The exact variational constant for empirical Wiener transport

Scott Schubr
Luminary Oddities

Abstract

For the equal-weight empirical law of n independent d-dimensional Brownian paths on $[0,1]$, the Wasserstein distance to Wiener measure in the supremum metric is asymptotic to $\sqrt{\Gamma_d/\log n}$ for every finite Wasserstein order. The convergence after multiplication by $\sqrt{\log n}$ holds almost surely and in all finite sample moments. We characterize Γ_d as the long-time minimum mutual information per unit time for coupling two exact Brownian bridge laws within a unit tube. In dimension one, we prove $\sqrt{\Gamma_1} > \sqrt{\kappa_1}$, where $\sqrt{\kappa_1}$ is the nearest-sample constant. Thus requiring equal sample masses increases the leading-order transport cost. The strict comparison follows by localizing the entropy defect and deriving a contradiction with the Gaussian bridge midpoint second moment. The same empirical constant holds under moment-controlled importance weighting.

Keywords. Wiener measure, empirical Wasserstein distance, Brownian bridges, small-ball probabilities, relative entropy.
MSC 2020. Primary 60J65, 60F25. Secondary 49Q22.

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 $[0,1]$, and let W_p use the supremum distance between paths. Let W_1 denote their common Wiener law. Boissard and Le Gouic [1, p. 544] ask for the sharp mean asymptotic for Gaussian path laws, including Wiener measure in the supremum norm. Coste [2, Theorem 1.1] proves the order $(\log n)^{-1/2}$ and treats convergence of the normalized error as conjectural [2, §§1.5–1.5.3]. His numerical study of scalar W_1 leaves open whether the limit equals the nearest-sample constant $\sqrt{\kappa_1}$.

Existence and characterization of the limit. Theorem 1 proves that $\sqrt{\log n}\, W_p(\widehat{\mathsf{W}}_n, \mathsf{W}_1)$ converges to $\sqrt{\Gamma_d}$ for every finite $p \geq 1$, almost surely and in every finite sample moment. It identifies Γ_d as the limiting mutual information per unit time required to couple two exact Brownian bridge laws within distance one throughout. A finite-block estimate quantifies the approximation of this rate by the relaxed bridge problem. In particular the theorem gives the sharp mean asymptotic for the Wiener case in [1].

Strict separation from nearest-sample transport. Theorem 2 proves $\Gamma_1 > \kappa_1$. Assigning each path to its nearest sample allows the sample masses to vary. Requiring equal sample masses therefore increases the leading-order cost in dimension one. The proof shows that equality would force mutually compatible local tube conditionals whose joint law violates the Brownian bridge midpoint second moment.

Importance weighting. Theorem 3 gives the same normalized limit for self-normalized importance sampling between moment-controlled Wiener tilts. This includes bounded adapted drift proposals and targets with common unit volatility and addresses the question of non-Brownian proposals raised in [2, §2.2].

1.1 The bridge entropy formulation

Fix $d \geq 1$ 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 $r > 0$,

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

For the relaxed problem let

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

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*

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

For every finite $p \geq 1$,

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

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

The exponent κ_d is the randomly centered Wiener small-ball constant of Dereich and Lifshits [3, Theorem 6.1]. Their nearest-sample asymptotic implies $\Gamma_d \geq \kappa_d$: the expectation of the W_1 cost dominates the mean nearest-sample distance [3, Theorems 5.1 and 6.1]. In dimension one the inequality is strict by Theorem 2. Lv [4, Theorem 2.1] studies the related quenched tube exponent with restrictions on starts and returns. He and Yuan [5, Theorems 2.1–2.2] establish supremum quantization and empirical order bounds for a wider class of Itô processes. Dereich and Scheutzow [6] study deterministic entropy coding of Wiener paths. Our use of couplings with prescribed Gaussian marginals is also related to output-constrained coding under additive distortion [7].

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 5–6 translate this equality into matching empirical bounds. Sections 7–8 use the exact-marginal couplings to establish the scalar gap and the weighted limit, respectively.

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 $p_j = \beta_T(J = j)$,

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

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 [6, 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

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

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 $p \ge 1$ 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 $L \ge 1$. 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 $\mathbb{P}_f(\log f \le -t) \le e^{-t}$. 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

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

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 scalar gap from nearest-sample transport

Theorem 2. *For scalar Wiener measure in the supremum metric, $\Gamma_1 > \kappa_1$. Consequently the leading equal-weight transport constant $\sqrt{\Gamma_1}$ is strictly larger than the nearest-sample constant $\sqrt{\kappa_1}$.*

Proof. Work in dimension one. Let $\mathcal{A}_T = \{\|x - y\|_\infty \le 1\}$ 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 [3, proof of Theorem 6.1, equation (6.1)] give $T^{-1}\mathbb{E}[-\log p_T(X)] \to \kappa_1$.

Choose symmetric bridge pairs of radius one with entropy at most $C_T(1) + 1$. Independently choose a common endpoint cell of length $\eta = T^{-1/4}$ with the rounded $N(0, T)$ 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 $1+\eta$. As in the skeleton construction, the added entropy is the entropy of the rounded endpoint, which is $O(\log T)$ by its

second-moment bound. Scale space by $(1+\eta)^{-1}$ and time by $(1+\eta)^{-2}$. Writing T for the resulting durations, we obtain symmetric radius-one Wiener couplings P_T with

$$\limsup_{T\to\infty} T^{-1}\,\mathrm{KL}(P_T\|\mathsf{W}_T\otimes\mathsf{W}_T)\le\Gamma_1.$$

If $\Gamma_1=\kappa_1$, 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 $d=(d_0,d_1)$ for the values of $Y-X$ at the window endpoints and let ℓ_d interpolate these values. Writing $\beta=\beta_L$, define

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

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

For a partition into length-L windows and any partial exterior intervals, condition on the entire X path and all Y knot values. Under Q_T the conditional Y residuals factor as independent reference bridges clipped to their respective tubes. Indeed the likelihood factor $1/p_T(X)$ is fixed under the conditioning and the tube event factors over the partition. The length-L factors are $K(d,U;\cdot)$. Relative entropy against a product law dominates the sum of the marginal relative entropies. Applying this fact after the entropy chain rule shows that the sum of the conditional entropy defects on the blocks is at most $\mathrm{KL}(P_T\|Q_T)$. Coarsening the conditioning on each block to (d,U) preserves the reference kernel and decreases the defect by data processing.

Let M_T be the local law of (d,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

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

Exchange symmetry also gives the reverse conditional bound. If N_T denotes the law of (d_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 $d_0\in[-1,1]$, the local laws are tight. Take a weak subsequential limit N and denote by M its image in the coordinates (d,U,V), with $d_1=d_0+J(L)-I(L)$. 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 Brownian motion of variance rate two. Both $Z(t)$ and $Z(L)-Z(L-t)$ take both signs arbitrarily close to zero almost surely. A boundary offset at either endpoint would force one of these two increment paths to stay on one side of zero. Consequently M gives full mass to offsets $d\in(-1,1)^2$.

For a fixed interior d and residual u, the clipped probability is positive by full support of β around $v=u$. The distribution function of $\|V-u+\ell_d\|_\infty$ under β is log-concave in the radius by the finite-dimensional Prékopa inequality [8] on nested dense time grids and continuity from above. It is positive for all radii larger than $\|\ell_d\|_\infty<1$, and therefore continuous at one. Thus

under convergence of (d, u) to an interior point, the tube indicators converge β-almost surely. The probability kernel K is continuous in total variation at those points by dominated convergence, including convergence of its denominator. In (10) the regenerated laws $(M_T)_{d,U}K$ consequently converge weakly to $M_{d,U}K$ by testing against bounded continuous functions. The exceptional boundary has zero limit mass. Pinsker's inequality gives $M = M_{d,U}K$, and exchange symmetry gives the reverse identity.

Fix an interior d 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 $a_d(u)1_{A_d}(u, v) = b_d(v)1_{A_d}(u, v)$. All pairs of residuals within distance $1 - \|\ell_d\|_\infty$ of each other belong to A_d. 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_d and b_d 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 $(\beta \otimes \beta)(\cdot \mid A_d)$.

For independent bridges, $S = (U+V)/2$ and $V-U$ are independent Gaussian bridges of variance rates $1/2$ and 2, respectively. The tube event depends only on $V - U$, so the law and independence of S persist under this conditional pair. The tube also gives $\|V - U\|_\infty \le 2$. In particular,

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

But the unconditional U marginal is exactly β, with midpoint second moment $L/4$. Averaging the displayed bound contradicts $L > 8$. This proves strict inequality. □

8 Universality under importance weighting

Let P be equivalent to W_1 and let $Q \ll \mathsf{W}_1$. Use densities $p_0 = dP/d\mathsf{W}_1$ and $q_0 = dQ/d\mathsf{W}_1$. Suppose p_0, q_0 have every finite positive moment under W_1 and $d\mathsf{W}_1/dP$ has all such moments under P. Then $Z = dQ/dP$ has all finite positive moments under P, by Hölder. For iid $Y_i \sim P$ let

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

with $\widehat{Q}_n = \delta_0$ when $D_n = 0$. The leading upper and lower bounds for this model were studied in [9]. The exact limit is as follows.

Theorem 3. *For every finite $p \ge 1$, $\sqrt{\log n}\, W_p(\widehat{Q}_n, Q) \to \sqrt{\Gamma_d}$ almost surely and in L^q for every finite $q \ge 1$. In particular this applies to any two bounded-drift laws with common unit volatility. For common deterministic scalar volatility bounded above and bounded away from zero, with total variance V, the constant is $\sqrt{V\Gamma_d}$.*

Proof. For the lower bound, apply the typical-set and collision argument from the Wiener case. Each collision probability for an independent target and proposal path is bounded by a constant times the $\mathsf{W}_1 \otimes \mathsf{W}_1$ collision probability to the power $1 - 1/u$, by Hölder for any fixed $u > 1$. This also controls the atypical target mass. Choose u large enough that the strict exponent gap when $a^2 < R_L/L$ survives. The expected unnormalized sample mass of atypical codewords is the Q mass of that event since $\mathbb{E}_P[Z1_B] = Q(B)$. With the same dyadic envelopes as before, Markov and Borel–Cantelli, followed by $D_n \to 1$ almost surely, show that the atypical weighted mass vanishes. Thus the covered mass and the two atypical masses again vanish and give the lower bound $\sqrt{\Gamma_d}$.

For the upper bound use the symmetric Wiener coupling from Section 6 after rescaling it to unit time. Denote it by π_n, its density relative to $\mathsf{W}_1 \otimes \mathsf{W}_1$ by f_n, and its hard radius by $r_n = (1+3\eta)/\sqrt{T}$. For fixed block parameters chosen there, $\log n/T \to b' > b$ and under π_n the event $\log f_n > Tb$ has probability $O((\log n)^{-K})$ for any prescribed K, by the choice of the fixed block partition. Weight this Wiener coupling by

$$a(x,y) = \min(q_0(x), q_0(y)).$$

The resulting subcoupling has identical marginals $\mu_{0,n} \le Q$. Write $\rho_n = Q - \mu_{0,n}$, of mass δ_n. Exact Wiener marginals give the identity

$$\delta_n = \tfrac{1}{2} \int |q_0(x) - q_0(y)| \, d\pi_n(x,y) = o(1).$$

Indeed bounded Lipschitz functions are dense in $L^1(\mathsf{W}_1)$ by regularity of Borel probability measures on a metric space. Approximate q_0 by one such function, then use the hard radius $r_n \to 0$ and the exact two marginals to control the integral. This argument requires no rate for δ_n.

For codeword y let K_y be the finite positive kernel in the x variable with density

$$F_n(x,y) = f_n(x,y)a(x,y)/p_0(y)$$

against W_1. Here and below the kernel may have mass up to $Z(y)$, and its average under P is $\mu_{0,n}$. In particular it couples x to y at distance at most r_n where it has mass. Write $w_0(y)$ for the total mass of K_y and set $w_{\mathrm{res}}(y) = Z(y) - w_0(y) \ge 0$. The average under P of $w_{\mathrm{res}}(y)\delta_y$ equals ρ_n by symmetry of the subcoupling.

We need a fast accuracy bound for the unnormalized average $\widehat{\mu}_{0,n} = n^{-1}\sum_i K_{Y_i}$. The truncation argument of (8) works with reference $\mathsf{W}_1(dx)P(dy)$ for the subdensity F_n, now comparing the average to $\mu_{0,n}$. Its total mass expectation is at most one. Choose b_0 strictly between b and b' and truncate at e^{Tb_0}. The mass of the discarded part is at most the mass under $a\pi_n$ of the union

$$\{\log f_n > Tb\} \ \cup \ \{Z(y) > e^{T(b_0-b)}\}.$$

Indeed $a(x,y)/p_0(y) \le Z(y)$. The first event has arbitrarily small negative powers of $\log n$ as mass, by Hölder with $a \le q_0(y)$ and the moment bound under π_n. The second is exponentially small in T by the Q moments of Z. The truncated variance term is $O(\sqrt{e^{Tb_0}/n})$. Hence the expected total variation norm of the difference of the finite measures is $O((\log n)^{-K})$ for arbitrarily large prescribed K.

For an all-time almost-sure bound, replace K_y by zero on $\{Z(y) > n^{1/4}\}$. The expected omitted mass is arbitrarily polynomially small, and almost surely no sampled weight exceeds the cutoff for all sufficiently large n, by a union bound with a high moment. The bounded-differences estimate for the truncated average has per-codeword bound $2n^{-3/4}$ for the total variation norm. This gives the same $O((\log n)^{-K})$ bound almost surely, using summable tails at that threshold. For any finite sample moment the bound follows from the expected bound and interpolation with the uniform higher moments of $D_n + 1$, which dominate the total variation norm. Choosing K larger in the expectation gives any required exponent after interpolation. Also $D_n - 1$ is polynomially small almost surely and in all fixed sample moments by the higher moments of Z.

It remains to transport the leftover empirical atom masses without spending the main radius on them. Use a deterministic Wiener code of $M = \lfloor n^{1/4} \rfloor + 1$ points with nearest projection ξ_n and error $e_n(x) = \|x - \xi_n(x)\|_\infty$ of order $(\log n)^{-1/2}$ in $L^{2p}(Q)$. Its existence follows from the deterministic codes of [6] and Hölder. Include zero and discard centers of norm bigger than $C\sqrt{\log n}$. For $\|x\|_\infty \le C\sqrt{\log n}/2$, zero is at least as close as every discarded center. Outside this

event, the new error is bounded by $\|x\|_\infty$. The Gaussian tail of $\|x\|_\infty$ under W_1 and Hölder with q_0 show that the change is polynomially small in each required $L^u(Q)$ norm. Let B_j be the code cells, z_j the centers, and set

$$c_j = \rho_n(B_j), \qquad \widehat{c}_j = n^{-1}\sum_i w_{\mathrm{res}}(Y_i)1_{B_j}(Y_i).$$

These errors sum in absolute value to a polynomially small quantity almost surely and in all finite sample moments. For instance each centered error has L^u norm $O_u(n^{-1/2})$ by the moments of Z, so their sum is $O_{L^u}(n^{-1/4})$ and Markov with large u gives an almost-sure power bound. The projected residual costs satisfy

$$\int e_n^p\, d\rho_n \le \delta_n^{1/2}(Qe_n^{2p})^{1/2} = o((\log n)^{-p/2}), \qquad n^{-1}\sum_i w_{\mathrm{res}}(Y_i)e_n(Y_i)^p = \int e_n^p\, d\rho_n + O(n^{-c})$$

almost surely and in all fixed sample moments, for a positive power c which may be chosen sufficiently small. The last statement follows from the independent moment bound, $w_{\mathrm{res}} \le Z$, and $e_n(x) \le \|x\|_\infty$ because zero is in the code.

On each cell retain the fraction of its target residual with mass $\min(c_j, \widehat{c}_j)$, and if there is excess sample mass add $\widehat{c}_j - c_j$ at z_j. Denote the resulting modified residual by $\widetilde{\rho}_n$. It has total mass $\sum_j \widehat{c}_j$ and differs from ρ_n in total variation norm by at most $\sum_j |\widehat{c}_j - c_j|$. Its within-cell projection cost to the specified centers is no greater than that of ρ_n. Match it cellwise to the raw empirical residual weights using each cell center. The pth power cost of this subcoupling is at most 2^{p-1} times the sum of the preceding two residual projection costs.

Define the auxiliary probability $Q'_n = (\widehat{\mu}_{0,n} + \widetilde{\rho}_n)/D_n$ when $D_n > 0$. Its total variation error from Q decays at any prescribed power of $(\log n)^{-1}$, almost surely and in finite sample moments on $\{D_n \ge 1/2\}$, by the bounds on its two raw components and $D_n - 1$. The high path moments of Q'_n are bounded almost surely and in every fixed sample moment on this event: $\widehat{\mu}_{0,n}$ is coupled within r_n to weighted codewords, and $\widetilde{\rho}_n$ is dominated by ρ_n apart from polynomially small mass at clipped codepoints. Thus the same common-measure and Hölder argument as in the Wiener upper bound gives $W_p(Q'_n, Q) = o((\log n)^{-1/2})$. The coupling of Q'_n to $\widehat{Q}_n$ uses the hard radius r_n for the main weights and the negligible cellwise cost for the residual. Its total pth power cost is at most $r_n^p + o((\log n)^{-p/2})$. These statements hold almost surely and in any required finite sample moment on $\{D_n \ge 1/2\}$. The complementary event has arbitrarily small polynomial probability by moments of Z, and the empirical error there is bounded by the target moment plus the maximum proposal path norm, with Gaussian-tail moment bounds. It contributes negligibly to all finite sample norms. Now send the fixed block parameters to the limit as in the Wiener upper bound. This proves the upper constant and, together with higher sample moments and the almost-sure lower bound, the claimed convergence.

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

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

makes X Brownian. For every real a, the exponential martingale with integrand $-ab$ gives $\mathbb{E}H^a \le \exp(\frac{1}{2}|a(a-1)|\|b\|_\infty^2)$. 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 hypotheses for each bounded-drift law, including when the drift is adapted to a larger filtration. With common deterministic scalar volatility σ, the time change $u = \int_0^t \sigma_s^2\, ds$ is an isometry in supremum norm and gives unit volatility on $[0, V]$. The drift in the new time is b_t/σ_t^2, which is bounded. Brownian rescaling from $[0, V]$ to $[0, 1]$ scales all distances by $V^{-1/2}$, giving the stated factor $\sqrt{V}$ for the original paths. □

9 Conclusion

Theorem 1 gives the exact leading asymptotic for equal-weight empirical Wiener transport in the supremum metric. Theorem 2 identifies the effect of the mass constraint: in dimension one, the equal-weight constant strictly exceeds the nearest-sample constant. The local midpoint argument traces this increase to the requirement that both coupled paths have exact Wiener marginals. The finite-block estimate gives control of the bridge entropy approximation, and Theorem 3 shows that the leading constant persists under the stated importance sampling assumptions.

References

[1] 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), 539–563. Article.

[2] S. Coste, *The Wasserstein cost of the Girsanov Theorem*, preprint, arXiv:2611.16043v1 (2026). Article.

[3] S. Dereich and M. A. Lifshits, *Probabilities of randomly centered small balls and quantization in Banach spaces*, Ann. Probab. 33 (2005), 1397–1421. Article.

[4] Y. Lv, *Brownian motion between two random trajectories*, Markov Process. Related Fields 25 (2019), no. 2, 359–377. Article.

[5] X. He and F. Yuan, *Sharp functional quantization and empirical Wasserstein rates for Itô processes*, preprint, arXiv:2608.07879v3 (2026). Article.

[6] S. Dereich and M. Scheutzow, *High-resolution quantization and entropy coding for fractional Brownian motion*, Electron. J. Probab. 11 (2006), 700–722. Article.

[7] N. Saldi, T. Linder and S. Yüksel, *Output constrained lossy source coding with limited common randomness*, IEEE Trans. Inform. Theory 61 (2015), 4984–4998. Article.

[8] A. Prékopa, *On logarithmic concave measures and functions*, Acta Sci. Math. (Szeged) 34 (1973), 335–343. Article.

[9] S. Schubr, *Importance-weighted path transport with non-Brownian proposals*, Luminary Oddities research manuscript (2026).