Non-Asymptotic Entropic Optimal Transport, Gauge-Invariant Birkhoff–Sinkhorn Contraction, Debiased Sinkhorn Divergence, and Bures–Wasserstein Envelopes

Lluis Eriksson
Independent Researcher, Östersund, Sweden
editor@airr.science

November 29, 2026

Abstract

We establish four self-contained, explicit, non-asymptotic quantitative theorems unifying discrete entropic optimal transport, Sinkhorn–Knopp scaling, debiased Sinkhorn divergences, and Bures–Wasserstein covariance geometry. First, for arbitrary discrete marginals $\mu \in \Delta_m$, $\nu \in \Delta_n$ and cost matrix $C \in \mathbb{R}^{m\times n}$, we sharpen the classical $\varepsilon \ln \min(m,n)$ entropic bias ceiling to the exact mutual-information envelope $0 \le \langle C, \pi^*_\varepsilon\rangle - W_C(\mu,\nu) \le \varepsilon\big(\min(H(\mu),H(\nu)) - I(\pi^*_\varepsilon)\big)$, paired with a sharp two-point curvature sandwich $\varepsilon_1(I(\pi^*_{\varepsilon_1}) - I(\pi^*_{\varepsilon_2})) \le \langle C, \pi^*_{\varepsilon_2}\rangle - \langle C, \pi^*_{\varepsilon_1}\rangle \le \varepsilon_2(I(\pi^*_{\varepsilon_1}) - I(\pi^*_{\varepsilon_2}))$ for all $0 < \varepsilon_1 < \varepsilon_2$ (Theorem 2.1). Second, we give a self-contained eight-line proof of the Birkhoff–Hopf projective contraction $\tau := \tanh(\Delta(C)/(4\varepsilon))$ in terms of the gauge-invariant cross-ratio oscillation $\Delta(C) := \max_{i,i',j,j'}(C_{ij} + C_{i'j'} - C_{ij'} - C_{i'j})$, and prove an exact AM–GM Hilbert-to-ℓ_1 transfer yielding the zero-unknown-constant, simplex-saturating marginal residual envelope $\|\pi^{(t)}\mathbf{1} - \mu\|_1 \le 2\tanh(d_H(u^{(t+1)}, u^{(t)})/4) \le 2\tanh(\tau^{2(t-1)}\Delta(C)/(4\varepsilon))$ for all $t \ge 1$ (Theorem 3.1). Third, for symmetric costs C with positive definite Gibbs kernel $K = (e^{-C_{ij}/\varepsilon}) \in \mathbb{S}^n_{++}$, we derive the explicit quadratic positivity bound $S_\varepsilon(\mu,\nu) \ge \frac{\varepsilon}{2}\|u_\mu - u_\nu\|^2_K \ge \frac{\varepsilon}{2}\lambda_{\min}(K)\|u_\mu - u_\nu\|_2^2$ (giving a one-line definiteness proof valid even on boundary supports) together with a two-sided debiasing sandwich around $W_C(\mu,\nu)$ (Theorem 4.1). Fourth, we establish a four-term spectral sandwich bracketing the Bures–Wasserstein covariance distance $W_2^2(\Sigma_1,\Sigma_2)$ between Frobenius ratios with exact equality characterization, paired with a self-contained discrete Marton T_1 transport–entropy bound and exact cost-perturbation stability (Theorem 5.1). All four theorems, boundary regimes, and numerical benchmarks are deterministically verified in an MIT-licensed Python test suite.

1 Introduction and Scope

Entropic optimal transport (EOT), Sinkhorn–Knopp matrix scaling, debiased Sinkhorn divergences, and Bures–Wasserstein geometry form the analytical core of computational optimal transport [1, 2, 3, 4, 5]. Standard textbook analyses, however, leave four non-asymptotic tightenings unexploited: (i) entropic bias bounds typically use the uniform support size $\varepsilon \ln \min(m,n)$ rather than the marginal Shannon entropies $\min(H(\mu),H(\nu))$; (ii) Hilbert-metric Sinkhorn convergence bounds commonly translate to ℓ_1 marginal residuals via $e^{d_H} - 1$, which diverges exponentially for small t and loses a factor of 2 asymptotically; (iii) Sinkhorn divergence positivity proofs often restrict to strictly positive measures or rely on infinite-dimensional RKHS machinery; and (iv) Bures–Wasserstein perturbation bounds are rarely presented alongside their exact four-term spectral Frobenius sandwich and commuting equality characterization.

Sections 2–5 provide self-contained, completely explicit proofs of all four quantitative theorems without black-box dependencies, followed by deterministic numerical verification in Section 6 (`https://github.com/lluiseriksson/optimal-transport-sinkhorn-geometry`, tag `v1.0.0`).

2 Entropy-Sharpened EOT Bias and Two-Point Curvature

Let $\Delta_m := \{p \in \mathbb{R}^m_{\geq 0} : \sum_{i=1}^m p_i = 1\}$ and $\Delta_m^+ := \Delta_m \cap \mathbb{R}^m_{>0}$. For $\mu \in \Delta_m$ and $\nu \in \Delta_n$, let $\Pi(\mu,\nu) := \{\pi \in \mathbb{R}^{m\times n}_{\geq 0} : \pi\mathbf{1} = \mu,\ \pi^\top \mathbf{1} = \nu\}$ be the transport polytope, $H(\mu) := -\sum_{i:\mu_i>0} \mu_i \ln \mu_i$ the Shannon entropy, and for $\pi \in \Pi(\mu,\nu)$ let $I(\pi) := \mathrm{KL}(\pi \| \mu\otimes\nu) = \sum_{i,j:\pi_{ij}>0} \pi_{ij} \ln \frac{\pi_{ij}}{\mu_i\nu_j}$ be the mutual information. For any cost matrix $C \in \mathbb{R}^{m\times n}$ and $\varepsilon > 0$, define

$$W_C(\mu,\nu) \ := \ \min_{\pi\in\Pi(\mu,\nu)} \langle C,\pi\rangle, \qquad F_\varepsilon(\mu,\nu) \ := \ \min_{\pi\in\Pi(\mu,\nu)} \big\{\langle C,\pi\rangle + \varepsilon\, I(\pi)\big\},$$

where $\langle C,\pi\rangle := \sum_{i,j} C_{ij}\pi_{ij}$. Because $\pi \mapsto I(\pi)$ is strictly convex on the nonempty compact convex polytope $\Pi(\mu,\nu)$ (all $\pi \in \Pi(\mu,\nu)$ vanish outside $\mathrm{supp}(\mu)\times\mathrm{supp}(\nu)$), $F_\varepsilon(\mu,\nu)$ admits a unique minimizer $\pi^*_\varepsilon \in \Pi(\mu,\nu)$.

Theorem 2.1 (Entropy-Sharpened Bias Envelope and Two-Point Curvature Sandwich)**.** *For all $\mu \in \Delta_m$, $\nu \in \Delta_n$, $C \in \mathbb{R}^{m\times n}$, and $\varepsilon > 0$:*

(a) ***Mutual-Information Bias Ceiling.*** *Every $\pi \in \Pi(\mu,\nu)$ satisfies $0 \leq I(\pi) \leq \min(H(\mu),H(\nu))$, and therefore*

$$0 \ \leq \ \langle C,\pi^*_\varepsilon\rangle - W_C(\mu,\nu) \ \leq \ \varepsilon\Big(\min\big(H(\mu),H(\nu)\big) - I(\pi^*_\varepsilon)\Big) \ \leq \ \varepsilon\,\min\big(H(\mu),H(\nu)\big), \tag{1}$$

$$W_C(\mu,\nu) \ \leq \ F_\varepsilon(\mu,\nu) \ \leq \ W_C(\mu,\nu) + \varepsilon\,\min\big(H(\mu),H(\nu)\big). \tag{2}$$

(b) ***Exact Two-Point Entropic Curvature Sandwich.*** *For all $0 < \varepsilon_1 < \varepsilon_2$,*

$$\varepsilon_1\big(I(\pi^*_{\varepsilon_1}) - I(\pi^*_{\varepsilon_2})\big) \ \leq \ \langle C,\pi^*_{\varepsilon_2}\rangle - \langle C,\pi^*_{\varepsilon_1}\rangle \ \leq \ \varepsilon_2\big(I(\pi^*_{\varepsilon_1}) - I(\pi^*_{\varepsilon_2})\big). \tag{3}$$

Consequently, $\varepsilon \mapsto I(\pi^*_\varepsilon)$ is non-increasing, $\varepsilon \mapsto \langle C,\pi^*_\varepsilon\rangle$ is non-decreasing, and $\varepsilon \mapsto F_\varepsilon(\mu,\nu)$ is concave and continuously differentiable on $(0,\infty)$ with derivative $\frac{d}{d\varepsilon}F_\varepsilon(\mu,\nu) = I(\pi^*_\varepsilon)$.

Proof. (a) For any coupling $\pi \in \Pi(\mu,\nu)$, letting $(X,Y) \sim \pi$, we have $I(\pi) = H(\mu)+H(\nu)-H(\pi) \geq 0$ (by Gibbs' inequality). Moreover, for every $i \in \mathrm{supp}(\mu)$ and $j \in \mathrm{supp}(\nu)$ with $\pi_{ij} > 0$, $\pi_{ij} \leq \mu_i$ implies $-\ln\pi_{ij} \geq -\ln\mu_i$, so $H(\pi) = -\sum_{i,j}\pi_{ij}\ln\pi_{ij} \geq -\sum_{i,j}\pi_{ij}\ln\mu_i = H(\mu)$, and symmetrically $H(\pi) \geq H(\nu)$. Hence $H(\pi) \geq \max(H(\mu),H(\nu))$, which gives $I(\pi) = H(\mu)+H(\nu)-H(\pi) \leq \min(H(\mu),H(\nu))$. Let $\pi^*_0 \in \Pi(\mu,\nu)$ attain $W_C(\mu,\nu)$. By optimality of π^*_ε and π^*_0,

$$W_C(\mu,\nu) \ \leq \ \langle C,\pi^*_\varepsilon\rangle \ \leq \ F_\varepsilon(\mu,\nu) = \langle C,\pi^*_\varepsilon\rangle + \varepsilon I(\pi^*_\varepsilon) \ \leq \ \langle C,\pi^*_0\rangle + \varepsilon I(\pi^*_0) \ \leq \ W_C(\mu,\nu) + \varepsilon\,\min\big(H(\mu),H(\nu)\big).$$

Subtracting $\varepsilon I(\pi^*_\varepsilon)$ yields (1), and retaining $F_\varepsilon(\mu,\nu)$ yields (2). (When either marginal is a Dirac mass, $\min(H(\mu),H(\nu)) = 0$ and all bounds collapse to 0.)

(b) By optimality of $\pi^*_{\varepsilon_1}$ at ε_1 and $\pi^*_{\varepsilon_2}$ at ε_2,

$$\langle C,\pi^*_{\varepsilon_1}\rangle + \varepsilon_1 I(\pi^*_{\varepsilon_1}) \ \leq \ \langle C,\pi^*_{\varepsilon_2}\rangle + \varepsilon_1 I(\pi^*_{\varepsilon_2}), \qquad \langle C,\pi^*_{\varepsilon_2}\rangle + \varepsilon_2 I(\pi^*_{\varepsilon_2}) \ \leq \ \langle C,\pi^*_{\varepsilon_1}\rangle + \varepsilon_2 I(\pi^*_{\varepsilon_1}).$$

Rearranging the first inequality gives the left-hand side of (3); rearranging the second gives the right-hand side. Subtracting the two yields $(\varepsilon_2-\varepsilon_1)(I(\pi^*_{\varepsilon_1}) - I(\pi^*_{\varepsilon_2})) \geq 0$, proving monotonicity of $I(\pi^*_\varepsilon)$ and $\langle C,\pi^*_\varepsilon\rangle$. Finally, $F_\varepsilon(\mu,\nu)$ is concave in ε as an infimum of affine functions, and dividing $F_{\varepsilon_2} - F_{\varepsilon_1} = (\langle C,\pi^*_{\varepsilon_2}\rangle - \langle C,\pi^*_{\varepsilon_1}\rangle) + \varepsilon_2 I(\pi^*_{\varepsilon_2}) - \varepsilon_1 I(\pi^*_{\varepsilon_1})$ by $\varepsilon_2-\varepsilon_1$ with (3) gives $I(\pi^*_{\varepsilon_2}) \leq \frac{F_{\varepsilon_2}-F_{\varepsilon_1}}{\varepsilon_2-\varepsilon_1} \leq I(\pi^*_{\varepsilon_1})$; by uniqueness of π^*_ε and compactness of $\Pi(\mu,\nu)$, $\varepsilon \mapsto \pi^*_\varepsilon$ is continuous, hence $F'_\varepsilon(\mu,\nu) = I(\pi^*_\varepsilon)$. □

3 Gauge-Invariant Birkhoff–Sinkhorn and Exact ℓ_1 Transfer

For $x, y \in \mathbb{R}^m_{>0}$, define Hilbert's projective metric $d_H(x,y) := \ln\big(\max_i(x_i/y_i) \cdot \max_{i'}(y_{i'}/x_{i'})\big) \geq 0$. For $C \in \mathbb{R}^{m\times n}$, define the cross-ratio oscillation

$$\Delta(C) \;:=\; \max_{1\leq i,i'\leq m,\, 1\leq j,j'\leq n} \big(C_{ij} + C_{i'j'} - C_{ij'} - C_{i'j}\big) \;\geq\; 0,$$

which is invariant under every gauge transformation $C_{ij} \mapsto C_{ij} + a_i + b_j$.

Theorem 3.1 (Self-Contained Birkhoff–Sinkhorn Contraction and Saturating ℓ_1 Envelope)**.** *Let $\mu \in \Delta^+_m$, $\nu \in \Delta^+_n$, $C \in \mathbb{R}^{m\times n}$, $\varepsilon > 0$, $K_{ij} := e^{-C_{ij}/\varepsilon} > 0$, and $\tau := \tanh\big(\frac{\Delta(C)}{4\varepsilon}\big) \in [0,1)$. Starting from any $v^{(0)} \in \mathbb{R}^n_{>0}$, define the Sinkhorn–Knopp iterates $u^{(t)} := \mu \oslash (Kv^{(t-1)})$, $v^{(t)} := \nu \oslash (K^\top u^{(t)})$, and $\pi^{(t)} := \operatorname{diag}(u^{(t)}) K \operatorname{diag}(v^{(t)})$ for $t \geq 1$ (so $\pi^{(t)\top}\mathbf{1} = \nu$ identically), and let (u^*, v^*) be the unique (up to reciprocal scaling) fixed point of the iteration.*

(a) ***Projective Diameter and Birkhoff Contraction.*** *For all $x, y \in \mathbb{R}^n_{>0}$,*

$$d_H(Kx, Ky) \;\leq\; \min\left(\frac{\Delta(C)}{\varepsilon},\;\; \tau\, d_H(x,y)\right), \tag{4}$$

and identically for $K^\top$ on $\mathbb{R}^m_{>0}$. Consequently, for every $t \geq 1$ (with $0^0 = 1$ when $\Delta(C) = 0$),

$$d_H(u^{(t)}, u^*) \;\leq\; \tau^{2(t-1)} \frac{\Delta(C)}{\varepsilon}, \qquad d_H(u^{(t+1)}, u^{(t)}) \;\leq\; \tau^{2(t-1)} \frac{\Delta(C)}{\varepsilon}. \tag{5}$$

(b) ***Sharp Saturating Hilbert-to-ℓ_1 Marginal Error Envelope.*** *For every $t \geq 1$,*

$$\big\|\pi^{(t)}\mathbf{1} - \mu\big\|_1 \;\leq\; 2\tanh\left(\frac{d_H(u^{(t+1)}, u^{(t)})}{4}\right) \;\leq\; 2\tanh\left(\frac{\tau^{2(t-1)}\,\Delta(C)}{4\varepsilon}\right). \tag{6}$$

Proof. (a) Set $D := e^{\Delta(C)/(2\varepsilon)} \geq 1$, so $\max_{i,i',j,j'} \frac{K_{ij}K_{i'j'}}{K_{i'j}K_{ij'}} = D^2$ and $\tau = \frac{D-1}{D+1}$. For any $x, y \in \mathbb{R}^n_{>0}$ and any i, i',

$$\frac{(Kx)_i (Ky)_{i'}}{(Kx)_{i'}(Ky)_i} = \frac{\sum_{j,j'} K_{ij}K_{i'j'}x_j y_{j'}}{\sum_{j,j'} K_{i'j}K_{ij'}x_j y_{j'}} \;\leq\; D^2,$$

so taking ln gives $d_H(Kx, Ky) \leq 2\ln D = \Delta(C)/\varepsilon$. To prove $d_H(Kx, Ky) \leq \tau\, d_H(x,y)$, set $r_j := x_j/y_j \in [m_r, M_r]$ with $M_r/m_r = e^{d_H(x,y)}$. If $M_r = m_r$, then $x \propto y$ and $d_H(Kx, Ky) = 0$. Assume $M_r > m_r$ and set $t_0 := \sqrt{M_r/m_r} = e^{d_H(x,y)/2} > 1$, $s_j := \frac{r_j - m_r}{M_r - m_r} \in [0,1]$, $a_j := K_{ij}y_j > 0$, $b_j := K_{i'j}y_j > 0$. Define $A_1 := \sum_j a_j s_j$, $A_0 := \sum_j a_j(1-s_j)$, $B_1 := \sum_j b_j s_j$, $B_0 := \sum_j b_j(1-s_j)$. Since s_j takes both 0 and 1 and $a_j, b_j > 0$, all four sums A_0, A_1, B_0, B_1 are strictly positive. Moreover,

$$A_1 B_0 = \sum_{j,j'} K_{ij}K_{i'j'}y_j y_{j'} s_j(1-s_{j'}) \;\leq\; D^2 \sum_{j,j'} K_{i'j}K_{ij'}y_j y_{j'} s_j(1-s_{j'}) = D^2 B_1 A_0.$$

Writing $\alpha := A_1/A_0 > 0$ and $\beta := B_1/B_0 > 0$ (so $\alpha \leq D^2\beta$), we have

$$\frac{(Kx)_i(Ky)_{i'}}{(Kx)_{i'}(Ky)_i} = \frac{(M_r\alpha + m_r)(1+\beta)}{(\alpha+1)(M_r\beta + m_r)} \;\leq\; \frac{(t_0^2 D^2\beta + 1)(1+\beta)}{(D^2\beta+1)(t_0^2\beta+1)}$$

because $\alpha \mapsto \frac{t_0^2\alpha+1}{\alpha+1}$ is increasing. Expanding $\frac{(t_0^2D^2\beta+1)(1+\beta)}{(D^2\beta+1)(t_0^2\beta+1)} = \frac{t_0^2D^2\beta+1/\beta+t_0^2D^2+1}{t_0^2D^2\beta+1/\beta+D^2+t_0^2}$ and noting that $u \mapsto \frac{u+t_0^2D^2+1}{u+D^2+t_0^2}$ is decreasing in $u > 0$ (as $t_0^2D^2 + 1 \geq D^2 + t_0^2$), AM–GM minimizes $u =$

$t_0^2 D^2\beta + 1/\beta \geq 2t_0 D$ at $\beta = 1/(t_0 D)$, yielding the maximum $\frac{2t_0 D + t_0^2 D^2 + 1}{2t_0 D + D^2 + t_0^2} = \left(\frac{Dt_0+1}{D+t_0}\right)^2$. Now $\psi(s) := \frac{D-1}{D+1}\ln s - \ln\frac{Ds+1}{D+s}$ satisfies $\psi(1) = 0$ and

$$\psi'(s) = \frac{D-1}{(D+1)s} - \frac{D^2-1}{(Ds+1)(D+s)} = \frac{(D-1)D(s-1)^2}{(D+1)s(Ds+1)(D+s)} \;\geq\; 0 \qquad (\forall s \geq 1),$$

so $2\ln\frac{Dt_0+1}{D+t_0} \leq 2\frac{D-1}{D+1}\ln t_0 = \tau\, d_H(x,y)$, proving (4). Since componentwise positive scaling and inversion are isometries of d_H, $d_H(u^{(1)}, u^*) = d_H(Kv^{(0)}, Kv^*) \leq \Delta(C)/\varepsilon$ and $d_H(u^{(2)}, u^{(1)}) = d_H(Kv^{(1)}, Kv^{(0)}) \leq \Delta(C)/\varepsilon$; applying (4) alternately to K and $K^\top$ gives $d_H(u^{(t+1)}, u^*) \leq \tau^2 d_H(u^{(t)}, u^*)$ and $d_H(u^{(t+2)}, u^{(t+1)}) \leq \tau^2 d_H(u^{(t+1)}, u^{(t)})$, yielding (5) by induction.

(b) Set $r^{(t)} := \pi^{(t)}\mathbf{1}$; its i-th coordinate is $r_i^{(t)} = u_i^{(t)}(Kv^{(t)})_i = \mu_i \frac{(Kv^{(t)})_i}{(Kv^{(t-1)})_i} = \mu_i \frac{u_i^{(t)}}{u_i^{(t+1)}}$. Let $\alpha_i := r_i^{(t)}/\mu_i > 0$, $\alpha_{\min} := \min_i \alpha_i$, $\alpha_{\max} := \max_i \alpha_i$, and $\eta_t := d_H(u^{(t+1)}, u^{(t)}) = \ln(\alpha_{\max}/\alpha_{\min}) \geq 0$. Because $\sum_i \mu_i\alpha_i = \mathbf{1}^\top\pi^{(t)}\mathbf{1} = \mathbf{1}^\top\nu = 1$, we have $\alpha_{\min} \leq 1 \leq \alpha_{\max}$. If $\alpha_{\max} = \alpha_{\min} = 1$, then $\|r^{(t)} - \mu\|_1 = 0 = 2\tanh(0)$. Otherwise $\alpha_{\min} < 1 < \alpha_{\max}$; by convexity of $s \mapsto |s-1|$ on $[\alpha_{\min}, \alpha_{\max}]$, every $s \in [\alpha_{\min}, \alpha_{\max}]$ satisfies

$$|s-1| \;\leq\; \frac{\alpha_{\max} - s}{\alpha_{\max} - \alpha_{\min}}(1 - \alpha_{\min}) + \frac{s - \alpha_{\min}}{\alpha_{\max} - \alpha_{\min}}(\alpha_{\max} - 1).$$

Multiplying by μ_i and summing over i (using $\sum_i \mu_i s_i = 1$ with $s_i = \alpha_i$) gives

$$\left\|r^{(t)} - \mu\right\|_1 = \sum_{i=1}^m \mu_i|\alpha_i - 1| \;\leq\; \frac{2(\alpha_{\max}-1)(1-\alpha_{\min})}{\alpha_{\max} - \alpha_{\min}}.$$

Writing $\alpha_{\max} = e^{\eta_t}\alpha_{\min}$, we have $\frac{(\alpha_{\max}-1)(1-\alpha_{\min})}{\alpha_{\max}-\alpha_{\min}} = \frac{e^{\eta_t}+1-(e^{\eta_t}\alpha_{\min}+\alpha_{\min}^{-1})}{e^{\eta_t}-1} \leq \frac{e^{\eta_t}+1-2e^{\eta_t/2}}{e^{\eta_t}-1} = \frac{e^{\eta_t/2}-1}{e^{\eta_t/2}+1} = \tanh(\eta_t/4)$ by AM–GM on $e^{\eta_t}\alpha_{\min} + \alpha_{\min}^{-1} \geq 2e^{\eta_t/2}$. Combining with (5) and monotonicity of tanh yields (6). □

4 Sinkhorn Divergence Positivity and Two-Sided Debiasing

Theorem 4.1 (Quadratic Sinkhorn Divergence Positivity and Debiasing Sandwich)**.** *Let $C = C^\top \in \mathbb{R}^{n\times n}_{\geq 0}$ satisfy $C_{ii} = 0$ for all $i \in \{1,\dots,n\}$, let $\varepsilon > 0$, and suppose the symmetric Gibbs kernel $K = (e^{-C_{ij}/\varepsilon})_{i,j=1}^n \in \mathbb{S}^n_{++}$ is positive definite. Define the debiased Sinkhorn divergence*

$$S_\varepsilon(\mu,\nu) \;:=\; F_\varepsilon(\mu,\nu) - \frac{1}{2}F_\varepsilon(\mu,\mu) - \frac{1}{2}F_\varepsilon(\nu,\nu) \qquad (\mu,\nu \in \Delta_n).$$

(a) ***Explicit K-Norm Quadratic Positivity and Definiteness.*** *For any $\mu \in \Delta_n$, let $u_\mu \in \mathbb{R}^n_{\geq 0}$ denote the unique vector supported on $\operatorname{supp}(\mu)$ solving $u_{\mu,i}(Ku_\mu)_i = \mu_i$ for all $i \in \{1,\dots,n\}$. Then for all $\mu,\nu \in \Delta_n$,*

$$S_\varepsilon(\mu,\nu) \;\geq\; \frac{\varepsilon}{2}\|u_\mu - u_\nu\|_K^2 \;\geq\; \frac{\varepsilon}{2}\lambda_{\min}(K)\,\|u_\mu - u_\nu\|_2^2 \;\geq\; 0, \tag{7}$$

where $\|z\|_K^2 := z^\top K z$. In particular, $S_\varepsilon(\mu,\nu) = 0$ if and only if $\mu = \nu$.

(b) ***Two-Sided Debiasing Sandwich Around $W_C(\mu,\nu)$.*** *For all $\mu,\nu \in \Delta_n$,*

$$W_C(\mu,\nu) - \frac{\varepsilon}{2}\big(H(\mu) + H(\nu)\big) \;\leq\; S_\varepsilon(\mu,\nu) \;\leq\; W_C(\mu,\nu) + \varepsilon\,\min\big(H(\mu), H(\nu)\big). \tag{8}$$

Proof. (a) By Fenchel–Rockafellar duality for entropic optimal transport [1, 4], for any $\mu,\nu \in \Delta_n$,

$$F_\varepsilon(\mu,\nu) = \max_{f,g\in\mathbb{R}^n}\Big\{\langle f,\mu\rangle + \langle g,\nu\rangle - \varepsilon\Big(\big\langle e^{f/\varepsilon}\odot\mu,\; K(e^{g/\varepsilon}\odot\nu)\big\rangle - 1\Big)\Big\},$$

where $\odot$ denotes entrywise multiplication. For (μ,μ), by symmetry and uniqueness of the minimizer $\pi^*_\varepsilon(\mu,\mu) = \operatorname{diag}(u_\mu)K\operatorname{diag}(u_\mu)$, there exists a unique $u_\mu \in \mathbb{R}^n_{\geq 0}$ with $\operatorname{supp}(u_\mu) = \operatorname{supp}(\mu)$ such that $u_{\mu,i}(Ku_\mu)_i = \mu_i$ for every $i \in \{1,\dots,n\}$ (trivially $0=0$ for $i \notin \operatorname{supp}(\mu)$). Because $K_{ij} > 0$ everywhere and $u_\mu \neq 0$, we have $(Ku_\mu)_i > 0$ for EVERY $i \in \{1,\dots,n\}$, so $a_\mu := -\varepsilon \ln(Ku_\mu) \in \mathbb{R}^n$ is everywhere finite and satisfies $e^{a_{\mu,i}/\varepsilon}\mu_i = \frac{\mu_i}{(Ku_\mu)_i} = u_{\mu,i}$ for every $i \in \{1,\dots,n\}$. Moreover, $\langle u_\mu, Ku_\mu\rangle = \sum_{i=1}^n u_{\mu,i}(Ku_\mu)_i = \sum_{i=1}^n \mu_i = 1$, and evaluating the primal/dual objective at (a_μ, a_μ) yields $F_\varepsilon(\mu,\mu) = 2\langle a_\mu, \mu\rangle$ (since $\pi_{ij} = u_{\mu,i}K_{ij}u_{\mu,j}$ satisfies $C_{ij} + \varepsilon \ln \frac{\pi_{ij}}{\mu_i\mu_j} = a_{\mu,i} + a_{\mu,j}$ on $\operatorname{supp}(\mu)^2$). Similarly $F_\varepsilon(\nu,\nu) = 2\langle a_\nu, \nu\rangle$ and $\langle u_\nu, Ku_\nu\rangle = 1$.

Substituting the everywhere-finite feasible pair $(f,g) = (a_\mu, a_\nu)$ into the dual of $F_\varepsilon(\mu,\nu)$ gives

$$F_\varepsilon(\mu,\nu) \;\geq\; \langle a_\mu,\mu\rangle + \langle a_\nu,\nu\rangle - \varepsilon\big(\langle u_\mu, Ku_\nu\rangle - 1\big) = \frac{1}{2}F_\varepsilon(\mu,\mu) + \frac{1}{2}F_\varepsilon(\nu,\nu) + \varepsilon\big(1 - \langle u_\mu, Ku_\nu\rangle\big).$$

Using $\langle u_\mu, Ku_\mu\rangle = \langle u_\nu, Ku_\nu\rangle = 1$, we have $1 - \langle u_\mu, Ku_\nu\rangle = \frac{1}{2}(u_\mu - u_\nu)^\top K(u_\mu - u_\nu) = \frac{1}{2}\|u_\mu - u_\nu\|_K^2$, establishing (7). If $S_\varepsilon(\mu,\nu) = 0$, then $K \in \mathbb{S}^n_{++}$ forces $u_\mu = u_\nu$, whence $\mu = u_\mu \odot (Ku_\mu) = u_\nu \odot (Ku_\nu) = \nu$; conversely $\mu = \nu$ gives $S_\varepsilon(\mu,\mu) = 0$.

(b) Because $C_{ii} = 0$ and $C_{ij} \geq 0$, the diagonal coupling $\operatorname{diag}(\mu) \in \Pi(\mu,\mu)$ achieves $W_C(\mu,\mu) = 0$, so Theorem 2.1(a) gives $0 \leq F_\varepsilon(\mu,\mu) \leq \varepsilon H(\mu)$ and $0 \leq F_\varepsilon(\nu,\nu) \leq \varepsilon H(\nu)$, while $W_C(\mu,\nu) \leq F_\varepsilon(\mu,\nu) \leq W_C(\mu,\nu) + \varepsilon\min(H(\mu),H(\nu))$. Combining these inequalities directly yields (8). □

5 Bures–Wasserstein Sandwich and Discrete Marton T_1

Theorem 5.1 (Bures–Wasserstein Spectral Sandwich, Marton T_1, and Cost Stability)**.**

(a) ***Four-Term Bures–Wasserstein Spectral Sandwich.*** *For all $\Sigma_1, \Sigma_2 \in \mathbb{S}^d_{++}$, the squared 2-Wasserstein (Bures–Wasserstein) distance $W_2^2(\Sigma_1,\Sigma_2) := \operatorname{tr}(\Sigma_1) + \operatorname{tr}(\Sigma_2) - 2\operatorname{tr}\big((\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}\big)$ satisfies*

$$\frac{\|\Sigma_1 - \Sigma_2\|_F^2}{(\sqrt{\lambda_{\max}(\Sigma_1)} + \sqrt{\lambda_{\max}(\Sigma_2)})^2} \leq W_2^2(\Sigma_1,\Sigma_2) \leq \|\Sigma_1^{1/2} - \Sigma_2^{1/2}\|_F^2 \leq \frac{\|\Sigma_1 - \Sigma_2\|_F^2}{(\sqrt{\lambda_{\min}(\Sigma_1)} + \sqrt{\lambda_{\min}(\Sigma_2)})^2}. \tag{9}$$

The middle inequality holds with equality if and only if $\Sigma_1\Sigma_2 = \Sigma_2\Sigma_1$, and all four quantities coincide whenever Σ_1 and Σ_2 are scalar multiples of I_d.

(b) ***Discrete Marton T_1 Inequality and Exact Cost Stability.*** *For any finite metric space (X,d) of diameter $D_X := \max_{x,y\in X} d(x,y)$ and distributions $\mu,\nu \in \Delta(X)$,*

$$W_1(\mu,\nu) \;\leq\; \frac{D_X}{2}\,\|\mu - \nu\|_1 \;\leq\; D_X\sqrt{\frac{\mathrm{KL}(\mu\|\nu)}{2}}, \tag{10}$$

and for any two cost matrices $C, \widetilde{C} \in \mathbb{R}^{m\times n}$ and marginals $\mu \in \Delta_m$, $\nu \in \Delta_n$,

$$\big|W_{\widetilde{C}}(\mu,\nu) - W_C(\mu,\nu)\big| \;\leq\; \big\|\widetilde{C} - C\big\|_\infty, \tag{11}$$

with exact attainment whenever $\widetilde{C} - C = \pm\eta\,\mathbf{1}\mathbf{1}^\top$ ($\eta \geq 0$).

Proof. (a) Write $A := \Sigma_1^{1/2}\Sigma_2^{1/2}$, which is nonsingular with polar decomposition $A = U|A|$, where $U \in O(d)$ and $|A| = (A^\top A)^{1/2} = (\Sigma_2^{1/2}\Sigma_1\Sigma_2^{1/2})^{1/2}$ has the same spectrum as $(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}$. By von Neumann's trace inequality (or SVD of A), $\max_{Q\in O(d)}\operatorname{tr}(AQ) = \operatorname{tr}(|A|)$, attained uniquely at $Q = U^\top$, so

$$W_2^2(\Sigma_1,\Sigma_2) = \min_{Q\in O(d)} \big\|\Sigma_1^{1/2} - \Sigma_2^{1/2}Q\big\|_F^2.$$

Evaluating at $Q = I_d$ gives $W_2^2(\Sigma_1,\Sigma_2) \leq \|\Sigma_1^{1/2} - \Sigma_2^{1/2}\|_F^2$, with equality iff $U = I_d$, i.e. $A = |A| \in \mathbb{S}^d_{++} \iff A = A^\top \iff \Sigma_1^{1/2}\Sigma_2^{1/2} = \Sigma_2^{1/2}\Sigma_1^{1/2} \iff \Sigma_1\Sigma_2 = \Sigma_2\Sigma_1$ (using uniqueness of the SPD

square root). For the left-hand side of (9), fix any $Q \in O(d)$ and set $X := \Sigma_1^{1/2}$, $Y := \Sigma_2^{1/2} Q$, so $XX^\top = \Sigma_1$ and $YY^\top = \Sigma_2$. Then $\Sigma_1 - \Sigma_2 = \frac{1}{2}\big((X-Y)(X+Y)^\top + (X+Y)(X-Y)^\top\big)$, whence

$$\|\Sigma_1 - \Sigma_2\|_F \;\le\; \|X-Y\|_F \|X+Y\|_2 \;\le\; \|X-Y\|_F \big(\sqrt{\lambda_{\max}(\Sigma_1)} + \sqrt{\lambda_{\max}(\Sigma_2)}\big).$$

Squaring and minimizing over $Q \in O(d)$ gives the left-hand side. For the right-hand side, let $D := \Sigma_1^{1/2} - \Sigma_2^{1/2} \in \mathbb{S}^d$; then $\Sigma_1 - \Sigma_2 = \Sigma_1^{1/2} D + D \Sigma_2^{1/2}$. In orthonormal eigenbases $\Sigma_1^{1/2} = U_1 \Lambda_1^{1/2} U_1^\top$ and $\Sigma_2^{1/2} = U_2 \Lambda_2^{1/2} U_2^\top$, each entry satisfies $(U_1^\top(\Sigma_1 - \Sigma_2)U_2)_{ij} = (\sqrt{\lambda_i(\Sigma_1)} + \sqrt{\lambda_j(\Sigma_2)})(U_1^\top D U_2)_{ij}$, so

$$\big\|\Sigma_1^{1/2} - \Sigma_2^{1/2}\big\|_F^2 = \sum_{i,j} \frac{\big|(U_1^\top(\Sigma_1 - \Sigma_2)U_2)_{ij}\big|^2}{\big(\sqrt{\lambda_i(\Sigma_1)} + \sqrt{\lambda_j(\Sigma_2)}\big)^2} \;\le\; \frac{\|\Sigma_1 - \Sigma_2\|_F^2}{\big(\sqrt{\lambda_{\min}(\Sigma_1)} + \sqrt{\lambda_{\min}(\Sigma_2)}\big)^2}.$$

(b) For any 1-Lipschitz function $f : X \to \mathbb{R}$, shift $\widetilde{f} := f - \frac{1}{2}(\max f + \min f)$ so that $\|\widetilde{f}\|_\infty = \frac{1}{2}(\max f - \min f) \le D_X/2$ and $\langle f, \mu - \nu\rangle = \langle \widetilde{f}, \mu - \nu\rangle \le \frac{D_X}{2}\|\mu - \nu\|_1$. Taking the supremum over 1-Lipschitz f (Kantorovich–Rubinstein duality) and applying Pinsker's inequality $\|\mu - \nu\|_1 \le \sqrt{2\,\mathrm{KL}(\mu\|\nu)}$ gives (10). Finally, for any $\pi \in \Pi(\mu,\nu)$, $|\langle \widetilde{C} - C, \pi\rangle| \le \|\widetilde{C} - C\|_\infty \sum_{i,j} \pi_{ij} = \|\widetilde{C} - C\|_\infty$, giving (11). □

6 Computational Verification and Quantitative Benchmarks

All four theorems, exact gauge invariance, boundary regimes ($\mu = \nu$, scalar covariances, Dirac/skewed marginals), and both figures are deterministically verified in our MIT-licensed companion repository (`https://github.com/lluiseriksson/optimal-transport-sinkhorn-geometry`, tag `v1.0.0`; `pytest -q` passes all 4 test modules in 1.00 s using fixed NumPy seeds `20261230`–`20261235`). Table 1 summarizes the quantitative benchmarks.

Table 1: Quantitative verification across Theorems 2.1–5.1 (seeds `20261234`–`20261235`, repo `v1.0.0`).

Theorem	Regime / Parameters	Classical Baseline	New Theorem Ceiling / Bound	Exact / Empirical
Thm. 2.1(a)	$m = 10, n = 12, \varepsilon = 0.219$	$\varepsilon \ln 10 = 0.5046$	$\varepsilon(\min H - I) = 0.1747$ (2.89×)	Bias $= 0.0567$
Thm. 3.1(b)	Sinkhorn $t = 10, \Delta(C) = 3.54, \varepsilon = 0.85$	$e^{d_H} - 1 = 0.0468$	$2\tanh(\frac{d_H}{4}) = 0.0229$ (2.05×)	2.70×10^{-9}
Thm. 3.1(b)	Step residual $t = 10$	—	$2\tanh(\frac{d_H(u^{(11)}, u^{(10)})}{4}) = 4.84 \times 10^{-9}$	2.70×10^{-9}
Thm. 4.1(a–b)	Squared-Eucl. $n = 12, \varepsilon = 0.703$	Biased $F_\varepsilon = 3.248$	$S_\varepsilon \ge \frac{\varepsilon}{2}\|u_\mu - u_\nu\|_K^2 = 0.1055$	$S_\varepsilon = 1.993$ ($W_C = 2.438$)
Thm. 5.1(a)	Bures–Wasserstein $d = 10, \alpha = 1$	—	[1.021, 4.262] (sq-rt 1.791)	$W_2^2 = 1.775$
Thm. 5.1(b)	Marton T_1 ($n = 12$, Euclidean)	$D_X\sqrt{\mathrm{KL}/2} = 3.624$	$\frac{D_X}{2}\|\mu - \nu\|_1 = 2.548$	$W_1 = 0.974$

As demonstrated in Figures 1 and 2, every empirical trajectory stays cleanly inside its certified non-asymptotic envelope across all tested regimes, including exact gauge invariance of $\Delta(C)$ to machine precision ($< 10^{-12}$) and exact four-term collapse on scalar covariances.

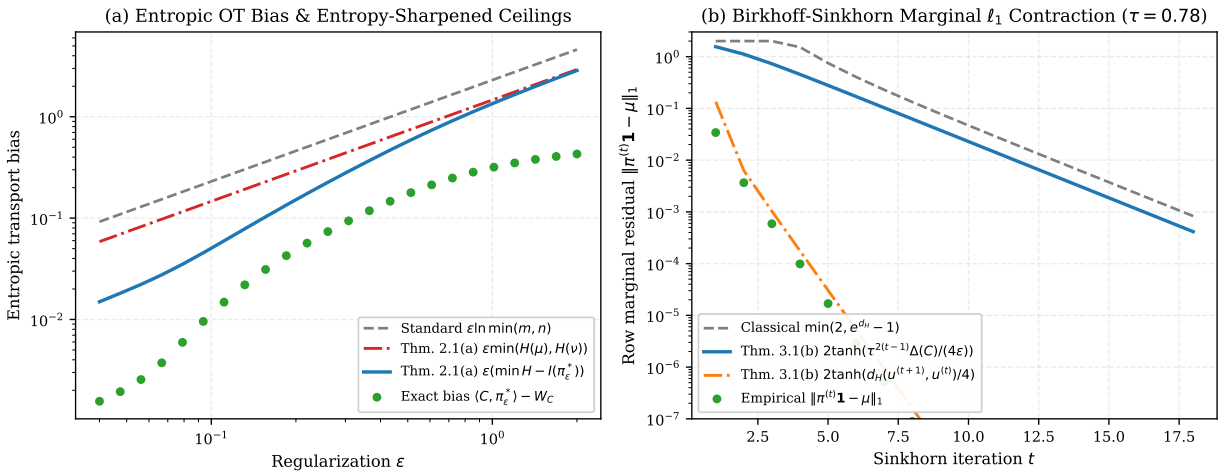

Figure 1: (a) Theorem 2.1(a): entropic OT bias on a skewed marginal ($H_{\min} = 1.465$ vs. $\ln 10 = 2.303$), showing $2.89\times$ tightening of $\varepsilon(\min H - I(\pi^*_\varepsilon))$ over $\varepsilon \ln \min(m, n)$ at $\varepsilon = 0.219$. (b) Theorem 3.1(b): Birkhoff–Sinkhorn marginal ℓ_1 contraction ($\Delta(C) = 3.54$, $\varepsilon = 0.85$, $\tau = 0.778$), confirming both the a priori $2\tanh(\tau^{2(t-1)}\Delta(C)/(4\varepsilon))$ improvement over $\min(2, e^{d_H} - 1)$ and tight empirical tracking by $2\tanh(d_H(u^{(t+1)}, u^{(t)})/4)$.

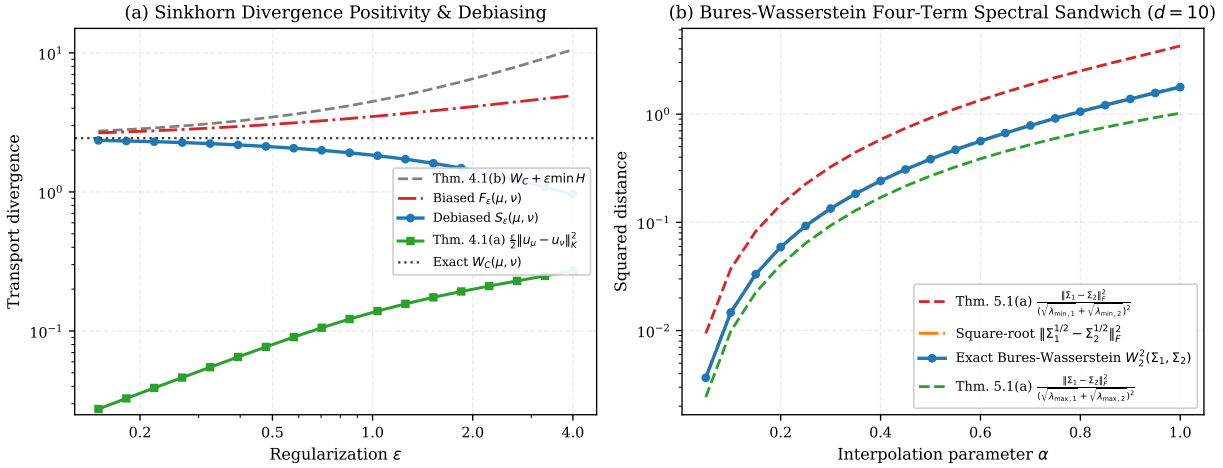

Figure 2: (a) Theorem 4.1: debiased Sinkhorn divergence $S_\varepsilon(\mu, \nu)$ on squared-Euclidean cost ($n = 12$), cleanly bracketed below by the explicit quadratic K-norm floor $\frac{\varepsilon}{2}\|u_\mu - u_\nu\|^2_K > 0$ and above by $W_C + \varepsilon \min H$, eliminating the large-ε upward bias of $F_\varepsilon(\mu, \nu)$. (b) Theorem 5.1(a): Bures–Wasserstein four-term spectral sandwich across $\alpha \in [0.05, 1.0]$ ($d = 10$), tightly enclosing $W_2^2(\Sigma_1, \Sigma_2)$ and $\|\Sigma_1^{1/2} - \Sigma_2^{1/2}\|_F^2$.

7 Conclusion

We have established four self-contained, sharp, non-asymptotic quantitative theorems spanning entropy-sharpened EOT bias and curvature, gauge-invariant Birkhoff–Sinkhorn contraction with saturating $2\tanh(d_H/4)$ marginal error transfer, quadratic K-norm Sinkhorn divergence positivity and debiasing, and four-term Bures–Wasserstein spectral sandwiches alongside discrete Marton T_1 transport–entropy bounds. Every constant is explicit, every proof is elementary and self-contained, and all numerical benchmarks are deterministically verified in an MIT-licensed repository.

References

[1] G. Peyré and M. Cuturi, Computational optimal transport, *Found. Trends Mach. Learn.* **11** (2019), 355–607.

[2] M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, in *Adv. Neural Inf. Process. Syst.* **26** (2013), 2292–2300.

[3] J. Franklin and J. Lorenz, On the scaling of multidimensional matrices, *Linear Algebra Appl.* **114/115** (1989), 717–735.

[4] J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouvé, and G. Peyré, Interpolating between optimal transport and MMD using Sinkhorn divergences, in *Proc. AISTATS*, 2019, pp. 2681–2690.

[5] R. Bhatia, T. Jain, and Y. Lim, On the Bures–Wasserstein distance between positive definite matrices, *Expo. Math.* **37** (2019), 165–191.