Wasserstein transport for Kloosterman paths: optimal rates, prime powers and transport constants

Scott Schubr
Luminary Oddities

November 2026

Abstract

We prove that prime-modulus Kloosterman paths converge in law in the space of continuous paths when one parameter is fixed and only the other varies, the case identified by Kowalski and Sawin as a key open problem, and we determine the rate exactly. For every fixed finite order $s \geq 1$, every fixed nonzero inverse coefficient and large odd primes p, the path law satisfies $c/\sqrt{\log p} \leq W_s \leq C_s/\sqrt{\log p}$ in the uniform path norm, and the same order holds when both parameters vary; this removes the $\sqrt{\log\log p}$ loss of the earlier two-parameter rate. The fixed-parameter short-interval estimate also verifies Lebreton's tightness assumption, so his results on large partial sums hold with one parameter fixed. The same order holds on most translates of any growing parameter set, on long intervals and progressions, and at prime powers, where hyper-Kloosterman paths of every fixed rank approach a parity-dependent cyclotomic model and fixed-prime observations converge at explicit depth rates. We then identify the leading constants: finite-support quantization, random codebooks and empirical transport give the strict hierarchy $C_F < C_G < C_B$ for the L^2 path limit, with exact empirical constants also for Gaussian measures on Hilbert space and for unequal Fourier laws, and a character-moment construction shows that moment inequalities alone cannot fix the arithmetic constant. Two-shift, rational stationary and resonant circle projections have sharp fixed-prime orders and explicit constants. Finally, the finite-support constant extends to weighted Hilbert-valued series with stationary ergodic coefficients: the leading constant is an integral of the process rate-distortion function, one universal codebook sequence attains the ergodic-component constants for every stationary law, and dependence can strictly lower the constant while the one-coordinate law stays fixed.

Contents

1 Introduction **4**

2 Optimal Kloosterman path Wasserstein rates **4**
2.1 The uniform path law 5
2.2 Character comparison and exact mesh completion 5
2.3 Time samples and correlated noise 7
2.4 Comparison on selected marginals 8
2.5 Sparse comparison and transport 9

3 Fixed-parameter Kloosterman paths **11**
3.1 Introduction 11
3.2 Completion and integrated block moments 12
3.3 Fixed-parameter convergence 13
3.4 Polynomial transfer for small balls 14

4 Kloosterman paths on translated parameter sets **16**
4.1 Statements . . . 17
4.2 Additive twists and character norms . . . 18
4.3 Completed tails and finite support . . . 19
4.4 Observation comparison and conclusions . . . 20
.1 Correlated observation comparison . . . 20
.2 Comparison on selected marginals . . . 22
.3 Sparse comparison and transport . . . 23

A Sparse Kloosterman path transport **24**
A.1 Statement and inputs . . . 25
A.2 Observations and stable reconstruction . . . 26
A.3 Polynomial sample channels . . . 27
A.4 Truncated comparison of the observation laws . . . 29
A.5 Transport cost . . . 30

B Kloosterman path transport at prime powers **30**
B.1 Path laws . . . 30
B.2 Stationary phase and derivative roots . . . 31
B.3 Completed Fourier blocks . . . 33
B.4 Transport at large base prime . . . 34
B.5 Fixed-prime depth convergence . . . 35
.1 Correlated observation coupling . . . 36
.2 Comparison on selected marginals . . . 37
.3 Sparse comparison and transport . . . 39

A Hyper-Kloosterman paths at prime powers **40**
A.1 Path laws . . . 40
A.2 Complete coefficients and torus modes . . . 42
A.3 Arithmetic comparison . . . 43
A.4 Fourier blocks and transfer to whole paths . . . 44
A.5 The lower bounds and the single missing residue . . . 46
.1 Correlated observation coupling . . . 47
.2 Comparison on selected marginals . . . 48
.3 Sparse comparison and transport . . . 49

A Fixed-prime transport for observations of Kloosterman paths **51**
A.1 Paths and observations . . . 51
A.2 Arithmetic comparison and a Fourier tail . . . 52
A.3 Characteristic functions of the observations . . . 54
A.4 What the individual phase bounds imply . . . 55

B The finite-support approximation constant of the Kloosterman limit **57**
B.1 Exact quantization for weighted coordinates . . . 57
B.2 Scalar objective and coding bounds . . . 58
B.3 Weighted bands and the strong converse . . . 59
B.4 Kloosterman paths and arithmetic lower bounds . . . 60
B.5 Bounds and evaluation of the Sato–Tate constant . . . 61

C Random codebooks and empirical transport for the Kloosterman limit 63
C.1 Three approximation constants . . . 63
C.2 Scalar tilts and weighted bands . . . 64
C.3 Conditional small balls and nearest codewords . . . 65
C.4 The equal-mass empirical law . . . 66
C.5 The ordinary Kloosterman path metric . . . 67
C.5.1 Converse in the ordinary metric . . . 67
C.5.2 Conditional small-ball lower estimate with pinning . . . 68
C.5.3 Equal masses with a small pinning error . . . 69
C.6 A distribution-constrained arithmetic lower bound . . . 70

D A character-moment obstruction to the Kloosterman transport constant 72
D.1 The obstruction . . . 72
D.2 The permitted moment tests and finite bins . . . 73
D.3 The exact cost of repeated random permutation groups . . . 74
D.4 All labels distinct using high-frequency permutations . . . 76

E Affine Kloosterman correlations and the empirical transport constant 77
E.1 Affine path ensembles . . . 77
E.2 Preserving likelihoods and the random affine theorem . . . 78
E.3 Exact completed arithmetic moments . . . 80
E.4 Sharp averaged character correlation . . . 81
E.5 A random labeling of the genuine arithmetic seed . . . 82
E.6 The short-window covariance needed for the arithmetic constant . . . 83

F Empirical transport excess for unequal random Fourier laws 86
F.1 Population transport and its empirical excess . . . 86
F.2 A slack decomposition and scalar couplings . . . 87
F.3 The exact weighted slack cost . . . 89
F.4 Two residue gates at a fixed prime . . . 90
F.5 An arithmetic excess lower bound . . . 91

G Sharp empirical Wasserstein constants for Gaussian measures on Hilbert space 92
G.1 The empirical constant . . . 92
G.2 Gaussian preserving couplings . . . 93
G.3 Soft covering in one and two samples . . . 94
G.4 The exact limit . . . 94
G.5 Brownian paths and two Kloosterman samples . . . 96
G.6 Excess transport between Gaussian laws . . . 96
G.7 An explicit noncommuting path application . . . 97
G.8 Paired Gaussian samples . . . 98

H Two-shift Kloosterman transport at fixed prime 99
H.1 Statements . . . 99
H.2 The stationary Fourier coefficient . . . 100
H.3 Transport on the phase torus . . . 101
H.4 Scalar traces and cosine folding . . . 103

I Rational stationary phases in Kloosterman transport **104**
I.1 The three-coordinate problem . . . 105
I.2 The exact cone and the remaining levels . . . 106
I.3 A strong negative Sobolev limit . . . 107
I.4 First variation of the transport distance . . . 108
I.5 An obstruction for arbitrarily many shifts . . . 109

J Transport constants for resonant circle projections **111**
J.1 The valuation model . . . 111
J.2 Exact circle transport and shell decomposition . . . 112
J.3 A Legendre path at odd depths . . . 113
J.4 Exceptional prime sequences . . . 114
J.5 What survives a cosine projection . . . 115

K Exact functional quantization for weighted stationary sequences **117**
K.1 The constant . . . 118
K.2 Rate and distortion profiles . . . 118
K.3 One code budget across several bands . . . 119
K.4 Matching weighted codes . . . 120
K.5 Arbitrary stationary laws . . . 121
K.6 Dependence and the role of the hypotheses . . . 123

1 Introduction

Kowalski and Sawin showed that Kloosterman paths converge to a random Fourier series with Sato–Tate coefficients, and the Luminary Research Archive article *A Wasserstein rate for Kloosterman paths, with sharp small-ball estimates* [2] gave a rate $O(\sqrt{\log\log p/\log p})$ in the Wasserstein distance together with sharp small-ball estimates, answering a question of Kowalski and Untrau. Kowalski and Sawin [4] proved convergence in law on the path space when both parameters vary and finite-dimensional convergence when one is fixed; they isolated tightness for a single varying parameter as a key open problem, and Ricotta, Royer and Shparlinski [3] record that this prime-modulus case remained open after the prime-power analogue was settled. Section 3 proves tightness and convergence in law with one parameter fixed, uniformly in the fixed parameter, and Section 2 gives the exact order $(\log p)^{-1/2}$ in that setting. This article thereby sharpens and extends the earlier work. Sections 4–A treat translated parameter sets and the sparse refinement. Sections B–A move to prime powers and higher ranks. Sections B–G determine the transport constants and their limits. Sections H–J treat fixed-prime projections with several shifts. Section K extends the finite-support constant of Section B from independent Sato–Tate coefficients to weighted series with stationary ergodic coefficients. The constant becomes an integral of the process rate-distortion function, the non-Gaussian and dependent counterpart of the sharp Gaussian formula of Luschgy and Pagès [1]; slowly switching binary Markov coefficients give an exact telegraph-process coefficient.

This is the second edition of this article. The second edition adds Section K.

2 Optimal Kloosterman path Wasserstein rates

We prove that the uniform-path Wasserstein distance between prime Kloosterman paths and their Sato–Tate Fourier limit has exact order $(\log p)^{-1/2}$. The estimate holds for every fixed

nonzero inverse-term coefficient, uniformly in that coefficient, and for the two-parameter ensemble. The upper bound uses correlated smoothing of sampled Fourier blocks. The smoothing law is close to a bounded product perturbation and has small joint tails on prescribed coordinate subsets. Its selected marginals have compact Fourier support and controlled mixed derivatives. These properties allow a sparse comparison of observation laws with weighted character norm $\exp(O(\log p))$. The lower bound follows from uniform small-ball estimates and finite support.

Keywords Kloosterman paths, Wasserstein distance, Fourier smoothing, sparse observations.
Mathematics Subject Classification 11L05, 11T23, 60F17.

2.1 The uniform path law

For an odd prime p, take $B \in \mathbb{F}_p^\times$ fixed and A uniform on $\mathbb{F}_p^\times$. Let $K_p(A, B)$ be the polygon formed by the successive normalized partial sums

$$p^{-1/2} \sum_{x=1}^{j} e((Ax + B/x)/p), \qquad e(t) = e^{2\pi i t}, \qquad 0 \le j \le p-1,$$

at times $j/(p-1)$. Here division in the phase is modulo p. Write $\mu_{p,B}$ for its law and μ_p for the law with both parameters independent uniform nonzero residues. Write μ for the law of

$$K(t) = tX_0 + \sum_{h \neq 0} X_h \frac{e(ht) - 1}{2\pi i h},$$

where the X_h are independent Sato–Tate variables of density $\sqrt{4 - x^2}/(2\pi)$ on $[-2, 2]$. The series converges almost surely uniformly. We equip the path space with the supremum norm.

Kowalski and Sawin [5] established the finite-dimensional limit when $B = 1$ and the functional limit when both parameters vary. Kowalski and Untrau [6, Section 1.3(4)] asked for a quantitative path Wasserstein estimate. The two-parameter estimate in [7] gives the upper rate $O(\sqrt{\log\log p/\log p})$ and the lower order $(\log p)^{-1/2}$. Its Fourier cutoff of order $\log p$ pays the expected supremum of the omitted random Fourier tail. We retain higher frequencies and compare observations of those frequencies in the time variable. Only the high-frequency samples that exceed a threshold need affect the reconstruction.

Theorem 2.1. *For every fixed finite $s \ge 1$ there are $c, C_s > 0$ such that for all sufficiently large primes, uniformly over $B \in \mathbb{F}_p^\times$,*

$$\frac{c}{\sqrt{\log p}} \le W_s(\mu_{p,B}, \mu) \le \frac{C_s}{\sqrt{\log p}}, \qquad \frac{c}{\sqrt{\log p}} \le W_s(\mu_p, \mu) \le \frac{C_s}{\sqrt{\log p}}.$$

The argument uses three estimates for the same positive noise law. The sampled Fourier blocks can be reconstructed with high probability at the target accuracy. The probability that many specified noisy coordinates leave their central ranges is small enough to truncate an inclusion–exclusion expansion. Fourier support and mixed derivatives of selected marginals are controlled together, allowing total variation comparison for each term of this expansion.

2.2 Character comparison and exact mesh completion

Use representatives $|h| \le (p-1)/2$ and set

$$W_h = p^{-1/2} \sum_{x \neq 0} e(((A - h)x + B/x)/p).$$

These variables are real and bounded by 2. Let $\|P\|_{\mathrm{ch}}$ be the sum of absolute coefficients in the product basis of the characters $\chi_k(2\cos\theta) = \sin((k+1)\theta)/\sin\theta$, weighted by the product of the dimensions $k+1$. The norm is submultiplicative by the Clebsch–Gordan formula and $\|x_h\|_{\mathrm{ch}} = 2$. The translated Kloosterman sheaf comparison [7, Lemma 6.3] gives

$$|\mathbb{E}P(W) - \mathbb{E}P(X)| \le 4(d+1)p^{-1/2}\,\|P\|_{\mathrm{ch}} \tag{1}$$

for polynomials on d distinct shifts. The bound is uniform in the fixed B. It extends to series absolutely convergent in this character norm, by the dimension bound on pointwise character values.

If P is bounded by one on the cube and has separate degrees n_h, then $\|P\|_{\mathrm{ch}} \le \prod_h (n_h+2)^2$. Under angular substitution $x_h = 2\cos\theta_h$ its Fourier coefficients are bounded by one. Grouping signs and converting cosines to characters costs weight at most $j+1$ per frequency j in one coordinate, which gives the bound. If instead the total degree is at most r in d variables, the number of angular modes is at most $2^r\binom{d+r}{r}$ and their product weight is at most 2^r. Thus for general cube bound U,

$$\|P\|_{\mathrm{ch}} \le U\,4^r\binom{d+r}{r}. \tag{2}$$

On the full mesh j/p, add a constant final segment and interpolate. The resulting path coincides at the mesh points with

$$tW_0 + \sum_{0<|h|\le(p-1)/2} b_{p,h}W_h(e(ht)-1), \qquad b_{p,h} = \frac{1}{p(1-e(-h/p))}.$$

This follows by summing the inverse discrete Fourier transform of the normalized summands, including the zero summand at $x = p$. The coefficients are $O(|h|^{-1})$ and differ from $b_h = (2\pi i h)^{-1}$ by $O(p^{-1})$. If frequencies through H_* are retained, the replacement of the retained multipliers costs $O(H_*/p)$. Interpolating that retained polynomial also costs $O(H_*/p)$ by its deterministic derivative bound. The mesh change from $j/(p-1)$ adds at most $p^{-1/2}$.

In an unpinned block $T_H(t) = \sum_{H/2<|h|\le H} a_h W_h e(ht)$ with $|a_h| \ll H^{-1}$, Parseval on T_H^r gives $\|T_H\|_\infty^{2r} \le (2rH+1)\int_0^1 |T_H|^{2r}$. Integration retains only tuples with equal sums of the first and last r frequencies. Their absolute coefficient mass is bounded by $\max|a_h|(\sum|a_h|)^{2r-1} \ll C^{2r}/H$, because all but one frequency determine the last. Each monomial comparison costs at most $4(2r+1)2^{2r}p^{-1/2}$ by (1). The factor $1/H$ pays for the supremum factor H. Under independent coefficients the fixed-time $2r$th moment root is $O(\sqrt{r/H})$ by the variance proxy $O(H^{-1})$. Therefore

$$(\mathbb{E}\,\|T_H\|_\infty^{2r})^{1/(2r)} \ll (rH)^{1/(2r)}\sqrt{r/H} + p^{-1/(4r)}. \tag{3}$$

The second term is absent under the independent law. Summing dyadic blocks above H_* with $r = \lceil \log(16H_*)\rceil$, pinning and interpolating the mesh points as above give an L^{2r} tail bound

$$O\left(\sqrt{\frac{r}{H_*}} + \log p\, p^{-1/(4r)} + H_*/p + p^{-1/2}\right). \tag{4}$$

The independent tail uses just the first term. Taking a union over at most $O(H)$ partial cutoffs per dyadic block with $r = \lceil\log(16H)\rceil$ adds a bounded factor to its moment root, so summability of the expected maxima gives almost sure uniform convergence.

2.3 Time samples and correlated noise

Let m be a large integer, later chosen to be a small constant times $\log p$. Put $H_0 = \lfloor m^{1/4} \rfloor$. Independently round coefficients $|h| \le H_0$ to $4\,\mathrm{Bin}(\lceil m^2 \rceil, (x_h+2)/4)/\lceil m^2 \rceil - 2$. The L^s path error is $O_s(\log m/m)$ by the individual binomial moments and the absolute multiplier sum. There are at most $\exp(O(H_0 \log m))$ low output vectors, and the conditional probability polynomial of each has character norm at most the same exponential order by the separate-degree estimate.

Take blocks $H = 2^j H_0$, $j \ge 1$, through the first $H \ge m \log m$. In such a block let $y = A_H x$ be the $I_H = 16H$ real and imaginary parts of $\sqrt{H} \sum b_h x_h e(ht)$ on the $8H$ equally spaced mesh samples. Discrete Parseval gives $\|A_H\|_{2\to 2} \ll 1$ and $\|y\|_2^2 \ll H$ on the cube. In the independent model, the sample vector has the directional subgaussian bound $\mathbb{E} \exp(t \cdot y) \le \exp(C \|t\|_2^2)$. For every subset S of samples, Gaussian integration then gives

$$\mathbb{E} \exp\left(\lambda \sum_{i \in S} y_i^2 \right) \le D^{|S|} \tag{5}$$

for absolute constants $\lambda > 0, D$.

For reconstruction, choose a smooth Fourier multiplier equal to one on $[-H, H]$, zero outside $[-2H, 2H]$, with kernel V_H. Discrete orthogonality on the $8H$ grid recovers the block as the sum of sampled values against weights $(8H)^{-1} V_H(t - j/(8H))$. The kernel bound $|V_H(u)| \ll H/(1+H^2 \operatorname{dist}(u,\mathbb{Z})^2)$, obtained by two summations by parts, makes the absolute sum of the weights uniformly bounded. Thus if normalized samples are changed by at most τ in each component, the sup error after reconstruction and pinning is $O(\tau/\sqrt{H})$.

Use scale $\tau_H = (H/m)^{3/4}$ for $H \le m$, and $\tau_H = A(H/m)^{1/4}$ for $H > m$, with large absolute A. The former blocks will be compared on all slots, and the latter on sparse subsets. Both parts of the sum $\sum_H \tau_H/\sqrt{H}$ are geometric and give $O(m^{-1/2})$.

Lemma 2.2. *Fix a positive constant J_0. For I real coordinates and noise scale $\tau > 0$, put $T = C_0 \max(1, \tau^4)$. There is a probability density on $\mathbb{R}^I$, with noise vector η, with these properties:*

1. $\mathbb{P}(\max_i |\eta_i| > \tau/8) = O(I^{-J_0})$.
2. *For each slot set S, the probability that all $|\eta_i| > \tau/8$ in S is at most $2(Ce^{-c\sqrt{T}})^{2|S|}$.*
3. *The characteristic function of any D-coordinate marginal, $D \ge 1$, has support in a ball of radius $R(D) = CTI^{1/4}D^{1/4}/\tau$. Every mixed first derivative, once in each of r distinct coordinates, is bounded by $(C\tau)^r$ up to an absolute constant. These derivatives vanish outside the same ball.*

Proof. Take a real even smooth function f supported in $[-1/8, 1/8]$, with $\|f\|_2 = 1$ and $|\widehat{f}(u)| + |\widehat{f}'(u)| \le Ce^{-c\sqrt{|u|}}$. For example, use the normalized density of the sum of independent uniforms on $[-c_*/j^2, c_*/j^2]$, $j \ge 1$, with c_* small. Its sinc product has at least $c\sqrt{|u|}$ factors bounded by $1/2$ for large $|u|$, and a differentiated term loses at most one such factor and contributes $O(j^{-2})$. This also proves smoothness and the requisite derivative decay.

Take a smooth even cutoff χ decreasing in absolute argument, equal to one on $[-1, 1]$ and zero outside $[-2, 2]$. Form a partition with $\phi_0(u) = \chi(u/T)$ and $\phi_j(u) = \chi(u/(2^j T)) - \chi(u/(2^{j-1}T))$ for $j \ge 1$. Its terms are nonnegative, sum to one, and $\sum_j |\phi_j'| \ll 1$. The upper radii are $U_j = 2^{j+1}T$, and if $\phi_j(u) > 0$ for $j \ge 1$, then $|u| > U_j/4$.

Retain band configurations $(j_1, \dots, j_I)$ with $\sum_i U_{j_i}^4 \le C_1 I T^4$. Let w be the sum of their product partition functions, so $0 \le w \le 1$. We use the squared-amplitude device for positive Fourier

smoothing as in [8, Section 7]. The restriction on the joint amplitude will control the Fourier support of every selected marginal. Define $\widehat{P}(u) = w(u)\prod_i \widehat{f}(u_i)$. Take the density $|P|^2/\|P\|_2^2$ and dilate by τ in each coordinate. As usual, the Fourier normalization can be taken unitary, with absolute constants adjusted for the characteristic-function convention.

For a rejected configuration the bounds $U_{j_i} \le 4\max(T, |u_i|)$ imply $\sum_i |u_i|^4 > C_2 I T^4$, with C_2 as large as needed by choosing C_1. Under the iid $|\widehat{f}|^2$ probability law, $|u_i|^4$ has all moments. For each fixed integer k, the $2k$th moment of the sum of centered copies is $O_k(I^k)$ by independence and expansion, since singleton factors vanish. Thus the rejected frequency mass is $O(I^{-2J_0})$ on taking k large enough. Choosing C_2 above four times the scalar fourth moment also bounds this mass by $1/4$ for every I by Markov's inequality. It follows that the amplitude P is $O(I^{-J_0})$ from the product amplitude in L^2, and hence the normalized densities are within the same order in total variation by Cauchy–Schwarz. This proves the first property; the uniform mass bound gives $\|P\|_2^2 \ge 3/4$.

For the joint bound, write f_j for the inverse transform of $\phi_j \widehat{f}$. Fix the band labels on S. The corresponding partial sum of the amplitude factors as the selected f_j times an amplitude in the complementary coordinates with L^2 norm at most one, because its partition multiplier lies between zero and one. On $|x| > 1/8$ we have $f_0 = -\sum_{j\ge1} f_j$ in L^2. Moreover $\sum_{j\ge1} \|f_j\|_2 \ll e^{-c\sqrt{T}}$ by the lower band radii and the decay of $\widehat{f}$. Summing the restricted L^2 bounds over selected band labels gives $(Ce^{-c\sqrt{T}})^{|S|}$. Squaring and normalizing proves the second property.

The support of $\widehat{P}$ has fourth power sum at most $C_1 I T^4$. Its projection to D coordinates has Euclidean radius at most $CTI^{1/4}D^{1/4}$ by Cauchy–Schwarz. The Fourier transform of $|P|^2$ is an autocorrelation of $\widehat{P}$, so the support of the selected characteristic function is in the corresponding difference set, after accounting for dilation. Finally a mixed derivative of $\widehat{P}$ on distinct slots is pointwise bounded by the product of $|\widehat{f}'| + C|\widehat{f}|$ there and $|\widehat{f}|$ elsewhere, by the partition derivative estimate. Its L^2 norm is $O(C^r)$. By Plancherel and Cauchy–Schwarz this bounds the product absolute moment of the corresponding η_i by $O((C\tau)^r)$. These moments justify differentiating the characteristic function, with the stated support. □

Draw the noises independently across blocks and independently of every input. In the dense blocks observe the noisy samples $y_i + \eta_i$. In each sparse block replace a noisy sample by zero whenever its absolute value is at most $\tau_H/2$. Clip the final output in either case to a centered interval of radius $C\sqrt{H} + \tau_H$, containing the deterministic range of y_i and its small perturbation. On the good event from Lemma 2.2(1) the sample error is $O(\tau_H)$.

2.4 Comparison on selected marginals

Select every dense slot. In each sparse block H select D_H slots, with D_H bounded below by zero and above by the cutoff specified in the next section. We first compare the noisy real vectors before thresholding or clipping, together with the rounded low output. Deterministic maps can only decrease total variation. Let g be the difference density of the selected noisy coordinates for one fixed low output label. Its Fourier transform factors as the product of the noise marginal transforms times the difference of the two input characteristic expectations, weighted by the conditional low label polynomial. The smoothing ensures a density, and the selected characteristic function is compactly supported.

For a Fourier test vector t_H, the character norm of $\exp(it_H \cdot A_H x)$ is at most $\exp(C\sqrt{H}\,\|t_H\|_2)$ by its absolutely convergent power series and $\left\|A_H^T t_H\right\|_1 \ll \sqrt{H}\,\|t_H\|_2$. A mixed derivative in r_H distinct variables in block H inserts a product of r_H sample coordinates. Its cube bound is at most $(CH/r_H)^{r_H/2}$ by the energy estimate and the arithmetic–geometric mean inequality. Its character

norm is bounded by $\exp(Cr_H \log(eCH/r_H))$ by (2), since the block uses $O(H)$ Fourier inputs. The bound is one for $r_H = 0$. Taking C sufficiently large, one may bound this exponent using D_H in place of r_H for all $r_H \le D_H \le I_H$. The low label costs $\exp(o(m))$. With the derivative bounds for the noise in Lemma 2.2, these estimates control all mixed first derivatives of $\widehat{g}$ by (1) and the product rule.

To bound the L^1 norm of g, assign $s_i = 1 + \tau_H$ in block H. Multiply and divide by the square root of $\prod_i (1 + (x_i/s_i)^2)$, then use Cauchy–Schwarz and Plancherel. If $D = \sum_H D_H$, this gives

$$\|g\|_1 \le \Big(\prod_i \pi s_i\Big)^{1/2} \left[C^D \sum_{U \subseteq \{1,\dots,D\}} \Big(\prod_{i \in U} s_i^{-2}\Big) \|\partial_U \widehat{g}\|_2^2\right]^{1/2}.$$

The weights absorb the powers of τ_H from the noise derivatives. The support is a product of balls. A ball of radius R in D_H dimensions has volume at most $(CR/\sqrt{D_H})^{D_H}$. Combining these facts with (1) gives, after summing over the low labels, a total variation upper bound of

$$\begin{aligned} &p^{-1/2} \exp\big(o(m) + C \textstyle\sum_H E_H(D_H)\big), \qquad \text{where} \\ &E_H(D_H) = \sqrt{H} R_H(D_H) + D_H \log(eCH/D_H) \\ &\qquad\qquad + D_H \log\left(\frac{C(1+\tau_H)T_H}{\tau_H}(I_H/D_H)^{1/4}\right), \end{aligned} \tag{6}$$

where $T_H = C_0 \max(1, \tau_H^4)$ and R_H is as in Lemma 2.2. Terms with $D_H = 0$ are omitted. The exponent includes the 2^D derivative sum and the polynomial factor for the number of Fourier inputs.

For dense blocks $H \le m$ take $D_H = I_H$. The contribution is $O(H/\tau_H + H[1 + \log(1/\tau_H)])$, summable to $O(m)$ because $H/\tau_H = m^{3/4} H^{1/4}$. In a sparse block $H > m$ define

$$K_H = \min\left(I_H, \lceil \max(m^{1/2}, C_2 I_H e^{-c_2 \tau_H^2}) \rceil\right)$$

with c_2 sufficiently small and C_2 sufficiently large. We allow $D_H \le \min(I_H, K_H + 1)$. Since $H \le 2m \log m$ up to an absolute factor and $\tau_H^2 \ll \sqrt{\log m}$, the $m^{1/2}$ term is eventually smaller than the exponential term throughout this range. Thus $x = D_H/I_H \le Ce^{-c_2\tau_H^2}$. The first contribution of this block to (6) is $O(H(T_H/\tau_H)x^{1/4})$. The others are bounded by $O(Hx[1 + \log(1/x) + \log T_H + \log(1 + 1/\tau_H)])$. Together they are $O(H(1+\tau_H)^C e^{-c_3\tau_H^2})$. Since $\tau_H^2 = A^2\sqrt{H/m}$, their dyadic sum is $O(m)$. In particular, every selected marginal compares uniformly in total variation at cost $p^{-1/2}\exp(Cm)$.

2.5 Sparse comparison and transport

Under the independent model, a specified set of D_H sparse slots is simultaneously nonzero after thresholding only if in each of those slots either $|y_i| \ge \tau_H/4$ or $|\eta_i| > \tau_H/8$. Split the slots between these alternatives. For the first use (5) and Markov's inequality, and for the second use Lemma 2.2(2). Independence of noise and input gives the product bound, summed over at most 2^{D_H} allocations. With A large the joint activity probability is at most $\exp(-c_4\tau_H^2 D_H)$. These bounds multiply across blocks, since the independent Fourier coordinates and the noises are independent across blocks.

The argument uses truncated inclusion–exclusion with a factorial error bound, as in read-once Boolean tests [9, Section 3], here for arbitrary bounded tests of multiple observed values. For a realized sparse block output with N_H nonzero entries, expand its point mass as a product of

zero point masses plus coordinate differences. Retain terms involving at most K_H differences. If $N_H \le K_H$ this is exact. Otherwise the signed variation error is at most

$$B_H = 2 \cdot 3^{K_H+1} \binom{N_H}{K_H+1},$$

by the binomial bounds $\binom{N_H}{j} \le \binom{N_H}{K_H+1}\binom{K_H+1}{j}$ for $j \le K_H$ and the factor 2^j for signed differences. Take $B_H = 0$ if the cap keeps all coordinates. For a product across sparse blocks, telescoping bounds the aggregate variation error by $\prod_H (1 + B_H) - 1$.

Expand this last product to estimate its expectation. Each factorial count is a sum over sets of prescribed active slots. Under the independent law the term for a chosen block is bounded by

$$\left(\frac{C I_H e^{-c_4 \tau_H^2}}{K_H + 1} \right)^{K_H+1}.$$

Take $c_2 < c_4$ and C_2 large in the definition of K_H. This is at most $\exp(-cK_H)$, so the whole expected error is smaller than any fixed power of m^{-1}. Under arithmetic input compare each selected joint activation by the selected-marginal bound (6). The sum of all weights and subset counts is $\exp(O(m))$, because $\sum_H K_H \log(eI_H/K_H) = O(m)$ by the same sparse budget estimate. Hence the arithmetic expected truncation error is bounded by $O_{J_0}(m^{-J_0}) + p^{-1/2}\exp(Cm)$.

For any test of the full observations bounded by one, the truncated expectation is a sum of tests on all the dense outputs, the low rounded vector, and at most K_H selected sparse coordinates in each block. Evaluate the other sparse outputs at zero in each term of the signed expansion. The same subset count and signed amplification is $\exp(O(m))$, so (6) compares these truncated expectations with error $p^{-1/2}\exp(Cm)$. Combining the two truncation errors shows that the full genuine observation laws differ in total variation by

$$O_{J_0}(m^{-J_0}) + p^{-1/2}\exp(Cm)$$

uniformly in the fixed inverse parameter.

Take $m = \lfloor a \log p \rfloor$ for sufficiently small $a > 0$ (depending on s through the chosen J_0). Maximal coupling of the observations gives a common reconstructed path except on an event of that probability. Decoded paths have sup norm $O(\log m)$ by clipping and stable interpolation. For either input the reconstruction error on the compact noise event is $O(\sum_H \tau_H/\sqrt{H}) = O(m^{-1/2})$ in addition to low binomial rounding. Its complementary probability is $O(\sum_H I_H^{-J_0}) = O(m^{-J_0/4})$, by the first noise bound, and the deterministic sup norms of both the decoded and truncated true paths are $O(\log m)$. Choose J_0 large depending on s. These events have negligible L^s cost. At $H_* \asymp m \log m$, the tails in (4) cost $O(m^{-1/2})$ in L^s. This proves the upper bound in Theorem 2.1, including the mixture over B.

By [7], the limit law assigns probability at most $\exp(-c_0\epsilon^{-2})$ in any uniform ball of sufficiently small radius ϵ, uniformly over its center. The union over at most $(p-1)^2$ arithmetic support points has mass at most $1/2$ for $\epsilon = c_1/\sqrt{\log p}$ with c_1 small. In any coupling the independent path is therefore at distance at least ϵ from the arithmetic support with probability at least $1/2$. This gives the W_1 lower bound, and hence the stated lower bound for every $s \ge 1$.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, proof drafting, and internal checks. No datasets or numerical certificates are required for the proof.

3 Fixed-parameter Kloosterman paths

In a normalized prime Kloosterman path, fix the coefficient of x^{-1} at any nonzero residue and sample the coefficient of x uniformly among the nonzero residues. We prove convergence in law in the uniform path metric, with W_1 rate $O(\sqrt{\log\log p/\log p})$ uniformly over the fixed parameter. The main estimate integrates the moment of each completed Fourier block before bounding its maximum. Integration imposes a frequency balance that cancels the block-length factor in the estimate for the maximum. Already at fourth order the argument gives the short-interval maximal bound used in the study of large Kloosterman sums. A polynomial comparison based on binomial observations and Sato–Tate conditional moments also gives shrinking-target proportions whose logarithms lie between negative constant multiples of ϵ^{-2}, for radii down to a constant times the Wasserstein rate.

Keywords Kloosterman sums, Kloosterman paths, Fourier completion, Wasserstein distance, shrinking targets.
Mathematics Subject Classification 11L05, 11T23, 60F17.

3.1 Introduction

For an odd prime p, write $e(t) = \exp(2\pi i t)$ and $e_p(x) = e(x/p)$. For $A, B \in \mathbb{F}_p^\times$ let $K_p(A, B)$ be the polygonal path with successive vertices

$$p^{-1/2} \sum_{x=1}^{j} e_p(Ax + B/x), \qquad 0 \le j \le p-1,$$

at times $j/(p-1)$. We fix B and take A uniformly distributed over $\mathbb{F}_p^\times$. Write $\mu_{p,B}$ for this path law and μ for the law of

$$K(t) = tX_0 + \sum_{h\neq 0} X_h v_h(t), \qquad v_h(t) = \frac{e(ht)-1}{2\pi i h},$$

where $(X_h)_{h\in\mathbb{Z}}$ are independent Sato–Tate variables with density $\sqrt{4-x^2}/(2\pi)$ on $[-2,2]$. The series converges almost surely in the uniform metric. We write $\|\cdot\|_\infty$ for the supremum norm and W_1 for its Wasserstein distance.

Kowalski and Sawin [5] proved finite-dimensional convergence for $B = 1$ and convergence in law on the whole path space when both parameters vary. They isolated a short-interval estimate as the missing input for fixed-B tightness. Lebreton [12, Section 1.2] uses a related maximal short-interval estimate to study the distribution and location of large partial sums and identifies the same outstanding fixed-parameter case. Kowalski and Untrau [6, Section 1.3(4)] asked for a quantitative Wasserstein version of path convergence. The two-parameter rate in [7] uses a fourth-moment estimate that averages both parameters for the highest-frequency remainder.

Theorem 3.1. *Uniformly in $B \in \mathbb{F}_p^\times$,*

$$W_1(\mu_{p,B}, \mu) \ll \sqrt{\frac{\log\log p}{\log p}}$$

as p tends to infinity through primes. For every fixed $0 < \eta < 1/4$, uniformly in the same B,

$$\mathbb{E}_A \max_{\substack{I\subseteq\mathbb{F}_p \text{ interval} \\ |I|\le p^{1/2+\eta}}} \left| p^{-1/2} \sum_{x\in I\setminus\{0\}} e_p(Ax + B/x) \right|^4 \ll p^{-1/2+\eta} + p^{-1/2}(\log p)^4.$$

In particular the fourth moment on the left is $O_\eta(p^{-\eta})$.

The short-interval bound verifies Lebreton's Assumption (Tightness). His Assumption (Law) holds for the same families by the character correlation bound below, since the Sato–Tate density has power decay at its endpoints. His Theorems 1.1 and 1.2, on the upper tail of the maximum and on its typical location among large paths, therefore apply with B fixed, including $B = 1$. The short-interval bound also follows under the fourth-order part of Assumption (Law) for the general families of [12]; we spell out this deduction after the proof.

For $L \geq 1$, let $\mathcal{H}_L$ consist of continuous paths g with $g(0) = 0$, $g(1) \in [-2, 2]$, $|g(t) - g(u)| \leq L\,|t - u|^{1/2}$, and

$$2\pi i h \int_0^1 [g(t) - t g(1)] e(-ht)\, dt \in [-2, 2] \qquad (h \neq 0).$$

These coefficient conditions describe the support of μ [10]. The limit small-ball theorem in [7] gives $\mu(B_\infty(g, \epsilon)) \geq \exp(-C_L \epsilon^{-2})$ for $g \in \mathcal{H}_L$. Kyselov [11] obtained a lower bound with an additional squared logarithm for an interior class, together with a fixed-radius arithmetic consequence. The direct result of [7] for the two-parameter family reaches radii of order $(\log p)^{-1/4}$. The next result preserves a lower bound of the form $\exp(-C_L \epsilon^{-2})$ at the smaller radius.

Theorem 3.2. *For each $L \geq 1$ there are constants $D_L, C_L > 0$ such that for all sufficiently large odd primes, every $B \in \mathbb{F}_p^\times$, every $g \in \mathcal{H}_L$ and*

$$D_L \sqrt{\log\log p / \log p} \leq \epsilon \leq 1/4,$$

one has $\mu_{p,B}(B_\infty(g, \epsilon)) \geq \exp(-C_L \epsilon^{-2})$. There are absolute $d, c, \epsilon_0 > 0$ such that, for all sufficiently large odd primes, every $B \in \mathbb{F}_p^\times$ and every continuous g, one has

$$\mu_{p,B}(B_\infty(g, \epsilon)) \leq \exp(-c\epsilon^{-2}) \qquad \left(d/\sqrt{\log p} \leq \epsilon \leq \epsilon_0\right).$$

3.2 Completion and integrated block moments

Put $W_h = p^{-1/2} \sum_{x \neq 0} e_p((A - h)x + B/x)$ and let $J_p = \{h \in \mathbb{Z} : |h| \leq (p-1)/2\}$. Then $W_h = \mathrm{Kl}_2(B(A-h); p)$ is real, $|W_h| \leq 2$, and $W_h = -p^{-1/2}$ if $A = h$. Let Q_p be the path using mesh j/p and a constant final segment. Thus $K_p(t) = Q_p((p-1)t/p)$ and $\|K_p - Q_p\|_\infty \leq p^{-1/2}$. Write I_p for linear interpolation at the $p + 1$ points j/p. Discrete Fourier inversion, followed by summation of a geometric progression, gives

$$Q_p = I_p R_p, \qquad R_p(t) = tW_0 + \sum_{h \in J_p \setminus \{0\}} b_{p,h} W_h (e(ht) - 1), \qquad b_{p,h} = \frac{1}{p(1 - e_p(-h))}. \tag{7}$$

Indeed inversion of W_h recovers each normalized increment, including the zero increment at $x = p$. In the indicated representatives $|b_{p,h}| \ll |h|^{-1}$, and $b_{p,h} = (2\pi i h)^{-1} + O(p^{-1})$. Interpolation is a contraction in the supremum norm.

We use the following character correlation estimate, with its dependence on representation degrees as in [7, Lemma 6.3]. For distinct shifts h_j modulo p and positive degrees k_j, $1 \leq j \leq s$,

$$\left| \mathbb{E}_A \prod_{j=1}^{s} \chi_{k_j}(W_{h_j}) \right| \leq 4(s+1) p^{-1/2} \prod_{j=1}^{s} (k_j + 1), \qquad \chi_k(2\cos\theta) = \frac{\sin((k+1)\theta)}{\sin\theta}. \tag{8}$$

Here B can be any nonzero residue. To recall the uniformity, the change $y = BA$ gives translates by Bh_j of the normalized rank-two Kloosterman sheaf. Product geometric monodromy follows from the

independence argument in [5, Lemma 2.4]. At the distinct finite singularities the nontrivial tame unipotent monodromy precludes twisted isomorphisms of different translates. Their symmetric-power tensor has rank $D = \prod_j (k_j + 1)$ and no geometric invariants. Its finite ramification is tame and the break at infinity is at most $1/2$, since the original factors have that break and upper ramification groups acting trivially remain trivial under tensor operations. On the affine line minus the s finite punctures, the Euler characteristic formula bounds $\dim H_c^1$ by $D(s-1) + D/2 \le sD$, with $H_c^0 = H_c^2 = 0$. The trace formula and the weight bound give an unnormalized sum bounded by $sD\sqrt{p}$. Changing the at most $s+1$ omitted or added points and normalizing gives (8). Each nontrivial character has Sato–Tate mean zero.

The weighted character norm $\|P\|_{\mathrm{ch}}$ of a polynomial in real variables is the sum of the absolute coefficients in the product character basis, with each weighted by the product of its representation dimensions. It is submultiplicative by the Clebsch–Gordan formula and dimension additivity. In particular $\|x_h^m\|_{\mathrm{ch}} = 2^m$. Thus any monomial of degree $2r$ in the W_h differs in expectation from its independent counterpart by at most $4(2r+1)2^{2r}/\sqrt{p}$. The corresponding estimate holds for complex linear combinations by the triangle inequality.

Lemma 3.3. *Let $T_H(t) = \sum_{h \in \mathcal{I}} b_h W_h e(ht)$, where $\mathcal{I} \subseteq J_p$ and $H/2 < |h| \le H$ on $\mathcal{I}$, with $|b_h| \le C/H$. Then for every integer $r \ge 1$,*

$$\left(\mathbb{E}_A \|T_H\|_\infty^{2r}\right)^{1/(2r)} \ll (rH)^{1/(2r)} \sqrt{r/H} + p^{-1/(4r)}.$$

In the independent Sato–Tate model the second term is absent. Constants depend only on C.

Proof. Parseval and Cauchy–Schwarz applied to T_H^r give $\|T_H\|_\infty^{2r} \le (2rH+1) \int_0^1 |T_H|^{2r}$. At a fixed time, the independent centered bounded model has moment root $O(\sqrt{r/H})$ by the subgaussian estimate for a linear combination, since $\sum |b_h|^2 \ll H^{-1}$. In the integrated arithmetic moment expansion only terms with $h_1 + \cdots + h_r = h_{r+1} + \cdots + h_{2r}$ remain. Any $2r-1$ indices determine the last, giving a sum of absolute coefficients at most

$$(\max_h |b_h|) \left(\sum_h |b_h| \right)^{2r-1} \ll C_1^{2r}/H$$

for a constant C_1. This cancels H in the bound for the maximum. The discrepancy contribution to the $2r$th moment is at most a constant to power $2r$ times $(r+1)^2 p^{-1/2}$, which gives the claimed second term after taking roots. □

3.3 Fixed-parameter convergence

Let $M = \lfloor a \log p \rfloor$ for sufficiently small absolute $a > 0$ and $r = \lceil \log(16M) \rceil$. Applying Lemma 3.3 to dyadic blocks above M, pinning at zero and summing by Minkowski yields a tail of L^{2r} norm at most

$$O(\sqrt{r/M}) + O(\log p \, p^{-1/(4r)}). \tag{9}$$

Indeed $M^{1/(2r)}$ is bounded and the first terms decrease geometrically with the block index. The independent tail has the same first term. To check almost sure convergence along all symmetric cutoffs, apply the independent block bound to every partial block of a dyadic interval and take the maximum over at most $O(H)$ cutoffs. With $r = \lceil \log(16H) \rceil$ this adds only a bounded factor to the moment root, and the expected maxima are summable over dyadic H. By (7) we compare the mesh values with the interpolant of the low polynomial. The low polynomial has derivative

bounded by $O(M)$ deterministically. Its interpolation error is $O(M/p)$, as is the coefficient error from replacing $b_{p,h}$ by $(2\pi ih)^{-1}$ for $|h| \le M$.

It remains to compare the two laws of the low path $x_0 t + \sum_{0<|h|\le M} x_h v_h(t)$. The tensor Bernstein argument of [7] applies with sharp cutoff: take degrees $n_0 = M$, $n_h = \lceil (M/|h|)^2 \rceil$ for $h \neq 0$, and given x replace x_h by $4\,\mathrm{Bin}(n_h, (x_h+2)/4)/n_h - 2$ independently. This replacement is centered about x_h with a subgaussian variance proxy $O(1/n_h)$. In the low path the nonzero-frequency variance proxy is $O(1/M)$, since $|v_h| \ll 1/|h|$. A Bernstein grid of $O(M)$ points bounds the expected sup norm of the replacement error by $O(\sqrt{\log M/M})$ through its pointwise subgaussian tails. The endpoint has expected error $O(M^{-1/2})$.

An arbitrary function of the rounded vector bounded by 1 becomes a polynomial of separate degrees at most n_h when averaged conditionally on x. Its character norm is at most $\prod_h (n_h+2)^2$. Indeed the angular substitution $x_h = 2\cos\theta_h$ gives Fourier coefficients bounded by 1. Group the signs into cosine products and use $\cos(j\theta) = (\chi_j - \chi_{j-2})/2$ for $j \ge 2$, with $\cos\theta = \chi_1/2$, and bound each cosine's weighted character norm by $j+1$. Summing over modes gives the displayed bound. Its logarithm is $O(M)$, by $\log(M!) \ge M\log M - M$. By (8), the two rounded vector laws therefore have total variation distance $O(Mp^{-1/2}\exp(CM))$. A maximal coupling and the deterministic $O(\log M)$ bound on the low path norm bound their transport contribution by a power of p^{-1} for sufficiently small a. Combining this comparison, the rounding errors, (9) and the mesh error proves the first assertion of Theorem 3.1.

For the second assertion take $r = 2$ and apply Bernstein's derivative inequality and Lemma 3.3 to all dyadic blocks. An increment over distance at most δ in a block of degree H is bounded by $C\min(1, H\delta)\,\|T_H\|_\infty$. The geometric series above and below δ^{-1} gives

$$\left(\mathbb{E}_A \omega(R_p, \delta)^4\right)^{1/4} \ll \delta^{1/4} + p^{-1/8}\log p, \tag{10}$$

where the affine term costs $O(\delta)$. At mesh times, interval sums equal increments of R_p by (7). An interval wrapping past zero splits into at most two such increments. Taking $\delta = p^{-1/2+\eta}$ and fourth powers proves the result.

The same proof uses only bounded real Fourier coefficients and a uniform fourth-monomial discrepancy $O(m^{-1/2})$ from a bounded independent centered model. The Fourier completion is identical for a family of m-periodic summands using the normalized discrete Fourier transform. For even m, include one Nyquist representative. The zero index is the affine slope. With the first mesh vertex at zero, use $Q(j/m) = m^{-1/2}\sum_{x=1}^{j}\varphi_a(x)$ and set $\varphi_a(m) = \varphi_a(0)$. Every circular interval splits into at most two mesh intervals, so the same increment estimate applies. Lebreton's Assumption (Law) includes this fourth-order hypothesis. It therefore implies his Assumption (Tightness), with fourth moment and any fixed $0 < \eta < 1/4$.

3.4 Polynomial transfer for small balls

The observations serve two purposes. Given the original path coefficients, they give a smoothed reference that reaches the target with sufficient probability. Given the observations, the independent Sato–Tate model has centered residuals, which lets us compare the moments of the distance to the reference.

Keep M, r as above. Take an even smooth function γ supported in $[-1,1]$, with values in $[0,1]$, equal to $(1+3s^2)^{-1}$ on $|s| \le 1/2$ and less than 1 for $s \neq 0$. For $0 < |h| \le M$ put $\nu_h = 3\gamma(h/M)/(1-\gamma(h/M))$, and put $\nu_0 = M$. Independently for each coordinate, randomize D_h between the neighboring integers of ν_h so that $\mathbb{E}[D_h/(D_h+3)] = \nu_h/(\nu_h+3)$. Given x_h, D_h, observe $k_h \sim \mathrm{Bin}(D_h, (x_h+2)/4)$ independently. Let $Y = (D_h, k_h)_{|h|\le M}$, set $z_h = (4k_h - 2D_h)/(D_h+3)$,

and set $z_h = 0$ outside the observation window. The reference path is

$$S(Y)(t) = tz_0 + \sum_{0<|h|\le M} b_{p,h} z_h (e(ht) - 1).$$

In the independent model the transformed Sato–Tate coordinate $(X_h + 2)/4$ is $\mathrm{Beta}(3/2, 3/2)$. Given Y, it remains independent of the other coordinates and has law $\mathrm{Beta}(k_h + 3/2, D_h - k_h + 3/2)$. The mean of X_h given Y is z_h. Its centered conditional variance proxy is $O(1/(1 + D_h))$. In fact the negative log-density on $[-2, 2]$ has second derivative at least $(1 + D_h)/16$, unchanged under exponential tilts. For an exponentially tilted variable U with negative log-density V, integration by parts gives $\mathrm{Cov}(U, V'(U)) = 1$. Using an independent copy and the curvature bound then gives $\mathrm{Var}(U) \le 16/(1 + D_h)$. The centered cumulant bound follows by integrating this variance under the tilt. The integration by parts is valid because both beta shape parameters exceed one even after exponential tilting.

In the other direction the mean of z_h given x_h is $\gamma(h/M)x_h$ when $h \neq 0$, and $Mx_0/(M + 3)$ when $h = 0$. The fluctuations are subgaussian with variance proxy $O(1/(1+\nu_h))$. For fixed D_h this follows from the centered binomial bound, and the two possible conditional means differ by at most $6/(\lfloor \nu_h \rfloor + 3)^2$. Thus the conditional expectation of $\|S - \mathbb{E}[S \mid X]\|_\infty$ is $O(\sqrt{r/M})$ uniformly. To check this, the nonzero-frequency pointwise variance proxy is $O(1/M)$ because $1/(1+\nu_h) \ll (h/M)^2$, and then apply a Bernstein grid and integrate its union bound as for rounding. The endpoint contributes $O(M^{-1/2})$.

The conditional mean $\mathbb{E}[S \mid X]$ is the smooth cutoff of the limit path with symbol $\gamma(h/M)$ on the pinned periodic part, up to deterministic error $O(M/p + 1/M)$. Its operator norm in the supremum metric is bounded. The kernel has size $O(M)$ and $O(1/(M \operatorname{dist}(t, \mathbb{Z})^2))$, by summation by parts twice, so its absolute half-moment is $O(M^{-1/2})$. Applying this estimate to the pinned periodic part of a path $g \in \mathcal{H}_L$ gives approximation error $O((L + 2)M^{-1/2})$. It follows that when K is within a sufficiently small absolute multiple of ϵ of g, its conditional reference mean is within $\epsilon/4$ of g for ϵ in Theorem 3.2, by increasing D_L. The uniform noise bound and Markov's inequality now give, for $F(Y) = \mathbf{1}_{\{\|S(Y)-g\|_\infty < \epsilon/2\}}$,

$$v := \mathbb{E}_X \mathbb{E}[F \mid X] \ge \tfrac{1}{2} \exp(-C_L \epsilon^{-2}). \tag{11}$$

Here we used the limit small-ball theorem of [7] at the smaller radius.

For any bounded function $G(Y)$, its conditional average given an input vector x is a polynomial with separate degrees at most $\lceil \nu_h \rceil$, bounded by $\sup |G|$ on the cube. The Bernstein character norm estimate used earlier yields $\|\mathbb{E}[G \mid x]\|_{\mathrm{ch}} \le D \sup |G|$ with

$$D = \prod_{|h|\le M} (\lceil \nu_h \rceil + 2)^2 \le \exp(CM).$$

This uses the factorial estimate and $\lceil \nu_h \rceil + 2 \ll (M/|h|)^2$ for $h \neq 0$. In particular, $v_B = \mathbb{E}_A \mathbb{E}[F \mid W]$ differs from v by at most $4(2M + 2)Dp^{-1/2}$. Choose a so $D \le p^{1/16}$. For the radii in the theorem, $v \ge p^{-1/128}$ for all sufficiently large p, since $\epsilon^{-2} = O_L(\log p / \log\log p)$. Thus $v_B \ge v/2$.

We next control the residual $R_p - S$ under the joint arithmetic law conditioned on $F = 1$. Write its trigonometric blocks with coefficients $b_{p,h}(W_h - z_h)$ and remove the value at zero to pin them. For any tuple of length $2r$, the polynomial

$$\mathbb{E}\left[F(Y) \prod_{j=1}^{2r} (x_{h_j} - z_{h_j}) \,\middle|\, x \right]$$

has character norm at most $D4^{2r}$. Expand the product: each factor x_{h_j} costs 2 in character norm, each z_{h_j} costs at most 2 in the bound for the coefficient function of Y, and the conditional polynomial bound contributes D. Each character uses at most $2M+1+2r$ shifts. Applying (8) and dividing by v_B gives discrepancy at most $4(2M+2r+2)D4^{2r}p^{-1/2}/v_B$ per tuple.

Conditionally on Y under the independent law, the centered residual has pointwise variance proxy $O(1/M)$ in the block $0<|h|\le M$, and $O(1/H)$ in each block $H/2<|h|\le H$ above M. This follows from the posterior variance estimates and $|b_{p,h}|\ll 1/|h|$. The integrated independent moments thus give the first term of Lemma 3.3 for the high blocks, and its analogue with $H=M$ for the low block. The arithmetic conditional moment is bounded by the independent conditional moment times $v/v_B\le 2$, plus the integrated discrepancy divided by v_B. For the high blocks the same frequency balance cancels H in the discrepancy bound, since neither the observations nor F depends on time. After taking roots the comparison error is $O(p^{-3/(16r)})$ per high block. For the low block one can discard frequency balance and bound the sum of absolute coefficients before taking roots by $O(\log M)^{2r}$. The low-block error is therefore $O((rM)^{1/(2r)}\log M\, p^{-3/(16r)})$, which is smaller than $M^{-1/2}$ for $r=\lceil\log(16M)\rceil$. The endpoint follows by the same moment comparison with the single coordinate x_0-z_0. Summing the bounds gives $\mathbb{E}[\|R_p-S\|_\infty \mid F=1,B]\ll\sqrt{r/M}$.

Since $\|I_pS-S\|_\infty\ll M/p$ deterministically, interpolation and the mesh bound give the same conditional expected residual for K_p-S, up to $O(M/p+p^{-1/2})$. By increasing D_L, conditional on $F=1$ at least half of the arithmetic paths lie in the open ϵ-ball about g. Thus $\mu_{p,B}(B_\infty(g,\epsilon))\ge v_B/2\ge v/4$. This proves the lower bound of Theorem 3.2 after absorbing the constant prefactor.

For the upper bound choose $m=\lfloor k\epsilon^{-2}\rfloor$ with $k>0$ sufficiently small. In the stated radius range $m<p/2$ for all sufficiently large p. On the event $\|K_p-g\|_\infty<\epsilon$, pin Q_p and g by subtracting their affine endpoint interpolants and apply Parseval over $1\le h\le m$ to their difference. The normalized Fourier coefficient of pinned Q_p is $c_{p,h}W_h$ with $c_{p,h}=e^{\pi ih/p}\sin(\pi h/p)/(\pi h/p)$, obtained by integrating the piecewise constant derivative. Let a_h be the Euclidean projection of the normalized coefficient of pinned g onto $[-2,2]$. Projection onto $[-2,2]$ reduces the distance from the real $W_h\in[-2,2]$ to the target coefficient. Since $|c_{p,h}-1|\ll h/p$, Parseval and the mesh error give

$$\sum_{h=1}^{m}(W_h-a_h)^2\ll m^2(\epsilon+p^{-1/2})^2+m^3/p^2\ll km$$

in the stated range. The polynomial $\prod_{h=1}^m(1-(x_h-a_h)^2/32)$ is then at least e^{-Ckm} on the event, whereas its mean under the independent law is at most $(31/32)^m$, by the first two Sato–Tate moments. The arithmetic discrepancy is $O(m\exp(C_0m)p^{-1/2})$ by (8). Take k small, then d large so that $m\le(k/d^2)\log p$. Both terms after multiplication by e^{Ckm} are $O(e^{-c_0m})$. Reducing ϵ_0 absorbs the constant, proving the upper bound.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, proof drafting, and internal checks. No datasets or numerical certificates are required for the proofs.

4 Kloosterman paths on translated parameter sets

Let a parameter of a prime Kloosterman path range over a translate of a deterministic set of size L. We prove that on all but an $O(L^{-c})$ fraction of translates the entire path law converges at the sharp uniform-metric Wasserstein order $(\log L)^{-1/2}$. The estimate also holds in L^s over uniform translates for each fixed finite Wasserstein order s. The set may have arbitrary shape

and grow subpolynomially in the prime. For every interval or progression of length at least $p^{1/2+\delta}$, the corresponding sharp order holds uniformly without removing translates. The results cover classical Kloosterman paths and a fixed-rank twisted hyper-Kloosterman family, with symplectic and special unitary trace laws. The proof uses additive twists of character sums, followed by a comparison of selected marginals of correlated Fourier-block observations.

Keywords Kloosterman paths, translated sets, Wasserstein distance, trace functions.
Mathematics Subject Classification 11L05, 11T23, 60F17.

4.1 Statements

For an odd prime p and $j \geq 1$, write $\mathrm{Kl}_j(c;p)$ for the normalized rank-j hyper-Kloosterman sum

$$p^{-(j-1)/2} \sum_{x_1 \cdots x_j = c} e_p(x_1 + \cdots + x_j), \qquad e_p(x) = \exp(2\pi i x/p),$$

with all variables nonzero. In particular $\mathrm{Kl}_1(c;p) = e_p(c)$. Fix the completed coefficient rank $R \geq 2$. For $a \in \mathbb{F}_p^\times$ and $b \in \mathbb{F}_p$, set

$$\varphi_{a,b}(x) = \mathrm{Kl}_{R-1}((ax)^{-1};p) e_p(bx) \quad (x \neq 0), \qquad \varphi_{a,b}(0) = 0.$$

Let $K_p(a,b)$ be the polygon with vertices $p^{-1/2} \sum_{x=1}^{j} \varphi_{a,b}(x)$ at times $j/(p-1)$, $0 \leq j \leq p-1$. At $R = 2$ these are ordinary Kloosterman paths with inverse coefficient a^{-1} and linear coefficient b. For $S \subseteq \mathbb{F}_p$ write $\mu_{a,S}$ for the path law with b uniform on S.

Put $G_R = \mathrm{USp}_R$ for even R and $G_R = \mathrm{SU}_R$ for odd R. Let X_h be independent traces of Haar elements of G_R. Write μ_R for the law of

$$K_R(t) = tX_0 + \sum_{h \neq 0} X_h \frac{e(ht) - 1}{2\pi i h}, \qquad e(t) = \exp(2\pi i t).$$

The series converges almost surely uniformly by the independent block estimate below. We take W_s with the supremum norm on paths.

Theorem 4.1. *Fix a finite $s \geq 1$ and $R \geq 2$. There are $c, C > 0$ depending at most on R, s such that for all odd primes and all deterministic $S \subseteq \mathbb{F}_p$ of sufficiently large size L, uniformly in $a \in \mathbb{F}_p^\times$,*

$$W_s(\mu_{a,S+u}, \mu_R) \geq \frac{c}{\sqrt{\log L}} \quad \text{for every } u, \qquad (\mathbb{E}_u W_s(\mu_{a,S+u}, \mu_R)^s)^{1/s} \leq \frac{C}{\sqrt{\log L}}.$$

Moreover $W_s(\mu_{a,S+u}, \mu_R) \leq C/\sqrt{\log L}$ outside a fraction $O_{R,s}(L^{-c})$ of the translates. The exceptional set may depend on a and S.

Theorem 4.2. *Fix $\delta > 0$. For every cyclic interval or arithmetic progression I of length $L \geq p^{1/2+\delta}$ with nonzero step, uniformly in $a \in \mathbb{F}_p^\times$ and I,*

$$W_s(\mu_{a,I}, \mu_R) \asymp_{R,s,\delta} (\log p)^{-1/2}.$$

Kowalski and Sawin [5] established the classical path limit.

Autissier, Bonolis and Lamzouri [13, Section 9.4.2] use the inverse-input twisted summands and identify the symplectic completed trace model in their two-parameter maximum problem. Its summand rank is one less than the completed rank, as is also the case in the convention of [12].

Short sums and sliding-sum estimates for trace functions address related distributions of scalar quantities [14, 15, 16]. Here the sampled object is the entire path at each parameter. We compare an empirical path law on a translated parameter set, rather than the distribution of a sum of traces over that set. Xi's bilinear Sato–Tate theorem [17] estimates scalar trace distributions over products of sets, under size conditions on both sets.

The exceptional set in Theorem 4.1 is necessary for arbitrary S. For example, choose S by retaining the parameters with real endpoint at least one. The complete trace law gives a positive fraction of such parameters for large p, since the Haar trace exceeds one in a neighborhood of the identity. On this translate ($u = 0$) the mean real endpoint is at least one, whereas the limit has mean zero. The endpoint test is one-Lipschitz in the uniform norm.

4.2 Additive twists and character norms

The completed coefficients of the normalized increments are

$$W_h = p^{-1/2} \sum_{x \neq 0} \varphi_{a,b}(x) e_p(-hx).$$

For $b \neq h$, introducing the last multiplicative coordinate $(b-h)x$ gives exactly $\mathrm{Kl}_R((b-h)/a; p)$. At $b = h$ the sum factors into $R-1$ complete nonzero additive sums, giving $(-1)^{R-1} p^{-(R-1)/2}$. This small real number belongs to the compact trace image: use conjugate eigenvalue pairs, and for odd R also an eigenvalue one. Thus compact group characters may be evaluated at a chosen class with this trace at the exceptional point. At ordinary points use the semisimple Frobenius class of the normalized Kloosterman sheaf. For even R multiply its trace class by the central element $-I$, which preserves the Haar model. For odd R the trace convention already has positive sign.

For distinct active shifts, let $F(b)$ be a product of nontrivial irreducible compact characters of those classes, with d_0 active shifts and product of dimensions D. We claim

$$\left| \sum_{b \in \mathbb{F}_p} F(b) e_p(kb) \right| \ll_R (d_0 + 1) D \sqrt{p} \qquad (k \in \mathbb{F}_p). \tag{12}$$

The normalized single-factor sheaf has arithmetic and geometric group Sp_R or SL_R, tame regular unipotent monodromy at zero and a single infinity break $1/R$, by Katz's results as recorded in [18, Prop. 4.6]. In the product over translates, inertia at $b = h$ acts nontrivially and unipotently in that factor alone. Projection onto each factor is surjective, and the normal closure of a noncentral unipotent in its connected simple group is the full factor. Hence the geometric group is the product. Applying the irreducible representations and tensoring gives rank D, no invariants, and infinity breaks at most $1/R$. For a nontrivial additive twist the break at infinity is one, since the original breaks are strictly smaller. In particular the twisted tensor has no invariants. All finite ramification is tame. The Euler characteristic on the affine line with the active punctures removed gives $\dim H^1_c \leq d_0 D$ and $H^0_c = H^2_c = 0$. Purity and the trace formula bound the sum there by $d_0 D \sqrt{p}$. The exceptional points contribute at most $d_0 D$. This proves (12). The all-trivial product has the exact full or restricted expectation one.

We use the norm $\|P\|_{\mathrm{ch}}$ of an expansion into products of characters, with absolute coefficients weighted by product dimensions. It is submultiplicative, and both the trace and its complex conjugate have norm R. For a polynomial in traces and dual traces of total degree r on d factors, with bound U on the product of compact groups,

$$\|P\|_{\mathrm{ch}} \leq U \exp(C_R r \log(eC_R(d+r)/r)) \quad (r \geq 1). \tag{13}$$

Indeed Haar character coefficients are bounded by U by orthonormality. Each possible highest weight in one factor has weight norm bounded by C_R times its trace degree, since it occurs in a tensor product of the standard representation and its dual. The total highest-weight norm is at most $C_R r$. Stars and bars count the products of possible highest weights, and the product of dimensions is at most $\exp(C_R r)$ by the Weyl dimension formula. For separate degrees r_h, the same argument bounds the norm by $U \prod_h [C_R(1+r_h)^{C_R}]$ on active factors, with factor one when $r_h = 0$. These estimates also apply to real and imaginary components expressed using the trace and its dual.

Let $m_S(k) = L^{-1} \sum_{b \in S} e_p(kb)$ and $\widehat{F}(k) = \sum_b F(b) e_p(-kb)$. For each nontrivial product F, Parseval over translates gives

$$\left\| \mathbb{E}_{b \in S+u} F(b) - \mathbb{E}_{b \in \mathbb{F}_p} F(b) \right\|_{L^2_u}^2 = \frac{1}{p^2} \sum_{k \neq 0} |\widehat{F}(k)|^2 |m_S(k)|^2.$$

Use (12) and $\sum_k |m_S(k)|^2 = p/L$, and add the error of the full average relative to the independent Haar characters. Thus for a test on d shifts,

$$\left\| \mathbb{E}_{b \in S+u} P(W) - \mathbb{E} P(X) \right\|_{L^2_u} \ll_R (d+1) L^{-1/2} \left\| P \right\|_{\mathrm{ch}}. \tag{14}$$

For an interval or a nonzero-step progression, Fourier inversion and $\sum_k |m_I(k)| \ll p \log p / L$ give the pointwise bound

$$\left| \mathbb{E}_{b \in I} P(W) - \mathbb{E} P(X) \right| \ll_R (d+1)(\sqrt{p} \log p / L) \left\| P \right\|_{\mathrm{ch}}. \tag{15}$$

Both extend to series convergent absolutely in this norm.

4.3 Completed tails and finite support

Exact summation of the discrete Fourier inversion formula on the mesh j/p gives the completed path polynomial at every mesh vertex:

$$tW_0 + \sum_{0 < |h| \leq (p-1)/2} b_{p,h} W_h (e(ht) - 1), \qquad b_{p,h} = \frac{1}{p(1 - e(-h/p))}.$$

The completion coefficients are $O(|h|^{-1})$, and $b_{p,h} = (2\pi i h)^{-1} + O(p^{-1})$. Replacing and interpolating a retained polynomial through H_* costs $O_R(H_*/p)$, and the original mesh change costs $O_R(p^{-1/2})$. These errors are deterministic.

For coefficients c_h bounded by C/H on $H/2 < |h| \leq H$, set $T(t) = \sum c_h W_h e(ht)$. Parseval on T^r bounds its supremum to power $2r$ by $(2rH+1) \int |T|^{2r}$. Only balanced tuples of indices survive under the integral, even for the complex coefficients, so their absolute multiplier mass is $O(C^{2r}/H)$ by fixing all but one index. Each trace monomial of degree $2r$ has character norm at most R^{2r}. Under independence the centered real projections have subgaussian variance proxy $O_R(H^{-1})$, so the moment root at a fixed time is $O_R(\sqrt{r/H})$. It follows that:

1. For the full parameter average, the supremum moment root is $O_R((rH)^{1/(2r)} \sqrt{r/H} + p^{-1/(4r)})$.
2. For intervals, replace the second term by $O_R((\sqrt{p} \log p / L)^{1/(2r)})$.
3. In a translated set, the L^2_u norm of the integrated moment discrepancy is $O_R(C_R^{2r}/(H\sqrt{L}))$ by (14).

These estimates include partial blocks. Pinning multiplies the sup estimates by at most two. The independent block estimates, with $r \asymp \log H$ and a maximum over the partial cutoffs, give almost sure uniform convergence.

Write $m = \lfloor a_1 \log L \rfloor$ for small fixed $a_1 > 0$ and $H_* \asymp m \log m$. For the translated case, put $N = \lfloor L^{1/8} \rfloor$ and $r = \lceil \log(16H_*) \rceil$. In the $O(\log L)$ blocks from H_* through N, apply Chebyshev to item 3 at threshold $C_R^{2r} L^{-1/4}/H$. Outside a fraction $O((\log L)L^{-1/2})$ of u, this holds for every such block. Summing the resulting L^s supremum estimates for these u gives $O_R(\sqrt{r/H_*} + \log L\, L^{-1/(8r)}) = O_R(m^{-1/2})$ for fixed s. Beyond N, use item 1 with fixed $r_0 \geq \max(2, \lceil s/2 \rceil)$. The full-parameter L^s tail bound is

$$O_{R,s}(N^{-1/2+1/(2r_0)} + \log p\, p^{-1/(4r_0)}) \ll_{R,s} L^{-c}.$$

Averaging its sth moment over translates gives the full-parameter moment exactly, by Fubini. Markov at $m^{-s/2}$ bounds this terminal tail by $m^{-1/2}$ outside a fraction $O_{R,s}(L^{-c'})$ of u. For the mean tail from H_*, use item 1 directly with $r = \lceil \log(16H_*) \rceil$, then average by the same Fubini identity. In the uniform interval case take $m = \lfloor a_{1,\delta} \log p \rfloor$ and sum item 2 above H_*; the second term is negligible uniformly in I.

For the lower bounds, Haar traces have mean zero and squared norm expectation one, by character orthogonality. Hence $\sup_{z \in \mathbb{C}} \mathbb{E} e^{-|X_h - z|^2} < 1$, since the law is nondegenerate and bounded. A uniform ball of radius ϵ around any path, after endpoint pinning, forces the sum of squared deviations in the normalized Fourier coefficients over $H \leq h < 2H$ to be at most $CH^2\epsilon^2$, by Parseval. With H a sufficiently small constant times ϵ^{-2}, exponential Markov and independence bound the ball mass by $e^{-c_R \epsilon^{-2}}$ uniformly over its center. The arithmetic law has at most L atoms. Taking the union of balls of radius $c'_R/\sqrt{\log L}$ around them covers at most half the target law. This proves the stated lower bounds.

4.4 Observation comparison and conclusions

Appendix .1 constructs a path observation at reconstruction error $O(m^{-1/2})$ in L^s on the retained frequencies. Its dense and selected sparse marginals compare by applying (14) to the Fourier test and each mixed derivative, before proceeding from Fourier comparison to total variation. The rank norm estimates have exponential cost $\exp(O_R(m))$, including all subset counts and truncation amplification factors. In the translated case this gives a nonnegative error function of u of expectation at most $L^{-1/2} \exp(C_R m)$. Apart from this error term the observation total variation bound is deterministically smaller than any fixed power of m^{-1}, as follows from the independent sparse factorial moments. Taking a_1 small and then Markov makes the arithmetic comparison error polynomially small in L outside a further exceptional fraction $O_{R,s}(L^{-c})$. For the remaining translates the observation couplings and the tails from Section 3 give $W_s = O_{R,s}(m^{-1/2})$. For the mean assertion, average the sth coupling costs, using the mean total variation estimate and the Fubini identity for the arithmetic tails. No maximum path bound on the exceptional set is needed.

For intervals use (15) pointwise instead of (14). The selected-marginal error is at most $(\sqrt{p} \log p/L) \exp(C_R m)$ after the same comparison. When $L \geq p^{1/2+\delta}$ and m is a sufficiently small constant times $\log p$, this is polynomially small in p. Combining the interval tail estimate, reconstruction bound and lower bound proves Theorem 4.2.

.1 Correlated observation comparison

Let m be a large integer, chosen in Section 3. Put $H_0 = \lfloor m^{1/4} \rfloor$. Independently round the real and imaginary parts of coefficients $|h| \leq H_0$ by binomial rounding on $[-R, R]$ with degree $\lceil m^2 \rceil$. For a component x, use $2R\,\mathrm{Bin}(\lceil m^2 \rceil, (x+R)/(2R))/\lceil m^2 \rceil - R$. The L^s path error is $O_{R,s}(\log m/m)$ by the

individual binomial moments and the absolute multiplier sum. There are at most $\exp(O(H_0 \log m))$ low output vectors, and the conditional probability polynomial of each has character norm at most the same exponential order by the separate-degree estimate.

Take blocks $H = 2^j H_0$, $j \geq 1$, through the first $H \geq m \log m$. In such a block let $y = A_H x$ be the $I_H = 16H$ real and imaginary parts of $\sqrt{H} \sum b_h x_h e(ht)$ on the $8H$ equally spaced mesh samples. Discrete Parseval gives $\|A_H\|_{2\to 2} \ll 1$ and $\|y\|_2^2 \ll H$ on the cube. In the independent model, the complex coefficients are independent across indices. Applying Hoeffding to their bounded real projections gives the directional subgaussian bound $\mathbb{E} \exp(t \cdot y) \leq \exp(C \|t\|_2^2)$. For every subset S of samples, Gaussian integration then gives

$$\mathbb{E} \exp\left(\lambda \sum_{i \in S} y_i^2\right) \leq D^{|S|} \tag{16}$$

for constants $\lambda > 0, D$ depending only on R.

For reconstruction, choose a smooth Fourier multiplier equal to one on $[-H, H]$, zero outside $[-2H, 2H]$, with kernel V_H. Discrete orthogonality on the $8H$ grid recovers the block as the sum of sampled values against weights $(8H)^{-1} V_H(t - j/(8H))$. The kernel bound $|V_H(u)| \ll H/(1+H^2 \operatorname{dist}(u, \mathbb{Z})^2)$, obtained by two summations by parts, makes the absolute sum of the weights uniformly bounded. Thus if normalized samples are changed by at most τ in each component, the sup error after reconstruction and pinning is $O(\tau/\sqrt{H})$.

Use scale $\tau_H = (H/m)^{3/4}$ for $H \leq m$, and $\tau_H = A(H/m)^{1/4}$ for $H > m$, with large absolute A. The former blocks will be compared on all slots, and the latter on sparse subsets. Both parts of the sum $\sum_H \tau_H/\sqrt{H}$ are geometric and give $O(m^{-1/2})$.

Lemma .3. *Fix a positive constant J_0. For I real coordinates and noise scale $\tau > 0$, put $T = C_0 \max(1, \tau^4)$ with $C_0 \geq 1$. There is a probability density on $\mathbb{R}^I$, with noise vector η, with these properties:*

1. $\mathbb{P}(\max_i |\eta_i| > \tau/8) = O(I^{-J_0})$.
2. *For each slot set S, the probability that all $|\eta_i| > \tau/8$ in S is at most $2(Ce^{-c\sqrt{T}})^{2|S|}$.*
3. *The characteristic function of any D-coordinate marginal, $D \geq 1$, has support in a ball of radius $R(D) = CTI^{1/4}D^{1/4}/\tau$. Every mixed first derivative, once in each of r distinct coordinates, is bounded by $(C\tau)^r$ up to an absolute constant. These derivatives vanish outside the same ball.*

Proof. Take a real even smooth function f supported in $[-1/8, 1/8]$, with $\|f\|_2 = 1$ and $|\widehat{f}(u)| + |\widehat{f}'(u)| \leq Ce^{-c\sqrt{|u|}}$. For example, use the normalized density of the sum of independent uniforms on $[-c_*/j^2, c_*/j^2]$, $j \geq 1$, with c_* small. Its sinc product has at least $c\sqrt{|u|}$ factors bounded by $1/2$ for large $|u|$, and a differentiated term loses at most one such factor and contributes $O(j^{-2})$. This also proves smoothness and the requisite derivative decay.

Take a smooth even cutoff χ decreasing in absolute argument, equal to one on $[-1, 1]$ and zero outside $[-2, 2]$. Form a partition with $\phi_0(u) = \chi(u/T)$ and $\phi_j(u) = \chi(u/(2^j T)) - \chi(u/(2^{j-1}T))$ for $j \geq 1$. Its terms are nonnegative, sum to one, and $\sum_j |\phi_j'| \ll 1$. The upper radii are $U_j = 2^{j+1}T$, and if $\phi_j(u) > 0$ for $j \geq 1$, then $|u| > U_j/4$.

Retain band configurations $(j_1, \ldots, j_I)$ with $\sum_i U_{j_i}^4 \leq C_1 I T^4$. Let w be the sum of their product partition functions, so $0 \leq w \leq 1$. We use the squared-amplitude device for positive Fourier smoothing as in [19, Section 7]. The restriction on the joint amplitude will control the Fourier support of every selected marginal. Define $\widehat{P}(u) = w(u) \prod_i \widehat{f}(u_i)$. Take the density $|P|^2/\|P\|_2^2$ and

dilate by τ in each coordinate. As usual, the Fourier normalization can be taken unitary, with absolute constants adjusted for the characteristic-function convention.

For a rejected configuration the bounds $U_{j_i} \le 4\max(T, |u_i|)$ imply $\sum_i |u_i|^4 > C_2 I T^4$, with C_2 as large as needed by choosing C_1. Under the iid $|\widehat{f}|^2$ probability law, $|u_i|^4$ has all moments. For each fixed integer k, the $2k$th moment of the sum of centered copies is $O_k(I^k)$ by independence and expansion, since singleton factors vanish. Thus the rejected frequency mass is $O(I^{-2J_0})$ on taking k large enough. Choosing C_2 above four times the scalar fourth moment also bounds this mass by $1/4$ for every I by Markov's inequality. It follows that the amplitude P is $O(I^{-J_0})$ from the product amplitude in L^2, and hence the normalized densities are within the same order in total variation by Cauchy–Schwarz. This proves the first property; the uniform mass bound gives $\|P\|_2^2 \ge 3/4$.

For the joint bound, write f_j for the inverse transform of $\phi_j \widehat{f}$. Fix the band labels on S. The corresponding partial sum of the amplitude factors as the selected f_j times an amplitude in the complementary coordinates with L^2 norm at most one, because its partition multiplier lies between zero and one. On $|x| > 1/8$ we have $f_0 = -\sum_{j \ge 1} f_j$ in L^2. Moreover $\sum_{j \ge 1} \|f_j\|_2 \ll e^{-c\sqrt{T}}$ by the lower band radii and the decay of $\widehat{f}$. Summing the restricted L^2 bounds over selected band labels gives $(Ce^{-c\sqrt{T}})^{|S|}$. Squaring and normalizing proves the second property.

The support of $\widehat{P}$ has fourth power sum at most $C_1 I T^4$. Its projection to D coordinates has Euclidean radius at most $CTI^{1/4}D^{1/4}$ by Cauchy–Schwarz. The Fourier transform of $|P|^2$ is an autocorrelation of $\widehat{P}$, so the support of the selected characteristic function is in the corresponding difference set, after accounting for dilation. Finally a mixed derivative of $\widehat{P}$ on distinct slots is pointwise bounded by the product of $|\widehat{f}'| + C|\widehat{f}|$ there and $|\widehat{f}|$ elsewhere, by the partition derivative estimate. Its L^2 norm is $O(C^r)$. By Plancherel and Cauchy–Schwarz this bounds the product absolute moment of the corresponding η_i by $O((C\tau)^r)$. These moments justify differentiating the characteristic function, with the stated support. □

Draw the noises independently across blocks and independently of every input. In the dense blocks observe the noisy samples $y_i + \eta_i$. In each sparse block replace a noisy sample by zero whenever its absolute value is at most $\tau_H/2$. Clip the final output in either case to a centered interval of radius $C\sqrt{H} + \tau_H$, containing the deterministic range of y_i and its small perturbation. On the good event from Lemma .3(1) the sample error is $O(\tau_H)$.

.2 Comparison on selected marginals

Select every dense slot. In each sparse block H select D_H slots, with D_H bounded below by zero and above by the cutoff specified in the next section. We first compare the noisy real vectors before thresholding or clipping, together with the rounded low output. Deterministic maps can only decrease total variation. Let g_u be the difference density of the selected noisy coordinates for one fixed low label on translate u. In the interval case there is a single such density. Its Fourier transform factors as the product of the noise marginal transforms times the difference of the two input characteristic expectations, weighted by the conditional low label polynomial. The smoothing ensures a density, and the selected characteristic function is compactly supported.

For a Fourier test vector t_H, the character norm of $\exp(it_H \cdot A_H x)$ is at most $\exp(C\sqrt{H}\,\|t_H\|_2)$ by its absolutely convergent power series and $\left\|A_H^T t_H\right\|_1 \ll \sqrt{H}\,\|t_H\|_2$. A mixed derivative in r_H distinct variables in block H inserts a product of r_H sample coordinates. Its cube bound is at most $(CH/r_H)^{r_H/2}$ by the energy estimate and the arithmetic–geometric mean inequality. Its character norm is bounded by $\exp(Cr_H \log(eCH/r_H))$ by (13), since the block uses $O(H)$ Fourier inputs. The bound is one for $r_H = 0$. Taking C sufficiently large, one may bound this exponent using D_H

in place of r_H for all $r_H \le D_H \le I_H$. The low label costs $\exp(o(m))$. With the derivative bounds for the noise in Lemma .3, these estimates and the product rule control all mixed first derivatives of $\widehat{g}_u$ in L^2_u by (14). For the interval case apply (15) pointwise.

To bound expected L^1 norm, assign $s_i = 1 + \tau_H$ in block H. Multiply and divide by the square root of $\prod_i (1 + (x_i/s_i)^2)$, then apply Cauchy–Schwarz with the average over u and Plancherel in the observation coordinates. For $D = \sum_H D_H$ this gives

$$\mathbb{E}_u \|g_u\|_1 \le \Big(\prod_i \pi s_i\Big)^{1/2} \left[C^D \sum_{U \subseteq \{1,\dots,D\}} \Big(\prod_{i \in U} s_i^{-2}\Big) \int \mathbb{E}_u |\partial_U \widehat{g}_u(t)|^2 \, dt \right]^{1/2}.$$

The weights absorb the powers of τ_H from the noise derivatives. The support is a product of balls. A ball of radius R in D_H dimensions has volume at most $(CR/\sqrt{D_H})^{D_H}$. For a single interval the same inequality omits the average over u. After summing over low labels, the total variation upper bound (averaged over translates, or pointwise for an interval) is

$$\begin{aligned} &\varepsilon \exp\big(o(m) + C \textstyle\sum_H E_H(D_H)\big), \qquad \text{where} \\ &E_H(D_H) = \sqrt{H} R_H(D_H) + D_H \log(eCH/D_H) \\ &\qquad\qquad + D_H \log\left(\frac{C(1+\tau_H) T_H}{\tau_H} (I_H/D_H)^{1/4} \right), \end{aligned} \tag{17}$$

Here $\varepsilon = L^{-1/2}$ in (14) and $\varepsilon = \sqrt{p} \log p / L$ in (15). Also $T_H = C_0 \max(1, \tau_H^4)$ and R_H is as in Lemma .3. Terms with $D_H = 0$ are omitted. The exponent includes the 2^D derivative sum and the polynomial factor for the number of Fourier inputs.

For dense blocks $H \le m$ take $D_H = I_H$. The contribution is $O(H/\tau_H + H[1 + \log(1/\tau_H)])$, summable to $O(m)$ because $H/\tau_H = m^{3/4} H^{1/4}$. In a sparse block $H > m$ define

$$K_H = \min\big(I_H, \lceil \max(m^{1/2}, C_2 I_H e^{-c_2 \tau_H^2}) \rceil\big)$$

with c_2 sufficiently small and C_2 sufficiently large. We allow $D_H \le \min(I_H, K_H + 1)$. Since $H \le 2m \log m$ up to an absolute factor and $\tau_H^2 \ll \sqrt{\log m}$, the $m^{1/2}$ term is eventually smaller than the exponential term throughout this range. Thus $x = D_H/I_H \le C e^{-c_2 \tau_H^2}$. The first contribution of this block to (17) is $O(H(T_H/\tau_H) x^{1/4})$. The others are bounded by $O(Hx[1 + \log(1/x) + \log T_H + \log(1 + 1/\tau_H)])$. Together they are $O(H(1+\tau_H)^C e^{-c_3 \tau_H^2})$. Since $\tau_H^2 = A^2 \sqrt{H/m}$, their dyadic sum is $O(m)$. In particular, every selected marginal compares in mean total variation over translates (pointwise for intervals) at cost $\varepsilon \exp(Cm)$.

.3 Sparse comparison and transport

Under the independent model, a specified set of D_H sparse slots is simultaneously nonzero after thresholding only if in each of those slots either $|y_i| \ge \tau_H/4$ or $|\eta_i| > \tau_H/8$. Split the slots between these alternatives. For the first use (16) and Markov's inequality, and for the second use Lemma .3(2). Independence of noise and input gives the product bound, summed over at most 2^{D_H} allocations. With A large the joint activity probability is at most $\exp(-c_4 \tau_H^2 D_H)$. These bounds multiply across blocks, since the independent Fourier coordinates and the noises are independent across blocks.

The argument uses truncated inclusion–exclusion with a factorial error bound, as in read-once Boolean tests [9, Section 3], here for arbitrary bounded tests of multiple observed values. For a realized sparse block output with N_H nonzero entries, expand its point mass as a product of

zero point masses plus coordinate differences. Retain terms involving at most K_H differences. If $N_H \le K_H$ this is exact. Otherwise the signed variation error is at most

$$B_H = 2 \cdot 3^{K_H+1} \binom{N_H}{K_H+1},$$

by the binomial bounds $\binom{N_H}{j} \le \binom{N_H}{K_H+1}\binom{K_H+1}{j}$ for $j \le K_H$ and the factor 2^j for signed differences. Take $B_H = 0$ if the cap keeps all coordinates. For a product across sparse blocks, telescoping bounds the aggregate variation error by $\prod_H (1 + B_H) - 1$.

Expand this last product to estimate its expectation. Each factorial count is a sum over sets of prescribed active slots. Under the independent law the term for a chosen block is bounded by

$$\left(\frac{C I_H e^{-c_4 \tau_H^2}}{K_H + 1} \right)^{K_H+1}.$$

Take $c_2 < c_4$ and C_2 large in the definition of K_H. This is at most $\exp(-cK_H)$, so the whole expected error is smaller than any fixed power of m^{-1}. Under arithmetic input compare each selected joint activation by the selected-marginal bound (17). The sum of all weights and subset counts is $\exp(O(m))$, because $\sum_H K_H \log(eI_H/K_H) = O(m)$ by the same sparse budget estimate. For each translate, retain the sum of the pertinent selected marginal total variation errors, with these weights. It is a nonnegative function of u with expectation at most $\varepsilon \exp(Cm)$ and controls the arithmetic part of the truncation error on that translate. For a single interval the same bound is pointwise.

For any test of the full observations bounded by one, the truncated expectation is a sum of tests on all the dense outputs, the low rounded vector, and at most K_H selected sparse coordinates in each block. Evaluate the other sparse outputs at zero in each term of the signed expansion. The same subset count and signed amplification is $\exp(O(m))$. Bounding each such test by its selected marginal total variation, and adding the bounds for the arithmetic truncation errors, gives a single nonnegative error function $J(u)$ independent of the test, with $\mathbb{E}_u J(u) \le \varepsilon \exp(Cm)$. In the interval case J has this bound pointwise. Thus the full genuine observation laws differ in total variation by at most

$$O_{J_0}(m^{-J_0}) + J(u).$$

A maximal coupling of the observation laws costs the total variation bound to power $1/s$ times $O(\log m)$ in L^s, since clipping and stable interpolation bound the decoded paths. The reconstruction error on the compact noise event is $O(m^{-1/2})$. Its complement has probability $O(\sum_H I_H^{-J_0}) = O(m^{-J_0/4})$, with deterministic $O(\log m)$ costs on the retained paths and outputs. Taking J_0 sufficiently large for the fixed s makes this contribution and the independent truncation term negligible. The remaining term $J(u)$ has the stated mean bound or the pointwise interval bound. This is the observation comparison used in Section 4.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, proof drafting, and internal checks. No datasets or numerical certificates are required for the proofs.

A Sparse Kloosterman path transport

We improve the upper bound for uniform-path W_1 convergence of prime Kloosterman paths from $O(\sqrt{\log\log p/\log p})$ to $O(\sqrt{\log\log\log p/\log p})$. The estimate holds when the coefficient of the

inverse term is any fixed nonzero residue. A binomial observation of the low Fourier coefficients leaves centered conditional residuals under the Sato–Tate model. We reconstruct these residuals and the intermediate Fourier blocks from sparse trigonometric samples. A truncated inclusion–exclusion expansion compares the full observation laws using polynomial tests of controlled character norm. The resulting upper bound differs from the finite-support lower order by a factor of order $\sqrt{\log\log\log p}$.

Keywords Kloosterman paths, Wasserstein distance, polynomial approximation.
Mathematics Subject Classification 11L05, 11T23, 60F17.

A.1 Statement and inputs

Let p be an odd prime, A be uniform in $\mathbb{F}_p^\times$, and $B \in \mathbb{F}_p^\times$ be fixed. Let $K_p(A,B)$ be the polygon with vertices

$$p^{-1/2}\sum_{x=1}^{j}\exp(2\pi i(Ax+B/x)/p), \qquad 0 \le j \le p-1,$$

at times $j/(p-1)$. Write $\mu_{p,B}$ for its law, and let μ be the law of

$$K(t) = tX_0 + \sum_{h\neq 0} X_h \frac{e(ht)-1}{2\pi i h}, \qquad e(u)=\exp(2\pi i u),$$

with independent Sato–Tate variables X_h of density $\sqrt{4-x^2}/(2\pi)$ on $[-2,2]$. We use W_1 for the Wasserstein distance with the supremum norm on paths.

Kowalski and Sawin [5] proved the finite-dimensional law for $B=1$ and functional convergence for the two-parameter family. Kowalski and Untrau [6, Section 1.3(4)] asked for quantitative Wasserstein convergence on path space. The two-parameter upper bound in [7] comes from a Fourier truncation with graded coefficient rounding, and the small-ball upper bound there yields $W_1 \gg (\log p)^{-1/2}$ for every law supported on at most $(p-1)^2$ paths. To move closer to that scale we sample the Fourier residual in time and retain the samples that exceed a scale-dependent threshold.

Theorem A.1. *There are absolute constants C, p_0 such that for every prime $p \ge p_0$ and every $B \in \mathbb{F}_p^\times$,*

$$W_1(\mu_{p,B},\mu) \le C\sqrt{\frac{\log\log\log p}{\log p}}.$$

The same bound holds for the law with both A and B independent uniform nonzero residues.

Put

$$W_h = p^{-1/2}\sum_{x\neq 0}\exp(2\pi i((A-h)x+B/x)/p), \qquad |h| \le (p-1)/2.$$

The variables are real and bounded by 2. For polynomials in the coordinates let $\|P\|_{\mathrm{ch}}$ be the sum of absolute coefficients in the product basis of the characters

$$\chi_k(2\cos\theta) = \sin((k+1)\theta)/\sin\theta,$$

weighted by the product of the dimensions $k+1$. It is submultiplicative by the Clebsch–Gordan identity, and $\|x_h\|_{\mathrm{ch}} = 2$. The uniform character estimate of [7, Lemma 6.3] gives, for variables at distinct shifts,

$$|\mathbb{E}P(W) - \mathbb{E}P(X)| \le 4(d+1)p^{-1/2}\,\|P\|_{\mathrm{ch}} \tag{18}$$

if there are at most d variables. We will also use the bound $\|P\|_{\mathrm{ch}} \le \sup|P| \prod_h (n_h+2)^2$ on the cube $[-2,2]^d$ when P has separate degrees n_h. Indeed after $x_h = 2\cos\theta_h$ the complex Fourier coefficients are bounded by $\sup|P|$. Group the sign orbits and express each cosine via $\cos(j\theta) = (\chi_j - \chi_{j-2})/2$ for $j \ge 2$, and $\cos\theta = \chi_1/2$. The weighted norm is at most $j+1$ for each cosine. Summing over modes gives the bound.

Lemma A.2. *If $2 \le H \le (p-1)/2$, the sharp low path formed from $(W_h)_{|h|\le H}$ with multipliers $(2\pi i h)^{-1}$ differs from K_p in expected uniform norm by*

$$O\left(\sqrt{\frac{\log(16H)}{H}} + \log p\, p^{-1/(4\lceil \log(16H)\rceil)} + H/p + p^{-1/2}\right).$$

The corresponding Sato–Tate tail is bounded by the first term.

Proof. On the mesh j/p, add a constant last segment and complete the summand by discrete Fourier inversion. The resulting path equals at mesh vertices

$$tW_0 + \sum_{0<|h|\le (p-1)/2} \frac{W_h(e(ht)-1)}{p(1-e(-h/p))}.$$

The completion coefficients are $O(1/|h|)$ and differ from $(2\pi i h)^{-1}$ by $O(p^{-1})$. Only the retained polynomial is interpolated in estimating the error, at cost $O(H/p)$ from its derivative. The mesh change from $j/(p-1)$ costs at most $p^{-1/2}$.

For a block $T(t) = \sum_{D/2<|h|\le D} b_h W_h e(ht)$ with $|b_h| \ll 1/D$, Parseval on T^r gives $\|T\|_\infty^{2r} \le (2rD+1)\int |T|^{2r}$. In the independent model, the linear form at fixed t is subgaussian with variance proxy $O(1/D)$. In comparing the integrated moments, only tuples with the sum of the first r shifts equal to that of the last r survive. The sum of absolute products of the multipliers of these tuples is bounded by $\max|b_h|(\sum|b_h|)^{2r-1} \ll C^{2r}/D$. Each monomial comparison costs at most $4(2r+1)2^{2r}p^{-1/2}$ by (18). The factor D^{-1} cancels the factor D from Parseval. Taking roots, the block error is bounded by $C(rD)^{1/(2r)}\sqrt{r/D} + Cp^{-1/(4r)}$. Pinning costs a factor at most two. Take $r = \lceil \log(16H)\rceil$ and sum the dyadic blocks. The independent bound uses no comparison error, and taking a maximum over $O(D)$ partial-block cutoffs adds at most $O(D)^{1/(2r)}$ to the moment bound as a multiplicative factor. This is bounded for $r = \lceil \log(16D)\rceil$. The expected dyadic maxima are summable, which also gives almost sure uniform convergence of the independent series. □

A.2 Observations and stable reconstruction

For an error of order $m^{-1/2}$, the proof uses a low observation window of size $m \log\log m$ and retains frequencies through a cutoff of order $m \log m$. We compare the observations by polynomial tests that use at most $m/(\log m)^{10}$ sampled coordinates in each term. The successive estimates control reconstruction error, the discarded terms of the expansion, and arithmetic comparison.

Let m be large, $L = \log m$ and $\ell = \log L$. Take $H_0 = \lceil A_0 m\ell \rceil$ with a sufficiently large absolute A_0. Let $H_j = 2^j H_0$ up to the first index J with $H_J \ge mL$, so $H_J \le 2mL$ for sufficiently large m. We compare the sharp low paths through H_J. Lemma A.2 bounds the discarded part by $O(m^{-1/2})$ for the choice of m made below.

At frequencies $|h| \le H_0$, take $n_0 = m$ and $n_h = \lceil H_0^2/h^2 \rceil$ for $h \ne 0$. Given an input x, independently observe

$$k_h \sim \mathrm{Bin}(n_h, (x_h+2)/4), \qquad z_h = \frac{4k_h - 2n_h}{n_h+3}.$$

Write $Y = (k_h)_{|h|\le H_0}$. Any bounded function of Y averaged given x has character norm at most D times its bound, where

$$D = \prod_{|h|\le H_0} (n_h + 2)^2 \le \exp(CH_0). \tag{19}$$

The factorial estimate $\log(H_0!) \ge H_0 \log H_0 - H_0$ gives this bound. In the independent model, $(X_h + 2)/4$ has the beta law with parameters $(3/2, 3/2)$. Given Y, its coordinates remain independent, with beta parameters $k_h + 3/2, n_h - k_h + 3/2$ after observing each count. Thus $X_h - z_h$ is centered and subgaussian with variance proxy $C/(n_h + 1)$. One proof of this estimate uses the lower bound $(n_h + 1)/16$ on the second derivative of the negative log-density on $[-2, 2]$. For any exponential tilt with negative log-density V, integration by parts gives $\mathrm{Cov}(X, V'(X)) = 1$. Symmetrizing with an independent copy bounds the tilted variance by $16/(n_h + 1)$. Integrating the second derivative of the log moment-generating function gives the stated subgaussian estimate.

We retain the reference path with coefficients z_h at $|h| \le H_0$. At the endpoint, $\mathbb{E}|X_0 - z_0|^2 \ll 1/m$. The same bound with an additive $O(Dp^{-1/2})$ holds for W_0 by expanding the square and applying (18) and (19). We encode the other low residuals $x_h - z_h$ by trigonometric sampling, and encode each subsequent band in the same way. Set $b_h = (2\pi i h)^{-1}$. In the first block $H = H_0$ take $T_H(t) = \sum_{0<|h|\le H_0} b_h(x_h - z_h)e(ht)$, and in each later block H sum $b_h x_h e(ht)$ over $H/2 < |h| \le H$.

Use $8H$ equally spaced samples in each block. For a fixed smooth cutoff ϕ equal to one on $[-1, 1]$ and supported on $[-2, 2]$, let V_H have Fourier symbol $\phi(h/H)$. Then

$$T_H(t) = \frac{1}{8H} \sum_{k=0}^{8H-1} T_H(k/(8H)) V_H(t - k/(8H)).$$

Discrete orthogonality excludes aliasing, and $|V_H(t)| \ll H/(1 + H^2 \operatorname{dist}(t, \mathbb{Z})^2)$ by summation by parts. Hence the sum of the absolute interpolation weights is bounded independently of H, t. Uniform error on the samples gives uniform reconstruction error up to an absolute constant, and pinning by subtracting the value at zero preserves this.

Write y_i for the real or imaginary part of a normalized sample $\sqrt{H}\, T_H(k/(8H))$. There are $I = O(mL)$ such entries. In the independent model, conditionally on Y, their joint linear forms are subgaussian:

$$\mathbb{E}\left[\exp\left(\sum_i u_i y_i\right) \; middle |\, Y\right] \le \exp(C\,\|u\|_2^2). \tag{20}$$

To see this, use discrete Parseval within each block. The normalized Fourier-to-sample map has bounded Euclidean operator norm when the inputs are scaled by their subgaussian widths. For the first block this follows from $\sqrt{H_0}|b_h|/\sqrt{n_h + 1} \ll H_0^{-1/2}$, and the other blocks use $\sqrt{H}|b_h| \ll H^{-1/2}$. The blocks use independent coefficients conditionally on Y. Taking real and imaginary parts does not affect the bounded operator norm. Gaussian integration in (20) shows that for some absolute $\lambda > 0, D_1 < \infty$,

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in S} y_i^2\right) \,\middle|\, Y\right] \le D_1^{|S|} \tag{21}$$

for every subset of entries.

A.3 Polynomial sample channels

For a sample in block H let

$$\tau = \sqrt{H/m}(H/H_0)^{-1/4}, \qquad R = C_R L.$$

The squared thresholds are at least $A_0\ell$ and at most $C\sqrt{L\ell}$. Their contributions to reconstruction satisfy $\sum_H \tau_H/\sqrt{H} = O(m^{-1/2})$.

For a scalar $y \in [-R, R]$, write $y = R\cos\theta$. Draw an angle U with even periodic density proportional to

$$\left(\frac{\sin(kU/2)}{k\sin(U/2)}\right)^{2s}, \qquad k = \lceil C_k R/\tau \rceil, quads = \lceil C_s L \rceil.$$

Approximate $R\cos(\theta+U)$ by a multiple of $\tau/4$, using the special output zero when the value lies in $[-\tau/2, \tau/2]$. The normalizing integral of the kernel is $\gg 1/(k\sqrt{s})$, by restricting to $|U| \le c/(k\sqrt{s})$. For $\tau/(4R) \le |U| \le \pi$ the sine ratio is bounded by $1/2$ when C_k is sufficiently large. The probability of this set is at most m^{-100} for sufficiently large C_s. Off this set the observation error is at most τ, and the output can be nonzero only if $|y| \ge \tau/4$.

The probability vector of the outputs has entries $Q_{i,z}(y)$ that are polynomials of degree at most $d_i = s(k-1)$. Convolution by the angular kernel yields an even trigonometric polynomial for each output, since the output partition depends only on cosine. The entries sum to one as polynomials. For arbitrary y the genuine channel $P_i(y)$ first clips y to $[-R, R]$. The signed polynomial vector $Q_i(y)$ is used only in polynomial tests. For its difference from the genuine channel, and for the genuine probability of a nonzero output, use respectively the envelopes

$$b_i(y) = (C_0 y/R)^{2d_i}, \qquad \pi_i(y) = m^{-100} + (4y/\tau_i)^{2\lfloor c_0 \tau_i^2 \rfloor}.$$

The difference is zero on $[-R, R]$. Outside, test it against an arbitrary function of the output bounded by one. The polynomial expectation is bounded by one on $[-R, R]$ and hence by $(2|y|/R)^{d_i}$ outside, by the extremal property of Chebyshev polynomials. The clipped expectation is bounded by one, so the total variation norm is at most $b_i(y)$ with C_0 sufficiently large. The bound by π_i follows from the angular tail inside $[-\tau_i/4, \tau_i/4]$, and from $\pi_i \ge 1$ outside. The indicator of $|y| > R$ is also bounded by $b_i(y)$.

For arbitrary subsets S of distinct entries, (21) yields

$$\mathbb{E}_X \prod_{i\in S} b_i(y_i) \le m^{-100|S|}, \qquad \mathbb{E}_X \prod_{i\in S} (b_i(y_i) + \pi_i(y_i)) \le L^{-25|S|}. \tag{22}$$

In fact $x^{2v} \le (v/(\lambda e))^v \exp(\lambda x^2)$. In the b_i factors, $d_i/R^2 \ll 1/\tau_i$ tends to zero uniformly and $d_i \gg L^{7/4}\ell^{-1/4}$, so these factors, including D_1 from (21), cost at most m^{-100} each. For the power term in π_i, taking c_0 small bounds its factor by $D_1 e^{-c_1\tau_i^2}$. With A_0 large this is at most a sufficiently small constant times L^{-25}. Expansion of the sums proves (22), even conditional on Y.

We explain the arithmetic cost of the same tests. Every scalar polynomial $Q_{i,z}$ is bounded by one on $[-R, R]$, so expanding in Chebyshev polynomials and then powers of y/R bounds the absolute coefficient sum by C^{d_i}. The envelope polynomials have the same coarse bound $\exp(O(L^2))$ in normalized powers, since $d_i = O(L^2)$ and $\tau_i \to \infty$. Each y_i is a linear form $U_i(x) - V_i(Y)$. The first term has character norm $O(\sqrt{H}\log(2H))$ and the second has supremum at most the same quantity, by summing $|b_h|$ and using $|z_h| \le 2$. For blocks above H_0 the second term is zero. Expand powers of the difference, multiply, and average coefficient functions of Y using (19). This gives the bound

$$D\exp(CjL^3) \tag{23}$$

for the character norm of the conditional average of any product of j output or envelope polynomials, multiplied by any function of Y bounded by one. The factor L^3 allows $O(L^2)$ powers per entry and a factor $\exp(O(L))$ per power. The bound is uniform over output labels and includes the replacement $Q_i - \delta_0$ entrywise.

A.4 Truncated comparison of the observation laws

Set $K_* = \lfloor m/L^{10} \rfloor$. This is the truncation order of our test polynomial. The number of nonzero observations is random, and we control terms beyond this order by factorial moments. Inclusion–exclusion bounds of this type also appear in the study of read-once Boolean formulas [9, Section 3]. Here the observations have finite alphabets and the underlying sample values can be dependent. We compare the expectation of any function F of the full observations bounded by one. Fix x, Y first. The genuine channels P_i are conditionally independent. Expand their product around the all-zero outputs and keep at most K_* signed differences $P_i - \delta_0$. For deterministic observations with N nonzero entries, this is an inclusion–exclusion expansion of F around zero, exact if $N \le K_*$. Otherwise its error is at most $2 \cdot 3^{K_*+1} \binom{N}{K_*+1}$, since each j-entry difference has norm at most 2^j and $\binom{N}{j} \le \binom{N}{K_*+1}\binom{K_*+1}{j}$ for $j \le K_*$. Averaging the factorial count under the genuine product channel bounds the error by

$$2 \cdot 3^{K_*+1} \sum_{|S|=K_*+1} \prod_{i \in S} \pi_i(y_i).$$

Now replace the genuine channels in the truncated expansion by Q_i. Expand first in the differences $Q_i - P_i$. For a nonempty subset B' of these differences, of size $b \le K_*$, keep its outputs fixed and expand the remaining genuine channels to order $K_* - b$. Comparing this inner truncation to a full genuine expectation gives the bound

$$\prod_{i \in B'} b_i(y_i) \left(1 + 2 \cdot 3^{K_*-b+1} \sum_{\substack{S \cap B' = \varnothing \\ |S| = K_*-b+1}} \prod_{i \in S} \pi_i(y_i) \right).$$

This follows from the previous factorial estimate applied to the remaining entries and then from the variation bound on the signed measures. Grouping the second terms by $B' \cup S$, the total error between the true expectation and the truncated polynomial expectation is bounded by

$$\sum_{1 \le |S| \le K_*} \prod_{i \in S} b_i(y_i) + C_1^{K_*+1} \sum_{|S|=K_*+1} \prod_{i \in S} (b_i(y_i) + \pi_i(y_i)). \tag{24}$$

This comparison controls the polynomial continuation only through products over small subsets.

Under independent inputs, the expectation of the first sum in (24) is $O(m^{-98})$ by (22) and $I = O(mL)$. The second is bounded by $(C_2 I L^{-25}/(K_* + 1))^{K_*+1}$, hence is $o(m^{-100})$. Under arithmetic inputs, both estimates hold with an additional discrepancy bounded by

$$p^{-1/2} \exp(CH_0 + C(K_* + 1)L^3 + C(K_* + 1)L).$$

Indeed each term uses (23), the number of terms is at most $\exp(O((K_*+1)L))$, and (18) contributes a number of shifts factor $O(mL)$. The same calculation bounds the sum of arithmetic probabilities that any $|y_i| > R$, using the first-order b_i envelopes.

The truncated expectation using Q_i is a polynomial after averaging over Y. Its character norm is bounded by

$$\exp(CH_0 + CK_*L^3 + CK_*L).$$

For each choice of up to K_* slots and their output labels, apply (23) with the appropriate bounded function of the low counts. There are $O(1 + R/\tau_i)$ labels per slot, and the sums over labels and subsets cost at most $\exp(O(K_*L))$. Take

$$m = \left\lfloor \frac{a \log p}{\log \log \log p} \right\rfloor$$

with a sufficiently small absolute $a > 0$. Since $H_0 = O(m\ell)$ and $K_* L^3 = o(m)$, all arithmetic character-comparison errors in this section are $O(p^{-c})$ for an absolute $c > 0$. The shifts involved are distinct modulo p because $2H_J < p$ for all sufficiently large p. We have proved that the genuine full observation laws have total variation distance $O(m^{-98} + p^{-c})$, uniformly in B.

A.5 Transport cost

Map each observation outcome to a path via the reference low coefficients z_h, stable interpolation of the coded residual blocks, and pinning. The resulting path norm is at most $C(\log H_0 + \sum_H (R + \tau_H)/\sqrt{H}) = O(\log m)$ uniformly. Under either input law the probability of a high-channel fidelity failure is $O(m^{-98} + p^{-c})$, by the clipping estimate and the union of angular tail bounds over $O(mL)$ entries. Off this event the residual reconstruction error is bounded by $C \sum_H \tau_H/\sqrt{H} = O(m^{-1/2})$. On the exceptional event both truncated input paths and reconstructed paths have deterministic norm $O(\log m)$, so its expected cost is negligible. The endpoint error was bounded in Section 2. A maximal coupling of the full observations costs $O(\log m)$ times their total variation distance. Finally add the tails from Lemma A.2 and use the triangle inequality. This proves $W_1 = O(m^{-1/2})$, and hence Theorem A.1. Averaging the couplings over the finitely many B gives the two-parameter conclusion.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, proof drafting, and internal checks. No datasets or numerical certificates are required for the proof.

B Kloosterman path transport at prime powers

We study prime-power Kloosterman paths with the inverse-term coefficient fixed and the linear coefficient averaged. In the large-prime limit the uniform-path Wasserstein distance to the mixed-arcsine Fourier law is $O((\log p)^{-1/2})$, uniformly over all depths $n \geq 2$ in the modulus p^n. A lower bound of order $(\log p)^{-1/2}$ holds uniformly in the depth and follows from the at most $p - 1$ patterns of forced Fourier zeros. For fixed odd p we also give the depth-aspect rate $O_p(n^{-1/4})$ for the residue-conditioned limit. The proof combines stationary signed sums of square roots, integrated Fourier-block moments, and a sparse comparison of correlated noisy observations. At fixed prime, a local divided-difference estimate controls growing numbers of shifted coefficients.

Keywords Kloosterman paths, prime powers, p-adic stationary phase, Wasserstein distance.
Mathematics Subject Classification 11L05, 11T23, 60F17.

B.1 Path laws

Let $q = p^n$ for odd p and $n \geq 2$, and write $e_q(u) = \exp(2\pi i u/q)$ and $e(t) = \exp(2\pi i t)$. For unit residues A, B the normalized summands are $q^{-1/2} e_q(Ax + B/x)$ for unit x. Define $K_q(A, B)$ by joining the partial sums in increasing order of unit x, starting from zero, with equal time steps. Using the convention that starts at the first nonzero vertex and gives that vertex time zero changes the uniform norm by $O(q^{-1/2})$, as shown below. We write $\mu_{q,B}$ for this path law when A is uniform over the units, and $\mu_{q,a_0,B}$ when A is uniform over the class $a_0 \in \mathbb{F}_p^\times$.

For a sequence of independent centered coefficients $X_h \in [-2, 2]$ let $\mu(X)$ denote the law of

$$tX_0 + \sum_{h \neq 0} X_h \frac{e(ht) - 1}{2\pi i h}.$$

The series is summed symmetrically and converges almost surely uniformly, by the independent block bound below. Let σ be the law of $2\cos\Theta$ for Θ uniform on $[0,\pi]$. Define $\mu_{\mathrm{mix}} = \mu(X)$ for identically distributed X_h with law $(\delta_0+\sigma)/2$. For $a_0, b_0 \in \mathbb{F}_p^\times$ define $\mu^{\mathrm{sq}}_{p,a_0,b_0} = \mu(X)$ with law σ for each X_h such that $b_0(a_0-h)$ is a nonzero square modulo p, and $X_h = 0$ otherwise.

Ricotta and Royer [20] proved finite-dimensional convergence for fixed inverse coefficient and functional convergence with both parameters averaged as $p \to \infty$ for fixed $n \geq 2$. Ricotta, Royer and Shparlinski [21] proved fixed-parameter functional convergence for fixed $n \geq 31$ using a short-sum estimate. Milićević and Zhang [22] established the fixed-p depth limit with the square-gated law above, and proved that the law obtained by mixing these limiting laws converges as $p \to \infty$. Quantitative control at finite depth uniformly in p, n also requires paying for many close p-adic shifts in the completed Fourier sums.

Theorem B.1. *For all sufficiently large primes p, all $n \geq 2$ and all units B modulo p^n,*

$$\frac{c}{\sqrt{\log p}} \leq W_1(\mu_{p^n,B}, \mu_{\mathrm{mix}}) \leq \frac{C}{\sqrt{\log p}},$$

where W_1 uses the supremum norm and $c, C > 0$ are absolute.

Theorem B.2. *Fix an odd prime p and $a_0, b_0 \in \mathbb{F}_p^\times$. As $n \to \infty$, uniformly over all lifts B of b_0,*

$$c_p n^{-1/2} \leq W_1(\mu_{p^n,a_0,B}, \mu^{\mathrm{sq}}_{p,a_0,b_0}) \leq C_p n^{-1/4}.$$

For Theorem B.1, the middle frequencies use the last digit of A, whereas the high frequencies use a longer digit average. In a completed block, integration over time matches the frequency sum from the unconjugated factors with the frequency sum from the conjugated factors. This offsets the dimension factor in the supremum estimate. The longer digit average also works with growing numbers of shifts for Theorem B.2, through a local derivative estimate and a divided-difference count. Finally a sparse comparison of noisy time samples matches the lower order from the forced-zero patterns. Related stationary-phase methods for products of shifted sums appear in [22, 23].

B.2 Stationary phase and derivative roots

Put

$$W_h = q^{-1/2} \sum_{x \in (\mathbb{Z}/q\mathbb{Z})^\times} e_q((A-h)x + B/x).$$

When $A-h$ is a unit, change variables to identify it with $q^{-1/2}\sum_{x\in(\mathbb{Z}/q\mathbb{Z})^\times} e_q(x + C/x)$ for $C = B(A-h)$. If $A-h$ is not a unit, the original sum vanishes by translating x by a multiple of p^{n-1} and using the nonzero derivative modulo p. The transformed sum with nonunit C vanishes by the same argument, since $1 - C/x^2$ is a unit.

The transformed sum is zero unless C is a nonzero square modulo p. In that case let s be either Hensel root of C modulo q. For even $n = 2k$, partition x modulo p^k and sum the higher digits. A class contributes only when $1 - C/x^2$ vanishes modulo p^k, and the two normalized terms are $e_q(2s)$ and its conjugate. For odd $n = 2k+1$, first partition modulo p^{k+1}, which requires stationarity modulo p^k. Around a root s, use the p representatives $s + p^k y$, y modulo p. The phase is $2s + p^{2k}y^2/s$ modulo q, since $3k \geq 2k+1$. After normalization the root contributes $e_q(2s)G(s^{-1})$, where $G(d) = p^{-1/2}\sum_{y \bmod p} e_p(dy^2)$ has modulus one for d a unit. The other term is its conjugate. This is the usual prime-power stationary-phase formula (compare [22, Lemma 2]). Thus always $|W_h| \leq 2$, and on the square condition $W_h = 2\cos\theta_h$, where the additional Gauss phase depends just on the root modulo p.

Write $C_k(x) = \cos(k \arccos(x/2))$. A product of these polynomials on the square coordinates expands into exponential sums with phase $2\sum_i k_i s_i(A)$ for signed nonzero integers k_i on the active shifts. The additional Gauss factor is constant on each residue class of A modulo p. At nonsquare coordinates, $C_k(W_h) = \cos(k\pi/2)$ is deterministic. For the moment of a monomial in the W_h one may also first impose that all its shifts are square, since the monomial is zero otherwise.

For distinct active shifts $h_1, \dots, h_t$, take formal square roots $s_i(A)^2 = B(A - h_i)$. Over $\overline{\mathbb{F}}_p(A)$ the square classes of the $B(A - h_i)$ are independent, as shown by their valuations at the distinct places $A = h_i$. In the multiquadratic field, define

$$N(A) = \operatorname{Norm}\left(\sum_{j=1}^{t} k_j \prod_{i \neq j} s_i(A)\right),$$

where the norm is the product over the 2^t sign changes. This is nonzero if each k_i is nonzero in the field. Indeed the rational expression $\sum k_i/s_i$ is nonzero, by subtracting its image under a single sign change. The norm is a polynomial of degree at most $(t-1)2^{t-1}$, since expanding the product and cancelling the odd powers of each square root leaves that degree bound.

Lemma B.3. *For s shifts with distinct reductions modulo p, and integers $1 \le k_j < p$, the discrepancy between the arithmetic expectation of $\prod_{j=1}^{s} C_{k_j}(W_{h_j})$ and its expectation for independent $(\delta_0 + \sigma)/2$ variables is*

$$O(sp^{-1/2} + s3^s/p),$$

uniformly over all $n \ge 2$ and B.

Proof. On each class of A modulo p^{n-1}, average the last digit. A lift of a chosen root changes by $Bp^{n-1}u/(2s_i)$ when A changes by $p^{n-1}u$. Thus a nonempty square subset I in the product contributes only when a signed relation $\sum_{i \in I} \pm k_i/s_i = 0$ holds modulo p. As the roots are nonzero, its presence forces a zero of the norm polynomial in A modulo p. The exceptional sets over all subsets have total size $O(s3^s)$, since a subset of size t has norm degree at most $(t-1)2^{t-1}$. The exceptional residues $A = h_i$ cost another $O(s/p)$. On the event all coordinates are nonsquares, the product is identically $\prod \cos(k_i\pi/2)$. Its probability is $2^{-s} + O(s/\sqrt{p})$ by expansion of the product of $(1 - \chi(B(A - h_i)))/2$, with the quadratic character χ, using Weil's bound for squarefree products and excluding the at most $s+1$ singular or nonunit residues. For the independent law the positive-degree cosine means are $\cos(k_i\pi/2)/2$, which gives the matching main term. □

Lemma B.4. *Suppose t active shifts are distinct modulo p^u, with $u \ge 1$, and the signed coefficients k_i are nonzero integers. Put*

$$k = 2 + \lceil \log_p(2t) \rceil, \qquad V = \lfloor n/2 \rfloor - \min_i v_p(k_i) - (t-1)u - kt.$$

If $V > 0$, the normalized signed phase contribution on the square mask is $O(tp^{-V/(t-1)})$ for $t \ge 2$, and zero for $t = 1$. The same estimate holds in any specified unit residue class of A modulo p. Constant factors in the phase on each residue class, of absolute value at most one, may be included.

Proof. Sum the top digits of A in each class modulo $p^{\lceil n/2 \rceil}$. The Taylor expansion of the square roots in that class is linear modulo p^n, since the higher-order terms vanish. Orthogonality gives a zero contribution unless the derivative sum $D(A) = \sum_i k_i/s_i(A)$ is zero modulo $p^{\lfloor n/2 \rfloor}$. We bound the fraction of points satisfying this condition inside each class modulo p^k on which the gate is active. The condition $V > 0$ guarantees $k \le \lfloor n/2 \rfloor$.

Fix such a class centered at $A = a$. Let $c_i = k_i/s_i(a)$, $z_i = (a - h_i)^{-1}$, and $M_j = \sum_i c_i z_i^j$. The z_i are units and $v_p(z_i - z_l) = v_p(h_i - h_l) < u$. For i minimizing $v_p(k_i)$, the coefficients of $\prod_{l \neq i}(z - z_l)$ give a linear combination of $M_0, \dots, M_{t-1}$ equal to $c_i \prod_{l \neq i}(z_i - z_l)$. Consequently

$$b = \min_{0 \leq j < t} v_p(M_j) \leq \min_i v_p(k_i) + (t-1)u.$$

All later M_j have valuation at least b, by the integral recurrence with characteristic polynomial $\prod_i (z - z_i)$.

Expand on $A = a + p^k z$. The jth coefficient of $D(a + p^k z)$ is $p^{kj}\binom{-1/2}{j} M_j$. The binomial is integral for odd p and its valuation for $j \geq 1$ is the number of carries in $j + j$ in base p, at most $1 + \lfloor \log_p(2j) \rfloor$. For $j = 0$ its valuation is zero. Thus some index $j_0 < t$ attaining b gives a rescaled coefficient of valuation strictly below $b + kt$, while all $j \geq t$ have valuation at least $b + kt$. Divide the series by p^κ, where κ is its minimum coefficient valuation. The resulting restricted power series has last unit coefficient at some index $d \leq t - 1$, and $\kappa \leq b + kt$.

If $d = 0$ this series is a unit everywhere. Otherwise, setting $L = \lceil V/d \rceil$, suppose it takes values of valuation at least V at $d + 1$ points distinct modulo p^L. Their divided difference of order d then has valuation at least $V - d(L-1) > 0$, by the interpolation formula. On monomials of degree below d this divided difference is zero, at degree d it is one, and above d it is an integral homogeneous symmetric polynomial in the points. The last identity follows by expanding $\prod_i (1 - z_i X)^{-1}$. Passing to the restricted series shows that the divided difference equals the unit coefficient at d modulo p, a contradiction. There are thus at most d solution classes modulo p^L in the z variable. Here $L \leq n - k$. The relative count per class of A is at most $dp^{-V/d}$, which implies the result after averaging. □

B.3 Completed Fourier blocks

Extend the summand by zero at the nonunits and form Q_q on the full mesh j/q, $0 \leq j \leq q$. Exact Fourier inversion gives

$$Q_q(j/q) = \frac{j}{q} W_0 + \sum_{0 < |h| \leq (q-1)/2} b_{q,h} W_h(e(hj/q) - 1), \qquad b_{q,h} = \frac{1}{q(1 - e_q(-h))}. \tag{25}$$

Here $|b_{q,h}| \ll 1/|h|$ and $b_{q,h} = (2\pi i h)^{-1} + O(q^{-1})$. The jth unit $x_j = j + \lfloor (j-1)/(p-1) \rfloor$ has $x_j/q = j/\varphi(q) + O(1/q)$. The two unit time conventions differ uniformly from these mesh times by $O(1/q)$. Consecutive unit mesh times differ by at most $2/q$. Since Q_q has Lipschitz constant $\sqrt{q}$, this shows $\|K_q - Q_q\|_\infty \ll q^{-1/2}$, including the initial interval. A sharp retained polynomial of degree H has Lipschitz constant $O(H)$, so interpolation and its coefficient correction each cost $O(H/q)$.

For a dyadic block $H/2 < |h| \leq H$, let $T_H(t) = \sum b_h W_h e(ht)$ with $|b_h| \ll H^{-1}$. Parseval on T_H^r bounds its $2r$th sup moment by $(2rH + 1) \int |T_H|^{2r}$. Integrating retains only frequency tuples with

$$h_1 + \cdots + h_r = h_{r+1} + \cdots + h_{2r}. \tag{26}$$

Fixing $2r - 1$ slots determines the last, so the sum of their absolute coefficient products is at most C^{2r}/H. When two distinct integer shifts in a tuple agree modulo P and $r \geq 2$, mark that pair, choose their values in $O(H)$ and $O(H/P)$ ways, fix $2r - 3$ other values, and determine the last by (26). Collision implies $2H \geq P$, so the bound for the second marked value is uniform. The sum of absolute coefficient products of such tuples is thus $O(r^2 C^{2r}/(PH))$. Independent centered bounded coefficients give the usual subgaussian block moment root $O((rH)^{1/(2r)}\sqrt{r/H})$,

since $\sum |b_h|^2 \ll 1/H$. The same holds conditionally on any square gate, with a uniform constant. It gives uniform convergence of the independent series by summing the dyadic norm bounds with $r = \lceil \log(16H) \rceil$, and taking a maximum over partial block cutoffs with a bounded extra factor $O(H)^{1/(2r)}$.

For the terminal arithmetic tail in Theorem B.1, take $r = 2$ and $p \geq 5$. For $n \geq 400$ put $u = \lfloor n/200 \rfloor$. In a monomial on good tuples (distinct integer shifts distinct modulo p^u), expand cosine powers after fixing the square mask. The terms with signed count zero at each shift give the independent moment conditioned on that mask. All other phases have $t \leq 4$ and $|k_i| \leq 4$. The k_i are p-units and the loss $(t-1)u + kt$ in Lemma B.4 is less than $n/8$. So these terms cost $O(q^{-c_0})$. Use the balanced collision bound with $P = p^u$ for bad tuples, where $u \geq n/400$. Multiplying by the Parseval factor and taking fourth roots gives a block sup bound $O(H^{-1/4} + q^{-c_1})$. For $2 \leq n < 400$ instead apply Lemma B.3 to the good tuples modulo p, expanding powers as cosine polynomials of degree at most four. Its error is $O(p^{-1/2})$, and the collision bound uses $P = p$. This gives $O(H^{-1/4} + p^{-1/8})$, which also has the form $O(H^{-1/4} + q^{-c_1})$ for small enough c_1. After pinning and summing, the terminal tail beyond N has expected uniform norm

$$O(N^{-1/4} + (1 + \log q) q^{-c_1}) \tag{27}$$

uniformly in the depth.

For intermediate blocks through $N = \lfloor p^{1/8} \rfloor$, all distinct integer shifts are distinct modulo p. Expand powers in cosines and use Lemma B.3 for moments of degree $2r < p$. The sum of absolute cosine coefficients of $x^j = (2\cos\theta)^j$ is 2^j. Thus the error for each monomial is $C^{2r} p^{-1/2}$ uniformly, incorporating the 3^{2r} term of the cosine estimate. By (26), the arithmetic contribution to the block sup moment root is $O(p^{-1/(4r)})$. From a cutoff $H_* \asymp m \log m$ with $m = \lfloor a \log p \rfloor$ to N, take $r = \lceil \log(16 H_*) \rceil$ and sum by Minkowski, giving $O(\sqrt{r/H_*} + \log p \, p^{-1/(4r)})$. At this N, (27) is uniformly a negative power of p, since $q \geq p^2$. Hence truncation of both paths at H_* costs $O(m^{-1/2})$.

B.4 Transport at large base prime

We use the correlated observation construction of Appendix .1. Its arithmetic input is a comparison of characteristic functions of the retained normalized time samples, multiplied by the conditional likelihood of a low-frequency rounding outcome. Mixed first derivatives insert products of specified sample coordinates. The Fourier support and test sizes in the appendix are expressed by a block budget whose sum is $O(m)$. We verify the arithmetic comparison on that support.

Use the product cosine coefficient norm with weight 1 at zero frequency and 3 at each positive degree in one coordinate. The product formula for cosines makes it submultiplicative, and $\|x_h\|_{\mathrm{ch}} = 6$ in this norm. Lemma B.3 compares polynomials whose separate degrees are $< p$, on the retained $O(m \log m)$ shifts, with error $O(m \log m) p^{-1/2}$ times their cosine norm. If a polynomial has separate degrees d_h and bound U on the cube, its cosine norm is bounded by $U \prod_h (6 d_h + 1)$: apply angular substitution, bound each Fourier coefficient by U, and group signs. For total degree at most r in d coordinates the same argument gives

$$\|P\|_{\mathrm{ch}} \leq U \, 6^r \binom{d+r}{r}. \tag{28}$$

Indeed the number of integer angular modes is at most $2^r \binom{d+r}{r}$ and the weight is at most 3^r.

In every mixed derivative of an input characteristic expectation, replace the exponential of the linear phase by its Taylor polynomial through degree $\lfloor p/4 \rfloor$. On the product support the linear

phase has absolute coefficient sum $O(m)$, as checked in Appendix .1. The remainder on the coefficient cube is at most $\exp(-cp\log p)$. The additional sample-coordinate and low-rounding polynomial factors have degree at most $O(m^2 + m\log m)$ per coordinate, so all separate degrees in the comparison polynomial are less than p. Its weighted norm is bounded by the absolute exponential series norm times those of the polynomial factors. The remainder, even after multiplication by the polynomial cube bound and after all subsequent test-volume and subset-count factors, stays $\exp(-c'p\log p)$. Thus the coefficient $p^{-1/2}$ in the appendix's characteristic comparison can be used here, with the exponential factor quantified there. Taking $m = \lfloor a\log p\rfloor$ for sufficiently small fixed $a > 0$ gives a negative power of p as the arithmetic comparison error. Interpolation and the omitted frequencies were paid in Section 3. The observation coupling gives the upper bound of Theorem B.1.

For the lower bound take $H = \lceil C_2\log p\rceil$ and the shifts $H \le h < 2H$. For a fraction $1 - O(1/H + 1/p)$ of the unit residues A modulo p, at least $H/3$ of the W_h are forced zero by the square condition. Indeed this zero count has mean $H/2 + O(H/p)$ and variance $O(H + H^2/p)$, by the quadratic character product sum -1 at distinct shifts and exclusion of the zero residue. For each such fixed pattern, an independent mixed-arcsine vector has $|X_h| \ge 1$ on at least c_2H of the forced zero positions except with probability e^{-c_3H}, by the binomial tail bound, since $\mathbb{P}(|X_h| \ge 1) = 1/3$. There are at most $p-1$ patterns. Choose C_2 large, so with probability at least $3/4$ all these patterns satisfy that bound for the same independent path.

In any coupling, arithmetic typicality and this property of the limit have joint probability at least $1/2$ for large p. The pinned mesh path $Q_q(t) - tQ_q(1)$ has Fourier coefficient $e(h/(2q))\operatorname{sinc}(\pi h/q)W_h/(2\pi ih)$ at $h \ne 0$, by integrating its piecewise constant derivative. Here $\operatorname{sinc}(z) = \sin(z)/z$. It vanishes at each forced zero position. The corresponding coefficient of the pinned independent path is $X_h/(2\pi ih)$. Parseval on the selected positions now bounds the sup norm of their pinned difference from below by $c_4H^{-1/2}$. Pinning costs only a factor two, and K_q differs by $O(q^{-1/2})$. Taking expectation and the infimum over couplings proves the lower bound.

B.5 Fixed-prime depth convergence

Fix p and condition on $A \equiv a_0 \pmod p$, with $B \equiv b_0 \pmod p$. The square masks are deterministic, and the zero signed phases give the moments of the square-gated law. Set $m = \lfloor a\sqrt{n}\rfloor$ with small fixed $a > 0$.

Apply the same observation construction. In every mixed derivative of the characteristic expectation truncate the exponential at degree $\lfloor C_*m\rfloor$ for sufficiently large fixed C_*. Since the phase has absolute coefficient sum $O(m)$, the cube remainder is at most $\exp(-Lm)$, where L can be any fixed constant by taking C_* large enough. Choose it to exceed the remaining exponential amplification in the density and subset estimates. With C_* now fixed, use the same weighted cosine norm as in Section 4. Each monomial in the tested polynomial uses $O(m)$ distinct shifts: the low-label polynomial uses $O(m^{1/4})$, the derivatives use at most the selected dimension $O(m)$, and the exponential polynomial uses at most C_*m. Individual degrees are at most $O(m^2)$. All retained shifts are distinct modulo p^u for $u = 1 + \lceil\log_p(4H_*)\rceil$. The loss in Lemma B.4 is $O_p(m\log n) = o(n)$. The discrepancy of each nonzero cosine phase is $O_p(mp^{-cn/m})$. For small a the discrepancy remains exponentially small after the $\exp(O(m))$ norm and test-volume factors, since $n/m \asymp m/a^2$. The resulting comparison error, including the Taylor remainder, is exponentially small in m.

For the omitted frequencies take $r = \lceil\log(16H_*)\rceil$, $H_* \asymp m\log m$, and $u = \lfloor n/(40(2r))\rfloor$. On good moment tuples the loss in Lemma B.4 is $O(ru + r\log r)$ and is below $n/8$ for large n, allowing coefficients divisible by p. Colliding tuples use the bound with $P = p^u$ in Section 3. After taking $2r$th roots, each arithmetic error is $O_p(e^{-cn/r^2})$. These errors are negligible when summed over $O(n\log p)$ blocks, since $r = O(\log n)$. The subgaussian terms give $O(\sqrt{r/H_*}) = O(m^{-1/2})$. Mesh

and interpolation errors are negligible. Thus the coupling gives the stated upper bound.

For the lower bound the active indices of the gated law have positive periodic density in h. In any block $[H, 2H)$ with H large there are at least $c_p H$ active indices. For any prescribed complex number a and an arcsine coordinate X, the event that X has sign opposite to $\Re(a)$ and modulus at least one has probability at least $1/3$, and implies $|X - a| \geq 1$. Pinning and Parseval show that a uniform ball of radius ϵ forces the squared deviations over these coordinates to sum to at most $CH^2\epsilon^2$. With $H \asymp_p \epsilon^{-2}$ and a small constant, a binomial tail bound therefore gives ball probability at most $\exp(-c_p\epsilon^{-2})$, uniformly over the center. The arithmetic law on the fixed class has at most p^{n-1} points. The union of their balls at radius $c'_p n^{-1/2}$ covers at most half the limit law for small c'_p, proving the lower bound.

.1 Correlated observation coupling

Let m be a large integer, as chosen in the two arithmetic regimes above. Put $H_0 = \lfloor m^{1/4} \rfloor$. Independently round coefficients $|h| \leq H_0$ to $4\,\mathrm{Bin}(\lceil m^2 \rceil, (x_h + 2)/4)/\lceil m^2 \rceil - 2$. The expected path error is $O(\log m/m)$ by the binomial variance estimates and the absolute multiplier sum. There are at most $\exp(O(H_0 \log m))$ low output vectors, and the conditional probability polynomial of each has character norm at most the same exponential order by the separate-degree estimate.

Take blocks $H = 2^j H_0$, $j \geq 1$, through the first $H_* \geq m \log m$. In such a block let $y = A_H x$ be the $I_H = 16H$ real and imaginary parts of $\sqrt{H} \sum b_h x_h e(ht)$ on the $8H$ equally spaced mesh samples. Discrete Parseval gives $\|A_H\|_{2\to 2} \ll 1$ and $\|y\|_2^2 \ll H$ on the cube. In the independent model, the sample vector has the directional subgaussian bound $\mathbb{E} \exp(t \cdot y) \leq \exp(C \|t\|_2^2)$. For every subset S of samples, Gaussian integration then gives

$$\mathbb{E} \exp\left(\lambda \sum_{i \in S} y_i^2\right) \leq D^{|S|} \tag{29}$$

for absolute constants $\lambda > 0, D$.

For reconstruction, choose a smooth Fourier multiplier equal to one on $[-H, H]$, zero outside $[-2H, 2H]$, with kernel V_H. Discrete orthogonality on the $8H$ grid recovers the block as the sum of sampled values against weights $(8H)^{-1} V_H(t - j/(8H))$. The kernel bound $|V_H(u)| \ll H/(1 + H^2 \operatorname{dist}(u, \mathbb{Z})^2)$, obtained by two summations by parts, makes the absolute sum of the weights uniformly bounded. Thus if normalized samples are changed by at most τ in each component, the sup error after reconstruction and pinning is $O(\tau/\sqrt{H})$.

Use scale $\tau_H = (H/m)^{3/4}$ for $H \leq m$, and $\tau_H = A(H/m)^{1/4}$ for $H > m$, with large absolute A. The former blocks will be compared on all slots, and the latter on sparse subsets. Both parts of the sum $\sum_H \tau_H/\sqrt{H}$ are geometric and give $O(m^{-1/2})$.

Lemma .5. *Fix a positive constant J_0. For I real coordinates and noise scale $\tau > 0$, put $T = C_0 \max(1, \tau^4)$ with $C_0 \geq 1$. There is a probability density on $\mathbb{R}^I$, with noise vector η, with these properties:*

1. $\mathbb{P}(\max_i |\eta_i| > \tau/8) = O(I^{-J_0})$.
2. *For each slot set S, the probability that all $|\eta_i| > \tau/8$ in S is at most $2(Ce^{-c\sqrt{T}})^{2|S|}$.*
3. *The characteristic function of any D-coordinate marginal, $D \geq 1$, has support in a ball of radius $R(D) = CTI^{1/4}D^{1/4}/\tau$. Every mixed first derivative, once in each of r distinct coordinates, is bounded by $(C\tau)^r$ up to an absolute constant. These derivatives vanish outside the same ball.*

Proof. Take a real even smooth function f supported in $[-1/8, 1/8]$, with $\|f\|_2 = 1$ and $|\widehat{f}(u)| + |\widehat{f}'(u)| \le Ce^{-c\sqrt{|u|}}$. For example, use the normalized density of the sum of independent uniforms on $[-c_*/j^2, c_*/j^2]$, $j \ge 1$, with c_* small. Its sinc product has at least $c\sqrt{|u|}$ factors bounded by $1/2$ for large $|u|$, and a differentiated term loses at most one such factor and contributes $O(j^{-2})$. This also proves smoothness and the requisite derivative decay.

Take a smooth even cutoff χ decreasing in absolute argument, equal to one on $[-1, 1]$ and zero outside $[-2, 2]$. Form a partition with $\phi_0(u) = \chi(u/T)$ and $\phi_j(u) = \chi(u/(2^jT)) - \chi(u/(2^{j-1}T))$ for $j \ge 1$. Its terms are nonnegative, sum to one, and $\sum_j |\phi_j'| \ll 1$. The upper radii are $U_j = 2^{j+1}T$, and if $\phi_j(u) > 0$ for $j \ge 1$, then $|u| > U_j/4$.

Retain band configurations $(j_1, \dots, j_I)$ with $\sum_i U_{j_i}^4 \le C_1 IT^4$. Let w be the sum of their product partition functions, so $0 \le w \le 1$. We use the squared-amplitude device for positive Fourier smoothing as in [19, Section 7]. The restriction on the joint amplitude will control the Fourier support of every selected marginal. Define $\widehat{P}(u) = w(u) \prod_i \widehat{f}(u_i)$. Take the density $|P|^2 / \|P\|_2^2$ and dilate by τ in each coordinate. As usual, the Fourier normalization can be taken unitary, with absolute constants adjusted for the characteristic-function convention.

For a rejected configuration the bounds $U_{j_i} \le 4 \max(T, |u_i|)$ imply $\sum_i |u_i|^4 > C_2 IT^4$, with C_2 as large as needed by choosing C_1. Under the iid $|\widehat{f}|^2$ probability law, $|u_i|^4$ has all moments. For each fixed integer k, the $2k$th moment of the sum of centered copies is $O_k(I^k)$ by independence and expansion, since singleton factors vanish. Thus the rejected frequency mass is $O(I^{-2J_0})$ on taking k large enough. Choosing C_2 above four times the scalar fourth moment also bounds this mass by $1/4$ for every I by Markov's inequality. It follows that the amplitude P is $O(I^{-J_0})$ from the product amplitude in L^2, and hence the normalized densities are within the same order in total variation by Cauchy–Schwarz. This proves the first property; the uniform mass bound gives $\|P\|_2^2 \ge 3/4$.

For the joint bound, write f_j for the inverse transform of $\phi_j \widehat{f}$. Fix the band labels on S. The corresponding partial sum of the amplitude factors as the selected f_j times an amplitude in the complementary coordinates with L^2 norm at most one, because its partition multiplier lies between zero and one. On $|x| > 1/8$ we have $f_0 = -\sum_{j \ge 1} f_j$ in L^2. Moreover $\sum_{j \ge 1} \|f_j\|_2 \ll e^{-c\sqrt{T}}$ by the lower band radii and the decay of $\widehat{f}$. Summing the restricted L^2 bounds over selected band labels gives $(Ce^{-c\sqrt{T}})^{|S|}$. Squaring and normalizing proves the second property.

The support of $\widehat{P}$ has fourth power sum at most $C_1 IT^4$. Its projection to D coordinates has Euclidean radius at most $CTI^{1/4}D^{1/4}$ by Cauchy–Schwarz. The Fourier transform of $|P|^2$ is an autocorrelation of $\widehat{P}$, so the support of the selected characteristic function is in the corresponding difference set, after accounting for dilation. Finally a mixed derivative of $\widehat{P}$ on distinct slots is pointwise bounded by the product of $|\widehat{f}'| + C|\widehat{f}|$ there and $|\widehat{f}|$ elsewhere, by the partition derivative estimate. Its L^2 norm is $O(C^r)$. By Plancherel and Cauchy–Schwarz this bounds the product absolute moment of the corresponding η_i by $O((C\tau)^r)$. These moments justify differentiating the characteristic function, with the stated support. □

Draw the noises independently across blocks and independently of every input. In the dense blocks observe the noisy samples $y_i + \eta_i$. In each sparse block replace a noisy sample by zero whenever its absolute value is at most $\tau_H/2$. Clip the final output in either case to a centered interval of radius $C\sqrt{H} + \tau_H$, containing the deterministic range of y_i and its small perturbation. On the good event from Lemma .5(1) the sample error is $O(\tau_H)$.

.2 Comparison on selected marginals

Select every dense slot. In each sparse block H select D_H slots, with D_H bounded below by zero and above by the cutoff specified in the next section. We first compare the noisy real vectors before

thresholding or clipping, together with the rounded low output. Deterministic maps can only decrease total variation. Let g be the difference density of the selected noisy coordinates for one fixed low output label. Its Fourier transform factors as the product of the noise marginal transforms times the difference of the two input characteristic expectations, weighted by the conditional low label polynomial. The smoothing ensures a density, and the selected characteristic function is compactly supported.

For a Fourier test vector t_H, the character norm of $\exp(it_H \cdot A_H x)$ is at most $\exp(C\sqrt{H}\,\|t_H\|_2)$ by its absolutely convergent power series and $\left\|A_H^T t_H\right\|_1 \ll \sqrt{H}\,\|t_H\|_2$. A mixed derivative in r_H distinct variables in block H inserts a product of r_H sample coordinates. Its cube bound is at most $(CH/r_H)^{r_H/2}$ by the energy estimate and the arithmetic–geometric mean inequality. Its character norm is bounded by $\exp(Cr_H \log(eCH/r_H))$ by (28), since the block uses $O(H)$ Fourier inputs. The bound is one for $r_H = 0$. Taking C sufficiently large, one may bound this exponent using D_H in place of r_H for all $r_H \le D_H \le I_H$. The low label costs $\exp(o(m))$. With the derivative bounds for the noise in Lemma .5, these estimates control all mixed first derivatives of $\widehat{g}$ through the arithmetic comparison of Sections 4 and 5 and the product rule.

To bound the L^1 norm of g, assign $s_i = 1 + \tau_H$ in block H. Multiply and divide by the square root of $\prod_i (1 + (x_i/s_i)^2)$, then use Cauchy–Schwarz and Plancherel. If $D = \sum_H D_H$, this gives

$$\|g\|_1 \le \Big(\prod_i \pi s_i\Big)^{1/2} \left[C^D \sum_{U \subseteq \{1,\dots,D\}} \Big(\prod_{i \in U} s_i^{-2}\Big)\, \|\partial_U \widehat{g}\|_2^2\right]^{1/2}.$$

The weights absorb the powers of τ_H from the noise derivatives. The support is a product of balls. A ball of radius R in D_H dimensions has volume at most $(CR/\sqrt{D_H})^{D_H}$. Combining these facts with the arithmetic comparison gives, after summing over the low labels, a total variation upper bound of

$$\begin{aligned} &\varepsilon_m \exp\big(o(m) + C\textstyle\sum_H E_H(D_H)\big), \qquad \text{where} \\ &E_H(D_H) = \sqrt{H} R_H(D_H) + D_H \log(eCH/D_H) \\ &\qquad\qquad + D_H \log\left(\frac{C(1+\tau_H)T_H}{\tau_H}(I_H/D_H)^{1/4}\right), \end{aligned} \tag{30}$$

where the arithmetic factor ε_m is $p^{-1/2}$ for the large-prime argument and may be taken as $\exp(-Lm)$ for arbitrary fixed L in the fixed-prime argument. Also, $T_H = C_0 \max(1, \tau_H^4)$ and R_H is as in Lemma .5. Terms with $D_H = 0$ are omitted. The exponent includes the 2^D derivative sum and the polynomial factor for the number of Fourier inputs.

For dense blocks $H \le m$ take $D_H = I_H$. The contribution is $O(H/\tau_H + H[1 + \log(1/\tau_H)])$, summable to $O(m)$ because $H/\tau_H = m^{3/4}H^{1/4}$. In a sparse block $H > m$ define

$$K_H = \min\left(I_H, \lceil \max(m^{1/2}, C_2 I_H e^{-c_2 \tau_H^2}) \rceil\right)$$

with c_2 sufficiently small and C_2 sufficiently large. We allow $D_H \le \min(I_H, K_H + 1)$. Since $H \le 2m \log m$ up to an absolute factor and $\tau_H^2 \ll \sqrt{\log m}$, the $m^{1/2}$ term is eventually smaller than the exponential term throughout this range. Thus $x = D_H/I_H \le Ce^{-c_2\tau_H^2}$. The first contribution of this block to (30) is $O(H(T_H/\tau_H)x^{1/4})$. The others are bounded by $O(Hx[1 + \log(1/x) + \log T_H + \log(1 + 1/\tau_H)])$. Together they are $O(H(1+\tau_H)^C e^{-c_3\tau_H^2})$. Since $\tau_H^2 = A^2\sqrt{H/m}$, their dyadic sum is $O(m)$. In particular, every selected marginal compares uniformly in total variation at cost $\varepsilon_m \exp(Cm)$.

.3 Sparse comparison and transport

Under the independent model, a specified set of D_H sparse slots is simultaneously nonzero after thresholding only if in each of those slots either $|y_i| \geq \tau_H/4$ or $|\eta_i| > \tau_H/8$. Split the slots between these alternatives. For the first use (29) and Markov's inequality, and for the second use Lemma .5(2). Independence of noise and input gives the product bound, summed over at most 2^{D_H} allocations. With A large the joint activity probability is at most $\exp(-c_4\tau_H^2 D_H)$. These bounds multiply across blocks, since the independent Fourier coordinates and the noises are independent across blocks.

The argument uses truncated inclusion–exclusion with a factorial error bound, as in read-once Boolean tests [9, Section 3], here for arbitrary bounded tests of multiple observed values. For a realized sparse block output with N_H nonzero entries, expand its point mass as a product of zero point masses plus coordinate differences. Retain terms involving at most K_H differences. If $N_H \leq K_H$ this is exact. Otherwise the signed variation error is at most

$$B_H = 2 \cdot 3^{K_H+1} \binom{N_H}{K_H+1},$$

by the binomial bounds $\binom{N_H}{j} \leq \binom{N_H}{K_H+1}\binom{K_H+1}{j}$ for $j \leq K_H$ and the factor 2^j for signed differences. Take $B_H = 0$ if the cap keeps all coordinates. For a product across sparse blocks, telescoping bounds the aggregate variation error by $\prod_H (1 + B_H) - 1$.

Expand this last product to estimate its expectation. Each factorial count is a sum over sets of prescribed active slots. Under the independent law the term for a chosen block is bounded by

$$\left(\frac{C I_H e^{-c_4\tau_H^2}}{K_H + 1} \right)^{K_H+1}.$$

Take $c_2 < c_4$ and C_2 large in the definition of K_H. This is at most $\exp(-cK_H)$, so the whole expected error is smaller than any fixed power of m^{-1}. Under arithmetic input compare each selected joint activation by the selected-marginal bound (30). The sum of all weights and subset counts is $\exp(O(m))$, because $\sum_H K_H \log(eI_H/K_H) = O(m)$ by the same sparse budget estimate. Hence the arithmetic expected truncation error is bounded by $O_{J_0}(m^{-J_0}) + \varepsilon_m \exp(Cm)$.

For any test of the full observations bounded by one, the truncated expectation is a sum of tests on all the dense outputs, the low rounded vector, and at most K_H selected sparse coordinates in each block. Evaluate the other sparse outputs at zero in each term of the signed expansion. The same subset count and signed amplification is $\exp(O(m))$, so (30) compares these truncated expectations with error $\varepsilon_m \exp(Cm)$. Combining the two truncation errors shows that the full genuine observation laws differ in total variation by

$$O_{J_0}(m^{-J_0}) + \varepsilon_m \exp(Cm)$$

uniformly in the permitted inverse coefficient for the respective arithmetic comparison.

In the large-prime argument take $m = \lfloor a \log p \rfloor$ with a small. In the fixed-prime argument, take $m = \lfloor a\sqrt{n} \rfloor$ and choose L in the arithmetic estimate large. In either case the arithmetic part of the total variation bound is exponentially small in m. A maximal coupling of the observations gives a common reconstructed path except on an event of the bounded probability. All clipped paths have sup norm $O(\log m)$ by stable interpolation. On the compact noise event the reconstruction error for either input is $O(\sum_H \tau_H/\sqrt{H}) = O(m^{-1/2})$ in addition to binomial rounding. The complementary probability is $O(\sum_H I_H^{-J_0}) = O(m^{-J_0/4})$. Choosing J_0 sufficiently large and using deterministic $O(\log m)$ bounds on reconstructed and truncated true paths makes the expected exceptional cost negligible. Together with the tail estimates in the main text this gives $W_1 = O(m^{-1/2})$.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, proof drafting, and internal checks. No datasets or numerical certificates are required for the proofs.

A Hyper-Kloosterman paths at prime powers

We study prime-power paths whose Fourier coefficients are hyper-Kloosterman sums of fixed rank R. Their limits depend on $d = \gcd(R, p-1)$ and can also depend on the depth parity. When $d \geq 2$ the Wasserstein distance in the uniform norm to the independent coefficient model has order $(\log p)^{-1/2}$ as p grows, uniformly in depth. When $d = 1$, the model uses circle Haar coefficients. At sufficiently large depth the distance then has order $(\log p)/p$ for W_1 and $p^{-1/s}$ for W_s, $s > 1$, due to a single progression of missing Fourier coefficients. Conditioned on a residue class, the distance to its gated limit is $O_{R,s}(J^{-1/2})$ with $J = \min(n, \sqrt{n \log p})$. The estimates hold for each fixed finite s. In some ranks the even- and odd-depth laws at a fixed prime have distinct limits.

Keywords hyper-Kloosterman sums, prime powers, random Fourier series, optimal transport.
MSC 2020 11L05, 11L07, 11T23, 60G17.

A.1 Path laws

Fix $R \geq 2$ and $q = p^n$ with $p > R$ an odd prime and $n \geq 2$. Write $e_q(x) = \exp(2\pi i x/q)$ and

$$\mathrm{Kl}_j(c; q) = q^{-(j-1)/2} \sum_{x_1 \cdots x_j = c} e_q(x_1 + \cdots + x_j),$$

with all variables units, so the sum vanishes at nonunit c. Keep a a unit. Form $K_q(a, b)$ by accumulating the increments $q^{-1/2}\,\mathrm{Kl}_{R-1}((ax)^{-1}; q)e_q(bx)$ in increasing unit order x with equal time steps. Let $\mu_{q,a}$ be its law with b uniform over units, and μ_{q,a,b_0} the conditional law at a nonzero residue b_0 of b modulo p.

Put $d = \gcd(R, p-1)$, $L = (p-1)/d$, and let ζ_d be a primitive complex dth root of unity. Use the compact torus $\mathcal{T}_d$ in $\mathbb{T}^d$ specified by

$$\prod_{j=0}^{d-1} Z_j^{m_j} = 1 \quad \text{whenever} \quad \sum_{j=0}^{d-1} m_j \zeta_d^j = 0, qquad m_j \in \mathbb{Z}.$$

This is the image of Haar on a $\varphi(d)$-torus via the cyclotomic Laurent powers [27]. We always take $Z = (Z_j)$ Haar on $\mathcal{T}_d$. Its coordinate functions are distinct nontrivial characters.

For the model at a residue b_0, write $a_0 = a \pmod p$ and $c_h = (b_0 - h)/a_0$. The coefficient at h is zero if c_h is not a nonzero dth power. At an active index use independent copies of the following variables. Write χ for the quadratic character, $\varepsilon_p = 1$ if $p \equiv 1 \pmod 4$ and i if $p \equiv 3 \pmod 4$, and put

$$\kappa = \varepsilon_p^{R-1} \chi(2^{-(R-1)} R), \qquad \tau = (-1)^L.$$

- If n is even, use $\sum_j Z_j$.
- If n and R are odd, use $\kappa \sum_j Z_j$.
- If n is odd and R even with $\tau = -1$, use $\kappa \sum_j (-1)^j Z_j$.

- If n is odd and R even with $\tau = 1$, use $\kappa\sigma_h \sum_j Z_j$, where $\sigma_h = 1$ if c_h is a $(2d)$th power and -1 otherwise.

The complete independent model chooses the gates independently, with active probability $1/d$, and draws a fair independent sign at each active coordinate in the last case. All active coefficients have mean zero and second absolute moment d by character orthogonality.

For either model the path from its coefficients is

$$tX_0 + \sum_{h\neq 0} X_h \frac{e(ht)-1}{2\pi i h}, \qquad e(t) = e^{2\pi i t}.$$

Write $\nu_{p,n}$ for the complete independent model and ν_{p,n,a_0,b_0} for the residue model. Put $J = \min(n, \sqrt{n\log p})$ and use W_s in the uniform path norm. When $d = 1$, the complete model is just the path law $\nu_{\mathbb{T}}$ with independent circle Haar coefficients. Write

$$G_s(p) = \begin{cases} (\log p)/p, & s = 1, \\ p^{-1/s}, & s > 1. \end{cases}$$

Theorem A.1. *For fixed R and $1 \le s < \infty$, uniformly in $p > R$, depths and units,*

$$W_s(\mu_{q,a,b_0}, \nu_{p,n,a_0,b_0}) \ll_{R,s} J^{-1/2}.$$

As $p \to \infty$, when $d \ge 2$,

$$W_s(\mu_{q,a}, \nu_{p,n}) \asymp_{R,s} (\log p)^{-1/2}$$

uniformly in $n \ge 2$. When $d = 1$,

$$\max\{(n\log p)^{-1/2}, G_s(p)\} \ll_{R,s} W_s(\mu_{q,a}, \nu_{\mathbb{T}}) \ll_{R,s} \min\{(\log p)^{-1/2},\, J^{-1/2} + G_s(p)\}.$$

In particular the latter distance is of order $G_s(p)$ whenever $J^{-1/2} \ll_{R,s} G_s(p)$, and is of order $(\log p)^{-1/2}$ at bounded depths, with constants allowed to depend on that depth bound.

At rank three, $d = 3$ for $p \equiv 1 \pmod 3$, where $\kappa = 1$ and the active trace is $Z_1 + Z_2 + (Z_1 Z_2)^{-1}$ for two independent circle variables. For $p \equiv 2 \pmod 3$, $d = 1$. A different behavior occurs at $R = 15, p = 37$: here $d = 3$ and $\kappa = \chi(-15) = -1$. At fixed p and a_0, the full arithmetic law converges along each parity to the mixture of the corresponding residue models, with error $O_{R,s,p}(n^{-1/4})$ by coupling on b_0 and using the conditioned bound. The lifts of a_0 can vary with n. Their endpoint third moments are $+2$ and -2, since the torus trace has third moment 6 and is active at the endpoint in exactly one third of the residues. Thus they are distinct limits. An even-rank example is $R = 6, p = 7$, where the even-depth endpoint mixture has third moment 2 from the two equilateral triples of sixth roots, and the odd-depth trace is symmetric by cyclic rotation.

Explicit tame hyper-Kloosterman evaluations via stationary phase go back to Smith and Dbrowski–Fisher [24, 25, 26]. For rank three see [25, Example 1.16]. The active unweighted scalar torus law occurs in subgroup equidistribution, as described by Untrau [27, Theorem 1.3 and Section 5.3]. We use the same phases to compare entire paths, allowing odd-depth Gauss weights and the changing gate density. For rank two, Ricotta and Royer [20], Ricotta, Royer and Shparlinski [21], and Milićević and Zhang [22] established qualitative path laws in different regimes. The estimates here use growing translated phase comparisons in higher rank.

A.2 Complete coefficients and torus modes

By completing the Fourier sum at shift h, if $b-h$ is a unit, the change $x \mapsto (b-h)x$ gives $W_h = \mathrm{Kl}_R((b-h)/a; q)$. If $b-h$ is a nonunit, write the rank-$(R-1)$ summand with $R-2$ free unit variables, of product v. The phase in x is $(b-h)x + (avx)^{-1}$. Its derivative is a unit, so the last digit of x sums to zero. For $R=2$ the empty product is one. The classical scalar evaluations are discussed in [24, 25, 26]. For clarity we derive the weights. The stationary points of the phase $\sum_i x_i + c/\prod_i x_i$, with $R-1$ summation variables, are all $x_i = s$ with $s^R = c$. The Hessian there is $s^{-1}(I + \mathbf{1}\mathbf{1}^{\mathsf{T}})$, with determinant $Rs^{-(R-1)}$ a unit. At even depth the top half-digit sum imposes stationarity and gives $e_q(Rs)$ from each root. At $n = 2k_0+1$ the last remaining lower digit deviations yield a normalized Gaussian sum on $R-1$ variables at the half-Hessian; higher terms vanish since $3k_0 \geq 2k_0 + 1$. The Gauss factor is $\kappa\chi(s)^{R-1}$. Thus when active the terms use the phase tuple

$$\big(e_q(Rs\zeta^j)\big)_{0\leq j<d},$$

where $\zeta \in \mathbb{Z}_p$ is a primitive dth root and s is a chosen branch. At odd R the coefficient factor is the constant $\kappa = \chi((-1)^{(R-1)/2}R)$. At even R, the factor on the jth root is $\kappa\chi(s)\tau^j$. If $\tau = -1$, choose an Rth root branch that is a square modulo p; half the Rth root branches are squares in this case. A cyclic rotation of the torus preserves Haar and changes the weighted trace by a sign, so the model does not depend on which branch is chosen. If $\tau = 1$, all roots have the same quadratic character. In terms of a multiplicative generator of $\mathbb{F}_p^\times$, write $s = g^r$ and $c = g^{dj}$. Then $j \equiv (R/d)r \pmod{L}$, and when L is even, R/d is odd and the parity of r equals the parity of j. This gives exactly the claimed sign labels.

A phase mode with integer exponents m_j gives $\gamma = \sum_j m_j\zeta^j$. It is zero in $\mathbb{Z}_p$ precisely for a trivial torus mode, and a nonzero mode satisfies

$$v_p(\gamma) \leq \log_p |N_{\mathbb{Q}(\zeta_d)/\mathbb{Q}}(\gamma)| \ll_R \log_p(2 + \max_j |m_j|).$$

All conjugates are integral and the absolute norm is at most a fixed power of the maximum mode size. The finite label at a nonsingular shift is a multiplicative power class of order d, or of order $2d$ in the sign-labeled case. A gate or sign restriction thus expands in multiplicative characters of bounded order. At singular shifts the coefficient is zero. If needed augment inactive coordinates by unused independent Haar phases on the torus, both in arithmetic and in the model, so nontrivial phase modes vanish there. For $d = 1$ use a separate fixed inactive symbol for the singular residue, absent in the complete independent model.

Use the product of each label space with the cyclotomic torus. A finite label can be Fourier expanded on a cyclic set of bounded size, with fixed numerical values for the trace at its labels. A trace or its conjugate has Fourier norm $O_R(1)$, where the norm is the sum of absolute Fourier coefficients on each product group. It is submultiplicative. A polynomial in bounded traces and their duals, uniformly bounded by one and of total degree D_0 over d_0 shifts, has norm

$$\exp\big(C_R D_0 \log(eC_R(d_0 + D_0)/D_0)\big) \quad (D_0 \geq 1).$$

Indeed each Fourier coefficient is bounded by orthogonality. Only $O(D_0)$ shift factors may be nontrivial in a monomial, the torus mode total size is $O_R(D_0)$, and a nontrivial label mode can occur only at a used factor. Counting those modes with stars and bars gives the bound. The separate-degree bound is $\prod_h C_R(1 + D_h)^{C_R}$ over nonzero degrees. Both norms also apply to real and imaginary parts.

A.3 Arithmetic comparison

First use distinct shift residues modulo p. Fix the active labels at the slots that carry nontrivial phase modes and average the last digit of b. The phase survives only if $\sum_i \gamma_i / s_i^{R-1} = 0 \pmod p$, since its derivative is a^{-1} times this sum. When each nontrivial γ_i is a unit, this forces a zero of a nonzero polynomial of degree $O_R(tR^t)$, with t the number of nontrivial phase slots. Over $\overline{\mathbb{F}}_p(b)$ the Rth roots of the linear functions $(b - h_i)/a$ generate a Kummer extension with group of order R^t. The distinct simple zeros give independence by valuations. Rotating just one chosen root changes its term by a nontrivial root of unity, because $\gcd(R, R-1) = 1$, so the rational derivative sum is nonzero. Clearing the denominator and taking the norm gives the degree and nonvanishing. This accounts for all possible choices of branches.

For terms with trivial torus modes the comparison reduces to the labels at the tested shifts. The character of order d, or $2d$ in the sign-labeled case, gives their joint product probabilities up to $O_R(p^{-1/2} e^{C_R t})$. The trivial character terms give the independent main term after discarding the at most $t + 1$ singular and excluded residues. Each other term has a distinct linear factor with exponent not divisible by the character order, so the multiplicative Weil bound applies. If $d = 1$, discard the same exceptional residues and all other labels are active. For terms with a nontrivial torus mode, the choice of unused Haar variables on inactive coordinates makes those cases vanish. If all the nontrivial modes are active, the derivative argument bounds the arithmetic contribution while the Haar main term is zero. Hence the complete-family discrepancy for an individual Fourier mode on the product of labels and tori is $O_R(p^{-1/2} e^{C_R t})$ uniformly in depth. Here t can count all participating shifts. This applies whenever all nontrivial γ_i are units, which includes modes of size $O((\log p)^C)$ for any fixed C as p grows.

For the depth comparison we use arbitrary nonzero modes, retaining their valuations. Suppose t active shifts are distinct modulo p^u, $u \geq 1$, and the mode coefficients $\gamma_i \in \mathbb{Z}_p$ are nonzero. Put

$$k_1 = 2 + \lceil \log_p(Rt + R) \rceil, qquad V = \lfloor n/2 \rfloor - \min_i v_p(\gamma_i) - (t-1)u - k_1 t.$$

If $V > 0$, the normalized contribution of the corresponding product phase, conditional on a fixed active residue class, is $O(tp^{-V/(t-1)})$ when $t \geq 2$ and is zero when $t = 1$.

To see this, first sum the top digits after fixing b modulo $p^{\lceil n/2 \rceil}$. Taylor expansion is linear in the summed digits, so only points where $D(b) = \sum_i \gamma_i / s_i(b)^{R-1}$ is divisible by $p^{\lfloor n/2 \rfloor}$ contribute. On a ball $b = b_1 + p^{k_1} z$ set $c_i = \gamma_i / s_i(b_1)^{R-1}$, $y_i = (b_1 - h_i)^{-1}$ and $M_j = \sum_i c_i y_i^j$. There is an index $0 \leq j < t$ with

$$v_p(M_j) \leq \min_i v_p(\gamma_i) + (t-1)u.$$

Indeed, for i attaining the minimum, use the coefficients of $\prod_{\ell \neq i}(X - y_\ell)$ as a linear combination of the first t moments. It takes value $c_i \prod_{\ell \neq i}(y_i - y_\ell)$, and $v_p(y_i - y_\ell) = v_p(h_i - h_\ell) < u$. If $b_2 = \min_{j<t} v_p(M_j)$, all later M_j have valuation at least b_2 by the integral recurrence with roots y_i.

The coefficient of z^j in the restricted expansion of D is $p^{k_1 j} \binom{-(R-1)/R}{j} M_j$. The binomial is p-integral and its valuation is at most $\log_p(Rj + R)$: in the numerator product of the positive progression $R-1, 2R-1, \ldots, Rj-1$, the number of multiples of each p^i differs by at most one from the denominator $j!$, and the difference is nonnegative. Some index $j < t$ attaining b_2 has a coefficient valuation strictly smaller than $b_2 + k_1 t$, while all later indices have valuation at least $b_2 + k_1 t$. Divide by the smallest coefficient valuation. The restricted series then has last unit coefficient at an index $d_1 \leq t - 1$, and surviving points are zeros modulo at least p^V of this series. If $d_1 = 0$ there are none. Otherwise they occupy at most d_1 residue classes modulo $p^{\lceil V/d_1 \rceil}$. Indeed $d_1 + 1$ such distinct classes would give a divided difference of valuation at least $V - d_1(\lceil V/d_1 \rceil - 1) > 0$. But on the restricted power series this divided difference is the unit coefficient at d_1 modulo p: monomials of

lower degree vanish, and higher divided differences are integral symmetric polynomials, so their nonunit coefficients vanish modulo p. This contradiction bounds the proportion within each ball by $d_1 p^{-V/d_1}$. Averaging such balls inside the specified residue incurs no additional factor. This is the same local power-series mechanism as in high-depth estimates for rank-two shifts [22, 28].

Take $m = \lfloor a_1 J \rfloor$ for sufficiently small a_1. In the depth aspect we can compare tests involving shifts up to $N = \lfloor e^J \rfloor$ by choosing $u = 1 + \lceil \log_p(4N) \rceil$. For a term with $O(m)$ active shifts and mode size $O(m^2)$, the valuations cost $O(\log(2J)/\log p)$ and the other losses in V are $O(a_1 n) + o(n)$ uniformly in p, since

$$J \le n, \qquad J^2 \le n \log p, \qquad J \log(2J)/\log p = O(n).$$

For a_1 small and J large, $V \ge n/4$. The discrepancy is exponentially small in $n \log p/m \ge J/a_1$, which pays any fixed exponential cost in m. The model here is the deterministic gate and label law on each residue; zero modes give its exact main terms.

A.4 Fourier blocks and transfer to whole paths

Extend the summand by zero at nonunits to form the mesh path Q_q at times j/q. Discrete Fourier inversion and summing a geometric progression give at mesh vertices

$$Q_q(t) = tW_0 + \sum_{0<|h|\le (q-1)/2} \frac{W_h(e(ht)-1)}{q(1-e(-h/q))}. \tag{31}$$

The normalized rank-$(R-1)$ summand is bounded by $R-1$ by the same stationary phase formula, using $p > R$ (for rank one the summand has modulus one). The jth unit has position $j + \lfloor (j-1)/(p-1) \rfloor$, so its normalized mesh time differs from the equal-unit time $j/\varphi(q)$ by $O(1/q)$. Thus $\|K_q - Q_q\|_\infty = O(q^{-1/2})$. The coefficient in (31) is $O(1/|h|)$ and differs from $(2\pi i h)^{-1}$ by $O(1/q)$. Retaining degree N and interpolating its polynomial has cost $O(N/q)$.

For a block of size U with coefficients $|b_h| \ll U^{-1}$, integration of $|\sum_h b_h W_h e(ht)|^{2r}$ keeps only frequency tuples with $h_1 + \cdots + h_r = h_{r+1} + \cdots + h_{2r}$. Their total absolute coefficient weight is $O(C^{2r}/U)$, since fixing all but the last index determines the last. The supremum to power $2r$ is at most $(2rU+1)$ times this integral by Parseval and Cauchy–Schwarz on the rth power. For fixed $r \ge 2$, tuples with a collision between distinct shifts modulo P_0 have coefficient weight $O_r(1/(UP_0))$: mark the pair, with $O(U^2/P_0)$ choices, fix $2r-3$ other shifts and use the frequency equality for the last. The model block from independent centered bounded traces has moment root

$$O((2rU+1)^{1/(2r)}\sqrt{r/U})$$

by the pointwise subgaussian bound. This applies also after any deterministic gates and gives almost sure uniform convergence of the independent Fourier series by dyadic summation with $r \asymp \log(2U)$, including partial blocks.

For the terminal arithmetic tail fix $2r \ge \max(4, s)$. At depths n sufficiently large in terms of R, r, take $P_0 = p^{\lfloor n/C_r \rfloor}$ with C_r large in terms of R, r. The local phase bound on good tuples gives an error $O_r(q^{-c_r})$ before roots, since $t \le 2r$ and cyclotomic valuations of bounded modes are uniformly bounded. The collisions pay P_0^{-1}. In the comparison one conditions on a residue and uses the gated model, whose pointwise moment bound is uniform. For the complete family at bounded depths take $P_0 = p$ and use the last-digit estimate above the fixed prime threshold for the chosen moment order. There are finitely many remaining primes and shallow depths, where the

trivial block bound is absorbed into the constants. Summing after pinning gives an L^s terminal tail beyond N of size

$$O_s(N^{-1/2+1/(2r)} + (1+\log q)q^{-c_r} + N/q)$$

in the complete family and the sufficiently deep conditioned family.

For shallow depths with a fixed residue b_0, only a uniform $O_{R,s}(1)$ bound on the path norm is needed, since J is bounded in this depth range. We count resonances over the frequency tuples instead of over the parameter residues. Use fixed $2r \geq \max(4, s)$. On balanced tuples, a repeated integer shift costs $O_r(U^{-2})$ in total coefficient weight, because its equation is linearly independent of frequency balance for $2r \geq 4$. Distinct shifts that collide modulo p cost $O_r((pU)^{-1})$. Now take the other tuples. Each shift occurs once, so if active its individual trace mode has unit coefficient. If an inactive trace appears, the product vanishes. Otherwise averaging the last digit of b gives zero unless the derivative sum vanishes modulo p. View the shifts as variables on the frequency-balance hyperplane, eliminate one by balance, and use the norm polynomial argument of Section 3 with linear forms $b_0 - h_i$. They are pairwise nonproportional after this elimination for $2r \geq 4$, so the Kummer norm is a nonzero polynomial of bounded degree in $2r-1$ free shift variables. A polynomial of bounded degree in that many variables vanishes on at most $O_r(U^{2r-1}/\min(U,p))$ points of the integer box: for $U \leq p$ this is the elementary root count on a grid, splitting the box into boundedly many boxes without repeated residues if necessary, and for $U \geq p$ count each residue tuple with multiplicity $O((U/p)^{2r-1})$. This also holds for the union of the finitely many torus modes. The added moment discrepancy is therefore $O_r((U\min(U,p))^{-1})$. The same Parseval bound yields a block norm $O_r(U^{-1/(2r)} + p^{-1/(2r)})$ after taking roots. The dyadic sum through q is bounded because n is bounded. This proves the required uniform shallow conditional bound.

For complete comparison now take $m = \lfloor a_1 \log p \rfloor$ and $N = \lfloor p^{1/8} \rfloor$. For the conditioned depth case use $m = \lfloor a_1 J \rfloor$ and $N = \lfloor e^J \rfloor$ as above. The terminal beyond N is negligible. For bands from $H_* \asymp m \log m$ through N use $r = \max(\lceil s/2 \rceil, \lceil \log(16m) \rceil)$, which has logarithmic order for large m. All shifts are distinct modulo p in the base aspect and modulo p^u with the chosen u in the depth aspect, unless equal as integers. Expanding the individual trace powers into torus modes bounds the sum of absolute coefficients by R^{2r}. Their errors before roots are exponentially small in $\log p$ in the base aspect and in $n \log p/r$ in the depth aspect, after the frequency balance cancels the factor U. Summing after roots gives a negligible discrepancy across the $O(\log p)$ or $O(J)$ blocks. The model terms sum to $O(\sqrt{r/H_*}) = O(m^{-1/2})$. This pays the tail above H_* in L^s in both aspects.

For the frequencies retained below H_* apply the sparse comparison of correlated observations in Appendix .1. Its input is the comparison of input characteristic functions and their mixed derivatives on the selected Fourier supports, multiplied by the low-label rounding probability. The appendix shows that the absolute coefficient sum of the phase linear form is $O(m)$ there. Expand its exponential through degree $C_* m$. By taking fixed C_* large the uniform remainder on the trace cube is e^{-Am} with A larger than the subsequent exponential costs. The expansion uses only $O(m)$ active shifts per term, even though H_* itself is larger than m. Individual polynomial degrees are $O(m^2)$ including low rounding; the mixed derivative adds a product of selected sample coordinates with the total-degree norm bound from Section 2. The Laurent norm of the exponential expansion is at most e^{Cm} by the exponential series and the bounded Fourier norm of each trace. Thus all modes in the base comparison are units when p is large and the comparison error remains $e^{-A'm}$ after choosing a_1 small. The same holds in the depth aspect by its local comparison and the cyclotomic valuations. The base input error is $p^{-1/2}e^{Cm} + e^{-Am}$ before the subsequent exponential costs, and the prescribed A absorbs those costs. This verifies the arithmetic input to the appendix. Combining its $O_{R,s}(m^{-1/2})$ reconstruction and coupling with the tails proves the upper bounds against the complete independent models and the residue-conditioned models. If m is bounded,

the terminal block bounds yield a uniform $O_{R,s}(1)$ estimate, which suffices for those cases.

A.5 The lower bounds and the single missing residue

We first recall the finite-support argument in these models. For any center, a uniform ball of radius ϵ forces the sum of squared deviations of the normalized pinned coefficients over $H \le h < 2H$ to be $O(H^2\epsilon^2)$ by Parseval. In either nondegenerate independent bounded model, $\sup_z \mathbb{E}e^{-|X-z|^2} \le 1-c_R$. Taking H a sufficiently small constant times ϵ^{-2} and using exponential Markov bounds the ball mass by $e^{-c\epsilon^{-2}}$. A law with at most q atoms therefore has W_s at least $c_s(\log q)^{-1/2}$ from either model: the union of one ball about each atom covers at most half the independent law at sufficiently small radius of that order.

For $d \ge 2$, let H be a large fixed multiple of $\log p$. On $H \le h < 2H$ all but $o(1)$ of the residues b_0 have at least cH inactive slots, by the pairwise power-class gate comparison and Chebyshev (variance $O(H + H^2p^{-1/2})$). There are at most $p-1$ such masks. For any one of them, on a fixed positive fraction of its inactive slots the independent coefficient has modulus at least a fixed positive constant except for probability $e^{-c'H}$. This follows uniformly from $\mathbb{E}|X_h|^2 = 1$ and $|X_h| \le R$ in the complete independent law. A union over the masks makes this event simultaneous with probability $1 - o(1)$, for the chosen large multiple. In any coupling one can lift the arithmetic path to its generating b by its conditional law given the path, even if b is not determined by the path. For the pinned mesh path the normalized Fourier coefficient at h is

$$2\pi i h\, \widehat{Q_q - tW}_0(h) = e(h/(2q))\,\mathrm{sinc}(\pi h/q) W_h, qquad \mathrm{sinc}(x) = \sin(x)/x,$$

by integrating the piecewise constant derivative. Thus it vanishes at the inactive slots. In every coupling, on the two typical events the pinned squared distance by Parseval is at least c/H. Pinning costs at most a factor two and K_q differs from the mesh by $O(q^{-1/2})$. This gives the matching $(\log p)^{-1/2}$ lower uniformly in depth.

For $d = 1$, couple the gated independent law at b_0 to the circle law by deleting the coefficients with $h \equiv b_0 \pmod p$. Let h_0 be the representative of least absolute value, nonzero since b_0 is a unit. Its contribution to the difference has norm $O(1/|h_0|)$. The other terms, before pinning, are

$$e(h_0 t) \sum_{k \ne 0} Z_{h_0+kp} \frac{e(kpt)}{2\pi i(h_0 + kp)}.$$

In a dyadic k band of size U, the sum inside the modulation is a trigonometric polynomial in pt with pointwise subgaussian width $O(1/(p\sqrt{U}))$. A Bernstein grid of $O(U)$ points in that variable gives an L^s uniform norm of at most $C_s\sqrt{\log(2U)}/(p\sqrt{U})$. Summing and pinning costs $O_{R,s}(1/p)$. Averaging the sth powers over uniformly distributed nonzero b_0 therefore gives a coupling of the gated mixture to $\nu_{\mathbb{T}}$ of cost $O_{R,s}(G_s(p))$. The conditioned comparison already proved, followed by this coupling, gives the second upper bound in Theorem A.1.

Conversely, for every arithmetic path at residue b_0, its pinned mesh Fourier coefficient vanishes at h_0, while $\nu_{\mathbb{T}}$ has a normalized coefficient of modulus one there almost surely, simultaneously for all indices. Lift the coupling to the generating residue as above. In any coupling the uniform distance is at least $c/|h_0| - O(q^{-1/2})$, because a Fourier coefficient is bounded by the pinned uniform norm. Restrict to $1 \le |h_0| \le c_0 p$ with a sufficiently small fixed c_0 to absorb the mesh error, uniformly in $n \ge 2$. Then averaging $1/|h_0|$ gives $\gg (\log p)/p$ for W_1 and the first few h_0 already give $\gg_s p^{-1/s}$ for $s > 1$. For the finitely many small primes for which the stated restriction is empty, use $|h_0| = 1$ for sufficiently large n and absorb the finitely many shallow cases into the constants by the finite-support bound. Together with that bound this proves the asserted lower bounds.

.1 Correlated observation coupling

Let m be a large integer, as chosen in the two arithmetic regimes above. Put $H_0 = \lfloor m^{1/4} \rfloor$. Independently round the real and imaginary parts x of coefficients $|h| \le H_0$ to $2R\,\mathrm{Bin}(\lceil m^2 \rceil, (x + R)/(2R))/\lceil m^2 \rceil - R$. The L^s path error is $O_{R,s}(\log m/m)$ by the centered binomial subgaussian bounds and the absolute multiplier sum. There are at most $\exp(O(H_0 \log m))$ low output vectors, and the conditional probability polynomial of each has character norm at most the same exponential order by the separate-degree estimate.

Take blocks $H = 2^j H_0$, $j \ge 1$, through the first $H_* \ge m \log m$. In such a block let $y = A_H x$ be the $I_H = 16H$ real and imaginary parts of $\sqrt{H} \sum b_h x_h e(ht)$ on the $8H$ equally spaced mesh samples. Discrete Parseval gives $\|A_H\|_{2\to 2} \ll 1$ and $\|y\|_2^2 \ll_R H$ on the cube. In the independent model, the sample vector has the directional subgaussian bound $\mathbb{E} \exp(t \cdot y) \le \exp(C_R \|t\|_2^2)$. Apply Hoeffding to each independent complex coefficient as one real two-dimensional block. For every subset S of samples, Gaussian integration then gives

$$\mathbb{E} \exp\left(\lambda \sum_{i \in S} y_i^2\right) \le D^{|S|} \tag{32}$$

for constants $\lambda > 0, D$ depending at most on R.

For reconstruction, choose a smooth Fourier multiplier equal to one on $[-H, H]$, zero outside $[-2H, 2H]$, with kernel V_H. Discrete orthogonality on the $8H$ grid recovers the block as the sum of sampled values against weights $(8H)^{-1} V_H(t - j/(8H))$. The kernel bound $|V_H(u)| \ll H/(1+H^2 \operatorname{dist}(u, \mathbb{Z})^2)$, obtained by two summations by parts, makes the absolute sum of the weights uniformly bounded. Thus if normalized samples are changed by at most τ in each component, the sup error after reconstruction and pinning is $O(\tau/\sqrt{H})$.

Use scale $\tau_H = (H/m)^{3/4}$ for $H \le m$, and $\tau_H = A(H/m)^{1/4}$ for $H > m$, with a large constant A depending at most on R. The former blocks will be compared on all slots, and the latter on sparse subsets. Both parts of the sum $\sum_H \tau_H/\sqrt{H}$ are geometric and give $O(m^{-1/2})$.

Lemma .2. *Fix a positive constant J_0. For I real coordinates and noise scale $\tau > 0$, put $T = C_0 \max(1, \tau^4)$ with $C_0 \ge 1$. There is a probability density on $\mathbb{R}^I$, with noise vector η, with these properties:*

1. $\mathbb{P}(\max_i |\eta_i| > \tau/8) = O(I^{-J_0})$.
2. *For each slot set S, the probability that all $|\eta_i| > \tau/8$ in S is at most $2(Ce^{-c\sqrt{T}})^{2|S|}$.*
3. *The characteristic function of any D-coordinate marginal, $D \ge 1$, has support in a ball of radius $R(D) = CTI^{1/4}D^{1/4}/\tau$. Every mixed first derivative, once in each of r distinct coordinates, is bounded by $(C\tau)^r$ up to an absolute constant. These derivatives vanish outside the same ball.*

Proof. Take a real even smooth function f supported in $[-1/8, 1/8]$, with $\|f\|_2 = 1$ and $|\widehat{f}(u)| + |\widehat{f}'(u)| \le Ce^{-c\sqrt{|u|}}$. For example, use the normalized density of the sum of independent uniforms on $[-c_*/j^2, c_*/j^2]$, $j \ge 1$, with c_* small. Its sinc product has at least $c\sqrt{|u|}$ factors bounded by $1/2$ for large $|u|$, and a differentiated term loses at most one such factor and contributes $O(j^{-2})$. This also proves smoothness and the requisite derivative decay.

Take a smooth even cutoff χ decreasing in absolute argument, equal to one on $[-1, 1]$ and zero outside $[-2, 2]$. Form a partition with $\phi_0(u) = \chi(u/T)$ and $\phi_j(u) = \chi(u/(2^j T)) - \chi(u/(2^{j-1}T))$ for $j \ge 1$. Its terms are nonnegative, sum to one, and $\sum_j |\phi_j'| \ll 1$. The upper radii are $U_j = 2^{j+1}T$, and if $\phi_j(u) > 0$ for $j \ge 1$, then $|u| > U_j/4$.

Retain band configurations $(j_1, \dots, j_I)$ with $\sum_i U_{j_i}^4 \le C_1 I T^4$. Let w be the sum of their product partition functions, so $0 \le w \le 1$. We use the squared-amplitude device for positive Fourier smoothing as in [19, Section 7]. The restriction on the joint amplitude will control the Fourier support of every selected marginal. Define $\widehat{P}(u) = w(u) \prod_i \widehat{f}(u_i)$. Take the density $|P|^2 / \|P\|_2^2$ and dilate by τ in each coordinate. As usual, the Fourier normalization can be taken unitary, with absolute constants adjusted for the characteristic-function convention.

For a rejected configuration the bounds $U_{j_i} \le 4 \max(T, |u_i|)$ imply $\sum_i |u_i|^4 > C_2 I T^4$, with C_2 as large as needed by choosing C_1. Under the iid $|\widehat{f}|^2$ probability law, $|u_i|^4$ has all moments. For each fixed integer k, the $2k$th moment of the sum of centered copies is $O_k(I^k)$ by independence and expansion, since singleton factors vanish. Thus the rejected frequency mass is $O(I^{-2J_0})$ on taking k large enough. Choosing C_2 above four times the scalar fourth moment also bounds this mass by $1/4$ for every I by Markov's inequality. It follows that the amplitude P is $O(I^{-J_0})$ from the product amplitude in L^2, and hence the normalized densities are within the same order in total variation by Cauchy–Schwarz. This proves the first property; the uniform mass bound gives $\|P\|_2^2 \ge 3/4$.

For the joint bound, write f_j for the inverse transform of $\phi_j \widehat{f}$. Fix the band labels on S. The corresponding partial sum of the amplitude factors as the selected f_j times an amplitude in the complementary coordinates with L^2 norm at most one, because its partition multiplier lies between zero and one. On $|x| > 1/8$ we have $f_0 = -\sum_{j \ge 1} f_j$ in L^2. Moreover $\sum_{j \ge 1} \|f_j\|_2 \ll e^{-c\sqrt{T}}$ by the lower band radii and the decay of $\widehat{f}$. Summing the restricted L^2 bounds over selected band labels gives $(Ce^{-c\sqrt{T}})^{|S|}$. Squaring and normalizing proves the second property.

The support of $\widehat{P}$ has fourth power sum at most $C_1 I T^4$. Its projection to D coordinates has Euclidean radius at most $CTI^{1/4}D^{1/4}$ by Cauchy–Schwarz. The Fourier transform of $|P|^2$ is an autocorrelation of $\widehat{P}$, so the support of the selected characteristic function is in the corresponding difference set, after accounting for dilation. Finally a mixed derivative of $\widehat{P}$ on distinct slots is pointwise bounded by the product of $|\widehat{f}'| + C|\widehat{f}|$ there and $|\widehat{f}|$ elsewhere, by the partition derivative estimate. Its L^2 norm is $O(C^r)$. By Plancherel and Cauchy–Schwarz this bounds the product absolute moment of the corresponding η_i by $O((C\tau)^r)$. These moments justify differentiating the characteristic function, with the stated support. □

Draw the noises independently across blocks and independently of every input. In the dense blocks observe the noisy samples $y_i + \eta_i$. In each sparse block replace a noisy sample by zero whenever its absolute value is at most $\tau_H/2$. Clip the final output in either case to a centered interval of radius $C\sqrt{H} + \tau_H$, containing the deterministic range of y_i and its small perturbation. On the good event from Lemma .2(1) the sample error is $O(\tau_H)$.

.2 Comparison on selected marginals

Select every dense slot. In each sparse block H select D_H slots, with D_H bounded below by zero and above by the cutoff specified in the next section. We first compare the noisy real vectors before thresholding or clipping, together with the rounded low output. Deterministic maps can only decrease total variation. Let g be the difference density of the selected noisy coordinates for one fixed low output label. Its Fourier transform factors as the product of the noise marginal transforms times the difference of the two input characteristic expectations, weighted by the conditional low label polynomial. The smoothing ensures a density, and the selected characteristic function is compactly supported.

For a Fourier test vector t_H, the character norm of $\exp(it_H \cdot A_H x)$ is at most $\exp(C\sqrt{H}\,\|t_H\|_2)$ by its absolutely convergent power series and $\left\|A_H^T t_H\right\|_1 \ll \sqrt{H}\,\|t_H\|_2$. A mixed derivative in r_H

distinct variables in block H inserts a product of r_H sample coordinates. Its cube bound is at most $(CH/r_H)^{r_H/2}$ by the energy estimate and the arithmetic–geometric mean inequality. Its character norm is bounded by $\exp(Cr_H \log(eCH/r_H))$ by the total-degree bound in Section 2, since the block uses $O(H)$ Fourier inputs. The bound is one for $r_H = 0$. Taking C sufficiently large, one may bound this exponent using D_H in place of r_H for all $r_H \le D_H \le I_H$. The low label costs $\exp(o(m))$. With the derivative bounds for the noise in Lemma .2, these estimates control all mixed first derivatives of $\widehat{g}$ through the arithmetic comparison of Section 4 and the product rule.

To bound the L^1 norm of g, assign $s_i = 1 + \tau_H$ in block H. Multiply and divide by the square root of $\prod_i (1 + (x_i/s_i)^2)$, then use Cauchy–Schwarz and Plancherel. If $D = \sum_H D_H$, this gives

$$\|g\|_1 \le \Big(\prod_i \pi s_i\Big)^{1/2} \left[C^D \sum_{U \subseteq \{1,\dots,D\}} \Big(\prod_{i \in U} s_i^{-2}\Big) \|\partial_U \widehat{g}\|_2^2 \right]^{1/2}.$$

The weights absorb the powers of τ_H from the noise derivatives. The support is a product of balls. A ball of radius R in D_H dimensions has volume at most $(CR/\sqrt{D_H})^{D_H}$. Combining these facts with the arithmetic comparison gives, after summing over the low labels, a total variation upper bound of

$$\begin{aligned} &\varepsilon_m \exp\big(o(m) + C \textstyle\sum_H E_H(D_H)\big), \qquad \text{where} \\ &E_H(D_H) = \sqrt{H} R_H(D_H) + D_H \log(eCH/D_H) \\ &\qquad\qquad + D_H \log\left(\frac{C(1+\tau_H) T_H}{\tau_H} (I_H/D_H)^{1/4} \right), \end{aligned} \tag{33}$$

Here $\varepsilon_m = p^{-1/2} + e^{-A_0 m}$ for the large-prime argument and $\varepsilon_m = e^{-Lm} + e^{-A_0 m}$ in the conditioned depth argument. The Taylor tolerance A_0 and depth saving L can be chosen larger than the fixed amplification constants. Implicit constants in the observation comparison may depend on R. Also, $T_H = C_0 \max(1, \tau_H^4)$ and R_H is as in Lemma .2. Terms with $D_H = 0$ are omitted. The exponent includes the 2^D derivative sum and the polynomial factor for the number of Fourier inputs.

For dense blocks $H \le m$ take $D_H = I_H$. The contribution is $O(H/\tau_H + H[1 + \log(1/\tau_H)])$, summable to $O(m)$ because $H/\tau_H = m^{3/4} H^{1/4}$. In a sparse block $H > m$ define

$$K_H = \min\big(I_H, \lceil \max(m^{1/2}, C_2 I_H e^{-c_2 \tau_H^2}) \rceil\big)$$

with c_2 sufficiently small and C_2 sufficiently large. We allow $D_H \le \min(I_H, K_H + 1)$. Since $H \le 2m \log m$ up to an absolute factor and $\tau_H^2 \ll \sqrt{\log m}$, the $m^{1/2}$ term is eventually smaller than the exponential term throughout this range. Thus $x = D_H/I_H \le C e^{-c_2 \tau_H^2}$. The first contribution of this block to (33) is $O(H(T_H/\tau_H) x^{1/4})$. The others are bounded by $O(Hx[1 + \log(1/x) + \log T_H + \log(1 + 1/\tau_H)])$. Together they are $O(H(1+\tau_H)^C e^{-c_3 \tau_H^2})$. Since $\tau_H^2 = A^2 \sqrt{H/m}$, their dyadic sum is $O(m)$. In particular, every selected marginal compares uniformly in total variation at cost $\varepsilon_m \exp(Cm)$.

.3 Sparse comparison and transport

Under the independent model, a specified set of D_H sparse slots is simultaneously nonzero after thresholding only if in each of those slots either $|y_i| \ge \tau_H/4$ or $|\eta_i| > \tau_H/8$. Split the slots between these alternatives. For the first use (32) and Markov's inequality, and for the second use Lemma .2(2). Independence of noise and input gives the product bound, summed over at most 2^{D_H} allocations. With A large the joint activity probability is at most $\exp(-c_4 \tau_H^2 D_H)$. These bounds multiply across blocks, since the independent Fourier coordinates and the noises are independent across blocks.

The argument uses truncated inclusion–exclusion with a factorial error bound, as in read-once Boolean tests [9, Section 3], here for arbitrary bounded tests of multiple observed values. For a realized sparse block output with N_H nonzero entries, expand its point mass as a product of zero point masses plus coordinate differences. Retain terms involving at most K_H differences. If $N_H \le K_H$ this is exact. Otherwise the signed variation error is at most

$$B_H = 2 \cdot 3^{K_H+1} \binom{N_H}{K_H+1},$$

by the binomial bounds $\binom{N_H}{j} \le \binom{N_H}{K_H+1}\binom{K_H+1}{j}$ for $j \le K_H$ and the factor 2^j for signed differences. Take $B_H = 0$ if the cap keeps all coordinates. For a product across sparse blocks, telescoping bounds the aggregate variation error by $\prod_H (1 + B_H) - 1$.

Expand this last product to estimate its expectation. Each factorial count is a sum over sets of prescribed active slots. Under the independent law the term for a chosen block is bounded by

$$\left(\frac{C I_H e^{-c_4 \tau_H^2}}{K_H + 1} \right)^{K_H+1}.$$

Take $c_2 < c_4$ and C_2 large in the definition of K_H. This is at most $\exp(-cK_H)$, so the whole expected error is smaller than any fixed power of m^{-1}. Under arithmetic input compare each selected joint activation by the selected-marginal bound (33). The sum of all weights and subset counts is $\exp(O(m))$, because $\sum_H K_H \log(eI_H/K_H) = O(m)$ by the same sparse budget estimate. Hence the arithmetic expected truncation error is bounded by $O_{J_0}(m^{-J_0}) + \varepsilon_m \exp(Cm)$.

For any test of the full observations bounded by one, the truncated expectation is a sum of tests on all the dense outputs, the low rounded vector, and at most K_H selected sparse coordinates in each block. Evaluate the other sparse outputs at zero in each term of the signed expansion. The same subset count and signed amplification is $\exp(O(m))$, so (33) compares these truncated expectations with error $\varepsilon_m \exp(Cm)$. Combining the two truncation errors shows that the full genuine observation laws differ in total variation by

$$O_{J_0}(m^{-J_0}) + \varepsilon_m \exp(Cm)$$

uniformly in the permitted inverse coefficient for the respective arithmetic comparison.

In the large-prime argument take $m = \lfloor a \log p \rfloor$ with a small. In the conditioned depth argument, take $m = \lfloor aJ \rfloor$ and choose L in the arithmetic estimate large. In either case the arithmetic part of the total variation bound is exponentially small in m. A maximal coupling of the observations gives a common reconstructed path except on an event of the bounded probability. All clipped paths have sup norm $O(\log m)$ by stable interpolation. On the compact noise event the reconstruction error for either input is $O(\sum_H \tau_H/\sqrt{H}) = O(m^{-1/2})$ in addition to binomial rounding. The complementary probability is $O(\sum_H I_H^{-J_0}) = O(m^{-J_0/4})$. Choosing J_0 sufficiently large and using deterministic $O(\log m)$ bounds on reconstructed and truncated true paths makes the exceptional cost negligible in each fixed L^s. Together with the tail estimates in the main text this gives $W_s = O_{R,s}(m^{-1/2})$.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, proof drafting and internal checks. No numerical certificates or data are required.

A Fixed-prime transport for observations of Kloosterman paths

For a fixed prime and increasing prime-power depth, we obtain Wasserstein rates for any fixed number of bounded linear observations of a residue-conditioned tame hyper-Kloosterman path. The rate is $O(n^{-1/2})$ for arbitrary bounded observations, improves with Fourier decay, and becomes exponential for observations involving finitely many frequencies. We also determine the support entropy compatible with an abstract form of the phase estimates used in full-path transport. Laws obeying those phase estimates and all fixed-order block bounds can still have uniform-path Wasserstein distance at least a constant times $(n \log n)^{-1/4}$.

Keywords: hyper-Kloosterman paths, prime powers, Wasserstein distance, phase moments.
MSC 2020: 11L05, 11L07, 60G17, 60B05.

A.1 Paths and observations

Fix $R \geq 2$ and an odd prime $p > R$. Put $q = p^n$, $n \geq 2$, and

$$\mathrm{Kl}_j(c; q) = q^{-(j-1)/2} \sum_{x_1 \cdots x_j = c} \exp(2\pi i(x_1 + \cdots + x_j)/q),$$

with unit variables. For a unit a and a nonzero residue $b_0 \pmod p$, take b uniform on $b \equiv b_0 \pmod p$. Let $K_q(t)$ accumulate the jumps $q^{-1/2}\,\mathrm{Kl}_{R-1}((ax)^{-1}; q)e_q(bx)$ at equal times in increasing unit order x. We use $e_q(v) = e^{2\pi i v/q}$ and $e(t) = e^{2\pi i t}$. Both a and the residue may be chosen anew at each depth.

The model $K_*(t)$ has independent coefficients X_h in the expansion

$$tX_0 + \sum_{h \neq 0} X_h \frac{e(ht) - 1}{2\pi i h}.$$

Here is the coefficient definition. Write $d = \gcd(R, p-1)$, and use a torus with characters Z_j corresponding to ζ_d^j, $0 \leq j < d$, in the power lattice $\mathbb{Z}[\zeta_d]$. All Z_j are sampled simultaneously at one independent Haar point per index. A shift h is active if $(b_0 - h)/a$ is a nonzero Rth power modulo p, and inactive coefficients are zero. At an active shift choose the same root branch s_h throughout the residue ball, with $s_h^R = (b-h)/a$ and specified root modulo p. At even n set $X_h = \sum_j Z_j$. At odd n use the weights from the local Hessian:

$$X_h = \varepsilon_p^{R-1} \chi(2^{-(R-1)} R) \sum_j \chi(s_h \zeta^j)^{R-1} Z_j,$$

where $\zeta \in \mathbb{Z}_p$ is a primitive dth root, χ denotes the quadratic character mod p, and $\varepsilon_p = 1$ for $p \equiv 1 \pmod 4$ and i otherwise. The weights are constant on the residue ball. Replacing a root branch permutes characters and preserves each model law. Active variables are centered, bounded by R, and have second absolute moment d.

At rank two the qualitative fixed-prime path limit is the theorem of Milićević and Zhang [22, Theorem 1]. Their complete-sum analysis in [22, §4] addresses possible coincidences of shifted stationary points. Milićević, Robinson and Shupe [29, Theorem 1.1] give sharper fixed-order rank-two product estimates with a finer alignment measure among the shifts. The tame all-rank local comparison and whole-path $O(n^{-1/4})$ upper bound at fixed p appear in Section A. The following result concerns the finer resolution of a fixed number of observations, as well as a precise limitation of the available individual phase bounds.

For $h \neq 0$ put $\phi_h(t) = (e(ht) - 1)/(2\pi i h)$ and put $\phi_0(t) = t$. Let $\mathcal{L} = (L_1, \dots, L_D)$ be a fixed number of real-linear maps $C([0,1]; \mathbb{C}) \to \mathbb{R}$ with uniformly bounded operator norms. For each h define

$$w_h = \max_{1 \le j \le D} \sup_{|z| \le 1} |L_j(z\phi_h)|.$$

In particular $w_h \ll (1 + |h|)^{-1}$ for arbitrary bounded observations. The hypotheses allow $\mathcal{L}$ to vary with n if the number and bounds stay fixed.

Theorem A.1. *For fixed D and $1 \le s < \infty$, the W_s distance between $\mathcal{L}(K_q)$ and $\mathcal{L}(K_*)$ in $\mathbb{R}^D$ satisfies the following bounds, uniformly in a, b_0 and depth parity.*

1. *If $w_h \ll (1 + |h|)^{-b}$ with $b \ge 1$, the distance is $O(n^{-b/2})$.*
2. *If $w_h \ll \exp(-c_0|h|^\alpha)$ for some $c_0, \alpha > 0$, it is $O(\exp(-cn^{\alpha/(\alpha+2)}))$.*
3. *If $w_h = 0$ outside a fixed finite set, it is $O(\exp(-cn))$.*

Constants may depend on the displayed weight bounds, p, R, D, s, and the probe norm bound. In particular, any fixed finite collection of complex path evaluations has a joint $O(n^{-1/2})$ rate.

For example, if $L(f)$ integrates a smooth real or imaginary part of the pinned path $f(t) - tf(1)$ against a periodic test function, its nonconstant Fourier weights have an extra factor $1/|h|$. The theorem then gives rates of any polynomial power for fixed smooth tests, and an exponential rate for a fixed trigonometric polynomial test.

A.2 Arithmetic comparison and a Fourier tail

Write $W_h = \mathrm{Kl}_R((b-h)/a; q)$ for the complete coefficient, with zero at a singular shift. For each active shift its stationary evaluation is the same weighted trace as in the model but on phases $e_q(Rs_h\zeta^j)$, using the reduction of the p-adic roots modulo q. This follows by expanding the rank-$(R-1)$ summation at its stationary points, where every free unit equals s_h. They exist precisely when the coefficient is active. The Hessian is $s_h^{-1}(I + \mathbf{1}\mathbf{1}^\mathsf{T})$ with determinant $Rs_h^{-(R-1)}$, a unit. Half-depth digit orthogonality gives the phase value and, at odd depth, the remaining quadratic Gauss factor displayed above. This is the tame stationary formula of Smith and Dąbrowski–Fisher [24, 31]. At an inactive shift give the unused latent coordinates independent Haar laws in both ensembles.

A torus character at one active shift is represented by an algebraic integer $\gamma = \sum_{j=0}^{\varphi(d)-1} m_j \zeta_d^j$, where m_j are integers. In the selected p-adic embedding its valuation is $O_{p,R}(\log(2 + \max_j |m_j|))$ if it is nonzero, by bounding its algebraic norm. Its arithmetic phase is $e_q(R\gamma s_h)$.

Lemma A.2. *Suppose a nontrivial product character has t occupied active shifts, distinct integers with $|h| \le H$, and all slot frequencies bounded by M. There are $c, C > 0$ depending only on p, R such that, if*

$$t \log(2 + H + t) + \log(2 + M) \le cn,$$

then its arithmetic mean has modulus at most $Ct \exp(-cn/t)$. Its independent-model mean is zero. When $t = 1$ the arithmetic mean is zero under the same hypothesis.

Proof. Only the active case needs consideration. Sum the upper half of the digits of b first. Taylor expansion there is linear, so contributions can survive only when $D(b) = \sum_i \gamma_i s_{h_i}(b)^{-(R-1)}$ is divisible by $p^{\lfloor n/2 \rfloor}$, since a is a unit. To bound this sublevel set, work on residue subballs $b = b_1 + p^{k_1} y$,

where $k_1 = 2+\lceil \log_p(Rt+R)\rceil$. Let $y_i = (b_1-h_i)^{-1}$ and $M_j = \sum_i \gamma_i s_{h_i} (b_1)^{-(R-1)} y_i^j$. The polynomial $\prod_{i'\neq i}(X - y_{i'})$ isolates the ith coefficient from the first t moments. In particular

$$b_2 = \min_{0\le j<t} v_p(M_j) \le \min_i v_p(\gamma_i) + (t-1)\lceil \log_p(2H+1)\rceil.$$

All later moments have valuation at least b_2 by the recurrence with roots y_i. The jth coefficient of $D(b_1 + p^{k_1}y)$ is $p^{k_1 j}\binom{-(R-1)/R}{j} M_j$. These binomials are p-integral. Their valuations are at most $\log_p(Rj+R)$, since at each power of p the numerator progression and $j!$ differ in their count of multiples by at most one. Thus after division by the smallest coefficient power of p, the last unit coefficient has degree $j_0 < t$, and the normalization uses at most $b_2 + k_1 t$ digits. Under the stated hypothesis the required remaining divisibility is p^V with $V \ge n/4$ for large n.

For $j_0 = 0$ there are no solutions. For $j_0 > 0$, $j_0 + 1$ solutions in distinct classes modulo $p^{\lceil V/j_0\rceil}$ would give a divided difference of order j_0 divisible by p. Its value modulo p is instead the unit coefficient, as the coefficients at greater degrees are divisible by p and monomial divided differences are integral. Thus the solutions occupy at most j_0 of these classes. This gives the bound on each subball and then on the conditioned class. The argument is also the local comparison used in Section A. □

Lemma A.3. *For fixed $s < \infty$ and fixed integer $2u \ge \max(4, s)$, replacement of either path by $\sum_{|h|\le N} Y_h \phi_h$, where $Y_h = W_h$ or X_h, has L^s uniform-norm error*

$$O\left(N^{-1/2+1/(2u)} + (1+\log q) q^{-c_u} + N/q\right) \qquad (1 \le N \le (q-1)/2).$$

Proof. For arithmetic, insert zeros at nonunits and use mesh times j/q. The original equal-unit times differ at the corresponding units by $O(1/q)$, and every jump is $O_R(q^{-1/2})$ by tame stationary phase (or rank one). The resulting uniform error is $O_R(q^{-1/2})$. Fourier inversion on the mesh and summing its first j increments give coefficients $W_h/[q(1-e(-h/q))]$ for the terms $e(ht)-1$ at nonzero h, and tW_0 at zero. The coefficients follow from scaling the completed sum by $b-h$ when it is a unit. When $b-h$ is a nonunit the last digit in the summation over x averages an affine phase with unit slope and gives zero. For retention below N the multiplier error and the difference between the polynomial and its mesh interpolation are $O_R(N/q)$.

For a dyadic block of size U the nonzero multipliers are $O(1/U)$. The integrated $2u$th moment before pinning only contains tuples with the frequency balance $h_1+\cdots+h_u = h_{u+1}+\cdots+h_{2u}$. Each balanced tuple has coefficient size $O_u(U^{-2u})$. There are $O_u(U^{2u-1})$ such tuples. For $P_0 = p^{\lfloor n/C_u\rfloor}$ with C_u large, balanced tuples with two distinct shifts colliding modulo P_0 number $O_u(U^{2u-1}/P_0)$. In fact, if any distinct collision is possible the block has length comparable to or exceeding P_0, and one marks the pair, counts it with $O(U^2/P_0)$ possibilities, then solves the balance using a third index. In the remaining tuples Lemma A.2 applies with $\log H$ in its proof replaced by $\log P_0$ for bounding the pair valuations. Their arithmetic moments agree with the independent-model moments up to $O_u(q^{-c_u})$ per tuple. The fixed-degree trace expansion has bounded coefficient norm, and its constant characters agree exactly. The independent pointwise block moment is $O_u(U^{-u})$ by the moment bound for independent centered bounded variables. Taking the supremum of the $2u$th power costs an $O_u(U)$ factor by Cauchy–Schwarz on the coefficients of the uth power. Thus the L^s uniform block error bound in either model is $O_u(U^{-1/2+1/(2u)} + q^{-c_u})$, with no arithmetic term needed for the independent model. Pinning each block at zero costs at most two, and the dyadic sum gives the result. □

A.3 Characteristic functions of the observations

We first treat polynomial weights in Theorem A.1. Choose $N = \lfloor n^B \rfloor$ with B sufficiently large that Lemma A.3 allows both paths to be truncated with error $O(n^{-b/2})$. For a real test vector $v \in \mathbb{R}^D$, its projection on the retained observations is a sum $\sum_h Z_h$ of real phase terms $Z_h = \sum_i v_i L_i(Y_h \phi_h)$. A term has absolute value and Fourier norm $O(\|v\| w_h)$ on its slot torus. Choose $H_0 = \lfloor a\sqrt{n} \rfloor$, with a small $a > 0$ fixed below, and work up to test frequencies $\|v\| \le T = H_0^b$. Constants in all budgets may depend on the fixed dimension and weight and norm bounds.

The Fourier norm of e^{iZ_h} on the slot torus is at most $C(1 + \|v\| w_h)^C$. Indeed weighted Cauchy–Schwarz bounds the Fourier norm by a Sobolev norm with fixed order exceeding half the torus dimension, and differentiation bounds that norm by the indicated power. Each Z_h is a trigonometric polynomial of fixed degree in the latent variables. On any fixed complex strip the bound for the absolute value of e^{iZ_h} is $e^{C\|v\| w_h}$. Shifting the Fourier contour gives a uniform error $e^{-cL + C\|v\| w_h}$ for cutting each slot to Fourier height L, with adjusted c, C. For $|h| \le H_0$ use this cut with a sufficiently large polynomial L in n. The low exponential is approximated uniformly with error smaller than every fixed inverse power of n, and its total Fourier norm is $\exp(O(H_0))$. To see the latter, sum $\log(1 + (H_0/(1 + |h|))^b)$ over the low slots and use $\sum_{1 \le h \le H_0} \log(H_0/h) = O(H_0)$. This bound includes the zero slot.

For each high dyadic block with $H \le |h| < 2H$ put $Z_H = \sum_h Z_h$. Its Fourier norm is bounded by $L_H = CTH^{1-b}$. In the independent model its variance proxy in the centered bounded-variable moment bound is $\sigma_H^2 = CT^2 H^{1-2b}$. Take

$$d_H = \left\lceil K(\sigma_H^2 + \log n) \right\rceil, \qquad P_H = \sum_{j=0}^{d_H - 1} \frac{(iZ_H)^j}{j!}.$$

The constant K can be chosen to give any prescribed inverse power of n in the errors. For real Z_H the Taylor remainder has absolute value at most $|Z_H|^{d_H}/d_H!$. The independent L^2 remainder is at most $(C\sigma_H/\sqrt{d_H})^{d_H}$ by the subgaussian moment bound and the factorial. It is $O(n^{-A})$ for arbitrarily large fixed A by choice of K.

For the arithmetic L^2 remainder compare the expectation of the polynomial $|Z_H|^{2d_H}/(d_H!)^2$. Its Fourier norm is at most $\exp(O(d_H \log(e + L_H/d_H)))$, and the same bound holds for the square of the modulus of P_H. This follows directly from the powers of L_H and the factorials. Each monomial in any product of these high block squares and the low approximation uses at most $O(H_0 + (\log n)^2)$ occupied slots, has polynomial Fourier height, and the total Fourier norm is at most $\exp(O(H_0 + (\log n)^3))$. In detail, $\sum_H d_H = O_K(H_0 + (\log n)^2)$, and for H comparable to $2^j H_0$ the ratio L_H/σ_H^2 is $O(2^{bj})$. Therefore the terms contributed by σ_H^2 sum to $O(H_0)$, while each term contributed by the extra $K \log n$ is $O((\log n)^2)$ over $O(\log n)$ blocks. Squaring products, or using a single extra $2d_H$th power, changes only the fixed constants.

As $|h| \le N$, the hypotheses of Lemma A.2 hold throughout these products. Its exponential saving $\exp(-cn/(C_K H_0 + O_K((\log n)^2)))$ pays the Fourier norms if a is small enough, even including the polynomial factor in the phase bound. The total expectation errors are smaller than every fixed inverse power of n. Thus the arithmetic Taylor remainders as well are $O(n^{-A})$ in L^2, by choosing the independent bound with a larger exponent if necessary. Also the products of any initial collection of high-block P_H have L^2 norms $O(1)$ in both laws. For the independent model this follows by independence of the blocks and $\|P_H\|_2 \le 1 + O(n^{-A})$. For the arithmetic law compare the moments of the product with those of the independent model by the same Fourier budgets.

Join the block exponentials in order. In the telescoping identity replacing them by P_H, each single remainder is multiplied by an earlier product of P_H and by later phase factors of absolute

value one. Cauchy–Schwarz bounds the expected error by the product of the L^2 norms just obtained. The low exponential is replaced by its uniform approximation. Then the full product comparison gives, for any prescribed A,

$$\sup_{\|v\|\le T} |\widehat{\mu}_N(v) - \widehat{\nu}_N(v)| = O(n^{-A}), \tag{34}$$

where μ_N, ν_N are the truncated observation laws and a can depend on A.

To translate (34), choose a probability density on $\mathbb{R}^D$ which is Schwartz and whose Fourier transform is supported in the unit ball. For example square the modulus of the inverse Fourier transform of a smooth compactly supported function, and normalize and dilate. Add an independent random vector with this density scaled by $1/T$. This changes each law by $O_s(1/T)$ in W_s. Both truncated laws are deterministically supported in a box of radius $O(1+\log N)$. The smoothed densities differ pointwise by $O(T^D n^{-A})$ by Fourier inversion, so their difference mass on a fixed enlargement of the box is $O((1+\log N)^D T^D n^{-A})$. Each tail outside the enlargement is smaller than any fixed power of T^{-1}. Clipping the smoothed laws to the enlarged box does not increase total variation, and can be coupled with the unsmoothed laws with $O_s(1/T)$ cost by coordinatewise contraction. The total variation term on the clipped box contributes at most its diameter times the $1/s$ power of the discrepancy by a common-part coupling. Choosing A large proves $W_s = O(1/T) = O(n^{-b/2})$, including the initial path truncations.

For stretched-exponential weights take $T = \exp(an^{\alpha/(\alpha+2)})$ and $N = \lfloor T^B \rfloor$ with B large. Lemma A.3 gives truncation error $O(T^{-1})$. On the frequency ball $\|v\| \le T$, the contribution of $|h| > H = C_A(\log T)^{1/\alpha}$ to the characteristic phase is deterministically $O(T^{-A})$, for any predetermined A. This follows by summing Tw_h. On the remaining slots use the exact exponential Fourier bounds with Fourier cutoff T^B for a sufficiently large B. The norm of the product is $\exp(O(H\log T))$, with slot heights $T^{O(1)}$, support $O(H)$, and negligible uniform cutoff error. The hypotheses of Lemma A.2 hold since $H\log(2+H) + \log T = o(n)$. Its error $\exp(-cn/H)$ pays the Fourier norm for small enough a and leaves $O(T^{-A})$. The same smoothing argument proves the second assertion. For finitely supported weights the same proof uses a fixed H and $T = \exp(an)$. First choose a truncation $N \asymp \exp(b_1 n)$ with $0 < b_1 < \log p$ in Lemma A.3, and then make a small enough to absorb the truncation and all Fourier costs. This proves the third assertion of Theorem A.1.

A.4 What the individual phase bounds imply

Let $H = \lfloor n^B \rfloor$ for a fixed $B > 1$, and let $B_1 > 0$ be fixed. Consider laws on H phase slots, each a torus of a fixed dimension $k_0 \ge 1$. For a nontrivial character ξ, let $t(\xi)$ be its number of occupied slots and let its height be its maximum absolute integer coordinate. The following condition abstracts the individual phase saving:

$$|\widehat{\mu}(\xi)| \le \exp(-cn/t(\xi)) \qquad (\text{height of } \xi \le n^{B_1},\ \xi \ne 0). \tag{35}$$

The arithmetic estimate includes a polynomial factor and imposes a restriction when growing supports and shifts are too large. Requiring (35) on the whole window strengthens the assumptions used in the obstruction below.

Theorem A.4. *For fixed $c > 0$, the minimum finite support size of a probability law satisfying* (35) *is*

$$\exp\left(\Theta(\sqrt{n\log n})\right).$$

For residue-conditioned path models as above, one can replace the latent phases in the active shifts with $|h| \le \lfloor n^B \rfloor$ by such a finite-window law, and take independent Haar phases outside that window.

For each fixed order and sufficiently large n, the resulting laws have the same integrated polynomial block moments as the independent paths, and each nontrivial character of bounded height outside the window has zero mean if it occupies any exterior slot. Yet their uniform-path W_1 distance from the independent model is at least $c'(n \log n)^{-1/4}$.

Proof. First consider the lower bound on support size. Take $r = \lfloor a\sqrt{n/\log n} \rfloor$ with $a > 0$ small. Use the $M = \binom{H}{r}$ functions obtained by multiplying one fixed first-coordinate character over each r-element subset of slots. The Gram matrix in $L^2(\mu)$ has diagonal one. Off-diagonal entries are bounded by $\delta = \exp(-cn/(2r))$ by (35), since the ratio characters have support at most $2r$ and height one. The rank bound by its trace and squared Hilbert–Schmidt norm gives

$$\text{rank} \geq \frac{M}{1 + (M-1)\delta^2} \geq M/2$$

for a sufficiently small. The dimension of $L^2(\mu)$ is at most the support size, and $\log M \asymp \sqrt{n \log n}$.

For the upper bound, choose a prime L exceeding H and $2n^{B_1}$ with $\log L = O(\log n)$, and use distinct evaluation points $x_1, \ldots, x_H \in \mathbb{F}_L$. Evaluate a uniform polynomial of degree less than r at the x_i, using one independent polynomial for each of the k_0 torus coordinates. This is the Reed–Solomon evaluation construction [30]. The resulting linear array is uniform on every set of r or fewer slots by polynomial interpolation. Embed every residue $u \in \mathbb{F}_L$ as $e^{2\pi iu/L}$ in a phase coordinate. Each nontrivial character of height at most n^{B_1} on at most r slots therefore has mean zero. On other slots, the mean of a character on the finite code is either zero or one. Translate the code by $J = \lceil \exp(Cn/r) \rceil$ independently uniform ambient shifts and mix the resulting cosets with equal weight. For a character trivial on the code, its mean across these shifts is an average of J independent mean-zero unit complex variables. The actual character is nontrivial on the ambient group because of the height cutoff. Hoeffding's inequality for real and imaginary parts bounds the probability its absolute mean exceeds $e^{-cn/r}$ by $4\exp(-c_2 J e^{-2cn/r})$. The number of characters has logarithm $O(H \log n)$, so choosing C sufficiently large lets a finite union bound give all the desired inequalities at once. The support is at most $JL^{k_0 r}$, giving the other entropy bound.

For the path application take a window of order n^B containing zero and every shift $|h| \leq n^B$. Restrict the finite construction to the active slots there, whose number is comparable to n^B for fixed p. Take independent Haar elsewhere, including in unused inactive slots. Every fixed-order polynomial in trace coefficients expands into characters of bounded height occupying a bounded number of slots. Consequently all such expectations agree with those in the independent model for sufficiently large n. This includes the balanced block moments used in Lemma A.3. The independence of the exterior coefficients gives a continuous path via the same uniformly convergent Fourier tails as for the model.

For $U \leq H/4$ project a path f to the Fourier coordinates with $U \leq h < 2U$ of $f(t) - (1-t)f(0) - tf(1)$. This projection is 2-Lipschitz from the uniform norm to the Euclidean norm by Parseval. In these coordinates, the independent path has entries $X_h/(2\pi ih)$ and the finite-window construction has a law with at most $JL^{k_0 r}$ atoms. Every Euclidean ball of radius ϵ has independent-model mass $\leq e^{-c_3 U}$ if U is comparable to a sufficiently small constant times ϵ^{-2}. Indeed there is a fixed positive density of active shifts. In each of those slots $\sup_z \mathbb{E}\exp(-|X_h - z|^2)$ is bounded strictly below one by boundedness and nondegeneracy, uniformly among the finitely many labels. The squared Euclidean ball condition implies $\sum |X_h - z_h|^2 \ll U^2\epsilon^2$ over the active shifts, so exponential Markov gives the bound. Take U a sufficiently large constant times $\sqrt{n \log n}$ and ϵ a sufficiently small multiple of $U^{-1/2}$. The union of all projected balls centered at the atom set then covers at most half the model law. Any coupling has uniform-path cost at least $\epsilon/2$ with probability at least $1/2$, proving the lower bound. □

The example can also be made finitely supported on the entire retained path, with at most p^{n-1} atoms. Fix $c < (\log p)/4$ in (35). Keep the finite polynomial window and retain exterior slots only for $|h| \le (q-1)/2$. For the exterior latent vectors, mix $\lceil e^{4cn} \rceil$ independent product-Haar samples. With positive probability all nonzero exterior characters of support at most n and height at most n^{B_1} have means at most e^{-cn}. Their number has logarithm $O(n^2)$, and Hoeffding gives this by a union bound. Sample this exterior ensemble independently of the finite polynomial-window law. The resulting retained path has at most $\exp(4cn + O(\sqrt{n \log n})) \le p^{n-1}$ atoms. Its relevant exterior phase bounds, including all fixed-order moments up to $O(e^{-c'n})$ errors per monomial, follow by independence between the two regions. The terminal L^s block bounds follow from the same balanced-tuple and supremum argument of Lemma A.3, with these errors in place of arithmetic errors. Low-band support is unaltered, so the uniform-path lower bound still holds. These constructions are abstract laws of Fourier coefficients. The fixed-prime arithmetic full-path distance remains between its finite-support lower bound of order $n^{-1/2}$ and the upper bound $O(n^{-1/4})$ of Section A.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, independent verification, and drafting. The results require no datasets.

B The finite-support approximation constant of the Kloosterman limit

We determine the leading constant in optimal finite-support approximation of the Kloosterman limit in the $L^2[0,1]$ path metric. The formula is an integral of the scalar Sato–Tate rate-distortion objective. It holds for every fixed finite distortion moment and every fixed excess-distortion probability, and supplies a leading-constant lower bound for arithmetic Wasserstein approximation. More generally, we prove a strong converse and matching construction for bounded independent coordinates with regularly varying squared weights and finitely many coefficient laws.

Keywords: functional quantization, rate-distortion, Kloosterman paths, independent random series.
MSC 2020: 60B05, 60G17, 94A29, 11L05.

B.1 Exact quantization for weighted coordinates

For a random element Z of a real or complex Hilbert space define

$$e_{N,s}(Z) = \inf_{|\Gamma| \le N} \left(\mathbb{E} \min_{y \in \Gamma} \|Z - y\|^s \right)^{1/s}, \qquad 1 \le s < \infty.$$

This equals the infimum of W_s from its law to laws supported on at most N points, by nearest-point encoding. The squared-error threshold $d_N(\epsilon; Z)$ is the infimum of $t \ge 0$ for which some such Γ satisfies $\mathbb{P}\{\min_{y \in \Gamma} \|Z - y\|^2 > t\} \le \epsilon$.

Let $(X_j)_{j \ge 1}$ be independent in one fixed finite-dimensional Euclidean space, with laws in a finite list $P_1, \dots, P_v$ of compactly supported measures. Write $\iota(j)$ for the law index and assume each index has an asymptotic density θ_i. For a positive function ϕ regularly varying with index $-b$,

$b > 1$, let $\lambda_j > 0$ satisfy $\lambda_j \sim \phi(j)$. Consider the Hilbert-valued vector $Y = (\sqrt{\lambda_j} X_j)_j$ with the usual square-sum norm. For $X \sim P_i$ and $a > 0$ set

$$F_i(a) = \inf_{X'|X} \{I(X; X') + a\mathbb{E}\,|X - X'|^2\}, \qquad J = \sum_i \theta_i \int_0^\infty F_i(u^{-b})\, du. \tag{36}$$

Mutual information is measured with natural logarithms.

Theorem B.1. *Suppose at least one nonconstant law has positive density, so $J > 0$. Put*

$$C_b = \frac{(b-1)^{b-1}}{b^b} J^b.$$

Then for every fixed $s \in [1, \infty)$ and every fixed $\epsilon \in (0, 1)$,

$$e_{N,s}(Y)^2 \sim C_b (\log N)\phi(\log N), \qquad d_N(\epsilon; Y) \sim C_b (\log N)\phi(\log N).$$

For any fixed multiplicative decrease of this squared distortion level, the mass covered by any N balls is at most $e^{-c \log N}$ for all large N.

The Gaussian case of the sharp functional quantization formula with regularly varying eigenvalues is due to Luschgy and Pagès [32, Theorem 2.2]. The formula here recovers their eigenvalue-based coefficient from the Gaussian scalar objective in (36). Luschgy and Pagès also obtained rates for broader classes of processes from mean regularity [33]. The argument here uses the full scalar rate-distortion objectives of bounded non-Gaussian coordinate laws. The fixed-size source-coding ingredients are classical. We include their application to growing weighted bands to obtain the exact coefficient and the strong converse. The converse uses the same tilted information inequality that underlies finite-blocklength lossy compression [34].

B.2 Scalar objective and coding bounds

We give details for one compactly supported law P, writing $F = F_i$. Restrict reproductions to the compact convex hull of the support by metric projection. If Q is a probability measure on that hull, put

$$g_{a,Q}(x) = -\log \int e^{-a|x-y|^2}\, dQ(y).$$

The decomposition of conditional relative entropy against Q into mutual information plus the relative entropy of the output law against Q gives

$$F(a) = \min_Q \mathbb{E}\, g_{a,Q}(X). \tag{37}$$

Indeed, after minimizing over Q, one may minimize separately at each x. The Gibbs conditional distribution with density $e^{g_{a,Q}(x) - a|x-y|^2}$ relative to Q minimizes the conditional entropy plus the distortion penalty. The minimum over Q is attained by compactness and positivity and continuity of the kernel. For a minimizing Q the Gibbs conditional has output marginal Q, again by the relative entropy decomposition. Denote its distortion and mutual information by $D(a)$ and $I(a)$. Then $F(a) = I(a) + aD(a)$ and $D(a) = F'(a)$ at each differentiability point because the optimizing channel supplies a supporting affine function of a.

For a minimizing Q, differentiate (37) in the direction of a point mass at y. Nonnegativity of the directional derivative gives

$$\mathbb{E} \exp(g_{a,Q}(X) - a\,|X - y|^2) \le 1 \tag{38}$$

for every y in the compact hull. Also $g_{a,Q}$ is bounded between zero and a times the squared diameter. With the same channel, the information density $g_{a,Q}(X) - a|X - X'|^2$ and the distortion are bounded.

For later use, for each fixed a and $\eta > 0$ there are deterministic codes for l independent copies of X with at most $\exp(l(I(a)+\eta)+O(1))$ reproductions and failure probability e^{-cl} at average squared distortion $D(a) + \eta$. To see this, draw a codebook from Q^l. Under the joint Gibbs distribution the event of distortion at most $D(a) + \eta$ and information density sum at most $l(I(a) + \eta/2)$ has probability at least $1-e^{-cl}$, by the bounded-variable concentration inequality. For all source vectors except a set of probability at most $2e^{-cl}$ its conditional probability is at least $1/2$. On each such vector its probability under Q^l is at least $\frac{1}{2}e^{-l(I(a)+\eta/2)}$ by the change of measure. The sampled codebook fails on those vectors with doubly exponentially small probability. Averaging shows the existence of a deterministic codebook with the claimed error probability. Every reproduction lies in the hull, so its distortion is also deterministically bounded.

Finally, F is increasing and concave as an infimum of affine objectives with nonnegative slopes. A constant reconstruction gives $F(a) \ll a$, and a deterministic grid in the hull of mesh $(1 + a)^{-1/2}$ gives $F(a) \ll 1 + \log(1 + a)$. Thus the integrals defining J are finite. If $F(a) = 0$ at $a > 0$, its optimizing channel would have both zero mutual information and zero distortion, which forces the law to be constant. For each law integration by parts gives

$$\int_0^\infty u^{-b} F_i'(u^{-b})\, du = \frac{1}{b} \int_0^\infty F_i(u^{-b})\, du, \tag{39}$$

$$\int_0^\infty [F_i(u^{-b}) - u^{-b} F_i'(u^{-b})]\, du = \left(1 - \frac{1}{b}\right) \int_0^\infty F_i(u^{-b})\, du. \tag{40}$$

Concavity gives local absolute continuity, and the boundary terms vanish by the stated growth bounds. These integrands are Riemann integrable on closed subintervals of $(0, \infty)$ when the derivative is given either one-sided convention. The derivative is monotone with at most countably many discontinuities, and representative choices may be made at differentiability points.

B.3 Weighted bands and the strong converse

Write $R = \log N$ and $B = (1 - 1/b)J$. On a central band $\epsilon L \le j \le TL$, with fixed $0 < \epsilon < T < \infty$, put $a_j = \lambda_j/\phi(L)$. The uniform convergence theorem for regularly varying functions [35] gives $a_j = (j/L)^{-b} + o(1)$ uniformly on this band. The density assumption on the types then gives

$$\frac{1}{L} \sum_{\epsilon L \le j \le TL} F_{\iota(j)}(a_j) \longrightarrow \sum_i \theta_i \int_\epsilon^T F_i(u^{-b})\, du. \tag{41}$$

Indeed the density assumption gives the limiting count of each type on each fixed interval of j/L, and the functions are continuous. Subsequent band approximations also apply to the Riemann integrable functions in (40).

For the lower bound take $L = R/B$. For each central slot put $G_j = g_{a_j,Q_j}(X_j)$ where Q_j minimizes (37) for that law and penalty. These variables are independent and uniformly bounded on the central band. Given $\eta > 0$, choose ϵ small and T large so that the right side of (41) exceeds $J - \eta/4$. Then

$$\sum_{\epsilon L \le j \le TL} G_j \ge (J - \eta/2)L$$

except on an event of probability at most e^{-cL}. Fix any reproduction in the Hilbert space, and write its coordinates as $\sqrt{\lambda_j} y_j$. For the lower bound one may project y_j to the convex hull of the

jth law, since this decreases every distortion. Independence and (38) show that the mean of

$$\exp\left(\sum_{\epsilon L\le j\le TL} G_j - \frac{1}{\phi(L)}\sum_{\epsilon L\le j\le TL}\lambda_j|X_j-y_j|^2\right)$$

is at most one. Thus for an arbitrary codebook of size N the probability of squared error at most $(J/b-\eta)L\phi(L)$, on the displayed high-probability event, is at most

$$N\exp((J/b-\eta)L-(J-\eta/2)L) = e^{-\eta L/2}.$$

This proves the strong converse because $R=BL$ and $L\phi(L)/(R\phi(R))\to B^{b-1}$. It gives the moment and fixed-quantile lower bounds.

For the upper bound use $L=R/(B+\gamma)$ with small $\gamma>0$. Choose the cutoffs ϵ, T as above and subdivide $[\epsilon,T]$ into finitely many short intervals. Within one interval, group its indices of each type and apply the deterministic iid codes proved after (38). Take the penalty at a differentiability point within the interval and use a sufficiently small fixed coding slack η. For each positive-density type the group length at a nonempty interval is asymptotic to a positive constant times L. Thus failure probabilities tend to zero exponentially in L. Each zero-density type uses only $o(L)$ indices on the central band and can use a fixed reconstruction on each of them, at $o(L\phi(L))$ total distortion. The weights normalized by $\phi(L)$ change by an arbitrarily small amount within each interval when the mesh is small. They are uniformly bounded there, as are distortions. The resulting central rate and high-probability distortion are bounded by

$$\log N_{\text{central}} \le L(B+\gamma/3), \qquad D_{\text{central}} \le (J/b+\gamma/3)L\phi(L)$$

for all large L, by (40), with the subdivision and slack chosen after the cutoffs.

The low indices $j<\epsilon L$ are encoded individually at mesh comparable to $\min(1,\sqrt{\phi(L)/\lambda_j})$ in the bounded coefficient space. Their total squared error is $O(\epsilon L\phi(L))$ pointwise. The total code rate is $O(\epsilon L(1+\log(1/\epsilon)))+o(L)$, since the logarithm of the code size per index is $O(1+\log^+(\lambda_j/\phi(L)))$. Potter's bound reduces the sum to a constant times $\sum_{j<\epsilon L}(1+\log(L/j))$ outside finitely many initial indices, and those initial indices cost $O(\log L)=o(L)$. For $j>TL$ use a constant vector at each slot. Independent of the source, this has squared error $O(\sum_{j>TL}\lambda_j)$, which is $O(T^{1-b}L\phi(L))$ for fixed T in the limit by regular variation, or by dyadic summation with a Potter bound. Choose cutoffs so that low and high contributions fit in the remaining slack. The product codebook has size at most N and squared error at most $(J/b+\gamma)L\phi(L)$ except on an event of probability tending to zero. The worst squared error everywhere is $O_{\epsilon,T}(L\phi(L))$, since the weighted central band has this total weight and all reproductions are bounded. This converts the upper bound to every fixed finite moment and fixed quantile. Letting $\gamma\downarrow 0$ gives $C_b=(J/b)B^{b-1}$ and completes the proof.

B.4 Kloosterman paths and arithmetic lower bounds

For the prime Kloosterman limit of Kowalski and Sawin [36], the path is

$$K(t)=tX_0+\sum_{h\ne 0}X_h\frac{e^{2\pi iht}-1}{2\pi ih},$$

where the X_h are independent with the Sato–Tate density $\pi^{-1}\sqrt{1-x^2/4}$ on $[-2,2]$. Write F_{ST} for its objective in (36) and put

$$J_{\mathrm{ST}}=\int_0^\infty F_{\mathrm{ST}}(u^{-2})\,du, \qquad C_{\mathrm{ST}}=\frac{J_{\mathrm{ST}}^2}{4\pi^2}. \tag{42}$$

Corollary B.2. *In the $L^2[0,1]$ metric of the Kloosterman limit,*

$$e_{N,s}(K)^2 \sim \frac{C_{\mathrm{ST}}}{\log N}, \qquad d_N(\epsilon; K) \sim \frac{C_{\mathrm{ST}}}{\log N}$$

for all fixed finite $s \geq 1$ and $\epsilon \in (0,1)$. If μ_p is any prime Kloosterman path law supported on at most $(p-1)^2$ paths, then

$$\liminf_{p\to\infty} \sqrt{\log p}\, W_s(\mu_p, \mathrm{law}(K)) \geq \sqrt{C_{\mathrm{ST}}/2}$$

in either the L^2 or uniform path metric. With support at most $p-1$, the lower bound is $\sqrt{C_{\mathrm{ST}}}$.

Proof. First use only the nonzero-frequency periodic series $Y(t) = \sum_{h\neq 0} X_h e^{2\pi i h t}/(2\pi i h)$. By Parseval its Hilbert weights are $1/(2\pi h)^2$ on both signs of h. Ordered by absolute frequency, they are asymptotic to $1/(\pi^2 j^2)$, giving precisely (42). The subspace of complex L^2 with those rotated real Fourier coefficients is a real Hilbert subspace, so allowing arbitrary complex codewords does not improve distortion by orthogonal projection.

To restore K, its difference from Y is the finite-rank random path $tX_0 - Y(0)$. The variable $Y(0)$ is well defined by the bounded independent centered series with square-summable weights and has subgaussian tails, obtained by applying the bounded-variable exponential inequality to partial sums and passing to the limit. A fixed-dimensional grid for the difference path, on a box of radius $O(\log N)$ and with mesh $(\log N)^{-1}$, needs only a polynomial in $\log N$ many points and has error $o((\log N)^{-1/2})$ in every fixed finite moment and with probability tending to one. Taking sums or differences of codebooks compares K and Y in both directions at this extra rate of $O(\log\log N)$. The triangle inequality gives the moment statements. For the quantiles use the high-probability upper construction and the vanishing covered-mass lower bound for Y, together with the same high-probability finite-dimensional grid approximation. This reasoning does not use independence of the finite-rank term from Y.

Finally, any law supported on at most N paths has distance at least $e_{N,s}(K)$. The uniform metric dominates the L^2 metric on these paths, and the two given support budgets yield the claimed lower constants. □

The same formula applies to fixed-prime residue-conditioned hyper-Kloosterman limits in Section A. Here p and tame rank R_0 are fixed, $d = \gcd(R_0, p-1)$, and the active Fourier slots have density $\alpha = (p-1)/(dp)$. Apart from deterministic rotations, all active coefficients in a fixed parity have the same bounded trace law P_{act}. Coordinates of the other type are identically zero. Write $J_{\mathrm{act}} = \int_0^\infty F_{\mathrm{act}}(u^{-2})\, du$. Replacing J_{ST} by αJ_{act} in (42) gives the exact L^2 quantization constant. Rotations at individual coordinates are harmless, since they act by a Hilbert isometry. The finite-rank argument applies as before by boundedness and centering. In particular the arithmetic path law in one residue class, with at most p^{n-1} paths, satisfies the lower bound $\liminf_{n\to\infty} \sqrt{n}\, W_s \geq \alpha J_{\mathrm{act}}/(2\pi\sqrt{\log p})$ along the chosen parity, in either path metric. A finite mixture with identical constants has the same asymptotic, by lower bounding the distortion in each component and using a union of component codebooks for the upper bound.

B.5 Bounds and evaluation of the Sato–Tate constant

For a centered scalar Gaussian with variance $V > 0$ the objective is

$$F_{G,V}(a) = \begin{cases} aV, & a \leq 1/(2V), \\ (1 + \log(2aV))/2, & a > 1/(2V). \end{cases}$$

Its integral at exponent b is $(2V)^{1/b}b^2/[2(b-1)]$. Substitution in the coefficient of Theorem B.1 gives $Vb^b/[2^{b-1}(b-1)]$, the Gaussian coefficient of [32] in this normalization.

For a non-Gaussian bounded scalar law of variance V, one has $F(a) < F_{G,V}(a)$ for $a > 1/(2V)$. In fact let $D = 1/(2a) < V$ and take $X' = (1 - D/V)X + Z$, where Z is an independent centered Gaussian of variance $D(1 - D/V)$. The distortion is D. By the Gaussian maximum-entropy inequality, $I(X; X') < \frac{1}{2}\log(V/D)$, with strictness because X' is non-Gaussian. The latter follows by dividing characteristic functions, since the Gaussian factor never vanishes. An arbitrary mean can be removed by translation. This gives the strict upper bound for the objective on the stated interval.

A lower bound follows from differential entropy. For a scalar source with entropy power $P = (2\pi e)^{-1}e^{2h(X)}$, the conditional Gaussian entropy bound gives $I(X; X') \geq \max(0, \frac{1}{2}\log(P/D))$ when the distortion is D. Thus $F(a) \geq F_{G,P}(a)$. For a bounded source with finite differential entropy the inequality is strict at $a > 1/(2P)$. Equality would force the minimizing channel to have $D = 1/(2a) > 0$, and equality in the entropy bound would force a Gaussian reconstruction error independent of the reconstruction. Its sum with the reconstruction cannot be bounded. One can obtain the same equality condition by comparing each conditional source law to the Gaussian with mean X' and variance D in relative entropy.

In this example the Sato–Tate law has $V = 1$, differential entropy $h(X) = \log(2\pi) - 1/2$ and entropy power $P = 2\pi/e^2$. The entropy calculation follows by writing $X = 2U$, since U^2 has beta distribution with parameters $1/2, 3/2$ and $\mathbb{E}\log(1 - U^2) = 1 - 2\log 2$.

The low-penalty branch is exactly $F_{\mathrm{ST}}(a) = a$ for $a \leq 1/2$. Indeed its even moments are the Catalan numbers, which are at most the corresponding standard Gaussian moments since $\mathbb{E}X^{2j}/(2j-1)!! = 2^j/(j+1)! \leq 1$. Hence X is subgaussian with variance proxy one. The entropy variational bound relative to the source then gives $I(X; X') \geq \frac{1}{2}\mathbb{E}|\mathbb{E}[X \mid X']|^2$. Meanwhile $D \geq 1 - \mathbb{E}|\mathbb{E}[X \mid X']|^2$. The objective is at least a, attained by the constant reconstruction. For $a > 1/2$ a small binary observation already gives $F_{\mathrm{ST}}(a) < a$: choose a sign with conditional mean tX and reconstruct by t times that sign. The distortion is $1 - t^2$ and its mutual information is $t^2/2 + O(t^4)$.

Consequently

$$\frac{(1 + \sqrt{2\pi/e^2})^2}{2\pi^2} < C_{\mathrm{ST}} < \frac{2}{\pi^2}.$$

For the lower bound on J_{ST} use $F_{G,P}(u^{-2})$ up to $u = \sqrt{2P}$, with strictness on that interval. On $[\sqrt{2P}, \sqrt{2}]$ use $F_{\mathrm{ST}}(u^{-2}) \geq 1/2$ by monotonicity. Beyond $\sqrt{2}$ use $F_{\mathrm{ST}}(u^{-2}) = u^{-2}$. The integral of these lower bounds is $\sqrt{2}(1 + \sqrt{P})$. For the upper bound integrate the variance-one Gaussian objective, with strictness on its high-penalty branch.

Formula (37) also gives a way to evaluate the constant with controlled errors. For any proposed reproduction measure Q on the compact hull let $H_Q(y) = \mathbb{E}e^{g_{a,Q}(X) - a|X-y|^2}$. Jensen's inequality applied to ratios of the Gibbs integrals gives

$$\mathbb{E}g_{a,Q}(X) - \log \sup_y H_Q(y) \leq F(a) \leq \mathbb{E}g_{a,Q}(X).$$

Together with the bounded-hull Lipschitz estimates for the logarithm of each Gibbs integral, this gives convergent upper and lower approximations by finite grids. The integrable endpoint bounds for $F(u^{-b})$ control the omitted integration tails. Thus (42) depends on the entire coefficient law rather than only its variance.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, internal independent checks and drafting. The proofs use no datasets.

C Random codebooks and empirical transport for the Kloosterman limit

We determine the exact leading constants for approximating the Kloosterman limit by independent sample paths, both by nearest-neighbor encoding and by equal-weight Wasserstein transport. The constants differ strictly from each other and from unrestricted finite-support approximation. We obtain them from three scalar rate-distortion objectives on regularly varying independent Fourier coordinates. A constant-composition coupling preserves both path laws and controls the pinning term in the ordinary L^2 metric. The equal-mass converse also strengthens the leading lower constant for actual prime Kloosterman path laws, including in the uniform metric.

Keywords: Kloosterman paths, empirical Wasserstein distance, random codebooks, rate-distortion.
MSC 2020: 11L05, 60B05, 60G17, 94A29.

C.1 Three approximation constants

Let $(X_j)_{j\ge1}$ be independent bounded vectors in a fixed Euclidean space, with laws from a finite list $P_1,\dots,P_v$. Suppose the indices of type i have asymptotic density θ_i. Let $\lambda_j>0$ satisfy $\lambda_j\sim\phi(j)$ for a positive regularly varying function ϕ of index $-b$, with $b>1$. Write $Y=(\sqrt{\lambda_j}X_j)_j$ and $\mu=\mathrm{law}(Y)$ on the square-sum Hilbert space. At least one nonconstant type is assumed to have positive density. Let $Y_1,Y_2,\dots$ be independent samples with law μ, and $\mu_N=N^{-1}\sum_{i=1}^N\delta_{Y_i}$. When measuring the error of this particular codebook, use a fresh independent Y with law μ. Its nearest-neighbor distortion moment is

$$D_{N,s}=\left(\int\min_{1\le i\le N}\|y-Y_i\|^s\,d\mu(y)\right)^{1/s}.$$

In contrast $W_s(\mu,\mu_N)$ forces every sampled point to receive mass $1/N$, with multiplicity.

For one coefficient law P and a penalty $a>0$, write

$$\begin{aligned}F_P(a)&=\inf_{\pi_X=P}[I_\pi(X;X')+a\mathbb{E}_\pi|X-X'|^2],\\ G_P(a)&=\mathbb{E}_{X\sim P}\left[-\log\int e^{-a|X-y|^2}\,dP(y)\right],\\ B_P(a)&=\inf_{\pi\in\Pi(P,P)}[I_\pi(X;X')+a\mathbb{E}_\pi|X-X'|^2].\end{aligned}$$

For $U=F,G,B$ define

$$J_U=\sum_i\theta_i\int_0^\infty U_{P_i}(u^{-b})\,du,\qquad C_b(J)=\frac{(b-1)^{b-1}}{b^b}J^b.$$

All logarithms use the natural base. Unrestricted approximation has leading squared constant $C_b(J_F)$ over $(\log N)\phi(\log N)$ by Section B. The two sampling constants are as follows.

Theorem C.1. *Under the preceding assumptions,*

$$D_{N,s}^2\sim C_b(J_G)(\log N)\phi(\log N),\qquad W_s(\mu,\mu_N)^2\sim C_b(J_B)(\log N)\phi(\log N)$$

for every fixed finite $s \geq 1$, almost surely for a nested sequence of sampled paths. They also hold for $(\mathbb{E}D_{N,s}^s)^{2/s}$ and $(\mathbb{E}W_s(\mu,\mu_N)^s)^{2/s}$, respectively. Every fixed nearest-neighbor excess-distortion probability has the same squared threshold as the first formula. For μ-almost every center,

$$-\log \mu\{z : \|Y-z\|^2 \leq (J_G/b)L\phi(L)\} \sim (1-1/b)J_G L. \tag{43}$$

In particular $J_F < J_G \leq J_B$.

The distinction between conditional nearest neighbors and prescribed reproduction distributions is classical in source coding. Dembo and Kontoyiannis establish a generalized asymptotic equipartition property for stationary sources and mismatched codebooks [37, Theorems 1–2]. Dereich and Lifshits link randomly centered Gaussian small balls to random quantizers [38, Theorem 5.1]. Output-constrained coding and soft covering appear in work of Saldi, Linder and Yüksel [39] and Cuff [40]. Recent work of He and Yuan [41] obtains empirical order bounds for Itô paths under the supremum norm. Here the weights vary across the growing band of active coordinates. Both the exact coefficient and the distinction between the three objectives enter the path problem.

C.2 Scalar tilts and weighted bands

For G, let $g_a(x) = -\log \int e^{-a|x-y|^2} dP(y)$. The conditional Gibbs kernel $\exp(g_a(x) - a|x-y|^2)dP(y)$ has reference conditional law P for reproduction. Differentiation under the integral gives expected squared distortion $G_P'(a)$, and the averaged log likelihood relative to that reference is $G_P(a) - aG_P'(a)$. No output constraint is imposed on this Gibbs kernel.

For B, let $K_a(x,y) = e^{-a|x-y|^2}$ on the compact support. There is a positive continuous scaling u with $u(x)\int K_a(x,y)u(y)dP(y) = 1$. For completeness normalize functions v by $\int v\, dP = 1$. The map

$$v \longmapsto \frac{(K_a v)^{-1}}{\int (K_a v)^{-1}\, dP}$$

is continuous with relatively compact image in the uniform norm, since K_a has a positive lower bound and is uniformly continuous. It maps the convex set of normalized continuous functions between $\inf K_a$ and $(\inf K_a)^{-1}$ into itself. Schauder's theorem gives a fixed point, and a scalar rescaling gives u. Thus, writing $f_a = \log u$, the density $e^{f_a(x)+f_a(y)}K_a(x,y)$ relative to $P \otimes P$ has both marginals P. The relative entropy identity against this kernel proves that it minimizes the objective and gives

$$B_P(a) = 2\mathbb{E}f_a(X), \qquad \int e^{f_a(x)+f_a(y)-a|x-y|^2}\, dP(x) = 1. \tag{44}$$

At differentiability points the expected distortion is $B_P'(a)$, by the supporting affine function property, and its information cost is $B_P(a) - aB_P'(a)$. On compact penalty bands the potentials are uniformly bounded and Lipschitz on the support. The Lipschitz assertion follows by writing $f_a = -\log(K_a e^{f_a})$ and comparing logarithms of the integrands. The information density under the Gibbs pair is bounded on such bands.

Both G_P and B_P are increasing and concave as functions of a. To see this for G_P, it is the minimum, without constraining the second marginal, of relative entropy against $P \otimes P$ plus penalized distortion, with source marginal P. Both objectives are $O(a)$ at zero by the independent coupling. To bound them at large a, partition the support into M cells of diameter $O(a^{-1/2})$, with M polynomial in $1+a$. Draw both coordinates independently inside the same cell whose index is chosen with its P-probability. This has both marginals P, distortion $O(a^{-1})$ and information at most $\log M$. Thus they grow at most logarithmically. At every a, $F_P(a) \leq G_P(a) \leq B_P(a)$. At

zero $G'_P(0) = 2\mathbb{E}|X - \mathbb{E}X|^2$, while a constant reconstruction gives $F_P(a) \le a\mathbb{E}|X - \mathbb{E}X|^2$. This proves the indicated strict inequality of integrals when a nonconstant type has positive density.

For $U = G$ or B integration by parts yields

$$\sum_i \theta_i \int_0^\infty u^{-b} U'_{P_i}(u^{-b})\, du = J_U/b, \qquad \sum_i \theta_i \int_0^\infty [U_{P_i}(u^{-b}) - u^{-b} U'_{P_i}(u^{-b})]\, du = (1 - 1/b) J_U. \quad (45)$$

The boundary terms vanish by the bounds just given. Central bands $\epsilon L \le j \le TL$ are chosen with fixed ϵ, T and $a_j = \lambda_j/\phi(L)$. Uniform regular variation and the type-density assumptions give Riemann sums for these integrals on the band. For derivatives of B one may either use one-sided derivatives or approximate by small bands and choose a differentiability point in each, since the derivative is monotone. We use the latter construction for upper bounds.

For $j < \epsilon L$ use common cells of mesh $\min(1, \sqrt{\phi(L)/\lambda_j})$. It preserves both marginals and has pointwise distortion $O(\epsilon L \phi(L))$. Its information is a sum of cell surprisals $-\log P(C(X_j))$. If a partition has M cells, then for $0 < t < 1$ the exponential moment of that surprisal is $\sum_C P(C)^{1-t} \le M^t$. Potter's bound gives an aggregate log cell count of order $\epsilon L(1+\log(1/\epsilon))$ plus $o(L)$. To see this, sum $1+\log(L/j)$ for $j < \epsilon L$, omitting finitely many initial indices that cost $O(\log L)$. Consequently the information is at most this deterministic bound plus any fixed positive multiple of L, except with probability e^{-cL}. For $j > TL$ use independent reproduction. Its squared error is bounded pointwise by $O(T^{1-b} L\phi(L))$ in the limit, by regular variation. The same constructions may be used for conditional small-ball probabilities, viewing the low coupling as a change of sampling law relative to the original independent reproductions.

C.3 Conditional small balls and nearest codewords

Fix a small $\eta > 0$, take ϵ small and T large, and set $A_G = (1 - 1/b)J_G$. For a typical center the sum of $g_{a_j}(X_j)$ on the central band exceeds $(J_G - \eta)L$, outside a center event of probability e^{-cL}, by bounded-variable concentration and the central Riemann sum. Conditional on that center, Chernoff's inequality on the central band bounds the small-ball probability at squared radius $(J_G/b)L\phi(L)$ by $e^{-(A_G - \eta)L}$. The discarded coordinates only increase squared distance.

For the reverse bound, tilt each central reproduction by its conditional Gibbs kernel, independently given the center. On the low band require the same cell, as above, and on the high band leave reproduction independent. Under the joint law of center and modified reproduction, the central squared distortion divided by $\phi(L)$ has mean L times the truncated first integral in (45), plus $o(L)$. The central log likelihood has mean L times the truncated second integral, plus $o(L)$. Both sums are independent across slots and have bounded central summands, so the two quantities stay below these means plus ηL except with exponential probability. The low information has the exponential bound above, while its distortion and the high distortion are pointwise small. Thus, after shrinking the cutoffs as needed, with exponentially high probability the joint tilted pair has squared distance at most $(J_G/b + \eta)L\phi(L)$ and log likelihood at most $(A_G + \eta)L$. Constants in the slack may be absorbed by starting with a smaller η. An application of Markov's inequality to conditional failure probability shows that, outside an exponentially small set of centers, the good event has conditional tilted probability at least $1/2$. The reverse change of measure gives conditional original small-ball probability at least $\frac{1}{2} e^{-(A_G + \eta)L}$ at that slightly larger radius.

These bounds give (43). For the lower bound on the small-ball probability at the exact radius, apply the tilted estimate with an index $(1+c_\eta)L$ for a fixed small $c_\eta > 0$, taking η small enough that regular variation makes the shifted squared radius smaller. Then let both discrepancies tend to zero. Exponential exceptional sets are summable along integer indices. The squared-radius scale $L\phi(L)$

has negative index, so it has an asymptotically equivalent decreasing version, and consecutive scale ratios tend to one. Monotonicity of balls extends the resulting almost-sure equivalence to all L.

Conditional on the center, the probability that none of N samples lies in a ball of mass v is $(1-v)^N$. Take L asymptotic to $(\log N)/A_G$. Applying the two bounds just obtained at indices $(1 \pm c)L$ for arbitrary fixed $c > 0$ yields convergence of the normalized squared nearest distance to $(J_G/b)A_G^{b-1} = C_b(J_G)$ in probability. The probabilities of failure at any fixed relative deviation are exponentially small in $\log N$; at the upper threshold they combine the exceptional centers and the doubly exponentially small sample failure on the other centers. The Hilbert norm here is bounded deterministically. Thus all fixed finite moments converge after normalization. The same reasoning gives almost-sure statements for a nested sample sequence: conditional failure probabilities or conditional distortion errors converge along $N_j = \lfloor e^j \rfloor$ by Markov's inequality and Borel–Cantelli, and nearest-neighbor distances decrease with N. The scale changes by a factor tending to one between adjacent N_j. This also proves the fixed nearest-neighbor quantile assertion.

C.4 The equal-mass empirical law

Write $A_B = (1 - 1/b)J_B$. For the lower bound take $L = (\log N)/A_B$ and a central band as above. Put $S_X = \sum_{\epsilon L \le j \le TL} f_{a_j}(X_j)$ and $M_L = \mathbb{E}S_X$. Then $2M_L/L$ tends to the central integral of B. The source mass outside $|S_X - M_L| \le \eta L$ is exponentially small. The expected empirical mass outside the identical condition on the sample vectors is also exponentially small. For every sample vector y satisfying that condition, (44) and independence give

$$\mathbb{E}\exp\left(S_X + S_y - \sum_{\epsilon L \le j \le TL} a_j |X_j - y_j|^2\right) = 1.$$

Therefore the source probability that it is typical and has central squared error at most $(J_B/b - c\eta)L\phi(L)$ to any typical sample point is exponentially small, by the union bound over $N = \exp(A_B L)$ and cutoffs with $2M_L \ge (J_B - \eta)L$. Here c may be any fixed constant larger than three. In every coupling the empirical mass of the remaining sample vectors and the source exceptional mass have the bounds just described. This gives a lower bound on every fixed finite Wasserstein moment, in expectation over the codebook and with high probability.

For the upper bound use $L = (\log N)/(A_B + \gamma)$ with fixed small $\gamma > 0$. Construct a joint law of source and reproduction with both marginals μ. Use the same-cell and independent constructions for the low and high indices. On each interval of a fine partition of the central band use the independent symmetric Gibbs channel at a representative penalty where B is differentiable, separately for each coefficient type. Types of zero density can use independent reproduction there at $o(L\phi(L))$ error. As in (45), the central expectations of normalized squared distortion and information per L approximate the truncated integrals. The central summands are bounded, and the low information has the exponential bound already proved. Choose cutoffs and subdivisions so the information density $i(X,Y) = \log(d\pi/d(\mu \otimes \mu))$ is at most $\log N - c\gamma L$ except with probability $e^{-c'L}$. The central squared distortion, divided by $L\phi(L)$, is at most $J_B/b + O(\gamma)$ with probability tending to one by concentration, including the head and tail. Its worst value is bounded by a constant depending on the cutoffs. It therefore has the same stated upper bound in every fixed finite distortion moment.

We give the soft-covering step explicitly. For N independent draws $Y_i \sim \mu$, let $\widetilde{\mu}_N = N^{-1}\sum_i P_{X|Y=Y_i}$, where $P_{X|Y}$ is the conditional source law of the constructed pair. This couples to the empirical measure by the sample index. For $t > 0$, its expected total variation distance to μ is at most

$$\mathbb{P}_\pi\{i(X,Y) > \log N - t\} + \tfrac{1}{2}e^{-t/2}. \tag{46}$$

Indeed truncate the conditional density at the event that it is at most Ne^{-t}. The expected omitted integral and the integrated bias of the truncated estimator each equal the displayed probability term. The integrated variance of the average truncated density is at most e^{-t}, since the untruncated average expectation at each source point is one. Cauchy–Schwarz bounds the remaining integrated absolute error. This proves (46) for total variation defined as half the integrated absolute difference.

When t is a small fixed multiple of L the right side is exponentially small. A common-part coupling corrects the source marginal from $\widetilde{\mu}_N$ to μ at negligible W_s cost, because the Hilbert diameter is bounded. The expected sth power distortion to the empirical sample before that correction is bounded by the joint channel distortion, by the sample-index coupling. Thus the upper bound in expectation matches the lower bound as $\gamma \downarrow 0$, after normalizing using regular variation.

For the almost-sure assertion, replacing one sample vector changes the optimal cost W_s^s by at most the diameter to the power s divided by N. One can reroute the mass assigned to that sample index in any coupling, and apply the same argument in the reverse direction. Bounded differences concentration therefore has summable probabilities of deviation from the expected cost by any fixed multiple of the target normalized scale. This proves the almost-sure asymptotic for all N and finishes Theorem C.1.

C.5 The ordinary Kloosterman path metric

Let $P = \mathrm{ST}$ denote Sato–Tate on $[-2, 2]$ with variance one. Let

$$K_X(t) = tX_0 + \sum_{h \neq 0} X_h \frac{e^{2\pi i h t} - 1}{2\pi i h},$$

where all coefficients have independent law P. This is the Kloosterman limit of Kowalski–Sawin [36]. Denote its law by ν, and let $K_1, K_2, \ldots$ be independent samples. Use the same definitions of $D_{N,s}$ and ν_N for these paths with metric $L^2[0,1]$. In this section write

$$C_U = \frac{1}{4\pi^2} \left(\int_0^\infty U_{\mathrm{ST}}(u^{-2})\, du \right)^2, \qquad U \in \{F, G, B\}.$$

Theorem C.2. *In the ordinary $L^2[0,1]$ metric, for every fixed finite $s \geq 1$,*

$$D_{N,s}^2 \sim \frac{C_G}{\log N}, \qquad W_s(\nu, \nu_N)^2 \sim \frac{C_B}{\log N}$$

almost surely and at the corresponding joint moment scales. For almost every Kloosterman path K,

$$-r^2 \log \nu\{f : \|K - f\|_{L^2} \leq r\} \longrightarrow C_G \qquad (r \downarrow 0).$$

The three constants satisfy $C_F < C_G < C_B$.

Ordered by absolute frequency, each sign included, the periodic unpinned component has squared Fourier weights $\lambda_j \sim 1/(\pi^2 j^2)$. This gives the normalization of the constants. In the ordinary path, the constant Fourier component and the endpoint term must also be controlled.

C.5.1 Converse in the ordinary metric

For two paths the nonzero Fourier difference at index h is

$$\frac{X_h - Y_h - (X_0 - Y_0)}{2\pi i h}.$$

For nearest-neighbor small balls, condition on the source and the sampled endpoint Y_0. Put $v = X_0 - Y_0 \in [-4, 4]$. The Sato–Tate density is even and log-concave, and its convolution with a Gaussian kernel is log-concave by Prékopa's theorem [42]. Thus the function $g_a(x)$ is convex and even on $\mathbb{R}$, and $\mathbb{E}g_a(X - v) \geq \mathbb{E}g_a(X)$ for every v, by symmetry and convexity. On a central penalty band these translated functions are uniformly bounded and Lipschitz over the compact ranges. Hoeffding's inequality on a finite fine mesh of v, followed by Lipschitz interpolation, gives the central lower bound of the small-ball proof simultaneously for $|v| \leq 4$, outside an exponentially small set of centers. Conditional on the endpoint the other reproduction slots are independent. The same Chernoff estimate is therefore valid in the ordinary metric.

For the equal-mass lower bound require in addition that source and sample points each have central weighted coefficient sum of size at most ηL, with weights $a_h = \lambda_h/\phi(L)$. This is true off exponentially small mass by centering, independence and bounded weights. On these typical points the end correction satisfies

$$\sum_{h\,\text{central}} a_h|X_h - Y_h - v|^2 \geq \sum_{h\,\text{central}} a_h|X_h - Y_h|^2 - O(\eta L),$$

because the cross term is controlled by the two weighted sums and the v^2 term is nonnegative. The constant Fourier coefficient also only increases squared error. After adjusting the fixed slack in the equal-mass converse, the bound goes through in the ordinary metric.

C.5.2 Conditional small-ball lower estimate with pinning

Use the product change of measure for nearest-neighbor balls, with $b = 2$, but reserve all coefficients at $|h| \leq H = \lfloor L^{3/4} \rfloor$ and the endpoint. On the unreserved low coordinates use the same-cell condition, on the central coordinates use the G-tilt, and on the high ones leave reproduction independent. Under the joint tilted law of exterior pairs, each coefficient difference has mean zero. The low pairs are exchangeable in their chosen cell, the central pairs are invariant under simultaneous sign reversal, and the high pairs are iid. All exterior pairs are independent, so

$$S = \sum_{|h|>H} (X_h - Y_h)/h$$

has a subgaussian proxy $O(1/H)$. Thus $|S| \leq H^{-1/4}$ outside a joint event of stretched-exponentially small probability. Conditional on the center, the good joint events for this sum, central distortion and log likelihood have probability at least $1/2$ outside a stretched-exponentially small set of centers, by Markov's inequality. The omission of the first H slots from the low construction can only decrease its worst information and distortion costs.

For typical centers there are at least cH slots $h \in [H/2, H]$ with $|X_h| \leq 1$. Given the exterior reproduction, let M be the number of these slots. Aim at $Y_h = X_h + hS/M$ on them and at $Y_h = X_h$ on all other reserved slots. These targets remain in $[-2, 2]$ for large L, and cancel the harmonic sum exactly at the target values. Require that each reserved reproduction lie within L^{-2} of its target, and that Y_0 lie within L^{-2} of X_0. Every interval of this radius centered in the support has Sato–Tate probability at least a fixed positive multiple of L^{-3}. By independence the cost is at most $O(H \log L) = o(L)$ in negative log probability, uniformly in the exterior. The squared Fourier distortion added by the intentional shifts is $O(S^2/H) = o(1/L)$. The final harmonic residual is $O(L^{-2} \log L)$, and the endpoint error is $O(L^{-2})$. The triangle inequality now gives the same squared-radius threshold up to $o(1/L)$ in the ordinary path metric.

This supplies the small-ball lower probability bound for each fixed slack. The same change-of-index argument as for (43) yields the stated almost-sure small-ball exponent. Replacing a deterministic Hilbert diameter by moment bounds does not affect the random-code conclusions. In fact the

periodic component is deterministically bounded in L^2, the endpoint is bounded, and the pinned mean is subgaussian because its coefficient weights have a finite square sum. On the exceptional joint event use the distance to the first sample point and Cauchy–Schwarz with its bounded higher moments. The stretched-exponentially small probabilities are still summable on the logarithmic sequence N_j used in the nearest-neighbor proof.

C.5.3 Equal masses with a small pinning error

The preserving-product upper bound of the previous section must be adjusted so that the pinned means of paired paths are close. The following construction gives that adjustment while retaining both exact path marginals.

Fix the central cutoffs ϵ, T. Divide the central frequencies of each sign into bands of small relative width κ, so that the penalty a_h varies little and the oscillation of $1/h$ within a band of magnitudes comparable to U is $O(\kappa/U)$. In one band of m_0 slots choose a differentiability penalty and its symmetric optimal scalar pair π_a. Partition $[-2, 2]$ into finitely many cells of mesh Δ, each with strictly positive probability p_i. Write $q = (q_{ij})$ for the joint label law of π_a, whose marginals are both $p = (p_i)$. Replacing actual values by their cell means changes bounded squared distortions by $O(\Delta)$. The label mutual information $I(q)$ is at most the original one by data processing.

Draw a type $t = (t_i)$ of counts from the multinomial distribution of m_0 independent p-labels. For each t choose integer joint counts $C(t) = (c_{ij})$ with both marginals t. They can be chosen with $C(t)/m_0$ close to q whenever t/m_0 is close to p. To construct them, first transport each marginal of q to t/m_0 by a coupling with mismatch probability equal to its total variation distance. This changes the joint law by at most twice that distance. Multiply by m_0, round down all entries, then fill the row and column integer deficits. Their sums agree and their total is bounded by the square of the fixed number of cells. This changes the normalized joint law by $O(1/m_0)$.

Given t, draw uniformly a pair of label sequences with those joint counts. Each marginal sequence is uniform on the type class of t. Since t itself is multinomial, each full marginal sequence is iid with law p. Independently sample continuous values in their cells from the corresponding conditionals of P. Thus the resulting two vector marginals are exactly iid Sato–Tate. Repeat independently over the bands. The pair on one band has log density relative to its two independent marginals given by

$$\log \frac{m_0! / \prod_i t_i!}{m_0! / \prod_{ij} c_{ij}!} - \sum_i t_i \log p_i = m_0 \left[I(C(t)/m_0) + D(t/m_0 \| p) \right] + O(\log m_0).$$

The continuous conditional densities cancel. Typical multinomial types are within any prescribed fixed tolerance of p with exponentially high probability. On that set the log density per slot is at most $I(\pi_a)$ plus an arbitrarily small slack, by continuity of the finite-alphabet expressions and the count construction. The squared distortion per slot, after weighting within the band, is at most the chosen channel distortion plus a slack tending to zero with the bin size, band width and type tolerance. This follows directly from the fixed joint counts and the uniform within-cell error, without additional moment assumptions. Consequently total information and weighted periodic distortion obey the same bounds used in the preserving-product upper proof.

For the pinning difference let d_h be the difference of the cell means in a given band. By the equal label counts, $\sum d_h = 0$. Conditional on the joint counts, the d_h are a uniformly permuted fixed multiset. To bound $\sum d_h/h$, reveal this permutation sequentially. For deterministic weights w_h the martingale increment is the difference between the next revealed weight and the average of the remaining weights, times the difference between the revealed value and the previous average of the remaining values. Thus its absolute value is bounded by $C \operatorname{osc}(w)$, as the values are bounded. The

bounded-difference martingale inequality gives a subgaussian proxy $Cm_0 \operatorname{osc}(w)^2$. Taking $w_h = 1/h$ yields $O(\kappa^2 m_0/U^2)$. The within-cell errors are centered and conditionally independent with range $O(\Delta)$, giving proxy $O(\Delta^2 \sum h^{-2})$ on that band. Summing over independent bands, the total central difference has L^s norm

$$O_s((\kappa + \Delta)/\sqrt{\epsilon L}).$$

For the low band use independent pairs in common cells with cell size at most $C|h|/L$. The differences are centered, and their harmonic sum has proxy $O(\epsilon/L)$. The high independent tail has proxy $O(1/(TL))$. Couple the two endpoints by a common cell of mesh L^{-2}, also preserving both endpoint marginals, at information cost $O(\log L)$ with exponentially small probability of exceeding any fixed multiple of L. For fixed cutoff slack, choose ϵ small and T large, and then choose κ, Δ sufficiently small. The L^s norm of the finite-rank path correction is then arbitrarily small relative to $L^{-1/2}$. This implies the desired ordinary L^2 path distortion and information bounds under a joint law with both exact path marginals ν.

Apply the soft-covering inequality (46) on the coefficient sequence space and push the coupling to path space. The path norm has finite moments of every order as noted above. To repair the small source discrepancy by a common-part coupling in the unbounded metric, its cost to the power s is at most

$$C_s \operatorname{TV}(\widetilde{\nu}_N, \nu)^{1/2} \left(\int \|x\|^{2s} \, d(\widetilde{\nu}_N + \nu)(x) \right)^{1/2}.$$

This is Cauchy–Schwarz on the residual measures, after bounding the distance by the sum of norms. Its expectation is exponentially small by (46), Cauchy–Schwarz again, and $\mathbb{E}\widetilde{\nu}_N = \nu$. The triangle inequality for the arrangement of these couplings proves the matching expected equal-mass bound in the ordinary metric.

For an almost-sure asymptotic first clip the source and every sample radially to a Hilbert ball of radius $\log N$. Because the path norm has a subgaussian tail, the clipping costs are smaller than every inverse power of $\log N$ in the expected sth moment. Moreover all sample points are unchanged for all large N almost surely, by the union bound and Borel–Cantelli. The clipped optimal cost to the power s has coordinatewise bounded differences $O_s((\log N)^s/N)$ and hence concentrates at the target scale with summable exceptional probabilities. Removing the clipping proves the almost-sure assertion.

Finally, strictness $G_P(a) < B_P(a)$ holds for Sato–Tate at every positive a. The unconstrained one-marginal Gibbs kernel attaining $G_P(a)$ is unique by conditional relative entropy. Its output, to have marginal P, would require $\int e^{g_a(x) - a|x-y|^2} dP(x) = 1$ for P-almost every y. The left side is real analytic, positive and tends to zero at infinity. It cannot be one on the whole support interval. The optimizer of B_P was constructed in (44), so equality is impossible. The earlier strictness $J_F < J_G$ completes Theorem C.2.

C.6 A distribution-constrained arithmetic lower bound

Let $\mu_{p,B}$ be the path obtained from the increments $p^{-1/2} e_p(Ax + B/x)$, $x = 1, \dots, p-1$, for uniform nonzero A and fixed nonzero B. Let μ_p also average B over the nonzero residues.

Corollary C.3. *For every fixed finite $s \geq 1$, in either the L^2 or uniform path metric,*

$$\liminf_{p \to \infty} \sqrt{\log p}\, W_s(\mu_p, \nu) \geq \sqrt{C_B/2}, \qquad \liminf_{p \to \infty} \inf_{B \neq 0} \sqrt{\log p}\, W_s(\mu_{p,B}, \nu) \geq \sqrt{C_B},$$

where C_B is the strictly larger empirical constant of Theorem C.2.

Proof. The exact Fourier coefficient on the prime full mesh uses $W_h = p^{-1/2} \sum_{x \neq 0} e_p((A-h)x + B/x)$. For each fixed degree, the one- and two-shift polynomial comparisons with independent Sato–Tate coefficients have error $O(p^{-1/2})$, uniformly in distinct shifts and in fixed nonzero B. This is the translated Kloosterman sheaf estimate of [43, Lemma 6.3], with the character norms bounded at fixed degree. It also holds after averaging B.

Use the central band from the equal-mass converse with $L = R/A_B$, where $R = 2\log(p-1)$ or $\log(p-1)$, respectively, and $A_B = J_B/2$ refers to the scalar objective, not the arithmetic parameter. The central coefficients have distinct shifts for large p. The potentials f_{a_h} are uniformly bounded and Lipschitz on $[-2, 2]$. They can be uniformly approximated to any fixed tolerance by polynomials of one fixed degree and uniformly bounded coefficients, for instance by Bernstein approximation on this interval. Their one- and two-shift comparisons imply

$$L^{-1} \sum_{h \,\text{central}} [f_{a_h}(W_h) - \mathbb{E}_P f_{a_h}(X)] \longrightarrow 0 \quad \text{in probability,} \qquad L^{-1} \sum_{h \,\text{central}} a_h W_h \longrightarrow 0 \quad \text{in probability.}$$

Indeed, first use fixed polynomials, for which the normalized mean and variance bounds are $O(p^{-1/2})$ and $O(1/L + p^{-1/2})$, apart from any approximation error. Then use arbitrarily small fixed approximation tolerance. All convergence is uniform in B. The model has the same typical conditions by independence.

For clarity, if Q_p is the mesh path at j/p with last increment zero, then

$$2\pi i h \, \widehat{Q_p - t}(h) = e^{\pi i h/p} \operatorname{sinc}(\pi h/p) W_h \qquad (h \neq 0).$$

This follows by integrating the piecewise constant derivative against the exponential. The factor differs from one by $O(|h|/p)$ on the central band. Its contribution to the normalized squared error sum there is $O(L^2/p) = o(L)$. The original equal-step path differs from Q_p by $O(p^{-1/2})$ in uniform norm, since the time change is $O(1/p)$ and each increment is $O(p^{-1/2})$. The resulting metric discrepancy is negligible at scale $L^{-1/2}$. By the ordinary-path cross-term argument, the endpoint shift is also negligible in the lower bound on the normalized central cost for typical pairs.

For any coupling, all but $o(1)$ of mass belongs to typical source and arithmetic targets as above. The latter form a deterministic set of at most e^R paths with lifted generating parameters. The tilted union bound of the equal-mass converse applies to this set, since the exponent at each target uses only its typical central potential sum. It shows that the coupled mass below a squared threshold $(J_B/2 - \eta)L\phi(L)$ tends to zero for every fixed $\eta > 0$, allowing the other tolerances to be small enough. The argument also applies if two labels produce the same path, by lifting the coupling to the given parameter probabilities. This proves the lower bounds in L^2, and the uniform statements follow since the uniform norm dominates L^2. □

The same marginal argument applies at a fixed tame prime and along either parity of depth in a fixed residue class of the hyper-Kloosterman paths of Section A. The fixed one- and two-shift torus character comparisons there give convergence uniformly for shifts $O(n)$ and fixed Fourier height. Uniform polynomial approximation on the bounded trace ranges, or approximation of the phase test by bounded-height torus characters, yields the same central typicality. The mesh factor is the displayed one with modulus p^n, and the endpoint cross-term uses a real part of a complex weighted sum when necessary. Write $d = \gcd(R_0, p-1)$ for tame rank R_0. Let P_{act} be the common active coefficient law up to isometries, $\alpha = (p-1)/(dp)$, and $\mu_{p^n,\text{res}}, \nu_{p,\text{res}}$ the arithmetic and limiting residue laws for the given parity. The conditional arithmetic bound is

$$\liminf_{n \to \infty} \sqrt{n}\, W_s(\mu_{p^n,\text{res}}, \nu_{p,\text{res}}) \geq \frac{\alpha}{2\pi\sqrt{\log p}} \int_0^\infty B_{P_{\text{act}}}(u^{-2})\, du.$$

Indeed the inactive coordinates have a degenerate coefficient law and contribute zero to the objective, the deterministic gates have density α, and the conditional support budget is p^{n-1}. This is a lower bound for arithmetic transport. The empirical theorems concern independent sample paths, for which there is a matching construction.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, internal independent checks and drafting. The proofs use no datasets.

D A character-moment obstruction to the Kloosterman transport constant

The prime Kloosterman path is governed by simultaneous translated character estimates. We construct arrays satisfying the same estimates separately at every inverse-parameter label and at every representation degree. They have the full number of distinct paths, the exact singular diagonal, and the exact row histograms. Their squared empirical transport constants fill the interval between those associated with p and p^2 independent paths. Thus the published character-moment inequalities, even with these structural constraints, leave the exact arithmetic transport constant undetermined. The construction uses column permutations of a single arithmetic table and a two-row likelihood calculation in the ordinary L^2 path metric.

D.1 The obstruction

For an odd prime p, let $H_p = \{-(p-1)/2, \dots, (p-1)/2\}$ represent $\mathbb{F}_p$. Put

$$\mathrm{Kl}_2(a;p) = p^{-1/2} \sum_{x \in \mathbb{F}_p^\times} e^{2\pi i (x + a/x)/p}.$$

Let P be the Sato–Tate law on $[-2,2]$, with density $\sqrt{4-x^2}/(2\pi)$. For a real vector $w = (w_h)_{h \in H_p}$ define its pinned Fourier path by

$$\mathcal{K}_p(w)(t) = t w_0 + \sum_{h \in H_p \setminus \{0\}} \frac{e^{2\pi i h t} - 1}{2\pi i h}\, w_h.$$

Write ν for the law of the corresponding infinite series with independent P coefficients, and always use the ordinary complex-valued $L^2[0,1]$ norm. The preserving scalar objective and ordinary L^2 empirical constant of Section C are

$$B_P(a) = \inf_{\pi \in \Pi(P,P)} \left[I_\pi(X;Y) + a \mathbb{E}_\pi (X-Y)^2 \right], \qquad J = \int_0^\infty B_P(u^{-2})\, du, \qquad C_B = \frac{J^2}{4\pi^2}.$$

Here I_π is mutual information with natural logarithms. In particular, independent equal-mass empirical samples of size N from ν have leading squared W_2 cost $C_B/\log N$.

Theorem D.1. *Fix, for each prime $p \geq 5$, any nonzero B_0 and any integer $1 \leq G \leq p-1$. There exist permutations π_b of $\mathbb{F}_p$, indexed by $b \in \mathbb{F}_p^\times$, such that the $(p-1)^2$ paths assigned to*

$$w_h^{A,b} = \mathrm{Kl}_2(B_0[\pi_b(A) - \pi_b(h)]; p), \qquad A, b \in \mathbb{F}_p^\times, \quad h \in H_p$$

are all distinct and have uniform empirical law $\mu_{p,G}$ satisfying

$$W_2^2(\mu_{p,G},\nu)\sim\frac{C_B}{\log((p-1)G)}$$

as $p\to\infty$ along any prescribed choice $G=G(p)$. For each fixed b, every set of distinct shifts h_i, and all nonnegative degrees k_i with $s=\#\{i:k_i>0\}>0$,

$$\left|\frac{1}{p-1}\sum_{A\neq0}\prod_i\chi_{k_i}(w_{h_i}^{A,b})\right|\le\frac{4(s+1)}{\sqrt{p}}\prod_i(k_i+1).$$

Here $\chi_k(2\cos\theta)=\sin((k+1)\theta)/\sin\theta$ with continuous endpoint values. Every row has $w_A^{A,b}=-p^{-1/2}$ and the multiset of entries $\{\mathrm{Kl}_2(a;p):a\in\mathbb{F}_p\}$.

The character bound is exactly the one in the published edition [44, Lemma 6.3]. Taking $G=p^{\vartheta+o(1)}$ with $0\le\vartheta\le1$ yields the continuum $C_B/(1+\vartheta)$ of limits of $(\log p)W_2^2$. All constructions use $(p-1)^2$ distinct equally weighted paths. The arrays may differ from the actual Kloosterman family in the coupling of the columns across labels b. In the actual family the full array has the affine action $h\mapsto B(A-h)$ over one fixed trace function, whereas the theorem allows column permutations. The obstruction concerns a deduction from the stated inequalities and the listed row structure.

Small-bias constructions such as [45] separate character tests from support size in finite binary spaces. Here the test class consists of continuous Sato–Tate characters, the inequalities hold within each inverse-parameter group, and the separation concerns the exact leading ordinary path transport costs with identical cardinalities. The proof uses the ordinary L^2 preserving-coupling and converse estimates in Section C; we state the precise inputs where needed.

D.2 The permitted moment tests and finite bins

Set $s_p(x)=\mathrm{Kl}_2(B_0x;p)$. For any permutation π, the images $\pi(h_i)$ are distinct, and as A ranges over $\mathbb{F}_p^\times$, $\pi(A)$ ranges over $\mathbb{F}_p\setminus\{\pi(0)\}$. The proof of [44, Lemma 6.3] bounds the complete active trace-product sum by $Ds\sqrt{p}$ with $D=\prod_i(k_i+1)$, and changing the active finite domain to the domain with one point omitted adds at most $(s+1)D$. The omitted point in that proof can be arbitrary. Dividing by $p-1$ gives the stated $4(s+1)D/\sqrt{p}$ bound. For $k_i=0$ throughout the empirical and model values agree. The row histogram and singular diagonal are immediate from permutation and $B_0\neq0$.

We give a bin approximation uniform over rows and row pairs. Use an equal-interval partition of $[-2,2]$ of mesh $\Delta\asymp(\log p)^{-D_0}$ with $D_0>2$, and denote the probabilities of the bins under P by q_i. Their minimum is at least $c\Delta^{3/2}$, by the semicircle density at the endpoints. For arbitrary distinct shifts $a,a'\in\mathbb{F}_p$, the histogram of $(s_p(a-u),s_p(a'-u))$ for uniform $u\in\mathbb{F}_p$ assigns each pair of bins probability

$$q_iq_j(1+O_K((\log p)^{-K})) \tag{47}$$

for every fixed K, uniformly over bins, shifts and B_0. The analogous one-shift statement has reference probability q_i. Both remain valid if up to two specified elements of the sampling pool are removed.

To see this, sandwich each indicator between bounded Lipschitz functions differing from it only in collars of width $(\log p)^{-D_1}$ at the bin boundaries. Their integrals under P differ by $O((\log p)^{-D_1})$. Bernstein polynomials of degree M approximate either function uniformly with error at most a fixed

multiple of its Lipschitz constant times $M^{-1/2}$. Choose M a sufficiently high fixed power of $\log p$, so the error has arbitrary inverse polylogarithmic order. A polynomial of degree M bounded by one has coefficients of modulus at most one in the orthonormal χ_k basis on $L^2(P)$. Its summed character weights $\sum_{k\le M}(k+1)|c_k|$ are $O(M^2)$. Thus the one- and two-shift moment estimates, with at most $O(1/p)$ corrections for the complete pool and the terms removed, bound the polynomial comparison error by $p^{-1/2}$ times a fixed power of $\log p$. Using products of the bounding approximations gives arbitrary inverse polylogarithmic absolute error for every pair cell. Dividing by the minimum cell probability, which is only inverse polylogarithmic, proves (47).

Suppose π is uniformly random, and consider a fixed set $\mathcal{H}$ of $m = O(\log p)$ retained frequencies containing 0. For row $A \notin \mathcal{H}$, condition on $\pi(A)$. Then $\pi(h)$ for $h \in \mathcal{H}$ samples without replacement from the remaining $p-1$ elements. For distinct $A, A' \notin \mathcal{H}$, condition on both images; the two-row binned pattern samples from the pair histogram with two elements removed. At draw j, at most m elements of any cell have been used. Therefore the conditional bin probabilities are the independent P-bin probabilities, or their independent two-row products, each with relative error bounded by

$$O_K((\log p)^{-K}) + O\left(\frac{m}{p \min_i q_i^2}\right).$$

Multiplying over the m draws shows that each entire one- or two-row retained pattern has relative density $1 + O_K((\log p)^{-K})$ with respect to the corresponding independent bin law, with arbitrarily large fixed K. Within each group the excluded rows have total fraction $O(m/p)$.

D.3 The exact cost of repeated random permutation groups

Take G independent uniform permutations $\pi_1, \dots, \pi_G$ and distribute the $p-1$ inverse labels across these groups as evenly as possible. For the moment allow repetition. Write $N_0 = (p-1)G$ and $R = \log N_0$. The effective codebook contains N_0 labelled points, each of mass at most $2/N_0$. Each of the G groups has mass at most $2/G$. We claim

$$\limsup_{p\to\infty} R\,\mathbb{E}W_2^2(\mu,\nu) \le C_B \tag{48}$$

for the resulting random law μ.

We recall the precise preserving input from Section C. Given any small $\eta > 0$, select a large cutoff constant T and couple two iid P coefficient sequences, each with exact product marginal. Use endpoint and head cell couplings and a fixed-composition coupling in short central relative bands. This yields a density ℓ relative to the independent marginals, depending only on $|h| \le TR$, with mean squared ordinary path distortion at most $(C_B + \eta)/R$ and

$$\mathbb{P}_\ell\{\log \ell > (1-\delta)R\} \le e^{-cR}$$

for some $\delta, c > 0$. The cutoff T may be taken arbitrarily large. The argument first leaves slack in the leading information per unit rate; low cell matching and central types preserve both marginals and pay the pinning term. The tail outside the retained band is independent of the finite coupling. In what follows use $\mathcal{H} = \{|h| \le TR\}$ with integral endpoints and include the endpoint coordinate. Frequencies beyond $\mathcal{H}$ in the reference sequences remain independent.

Write P^∞ for the infinite product law, let $\ell_0 = \ell\,\mathbf{1}_{\log\ell \le (1-\delta)R}$, and put $L_{\max} = e^{(1-\delta)R}$. If z denotes the collection of fine bins of the retained second vector, average over its independent reference conditional to obtain

$$k_0(x,z) = \int \ell_0(x,y)\,P^\infty(dy \mid z), \qquad m_x = \mathbb{E}_z k_0(x,z) \le 1.$$

Then $k_0 \le L_{\max}$ and $\int k_0(x,z)P^\infty(dx) \le 1$ for each bin pattern, by the exact second marginal of ℓ. Moreover $\int (1-m_x)\,dP^\infty(x) \le e^{-cR}$.

Omit the special rows $A \in \mathcal{H}$ from the partial construction. With source P^∞, assign to each remaining row of mass w_i and bins z_i the partial measure $w_i k_0(x,z_i)P^\infty(dx)$. Its mass is at most w_i. The projected source density is $S(x) = \sum_i w_i k_0(x,z_i)$. Let $\epsilon_p = (\log p)^{-K}$ for sufficiently large fixed K in the relative pattern estimates. For each fixed x, the comparison of one-row laws bounds the mean bias relative to m_x by $C(\epsilon_p + m/p)$. The diagonal part of the variance is at most $CL_{\max}/N_0$ because $k_0^2 \le L_{\max}k_0$. For distinct retained rows in one group, their joint patterns have uniform relative error $O(\epsilon_p)$ against independence. Their covariance is therefore at most $C\epsilon_p m_x^2$. After weighting by the masses, the within-group contribution is at most $C\epsilon_p/G$. Rows in distinct groups are independent. Integrating the first absolute moment of the source discrepancy yields

$$\mathbb{E}\int |S-1|\,dP^\infty \le C\big[e^{-cR} + \epsilon_p + m/p + (L_{\max}/N_0 + \epsilon_p/G)^{1/2}\big]. \tag{49}$$

Trim excess source density to one, then couple the remaining source and target deficits. Their common mass is at most $\int |S-1|\,dP^\infty$. A full target path has norm bounded by $C\log p$ deterministically, and the source path law has bounded fourth norm moment. Cauchy–Schwarz thus bounds the squared repair cost by $C(z^{1/2} + (\log p)^2 z)$ if the deficit mass is z. By (49) its expected value is $o(1/R)$ on taking K sufficiently large, including when $G = 1$. To estimate the retained cost, lift each partial row measure through the reference second vector y in its bin pattern, using density $\ell_0(x,y)$. The one-row relative pattern comparison bounds its expected source-to-y squared cost, summed over the rows and their weights, by $(1+O(\epsilon_p))(C_B+\eta)/R$. On the retained frequencies, the second reference path and the assigned row agree up to bin errors. The corresponding ordinary norm discrepancy is at most $C\Delta(1+\log(m+1))$ deterministically, including the endpoint and pinned constant. The high tail of the reference path has squared mean $O(1/(TR))$ under the lifted mass because ℓ_0 is independent of its high coordinates.

For a permutation row, its high periodic part has squared norm $O(1/(TR))$ deterministically by Parseval and $|s_p| \le 2$. Its high pinned constant is a weighted sum of uniformly sampled entries from the seed, conditional on the image of A, except for the potentially deterministic value $s_p(0)$ at $h = A$. The signed weights $1/h$ sum to zero on the symmetric high band. If A belongs to that band, the mean of the pinned constant is consequently $O(1/|A|) = O(1/(TR))$; otherwise the mean vanishes. The usual formula for sampling without replacement gives variance $O(\sum_{|h|>TR} h^{-2})$: the individual variances are bounded and common covariances are minus the variance divided by the population size minus one. Thus the row's high tail has squared mean $O(1/(TR))$, uniformly in the row. The partial row mass is at most its assigned weight, so this bound also applies to the expected cost under the partial coupling. The triangle inequality in the joint mean-square cost shows

$$\big(\mathbb{E}W_2^2(\mu,\nu)\big)^{1/2} \le (1+o(1))\sqrt{(C_B+\eta)/R} + O((TR)^{-1/2}) + o(R^{-1/2}).$$

Choosing T arbitrarily large and η arbitrarily small proves (48).

For the matching lower bound, any such collection of G permutations, with the above row weights, satisfies

$$\liminf_{p\to\infty} RW_2^2(\mu,\nu) \ge C_B. \tag{50}$$

The reason is the distribution-constrained converse of Section C, which only needs uniform one- and two-shift comparisons for fixed-degree polynomials and the upper bound N_0 on the support count. Here are the details of that implication. Retain a central band of order R and use the scalar preserving dual potentials f_{a_h} for $B_P(a_h)$, where the parameters a_h are proportional to R^2/h^2 on

the band. They are uniformly bounded and Lipschitz on the central range. For any fixed approximation tolerance they admit uniform polynomial approximations of one fixed degree. The first two character-moment comparisons then show that the normalized sum of these potentials over the central band converges in probability to its Sato–Tate mean, uniformly over the G groups. Indeed the diagonal contributions to the variance divided by R^2 are $O(R^{-1})$, and the off-diagonal contributions tend to zero by the two-shift bound. The same argument gives a negligible normalized linear sum of the central coefficients with weights a_h. The independent source has the same typicality by its product structure.

For typical pairs the squared distance in the ordinary path controls the squared central coefficient distance, with loss at most ε/R for any prescribed $\varepsilon > 0$ on suitably chosen typical sets. In the nonzero Fourier modes the coefficient of the endpoint difference subtracts the same bounded scalar $X_0 - Y_0$ from each $X_h - Y_h$. The linear cross term is controlled by the two negligible normalized weighted sums, and the pure endpoint square is nonnegative. The constant Fourier mode also adds a nonnegative term. On a fixed typical target the conditional probability of a typical source within squared cost less than $(C_B - \eta)/R$, for any fixed $\eta > 0$, is at most $e^{-(1+c_\eta)R}$ under the source, after central cutoffs and tolerances are chosen. This is the preserving-dual exponential estimate in Section C: the conditional density is $\exp(f_{a_h}(x) + f_{a_h}(y) - a_h(x-y)^2)$ in each coordinate, and its integral over the source is one. The band scaling makes the two typical potential sums minus the allowed cost exceed $(1+c_\eta)R$, with the normalization $C_B = J^2/(4\pi^2)$. A union bound over at most $N_0 = e^R$ centers shows that typical pairs below this cost carry vanishing mass in any transport plan. The atypical marginal mass tends to zero as shown above, proving (50).

D.4 All labels distinct using high-frequency permutations

Injectivity of s_p for $p \geq 5$ follows directly from a cyclotomic calculation. If two unnormalized Kloosterman sums agree, let $n_a(t)$ count the appearances of t in the phases $x + a/x$, over $x \neq 0$. The polynomial with coefficients $n_a(t) - n_{a'}(t)$ for $0 \leq t < p$ is a multiple of the cyclotomic polynomial $1 + z + \cdots + z^{p-1}$. Its coefficient sum is zero, so this multiple is zero. The two phase histograms agree. But in $\mathbb{F}_p$,

$$\sum_t t^2 n_a(t) = \sum_{x \neq 0} (x + a/x)^2 = -2a$$

because $\sum_{x \neq 0} x^{\pm 2} = 0$ for $p \geq 5$. Therefore $a = a'$. Since $B_0 \neq 0$, s_p is injective as well.

Reserve a set $\mathcal{R}$ of $M = \lceil 3 \log p \rceil$ frequencies with $p/4 \leq h \leq (p-1)/2$, possible for all large p. The probability that two independent base permutations differ by a translation on its complement is at most $p\, M!/p!$, because every fixed translation of the restriction has at most $M!$ extensions. The union over at most p^2 pairs still has probability $o(1)$. For all sufficiently large p, condition on the event that no pair has translation-coincident restrictions. The conditional mean in (48) increases by at most a factor $1 + o(1)$, so one can choose deterministic base permutations satisfying the distinct-restriction property and the upper asymptotic simultaneously. To make this choice for every p, take a minimizing realization on this event, whose cost is bounded above by its conditional mean. Bound (50) applies to all base choices.

Within each base, replace its repeated copies by distinct right compositions with permutations supported on $\mathcal{R}$. There are $M!$ available choices, enough to distinguish every copy. A modified permutation π' equals its base π off $\mathcal{R}$. Relabel its row A as $A' = (\pi')^{-1}\pi(A)$, a permutation of $\mathbb{F}_p^\times$ since $0 \notin \mathcal{R}$. Then every unreserved coefficient is unchanged, including $h = 0$. The uniform path discrepancy is $O(M/p)$ by summing the harmonic terms over $\mathcal{R}$. Coupling these relabelled

paths gives the same ordinary L^2 rate up to $O(M/p) = o(R^{-1/2})$. All character tests and exact histograms continue to hold, since every modified group is a permuted arithmetic seed.

Finally, equality of two paths would force equality of their endpoints and all pinned Fourier coefficients. By injectivity of s_p their two permutations would then differ by a translation at every h. For different base groups this is excluded by their restrictions. Within a base the permutations agree off $\mathcal{R}$, so the translation must be zero. Different copies are distinct by construction, and injectivity also forces the same row index when the permutation agrees. For finitely many small primes the asymptotic argument is unaffected, and one may use the original affine family of distinct rows: take $\pi_b(h) = (b/B_0)h$, whose restrictions on the whole field differ by more than a translation. This completes the proof of Theorem D.1.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, independent internal checks and drafting. The proof uses no data.

E Affine Kloosterman correlations and the empirical transport constant

The two-parameter prime Kloosterman coefficient array is an affine orbit of one scalar trace function. We show that an affine orbit of one independent Sato–Tate seed attains the same exact transport constant as a fully independent codebook, despite having far fewer independent inputs. A common random relabeling of the genuine arithmetic seed also attains this constant. The unpermuted arithmetic orbit has explicit completed third- and fourth-moment resonances and sharp averaged square-root-scale cancellation in the number of affine rows for uniformly sampled patterns of logarithmic size. For the exact arithmetic constant we give a reduction to a one-sided short-window covariance estimate of a clipped likelihood and identify the representation cost of a uniform absolute-error approach. These results distinguish randomness in the seed from the role of the affine action.

E.1 Affine path ensembles

Let p tend through odd primes and put

$$s_p(z) = p^{-1/2} \sum_{x \in \mathbb{F}_p^\times} e_p(x + z/x), \qquad e_p(z) = \exp(2\pi i z/p).$$

Write $\mathcal{G} = \mathrm{AGL}_1(\mathbb{F}_p)$, with elements $\gamma h = a + bh$, $b \neq 0$, and size $p(p-1)$. The complete arithmetic coefficient array is $w_h^\gamma = s_p(\gamma h)$. In the prime Kloosterman family $a = AB$ and $b = -B$ with $A, B \neq 0$, so the complete array adds exactly the rows $a = 0$. Let $H_p = \{-(p-1)/2, \dots, (p-1)/2\}$. Represent any bounded real seed $(S_z)_{z \in \mathbb{F}_p}$ by its affine path orbit

$$\mathcal{K}_p(S, \gamma)(t) = tS_{\gamma 0} + \sum_{h \in H_p \setminus \{0\}} \frac{e^{2\pi i h t} - 1}{2\pi i h} S_{\gamma h}.$$

Write ν for the limiting law with independent Sato–Tate coefficients $P(dx) = \sqrt{4 - x^2}\, dx/(2\pi)$ in the corresponding infinite series. The metric is the ordinary complex-valued $L^2[0,1]$ distance. For the preserving Sato–Tate scalar objective and the empirical constant let

$$B_P(a) = \inf_{\pi \in \Pi(P,P)} [I_\pi(X;Y) + a\mathbb{E}_\pi(X - Y)^2], \qquad J = \int_0^\infty B_P(u^{-2})\, du, \qquad C_B = J^2/(4\pi^2).$$

The limiting ordinary-path constant and its fixed-marginal coding proof are in Section C. Here I_π is mutual information with natural logarithms.

Theorem E.1. *Draw the p seed values S_z independently from P, and let ν_p^{aff} be the equal-mass law of the $p(p-1)$ affine paths. For each fixed finite Wasserstein order $q \geq 1$, with $R = \log[p(p-1)]$,*

$$RW_q^2(\nu, \nu_p^{\mathrm{aff}}) \longrightarrow C_B \quad \textit{almost surely,} \qquad R\big(\mathbb{E}W_q^q(\nu, \nu_p^{\mathrm{aff}})\big)^{2/q} \longrightarrow C_B.$$

Almost sure convergence holds under any joint realization of the seed arrays across primes.

The result uses the exact affine action from the arithmetic array, with a random seed. A previous obstruction Section D preserves the published conditional character estimates and row histograms by permutations; its rows are allowed to break the affine link between inverse parameters. Theorem E.1 treats a separate aspect of the question. It shows that the affine action alone permits the exact independent-codebook constant. The arithmetic seed's short-window behavior is the remaining issue. The original convergence of Kloosterman paths is due to Kowalski–Sawin [5]; the scalar trace estimates and path-rate question of Kowalski–Untrau [46] and the published two-parameter estimates [44] give the surrounding arithmetic context.

E.2 Preserving likelihoods and the random affine theorem

We use the following precise input of the ordinary-path argument in Section C. For each prescribed distortion slack $\eta > 0$, choose a large fixed coefficient cutoff T and a preserving coupling of two independent-law coefficient vectors, with joint density ℓ relative to $P^\infty \otimes P^\infty$. Here preserving means that both marginals of the coupling are exactly P^∞. For R large, ℓ depends only on the coordinates in $H = \{|h| \leq TR\}$, including 0. Its ordinary path qth distortion moment, raised to power $2/q$, is at most $(C_B + \eta)/R$. The clipped likelihood

$$k_0 = \ell \, \mathbf{1}_{\log \ell \leq (1-\delta)R}$$

for some fixed $\delta > 0$ discards exponentially small planted mass. The joint law in each central relative band is a constant-composition coupling, and on the head and endpoint it uses common-cell couplings. Thus as a function of each output coordinate, k_0 is constant on the cells of that coordinate's partition. The minimum P probability of any of these cells is at least R^{-C} for a fixed C. In the central bands the partitions have fixed finite size. On the head the cell sizes may be chosen of order $|h|/R$ and at the endpoint of order R^{-2}, so the inverse polynomial bound follows from the semicircle density. The low and central coupling parameters and its information slack can be taken as in the companion proof. The reference coordinates outside H remain independent.

For fixed source vector x, write $m(x) = \int k_0(x, y) dP^\infty(y) \leq 1$. Let

$$S(x) = \frac{1}{|\mathcal{G}|} \sum_{\gamma \in \mathcal{G}} k_0(x, w^\gamma)$$

on the random affine orbit. Here k_0 uses only the observed finite coordinates of each row. Every row has independent P entries, hence $\mathbb{E}S(x) = m(x)$. For two rows γ and $\gamma\sigma$, their coefficient vectors restricted to H are jointly independent outside the common sites of H and σH, with matching seed values at common sites.

There are $O(|H|^4)$ relative maps σ with at least two common sites: two distinct matched pairs determine an affine map. There are $O(p|H|^2)$ sharing at least one site. For no common sites the covariance at fixed x is zero. For one shared coordinate the product expectation is the integral of

the product of two functions of that one common input, after averaging the remaining coordinates. Each function has integral at most one. Because it is constant on the cell of that coordinate, it is bounded by the reciprocal of its smallest cell probability, so the product integral is at most R^C. For at least two common sites, the product expectation is at most $e^{(1-\delta)R}m(x)$ by the cap and the exact one-row marginal. Uniformly in x this bounds

$$\operatorname{Var} S(x) \le C' \left(R^C |H|^2/p + e^{(1-\delta)R}|H|^4/|\mathcal{G}| \right).$$

Both terms decrease at least as a fixed positive power of p up to logarithmic factors. The integral of $1-m(x)$ under P^∞ is exponentially small in R. Thus the expected integrated source discrepancy is a negative power of p up to logarithmic factors.

For row γ assign the subprobability source mass $|\mathcal{G}|^{-1}k_0(x, w^\gamma)dP^\infty(x)$ to its path. Each row gets at most its allotted mass since the planted second marginal was exact. The conditional integrals are constant on the output cells, so their bounds hold for every row with the fixed assignment of cell boundaries. Trim excess source density and fill the two deficits. The fill cost to power q is bounded by $C_q(z^{1/2} + (\log p)^q z)$ if the missing mass is z: the source has finite $2q$ moment and each actual path has norm at most $C \log p$ deterministically. Its expected cost is negligible by the variance bound, with $z \le \int |S-1|$.

For the retained cost, supplement each row by independent reference coordinates beyond H_p. Under the seed randomness each supplemented output is exactly P^∞, so the expected partial cost to it is bounded by the distortion of the planted coupling. Removing the supplemental tail costs $O_q(p^{-1/2})$ in root moment of order q: the periodic part has squared weights with sum $O(1/p)$, and the harmonic pinned constant is a sum of centered independent bounded variables with the same square-sum bound. The cap can only lower each integrated conditional mass. This proves the upper bound in expectation after η tends to zero.

For the lower bound, uniform independent coefficients within every individual row give the usual typicality of the central preserving potentials and weighted coefficient sums for the target marginal, in expectation over the seed. The ordinary-path converse in Section C requires only this typicality and cardinality at most e^R. For a typical row the conditional probability of a typical source at squared distance at most $(C_B - \eta)/R$ is at most $e^{-(1+c_\eta)R}$ by the preserving dual. Indeed the endpoint shift in each central Fourier component changes the central squared coefficient comparison only through the negligible two weighted coefficient sums and a nonnegative square. A union bound over the rows makes cheap coupled mass negligible, with the remaining atypical mass tending to zero in expectation. Thus the expected moment lower asymptotic matches the upper.

For almost sure convergence, radially clip the source path to norm $C' \log p$, larger than the deterministic bound for all finite target rows. The independent-series norm of the source has a subgaussian tail, by orthogonality for the periodic part and bounded-variable concentration for its pinned constant. The clipping cost is smaller than any inverse power of $\log p$. Changing one seed value changes just one frequency in every target row. Every frequency is that changed position for a fraction $1/p$ of the rows. Therefore the average path displacement under the natural row matching is $O((\log p)/p)$ from summing the harmonic weights and the endpoint contribution. The clipped optimal qth cost changes by at most $C_q(\log p)^q/p$, by rerouting row masses and the mean value bound for the power cost on bounded distances. Bounded differences at scale $(\log p)^{-q/2}$ give summable exceptional probabilities, so Borel–Cantelli, the mean limit, and unclipping prove Theorem E.1.

E.3 Exact completed arithmetic moments

For distinct $h_1, \dots, h_m \in \mathbb{F}_p$ write

$$M(h_1, \dots, h_m) = \frac{1}{|\mathcal{G}|} \sum_{\gamma \in \mathcal{G}} \prod_{i=1}^{m} s_p(\gamma h_i).$$

For any such list let S_1 be the sum of $e_p(\sum_i z_i^{-1})$ over $z_i \neq 0$ with $\sum_i z_i = 0$, and S_2 the same sum with $\sum_i h_i z_i = 0$ also imposed. Then

$$M(h_1, \dots, h_m) = p^{-m/2}(pS_2 - S_1)/(p-1). \tag{51}$$

Indeed, the scalar Fourier expansion is $s_p(x) = p^{-1/2} \sum_{z \neq 0} e_p(z^{-1} + xz)$. Averaging the translation $\gamma(0)$ enforces the first linear relation; averaging the nonzero slope gives $(p\mathbf{1}_{\sum h_i z_i = 0} - 1)/(p-1)$. On every projective direction of admissible nonzero tuples the radial sum is $p-1$ or -1 as $\sum_i z_i^{-1}$ vanishes or does not. Hence each $S_j = pN_j - T_j$, where T_j is the number of admissible projective directions and N_j counts those with reciprocal sum zero.

Proposition E.2. *For three distinct shifts define $\mathcal{E}$ by*

$$(h_1 - h_2)^2 + (h_2 - h_3)^2 + (h_3 - h_1)^2 = 0 \quad \text{in } \mathbb{F}_p,$$

and let ν_p count the roots of $z^2 + z + 1$ in $\mathbb{F}_p$. Then

$$M(h_1, h_2, h_3) = \frac{p^{-3/2}}{p-1}(p^2 \mathbf{1}_{\mathcal{E}} - p\nu_p - 2).$$

For four distinct shifts, let $\mathcal{P}$ be the event that one pairing satisfies $h_i + h_j = h_k + h_\ell$. Then

$$M(h_1, h_2, h_3, h_4) = \frac{p^{-2}}{p-1}(6p + 3 - p^2 \mathbf{1}_{\mathcal{P}}).$$

In particular the resonant triple can have size $p^{-1/2}$ and the resonant quadruple has a leading term of size $-p^{-1}$. The nonresonant completed fourth moment is $O(p^{-2})$.

Proof. For three shifts the second linear kernel has just one projective direction, represented by $(h_2 - h_3, h_3 - h_1, h_1 - h_2)$, with all entries nonzero. Its reciprocal sum vanishes exactly when the sum of their squares vanishes, since their sum is zero and p is odd. Thus $S_2 = p\mathbf{1}_{\mathcal{E}} - 1$. For the first kernel write $(1, t, -1-t)$, with $t \neq 0, -1$. The reciprocal condition is $t^2 + t + 1 = 0$, so $S_1 = p\nu_p - (p-2)$.

For four shifts the first linear kernel has $p^2 - 3p + 3$ projective directions with nonzero entries, by inclusion and exclusion of the four coordinate lines in that projective plane. The reciprocal condition is $e_3(z_1, z_2, z_3, z_4) = 0$, which factors on $\sum z_i = 0$ into the three pairing conditions $z_i = -z_j$, $z_k = -z_\ell$. Each pairing line has $p-1$ allowed points and each intersection of two different pairing lines gives one allowed point, with no triple intersection. Thus the reciprocal zeros total $3p - 6$ and $S_1 = 2p^2 - 3p - 3$.

The second linear kernel is a projective line. Its four excluded coordinate points are distinct, since two zero coordinates would force the remaining two coordinates to vanish by distinctness of the shifts. Thus there are $p-3$ allowed points. Each pairing gives one allowed solution of the second linear equation. Two such points coincide exactly when $h_i + h_j = h_k + h_\ell$ for a pairing of the shifts, since a coincident direction has entries $(z, z, -z, -z)$ in some order. With distinct shifts in odd characteristic there is at most one such shift pairing, and there is no triple coincidence. Hence $S_2 = p(3 - \mathbf{1}_{\mathcal{P}}) - (p-3)$. Equation (51) gives both formulas. □

For shifts in an integer window of length $O(\log p)$, the triple condition is absent for large p: the positive sum of three squared integer differences is below p. It can occur in general patterns when -3 is a square, by solving the quadratic shape equation after an affine normalization of two sites. Additive quadruples can occur in the short windows. The reciprocal count works at every degree as written. On the projective torus its Jacobian can lose rank relative to the two linear constraints only where $z_i^{-2} = \alpha + \beta h_i$ for constants α, β; this also forces the reciprocal equation on the linear kernel. If $\beta = 0$, these are balanced sign directions with both signed linear sums zero. If $\beta \neq 0$, write $t = \alpha/\beta$ and choose signed square roots of $t + h_i$ in the algebraic closure. The condition is equivalently that the signed sum of the square roots and its derivative with respect to t both vanish at this nonsingular local branch. This follows by expressing the two linear equations in the signed reciprocal square roots.

The excluded rows $a = 0$ in the actual $A, B \neq 0$ family cost $O_m(p^{-3/2})$ in these moment comparisons. Indeed the omitted vertical moment is the one-parameter product of distinct multiplicative translates of s_p, which is $O_m(p^{-1/2})$ by the product trace estimate [47, Corollaries 1.6 and 3.3]. If one shift is zero, its fixed factor $s_p(0) = -p^{-1/2}$ gives an extra $p^{-1/2}$ saving for the remaining distinct nonzero shifts in the two cases of the proposition. The difference between the completed and uncompleted averages also has the harmless rescaling factor $p/(p-1)$. The older one-slope mixed moments and their curve counts in [48] serve as a comparator for the vertical terms. The additional translation in (51) gives the affine pattern dependence.

E.4 Sharp averaged character correlation

Use the same Sato–Tate characters χ_k as in [44, Lemma 6.3], normalized to be orthonormal under P. For positive degrees $k_1, \dots, k_m$ define

$$M_{\boldsymbol{k}}(\boldsymbol{h}) = \frac{1}{|\mathcal{G}|} \sum_{\gamma \in \mathcal{G}} \prod_{i=1}^{m} \chi_{k_i}(s_p(\gamma h_i)).$$

Here $\boldsymbol{h}$ is an ordered pattern of distinct sites. In the next theorem it is sampled uniformly without replacement.

Theorem E.3. *Fix $A, C > 0$. Uniformly for $5 \le m \le C \log p$ and $1 \le k_i \le (\log p)^A$,*

$$\mathbb{E}_{\boldsymbol{h}} M_{\boldsymbol{k}}(\boldsymbol{h})^2 = \frac{1 + o(1)}{p(p-1)}.$$

The averaging is over full uniformly sampled patterns. A fixed arithmetic progression window is a separate family of patterns.

Proof. Write $L = 1 + \max_i k_i$. On taking the square and fixing the relative map σ of the two affine rows, the average becomes

$$\frac{1}{|\mathcal{G}|} \sum_{\sigma \in \mathcal{G}} \mathbb{E}_{\boldsymbol{h} \text{ distinct}} \prod_{i=1}^{m} f_{i,\sigma}(h_i), \qquad f_{i,\sigma}(x) = \chi_{k_i}(s_p(x)) \chi_{k_i}(s_p(\sigma x)).$$

Each function is bounded by L^2. For nonidentity σ,

$$|\mathbb{E}_x f_{i,\sigma}(x)| \le C L^2 p^{-1/2}. \tag{52}$$

The rank-two Kloosterman sheaf has trivial projective automorphism group and SL_2 monodromy [47, §3]. Its two distinct affine pullbacks thus have product geometric monodromy on their common

nonsingular open set. Their nontrivial symmetric powers in the two factors have no tensor-product invariants. The test has rank at most L^2, tame finite ramification at at most two points, and maximum upper break at infinity at most $1/2$: affine changes preserve the break bound, and tensor products and symmetric powers preserve triviality of the inertia groups above that break. The same trace and conductor argument as in [44, Lemma 6.3] gives $O(L^2\sqrt{p})$ for the complete sum, including the $O(L^2)$ exceptional entries. On the identity map, the mean is $1 + O(L^2/\sqrt{p})$, by expanding $\chi_{k_i}^2 = \sum_{j=0}^{k_i} \chi_{2j}$ and applying the one-site moment bound. For raw characters, (52) also has the exact elementary formula

$$\mathbb{E}_x s_p(x)s_p(vx+w) = p^{-1}\sum_{z\neq 0} e_p\big(wz + (1-v^{-1})/z\big)$$

for $\sigma x = vx + w$, by Fourier orthogonality. On the identity it is $1 - 1/p$.

For a nonidentity relative map, use inclusion and exclusion on coincidences among the m coordinates. The average over distinct ordered coordinates is

$$\frac{p^m}{(p)_m}\sum_{\pi}\mu(\pi)p^{|\pi|-m}\prod_{B\in\pi}\mathbb{E}_x\prod_{i\in B}f_{i,\sigma}(x), \qquad \mu(\pi) = \prod_{B\in\pi}(-1)^{|B|-1}(|B|-1)!,$$

where π runs over set partitions and $(p)_m = p(p-1)\cdots(p-m+1)$. The factor $p^m/(p)_m$ is bounded for the stated m. A singleton block supplies $p^{-1/2}$ by (52); a size-r block for $r \geq 2$ supplies p^{1-r} by its coincidence constraint and a supremum bound. Since $r - 1 \geq r/2$, the total factor in p is bounded by $p^{-m/2}$. The sum of the absolute Möbius coefficients equals $m!$ (count permutations by cycle partitions), so the total is at most $(CL^2)^m m! p^{-m/2} = o(p^{-2})$ uniformly for $m \geq 5$ in the stated range.

For the identity map all factors are nonnegative. Without the distinctness constraint the value is the product of the one-coordinate means, $1 + o(1)$. Reweight each coordinate's uniform law by its nonnegative factor divided by its mean. Every resulting atom has mass at most CL^2/p, so the probability of a collision under the reweighted independent sample is $O(m^2L^2/p)$. Conditioning on distinctness therefore preserves the value $1 + o(1)$. This single term is weighted by $1/[p(p-1)]$ in the relative-map average, proving the result. □

E.5 A random labeling of the genuine arithmetic seed

The physical frequencies can also be randomly assigned to the genuine arithmetic orbit, while leaving every seed value and using the same assignment in every row.

Theorem E.4. *Let π be a uniform permutation of $\mathbb{F}_p$, and take the $p(p-1)$ paths with coordinates $w_h^\gamma = s_p(\gamma\pi(h))$, using the same π throughout. Write μ_p^π for the uniform law on this orbit and $R = \log[p(p-1)]$. Then*

$$R\,\mathbb{E}_\pi W_2^2(\mu_p^\pi, \nu) \longrightarrow C_B, \qquad RW_2^2(\mu_p^\pi, \nu) \longrightarrow C_B \quad \textit{in probability.}$$

All path labels are distinct when $p \geq 5$.

Proof. Fix an equal-interval binning of $[-2, 2]$ of mesh $\Delta = (\log p)^{-D}$ with $D > 2$, and denote the Sato–Tate bin probabilities by q_i. Their minimum is bounded below by $c\Delta^{3/2}$. For one row, the histogram of its seed values has pointwise bin probabilities $q_i(1 + O_K((\log p)^{-K}))$ for every fixed K. For two distinct affine rows, the joint histogram after changing the seed variable by the first row is that of $(s_p(x), s_p(\sigma x))$ for a nonidentity σ. It has joint bin probabilities $q_iq_j(1 + O_K((\log p)^{-K}))$

uniformly over σ. To verify these estimates, approximate bin indicators from above and below by Lipschitz functions with collars of inverse-polylogarithmic width, then approximate those functions uniformly by Bernstein polynomials of sufficiently high polylogarithmic degrees. The reference collar mass is at most a constant times its width. One-variable basis coefficients of the bounded polynomials are bounded by one in the orthonormal χ_k basis. The product monodromy argument for (52) also applies to distinct degrees, including degree zero in one factor, and bounds the error in their expectation by $p^{-1/2}$ times a fixed power of $\log p$. Both degrees zero have exact expectation. Absolute errors of arbitrary inverse-logarithmic order follow, and dividing by the minimum bin masses gives the relative estimates.

On a retained window $H = O(\log p)$, the random π samples these histograms without replacement, jointly for each fixed pair of distinct rows. Each required cell has at least $p(\log p)^{-C}$ initial entries, and at most $|H|$ entries are used previously. Multiplying the conditional cell comparisons gives relative error $O_K((\log p)^{-K})$ for every complete one-row or pair pattern, versus the independent Sato–Tate label vectors. This statement is uniform over the row pair. For the likelihood construction average $k_0(x, y)$ of Section 2 over the ideal second vector within its fine bins on H. This kernel still has the cap $e^{(1-\delta)R}$, row and column integrals at most one, and exponentially small lost mass under the ideal coupling. The one-row relative error gives the same source bias up to $O_K((\log p)^{-K})$. Distinct row pairs have covariance at most $O_K((\log p)^{-K})m(x)^2$, by the two-row relative bounds; the diagonal contribution is at most $Ce^{(1-\delta)R}/|\mathcal{G}|$. Thus the integrated expected source deficit is smaller than any fixed inverse power of $\log p$.

Lift the retained subcoupling through the ideal second vector with the observed fine bins. The source-to-ideal output mean cost is at most $(1+o(1))(C_B+\eta)/R$ by the one-row relative comparison. In retained modes the mesh Δ gives ordinary path difference at most $C\Delta(1+\log|H|)$. The ideal high modes contribute mean squared $O(1/(TR))$ by independence of the clipped coupling. A genuine high row is a sampling without replacement from the bounded seed. Its periodic squared cost is deterministically $O(1/(TR))$, and its high pinned constant has variance at most $C\sum_{|h|>TR} h^{-2}$ and zero mean, since the harmonic weights sum to zero on symmetric slots. The retained row mass is at most $1/|\mathcal{G}|$, so the same bound applies under the partial coupling. Repair costs are negligible with the arbitrary inverse-logarithmic source deficit, as in Section 2. Taking T large and η small yields the expected upper bound.

For the lower bound, for every fixed π the completed array has uniform one- and two-shift character comparisons for fixed degrees, since the shifts $\pi(h_i)$ are distinct and the translated trace estimate applies before averaging the nonzero slope. The uniform ordinary-path converse of Section C, using only these moment comparisons and the support bound, shows $\liminf RW_2^2 \geq C_B$ uniformly in π. Together with the expected upper bound this also gives convergence in probability. Finally, s_p is injective for $p \geq 5$. Equality of two raw Kloosterman sums implies equality of their phase histograms: their difference polynomial in $e^{2\pi i/p}$ is a multiple of $1 + z + \cdots + z^{p-1}$ with coefficient sum zero. The second moment of the histogram in $\mathbb{F}_p$ equals $-2a$ for the phase $x + a/x$, since $\sum_{x\neq 0} x^{\pm 2} = 0$. Hence the two seed inputs are equal. Equality of paths would force equality of their endpoint and all Fourier coefficients, then injectivity and the permutation would force equality of the affine maps. This proves distinctness. □

E.6 The short-window covariance needed for the arithmetic constant

Here the physical frequencies remain unpermuted. Use the likelihood of Section 2, with rate $R = \log|\mathcal{G}|$, and let $w_h^\gamma = s_p(\gamma h)$. Define

$$J_\sigma = \mathbb{E}_\gamma \int k_0(x, w^\gamma) k_0(x, w^{\gamma\sigma})\, dP^\infty(x).$$

Let J_σ^* denote the same product integral when the variables at distinct positions of $H \cup \sigma H$ are independent Sato–Tate values and shared positions use the same value. This expectation is determined by σ and the two retained windows.

Proposition E.5. *Suppose that for every positive distortion slack the corresponding kernels of Section 2 can be chosen so that*

$$\frac{1}{|\mathcal{G}|} \sum_{\sigma:\, |H \cap \sigma H| \le 1} (J_\sigma - J_\sigma^*) \le o(R^{-6}). \tag{53}$$

Then for the actual uncompleted prime Kloosterman path law μ_p,

$$(\log p) W_2^2(\mu_p, \nu) \longrightarrow C_B/2.$$

Proof. Let $S(x) = |\mathcal{G}|^{-1} \sum_\gamma k_0(x, w^\gamma)$. Its squared integral is the average of J_σ. The affine incidence counts of Section 2 control the heavy-overlap part: every J_σ is at most $e^{(1-\delta)R}$ because the row integral of $k_0(x, w^\gamma)$ is at most one. At most $|H|^4$ maps have two overlaps. Disjoint J_σ^* is at most one by the ideal row integrals. When there is exactly one overlap, integrating the other coordinates gives at most a fixed power of R by the smallest cell probability, as before. The number of these maps is $O(p|H|^2)$. Thus (53) implies $\int S^2 \le 1 + o(R^{-6})$.

We also have $1 - \int S = o(R^{-6})$ by the marginal type control in the planted construction. Here are details for the arithmetic use. On the head and endpoint the information is bounded deterministically by the total log inverse probabilities of their chosen cells, which is $O(\epsilon R(1 + |\log \epsilon|) + \log R)$ for sufficiently small head cutoff ϵ. In central bands the contribution depends on label counts and is within its typical value up to small fixed relative slack if every cell frequency is within a fixed tolerance. This follows from the constant-composition count formula in Section C. For the complete arithmetic row, the failure fraction in any of these fixed finitely many tests is $O_K(R^{-K})$ for arbitrarily large fixed K. For each fixed even moment $2K$ of a centered frequency-band sum, expand by distinct slots. Approximate fixed cell indicators at inverse-polylogarithmic boundary collars by bounded Bernstein polynomials of sufficiently high polylogarithmic degree as in Section 5. The one-site bound controls the collar masses and the polylogarithmically bounded rank-weighted character norms give the independent comparison for any fixed number of distinct slots by [44, Lemma 6.3]. With sufficiently small approximation error the normalized moment is $O_K(R^{-K})$ by the independent moment bound. Moment Markov gives the stated failure fraction. Repeated slot indices in the expansion are handled by combining their bounded factors before the same fixed degree comparison. For typical output rows the planted conditional likelihood is below the cap, because its central log density within each band is determined by the common label type and the selected joint counts. The coupling has fixed information slack, and the head and endpoint bounds fit inside it. Hence the source total mass loss is at most the atypical row fraction, which can be made $o(R^{-6})$.

It follows that $\int (S-1)^2 = o(R^{-6})$. Trim and fill the partial transport as in Section 2. The missing mass is $o(R^{-3})$, so the squared fill cost is $o(R^{-1})$, including the deterministic $O(\log p)$ target path norm.

We verify that the retained squared cost has the required ordinary-path bound. The periodic weighted coefficient errors in the head are $O(\epsilon/R)$ by same-cell matching, and the endpoint is matched in a cell of mesh R^{-2}. In a central band, given an output row of typical label histogram, the coupling draws the input by the prescribed joint counts with equal marginals, then independently inside cells. The empirical pair counts give the same limiting central periodic cost as the reference by continuity of the finite-alphabet distortion. Error from replacing values by cell means tends to

zero with the fixed cell mesh. Atypical rows can be discarded at negligible cost by the preceding mass bound. Source high coefficients are independent and the target weighted periodic tail is deterministically $O(1/(TR))$.

To control the pinned constant, consider one central band of m_0 slots with weights $v_h = 1/h$ of oscillation $O(\kappa/U_0)$, where the band's relative width is κ and its magnitudes are comparable to U_0. Conditional on the arithmetic row's bin labels, the source labels are drawn by uniform allocations within each row-label class with prescribed counts. The residual permutation noise in the pinned sum has variance at most $Cm_0 \operatorname{osc}(v)^2$ by the finite-population variance formula, and the continuous within-cell residual contributes at most $C\Delta^2 \sum v_h^2$. The conditional mean of the between-cell component has the form

$$\sum_j b_j \sum_{h \text{ in band}} (v_h - \bar{v}) \mathbf{1}_{\{w_h \in \text{cell } j\}}, \qquad |b_j| \le 4.$$

Subtracting $\bar{v}$ is valid because the two label histograms coincide. Averaging over the affine orbit, each inner sum has zero mean and variance at most $Cm_0 \operatorname{osc}(v)^2$, by the exact two-transitivity of the affine group and the usual variance for sampling without replacement. There are a fixed finite number of cells when the mesh is fixed, so the full conditional mean is controlled in squared expectation by a fixed multiple of this bound. One may take the same central partition in every band. The remaining mean within-cell function $\bar{y}_j - y$ has range $O(\Delta)$ and Sato–Tate mean zero, hence its arithmetic scalar mean is smaller than any fixed inverse logarithm by the same bin smoothing and one-site argument. Its weighted contribution has variance $O(\Delta^2 \sum v_h^2)$ by two-transitivity. These estimates bound the central pinning error by an arbitrarily small multiple of $R^{-1/2}$ in root mean square, by combining the within-cell variances across all bands first, taking a fine common partition, and then choosing the relative widths small for that fixed partition. For instance the sum over relative bands of $\sqrt{m_0} \operatorname{osc}(v)$ is $O(\sqrt{\kappa}/\sqrt{\epsilon R})$ for fixed lower and upper cutoffs and fixed central partition.

In the head the cell means minus row values have slotwise range $O(|h|/R)$ and Sato–Tate mean zero. Their arithmetic means are negligible after the uniform polylogarithmic polynomial approximation. For two different sites the covariance of any such slot functions evaluated on the uniform affine row is bounded by $O(1/p)$ times the product of their ranges, since the sites form a uniformly ordered distinct pair of seed positions. Thus the head harmonic contribution has squared mean $O(\epsilon/R)$ and negligible off-diagonal error. The independent conditional source residual has the same bound by summing squared ranges. Finally the high signed weights have zero sum and the exact affine two-transitivity controls their squared mean by $O(1/(TR))$, also for the arithmetic pinning. Clipping only decreases integrated nonnegative squared costs. With cutoffs chosen to pay the slack in the companion coupling, the limsup of the retained root mean square cost multiplied by $\sqrt{R}$ is thus at most $\sqrt{C_B + \eta}$ for arbitrary $\eta > 0$.

To pass to the usual path mesh, its nonzero unpinned Fourier coefficients after division by $2\pi i h$ are $c_{p,h} w_{\bar{h}}^{\gamma}$, where $c_{p,h} = e^{\pi i h/p} \operatorname{sinc}(\pi h/p)$ and $\bar{h}$ is the residue of h. The endpoint is w_0^{γ}. The periodic squared difference from the truncated path above is $O(1/p)$ since $|c_{p,h} - 1| \le C|h|/p$ on H_p, and the squared tail weights are $O(1/p)$. In the pinned constant the difference is a linear combination of residue coefficients with weights $O(1/p)$ each: the main-frequency correction has this size and the aliases sum $O(p \sum_{k \ge 1} (kp)^{-2}) = O(1/p)$, using $|c_{p,h}| \le Cp/|h|$. The seed has mean zero and bounded second moment. Two-transitivity gives variance at most a constant times the sum of squared weights, $O(1/p)$. Thus ordinary mesh completion costs $O(p^{-1/2})$ in W_2 on the complete family. Its conversion to the original linked mesh is the $O(p^{-1/2})$ time-change estimate of [44]. Removing the $a = 0$ rows costs $O((\log p)/\sqrt{p})$ because they have fraction $1/p$ and the deterministic path norm is $O(\log p)$. All these costs are negligible on the logarithmic scale. The opposite inequal-

ity follows from the arithmetic lower bound in Section C, which gives $\liminf(\log p)W_2^2 \geq C_B/2$. This proves the conditional assertion. □

Absolute character control and its coefficient cost. A constraint in trying to prove (53) by absolute termwise bounds can be made explicit. Work on the finite window and put $Q = P^H$. In the product orthonormal character basis (ϕ_α) of $L^2(Q)$, the positive operator with kernel $J(y,z) = \int k_0(x,y)k_0(x,z)\,dQ(x)$ has diagonal coefficients $j_{\alpha\alpha} = \|\int k_0(\cdot,y)\phi_\alpha(y)dQ(y)\|_2^2 \geq 0$ and trace $\sum_\alpha j_{\alpha\alpha} = \iint k_0^2$. The constant coefficient is at most one. If planted mass bounded below lies where $k_0 \geq e^{(1-\eta)R}$, then the nonconstant diagonal trace is at least $ce^{(1-\eta)R} - 1$. In the preserving construction, for any fixed small η one can have such a band by choosing the rate slack, cutoffs and central band discretizations sufficiently small: the summed central typical information tends to the corresponding scalar preserving integral, and the lost low and tail information can be made arbitrarily small. Equivalently it is sufficient in this assertion to impose the displayed planted mass hypothesis directly.

In particular the absolute coefficient budget of this complete kernel expansion is at least of order $p^{2-2\eta}$ even in unweighted orthonormal characters. An argument multiplying every coefficient by the same absolute discrepancy bound $p^{-\gamma}$ cannot make the resulting absolute budget small for $\gamma < 2 - 2\eta$. This is a restriction on that termwise argument, including expansions that retain this trace. It does not give a lower bound for arbitrary approximants that discard the trace or for the signed arithmetic covariance itself. The exact resonances and the sampled-pattern average above show why the short-window covariance problem requires its own cancellation estimate.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, independent internal checks and drafting. All main results are analytic.

F Empirical transport excess for unequal random Fourier laws

We determine the leading excess over the population squared Wasserstein cost when compact independent-coordinate laws with unequal coefficient distributions are replaced by equal-weight empirical measures. The rate and coefficient use a scalar two-marginal information objective applied to the slack in ordinary optimal transport. We apply the formula to periodic Fourier projections of Kloosterman paths at a fixed prime, where two residue gates give an explicit population cost and an excess coefficient depending on their overlap. The same overlap gives a quantitative lower bound when one of the empirical laws is replaced by the actual arithmetic family.

Keywords: empirical Wasserstein distance, random Fourier series, rate-distortion, Kloosterman paths.
MSC 2020: 60B10, 60G17, 94A29, 11L05.

F.1 Population transport and its empirical excess

Consider two independent-coordinate laws μ and ν on the same Hilbert sequence space, with realizations of the form

$$X = (\sqrt{\lambda_j}x_j)_{j\geq 1}, \qquad Y = (\sqrt{\kappa_j}y_j)_{j\geq 1}.$$

Here each x_j, y_j is in a fixed real Euclidean space, and the marginal laws of the coordinates are drawn from a finite collection of pairs (P_i, Q_i). The indices of each pair type have asymptotic

density θ_i. Vectors within each law have mutually independent coordinates. Coefficient supports are compact, and the positive weights (λ_j) and (κ_j) are summable. Put $w_j = \sqrt{\lambda_j \kappa_j}$ and suppose $w_j \sim \phi(j)$ for a function ϕ regularly varying with index $-b$, $b > 1$.

For a type (P_i, Q_i) let u_i, v_i be scalar Kantorovich potentials on the two supports for the continuous cost $c(x, y) = -2\langle x, y\rangle$. They can be chosen bounded and Lipschitz by the compact-space duality and the cost-transform formula [49]. Write

$$s_i(x, y) = c(x, y) - u_i(x) - v_i(y) \geq 0.$$

This slack vanishes almost surely under some optimal pair. Define the scalar preserving objective and its accumulated constant by

$$B_i(a) = \inf_{\pi \in \Pi(P_i, Q_i)} \left[I_\pi(x; y) + a\mathbb{E}_\pi s_i(x, y) \right], \qquad J = \sum_i \theta_i \int_0^\infty B_i(t^{-b})\, dt, \qquad C_b(J) = \frac{(b-1)^{b-1}}{b^b} J^b.$$

The scalar objective is independent of the chosen optimal dual pair, since the expected potentials are fixed by the two marginals.

Theorem F.1. *Assume $J > 0$. Let ν_N be the equal-weight empirical law from N iid samples of ν. Then*

$$W_2^2(\mu, \nu_N) - W_2^2(\mu, \nu) \sim C_b(J) R \phi(R), \qquad R = \log N$$

almost surely and in expectation. If both laws are replaced by independent empirical samples of sizes N, M, then the same result holds with $R = \log \min(N, M)$, almost surely for nested clouds jointly over all pairs of sizes with $\min(N, M) \to \infty$. The one-sample result is also valid with the first law sampled instead.

This is the unequal-law counterpart to the independent-coordinate equal-mass estimates of Section C and the Gaussian cross-law excess formula of Section G. Earlier constrained-output rate-distortion results, such as [50], use fixed-block source coding with stationary marginal laws. The distinction here is that a decaying spectrum yields a weighted excess over population transport, with both true or empirical target marginals specified. The proof directly uses the slack of the ordinary population optimum.

F.2 A slack decomposition and scalar couplings

For any coupling of the two product laws, the squared Hilbert cost decomposes into a sum of two marginal terms and a nonnegative slack:

$$\|X - Y\|^2 = A(X) + D(Y) + \sum_j w_j s_{\iota(j)}(x_j, y_j), \tag{54}$$

$$A(X) = \sum_j [\lambda_j |x_j|^2 + w_j u_{\iota(j)}(x_j)], \tag{55}$$

$$D(Y) = \sum_j [\kappa_j |y_j|^2 + w_j v_{\iota(j)}(y_j)]. \tag{56}$$

All series converge absolutely and uniformly on the marginal supports. The population optimum is $\mathbb{E}A(X) + \mathbb{E}D(Y)$, since independent optimal pairs across coordinates make the total slack zero. For empirical marginals the analogous identity holds for their optimal cost, with empirical averages of A, D plus the minimum mean slack over couplings. The two centered marginal corrections are

negligible compared with any fixed power of $1/\log N$, by bounded-variable concentration. This is true in mean absolute value and almost surely along nested samples, and likewise for M.

Fix a type, suppress its index, and write $k_a(x,y) = e^{-as(x,y)}$. This kernel is positive and continuous on the two compact supports. It has a two-sided scaling giving a coupling

$$d\pi_a(x,y) = e^{f_a(x)+g_a(y)-as(x,y)}dP(x)dQ(y), \qquad B(a) = \mathbb{E}_P f_a + \mathbb{E}_Q g_a. \tag{57}$$

To construct it, start from a positive continuous normalized function u on the first support, put $v = (k_a^* u)^{-1}$ on the second, and map u to the normalized function $(k_a v)^{-1}/\int (k_a v)^{-1} dP$. Positivity and uniform continuity of the kernel show that the image is equicontinuous and uniformly bounded above and below. Indeed, if $k_a \in [\zeta, 1]$ and $\int u\, dP = 1$, then $v \in [1, 1/\zeta]$, so $k_a v \in [\zeta, 1/\zeta]$ and the normalized map takes values in $[\zeta^2, \zeta^{-2}]$. Schauder's theorem gives a fixed point on this convex closed normalized set. The column constraint $v k_a^* u = 1$ implies $\int u k_a v\, dP = 1$. The fixed point of the normalized map then already obeys the row constraint $u k_a v = 1$. Let $f_a = \log u$ and $g_a = \log v$. Equation (57) minimizes the penalized cost by the relative entropy identity, and the minimum is its displayed marginal mean. The two tilts are bounded uniformly for penalties in a compact interval, with the chosen normalization. The identities $f_a = -\log(k_a v)$ and $g_a = -\log(k_a^* u)$ imply that the Lipschitz constant of each tilt is at most a times the Lipschitz constant of s in the corresponding variable. This also bounds the information density there. At differentiability points the slack distortion is $B'(a)$ and the information is $B(a) - aB'(a)$.

At small a the independent pair bounds $B(a)$ by $O(a)$. For large a, take an original optimal pair for the scalar squared cost, partition the two supports into cells of diameter $O(a^{-1})$, and retain its joint cell probabilities p_{ij}. Conditional on the cell pair, sample the two values independently from their corresponding marginal cell conditionals. The marginals are unchanged, and the slack is $O(a^{-1})$ everywhere on the resulting support by its Lipschitz bound and zero slack in every used optimal cell pair. Its information is at most the logarithm of the total number of cell pairs, which is $O(1+\log(1+a))$. More precisely the likelihood ratio in these cells is $p_{ij}/(p_i q_j)$ and for $0 < t < 1$,

$$\mathbb{E}e^{ti(x,y)} = \sum_{ij} \frac{p_{ij}^{1+t}}{p_i^t q_j^t} \le \sum_{ij} p_{ij}^{1-t} \le M^t, \tag{58}$$

where M is the number of possible cell pairs. This proves the logarithmic bound and supplies a tail estimate at low coordinates. The function B is concave and increasing. The integrals are finite, and integration by parts gives

$$\sum_i \theta_i \int_0^\infty t^{-b} B_i'(t^{-b})\, dt = J/b, qquad \sum_i \theta_i \int_0^\infty [B_i(t^{-b}) - t^{-b} B_i'(t^{-b})]\, dt = (1-1/b)J.$$

For classification, B_i vanishes identically exactly when

$$\text{span}(\text{supp}\, P_i - \text{supp}\, P_i) \quad \text{is orthogonal to} \quad \text{span}(\text{supp}\, Q_i - \text{supp}\, Q_i).$$

Indeed zero objective at positive a forces independence and zero slack by the attained minimum. The continuous slack then vanishes on the full support product. The rectangular differences of the cost are $-2\langle x-x', y-y'\rangle$, so this is equivalent to the stated orthogonality. Conversely orthogonality makes the scalar cost additive on the support product, and the independent coupling is optimal with zero information.

F.3 The exact weighted slack cost

The proof concerns the infimum over couplings of the weighted slack in the decomposition above. Put $A_0 = (1 - 1/b)J$ and temporarily use $a_j = w_j/\phi(L)$. Central bands are $\epsilon L \le j \le TL$ for fixed $0 < \epsilon < T < \infty$. Uniform regular variation and the type densities imply the corresponding central Riemann sums, as in Sections C, G. For the derivatives of B_i one may partition the band into small subintervals and choose a differentiability point for each type in each interval, because these derivatives are monotone.

For the lower bound let $L = R/A_0$. On the central band write $F(x) = \sum_j f_{a_j}(x_j)$ and $G(y) = \sum_j g_{a_j}(y_j)$, where the potentials include their slot type. A point in either law is typical if its sum is within ηL of its expectation. Each atypical marginal mass is exponentially small, by independence and bounded central potentials. By choosing ϵ small and T large the sum of the two expectations is at least $(J - \eta)L$ for large L. For a fixed typical y, integrating the product of (57) against the first marginal gives one. Hence the probability that X is typical and its central weighted slack to this y is at most $(J/b - 4\eta)L\phi(L)$ is at most $\exp(-R - \eta L)$. The estimate is symmetric for a typical fixed x.

If one cloud is empirical, the union over its typical points gives an exponentially small bound on the true marginal mass that can be coupled at this low cost. The sampled marginal's atypical mass is exponentially small in expectation. If both clouds are empirical, the expected number of cheap typical pairs is at most $NM \exp(-R - \eta L)$. In a coupling each pair can carry at most $1/\max(N, M)$, so the total expected mass below this cost has the same exponentially small bound, apart from the two atypical marginal masses. Positivity of the slack gives the moment lower bound.

For the upper bound take $L = R/(A_0 + \gamma)$ with fixed $\gamma > 0$. Construct a joint law with both exact product marginals. On each short central subinterval use the independent optimal tilt (57) at a representative penalty for each positive-density type. Zero-density types can be sampled independently there at $o(L\phi(L))$ slack. For $j < \epsilon L$ use the cell-pair construction with mesh $\min(1, \phi(L)/w_j)$. Its deterministic slack contribution is $O(\epsilon L\phi(L))$. Potter's bound shows that the sum of the logarithms of the numbers of cell pairs is $O(\epsilon L(1 + \log(1/\epsilon))) + o(L)$. Equation (58) gives the same bound for the head information up to any fixed positive multiple of L, except with exponential probability. For $j > TL$, use independent pairs and the bounded slack to pay at most $O(T^{1-b}L\phi(L))$ in the limit, with zero information.

The central channel information and expected normalized slack tend to the corresponding truncated integrals. In particular, with sufficiently small mesh in the central band and after selecting the cutoffs, the joint information density exceeds $R - c_\gamma L$ only with probability $e^{-c'L}$. This follows from independence, the bounded central likelihoods, and the head estimate. Its mean weighted slack is at most $(J/b + O(\gamma))L\phi(L)$, by the allocation integrals. There is no likelihood contribution from the far tail. Use the one- or two-sample soft-covering coupling of Section G: truncate the joint likelihood at $\exp(R - c_\gamma L)$, use it as the density of a partial coupling, and trim and fill the marginal deficits. In the two-sample case each row and column averages the corresponding independent cloud. The expected sum of absolute row and column errors is exponentially small, by the conditional variance bound and the omitted likelihood probability. In the one-sample case take the integral version on the unsampled side. Boundedness of the full weighted slack makes the cost of filling the remaining mass exponentially small. Thus the upper bound matches the lower as $\gamma, \eta \downarrow 0$ and $L\phi(L)/(R\phi(R)) \to A_0^{b-1}$. This proves the expected optimum slack asymptotic.

For the almost-sure version, changing one sample changes the minimum slack by at most its global bound divided by the sample size, by rerouting that atom's mass. Thus bounded differences gives summable deviation probabilities for the one-sample limit and for the two-sample limit throughout the ranges $m \le n \le \exp((\log m)^K)$, where m, n are the smaller and larger sample sizes

and $K > 1$ is fixed. The union bound over these ranges is valid since their number of sizes is subexponential in m, whereas the concentration exponent is of order m times a negative power of $\log m$.

For more disparate sizes use the slack estimate rather than the original quadratic transport metric. Because each scalar slack is uniformly Lipschitz, changing the second marginal from ν_M to ν changes the optimal mean slack by at most a constant times the W_1 distance between their laws in coordinates $(\sqrt{w_j}y_j)_j$. This follows by gluing and Cauchy–Schwarz, since $\sum_j w_j < \infty$. The same holds on the first side. The empirical same-law theorem of Section C gives an almost-sure bound of order $[\log n\, \phi(\log n)]^{1/2}$ for this correction. Indeed $J > 0$ forces a positive-density type with both marginals nonconstant, by the orthogonality criterion. For $\log n > (\log m)^K$ with K sufficiently large, Potter's bound makes this error $o(\log m\, \phi(\log m))$. The one-sample limit therefore implies the all-size two-sample limit. The centered marginal terms in the slack decomposition are negligible in the same ranges by their concentration. This proves Theorem F.1.

F.4 Two residue gates at a fixed prime

We apply the theorem to the periodic Fourier component of the classical Kloosterman path at a fixed odd prime p. The Hilbert metric in this section is explicitly on the nonzero Fourier component after subtracting the endpoint trend. For a continuous path f starting at zero put

$$\mathcal{P}f(t) = f(t) - tf(1) - \int_0^1 (f(v) - vf(1))\, dv.$$

On the Fourier path model this is the periodic series with coefficients $X_h/(2\pi ih)$ at $h \neq 0$. In particular the squared metric is the sum of squared nonzero Fourier differences. It is the ordinary L^2 metric on these projected paths.

Let χ be the quadratic character mod p. For gate i prescribe $b_i \in \mathbb{F}_p^\times$ and an inverse parameter with quadratic sign $\epsilon_i \in \{\pm 1\}$. In the pinned limit the coefficient at h is active precisely when $\chi(b_i - h) = \epsilon_i$. Active coefficients have the arcsine law P_{A} of $2\cos(2\pi U)$ for uniform $U \in [0,1]$, and inactive coefficients are zero. The coefficients are mutually independent for the limit. This is the classical fixed-prime law [51] in translated-coordinate notation. Let ν_1, ν_2 here denote the corresponding projected laws. If $\mathcal{M} \subset \mathbb{F}_p$ is the set on which the two gates disagree, the population squared distance is

$$d_0^2 = W_2^2(\nu_1, \nu_2) = \frac{1}{2p^2} \sum_{r \in \mathcal{M}\setminus\{0\}} \csc^2(\pi r/p) + \frac{\mathbf{1}_{\{0\in\mathcal{M}\}}}{6p^2}. \tag{59}$$

Indeed the one-coordinate squared cost is zero for a matched type and is the arcsine second moment 2 for a mismatched coordinate before Fourier weighting. Summing $(2\pi h)^{-2}$ over each residue gives (59), using the usual partial fractions for $\csc^2$ and omitting $h = 0$.

Let α be the fraction of residues active in both gates. For $b_1 \neq b_2$ one has

$$p\alpha = \frac{1}{4}\big[p - 2 - \epsilon_1\chi(b_1 - b_2) - \epsilon_2\chi(b_2 - b_1) - \epsilon_1\epsilon_2\big]. \tag{60}$$

To check this, sum $(1 + \epsilon_1\chi(b_1 - h))(1 + \epsilon_2\chi(b_2 - h))/4$ over $h \neq b_1, b_2$. The two linear sums omit exactly one nonzero term each, and the quadratic character sum of the two distinct linear factors is -1. If $b_1 = b_2$, the overlap is $(p-1)/2$ when the signs agree and zero otherwise.

On both-active slots use slack $(x - y)^2$ for $P_{\mathrm{A}}, P_{\mathrm{A}}$, by taking the potentials $-x^2, -y^2$ for the cross cost. On all other slot types the slack is identically zero, with potentials zero. Write $J_{\mathrm{A}} =$

$\int_0^\infty B_{P_A}(u^{-2})\,du$, where B_{P_A} is the scalar preserving objective with squared difference. Ordered by $|h|$ with both signs included, the weights are asymptotic to $1/(\pi^2 j^2)$. Thus if $\alpha > 0$, Theorem F.1 gives

$$W_2^2((\nu_1)_N, (\nu_2)'_M) - d_0^2 \sim \frac{\alpha^2 J_A^2}{4\pi^2 \log \min(N, M)}. \tag{61}$$

In a one-sample version use that sample's log size. If $\alpha = 0$, the slack is zero pointwise for these gates, so the excess is exactly the corresponding centered empirical second-moment correction. In particular it has zero expectation for every sample size. Thus the population mismatch profile and the leading empirical correction use different sets of residue frequencies.

F.5 An arithmetic excess lower bound

Keep the first projected law ν_1 and replace the second by an arithmetic depth-n path at the same prime in its residue class, with the inverse parameter fixed at any lift of its prescribed sign. Let its projected law be $\mu_{q,2}$ at modulus $q = p^n$.

Corollary F.2. *In the periodic Fourier metric just specified,*

$$\liminf_{n\to\infty} n[W_2^2(\nu_1, \mu_{p^n,2}) - d_0^2] \geq \frac{\alpha^2 J_A^2}{4\pi^2 \log p}.$$

If $\alpha = 0$ then $W_2^2(\nu_1, \mu_{p^n,2}) = d_0^2 + O_p(p^{-n/2})$.

Proof. Complete the path Fourier coefficients on the full mesh. At a nonsingular active frequency h the complete coefficient W_h is a signed cosine or sine of an angular root lift, with amplitude two. The signs and possible phase shift depend only on the base residue and parity; this follows from the tame stationary formula Section A. The chosen Hensel root, as the path parameter ranges uniformly over its fixed residue class, is uniform over the lifts of a unit root at that frequency. Thus $\mathbb{E}W_h^2 = 2$ exactly for every $n \geq 2$: the nonconstant double-angle term is a complete additive sum over a residue class with unit numerator. At inactive frequencies the coefficient vanishes.

Consider first the truncated arithmetic Fourier series with $W_h/(2\pi i h)$ at $0 < |h| \leq (q-1)/2$ and zero beyond. In the population slack decomposition the only variable marginal gauge on the arithmetic side occurs where its gate is active and the other is inactive. At those retained coordinates the expectation is exactly the population term just computed. On both-active slots its marginal gauge is identically zero. Omitted coordinates change the expected marginal gauge by $O(q^{-1})$ because the weights have a square-summable tail of that order. These facts show that the population cost d_0^2 is the correct baseline up to $O(q^{-1})$ in the truncated series. If no gates overlap the weighted slack vanishes identically as well.

Now suppose $\alpha > 0$ and run the lower proof for the central band with $R = (n-1)\log p$ and $J = \alpha J_A$. These shifts have $|h| = O(n)$ and are below the mesh cutoff. The source potentials are typical by independence. Uniform typicality of the arithmetic target potentials follows from the fixed one- and two-shift moment comparisons for root phases on a residue ball. More explicitly, the p-adic comparison of Sections A, A applies to two distinct active shifts of size $O(n)$ and fixed Fourier height, giving error $O_p(e^{-cn})$. The valuation losses from the two shift differences are $O_p(\log n)$. Approximate the bounded central potentials uniformly by trigonometric polynomials on the coefficient circles to any fixed tolerance, with height independent of n. The fixed-height result then gives the needed first and pair estimates, and the normalized central potential sum converges in probability to its marginal expectation. This argument is uniform in the arithmetic lifts and for both parity laws. Inactive or mismatched slots have zero slack, so need no comparison.

There are at most e^R arithmetic paths at this residue. The typical arithmetic points may therefore be inserted into the same tilted union bound as in Theorem F.1. Every coupling places all but $o(1)$ mass on typical targets and on typical sources. The coupled mass below the weighted slack threshold $(J/2-\eta)L\phi(L)$ tends to zero for $L = 2R/J$, after slack parameters are chosen small enough. This proves the asserted lower bound for the truncated Fourier law, using (61)'s normalization.

Finally, the full mesh gives the nonzero projected coefficient

$$\widehat{\mathcal{P}f}(h) = \frac{e^{\pi i h/q}\operatorname{sinc}(\pi h/q) W_{h \bmod q}}{2\pi i h}.$$

This follows by integrating the constant derivative of each spline segment. Replacing the factor by one on the retained range has L^2 error $O(q^{-1/2})$, by Parseval and the pointwise coefficient difference $O(q^{-1})$. The omitted Fourier tail has the same bound by summing h^{-2} and using $|W_h| \le 2$. The equal-unit mesh change costs another $O(q^{-1/2})$ in uniform norm before projection, since corresponding times differ by $O(q^{-1})$ and the increments have size $O(q^{-1/2})$. The projected path norm is uniformly bounded by Parseval on the displayed coefficients. Hence these changes alter squared Wasserstein cost by $O(q^{-1/2})$, negligible on the depth scale. This also proves the no-overlap statement. □

The Fourier projection is material here. Without removing the constant coefficient and endpoint trend the population optimum may couple different Fourier slots. The theorem above addresses the projected law, where the squared cost separates into the scalar gate contributions.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, internal independent checks and drafting. The proofs use no datasets.

G Sharp empirical Wasserstein constants for Gaussian measures on Hilbert space

We give an exact beta-function constant for empirical Wasserstein convergence of a Gaussian law on Hilbert space when its covariance eigenvalues are regularly varying. The result holds for every fixed finite Wasserstein order, for one empirical cloud or two independent clouds. For Brownian motion in $L^2[0,1]$, the constant multiplying $(\log N)^{-1/2}$ is $\mathrm{B}(3/4,1/2)/(\sqrt{2}\pi)$. An exact spectral identity also gives the leading empirical squared W_2 excess between different Gaussian laws without requiring their covariance operators to commute.

Keywords: Gaussian measures, empirical Wasserstein distance, Brownian motion, regular variation.
MSC 2020: 60B10, 60G15, 94A29.

G.1 The empirical constant

Let μ be a Gaussian probability measure on a real separable Hilbert space with positive covariance eigenvalues λ_j listed in decreasing order, with multiplicities. Suppose that $\lambda_j \sim \phi(j)$ for a positive regularly varying function ϕ of index $-b$, $b > 1$. Let μ_N be the uniform empirical measure of N independent samples from μ. Use μ'_M for an independent empirical measure of size M. Write

$$K_b = \frac{\mathrm{B}(1-1/(2b),1/b)^b}{2^{b-1}(b-1)}, \qquad \mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt.$$

Theorem G.1. *For each fixed $1 \le s < \infty$ let $R = \log N$ in the one-sample problem and $R = \log \min(N, M)$ in the two-sample problem. Then*

$$\left(\mathbb{E}W_s(\mu, \mu_N)^s\right)^{2/s} \sim K_b R\phi(R), \qquad \left(\mathbb{E}W_s(\mu_N, \mu'_M)^s\right)^{2/s} \sim K_b R\phi(R).$$

The normalized squared errors converge to K_b in probability. Almost-sure convergence holds for nested samples, in the two-sample case jointly over all pairs of sizes with $\min(N, M) \to \infty$.

The exact quadratic optimal finite-support constants for Gaussian Hilbert laws were found by Luschgy and Pagès [32, Theorem 2.2]. Nearest-neighbor errors of Gaussian random codebooks are related to randomly centered small balls by Dereich [53] and Dereich and Lifshits [52, Theorem 5.1]. Equal-weight empirical Wasserstein transport imposes a separate mass constraint. Mean empirical Wasserstein rates in infinite-dimensional Gaussian settings were studied by Boissard and Le Gouic [54] and by Lei [55, Section 4]. He and Yuan [41] and Coste [56] treat logarithmic order bounds for path measures under the supremum norm. The theorem here gives a leading constant for the Hilbert norm, using a coupling that preserves both marginals and a matching typicality converse.

G.2 Gaussian preserving couplings

Translation lets us center μ. In eigen-coordinates it is the law of $(\sqrt{\lambda_j} X_j)_{j \ge 1}$ for independent $X_j \sim N(0, 1)$. At a scalar penalty $a > 0$ define

$$B(a) = \inf\{I(X; Y) + a\mathbb{E}(X - Y)^2 : \ X, Y \sim N(0, 1)\}.$$

For fixed correlation ρ the joint Gaussian minimizes the mutual information by maximum entropy. Minimization gives

$$\frac{\rho}{1 - \rho^2} = 2a, \qquad B(a) = -\tfrac{1}{2}\log(1 - \rho^2) + 2a(1 - \rho), \qquad B'(a) = 2(1 - \rho). \tag{62}$$

This is the scalar Gaussian instance of entropy-regularized transport [57]. For the minimizing pair its density relative to two independent standard Gaussians is

$$e^{f_a(x) + f_a(y) - a(x-y)^2}, \qquad f_a(x) = -\tfrac{1}{4}\log(1 - \rho^2) + a(1 - \rho)x^2. \tag{63}$$

In particular $2\mathbb{E}f_a(X) = B(a)$, and integration of this density against either marginal gives one. The coefficient $a(1 - \rho) = \rho/(2(1 + \rho))$ is at most $1/4$ at every penalty.

There is uniform concentration even for growing penalties. Under the joint pair, $U = (X + Y)/\sqrt{2(1 + \rho)}$ and $V = (X - Y)/\sqrt{2(1 - \rho)}$ are independent standard Gaussians. The information density equals its mean $-\frac{1}{2}\log(1 - \rho^2)$ plus $(\rho/2)(U^2 - V^2)$. The normalized squared distortion is $a(X - Y)^2 = 2a(1 - \rho)V^2$. Thus both centered quantities have exponential moments on a fixed neighborhood of zero, with uniform bounds, even when a is large. On a central band of bounded penalties the same applies to the centered potential $f_a(X)$ in the marginal law.

Put $J = \int_0^\infty B(u^{-b})\, du$ and $A = (1 - 1/b)J$. The scalar bounds $B(a) = O(a)$ at zero and $B(a) = O(1 + \log(1 + a))$ at infinity make the integral finite. Integration by parts gives

$$\int_0^\infty u^{-b} B'(u^{-b})\, du = J/b, \qquad \int_0^\infty [B(u^{-b}) - u^{-b}B'(u^{-b})]\, du = A. \tag{64}$$

On the central band $\epsilon L \le j \le TL$, the penalties $a_j = \lambda_j/\phi(L)$ approach $(j/L)^{-b}$ uniformly, so the normalized sums converge to their central integrals. Potter's bound [35] gives

$$\sum_{j < \epsilon L} (1 + \log(1 + a_j)) = O(\epsilon L(1 + \log(1/\epsilon))) + o(L).$$

Indeed outside the finitely many initial indices $\log(1+a_j)$ is bounded by a constant times $1+\log(L/j)$. The initial indices total $O(\log L)$. This handles all low information terms, and the corresponding expected normalized distortion is $O(\epsilon L)$ by (62). At $j > TL$ independent pairs have total expected normalized squared distortion $O(T^{1-b}L)$ in the limit, by regular variation. Thus no separate low-resolution approximation of large Gaussian coordinates is needed.

G.3 Soft covering in one and two samples

We first record a general sample coupling bound. Let π have marginals μ, ν with finite $2s$th moments, and density k relative to $\mu \otimes \nu$. For independent empirical samples of sizes N, M put $R = \log \min(N, M)$ and truncate k to $k_0 = k\mathbf{1}_{\{\log k \le R-t\}}$. If $\zeta = \mathbb{P}_\pi(\log k > R - t)$, there is a coupling of the two empirical laws whose expected sth power cost is at most

$$\mathbb{E}_\pi \|X - Y\|^s + C_s(\zeta + e^{-t/2})^{1/2}, \tag{65}$$

where the constant may depend on the two marginal $2s$th moments. The same bound holds for one true marginal and an empirical sample on the other side, with R equal to its log size.

For two samples first form the partial matrix of masses $k_0(X_i, Y_j)/(NM)$. If $X_i = x$ is fixed, the row sum multiplied by N has conditional expectation $\mathbb{E}_Y k_0(x, Y) \le 1$ and variance at most e^{-t}, since $k_0 \le e^{R-t}$ and $M \ge e^R$. Averaging the absolute conditional bias over x gives ζ. Thus the expected sum of absolute row discrepancies is at most $\zeta + e^{-t/2}$. The columns have the same bound. Trim any excess row mass down to $1/N$, then any excess column mass down to $1/M$. The missing mass after both operations is at most the sum of the original absolute discrepancies. Fill it by coupling the two deficits, for instance by their normalized product. The retained matrix cost is no greater than that of the original matrix, with expected cost at most $\mathbb{E}_\pi \|X - Y\|^s$. If the fill mass is z, its cost is bounded by $C_s z^{1/2}$ times the square root of the sum of the empirical $2s$th moments, by Cauchy–Schwarz on the deficits. A second use of Cauchy–Schwarz in expectation gives (65). This reasoning treats the sample points with labels, so it is valid even if some sample values agree.

For one true marginal use the integral matrix $N^{-1}k_0(x, Y_i)d\mu(x)$ instead. The first marginal discrepancy is the usual mean soft-covering estimate, from the identical conditional variance and bias calculation. The second marginal has missing mass with expectation ζ and no excess. The same trim and fill argument applies. The relation to the usual one-sample soft-covering lemma is described in [40].

G.4 The exact limit

For the upper bound let $L = R/(A + \gamma)$, $\gamma > 0$. On $j \le TL$ take independent preserving Gaussian pairs (63) with penalties $a_j = \lambda_j/\phi(L)$. On the remaining coordinates take independent standard Gaussian pairs. The joint law has both product marginals and a density relative to their product, since only finitely many coordinates are coupled. The mean information is at most $(A + o(1))L$ as $L \to \infty$, up to an arbitrarily small prescribed multiple of L. This follows from the central Riemann sum and the low information bound above. Its fluctuation is a sum of independent uniformly subexponential random variables over $O_T(L)$ coordinates. The sample-density tail in (65) is therefore exponentially small for $t = c_\gamma L$ with c_γ small enough.

The mean normalized squared distortion, divided by $L\phi(L)$, has limit superior at most $J/b + O(T^{1-b})$. We take T arbitrarily large. For $j \le TL$, the normalized individual distortions after division by $\phi(L)$ have the bounded Gaussian-square scale in (62). For $j > TL$, they equal $2a_j$ times independent Gaussian squares, with $\sup_{j>TL} a_j = O(1)$ and $\sum_{j>TL} a_j = O(L)$ by regular

variation and Potter's bound. The variance of the total divided by L^2 tends to zero, and its moments of each fixed finite order are bounded by Minkowski's inequality and the sum of scales. This yields the same upper bound J/b to arbitrarily small slack for the fixed finite distortion moments. Apply (65) and let $\gamma \downarrow 0$. Since $L\phi(L)/(R\phi(R)) \to (A+\gamma)^{b-1}$, the moment upper coefficient is $(J/b)A^{b-1}$.

For the lower bound take $L = R/A$. On $\epsilon L \le j \le TL$ let $S(x) = \sum_j f_{a_j}(x_j)$ and $m_L = \mathbb{E}S(X)$. Choose the cutoffs so that $2m_L \ge (J-\eta)L$ for all large L, with a small fixed $\eta > 0$. A point is typical if $|S(x) - m_L| \le \eta L$. The marginal atypical probability is exponentially small, by the bounded Gaussian-square coefficients of the centered potentials and independence. For a fixed typical point y, integrating (63) on the central coordinates gives

$$\mathbb{E}\exp\left(S(X) + S(y) - \sum_j a_j (X_j - y_j)^2\right) = 1.$$

Hence the probability that X is typical and its squared Hilbert distance to y is at most $(J/b - 4\eta)L\phi(L)$ is at most $e^{-(R+\eta L)}$. The same holds with source and target reversed.

For one empirical law, a union over its typical points bounds the typical mass that can couple at this small cost by $e^{-\eta L}$. The atypical empirical mass is negligible in expectation because every sample has the true law. For two empirical clouds, the expected number of typical pairs at this small cost is at most $NMe^{-(R+\eta L)}$. Each pair in any coupling can carry at most $1/\max(N,M)$ mass, so the expected cheap mass is at most $e^{-\eta L}$ plus the two atypical marginal masses. These estimates imply the moment lower bound with the same coefficient after $\eta \downarrow 0$.

To turn the expected asymptotic into an almost-sure one, first project all points and laws radially onto a ball of radius $\log N$ for a sample of size N, or $\log\min(N,M)$ in the two-sample problem. A Gaussian trace-class law has a subgaussian norm tail, since $\mathbb{E}\exp(c\,\|X\|^2) < \infty$ for small $c > 0$ by the eigenvalue product. Clipping therefore costs less than any inverse power of R in joint finite moments. Changing one clipped point changes the optimal cost to the power s by at most $(2R)^s/N$ or $(2R)^s/M$, by rerouting that point's mass. Bounded differences concentration gives probabilities summable in the smaller sample size for deviations by a fixed multiple of the target scale, including when compared to the clipped mean. For two samples with $m = \min(N,M)$ and $\max(N,M) \le m^K$ for fixed $K > 1$, a polynomial union over the other sample size preserves summability, and the clipping is almost surely unnecessary for all large m by the Gaussian tail and the same union bound. The expected asymptotic was uniform in these sizes. Thus the limit holds almost surely throughout every such comparable-size range.

For more disparate sizes let $m = \min(N,M)$ and $n = \max(N,M) > m^K$. By the triangle inequality the difference between two-sample distance and the smaller cloud's distance to μ is at most the larger cloud's distance to μ. One-sample convergence and Potter's bound show that this error, divided by the smaller target scale, is $O(K^{-(b-1)/4})$ almost surely for all sufficiently large m, since $\log n > K \log m$. Let K tend to infinity along the integers. This also gives almost-sure convergence over all pairs of sizes tending to infinity in nested samples.

Finally compute J. With $c = 1 - 1/b$, change variables $a = u^{-b}$ in the first integral of (64). Then integrate by parts in a, and use (62), to obtain

$$J = \int_0^\infty a^{-1/b} B'(a)\, da = \frac{2}{c}\int_0^1 \left(\frac{\rho}{2(1-\rho^2)}\right)^c d\rho = \frac{2^{-c}}{c}\,\mathrm{B}(1 - 1/(2b), 1/b).$$

The boundary terms vanish since $B'(a)$ is bounded at zero and $O(a^{-1})$ at infinity. Substituting gives $(J/b)A^{b-1} = K_b$. This proves Theorem G.1, including the almost-sure two-sample statement.

G.5 Brownian paths and two Kloosterman samples

The covariance operator of standard Brownian motion in $L^2[0,1]$ has eigenvalues $(\pi(j-\frac{1}{2}))^{-2}$, and the Brownian bridge has eigenvalues $(\pi j)^{-2}$. These follow by twice differentiating the covariance eigenfunction equation with the endpoint conditions for each kernel. Both have the same empirical coefficient.

Corollary G.2. *For either of these path laws μ, all fixed finite orders $s \geq 1$ satisfy*

$$\sqrt{\log N}\, W_s(\mu, \mu_N) \longrightarrow \frac{\mathrm{B}(3/4, 1/2)}{\sqrt{2}\pi}$$

almost surely and with the matching expected sth power asymptotic, in the $L^2[0,1]$ metric. The two-sample version uses $\log \min(N, M)$.

Here the optimal deterministic finite-support coefficient is $\sqrt{2}/\pi$ by [32]. The nearest-neighbor random Gaussian coefficient is $3\sqrt{2}/8$ in this normalization. To compute the latter using the randomly centered small-ball method of [52], the scalar fixed-reference objective is $G(a) = \frac{1}{2}\log(1+2a) + a/(1+2a)$. Its integral over $a = u^{-2}$ is $3\pi\sqrt{2}/4$. The conditional small-ball saddle gives the squared coefficient $J_G^2/(4\pi^2)$, hence the stated comparison. The equal-weight constant is strictly larger. The unique Gibbs conditional attaining $G(a)$ with its standard Gaussian reference has reproduction variance $1 - 2a/(1+2a)^2$, which is smaller than one for every $a > 0$. It cannot attain the preserving minimum in (62).

For the Sato–Tate Kloosterman path, the same two-sample reasoning gives the equal-weight constant C_B of Section C. In fact the preserving path coupling of that work, which controls the pinned mean by equal histograms in short central Fourier bands, is exactly a coupling to which (65) applies. Its two marginal $2s$th moments are finite. For the lower bound the two-cloud version of the typical pair count uses the central Fourier potentials of Section C, including their endpoint cross-term estimate in the ordinary metric. Each pair still carries at most $1/\max(N, M)$ mass. Together these give

$$W_s(\nu_N, \nu'_M)^2 \sim \frac{C_B}{\log \min(N, M)}$$

almost surely for nested independent clouds, with $(\mathbb{E}W_s(\nu_N, \nu'_M)^s)^{2/s}$ having the same asymptotic. For the almost-sure limit use the clipping and unequal-size argument above. The Kloosterman L^2 norm is subgaussian because the periodic component is bounded and the pinned constant is a sum of centered bounded coefficients with square-summable weights. This application concerns the independent empirical path laws.

G.6 Excess transport between Gaussian laws

Let $\mu = \mathcal{N}(0, C)$ and $\nu = \mathcal{N}(0, D)$ on the same real separable Hilbert space, where C, D are injective trace-class covariance operators. Let w_j be the singular values, with multiplicities and in decreasing order, of $C^{1/2}D^{1/2}$. Let κ be the standard diagonal Gaussian law with covariance eigenvalues w_j. The following reduction does not require commuting operators.

Theorem G.3. *Suppose $w_j \sim \psi(j)$ for a positive regularly varying function of index $-b$, $b > 1$. Let ν_N be the uniform empirical law of a nested iid sample from ν. Then almost surely and in expectation,*

$$W_2^2(\mu, \nu_N) - W_2^2(\mu, \nu) \sim K_b(\log N)\psi(\log N).$$

The same statement holds with two independent empirical laws μ_N, ν_M, with $\log \min(N, M)$ *in place of* $\log N$. *The assertions are unchanged for arbitrary fixed means.*

Proof. Let $S = C^{1/2}D^{1/2}$. This is trace class by the product inequality for Hilbert–Schmidt operators, and is injective with dense range. Take orthonormal left and right singular bases e_j, f_j with $Sf_j = w_j e_j$. Standard Gaussian coordinates in these bases give representations

$$X = \sum_j C^{1/2} e_j U_j, qquad Y = \sum_j D^{1/2} f_j V_j.$$

The series converge in Hilbert norm almost surely and in mean square, by trace class covariance and independence in each marginal. The standard coordinates are recoverable measurably from almost every X or Y. For example, whiten the usual eigen-coordinates of X by dividing each by the square root of its positive eigenvalue, and use the mean-square convergent orthogonal change of basis to (e_j). Coordinatewise almost-sure convergence follows from square-summable coefficients of the independent standard Gaussians. Applying this simultaneously to the countable coordinates gives the inverse on a set of full measure. It also works for all the countably many empirical points, almost surely.

For any points in the two full-measure convergence sets, take simultaneous finite sums in both representations. Orthogonality of singular vectors gives the exact identity

$$\langle X, Y \rangle = \sum_j w_j U_j V_j = \left\langle (\sqrt{w_j} U_j)_j, (\sqrt{w_j} V_j)_j \right\rangle .$$

The right side converges by the square-sum property, since $\sum_j w_j < \infty$. The left side converges by Hilbert norm convergence of the representations. Thus cross inner products, and consequently their coupling suprema, are preserved under the two measurable coordinate maps. Each underlying true law pushes to κ and each empirical cloud pushes to a corresponding empirical sample from κ. For two true laws the cross inner-product maximum is $\sum_j w_j$, attained by taking $U_j = V_j$. This recovers the population Gaussian transport formula of Gelbrich [58].

For one cloud write $Y_i = \sum_j D^{1/2} f_j V_{ij}$ and $Z_i = (\sqrt{w_j} V_{ij})_j$. Let κ_N be the empirical law of the Z_i. Expanding the quadratic costs and using the identical cross-term optimum gives the exact identity

$$W_2^2(\mu, \nu_N) - W_2^2(\mu, \nu) = W_2^2(\kappa, \kappa_N) + \frac{1}{N} \sum_i (\|Y_i\|^2 - \|Z_i\|^2) - \left(\operatorname{tr} D - \sum_j w_j \right).$$

For two empirical clouds there are identical centered square-norm corrections from both sides, with the true-to-empirical term for κ replaced by the two-sample term. Each correction has mean zero and a subexponential single-sample tail, since both individual square norms have exponential moments near zero. Bernstein concentration gives sample corrections $O(\sqrt{\log N / N})$ almost surely and $O(N^{-1/2})$ in mean absolute value, and similarly for M. They are negligible on the regularly varying logarithmic scale. Apply Theorem G.1 to κ.

For nonzero means first subtract the two fixed means. In the squared cost, the change consists of the population squared mean difference and a centered linear sample-mean term depending only on the two marginals. The latter has the same negligible bounds by Gaussian concentration. Hence the result holds as stated. □

G.7 An explicit noncommuting path application

Let σ and τ be continuous strictly positive functions on $[0, 1]$. Consider the centered Gaussian paths $X_\sigma(t) = \int_0^t \sigma(u) dW_u$ and $X_\tau(t) = \int_0^t \tau(u) dW_u$ as laws on $L^2[0, 1]$, with covariance operators

C_σ, C_τ. Their cross singular values satisfy

$$w_j(C_\sigma^{1/2} C_\tau^{1/2}) \sim \frac{1}{\pi^2 j^2} \left(\int_0^1 \sqrt{\sigma(t)\tau(t)}\, dt \right)^2 . \tag{66}$$

Consequently the coefficient multiplying $1/\log N$ in the squared empirical excess of Theorem G.3 is

$$\frac{\mathrm{B}(3/4, 1/2)^2}{2\pi^2} \left(\int_0^1 \sqrt{\sigma(t)\tau(t)}\, dt \right)^2 .$$

We prove (66) directly. Let A be the Volterra integral operator and M_σ the multiplier by σ, so $C_\sigma = (AM_\sigma)(AM_\sigma)^*$. Since AM_σ is injective with dense range, its left polar decomposition uses a unitary right factor. Hence the cross singular values are those of $M_\sigma A^* A M_\tau$. The kernel of A^*A is $1 - \max(s,t)$. First use positive step functions on a common finite partition into intervals $I_i = [a_i, b_i]$. Off-diagonal blocks of this kernel depend on only one variable and are rank one. On each diagonal block the difference between this kernel and $b_i - \max(s,t)$ is the constant $1 - b_i$, also rank one. Thus, up to finite rank, the operator is block diagonal with singular values

$$\frac{\sigma_i \tau_i |I_i|^2}{\pi^2 (j - \frac{1}{2})^2}, \qquad j \geq 1,$$

on block i. The unweighted eigenvalues in this expression are those of the local Volterra square, by the same endpoint calculation as for Brownian covariance. Counting all block singular values above a small threshold gives (66) with $\sum_i |I_i| \sqrt{\sigma_i \tau_i}$ in place of the integral. A finite-rank perturbation shifts the indices in the singular-value inequalities by at most its rank and so leaves the leading asymptotic unchanged.

For continuous positive functions, approximate both by such step functions with uniformly small relative errors. The true operator is obtained by multiplying the step operator on the left and the right by ratios tending uniformly to one as the partition is refined. The singular-value bounds by the two operator norms and those of the inverses give the asymptotic constant by squeezing, since the step sums converge to the displayed integral.

One can also compare Brownian motion and Brownian bridge. Their covariance operators are C and $D = C - t \otimes t$ on $L^2[0,1]$, where t denotes the function $t \mapsto t$. The squared cross singular values are the eigenvalues of the rank-one perturbation $C^2 - (C^{1/2}t) \otimes (C^{1/2}t)$. They interlace the squared eigenvalues of C and have the same leading coefficient. These covariance operators do not commute, since $Ct = t/2 - t^3/6$ is not a multiple of t. The cross empirical excess again has the coefficient $\mathrm{B}(3/4, 1/2)^2/(2\pi^2)$.

G.8 Paired Gaussian samples

An extension applies when two equally sized empirical clouds use the same observations of underlying noise. First let P be the common reduced diagonal Gaussian law of Theorem G.1. Draw iid pairs (X_i, Y_i) with both marginals P, allowing dependence within a pair. If $\mathbb{P}(X_i = Y_i) = 0$, the same two-sample limit holds for their two equal-size empirical clouds.

For the upper proof omit the paired diagonal of the likelihood matrix and use normalization $N(N-1)$ instead of N^2. A row conditional on X_i still averages $N - 1$ independent P-points, and likewise for each column conditional on Y_j. Use the cutoff $\log(N-1) - t$ in (65); the row and column discrepancy bounds and the expected cost bound are unchanged. For the lower bound the off-diagonal expected close-pair count has the independent estimate. A diagonal pair has capacity $1/N$, so its total expected cheap mass is bounded by $\mathbb{P}(\|X_1 - Y_1\| \leq d_N)$ at the small-distance

threshold $d_N \to 0$. This tends to zero by the assumption. The marginal atypical masses have the same bounds as before. For the almost-sure conclusion apply bounded differences after clipping, now altering both points of one iid pair, which costs at most twice the previous amount. For jointly Gaussian pairs, either $X_i = Y_i$ almost surely or their difference has a nondegenerate scalar Gaussian projection and the assumption holds. If equality holds almost surely, the two reduced empirical measures coincide exactly.

Applying this argument to the cross-coordinate maps in Theorem G.3 gives the identical squared-excess formula for paired Gaussian samples whenever the reduced pair is not identical. This covers the two deterministic-volatility path laws above when paired at each label using the same Brownian motion and σ, τ are not proportional. In fact, in white-noise coordinates the paired reduced variables are proportional coordinatewise to the noise evaluated in the left and right singular bases of $M_\sigma A^* A M_\tau$. They coincide almost surely only if these bases coincide, in which case that operator is positive self-adjoint. Its kernel is $\sigma(s)(1 - \max(s,t))\tau(t)$, and symmetry would force σ/τ constant. Conversely, if the volatilities are proportional the kernel is positive self-adjoint, one can take identical singular bases, and the reduced empirical distance is zero. The original excess then consists only of the centered empirical norm corrections in the exact cost identity.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, internal independent checks and drafting. The proofs use no datasets.

H Two-shift Kloosterman transport at fixed prime

For two distinct active shifts at depth n over a fixed odd prime, we determine the quadratic Wasserstein order of their stationary-phase coordinates for every tame rank. When the inverse coefficient is fixed, the order is $q^{-1/2}\sqrt{\log q}$ in two angle dimensions and $q^{-1/(2k)}$ in $2k > 2$ angle dimensions, where $q = p^n$ and k is the cyclotomic phase dimension of one root packet. The critical logarithm persists for the classical real trace pair after cosine folding, including the mixture of residue classes. When both coefficients vary within prescribed residue classes, the phase law is a finite mixture of affine product grids and has order $q^{-1/k}$, even for bottleneck transport. The proof uses an exact stationary Fourier magnitude and, for the real scalar lower bound, an endpoint-vanishing multiplier whose nonconstant Fourier shifts are all nonstationary.

Keywords: prime-power Kloosterman sums, modular hyperbola, Wasserstein distance, stationary phase.
MSC 2020: 11L05, 11L07, 60B15.

H.1 Statements

Fix an odd prime p, a rank $R \geq 2$ with $p \nmid R$, a modulus $q = p^n$, and fixed distinct integer shifts h_1, h_2. Put $v = v_p(h_2 - h_1)$ and use $e_q(t) = \exp(2\pi i t/q)$. For a unit a put $c_i = (b - h_i)/a$ and use normalized complete sums

$$K_R(c_i; q) = q^{-(R-1)/2} \sum_{z_1 \cdots z_R = c_i} e_q(z_1 + \cdots + z_R),$$

with unit variables. In rank two $K_2((b - h_i)/a; q)$ equals the normalized classical sum with linear coefficient $b - h_i$ and inverse coefficient a^{-1} at every active residue. Write the dependence on the fixed shifts as v in the distance constants.

Let $d = \gcd(R, p-1)$, $k = \varphi(d)$, with $k = 1$ at $d = 1$. Use a primitive dth root $\zeta \in \mathbb{Z}_p$ (take $\zeta = 1$ at $d = 1$). Fix $a_0 \in \mathbb{F}_p^\times, b_0 \in \mathbb{F}_p$ with both c_i nonzero Rth powers modulo p. Fix root residues $x_0^R = (b_0 - h_1)/a_0$ and $y_0^R = (b_0 - h_2)/a_0$. For each lift of a, b let x, y be the unique respective lifts of these roots modulo q. The phase vector in power-basis coordinates is

$$\Theta(a,b) = \frac{R}{q}(x, x\zeta, \dots, x\zeta^{k-1}, y, y\zeta, \dots, y\zeta^{k-1}) \in \mathbb{T}^{2k}.$$

Write m_j for normalized Haar on $\mathbb{T}^j$. The torus metric is the Euclidean quotient metric, and trace pairs use the Euclidean metric.

Theorem H.1. *Let μ_q be the law of $\Theta(a,b)$ with a fixed and b uniform over its prescribed residue class. Uniformly in lifts of the fixed residue data, as $n \to \infty$,*

$$W_2(\mu_q, m_{2k}) \asymp_{p,R,v} \begin{cases} q^{-1/2}\sqrt{\log q}, & k = 1, \\ q^{-1/(2k)}, & k > 1. \end{cases}$$

When instead a, b are independent uniform lifts of a_0, b_0, the law $\widetilde{\mu}_q$ satisfies $W_s(\widetilde{\mu}_q, m_{2k}) \asymp_{p,R,v} q^{-1/k}$ for every $1 \le s \le \infty$. For shifts distinct modulo p it is an exact product of two affine grids.

Theorem H.2. *For the residue-conditioned trace pair and the independent Haar phase trace pair, the same fixed-a orders hold when $d = 1, 2, 3$. At $d = 2$, the two comparison marginals are arcsine laws on $[-2, 2]$. At $d = 3$, the comparison marginals are planar three-root trace laws, with the appropriate common rotation. The two-parameter orders also hold for these d and all $1 \le s \le \infty$.*

For the classical rank-two family, let b range over the lifts of any fixed nonempty union of residue classes modulo p. Use the mixture with the same residue weights and, on each class, independent arcsine traces at the active shifts and zero at the other shifts. If any sampled residue has both shifts active, the fixed-a quadratic distance to this law is of order $q^{-1/2}\sqrt{\log q}$. If at most one shift is active in each class and some shift is active, the order is q^{-1}; when neither is ever active both laws are the same point mass.

The scalar theorem concerns two completed coordinates. Milićević and Zhang [59, Section 4.2] use complete sums of shifted root phases to prove the qualitative fixed-prime path limit. Their estimate allows many shifts and has a bounded-frequency hypothesis. Milićević, Robinson and Shupe [60] study correlations of products and the resulting alignment constraints for prime-power Kloosterman sums. In the two-shift problem both frequencies can range up to the transport scale, and the stationary magnitude has a particularly simple exact form. Blomer, Risager and Shparlinski [61] study discrepancy of modular inverses collected over a growing interval of moduli. The fixed-prime quadratic transport question uses the critical $\dot{H}^{-1}$ energy of one modulus. Peyre's comparison [62] supplies the general transport inequality used below.

H.2 The stationary Fourier coefficient

With a fixed, set $\Delta = (h_2 - h_1)/a$, regarding a as any unit lift. The parameterization is $x^R - y^R = \Delta$, with x ranging uniformly over its root residue class. The implicit function $y = y(x)$ has integral Taylor coefficients on that disk, as follows by solving $y^R = x^R - \Delta$ recursively at any center, using the unit Ry^{R-1}. Put $w = x/y$. The derivative identities are

$$y' = w^{R-1}, \qquad w' = -\Delta/y^{R+1}.$$

For x, x' in the root disk, subtracting their values of $w^R = 1 + \Delta/(x^R - \Delta)$ gives $v_p(w(x) - w(x')) = v + v_p(x - x')$. Here taking Rth powers preserves difference valuations on any fixed unit residue

disk by $p \nmid R$. Thus w maps the root disk onto $w_0 + p^{v+1}\mathbb{Z}_p$: after division by the initial scale, the isometry is bijective on each finite quotient. Take w_0 to be the image of any one chosen lift in the root disk.

Lemma H.3. *For $A, B \in \mathbb{Z}_p$, let $u = \min(v_p(A), v_p(B))$ with $v_p(0) = \infty$, and write $A = p^u A_1$, $B = p^u B_1$ when $u < n$. Put $e = v_p(R-1) + v$ and $N = n - u$. If N exceeds a constant depending only on e, the normalized residue average of $e_q(R(Ax + By))$ is zero unless*

$$A_1 + B_1 w_0^{R-1} \equiv 0 \pmod{p^{e+1}}. \tag{67}$$

On the stationary congruence both A_1, B_1 are units. The average equals $p^{1-(N-e)/2}\gamma\, e_{p^N}(f(x_*))$ with $f(x) = R(A_1 x + B_1 y(x))$, its unique critical point x_* in the disk, and a factor γ that is 1 for $N - e$ even and is ± 1 or $\pm i$ for $N - e$ odd according as $p \equiv 1$ or $3 \pmod 4$.

Proof. The power map on $w_0(1 + p^{v+1}\mathbb{Z}_p)$ has image $w_0^{R-1}(1 + p^{e+1}\mathbb{Z}_p)$ by the p-adic logarithm. On this disk raising to the $(R-1)$st power adds $v_p(R-1)$ to the difference valuation, by the same logarithm or the lifting-the-exponent identity. The derivative $f' = R(A_1 + B_1 w^{R-1})$ therefore has its unique zero exactly under (67), and then $v_p(f'(x)) = e + v_p(x - x_*)$.

Set $l = \lfloor N/2 \rfloor$. Averaging over the top l digits of x cancels all terms with $f'(x) \not\equiv 0 \pmod{p^l}$, since the second-order remainder vanishes at depth N. For N large, only the disk $x_* + p^r\mathbb{Z}_p$, $r = l - e \geq 1$, survives; if there is no critical point all terms cancel. On that disk all Taylor terms of degree at least three vanish modulo p^N since $3r \geq N$. The coefficient of $(x - x_*)^2$ has valuation e, because $f'' = RB_1(R-1)w^{R-2}w'$ and 2 is a unit. The normalized quadratic Gauss average on the disk has modulus $p^{-(N-e-2r)/2}$, and its factor is as stated since subtracting $2r$ preserves parity. Multiplying by the disk's relative measure p^{1-r} proves the formula. A zero quadratic modulus exponent is interpreted as an average of one. □

In rank two with $p \nmid \Delta$, a change of variables also identifies the exact modular hyperbola. Write $t = x - y$; then $x + y = \Delta/t$ and t maps its residue disk bijectively onto a unit residue disk since $dt/dx = 1 - x/y$ is a unit. The pair $(2x, 2y)$ is obtained from $(t, \Delta/t)$ by the integral linear map $(t, z) \mapsto (t+z, z-t)$, whose determinant is a unit modulo p. Lemma H.3 can in this case be obtained from the ordinary quadratic stationary sum of a restricted modular inverse pair.

H.3 Transport on the phase torus

For a signed measure of mean zero define

$$\|\sigma\|_{\dot{H}^{-1}}^2 = \sum_{m \neq 0} \frac{|\widehat{\sigma}(m)|^2}{4\pi^2 |m|^2}, \qquad \widehat{\sigma}(m) = \int e^{-2\pi i m \cdot \theta}\, d\sigma(\theta).$$

Convolution with a periodic Gaussian with Fourier multiplier $e^{-4\pi^2|m|^2/T^2}$ costs $O_k(T^{-1})$ in W_2, by using the same draw from the periodized Gaussian. It contracts W_2 by common additive noise. For a smooth probability density ρ relative to Haar, $W_2(\rho, m_j) \leq 2\|\rho - m_j\|_{\dot{H}^{-1}}$ [62, Theorem 2.1]. Indeed the curve of densities $1 + (1 - (1-t)^2)(\rho - 1)$ with flux given by the corresponding rescaling of the inverse Laplacian has action at most $4\|\rho - m_j\|_{\dot{H}^{-1}}^2$. If both densities are bounded above by M on the flat torus, $\|\rho - m_j\|_{\dot{H}^{-1}} \leq \sqrt{M} W_2(\rho, m_j)$ [62, Theorem 2.5].

For $m = (m^{(1)}, m^{(2)}) \in \mathbb{Z}^{2k}$ the frequency coefficients are $A = \sum_{j=0}^{k-1} m_j^{(1)} \zeta^j$ and $B = \sum_{j=0}^{k-1} m_j^{(2)} \zeta^j$. If $p^u \mid A, B$, the cyclotomic norm of every nonzero block is divisible by p^u. Its absolute value is at most a rank-dependent constant times the kth power of the block's Euclidean

norm. Thus the frequency sublattice at level u has shortest nonzero length $\gg_R p^{u/k}$. Its points in a ball of radius H number $O_R((H/p^{u/k})^{2k})$ once $H \gg_R p^{u/k}$, by disjoint small balls. For $k > 1$, dyadic summation yields, at every $T \geq 1$,

$$\sum_{\substack{m \neq 0 \\ p^u | A,B}} \frac{e^{-8\pi^2|m|^2/T^2}}{|m|^2} \ll_R p^{-2u} T^{2k-2}.$$

This bound also holds when $T < p^{u/k}$ by the same summation from the first allowable shell and Gaussian decay. For frequencies with $u < n - O_{R,v}(1)$ use the upper bound $O_{p,R,v}(p^u/q)$ for the squared Fourier coefficients from Lemma H.3. The other frequencies have gap $\gg_{p,R,v} q^{1/k}$ and negligible weight at $T = q^{1/(2k)}$ even with coefficients bounded by one, by Gaussian decay. The smoothed $\dot{H}^{-1}$ energy is thus $O_{p,R,v}(q^{-1}T^{2k-2})$. This and the smoothing cost prove the upper at $k > 1$. Every coupling from Haar to a measure on at most q/p atoms moves a fixed fraction of mass by at least a constant times $q^{-1/(2k)}$, by the volume of the union of the balls around the atoms. This proves its sharpness.

For $k = 1$ we have $A = m_1, B = m_2$. The number of primitive residue pairs obeying (67) has density $(p-1)/p^{e+2}$ among integer pairs. In fact one coordinate may be any unit modulo p^{e+1} and the other is prescribed. By elementary lattice-point counting in these residue classes followed by partial summation, their sum of $(4\pi^2|m|^2)^{-1}$ over $0 < |m| \leq H$ is this density times $(2\pi)^{-1} \log H$, up to $O_{p,R,v}(1)$ for $H \geq 1$. At valuation u, the weight from Lemma H.3 and the change of variables $m \mapsto p^u m$ multiply to $q^{-1}p^{2+e-u}$. Summing the geometric series and the errors gives, for $2 \leq T \leq q/p^{N_0}$ with a fixed sufficiently large N_0,

$$\sum_{0<|m|\leq T} \frac{|\widehat{\mu}_q(m)|^2}{4\pi^2|m|^2} = \frac{p}{2\pi q} \log T + O_{p,R,v}(q^{-1}). \tag{68}$$

The same leading term holds with Gaussian smoothing at $T = q^{1/2}$ or at $T = q^{1/4}/\log q$. This follows by partial summation of (68), with exponentially small mass beyond q/p^{N_0}. The upper for $k = 1$ follows at the first bandwidth. At the second bandwidth, the sum of absolute nonzero smoothed Fourier coefficients is $O_{p,R,v}(q^{-1/2}T^2)$. Indeed at valuation u their magnitude bound is $O_{p,R,v}(q^{-1/2}p^{u/2})$, and the Gaussian sum over nonzero points of $p^u\mathbb{Z}^2$ is $O(T^2p^{-2u})$. The alias frequencies are again negligible. The resulting density is bounded by $1 + o(1)$. The lower transport comparison and contraction show the lower order and give

$$\liminf_{n\to\infty} \frac{q}{\log q} W_2(\mu_q, m_2)^2 \;\geq\; \frac{p}{8\pi}.$$

If both parameters vary, the root pairs are uniform over those satisfying their two root-residue conditions and $x^R - y^R \equiv (h_2 - h_1)/a_0 \pmod{p^{v+1}}$. The latter quotient is independent of the lift of a_0. For every such pair, divide $a(x^R - y^R) = h_2 - h_1$ by p^v to get one solution for a modulo p^{n-v}, hence p^v lifts modulo q, each with $b = h_1 + ax^R$. For $n > v + 1$ this makes the phase law an equally weighted mixture of p^v products of affine phase grids: each lift of x_0 modulo p^{v+1} determines one lift of y_0 there by Hensel's lemma. When $v = 0$ the root-residue conditions already enforce the difference condition, so there is one product law.

For each block, put $Q = q/p^{v+1}$. The grid is the projection of a translate of

$$\Lambda = \mathbb{Z}^k + Q^{-1}\mathbb{Z}(1, \zeta, \ldots, \zeta^{k-1}).$$

Its dual has determinant Q and first minimum $\gg_R Q^{1/k}$ by the norm argument. Minkowski's second theorem [63] bounds all its successive minima by $O_R(Q^{1/k})$. The independent short dual

vectors span a sublattice of bounded index. Their dual basis has lengths $O_R(Q^{-1/k})$ by the cofactor bound, since the determinant is at least Q. Multiplying by the index puts these dual-basis vectors in Λ. They supply a covering radius $O_R(Q^{-1/k})$. Equal-volume Voronoi coupling gives this upper bound in W_∞ for each block, hence for the product and mixture. The lower bound follows from the number of atoms by the same volume argument as above. This proves Theorem H.1.

H.4 Scalar traces and cosine folding

For $n \geq 2$, tame stationary phase for K_R gives a sum over the d roots $s\zeta^j$ with phases $e_q(Rs\zeta^j)$. At even depth all weights are one. At odd depth their weights are a single unit times $\chi(s\zeta^j)^{R-1}$, with χ the quadratic character modulo p [24, 31]. The rank-$R-1$ Hessian is $s^{-1}(I+\mathbf{1}\mathbf{1}^{\mathsf{T}})$, with determinant $Rs^{-(R-1)}$, which gives the weight on each branch. All weights are fixed on each prescribed residue. The trace maps are uniformly Lipschitz on the phase tori, so all stated upper bounds follow. When $d = 1$ the map is the circle embedding up to rotation, which is bi-Lipschitz for the quotient metric. When $d = 3$, R is odd, and the three-root trace $e(x) + e(y) + e(-x - y)$ has real differential of rank two wherever these three phases are all distinct. A regular chart pushes Haar onto a positive bounded-below density on a planar ball. The pair of independent such charts gives a positive bounded-below density on a four-dimensional open set, and the support lower bound proves the scalar rates for $d = 3$.

For $d = 2$ one may rotate each trace by a fixed unit and write it as $2\cos(2\pi\theta_j)$, where $\theta_1 = Rx/q + \delta_1$, $\theta_2 = Ry/q + \delta_2$, with $4\delta_j \in \mathbb{Z}$. Indeed the two opposite roots contribute a sum of exponentials or their difference, times a unit. Let ν be the law of this translated angle vector, averaged under the two independent coordinate reflections. Its cosine pushforward is unchanged. Haar on the angle torus pushes to the product arcsine law. Reflection changes neither the support of the test functions below nor the effect of the fixed rotations.

Choose $T = q^{1/10}$. Among $m \in 4p^2\mathbb{Z}^2$ with $0 < |m| \leq T$ take those of exact common p-valuation 2 satisfying the primitive stationary congruence (67). This is a positive-density collection in $\mathbb{Z}^2$ for fixed p, R. For a primitive stationary pair, changing the sign of just one coordinate destroys the congruence, since both entries are units and p is odd. On frequencies divisible by 4 the shifts δ_j give no Fourier multiplier. Hence $\widehat{\nu}(m)$ at these frequencies is one half the real part of the stationary coefficient (the conjugate comes from simultaneous reflection).

For each base frequency m at least one of $m, 2m, pm$ has $|\widehat{\nu}| \gg_{p,R,v} q^{-1/2}$. If the stationary factor at m is real, doubling the phase keeps the effective depth unchanged, and $\cos^2\phi + \cos^2(2\phi)$ has a positive minimum. If it is imaginary, passing to pm decreases the effective depth by one and keeps the primitive phase and critical point. It raises the exponential to its pth power and makes the Gaussian factor real. The continuous periodic function $\sin^2\phi + \cos^2(p\phi)$ has a positive minimum. In each case the possible Gaussian signs are irrelevant. Use the finite reflection-invariant set S of all the base frequencies and their doubles and p-multiples and their reflections. They have bounded possible multiplicity when generated this way. Set

$$E = \sum_{m \in S} \frac{|\widehat{\nu}(m)|^2}{4\pi^2|m|^2}, \qquad \psi(\theta) = \sum_{m \in S} \frac{\widehat{\nu}(m) e^{2\pi i m \cdot \theta}}{4\pi^2|m|^2}.$$

The base lattice sum and Lemma H.3 give $E \asymp_{p,R,v} q^{-1}\log q$. The polynomial ψ is even in each coordinate and $\|\psi\|_{H^1(m_2)} \ll \sqrt{E}$.

Use the multiplier $\eta(\theta) = \sin^2(4\pi\theta_1)\sin^2(4\pi p\theta_2)$. Its Fourier support consists of pairs with entries 0 or ± 4 in the first coordinate and 0 or $\pm 4p$ in the second. Every nonzero such shift added to a frequency of S gives a pair with unequal p-valuations. It is nonstationary in every coordinate

reflection by Lemma H.3, because its primitive pair has exactly one unit. Also it is nonzero as an integer vector. It follows exactly that

$$\int \eta\psi \, d(\nu - m_2) = \tfrac{1}{4}E. \tag{69}$$

The multiplier vanishes quadratically on the fold lines and also vanishes where either cosine is zero. The product polynomial descends to a polynomial F on the trace square. The chain rule shows $\|\nabla F\|_{L^2(\arcsine^{\otimes 2})} \ll_p \sqrt{E}$: for each derivative, both η and its corresponding partial derivative divided by $\sin(2\pi\theta_j)$ are bounded. By the Chebyshev derivative bounds (or the Markov inequality applied twice), a cosine mode of index m on $[-2,2]^2$ has second derivatives bounded by $O((1+|m|)^4)$. Since each $|\widehat{\nu}(m)| \ll_{p,R,v} q^{-1/2}$ on S, this gives $\|D^2F\|_\infty \ll_{p,R,v} q^{-1/2}T^4 \ll_{p,R,v} 1$. Taylor expansion from the reference point in any coupling between reference and arithmetic traces gives

$$\tfrac{1}{4}E \ll_{p,R,v} \sqrt{E}\, W_2 + W_2^2.$$

This proves the scalar lower at $d = 2$ for the fixed parameter. With both coefficients varying the scalar lower follows directly from the two-dimensional support bound, since the arcsine product density is bounded below on a fixed interior square.

For the classical mixture version, fix one sampled residue with both shifts active and use its stationary slope in the construction of S. In a primitive stationary pair at another active residue the square of the slope must equal $w^2 = (b-h_1)/(b-h_2)$ modulo p^{v+1}, including after a coordinate reflection. This fractional linear expression is injective in the both-active residues: the difference for distinct residues b, b' has exact valuation v because $b-h_2, b'-h_2$ and $b-b'$ are units. Thus all frequencies of S give zero on the other both-active residues, and all nonconstant multiplier shifts give zero on every both-active residue. Formula (69) now contributes $E/4$ times the fixed positive weight of the chosen residue. On the one-active or zero-active residues, F vanishes on the relevant coordinate axes. The gradient bound for Taylor still holds relative to the mixture reference: ∇F vanishes on those axes as well, since η, as a polynomial in the scalar traces, has a factor given by the square of each trace. In unconditional rank two all traces are real, and the residue-dependent sine/cosine signs are realized by the allowed translations. Inactive shifts give zero by stationary-phase cancellation, including a nonunit linear coefficient. Mixing the conditional couplings gives the corresponding upper. If there is only one active shift per residue, the active angle is uniform on an affine circle grid and its scalar transport is $O_p(q^{-1})$. For the lower, a positive-weight arcsine component gives a reference submeasure with density bounded below on an interval; at most q points in the target require a positive mass there to move distance $\gg_p q^{-1}$. This proves the remaining assertions of Theorem H.2.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, including the frequency calculation and scalar multiplier, and with drafting and proof checking. No data are needed for the results.

I Rational stationary phases in Kloosterman transport

Fix three shifts distinct modulo an odd prime p, and vary both Kloosterman coefficients over lifts of a residue pair on which the three shifts are active. Over $q = p^n$, their angle and real trace laws have Wasserstein order $q^{-1/2}$ for exponents between one and two. A rational quadratic cone gives the exact leading quadratic angle constant and dual gradient formulas for the trace constants along each parity. For any prescribed number $m \geq 3$ of shifts and fixed rank we also

construct fixed shift configurations with a bounded-frequency lower bound of order $q^{-1/(m-1)}$. A circle projection of the angle law attains this order in every Wasserstein distance and has arithmetic mass of that order on a single character fiber. In the classical rank the lower bound persists after cosine folding.

Keywords: Kloosterman sums, rational conics, Wasserstein transport, prime powers.
MSC 2020: 11L05, 11L07, 60B15.

I.1 The three-coordinate problem

Fix an odd prime p and integers h_1, h_2, h_3 pairwise distinct modulo p. Set $q = p^n$ and $e_q(t) = \exp(2\pi i t/q)$. For a a unit and b arbitrary modulo q, let

$$K_i(a,b) = q^{-1/2} \sideset{}{^\times}\sum_{t \bmod q} e_q\big((b-h_i)t + a^{-1}t^{-1}\big).$$

Choose $a_0 \in \mathbb{F}_p^\times$ and $b_0 \in \mathbb{F}_p$ so that the three values $(b_0 - h_i)/a_0$ are nonzero squares. Fix one square-root residue of each, denoted x_0, y_0, z_0. Take a, b independent uniform lifts of a_0, b_0 modulo q and let x, y, z be the corresponding square-root lifts of $(b-h_i)/a$. Define the angle law M_n on the Euclidean quotient torus $\mathbb{T}^3$ by $(2x/q, 2y/q, 2z/q)$, and let m_3 be its normalized Haar comparison. The real trace law is $L_n = \mathrm{law}(K_1, K_2, K_3)$ on $[-2,2]^3$ for $n \geq 2$. Let ρ be the product of three arcsine laws on $[-2,2]$.

Define

$$\lambda = \frac{h_2 - h_3}{h_2 - h_1}, \qquad \kappa = \frac{h_3 - h_1}{h_2 - h_1}, \qquad D(A,B,C) = \frac{A^2}{\lambda} + \frac{B^2}{\kappa} - C^2.$$

The two fractions are p-units and have sum one. Set $w_0 = y_0/x_0$ and $s_0 = z_0/x_0$ in $\mathbb{F}_p$. Write $\mathcal{C}$ for the nonzero integer vectors $j = p^u(A,B,C)$, with at least one of A, B, C a unit at p, for which

$$D(A,B,C) = 0, \qquad A + Bw_0 + Cs_0 = 0, \qquad B + C\kappa w_0/s_0 = 0 \pmod p. \tag{70}$$

Here the first equation is over $\mathbb{Q}$ and the other two are in $\mathbb{F}_p$. All three primitive entries are units at p on $\mathcal{C}$.

Theorem I.1. *As $n \to \infty$ the angle law satisfies*

$$qW_2(M_n, m_3)^2 \longrightarrow \sum_{j \in \mathcal{C}} \frac{p^{2+u(j)}}{4\pi^2 |j|^2} \ \in (0, \infty). \tag{71}$$

For each $1 \leq r \leq 2$ the distances $W_r(M_n, m_3)$ and $W_r(L_n, \rho)$ have order $q^{-1/2}$, and after multiplication by $\sqrt{q}$ have positive finite limits along each depth parity. The constants for each r and parity are the dual gradient norms in (75) *below.*

The pair-lift law uses $(q/p)^2$ parameter pairs. The volume lower bound from that support size in three dimensions gives $q^{-2/3}$, while fixed Fourier modes produce the larger order in Theorem I.1. The mechanism differs from the product affine grids that occur with two shifts. Milićević and Zhang [59, Section 4.2] prove fixed-prime path limits by stationary analysis of bounded-order shifted root sums with one free coefficient. Milićević, Robinson and Shupe [60, Proposition 5.3] estimate general prime-power product correlations by means of alignment and stationary-point bounds. The two-variable triple here has a direct rational description of its entire leading Fourier variation.

I.2 The exact cone and the remaining levels

The first two roots determine the coefficients by $a = (h_2 - h_1)/(x^2 - y^2)$ and $b = h_1 + ax^2$. The denominator is a unit, so x, y are independent uniform lifts of their residues. Also $z^2 = \lambda x^2 + \kappa y^2$. Put $w = y/x$ and let $s(w)$ be the branch of $\sqrt{\lambda + \kappa w^2}$ on $w_0 + p\mathbb{Z}_p$ with residue s_0. Then $z = xs(w)$ and x, w are independent uniform lifts. The function s has integral Taylor coefficients at every center of this disk, by successive coefficient solution and the unit $2s$.

Use the Fourier convention $\widehat{M}_n(j) = \int e^{2\pi i j\cdot\theta}\, dM_n(\theta)$. For $j = p^u(A, B, C)$ put $N = n - u$ and $F(w) = A + Bw + Cs(w)$. If $N \geq 2$, last-digit summation in x and w gives zero unless $F(w_0) = F'(w_0) = 0 \pmod p$. These equations give $B = -C\kappa w_0/s_0$ and $A = -C\lambda/s_0$ modulo p, so all entries must be units for a primitive vector. As $s'' = \lambda\kappa/s^3$, the stationary derivative is simple and gives one critical point w_* in the disk. The normalized average over w, for every x in its residue class, is

$$p^{1-N/2}\,\gamma_N\, e_{p^N}(2xF(w_*)). \tag{72}$$

For even N the Gaussian factor is 1. For odd N it is the normalized Gauss phase of the unit $xF''(w_*)$ modulo p, so it depends only on x_0 and the primitive residues. For (72), group the top $\lfloor N/2 \rfloor$ digits of w. The sum restricts to the stationary ball of radius $p^{-\lfloor N/2 \rfloor}$. On this ball the Taylor phase modulo p^N is quadratic (also at $N = 2, 3$), which gives the factor and magnitude displayed.

Average (72) over x. Since the Gaussian factor is fixed, the result is zero unless p^{N-1} divides $F(w_*)$, and otherwise the phase after averaging is $e_{p^N}(2x_0F(w_*))$. To convert the condition to an integer congruence, write $V = C\lambda/s(w_*)$. The critical equation gives $F(w_*) = A + V$ and

$$D(A, B, C) = \lambda^{-1}(A - V)(A + V).$$

As $A \equiv -V \pmod p$, the first factor $A - V$ is a unit. In the stationary residue conditions we therefore obtain exactly

$$|\widehat{M}_n(j)|^2 = \frac{p^{2+u}}{q}\,\mathbf{1}_{D(A,B,C)\equiv 0\ (p^{N-1})}. \tag{73}$$

If the residue conditions fail the coefficient is zero. At an exact cone mode $j \in \mathcal{C}$, $F(w_*) = 0$, so $\sqrt{q}\,\widehat{M}_n(j) = p^{1+u/2}\gamma_N$ depends only on the parity of n once $N \geq 2$.

Every admitted residue class has exact cone modes. The rational conic $s^2 = \lambda + \kappa w^2$ contains $(1, 1)$. Its given residue point (w_0, s_0) has $w_0^2 \neq 1$, because the first two shift values are distinct and nonzero. The line from $(1, 1)$ to that residue point is a secant. Choose a rational slope lifting its slope modulo p. Substituting $s = 1 + t(w - 1)$ and factoring off $w - 1$ shows that the other intersection is rational and reduces to (w_0, s_0). In fact $t^2 - \kappa$ is a unit: otherwise the remaining linear factor, at the residue slope, would vanish identically, forcing $t = \kappa$ and $\kappa = 1$, which is impossible. For this rational point (w_*, s_*), the rational vector $(-\lambda/s_*, -\kappa w_*/s_*, 1)$ satisfies the exact and residue conditions of (70). Clearing denominators that are units at p produces a fixed integer mode in $\mathcal{C}$ primitive at p.

Lemma I.2. *Let D_I be the integral ternary form obtained by multiplying D by a fixed common denominator, chosen prime to p. For every $\eta > 0$, uniformly for integers l and $H \geq 1$,*

$$\#\{(A, B, C) \in \mathbb{Z}^3 : |(A, B, C)| \leq H,\ D_I(A, B, C) = l\} \ll_{h_1,h_2,h_3,\eta} H^{1+\eta}.$$

Proof. The quadratic form is nonsingular and rationally isotropic, since $(\lambda, \kappa, 1)$ has value zero. In a rational basis it takes the form of a hyperbolic plane and a nonzero one-variable square. To see this, pair one isotropic vector with a vector having nonzero polarized pairing, subtract a multiple of the

isotropic vector to make the second isotropic, and take their orthogonal complement. After clearing denominators there is an injective fixed integral linear change of coordinates carrying $D_I = l$ into an equation $UV = cW^2 + c'l$ with fixed nonzero integers c, c' and $|U|, |V|, |W| \ll H$. The left and right sides determine each other exactly on integer points under this change. For each W with nonzero right side there are $O_\eta(H^\eta)$ signed factor pairs, since $|l| \ll H^2$ on the original radius and by the divisor bound (with exponent $\eta/2$). The right side can vanish for at most two W; in either such case $UV = 0$ contributes $O(H)$ pairs. This proves the estimate. □

I.3 A strong negative Sobolev limit

For a mean-zero signed distribution on $\mathbb{T}^3$ define

$$\|\sigma\|_{\dot{H}^{-1}}^2 = \sum_{j \neq 0} \frac{|\widehat{\sigma}(j)|^2}{4\pi^2 |j|^2}.$$

For $\epsilon \in \{0, 1\}$ let σ_ϵ be the real mean-zero distribution whose coefficient at $j \in \mathcal{C}$ is $p^{1+u(j)/2}\gamma_N$ for $N \equiv \epsilon - u(j) \pmod 2$, and is zero off $\mathcal{C}$. Here γ_N denotes just the residue Gaussian phase in (72), defined by that parity; the coefficients have conjugate symmetry by (72). Its $\dot{H}^{-1}$ norm is finite. Indeed strip the common valuation u, use Lemma I.2 at level zero and sum $H^{-1+\eta}$ over dyadic radii in the primitive coordinates, then sum p^{-u} over valuations. The norm is independent of parity and nonzero. Equivalently, its square in (71) is

$$\frac{p(p+1)}{12} \sum_{v \in \mathcal{P}} |v|^{-2},$$

where $\mathcal{P}$ contains one representative up to sign of each slope in $\mathcal{C}$ with primitive integer coordinates. This uses $\sum_{l \geq 1} p^{v_p(l)}/l^2 = (1 + p^{-1})\zeta(2)$.

Take $T = \sqrt{q} \log q$ and smooth M_n on the torus by the periodized Gaussian with multiplier $\exp(-4\pi^2|j|^2/T^2)$, giving $M_n^{(T)}$. The coupling using this Gaussian has $W_2(M_n, M_n^{(T)}) = O(T^{-1}) = o(q^{-1/2})$. We prove

$$\left\| \sqrt{q}\,(M_n^{(T)} - m_3) - \sigma_\epsilon \right\|_{\dot{H}^{-1}} \longrightarrow 0 \qquad (n \equiv \epsilon \pmod 2). \tag{74}$$

On the exact cone this is dominated convergence with the summable weights just proved, for effective depth $N \geq 2$. Frequencies with $N < 2$ lie in $p^{n-1}\mathbb{Z}^3$ and are negligible even with coefficients bounded by one, by Gaussian decay. It remains to treat the nonzero congruence levels from (73). At a valuation $u \leq n - 2$ let $P = p^{n-u-1}$ and $B_0 = T/p^u$. On a primitive shell of radius H there are $O(1 + H^2/P)$ possible nonzero levels $D_I = lP$, and none for $H < c\sqrt{P}$. After multiplication by q, Lemma I.2 bounds its contribution to squared energy by

$$C_\eta p^{-u} H^{-1+\eta}(1 + H^2/P) \exp(-c'H^2/B_0^2).$$

Choose $0 < \eta < 1$. The dyadic sum at $H \geq c\sqrt{P}$ is bounded by

$$C_\eta p^{-u} P^{(-1+\eta)/2}\big(1 + (B_0/\sqrt{P})^{1+\eta}\big).$$

For the term without H^2/P this follows from the decreasing geometric sum. For the other term sum the increasing power up to B_0 and then use Gaussian decay; the same bound holds when $B_0 < \sqrt{P}$. Since $B_0/\sqrt{P} = p^{(1-u)/2} \log q$, the sum over valuations is $O_{p,h_1,h_2,h_3,\eta}(q^{(-1+\eta)/2}(\log q)^{1+\eta}) = o(1)$. This proves (74).

I.4 First variation of the transport distance

For a full-support probability measure β on a flat torus or a Euclidean cube, let

$$\|\xi\|_{-1,r;\beta} = \sup_{\|\nabla f\|_{L^{r'}(\beta)} \le 1} |\xi(f)|, \qquad 1 \le r \le 2,$$

where the tests are smooth, r' is the conjugate exponent, and the smooth tests on a cube are defined on a neighborhood of the cube. For $r = 1$ this is the Lipschitz dual norm. For $r = 2$ and Haar it is the Fourier $\dot{H}^{-1}$ norm. The equality with the Lipschitz dual at $r = 1$ follows by periodic smoothing on the torus or extension and smoothing on the cube.

Lemma I.3. *Suppose $\xi = g\beta$ with bounded real g of mean zero and $\|\xi\|_{-1,r;\beta} < \infty$. Then*

$$W_r(\beta + t\xi, \beta) = t\|\xi\|_{-1,r;\beta} + o(t), \qquad t \downarrow 0.$$

Proof. For $r = 1$ this is exactly Kantorovich–Rubinstein duality. For $1 < r \le 2$, the functional on smooth gradients given by $\nabla f \mapsto \xi(f)$ has norm $\|\xi\|_{-1,r;\beta}$ in the $L^{r'}(\beta)$ norm. Hahn–Banach and duality give a vector field $V \in L^r(\beta)$ of that norm with distributional flux $-\operatorname{div}(V\beta) = \xi$. The relative densities $1 + stg$, $0 \le s \le 1$, with flux $tV\beta$ solve the continuity equation. Its transport action gives an upper bound $(1 + O(t\|g\|_\infty))t^r\|V\|_r^r$ for W_r^r by the dynamic transport formulation [64, Theorem 8.3.1], used periodically on the torus. In the cube one can regard the flux as zero off the cube and use the equation in Euclidean space. The lower bound follows by Taylor expansion from the reference in a coupling. For each fixed smooth f,

$$t|\xi(f)| \le \|\nabla f\|_{L^{r'}(\beta)} W_r + O_f(W_r^r).$$

The compact diameter bounds the quadratic Taylor remainder by a constant times distance to the power r. On the torus use a shortest representative of the displacement. Apply the just-proved $W_r = O(t)$ and then take the supremum over f after letting $t \to 0$. □

For each parity, the classical evaluation of K_i writes the trace vector as the coordinatewise map

$$P_\epsilon(\theta) = \big(2\cos(2\pi(\theta_i + \delta_{i,\epsilon}))\big)_{i=1}^3, \qquad 4\delta_{i,\epsilon} \in \mathbb{Z},$$

with fixed shifts depending only on the selected residues and parity [59]. At even depth all shifts can be zero. At odd depth the two opposite square-root phases give either a signed cosine or a signed sine, from the quadratic Gauss phase. In particular $P_{\epsilon *} m_3 = \rho$ and $L_n = P_{\epsilon *} M_n$ along that parity. The signed pushforward $P_{\epsilon *}\sigma_\epsilon$ is defined by composing smooth scalar tests with P_ϵ. The chain rule bounds its dual gradient norm for $r = 2$ by the Lipschitz constant of P_ϵ times $\|\sigma_\epsilon\|_{\dot{H}^{-1}}$. The same bound works for $r \le 2$ since $r' \ge 2$ and the measures are probabilities.

Let $\sigma_{\epsilon,K}$ be the Fourier truncation at $|j| \le K$. For large q the density $m_3 + q^{-1/2}\sigma_{\epsilon,K}$ is at least $\frac{1}{2}m_3$. Peyre's inequality [62, Theorem 2.1] then bounds its distance in W_2 to $M_n^{(T)}$ by $2\sqrt{2}$ times their $\dot{H}^{-1}$ difference. By (74), the scaled limsup of the residual is bounded by $2\sqrt{2}\|\sigma_\epsilon - \sigma_{\epsilon,K}\|_{\dot{H}^{-1}}$. Returning to M_n costs $o(q^{-1/2})$, and the same residual after cosine pushforward is bounded up to the Lipschitz constant. Each truncated pushforward has a bounded density variation relative to ρ, since it is the pushforward of a bounded signed trigonometric polynomial times Haar. Apply Lemma I.3 to each fixed truncation, on both spaces. By Sobolev convergence of the truncation and the dual gradient bounds, one can let $K \to \infty$. We obtain the formulas

$$\begin{aligned} \lim_{n \equiv \epsilon\ (2)} \sqrt{q}\, W_r(M_n, m_3) &= \|\sigma_\epsilon\|_{-1,r;m_3}, \\ \lim_{n \equiv \epsilon\ (2)} \sqrt{q}\, W_r(L_n, \rho) &= \|P_{\epsilon *}\sigma_\epsilon\|_{-1,r;\rho}. \end{aligned} \tag{75}$$

For phase $r = 2$ the first limit squared is (71) and does not depend on ϵ.

Finally the scalar limit cannot vanish. Choose the fixed exact-cone mode j constructed above, and multiply it by 4, and, if needed, by p to make the effective depth even along the given parity. All its coordinates remain nonzero. In the expectation of the product of three Chebyshev polynomials of degrees $|j_i|$ applied to $K_i/2$, independent cosine signs average the eight sign choices for j. Only two can satisfy the stationary residue equations, because those equations force the primitive pairwise ratios and all entries are units, so only the chosen stationary vector and its simultaneous negative survive. The quarter turns give multipliers one. Both surviving exact-cone coefficients have Gaussian factor one and hence positive scaled value. The product test has zero reference mean and finite Euclidean Lipschitz constant, proving positivity for $r = 1$ and then for all $r \geq 1$. An exact-cone character similarly proves positivity on the phase torus. This completes the proof of Theorem I.1.

I.5 An obstruction for arbitrarily many shifts

For three shifts the rational conic gives a stationary direction on every active residue disk. For more shifts one can prescribe the rational point first and choose the shifts through it. The shifted-root derivative matrix in Milićević–Zhang [59, Proposition 9] uses the corresponding Vandermonde structure to control cancellation. Milićević, Robinson and Shupe [60, Remark 2 and Section 5, Example 1] construct highly vanishing one-parameter phases with congruent shifts. The construction below chooses fixed, pairwise distinct shifts and a bounded set of witness frequencies for the two-parameter transport problem. Fix $R \geq 2$ and $m \geq 3$, and put $k = m - 1$. The data in the following theorem are independent of the prime. For each modulus define the normalized rank-R trace by

$$K_R(c; q) = q^{-(R-1)/2} \sideset{}{^{\times}}\sum_{z_1 \cdots z_R = c} e_q(z_1 + \cdots + z_R).$$

Theorem I.4. *There exist distinct fixed integers $h_1, \ldots, h_m$ of the form $h_i = -r_i^R$ for positive integers r_i, such that for every sufficiently large prime p the following holds. Let a, b be independent uniform lifts of $1, 0$ modulo p, and choose the root lift x_i of $(b - h_i)/a$ in the root residue r_i modulo p. Let $\mathcal{A}_n$ be the law of the angles Rx_i/q in $\mathbb{T}^m$. There is a fixed nonzero integer vector J, independent of p, n, for which*

$$W_r(\mathcal{A}_n, m_m) \gg_{p,R,m,h} q^{-1/k} \qquad (1 \leq r \leq \infty).$$

The circle projection $\theta \mapsto J \cdot \theta \pmod 1$ has distance of order $q^{-1/k}$ from circle Haar for every $1 \leq r \leq \infty$. Its atom at the origin has mass $p^{1-\lceil n/k \rceil}$ for all sufficiently large n. If $d = \gcd(R, p-1)$ is 1 or 2, the same lower bound holds for the scalar trace vector $(K_R((b - h_i)/a; q))_i$ relative to its independent Haar phase comparison. For $R = 2$ this is the real product arcsine comparison at every allowed prime.

Proof. Take positive distinct integers r_i, set $h_i = -r_i^R$, and denote $X_i = r_i^{-R}$. Let $c_i = \prod_{j \neq i}(X_i - X_j)^{-1}$. Lagrange interpolation of monomials in degree at most k gives

$$\sum_{i=1}^{m} c_i X_i^{\ell} = 0 \quad (0 \leq \ell < k), \qquad \sum_{i=1}^{m} c_i X_i^{k} = 1.$$

Choose the r_i so that $\sum_{i \in S} c_i \neq 0$ for every proper nonempty subset S. For each such S this rational function is nonzero: move some X_i with $i \in S$ toward X_j with $j \notin S$, holding the other distinct variables fixed. Only the included c_i has a pole at that collision. The product of the

nonzero numerators and of all pairwise differences is a nonzero polynomial. It cannot vanish on the product of m copies of the infinite set $\{r^{-R} : r \in \mathbb{N}\}$, by the univariate root bound successively in each coordinate. This proves the availability of the choice.

Choose a positive integer L such that $J_i = Lc_i/r_i$ are integers. There are only finitely many excluded primes: those that divide $2Rk$ or a numerator or denominator needed below, including the nonzero subset sums, the shifts' differences and the leading coefficient. Write $t = a^{-1/R}$ on the branch $t \equiv 1 \pmod{p}$. Hensel lifting makes t uniform in its residue disk, independently of b. Choose $x_i = t(r_i^R + b)^{1/R}$ with the root branch reducing to r_i. For

$$F(b) = \sum_i J_i (r_i^R + b)^{1/R}$$

the binomial expansion and the interpolation identities give exactly $F(b) = \beta b^k (1 + bG(b))$, with $\beta = L\binom{1/R}{k} \neq 0$, a p-unit. The series G has coefficients integral at a permitted prime. Indeed $1/R \in \mathbb{Z}_p$ has p-integral binomial coefficients, by integer approximation, and r_i, β are units. The change of variable $y = b(1 + bG(b))^{1/k}$ on $p\mathbb{Z}_p$ is a measure-preserving bijection: its power series is b plus higher integral terms, so it is an isometry and lifts digits bijectively.

On the shell $v_p(y) = j$ with $j \geq 1$ and $kj < n$, the projected angle variable is $R\beta t y^k/q$ modulo one. For each fixed residue of the unit part of y, uniformity of t makes this variable uniform on an affine circle grid of mesh $p^{-(n-kj-1)}$. All remaining y project to zero with exact relative probability $p^{1-\lceil n/k \rceil}$. To bound the distribution function error on $[0, 1)$ against uniform measure, use the shell probability $(p-1)p^{-j}$ and the mesh error of each grid. This yields

$$p^{1-\lceil n/k \rceil} + \sum_{1 \leq j < \lceil n/k \rceil} (p-1) p^{-j} p^{-(n-kj-1)} \ll_{p,k} q^{-1/k}.$$

The quantile coupling on the interval has displacement bounded by this distribution function error, giving the asserted circle upper bound, also for the quotient metric.

For the lower bound choose $u \in \{0, \ldots, k-1\}$ with $N = n - u$ a multiple of k, and test the projected frequency p^u. The map above reduces its phase to $e_{p^N}(R\beta t y^k)$. If $kv_p(y) < N$, the effective depth of the unit variable is at least $k \geq 2$. Its last digit cancels the sum since the derivative of the kth power is a unit there. Here one can use the normalized uniform variable y/p^j on the units at every such shell, independently of t. The ball $v_p(y) \geq N/k$ contributes one pointwise. Thus the Fourier average is the positive real number $p^{1-N/k}$. The frequencies p^u lie in a fixed finite set as n varies for fixed p. A real character test therefore gives the corresponding W_1 lower bound on the circle and, by projection, on the angle torus. Monotonicity proves all claimed lower bounds for the phase law.

At $d = 1$ the scalar trace is a circle phase up to a unit depending only on the residue and parity, so the same test applies. At $d = 2$ the two stationary branches make each trace a translated cosine times 2, up to a fixed unit rotation. This follows from the tame Hessian weights of the root branches [31]. Use the product Chebyshev test at frequency vector $4p^u J$, so all quarter-turn shifts disappear. The angle phase for a sign vector (ϵ_i) replaces F by $\sum_i \epsilon_i J_i (r_i^R + b)^{1/R}$. If the signs are mixed, its constant term is a nonzero fixed multiple of a proper subset sum of the c_i. This is a unit at the permitted primes, and the average over t in its fixed unit residue vanishes exactly at depth $N \geq 2$. For both constant sign vectors the average is $p^{1-N/k}$, since the additional factor 4 and a common sign do not affect last-digit cancellation. The Chebyshev test therefore has arithmetic mean $2^{1-m} p^{1-N/k}$ and Haar mean zero. Its Lipschitz constant is bounded as n varies. This proves the scalar claim. □

The lower bound of Theorem I.4 exceeds the size-only scale $q^{-2/m}$ for the single-branch phase vector, for every $m \geq 3$. The exact atom also gives a local counting obstruction: the arithmetic mass of any tube containing the origin fiber $J \cdot \theta = 0$ is at least $p^{1-\lceil n/k \rceil}$, while Haar mass of the tube $\{\|J \cdot \theta\|_{\mathbb{T}} \leq \delta\}$ is 2δ for $0 < \delta < 1/2$.

For example, in rank two take shifts $(-1, -4, -9, -16)$ and $J = (7, -112, 243, -128)$. The weighted derivative identities at $b = 0$ are immediate from these four squares, and the proper subset sums of $(7, -224, 729, -512)$ are all nonzero. At $p = 11$ all four squares are distinct modulo the prime, the leading cubic coefficient is a unit, and none of the proper subset sums is divisible by 11. This gives a concrete four-coordinate scalar obstruction of order $q^{-1/3}$.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, including the Fourier support calculation and proof of the transport limit, and with drafting and checking the argument. No data are needed for the results.

J Transport constants for resonant circle projections

Consider the phase $\beta t y^k / p^n$ modulo one, where t lifts 1 and y lifts 0 modulo p. This law occurs in character projections of shifted Kloosterman phase vectors near a rational stationary point. A convergent series of distribution-function defects gives the transport constants to circle Haar for every Wasserstein order along each depth class. For $k = 2$, the quadratic distance is exactly $W_2^2 = p(p+1)/(12q) - p^3/(12q^2)$. In other orders, even depths have a deterministic large-prime profile while odd depths are governed by the Legendre path, giving joint limiting laws and exceptional bottleneck growth. Taking the cosine removes the quadratic-residue contribution at primes 3 modulo 4 and yields a folded odd-depth scale of $\sqrt{p/q}$.

Keywords: stationary phase, Wasserstein distance, Legendre paths, prime powers.
MSC 2020: 11L40, 11K65, 60B15.

J.1 The valuation model

Let p be an odd prime, let $k \geq 2$ be an integer, and fix $\beta \in \mathbb{Z}_p^\times$. For $n \geq 1$ and $q = p^n$ let $\mu_{p,n}$ be the circle law of $\beta t y^k / q$, with t and y independent and uniform on the respective lifts of 1 and 0 modulo p. All division by q in such a circle law uses reduction modulo q. Write m_1 for Haar measure on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, with its geodesic distance. For a bounded real function f on $[0,1]$ use the quotient norm

$$\mathcal{N}_r(f) = \inf_{a \in \mathbb{R}} \|f - a\|_{L^r([0,1])}, \qquad 1 \leq r \leq \infty.$$

Let $g(x) = 1/2 - x$ on $(0,1)$, extended periodically. Let $f = f_{p,k,\beta}$ be the centered distribution function minus x of the circle law uniform on the residue coset $\beta(\mathbb{F}_p^\times)^k$, embedded as points at the representative residues divided by p. Centering always means subtracting the integral on $[0,1]$.

Theorem J.1. *Put $C = \lceil n/k \rceil$ and $d = kC - n$. For $0 \leq d \leq k-1$ define the uniformly convergent series*

$$D_d(x) = p^{-d/k} \left[pg(x) + (p-1)p^{d+1} \sum_{\ell \geq 1} p^{-(k-1)\ell} f\big(p^{k\ell - d - 1} x\big) \right]. \tag{76}$$

Let $D_{d,n}$ denote the same expression truncated at $1 \leq \ell < C$. Then for every $1 \leq r \leq \infty$,

$$W_r(\mu_{p,n}, m_1) = q^{-1/k} \mathcal{N}_r(D_{d,n}) = q^{-1/k} \mathcal{N}_r(D_d) + O(p^2/q). \tag{77}$$

The error bound is uniform in the displayed parameters, and each $\mathcal{N}_r(D_d)$ is positive. In particular the constants at fixed p, k, β are explicit and indexed by $n \bmod k$.

If $\gcd(k, p-1) = 1$, then every unit lies in the residue coset and the distances reduce to the exact formula

$$W_r(\mu_{p,n}, m_1) = p^{1-C}\|g\|_r \qquad (1 \le r \le \infty). \tag{78}$$

The model also has a local stationary-phase interpretation. If an integral analytic function on $p\mathbb{Z}_p$ satisfies $F(b) = \beta b^k(1 + bG(b))$ with $p \nmid k$, then $b \mapsto y = b(1 + bG(b))^{1/k}$ preserves the Haar measure of $p\mathbb{Z}_p$. Its linear term is one and its higher terms are integral, so it is a bijective isometry by digit lifting. Hence the circle law of $tF(b)/q$ with $b \equiv 0$ and $t \equiv 1$ is $\mu_{p,n}$. In the shifted-root phase of a rank-R Kloosterman family with a free coefficient $a \equiv 1$, the factor $a^{-1/R}$ is exactly such a uniform t when $p \nmid R$. For an explicit quadratic example, take the classical shift triple $(-1, -4, -9)$ and mode $(5, -16, 9)$. At $b = 0$ the branch roots are $1, 2, 3$, and this linear combination of their three square-root functions has zero constant and first derivative and second-order coefficient $-5/12$. The factor 2 in the angles gives $\beta = -5/6$, so the model holds for every prime $p > 5$ on the selected residues.

J.2 Exact circle transport and shell decomposition

Lemma J.2. *For a circle probability law μ with distribution function F on $[0, 1)$, one has $W_r(\mu, m_1) = \mathcal{N}_r(F - \mathrm{id})$ for every $1 \le r \le \infty$.*

Proof. Extend F by $F(x + 1) = F(x) + 1$ and use a right-continuous version and the lift $Q(u) = \inf\{x : F(x) \ge u\}$. The circular quantile formula of Rabin, Delon and Gousseau [65, Theorem 1 and Eq. (9)], or the general measure theory in Delon, Salomon and Sobolevski [66], gives the infimum over constants of $\|Q(u) - u - a\|_r$ for each finite r. For uniform U on $[0, 1)$, the inequality $Q(U) - U > s$ is equivalent almost surely to $F(U + s) < U$. By the degree-one periodicity this gives the law of $-(F(U) - U)$ for $Q(U) - U$. This argument allows atoms. It follows that the two quotient norms agree. For $r = \infty$ increase finite r to infinity: on the compact circle weak limits of optimal couplings give $W_r \uparrow W_\infty$, and minimizers in the quotient norms lie in a compact interval between the essential endpoints of $F - \mathrm{id}$, so the quotient norms have the same limit property. □

If $j = v_p(y) \ge 1$ and $kj < n$, the circle law conditioned on this shell is the distribution η_N uniform on all lifts of the same residue coset modulo p^N, with $N = n - kj$. For any fixed unit part of y modulo p, multiplication by t is uniform over all lifts of its resulting unit residue. Every coset element has the same number of preimages by the kth-power map on $\mathbb{F}_p^\times$. Thus η_N consists of p^{N-1} equal-weight copies of η_1 rescaled to successive intervals of length p^{1-N}, and its centered distribution-function discrepancy is $p^{1-N}f(p^{N-1}x)$. The shell has weight $(p-1)p^{-j}$. The complement is the atom at zero, of mass p^{1-C}, with centered discrepancy g. Put $\ell = C - j$ on the shells. Multiplying the resulting centered defect by $q^{1/k}$ yields exactly $D_{d,n}$. This proves the first identity in (77).

The bound $\|f\|_\infty \le 2$ and the geometric series from $\ell = C$ give $q^{-1/k}\|D_{d,n} - D_d\|_\infty \ll p^2/q$. The maps $\mathcal{N}_r$ are all 1-Lipschitz in the supremum norm, giving the error claimed. The one-sided limits of each replica term from the two ends of $[0, 1]$ agree, since η_1 has no mass at zero. This persists under uniform summation. The term from g gives a jump of height $p^{1-d/k}$, so D_d is nonconstant even modulo null sets. This proves Theorem J.1.

For (78), let f_{u} be the centered defect of the uniform nonzero p-grid. Direct counting gives $(p-1)f_{\mathrm{u}}(x) = g(px) - g(x)$ almost everywhere. Put $N_\ell = k\ell - d$. For uniform x, write $g(x) =$

$\sum_{j\ge1} p^{-j}Z_j$, where the Z_j are its independent centered digit complements, uniform on $\{(p-1)/2-a : 0 \le a < p\}$. The expression $D_{d,n}/p^{1-d/k}$ is then

$$g(x) + \sum_{1\le\ell<C} p^{d-(k-1)\ell}(g(p^{N_\ell}x) - g(p^{N_\ell-1}x)).$$

It deletes precisely the digits in positions N_ℓ. Indeed the bracket multiplying $p^{-j}Z_j$ starts at one. Each level with $N_\ell < j$ adds $(p-1)p^{\ell-1}$, and if $N_\ell = j$ the level instead subtracts $p^{\ell-1}$, canceling the current bracket. A surviving digit j has weight $p^{-j+\#\{\ell:N_\ell<j\}}$. Deleting those fixed positions preserves the independent uniform law of the remaining digits, so $D_{d,n}/p^{1-d/k}$ has exactly the same value distribution as g. The first identity of (77) gives (78).

There is also a finite formula for the quadratic constant. Write $A = \|f\|_2^2$ and $B = \langle f, g\rangle$. With U, V independent with law η_1 and $B_2(x) = x^2 - x + 1/6$ for $0 \le x \le 1$, the values are

$$A = \tfrac{1}{2}\mathbb{E}B_2(\{U-V\}), \qquad B = \tfrac{1}{2}\mathbb{E}B_2(U).$$

Indeed the centered CDF has Fourier coefficients $\widehat{\eta}_1(j)/(2\pi ij)$ for $j \ne 0$, taking Fourier transforms with $e^{-2\pi ijx}$. The law η_1 lies on the p-grid, so $\widehat{f}(pj) = \widehat{g}(j)/p$. The Bernoulli identities follow from Parseval and the Fourier series of B_2. Set $b = (p-1)p^{d+1}$ and $a = p^{-(k-1)}$. Then

$$\mathcal{N}_2(D_d)^2 = p^{-2d/k}\left[\frac{p^2}{12} + \frac{b^2a^2}{1-a^2}\left(A + \frac{2B}{p-1}\right)\right]. \tag{79}$$

For the finite formula replace $b^2a^2/(1-a^2)$ by $\sum_{1\le\ell<C} b^2a^{2\ell}$. To see this, the correlation of $f(p^sx)$ with $f(p^{s'}x)$ for $s' > s$ is $p^{-(s'-s)}B$, and that of $g(x)$ with $f(p^sx)$ is $p^{-s}B$. With $s_\ell = k\ell - d - 1$ and $b_\ell = ba^\ell$, a geometric sum gives

$$p^{1-s_\ell} + \sum_{j<\ell} b_j p^{-k(\ell-j)} = \frac{b_\ell}{p-1}.$$

Hence each new replica adds $b_\ell^2(A + 2B/(p-1))$ to the variance inside the bracket of (76), as claimed.

J.3 A Legendre path at odd depths

We now let $k = 2$ and let the prime vary. Denote by P_p the polygonal Legendre path, whose partial-sum vertices have heights $p^{-1/2}\sum_{a\le j}(a/p)$, with the usual normalization on $[0,1]$ [67]. The choice of using p or $p-1$ for the time mesh changes the function by $O(p^{-1/2})$ uniformly. The following estimates apply uniformly to every unit β and transport order $1 \le r \le \infty$.

Theorem J.3. *For even $n \ge 2$ and odd $n \ge 3$, respectively,*

$$\frac{\sqrt{q}}{p}W_r(\mu_{p,n}, m_1) = \|g\|_r + O(p^{-1/2}\log p), \tag{80}$$

$$\frac{\sqrt{q}}{p}W_r(\mu_{p,n}, m_1) = \mathcal{N}_r(P_p) + O(p^{-1/2}). \tag{81}$$

The ranges and errors also hold if n or β vary with p. For all odd primes and all $n \ge 1$ the quadratic distance satisfies the exact identity

$$W_2(\mu_{p,n}, m_1)^2 = \frac{p(p+1)}{12q} - \frac{p^3}{12q^2}. \tag{82}$$

In particular along all primes and depths $n \geq 2$ the normalized W_2 tends to $1/\sqrt{12}$. For any odd depth choice $n = n(p) \geq 3$ and primes uniform on $[X, 2X]$, the random curve of normalized distances indexed by $r \in [1, \infty]$ converges in law, with the supremum norm on these curves, to $r \mapsto \mathcal{N}_r(P)$ for the Hussain–Lamzouri Legendre limit P. One can restrict primes to either class modulo 4 and use their corresponding limit. The even-depth curve instead tends uniformly to $r \mapsto \|g\|_r$.

Proof. At $k = 2$, let χ_p be the Legendre symbol. The indicator of the coset is $(1 + \chi_p(\beta)\chi_p(a))/2$ on $a \neq 0$. Its cardinality is $(p-1)/2$. Counting the initial interval of representatives therefore gives, modulo an additive constant,

$$\sqrt{p}\, f_{p,2,\beta} = \chi_p(\beta) P_p + O(p^{-1/2}) \tag{83}$$

uniformly. Indeed the uniform-unit part of the counting discrepancy is $O(p^{-1})$ before scaling, and the path interpolation crosses at most a bounded number of increments of size $p^{-1/2}$. The change from $\sqrt{p}/(p-1)$ to $1/\sqrt{p}$ is absorbed by the trivial character-sum bound p.

For even depths $d = 0$, the coefficient of g in $D_{0,n}/p$ is one, and the sum of the absolute replica coefficients is at most one. The ordinary finite Fourier completion of an interval character sum gives $\|f\|_\infty \ll (\log p)/\sqrt{p}$: the Gauss coefficients have modulus $\sqrt{p}$ and the sum of the moduli of the finite geometric sums over the nonzero frequencies is $O(p \log p)$. This proves (80), since centering g minimizes its quotient norm by symmetry for all r. For odd depths $d = 1$ the $\ell = 1$ term is included. Its coefficient in $D_{1,n}/p$ is $(1 - 1/p)\sqrt{p}$, the cumulative coefficient of later terms is at most $p^{-1/2}$, and that of g is $p^{-1/2}$. Hence $D_{1,n}/p = \sqrt{p}\, f + O(p^{-1/2})$ uniformly, using the trivial bound on f. This yields (81) by (83) and invariance of $\mathcal{N}_r$ under constants and sign.

For the exact W_2 formula let A, B be as in (79). At Fourier frequencies prime to p, the quadratic Gauss sum gives $\widehat{\eta}_1(j) = (-1 + Z_j)/(p-1)$ with $|Z_j| = \sqrt{p}$. Hence $|\widehat{\eta}_1(j)|^2 + 2\Re\widehat{\eta}_1(j)/(p-1) = 1/(p-1)$. At multiples of p the coefficient is one. Summation with the $1/(4\pi^2 j^2)$ weight gives $A + 2B/(p-1) = (p+1)/(12p(p-1))$. The finite form of (79) therefore gives

$$qW_2^2 = \frac{p^{2-d} + p^{d+1}(1 - p^{-2(C-1)})}{12} = \frac{p^2 + p - p^3/q}{12}.$$

This proves (82); the cancellation is independent of the Gauss phase and of the chosen coset.

Finally apply Hussain–Lamzouri [67, Theorem 1.2 and the parity discussion on p. 595]. With topology induced by $1/r$, the interval $[1, \infty]$ is compact. For a continuous bounded path P, the function $r \mapsto \mathcal{N}_r(P)$ is continuous, including at ∞ by compactness of the minimizing constants. The map from paths to these curves is 1-Lipschitz in the supremum norm. Thus their uniform-topology convergence in law and (81) give the stated joint conclusion, including bottleneck transport. □

J.4 Exceptional prime sequences

The odd-depth law also admits a deterministic classification for finite transport orders. Let s be any completely multiplicative function on the positive integers with values ± 1. For $\epsilon \in \{1, -1\}$ extend it to negative integers by $s(-m) = \epsilon s(m)$. Define a mean-zero real function $V_{\epsilon,s}$ by its Fourier coefficients at $m \neq 0$

$$\widehat{V}_{\epsilon,s}(m) = \frac{\gamma_\epsilon s(m)}{2\pi i m}, \qquad \gamma_1 = 1, \quad \gamma_{-1} = -i. \tag{84}$$

It lies in every finite L^r, by inverse Hausdorff–Young and Parseval.

Theorem J.4. *For primes restricted by $\chi_p(-1) = \epsilon$, every subsequential limit of the odd-depth normalized finite-order transport curve on $[1, R_0]$, for any fixed finite $R_0 \geq 1$, has the form $r \mapsto \mathcal{N}_r(V_{\epsilon,s})$ for some s as above. Every such curve occurs. In particular the all-positive patterns give normalized W_1 limits $1/4$ at $p \equiv 1 \pmod 4$ and $2G/\pi^2$ at $p \equiv 3 \pmod 4$, where G is Catalan's constant.*

For the odd-depth normalized bottleneck distance the general upper bound is $O(\log p)$, and there is a sequence of primes $p \equiv 3 \pmod 4$ on which it is at least $(1/(2\pi) + o(1)) \log\log p$, uniformly at all odd depths $n \geq 3$.

Proof. For $p \nmid m$, the Fourier coefficient of η_1 is

$$\widehat{\eta}_1(m) = \frac{-1 + \chi_p(\beta)\chi_p(m)\gamma_{\chi_p(-1)}\sqrt{p}}{p-1}.$$

Here the Gauss phase uses the negative Fourier convention. For $p \mid m$ the coefficient is one. Multiply the centered function $\sqrt{p}f$ by $\chi_p(\beta)$, a sign that leaves its quotient norms unchanged. Its fixed nonzero Fourier coefficients tend to (84) if the positive Legendre symbols stabilize pointwise. All tails above H in finite L^r are uniformly $O_r(H^{-1/r} + p^{-1/2})$ when $r \geq 2$, by inverse Hausdorff–Young on the nonalias spectrum and the separate $O_r(p^{-1/2})$ alias norm. For $1 \leq r \leq 2$ use the L^2 bound $O(H^{-1/2} + p^{-1/2})$. Thus stabilization gives convergence in $L^{\max(R_0,2)}$, which gives uniform convergence of all the quotient norms on $[1, R_0]$. Diagonal compactness gives a stabilized completely multiplicative sign function along a further subsequence of any sequence of primes. Conversely prescribe any finite set of signs on primes ℓ. Quadratic reciprocity translates these into reduced residue conditions on p modulo the odd ℓ, and the sign at 2 and parity at -1 can be prescribed together modulo 8. The Chinese remainder theorem and Dirichlet's theorem yield primes realizing the finite list, and a diagonal sequence realizes s.

For $s \equiv 1$ and $\epsilon = 1$, the function in (84) is g and $\mathcal{N}_1(g) = 1/4$. For $\epsilon = -1$ it is $\pi^{-1}\log(2\sin\pi x)$ on $0 < x < 1$, by the usual logarithmic Fourier series. Its median is $\pi^{-1}\log\sqrt{2}$, and its mean is zero. Hence its quotient L^1 norm is $-4\pi^{-1}\int_0^{1/4}\log(2\sin\pi x)\,dx = 2G/\pi^2$, from integration of the Fourier series and $G = \sum_{j\geq 0}(-1)^j/(2j+1)^2$.

For the bottleneck lower, prescribe $\chi_p(-1) = -1$ and $\chi_p(m) = 1$ at every $1 \leq m \leq H$. The residue conditions are imposed modulo $8\prod_{3\leq \ell \leq H,\ \ell \text{ prime}} \ell$, whose logarithm is $O(H)$ by Chebyshev. Linnik's theorem (for example [69, Theorem 1.1]) permits a choice $H < p \leq \exp(O(H))$ since every prime at most H is excluded by coprimality. Convolve the centered, sign-oriented $\sqrt{p}f$ with the Fejér kernel of degree H and evaluate at zero. On these positive and negative frequencies there are no aliases, and (84) with the elementary Gauss formula gives the value $-\pi^{-1}\log H + O(1)$. The mean is zero and the kernel is positive with mass one, so the oscillation of the original function is at least $\pi^{-1}\log H - O(1)$. Half its essential oscillation is $\mathcal{N}_\infty$. The normalized distance differs from this norm by $O(p^{-1/2})$ by the odd-depth argument. Thus the lower follows using $\log H \geq \log\log p - O(1)$, and the upper is the same finite Fourier completion as in Theorem J.3. This gives a Paley-type bottleneck phenomenon, consistent with the distribution of large quadratic character sums studied by Lamzouri [68]. □

J.5 What survives a cosine projection

Let $\Psi(x) = 2\cos(2\pi x)$ and let $\alpha = \Psi_* m_1$, the arcsine law on $[-2, 2]$. This section concerns the cosine of the one-dimensional resonant character. Write $I = [0, 1/2]$, let m_I be its normalized Lebesgue measure and put $w(u) = 4\pi\sin(2\pi u)$ on I. We continue with $k = 2$.

Theorem J.5. *At all even depths $n \geq 4$, as $p \to \infty$,*

$$\frac{\sqrt{q}}{p} W_r(\Psi_* \mu_{p,n}, \alpha) = \|wg\|_{L^r(m_I)} + O(p^{-1/2} \log p)$$

uniformly for $1 \leq r \leq \infty$. At odd depths $n \geq 3$ the prime classes have different scalings. When $p \equiv 1 \pmod 4$,

$$\frac{\sqrt{q}}{p} W_r(\Psi_* \mu_{p,n}, \alpha) = \|wP_p\|_{L^r(m_I)} + O(p^{-1/2} \log^2 p)$$

uniformly in r. Thus the conditioned Legendre limit gives the simultaneous limiting law for these weighted norms. When $p \equiv 3 \pmod 4$, for each fixed $1 \leq r \leq \infty$,

$$\frac{\sqrt{q}}{\sqrt{p}} W_r(\Psi_* \mu_{p,n}, \alpha) \quad \longrightarrow \quad \|w\|_{L^r(m_I)} \|g\|_{L^r([0,1])}.$$

The limits and bounds are uniform in the specified depths and in the unit β. In particular the last bottleneck limit is 2π.

At fixed $p \equiv 3 \pmod 4$ and odd $n \geq 3$, the quadratic distance satisfies the exact identity

$$W_2(\Psi_* \mu_{p,n}, \alpha)^2 = 4 \left(1 - \operatorname{sinc} \left(\pi \sqrt{p/q} \right) \right), \qquad \operatorname{sinc}(z) = \frac{\sin z}{z}.$$

Proof. Let $H(u)$ be the distribution function of $U = \min(X, 1 - X)$ for $X \sim \mu_{p,n}$. The reference angular law is m_I, with distribution function $2u$. Up to endpoint conventions irrelevant to Lebesgue norms, $H(u) - 2u$ is the difference of the two original centered defects at u and $1 - u$. Set $M = \|H - 2\,\mathrm{id}\|_\infty$. For a general law H on I, monotone transport on the real line and the quantile inverse $Q_H(t)$ give

$$W_r(\Psi_* H, \alpha) = \tfrac{1}{2} \|w(u)(H(u) - 2u)\|_{L^r(m_I)} + O(M^2). \tag{85}$$

Here $\Psi_* H$ denotes the pushforward of the law with distribution function H. To prove (85), match $Q_H(t)$ to $t/2$ and use that Ψ is decreasing. The supremum of $|Q_H(t) - t/2|$ is at most $M/2$, so Taylor expansion replaces the transported trace displacement by $w(t/2)|Q_H(t) - t/2|$ with uniform error $O(M^2)$. For finite r, raise to power r inside its norm and use the measure on the region between $u = Q_H(t)$ and $u = t/2$ with density $r|u - t/2|^{r-1}\, du\, dt$. Its total mass is at most $(M/2)^r$. Replacing the weight $w(t/2)$ by $w(u)$ in the resulting norm costs $O(M^2)$, since w is Lipschitz and $|u - t/2| \leq M/2$. For a fixed u the vertical slice runs between $t = 2u$ and $t = H(u)$, up to null sets. Integrating this slice gives $2(|H(u) - 2u|/2)^r$, which gives (85). All the norm errors are uniform in r, and the $r = \infty$ case follows by increasing finite r.

At even depth, $D_{0,n}/p = g + O(p^{-1/2} \log p)$ uniformly as in the proof of Theorem J.3. Its reflected defect is therefore $2g + O(p^{-1/2} \log p)$ under the same normalization. Since $M = O(p/\sqrt{q})$, the normalized quadratic error in (85) is $O(p/\sqrt{q})$, which is $O(p^{-1})$ for $n \geq 4$. This gives the even assertion.

When the depth is odd and $p \equiv 1 \pmod 4$, the unit coset is invariant under reflection. The polygonal Legendre path in the mesh of Hussain–Lamzouri is odd under $u \mapsto 1 - u$ [67, Remark 1.1], and has mean zero. Formula (83) and the odd-depth estimate of $D_{1,n}$ show that the reflected defect, under the normalization $\sqrt{q}/p$, is twice this path up to sign and an error $O(p^{-1/2})$. The completion bound gives $M = O(p \log p/\sqrt{q})$. The normalized quadratic error in (85) is therefore $O(p \log^2 p/\sqrt{q})$, at most $O(p^{-1/2} \log^2 p)$ at odd depths $n \geq 3$. As before, the map taking the path to all its weighted norms is continuous in the supremum norm, so the parity-conditioned prime limit applies.

Finally when $p \equiv 3 \pmod 4$, the cosine identifies the two opposite quadratic residue cosets. Symmetrizing the phase law gives, at each shell, the uniform law on all unit residues modulo p^N. Thus one can replace f by the centered defect f_{u} of the uniform nonzero p-grid in (76). It satisfies $\|f_{\mathrm{u}}\|_\infty = O(p^{-1})$ and $(p-1)f_{\mathrm{u}}(x) = \lfloor px \rfloor - (p-1)x$ almost everywhere, since the discrepancy is already centered. At an odd depth $d = 1$, dividing the truncated $D_{1,n}$ by $\sqrt{p}$ leaves $g + (p-1)f_{\mathrm{u}}(x)$ plus $O(p^{-1})$ from later replicas, uniformly. The first two terms are exactly $g(px)$ almost everywhere. This is odd under reflection up to null points, so (85) gives

$$\frac{\sqrt{q}}{\sqrt{p}} W_r(\Psi_* \mu_{p,n}, \alpha) = \|w(u) g(pu)\|_{L^r(m_I)} + O(p^{-1}),$$

using $M = O(\sqrt{p}/\sqrt{q})$ and $n \geq 3$. For any finite r, partitioning into intervals of length $1/p$ and using uniform continuity of the fixed weight w^r proves that the last norm tends to $\|w\|_r \|g\|_r$. For $r = \infty$, the product has upper bound $\|w\|_\infty \|g\|_\infty$, and arbitrarily near a maximizer of w there are full periods of $g(pu)$ for all sufficiently large p. This gives the matching limit.

□

To prove the exact identity, suppose more generally that the phase law after symmetrization has the full-unit shell law and that $n \equiv 1 \pmod k$ with $n > k$. Then $N_1 = 1$ is deleted. Let $F^\sharp(x)$ be the CDF lift with centered defect of this symmetrized law, and $Q^\sharp$ its quantile lift. By symmetry $H(u) = 2F^\sharp(u)$ on the half circle up to null points: subtracting half the atom at zero centers the defect. Thus $Q_H(t) = Q^\sharp(t/2)$ almost everywhere. The displacement $E(x) = Q^\sharp(x) - x$ is $1/p$-periodic, since the deleted first digit makes $F^\sharp(x) - x$ $1/p$-periodic. It is reflection odd almost everywhere, by the symmetry $F^\sharp(x) + F^\sharp(1-x) = 1$ almost everywhere. By the quantile argument of Lemma J.2 and digit deletion, its value at a uniform x is distributed as V uniform on $[-h, h]$, where $h = p^{1-C}/2$.

By monotone transport for the folded law and reflection, its rth power distance is the integral over the full circle of $|\Psi(x + E(x)) - \Psi(x)|^r$. That integrand is

$$4^r |\sin(\pi E(x))|^r |\sin(2\pi x + \pi E(x))|^r.$$

For an even integer $r < 2p$ the last sine power has only even harmonics of frequencies at most r in x, with coefficients depending $1/p$-periodically on $E(x)$. Each nonconstant harmonic integrates to zero against the other factors because its frequency is not divisible by p. Consequently, with U uniform on the circle, one obtains the exact formula

$$W_r(\Psi_* \mu_{p,n}, \alpha) = 4 \|\sin(\pi V)\|_r \|\sin(2\pi U)\|_r \qquad (r \text{ even},\ 2 \leq r < 2p).$$

For $k = 2$, odd $n \geq 3$ and $p \equiv 3 \pmod 4$ the hypotheses hold and $h = \sqrt{p/q}/2$. Evaluation at $r = 2$ gives the claimed sinc formula.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, including the valuation and transport calculations, and with drafting and checking the proof. No data are needed for the results.

K Exact functional quantization for weighted stationary sequences

We determine the leading finite-support approximation constant for Hilbert-valued series with regularly varying squared weights and compact stationary ergodic coefficients. The constant is

an integral of the process rate-distortion objective. It holds for every fixed distortion moment and every fixed excess-distortion probability. The converse controls one global codebook by conditional block entropies across several frequency bands. For arbitrary stationary laws, one universal sequence of codebooks attains the distribution of component constants. For bounded Fourier series, dependence can strictly lower the constant while keeping the one-coordinate law fixed.

K.1 The constant

Let $X = (X_j)_{j\ge 1}$ be a stationary ergodic sequence in a fixed compact subset K of $\mathbb{R}^d$. Let C denote its closed convex hull. Let $\lambda_j > 0$ with $\lambda_j \sim \phi(j)$, where ϕ is positive and regularly varying of index $-b$, $b > 1$. The weighted vector $Y = (\sqrt{\lambda_j} X_j)_{j\ge 1}$ belongs to $\ell^2(\mathbb{R}^d)$. For $s \in [1,\infty)$ let $e_{N,s}(Y)$ be the infimum over codebooks of size at most N of the L^s norm of distance to the codebook. Let $d_N(\eta; Y)$ be the optimal squared-distance threshold exceeded with probability at most $\eta \in (0,1)$.

With logarithms natural, put

$$F_m(a) = \frac{1}{m} \inf \left\{ I(X_1^m; V_1^m) + a\,\mathbb{E} \sum_{j=1}^{m} \|X_j - V_j\|^2 \right\}, \qquad a > 0,$$

where $X_1^m = (X_1, \dots, X_m)$, and the infimum is over channels with V_1^m in C^m. Write $F(a) = \lim_{m\to\infty} F_m(a)$ and

$$J = \int_0^\infty F(u^{-b})\, du, \qquad \kappa_b = \frac{(b-1)^{b-1}}{b^b} J^b.$$

Theorem K.1. *The displayed limit $F(a)$ exists for every $a > 0$, and $J < \infty$. For every fixed $s \in [1,\infty)$ and $\eta \in (0,1)$,*

$$\lim_{N\to\infty} \frac{e_{N,s}(Y)^2}{(\log N)\phi(\log N)} = \lim_{N\to\infty} \frac{d_N(\eta; Y)}{(\log N)\phi(\log N)} = \kappa_b.$$

If $J > 0$, then for any $c < \kappa_b$ the largest mass covered by N balls of squared radius $c(\log N)\phi(\log N)$ tends to zero.

When the coefficients are independent and identically distributed, $F_m = F_1$ by information tensorization and product channels. The theorem then recovers the exact bounded-coordinate formula for the Kloosterman limit in Section B. Luschgy and Pagès [70] proved the corresponding sharp Gaussian Hilbert-space formula from the covariance eigenvalues. In the stationary setting the objective here depends on the law of the sequence. The coding input is the classical stationary ergodic block theorem [71, Section 11.4]. The converse uses the block-entropy method underlying sample converses for lossy coding [72, 74], with one code budget across unequal coordinate penalties.

K.2 Rate and distortion profiles

The sequence $mF_m(a)$ is subadditive. Indeed, use a product of two channels on the adjacent blocks of lengths m and n. Because the outputs are conditionally independent given the input blocks, the joint mutual information is at most the sum of the individual mutual informations. For example, bound it by the conditional relative entropy against the product of the two output marginals. Stationarity makes the second input block law the length-n law. The distortion terms add. Fekete's lemma gives the limit. Projection of reproductions onto C is harmless. A constant reproduction and a fixed Euclidean grid give, uniformly in m,

$$0 \le F_m(a) \le C_0 \min\{a, 1 + \log(1+a)\}.$$

For the latter bound use a one-coordinate grid of mesh $(1+a)^{-1/2}$. The information per coordinate is at most the logarithm of the grid size. The same bounds hold for F and make J finite.

The limit F is increasing and concave, and locally absolutely continuous on $(0,\infty)$. Where it is differentiable, set $D(a) = F'(a)$ and $I(a) = F(a) - aD(a)$. If a sequence of length-m channels has cost at a within $o(1)$ per coordinate of $F_m(a)$, comparison at $a+h$ and $a-h$ confines its distortion to the limiting difference quotients of F. Thus its normalized distortion and information converge to $D(a)$ and $I(a)$. Both are nonnegative. The stationary block coding theorem gives, for each fixed slack $\gamma > 0$, deterministic length-n codebooks of cardinality at most $\exp(n(I(a)+\gamma))$ whose average squared distortion exceeds $D(a)+\gamma$ with probability tending to zero. Indeed the finite-block information-distortion infimum is at most the distortion of any block channel with that information budget, and subadditivity extends the pair to the process infimum. The block theorem with slack then applies. Its random-coding proof gives the stated probability form directly [71, Equation (11.14)].

The integrands $I(u^{-b})$ and $u^{-b}D(u^{-b})$ are Riemann integrable on compact subintervals of $(0,\infty)$ using one-sided slopes at exceptions. Integration by parts and the bounds on F give

$$\int_0^\infty u^{-b} D(u^{-b})\, du = J/b, \qquad \int_0^\infty I(u^{-b})\, du = (1-1/b)J =: B_0. \tag{86}$$

K.3 One code budget across several bands

We first give the converse at an arbitrary scale $L \to \infty$. Fix $0 < \alpha < T < \infty$ and split $[\alpha, T]$ into finitely many intervals I_i of lengths $\ell_i > 0$. For each choose a positive $a_i < \inf_{u \in I_i} u^{-b}$. Write B_i for the consecutive coordinate indices j with $j/L \in I_i$, using any disjoint endpoint convention. By uniform convergence of regular variation, for all large L, $\lambda_j/\phi(L) \geq a_i$ on B_i. The union B is a consecutive block of size $(T-\alpha)L + O(1)$.

Suppose along a sequence of codebooks there is probability bounded below by $q_* > 0$ of squared error at most t_L. On this event E_L let $\widehat{X}_j$ be the selected codeword coordinate divided by $\sqrt{\lambda_j}$ and projected to C. This does not increase the errors. The projected vector still takes at most N possible values. Round both X_j and $\widehat{X}_j$ to a common finite δ-net of C. Denote the rounded coefficients by Z_j, and the rounded reproductions by U_j. Pointwise on E_L,

$$\sum_i a_i \sum_{j \in B_i} \|Z_j - U_j\|^2 \leq t_L/\phi(L) + C_{\alpha,T}\delta L. \tag{87}$$

This follows from the bounded support and the $O(\delta)$ change of each squared distance. The rounded source is stationary and ergodic. Let h_δ be its entropy rate and let P_m be its length-m marginal. The entropy of Z_B conditional on E_L satisfies

$$H(Z_B \mid E_L) \geq h_\delta |B| - o(L).$$

In fact its conditional probability at any word is at most its original probability divided by q_*. The Shannon–McMillan–Breiman theorem gives typical original word probabilities at most $\exp[-(h_\delta - o(1))|B|]$, using stationarity for this moving block and convergence in probability. Conditioning on events of probability at least q_* preserves the typical-set probability tending to one. Taking the conditional expected negative log probability proves the entropy bound, with at most $\log(1/q_*)$ lost in changing the point probabilities.

Fix m. For each B_i, let $\pi_{i,L}$ be the mean, over all starts of length-m subblocks lying in B_i, of the joint law of the corresponding Z, U blocks conditional on E_L. The source marginal of $\pi_{i,L}$ tends to P_m. Indeed the empirical frequencies of all overlapping m-words tend to their stationary

probabilities on B_i, by subtracting two prefix ergodic sums. Again conditioning on E_L preserves convergence in probability and the bounded frequencies also converge in conditional mean.

For each offset from 0 through $m-1$, cover each B_i by disjoint length-m subblocks of that offset, leaving at most $2m$ boundary letters. Use the conditional entropy chain rule, since conditioning on all U_B is at least as informative as conditioning on the subblock. Then average over the m offsets. Every full subblock appears once, and concavity of conditional entropy yields

$$H(Z_B \mid U_B, E_L) \le \sum_i \frac{|B_i| - m + 1}{m} H_{\pi_{i,L}}(Z^m \mid U^m) + O(m).$$

All entropy in this argument is for finite alphabets with the net, partition and m fixed. The $O(m)$ term includes boundary letters in all bands. Each coordinate is included at most m times in the overlapping blocks. The conditional expectation of the left side of (87) is therefore at least

$$\sum_i a_i \frac{|B_i| - m + 1}{m} \mathbb{E}_{\pi_{i,L}} \sum_{j=1}^{m} \|Z_j - U_j\|^2.$$

Add this inequality to $\log N \ge H(Z_B \mid E_L) - H(Z_B \mid U_B, E_L)$. Continuity on the finite joint simplex and the source marginal convergence show that

$$\liminf_{L\to\infty} \frac{\log N + t_L/\phi(L)}{L} \ge \sum_i \ell_i \left[F_m^\delta(a_i) + h_\delta - \frac{H(P_m)}{m} \right] - C_{\alpha,T}\delta, \tag{88}$$

where F_m^δ is the normalized objective with both alphabets the rounded finite net and source Z.

As $m \to \infty$, $H(P_m)/m \to h_\delta$. The rounding approximation gives

$$|F_m^\delta(a) - F_m(a)| \le Ca\delta$$

uniformly in m. In one direction round both arguments of a channel and use data processing. In the other generate the rounded reproduction from the rounded input, for which the mutual information is unchanged by conditioning on the original input. The squared distortions differ by at most $C\delta$ per coordinate. Let first $m \to \infty$, then $\delta \downarrow 0$ in (88). Refine the intervals and bring the a_i up to their infimum weights. Finally take $\alpha \downarrow 0$ and $T \uparrow \infty$. This gives

$$\liminf_{L\to\infty} \frac{\log N + t_L/\phi(L)}{L} \ge J \tag{89}$$

whenever the covered mass stays bounded away from zero.

For $J > 0$, use $L = (\log N)/B_0$. Regular variation implies $L\phi(L)/[(\log N)\phi(\log N)] \to B_0^{b-1}$. Thus squared radius $c(\log N)\phi(\log N)$ for $c < (J/b)B_0^{b-1}$ would contradict (89) if a subsequence of codebooks covered mass bounded away from zero. This proves the strong converse and both lower limits in Theorem K.1.

K.4 Matching weighted codes

For $J > 0$, let $R = \log N$ and $L = R/(B_0 + \gamma)$ with fixed small $\gamma > 0$. As in the independent case of Section B, we split into a low part $j < \alpha L$, a central part $\alpha L \le j \le TL$, and a tail $j > TL$. On each of finitely many short central intervals choose a representative u where F is differentiable at u^{-b}. Use the stationary block codes of the preceding section, with per-coordinate rate and distortion $I(u^{-b})+\eta$ and $D(u^{-b})+\eta$. Stationarity gives the proper block distribution even at moving

starting indices. No independence is needed to combine the finitely many failure probabilities. For sufficiently fine subdivision and small η, (86) and uniform regular variation give total central log cardinality at most $(B_0 + \gamma/3)L$ and weighted squared error at most $(J/b + \gamma/3)L\phi(L)$ with probability tending to one.

On each low coordinate use a Euclidean grid of mesh comparable to $\min(1, \sqrt{\phi(L)/\lambda_j})$. Its squared weighted error per coordinate is $O(\phi(L))$. The total log cardinality on the low part is at most

$$O\left(\sum_{j<\alpha L} [1 + \log^+(\lambda_j/\phi(L))]\right) = O(\alpha L(1 + |\log \alpha|)) + o(L).$$

This follows from Potter bounds on all but finitely many initial indices, which together cost $O(\log L)$. On the tail reproduce a fixed point, with squared error at most $C \sum_{j>TL} \lambda_j$. This is $O(T^{1-b} L\phi(L))$ by regular variation. First take α small and T large so that both errors and the head rate fit the remaining slack, then choose the central subdivision and rate slack. The product codebook has size at most N and squared error at most $(J/b + \gamma)L\phi(L)$ with probability tending to one. Its pointwise squared error is $O_{\alpha,T}(L\phi(L))$, because the low part has a pointwise mesh bound and all other reproductions are bounded. This proves the upper limits for every fixed moment and quantile on letting $\gamma \downarrow 0$.

If $J = 0$, then F vanishes identically. Use $L = R$, choose the two cutoffs to give arbitrarily small normalized low and tail distortion and less than half the rate in the low part, and use central codes with arbitrarily small rates and distortions. The same argument proves the zero limit. This completes Theorem K.1.

K.5 Arbitrary stationary laws

Let the compact coefficient law now be stationary, with ergodic decomposition $\mathsf{P} = \int \mathsf{P}_\theta \, d\vartheta(\theta)$. Write $\kappa_b(\theta)$ for the constant of Theorem K.1 computed on P_θ, and G for its distribution function under ϑ. These constants are bounded uniformly. They are measurable: the finite-grid finite-block objectives are continuous functions of the finite source word probabilities. Take the block limits, then the grid limit, then the integral.

Theorem K.2. *Under any stationary law on K, for every fixed finite $s \geq 1$,*

$$\lim_{N\to\infty} \frac{e_{N,s}(Y)^2}{(\log N)\phi(\log N)} = \left(\int \kappa_b(\theta)^{s/2} \, d\vartheta(\theta)\right)^{2/s}.$$

At each continuity point x of G, the greatest mass covered by N balls of squared radius $x(\log N)\phi(\log N)$ converges to $G(x)$. One sequence of codebooks independent of the stationary law attains both conclusions. The lower and upper limits of the normalized fixed-level quantile are bounded respectively by

$$\inf\{x : G(x) \geq 1 - \eta\}, \qquad \inf\{x : G(x) > 1 - \eta\}.$$

If these bounds coincide, the quantile converges to their common value.

Gray [71, Theorem 11.3.1] gives the corresponding ergodic decomposition of the fixed-rate distortion function for ordinary additive distortion. Universal stationary lossy coding goes back to the work of Yang and Kieffer [73]. We include a block construction to explain how the code uses different rate allocations for different components without observing the component.

Lemma K.3. *On a finite alphabet with bounded one-letter distortion, at each fixed rate $r > 0$ there are common length-n codebooks of size at most e^{nr} with the following property. For any stationary ergodic source and any fixed m-block channel whose information per letter is strictly less than r, the code's average distortion is at most the channel distortion per letter plus any fixed slack with probability tending to one.*

Proof. For a fixed m, split the word into m consecutive chunks of nearly equal length. In chunk j, tile with length-m blocks whose start indices have residue j modulo m, discarding at most $2m$ letters per chunk. For each residue the shift by m has ergodic averages that need not be constant. Its average in that chunk tends to the same value as in a prefix of that residue, by taking differences of the two prefix sums. Averaging the m equally represented residues gives the ordinary overlapping block frequencies, which tend to the stationary m-block law by ergodicity of the one-step shift. Thus the empirical type of the selected blocks tends to the m-block law.

For each possible type P of the selected input blocks, and each fixed internal rate $r' < r$, choose a joint type with that input marginal, minimizing distortion under normalized mutual information at most r'. Such a joint type always exists, by using a constant output. For a fixed input word of type P, a random word drawn uniformly from the corresponding output type has the chosen joint type with probability at least $\exp(-MI - O_m(\log n))$, where M is the number of selected blocks and I the mutual information per block. This follows by dividing the multinomial conditional-type count by the output-type count. With $\exp(MI + O_m(\log n))$ independent output words and a sufficiently large coefficient of $\log n$, the failure probability is less than the reciprocal of the number of words of the input type. A union bound then gives a deterministic cover of the entire input type. The polynomial number of types and the fixed reproductions on discarded letters thus cost total log cardinality at most $nr' + O_m(\log n)$.

Enumerate all pairs m, r' with rational $0 < r' < r$. At length n include from the first n pairs those codebooks with at most $\lfloor e^{nr}/n \rfloor$ words. Use a singleton if no such codebook is included. Every fixed pair is included for all sufficiently large n. Given any m-block channel as in the statement, choose r' above its information per letter. When the observed input type approaches the correct m-block law, rounding the channel row by row to a compatible joint type gives convergence of mutual information and distortion. The strict rate slack makes these types feasible. The covering code then attains the stated distortion. □

To construct common weighted codes, set $R = \log N$. Enumerate profiles consisting of a rational multiple $L/R > 0$, rational cutoffs and central partition endpoints in j/L, a finite grid resolution, and positive rational rates for the central intervals. For each profile use the exact same low coordinate grids and deterministic tail as in the proof of Theorem K.1. On every central band use Lemma K.3 on the selected finite grid at its assigned rate. The rounded source in each ergodic component is again stationary ergodic, and de-rounding the input affects the squared error by $O(\delta)$ per coordinate before weighting. For any fixed component and goal $\kappa_b(\theta) + \xi$, $\xi > 0$, select a profile approximating the code allocation from the preceding section. The slack there permits rational choices and a total log-cardinality bound $(1 - c_\theta)R$ for some $c_\theta > 0$. Taking the central grid fine makes the cost perturbation negligible. The information-distortion pairs from the preceding section can be approximated by finite-block channels and then rounded, which decreases their information and changes their distortion by $O(\delta)$. Thus the selected profile gives normalized distortion at most $\kappa_b(\theta) + \xi$ with conditional probability tending to one. If $\kappa_b(\theta) = 0$, use the same small-rate allocation as in the zero case above.

For each N, include the realizable codes from the first $\lfloor R \rfloor$ profiles whose actual sizes are at most $\lfloor N/[2(1 + \lfloor R \rfloor)] \rfloor$. Every fixed profile with a strict rate slack is eventually included. Also

include a deterministic fallback of size at most $N/2$ and pointwise squared error $O(R\phi(R))$. For the fallback, take $L = cR$, use only low-coordinate grids up to $j \le L$, and use a constant tail. The preceding low-coordinate estimate with cutoff one gives total cost $O(cR)$, so a small fixed $c > 0$ suffices. All error bounds of this fallback are uniform on $K^{\mathbb{N}}$.

For any one sequence of N-point codebooks, the ergodic strong converse applies conditionally on θ. For $\kappa_b(\theta) > 0$ the probability of normalized error below any strictly smaller threshold tends to zero. For the common codebooks constructed above, the conditional probability of exceeding any strictly larger threshold also tends to zero. When the component constant is zero the latter statement alone suffices. Thus the normalized squared error of the common codebooks converges in probability to the component constant under the ergodic decomposition, and is uniformly bounded by the fallback. This gives the limiting moments and coverage lower bound. For arbitrary competitors the conditional converse gives the coverage upper bound by bounded convergence at continuity points, and the respective moment lower bound by Fatou's lemma. Taking near-optimal sequences transfers these bounds to the suprema and infima. The quantile statements follow by taking continuity-point thresholds on either side of the generalized inverse.

K.6 Dependence and the role of the hypotheses

When the constant vanishes. For a compact finite-dimensional alphabet, $J = 0$ exactly when the shift has zero Kolmogorov–Sinai entropy. If the entropy is zero, a finite coordinate quantizer of mesh δ has zero output entropy rate. Use its output as the reproduction, giving $F(a) \le a\delta^2$, and send δ to zero. For the converse, take nested translated dyadic coordinate partitions whose boundaries have marginal probability zero. They and their shifts generate the process sigma field, so positive entropy implies positive entropy rate h_0 for some one-coordinate partition. Indeed, a finite partition of paths can be approximated in probability by a function of finitely many rounded coordinates. Its entropy rate differs from that approximation by at most the finite-alphabet conditional entropy bound for their disagreement probability. The approximating partition has entropy rate at most that of the rounded coordinate process. Given a reproduction of average squared distortion D, classify each source cell by the cell containing the reproduction. Its average error probability is at most the marginal mass within distance ζ of the cell boundaries plus D/ζ^2. Choose small fixed ζ and then sufficiently small D. The conditional entropy bound for a finite alphabet gives information per coordinate at least $h_0/2$, by data processing from the coordinate cells. If D is larger the penalty aD is bounded below for $a > 0$. Thus $F(a) > 0$ and $J > 0$.

Binary sources and Fourier series. Consider independent uniform signs and a stationary symmetric binary Markov source with the same marginal and flip probability $\rho \in (0,1)$, $\rho \ne 1/2$. For one independent sign, squared-error reproduction can be replaced by the posterior mean $v \in [-1,1]$ given the reproduction. Its distortion is $1-\mathbb{E}v^2$, and its information is $\log 2 - \mathbb{E}H_2((1+v)/2)$, where H_2 is binary entropy. Minimizing gives

$$F_{\mathrm{iid}}(a) = \min_{0 \le z \le 1} \left[\log 2 - H_2((1+z)/2) + a(1-z^2) \right].$$

The lower bound is pointwise and equality is attained by posterior means $\pm z$ with equal probabilities. For $a > 1/2$ a minimizer has $0 < z < 1$: the quadratic coefficient at zero is $1/2 - a$, and the entropy term has infinite slope at one. Apply the associated binary symmetric channel independently at all coordinates of the Markov source, with output signs of conditional correlation z and reconstruction magnitude z. The one-coordinate distortion is unchanged. The conditional output entropy per coordinate is $H_2((1-z)/2)$. The output entropy rate is strictly less than $\log 2$, since adjacent

output signs have correlation $z^2(1-2\rho) \neq 0$. Thus the per-coordinate information rate is strictly smaller, and $F_{\mathrm{Markov}}(a) < F_{\mathrm{iid}}(a)$ for $a > 1/2$. The same channel gives the weak inequality at all $a > 0$. Integrating proves the strict inequality between the two quantization constants.

There is also an order comparison as $\rho \downarrow 0$. For $0 < \rho \leq 1/4$,

$$c \min(a, \rho) \leq F_{\mathrm{Markov}}(a) \leq C \min\{a, \rho[1 + \log_+(a/\rho)]\}. \tag{90}$$

For the lower bound, if the average squared distortion is smaller than a fixed small constant, keep a residue class of indices modulo $\lceil 1/\rho \rceil$ with at most a constant multiple of that distortion. Taking signs of the reproductions gives bit error probability at most the squared distortion on those indices. The entropy per retained source symbol is bounded below by a constant, since its flip probability is $[1 - (1-2\rho)^{\lceil 1/\rho \rceil}]/2$. The binary conditional entropy bound gives information at least a fixed positive constant per retained symbol. Data processing gives information at least $c\rho$ per original coordinate in the block limit. If the distortion is larger, the penalty is at least ca. For the upper bound when $\rho \leq a \leq 1$, transmit every ℓth sign for $\ell = \lfloor 1/a \rfloor$ and repeat the previous transmitted sign until the next one. This costs at most $C\rho\ell$ per letter in squared distortion. The entropy rate of the transmitted signs, divided by ℓ, is at most $C\rho[1 + \log(1/(\rho\ell))]$. For $a > 1$ transmit all the signs, and for $a < \rho$ use the zero reproduction. This proves (90). Integration at $a = u^{-b}$ shows $J \asymp \rho^{1-1/b}$ and $\kappa_b \asymp \rho^{b-1}$.

The exact slow-switching coefficient. Let Z be the stationary continuous-time sign process that flips at rate one. Define $F_{\mathrm{tel}}(a)$ to be the large-T limit of

$$\frac{1}{T} \inf_V \left[I(Z_{[0,T]}; V) + a\, \mathbb{E} \int_0^T |Z(t) - V(t)|^2 \, dt \right]$$

over reproductions in $L^2[0,T]$ taking values in $[-1,1]$ almost everywhere. The limit exists by subadditivity of the unnormalized costs, using product channels on adjacent time windows and stationarity. For any fixed window length, a remainder window has bounded cost by the zero reproduction, giving the usual subadditive limit over real lengths. Set $\Delta = -\frac{1}{2}\log(1-2\rho)$ for $0 < \rho < 1/2$. Then the left-endpoint Δ-skeleton of Z is the binary Markov chain with flip probability ρ.

For $T = m\Delta$, a discrete channel on the skeleton lifts to the path by using piecewise constant reproductions. Its information is unchanged because its channel depends only on the skeleton, which is read from the càdlàg path. Conversely, from any path reproduction take its cell averages as discrete reproductions. Jensen's inequality and data processing control their distortion and information. In either direction replacing $Z(t)$ by its left-endpoint skeleton on each cell changes the average squared distortion by $O(\Delta)$, since the values are bounded and the chance of a jump in a cell is $O(\Delta)$. Taking the block and time limits gives

$$|F_{\mathrm{Markov}}(\Delta a)/\Delta - F_{\mathrm{tel}}(a)| \leq Ca\Delta.$$

The integrable envelope in (90) and $\Delta \sim \rho$ then give, by the change of variable $u = \Delta^{-1/b} v$,

$$\lim_{\rho \downarrow 0} \frac{\kappa_b(\rho)}{\rho^{b-1}} = \frac{(b-1)^{b-1}}{b^b} \left(\int_0^\infty F_{\mathrm{tel}}(v^{-b})\, dv \right)^b.$$

The integral is positive by the lower bound in (90). The convergence here concerns the weighted quantization constant after taking the code-size limit at each fixed ρ.

By Parseval, the constants apply without change to $\sum_{j \geq 1} \sqrt{\lambda_j} X_j e^{2\pi i j t}$ in $L^2[0,1]$.

The rate of the strong converse. Let X_j indicate a renewal in a stationary integer renewal process with gap probabilities proportional to m^{-a_0} for $m \geq 1$ and $a_0 > 2$. This is stationary ergodic by the finite mean regenerative construction. Its entropy rate is positive, since conditional on a renewal at the preceding time there is a nontrivial probability of a unit gap. For a discrete source of positive entropy rate, $F(a) > 0$ for every $a > 0$: small squared distortion on a finite alphabet forces small symbol error under nearest reconstruction and hence positive information rate by the finite-alphabet conditional entropy bound. Thus this renewal source has $J > 0$. The stationary probability of no renewal among the first m sites is bounded below by a constant times m^{2-a_0}. Indeed the stationary gap law is size-biased, and for gaps of lengths between $2m$ and $3m$ a fixed fraction of their positions have residual length at least m. For $\lambda_j = j^{-b}$, on a renewal-free initial segment of length $T \log N$ the zero codeword already has squared error $O(T^{1-b}(\log N)^{1-b})$. Taking T fixed and sufficiently large puts this error below any positive fraction of the asymptotic constant, with probability bounded below by a negative power of $\log N$. The strong converse therefore holds at a rate that depends on the source, here polynomial in $\log N$.

Why finite-dimensional coefficient space matters. The finite dimension of the coefficient space is essential. For iid coefficients in an infinite-dimensional Hilbert space, compactness and $0 < J < \infty$ do not by themselves determine the error. Fix $b > 1$, let $\lambda_j = j^{-b}$, and take orthogonal points of norm $r_i = 2^{-i}$ in each of countably many disjoint groups, with M_i points in group i and $L_i = \log M_i \asymp 2^{i^2}$. The ith group has total probability p_i, uniform within its points, where

$$p_i = 2^{-i} L_i^{-(b-1)/b}.$$

The remaining mass is placed at zero. The support is compact. The entropy of the group label is finite. Sending the group label and optionally its within-group index shows

$$F(a) \leq H(\text{label}) + \sum_i p_i \min(L_i, a r_i^2), \qquad F(a) \leq a\, \mathbb{E}\|X\|^2.$$

Integrate the first bound on $0 < u < 1$ at $a = u^{-b}$ and the second on $u \geq 1$. The integral of each minimum over all $u > 0$ is $O(r_i^{2/b} L_i^{1-1/b})$, and the weighted sum is finite. Positivity of J follows by detecting whether the input equals an isolated atom of strictly intermediate probability: sufficiently small squared distortion makes its binary classification error small, forcing positive mutual information. On the other hand any codebook with $N_i = \lfloor M_i/2 \rfloor$ points misses at least half the ith group in its first-coordinate projections at radius $r_i/3$, by orthogonality of the group points. Hence $e_{N_i,2}^2 \geq c p_i r_i^2$. Dividing by $(\log N_i)^{1-b}$ makes this lower bound tend to infinity, because $p_i r_i^2 L_i^{b-1} = 2^{-3i} L_i^{(b-1)^2/b}$.

Statements and Declarations

Funding. This research was self-funded.

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

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of each part. Anthropic Claude edited, consolidated and cross-referenced the parts and checked computations. Scott Schubr directed the research and is responsible for the content.

References

[1] H. Luschgy and G. Pagès, Sharp asymptotics of the functional quantization problem for Gaussian processes, *Ann. Probab.* 32 (2004), 1574–1599. https://doi.org/10.1214/009117904000000324.

[2] S. Schubr, A Wasserstein rate for Kloosterman paths, with sharp small-ball estimates, Luminary Research Archive, edition 6 (2026). https://research.luminaryoddities.org/small-ball-estimates-holder-centers-kloosterman-process.

[3] G. Ricotta, E. Royer and I. Shparlinski, Kloosterman paths of prime powers moduli, II, *Bull. Soc. Math. France* **148** (2020), 173–188. https://arxiv.org/abs/1810.01150.

[4] E. Kowalski and W. F. Sawin, Kloosterman paths and the shape of exponential sums, *Compos. Math.* **152** (2016), 1489–1516. https://arxiv.org/abs/1410.7892.

[5] E. Kowalski and W. Sawin, *Kloosterman paths and the shape of exponential sums*, Compos. Math. **152** (2016), 1489–1516.

[6] E. Kowalski and T. Untrau, *Wasserstein metrics and quantitative equidistribution of exponential sums over finite fields*, arXiv:2505.22059.

[7] S. Schubr, *A Wasserstein rate for Kloosterman paths, with sharp small-ball estimates*, Luminary Research Archive, ed. 6 (2026), doi:10.5281/zenodo.24409094.

[8] S. G. Bobkov and M. Ledoux, *Transport inequalities on Euclidean spaces for non-Euclidean metrics*, J. Fourier Anal. Appl. **26** (2020), Paper No. 60, doi:10.1007/s00041-020-09766-2.

[9] A. De, O. Etesami, L. Trevisan, and M. Tulsiani, *Improved pseudorandom generators for depth 2 circuits*, APPROX–RANDOM, Lecture Notes in Comput. Sci. **6302**, Springer, 2010.

[10] E. Kowalski and W. Sawin, *On the support of the Kloosterman paths*, arXiv:1709.05192.

[11] D. Kyselov, *Space-filling curves and small-ball probabilities for the Kloosterman process*, arXiv:2610.22645v1.

[12] K. Lebreton, *The distribution and the structure of the maximum of partial sums in families of trace functions*, arXiv:2608.04249v2.

[13] P. Autissier, D. Bonolis, and Y. Lamzouri, *The distribution of the maximum of partial sums of Kloosterman sums and other trace functions*, Compos. Math. **157** (2021), 1610–1651.

[14] E. Fouvry, E. Kowalski, and P. Michel, *The sliding-sum method for short exponential sums*, arXiv:1307.0135.

[15] É. Fouvry, E. Kowalski, P. Michel, C. S. Raju, J. Rivat, and K. Soundararajan, *On short sums of trace functions*, Ann. Inst. Fourier **67** (2017), doi:10.5802/aif.3087.

[16] C. Perret-Gentil, *Gaussian distribution of short sums of trace functions over finite fields*, arXiv:1609.09334.

[17] P. Xi, *Bilinear forms with trace functions over arbitrary sets and applications to Sato–Tate*, Sci. China Math. **66** (2023), 2819–2834.

[18] E. Kowalski, P. Michel, and W. Sawin, *Bilinear forms with Kloosterman sums and applications*, Ann. of Math. **186** (2017), 413–500.

[19] S. G. Bobkov and M. Ledoux, *Transport inequalities on Euclidean spaces for non-Euclidean metrics*, J. Fourier Anal. Appl. **26** (2020), Paper No. 60.

[20] G. Ricotta and E. Royer, *Kloosterman paths of prime powers moduli*, Comment. Math. Helv. **93** (2018), 493–532.

[21] G. Ricotta, E. Royer, and I. E. Shparlinski, *Kloosterman paths of prime powers moduli, II*, Bull. Soc. Math. France **148** (2020), 173–188.

[22] D. Milićević and S. Zhang, *Distribution of Kloosterman paths to high prime power moduli*, Trans. Amer. Math. Soc. Ser. B **10** (2023), 636–669, doi:10.1090/btran/98.

[23] D. Milićević, C. Robinson, and C. Shupe, *Sums of products of Kloosterman sums to prime power moduli*, Anal. Math. (2026), doi:10.1007/s10476-026-00197-7.

[24] R. A. Smith, *On n-dimensional Kloosterman sums*, J. Number Theory **11** (1979), 324–343, doi:10.1016/0022-314X(79)90006-4.

[25] R. Dbrowski and B. Fisher, *A stationary phase formula for exponential sums over $\mathbb{Z}/p^m\mathbb{Z}$ and applications to GL(3)-Kloosterman sums*, Acta Arith. **80** (1997), 1–48, doi:10.4064/aa-80-1-1-48.

[26] J. Van Order, *Dirichlet twists of* GL_n*-automorphic L-functions and hyper-Kloosterman Dirichlet series*, Ann. Fac. Sci. Toulouse Math., doi:10.5802/afst.1687.

[27] T. Untrau, *Equidistribution of exponential sums indexed by a subgroup of fixed cardinality*, Math. Proc. Cambridge Philos. Soc. **176** (2024), 65–94, doi:10.1017/S0305004123000361.

[28] D. Milićević, C. Robinson, and C. Shupe, *Sums of products of Kloosterman sums to prime power moduli*, Anal. Math. **52** (2026), 1135–1173, doi:10.1007/s10476-026-00197-7.

[29] D. Milićević, C. Robinson and C. Shupe, *Sums of products of Kloosterman sums to prime power moduli*, arXiv:2608.21346, 2026.

[30] I. S. Reed and G. Solomon, *Polynomial codes over certain finite fields*, J. Soc. Indust. Appl. Math. **8** (1960), 300–304, doi:10.1137/0108018.

[31] R. Dąbrowski and B. Fisher, *A stationary phase formula for exponential sums over $\mathbb{Z}/p^m\mathbb{Z}$ and applications to* GL(3)*-Kloosterman sums*, Acta Arith. **80** (1997), 1–48, doi:10.4064/aa-80-1-1-48.

[32] H. Luschgy and G. Pagès, *Sharp asymptotics of the functional quantization problem for Gaussian processes*, Ann. Probab. **32** (2004), 1574–1599, doi:10.1214/009117904000000324.

[33] H. Luschgy and G. Pagès, *Functional quantization rate and mean regularity of processes with an application to Lévy processes*, Ann. Appl. Probab. **18** (2008), 427–469, arXiv:math/0601774.

[34] V. Kostina and S. Verdú, *Fixed-length lossy compression in the finite blocklength regime*, IEEE Trans. Inform. Theory **58** (2012), 3309–3338, arXiv:1102.3944.

[35] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular Variation*, Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press, 1987.

[36] E. Kowalski and W. F. Sawin, *Kloosterman paths and the shape of exponential sums*, Compos. Math. **152** (2016), 1489–1516, doi:10.1112/S0010437X16007351.

[37] A. Dembo and I. Kontoyiannis, *Source coding, large deviations, and approximate pattern matching*, IEEE Trans. Inform. Theory **48** (2002), 1590–1615, arXiv:math/0103007.

[38] S. Dereich and M. A. Lifshits, *Probabilities of randomly centered small balls and quantization in Banach spaces*, Ann. Probab. **33** (2005), 1397–1421, arXiv:math/0402220.

[39] N. Saldi, T. Linder and S. Yüksel, *Randomized quantization and source coding with constrained output distribution*, IEEE Trans. Inform. Theory **61** (2015), 91–106, arXiv:1309.2915.

[40] P. Cuff, *Distributed channel synthesis*, IEEE Trans. Inform. Theory **59** (2013), 7071–7096, arXiv:1208.4415.

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

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

[43] S. Schubr, *A Wasserstein rate for Kloosterman paths, with sharp small-ball estimates*, Luminary Research Archive, edition 6, November 21, 2026, edition 6.

[44] S. Schubr, *A Wasserstein rate for Kloosterman paths, with sharp small-ball estimates*, Luminary Research Archive, edition 6, 2026.

[45] J. Naor and M. Naor, *Small-bias probability spaces: Efficient constructions and applications*, SIAM J. Comput. 22 (1993), 838–856.

[46] E. Kowalski and T. Untrau, *Wasserstein metrics and quantitative equidistribution of exponential sums over finite fields*, arXiv:2505.22059v2, 2025.

[47] É. Fouvry, E. Kowalski and P. Michel, *A study in sums of products*, Philos. Trans. Roy. Soc. A 373 (2015), 20140309.

[48] Á. Chávez and G. Todd, *Supercharacters and mixed moments of Kloosterman sums*, Int. J. Number Theory 14 (2018), 1023–1032.

[49] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. **338**, Springer, 2009.

[50] N. Saldi, T. Linder and S. Yüksel, *Randomized quantization and source coding with constrained output distribution*, IEEE Trans. Inform. Theory **61** (2015), 91–106, doi:10.1109/TIT.2014.2373382.

[51] D. Milievi and S. Zhang, *Distribution of Kloosterman paths to high prime power moduli*, Trans. Amer. Math. Soc. Ser. B **10** (2023), 636–669, doi:10.1090/btran/98.

[52] S. Dereich and M. A. Lifshits, *Probabilities of randomly centered small balls and quantization in Banach spaces*, Ann. Probab. **33** (2005), 1397–1421, doi:10.1214/009117905000000161.

[53] S. Dereich, *Small ball probabilities around random centers of Gaussian measures and applications to quantization*, J. Theoret. Probab. **16** (2003), 427–449, doi:10.1023/A:1023578812641.

[54] 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, doi:10.1214/12-AIHP517.

[55] J. Lei, *Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces*, Bernoulli **26** (2020), 767–798, doi:10.3150/19-BEJ1151.

[56] S. Coste, *The Wasserstein cost of the Girsanov theorem*, arXiv:2611.16043, 2026.

[57] H. Q. Minh, *Entropic regularization of Wasserstein distance between infinite-dimensional Gaussian measures and Gaussian processes*, J. Theoret. Probab. **36** (2023), 201–296, doi:10.1007/s10959-022-01165-1.

[58] M. Gelbrich, *On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces*, Math. Nachr. **147** (1990), 185–203, doi:10.1002/mana.19901470121.

[59] D. Milićević and S. Zhang, *Distribution of Kloosterman paths to high prime power moduli*, Trans. Amer. Math. Soc. Ser. B **10** (2023), 636–669, arXiv:2005.08865.

[60] D. Milićević, C. Robinson and C. Shupe, *Sums of products of Kloosterman sums to prime power moduli*, Anal. Math. (2026), arXiv:2608.21346.

[61] V. Blomer, M. S. Risager and I. E. Shparlinski, *Triple sums of Kloosterman sums and the discrepancy of modular inverses*, J. Lond. Math. Soc. **112** (2025), e70291, doi:10.1112/jlms.70291.

[62] R. Peyre, *Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance*, ESAIM Control Optim. Calc. Var. **24** (2018), 1489–1501, doi:10.1051/cocv/2017050.

[63] J. W. S. Cassels, *An Introduction to the Geometry of Numbers*, Grundlehren Math. Wiss. **99**, Springer, 1959.

[64] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Birkhäuser, 2008, doi:10.1007/978-3-7643-8722-8.

[65] J. Rabin, J. Delon and Y. Gousseau, *Transportation distances on the circle*, J. Math. Imaging Vis. **41** (2011), 147–167, arXiv:0906.5499.

[66] J. Delon, J. Salomon and A. Sobolevski, *Fast transport optimization for Monge costs on the circle*, SIAM J. Appl. Math. **70** (2010), 2239–2258, arXiv:0902.3527.

[67] A. Hussain and Y. Lamzouri, *The limiting distribution of Legendre paths*, J. Èc. polytech. Math. **11** (2024), 589–611, doi:10.5802/jep.260.

[68] Y. Lamzouri, *The distribution of large quadratic character sums and applications*, arXiv:2212.03227.

[69] T. Xylouris, *On the least prime in an arithmetic progression and estimates for the zeros of Dirichlet L-functions*, Acta Arith. **150** (2011), 65–91, doi:10.4064/aa150-1-4.

[70] H. Luschgy and G. Pagès. Sharp asymptotics of the functional quantization problem for Gaussian processes. *Annals of Probability*, 32(2):1574–1599, 2004. `https://doi.org/10.1214/009117904000000324`.

[71] R. M. Gray. *Entropy and Information Theory*. Corrected first edition. https://ee.stanford.edu/~gray/it.pdf.

[72] J. C. Kieffer. Sample converses in source coding theory. *IEEE Transactions on Information Theory*, 37(2):263–268, 1991. https://doi.org/10.1109/18.75241.

[73] E.-H. Yang and J. C. Kieffer. Simple universal lossy data compression schemes derived from the Lempel–Ziv algorithm. *IEEE Transactions on Information Theory*, 42(1):239–245, 1996. https://doi.org/10.1109/18.481794.

[74] T. Weissman and E. Ordentlich. The empirical distribution of rate-constrained source codes. *IEEE Transactions on Information Theory*, 51(11):3718–3733, 2005. https://doi.org/10.1109/TIT.2005.856982.