrow \infty$ along which the quantity in the definition of A converges to A . Let $t_n = s_n + \tau_c(s_n)$. Then $t_n - s_n \rightarrow \infty$, $C(t_n, s_n) = c \in (0, q)$, and $\varepsilon_n \rightarrow \sqrt{A} < 1$. By Lemma 15.3 with $z = c$ and $\varepsilon = \sqrt{A}$, and since $A \leq \sqrt{A}$, we have

$$3\beta^2(2q-1)(q-c)^2 \leq (\ell + 9\beta^2)\sqrt{A} + 6\beta^2 A \leq (\ell + 9\beta^2)\sqrt{A} + 6\beta^2\sqrt{A} = (\ell + 15\beta^2)\sqrt{A}.$$

Squaring, we find $A \geq [3\beta^2(2q-1)(q-c)^2/(\ell + 15\beta^2)]^2 \geq \mathfrak{c}_\beta(c)$. This proves (15.4). Because $a_F \geq 0$, enlarging the domain of integration from $[s, s + \tau_c(s)]$ to $[s, \infty)$ can only increase the left side, and (15.5) follows. $\square$

Theorem 15.1 bounds the physical time needed to relax below the plateau in terms of the Fisher majorant. It does not determine the growth of $\tau_c(s)$, and it does not identify an aging clock; see Section 17.

PART V

The stationary approach to the plateau

Algebraic relaxation in the stationary sector

In this section, we prove Theorem 16.1. Together with the discussion of the exponent in Section 16.1, it proves Theorem 1.5. For $\beta \geq \beta_d$, it describes how the selected stationary profile c approaches its plateau q . The difference $c(t) - q$ decays like $A_\beta t^{-a}$, where the amplitude A_β is finite and positive, and the exponent $a \in (0, 1/2)$ is determined by q through the relation (16.2). Every fixed derivative of c has the corresponding power asymptotic.

The argument is an analysis of one scalar equation. It uses four results proved earlier: the regularity of a stationary cluster (Proposition 5.7), the scalar memory equation and the plateau identity (Proposition 6.1), the selection of the plateau together with marginality (Theorem 8.5), and the complete monotonicity of the profile (Corollary 7.7). The precise properties that we use are listed at the start of Section 16.1. The results of this section prove Theorem 1.5 and verify the power-law hypotheses in Corollaries 11.6 and 11.7. The section can be read once Part II is available.

The relation (16.2) between the exponent and the plateau is the classical exponent relation of mode-coupling theory. Götze analyzed the leading critical and beta-scaling asymptotics of schematic and full mode-coupling equations in [79, p. 144, Equation (11)] and [80, p. 195, Equation (1b), and Sections 3 and 5], the latter under a generic instability with an isolated critical stability eigenvalue. The cited critical power-law expansions concern an exponent parameter strictly between zero and one. An explicit statement of the relation is given in [81, Section 4, p. 881, paragraph following Equation (6)]. The liquid-theory review gives the exponent relation for $\lambda_{\text{MCT}} \in [1/2, 1)$ [83, Section 3.3, Equation (3.19), p. 306]. The general polynomial discussion in [82, Section 6.1.2, Equations (6.11d) and (6.17b), pp. 444–447] includes $0 < \lambda_{\text{MCT}} < 1$, but assumes that the constructed low-frequency solution determines the asymptote of the actual critical correlator; it explicitly states on p. 447 that a proof of this matching assumption is unavailable. An aging analogue, relating the approach and departure exponents to the fluctuation–dissipation ratio, was analyzed in [53, arXiv version 1, Sections III and VI.D–E] and [34, Section IV, Equation (IV.6)]; see also [49, arXiv version 2, Section 12.3.8, Equation (12.36)]. For the pure spherical model, its parameter is $\vartheta = (p - 2)(1 - q)/(2q)$, with $0 < \vartheta \leq 1/2$ for the selected aging profile, and it is half of the fluctuation–dissipation ratio $(p - 2)(1 - q)/q$ of the aging regime predicted in [51, arXiv version 1, paragraph after Equation (12), p. 6], and the relation is written in this form in [96, arXiv version 1, p. 4,

paragraph following Equation (12)]. The structural theory of scalar mode-coupling equations in [84, Section 2, Theorem 1 and Corollaries 1.1–1.2, pp. 232–233], its extension to Brownian mixtures with completely monotone solutions [71, arXiv version 1, Section III], and the regenerative-process viewpoint of [114, Sections 1 and 5.2 and Appendix C] are mathematical predecessors of the argument below. The critical power law, with an amplitude that depends on the transient dynamics, is a classical asymptotic result of mode-coupling theory [81, Section 4, Equation (6), pp. 880–881]. We give a self-contained proof for the stationary profile selected by the CK solution, including a finite positive amplitude and the asymptotics of all fixed derivatives. The proof does not assume a power expansion of the profile and does not differentiate an asymptotic equivalence. The tools are classical: comparison principles, beta integrals, the Schur test with its exact remainder, a factorization of gamma variables, the key renewal theorem [68], and the monotone density theorem [30].

16.1 Inputs and statement of the result

Throughout this section, we fix $p \geq 3$ and $\beta \geq \beta_d$. Let (c, r) be the stationary profile and $\ell = \ell_\beta$ the limiting multiplier given by Theorem 8.5, and let $q = q_\beta$ be the plateau of c . We use the following properties of (c, r, q, ℓ) , and no other information about the CK solution.

- (S1) (Proposition 5.7.) The function r is continuously differentiable on $[0, \infty)$, and $c' = -r$ and $c(0) = r(0) = 1$.
- (S2) (Proposition 6.1.) The function r is strictly positive on $[0, \infty)$. With

$$\phi(y) = \gamma + \beta^2 p y^{p-1}, \quad \gamma = \ell - \beta^2 p,$$

we have the scalar memory equation (6.1),

$$c'(t) = -1 + \int_0^t \phi(c(t-s)) r(s) \, ds, \quad t \geq 0,$$

and the plateau identity $\phi(q)(1-q) = 1$ from (6.2).

- (S3) (Theorem 8.5.) The plateau satisfies $q_d \leq q < 1$ and the marginality condition

$$\beta^2 p(p-1)q^{p-2}(1-q)^2 = 1. \tag{16.1}$$

Moreover, $q = q_d$ if and only if $\beta = \beta_d$.

- (S4) (Corollary 7.7.) There exists a probability measure ν on $[0, \infty)$ such that

$$c(t) = \int_{[0, \infty)} e^{-xt} \nu(dx), \quad t \geq 0, \quad \nu(\{0\}) = q.$$

The last sentence of (S3) follows from Theorem 8.5, which identifies q as the larger root of (16.1), and this root lies in $[(p-2)/p, 1)$. The function $y \mapsto y^{p-2}(1-y)^2$ is strictly decreasing on this interval, which implies that the larger root is strictly increasing in β . The point q_d lies in the same interval, and a direct substitution shows that it is a root when $\beta = \beta_d$.

The exponent. We define

$$\vartheta = \frac{(p-2)(1-q)}{2q}, \quad \frac{\Gamma(1-a)^2}{\Gamma(1-2a)} = \vartheta, \quad 0 < a < \frac{1}{2}, \quad (16.2)$$

where Γ is Euler's gamma function. We now check that the second relation determines a uniquely. Recall that $q \geq q_d = (p-2)/(p-1)$. Then $(1-q)/q \leq 1/(p-2)$, which gives $0 < \vartheta \leq 1/2$. Let $\psi = \Gamma'/\Gamma$ be the digamma function. For $0 < \alpha < 1/2$, we have

$$\frac{d}{d\alpha} \log \frac{\Gamma(1-\alpha)^2}{\Gamma(1-2\alpha)} = 2[\psi(1-2\alpha) - \psi(1-\alpha)] < 0,$$

since ψ is strictly increasing on $(0, \infty)$. The quotient equals 1 at $\alpha = 0$ and tends to 0 as $\alpha \uparrow 1/2$, because $\Gamma(1-2\alpha) \rightarrow \infty$. We conclude that it is a strictly decreasing bijection from $[0, 1/2)$ onto $(0, 1]$. In particular, there exists exactly one $a \in (0, 1/2)$ satisfying (16.2).

At $\beta = \beta_d$, we have $q = q_d$, and then $\vartheta = 1/2$ for all p . Solving $\Gamma(1-a)^2/\Gamma(1-2a) = 1/2$ numerically, by bisection in double-precision arithmetic with the logarithm of the gamma function, yields $a = 0.3952630630\dots$ at $\beta = \beta_d$. For $\beta > \beta_d$, we have $q > q_d$ by (S3). Then $\vartheta < 1/2$, and a lies strictly between this value and $1/2$. In the cubic model, $\vartheta = (1-q)/(2q)$. These numerical values are not used in the proofs.

The relation (16.2) has a simple meaning. We write $(f * g)(t) = \int_0^t f(t-s)g(s) ds$ for the convolution of two functions on $[0, \infty)$; we use this notation throughout the section. For the pure power $w(t) = t^{-a}$, the beta integral gives $(w * w)(t) = B(1-a, 1-a) t^{1-2a}$, where B is Euler's beta function. Using $\Gamma(2-2a) = (1-2a)\Gamma(1-2a)$, we find

$$(w * w)'(t) = (1-2a)B(1-a, 1-a) t^{-2a} = \frac{\Gamma(1-a)^2}{\Gamma(1-2a)} w(t)^2. \quad (16.3)$$

In other words, (16.2) says that $(w * w)' = \vartheta w^2$. This is the balance of the two quadratic terms in the normalized equation (16.6) below; see (16.26). Because both sides of this balance are quadratic, it fixes the exponent but not the amplitude.

The result.

Theorem 16.1 (Stationary critical relaxation). *Fix $p \geq 3$ and $\beta \geq \beta_d$. Let c be the stationary profile of Theorem 8.5, let q be its plateau, and let $a \in (0, 1/2)$ be defined by (16.2). Then there exists $A_\beta \in (0, \infty)$ such that, as $t \rightarrow \infty$,*

$$c(t) - q \sim A_\beta t^{-a}, \quad (-1)^n c^{(n)}(t) \sim A_\beta(a)_n t^{-a-n} \quad \text{for every fixed integer } n \geq 1, \quad (16.4)$$

where $(a)_n = a(a+1) \cdots (a+n-1)$. In particular, $r(t) \sim aA_\beta t^{-1-a}$.

Remark 16.2. The exponent a depends only on ϑ , that is, on p and q . The proof determines the amplitude as a limit, $A_\beta = (1-q)^{1+a}A_0$ with A_0 given in (16.41), and gives no closed formula for it. In line with the remark after (16.3), marginality determines the exponent, while the amplitude is determined by the whole profile c .

Outline of the proof. The proof occupies the rest of this section. We first rescale time and the profile so that marginality says that the coefficient of the linear memory term equals one (Lemma 16.3). The normalized profile u then solves $u' + u + M(u) * u' = 0$, where M is a polynomial with $M(y) = y + \vartheta y^2 + O(y^3)$. The proof then has three steps.

1. *Comparison profiles* (Sections 16.2 and 16.3). A comparison principle (Lemma 16.4) turns the sign of the residual of an explicit profile into an inequality between that profile and u . For the shifted power $g(t) = (1+t)^{-a}$, the choice of a makes the quadratic part of the residual positive but of smaller order than g^3 , and the sign of the residual is decided by cubic terms. We fix this sign by correcting g with a small multiple of g^2 (Lemmas 16.5 and 16.6). The comparison profiles keep u itself up to a fixed time and continue it by a rescaled corrected power. This yields $A_-(1+t)^{-a} \leq u(t) \leq A_+(1+t)^{-a}$ (Proposition 16.7). These bounds identify the exponent, but they still allow $t^a u(t)$ to oscillate between A_- and A_+ on a logarithmic time scale.
2. *The logarithmic Dirichlet form* (Section 16.4). The function $Q = u * u$ satisfies $Q' = \vartheta u^2 + \mathcal{E}$, where the error $\mathcal{E}$ is integrable against $t^{-(1-2a)}$ (Lemma 16.8). Integrating this balance against the same weight and comparing the result with the Schur test for a homogeneous kernel, we deduce that $D(x) = e^{ax}u(e^x)$ has a finite Dirichlet energy for an explicit kernel on the line (Lemma 16.9). This implies that $D(x+h) - D(x) \rightarrow 0$ as $x \rightarrow \infty$ for every fixed h . This still allows D to vary slowly without converging.
3. *The renewal equation* (Section 16.5). We take the Laplace transform of the balance at the point e^{-x} . A factorization of gamma random variables turns the result into a linear renewal equation $f - \mathcal{B}f = g_0$ for a smoothed version f of D^2 , with an integrable source g_0 . We control the variance part of g_0 by the Dirichlet energy of the second step, and the rest by the error $\mathcal{E}$. Using the key renewal theorem, we show that f converges. By the second step, $f - D^2 \rightarrow 0$, and by the first step the limit of D lies in $(0, \infty)$. Finally, complete monotonicity and the monotone density theorem give the asymptotics of every derivative.

16.2 Normalization and comparison

We first rescale the profile so that marginality becomes a statement about a single coefficient. Recall the function ϕ and the plateau identity $\phi(q)(1-q) = 1$ from (S2). We define

the polynomial

$$M(y) = (1 - q)[\phi(q + (1 - q)y) - \phi(q)], \quad y \in \mathbb{R}. \quad (16.5)$$

Lemma 16.3 (The normalized equation). *Define $u: [0, \infty) \rightarrow \mathbb{R}$ by*

$$u(z) = \frac{c((1 - q)z) - q}{1 - q}.$$

Then u is continuously differentiable, and

$$u' + u + M(u) * u' = 0, \quad u(0) = 1, \quad M(y) = y + \sum_{j=2}^{p-1} \lambda_j y^j, \quad \lambda_2 = \vartheta, \quad (16.6)$$

where

$$\lambda_j = \frac{1}{p-1} \binom{p-1}{j} \left(\frac{1-q}{q}\right)^{j-1} \geq 0, \quad 2 \leq j \leq p-1. \quad (16.7)$$

Further, there exists a probability measure μ on $(0, \infty)$ such that

$$u(z) = \int_{(0, \infty)} e^{-\eta z} \mu(d\eta), \quad z \geq 0. \quad (16.8)$$

In particular, u is strictly positive and completely monotone on $(0, \infty)$, and $u(z) \rightarrow 0$ as $z \rightarrow \infty$. The function $v = -u'$ is continuous, strictly positive, and nonincreasing on $[0, \infty)$, with $v(0) = 1$ and $\int_0^\infty v(z) dz = 1$.

Proof. Recall from (S1) that c is continuously differentiable with $c' = -r$. By the chain rule, $u'(z) = c'((1 - q)z) = -r((1 - q)z)$; the factor $1 - q$ produced by the chain rule cancels the normalization. In particular, u is continuously differentiable, $u(0) = 1$, and $v(z) = r((1 - q)z)$; by (S2), v is strictly positive, and $v(0) = r(0) = 1$.

Fix $z \geq 0$. We use (6.1) at the time $(1 - q)z$ and substitute $s = (1 - q)w$ in its integral. Using $r((1 - q)w) = v(w)$ and $c((1 - q)(z - w)) = q + (1 - q)u(z - w)$, we find

$$u'(z) = -1 + (1 - q) \int_0^z \phi(q + (1 - q)u(z - w)) v(w) dw.$$

By (16.5), we have $(1 - q)\phi(q + (1 - q)y) = (1 - q)\phi(q) + M(y) = 1 + M(y)$, where we used the plateau identity. Inserting this and $\int_0^z v(w) dw = 1 - u(z)$ into the previous display, we conclude that

$$u'(z) = -1 + (1 - u(z)) + \int_0^z M(u(z - w)) v(w) dw = -u(z) - (M(u) * u')(z),$$

which is the first relation in (16.6).

Next, we compute M . By the binomial theorem,

$$M(y) = (1-q)\beta^2 p [(q + (1-q)y)^{p-1} - q^{p-1}] = \beta^2 p \sum_{j=1}^{p-1} \binom{p-1}{j} q^{p-1-j} (1-q)^{j+1} y^j.$$

The coefficient of y is $\beta^2 p(p-1)q^{p-2}(1-q)^2$, which equals 1 by the marginality condition (16.1). Dividing the coefficient of y^j by this quantity, we find that it equals λ_j in (16.7). For $j = 2$, this gives $\lambda_2 = (p-2)(1-q)/(2q) = \vartheta$. All coefficients are nonnegative because $0 < q < 1$. For $p = 3$, we have simply $M(y) = y + \vartheta y^2$.

Finally, by (S4), we have $c(t) - q = \int_{(0,\infty)} e^{-xt} \nu(dx)$ and $\nu((0,\infty)) = 1 - q > 0$. Let μ be the image of the probability measure $(1-q)^{-1}\nu(\cdot \cap (0,\infty))$ under the map $x \mapsto (1-q)x$. Then (16.8) holds. By the representation, we have $u > 0$, since μ is a nonzero measure on $(0,\infty)$. By dominated convergence, $u(z) \rightarrow 0$ as $z \rightarrow \infty$. For $n \geq 1$ and $z_0 > 0$, the function $\eta^n e^{-\eta z'}$ is bounded uniformly in $\eta > 0$ and $z' \geq z_0$. By induction on n , this justifies differentiating n times under the integral sign on (z_0, ∞) , and we have

$$(-1)^n u^{(n)}(z) = \int_{(0,\infty)} \eta^n e^{-\eta z} \mu(d\eta) > 0, \quad z > 0, n \geq 1.$$

For $n = 2$, this shows that v is nonincreasing on $(0, \infty)$, and by continuity also on $[0, \infty)$. Lastly, $\int_0^\infty v = u(0) - \lim_{z \rightarrow \infty} u(z) = 1$. $\square$

From now on, we write t for the normalized time variable. The physical lag reappears only at the end of the proof of Theorem 16.1. We record two properties of M . Its coefficients are nonnegative, and in particular M is increasing and convex on $[0, \infty)$, with $M(0) = 0$. The number $L_M = M'(1)$ is a Lipschitz constant for M on $[0, 1]$.

For a locally absolutely continuous, nonincreasing function $P: [0, \infty) \rightarrow [0, 1]$ with $P(0) = 1$, we define its residual

$$\mathcal{R}_P = P' + P + M(P) * P'.$$

It is defined almost everywhere, and $\mathcal{R}_u = 0$ by (16.6). By the following lemma, a nonnegative residual gives an upper bound for u , and a nonpositive residual gives a lower bound.

Lemma 16.4 (Comparison). *Let $P: [0, \infty) \rightarrow [0, 1]$ be locally absolutely continuous and non-increasing, with $P(0) = 1$. If $\mathcal{R}_P \geq 0$ almost everywhere, then $u \leq P$. If $\mathcal{R}_P \leq 0$ almost everywhere, then $P \leq u$.*

Proof. We begin with the first assertion. Set $w = P - u$, and recall that $v = -u' \geq 0$. We have

$$M(P) * P' - M(u) * u' = M(P) * w' - [M(P) - M(u)] * v.$$

Subtracting (16.6) from the definition of $\mathcal{R}_P$ and using this identity, we find

$$w' + w + M(P) * w' = \mathcal{R}_P + [M(P) - M(u)] * v. \quad (16.9)$$

The functions $M \circ P$ and w are locally absolutely continuous, and $w(0) = 0$ and $P(0) = 1$. Integrating by parts in the convolution, we have, for all $t \geq 0$,

$$(M(P) * w')(t) = \int_0^t M(P(t-s)) w'(s) \, ds = M(1) w(t) - (k_P * w)(t),$$

where $k_P = -(M \circ P)'$. We have $k_P \geq 0$ almost everywhere and $\int_0^t k_P = M(1) - M(P(t)) \leq M(1)$, because M is increasing on $[0, 1]$ and P is nonincreasing with values in $[0, 1]$. Inserting this into (16.9), we obtain, with $b_0 = 1 + M(1)$,

$$w' + b_0 w = \mathcal{R}_P + [M(P) - M(u)] * v + k_P * w \quad \text{almost everywhere.} \quad (16.10)$$

We now estimate the negative part $w_- = \max\{-w, 0\}$. Recall that M is increasing on $[0, 1]$, with Lipschitz constant L_M there. Since u and P take values in $[0, 1]$, we have $M(P) - M(u) \geq -L_M w_-$. We also have $w \geq -w_-$, and v and k_P are nonnegative. Using also $\mathcal{R}_P \geq 0$, we deduce from (16.10) that

$$(e^{b_0 t} w(t))' \geq -e^{b_0 t} [L_M (w_- * v)(t) + (k_P * w_-)(t)] \quad \text{almost everywhere.}$$

The function $e^{b_0 t} w(t)$ is locally absolutely continuous and vanishes at $t = 0$. Integrating, and using $e^{-b_0(t-s)} \leq 1$, we find

$$w(t) \geq - \int_0^t [L_M (w_- * v)(s) + (k_P * w_-)(s)] \, ds.$$

Since the integral on the right is nonnegative, $w_-(t)$ is bounded by the same integral. For a nonnegative locally integrable function k , we have, by Tonelli's theorem,

$$\int_0^t (w_- * k)(s) \, ds = \int_0^t k(\sigma) \int_0^{t-\sigma} w_-(s') \, ds' \, d\sigma \leq \|k\|_{L^1(0,t)} \int_0^t w_-(s) \, ds.$$

We have $\|v\|_{L^1(0,t)} = 1 - u(t) \leq 1$ and $\|k_P\|_{L^1(0,t)} \leq M(1)$. Combining these bounds, we find

$$w_-(t) \leq (L_M + M(1)) \int_0^t w_-(s) \, ds, \quad t \geq 0.$$

The function w_- is continuous and nonnegative, and Gronwall's inequality implies that $w_- = 0$, that is, $u \leq P$.

For the second assertion, we set $w = u - P$ and $v_P = -P' \geq 0$. Repeating the computation above, with the roles of u and P exchanged and with $u(0) = 1$ in the integration by parts, we have

$$w' + b_0 w = -\mathcal{R}_P + [M(u) - M(P)] * v_P + k_u * w, \quad k_u = -(M \circ u)' \geq 0,$$

almost everywhere. Now $-\mathcal{R}_P \geq 0$, $\|v_P\|_{L^1(0,t)} = 1 - P(t) \leq 1$, and $\|k_u\|_{L^1(0,t)} \leq M(1)$. Arguing as in the previous paragraph, we conclude that $w_- = 0$, that is, $P \leq u$. $\square$

16.3 Comparison profiles at the critical exponent

To apply Lemma 16.4, we need explicit profiles whose residuals have a definite sign. The natural candidates are multiples of the shifted power

$$g(t) = (1+t)^{-a}.$$

We define

$$D_g = (g * g)' - \vartheta g^2, \quad H = (g * g^2)' - \vartheta g^3, \quad D_2 = (g^2 * g^2)' - \vartheta g^4,$$

and, for $|\varepsilon| \leq 1/4$,

$$f_\varepsilon = (1+\varepsilon)g - \varepsilon g^2.$$

The next lemma explains why an uncorrected power does not suffice. By the choice of a , the quadratic part D_g of the residual of g is positive but of smaller order than g^3 , and the sign of the residual is decided by cubic terms. The cross term H that appears when g is corrected by a multiple of g^2 is of order g^3 and has a definite sign. Unless stated otherwise, constants denoted by K , possibly with indices, depend only on p and q , through a , ϑ , and the coefficients λ_j , and on the indicated indices; their values may change from line to line. Constants that also depend on the profile u are indicated as such.

Lemma 16.5 (Shifted-power estimates). *For every $t \geq 0$, we have $D_g(t) > 0$. Moreover, as $t \rightarrow \infty$,*

$$D_g(t) = O((1+t)^{-a-1}) = o(g(t)^3), \quad \frac{H(t)}{g(t)^3} \longrightarrow (1-3a)B(1-a, 1-2a) - \vartheta < 0. \quad (16.11)$$

When $3a \neq 1$, the limit equals $\Gamma(1-a)\Gamma(1-2a)/\Gamma(1-3a) - \vartheta$. There exists K such that $|D_2| \leq Kg^3$ on $[0, \infty)$. For every ε with $|\varepsilon| \leq 1/4$, the profile f_ε is positive and strictly decreasing, $f_\varepsilon(0) = 1$, $f_\varepsilon \leq Kg$, and $|f'_\varepsilon(t)| \leq K(1+t)^{-a-1}$. Finally, for every integer $j \geq 2$ there exists K_j such that, for all $|\varepsilon| \leq 1/4$ and $t \geq 0$,

$$|(f_\varepsilon^j * f_\varepsilon)'(t)| \leq K_j g(t)^3. \quad (16.12)$$

Proof. The proof has four steps. We first express the derivative of a convolution of two shifted powers through a truncated beta integral. We then use this formula to prove that $D_g > 0$, and next to prove (16.11). Finally, we bound derivatives of convolutions by splitting them at the midpoint, without differentiating an asymptotic formula.

Step 1: a beta formula. Let $\alpha, \zeta \in (0, 1)$ and $g_\alpha(t) = (1+t)^{-\alpha}$. For $t \geq 0$, set $T = t+2$ and $x = 1/T \in (0, 1/2]$. Substituting $y = (1+s)/T$ in the convolution integral, we find

$$(g_\alpha * g_\zeta)(t) = \int_0^t (1+t-s)^{-\alpha} (1+s)^{-\zeta} ds = T^{1-\alpha-\zeta} \int_x^{1-x} y^{-\alpha} (1-y)^{-\zeta} dy,$$

where we used the symmetry $y \mapsto 1 - y$ of the interval $[x, 1 - x]$. We differentiate in t , using $dT/dt = 1$ and $dx/dt = -x^2$. The two endpoints of the integral contribute boundary terms, multiplied by $x^2 T^{1-\alpha-\zeta} = x T^{-\alpha-\zeta}$, and we arrive at

$$(g_\alpha * g_\zeta)'(t) = T^{-\alpha-\zeta} \left[(1 - \alpha - \zeta) \int_x^{1-x} y^{-\alpha} (1 - y)^{-\zeta} dy + x^{1-\alpha} (1 - x)^{-\zeta} + x^{1-\zeta} (1 - x)^{-\alpha} \right]. \quad (16.13)$$

Both boundary terms are positive.

Step 2: positivity of D_g . We take $\alpha = \zeta = a$ in (16.13). With $\omega(x) = [x(1 - x)]^{-a}$, the bracket in (16.13) is

$$J(x) = (1 - 2a) \int_x^{1-x} \omega(y) dy + 2x \omega(x).$$

Since $g(t)^2 = (T - 1)^{-2a} = T^{-2a} (1 - x)^{-2a}$, we have

$$\frac{(g * g)'(t)}{g(t)^2} = (1 - x)^{2a} J(x). \quad (16.14)$$

Using $\omega(1 - x) = \omega(x)$ and $\omega'(x) = -a(1 - 2x) \omega(x) / [x(1 - x)]$, a direct computation shows that $J'(x) = 2a \omega(x) / (1 - x)$, and then

$$\frac{d}{dx} [(1 - x)^{2a} J(x)] = 2a(1 - x)^{2a-1} L(x),$$

where

$$L(x) = (1 - 2x) \omega(x) - (1 - 2a) \int_x^{1-x} \omega(y) dy.$$

We have $L(1/2) = 0$, and

$$L'(x) = -\frac{a \omega(x)}{x(1 - x)} < 0, \quad 0 < x < \frac{1}{2},$$

where we used $(1 - 2x)^2 + 4x(1 - x) = 1$. It follows that $L > 0$ on $(0, 1/2)$, and that $(1 - x)^{2a} J(x)$ is strictly increasing on $(0, 1/2]$. As $x \downarrow 0$, we have $2x \omega(x) \rightarrow 0$ and $J(x) \rightarrow (1 - 2a)B(1 - a, 1 - a)$. As in (16.3), this limit equals $\Gamma(1 - a)^2 / \Gamma(1 - 2a) = \vartheta$. Every $t \geq 0$ corresponds to some $x \in (0, 1/2]$, and (16.14) implies that $(g * g)'(t) > \vartheta g(t)^2$, that is, $D_g(t) > 0$, for all $t \geq 0$, including $t = 0$.

Step 3: the limits. By (16.14), we have $D_g(t) = T^{-2a}[J(x) - \vartheta(1-x)^{-2a}]$. Using $\vartheta = (1-2a) \int_0^1 \omega$, we can write

$$J(x) - \vartheta(1-x)^{-2a} = -2(1-2a) \int_0^x \omega(y) dy + 2x\omega(x) - \vartheta[(1-x)^{-2a} - 1].$$

For $0 < y \leq x \leq 1/2$, we have $\omega(y) \leq 2^a y^{-a}$. Then the first two terms are bounded by Kx^{1-a} , and the last term is bounded by $Kx \leq Kx^{1-a}$. With $x = 1/T$ and $1+t \leq T \leq 2(1+t)$, this gives

$$|D_g(t)| \leq KT^{-2a}x^{1-a} = KT^{-1-a} \leq K(1+t)^{-1-a}.$$

This is the first estimate in (16.11). We also have $D_g = o(g^3)$, because $(1+t)^{-1-a} = g(t)^3(1+t)^{-(1-2a)}$ and $1-2a > 0$.

For the second limit, we take $(\alpha, \zeta) = (a, 2a)$ in (16.13), which is allowed since $2a < 1$. Using $g(t)^3 = T^{-3a}(1-x)^{-3a}$, we find

$$\begin{aligned} \frac{(g * g^2)'(t)}{g(t)^3} &= (1-x)^{3a} \left[(1-3a) \int_x^{1-x} y^{-a}(1-y)^{-2a} dy \right. \\ &\quad \left. + x^{1-a}(1-x)^{-2a} + x^{1-2a}(1-x)^{-a} \right]. \end{aligned}$$

As $t \rightarrow \infty$, we have $x \rightarrow 0$, the boundary terms tend to zero, and the integral tends to $B(1-a, 1-2a)$. This proves the convergence in (16.11). When $3a \neq 1$, the identity $\Gamma(2-3a) = (1-3a)\Gamma(1-3a)$ yields $(1-3a)B(1-a, 1-2a) = \Gamma(1-a)\Gamma(1-2a)/\Gamma(1-3a)$.

It remains to show that the limit is negative. If $3a \geq 1$, then $(1-3a)B(1-a, 1-2a) \leq 0 < \vartheta$. If $3a < 1$, consider

$$Y(\sigma) = \frac{\Gamma(1-a)\Gamma(1-\sigma)}{\Gamma(1-a-\sigma)}, \quad 0 < \sigma < 1-a.$$

Its logarithmic derivative is $\psi(1-a-\sigma) - \psi(1-\sigma) < 0$, and hence Y is strictly decreasing. We have $Y(a) = \vartheta$ by (16.2), and $2a < 1-a$. We conclude that $(1-3a)B(1-a, 1-2a) = Y(2a) < Y(a) = \vartheta$.

Step 4: derivative bounds. Let F and G be continuously differentiable on $[0, \infty)$. Splitting the convolution at $t/2$ and substituting $s \mapsto t-s$ on $[t/2, t]$, we have

$$(F * G)(t) = \int_0^{t/2} F(s)G(t-s) ds + \int_0^{t/2} G(s)F(t-s) ds.$$

Differentiating both integrals, we find the exact identity

$$\begin{aligned} (F * G)'(t) &= F(t/2)G(t/2) + \int_0^{t/2} F(s)G'(t-s) ds \\ &\quad + \int_0^{t/2} G(s)F'(t-s) ds, \end{aligned} \tag{16.15}$$

in which each of the two boundary terms contributed $F(t/2)G(t/2)/2$.

We next consider f_ε . We can write

$$f_\varepsilon = g[1 + \varepsilon(1 - g)], \quad f'_\varepsilon = g'[1 + \varepsilon(1 - 2g)],$$

and both brackets lie in $[3/4, 5/4]$ when $|\varepsilon| \leq 1/4$, because $0 < g \leq 1$. Recalling that $g'(t) = -a(1+t)^{-a-1} < 0$, we deduce that f_ε is positive and strictly decreasing, and that

$$\frac{3}{4}g \leq f_\varepsilon \leq \frac{5}{4}g, \quad \frac{3}{4}a(1+t)^{-a-1} \leq |f'_\varepsilon(t)| \leq \frac{5}{4}a(1+t)^{-a-1}. \quad (16.16)$$

Also $f_\varepsilon(0) = (1 + \varepsilon) - \varepsilon = 1$.

Fix $j \geq 2$, and apply (16.15) with $F = f_\varepsilon^j$ and $G = f_\varepsilon$. Using (16.16), we have $|F(t)| \leq K(1+t)^{-ja}$, $|F'(t)| = jf_\varepsilon^{j-1}|f'_\varepsilon| \leq K_j(1+t)^{-ja-1}$, $|G(t)| \leq K(1+t)^{-a}$, and $|G'(t)| \leq K(1+t)^{-a-1}$. For $0 \leq s \leq t/2$, we have $1+t-s \geq (1+t)/2$. Inserting these bounds into the three terms of (16.15), we obtain

$$\begin{aligned} |(f_\varepsilon^j * f_\varepsilon)'(t)| &\leq K_j \left[(1+t)^{-(j+1)a} + (1+t)^{-a-1} \int_0^{t/2} (1+s)^{-ja} ds \right. \\ &\quad \left. + (1+t)^{-ja-1} \int_0^{t/2} (1+s)^{-a} ds \right]. \end{aligned}$$

The last term is at most $K_j(1+t)^{-(j+1)a}$, since $\int_0^{t/2} (1+s)^{-a} ds \leq (1+t)^{1-a}/(1-a)$. The integral in the middle term is at most $(1+t)^{1-ja}/(1-ja)$ if $ja < 1$, at most $\log(2+t)$ if $ja = 1$, and at most $1/(ja-1)$ if $ja > 1$. Accordingly, the middle term is at most $K_j(1+t)^{-(j+1)a}$, $K(1+t)^{-a-1} \log(2+t)$, or $K_j(1+t)^{-a-1}$. Each of these bounds is at most $K_j(1+t)^{-3a} = K_j g(t)^3$: the first because $j \geq 2$, and the other two because $a+1 > 3a$ and $(1+t)^{-(1-2a)} \log(2+t)$ is bounded. This proves (16.12).

Finally, we take $(\alpha, \zeta) = (2a, 2a)$ in (16.13), which is allowed since $2a < 1$. The integral in the bracket is at most $B(1-2a, 1-2a)$, and the boundary terms are bounded on $(0, 1/2]$. Then $|(g^2 * g^2)'(t)| \leq KT^{-4a} \leq K(1+t)^{-3a}$. Together with $g^4 \leq g^3$, this shows that $|D_2| \leq Kg^3$. $\square$

We now correct g by a multiple of g^2 . The following quantity plays the role of the residual of a profile $hf(\cdot/\tau)$ attached at a time where u has fallen to the value h ; its precise relation to the residual $\mathcal{R}_p$ appears in (16.24) below. For a continuously differentiable function f on $[0, \infty)$ and $h \in (0, 1]$, we write $j_f = -f'$ and define the ideal residual

$$I_h(f) = hf - (1-h)M(hf) - h[M(hf) * j_f]. \quad (16.17)$$

Lemma 16.6 (Corrected critical profiles). *There exist constants $\kappa, \eta, h_0 > 0$, depending only on p and q , with $h_0 \leq 1$ and $\kappa h_0 \leq 1/4$, such that the following holds for every $h \in (0, h_0]$. The profiles*

$$f_+ = (1 + \kappa h)g - \kappa h g^2, \quad f_- = (1 - \kappa h)g + \kappa h g^2$$

are positive and strictly decreasing, with $f_{\pm}(0) = 1$. Further, $I_h(f_+) \geq \eta h^3 g^3$ on $[0, \infty)$, and there exists $X_h < \infty$ such that $I_h(f_-) \leq -\eta h^3 g^3$ on $[X_h, \infty)$.

Proof. The profiles are $f_{\pm} = f_{\pm \kappa h}$. Once $\kappa h_0 \leq 1/4$, the first assertion follows from Lemma 16.5. We begin by expanding $I_h(f)$ for a continuously differentiable f with $f(0) = 1$. Differentiating $(f^j * f)(t) = \int_0^t f(s)^j f(t-s) ds$ and using $f(0) = 1$, we have

$$(f^j * f)' = f^j - f^j * j_f, \quad j \geq 1.$$

Inserting $M(y) = \sum_{j=1}^{p-1} \lambda_j y^j$, with $\lambda_1 = 1$, into (16.17) and using this identity, we obtain

$$I_h(f) = hf - \sum_{j=1}^{p-1} \lambda_j h^j f^j + \sum_{j=1}^{p-1} \lambda_j h^{j+1} (f^j * f)'.$$

The term hf cancels against the term $j = 1$ of the first sum. Separating the terms with $j = 1$ and $j = 2$ from the rest, we have

$$\begin{aligned} I_h(f) &= h^2[(f * f)' - \vartheta f^2] + N_h(f), \\ N_h(f) &= \vartheta h^3(f^2 * f)' + \sum_{j=3}^{p-1} \lambda_j [-h^j f^j + h^{j+1} (f^j * f)']. \end{aligned} \quad (16.18)$$

We now bound $N_h(f_\varepsilon)$ for $|\varepsilon| \leq 1/4$ and $0 < h \leq 1$. By (16.12), we have $h^{j+1} |(f_\varepsilon^j * f_\varepsilon)'| \leq K_j h^3 g^3$ for $j \geq 2$. Using (16.16), we have $h^j f_\varepsilon^j \leq (5/4)^j h^3 g^3$ for $j \geq 3$, where we used $h \leq 1$ and $g \leq 1$. Then there exists K_0 such that

$$|N_h(f_\varepsilon)| \leq K_0 h^3 g^3, \quad |\varepsilon| \leq 1/4, \quad 0 < h \leq 1. \quad (16.19)$$

Next, expanding $f_\varepsilon * f_\varepsilon$ and f_ε^2 by bilinearity, we find the exact quadratic expansion

$$(f_\varepsilon * f_\varepsilon)' - \vartheta f_\varepsilon^2 = (1 + \varepsilon)^2 D_g - 2\varepsilon(1 + \varepsilon)H + \varepsilon^2 D_2. \quad (16.20)$$

We now fix the constants. Let

$$\Delta = \vartheta - (1 - 3a)B(1 - a, 1 - 2a),$$

which is positive by Lemma 16.5. By (16.11), there exists $t_* < \infty$ such that $H(t) \leq -(\Delta/2)g(t)^3$ for $t \geq t_*$. There also exist K_D and K_H such that $|D_2| \leq K_D g^3$ and $|H| \leq K_H g^3$ on $[0, \infty)$, by Lemma 16.5; for H we used that H/g^3 is continuous and has a finite limit. Let $\mu_* = \min_{[0, t_*]} D_g/g^3$, which is positive by Lemma 16.5. We set

$$\kappa = 1 + \frac{4K_0}{\Delta},$$

so that $K_0 \leq \kappa\Delta/4$. Let $h_0 \in (0, 1]$ be so small that

$$\kappa h_0 \leq \frac{1}{4}, \quad \kappa^2 K_D h_0 \leq \frac{\kappa\Delta}{4}, \quad \left(\frac{5}{2}\kappa K_H + \kappa^2 K_D + K_0\right)h_0 \leq \frac{\mu_*}{2},$$

and we set $\eta = \min\{\kappa\Delta/8, \mu_*/2\}$. Fix $h \in (0, h_0]$.

For the upper profile, we take $\varepsilon = \kappa h \in (0, 1/4]$ in (16.20). On $[t_*, \infty)$, we have $(1 + \varepsilon)^2 D_g \geq 0$, $-2\varepsilon(1 + \varepsilon)H \geq \varepsilon\Delta g^3$, and $\varepsilon^2 D_2 \geq -\kappa^2 h^2 K_D g^3$. Together with (16.18) and (16.19), these bounds imply that

$$I_h(f_+) \geq h^3 [\kappa\Delta - \kappa^2 h K_D - K_0] g^3 \geq \frac{\kappa\Delta}{2} h^3 g^3 \geq \eta h^3 g^3 \quad \text{on } [t_*, \infty),$$

where we used $\kappa^2 h K_D \leq \kappa\Delta/4$ and $K_0 \leq \kappa\Delta/4$. On $[0, t_*]$, we divide (16.18) by $h^2 g^3$ and use $(1 + \varepsilon)^2 \geq 1$ and $|2\varepsilon(1 + \varepsilon)| \leq 5\kappa h/2$. This yields

$$\frac{I_h(f_+)}{h^2 g^3} \geq \mu_* - \left(\frac{5}{2}\kappa K_H + \kappa^2 K_D + K_0\right)h \geq \frac{\mu_*}{2},$$

that is, $I_h(f_+) \geq (\mu_*/2)h^2 g^3 \geq \eta h^3 g^3$ on $[0, t_*]$, since $h \leq 1$.

For the lower profile, we take $\varepsilon = -\kappa h$. Then $1 - \kappa h \geq 3/4$, and on $[t_*, \infty)$ the cross term satisfies $2\kappa h(1 - \kappa h)H \leq -(3/4)\kappa h\Delta g^3$. Using also $(1 - \kappa h)^2 D_g \leq D_g$, $\varepsilon^2 D_2 \leq \kappa^2 h^2 K_D g^3$, and (16.19), we find

$$I_h(f_-) \leq h^2 D_g + h^3 \left[-\frac{3}{4}\kappa\Delta + \kappa^2 h K_D + K_0\right] g^3 \leq h^2 D_g - \frac{\kappa\Delta}{4} h^3 g^3 \quad \text{on } [t_*, \infty).$$

The ratio D_g/g^3 tends to zero by (16.11). Then there exists $X_h \geq t_*$ such that $D_g \leq (\kappa\Delta/8)h g^3$ on $[X_h, \infty)$, and there $I_h(f_-) \leq -(\kappa\Delta/8)h^3 g^3 \leq -\eta h^3 g^3$. This completes the proof. $\square$

The cutoff X_h depends on h ; since $D_g/g^3 \leq K(1+t)^{-(1-2a)}$, it can be taken of order $h^{-1/(1-2a)}$. In the next proof, h is fixed before X_h is used.

Proposition 16.7 (Critical endpoint bounds). *There exist constants $0 < A_- \leq A_+ < \infty$ such that*

$$A_-(1+t)^{-a} \leq u(t) \leq A_+(1+t)^{-a}, \quad t \geq 0. \quad (16.21)$$

For every integer $n \geq 1$, there exist constants $0 < \underline{K}_n \leq \bar{K}_n < \infty$ and $t_n < \infty$ such that

$$\underline{K}_n(1+t)^{-a-n} \leq (-1)^n u^{(n)}(t) \leq \bar{K}_n(1+t)^{-a-n}, \quad t \geq t_n.$$

Proof. The proof has three steps. First, we construct a profile P_+ with nonnegative residual. It equals u up to a time T_0 , is constant up to $T_1 = 2T_0$, and then follows a copy of f_+ on a long time scale τ ; we choose the time scale large enough to absorb the error caused by the memory of the initial segment of u . Second, we construct a profile P_-

with nonpositive residual by attaching a copy of f_- directly at T_0 , on a short time scale. In both cases, we then obtain a bound for u from Lemma 16.4. Third, we deduce the derivative bounds from complete monotonicity.

We will use the following consequence of Lemma 16.5. Let $f = f_{\pm}$ be one of the profiles of Lemma 16.6. For $0 \leq \delta \leq 1$ and $x \geq 0$, every $y \geq (x - \delta)_+$ satisfies $1 + y \geq (1 + x)/2$. By the mean value theorem and the bound on $|f'_\varepsilon|$ in Lemma 16.5, we have

$$0 \leq f((x - \delta)_+) - f(x) \leq K\delta(1 + x)^{-a-1}, \quad 0 \leq \delta \leq 1, x \geq 0. \quad (16.22)$$

We also use the convolution bound

$$\int_0^x (1 + z)^{-a-1} (1 + x - z)^{-a-1} dz \leq 2 \left(1 + \frac{x}{2}\right)^{-a-1} \int_0^\infty (1 + z)^{-a-1} dz \leq K(1 + x)^{-a-1}, \quad (16.23)$$

which follows from the symmetry $z \mapsto x - z$ and splitting the integral at $x/2$.

Step 1: the upper bound. Let κ, η, h_0 be as in Lemma 16.6. Since u decreases to zero, there exists $T_0 \geq 1$ with $h = u(T_0) \leq h_0$. We fix such a T_0 ; note that $h > 0$. Let $T_1 = 2T_0$, and let $\tau \geq T_1$ be a parameter, to be fixed below. We write $f = f_+$ and $j = -f'$, and define

$$P_+(t) = \begin{cases} u(t), & 0 \leq t \leq T_0, \\ h, & T_0 \leq t \leq T_1, \\ hf((t - T_1)/\tau), & t \geq T_1. \end{cases}$$

This function is continuous, since $u(T_0) = h = hf(0)$. It is locally absolutely continuous, nonincreasing, takes values in $[0, 1]$, and satisfies $P_+(0) = 1$. We check that $\mathcal{R}_{P_+} \geq 0$ on each of the three intervals.

On $(0, T_0)$, we have $\mathcal{R}_{P_+} = \mathcal{R}_u = 0$, since the residual at time t involves only the values of P_+ on $[0, t]$.

On (T_0, T_1) , we have $P'_+(t) = 0$ and $P_+(t) = h$, and $-P'_+$ vanishes on (T_0, t) . For $0 \leq s \leq T_0$, we have $t - s \geq T_0 - s$, which implies $P_+(t - s) \leq P_+(T_0 - s) = u(T_0 - s)$. Since M is increasing on $[0, 1]$, we have

$$\mathcal{R}_{P_+}(t) = h - \int_0^{T_0} M(P_+(t - s)) v(s) ds \geq h - (M(u) * v)(T_0) = v(T_0) > 0.$$

In the equality on the right, we used (16.6) at time T_0 , in the form $(M(u) * v)(T_0) = u(T_0) + u'(T_0) = h - v(T_0)$.

On (T_1, ∞) , we write $x = (t - T_1)/\tau$. Then $P_+(t) = hf(x)$ and $P'_+(t) = (h/\tau)f'(x)$. The measure $-P'_+(s) ds$ equals $v(s) ds$ on $[0, T_0]$, vanishes on (T_0, T_1) , and equals $(h/\tau)j((s - T_1)/\tau) ds$ on (T_1, ∞) . For $0 \leq s \leq T_0$, we have $t - s \geq T_1 - T_0 = T_0$ and $P_+(t - s) = hf((x - s/\tau)_+)$: this is the definition of P_+ if $t - s \geq T_1$, and if $T_0 \leq t - s < T_1$ both sides equal h . In the integral over (T_1, t) , we substitute $s = T_1 + \tau(x - z)$, for which

$t - s = \tau z$. Recalling that $\int_0^{T_0} v = 1 - u(T_0) = 1 - h$ and comparing with (16.17), we obtain the exact decomposition

$$\mathcal{R}_{P_+}(t) = \frac{h}{\tau} f'(x) + I_h(f)(x) - E_1(x) - E_2(x), \quad (16.24)$$

where

$$\begin{aligned} E_1(x) &= \int_0^{T_0} \left[M(hf((x - s/\tau)_+)) - M(hf(x)) \right] v(s) \, ds, \\ E_2(x) &= h \int_0^x [M(P_+(\tau z)) - M(hf(z))] j(x - z) \, dz. \end{aligned}$$

The terms E_1 and E_2 measure the difference between the actual memory of P_+ , which contains the initial segment of u , and the ideal memory in $I_h(f)$. Both are nonnegative. For E_1 , this holds because f is decreasing and M is increasing. For E_2 , we have $P_+(\tau z) \geq hf(z)$ for all $z \geq 0$: if $\tau z \leq T_1$, then $P_+(\tau z) \geq h \geq hf(z)$, and if $\tau z \geq T_1$, then $P_+(\tau z) = hf(z - T_1/\tau) \geq hf(z)$.

We now bound E_1 and E_2 from above, keeping the whole initial segment of u . For $s \leq T_0$, we have $s/\tau \leq T_0/\tau \leq 1$. By (16.22) with $\delta = s/\tau$, the Lipschitz bound for M on $[0, 1]$, and $\int_0^{T_0} v \leq 1$, we have

$$0 \leq E_1(x) \leq L_M K h \frac{T_0}{\tau} (1 + x)^{-a-1}.$$

For E_2 , we split the integral at $z = T_0/\tau$. On $0 \leq z \leq \min\{x, T_0/\tau\}$, the difference in brackets is at most $M(1)$, and $j(x - z) \leq K(1 + x - z)^{-a-1} \leq K(1 + x)^{-a-1}$, since $z \leq 1$. This part of E_2 is at most $K h M(1)(T_0/\tau)(1 + x)^{-a-1}$. On $T_0/\tau \leq z \leq x$, we have $\tau z \geq T_0$ and $P_+(\tau z) = hf((z - T_1/\tau)_+)$. By (16.22) with $\delta = T_1/\tau \leq 1$, the difference in brackets is at most $L_M K h (T_1/\tau)(1 + z)^{-a-1}$. Using also $j(x - z) \leq K(1 + x - z)^{-a-1}$ and (16.23), we find that this part of E_2 is at most $K h^2 (T_1/\tau)(1 + x)^{-a-1}$. Finally, $(h/\tau)f'(x) \geq -K(h/\tau)(1 + x)^{-a-1}$.

We insert these bounds and $I_h(f_+) \geq \eta h^3 g^3$ into (16.24). Using also $(1 + x)^{-a-1} \leq g(x)^3$, which holds because $a + 1 > 3a$, we conclude that

$$\mathcal{R}_{P_+}(t) \geq \left[\eta h^3 - \frac{K_1 h}{\tau} \right] g(x)^3, \quad t > T_1,$$

where K_1 depends on p, q , and T_1 , but not on $\tau \geq T_1$. We now fix

$$\tau = \max \left\{ T_1, \frac{2K_1}{\eta h^2} \right\}.$$

Then $\mathcal{R}_{P_+} > 0$ on (T_1, ∞) . We have shown that $\mathcal{R}_{P_+} \geq 0$ almost everywhere, and Lemma 16.4 implies that $u \leq P_+$. For $t \geq T_1$, we have $\tau + t - T_1 \geq t \geq (1 + t)/2$, since $\tau \geq T_1$ and $t \geq 1$. Using (16.16), we find

$$u(t) \leq hf_+\left(\frac{t - T_1}{\tau}\right) \leq \frac{5}{4} h \left(\frac{\tau + t - T_1}{\tau}\right)^{-a} \leq \frac{5}{4} h (2\tau)^a (1 + t)^{-a}, \quad t \geq T_1.$$

For $t \leq T_1$, we have $u(t) \leq 1 \leq (1 + T_1)^a(1 + t)^{-a}$. This proves the upper bound in (16.21).

Step 2: the lower bound. We now write $f = f_-$ and $j = -f'$. We fix T_0 with $h = u(T_0) \leq h_0$ and the cutoff X_h of Lemma 16.6. For a parameter $\tau > 0$, to be fixed below, we define

$$P_-(t) = \begin{cases} u(t), & 0 \leq t \leq T_0, \\ hf((t - T_0)/\tau), & t \geq T_0. \end{cases}$$

This function is continuous, locally absolutely continuous, nonincreasing, takes values in $[0, 1]$, and satisfies $P_-(0) = 1$. As before, $\mathcal{R}_{P_-} = 0$ on $(0, T_0)$.

For $t > T_0$, let $x = (t - T_0)/\tau$. Splitting the memory integral of $\mathcal{R}_{P_-}(t)$ at T_0 and substituting $s = T_0 + \tau(x - z)$ in the second part, we have

$$\mathcal{R}_{P_-}(t) = \frac{h}{\tau}f'(x) + hf(x) - \int_0^{T_0} M(P_-(t-s))v(s) ds - h \int_0^x M(P_-(\tau z))j(x-z) dz.$$

The actual values of P_- in the memory integrals are at least their ideal replacements. For $0 \leq s \leq T_0$, we have $P_-(t-s) \geq P_-(t) = hf(x)$, since P_- is nonincreasing. For $z \geq 0$, we have $P_-(\tau z) \geq hf(z)$: if $\tau z \leq T_0$, then $P_-(\tau z) = u(\tau z) \geq u(T_0) = h \geq hf(z)$, and if $\tau z \geq T_0$, then $P_-(\tau z) = hf(z - T_0/\tau) \geq hf(z)$. The function M is increasing, and the memory terms enter with a negative sign. Recalling also that $\int_0^{T_0} v = 1 - h$, we conclude that

$$\mathcal{R}_{P_-}(t) \leq \frac{h}{\tau}f'(x) + I_h(f)(x), \quad t > T_0.$$

In contrast with the upper bound, this inequality has no error term.

It remains to fix τ so that the right side is nonpositive for all $x \geq 0$. For $x \geq X_h$, both terms are nonpositive by Lemma 16.6. On $[0, X_h]$, the function $I_h(f)$ is continuous, and $L_h = \max_{[0, X_h]} I_h(f)_+$ is finite. By (16.16), we have $-f'(x) \geq \delta_h = (3/4)a(1 + X_h)^{-a-1} > 0$ there. Let

$$\tau = \frac{h\delta_h}{L_h + 1}.$$

Then $(h/\tau)f'(x) \leq -(L_h + 1) < -I_h(f)(x)$ for $0 \leq x \leq X_h$. We have shown that $\mathcal{R}_{P_-} \leq 0$ almost everywhere; the single jump of P'_- at T_0 does not affect the argument. By Lemma 16.4, we have $P_- \leq u$. For $t \geq T_0$, using (16.16) and $1 + (t - T_0)/\tau \leq 1 + t/\tau \leq \max\{1, 1/\tau\}(1 + t)$, we find

$$u(t) \geq hf_-\left(\frac{t - T_0}{\tau}\right) \geq \frac{3}{4}h\left(1 + \frac{t - T_0}{\tau}\right)^{-a} \geq \frac{3}{4}h \min\{1, \tau\}^a(1 + t)^{-a}.$$

For $t \leq T_0$, we have $u(t) \geq u(T_0) = h \geq h(1 + t)^{-a}$. This proves the lower bound in (16.21).

Step 3: the derivatives. For $n \geq 0$ and $t > 0$, let $F_n(t) = (-1)^n u^{(n)}(t)$. By Lemma 16.3, each F_n is positive. Since $F'_n = -F_{n+1} \leq 0$, each F_n is nonincreasing. Using (16.21), we have $\underline{k}_0 t^{-a} \leq F_0(t) \leq \bar{k}_0 t^{-a}$ for $t \geq 1$, with $\underline{k}_0 = 2^{-a} A_-$ and $\bar{k}_0 = A_+$.

Suppose that, for some $n \geq 0$ and $t'_n \geq 1$,

$$\underline{k}_n t^{-\alpha} \leq F_n(t) \leq \bar{k}_n t^{-\alpha}, \quad t \geq t'_n, \quad \alpha = a + n.$$

Using that F_{n+1} is nonincreasing, we have, for $t \geq 2t'_n$,

$$\frac{t}{2} F_{n+1}(t) \leq \int_{t/2}^t F_{n+1}(s) \, ds = F_n(t/2) - F_n(t) \leq F_n(t/2) \leq 2^\alpha \bar{k}_n t^{-\alpha},$$

that is, $F_{n+1}(t) \leq 2^{1+\alpha} \bar{k}_n t^{-\alpha-1}$. For the lower bound, let $\Lambda > 1$ be such that $\bar{k}_n \Lambda^{-\alpha} \leq \underline{k}_n/2$. For $t \geq t'_n$, we have

$$\frac{\underline{k}_n}{2} t^{-\alpha} \leq F_n(t) - F_n(\Lambda t) = \int_t^{\Lambda t} F_{n+1}(s) \, ds \leq (\Lambda - 1) t F_{n+1}(t),$$

that is, $F_{n+1}(t) \geq \underline{k}_n t^{-\alpha-1} / (2(\Lambda - 1))$. The same two-sided bound therefore holds for F_{n+1} , with α replaced by $\alpha + 1$ and $t'_{n+1} = 2t'_n$. By induction, it holds for all $n \geq 0$. This proves the derivative bounds, because $(1+t)/2 \leq t \leq 1+t$ for $t \geq 1$. $\square$

The function v is nonincreasing with $v(0) = 1$. Together with the case $n = 1$ of Proposition 16.7, this implies that, for a constant K depending on u ,

$$0 < v(t) \leq K(1+t)^{-a-1}, \quad t \geq 0. \quad (16.25)$$

16.4 The quadratic balance and its logarithmic Dirichlet form

The bounds (16.21) identify the exponent, but they do not show that $t^a u(t)$ has a limit. The goal of this section is to prove that $t^a u(t)$ varies more and more slowly on a logarithmic time scale. We first isolate the quadratic balance behind (16.3).

Lemma 16.8 (Integrable critical balance error). *Let $Q = u * u$ and $\rho = 1 - 2a$. Then the function $\mathcal{E} = Q' - \vartheta u^2$ satisfies*

$$Q' = \vartheta u^2 + \mathcal{E}, \quad |\mathcal{E}(t)| \leq K(1+t)^{-3a}, \quad \int_0^\infty |\mathcal{E}(t)| t^{-\rho} \, dt < \infty, \quad (16.26)$$

where K depends on u .

Proof. Recall that $v = -u'$. Differentiating $Q = u * u$ and using $u(0) = 1$, we have $Q' = u - u * v$. By (16.6), we have $u = v + M(u) * v = v + u * v + \sum_{j=2}^{p-1} \lambda_j u^j * v$. Inserting this into the expression for Q' , we find

$$Q' = v + \sum_{j=2}^{p-1} \lambda_j u^j * v.$$

In the same way, we have $(u^j * u)' = u^j - u^j * v$. Inserting this identity and separating the term $\lambda_2 u^2 = \vartheta u^2$, we obtain $Q' = \vartheta u^2 + \mathcal{E}$ with the exact error

$$\mathcal{E} = v + \sum_{j=3}^{p-1} \lambda_j u^j - \sum_{j=2}^{p-1} \lambda_j (u^j * u)'. \quad (16.27)$$

For $p = 3$, the first sum is empty.

We bound each term. Using (16.21) and (16.25), we have $u(t) \leq K(1+t)^{-a}$ and $v(t) \leq K(1+t)^{-a-1}$. The term v is at most $K(1+t)^{-3a}$, since $a+1 > 3a$, and so is u^j for $j \geq 3$. For $j \geq 2$, we apply (16.15) with $F = u^j$ and $G = u$, which are continuously differentiable. We have $|F| \leq K(1+t)^{-ja}$, $|F'| = ju^{j-1}v \leq K_j(1+t)^{-ja-1}$, $|G| \leq K(1+t)^{-a}$, and $|G'| = v \leq K(1+t)^{-a-1}$. These are the bounds used in Step 4 of the proof of Lemma 16.5, and the computation there yields $|(u^j * u)'(t)| \leq K_j(1+t)^{-3a}$, including the case $ja = 1$. This proves the pointwise bound in (16.26).

The error is bounded near zero and $\rho < 1$, which implies that $\int_0^1 |\mathcal{E}(t)| t^{-\rho} dt < \infty$. On $[1, \infty)$, we use $3a + \rho = 1 + a$ and find $\int_1^\infty |\mathcal{E}(t)| t^{-\rho} dt \leq K \int_1^\infty t^{-1-a} dt < \infty$. $\square$

We now pass to logarithmic time. In the following lemma, we convert the balance (16.26) into a finite Dirichlet energy for the function $e^{ax}u(e^x)$, which equals $t^a u(t)$ at $t = e^x$.

Lemma 16.9 (Logarithmic Dirichlet bound). *The function $D(x) = e^{ax}u(e^x)$, $x \in \mathbb{R}$, is bounded and Lipschitz on $\mathbb{R}$, it is bounded below by a positive constant on $[0, \infty)$, and*

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{[D(x) - D(y)]^2}{[2 \cosh((x-y)/2)]^{2-2a}} dx dy < \infty. \quad (16.28)$$

In particular, for every fixed $h \in \mathbb{R}$,

$$D(x+h) - D(x) \longrightarrow 0 \quad (x \rightarrow \infty), \quad (16.29)$$

and the convergence is uniform when h ranges over a compact interval.

Proof. The proof has four steps. We first record the elementary properties of D . We then integrate the balance (16.26) against $t^{-\rho}$ up to a finite time T and obtain a quantity J_T that is bounded above uniformly in T . Next, we use the identity behind the Schur test to bound a truncated version of the energy in (16.28) by J_T . Finally, we deduce (16.29). We keep the cutoff T finite until the last moment, because the separate integrals that appear in the second and third steps diverge as $T \rightarrow \infty$.

Step 1: properties of D . By (16.21), we have

$$A_- \left(\frac{e^x}{1+e^x} \right)^a \leq D(x) \leq A_+ \left(\frac{e^x}{1+e^x} \right)^a, \quad x \in \mathbb{R}.$$

In particular, $0 < D \leq A_+$, $D(x) \geq 2^{-a}A_-$ for $x \geq 0$, and $D(x) \rightarrow 0$ as $x \rightarrow -\infty$. The derivative of D is

$$D'(x) = ae^{ax}u(e^x) - e^{(a+1)x}v(e^x).$$

For $x \geq 0$, the bounds (16.21) and (16.25) show that both terms are bounded. For $x \leq 0$, both terms are bounded by 1, since $u \leq 1$ and $v \leq 1$. We conclude that D is Lipschitz on $\mathbb{R}$. Let L_D denote a Lipschitz constant.

Step 2: the balance in integrated form. Let

$$C_\rho = B\left(\frac{1+\rho}{2}, \frac{1+\rho}{2}\right) = B(1-a, 1-a),$$

so that $\vartheta = \rho C_\rho$ by (16.3). For $T > 0$, we define

$$I_T = \int_0^T Q(t)t^{-1-\rho} dt, \quad J_T = C_\rho \int_0^T u(t)^2 t^{-\rho} dt - I_T.$$

Note that $0 \leq Q(t) \leq t$, because $0 \leq u \leq 1$. Then both integrals are finite, and $Q(t)t^{-\rho} \rightarrow 0$ as $t \rightarrow 0$. The function Q is continuously differentiable. Integrating by parts, we find

$$\int_0^T Q'(t)t^{-\rho} dt = Q(T)T^{-\rho} + \rho I_T.$$

We insert (16.26) on the left side, use $\vartheta = \rho C_\rho$, and divide by ρ . We arrive at the exact identity

$$J_T = \frac{Q(T)T^{-\rho}}{\rho} - \frac{1}{\rho} \int_0^T \mathcal{E}(t)t^{-\rho} dt. \quad (16.30)$$

Using (16.21), we have

$$Q(T) \leq A_+^2 \int_0^T (1+T-s)^{-a}(1+s)^{-a} ds \leq A_+^2 \int_0^T (T-s)^{-a}s^{-a} ds = A_+^2 C_\rho T^\rho.$$

Together with Lemma 16.8, this shows that

$$\sup_{T>0} J_T \leq \frac{A_+^2 C_\rho}{\rho} + \frac{1}{\rho} \int_0^\infty |\mathcal{E}(t)|t^{-\rho} dt < \infty.$$

Step 3: the weighted Schur identity. We define the symmetric kernel

$$\Xi(s, t) = \frac{(st)^{-a}}{(s+t)^{2-2a}}, \quad s, t > 0.$$

Substituting $t = s\varsigma$ and using the beta integral $\int_0^\infty \varsigma^{-a}(1+\varsigma)^{2a-2} d\varsigma = B(1-a, 1-a)$, we find, for $s > 0$,

$$\int_0^\infty \Xi(s, t) dt = \frac{C_\rho}{s}. \quad (16.31)$$

Let $u_T = u\mathbf{1}_{[0,T]}$ and $\Phi_T(s) = s^a u_T(s)$. Using (16.31), $2a - 1 = -\rho$, and $2 - 2a = 1 + \rho$, we have

$$\int_0^\infty \int_0^\infty \Xi(s, t) \Phi_T(s)^2 \, ds \, dt = C_\rho \int_0^T u(s)^2 s^{-\rho} \, ds, \quad \Xi(s, t) \Phi_T(s) \Phi_T(t) = \frac{u_T(s) u_T(t)}{(s + t)^{1+\rho}}.$$

The first quantity is finite, since $u \leq 1$ and $\rho < 1$. By the Cauchy–Schwarz inequality with respect to the measure $\Xi(s, t) \, ds \, dt$, the function $\Xi(s, t) |\Phi_T(s) \Phi_T(t)|$ is integrable as well. Expanding the square and using the symmetry of Ξ , we obtain

$$\begin{aligned} & \frac{1}{2} \int_0^\infty \int_0^\infty \Xi(s, t) [\Phi_T(s) - \Phi_T(t)]^2 \, ds \, dt \\ &= C_\rho \int_0^T u(s)^2 s^{-\rho} \, ds - \int_0^T \int_0^T \frac{u(s) u(t)}{(s + t)^{1+\rho}} \, ds \, dt \leq J_T. \end{aligned} \quad (16.32)$$

For the inequality, we used the representation of I_T obtained by writing out Q and applying Tonelli's theorem,

$$I_T = \iint_{\{s, t > 0, s+t \leq T\}} \frac{u(s) u(t)}{(s + t)^{1+\rho}} \, ds \, dt,$$

together with the facts that the triangle $\{s + t \leq T\}$ is contained in the square $[0, T]^2$ and that the integrand is nonnegative. The identity in (16.32) is the sum-of-squares identity behind the Schur test [111, Section 7, Equation (1), pp. 23–24] for the kernel $(st)^{\rho/2} (s + t)^{-1-\rho}$ on $L^2(0, \infty)$ with the test function $s^{-1/2}$, written in the variables Φ_T .

As $T \rightarrow \infty$, the integrand on the left of (16.32) converges pointwise to $\Xi(s, t) [s^a u(s) - t^a u(t)]^2$; indeed, it equals this limit as soon as $T > \max\{s, t\}$. Fatou's lemma and Step 2 then imply that

$$\int_0^\infty \int_0^\infty \frac{(st)^{-a}}{(s + t)^{2-2a}} [s^a u(s) - t^a u(t)]^2 \, ds \, dt \leq 2 \sup_{T>0} J_T < \infty.$$

We now substitute $s = e^x$ and $t = e^y$. Using $e^x + e^y = e^{(x+y)/2} 2 \cosh((x-y)/2)$, we have

$$\begin{aligned} \Xi(e^x, e^y) e^{x+y} &= e^{-a(x+y)} e^{(a-1)(x+y)} [2 \cosh((x-y)/2)]^{2a-2} e^{x+y} \\ &= [2 \cosh((x-y)/2)]^{-(2-2a)}, \end{aligned}$$

and $s^a u(s) = D(x)$. This proves (16.28).

Step 4: vanishing increments. Suppose that (16.29) fails for some $h \in \mathbb{R}$. Then there exist $\epsilon > 0$ and a sequence $x_k \rightarrow \infty$ with $|D(x_k + h) - D(x_k)| \geq \epsilon$. Let $\delta = \epsilon / (6L_D)$, and consider the regions

$$\mathcal{S}_k = \{(x, y) : |x - x_k| \leq \delta, |y - x - h| \leq \delta\}.$$

For $(x, y) \in \mathcal{S}_k$, we have $|x - x_k| \leq \delta$ and $|y - (x_k + h)| \leq 2\delta$. Using the Lipschitz bound, we find

$$|D(y) - D(x)| \geq |D(x_k + h) - D(x_k)| - 3L_D\delta \geq \epsilon/2.$$

Further, since $|x - y| \leq |h| + \delta$ on $\mathcal{S}_k$, the kernel in (16.28) is at least $[2 \cosh((|h| + \delta)/2)]^{-(2-2a)} > 0$ there. Passing to a subsequence, we may assume that the intervals $[x_k - \delta, x_k + \delta]$ are disjoint; then the regions $\mathcal{S}_k$ are disjoint. Each of them has area $4\delta^2$ and contributes at least a fixed positive amount to the integral in (16.28). This contradicts the finiteness of the integral, and proves (16.29).

For the uniformity, let I be a compact interval and $\epsilon > 0$. There exist finitely many points $h_1, \dots, h_N \in I$ such that every $h \in I$ satisfies $|h - h_i| \leq \epsilon/(2L_D)$ for some i . Then $|D(x + h) - D(x)| \leq |D(x + h_i) - D(x)| + \epsilon/2$, and the claim follows from (16.29) for the finitely many h_i . $\square$

16.5 An integrable-source renewal equation

The vanishing of the increments in (16.29) does not by itself imply that D converges. In this section, we show that a smoothed version of D^2 satisfies a renewal equation with an integrable source and use the renewal theorem to prove that it converges.

The renewal theorem. We use the renewal theorem in the following form. A probability measure F on $(0, \infty)$ is called arithmetic if it is concentrated on a set of the form $\{h, 2h, 3h, \dots\}$ for some $h > 0$. A function $z: [0, \infty) \rightarrow [0, \infty)$ is called directly Riemann integrable if, for every $h > 0$, the upper and lower sums

$$\bar{\sigma}(h) = h \sum_{n \geq 1} \sup_{[(n-1)h, nh]} z, \quad \underline{\sigma}(h) = h \sum_{n \geq 1} \inf_{[(n-1)h, nh]} z$$

are finite, and they tend to the same limit as $h \rightarrow 0$. This limit is then $\int_0^\infty z$.

Theorem 16.10 (Key renewal theorem; [68, Chapter XI, Section 1, pp. 362–363]). *Let F be a probability measure on $(0, \infty)$ which is not arithmetic and has finite mean m_F , and let F^{*n} denote its n -fold convolution, with $F^{*0} = \delta_0$. Let $z: [0, \infty) \rightarrow [0, \infty)$ be Borel measurable and directly Riemann integrable, and define*

$$Z(x) = \sum_{n \geq 0} \int_{[0, x]} z(x - y) F^{*n}(\mathrm{d}y), \quad x \geq 0.$$

Then $Z(x) \rightarrow m_F^{-1} \int_0^\infty z(y) \mathrm{d}y$ as $x \rightarrow \infty$.

We extend the test z by zero to negative arguments. Borel measurability on $[0, \infty)$ is equivalent to Borel measurability of this extension and makes the integrals well-defined even for singular increment laws. The Darboux-sum condition alone does not imply

measurability for such laws. The application below uses a continuous nonnegative source.

In [68, Chapter XI, Section 1, pp. 362–363], the function Z is the solution of the renewal equation $Z(x) = z(x) + \int_{[0,x]} Z(x-y) F(dy)$, written as $Z = U \star z$ with $U = \sum_{n \geq 0} F^{*n}$, and the theorem is stated there as the alternative form of the renewal theorem. We check direct Riemann integrability through the following elementary criterion, which we prove for completeness.

Lemma 16.11 (A criterion for direct Riemann integrability). *Let $z: [0, \infty) \rightarrow [0, \infty)$ be continuous, and suppose that $\sum_{k \geq 0} s_k < \infty$, where $s_k = \sup_{[k, k+1]} z$. Then z is directly Riemann integrable.*

Proof. Let $S = \sum_k s_k$. For $0 < h \leq 1$ and $n \geq 1$, the interval $[(n-1)h, nh]$ lies in $[k, k+2]$ with $k = \lfloor (n-1)h \rfloor$, and its supremum is at most $s_k + s_{k+1}$. At most $2/h$ indices n share the same value of k , which gives $\bar{\sigma}(h) \leq 4S$. For $h > 1$, each interval $[(n-1)h, nh]$ is covered by the intervals $[k, k+1]$ that meet it, and each $[k, k+1]$ meets at most two of them. Then $\bar{\sigma}(h) \leq 2hS$. This shows that all upper and lower sums are finite. Now let $X \geq 2$ be an integer. The function z is continuous on $[0, X]$, and the parts of $\bar{\sigma}(h)$ and $\underline{\sigma}(h)$ coming from intervals contained in $[0, X]$ tend to $\int_0^X z$ as $h \rightarrow 0$. For $h \leq 1$, the remaining parts of both sums are at most $4 \sum_{k \geq X-2} s_k$, by the argument above. This tail and $\int_X^\infty z$ tend to zero as $X \rightarrow \infty$, and we conclude that both sums converge to $\int_0^\infty z$. $\square$

A renewal density. The renewal equation that we derive is driven by a logarithmic beta variable, whose density is unbounded near zero. The following lemma records the behavior of its renewal density.

Lemma 16.12 (Renewal density for a logarithmic beta variable). *Let $0 < a < 1$ and $\rho > 0$, and let $Y = -\log B$, where B has the beta distribution with parameters (ρ, a) . The density of Y is*

$$k(y) = \frac{e^{-\rho y} (1 - e^{-y})^{a-1}}{B(\rho, a)}, \quad y > 0,$$

*and its mean is $m = \psi(\rho + a) - \psi(\rho) > 0$. The renewal density $U(y) = \sum_{n \geq 1} k^{*n}(y)$ is locally integrable on $[0, \infty)$, bounded on $[1, \infty)$, and satisfies*

$$U(y) \longrightarrow \frac{1}{m} \quad (y \rightarrow \infty). \quad (16.33)$$

Here we set $k = 0$ on $(-\infty, 0]$, $k^{*1} = k$, and $k^{*(n+1)} = k^{*n} * k$. Each convolution power is a nonnegative Borel probability density supported in $(0, \infty)$. The parameter a in this lemma is arbitrary in $(0, 1)$; we apply it with the exponent a of (16.2) and $\rho = 1 - 2a$.

Proof. The density of B is $b^{\rho-1}(1-b)^{a-1}/B(\rho, a)$ on $(0, 1)$. The change of variables $b = e^{-y}$ gives the formula for k . Differentiating $\log B(\rho, a) = \log \int_0^1 b^{\rho-1}(1-b)^{a-1} db$ in ρ , we find

$\mathbb{E} \log B = \psi(\rho) - \psi(\rho + a)$; the differentiation under the integral sign is justified because $|\log b| b^{\rho'-1}(1-b)^{a-1}$ is integrable on $(0, 1)$, uniformly for ρ' near ρ . We conclude that $m = \mathbb{E}Y = \psi(\rho + a) - \psi(\rho)$, which is positive since ψ is strictly increasing.

We first bound the convolution powers. The function $(1 - e^{-y})/y$ decreases from 1 to $1 - e^{-1}$ on $(0, 1]$, and $y^{1-a}e^{-\rho y/2}$ is bounded on $[1, \infty)$. It follows that there exists K , depending only on a and ρ , such that $k(y) \leq Ky^{a-1}e^{-\rho y/2}$ for all $y > 0$. The beta integral

$$\int_0^y s^{na-1}(y-s)^{a-1} ds = \frac{\Gamma(na)\Gamma(a)}{\Gamma((n+1)a)} y^{(n+1)a-1}$$

and $e^{-\rho s/2}e^{-\rho(y-s)/2} = e^{-\rho y/2}$ imply, by induction on n , that

$$k^{*n}(y) \leq \frac{K^n \Gamma(a)^n}{\Gamma(na)} y^{na-1} e^{-\rho y/2}, \quad y > 0, \quad n \geq 1. \quad (16.34)$$

The series $\sum_n (K\Gamma(a))^n Y_0^{na} / \Gamma(na)$ converges for all $Y_0 > 0$, because $\Gamma(na)$ grows faster than any exponential in n . For $0 < y \leq 1$, we have $y^{na-1} \leq y^{a-1}$, so $U(y) \leq K'y^{a-1}$. In particular, U is finite on $(0, \infty)$ and locally integrable on $[0, \infty)$.

We next show that each k^{*n} is continuous on $(0, \infty)$. Substituting $s_i = yq_i$ in the $(n-1)$ -fold integral that defines $k^{*n}(y)$, we have

$$k^{*n}(y) = y^{n-1} \int_{\Sigma_{n-1}} \prod_{i=1}^n k(yq_i) dq_1 \cdots dq_{n-1}, \quad q_n = 1 - \sum_{i < n} q_i,$$

where $\Sigma_{n-1} = \{q_i > 0, \sum_{i < n} q_i < 1\}$. For y in a compact subset of $(0, \infty)$, the integrand is continuous in y and dominated by a constant times $\prod_i q_i^{a-1}$, which is integrable on Σ_{n-1} . The continuity follows by dominated convergence.

Now fix an integer j with $ja > 1$, and let $k_j = k^{*j}$. By (16.34), we have $k_j(y) \rightarrow 0$ as $y \downarrow 0$, and k_j extends to a continuous function on $[0, \infty)$ with $k_j(0) = 0$. Its zero extension is therefore Borel measurable. The same bound shows that $\sum_{l \geq 0} \sup_{[l, l+1]} k_j < \infty$. Then Lemma 16.11 implies that k_j is directly Riemann integrable. It is the density of the sum of j independent copies of Y . The corresponding probability measure is supported in $(0, \infty)$, it is not arithmetic because it has a density, and its finite positive mean is jm . Thus this law and the nonnegative test $z = k_j$ satisfy every hypothesis of Theorem 16.10. In this case, $Z = k_j + \sum_{n \geq 1} k_j^{*n} * k_j = \sum_{n \geq 1} k_j^{*n}$, where the interchange of sum and integral is justified by Tonelli's theorem. The theorem gives

$$U_j(y) = \sum_{n \geq 1} k_j^{*n}(y) \longrightarrow \frac{1}{jm} \quad (y \rightarrow \infty).$$

Note that $k_j^{*n} = k^{*jna}$. Using (16.34) with $jna - 1 > 0$, we see that the series defining U_j converges uniformly on every interval $[0, Y_0]$. Each term of this series is continuous on

$[0, \infty)$, and so is U_j . Since U_j is continuous and converges at infinity, it is bounded on $[0, \infty)$.

Finally, we decompose the indices $n \geq 1$ according to their residues modulo j , writing $n = i + jl$ with $1 \leq i \leq j$ and $l \geq 0$. Using Tonelli's theorem, we obtain

$$U = \sum_{i=1}^j k^{*i} + \sum_{i=1}^j k^{*i} * U_j.$$

By (16.34), each k^{*i} is bounded on $[1, \infty)$ and tends to zero at infinity. For each i , we have $(k^{*i} * U_j)(y) = \int_0^\infty k^{*i}(s) U_j(y-s) \mathbf{1}_{\{s \leq y\}} ds$. The integrand is bounded by $\|U_j\|_\infty k^{*i}(s)$, which is integrable, and it converges to $k^{*i}(s)/(jm)$ for every s . Then $(k^{*i} * U_j)(y) \rightarrow 1/(jm)$ by dominated convergence, and also $(k^{*i} * U_j) \leq \|U_j\|_\infty$. Summing over i , we find that U is bounded on $[1, \infty)$ and that $U(y) \rightarrow j/(jm) = 1/m$. This proves (16.33). $\square$

The gamma–beta factorization. The second ingredient is a classical identity between gamma and beta variables. For a shape parameter $\alpha > 0$, we say that a random variable has the gamma distribution with shape α if its density is $y^{\alpha-1}e^{-y}/\Gamma(\alpha)$ on $(0, \infty)$.

Lemma 16.13 (Gamma–beta factorization). *Let $\rho, a > 0$ and $b = \rho + a$. Let G_b and B be independent random variables, where G_b has the gamma distribution with shape b and B has the beta distribution with parameters (ρ, a) . Then BG_b has the gamma distribution with shape ρ .*

Proof. Let X and W be independent, with gamma distributions of shapes ρ and a . The map $(x, w) \mapsto (g, y) = (x + w, x/(x + w))$ is a bijection from $(0, \infty)^2$ onto $(0, \infty) \times (0, 1)$, with inverse $(g, y) \mapsto (gy, g(1 - y))$ and Jacobian g . The joint density of $(X + W, X/(X + W))$ is therefore

$$\frac{(gy)^{\rho-1}(g(1-y))^{a-1}e^{-g}}{\Gamma(\rho)\Gamma(a)} g = \frac{g^{b-1}e^{-g}}{\Gamma(b)} \cdot \frac{y^{\rho-1}(1-y)^{a-1}}{B(\rho, a)},$$

where we used $\Gamma(\rho)\Gamma(a) = \Gamma(b)B(\rho, a)$. This shows that $X + W$ and $X/(X + W)$ are independent, with the laws of G_b and B . Then BG_b has the law of $X = [X/(X + W)] \cdot (X + W)$, which is the gamma distribution with shape ρ . $\square$

The proof of the theorem. We now combine the three steps.

Proof of Theorem 16.1. We keep the normalized time variable until the last step. Recall that $D(x) = e^{ax}u(e^x)$ and $\rho = 1 - 2a$, and set $b = 1 - a = \rho + a$. Let G_b have the gamma distribution with shape b , and let B be an independent random variable with the beta distribution with parameters (ρ, a) . For bounded measurable functions ζ on $\mathbb{R}$, we define the averaging operators

$$(\mathcal{K}\zeta)(x) = \mathbb{E} \zeta(x + \log G_b), \quad (\mathcal{B}\zeta)(x) = \mathbb{E} \zeta(x + \log B) = \int_0^\infty k(y) \zeta(x - y) dy,$$

where k is the density of $-\log B$ from Lemma 16.12. By the independence of B and G_b and Fubini's theorem, $\mathcal{BK}\zeta(x) = \mathbb{E} \zeta(x + \log(BG_b))$. Recall from Lemma 16.13 that BG_b has the gamma distribution with shape ρ .

The proof has five steps. We first transform the balance (16.26) into an exact equation in the logarithmic variable. We then show that it is a renewal equation with an integrable source, solve this equation by iteration, and use Lemma 16.12 to find the limit of its solution. Next, we deduce that D converges. Finally, we treat the derivatives and return to the physical lag.

Step 1: the balance in logarithmic time. For $z > 0$, we write $\widehat{\zeta}(z) = \int_0^\infty e^{-zt} \zeta(t) dt$ for the Laplace transform of a bounded function ζ on $[0, \infty)$. By Lemma 16.8, the functions u , u^2 , and $\mathcal{E}$ are bounded. The function Q is continuously differentiable with $Q(0) = 0$ and $0 \leq Q(t) \leq t$. An integration by parts gives $\widehat{Q}'(z) = z\widehat{Q}(z)$, and $\widehat{Q} = \widehat{u}^2$ by Fubini's theorem. Taking the Laplace transform of (16.26), we obtain

$$z\widehat{u}(z)^2 = \vartheta\widehat{u}^2(z) + \widehat{\mathcal{E}}(z), \quad z > 0. \quad (16.35)$$

We evaluate the transforms at $z = e^{-x}$, $x \in \mathbb{R}$. We have $u(t) = t^{-a}D(\log t)$, and the substitution $t = e^x w$ gives

$$\widehat{u}(e^{-x}) = \int_0^\infty e^{-w} (e^x w)^{-a} D(x + \log w) e^x dw = e^{bx} \Gamma(b) (\mathcal{KD})(x),$$

where we used that $w^{b-1}e^{-w}/\Gamma(b)$ is the density of G_b . In the same way, since BG_b has the gamma distribution with shape $\rho = 1 - 2a$,

$$\widehat{u}^2(e^{-x}) = \int_0^\infty e^{-w} (e^x w)^{-2a} D(x + \log w)^2 e^x dw = e^{\rho x} \Gamma(\rho) \mathcal{BK}(D^2)(x).$$

Further, $e^{-x}\widehat{u}(e^{-x})^2 = e^{(2b-1)x}\Gamma(b)^2(\mathcal{KD})(x)^2 = e^{\rho x}\Gamma(b)^2(\mathcal{KD})(x)^2$. Inserting these three identities into (16.35) at $z = e^{-x}$ and dividing by $e^{\rho x}\Gamma(b)^2$, we obtain the exact equation

$$[\mathcal{KD}(x)]^2 = \mathcal{BK}(D^2)(x) + \sigma(x), \quad \sigma(x) = \frac{e^{-\rho x} \widehat{\mathcal{E}}(e^{-x})}{\Gamma(b)^2}, \quad (16.36)$$

where we used the exponent relation in the form $\vartheta\Gamma(\rho) = \Gamma(b)^2$, which is (16.2).

The source σ is integrable on $\mathbb{R}$. Indeed, for $t > 0$, the substitution $w = te^{-x}$ yields $\int_{\mathbb{R}} e^{-\rho x} e^{-te^{-x}} dx = t^{-\rho} \Gamma(\rho)$. By Tonelli's theorem and Lemma 16.8, we have

$$\int_{\mathbb{R}} |\sigma(x)| dx \leq \frac{\Gamma(\rho)}{\Gamma(b)^2} \int_0^\infty |\mathcal{E}(t)| t^{-\rho} dt < \infty. \quad (16.37)$$

Step 2: the linear renewal equation. We set

$$f = \mathcal{K}(D^2), \quad V = \mathcal{K}(D^2) - [\mathcal{K}D]^2.$$

For each x , $V(x)$ is the variance of $D(x + \log G_b)$; in particular, $V \geq 0$. We have $\mathcal{BK}(D^2) = \mathcal{B}f$ and $[\mathcal{K}D]^2 = f - V$, and (16.36) becomes the linear renewal equation

$$f - \mathcal{B}f = g_0, \quad g_0 = V + \sigma. \quad (16.38)$$

We now verify that V is integrable. Let $G^{(1)}$ and $G^{(2)}$ be independent copies of G_b . Then

$$V(x) = \frac{1}{2} \mathbb{E} [D(x + \log G^{(1)}) - D(x + \log G^{(2)})]^2.$$

Integrating out $G^{(1)}$, we find that the ratio $G^{(2)}/G^{(1)}$ has the density $\varsigma^{b-1}(1 + \varsigma)^{-2b}/\mathcal{B}(b, b)$ on $(0, \infty)$. Taking logarithms, we see that the difference $\log G^{(2)} - \log G^{(1)}$ has the density

$$\frac{e^{bh}}{\mathcal{B}(b, b)(1 + e^h)^{2b}} = \frac{[2 \cosh(h/2)]^{-2b}}{\mathcal{B}(b, b)}, \quad h \in \mathbb{R}.$$

By Tonelli's theorem and the translation $x \mapsto x - \log G^{(1)}$, and since $2b = 2 - 2a$, we obtain

$$\int_{\mathbb{R}} V(x) \, dx = \frac{1}{2\mathcal{B}(b, b)} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{[D(x) - D(y)]^2}{[2 \cosh((x - y)/2)]^{2-2a}} \, dx \, dy < \infty,$$

where the finiteness is (16.28). Together with (16.37), this shows that $g_0 \in L^1(\mathbb{R})$.

We next record the regularity of the terms in (16.38). Averages of translates preserve bounds and Lipschitz constants. By Lemma 16.9, D is bounded and Lipschitz. The functions D^2 , $\mathcal{K}D$, f , $[\mathcal{K}D]^2$, V , and $\mathcal{B}f$ inherit these properties. By (16.36) and (16.38), the same holds for σ and g_0 . A uniformly continuous integrable function on $\mathbb{R}$ tends to zero at $\pm\infty$: otherwise there would be infinitely many disjoint intervals of a fixed length on which its absolute value is bounded below by a fixed positive number. In particular, $g_0(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. Further, $f(x) \rightarrow 0$ as $x \rightarrow -\infty$, by dominated convergence, since D is bounded and $D(y) \rightarrow 0$ as $y \rightarrow -\infty$.

Step 3: the renewal representation and the limit of f . The source g_0 is bounded, continuous, and integrable, but it may be signed. We use the local integrability and bounded tail of the renewal density from Lemma 16.12 to handle this source. Let $Y_1, Y_2, \dots$ be independent copies of $Y = -\log B$, and let $S_n = Y_1 + \dots + Y_n$, with $S_0 = 0$. Then $\mathcal{B}^n \zeta(x) = \mathbb{E} \zeta(x - S_n)$, and for $n \geq 1$ the variable S_n has the density k^{*n} . Applying (16.38) repeatedly, we have, for all $n \geq 1$,

$$f(x) = \sum_{l=0}^{n-1} \mathbb{E} g_0(x - S_l) + \mathbb{E} f(x - S_n).$$

By the strong law of large numbers, $S_n/n \rightarrow m > 0$ almost surely, and in particular $S_n \rightarrow \infty$ almost surely. Then $\mathbb{E}f(x - S_n) \rightarrow 0$ by bounded convergence, because f is bounded and $f(-\infty) = 0$. For every fixed x , we have

$$\begin{aligned} \sum_{l \geq 1} \mathbb{E} |g_0(x - S_l)| &= \int_0^\infty U(y) |g_0(x - y)| \, dy \\ &\leq \|g_0\|_\infty \int_0^1 U(y) \, dy + \sup_{y \geq 1} U(y) \|g_0\|_{L^1(\mathbb{R})} < \infty, \end{aligned}$$

by Lemma 16.12 and Tonelli's theorem. This justifies summing the series term by term, and we obtain

$$f(x) = g_0(x) + \int_0^\infty U(y) g_0(x - y) \, dy, \quad x \in \mathbb{R}. \quad (16.39)$$

We now let $x \rightarrow \infty$ in (16.39). The term $g_0(x)$ tends to zero. The part of the integral with $0 < y < 1$ is at most $\sup_{[x-1, x]} |g_0| \int_0^1 U$, which tends to zero. In the remaining part, we substitute $s = x - y$:

$$\int_1^\infty U(y) g_0(x - y) \, dy = \int_{\mathbb{R}} U(x - s) g_0(s) \mathbf{1}_{\{s \leq x-1\}} \, ds.$$

The integrand is dominated by $\sup_{y \geq 1} U(y) |g_0(s)|$, which is integrable, and by (16.33) it converges to $g_0(s)/m$ for every s . We conclude by dominated convergence that

$$f(x) \rightarrow f_\infty = \frac{1}{m} \int_{\mathbb{R}} g_0(s) \, ds \quad (x \rightarrow \infty). \quad (16.40)$$

Step 4: the limit of D . For every fixed $h \in \mathbb{R}$, we have

$$D(x+h)^2 - D(x)^2 = [D(x+h) - D(x)] [D(x+h) + D(x)] \rightarrow 0$$

as $x \rightarrow \infty$, by (16.29) and the boundedness of D . These differences are bounded by $2\|D\|_\infty^2$. By dominated convergence with respect to the law of $\log G_b$, we find

$$f(x) - D(x)^2 = \mathbb{E}[D(x + \log G_b)^2 - D(x)^2] \rightarrow 0.$$

Combined with (16.40), this shows that $D(x)^2 \rightarrow f_\infty$. Recall from Lemma 16.9 that D is bounded, and bounded below by a positive constant on $[0, \infty)$. This implies that $f_\infty \in (0, \infty)$. Since $D > 0$, we conclude that

$$D(x) \rightarrow A_0 = \sqrt{f_\infty} \in (0, \infty), \quad \text{that is,} \quad u(t) \sim A_0 t^{-a} \quad (t \rightarrow \infty). \quad (16.41)$$

We did not need to know the sign of $\int g_0$ in advance: its positivity follows from the lower bound in (16.21).

Step 5: the derivatives and the physical lag. As in the proof of Proposition 16.7, let $F_n = (-1)^n u^{(n)}$ on $(0, \infty)$ for $n \geq 0$. Recall from Lemma 16.3 that each $F_{n+1} = -F'_n$ is positive and nonincreasing. We show by induction that

$$F_n(t) \sim A_0(a)_n t^{-a-n} \quad (t \rightarrow \infty), \quad n \geq 0, \quad (16.42)$$

where $(a)_0 = 1$. For $n = 0$, this is (16.41). Suppose that (16.42) holds for some $n \geq 0$, and write $\alpha = a + n$ and $A_n = A_0(a)_n$. For $\Lambda > 1$, since F_{n+1} is nonincreasing, we have

$$\frac{F_n(t) - F_n(\Lambda t)}{(\Lambda - 1)t} \leq F_{n+1}(t) \leq \frac{F_n(t/\Lambda) - F_n(t)}{(1 - 1/\Lambda)t}.$$

Multiplying by $t^{\alpha+1}$ and letting $t \rightarrow \infty$, we obtain

$$A_n \frac{1 - \Lambda^{-\alpha}}{\Lambda - 1} \leq \liminf_{t \rightarrow \infty} t^{\alpha+1} F_{n+1}(t) \leq \limsup_{t \rightarrow \infty} t^{\alpha+1} F_{n+1}(t) \leq A_n \frac{\Lambda^\alpha - 1}{1 - 1/\Lambda}.$$

Both bounds tend to αA_n as $\Lambda \downarrow 1$, and we conclude that $F_{n+1}(t) \sim (a + n) A_0(a)_n t^{-a-n-1} = A_0(a)_{n+1} t^{-a-n-1}$, which is (16.42) for $n + 1$. This argument is the monotone density theorem; see [30, Section 1.7.3, Theorem 1.7.2 and its following tail-integral variant, p. 39].

Finally, we return to the physical lag, which we again denote by t . By the definition of u , we have $c(t) = q + (1 - q)u(t/(1 - q))$ and $c^{(n)}(t) = (1 - q)^{1-n} u^{(n)}(t/(1 - q))$ for $n \geq 1$. Using (16.42), we obtain

$$c(t) - q \sim (1 - q) A_0 \left(\frac{t}{1 - q} \right)^{-a} = A_0 (1 - q)^{1+a} t^{-a},$$

and, for all $n \geq 1$,

$$(-1)^n c^{(n)}(t) \sim (1 - q)^{1-n} A_0(a)_n \left(\frac{t}{1 - q} \right)^{-a-n} = A_0 (1 - q)^{1+a} (a)_n t^{-a-n}.$$

This is (16.4) with $A_\beta = A_0 (1 - q)^{1+a} \in (0, \infty)$. The case $n = 1$ gives $r(t) \sim a A_\beta t^{-1-a}$, because $r = -c'$. This completes the proof. $\square$

Further questions

We have proved correlation aging for the uniformly initialized CK solution and identified its stationary endpoint. For every $p \geq 3$ and $\beta \geq \beta_d$, the solution satisfies the stationary fluctuation–dissipation theorem at fixed lag, selects a positive marginal plateau and an explicit limiting energy, and forgets every fixed earlier state (Theorem 1.1). For $\beta > \beta_d$, we have also proved an integrated violation of the fluctuation–dissipation theorem (Theorem 1.2). In the cubic model, we obtained quantitative restrictions on correlations at three late times for $\beta \geq \beta_d$ and on relaxation below the plateau for $\beta > \beta_d$ (Theorems 1.3 and 1.4). Several central parts of the physical picture of Cugliandolo and Kurchan [51] remain to be proved. We now formulate them in terms of observables of the CK solution itself, without an asymptotic ansatz.

17.1 The complete slow fluctuation–dissipation relation

Fix $\beta > \beta_d$, write $q = q_\beta$, and let $x = (p - 2)(1 - q)/q$. For a level $c \in (0, q)$, the predicted integrated relation in the aging regime is

$$\chi(t_n, s_n) \longrightarrow 1 - q + x(q - c) \quad \text{if } s_n \rightarrow \infty, t_n - s_n \rightarrow \infty, C(t_n, s_n) \rightarrow c. \quad (17.1)$$

This is the slow fluctuation–dissipation relation of Cugliandolo and Kurchan, with the ratio x given after [51, Equation (12)]. We write it in the conventions of Section 3, in which the fluctuation–dissipation theorem in the stationary regime reads $R(t, s) = \partial_s C(t, s)$. We refer to [35, Sections 3.3–3.4] for a review of the two time sectors and their response, and to [96] for a numerical test of the relation on the CK equations. The weighted ratio in (1.10) identifies one moment of the slow response. Proving (17.1) requires more information than this moment.

A natural intermediate target is the weak convergence, as finite measures on $[0, 1]$, of the response-weighted correlation measures

$$\nu_t = \int_0^t R(t, u) \delta_{C(t, u)} \, du$$

to the predicted measure

$$x \mathbf{1}_{(0, q)}(y) \, dy + \mathbf{1}_{(q, 1)}(y) \, dy.$$

This target includes the statement that no mass accumulates near correlation zero,

$$\lim_{\varepsilon \downarrow 0} \limsup_{t \rightarrow \infty} \nu_t([0, \varepsilon]) = 0,$$

and, since $\nu_t([0, 1]) = \chi(t, 0)$, it also includes the convergence of $\chi(t, 0)$ to the total mass $1 - q + xq$ of the predicted measure. Such a statement concerns the amount of response at each correlation level. To deduce (17.1) from it, we must also link the correlation levels to chronological order. Specifically, we would need the response masses

$$\int_0^{s_n} R(t_n, u) \mathbf{1}_{\{C(t_n, u) > C(t_n, s_n)\}} du \quad \text{and} \quad \int_{s_n}^{t_n} R(t_n, u) \mathbf{1}_{\{C(t_n, u) < C(t_n, s_n)\}} du$$

to vanish, together with control at the boundary level $C(t_n, s_n)$. These masses measure returns of the row $u \mapsto C(t_n, u)$ across the level $C(t_n, s_n)$. The product inequality (1.13) constrains the correlations at three times, but it does not exclude such returns. Note also that the vanishing of the squared defect in (1.9) does not control the integrated negative part of the defect over a growing interval.

17.2 The sharp susceptibility upper bound

In the cubic model with $\beta \geq \beta_d$, we proved in (1.14) that $\liminf_{t \rightarrow \infty} \chi(t, 0) \geq 2(1 - q)$. The expected matching estimate is

$$\limsup_{t \rightarrow \infty} \chi(t, 0) \leq 2(1 - q). \quad (17.2)$$

For $p = 3$, we have $xq = 1 - q$, so $2(1 - q)$ is the total mass $1 - q + xq$ of the predicted measure in Section 17.1. Together with the lower bound (1.14), the estimate (17.2) would determine the limit of the total response mass. This is much more precise than the bound $\chi(t, 0) = o(\sqrt{t})$ in (1.8). A related scalar target is

$$\int_0^t C(t, u) R(t, u) du \rightarrow \frac{(1 - q)(1 + 2q)}{2},$$

which is the first moment of the predicted measure when $p = 3$. The energy identity fixes the moment with C^2 in place of C , since $e_\beta(t) = 3\beta \int_0^t C(t, u)^2 R(t, u) du$ converges to the value $e_\beta(\infty)$ given by (1.4). It does not determine the first moment, nor the unweighted mass near correlation zero. In the proof of Theorem 14.7, we obtained the lower bound by positivity, through a trace argument. An upper bound requires a separate estimate on the response-weighted frame used in that proof. Even equality in (17.2) would leave open the question of chronological returns raised in Section 17.1.

17.3 A nondegenerate clock and physical relaxation rates

The noncommutation of limits in (1.6) does not specify the time scale on which correlations move below q . A stronger description would exhibit an increasing function $\mathfrak{h}: [0, \infty) \rightarrow (0, \infty)$ and a nonconstant profile $\mathcal{C}: (0, 1) \rightarrow [0, q]$ such that, for every $a \in (0, 1)$,

$$\mathcal{C}(t_n, s_n) \longrightarrow \mathcal{C}(a) \quad \text{whenever } s_n \rightarrow \infty, t_n - s_n \rightarrow \infty \text{ and } \frac{\mathfrak{h}(s_n)}{\mathfrak{h}(t_n)} \rightarrow a,$$

and such that sequences with these properties exist for every $a \in (0, 1)$. The last requirement excludes clocks for which the statement is vacuous, such as $\mathfrak{h}(t) = e^t$, for which $\mathfrak{h}(s_n)/\mathfrak{h}(t_n) \rightarrow 0$ whenever $t_n - s_n \rightarrow \infty$. Our results fix neither the shape of $\mathcal{C}$ nor the growth of $\mathfrak{h}$. In the physical theory, the asymptotic equations for the aging regime admit reparametrizations of time, as observed by Cugliandolo and Kurchan [52], so they do not select $\mathfrak{h}$ either. The clock has been studied by numerical integration of the CK equations [96], and by a matching calculation [3] that proposes a clock within a scaling expansion whose remaining exponent is estimated numerically. For $p = 3$ and $\beta \geq \beta_d$, the composition inequalities for last-crossing times in Corollary 14.4 give one-sided restrictions on passage times. A composition law in the limit that is compatible with these inequalities, such as the triangle relations of [52], would give much more information on the time scale.

For $p = 3$, $\beta > \beta_d$, and a level $c \in (0, q)$, the local Fisher cost (1.15) is a constraint in physical time: relaxation to the level c costs a positive amount of $\tau_c(s) \int_s^{s+\tau_c(s)} a_F(u) du$. It does not determine the growth of $\tau_c(s)$. We expect that quantitative estimates on the energy current or on the tail of the Fisher density, together with matching dynamical bounds on passage times, would help to determine this growth. The majorant a_F need not be unique, and we do not claim that it defines a physical aging clock. By contrast, the exponent a in (1.16) describes stationary relaxation in the lag variable after the waiting-time limit. It is not an exponent of the aging dynamics.

17.4 Selection below the dynamical transition

In the interval $\beta_m \leq \beta < \beta_d$, we proved convergence in Proposition 8.4, but we did not decide between the zero plateau and the upper marginal root $q_+(\beta)$. The physical picture of the dynamical transition [48, 51] predicts the zero plateau throughout the interval, and proving this requires excluding the positive alternative. One sufficient estimate would be $e_\beta(\infty) \leq \beta$ throughout the interval, since by Remark 8.7 the positive branch has $e_\beta(\infty) > \beta$. Such an estimate would complete the classification of the stationary plateau from below.

17.5 Other orders of limits

All of these questions concern a fixed positive temperature and the order of limits in which the dimension tends to infinity before the times do. The bridge of Proposition 3.3 transfers the conclusions about correlations and about primitives of the response to the Langevin dynamics in this order, and Proposition 4.1 gives the limit of the annealed energy. The annealed response means also converge uniformly on fixed causal time windows. The concentration statement in the bridge concerns the Brownian observables representing their primitives; see Remark 3.5. These fixed-window results give no information on observation times that grow with the dimension, and the low-temperature limit in Remark 8.6 is taken after all other limits. We regard a description of the dynamics on time scales that grow with the dimension, and a theorem at zero temperature, as separate problems.

References

- [1] Girish S. Agarwal. Fluctuation-dissipation theorems for systems in non-thermal equilibrium and applications. *Zeitschrift für Physik*, 252(1):25–38, 1972. doi: 10.1007/BF01391621. URL <https://link.springer.com/article/10.1007/BF01391621>.
- [2] Ada Altieri, Giulio Biroli, and Chiara Cammarota. Dynamical mean-field theory and aging dynamics. *Journal of Physics A: Mathematical and Theoretical*, 53:375006, 2020. doi: 10.1088/1751-8121/aba3dd.
- [3] Alexey Andreanov and Alexandre Lefèvre. Crossover from stationary to aging regime in glassy dynamics. *Europhysics Letters*, 76(5):919–925, 2006. doi: 10.1209/epl/i2006-10352-9.
- [4] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950. doi: 10.1090/S0002-9947-1950-0051437-7. URL <https://www.ams.org/journals/tran/1950-068-03/S0002-9947-1950-0051437-7/>.
- [5] Antonio Auffinger and Gérard Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *The Annals of Probability*, 41(6):4214–4247, 2013. doi: 10.1214/13-AOP862.
- [6] Antonio Auffinger and Wei-Kuo Chen. On properties of Parisi measures. *Probability Theory and Related Fields*, 161(3–4):817–850, 2015. doi: 10.1007/s00440-014-0563-y.
- [7] Antonio Auffinger, Gérard Ben Arous, and Jiří Černý. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2): 165–201, 2013. doi: 10.1002/cpa.21422. URL <https://arxiv.org/abs/1003.1129>.
- [8] Marco Baiesi, Christian Maes, and Bram Wynants. Fluctuations and response of nonequilibrium states. *Physical Review Letters*, 103(1):010602, 2009. doi: 10.1103/PhysRevLett.103.010602. URL <https://arxiv.org/abs/0902.3955>.
- [9] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In Jacques Azéma and Marc Yor, editors, *Séminaire de Probabilités XIX 1983/84*, volume 1123 of *Lecture Notes in Mathematics*, pages 177–206. Springer, 1985. doi: 10.1007/BFb0075847. URL <https://link.springer.com/chapter/10.1007/BFb0075847>.

- [10] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2014. doi: 10.1007/978-3-319-00227-9. URL <https://link.springer.com/book/10.1007/978-3-319-00227-9>.
- [11] Alain Barrat, Raffaella Burioni, and Marc Mézard. Ageing classification in glassy dynamics. *Journal of Physics A: Mathematical and General*, 29(7):1311–1330, 1996. doi: 10.1088/0305-4470/29/7/005. URL <https://arxiv.org/abs/cond-mat/9509142>.
- [12] Alain Barrat, Silvio Franz, and Giorgio Parisi. Temperature evolution and bifurcations of metastable states in mean-field spin glasses, with connections with structural glasses. *Journal of Physics A: Mathematical and General*, 30(16):5593–5612, 1997. doi: 10.1088/0305-4470/30/16/006.
- [13] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011. doi: 10.1109/TIT.2010.2094817.
- [14] Gérard Ben Arous. Aging and spin-glass dynamics. In *Proceedings of the International Congress of Mathematicians, Beijing 2002*, volume III, pages 3–14. Higher Education Press, Beijing, 2002.
- [15] Gérard Ben Arous and Jiří Černý. Dynamics of trap models. In *Mathematical Statistical Physics*, volume 83 of *Les Houches*, pages 331–394. Elsevier, 2006. doi: 10.1016/S0924-8099(06)80045-4.
- [16] Gérard Ben Arous and Jiří Černý. The arcsine law as a universal aging scheme for trap models. *Communications on Pure and Applied Mathematics*, 61(3):289–329, 2008. doi: 10.1002/cpa.20177.
- [17] Gérard Ben Arous and Alice Guionnet. Large deviations for Langevin spin glass dynamics. *Probability Theory and Related Fields*, 102(4):455–509, 1995. doi: 10.1007/BF01198846.
- [18] Gérard Ben Arous and Alice Guionnet. Symmetric Langevin spin glass dynamics. *The Annals of Probability*, 25(3):1367–1422, 1997. doi: 10.1214/aop/1024404517.
- [19] Gérard Ben Arous and Onur Gün. Universality and extremal aging for dynamics of spin glasses on subexponential time scales. *Communications on Pure and Applied Mathematics*, 65(1):77–127, 2012. doi: 10.1002/cpa.20372.
- [20] Gérard Ben Arous and Aukosh Jagannath. Shattering versus metastability in spin glasses. *Communications on Pure and Applied Mathematics*, 77(1):139–176, 2024. doi: 10.1002/cpa.22133.

- [21] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Aging of spherical spin glasses. *Probability Theory and Related Fields*, 120(1):1–67, 2001. doi: 10.1007/PL00008774. URL <https://link.springer.com/article/10.1007/PL00008774>.
- [22] Gérard Ben Arous, Anton Bovier, and Véronique Gayrard. Glauber dynamics of the random energy model. I. metastable motion on the extreme states. *Communications in Mathematical Physics*, 235(3):379–425, 2003. doi: 10.1007/s00220-003-0798-4.
- [23] Gérard Ben Arous, Anton Bovier, and Véronique Gayrard. Glauber dynamics of the random energy model. II. aging below the critical temperature. *Communications in Mathematical Physics*, 236(1):1–54, 2003. doi: 10.1007/s00220-003-0799-3.
- [24] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo–Kurchan equations for dynamics of spin-glasses. *Probability Theory and Related Fields*, 136(4): 619–660, 2006. doi: 10.1007/s00440-005-0491-y. URL <https://arxiv.org/abs/math/0409273>.
- [25] Gérard Ben Arous, Anton Bovier, and Jiří Černý. Universality of the REM for dynamics of mean-field spin glasses. *Communications in Mathematical Physics*, 282(3):663–695, 2008. doi: 10.1007/s00220-008-0565-7.
- [26] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Bounding flows for spherical spin glass dynamics. *Communications in Mathematical Physics*, 373(3): 1011–1048, 2020. doi: 10.1007/s00220-019-03649-4.
- [27] U. Bengtzelius, W. Götze, and A. Sjölander. Dynamics of supercooled liquids and the glass transition. *Journal of Physics C: Solid State Physics*, 17:5915–5934, 1984. doi: 10.1088/0022-3719/17/33/005.
- [28] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*, volume 100 of *Graduate Texts in Mathematics*. Springer, 1984. doi: 10.1007/978-1-4612-1128-0. URL <https://link.springer.com/book/10.1007/978-1-4612-1128-0>.
- [29] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 3 edition, 1995. ISBN 0-471-00710-2. URL <https://www.colorado.edu/amath/sites/default/files/attached-files/billingsley.pdf>.
- [30] Nicholas H. Bingham, Charles M. Goldie, and Jozef L. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1987. doi: 10.1017/CBO9780511721434. URL <https://doi.org/10.1017/CBO9780511721434>.

- [31] Giulio Biroli. Dynamical TAP approach to mean field glassy systems. *Journal of Physics A: Mathematical and General*, 32(48):8365–8388, 1999. doi: 10.1088/0305-4470/32/48/301. URL <https://arxiv.org/abs/cond-mat/9909415>.
- [32] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1): 333–366, 2014. doi: 10.1007/s00220-013-1862-3.
- [33] Jean-Philippe Bouchaud. Weak ergodicity breaking and aging in disordered systems. *Journal de Physique I*, 2(9):1705–1713, 1992. doi: 10.1051/jp1:1992238.
- [34] Jean-Philippe Bouchaud, Leticia F. Cugliandolo, Jorge Kurchan, and Marc Mézard. Mode-coupling approximations, glass theory and disordered systems. *Physica A: Statistical Mechanics and its Applications*, 226(3–4):243–273, 1996. doi: 10.1016/0378-4371(95)00423-8.
- [35] Jean-Philippe Bouchaud, Leticia F. Cugliandolo, Jorge Kurchan, and Marc Mézard. Out of equilibrium dynamics in spin-glasses and other glassy systems. In A. Peter Young, editor, *Spin Glasses and Random Fields*, pages 161–223. World Scientific, Singapore, 1998. doi: 10.1142/9789812819437_0006.
- [36] Anton Bovier and Véronique Gayraud. Convergence of clock processes in random environments and ageing in the p -spin SK model. *The Annals of Probability*, 41(2): 817–847, 2013. doi: 10.1214/11-AOP705.
- [37] R. H. Cameron and W. T. Martin. Transformations of Wiener integrals under translations. *Annals of Mathematics*, 45(2):386–396, 1944. doi: 10.2307/1969276. URL <https://www.jstor.org/stable/1969276>.
- [38] Andrea Cavagna, Irene Giardinà, and Giorgio Parisi. Stationary points of the Thouless–Anderson–Palmer free energy. *Physical Review B*, 57:11251–11257, 1998. doi: 10.1103/PhysRevB.57.11251.
- [39] Michael Celentano, Chen Cheng, and Andrea Montanari. The high-dimensional asymptotics of first order methods with random data, 2021. arXiv:2112.07572, version 3, April 2026.
- [40] Jiří Černý and Tobias Wassmer. Aging of the Metropolis dynamics on the random energy model. *Probability Theory and Related Fields*, 167(1–2):253–303, 2017. doi: 10.1007/s00440-015-0681-1.
- [41] Claudio Chamon, Malcolm P. Kennett, Horacio E. Castillo, and Leticia F. Cugliandolo. Separation of time scales and reparametrization invariance for aging systems. *Physical Review Letters*, 89:217201, 2002. doi: 10.1103/PhysRevLett.89.217201.

- [42] Wei-Kuo Chen. The Aizenman–Sims–Starr scheme and Parisi formula for mixed p -spin spherical models. *Electronic Journal of Probability*, 18:no. 94, 1–14, 2013. doi: 10.1214/EJP.v18-2580.
- [43] Wei-Kuo Chen and Arnab Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p -spin models. *Communications in Mathematical Physics*, 350(1):129–173, 2017. doi: 10.1007/s00220-016-2808-3.
- [44] Riccardo Cipolloni and Leticia F. Cugliandolo. Coupling spherical p -spin systems, 2026. arXiv:2609.17522, version 1, September 2026.
- [45] Vincenzo Citro and Federico Ricci-Tersenghi. Strong ergodicity breaking in dynamical mean-field equations for mixed p -spin glasses. *Physical Review Letters*, 135(24):247102, 2025. doi: 10.1103/lvmp-pydk.
- [46] Andrea Crisanti and Hans-Jürgen Sommers. The spherical p -spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter*, 87(3):341–354, 1992. doi: 10.1007/BF01309287. URL <https://link.springer.com/article/10.1007/BF01309287>.
- [47] Andrea Crisanti and Hans-Jürgen Sommers. Thouless-Anderson-Palmer approach to the spherical p -spin spin glass model. *Journal de Physique I*, 5(7):805–813, 1995. doi: 10.1051/jp1:1995164.
- [48] Andrea Crisanti, Heinz Horner, and Hans-Jürgen Sommers. The spherical p -spin interaction spin-glass model: the dynamics. *Zeitschrift für Physik B Condensed Matter*, 92(2):257–271, 1993. doi: 10.1007/BF01312184. URL <https://link.springer.com/article/10.1007/BF01312184>.
- [49] Leticia F. Cugliandolo. Course 7: Dynamics of glassy systems. In Jean-Louis Barrat, Mikhail Feigelman, Jorge Kurchan, and Jean Dalibard, editors, *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter*, volume 77 of *Les Houches*, pages 367–521. Springer, Berlin, Heidelberg, 2004. doi: 10.1007/978-3-540-44835-8_7.
- [50] Leticia F. Cugliandolo and David S. Dean. Full dynamical solution for a spherical spin-glass model. *Journal of Physics A: Mathematical and General*, 28(15):4213–4234, 1995. doi: 10.1088/0305-4470/28/15/003.
- [51] Leticia F. Cugliandolo and Jorge Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173–176, 1993. doi: 10.1103/PhysRevLett.71.173. URL <https://arxiv.org/abs/cond-mat/9303036>.

- [52] Leticia F. Cugliandolo and Jorge Kurchan. On the out-of-equilibrium relaxation of the Sherrington-Kirkpatrick model. *Journal of Physics A: Mathematical and General*, 27(17):5749–5772, 1994. doi: 10.1088/0305-4470/27/17/011.
- [53] Leticia F. Cugliandolo and Pierre Le Doussal. Large time nonequilibrium dynamics of a particle in a random potential. *Physical Review E*, 53(2):1525–1552, 1996. doi: 10.1103/PhysRevE.53.1525.
- [54] Leticia F. Cugliandolo, David S. Dean, and Jorge Kurchan. Fluctuation-dissipation theorems and entropy production in relaxational systems. *Physical Review Letters*, 79(12):2168–2171, 1997. doi: 10.1103/PhysRevLett.79.2168. URL <https://arxiv.org/abs/cond-mat/9705002>.
- [55] Leticia F. Cugliandolo, Jorge Kurchan, and Luca Peliti. Energy flow, partial equilibration, and effective temperatures in systems with slow dynamics. *Physical Review E*, 55(4):3898–3914, 1997. doi: 10.1103/PhysRevE.55.3898.
- [56] Yatin Dandi, David Gamarnik, Francisco Pernice, and Lenka Zdeborová. Rigorous asymptotics for first-order algorithms through the dynamical cavity method. In Steve Hanneke and Tor Lattimore, editors, *Proceedings of the Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pages 1611–1646. PMLR, 2026.
- [57] Amir Dembo and Reza Gheissari. Diffusions interacting through a random matrix: universality via stochastic Taylor expansion. *Probability Theory and Related Fields*, 180(3–4):1057–1097, 2021. doi: 10.1007/s00440-021-01027-7. URL <https://arxiv.org/abs/2006.13167>.
- [58] Amir Dembo and Eliran Subag. Dynamics for spherical spin glasses: disorder dependent initial conditions. *Journal of Statistical Physics*, 181(2):465–514, 2020. doi: 10.1007/s10955-020-02587-z. URL <https://arxiv.org/abs/1908.01126>.
- [59] Amir Dembo and Eliran Subag. Dynamics for spherical spin glasses: Gibbs distributed initial conditions, 2025. URL <https://arxiv.org/abs/2503.23342>. arXiv version 3, 8 July 2026.
- [60] Amir Dembo, Alice Guionnet, and Christian Mazza. Limiting dynamics for spherical models of spin glasses at high temperature. *Journal of Statistical Physics*, 128(4): 847–881, 2007. doi: 10.1007/s10955-006-9239-z. URL <https://arxiv.org/abs/math/0609546>.
- [61] Amir Dembo, Eyal Lubetzky, and Ofer Zeitouni. Universality for Langevin-like spin glass dynamics. *The Annals of Applied Probability*, 31(6):2864–2880, 2021. doi: 10.1214/21-AAP1665.

- [62] Bernard Derrida. Random-energy model: Limit of a family of disordered models. *Physical Review Letters*, 45:79–82, 1980. doi: 10.1103/PhysRevLett.45.79.
- [63] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *The Annals of Probability*, 49(6):2922–2960, 2021. doi: 10.1214/21-AOP1519.
- [64] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Shattering in pure spherical spin glasses. *Communications in Mathematical Physics*, 406(5):111, 2025. doi: 10.1007/s00220-025-05243-3. URL <https://arxiv.org/abs/2307.04659>.
- [65] Zhou Fan, Justin Ko, Bruno Loureiro, Yue M. Lu, and Yandi Shen. Dynamical mean-field analysis of adaptive Langevin diffusions: Propagation-of-chaos and convergence of the linear response, 2025. arXiv:2504.15556, version 2, November 2025.
- [66] Zhou Fan, Justin Ko, Bruno Loureiro, Yue M. Lu, and Yandi Shen. Dynamical mean-field analysis of adaptive Langevin diffusions: Replica-symmetric fixed point and empirical Bayes, 2025. arXiv:2504.15558, version 2, November 2025.
- [67] Zhou Fan, Theodor Misiakiewicz, Leda Wang, and Garrett G. Wen. Dynamical mean-field limit and replica-symmetric free energy for the orthogonally-invariant SK model, 2026. arXiv:2607.10102, version 2, September 2026.
- [68] William Feller. *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley & Sons, 2 edition, 1971. ISBN 9780471257097.
- [69] Giampaolo Folena and Francesco Zamponi. On weak ergodicity breaking in mean-field spin glasses. *SciPost Physics*, 15(3):109, 2023. doi: 10.21468/SciPostPhys.15.3.109.
- [70] Giampaolo Folena, Silvio Franz, and Federico Ricci-Tersenghi. Rethinking mean-field glassy dynamics and its relation with the energy landscape: The surprising case of the spherical mixed p -spin model. *Physical Review X*, 10:031045, 2020. doi: 10.1103/PhysRevX.10.031045.
- [71] Thomas Franosch and Thomas Voigtmann. Completely monotone solutions of the mode-coupling theory for mixtures. *Journal of Statistical Physics*, 109(1–2):237–259, 2002. doi: 10.1023/A:1019991729106. URL <https://arxiv.org/abs/cond-mat/0202238>.
- [72] Silvio Franz and Miguel A. Virasoro. Quasi-equilibrium interpretation of ageing dynamics. *Journal of Physics A: Mathematical and General*, 33(5):891–905, 2000. doi: 10.1088/0305-4470/33/5/306. URL <https://arxiv.org/abs/cond-mat/9907438>.

- [73] Silvio Franz, Marc Mézard, Giorgio Parisi, and Luca Peliti. Measuring equilibrium properties in aging systems. *Physical Review Letters*, 81:1758–1761, 1998. doi: 10.1103/PhysRevLett.81.1758.
- [74] Elizabeth Gardner. Spin glasses with p -spin interactions. *Nuclear Physics B*, 257: 747–765, 1985. doi: 10.1016/0550-3213(85)90374-8.
- [75] Véronique Gayraud. Aging in Metropolis dynamics of the REM: a proof. *Probability Theory and Related Fields*, 174(1–2):501–551, 2019. doi: 10.1007/s00440-018-0873-6.
- [76] Cédric Gerbelot, Emanuele Troiani, Francesca Mignacco, Florent Krzakala, and Lenka Zdeborová. Rigorous dynamical mean-field theory for stochastic gradient descent methods. *SIAM Journal on Mathematics of Data Science*, 6(2):400–427, 2024. doi: 10.1137/23M1594388.
- [77] Reza Gheissari and Aukosh Jagannath. On the spectral gap of spherical spin glass dynamics. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 55(2): 756–776, 2019. doi: 10.1214/18-AIHP897.
- [78] Federico Ghimenti, Ludovic Berthier, Jorge Kurchan, and Frédéric van Wijland. Clever algorithms for glasses work by time reparameterization. *Proceedings of the National Academy of Sciences*, 123(4):e2520818123, 2026. doi: 10.1073/pnas.2520818123. URL <https://doi.org/10.1073/pnas.2520818123>. Correction: *Proc. Natl. Acad. Sci. USA* 123(12), e2606723123 (2026), doi:10.1073/pnas.2606723123.
- [79] Wolfgang Götze. Some aspects of phase transitions described by the self consistent current relaxation theory. *Zeitschrift für Physik B Condensed Matter*, 56:139–154, 1984. doi: 10.1007/BF01469695.
- [80] Wolfgang Götze. Properties of the glass instability treated within a mode coupling theory. *Zeitschrift für Physik B Condensed Matter*, 60:195–203, 1985. doi: 10.1007/BF01304439.
- [81] Wolfgang Götze. The essentials of the mode-coupling theory for glassy dynamics. *Condensed Matter Physics*, 1(4):873–904, 1998. doi: 10.5488/CMP.1.4.873. URL <https://icmp.lviv.ua/journal/zbirnyk.16/008/art08.pdf>.
- [82] Wolfgang Götze. *Complex Dynamics of Glass-Forming Liquids: A Mode-Coupling Theory*, volume 143 of *International Series of Monographs on Physics*. Oxford University Press, 2009. doi: 10.1093/acprof:oso/9780199235346.001.0001. URL <https://academic.oup.com/book/10864>.

- [83] Wolfgang Götze and Lennart Sjögren. Relaxation processes in supercooled liquids. *Reports on Progress in Physics*, 55(3):241–376, 1992. doi: 10.1088/0034-4885/55/3/001. URL <https://doi.org/10.1088/0034-4885/55/3/001>.
- [84] Wolfgang Götze and Lennart Sjögren. General properties of certain non-linear integro-differential equations. *Journal of Mathematical Analysis and Applications*, 195(1):230–250, 1995. doi: 10.1006/jmaa.1995.1352. URL <https://doi.org/10.1006/jmaa.1995.1352>.
- [85] David J. Gross and Marc Mézard. The simplest spin glass. *Nuclear Physics B*, 240(4): 431–452, 1984. doi: 10.1016/0550-3213(84)90237-2.
- [86] Malte Grunwald. Sanov results for Glauber spin-glass dynamics. *Probability Theory and Related Fields*, 106(2):187–232, 1996. doi: 10.1007/s004400050062.
- [87] Alice Guionnet. Averaged and quenched propagation of chaos for spin glass dynamics. *Probability Theory and Related Fields*, 109(2):183–215, 1997. doi: 10.1007/s004400050130.
- [88] Alice Guionnet. Dynamics for spherical models of spin-glass and aging. In Erwin Bolthausen and Anton Bovier, editors, *Spin Glasses*, volume 1900 of *Lecture Notes in Mathematics*, pages 117–144. Springer, 2007. doi: 10.1007/978-3-540-40908-3_5.
- [89] Alice Guionnet and Christian Mazza. Long time behaviour of the solution to non-linear Kraichnan equations. *Probability Theory and Related Fields*, 131(4):493–518, 2005. doi: 10.1007/s00440-004-0382-7. URL <https://arxiv.org/abs/math/0409274>.
- [90] Martin Hairer. An introduction to stochastic PDEs. Author’s lecture notes, 2025. URL <https://www.hairer.org/notes/SPDEs.pdf>. Version of November 10, 2025.
- [91] Desmond J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001. doi: 10.1137/S0036144500378302.
- [92] Brice Huang and Mark Sellke. Algorithmic threshold for multi-species spherical spin glasses, 2023. URL <https://arxiv.org/abs/2303.12172>. arXiv version 2, 13 September 2023.
- [93] Brice Huang and Mark Sellke. Optimization algorithms for multi-species spherical spin glasses. *Journal of Statistical Physics*, 191:29, 2024. doi: 10.1007/s10955-024-03242-7.

- [94] Brice Huang and Mark Sellke. Tight Lipschitz hardness for optimizing mean field spin glasses. *Communications on Pure and Applied Mathematics*, 78(1):60–119, 2025. doi: 10.1002/cpa.22222. URL <https://arxiv.org/abs/2110.07847>.
- [95] Svante Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 1997. doi: 10.1017/CBO9780511526169. URL <https://doi.org/10.1017/CBO9780511526169>.
- [96] Bongsoo Kim and Arnulf Latz. The dynamics of the spherical p -spin model: from microscopic to asymptotic. *Europhysics Letters*, 53(5):660–666, 2001. doi: 10.1209/epl/i2001-00202-4.
- [97] T. R. Kirkpatrick and Devarajan Thirumalai. Dynamics of the structural glass transition and the p -spin-interaction spin-glass model. *Physical Review Letters*, 58: 2091–2094, 1987. doi: 10.1103/PhysRevLett.58.2091.
- [98] T. R. Kirkpatrick and Devarajan Thirumalai. p -spin-interaction spin-glass models: Connections with the structural glass problem. *Physical Review B*, 36:5388–5397, 1987. doi: 10.1103/PhysRevB.36.5388.
- [99] T. R. Kirkpatrick, Devarajan Thirumalai, and Peter G. Wolynes. Scaling concepts for the dynamics of viscous liquids near an ideal glassy state. *Physical Review A*, 40: 1045–1054, 1989. doi: 10.1103/PhysRevA.40.1045.
- [100] Ryogo Kubo. Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957. doi: 10.1143/JPSJ.12.570. URL https://www.jstage.jst.go.jp/article/jpsj1946/12/6/12_6_570/_article.
- [101] Hiroshi Kunita. Stochastic flows and stochastic partial differential equations. In Andrew M. Gleason, editor, *Proceedings of the International Congress of Mathematicians, Berkeley, 1986*, volume 2, pages 1021–1031. American Mathematical Society, 1987. URL <https://www.mathunion.org/fileadmin/ICM/Proceedings/ICM1986.2/ICM1986.2.ocr.pdf>.
- [102] Hiroshi Kunita. *Stochastic Flows and Stochastic Differential Equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.
- [103] Jorge Kurchan and Laurent Laloux. Phase space geometry and slow dynamics. *Journal of Physics A: Mathematical and General*, 29(9):1929–1948, 1996. doi: 10.1088/0305-4470/29/9/009.

- [104] Jorge Kurchan, Giorgio Parisi, and Miguel Angel Virasoro. Barriers and metastable states as saddle points in the replica approach. *Journal de Physique I*, 3(8):1819–1838, 1993. doi: 10.1051/jp1:1993217.
- [105] Johannes Lang, Subir Sachdev, and Sebastian Diehl. Numerical renormalization of glassy dynamics. *Physical Review Letters*, 135:247101, 2025. doi: 10.1103/z64g-nqs6.
- [106] Johannes Lang, Vincenzo Citro, Luca Leuzzi, and Federico Ricci-Tersenghi. Aging phase diagram and exact asymptotic energies of mixed spherical spin glasses, 2026. arXiv:2608.20623, version 1, August 2026.
- [107] E. Leutheusser. Dynamical model of the liquid-glass transition. *Physical Review A*, 29(5):2765–2773, 1984. doi: 10.1103/PhysRevA.29.2765.
- [108] Tengyuan Liang, Subhabrata Sen, and Pragya Sur. High-dimensional asymptotics of Langevin dynamics in spiked matrix models. *Information and Inference*, 12(4): 2720–2752, 2023. doi: 10.1093/imaiai/iaad042.
- [109] Lars Lundgren, Per Svedlindh, Per Nordblad, and Olov Beckman. Dynamics of the relaxation-time spectrum in a CuMn spin-glass. *Physical Review Letters*, 51:911–914, 1983. doi: 10.1103/PhysRevLett.51.911.
- [110] Andrea Montanari. Optimization of the Sherrington–Kirkpatrick hamiltonian. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science*, pages 1417–1433. IEEE, 2019. doi: 10.1109/FOCS.2019.00087.
- [111] Issai Schur. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 140: 1–28, 1911. doi: 10.1515/crll.1911.140.1. URL https://gdz.sub.uni-goettingen.de/id/PPN243919689_0140.
- [112] Mark Sellke. The threshold energy of low temperature Langevin dynamics for pure spherical spin glasses. *Communications on Pure and Applied Mathematics*, 77(11):4065–4099, 2024. doi: 10.1002/cpa.22197. URL <https://arxiv.org/abs/2305.07956>.
- [113] Mark Sellke. On marginal stability in low temperature spherical spin glasses. *Communications in Mathematical Physics*, 406(9):227, 2025. doi: 10.1007/s00220-025-05410-6. URL <https://arxiv.org/abs/2409.15728>. arXiv version 4, 23 August 2026.
- [114] Lennart Sjögren. Regenerative processes in supercooled liquids and glasses. *Physica A: Statistical Mechanics and its Applications*, 322:81–117, 2003. doi: 10.1016/S0378-4371(02)01832-0. URL <https://www.sciencedirect.com/science/article/abs/pii/S0378437102018320>.

- [115] Haim Sompolinsky and Annette Zippelius. Dynamic theory of the spin-glass phase. *Physical Review Letters*, 47:359–362, 1981. doi: 10.1103/PhysRevLett.47.359.
- [116] Haim Sompolinsky and Annette Zippelius. Relaxational dynamics of the Edwards–Anderson model and the mean-field theory of spin-glasses. *Physical Review B*, 25: 6860–6875, 1982. doi: 10.1103/PhysRevB.25.6860.
- [117] Eliran Subag. The complexity of spherical p -spin models—a second moment approach. *The Annals of Probability*, 45(5):3385–3450, 2017. doi: 10.1214/16-AOP1139.
- [118] Eliran Subag. Following the ground states of Full-RSB spherical spin glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021. doi: 10.1002/cpa.21922. URL <https://arxiv.org/abs/1812.04588>.
- [119] Eliran Subag. The free energy of spherical pure p -spin models: computation from the TAP approach. *Probability Theory and Related Fields*, 186(3–4):715–734, 2023. doi: 10.1007/s00440-023-01200-0.
- [120] Michel Talagrand. Free energy of the spherical mean field model. *Probability Theory and Related Fields*, 134(3):339–382, 2006. doi: 10.1007/s00440-005-0433-8.
- [121] Eric Vincent, Jacques Hammann, Miguel Ocio, Jean-Philippe Bouchaud, and Leticia F. Cugliandolo. Slow dynamics and aging in spin glasses. In Miguel Rubí and Conrado Pérez-Vicente, editors, *Complex Behaviour of Glassy Systems*, volume 492 of *Lecture Notes in Physics*, pages 184–219. Springer, Berlin, Heidelberg, 1997. doi: 10.1007/BFb0104827.
- [122] David Vernon Widder. *The Laplace Transform*, volume 6 of *Princeton Mathematical Series*. Princeton University Press, 1941. URL <https://www.jstor.org/stable/j.ctt183ptft>.