The exact variational constant for empirical Wiener transport

Scott Schubr
Luminary Oddities

November 26, 2026 (edition 3)

Abstract

Coste conjectured that the Wasserstein distance between the empirical law of n independent Brownian paths and Wiener measure, multiplied by $\sqrt{\log n}$, converges; he asked where the limit lies, whether it equals the nearest-sample constant $\sqrt{\kappa_d}$ of Dereich and Lifshits, and what happens when Girsanov importance sampling uses a non-Brownian proposal. For d-dimensional Brownian motion in the supremum metric we prove that, for every finite Wasserstein order, the normalized distance converges almost surely and in every sample moment to $\sqrt{\Gamma_d}$, where Γ_d is the long-time minimum mutual information per unit time for coupling two exact Brownian bridge laws within a unit tube. This gives, for Wiener measure, the sharp mean asymptotic that Boissard and Le Gouic left aside. In every dimension $\Gamma_d > \kappa_d$, so equal sample masses raise the leading constant by exactly $\sqrt{\Gamma_d} - \sqrt{\kappa_d}$. For self-normalized importance sampling the limit inferior is at least $\sqrt{\Gamma_d}$ for every target absolutely continuous with respect to Wiener measure and every proposal dominating it, and the almost-sure limit equals $\sqrt{\Gamma_d}$ whenever the target density lies in L^r and the likelihood ratio in L^s of the proposal for some $r, s > 1$. This includes Girsanov targets with drifts satisfying Novikov's condition with a margin; the constant does not depend on the target. Integrability of the density does not suffice: conditioning Schauder coefficients produces a target equivalent to Wiener measure, with all moments, for which the normalized error tends to infinity for every order, every proposal and every sample.

Keywords. Wiener measure, empirical Wasserstein distance, Brownian bridges, small-ball probabilities, relative entropy, importance sampling, Girsanov theorem.
MSC 2020. Primary 60J65, 60B10. Secondary 49Q22, 60F15, 60F25, 65C05, 94A34.

1 Introduction and main results

Let $\widehat{\mathsf{W}}_n$ be the equal-weight empirical law of n independent standard d-dimensional Brownian paths on $[0,1]$, let W_1 be their common Wiener law, and let W_p be the Wasserstein distance of order p for the supremum distance between paths. Coste [3, Theorem 1.1] proves that, almost surely and for every $p \geq 1$,

$$\sqrt{2\lambda_d} \leq \liminf_{n\to\infty} \sqrt{\log n}\, W_p(\widehat{\mathsf{W}}_n, \mathsf{W}_1) \leq \limsup_{n\to\infty} \sqrt{\log n}\, W_p(\widehat{\mathsf{W}}_n, \mathsf{W}_1) \leq 4\sqrt{2\lambda_d},$$

where λ_d is the principal Dirichlet eigenvalue of $-\frac{1}{2}\Delta$ on the Euclidean unit ball, so that $\lambda_1 = \pi^2/8$. This article determines the limit and answers the following questions.

1. *Convergence and the value of the limit* (Coste [3, §1.5]). Coste calls the convergence of $\sqrt{\log n}\, W_p(\widehat{\mathsf{W}}_n, \mathsf{W}_1)$ "purely conjectural" and asks where the limiting constant should lie. Theorem 1 proves the convergence for every $d \geq 1$ and every finite $p \geq 1$, almost surely and in every finite sample moment, and identifies the limit as $\sqrt{\Gamma_d}$, where Γ_d is the long-time

minimum mutual information per unit time for coupling two exact Brownian bridge laws within distance one throughout. A finite-block estimate quantifies the approximation of this rate by a relaxed bridge problem.

2. *Comparison with the nearest-sample constant* (Coste [3, §§1.5.1–1.5.3]). Sending each path to its nearest sample gives Coste's lower bound L_n, and $\mathbb{E}L_n \sim (\kappa_d/\log n)^{1/2}$, where κ_d is the randomly centered Wiener small-ball constant of Dereich and Lifshits [7, Theorems 5.1 and 6.1]. Coste asks whether the empirical limit equals $\sqrt{\kappa_1}$, and he reports in §1.5.3 that his scalar computations neither confirm nor exclude this value. Theorem 2 answers the question negatively in every dimension and for every order: $\Gamma_d > \kappa_d$.

3. *Non-Brownian proposals* (Coste [3, §2.2, third paragraph]). Coste treats Girsanov reweighting with Brownian proposals and Itô targets with bounded Lipschitz drifts [3, §1.4], and writes that what happens when the proposal is not itself Brownian is not clear. Theorem 3 answers this for every proposal P whose likelihood ratio dQ/dP lies in $L^s(P)$ for some $s > 1$ and every target Q whose density with respect to W_1 lies in $L^r(\mathsf{W}_1)$ for some $r > 1$: the self-normalized importance-weighted empirical law has the same limit $\sqrt{\Gamma_d}$, and the limit is $\sqrt{V\Gamma_d}$ for the corresponding class built on the law of $\int_0^{\cdot} \sigma_s\, dB_s$, for a common deterministic volatility σ with total variance V (Theorem 3(d)). For every target absolutely continuous with respect to W_1 and every proposal P with $Q \ll P$ the limit inferior is at least $\sqrt{\Gamma_d}$. The integrability exponents cannot be lowered to one: Proposition 2 gives a proposal with integrable but heavy-tailed weights for which the limit fails when $p > 2$, and Theorem 4 gives a target with integrable density for which it fails for every p.

4. *Dependence of the constant on the target* (Coste [3, §1.4, paragraph after Theorem 1.2]). Coste notes that his bounds do not match and that the asymptotics might be of the form $c/\sqrt{\log n}$ with c depending on the target in a non-trivial way, as for the finite-dimensional importance-sampling constant of Coste and Goldman [4], which is bracketed by multiples of an integral of the target density against a negative power of the proposal density. Theorem 3 and Corollary 2 give $c = \sqrt{\Gamma_d}$, the same for every target and every proposal, throughout the class of the previous item; this class contains every Girsanov target whose drift satisfies Novikov's condition with a margin, in particular all of Coste's targets. Outside it no finite constant need exist: Theorem 4 gives a target equivalent to W_1, with all moments, for which $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \to \infty$ for every order, every proposal and every sample, and Proposition 1 gives one with finite relative entropy for which this happens for every $p > 4$.

5. *The sharp mean asymptotic* (Boissard and Le Gouic [2, §1.2]). For a Gaussian measure μ on a separable Banach space with small-ball function $\psi(t) = -\log \mu(B(0,t))$ and empirical measure $\widehat{\mu}_n$ of n independent samples, Boissard and Le Gouic bound $\mathbb{E}W_2(\widehat{\mu}_n, \mu)$ by a constant times $\psi^{-1}(\log n)$ under assumptions on ψ, and they leave aside the question of finding c with $\mathbb{E}W_2(\widehat{\mu}_n, \mu) \sim c\,\psi^{-1}(\log n)$. For Wiener measure in the supremum norm, $\psi(t) = \lambda_d t^{-2} + O(1)$ as $t \downarrow 0$ [3, (3.2) and Appendix A.2], so $\psi^{-1}(\log n) \sim (\lambda_d/\log n)^{1/2}$, and Theorem 1 answers the question in this case with $c = (\Gamma_d/\lambda_d)^{1/2}$, for W_2 and for every finite order.

The answer to the second question has a direct quantitative meaning. Assigning each path to its nearest sample lets the sample masses vary. Corollary 1 shows that, for equal-weight samples and for moment-controlled importance weights, the L^p-mean distance $L_{n,p}$ to the nearest sample satisfies

$$\sqrt{\log n}\, L_{n,p} \longrightarrow \sqrt{\kappa_d},$$

so prescribing the sample masses adds exactly $(\sqrt{\Gamma_d} - \sqrt{\kappa_d})/\sqrt{\log n}$, a strictly positive amount, to the leading transport cost. The proof of Theorem 2 shows that equality $\Gamma_d = \kappa_d$ would force mutually compatible local tube conditionals whose joint law violates the Brownian bridge midpoint second moment.

Coste's Monte Carlo estimates of dual lower and primal upper bounds for the normalized scalar W_1 cost, about 1.60–1.62 and 1.70–1.72 for $64 \le n \le 512$ [3, §1.5.2], bracket the finite-n cost at these sample sizes. They are consistent with Theorem 2, which concerns the limit $\sqrt{\Gamma_1}$.

This third edition replaces the moment hypotheses of Theorem 3 for the almost-sure limit by one integrability exponent for each density, adds the universal lower bound and the Girsanov Corollary 2, and shows in Section 9 that integrability alone does not suffice.

1.1 The bridge entropy formulation

Fix $d \ge 1$ and use Euclidean length in $\mathbb{R}^d$. Let W_T be Wiener law on $[0,T]$ starting at zero, and β_T the corresponding Brownian bridge law pinned at both endpoints. Write $\|\cdot\|_\infty$ for the supremum path norm. If a pair law has marginals β_T, its mutual information is its relative entropy with respect to $\beta_T \otimes \beta_T$. Define, for $r > 0$,

$$C_T(r) = \inf \left\{ \mathrm{KL}(\pi \,\|\, \beta_T \otimes \beta_T) : \pi \in \Pi(\beta_T, \beta_T),\ \pi(\|u - v\|_\infty \le r) = 1 \right\}.$$

For the relaxed problem let

$$E_T = \left\{ (u,v) : \min_{|a|,|b| \le 1} \|u - v + (1 - t/T)a + (t/T)b\|_\infty \le 1 \right\}, \qquad R_T = \inf_{\substack{\pi \in \Pi(\beta_T,\beta_T) \\ \pi(E_T)=1}} \mathrm{KL}(\pi \,\|\, \beta_T \otimes \beta_T).$$

The affine parameters in E_T range over compact balls, so E_T is closed. In the empirical law $\widehat{\mathsf{W}}_n = n^{-1} \sum_{i=1}^n \delta_{W_i}$, the sequence (W_i) has iid law W_1.

Theorem 1. *The constant*

$$\Gamma_d = \sup_{T>0} \frac{R_T}{T} = \lim_{T \to \infty} \frac{R_T}{T} = \lim_{T \to \infty} \frac{C_T(1)}{T}$$

belongs to $(0, \infty)$. *For all sufficiently large* T *it satisfies*

$$0 \le \Gamma_d - \frac{R_T}{T} \le c_d T^{-1/9}.$$

For every finite $p \ge 1$,

$$\sqrt{\log n}\, W_p(\widehat{\mathsf{W}}_n, \mathsf{W}_1) \longrightarrow \sqrt{\Gamma_d}$$

almost surely and in L^q *of the sample for every finite* $q \ge 1$.

The nearest-sample asymptotic of Dereich and Lifshits [7, Theorems 5.1 and 6.1] gives $\Gamma_d \ge \kappa_d$, since the expectation of the W_1 cost dominates the mean nearest-sample distance; Theorem 2 makes the inequality strict. Lv [12, Theorem 2.1] proves the existence of the related quenched exponent for one-dimensional Brownian motion confined between two shifted copies of an independent Brownian path. He and Yuan [10, Theorems 2.1 and 2.2] establish supremum-norm quantization and empirical order bounds for a wider class of Itô processes. Dereich and Scheutzow [8] determine high-resolution quantization and entropy-coding constants of fractional Brownian motion, including Wiener paths, and their deterministic quantizers enter Section 2. For deterministic codebooks, Dereich [6] (see

[5, Theorem 1.2] for the d-dimensional statement) showed that the supremum-norm quantization constant of a diffusion with unit volatility, under his growth and regularity conditions, equals that of Wiener measure and so does not depend on the drift. Theorem 3 gives the corresponding statement for empirical and self-normalized importance-weighted laws, and Theorem 4 shows that for a merely integrable density even the optimal n-point quantization error can be of larger order than $(\log n)^{-1/2}$. Our use of couplings with prescribed Gaussian marginals is also related to output-constrained coding under additive distortion [16].

The empirical lower bound uses relaxed tubes because subtracting a block's endpoint chord introduces an affine correction. The upper bound uses strict tubes to concatenate bridges and control distance on the whole path. Sections 2–4 prove that the two tube problems have the same entropy rate. Sections 5 and 6 translate this equality into matching empirical bounds. Sections 7 and 8 use the exact-marginal couplings to establish the gap and the weighted limits, respectively. Section 9 conditions Schauder coefficients of Brownian paths to build the targets and the proposal for which the weighted limit fails.

2 Partitions and the relaxed bridge rate

We first give a quantitative cell partition fact. For any fixed $h > 0$ and finite q, each β_T has a countable measurable partition into cells of diameter at most h such that, if $J(x)$ is the cell containing x and $p_j = \beta_T(J = j)$,

$$\left\| \log \frac{1}{p_{J(x)}} \right\|_{L^q(\beta_T)} \le C_{d,h,q}(1+T). \tag{1}$$

This also holds for the law of the restriction of a bridge to any closed subinterval. To prove it, use the deterministic supremum Wiener quantizers of [8, Theorem 1.1 and Remark 1.2] in each coordinate. Brownian scaling, followed by subtracting the linear endpoint of each path and codepoint, gives bridge codebooks at rates $c(1+T)2^j$ whose L^v errors are $O_v(2^{-j/2})$ for any prescribed finite v. Subtracting the linear endpoint is a map of norm at most two. For restrictions, just restrict the codepoints. Assign each point to the first level where a codepoint is within $h/2$, using a fixed order within the level. The level tail is $O(2^{-jv/2})$ by Markov's inequality, and the number of cells through level j is at most $C\exp(C(1+T)2^j)$. The probability of being in a cell of mass below $e^{-u(1+T)}$ is at most the tail after level j plus $C\exp(C(1+T)2^j - u(1+T))$. For large u take 2^j comparable to a sufficiently small multiple of u. This gives (1) by choosing $v > 2q$. The unsuccessful set has measure zero and is irrelevant.

Choosing a common cell with its marginal probability and independent conditional bridges within that cell gives mutual information equal to the cell entropy. By scaling,

$$C_T(r) \le C_d(1 + T/r^2), \qquad C_T(r) = C_{T/r^2}(1). \tag{2}$$

The first inequality at $r = 1$ follows from (1), and the scaling identity is the Brownian time and space rescaling of both marginals.

For $L, M > 0$, an admissible relaxed pair on $[0, L+M]$ induces relaxed pairs of centered bridge residuals on each subinterval. Indeed subtract from each path its line joining the subinterval endpoints. The new affine correction has endpoint values equal to those of the original corrected difference on that subinterval. They have length at most one. For each single bridge law the two residuals are independent bridges of lengths L and M. Data processing and the relative entropy chain rule against the product reference therefore yield

$$R_{L+M} \ge R_L + R_M. \tag{3}$$

Since $0 \le R_T \le C_T(1)$ and R_T is uniformly linearly bounded for $T \ge 1$, (3) gives

$$\Gamma_d = \lim_{T\to\infty} R_T/T = \sup_{T>0} R_T/T < \infty.$$

One can see this by partitioning a long interval into copies of any fixed interval and one remainder. Also $R_1 > 0$: the original difference of an admissible pair has norm at most two, whereas the product bridge law assigns positive probability to a midpoint difference of length greater than two. The relative entropy of a law supported on an event is at least minus the logarithm of the reference probability of that event.

3 From relaxed tubes to strict tubes

We will show that, for large T,

$$C_T(1 + 3T^{-1/9}) \le R_T + C_d T^{8/9}. \tag{4}$$

Use $\delta = T^{-1/3}$ and $\eta = T^{-1/9}$. Take a relaxed coupling whose entropy is within one of R_T and retain the core $[\delta T, (1-\delta)T]$. The single-path core law, denoted by μ, is the product in natural coordinates of a centered core bridge and its two endpoints. The latter are Gaussian with covariance Λ_T on $\mathbb{R}^{2d}$. The eigenvalues of Λ_T are comparable to δT: the diagonal entries for each spatial coordinate are $\delta(1-\delta)T$, and the off-diagonal entry is $\delta^2 T$. The centered residual is independent of the endpoints. These statements follow directly from the Brownian bridge covariance or from its Markov decomposition.

For a relaxed pair choose affine labels (a, b) in a finite grid of the unit balls such that the corrected difference is bounded by $1+\eta$. A sufficiently fine deterministic grid and first acceptable choice make the labels measurable. Add half of this affine correction to the first core path and subtract half from the other. This only changes the core endpoints, by bounded shifts which may depend on the pair, while leaving the centered core bridge coordinates intact. To both endpoint vectors add the same independent Gaussian J with covariance $T^{4/9}I$. Also add independent endpoint jitters, one per path, with bounded density and each endpoint of length at most $\eta/2$. For example use a suitable rescaled uniform cube in $\mathbb{R}^{2d}$. The common Gaussian cancels in the difference. The resulting core pair has supremum difference at most $1 + 2\eta$.

The new single-path marginal is within $C_d T^{-2/9}$ in total variation of μ. First condition on the original pair. Translating the common Gaussian by the bounded adaptive shift changes its law in total variation by $O(T^{-2/9})$; this bound also applies after adding the independent small jitter. With the shift dropped, each marginal still has independent fine bridge coordinate and coarse endpoint vector. Adding J to that marginal endpoint Gaussian changes its law by $O(T^{4/9}/(\delta T)) = O(T^{-2/9})$ by the Gaussian entropy formula and Pinsker's inequality. The bounded independent jitter costs at most $C\eta/\sqrt{\delta T}$ by translation of this Gaussian. This proves the stated total variation bound, jointly with the centered core coordinate.

The entropy of the new core pair relative to $\mu \otimes \mu$ is at most the original relaxed entropy plus $O_d(\log T)$. To see this, disintegrate over the pair of centered core bridges. Their relative entropy is bounded by the original entropy by data processing. Conditionally on them the pair of endpoint vectors has bounded Lebesgue density at most $(C_d/\eta)^{4d}$, since the two independent jitters are added last. Against the product Gaussian endpoint reference, the conditional entropy is bounded using this density bound plus the expectation of minus the reference log density. The log determinant term is $O_d(\log T)$. The quadratic term, after averaging, is $O_d(1)$ because the original marginal endpoint covariance is Λ_T, the shifts are bounded, and $T^{4/9}/(\delta T)$ is bounded. This establishes the claimed entropy bound without requiring the adaptive shifts to be invertible.

Average the new core pair with its transpose, preserving the distance bound and the entropy bound. Its two marginals are now a common ν with $\|\nu - \mu\|_{\mathrm{TV}} \le C_d T^{-2/9}$. Multiply its density by $a(x)a(y)$, where $a = d(\mu \wedge \nu)/d\nu \in [0,1]$. The capped subprobability has equal marginals $\mu_0 \le a\nu = \mu \wedge \nu \le \mu$. The common deficit $\theta = \mu - \mu_0$ has mass $e \le 2\|\nu - \mu\|_{\mathrm{TV}}$, since $1 - a(x)a(y) \le (1 - a(x)) + (1 - a(y))$ and each marginal before capping is ν. Fill it by drawing two independent copies from θ conditional on the same cell of a diameter-one partition of the core law from (1), using θ's mass on that cell. This gives an exact coupling of both core marginals with distance at most $1 + 2\eta$.

For the entropy bound on this repair, write θ_j and μ_j for the two cell masses. The filling measure has subprobability entropy at most $\sum_j \theta_j \log(1/\theta_j)$, because $d\theta/d\mu \le 1$ and the density contributed on a cell pair is the product of two deficit densities divided by θ_j. By the nonnegativity of discrete relative entropy of the normalized deficit relative to the cell law, this is at most

$$e \log(1/e) + \sum_j \theta_j \log(1/\mu_j) \le e \log(1/e) + \sqrt{e}\, C_d(1 + T).$$

Capping contributes no more than the original core entropy plus 1/e, because the positive part of the density-log-density integral cannot increase and the negative part has magnitude at most 1/e under a probability reference. Adding the two subprobabilities costs at most $\log 2$ in addition, by convexity of relative entropy applied after normalizing their masses. Thus the repaired entropy is at most $R_T + O_d(T^{8/9})$.

Finally append left and right bridge collars of length δT. Given each core path, the conditional bridge pieces toward zero at the outer endpoints have independent centered bridge parts plus linear endpoint interpolation. Choose independent balanced couplings of these centered parts with distance at most η, using (2). Interpolating the paired core endpoints keeps their discrepancy at most $1 + 2\eta$. The collars thus raise the total radius to at most $1 + 3\eta$ and contribute entropy at most $2C_{\delta T}(\eta) + O(1)$ by the conditional entropy chain rule. This is $O_d(1 + \delta T/\eta^2) = O_d(T^{8/9})$, proving (4).

Since $R_T \le C_T(1)$, the limiting lower bound on $C_T(1)/T$ is Γ_d. For any large u, choose $T = u(1 + Ku^{-1/9})$ with a fixed sufficiently large K. Then $T/(1 + 3T^{-1/9})^2 \ge u$. The function $t \mapsto C_t(1)$ is nondecreasing, by scaling and monotonicity in the permitted radius. Apply (4) and (2) to bound $C_u(1) \le R_T + O_d(T^{8/9})$. Dividing by u proves the limiting upper bound, and hence the equality of the two long-bridge rates in Theorem 1.

4 Shared skeletons and a finite-block error bound

We use a general bridge concatenation bound. Divide a long bridge interval of length U into $M = \lceil U/T \rceil$ equal blocks. Round the path values at the knots to lattice cells of Euclidean diameter at most η. For a pair of bridges, choose one cell sequence with its exact skeleton distribution under β_U, then independently sample each skeleton conditional on that sequence. This gives exact skeleton marginals, and its mutual information is the entropy of the cell sequence. That entropy is at most

$$M c_d \log(2 + \sqrt{T}/\eta).$$

Indeed the sequence is encoded by successive integer cell increments (the initial cell is fixed). Their second moments are $O_d(1 + T/\eta^2)$ because each Gaussian bridge increment has coordinate variance at most the block length. The entropy bound follows, for example, by adding independent uniform noise in a unit cube to an integer vector and applying the Gaussian maximum differential entropy

bound at the resulting second moment. Entropy subadditivity applies even though the skeleton increments need not be independent.

Conditional on the two skeletons, couple independent centered bridge residuals on each block using near-minimizers for $C_T(r)$ scaled to the block length, which is at most T. Each marginal is still a full β_U bridge. Its conditional bridge decomposition, independent given its skeleton, is exact. By the entropy chain rule and the uniform bound η on the endpoint differences, this gives

$$C_U(r+\eta) \le M\{C_T(r) + c_d \log(2 + \sqrt{T}/\eta)\},$$

with an arbitrarily small added cost when an infimum is approximated. Divide by U, take $U \to \infty$, and apply Brownian scaling and the long-bridge limit. We get

$$\frac{\Gamma_d}{(r+\eta)^2} \le \frac{C_T(r) + c_d \log(2 + \sqrt{T}/\eta)}{T}. \tag{5}$$

Use $r = 1 + 3T^{-1/9}$ and $\eta = T^{-1/9}$ in (5), then apply (4). Since $R_T/T \le \Gamma_d$ and is uniformly bounded, it follows that $0 \le \Gamma_d - R_T/T \le c_d T^{-1/9}$.

5 The empirical lower bound

Fix a block length L with $R_L > 0$ and $a > 0$ with $a^2 < R_L/L$. We prove the almost-sure lower bound a for the normalized empirical distance. The argument takes both marginals of the empirical coupling into account.

For a radius ϵ split a unit Brownian path into $k = \lfloor 1/(L\epsilon^2) \rfloor$ complete blocks of length $L\epsilon^2$. In each block subtract the line joining the two endpoints, rescale time to $[0, L]$ and divide space by ϵ. The centered residuals are independent bridges of law β_L. For two paths of supremum distance at most ϵ, their paired bridge residuals lie in E_L on every block. The affine labels are the rescaled differences of the original block endpoints. For two independent Wiener paths the residual pairs are iid with law $\beta_L \otimes \beta_L$.

Call a path typical if the empirical measure of its k bridge residuals lies in a sufficiently small fixed open ball about β_L, for a metric inducing weak convergence. The reference mass of atypical paths is $O(e^{-ck})$ for some $c > 0$, by the closed-set bound in Sanov's theorem and compact entropy sublevel sets. For two independent Wiener paths, the probability that both are typical and have supremum distance at most ϵ is at most

$$\exp(-k(R_L - \xi)) \tag{6}$$

for all large k, with $\xi > 0$ arbitrarily prescribed by choosing the typical neighborhood sufficiently small. Indeed the joint empirical law of the residual pairs is supported on E_L and its two marginals lie in the closed neighborhoods. This is a closed condition on the joint law. The infimum of the product-reference relative entropy on these sets tends to R_L as the closed balls shrink, by compactness of the entropy sublevels, lower semicontinuity and the closedness of E_L. Sanov's theorem in the Polish path space then gives (6), with slack ξ.

For integer j, use the radius $\epsilon_j = a/\sqrt{\log 2^j}$, its fixed block extraction, and its typical set throughout the range $2^j \le n < 2^{j+1}$. The expected W_1 mass of typical paths covered within this radius by any typical sample among the first 2^{j+1} is bounded by $2^{j+1}\exp(-k_j(R_L - \xi))$, by the union bound and (6), where $k_j = \lfloor 1/(L\epsilon_j^2) \rfloor$. Choose ξ so this bound decays geometrically in j. By Markov's inequality and Borel–Cantelli the covered mass then tends to zero almost surely. The empirical mass of atypical codewords also tends to zero almost surely uniformly in the range: bound

their number by all atypical codewords among the first 2^{j+1}, divide by 2^j, and apply Markov and Borel–Cantelli to its exponentially decaying expectation. The target atypical mass tends to zero deterministically.

In any transport to the equal-weight empirical measure, the portion moving at distance at most ϵ_j has mass at most the sum of these three exceptional masses: atypical target paths, typical paths covered by typical codewords, and the prescribed mass of atypical codewords. Thus the distance exceeds ϵ_j on mass $1-o(1)$ in every coupling. This proves the asserted lower bound a for finite $p \geq 1$ and all sample sizes. Let a tend to $\sqrt{R_L/L}$, then use a countable sequence of fixed L approaching the supremum. This yields the lower bound $\sqrt{\Gamma_d}$.

6 The empirical upper bound

We construct a balanced coupling on a long Wiener interval whose information density is sufficiently concentrated for an iid empirical codebook. Fix $\eta \in (0,1)$ and $L \geq 1$. Take a bridge pair on $[0,L]$ of distance at most one and mutual information at most $C_L(1)+1$. Symmetrize it by averaging with its transpose. To control moments of information density, use a partition of β_L into cells of diameter at most η from (1). Keep the joint law of the two cell indices, then independently redraw each bridge from its reference law conditional on its own cell. By data processing this coupling has no larger mutual information. Its distance is at most $1+2\eta$, by comparison with an original pair in the same cells. This claim holds on every cell pair of positive original probability. If p_j is the cell mass and π_{jk} the joint cell law, its density on that cell pair is $\pi_{jk}/(p_j p_k)$. The positive part of its log under the joint law is at most $\log(1/p_j)+\log(1/p_k)$. The negative log tail under any joint law with density f is bounded by $\mathbb{P}_f(\log f \leq -t) \leq e^{-t}$. Thus the modified bridge information density has any prescribed finite moment, by choosing the partition with that moment in (1). The mean remains at most $C_L(1)+1$.

On $[0,T]$ with $T=ML$, use Wiener skeletons with knots of spacing L. Use regular lattice cubes of side $\eta/\sqrt{d}$ for the skeleton. As in the bridge concatenation construction, choose one rounded skeleton cell sequence and independently redraw both skeletons from their exact conditional laws. Given these skeletons use independent copies of the modified centered bridge pair on each block. The two complete marginals are exactly W_T and the coupled paths have distance at most $1+3\eta$. Write π_T for this pair law and i_T for its log density relative to $\mathsf{W}_T \otimes \mathsf{W}_T$. The density factors into its skeleton and independent bridge parts.

The skeleton information is the surprisal of the common cell sequence. It is at most

$$Mc_d \log(2+\sqrt{L}/\eta) + \frac{1}{2L}\sum_{i=1}^{M} \left(|X_{iL} - X_{(i-1)L}| + 2\eta\right)^2,$$

on a draw with first skeleton X. In fact the conditional probability of a next cell given the preceding cells is at least the cell volume times the Gaussian step density at the largest displacement between the previous and next cells. At step i this gives the lower bound

$$(\eta/\sqrt{d})^d (2\pi L)^{-d/2} \exp\left(-\frac{(|X_{iL} - X_{(i-1)L}| + 2\eta)^2}{2L}\right),$$

because the actual step and every possible step between these cells differ by at most twice the cell diameter. Multiplication along the sequence gives the bound. Since the first skeleton has the ordinary Wiener law, the squared steps divided by L have a chi-square sum. The upper bound on skeleton information is at most $Mc'_d \log(2+\sqrt{L}/\eta)$ outside an event with exponentially small

probability in M. The bridge information terms are iid with the mean and moments just established. The moment inequality for centered independent sums gives an arbitrarily high polynomial tail in M for their sum to exceed $M(C_L(1)+2)$. It follows that for each fixed

$$b > \frac{C_L(1) + 2 + c'_d \log(2 + \sqrt{L}/\eta)}{L} \tag{7}$$

the probability under π_T of $i_T > Tb$ decays faster than any prescribed fixed power of M, using the corresponding choice of fixed-block partition.

We give the required random-mixture estimate directly. For a pair coupling π with equal marginals μ and joint density f relative to $\mu \otimes \mu$, draw Y_i iid from μ. The conditional output mixture $\nu_n = n^{-1} \sum_i \pi(\cdot \mid Y_i)$ satisfies

$$\mathbb{E}\|\nu_n - \mu\|_{\mathrm{TV}} \le \pi(\log f > \tau) + \tfrac{1}{2}\sqrt{e^\tau/n}. \tag{8}$$

To prove this, truncate $f(x, Y_i)$ above e^τ. The integrated expected mass lost in the mixture is $\pi(\log f > \tau)$ and its expectation has the same mass deficit relative to μ. The integrated expected absolute centered fluctuation of the truncated mixture is at most $\sqrt{e^\tau/n}$ by Cauchy–Schwarz and the second-moment bound $f^2 1_{f \le e^\tau} \le e^\tau f 1_{f \le e^\tau}$. Dividing the total L^1 bound by two gives (8). This is the soft-covering estimate needed here.

Fix b as in (7) and $b' > b$. For sample size n take $M = \lfloor (\log n)/(Lb') \rfloor$ and $T = ML$ for all sufficiently large n. Rescale the unit Wiener paths to $[0, T]$, apply (8) with π_T and $\tau = Tb$, then rescale the conditional output mixture to unit paths again. Its expected total variation from W_1 is $O((\log n)^{-K})$ for any chosen fixed K, by the bridge moment choice above; the exponential term in (8) is polynomially small in n. The total variation, as a function of the empirical sample, changes by at most $1/n$ if one codeword is replaced. The independent bounded-differences inequality therefore shows that this total variation is $O((\log n)^{-K})$ almost surely for all n, with K any desired fixed order after choosing the construction's moment bound. Indeed deviations of size $(\log n)^{-K}$ have summable probabilities. Conditional kernels may be defined arbitrarily as probability kernels on null codewords, which do not affect the almost-sure distortion claim.

By construction the unit-time mixture couples to the equal-weight empirical measure at distance at most $(1 + 3\eta)/\sqrt{T}$. The Wasserstein distance from the mixture to W_1 is negligible at scale $(\log n)^{-1/2}$. To see this, couple their common measure identically, then couple the residual masses. Hölder bounds the pth power cost by

$$C_p \|\nu_n - \mathsf{W}_1\|_{\mathrm{TV}}^{1/2} \left(\nu_n(\|x\|_\infty^{2p}) + \mathsf{W}_1(\|x\|_\infty^{2p})\right)^{1/2}.$$

The moment factor is bounded almost surely by the hard distance coupling with the codewords and the strong law. For sample moments it is bounded in every fixed L^q by the same inequality and Gaussian moments, or by Jensen using $\mathbb{E}\nu_n = \mathsf{W}_1$. The high polynomial total-variation estimates therefore give the negligible residual also in every prescribed finite sample moment.

Since $\log n/T \to b'$, the triangle inequality gives an almost-sure upper limit at most $(1+3\eta)\sqrt{b'}$ for the normalized empirical error, and the same upper limit for its finite sample norms. Let b' and b decrease to the right side of (7), then let $L \to \infty$ using $C_L(1)/L \to \Gamma_d$, and finally let $\eta \downarrow 0$. Countable choices suffice for the almost-sure limits. The upper bound matches the empirical lower bound. Convergence in each finite sample moment follows from the almost-sure convergence and the bound in a strictly higher sample moment. This completes Theorem 1.

7 The gap from nearest-sample transport

Theorem 2. *For every $d \geq 1$ and Wiener measure on $\mathbb{R}^d$-valued paths in the Euclidean supremum metric, $\Gamma_d > \kappa_d$. Consequently the leading equal-weight transport constant $\sqrt{\Gamma_d}$ is strictly larger than the nearest-sample constant $\sqrt{\kappa_d}$.*

The boundary step of the proof uses the following property of Brownian paths near their starting point.

Lemma 1. *Let Z be a Brownian motion in $\mathbb{R}^d$ started at zero, with covariance $\sigma^2 t\, I_d$ for some $\sigma > 0$. Almost surely, for every $\varepsilon > 0$, the origin is an interior point of the convex hull of $Z([0,\varepsilon])$. In particular, no closed half-space $\{z : \langle z, e\rangle \leq 0\}$ with $|e| = 1$ contains $Z([0,\varepsilon])$.*

Proof. Let H_ε be the event that the origin is interior to the convex hull of $Z([0,\varepsilon])$. This hull is compact, so the origin is interior to it if and only if its support function $h(e) = \sup_{t\in\mathbb{Q}\cap[0,\varepsilon]}\langle Z(t), e\rangle$ is at least some rational $\delta > 0$ for every e in a countable dense subset of the unit sphere; hence H_ε is an event. Let $z_0, \dots, z_d$ be the vertices of a regular simplex centered at the origin. Choose $r > 0$ so small that every choice of points $w_k \in B(z_k, r)$ spans a simplex that still contains the origin in its interior. The vector $(Z(k/(d+1)))_{k=1}^{d+1}$ has a positive Gaussian density on $(\mathbb{R}^d)^{d+1}$, so with positive probability $Z(k/(d+1)) \in B(z_{k-1}, r)$ for every k. On this event H_1 holds, so $\mathbb{P}(H_1) > 0$. Brownian scaling shows that $\mathbb{P}(H_\varepsilon)$ does not depend on ε, and H_ε decreases as $\varepsilon \downarrow 0$. The intersection $\bigcap_k H_{1/k}$ is measurable with respect to the germ σ-field at time zero and has the same positive probability, so it has probability one by Blumenthal's zero–one law [13, §2.1]. A point interior to a convex hull lies in no closed half-space through it that contains the hull. □

Proof. Fix $d \geq 1$. Let $\mathcal{A}_T = \{\|x - y\|_\infty \leq 1\}$ on the product Wiener path space and set $p_T(x) = \mathsf{W}_T(\mathcal{A}_T(x)) > 0$. The joint law

$$Q_T(dx, dy) = \frac{1_{\mathcal{A}_T}(x,y)}{p_T(x)}\mathsf{W}_T(dx)\mathsf{W}_T(dy)$$

keeps the first marginal fixed and conditions the second path to lie in the tube around the first. For any finite-entropy pair P in the tube with first marginal W_T,

$$\mathrm{KL}(P\|Q_T) = \mathrm{KL}(P\|\mathsf{W}_T \otimes \mathsf{W}_T) - \mathbb{E}_{\mathsf{W}_T}[-\log p_T(X)]. \tag{9}$$

Brownian scaling and the expected small-ball limit in [7, proof of Theorem 6.1, equation (6.1)] give $T^{-1}\mathbb{E}[-\log p_T(X)] \to \kappa_d$.

Choose symmetric bridge pairs of radius one with entropy at most $C_T(1) + 1$. Independently choose a common endpoint cell, a cube of side $\eta = T^{-1/4}$, with the rounded $N(0, TI_d)$ law, then draw the two endpoints independently within that cell from their exact conditional laws. Adding the linear chords to the bridge pair gives exact Wiener marginals and radius at most $1 + \sqrt{d}\,\eta$. As in the skeleton construction, the added entropy is the entropy of the rounded endpoint, which is $O(d \log T)$ by its second-moment bound. Scale space by $(1 + \sqrt{d}\,\eta)^{-1}$ and time by $(1 + \sqrt{d}\,\eta)^{-2}$. Writing T for the resulting durations, we obtain symmetric radius-one Wiener couplings P_T with

$$\limsup_{T\to\infty} T^{-1}\,\mathrm{KL}(P_T\|\mathsf{W}_T \otimes \mathsf{W}_T) \leq \Gamma_d.$$

If $\Gamma_d = \kappa_d$, then (9) yields $\mathrm{KL}(P_T\|Q_T) = o(T)$. We now show that such a sublinear defect is impossible.

Fix $L > 8$. In a window $[s, s+L]$, denote by I, J the increments of X, Y from their respective left endpoint values. Write U, V for the pinned residuals after subtracting the two increment chords. Put $\omega = (\omega_0, \omega_1) \in (\mathbb{R}^d)^2$ for the values of $Y - X$ at the window endpoints and let ℓ_ω interpolate these values linearly. Writing $\beta = \beta_L$, define

$$A_\omega = \{(u, v) : \|v - u + \ell_\omega\|_\infty \le 1\}, \qquad K(\omega, u; dv) = \beta(dv \mid (u, v) \in A_\omega)$$

with an arbitrary probability-kernel version at a zero denominator.

Fix a partition of $[0, T]$ into consecutive length-L windows $w_1, \dots, w_m$ and at most two exterior intervals of length less than L. Let $\mathcal{G}$ be the σ-field generated by the whole X path and the values of Y at all partition points. The path Y is a measurable function of these values and of its pinned residuals $R_1, \dots, R_{m'}$ on the pieces of the partition; on the window w_k write ω^k, U^k and V^k for the window quantities defined above, so that V^k is one of the residuals R_j. All conditional laws below are regular versions, which exist on these Polish spaces.

Under $\mathsf{W}_T \otimes \mathsf{W}_T$ the residuals of Y are independent reference bridges, independent of $\mathcal{G}$. The density of Q_T is $1_{\mathcal{A}_T}/p_T(X)$, the factor $1/p_T(X)$ is $\mathcal{G}$-measurable, and the tube event is the intersection over the pieces of events each involving one residual of Y and $\mathcal{G}$-measurable data; on the window w_k it is $\|\ell_{\omega^k} + V^k - U^k\|_\infty \le 1$. Hence under Q_T, conditionally on $\mathcal{G}$, the residuals are independent and V^k has law $K(\omega^k, U^k; \cdot)$. The chain rule gives

$$\mathrm{KL}(P_T \| Q_T) = \mathrm{KL}(P_T|_{\mathcal{G}} \,\|\, Q_T|_{\mathcal{G}}) + \mathbb{E}_{P_T} \mathrm{KL}\left(P_T(R \in \cdot \mid \mathcal{G}) \,\middle\|\, Q_T(R \in \cdot \mid \mathcal{G})\right),$$

and the first term is nonnegative. For any law μ on a product space with marginals μ_k and any product law $\bigotimes \nu_k$, $\mathrm{KL}(\mu \| \bigotimes \nu_k) = \mathrm{KL}(\mu \| \bigotimes \mu_k) + \sum_k \mathrm{KL}(\mu_k \| \nu_k) \ge \sum_k \mathrm{KL}(\mu_k \| \nu_k)$. Dropping the exterior pieces therefore gives

$$\sum_{k=1}^{m} \mathbb{E}_{P_T} \mathrm{KL}\left(P_T(V^k \in \cdot \mid \mathcal{G}) \,\middle\|\, K(\omega^k, U^k; \cdot)\right) \le \mathrm{KL}(P_T \| Q_T).$$

Since (ω^k, U^k) is $\mathcal{G}$-measurable and the reference kernel depends on $\mathcal{G}$ only through it, conditional Jensen for the jointly convex relative entropy shows that replacing $\mathcal{G}$ by $\sigma(\omega^k, U^k)$ does not increase the kth term. The kth term then equals $\mathrm{KL}(\mathcal{L}(\omega^k, U^k, V^k) \,\|\, \mathcal{L}(\omega^k, U^k) \otimes K)$, where $\mathcal{L}$ denotes the law under P_T.

Let M_T be the local law of (ω, U, V) when s is uniform on $[0, T - L]$, independently of the pair. Averaging over stride-L partitions with offsets in $[0, L)$ covers all such starts once. Joint convexity of relative entropy then gives

$$\mathrm{KL}(M_T \| (M_T)_{\omega,U} K) \le \frac{L}{T - L} \mathrm{KL}(P_T \| Q_T) = o(1). \tag{10}$$

Exchange symmetry also gives the reverse conditional bound. If N_T denotes the law of (ω_0, I, J) under the same window choice, the increment pairs on a disjoint partition are independent under the product Wiener reference. Data processing, the same product-entropy bound, and offset averaging therefore give

$$\mathrm{KL}((N_T)_{I,J} \| \mathsf{W}_L \otimes \mathsf{W}_L) \le \frac{L}{T - L} \mathrm{KL}(P_T \| \mathsf{W}_T \otimes \mathsf{W}_T) = O_L(1). \tag{11}$$

The increment marginals of N_T are both exactly W_L. Since ω_0 lies in the closed unit ball, the local laws are tight. Take a weak subsequential limit N and denote by M its image in the coordinates (ω, U, V), with $\omega_1 = \omega_0 + J(L) - I(L)$. This map is continuous, and the limit is

supported on the closed tube. By lower semicontinuity in (11), the increment pair is absolutely continuous with respect to $\mathsf{W}_L \otimes \mathsf{W}_L$. Under the latter law, $Z = J - I$ is a Brownian motion with covariance $2tI_d$, and so is the reversed increment path $\widetilde{Z}(t) = Z(L-t) - Z(L)$. In the window the difference $Y - X$ equals $\omega_0 + Z(t)$ at time $s+t$ and $\omega_1 + \widetilde{Z}(t)$ at time $s+L-t$. If $|\omega_0| = 1$, the tube condition $|\omega_0 + Z(t)| \le 1$ gives $\langle Z(t), \omega_0 \rangle = (|\omega_0 + Z(t)|^2 - 1 - |Z(t)|^2)/2 \le 0$ for all $t \in [0, L]$, so $Z([0, L])$ lies in a closed half-space through the origin. The same holds for $\widetilde{Z}$ if $|\omega_1| = 1$. By Lemma 1 and absolute continuity, both events have N-probability zero. Consequently M gives full mass to offsets with $|\omega_0| < 1$ and $|\omega_1| < 1$, which we call interior.

For a fixed interior ω one has $\|\ell_\omega\|_\infty = \max(|\omega_0|, |\omega_1|) < 1$, since $\ell_\omega(t)$ is a convex combination of ω_0 and ω_1. For a residual u, the clipped probability is positive by full support of β on pinned continuous paths: the ball of radius $1 - \|\ell_\omega\|_\infty$ around $v = u$ lies in the tube. The distribution function $F(r) = \beta(\|V - u + \ell_\omega\|_\infty \le r)$ is log-concave in r. Indeed, on a finite time grid the sets $A_r = \{v : \max_t |v(t) - u(t) + \ell_\omega(t)| \le r\}$ satisfy $\lambda A_{r_1} + (1-\lambda) A_{r_2} \subset A_{\lambda r_1 + (1-\lambda) r_2}$, the grid marginal of β is a log-concave Gaussian law, and the Prékopa inequality [15] applies. Along nested dense grids, continuity of paths and continuity from above give F as a decreasing limit of log-concave functions. Since F is positive for all radii larger than $\|\ell_\omega\|_\infty$, it is continuous at one, so $\beta(\|V - u + \ell_\omega\|_\infty = 1) = 0$.

Let $(\omega_m, u_m) \to (\omega, u)$ in $(\mathbb{R}^d)^2 \times C$, with ω interior. For every v off the null set $\{\|v - u + \ell_\omega\|_\infty = 1\}$ the indicators $1\{\|v - u_m + \ell_{\omega_m}\|_\infty \le 1\}$ converge to $1\{\|v - u + \ell_\omega\|_\infty \le 1\}$, because $\|v - u_m + \ell_{\omega_m}\|_\infty \to \|v - u + \ell_\omega\|_\infty$. By dominated convergence the numerators converge in $L^1(\beta)$ and the denominators converge to a positive limit, so $K(\omega_m, u_m; \cdot) \to K(\omega, u; \cdot)$ in total variation. Thus the map $(\omega, u) \mapsto K(\omega, u; \cdot)$ is continuous in total variation at M-almost every point. For a bounded continuous g, the function $G(\omega, u) = \int g(\omega, u, v)\, K(\omega, u; dv)$ is bounded and continuous at those points: $|G(\omega_m, u_m) - G(\omega, u)|$ is at most $\sup|g|$ times the total variation distance plus $\int |g(\omega_m, u_m, v) - g(\omega, u, v)|\, K(\omega, u; dv)$, which tends to zero by dominated convergence. Along the chosen subsequence $(M_T)_{\omega,U} \to M_{\omega,U}$ weakly, as images of N_T under a continuous map. The mapping theorem [1, Theorem 2.7] therefore gives $\int G\, d(M_T)_{\omega,U} \to \int G\, dM_{\omega,U}$, that is, $(M_T)_{\omega,U} K \to M_{\omega,U} K$ weakly. Joint lower semicontinuity of relative entropy under weak convergence in (10) gives $\mathrm{KL}(M \| M_{\omega,U} K) = 0$, so $M = M_{\omega,U} K$, and exchange symmetry gives the reverse identity.

Fix an interior ω for which both conditional identities hold. Both conditional residual marginals are absolutely continuous with respect to β, as mixtures of clipped bridge laws. The conditional joint density with respect to $\beta \otimes \beta$ has both forms $a_\omega(u) 1_{A_\omega}(u, v) = b_\omega(v) 1_{A_\omega}(u, v)$. All pairs of residuals within distance $1 - \|\ell_\omega\|_\infty$ of each other belong to A_ω. Cover the pinned continuous path space by countably many open balls of diameter smaller than this threshold. On each ball times itself, Fubini's theorem makes a_ω and b_ω the same almost everywhere constant. Intersecting balls have positive intersection mass by full support, and the intersection graph of the cover is connected because the path space is connected. Thus the conditional pair is exactly $(\beta \otimes \beta)(\cdot \mid A_\omega)$.

For independent bridges, $S = (U + V)/2$ and $V - U$ are independent centered Gaussian bridges with covariance rates $\frac{1}{2} I_d$ and $2 I_d$, respectively. The tube event depends only on $V - U$, so the law of S, its independence from $V - U$ and its zero mean persist under this conditional pair. The tube also gives $\|V - U\|_\infty \le 2$, because $\|\ell_\omega\|_\infty \le 1$. Writing $U = S - (V - U)/2$, the cross term vanishes and

$$\mathbb{E}[|U(L/2)|^2 \mid \omega] = \frac{dL}{8} + \frac{1}{4} \mathbb{E}[|V(L/2) - U(L/2)|^2 \mid \omega] \le \frac{dL}{8} + 1.$$

But the unconditional U marginal is exactly β, with midpoint second moment $dL/4$. Averaging the displayed bound gives $dL/4 \le dL/8 + 1$, which contradicts $L > 8$. This proves strict inequality. $\square$

8 Universality under importance weighting

Let $C_0 = C_0([0,1];\mathbb{R}^d)$ be the space of continuous paths starting at the origin, with the supremum norm $\|\cdot\|_\infty$; then W_1 is a Borel probability measure on C_0. Let P and Q be Borel probability measures on C_0 with $Q \ll P$, and write $Z = dQ/dP$. For independent $Y_1, Y_2, \dots$ with law P let

$$D_n = \frac{1}{n}\sum_{i=1}^{n} Z(Y_i), \qquad \widehat{Q}_n = D_n^{-1}\frac{1}{n}\sum_{i=1}^{n} Z(Y_i)\,\delta_{Y_i}, \tag{12}$$

with $\widehat{Q}_n = \delta_0$ when $D_n = 0$. For $P = Q$, $\widehat{Q}_n$ is the equal-weight empirical law of n independent samples of Q. When $Q \ll \mathsf{W}_1$ we write $q_0 = dQ/d\mathsf{W}_1$, and p_0 denotes the density of the absolutely continuous part of P with respect to W_1.

Theorem 3. *Let $Q \ll \mathsf{W}_1$, let P be a Borel probability measure on C_0 with $Q \ll P$, and let $p \geq 1$.*

(a) Almost surely, $\liminf_{n\to\infty} \sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \geq \sqrt{\Gamma_d}$.

(b) If $q_0 \in L^r(\mathsf{W}_1)$ for some $r > 1$ and $Z \in L^s(P)$ for some $s > 1$, then $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \to \sqrt{\Gamma_d}$ almost surely.

(c) If in addition

$$\int (Z^v + \|y\|_\infty^v)\, dP(y) < \infty \qquad \text{for every } v < \infty, \tag{13}$$

the convergence in (b) also holds in L^q of the sample for every finite $q \geq 1$.

(d) The hypotheses of (b) and (c) hold when P is equivalent to W_1 and p_0, q_0 and $d\mathsf{W}_1/dP$ have finite moments of every order under W_1, W_1 and P, respectively. In particular (b) and (c) apply to any two bounded-drift laws with common unit volatility. More generally, let σ be a deterministic scalar volatility, bounded above and bounded away from zero, with total variance $V = \int_0^1 \sigma_s^2\, ds$, and let W^σ be the law of $\int_0^\cdot \sigma_s\, dB_s$ for a standard d-dimensional Brownian motion B. For Borel probability measures Q and P on C_0 with $Q \ll \mathsf{W}^\sigma$ (in place of $Q \ll \mathsf{W}_1$) and $Q \ll P$, if $dQ/d\mathsf{W}^\sigma \in L^r(\mathsf{W}^\sigma)$ and $Z \in L^s(P)$ for some $r, s > 1$, then (a) and (b) hold with $\sqrt{V\Gamma_d}$ in place of $\sqrt{\Gamma_d}$, and so does (c) under (13); this applies to any two bounded-drift laws with volatility σ.

Part (a) requires no integrability beyond $Q \ll \mathsf{W}_1$, and in (b) the proposal need not be equivalent to W_1: it may have a singular part, on which the weights vanish. For $P = Q$ the likelihood ratio is $Z = 1$ Q-almost everywhere, so the equal-weight empirical law of Q has the limit $\sqrt{\Gamma_d}$ almost surely and in every L^q whenever $q_0 \in L^r(\mathsf{W}_1)$ for some $r > 1$, by Lemma 2 below. Section 9 shows that the exponents $r > 1$ and $s > 1$ in (b) cannot be replaced by integrability.

The proof of (a) truncates both densities: targets with density at most Θ and samples with proposal density at most Θ interact only through Θ^2 times the Wiener collision probability of Section 5, and the other samples carry total weight close to $Q(q_0 > \Theta) + Q(p_0 > \Theta)$, because each sample carries only its own weight. The proof of (b) weights the couplings of Section 6 by $\min(q_0(x), q_0(y))$. One Hölder step against the polynomial decay of the information-density tail controls the soft-covering estimate, and the deficit of the weighted coupling has a density bounded by q_0 that tends to zero in $L^1(\mathsf{W}_1)$, hence in $L^r(\mathsf{W}_1)$; this makes the residual transport negligible without any rate. Almost-sure control uses Doob's maximal inequality on blocks of sample sizes on which the construction is fixed, and the von Bahr–Esseen inequality [17].

8.1 Tools

For a finite signed Borel measure μ on C_0 we write $\|\mu\|_{\mathrm{var}} = |\mu|(C_0)$ for its total variation norm. We write $\overline{\Phi}(t) = \int_t^\infty (2\pi)^{-1/2} e^{-u^2/2} du$ and, for $r \in (1,\infty)$, $r' = r/(r-1)$.

Definition 1 (Versions of the densities)**.** Let $Q \ll \mathsf{W}_1$ and $Q \ll P$. Let $P = P_a + P_\perp$ be the Lebesgue decomposition of P with respect to W_1, fix a W_1-null Borel set $\mathcal{S}$ with $P_\perp(C_0 \setminus \mathcal{S}) = 0$, and let p_0 be a version of $dP_a/d\mathsf{W}_1$ with $p_0 = 0$ on $\mathcal{S}$. Since $Q = Z \cdot P_a + Z \cdot P_\perp$ and $Q \ll \mathsf{W}_1$, the second term vanishes, so we choose $Z = 0$ on $\mathcal{S}$. Then $q_0 := Z p_0$ is a version of $dQ/d\mathsf{W}_1$, and $q_0 = 0$ on $\mathcal{S}$.

With these versions a sample $Y_i \in \mathcal{S}$ receives weight zero, and $q_0(y) > 0$ implies $p_0(y) > 0$.

Lemma 2 (Gaussian tails)**.** *For $u \ge 0$, $\mathsf{W}_1(\|x\|_\infty \ge u) \le 2d\, e^{-u^2/(2d)}$, and $\mathsf{W}_1(\|x\|_\infty \ge u) \le (1-2\theta)^{-d/2} e^{-\theta u^2}$ for every $\theta \in (0,1/2)$. Hence $\int \|x\|_\infty^v \, d\mathsf{W}_1 < \infty$ for every $v < \infty$, and $\int e^{a\|x\|_\infty^2} d\mathsf{W}_1 < \infty$ for $a < 1/2$. If $q_0 \in L^r(\mathsf{W}_1)$ for some $r > 1$, then $\int \|x\|_\infty^v \, dQ \le \|q_0\|_{L^r(\mathsf{W}_1)} (\int \|x\|_\infty^{vr'} d\mathsf{W}_1)^{1/r'} < \infty$ for every v.*

Proof. Let x^i be a coordinate of the path under W_1, a standard Brownian motion, and let $\theta > 0$. The process $e^{\theta x_t^i}$ is a nonnegative submartingale, so Doob's inequality [9, Theorem 4.4.2] on the dyadic times of level j gives $\mathbb{P}(\max_{t \in 2^{-j}\mathbb{Z} \cap [0,1]} x_t^i > u) \le e^{-\theta u} \mathbb{E} e^{\theta x_1^i} = e^{-\theta u + \theta^2/2}$. Letting $j \to \infty$, using continuity of paths, and taking $\theta = u$ gives $\mathbb{P}(\sup_{t \le 1} x_t^i > u) \le e^{-u^2/2}$, hence $\mathbb{P}(\sup_t |x_t^i| > u) \le 2e^{-u^2/2}$. If $\|x\|_\infty \ge u$ then some coordinate has supremum modulus at least $u/\sqrt{d}$; a union bound over the coordinates and continuity from above in u give the first claim. For the second, $|x_t|^2$ is a submartingale, and $e^{\theta |x_t|^2}$ is a nonnegative submartingale because $u \mapsto e^{\theta u}$ is convex and nondecreasing and $\mathbb{E} e^{\theta |x_t|^2} = (1-2\theta t)^{-d/2} < \infty$ for $t \le 1$. Doob's inequality on the dyadic times and continuity give $\mathbb{P}(\|x\|_\infty > u) \le e^{-\theta u^2} \mathbb{E} e^{\theta |x_1|^2} = (1-2\theta)^{-d/2} e^{-\theta u^2}$, and continuity from above in u gives the bound for $\|x\|_\infty \ge u$. The moment statements follow by integrating the tail bounds, choosing $\theta \in (a, 1/2)$ for the exponential moment, and the last one is Hölder's inequality. □

The next lemma collects what the proof of Theorem 1 established.

Lemma 3. *(i) Let $L > 0$ with $R_L > 0$ and let $\varpi \in (0, R_L)$. There are $c > 0$, $\epsilon_0 > 0$ and Borel sets $\mathcal{T}_\epsilon \subset C_0$, $0 < \epsilon \le \epsilon_0$, such that, with $k(\epsilon) = \lfloor 1/(L\epsilon^2) \rfloor$,*

$$\mathsf{W}_1(C_0 \setminus \mathcal{T}_\epsilon) \le e^{-c\,k(\epsilon)}, \qquad (\mathsf{W}_1 \otimes \mathsf{W}_1)(\{(x,y) \in \mathcal{T}_\epsilon^2 : \|x-y\|_\infty \le \epsilon\}) \le e^{-k(\epsilon)(R_L - \varpi)}.$$

(ii) Let $\eta \in (0,1)$ and $L \ge 1$, and let $b_*(L,\eta)$ be the right side of (7). Then $b_*(L,\eta) \to \Gamma_d$ as $L \to \infty$. For every $K > 0$ there are Borel probability measures π_{mL}, $m \in \mathbb{N}$, on $C_0([0,mL];\mathbb{R}^d)^2$ that are invariant under $(x,y) \mapsto (y,x)$, have both marginals equal to W_{mL}, satisfy $\pi_{mL}(\|x-y\|_\infty \le 1+3\eta) = 1$ and $\pi_{mL} \ll \mathsf{W}_{mL} \otimes \mathsf{W}_{mL}$, and satisfy, for every $b > b_*(L,\eta)$ and some $C < \infty$, $\pi_{mL}(i_{mL} > mLb) \le C m^{-K}$ for all m, where i_{mL} is the logarithm of the density.

Proof. Part (i) is the typical-set estimate of Section 5: $\mathcal{T}_\epsilon$ is the set of paths whose empirical law of block residuals lies in the neighbourhood chosen there, the first bound is the Sanov bound for atypical paths, and the second is (6), with ϖ in the role of the slack ξ there. Part (ii) is the construction of Section 6 with $M = m$ blocks, symmetrized as there, with the fixed-block partition chosen for the moment order K. Finally $b_*(L,\eta)$ is $C_L(1)/L$ plus terms of order $L^{-1}\log(2+\sqrt{L}/\eta)$, and $C_L(1)/L \to \Gamma_d$ by Theorem 1. □

Lemma 4 (Codes)**.** *For every $v \geq 1$ there is $C_v < \infty$ such that for every integer $N \geq 2$ there is a finite set $\mathcal{C} \subset C_0$ with $0 \in \mathcal{C}$, at most $N+1$ elements, all of norm at most $4\sqrt{d \log N}$, and $\int \operatorname{dist}(x, \mathcal{C})^v \, \mathsf{W}_1(dx) \leq C_v (\log N)^{-v/2}$.*

Proof. By [7, Theorem 6.1], $\epsilon^2 \ell_\epsilon \to \kappa_d$ almost surely, where $\ell_\epsilon = -\log \mathsf{W}_1(B(X, \epsilon))$ and $X \sim \mathsf{W}_1$, and the proof there shows $\epsilon^2 \mathbb{E}\ell_\epsilon \to \kappa_d$. The function $\bar{\ell}(\epsilon) = \mathbb{E}\ell_\epsilon$ is convex, one-to-one and onto [7, Remark 4.3], as required before [7, (5.1)]. Condition (5.1) there, $\bar{\ell}(\epsilon) \sim \ell_\epsilon$ in probability, holds because $\ell_\epsilon / \mathbb{E}\ell_\epsilon \to 1$ almost surely, and condition (5.2) there, $\bar{\ell}(\epsilon) \geq \tilde{\kappa}\, \bar{\ell}(2\epsilon)$ for small ϵ, holds for every $\tilde{\kappa} \in (1, 4)$ because $\epsilon^2 \bar{\ell}(\epsilon) \to \kappa_d$. Theorem 5.1 of [7] then states that for independent $X, Y_1, Y_2, \ldots$ with law W_1 and every $v > 0$,

$$\Big(\mathbb{E} \min_{i \leq e^\rho} \|X - Y_i\|_\infty^v \Big)^{1/v} \sim (\kappa_d / \rho)^{1/2} \qquad (\rho \to \infty).$$

Taking $\rho = \log N$, some realization $\mathcal{C}_0$ of $\{Y_1, \ldots, Y_N\}$ satisfies $\int \operatorname{dist}(x, \mathcal{C}_0)^v d\mathsf{W}_1 \leq C (\log N)^{-v/2}$ for all large N; for the finitely many remaining N take $\mathcal{C}_0 = \{0\}$ and enlarge C. Put $\varrho = 4\sqrt{d \log N}$ and $\mathcal{C} = \{0\} \cup \{z \in \mathcal{C}_0 : \|z\|_\infty \leq \varrho\}$. If $\|x\|_\infty \leq \varrho/2$ and the nearest point z of $\mathcal{C}_0$ to x has $\|z\|_\infty > \varrho$, then $\|x - z\|_\infty > \varrho/2 \geq \|x - 0\|_\infty$; hence $\operatorname{dist}(x, \mathcal{C}) \leq \operatorname{dist}(x, \mathcal{C}_0)$ when $\|x\|_\infty \leq \varrho/2$, and $\operatorname{dist}(x, \mathcal{C}) \leq \|x\|_\infty$ always. By Cauchy–Schwarz and Lemma 2,

$$\int_{\{\|x\|_\infty > \varrho/2\}} \|x\|_\infty^v \, d\mathsf{W}_1 \leq \Big(\int \|x\|_\infty^{2v} \, d\mathsf{W}_1 \Big)^{1/2} \big(2d\, e^{-\varrho^2/(8d)}\big)^{1/2} = O(N^{-1}),$$

which is of smaller order than $(\log N)^{-v/2}$. □

Lemma 5 (Transport through total variation)**.** *Let μ, ν be Borel probability measures on C_0 with $\int \|x\|_\infty^{2p} \, d(\mu + \nu) \leq \Upsilon$. Then $W_p(\mu, \nu)^p \leq 2^{p-1} \|\mu - \nu\|_{\mathrm{var}}^{1/2} \Upsilon^{1/2}$.*

Proof. Let $t = (\mu - \nu)_+(C_0)$. If $t = 0$ there is nothing to prove. Otherwise couple the common part $\mu \wedge \nu$ identically and the remainders by $(\mu - \nu)_+ \otimes (\nu - \mu)_+ / t$. Since $\|x - y\|_\infty^p \leq 2^{p-1}(\|x\|_\infty^p + \|y\|_\infty^p)$, the cost is at most $2^{p-1} \int \|x\|_\infty^p \, d|\mu - \nu|$, and Cauchy–Schwarz with $|\mu - \nu| \leq \mu + \nu$ gives the bound. □

Lemma 6 (Domination by the nearest-sample cost)**.** *If μ is a Borel probability measure with $\mu(F) = 1$ for a finite set $F \subset C_0$, then $\int \operatorname{dist}(x, F)^p \, Q(dx) \leq W_p(\mu, Q)^p$ for every Borel probability measure Q on C_0.*

Proof. Every coupling γ of μ and Q is carried by $F \times C_0$, and $\|x - y\|_\infty \geq \operatorname{dist}(x, F)$ for $y \in F$. Integrate against γ and minimize over γ. □

Lemma 7 (Maximal inequalities)**.** *Let $\zeta_1, \zeta_2, \ldots$ be independent and identically distributed random elements of a measurable space $\mathcal{Y}$, and let $\mathcal{F}_n$ be the σ-field generated by $\zeta_1, \ldots, \zeta_n$.*

(a) Let $(\mathcal{X}, \nu)$ be a σ-finite measure space and $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ jointly measurable with $\mathbb{E}g(\theta, \zeta_1) = 0$ for ν-almost every θ and $\int \mathbb{E}|g(\theta, \zeta_1)| \, \nu(d\theta) < \infty$. Put $U_n(\theta) = \sum_{i \leq n} g(\theta, \zeta_i)$. Then for $\lambda > 0$ and $N \in \mathbb{N}$, $\mathbb{P}(\max_{n \leq N} \int |U_n| \, d\nu \geq \lambda) \leq \lambda^{-1} \mathbb{E} \int |U_N| \, d\nu$.

(b) If $g : \mathcal{Y} \to \mathbb{R}$ satisfies $\mathbb{E}g(\zeta_1) = 0$ and $\mathbb{E}|g(\zeta_1)|^s < \infty$ for some $s \in [1, 2]$, then $U_n = \sum_{i \leq n} g(\zeta_i)$ satisfies $\mathbb{E}|U_N|^s \leq 2N \mathbb{E}|g(\zeta_1)|^s$ and $\mathbb{P}(\max_{n \leq N} |U_n| \geq \lambda) \leq 2N \mathbb{E}|g(\zeta_1)|^s \lambda^{-s}$.

Proof. (a) For ν-almost every θ, $(U_n(\theta))_n$ is a martingale, so $\mathbb{E}[|U_{n+1}(\theta)| \mid \mathcal{F}_n] \geq |U_n(\theta)|$. For $G \in \mathcal{F}_n$, Tonelli's theorem gives $\mathbb{E}[1_G \int |U_{n+1}| d\nu] = \int \mathbb{E}[1_G |U_{n+1}(\theta)|] \, \nu(d\theta) \geq \mathbb{E}[1_G \int |U_n| d\nu]$, so $\int |U_n| d\nu$ is a nonnegative integrable submartingale, and Doob's inequality [9, Theorem 4.4.2] applies. (b) The moment bound is the von Bahr–Esseen inequality [17] for independent centred summands. The process $|U_n|^s$ is a nonnegative submartingale, and Doob's inequality gives the second bound. □

8.2 Proof of the universal lower bound

Proof of Theorem 3(a). Use the versions of Definition 1. Fix a rational $a \in (0, \sqrt{\Gamma_d})$. By Theorem 1 there is $L > 0$ with $R_L/L > a^2$; fix it and choose $\varpi \in (0, R_L - La^2)$, so that

$$\vartheta := \frac{R_L - \varpi}{La^2} - 1 > 0.$$

Let $c, \epsilon_0, \mathcal{T}_\epsilon$ be as in Lemma 3(i). For $j \in \mathbb{N}$ put $\epsilon_j = a/\sqrt{j \log 2}$, $k_j = k(\epsilon_j) = \lfloor j \log 2/(La^2) \rfloor$ and $\mathcal{T}_j = \mathcal{T}_{\epsilon_j}$. For $2^j \le n < 2^{j+1}$ we have $a/\sqrt{\log n} \le \epsilon_j$.

Fix $\Theta \in \mathbb{N}$ and call a target path x *good* if $x \in \mathcal{T}_j$ and $q_0(x) \le \Theta$, and a sample path y *good* if $y \in \mathcal{T}_j$, $y \notin \mathcal{S}$ and $p_0(y) \le \Theta$. Let Cov_j be the set of good x for which some good Y_i with $i < 2^{j+1}$ satisfies $\|x - Y_i\|_\infty \le \epsilon_j$. By the union bound, Fubini's theorem and Lemma 3(i),

$$\begin{aligned}\mathbb{E}\, Q(\mathrm{Cov}_j) &\le 2^{j+1} \iint 1_{\{x,y \in \mathcal{T}_j,\ \|x-y\|_\infty \le \epsilon_j\}}\, q_0(x) 1_{\{q_0(x) \le \Theta\}}\, p_0(y) 1_{\{p_0(y) \le \Theta\}}\, \mathsf{W}_1(dx)\mathsf{W}_1(dy) \\ &\le 2^{j+1}\Theta^2 e^{-k_j(R_L - \varpi)}.\end{aligned}$$

Since $k_j \ge j \log 2/(La^2) - 1$, the right side is at most $2\Theta^2 e^{R_L} 2^{-j\vartheta}$. By Markov's inequality, $\mathbb{P}(Q(\mathrm{Cov}_j) > 2^{-j\vartheta/2}) \le 2\Theta^2 e^{R_L} 2^{-j\vartheta/2}$, which is summable, so $Q(\mathrm{Cov}_j) \to 0$ almost surely by the Borel–Cantelli lemma.

Next consider the weight of samples that are not good. Put $X_j = 2^{-j} \sum_{i < 2^{j+1}} Z(Y_i) 1\{Y_i \notin \mathcal{T}_j,\ q_0(Y_i) \le \Theta\}$. Since $\mathbb{E}_P[Z 1_B] = Q(B)$ for Borel B, $\mathbb{E}X_j \le 2Q(\{q_0 \le \Theta\} \setminus \mathcal{T}_j) \le 2\Theta \mathsf{W}_1(C_0 \setminus \mathcal{T}_j) \le 2\Theta e^{-ck_j}$, which decays geometrically in j; Markov's inequality and Borel–Cantelli give $X_j \to 0$ almost surely. By the strong law of large numbers, $D_n \to \mathbb{E}_P Z = 1$ and, for each fixed Borel set B, $n^{-1} \sum_{i \le n} Z(Y_i) 1_B(Y_i) \to Q(B)$ almost surely. Hence, almost surely, $D_n \ge 1/2$ for large n, and for $2^j \le n < 2^{j+1}$,

$$\widehat{Q}_n(\{y : y \text{ not good}\}) \le \frac{X_j}{D_n} + \widehat{Q}_n(\{q_0 > \Theta\}) + \widehat{Q}_n(\{p_0 > \Theta\}) \longrightarrow Q(q_0 > \Theta) + Q(p_0 > \Theta),$$

using that samples in $\mathcal{S}$ have weight zero. Similarly $Q(\{x \text{ not good}\}) \le \Theta \mathsf{W}_1(C_0 \setminus \mathcal{T}_j) + Q(q_0 > \Theta)$.

Let n be large, $2^j \le n < 2^{j+1}$, and let γ be any coupling of $\widehat{Q}_n$ and Q, with the first coordinate the sample path. If $\|x - y\|_\infty \le \epsilon_j$ with x good and y a good atom of $\widehat{Q}_n$, then $x \in \mathrm{Cov}_j$. Therefore

$$\gamma(\|x - y\|_\infty \le a/\sqrt{\log n}) \le Q(x \text{ not good}) + \widehat{Q}_n(y \text{ not good}) + Q(\mathrm{Cov}_j),$$

and almost surely the limit superior of the right side, uniformly over γ, is at most $2Q(q_0 > \Theta) + Q(p_0 > \Theta)$. Since p_0 and q_0 are finite W_1-almost everywhere and $Q \ll \mathsf{W}_1$, letting $\Theta \to \infty$ along the integers shows that almost surely $\sup_\gamma \gamma(\|x - y\|_\infty \le a/\sqrt{\log n}) \to 0$. As $W_p(\widehat{Q}_n, Q)^p \ge (a/\sqrt{\log n})^p \inf_\gamma \gamma(\|x - y\|_\infty > a/\sqrt{\log n})$, we obtain $\liminf_n \sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \ge a$ almost surely. Intersecting over rational $a < \sqrt{\Gamma_d}$ proves (a). □

8.3 Proof of the limit for L^r densities

Throughout this subsection the hypotheses of Theorem 3(b) hold, the versions of Definition 1 are used, and without loss of generality $s \le 2$. By Lemma 2, $\int \|x\|_\infty^v dQ < \infty$ for every v. In view of Theorem 3(a) it suffices to prove

$$\limsup_{n \to \infty} \sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \le (1 + 3\eta)\sqrt{b'} \quad \text{almost surely} \tag{14}$$

for fixed $\eta \in (0,1)$, $L \geq 1$ and $b' > b_*(L,\eta)$.

The construction. Fix η, L, b' as above and choose b, b_0 with $b_*(L,\eta) < b < b_0 < b'$. Fix $K > 2r'(p+1)$, put $K_1 = K/(2r') > p+1$, and let π_{mL} be the couplings of Lemma 3(ii) for this K. For $m \in \mathbb{N}$ let

$$T_m = mL, \qquad I_m = \{n \in \mathbb{N} : \ mLb' \leq \log n < (m+1)Lb'\}, \qquad N_m = \min I_m,$$

so that $N_m \geq e^{mLb'}$ and every $n \in I_m$ satisfies $n < e^{Lb'} N_m$. For $n \in I_m$ we write $m = m(n)$; then $m(n)$ is the number M of blocks used in Section 6, $m(n) \to \infty$ and $\log n / T_{m(n)} \to b'$. All deterministic objects below depend on n only through m.

Let S_T be the rescaling $(S_T x)(t) = T^{-1/2} x(Tt)$, which maps W_T to W_1 and multiplies supremum distances by $T^{-1/2}$. Let $\pi^{(m)} = (S_{T_m} \times S_{T_m})_{\#} \pi_{T_m}$, a symmetric coupling of W_1 with itself, and $r_m = (1+3\eta) T_m^{-1/2}$. Let f_m be a symmetric version of $d\pi^{(m)}/d(\mathsf{W}_1 \otimes \mathsf{W}_1)$ that vanishes when $\|x-y\|_\infty > r_m$. Rescaling preserves densities, so

$$\pi^{(m)}(f_m > e^{T_m b}) \leq C m^{-K}. \tag{15}$$

Let $a(x,y) = \min(q_0(x), q_0(y))$ and $\varsigma_m = a\,\pi^{(m)}$. Both marginals of ς_m equal the measure μ_m with density $\int f_m(x,y) a(x,y)\, \mathsf{W}_1(dy) \leq q_0(x)$, so $\mu_m \leq Q$. Let $\rho_m = Q - \mu_m$, with density $\bar{q}_m = d\rho_m / d\mathsf{W}_1$ and mass $\rho_m(C_0)$. Because the marginals of $\pi^{(m)}$ are exact, $q_0(x) - a(x,y) = (q_0(x) - q_0(y))_+$, and the two orderings average to

$$\rho_m(C_0) = \int (q_0(x) - a(x,y))\, d\pi^{(m)} = \tfrac{1}{2} \int |q_0(x) - q_0(y)|\, d\pi^{(m)}(x,y). \tag{16}$$

Let N' be the W_1-null set of y for which $\int f_m(x,y) \mathsf{W}_1(dx) \neq 1$. For $y \notin \mathcal{S} \cup N'$ with $p_0(y) > 0$ put $F_m(x,y) = f_m(x,y) a(x,y) / p_0(y)$, and put $F_m(\cdot, y) = 0$ otherwise. Let $\mathcal{K}_m(y, dx) = F_m(x,y) \mathsf{W}_1(dx)$, $w_m(y) = \mathcal{K}_m(y, C_0)$ and $\bar{w}_m(y) = Z(y) - w_m(y)$.

Let $\mathcal{C}_m$ be the code of Lemma 4 with $N = \lceil N_m^{1/4} \rceil$ and $v = pr'$, let $\Delta_{\mathcal{C}_m}(x) = \operatorname{dist}(x, \mathcal{C}_m)$, let $z_{m,1}, \dots, z_{m,J_m}$ be its distinct points, $J_m \leq 2N_m^{1/4}$, and let $B_{m,j}$ be the Voronoi cells: $x \in B_{m,j}$ if j is the least index with $\|x - z_{m,j}\|_\infty = \Delta_{\mathcal{C}_m}(x)$. Then $z_{m,j} \in B_{m,j}$ and $\|z_{m,j}\|_\infty \leq 4\sqrt{d \log N_m}$.

Lemma 8. *(a) $\rho_m(C_0) \to 0$ and $\|\bar{q}_m\|_{L^r(\mathsf{W}_1)} \to 0$ as $m \to \infty$.*

(b) $0 \leq w_m \leq Z$ and $\bar{w}_m \geq 0$ P-almost everywhere, and $\mathcal{K}_m(y, \cdot)$ is carried by $B(y, r_m)$.

(c) $\int P(dy)\, \mathcal{K}_m(y, \cdot) = \mu_m$, and for every Borel $g \geq 0$, $\int g\, w_m\, dP = \int g\, d\mu_m$ and $\int g\, \bar{w}_m\, dP = \int g\, d\rho_m$.

(d) $\int \Delta_{\mathcal{C}_m}^p\, d\rho_m = o(m^{-p/2})$.

(e) With $\tau_m(y) = \int F_m(x,y) 1\{F_m(x,y) > e^{T_m b_0}\}\, \mathsf{W}_1(dx)$, $\int \tau_m\, dP \leq C m^{-2K_1}$.

Proof. (a) Bounded Lipschitz functions are dense in $L^1(\mathsf{W}_1)$: Borel sets are approximated in measure by closed sets F, and 1_F is the decreasing limit of the Lipschitz functions $(1 - j \operatorname{dist}(\cdot, F))_+$. Given $\epsilon > 0$ choose a bounded Lipschitz g with $\int |q_0 - g|\, d\mathsf{W}_1 < \epsilon$. Since both marginals of $\pi^{(m)}$ are W_1 and $\|x-y\|_\infty \leq r_m$ on its support, (16) gives $\rho_m(C_0) \leq \epsilon + \frac{1}{2} \mathrm{Lip}(g)\, r_m$, so $\rho_m(C_0) \to 0$. Next, $0 \leq \bar{q}_m \leq q_0$ and $\int \bar{q}_m\, d\mathsf{W}_1 = \rho_m(C_0)$. For $\Theta > 0$, $\int \bar{q}_m^r d\mathsf{W}_1 \leq \int_{\{q_0 > \Theta\}} q_0^r d\mathsf{W}_1 + \Theta^{r-1} \rho_m(C_0)$, and letting $m \to \infty$ and then $\Theta \to \infty$ gives $\|\bar{q}_m\|_{L^r(\mathsf{W}_1)} \to 0$.

(b) For $y \notin \mathcal{S} \cup N'$ with $p_0(y) > 0$ we have $a(x,y) \leq q_0(y) = Z(y) p_0(y)$, so $F_m(x,y) \leq Z(y) f_m(x,y)$ and $w_m(y) \leq Z(y)$. The exceptional y have $w_m(y) = 0$. The set N' is W_1-null, hence

P_a-null, and $P_\perp(N'\backslash\mathcal{S}) = 0$, so these statements hold P-almost everywhere. The support statement holds because f_m vanishes off the tube.

(c) For $y \notin \mathcal{S}$ with $p_0(y) = 0$ we have $q_0(y) = 0$, hence $a(\cdot, y) = 0$. Therefore $\int P(dy)F_m(x,y) = \int f_m(x,y)a(x,y)\mathsf{W}_1(dy)$, which is the density of μ_m. By symmetry of f_m and a, the second marginal of ς_m is also μ_m, so $\int g\, w_m\, dP = \iint g(y)f_m(x,y)a(x,y)\mathsf{W}_1(dx)\mathsf{W}_1(dy) = \int g\, d\mu_m$. Since $\int gZ\, dP = \int g\, dQ$, the last identity follows by subtraction for bounded g, and then for every Borel $g \geq 0$ by monotone convergence, because $\bar{w}_m \geq 0$ P-almost everywhere.

(d) By Hölder's inequality and Lemma 4 with $v = pr'$,

$$\int \Delta_{\mathcal{C}_m}^p\, d\rho_m = \int \Delta_{\mathcal{C}_m}^p\, \bar{q}_m\, d\mathsf{W}_1 \leq \|\bar{q}_m\|_{L^r(\mathsf{W}_1)}\Big(\int \Delta_{\mathcal{C}_m}^{pr'} d\mathsf{W}_1\Big)^{1/r'} \leq C\|\bar{q}_m\|_{L^r(\mathsf{W}_1)}(\log N_m)^{-p/2},$$

because the code was built for $\lceil N_m^{1/4}\rceil$ points. Since $\log N_m \geq mLb'$, (a) gives (d).

(e) If $F_m(x,y) > e^{T_m b_0}$ then $f_m(x,y) > e^{T_m b}$ or $Z(y) > e^{T_m(b_0-b)}$, because $F_m \leq f_m Z(y)$. By (c), $\int \tau_m\, dP = \varsigma_m(F_m > e^{T_m b_0})$. First, $\varsigma_m(f_m > e^{T_m b}) \leq \int q_0(y)1\{f_m > e^{T_m b}\}\, d\pi^{(m)} \leq \|q_0\|_{L^r(\mathsf{W}_1)}\pi^{(m)}(f_m > e^{T_m b})^{1/r'} \leq Cm^{-K/r'}$ by Hölder's inequality, the second marginal W_1, and (15). Second, $\varsigma_m(Z(y) > t) \leq \int q_0(y)1\{Z(y) > t\}\mathsf{W}_1(dy) = Q(Z > t) = \int_{\{Z>t\}} Z\, dP \leq t^{1-s}\int Z^s dP$, which with $t = e^{T_m(b_0-b)}$ is exponentially small in m. Since $K/r' = 2K_1$, (e) follows. □

For $n \in I_m$ define the random measures

$$\widehat{\mathcal{K}}_n = \frac{1}{n}\sum_{i\leq n}\mathcal{K}_m(Y_i,\cdot), \qquad \mathcal{A}_n = \frac{1}{n}\sum_{i\leq n} w_m(Y_i)\delta_{Y_i}, \qquad \bar{\mathcal{A}}_n = \frac{1}{n}\sum_{i\leq n}\bar{w}_m(Y_i)\delta_{Y_i},$$

and the cell masses $\hat{c}_{n,j} = \bar{\mathcal{A}}_n(B_{m,j})$ and $c_{m,j} = \rho_m(B_{m,j})$. By Lemma 8(b), $\mathcal{A}_n + \bar{\mathcal{A}}_n = D_n\widehat{Q}_n$ when $D_n > 0$, and $\widehat{\mathcal{K}}_n(C_0) = \mathcal{A}_n(C_0)$.

Lemma 9. *Let $c_\star = (s-1)/(2(s+1))$. There are constants $M_\Omega, C < \infty$ such that the following hold. For $n \in I_m$ consider the events*

$$\Omega_1(n):\ \|\widehat{\mathcal{K}}_n - \mu_m\|_{\mathrm{var}} \leq Cm^{-K_1}, \qquad \Omega_2(n):\ |D_n - 1| \leq 2n^{-c_\star},$$
$$\Omega_3(n):\ \textstyle\sum_j |\hat{c}_{n,j} - c_{m,j}| \leq N_m^{-c_\star}, \qquad \Omega_4(n):\ \int \Delta_{\mathcal{C}_m}^p\, d\bar{\mathcal{A}}_n \leq e^{Lb'}\int \Delta_{\mathcal{C}_m}^p\, d\rho_m + N_m^{-c_\star},$$
$$\Omega_5(n):\ \tfrac{1}{n}\textstyle\sum_{i\leq n} Z(Y_i)(1 + \|Y_i\|_\infty^{2p}) \leq M_\Omega.$$

Almost surely all five events hold for all sufficiently large n. The probability that $\Omega_1(n)$, $\Omega_3(n)$ or $\Omega_4(n)$ fails for some $n \in I_m$ is at most Cm^{-K_1}. Under (13), $\mathbb{P}(\Omega_2(n)^c) + \mathbb{P}(\Omega_5(n)^c) \leq Cn^{-1}$.

Proof. Throughout, $n \in I_m$, so $N_m \leq n < e^{Lb'}N_m$, and N_m^+ denotes $\max I_m$.

Ω_1. Write $\bar{F}_m = F_m 1\{F_m \leq e^{T_m b_0}\}$ and $\bar{\mathcal{K}}_m(y, dx) = \bar{F}_m(x,y)\mathsf{W}_1(dx)$. Then

$$\|\widehat{\mathcal{K}}_n - \mu_m\|_{\mathrm{var}} \leq \frac{1}{n}\int |U_n(x)|\, \mathsf{W}_1(dx) + \frac{1}{n}\sum_{i\leq n}\tau_m(Y_i) + \int \tau_m\, dP,$$

where

$$U_n(x) = \sum_{i\leq n}\Big(\bar{F}_m(x,Y_i) - \int \bar{F}_m(x,y)\, P(dy)\Big).$$

By Cauchy–Schwarz, $\bar{F}_m^2 \leq e^{T_m b_0}F_m$, Lemma 8(c) and Jensen's inequality,

$$\mathbb{E}\int |U_N(x)|\, \mathsf{W}_1(dx) \leq \int \Big(Ne^{T_m b_0}\int F_m(x,y)P(dy)\Big)^{1/2}\mathsf{W}_1(dx) \leq \big(Ne^{T_m b_0}\big)^{1/2}.$$

Lemma 7(a) with $\lambda = N_m e^{-mL(b'-b_0)/4}$ and $N = N_m^+$ gives

$$\mathbb{P}\Big(\max_{n\in I_m} \tfrac{1}{n}\textstyle\int |U_n|\, d\mathsf{W}_1 > e^{-mL(b'-b_0)/4}\Big) \le \frac{(e^{Lb'}N_m e^{mLb_0})^{1/2}}{N_m e^{-mL(b'-b_0)/4}} \le e^{Lb'/2} e^{-mL(b'-b_0)/4},$$

using $N_m \ge e^{mLb'}$. Since $\tau_m \ge 0$, Markov's inequality and Lemma 8(e) give

$$\mathbb{P}\Big(\max_{n\in I_m} \frac{1}{n}\sum_{i\le n} \tau_m(Y_i) > m^{-K_1}\Big) \le \mathbb{P}\Big(\frac{1}{N_m}\sum_{i\le N_m^+} \tau_m(Y_i) > m^{-K_1}\Big) \le e^{Lb'} C m^{-2K_1} m^{K_1}.$$

Outside these two events every $n \in I_m$ satisfies $\Omega_1(n)$ with a suitable C.

Ω_3. For each j put $U_n^{(j)} = \sum_{i\le n}(\bar w_m(Y_i)1_{B_{m,j}}(Y_i) - c_{m,j})$, a sum of centred terms by Lemma 8(c). By Lemma 7(b), $|g - \mathbb{E}g|^s \le 2^{s-1}(|g|^s + |\mathbb{E}g|^s)$, Hölder's inequality over j and $\bar w_m \le Z$,

$$\mathbb{E}\sum_j |U_N^{(j)}| \le \sum_j \big(2^{s+1} N \textstyle\int \bar w_m^s 1_{B_{m,j}} dP\big)^{1/s} \le 4 N^{1/s} J_m^{1-1/s} \|Z\|_{L^s(P)}.$$

Lemma 7(a), applied with counting measure on the cells, gives for $\lambda > 0$

$$\mathbb{P}\Big(\max_{n\in I_m} \textstyle\sum_j |\hat c_{n,j} - c_{m,j}| > \lambda\Big) \le \frac{4(e^{Lb'}N_m)^{1/s}(2N_m^{1/4})^{1-1/s}\|Z\|_{L^s(P)}}{N_m\lambda} \le C\, N_m^{-\frac{3}{4}(1-\frac{1}{s})}\lambda^{-1}.$$

With $\lambda = N_m^{-c_\star}$ this is at most $CN_m^{-c_\star} \le Ce^{-c_\star mLb'}$, because $\frac{3}{4}(1-\frac{1}{s}) \ge 2c_\star$ for $s \le 2$.

Ω_4. Let $\Psi_m = \bar w_m \Delta_{\mathcal{C}_m}^p \ge 0$. Since $0 \in \mathcal{C}_m$, $\Psi_m(y) \le Z(y)\|y\|_\infty^p$. Put $s' = (1+s)/2 \in (1,s)$. By Hölder's inequality with exponents $(s-1)/(s'-1)$ and $(s-1)/(s-s')$,

$$\int \Psi_m^{s'} dP \le \int Z^{s'-1}\|y\|_\infty^{ps'}\, dQ \le \Big(\int Z^s dP\Big)^{\frac{s'-1}{s-1}} \Big(\int \|y\|_\infty^{ps'\frac{s-1}{s-s'}} dQ\Big)^{\frac{s-s'}{s-1}} =: G_\star < \infty,$$

uniformly in m. Let $\widetilde U_m = \sum_{i\le N_m^+}(\Psi_m(Y_i) - \int \Psi_m\, dP)$. By Lemma 7(b) and $|g-\mathbb{E}g|^{s'} \le 2^{s'}(g^{s'} + (\mathbb{E}g)^{s'})$, $\mathbb{P}(|\widetilde U_m| > N_m^{1-c_\star}) \le 2^{s'+2}e^{Lb'}N_m G_\star N_m^{-s'(1-c_\star)}$, and since $c_\star = (s'-1)/(2s')$ this is at most $CN_m^{-(s'-1)/2}$. Off this event, for every $n \in I_m$, Lemma 8(c) gives

$$\int \Delta_{\mathcal{C}_m}^p\, d\bar{\mathcal{A}}_n = \frac{1}{n}\sum_{i\le n}\Psi_m(Y_i) \le \frac{1}{N_m}\sum_{i\le N_m^+}\Psi_m(Y_i) \le e^{Lb'}\int \Delta_{\mathcal{C}_m}^p\, d\rho_m + N_m^{-c_\star}.$$

Ω_2 and Ω_5. For $2^{l-1} \le n < 2^l$, Lemma 7(b) with $g = Z-1$ and $\lambda = 2^{l(1-c_\star)}$ gives $\mathbb{P}(\max_{n<2^l}|n(D_n-1)| > 2^{l(1-c_\star)}) \le C2^{l(1-s(1-c_\star))}$, which is summable because $c_\star < 1 - 1/s$; off these events $|D_n - 1| \le 2^{1-l}2^{l(1-c_\star)} \le 2n^{-c_\star}$. This gives Ω_2 eventually. For Ω_5 take $M_\Omega = 2\int(1 + \|y\|_\infty^{2p})\, dQ$; the strong law of large numbers applies because $\int Z(1+\|y\|_\infty^{2p})\, dP = \int(1+\|y\|_\infty^{2p})\, dQ$. The events $\Omega_1(n)$, $\Omega_3(n)$ and $\Omega_4(n)$ hold for all $n \in I_m$ when m is large, by the Borel–Cantelli lemma, since their failure probabilities are summable in m. Under (13), $\mathbb{E}(D_n-1)^4 \le 3\mathbb{E}(Z-1)^4 n^{-2}$ and Chebyshev's inequality give $\mathbb{P}(\Omega_2(n)^c) \le Cn^{4c_\star-2} \le Cn^{-1}$, and $\mathbb{P}(\Omega_5(n)^c) \le Cn^{-1}$ by Chebyshev's inequality applied to $Z(1+\|Y\|_\infty^{2p})$, which has a finite second moment under (13). □

Lemma 10. *There are deterministic $\iota_m \to 0$ such that, for every sufficiently large $n \in I_m$ on the event $\Omega_1(n) \cap \cdots \cap \Omega_5(n)$,*

$$W_p(\widehat Q_n, Q) \le \big(r_m^p + \iota_m m^{-p/2}\big)^{1/p} + \iota_m m^{-1/2}.$$

Proof. On $\Omega_2(n)$ and for large n, $D_n \in [1/2, 2]$. On each cell put

$$\tilde\rho_n|_{B_{m,j}} = \frac{\min(c_{m,j}, \hat c_{n,j})}{c_{m,j}}\,\rho_m|_{B_{m,j}} + (\hat c_{n,j} - c_{m,j})_+\,\delta_{z_{m,j}},$$

with the first term read as zero when $c_{m,j} = 0$. Then $\tilde\rho_n(B_{m,j}) = \hat c_{n,j} = \bar{\mathcal{A}}_n(B_{m,j})$, $\|\tilde\rho_n - \rho_m\|_{\mathrm{var}} \le \sum_j |\hat c_{n,j} - c_{m,j}|$, and, since $\Delta_{\mathcal{C}_m}(z_{m,j}) = 0$, $\int \Delta^p_{\mathcal{C}_m}\, d\tilde\rho_n \le \int \Delta^p_{\mathcal{C}_m}\, d\rho_m$. Define the probability measure $Q'_n = D_n^{-1}(\widehat{\mathcal{K}}_n + \tilde\rho_n)$, whose mass is $D_n^{-1}(\mathcal{A}_n(C_0) + \bar{\mathcal{A}}_n(C_0)) = 1$.

From $\widehat{Q}_n$ to Q'_n. Couple $\mathcal{A}_n$ with $\widehat{\mathcal{K}}_n$ by $n^{-1}\sum_i \delta_{Y_i} \otimes \mathcal{K}_m(Y_i, \cdot)$, which by Lemma 8(b) moves mass at most r_m. On each cell with $\hat c_{n,j} > 0$ couple $\bar{\mathcal{A}}_n|_{B_{m,j}}$ with $\tilde\rho_n|_{B_{m,j}}$ by the normalized product measure; since $\|x - y\|^p_\infty \le 2^{p-1}(\|x - z_{m,j}\|^p_\infty + \|y - z_{m,j}\|^p_\infty)$ and $\|x - z_{m,j}\|_\infty = \Delta_{\mathcal{C}_m}(x)$ on $B_{m,j}$, its cost is at most $2^{p-1}(\int_{B_{m,j}} \Delta^p_{\mathcal{C}_m} d\bar{\mathcal{A}}_n + \int_{B_{m,j}} \Delta^p_{\mathcal{C}_m} d\tilde\rho_n)$. Dividing the sum of these couplings by D_n gives a coupling of $\widehat{Q}_n$ and Q'_n, so on $\Omega_4(n)$

$$W_p(\widehat{Q}_n, Q'_n)^p \le r_m^p + 2^p\Big((e^{Lb'} + 1)\int \Delta^p_{\mathcal{C}_m}\, d\rho_m + N_m^{-c_\star}\Big),$$

and the bracket is $o(m^{-p/2})$ by Lemma 8(d) and $N_m \ge e^{mLb'}$.

From Q'_n to Q. Since $Q = \mu_m + \rho_m$,

$$\|Q'_n - Q\|_{\mathrm{var}} \le |1 - D_n| + \|\widehat{\mathcal{K}}_n - \mu_m\|_{\mathrm{var}} + \|\tilde\rho_n - \rho_m\|_{\mathrm{var}} \le 2n^{-c_\star} + Cm^{-K_1} + N_m^{-c_\star}$$

on $\Omega_1(n) \cap \Omega_2(n) \cap \Omega_3(n)$. For the moments, Lemma 8(b) and $r_m \le 1$ give, on $\Omega_5(n)$,

$$\int \|x\|^{2p}_\infty d\widehat{\mathcal{K}}_n \le \frac{1}{n}\sum_i w_m(Y_i)(\|Y_i\|_\infty + 1)^{2p} \le 2^{2p-1} M_\Omega,$$

and $\int \|x\|^{2p}_\infty d\tilde\rho_n \le \int \|x\|^{2p}_\infty dQ + (4\sqrt{d\log N_m})^{2p} N_m^{-c_\star}$ on $\Omega_3(n)$. Hence $\int \|x\|^{2p}_\infty d(Q'_n + Q) \le M'_\Omega$ for a constant M'_Ω and all large m, and Lemma 5 gives $W_p(Q'_n, Q)^p \le 2^{p-1}(M'_\Omega)^{1/2}(3Cm^{-K_1})^{1/2} = o(m^{-p/2})$, because $K_1 > p$. The triangle inequality completes the proof. □

Proof of Theorem 3(b) and (c). By Lemmas 9 and 10, almost surely for all large n,

$$\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \le \sqrt{(m+1)Lb'}\,\Big[\big(r_m^p + \iota_m m^{-p/2}\big)^{1/p} + \iota_m m^{-1/2}\Big], \qquad m = m(n),$$

and the right side tends to $(1+3\eta)\sqrt{b'}$ because $r_m = (1+3\eta)(mL)^{-1/2}$. This is (14). Intersecting the almost-sure events over rational $b' > b_*(L, \eta)$, integers $L \ge 1$ and $\eta = 1/j$, and using $b_*(L, \eta) \to \Gamma_d$ (Lemma 3(ii)), we obtain $\limsup_n \sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \le \sqrt{\Gamma_d}$ almost surely. Together with (a), this proves (b).

Assume now (13), fix $q < \infty$ with $4q \ge p$ (larger q give stronger statements), and put $X_n = \sqrt{\log n}\, W_p(\widehat{Q}_n, Q)$. Since $X_n \to \sqrt{\Gamma_d}$ almost surely, convergence in L^q follows from $\sup_n \mathbb{E}X_n^{2q} < \infty$. Fix one parameter choice (η, L, b', K) with $K_1 \ge 2q$. By Lemmas 9 and 10, there is a constant X^* with $X_n \le X^*$ on $\Omega(n) = \Omega_1(n) \cap \cdots \cap \Omega_5(n)$ for all large n, and $\mathbb{P}(\Omega(n)^c) \le Cm(n)^{-K_1} + Cn^{-1} \le C(\log n)^{-2q}$. To bound $\mathbb{E}W_p(\widehat{Q}_n, Q)^{4q}$, note that

$$W_p(\widehat{Q}_n, Q) \le \Big(\int \|y\|^p_\infty\, d\widehat{Q}_n\Big)^{1/p} + \Big(\int \|x\|^p_\infty\, dQ\Big)^{1/p}.$$

On $\{D_n \ge 1/2\}$, $\int \|y\|^p_\infty d\widehat{Q}_n \le 2n^{-1}\sum_i Z(Y_i)\|Y_i\|^p_\infty$, whose moments of order $u \ge 1$ are at most $2^u \int (Z\|y\|^p_\infty)^u dP < \infty$ by convexity. On $\{D_n < 1/2\}$, $\int \|y\|^p_\infty d\widehat{Q}_n \le \max_{i \le n}\|Y_i\|^p_\infty$, and

$$\mathbb{E}\big[\max_{i \le n}\|Y_i\|^{pu}_\infty 1\{D_n < 1/2\}\big] \le \big(n\mathbb{E}\|Y_1\|^{2pu}_\infty\big)^{1/2}\big(16\,\mathbb{E}(D_n - 1)^4\big)^{1/2} = O(n^{-1/2}).$$

Hence $\sup_n \mathbb{E}W_p(\widehat{Q}_n, Q)^{4q} < \infty$, and

$$\mathbb{E}X_n^{2q} \le (X^*)^{2q} + (\log n)^q \big(\mathbb{E}W_p(\widehat{Q}_n, Q)^{4q}\big)^{1/2} \mathbb{P}(\Omega(n)^c)^{1/2}$$

is bounded. This proves (c). □

8.4 Moment-controlled tilts, drifts and volatility

Proof of Theorem 3(d). Suppose P is equivalent to W_1 and p_0, q_0 and $d\mathsf{W}_1/dP$ have finite moments of every order. Then $q_0 \in L^2(\mathsf{W}_1)$, and for every $v \ge 1$ Hölder's inequality gives

$$\int Z^v \, dP = \int q_0^v \Big(\frac{d\mathsf{W}_1}{dP}\Big)^v dP \le \Big(\int q_0^{2v} p_0 \, d\mathsf{W}_1\Big)^{1/2} \Big(\int \Big(\frac{d\mathsf{W}_1}{dP}\Big)^{2v} dP\Big)^{1/2} < \infty,$$

and $\int \|y\|_\infty^v \, dP = \int \|y\|_\infty^v p_0 \, d\mathsf{W}_1 < \infty$ by Hölder and Lemma 2. So the hypotheses of (b) and (c) hold.

For a bounded adapted drift φ, let B be Brownian motion under the law on an underlying filtered space, and set $X_t = B_t + \int_0^t \varphi_s \, ds$. The change of measure with density

$$H = \exp\Big(-\int_0^1 \varphi_s \cdot dB_s - \tfrac{1}{2}\int_0^1 |\varphi_s|^2 \, ds\Big)$$

makes X Brownian. For every real α, the exponential martingale with integrand $-\alpha\varphi$ gives $\mathbb{E}H^\alpha \le \exp(\frac{1}{2}|\alpha(\alpha-1)|\|\varphi\|_\infty^2)$. Both directions of the change of measure therefore have all positive moments under their respective reference laws. Passing to the path law of X replaces each density by a conditional expectation, preserving moments of order at least one by Jensen's inequality. This verifies the moment hypotheses for each bounded-drift law, including when the drift is adapted to a larger filtration, and for two such laws the likelihood ratio $dQ/dP = q_0/p_0$ inherits all moments by Hölder's inequality.

For the volatility statement let $A_\sigma(t) = \int_0^t \sigma_s^2 \, ds$, an increasing bijection of $[0,1]$ onto $[0,V]$, and define $(S^\sigma x)(w) = V^{-1/2} x(A_\sigma^{-1}(Vw))$ for $x \in C_0$ and $w \in [0,1]$. The map S^σ is a bijection of C_0 with continuous inverse $((S^\sigma)^{-1}y)(t) = V^{1/2} y(A_\sigma(t)/V)$, and $\|S^\sigma x - S^\sigma y\|_\infty = V^{-1/2}\|x-y\|_\infty$, because the time change is a bijection of the time intervals. Under W^σ the process $S^\sigma x$ is continuous, centred and Gaussian with covariance $V^{-1}A_\sigma(A_\sigma^{-1}(V(w\wedge u)))I = (w\wedge u)I$, so $S^\sigma_\# \mathsf{W}^\sigma = \mathsf{W}_1$. Couplings of two laws correspond under $S^\sigma \times S^\sigma$ to couplings of their images, so $W_p(S^\sigma_\#\mu, S^\sigma_\#\nu) = V^{-1/2} W_p(\mu,\nu)$. The laws $Q' = S^\sigma_\# Q$ and $P' = S^\sigma_\# P$ satisfy $Q' \ll \mathsf{W}_1$ with $dQ'/d\mathsf{W}_1 = (dQ/d\mathsf{W}^\sigma) \circ (S^\sigma)^{-1}$ and $dQ'/dP' = Z \circ (S^\sigma)^{-1}$, with the same L^r and L^s norms, and (13) for P implies it for P'. The samples $S^\sigma(Y_i)$ are independent with law P' and carry the same weights, so $S^\sigma_\# \widehat{Q}_n$ is the importance-weighted empirical law of Q', and $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) = \sqrt{V}\sqrt{\log n}\, W_p(S^\sigma_\#\widehat{Q}_n, Q')$. Parts (a)–(c) for (P', Q') give the statement with $\sqrt{V\Gamma_d}$. Finally, if $X_t = \int_0^t \varphi_s \, ds + \int_0^t \sigma_s \, dB_s$ with φ bounded, then $S^\sigma X$ has unit volatility and drift $V^{1/2}\varphi_{A_\sigma^{-1}(Vw)}/\sigma^2_{A_\sigma^{-1}(Vw)}$ at time w, which is bounded. So the image laws are bounded-drift laws with unit volatility, covered by the first part of (d). □

8.5 The nearest-sample limit and the cost of equal masses

Keep the setting of Theorem 3. Assigning each target path to its nearest sample, with whatever mass this produces, costs

$$L_{n,p} = \Big(\int \min_{1\le i\le n} \|x - Y_i\|_\infty^p \, Q(dx)\Big)^{1/p}.$$

The importance-weighted law $\widehat{Q}_n$ prescribes the sample masses. The next result identifies both leading constants and so measures exactly what the prescription costs.

Corollary 1. *Let P be equivalent to W_1, and let p_0, q_0 and $d\mathsf{W}_1/dP$ have finite moments of every order, as in Theorem 3(d). Then for every finite $p \geq 1$,*

$$\sqrt{\log n}\, L_{n,p} \longrightarrow \sqrt{\kappa_d} \quad \text{and} \quad \sqrt{\log n}\,(W_p(\widehat{Q}_n, Q) - L_{n,p}) \longrightarrow \sqrt{\Gamma_d} - \sqrt{\kappa_d} > 0$$

almost surely and in L^q of the sample for every finite $q \geq 1$. If instead P and Q are equivalent to W^σ for a volatility σ as in Theorem 3(d), with total variance V, and $dP/d\mathsf{W}^\sigma$, $dQ/d\mathsf{W}^\sigma$ and $d\mathsf{W}^\sigma/dP$ have finite moments of every order (for instance for two bounded-drift laws with volatility σ), the two limits are $\sqrt{V\kappa_d}$ and $\sqrt{V}(\sqrt{\Gamma_d} - \sqrt{\kappa_d})$.

Proof. Write $\Delta_n(x) = \min_{i \leq n} \|x - Y_i\|_\infty$, a jointly measurable function of x and the sample, non-increasing in n. Balls are closed supremum balls; open balls give the same limits below, by monotonicity in the radius.

Small balls under the proposal. Fix $q > 1$ with conjugate q'. For a ball B, Hölder's inequality gives

$$P(B) = \int_B p_0 \, d\mathsf{W}_1 \leq \|p_0\|_{L^q(\mathsf{W}_1)} \mathsf{W}_1(B)^{1/q'}, \qquad \mathsf{W}_1(B) = \int_B \frac{d\mathsf{W}_1}{dP}\, dP \leq \Big\| \frac{d\mathsf{W}_1}{dP} \Big\|_{L^q(P)} P(B)^{1/q'}.$$

By [7, Theorem 6.1], for W_1-almost every x, $-\varepsilon^2 \log \mathsf{W}_1(B(x,\varepsilon)) \to \kappa_d$ as $\varepsilon \downarrow 0$. The two bounds give, for such x,

$$\kappa_d/q' \leq \liminf_{\varepsilon \downarrow 0} -\varepsilon^2 \log P(B(x,\varepsilon)) \leq \limsup_{\varepsilon \downarrow 0} -\varepsilon^2 \log P(B(x,\varepsilon)) \leq q'\kappa_d.$$

Letting $q \to \infty$ along a sequence shows that $-\varepsilon^2 \log P(B(x,\varepsilon)) \to \kappa_d$ for W_1-almost every x, hence for Q-almost every x because $Q \ll \mathsf{W}_1$.

Pointwise limit. Fix such an x and write $\kappa = \kappa_d$. Let $0 < a < \sqrt{\kappa}$ and choose $\theta > 0$ with $(\kappa - \theta)/a^2 > 1$. For $2^j \leq n < 2^{j+1}$ one has $\Delta_n(x) \geq \Delta_{2^{j+1}}(x)$ and $a/\sqrt{\log n} \leq a_j := a/\sqrt{j \log 2}$. For large j, a union bound gives

$$\mathbb{P}(\Delta_{2^{j+1}}(x) \leq a_j) \leq 2^{j+1} P(B(x, a_j)) \leq 2^{j+1} \exp\big(- (\kappa - \theta) j \log 2/a^2\big),$$

which is summable in j. By Borel–Cantelli, $\sqrt{\log n}\,\Delta_n(x) > a$ for all large n almost surely. Now let $a > \sqrt{\kappa}$ and choose $\theta > 0$ with $\rho := (\kappa + \theta)/a^2 < 1$. For $2^j \leq n < 2^{j+1}$ one has $\Delta_n(x) \leq \Delta_{2^j}(x)$ and $a/\sqrt{\log n} \geq b_j := a/\sqrt{(j+1)\log 2}$. For large j,

$$\mathbb{P}(\Delta_{2^j}(x) > b_j) = \big(1 - P(B(x, b_j))\big)^{2^j} \leq \exp\big(- 2^j\, 2^{-(j+1)\rho}\big),$$

which is summable since $\rho < 1$. Hence $\sqrt{\log n}\,\Delta_n(x) < a$ for all large n almost surely. Taking rational a, $\sqrt{\log n}\,\Delta_n(x) \to \sqrt{\kappa}$ almost surely. By Fubini's theorem, almost surely this convergence holds for Q-almost every x.

Domination by the transport cost. Since $\mathbb{E}_P Z = 1$, the strong law gives $D_n > 0$ for all large n almost surely, and then Lemma 6 with $F = \{Y_1, \ldots, Y_n\}$ gives $\int \Delta_n^r \, dQ \leq W_r(\widehat{Q}_n, Q)^r$ for every $r \geq 1$; in particular $L_{n,p} \leq W_p(\widehat{Q}_n, Q)$.

Almost-sure limit. Put $g_n = (\log n)^{p/2} \Delta_n^p$. By the last display with $r = 2p$, $\int g_n^2 \, dQ \leq (\sqrt{\log n}\, W_{2p}(\widehat{Q}_n, Q))^{2p}$, which converges almost surely by Theorem 3 at order $2p$. Thus almost surely (g_n) is bounded in $L^2(Q)$ for large n, hence uniformly integrable, and $g_n \to \kappa^{p/2}$ Q-almost everywhere. Vitali's theorem gives $\int g_n \, dQ \to \kappa^{p/2}$, that is, $\sqrt{\log n}\, L_{n,p} \to \sqrt{\kappa}$ almost surely.

Sample moments. Always $L_{n,p} \le (\int \|x - Y_1\|_\infty^p \, Q(dx))^{1/p} \le m_p + \|Y_1\|_\infty$, where $m_p^p = \int \|x\|_\infty^p \, dQ$ is finite by Hölder with the moments of q_0. The variable $\|Y_1\|_\infty$ has all moments under P by Hölder with the moments of p_0. By the Marcinkiewicz–Zygmund inequality and the moments of Z, $\mathbb{P}(D_n = 0) \le \mathbb{E}|D_n - 1|^k = O(n^{-k/2})$ for every k. Splitting on $\{D_n > 0\}$ and using Hölder on the complement,

$$\mathbb{E}\big(\sqrt{\log n}\, L_{n,p}\big)^u \le \mathbb{E}\big(\sqrt{\log n}\, W_p(\widehat{Q}_n, Q)\big)^u + (\log n)^{u/2} \, \big\|m_p + \|Y_1\|_\infty\big\|_{L^{2u}}^u \, \mathbb{P}(D_n = 0)^{1/2}$$

for every $u \ge 1$. The first term is bounded by the convergence in L^u in Theorem 3 and the second tends to zero. So $\sqrt{\log n}\, L_{n,p}$ is bounded in every L^u, and with the almost-sure limit this gives convergence in every L^q.

Excess and volatility. The excess is the difference of two sequences that converge almost surely and in every L^q, by Theorem 3 and the preceding steps, and its limit is positive by Theorem 2. Under a common deterministic volatility, the map S^σ of the proof of Theorem 3(d) multiplies all supremum distances by $V^{-1/2}$ and carries (P, Q) to a pair satisfying the hypotheses above, with the same moments of the densities. Both $L_{n,p}$ and $W_p(\widehat{Q}_n, Q)$ therefore scale by $\sqrt{V}$. □

Remark. With $P = Q = \mathsf{W}_1$ the corollary recovers the Dereich–Lifshits nearest-sample limit in the supremum norm and strengthens it to the almost-sure form, conditional on the sample, and to all sample moments, together with the exact equal-mass excess $\sqrt{\Gamma_d} - \sqrt{\kappa_d}$.

8.6 Girsanov targets

The following consequence covers Coste's targets and unbounded drifts.

Corollary 2. *Let h be a progressively measurable $\mathbb{R}^d$-valued process on C_0 with*

$$\int \exp\Big(\big(\tfrac{1}{2} + \delta\big) \int_0^1 |h_t|^2 \, dt\Big) \, d\mathsf{W}_1 < \infty \qquad \text{for some } \delta > 0, \tag{17}$$

let $H = \exp(\int_0^1 h_t \cdot dx(t) - \frac{1}{2}\int_0^1 |h_t|^2 dt)$ and $Q = H \cdot \mathsf{W}_1$. Then $H \in L^r(\mathsf{W}_1)$ for some $r > 1$, and for every $p \ge 1$ the following hold.

- *(a) Let $P = \mathsf{W}_1$, so that (12) is Coste's estimator. Then $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q)$ converges to $\sqrt{\Gamma_d}$ almost surely. If (17) holds for every $\delta > 0$, for instance for bounded h, the convergence also holds in every L^q.*
- *(b) Let P be any proposal with $Q \ll P$ and $dQ/dP \in L^s(P)$ for some $s > 1$. Then $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q)$ converges to $\sqrt{\Gamma_d}$ almost surely.*

Condition (17) holds for every bounded progressive drift, and for every progressive drift with $|h_t(x)| \le c_1 + c_2 \sup_{u \le t} |x(u)|$ and $c_2 < 2^{-1/2}$.

Proof. Let $\mathcal{M}_t = \int_0^t h_u \cdot dx(u)$ under W_1, so $\langle \mathcal{M} \rangle_1 = \int_0^1 |h_t|^2 dt$. By (17) and Novikov's criterion [14], $\mathbb{E}_{\mathsf{W}_1} H = 1$, so Q is a probability measure. Let $q, r > 1$ and $q' = q/(q-1)$. Since

$$H^r = e^{r\mathcal{M}_1 - \frac{r}{2}\langle \mathcal{M} \rangle_1} = \Big(e^{qr\mathcal{M}_1 - \frac{q^2 r^2}{2}\langle \mathcal{M} \rangle_1}\Big)^{1/q} e^{\frac{qr^2 - r}{2}\langle \mathcal{M} \rangle_1},$$

Hölder's inequality with exponents q and q' gives

$$\mathbb{E}_{\mathsf{W}_1} H^r \le \Big(\mathbb{E}_{\mathsf{W}_1} e^{qr\mathcal{M}_1 - \frac{q^2 r^2}{2}\langle \mathcal{M} \rangle_1}\Big)^{1/q} \Big(\mathbb{E}_{\mathsf{W}_1} e^{\frac{q'(qr^2 - r)}{2}\langle \mathcal{M} \rangle_1}\Big)^{1/q'}.$$

The first factor is at most one, since the stochastic exponential of $qr\mathcal{M}$ is a nonnegative local martingale, hence a supermartingale. Take $q = 1+\varepsilon$ and $r = 1+\varepsilon^2$. Then

$$\frac{q'(qr^2-r)}{2} = \frac{(1+\varepsilon)(1+\varepsilon^2)(1+\varepsilon+\varepsilon^2)}{2} = \frac{1}{2} + \varepsilon + O(\varepsilon^2),$$

which is at most $\frac{1}{2}+\delta$ for small $\varepsilon > 0$, so $H \in L^{1+\varepsilon^2}(\mathsf{W}_1)$. Part (b) is Theorem 3(b) with this r. For (a), $P = \mathsf{W}_1$ gives $Z = H \in L^{1+\varepsilon^2}(P)$, so Theorem 3(b) applies; if (17) holds for every δ, the same Hölder estimate with arbitrary r and $q = 2$ shows $H \in L^v(\mathsf{W}_1)$ for every v, and (13) holds by Lemma 2, so Theorem 3(c) applies. For bounded h, (17) is immediate. Under the growth bound, $\int_0^1 |h_t|^2 dt \le 2c_1^2 + 2c_2^2\|x\|_\infty^2$, and $(1+2\delta)c_2^2 < 1/2$ for small δ, so (17) follows from Lemma 2. □

9 Limits of universality

This section shows that the integrability exponents in Theorem 3(b) cannot be replaced by integrability. We use the nearest-sample cost $L_{n,p}$ of Section 8.5. For every sample, $\widehat{Q}_n$ is carried by at most n points, namely by $\{Y_1, \dots, Y_n\}$ when $D_n > 0$ and by $\{0\}$ otherwise, and $L_{n,p} \le W_p(\widehat{Q}_n, Q)$ whenever $D_n > 0$, by Lemma 6.

Theorem 4. *There is a Borel probability measure Q on C_0 with the following properties.*

(i) Q and W_1 are equivalent, and $d\mathsf{W}_1/dQ \le 2$.

(ii) $\int \|x\|_\infty^v \, dQ < \infty$ for every $v < \infty$.

(iii) $dQ/d\mathsf{W}_1 \notin L^r(\mathsf{W}_1)$ for every $r > 1$.

(iv) For every $p \ge 1$,

$$\sqrt{\log n}\, \inf_\mu W_p(\mu, Q) \longrightarrow \infty \qquad (n \to \infty),$$

where the infimum is over all Borel probability measures μ carried by at most n points. More precisely, $\inf_\mu W_p(\mu, Q) \ge \frac{1}{16}\bigl(3/(2\pi^2 I(n))\bigr)^{1/p}$, where $I(n)$ is the least integer $i \ge 1$ with $15 \cdot 16^i \ge \log_2(2n)$, so that $I(n) \le 1 + \log_{16} \max\{1, \log_2(2n)/15\}$. Consequently, for every proposal P with $Q \ll P$, every realization of the sample and every $p \ge 1$, $\sqrt{\log n}\, L_{n,p} \to \infty$ and $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \to \infty$.

For this Q the normalized error therefore tends to infinity surely, for every proposal, and no finite constant c of Coste's form exists; for $P = \mathsf{W}_1$ this is Coste's estimator, with a target equivalent to W_1 whose density is integrable but in no L^r with $r > 1$. The proof bounds $\int \operatorname{dist}(x, F)^p \, Q(dx)$ from below for every set F of at most n points, so the optimal n-point quantization error of Q is of larger order than $(\log n)^{-1/2}$, in contrast with the drift independence of [6], [5, Theorem 1.2]; it also shows that the conclusion of Corollary 1 fails for this Q. The same construction, with smaller mixture weights, keeps the relative entropy finite.

Proposition 1. *There is a Borel probability measure Q on C_0 with properties (i) and (ii) of Theorem 4 and $\mathrm{KL}(Q \,\|\, \mathsf{W}_1) < \infty$, such that for every $p > 4$ the conclusions of Theorem 4(iv) hold: $\sqrt{\log n}\, \inf_\mu W_p(\mu, Q) \to \infty$ over measures μ carried by at most n points, and $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \to \infty$ for every proposal P with $Q \ll P$ and every realization of the sample.*

By the Girsanov theory under finite entropy [11, Theorems 2.1 and 2.3] (Wiener measure satisfies the uniqueness condition (U) there, and both laws start at the origin), a law Q with $\mathrm{KL}(Q\,\|\,\mathsf{W}_1) < \infty$ is the law of a Brownian motion with an adapted drift h of finite energy, $\mathbb{E}_Q\int_0^1 |h_t|^2\,dt = 2\,\mathrm{KL}(Q\,\|\,\mathsf{W}_1)$. Proposition 1 therefore shows that finite drift energy does not preserve the constant for $p > 4$, while Corollary 2 shows that Novikov's exponential condition with a margin does. The integrability of the weights is needed as well.

Proposition 2. *There is a proposal P equivalent to W_1, with $dP/d\mathsf{W}_1 \le 4$, such that for the Brownian target $Q = \mathsf{W}_1$ the likelihood ratio $Z = d\mathsf{W}_1/dP$ lies in no $L^s(P)$ with $s > 1$, and for every $p > 2$, almost surely $\limsup_n \sqrt{\log n}\, W_p(\widehat{Q}_n, \mathsf{W}_1) = \infty$.*

Remark. Absolute continuity cannot be dropped either. If $\sigma > 0$ and Q is the law of σx under W_1, then Q and W_1 are mutually singular for $\sigma \neq 1$, the map $x \mapsto \sigma x$ carries independent Brownian samples to independent Q-samples and multiplies all supremum distances by σ, and Theorem 1 gives $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \to \sigma\sqrt{\Gamma_d}$ for $P = Q$.

9.1 Wiener laws conditioned on large Schauder coefficients

Definition 2. For $x \in C_0$ with first coordinate x^1, integers $l \ge 0$ and $0 \le m < 2^l$, let

$$\xi_{l,m}(x) = 2^{l/2+1}\Big(x^1\big(\tfrac{2m+1}{2^{l+1}}\big) - \tfrac{1}{2}x^1\big(\tfrac{m}{2^l}\big) - \tfrac{1}{2}x^1\big(\tfrac{m+1}{2^l}\big)\Big),$$

and let $s_{l,m}(t) = \int_0^t \chi_{l,m}(u)\,du$, where $\chi_{l,m} = 2^{l/2}$ on $[m2^{-l}, (2m+1)2^{-l-1})$, $\chi_{l,m} = -2^{l/2}$ on $[(2m+1)2^{-l-1}, (m+1)2^{-l})$ and $\chi_{l,m} = 0$ elsewhere. The function $s_{l,m}$ is a tent of height $2^{-l/2-1}$ supported on $[m2^{-l}, (m+1)2^{-l}]$.

Lemma 11. *(a) $|\xi_{l,m}(x) - \xi_{l,m}(y)| \le 2^{l/2+2}\|x - y\|_\infty$ for all $x, y \in C_0$.*

(b) Under W_1 the family $(\xi_{l,m})_{l\ge 0,\, 0\le m<2^l}$ is independent and standard normal.

(c) Let Λ be a finite set of levels, $x_\Lambda = \sum_{l\in\Lambda}\sum_m \xi_{l,m}(x)s_{l,m}e_1$ and $x' = x - x_\Lambda$. Under W_1, x' is independent of $(\xi_{l,m})_{l\in\Lambda,m}$, and $\int\|x'\|_\infty^v\,d\mathsf{W}_1 \le \int\|x\|_\infty^v\,d\mathsf{W}_1$ for $v \ge 1$.

Proof. (a) The bracket in the definition changes by at most $2\|x - y\|_\infty$. (b) Under W_1, $\xi_{l,m} = \int_0^1 \chi_{l,m}\,dx^1$ is a Wiener integral, and the Haar functions $\chi_{l,m}$ are orthonormal in $L^2[0,1]$; a centred Gaussian family with identity covariance is independent and standard normal. (c) Since $s_{l,m}(t) = \langle 1_{[0,t]}, \chi_{l,m}\rangle_{L^2}$, the first coordinate of x' is $x'^1(t) = \int_0^1 \Pi^\perp 1_{[0,t]}\,dx^1$, where $\Pi^\perp$ is the orthogonal projection onto the complement of the span of $\{\chi_{l,m} : l \in \Lambda\}$; the other coordinates are unchanged and independent of x^1. Hence x' and $(\xi_{l,m})_{l\in\Lambda}$ are jointly Gaussian and uncorrelated, so independent; continuity of x' makes this an independence statement for random elements of C_0. As x_Λ is centred and independent of x', $x'(t) = \mathbb{E}[x(t) \mid x']$, so $|x'(t)| \le \mathbb{E}[\|x\|_\infty \mid x']$ for every rational t, and conditional Jensen gives $\|x'\|_\infty^v \le \mathbb{E}[\|x\|_\infty^v \mid x']$. □

Definition 3. For an integer $k \ge 4$ let $\Lambda_k = \{k, k+1, k+2, k+3\}$, $A_k = 2^{k/2}$,

$$\mathcal{E}_k = \big\{x \in C_0 :\ |\xi_{l,m}(x)| \ge A_k \text{ for all } l \in \Lambda_k,\ 0 \le m < 2^l\big\},$$

and $\mathsf{R}_k = \mathsf{W}_1(\cdot \mid \mathcal{E}_k)$. The event $\mathcal{E}_k$ constrains $2^k(1 + 2 + 4 + 8) = 15 \cdot 2^k$ coefficients.

Lemma 12. *Let $k \ge 4$.*

(a) $\mathsf{W}_1(\mathcal{E}_k) = (2\overline{\Phi}(A_k))^{15\cdot 2^k}$, and $\frac{15}{2}\,4^k \le \log(1/\mathsf{W}_1(\mathcal{E}_k)) \le 15 \cdot 4^k$.

(b) $\mathsf{R}_k(B(y,1/8)) \le 2^{-15\cdot 2^k}$ *for every* $y \in \mathcal{E}_k$.

(c) $\mathsf{R}_k(B(y,1/16)) \le 2^{-15\cdot 2^k}$ *for every* $y \in C_0$.

(d) For every integer $v \ge 1$, $\int \|x\|_\infty^v \, d\mathsf{R}_k \le 3^{v-1}\big(\int \|x\|_\infty^v d\mathsf{W}_1 + 1.29^v(1+15\,v!)\big)$.

Proof. For $t \ge 0$ we use the two elementary bounds

$$\overline{\Phi}(t) = e^{-t^2/2}\int_0^\infty \frac{e^{-u^2/2-tu}}{\sqrt{2\pi}}\,du \le \frac{1}{2}e^{-t^2/2}, \qquad \overline{\Phi}(t) \ge \int_t^{t+1}\frac{e^{-u^2/2}}{\sqrt{2\pi}}\,du \ge \frac{e^{-(t+1)^2/2}}{\sqrt{2\pi}}.$$

(a) The product formula is Lemma 11(b). The upper bound $2\overline{\Phi}(A_k) \le e^{-A_k^2/2}$ gives the lower bound on $\log(1/\mathsf{W}_1(\mathcal{E}_k))$. The lower bound on $\overline{\Phi}$ gives $\log(1/(2\overline{\Phi}(t))) \le (t+1)^2/2 + \frac{1}{4} \le t^2$ for $t \ge 3$, and $A_k \ge 4$.

(b) Let $x, y \in \mathcal{E}_k$ with $\|x-y\|_\infty \le 1/8$ and $l \in \Lambda_k$. By Lemma 11(a), $|\xi_{l,m}(x) - \xi_{l,m}(y)| \le 2^{l/2-1} \le 2^{(k+3)/2-1} = \sqrt{2}\,A_k < 2A_k$. Both coefficients have modulus at least A_k, so they have the same sign. Under W_1 the coefficients are independent and symmetric and $\mathcal{E}_k$ depends only on their moduli, so under R_k the $15\cdot 2^k$ signs $\operatorname{sgn}\xi_{l,m}$, $l \in \Lambda_k$, are independent fair signs. Hence $\mathsf{R}_k(B(y,1/8)) \le \mathsf{R}_k(\operatorname{sgn}\xi_{l,m} = \operatorname{sgn}\xi_{l,m}(y)\ \forall l \in \Lambda_k, m) = 2^{-15\cdot 2^k}$.

(c) Since $\mathsf{R}_k(\mathcal{E}_k) = 1$, the claim is trivial when $B(y,1/16)\cap\mathcal{E}_k = \emptyset$. Otherwise choose y' in this intersection. Every $x \in B(y,1/16)$ satisfies $\|x-y'\|_\infty \le \|x-y\|_\infty + \|y-y'\|_\infty \le 1/8$, so $B(y,1/16) \subset B(y',1/8)$, and (b) applied to $y' \in \mathcal{E}_k$ gives the bound.

(d) Write $x = x' + x_{\Lambda_k}$ as in Lemma 11(c). Since $\mathcal{E}_k$ depends only on $(\xi_{l,m})_{l\in\Lambda_k}$, the law of x' under R_k is its law under W_1. Tents of one level have disjoint supports, so $\|x_{\Lambda_k}\|_\infty \le \sum_{l\in\Lambda_k} 2^{-l/2-1}\max_m|\xi_{l,m}|$. Under R_k, the overshoots $|\xi_{l,m}|-A_k$ are independent with $\mathbb{P}(|\xi_{l,m}|-A_k > u) = \overline{\Phi}(A_k+u)/\overline{\Phi}(A_k) \le e^{-A_k u}$, because $\overline{\Phi}(A+u) = \int_A^\infty (2\pi)^{-1/2}e^{-(w+u)^2/2}dw \le e^{-Au}\overline{\Phi}(A)$. With $\Xi = \max_{l,m}(|\xi_{l,m}|-A_k)$ we get $\mathbb{E}\,\Xi^v \le 15\cdot 2^k v! A_k^{-v}$, and $\|x_{\Lambda_k}\|_\infty \le \sum_{j=0}^3 2^{-j/2-1}(1+2^{-k/2}\Xi) \le 1.29(1+2^{-k/2}\Xi)$. Since $\mathbb{E}(2^{-k/2}\Xi)^v \le 15\,v!\,2^{k(1-v)} \le 15\,v!$, the bound follows from $(\alpha+\beta+\gamma)^v \le 3^{v-1}(\alpha^v+\beta^v+\gamma^v)$. □

9.2 Proofs of Theorem 4 and Proposition 1

Lemma 13. *Let* $(c_i)_{i\ge1}$ *be positive with* $\sum_i c_i = 1$, *let* $k_i = 4i$, *and let*

$$Q = \tfrac{1}{2}\mathsf{W}_1 + \tfrac{1}{2}\sum_{i\ge1} c_i \mathsf{R}_{k_i}.$$

Then Q *satisfies (i) and (ii) of Theorem 4, and* $\mathrm{KL}(Q\,\|\,\mathsf{W}_1) \le \frac{15}{2}\sum_i c_i 4^{k_i}$. *For* $n \ge 1$ *let* $I(n)$ *be the least integer* $i \ge 1$ *with* $15\cdot 16^i \ge \log_2(2n)$. *Then* $16^{I(n)} \le 16\max\{1, \log_2(2n)/15\}$, *and for every finite set* $F \subset C_0$ *with at most* n *points and every* $p \ge 1$,

$$\int \operatorname{dist}(x,F)^p\, Q(dx) \;\ge\; \frac{16^{-p}}{4}\sum_{i\ge I(n)} c_i. \tag{18}$$

Proof. Property (i) holds because $Q \ge \mathsf{W}_1/2$ and each $\mathsf{R}_{k_i} \ll \mathsf{W}_1$, and (ii) follows from Lemma 12(d), whose bound does not depend on k. For the entropy, write $Q = \sum_{i\ge0}\alpha_i\nu_i$ with $\nu_0 = \mathsf{W}_1$, $\alpha_0 = \frac{1}{2}$, $\nu_i = \mathsf{R}_{k_i}$ and $\alpha_i = c_i/2$. Convexity of $t \mapsto t\log t$ gives, pointwise, $f\log f \le \sum_i \alpha_i f_i \log f_i$ for the densities $f = \sum_i \alpha_i f_i$ and $f_i = d\nu_i/d\mathsf{W}_1$, and Tonelli's theorem applied to $f_i\log f_i + e^{-1} \ge 0$ gives $\mathrm{KL}(Q\,\|\,\mathsf{W}_1) \le \sum_i \alpha_i \mathrm{KL}(\nu_i\,\|\,\mathsf{W}_1) = \frac{1}{2}\sum_i c_i \log(1/\mathsf{W}_1(\mathcal{E}_{k_i}))$, which is at most $\frac{15}{2}\sum_i c_i 4^{k_i}$ by Lemma 12(a).

If $I(n) = 1$ the bound on $16^{I(n)}$ is clear; otherwise $15 \cdot 16^{I(n)-1} < \log_2(2n)$ by minimality. For (18), fix i. The set $\{x : \mathrm{dist}(x, F) \le 1/16\}$ is covered by the balls $B(z, 1/16)$, $z \in F$, so Lemma 12(c) gives $\mathsf{R}_{k_i}(\mathrm{dist}(\cdot, F) \le 1/16) \le n 2^{-15 \cdot 2^{k_i}}$. For $i \ge I(n)$ we have $15 \cdot 2^{k_i} = 15 \cdot 16^i \ge \log_2(2n)$, so this is at most $1/2$. Since $Q \ge \frac{1}{2} \sum_i c_i \mathsf{R}_{k_i}$,

$$Q\big(\,\mathrm{dist}(\cdot, F) > 1/16\big) \ge \frac{1}{2} \sum_{i \ge I(n)} c_i \, \mathsf{R}_{k_i}\big(\,\mathrm{dist}(\cdot, F) > 1/16\big) \ge \frac{1}{4} \sum_{i \ge I(n)} c_i,$$

and (18) follows from $\int \mathrm{dist}(x, F)^p Q(dx) \ge 16^{-p} Q(\mathrm{dist}(\cdot, F) > 1/16)$. □

Proof of Theorem 4. Take $c_i = 6/(\pi^2 i^2)$ in Lemma 13. Properties (i) and (ii) hold. For (iii), $dQ/d\mathsf{W}_1 \ge \frac{c_i}{2} 1_{\mathcal{E}_{k_i}} / \mathsf{W}_1(\mathcal{E}_{k_i})$, so for $r > 1$, by Lemma 12(a),

$$\int \Big(\frac{dQ}{d\mathsf{W}_1}\Big)^r d\mathsf{W}_1 \ge \Big(\frac{c_i}{2}\Big)^r \mathsf{W}_1(\mathcal{E}_{k_i})^{1-r} \ge \Big(\frac{3}{\pi^2 i^2}\Big)^r \exp\Big(\frac{15(r-1)}{2}\, 4^{k_i}\Big) \longrightarrow \infty.$$

For (iv), let μ be carried by a set F of at most n points. Since $\sum_{i \ge I} i^{-2} \ge \sum_{i \ge I} (i(i+1))^{-1} = 1/I$, Lemmas 6 and 13 give

$$W_p(\mu, Q)^p \ge \int \mathrm{dist}(x, F)^p \, Q(dx) \ge \frac{16^{-p}}{4} \cdot \frac{6}{\pi^2 I(n)} = \frac{16^{-p} \cdot 3}{2\pi^2 I(n)}.$$

By Lemma 13, $I(n) \le 1 + \log_{16} \max\{1, \log_2(2n)/15\} = O(\log\log n)$, so $\sqrt{\log n}\, \inf_\mu W_p(\mu, Q) \ge \frac{1}{16}\big(3/(2\pi^2 I(n))\big)^{1/p} \sqrt{\log n} \to \infty$. For any proposal and any sample, $\widehat{Q}_n$ is carried by at most n points, and $L_{n,p} = (\int \mathrm{dist}(x, \{Y_1, \dots, Y_n\})^p Q(dx))^{1/p}$ obeys the same lower bound by (18). □

Proof of Proposition 1. Take $c_i = C_{\mathrm{e}}\, i^{-2} 4^{-k_i} = C_{\mathrm{e}}\, i^{-2} 16^{-2i}$ in Lemma 13, with C_{e} chosen so that $\sum_i c_i = 1$. Then $\mathrm{KL}(Q \,\|\, \mathsf{W}_1) \le \frac{15}{2} C_{\mathrm{e}} \sum_i i^{-2} < \infty$. For a set F of at most n points, (18), the term $i = I(n)$ and $16^{I(n)} \le 16 \max\{1, \log_2(2n)/15\}$ give, for $\log_2(2n) \ge 15$,

$$\int \mathrm{dist}(x, F)^p \, Q(dx) \ge \frac{16^{-p}}{4}\, C_{\mathrm{e}}\, I(n)^{-2} 16^{-2I(n)} \ge \frac{16^{-p} \cdot 225\, C_{\mathrm{e}}}{1024}\, I(n)^{-2} \big(\log_2(2n)\big)^{-2}.$$

Hence $(\sqrt{\log n}\, \inf_\mu W_p(\mu, Q))^p$ is at least a positive constant times $(\log n)^{p/2-2} I(n)^{-2}$, which tends to infinity for every $p > 4$ because $I(n) = O(\log\log n)$. The statements about $\widehat{Q}_n$ follow as in the proof of Theorem 4. □

Remark. At sample size n, every piece R_{k_i} with $15 \cdot 16^i \ge \log_2(2n)$ keeps half of its mass at distance at least $1/16$ from any n points, which is of order $\sqrt{\log n}$ times the typical scale. The first such piece has $16^i \asymp \log n$ and relative entropy $\log(1/\mathsf{W}_1(\mathcal{E}_{k_i})) \asymp 4^{k_i} = 16^{2i} \asymp (\log n)^2$. Its contribution to $(\sqrt{\log n}\, W_p)^p$ is of order $c_i (\log n)^{p/2}$, while its entropy contribution is of order $c_i (\log n)^2$; both results of this subsection follow from this balance.

9.3 Heavy-tailed weights

Lemma 14. *Let $c_0 = \pi \log 2/256$. For every $y \in C_0$ and $0 < \epsilon \le \sqrt{2\pi}/16$, $\mathsf{W}_1(B(y, \epsilon)) \le \exp(-c_0 \epsilon^{-2})$.*

Proof. If $x \in B(y, \epsilon)$, Lemma 11(a) gives $|\xi_{l,m}(x) - \xi_{l,m}(y)| \le 2^{l/2+2}\epsilon$ for all l, m. A standard normal variable lies in a given interval of length ℓ with probability at most $\ell(2\pi)^{-1/2}$. With $u_l = (2\pi)^{1/2}/(2^{l/2+3}\epsilon)$, Lemma 11(b) gives $\mathsf{W}_1(B(y, \epsilon)) \le \prod_{l \le L'} u_l^{-2^l}$ for every L' with $u_{L'} \ge 1$.

Let L' be the largest integer with $u_{L'} \geq 2$; it exists because $u_0 \geq 2$ and $u_l \to 0$. Then $u_l \geq 2$ for $l \leq L'$, so $-\log \mathsf{W}_1(B(y,\epsilon)) \geq 2^{L'} \log 2$. Since $u_{L'+1} < 2$, we have $2^{(L'+1)/2} > (2\pi)^{1/2}/(16\epsilon)$, that is, $2^{L'} > \pi/(256\epsilon^2)$. □

Proof of Proposition 2. Let $t_0 = e^e$ and $\beta_0 = \frac{1}{2}$. Let μ_Z be the probability measure on $(0,\infty)$ with $\mu_Z((t,\infty)) = \beta_0/(t \log t\,(\log\log t)^2)$ for $t \geq t_0$, and with the remaining mass $1 - \beta_0 e^{-e-1}$ at the point

$$z_- = \frac{1 - \beta_0(1+e^{-1})}{1 - \beta_0 e^{-e-1}} \in (0.31, 0.33).$$

The tail function is decreasing and continuous on $[t_0,\infty)$, the mass above t_0 is $\beta_0/(t_0 e) = \beta_0 e^{-e-1}$, and the substitution $u = \log\log t$ gives $\int_{t_0}^{\infty} \mu_Z((t,\infty))dt = \beta_0 \int_1^{\infty} u^{-2}du = \beta_0$, so $\int z\,\mu_Z(dz) = z_-(1-\beta_0 e^{-e-1}) + \beta_0 e^{-1} + \beta_0 = 1$. Moreover $\int z^s \mu_Z(dz) = \infty$ for every $s > 1$, since $\int^{\infty} t^{s-1}\mu_Z((t,\infty))\,dt = \infty$.

The measure $\nu_Z(dz) = z\,\mu_Z(dz)$ is a probability measure. Let g_Z be its quantile function and $Z(x) = g_Z(\Phi(x^1(1)))$, where Φ is the standard normal distribution function; under W_1, Z has law ν_Z and $Z \geq z_-$. Put $P = Z^{-1} \cdot \mathsf{W}_1$. Then $\int Z^{-1} d\mathsf{W}_1 = \int z^{-1}\nu_Z(dz) = 1$, $dP/d\mathsf{W}_1 \leq 1/z_- \leq 4$, P and W_1 are equivalent, $d\mathsf{W}_1/dP = Z$, and under P the law of Z is μ_Z. Hence $Z \notin L^s(P)$ for $s > 1$, and $D_n \to 1$ almost surely.

Let $t_n = n/(\log n\,(\log\log n)^2)$. For large n, $t_n \geq t_0$ and $t_n \log t_n (\log\log t_n)^2 \leq n$, so $\mathbb{P}(Z(Y_n) > t_n) \geq \beta_0/n$. The events $\{Z(Y_n) > t_n\}$ are independent with divergent probability sum, so by the second Borel–Cantelli lemma they occur for infinitely many n almost surely. Since P has no atoms, the samples are almost surely distinct. Consider such an n, large enough that $D_n \leq 2$. The atom at Y_n has $\widehat{Q}_n$-mass $w \geq t_n/(2n) = 1/(2\log n\,(\log\log n)^2)$. Let $\epsilon^2 = c_0/\log(2/w)$, so that $\epsilon \leq \sqrt{2\pi}/16$ for large n. In any coupling of $\widehat{Q}_n$ and W_1 the atom sends at most $\mathsf{W}_1(B(Y_n,\epsilon)) \leq e^{-c_0/\epsilon^2} = w/2$ of its mass into $B(Y_n,\epsilon)$ by Lemma 14, so mass at least $w/2$ travels farther than ϵ. Hence

$$(\sqrt{\log n}\, W_p(\widehat{Q}_n, \mathsf{W}_1))^p \geq \frac{w}{2}(\log n)^{p/2}\epsilon^p \geq \frac{c_0^{p/2}(\log n)^{p/2-1}}{4(\log\log n)^2\big(\log(4\log n(\log\log n)^2)\big)^{p/2}},$$

which tends to infinity for $p > 2$. □

10 Conclusion

The leading constant of equal-weight empirical Wiener transport in the supremum metric is the bridge entropy rate $\sqrt{\Gamma_d}$ of Theorem 1, the same for every finite order and, under importance weighting, for every target density in some L^r with $r > 1$ and every likelihood ratio in some L^s with $s > 1$ (Theorem 3); the constructions of Section 9 show that integrable densities and weights can destroy the limit. The local midpoint argument of Theorem 2 traces the excess of $\sqrt{\Gamma_d}$ over the nearest-sample constant $\sqrt{\kappa_d}$ to the requirement that both coupled paths have exact Wiener marginals, and Corollary 1 measures that excess exactly.

Statements and Declarations

Funding. This research was self-funded.

Competing interests. The author owns Luminary Oddities and has no other relevant competing interests.

Data and code availability. The verification scripts are included with this article. The script `verification/checks_ed3_midpoint.py` rechecks the midpoint identity of Theorem 2 and Lemma 1 by simulation, and `verification/checks_tilt.py` checks every numerical constant of Section 9 and Corollary 2 in high-precision arithmetic.

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of the parts of the first edition. Anthropic Claude assisted with derivations, proof development, computations, source checks and independent internal review of the parts added in later editions, and edited and consolidated the article. Scott Schubr directed the research and is responsible for the content.

Edition history. Edition 1 (November 22, 2026, DOI 10.5281/zenodo.24450849) proved Theorems 1 and 3, the latter for moment-controlled tilts, and the case $d = 1$ of Theorem 2. Edition 2 (November 25, 2026, DOI 10.5281/zenodo.24550506) extended Theorem 2 to every dimension and added Corollary 1.

References

[1] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley Series in Probability and Statistics, Wiley, New York, 1999. https://doi.org/10.1002/9780470316962.

[2] E. Boissard and T. Le Gouic, *On the mean speed of convergence of empirical and occupation measures in Wasserstein distance*, Ann. Inst. Henri Poincaré Probab. Stat. **50** (2014), no. 2, 539–563. https://doi.org/10.1214/12-AIHP517.

[3] S. Coste, *The Wasserstein cost of the Girsanov Theorem*, preprint, arXiv:2611.16043v1, 2026. https://arxiv.org/abs/2611.16043v1.

[4] S. Coste and M. Goldman, *The Wasserstein cost of Importance Sampling*, preprint, arXiv:2605.30055v1, 2026. https://arxiv.org/abs/2605.30055v1.

[5] S. Dereich, *The quantization complexity of diffusion processes*, preprint, arXiv:math/0411597v1, 2004. https://arxiv.org/abs/math/0411597v1.

[6] S. Dereich, *The coding complexity of diffusion processes under supremum norm distortion*, Stochastic Process. Appl. **118** (2008), no. 6, 917–937. https://doi.org/10.1016/j.spa.2007.07.003.

[7] S. Dereich and M. A. Lifshits, *Probabilities of randomly centered small balls and quantization in Banach spaces*, Ann. Probab. **33** (2005), no. 4, 1397–1421. https://doi.org/10.1214/009117905000000161.

[8] S. Dereich and M. Scheutzow, *High-resolution quantization and entropy coding for fractional Brownian motion*, Electron. J. Probab. **11** (2006), 700–722. https://doi.org/10.1214/EJP.v11-344.

[9] R. Durrett, *Probability: Theory and Examples*, 5th ed., Cambridge University Press, Cambridge, 2019. https://doi.org/10.1017/9781108591034.

[10] X. He and F. Yuan, *Sharp functional quantization and empirical Wasserstein rates for Itô processes*, preprint, arXiv:2608.07879v3, 2026. https://arxiv.org/abs/2608.07879v3.

[11] C. Léonard, *Girsanov theory under a finite entropy condition*, in Séminaire de Probabilités XLIV, Lecture Notes in Math. 2046, Springer, Berlin, 2012, 429–465. https://doi.org/10.1007/978-3-642-27461-9_20.

[12] Y. Lv, *Brownian motion between two random trajectories*, Markov Process. Related Fields **25** (2019), no. 2, 359–377. https://math-mprf.org/journal/articles/id1540/; theorem numbering as in arXiv:1802.03876v4.

[13] P. Mörters and Y. Peres, *Brownian Motion*, Cambridge Series in Statistical and Probabilistic Mathematics 30, Cambridge University Press, Cambridge, 2010. https://doi.org/10.1017/CBO9780511750489.

[14] A. A. Novikov, *On an identity for stochastic integrals*, Theory Probab. Appl. **17** (1973), no. 4, 717–720. https://doi.org/10.1137/1117088.

[15] A. Prékopa, *On logarithmic concave measures and functions*, Acta Sci. Math. (Szeged) **34** (1973), 335–343. https://acta.bibl.u-szeged.hu/14411/.

[16] N. Saldi, T. Linder and S. Yüksel, *Output constrained lossy source coding with limited common randomness*, IEEE Trans. Inform. Theory **61** (2015), no. 9, 4984–4998. https://doi.org/10.1109/TIT.2015.2450721.

[17] B. von Bahr and C.-G. Esseen, *Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$*, Ann. Math. Statist. **36** (1965), no. 1, 299–303. https://doi.org/10.1214/aoms/1177700291.