AIRR.SCIENCE · ARR-2026-462CJ3QN079CHS3H v1 [Mathematics] 28 Nov 2026

THE EXACT SCHOENBERG–FEYDY POSITIVITY TRICHOTOMY, FLAT ZERO-LOCI, AND CRITICAL REGULARIZATION PHASE TRANSITIONS OF SINKHORN DIVERGENCES

LLUIS ERIKSSON

Abstract. Debiased Sinkhorn divergences $S_\varepsilon(\alpha,\beta) = \mathrm{OT}_\varepsilon(\alpha,\beta) - \frac{1}{2}\mathrm{OT}_\varepsilon(\alpha,\alpha) - \frac{1}{2}\mathrm{OT}_\varepsilon(\beta,\beta)$ interpolate between optimal transport ($\varepsilon \to 0^+$) and maximum mean discrepancies ($\varepsilon \to \infty$). A foundational theorem of Feydy, Séjourné, Vialard, Amari, Trouvé, and Peyré (2019) proves that if the Gibbs kernel $K_\varepsilon = (e^{-C_{ij}/\varepsilon})$ is strictly positive definite, then S_ε is non-negative and strictly positive definite ($S_\varepsilon(\alpha,\beta) > 0$ for $\alpha \neq \beta$). Whether positive semidefiniteness of K_ε is also *necessary*, what happens on the singular boundary $\lambda_{\min}(K_\varepsilon) = 0$, and how positivity behaves as a function of ε for non-Euclidean and graph costs have remained uncharacterized.

We establish the complete converse and a sharp three-way positivity trichotomy for every finite symmetric cost matrix $C = C^\top \in \mathbb{R}^{n\times n}$ and every $\varepsilon > 0$. First, using gauge-fixed Legendre duality and the implicit function theorem, we derive the exact diagonal Hessian of S_ε along antisymmetric perturbations $(\mu + t\eta, \mu - t\eta)$ in terms of the normalized symmetric Sinkhorn self-coupling $S_\mu = D_\mu^{-1/2} P_\mu D_\mu^{-1/2}$:

$$\left.\frac{d^2}{dt^2}\right|_{t=0} S_\varepsilon(\mu + t\eta, \mu - t\eta) \;=\; 4\varepsilon\, \tilde\eta^{\,\top} \frac{S_\mu}{I - S_\mu^2}\, \tilde\eta \qquad (\tilde\eta := D_\mu^{-1/2}\eta \perp \sqrt{\mu}).$$

Second, combining this Hessian formula with Sylvester's law of inertia for the congruence between S_μ and K_ε, we prove the exact Schoenberg–Feydy trichotomy: (i) $S_\varepsilon(\alpha,\beta) > 0$ for all $\alpha \neq \beta$ if and only if $K_\varepsilon \succ 0$; (ii) $S_\varepsilon \geq 0$ with a non-trivial zero-locus if and only if $K_\varepsilon \succeq 0$ is singular, in which case $S_\varepsilon(\alpha,\beta) = 0$ holds if and only if $u_\alpha - u_\beta \in \ker(K_\varepsilon)$, generating flat affine zero-loci; and (iii) $\inf_{\alpha,\beta} S_\varepsilon(\alpha,\beta) < 0$ if and only if $\lambda_{\min}(K_\varepsilon) < 0$, in which case negativity already occurs infinitesimally along the diagonal at *every* strictly positive measure $\mu \in \Delta_n^+$. Third, by Schoenberg's theorem, universal positivity ($S_\varepsilon \geq 0$ for all $\varepsilon > 0$) holds if and only if C is conditionally negative definite; for complete bipartite graph metrics $K_{m,n}$ ($(m-1)(n-1) \geq 2$) we prove the exact critical threshold $\varepsilon_c(K_{m,n}) = 2/\log((m-1)(n-1))$, while for super-quadratic costs $|x-y|^p$ ($p > 2$) we prove $\varepsilon_c(p) = \frac{1}{(p-2)\log 2} + \frac{3}{2} + O(p-2)$ as $p \to 2^+$ on $\{0,1,2\}$ and deduce continuum collapse: on every compact interval $[0, L]$ ($L > 0$), S_ε attains strictly negative values for *every* $\varepsilon > 0$. Fourth, we prove that $\mathcal{B}_+(C) := \{\beta > 0 : S_{1/\beta} \geq 0\}$ is a closed additive subsemigroup of $(0,\infty)$ via the Schur product theorem, yet—contrary to naive intuition—positivity is *not* monotone in ε: using FitzGerald–Horn Hadamard powers of cosine Gram matrices, we construct explicit $n \times n$ costs ($n \geq 4$) exhibiting open re-entrant positivity islands separated from the low-ε positive phase by an intermediate negative-divergence gap.

1. Introduction

Entropic optimal transport (EOT) replaces the linear program of classical Monge–Kantorovich optimal transport by a strictly convex Kullback–Leibler regularized problem [1, 6]: on a finite space $X = \{1,\dots,n\}$ equipped with a symmetric cost matrix $C = C^\top \in \mathbb{R}^{n\times n}$ and regularization

2020 *Mathematics Subject Classification.* Primary 49Q22, 46E22; Secondary 15B48, 05C12, 90C25.

Key words and phrases. Sinkhorn divergence, entropic optimal transport, Schoenberg's theorem, conditionally negative definite cost, Gibbs kernel, Hadamard powers, re-entrant phase transition.

ORCID: 0009-0006-7494-9040. Code, 50-digit tables, figures, and verification suite archived at https://github.com/lluiseriksson/sinkhorn-positivity-trichotomy (release tag `v1.0.2`).

parameter $\varepsilon > 0$,

$$\mathrm{OT}_\varepsilon(\alpha,\beta) \;:=\; \min_{\pi\in\Pi(\alpha,\beta)} \Big\{ \langle C,\pi\rangle + \varepsilon\,\mathrm{KL}(\pi\|\alpha\otimes\beta) \Big\}, \qquad \alpha,\beta\in\Delta_n, \tag{1}$$

where $\Delta_n := \{\alpha\in\mathbb{R}^n_{\geq 0} : \sum_{i=1}^n \alpha_i = 1\}$ is the probability simplex, $\Pi(\alpha,\beta) := \{\pi\in\mathbb{R}^{n\times n}_{\geq 0} : \pi\mathbf{1} = \alpha,\ \pi^\top\mathbf{1} = \beta\}$, and $\mathrm{KL}(\pi\|\alpha\otimes\beta) := \sum_{i,j}\pi_{ij}\log\frac{\pi_{ij}}{\alpha_i\beta_j}$ (with $0\log 0 = 0$). Because $\mathrm{OT}_\varepsilon(\alpha,\alpha)$ is generally non-zero and $\alpha\mapsto\mathrm{OT}_\varepsilon(\alpha,\beta)$ is not minimized at $\alpha=\beta$, Genevay, Peyré, and Cuturi [4] and Feydy et al. [2] introduced the debiased *Sinkhorn divergence*:

$$S_\varepsilon(\alpha,\beta) \;:=\; \mathrm{OT}_\varepsilon(\alpha,\beta) - \frac{1}{2}\,\mathrm{OT}_\varepsilon(\alpha,\alpha) - \frac{1}{2}\,\mathrm{OT}_\varepsilon(\beta,\beta). \tag{2}$$

Sinkhorn divergences interpolate smoothly between classical optimal transport $\lim_{\varepsilon\to 0^+} S_\varepsilon(\alpha,\beta) = \mathrm{OT}_0(\alpha,\beta)$ (when $C\geq 0$ has zero diagonal, so $\mathrm{OT}_0(\alpha,\alpha) = \mathrm{OT}_0(\beta,\beta) = 0$; for arbitrary symmetric C, the $\varepsilon\to 0^+$ limit is $\mathrm{OT}_0(\alpha,\beta) - \frac{1}{2}\mathrm{OT}_0(\alpha,\alpha) - \frac{1}{2}\mathrm{OT}_0(\beta,\beta)$) and the squared energy distance $\lim_{\varepsilon\to\infty} S_\varepsilon(\alpha,\beta) = -\frac{1}{2}(\alpha-\beta)^\top C(\alpha-\beta)$. In [2] (see also [5] for the local Riemannian geometry when $K_\varepsilon \succ 0$), it was proved that if the Gibbs kernel

$$K_\varepsilon \;:=\; \big(e^{-C_{ij}/\varepsilon}\big)_{i,j=1}^n \;\in\; \mathbb{R}^{n\times n}_{>0} \tag{3}$$

is strictly positive definite ($K_\varepsilon \succ 0$), then S_ε is non-negative and strictly positive definite: $S_\varepsilon(\alpha,\beta)\geq 0$ with equality if and only if $\alpha=\beta$.

This sufficient condition naturally raises three fundamental mathematical questions:

(1) *Sharpness and Converse:* Is positive semidefiniteness of K_ε also *necessary* for $S_\varepsilon\geq 0$? Equivalently, can a non-PSD Gibbs kernel ever still generate a non-negative Sinkhorn divergence?
(2) *Boundary Geometry and Non-CND Metrics:* What happens at the exact singular boundary $\lambda_{\min}(K_\varepsilon) = 0$, and what are the exact critical regularization thresholds $\varepsilon_c(C)$ for canonical non-negative-type graph metrics (such as $K_{m,n}$) and super-quadratic Euclidean powers $|x-y|^p$ ($p>2$)?
(3) *Monotonicity in ε:* Since $\lim_{\varepsilon\to 0^+} S_\varepsilon(\alpha,\beta) = \mathrm{OT}_0(\alpha,\beta)\geq 0$ for every metric cost, if $S_{\varepsilon_0}\geq 0$ at some regularization $\varepsilon_0>0$, must $S_\varepsilon\geq 0$ hold for all smaller $\varepsilon\in(0,\varepsilon_0)$?

In this paper we answer all three questions completely and sharply.

Overview of Main Results.

- **Exact Diagonal Hessian (Theorem 2.3).** For any symmetric $C = C^\top\in\mathbb{R}^{n\times n}$, any $\varepsilon>0$, any $\mu\in\Delta_n^+ := \{\mu\in\Delta_n : \mu_i>0\,\forall i\}$, and any $\eta\in\mathbf{1}^\perp$,

$$\frac{d^2}{dt^2}\Big|_{t=0} S_\varepsilon(\mu+t\eta,\mu-t\eta) \;=\; 4\varepsilon\,\tilde\eta^\top\frac{S_\mu}{I-S_\mu^2}\,\tilde\eta,$$

  where $P_\mu = \mathrm{diag}(u_\mu)K_\varepsilon\,\mathrm{diag}(u_\mu)$ is the unique symmetric Sinkhorn self-coupling of μ, $S_\mu := D_\mu^{-1/2}P_\mu D_\mu^{-1/2}$, and $\tilde\eta := D_\mu^{-1/2}\eta\perp\sqrt{\mu}$. Because every non-principal eigenvalue $\lambda_k(S_\mu)$ lies in $(-1,1)$, the curvature factor $\lambda_k/(1-\lambda_k^2)$ has the exact same sign as $\lambda_k(S_\mu)$.
- **The Exact Schoenberg–Feydy Positivity Trichotomy (Theorem 3.1).** For every symmetric $C = C^\top\in\mathbb{R}^{n\times n}$ and every $\varepsilon>0$:
  (i) $S_\varepsilon(\alpha,\beta)>0$ for all $\alpha\neq\beta\in\Delta_n \iff K_\varepsilon\succ 0$;
  (ii) $S_\varepsilon\geq 0$ globally with $S_\varepsilon(\alpha,\beta)=0$ for some $\alpha\neq\beta \iff K_\varepsilon\succeq 0$ and $\ker(K_\varepsilon)\neq\{0\}$; moreover, for $K_\varepsilon\succeq 0$, the exact zero-locus is $\{(\alpha,\beta)\in\Delta_n\times\Delta_n : u_\alpha - u_\beta\in\ker(K_\varepsilon)\}$, which contains non-empty open 1-dimensional flat affine segments in $\Delta_n^+\times\Delta_n^+$ as well as disjoint-support boundary pairs;

(iii) $\inf_{\alpha,\beta\in\Delta_n} S_\varepsilon(\alpha,\beta) < 0 \iff \lambda_{\min}(K_\varepsilon) < 0$ (and negativity already occurs infinitesimally along the diagonal at *every* interior measure $\mu \in \Delta_n^+$ simultaneously, by Sylvester's law of inertia).

- **Universal Positivity, Exact Graph Thresholds, and Continuum Collapse (Theorem 4.2 and Theorem 4.3).** By Schoenberg's theorem [7], $S_\varepsilon \geq 0$ for all $\varepsilon > 0$ if and only if C is conditionally negative definite (CND). For complete bipartite graphs $K_{m,n}$ $(m,n \geq 2)$ with shortest-path metric, universal positivity holds if and only if $K_{2,2} = C_4$; for all $(m-1)(n-1) \geq 2$, the exact critical threshold is $\varepsilon_c(K_{m,n}) = 2/\log((m-1)(n-1))$, yielding the universal upper bound $\varepsilon_c(G) \leq 2/\log((m-1)(n-1))$ for any connected finite graph G containing $K_{m,n}$ $((m-1)(n-1) \geq 2)$ isometrically. For super-quadratic power costs $|x-y|^p$ $(p > 2)$ on $\{0,1,2\}$, we derive the exact transcendental phase curve $\varepsilon_c(p)$ with sharp asymptotics $\varepsilon_c(p) = \frac{1}{(p-2)\log 2} + \frac{3}{2} + O(p-2)$ as $p \to 2^+$ and $\varepsilon_c(p) = \frac{2}{\log 2} + O(2^{-2^{p-1}})$ as $p \to \infty$, and prove *continuum collapse*: on every compact subset $X \subset \mathbb{R}^d$ containing a non-degenerate line segment, S_ε with cost $\|x-y\|^p$ $(p > 2)$ attains strictly negative values for *every* $\varepsilon > 0$, despite converging pointwise to $W_p^p \geq 0$ as $\varepsilon \to 0^+$.
- **Additive Positivity Semigroup and FitzGerald–Horn Re-Entrant Islands (Theorem 5.1 and Theorem 5.2).** Via the Schur product theorem, $\mathcal{B}_+(C) := \{\beta > 0 : S_{1/\beta} \geq 0 \text{ on } \Delta_n \times \Delta_n\}$ is a closed additive subsemigroup of $(0,\infty)$ containing $[\beta_{\text{crit}}(C), \infty)$ whenever $C_{ii} = 0 < C_{ij}$ $(i \neq j)$. Strikingly, positivity is *not* monotone in ε: using FitzGerald–Horn Hadamard powers [3] of perturbed 4×4 cosine Gram matrices $A_\delta = (1-\delta)A_0 + \delta I_4$, we prove via the exact identity $A_\delta^{\circ\beta} = (1-\delta)^\beta A_0^{\circ\beta} + (1-(1-\delta)^\beta) I_4$ that for every $n \geq 4$ there exist symmetric zero-diagonal costs $C \geq 0$ whose positivity set $\{\varepsilon > 0 : S_\varepsilon > 0 \text{ off-diagonal}\}$ is disconnected, featuring an open high-ε positivity island strictly separated from the low-ε positive phase by a re-entrant negative-divergence gap.

2. Dual Structure and the Exact Diagonal Hessian of S_ε

Throughout, $C = C^\top \in \mathbb{R}^{n\times n}$ $(n \geq 2)$ is an arbitrary real symmetric matrix, $\varepsilon > 0$, and $K_\varepsilon = (e^{-C_{ij}/\varepsilon})_{i,j=1}^n \in \mathbb{R}_{>0}^{n\times n}$ is symmetric with strictly positive entries. Denote the Hadamard (entrywise) product by $x \odot y$, the all-ones vector by $\mathbf{1} \in \mathbb{R}^n$, and the Shannon entropy by $H(\alpha) := -\sum_{i=1}^n \alpha_i \log \alpha_i$ (with $0 \log 0 = 0$).

Lemma 2.1 (Dual representation and marginal-entropy cancellation)**.** *Define the convex conjugate pair on $\mathbb{R}^n \times \mathbb{R}^n$ and $\Delta_n \times \Delta_n$:*

$$\Phi_\varepsilon^*(f,g) := \varepsilon \left\langle e^{f/\varepsilon}, K_\varepsilon e^{g/\varepsilon} \right\rangle, \tag{4}$$

$$\Phi_\varepsilon(\alpha,\beta) := \sup_{f,g\in\mathbb{R}^n} \Big\{ \langle f,\alpha\rangle + \langle g,\beta\rangle - \Phi_\varepsilon^*(f,g) \Big\}. \tag{5}$$

Then for all $\alpha, \beta \in \Delta_n$,

$$\mathrm{OT}_\varepsilon(\alpha,\beta) = \Phi_\varepsilon(\alpha,\beta) + \varepsilon + \varepsilon H(\alpha) + \varepsilon H(\beta), \tag{6}$$

and consequently the Sinkhorn divergence satisfies the exact identity

$$S_\varepsilon(\alpha,\beta) = \Phi_\varepsilon(\alpha,\beta) - \frac{1}{2}\Phi_\varepsilon(\alpha,\alpha) - \frac{1}{2}\Phi_\varepsilon(\beta,\beta) \qquad \forall\, \alpha,\beta \in \Delta_n. \tag{7}$$

Moreover, for every $\alpha \in \Delta_n$ there exists a unique vector $u_\alpha \in \mathbb{R}_{\geq 0}^n$ with $\mathrm{supp}(u_\alpha) = \mathrm{supp}(\alpha)$ satisfying

$$u_\alpha \odot (K_\varepsilon u_\alpha) = \alpha, \qquad u_\alpha^\top K_\varepsilon u_\alpha = 1, \qquad \Phi_\varepsilon(\alpha,\alpha) = 2\varepsilon \sum_{i\in\mathrm{supp}(\alpha)} \alpha_i \log(u_\alpha)_i - \varepsilon. \tag{8}$$

Proof. For any $\pi \in \Pi(\alpha,\beta)$, expanding $\mathrm{KL}(\pi\|\alpha\otimes\beta) = \sum_{i,j}\pi_{ij}\log\pi_{ij} + H(\alpha) + H(\beta)$ yields

$$\mathrm{OT}_\varepsilon(\alpha,\beta) \;=\; \min_{\pi\in\Pi(\alpha,\beta)} \Big\{ \langle C,\pi\rangle + \varepsilon \sum_{i,j} \pi_{ij}(\log\pi_{ij} - 1) \Big\} + \varepsilon + \varepsilon H(\alpha) + \varepsilon H(\beta).$$

For any dual pair $f,g \in \mathbb{R}^n$, minimizing $\sum_{i,j}\big(C_{ij}\pi_{ij} + \varepsilon\pi_{ij}(\log\pi_{ij}-1) - f_i\pi_{ij} - g_j\pi_{ij}\big)$ over $\pi_{ij}\geq 0$ gives the unique stationarity condition $\pi_{ij} = e^{(f_i+g_j-C_{ij})/\varepsilon} = e^{f_i/\varepsilon}K_{\varepsilon,ij}e^{g_j/\varepsilon}$, yielding dual objective $\langle f,\alpha\rangle + \langle g,\beta\rangle - \Phi_\varepsilon^*(f,g)$. By Fenchel–Rockafellar strong duality on $\mathrm{supp}(\alpha)\times\mathrm{supp}(\beta)$ (noting that for $i\notin\mathrm{supp}(\alpha)$, sending $f_i\to-\infty$ drives $e^{f_i/\varepsilon}\to 0$ without affecting $\langle f,\alpha\rangle$, so the supremum over $\mathbb{R}^n\times\mathbb{R}^n$ coincides with the support-restricted dual), the minimization equals $\Phi_\varepsilon(\alpha,\beta)$, proving (6). Substituting (6) into (2), both ε and the marginal entropies $\varepsilon H(\alpha), \varepsilon H(\beta)$ cancel identically, establishing (7). Finally, for $\alpha=\beta$ with $I := \mathrm{supp}(\alpha)$, Sinkhorn's theorem on the strictly positive symmetric matrix $K_\varepsilon|_{I\times I}$ yields a unique positive scaling $(u_\alpha)_I \in \mathbb{R}^{|I|}_{>0}$ solving $(u_\alpha)_i\sum_{j\in I}K_{\varepsilon,ij}(u_\alpha)_j = \alpha_i$ for $i\in I$ (see [8, 6]). Extending by $(u_\alpha)_i := 0$ for $i\notin I$ gives $u_\alpha\odot(K_\varepsilon u_\alpha) = \alpha$ on all of $\{1,\dots,n\}$, hence $u_\alpha^\top K_\varepsilon u_\alpha = \sum_{i\in I}\alpha_i = 1$, and evaluating the dual optimum $(f_i,f_j) = (\varepsilon\log(u_\alpha)_i, \varepsilon\log(u_\alpha)_j)$ on I yields (8). □

Remark 2.2. Notice that the sign of the marginal entropy term in $\mathrm{OT}_\varepsilon(\alpha,\beta) = \Phi_\varepsilon(\alpha,\beta) + \varepsilon + \varepsilon H(\alpha) + \varepsilon H(\beta)$ is $+\varepsilon(H(\alpha)+H(\beta))$, and regardless of whether one regularizes with $\mathrm{KL}(\pi\|\alpha\otimes\beta)$ or negative Shannon entropy $\sum_{i,j}\pi_{ij}(\log\pi_{ij}-1)$, the debiased Sinkhorn divergence $S_\varepsilon(\alpha,\beta)$ is identically equal to $\Phi_\varepsilon(\alpha,\beta) - \frac{1}{2}\Phi_\varepsilon(\alpha,\alpha) - \frac{1}{2}\Phi_\varepsilon(\beta,\beta)$.

We now compute the exact second variation of S_ε along the diagonal of $\Delta_n^+\times\Delta_n^+$.

Theorem 2.3 (Exact diagonal Hessian of the Sinkhorn divergence)**.** *Let $C = C^\top \in \mathbb{R}^{n\times n}$ and $\varepsilon>0$. Fix any $\mu\in\Delta_n^+$ with symmetric Sinkhorn scaling $u_\mu\in\mathbb{R}^n_{>0}$ ($u_\mu\odot(K_\varepsilon u_\mu)=\mu$), symmetric optimal self-coupling $P_\mu := \mathrm{diag}(u_\mu)K_\varepsilon\,\mathrm{diag}(u_\mu)$, and normalized symmetric Markov operator*

$$S_\mu \;:=\; D_\mu^{-1/2}P_\mu D_\mu^{-1/2}, \qquad D_\mu := \mathrm{diag}(\mu).$$

Then:

(a) *S_μ is symmetric with simple principal eigenvalue $\lambda_1(S_\mu)=1$ (eigenvector $\sqrt{\mu}$), and all remaining eigenvalues satisfy $\lambda_2(S_\mu),\dots,\lambda_n(S_\mu)\in(-1,1)$.*

(b) *The function $(\alpha,\beta)\mapsto S_\varepsilon(\alpha,\beta)$ is C^∞ on $\Delta_n^+\times\Delta_n^+$, and for every tangent vector $\eta\in\mathbf{1}^\perp$ with $\tilde\eta := D_\mu^{-1/2}\eta \perp \sqrt{\mu}$,*

$$\frac{d^2}{dt^2}\Big|_{t=0} S_\varepsilon(\mu+t\eta,\mu-t\eta) \;=\; 4\varepsilon\,\tilde\eta^\top \frac{S_\mu}{I-S_\mu^2}\,\tilde\eta \;=\; 4\varepsilon\sum_{k=2}^{n}\frac{\lambda_k(S_\mu)}{1-\lambda_k(S_\mu)^2}\,\langle\tilde\eta,v_k\rangle^2, \tag{9}$$

where $\{v_2,\dots,v_n\}$ is an orthonormal eigenbasis of S_μ on $\sqrt{\mu}^\perp$.

Proof. Proof of (a). Since $P_\mu = P_\mu^\top$ has strictly positive entries and $P_\mu\mathbf{1}=\mu$, the matrix $S_\mu = D_\mu^{-1/2}P_\mu D_\mu^{-1/2}$ is symmetric, entrywise strictly positive, and satisfies $S_\mu\sqrt{\mu}=\sqrt{\mu}$. By the Perron–Frobenius theorem for strictly positive matrices, $\lambda_1(S_\mu)=1$ is a simple eigenvalue and every other eigenvalue satisfies $|\lambda_k(S_\mu)|<1$ (strict inequality <1 holds in particular against -1 because every entry of S_μ is strictly positive, so S_μ is primitive).

Proof of (b). Fix the $(2n-1)$-dimensional gauge hyperplane $\mathcal{G} := \{(f,g)\in\mathbb{R}^n\times\mathbb{R}^n : \langle f-g,\mathbf{1}\rangle = 0\}$. For any $(\alpha,\beta)\in\Delta_n^+\times\Delta_n^+$, the dual objective $\langle f,\alpha\rangle + \langle g,\beta\rangle - \Phi_\varepsilon^*(f,g)$ is invariant under $(f,g)\mapsto(f+c\mathbf{1},g-c\mathbf{1})$ and attains its unique maximizer $(f(\alpha,\beta),g(\alpha,\beta))\in\mathcal{G}$ as the unique solution in $\mathcal{G}$ of the first-order optimality system

$$\nabla\Phi_\varepsilon^*(f,g) \;=\; \Big(e^{f/\varepsilon}\odot(K_\varepsilon e^{g/\varepsilon}),\; e^{g/\varepsilon}\odot(K_\varepsilon e^{f/\varepsilon})\Big) \;=\; (\alpha,\beta). \tag{10}$$

At the symmetric point (μ,μ), the unique solution is (f_0,f_0) with $e^{f_0/\varepsilon}=u_\mu$. Differentiating $\nabla\Phi^*_\varepsilon$ at (f_0,f_0) gives the $2n\times 2n$ block Hessian

$$H_0 \;:=\; \nabla^2\Phi^*_\varepsilon(f_0,f_0) \;=\; \frac{1}{\varepsilon}\begin{pmatrix} D_\mu & P_\mu \\ P_\mu & D_\mu\end{pmatrix}.$$

In the orthogonal decomposition into symmetric vectors (x,x) and antisymmetric vectors $(y,-y)$,

$$H_0\begin{pmatrix}x\\x\end{pmatrix}=\frac{1}{\varepsilon}\begin{pmatrix}(D_\mu+P_\mu)x\\(D_\mu+P_\mu)x\end{pmatrix},\qquad H_0\begin{pmatrix}y\\-y\end{pmatrix}=\frac{1}{\varepsilon}\begin{pmatrix}(D_\mu-P_\mu)y\\-(D_\mu-P_\mu)y\end{pmatrix}.$$

By Part (a), $D_\mu+P_\mu=D_\mu^{1/2}(I+S_\mu)D_\mu^{1/2}\succ 0$ on all of $\mathbb{R}^n$, while $D_\mu-P_\mu=D_\mu^{1/2}(I-S_\mu)D_\mu^{1/2}\succeq 0$ has exact one-dimensional kernel $\operatorname{span}(\mathbf{1})$ and is strictly positive definite on $\mathbf{1}^\perp$. Hence $\ker(H_0)=\operatorname{span}\{(\mathbf{1},-\mathbf{1})\}=\mathcal{G}^\perp$, and $H_0|_{\mathcal{G}\times\mathcal{G}}\succ 0$ is strictly positive definite. More generally, at any $(\alpha,\beta)\in\Delta_n^+\times\Delta_n^+$ with optimal coupling $P_{ij}=e^{f_i/\varepsilon}K_{\varepsilon,ij}e^{g_j/\varepsilon}>0$, the Hessian quadratic form is

$$\begin{pmatrix}u\\v\end{pmatrix}^\top\nabla^2\Phi^*_\varepsilon(f,g)\begin{pmatrix}u\\v\end{pmatrix}\;=\;\frac{1}{\varepsilon}\sum_{i,j=1}^n P_{ij}(u_i+v_j)^2\;\geq\;0,$$

which vanishes if and only if $u_i+v_j=0$ for all i,j, i.e., $(u,v)\in\operatorname{span}\{(\mathbf{1},-\mathbf{1})\}=\mathcal{G}^\perp$; thus $\nabla^2\Phi^*_\varepsilon(f,g)|_{\mathcal{G}\times\mathcal{G}}\succ 0$ at every $(\alpha,\beta)\in\Delta_n^+\times\Delta_n^+$. Notice also that $\langle\nabla\Phi^*_\varepsilon(f,g),(\mathbf{1},-\mathbf{1})\rangle=e^{f/\varepsilon\top}K_\varepsilon e^{g/\varepsilon}-e^{g/\varepsilon\top}K_\varepsilon e^{f/\varepsilon}=0$ identically, so $\nabla\Phi^*_\varepsilon$ maps $\mathcal{G}$ into the hyperplane $\{(a,b):\langle a-b,\mathbf{1}\rangle=0\}$, which contains $\Delta_n^+\times\Delta_n^+$. By the implicit function theorem, $(f(\alpha,\beta),g(\alpha,\beta))$ is C^∞ on all of $\Delta_n^+\times\Delta_n^+$, with tangent derivative at (μ,μ) given on $\mathbf{1}^\perp\times\mathbf{1}^\perp\subset\mathcal{G}$ by H_0^+ (the inverse of $H_0|_{\mathcal{G}\times\mathcal{G}}$).

By the envelope theorem, the gradient of Φ_ε along any smooth curve $(\alpha(t),\beta(t))\in\Delta_n^+\times\Delta_n^+$ is $\frac{d}{dt}\Phi_\varepsilon(\alpha(t),\beta(t))=\langle f(\alpha(t),\beta(t)),\dot\alpha(t)\rangle+\langle g(\alpha(t),\beta(t)),\dot\beta(t)\rangle$. Differentiating once more at $t=0$ for linear curves $(\mu+t\delta\alpha,\mu+t\delta\beta)$ with $\delta\alpha,\delta\beta\in\mathbf{1}^\perp$ gives

$$\left.\frac{d^2}{dt^2}\right|_{t=0}\Phi_\varepsilon(\mu+t\delta\alpha,\mu+t\delta\beta)=\begin{pmatrix}\delta\alpha\\\delta\beta\end{pmatrix}^\top H_0^+\begin{pmatrix}\delta\alpha\\\delta\beta\end{pmatrix}.$$

To evaluate the antisymmetric direction $(\eta,-\eta)$ without assuming μ is uniform, let $x\in\mathbf{1}^\perp$ be the unique vector solving $(D_\mu-P_\mu)x=\eta$, so that $(D_\mu-P_\mu)^+\eta=x$ and $\eta^\top(D_\mu-P_\mu)^+\eta=\eta^\top x$. Choose $c\in\mathbb{R}$ such that $y:=D_\mu^{1/2}(x+c\mathbf{1})$ satisfies $\langle y,\sqrt{\mu}\rangle=\langle x+c\mathbf{1},\mu\rangle=0$ (explicitly $c=-\langle x,\mu\rangle$). Since $(D_\mu-P_\mu)\mathbf{1}=0$ and $\eta\in\mathbf{1}^\perp$, we have

$$(I-S_\mu)y\;=\;D_\mu^{-1/2}(D_\mu-P_\mu)(x+c\mathbf{1})\;=\;D_\mu^{-1/2}\eta\;=\;\tilde\eta$$

with $y,\tilde\eta\in\sqrt{\mu}^\perp$, whence $y=(I-S_\mu)|^{-1}_{\sqrt{\mu}^\perp}\tilde\eta$ and $\eta^\top(D_\mu-P_\mu)^+\eta=\eta^\top(x+c\mathbf{1})=\tilde\eta^\top y=\tilde\eta^\top(I-S_\mu)|^{-1}_{\sqrt{\mu}^\perp}\tilde\eta$. Together with the symmetric directions (η,η) and $(-\eta,-\eta)$ (where $D_\mu+P_\mu$ is invertible on all of $\mathbb{R}^n$), this gives:

$$\left.\frac{d^2}{dt^2}\right|_{t=0}\Phi_\varepsilon(\mu+t\eta,\mu-t\eta)=2\varepsilon\,\eta^\top(D_\mu-P_\mu)^+\eta\;=\;2\varepsilon\,\tilde\eta^\top(I-S_\mu)|^{-1}_{\sqrt{\mu}^\perp}\tilde\eta,$$
$$\left.\frac{d^2}{dt^2}\right|_{t=0}\Phi_\varepsilon(\mu\pm t\eta,\mu\pm t\eta)=2\varepsilon\,\eta^\top(D_\mu+P_\mu)^{-1}\eta\;=\;2\varepsilon\,\tilde\eta^\top(I+S_\mu)^{-1}\tilde\eta.$$

Substituting into (7), where $\frac12\Phi_\varepsilon(\mu+t\eta,\mu+t\eta)+\frac12\Phi_\varepsilon(\mu-t\eta,\mu-t\eta)$ has second derivative $2\varepsilon\,\tilde\eta^\top(I+S_\mu)^{-1}\tilde\eta$, yields

$$\left.\frac{d^2}{dt^2}\right|_{t=0}S_\varepsilon(\mu+t\eta,\mu-t\eta)=2\varepsilon\,\tilde\eta^\top\Big[(I-S_\mu)|^{-1}_{\sqrt{\mu}^\perp}-(I+S_\mu)^{-1}\Big]\tilde\eta=4\varepsilon\,\tilde\eta^\top\frac{S_\mu}{I-S_\mu^2}\,\tilde\eta.\qquad\square$$

3. The Exact Schoenberg–Feydy Positivity Trichotomy and Flat Zero-Loci

Theorem 2.3 unlocks the complete converse to Feydy et al. [2] together with an exact algebraic characterization of the zero-locus on the singular boundary $\lambda_{\min}(K_\varepsilon) = 0$.

Theorem 3.1 (Exact Schoenberg–Feydy Positivity Trichotomy)**.** *Let $C = C^\top \in \mathbb{R}^{n\times n}$ ($n \geq 2$) be any symmetric cost matrix, fix $\varepsilon > 0$, and let $K_\varepsilon = (e^{-C_{ij}/\varepsilon})_{i,j=1}^n$. For each $\alpha \in \Delta_n$, let $u_\alpha \in \mathbb{R}^n_{\geq 0}$ be the unique symmetric Sinkhorn scaling from Lemma 2.1. Then:*

(a) Universal RKHS lower bound. *For all $\alpha, \beta \in \Delta_n$ (regardless of whether K_ε is positive semidefinite),*

$$S_\varepsilon(\alpha, \beta) \;\geq\; \frac{\varepsilon}{2}\,(u_\alpha - u_\beta)^\top K_\varepsilon (u_\alpha - u_\beta). \tag{11}$$

(b) Strict positive definiteness. *$S_\varepsilon(\alpha, \beta) > 0$ for all distinct $\alpha \neq \beta \in \Delta_n$ if and only if $K_\varepsilon \succ 0$.*

(c) Degenerate non-negativity and exact zero-locus. *$S_\varepsilon(\alpha, \beta) \geq 0$ for all $\alpha, \beta \in \Delta_n$ while $S_\varepsilon(\alpha, \beta) = 0$ for some $\alpha \neq \beta$ if and only if $K_\varepsilon \succeq 0$ and $\ker(K_\varepsilon) \neq \{0\}$. Moreover, whenever $K_\varepsilon \succeq 0$, the exact zero-locus of S_ε on $\Delta_n \times \Delta_n$ is*

$$\big\{(\alpha, \beta) \in \Delta_n \times \Delta_n : S_\varepsilon(\alpha, \beta) = 0\big\} \;=\; \big\{(\alpha, \beta) \in \Delta_n \times \Delta_n : u_\alpha - u_\beta \in \ker(K_\varepsilon)\big\}. \tag{12}$$

In particular, for every $u \in \mathbb{R}^n_{>0}$ and every $w \in \ker(K_\varepsilon) \setminus \{0\}$, letting $Z := u^\top K_\varepsilon u > 0$, the flat affine family

$$\alpha_t := \frac{(u + tw) \odot (K_\varepsilon u)}{Z}, \qquad \beta_t := \frac{(u - tw) \odot (K_\varepsilon u)}{Z} \qquad (u \pm tw \in \mathbb{R}^n_{>0},\ \text{extending to } \mathbb{R}^n_{\geq 0}) \tag{13}$$

traces a non-empty open 1-dimensional affine segment in $\Delta_n^+ \times \Delta_n^+$ (extending continuously to the closed boundary endpoints where $u \pm tw \in \mathbb{R}^n_{\geq 0}$) along which $\alpha_t \neq \beta_t$ ($t \neq 0$) and $S_\varepsilon(\alpha_t, \beta_t) = 0$ identically. Moreover, decomposing $w = w_+ - w_-$ into its non-negative disjoint-support parts ($w_+, w_- \in \mathbb{R}^n_{\geq 0} \setminus \{0\}$) and setting $Z_w := w_+^\top K_\varepsilon w_+ = w_-^\top K_\varepsilon w_- > 0$, the disjoint-support boundary measures

$$\alpha_w := \frac{w_+ \odot (K_\varepsilon w_+)}{Z_w}, \qquad \beta_w := \frac{w_- \odot (K_\varepsilon w_-)}{Z_w}$$

satisfy $\operatorname{supp}(\alpha_w) \cap \operatorname{supp}(\beta_w) = \emptyset$ and $S_\varepsilon(\alpha_w, \beta_w) = 0$.

(d) Failure of non-negativity. *There exist $\alpha, \beta \in \Delta_n$ with $S_\varepsilon(\alpha, \beta) < 0$ if and only if $\lambda_{\min}(K_\varepsilon) < 0$. Furthermore, whenever $\lambda_{\min}(K_\varepsilon) < 0$, for* every *strictly positive base measure $\mu \in \Delta_n^+$ there exists a tangent vector $\eta \in \mathbf{1}^\perp$ such that $\frac{d^2}{dt^2}\Big|_{t=0} S_\varepsilon(\mu + t\eta, \mu - t\eta) < 0$, so $S_\varepsilon(\mu + t\eta, \mu - t\eta) < 0$ for all sufficiently small $t > 0$.*

Proof. Proof of (a). Fix any $\alpha, \beta \in \Delta_n$ with supports $I = \operatorname{supp}(\alpha)$ and $J = \operatorname{supp}(\beta)$. For any $R > 0$, define $f_\alpha^{(R)}, f_\beta^{(R)} \in \mathbb{R}^n$ by $(f_\alpha^{(R)})_i := \varepsilon \log(u_\alpha)_i$ for $i \in I$ and $-R$ for $i \notin I$, and $(f_\beta^{(R)})_j := \varepsilon \log(u_\beta)_j$ for $j \in J$ and $-R$ for $j \notin J$. Inserting the feasible pair $(f_\alpha^{(R)}, f_\beta^{(R)})$ into (5) gives

$$\Phi_\varepsilon(\alpha, \beta) \;\geq\; \langle f_\alpha^{(R)}, \alpha\rangle + \langle f_\beta^{(R)}, \beta\rangle - \varepsilon \left\langle e^{f_\alpha^{(R)}/\varepsilon},\, K_\varepsilon e^{f_\beta^{(R)}/\varepsilon}\right\rangle.$$

Note that $\langle f_\alpha^{(R)}, \alpha\rangle = \varepsilon \sum_{i\in I} \alpha_i \log(u_\alpha)_i$ is independent of R, and similarly for β, while $e^{f_\alpha^{(R)}/\varepsilon} \to u_\alpha$ and $e^{f_\beta^{(R)}/\varepsilon} \to u_\beta$ as $R \to \infty$. Sending $R \to \infty$ and subtracting $\frac{1}{2}\Phi_\varepsilon(\alpha, \alpha) + \frac{1}{2}\Phi_\varepsilon(\beta, \beta) = \varepsilon \sum_{i\in I} \alpha_i \log(u_\alpha)_i + \varepsilon \sum_{j\in J} \beta_j \log(u_\beta)_j - \varepsilon$ (from (8)) yields

$$S_\varepsilon(\alpha, \beta) \;\geq\; \varepsilon\big(1 - u_\alpha^\top K_\varepsilon u_\beta\big) \;=\; \frac{\varepsilon}{2}\,(u_\alpha - u_\beta)^\top K_\varepsilon (u_\alpha - u_\beta),$$

where the final equality uses $u_\alpha^\top K_\varepsilon u_\alpha = u_\beta^\top K_\varepsilon u_\beta = 1$.

Proof of sufficiency in (b) and (c). Suppose $K_\varepsilon \succeq 0$. Then (11) immediately gives $S_\varepsilon(\alpha,\beta) \geq 0$ for all $\alpha,\beta \in \Delta_n$. If $K_\varepsilon \succ 0$ and $\alpha \neq \beta$, then $u_\alpha \neq u_\beta$ (since $\alpha = u_\alpha \odot (K_\varepsilon u_\alpha)$ determines α uniquely from u_α), so (11) gives $S_\varepsilon(\alpha,\beta) \geq \frac{\varepsilon}{2}\lambda_{\min}(K_\varepsilon)\|u_\alpha - u_\beta\|_2^2 > 0$.

Now suppose $K_\varepsilon \succeq 0$ and let $\alpha,\beta \in \Delta_n$. If $S_\varepsilon(\alpha,\beta) = 0$, then (11) gives $0 \geq \frac{\varepsilon}{2}(u_\alpha - u_\beta)^\top K_\varepsilon(u_\alpha - u_\beta) \geq 0$, which forces $u_\alpha - u_\beta \in \ker(K_\varepsilon)$ because $K_\varepsilon \succeq 0$. Conversely, suppose $u_\alpha - u_\beta \in \ker(K_\varepsilon)$, so that $K_\varepsilon u_\alpha = K_\varepsilon u_\beta$. Define the non-negative coupling $\pi^* := \operatorname{diag}(u_\alpha) K_\varepsilon \operatorname{diag}(u_\beta) \in \mathbb{R}^{n\times n}_{\geq 0}$. Its marginals are

$$\pi^*\mathbf{1} = u_\alpha \odot (K_\varepsilon u_\beta) = u_\alpha \odot (K_\varepsilon u_\alpha) = \alpha, \qquad (\pi^*)^\top \mathbf{1} = u_\beta \odot (K_\varepsilon u_\alpha) = u_\beta \odot (K_\varepsilon u_\beta) = \beta,$$

so $\pi^* \in \Pi(\alpha,\beta)$. On $\operatorname{supp}(\alpha)\times\operatorname{supp}(\beta)$, $\pi^*_{ij} = (u_\alpha)_i K_{\varepsilon,ij}(u_\beta)_j = \exp((f_{\alpha,i} + f_{\beta,j} - C_{ij})/\varepsilon)$ (where $f_{\alpha,i} := \varepsilon\log(u_\alpha)_i$, $f_{\beta,j} := \varepsilon\log(u_\beta)_j$), so evaluating the primal functional gives

$$\langle C, \pi^*\rangle + \varepsilon \sum_{i,j} \pi^*_{ij}(\log \pi^*_{ij} - 1) = \sum_{i\in\operatorname{supp}(\alpha)} \alpha_i f_{\alpha,i} + \sum_{j\in\operatorname{supp}(\beta)} \beta_j f_{\beta,j} - \varepsilon.$$

Hence $\Phi_\varepsilon(\alpha,\beta) \leq \sum_{i\in\operatorname{supp}(\alpha)} \alpha_i f_{\alpha,i} + \sum_{j\in\operatorname{supp}(\beta)} \beta_j f_{\beta,j} - \varepsilon = \frac{1}{2}\Phi_\varepsilon(\alpha,\alpha) + \frac{1}{2}\Phi_\varepsilon(\beta,\beta)$, so $S_\varepsilon(\alpha,\beta) \leq 0$. Combining with $S_\varepsilon(\alpha,\beta) \geq 0$ proves $S_\varepsilon(\alpha,\beta) = 0$ and establishes (12).

For the explicit family (13): if $w \in \ker(K_\varepsilon)\setminus\{0\}$ and $u \in \mathbb{R}^n_{>0}$, then $K_\varepsilon(u \pm tw) = K_\varepsilon u \in \mathbb{R}^n_{>0}$, so $u_{\alpha_t} := (u+tw)/\sqrt{Z} \in \mathbb{R}^n_{\geq 0}$ satisfies $u_{\alpha_t} \odot (K_\varepsilon u_{\alpha_t}) = \alpha_t$ and $\sum_i (\alpha_t)_i = (u+tw)^\top K_\varepsilon u/Z = u^\top K_\varepsilon u/Z = 1$ (since $w^\top K_\varepsilon u = 0$). Likewise $u_{\beta_t} = (u-tw)/\sqrt{Z}$ is the symmetric scaling of β_t, and $u_{\alpha_t} - u_{\beta_t} = (2t/\sqrt{Z})w \in \ker(K_\varepsilon)$. For $t \neq 0$, $\alpha_t - \beta_t = (2t/Z)(w \odot K_\varepsilon u) \neq 0$ because $K_\varepsilon u > 0$ entrywise, so $\alpha_t \neq \beta_t$ while $S_\varepsilon(\alpha_t,\beta_t) = 0$. For the disjoint-support pair (α_w,β_w): since $K_\varepsilon > 0$ entrywise and $K_\varepsilon w = 0$ with $w \neq 0$, w must have both strictly positive and strictly negative entries, so $w_+, w_- \in \mathbb{R}^n_{\geq 0}\setminus\{0\}$ have disjoint non-empty supports and $K_\varepsilon w_+ = K_\varepsilon w_- \in \mathbb{R}^n_{>0}$. Hence $Z_w := w_+^\top K_\varepsilon w_+ = w_+^\top K_\varepsilon w_- = w_-^\top K_\varepsilon w_- > 0$, $\alpha_w,\beta_w \in \Delta_n$ have disjoint supports $\operatorname{supp}(w_+)$ and $\operatorname{supp}(w_-)$ with unique symmetric scalings $u_{\alpha_w} = w_+/\sqrt{Z_w}$ and $u_{\beta_w} = w_-/\sqrt{Z_w}$, and $u_{\alpha_w} - u_{\beta_w} = w/\sqrt{Z_w} \in \ker(K_\varepsilon)$, giving $S_\varepsilon(\alpha_w,\beta_w) = 0$ by (12).

Proof of (d) (the converse via Sylvester's law of inertia). If $\lambda_{\min}(K_\varepsilon) \geq 0$, then $S_\varepsilon \geq 0$ by (a), proving the "only if" direction. Conversely, suppose $\lambda_{\min}(K_\varepsilon) < 0$, and let $\mu \in \Delta_n^+$ be *any* strictly positive probability distribution (for instance $\mu = \frac{1}{n}\mathbf{1}$). Let $u_\mu \in \mathbb{R}^n_{>0}$ be its unique symmetric Sinkhorn scaling from Lemma 2.1, and let $S_\mu := D_\mu^{-1/2}\operatorname{diag}(u_\mu) K_\varepsilon \operatorname{diag}(u_\mu) D_\mu^{-1/2}$ be the normalized symmetric Sinkhorn operator from Theorem 2.3. Because $M_\mu := D_\mu^{-1/2}\operatorname{diag}(u_\mu)$ is a real invertible diagonal matrix, $S_\mu = M_\mu K_\varepsilon M_\mu^\top$ is congruent to K_ε. By Sylvester's law of inertia, S_μ and K_ε have the exact same number of negative eigenvalues; hence S_μ possesses at least one strictly negative eigenvalue $\lambda < 0$. By Theorem 2.3(a), $\lambda \in (-1,0)$ and the symmetric matrix S_μ has principal eigenvalue $+1 > 0$ at $\sqrt{\mu}$; consequently, any unit eigenvector $v \in \mathbb{R}^n$ of S_μ with eigenvalue $\lambda < 0$ is automatically orthogonal to $\sqrt{\mu}$. Define $\eta := D_\mu^{1/2} v$. Then $\langle \eta, \mathbf{1}\rangle = \langle D_\mu^{1/2} v, \mathbf{1}\rangle = \langle v, \sqrt{\mu}\rangle = 0$ (so $\eta \in \mathbf{1}^\perp$) and $\tilde{\eta} := D_\mu^{-1/2}\eta = v$. Applying Theorem 2.3(b) yields

$$\left.\frac{d^2}{dt^2}\right|_{t=0} S_\varepsilon(\mu + t\eta, \mu - t\eta) \;=\; 4\varepsilon\,\frac{\lambda}{1-\lambda^2} \;<\; 0.$$

Since $S_\varepsilon(\mu,\mu) = 0$ and $t \mapsto S_\varepsilon(\mu+t\eta, \mu-t\eta)$ is even and smooth near $t = 0$, Taylor's theorem gives $S_\varepsilon(\mu+t\eta,\mu-t\eta) = 2\varepsilon\frac{\lambda}{1-\lambda^2}t^2 + O(t^4) < 0$ for all sufficiently small $t > 0$. Since (b), (c), and (d) partition the sign of $\lambda_{\min}(K_\varepsilon)$, this also establishes the necessity directions of (b) and (c). □

4. Schoenberg Characterization, Graph Thresholds, and Continuum Collapse

Combining Theorem 3.1 with Schoenberg's classical theorem [7] immediately yields the complete characterization of universal positivity across all $\varepsilon > 0$ and exact critical thresholds for canonical non-Euclidean spaces.

Definition 4.1 (Conditionally negative definite cost and critical regularization)**.** A symmetric matrix $C = C^\top \in \mathbb{R}^{n\times n}$ is *conditionally negative definite (CND)* if $v^\top C v \le 0$ for all $v \in \mathbf{1}^\perp$. For a symmetric cost C, define its *critical Sinkhorn regularization* by

$$\varepsilon_c(C) \;:=\; \inf\{\varepsilon > 0 : \lambda_{\min}(K_\varepsilon) < 0\} \;=\; \inf\Big\{\varepsilon > 0 : \inf_{\alpha,\beta\in\Delta_n} S_\varepsilon(\alpha,\beta) < 0\Big\},$$

with $\inf \emptyset := +\infty$.

Theorem 4.2 (Schoenberg universal positivity and complete bipartite graph thresholds)**.** (a) Universal positivity equivalence. *For any symmetric $C = C^\top \in \mathbb{R}^{n\times n}$,*

$$S_\varepsilon(\alpha,\beta) \ge 0 \;\; \forall\, \varepsilon > 0,\; \forall\, \alpha,\beta \in \Delta_n \quad\Longleftrightarrow\quad C \text{ is conditionally negative definite.}$$

(b) Principal submatrix monotonicity. *If $X' \subseteq X$ with restricted cost $C' = C|_{X'\times X'}$, then $\varepsilon_c(C) \le \varepsilon_c(C')$.*

(c) Exact complete bipartite graph threshold. *Let $G = K_{m,n}$ $(m, n \ge 2)$ with shortest-path graph metric d_G. Then:*

- *For $K_{2,2} = C_4$, d_G is CND and $\varepsilon_c(K_{2,2}) = +\infty$ ($S_\varepsilon > 0$ off-diagonal for all $\varepsilon > 0$).*
- *For every (m, n) with $(m-1)(n-1) \ge 2$, S_ε is strictly positive definite for all $\varepsilon \in (0, \varepsilon_c)$, has a flat zero-locus containing the disjoint uniform bipartition measures (μ_A, μ_B) at $\varepsilon = \varepsilon_c$, and satisfies $S_\varepsilon(\mu_A, \mu_B) < 0$ for all $\varepsilon > \varepsilon_c$, where*

$$\varepsilon_c(K_{m,n}) \;=\; \frac{2}{\log\big((m-1)(n-1)\big)}. \tag{14}$$

In particular, $\varepsilon_c(K_{2,3}) = 2/\log 2 = 2.8853900818\ldots$, and for any connected finite graph G containing $K_{m,n}$ (with $(m-1)(n-1) \ge 2$) as an isometric subgraph, $\varepsilon_c(G) \le 2/\log((m-1)(n-1))$.

Proof. Proof of (a). By Theorem 3.1, $S_\varepsilon \ge 0$ on $\Delta_n \times \Delta_n$ for all $\varepsilon > 0$ if and only if $K_\varepsilon = (e^{-C_{ij}/\varepsilon}) \succeq 0$ for all $\varepsilon > 0$. By Schoenberg's theorem [7], this holds if and only if C is CND (for completeness: if $K_\varepsilon \succeq 0$ for all $\varepsilon > 0$ and $v \in \mathbf{1}^\perp$, then $0 \le \lim_{\varepsilon\to\infty} \varepsilon\, v^\top K_\varepsilon v = -v^\top C v$; conversely, if C is CND then K_ε is the entrywise exponential of a conditionally positive definite matrix, hence PSD by the Schur product theorem).

Proof of (b). If $\lambda_{\min}(K_\varepsilon(C')) < 0$, then by Cauchy interlacing $\lambda_{\min}(K_\varepsilon(C)) \le \lambda_{\min}(K_\varepsilon(C')) < 0$, so $\varepsilon_c(C) \le \varepsilon_c(C')$.

Proof of (c). On $K_{m,n}$ with bipartition $A \sqcup B$ ($|A| = m$, $|B| = n$), distances are $d(i,i) = 0$, $d(i,j) = 2$ within A or within B ($i \ne j$), and $d(i,j) = 1$ across $A \times B$. Setting $v := e^{-1/\varepsilon} \in (0,1)$ and $u := v^2 = e^{-2/\varepsilon}$, the Gibbs matrix has block structure

$$K_\varepsilon \;=\; \begin{pmatrix} (1-u)I_m + uJ_{m,m} & vJ_{m,n} \\ vJ_{n,m} & (1-u)I_n + uJ_{n,n} \end{pmatrix}.$$

On the orthogonal complements $\mathbf{1}_A^\perp$ (dimension $m-1$) and $\mathbf{1}_B^\perp$ (dimension $n-1$), K_ε acts as the positive scalar $(1-u) > 0$. On the invariant 2-dimensional subspace $\operatorname{span}\{\mathbf{1}_A/\sqrt{m},\, \mathbf{1}_B/\sqrt{n}\}$, K_ε reduces to the 2×2 matrix

$$M_2(v) \;=\; \begin{pmatrix} 1 + (m-1)v^2 & \sqrt{mn}\, v \\ \sqrt{mn}\, v & 1 + (n-1)v^2 \end{pmatrix}.$$

Its trace is strictly positive and its determinant factors identically as

$$\det M_2(v) \;=\; \big(1+(m-1)v^2\big)\big(1+(n-1)v^2\big) - mnv^2 \;=\; \big(1-v^2\big)\big(1-(m-1)(n-1)v^2\big).$$

Since $1-v^2>0$ for all $\varepsilon>0$, we have $K_\varepsilon \succ 0 \iff (m-1)(n-1)e^{-2/\varepsilon}<1$, $K_\varepsilon \succeq 0$ singular $\iff (m-1)(n-1)e^{-2/\varepsilon}=1$, and $\lambda_{\min}(K_\varepsilon)<0 \iff (m-1)(n-1)e^{-2/\varepsilon}>1$. For $(m,n)=(2,2)$, $(m-1)(n-1)=1$, so $1-v^2>0$ for all $\varepsilon>0$. For $(m-1)(n-1)\geq 2$, the unique sign change occurs at $\varepsilon_c(K_{m,n}) = 2/\log((m-1)(n-1))$. Finally, for the uniform bipartition measures $\mu_A = \frac{1}{m}\mathbf{1}_A$ and $\mu_B = \frac{1}{n}\mathbf{1}_B$, symmetry gives exact closed forms $\mathrm{OT}_\varepsilon(\mu_A,\mu_B)=1$, $\mathrm{OT}_\varepsilon(\mu_A,\mu_A) = -\varepsilon\log\frac{1+(m-1)u}{m}$, and $\mathrm{OT}_\varepsilon(\mu_B,\mu_B) = -\varepsilon\log\frac{1+(n-1)u}{n}$, whence

$$\begin{aligned} S_\varepsilon(\mu_A,\mu_B) &= 1+\frac{\varepsilon}{2}\log\!\left(\frac{(1+(m-1)u)(1+(n-1)u)}{mn}\right) \\ &= \frac{\varepsilon}{2}\log\!\left(\frac{(1+(m-1)u)(1+(n-1)u)}{mnu}\right). \end{aligned}$$

Since $(1+(m-1)u)(1+(n-1)u) - mnu = (1-u)(1-(m-1)(n-1)u)$, $S_\varepsilon(\mu_A,\mu_B)$ has the exact sign of $1-(m-1)(n-1)u$, vanishing at $\varepsilon=\varepsilon_c$ and strictly negative for all $\varepsilon>\varepsilon_c$. Applying (b) yields the isometric subgraph bound. □

We now analyze the super-quadratic Euclidean power cost $c_p(x,y)=|x-y|^p$ $(p>2)$, which is widely used in numerical optimal transport despite failing Schoenberg's CND criterion.

Theorem 4.3 (Super-quadratic $|x-y|^p$ phase curve and continuum collapse)**.** *Fix $p>2$.*

(a) Exact 3-point threshold. *On the scaled 3-point space $\{0,h,2h\}\subset\mathbb{R}$ $(h>0)$ with cost $c_p(x,y)=|x-y|^p$, the Sinkhorn divergence satisfies $S_\varepsilon>0$ off-diagonal for $\varepsilon<\varepsilon_c(p,h)$, $S_\varepsilon\geq 0$ with singular zero-locus at $\varepsilon=\varepsilon_c(p,h)$, and $\inf S_\varepsilon<0$ for $\varepsilon>\varepsilon_c(p,h)$, where*

$$\varepsilon_c(p,h) \;=\; \frac{h^p}{\beta_*(p)}, \tag{15}$$

and $\beta_*(p)>0$ is the unique positive root of $1+e^{-2^p\beta}=2e^{-2\beta}$.

(b) Exact two-term near-quadratic and large-p asymptotics. *Setting $\varepsilon_c(p):=\varepsilon_c(p,1)$,*

$$\varepsilon_c(p) \;=\; \frac{1}{(p-2)\log 2}+\frac{3}{2}+O(p-2) \qquad (p\to 2^+), \tag{16}$$

$$\varepsilon_c(p) \;=\; \frac{2}{\log 2}+\frac{2}{(\log 2)^2}\,2^{-2^{p-1}}\big(1+o(1)\big) \qquad (p\to\infty). \tag{17}$$

(c) Continuum collapse. *Let $X\subset\mathbb{R}^d$ be any compact set containing a non-degenerate line segment of length $L>0$, equipped with cost $c_p(x,y):=\|x-y\|_2^p$ $(p>2)$, where for Borel probability measures $\alpha,\beta\in\mathcal{P}(X)$ we define $\mathrm{OT}_\varepsilon^X(\alpha,\beta):=\inf_{\pi\in\Pi(\alpha,\beta)}\big\{\int_{X\times X}c_p\,d\pi+\varepsilon\,\mathrm{KL}(\pi\|\alpha\otimes\beta)\big\}$ and $S_\varepsilon^X(\alpha,\beta):=\mathrm{OT}_\varepsilon^X(\alpha,\beta)-\frac12\mathrm{OT}_\varepsilon^X(\alpha,\alpha)-\frac12\mathrm{OT}_\varepsilon^X(\beta,\beta)$. Then for* every *$\varepsilon>0$ (no matter how small), there exist finitely supported probability measures $\alpha,\beta\in\mathcal{P}(X)$ such that $S_\varepsilon^X(\alpha,\beta)<0$—even though $\lim_{\varepsilon\to0^+}S_\varepsilon^X(\alpha,\beta)=W_p^p(\alpha,\beta)\geq 0$ pointwise for every fixed pair (α,β).*

Proof. Proof of (a). By homogeneity $c_p(hx,hy)=h^pc_p(x,y)$, $\varepsilon_c(p,h)=h^p\varepsilon_c(p,1)$, so take $h=1$. With $\beta:=1/\varepsilon>0$ and $a:=2^p>4$, the 3×3 Gibbs matrix is

$$K_{1/\beta} \;=\; \begin{pmatrix} 1 & e^{-\beta} & e^{-a\beta} \\ e^{-\beta} & 1 & e^{-\beta} \\ e^{-a\beta} & e^{-\beta} & 1 \end{pmatrix}.$$

Its eigenvalues are $1-e^{-a\beta}>0$ (eigenvector $(1,0,-1)^\top$) and the two eigenvalues of the 2×2 block $\left(\begin{smallmatrix}1+e^{-a\beta} & \sqrt{2}e^{-\beta}\\ \sqrt{2}e^{-\beta} & 1\end{smallmatrix}\right)$, whose trace is positive and whose determinant is $g(\beta):=1+e^{-a\beta}-2e^{-2\beta}$. Observe that $e^{2\beta}g(\beta)=e^{2\beta}+e^{-(a-2)\beta}-2$ is strictly convex on $[0,\infty)$ (its second derivative is $4e^{2\beta}+(a-2)^2e^{-(a-2)\beta}>0$), vanishes at $\beta=0$, has derivative $2-(a-2)=4-a<0$ at $\beta=0$, and tends to $+\infty$ as $\beta\to\infty$. Therefore $g(\beta)$ has a unique positive root $\beta_*(p)>0$, with $g(\beta)<0$ on $(0,\beta_*(p))$ and $g(\beta)>0$ on $(\beta_*(p),\infty)$. Theorem 3.1 yields (a).

Proof of (b). For $p\to 2^+$, set $h:=(p-2)\log 2\to 0^+$ and $\delta:=a-4=4(e^h-1)=4h+2h^2+O(h^3)$. Expanding $g(\beta)=1+e^{-(4+\delta)\beta}-2e^{-2\beta}$ at $\beta=\delta/2$ gives

$$g(\delta/2)\;=\;-\delta(\delta/2)+\left(\frac{(4+\delta)^2}{2}-4\right)\frac{\delta^2}{4}+O(\delta^3)\;=\;\frac{1}{2}\delta^2+O(\delta^3)\;>\;0$$

for all sufficiently small $\delta>0$. Because $g(0)=0$, $g(\beta)<0$ for sufficiently small $\beta>0$, and g has a unique positive root β_*, it follows that $0<\beta_*<\delta/2$, so $\beta_*=O(\delta)$. Expanding $g(\beta_*)=0$ to third order and dividing by $\beta_*>0$ gives

$$\delta\;=\;\left(\frac{(4+\delta)^2}{2}-4\right)\beta_*-\left(\frac{(4+\delta)^3}{6}-\frac{8}{3}\right)\beta_*^2+O(\beta_*^3)\;=\;(4+4\delta)\beta_*-8\beta_*^2+O(\delta^3).$$

Iterating once from $\beta_*=\frac{\delta}{4}+O(\delta^2)$ yields $4\beta_*=\delta-4\delta(\frac{\delta}{4})+8(\frac{\delta}{4})^2+O(\delta^3)=\delta-\frac{1}{2}\delta^2+O(\delta^3)$, hence

$$\beta_*\;=\;\frac{\delta}{4}-\frac{\delta^2}{8}+O(\delta^3)\;=\;h+\frac{h^2}{2}-2h^2+O(h^3)\;=\;h\left(1-\frac{3}{2}h+O(h^2)\right).$$

Inverting gives $\varepsilon_c(p)=1/\beta_*=h^{-1}(1+\frac{3}{2}h+O(h^2))=\frac{1}{(p-2)\log 2}+\frac{3}{2}+O(p-2)$.

For $p\to\infty$, $a=2^p\to\infty$. Evaluating g at $\beta=\frac{\log 2}{4}$ gives $g(\frac{\log 2}{4})=1+2^{-a/4}-\sqrt{2}<0$ for all $p\ge 4$ (since $2^{-4}=0.0625<\sqrt{2}-1\approx 0.4142$). Because $g(\beta)<0$ on $(0,\beta_*)$ and $g(\beta)>0$ on (β_*,∞), this proves the uniform lower bound $\beta_*>\frac{\log 2}{4}$ for all $p\ge 4$, whence $e^{-a\beta_*}<2^{-a/4}$. From $2e^{-2\beta_*}=1+e^{-a\beta_*}\in(1,2)$, we obtain

$$0\;<\;\frac{\log 2}{2}-\beta_*\;=\;\frac{1}{2}\log(1+e^{-a\beta_*})\;\le\;\frac{1}{2}e^{-a\beta_*}\;<\;\frac{1}{2}\,2^{-a/4},$$

so $a(\frac{\log 2}{2}-\beta_*)=O(a\,2^{-a/4})=o(1)$. Consequently $e^{-a\beta_*}=e^{-(a\log 2)/2}(1+o(1))=2^{-2^{p-1}}(1+o(1))$, giving $\beta_*=\frac{\log 2}{2}-\frac{1}{2}2^{-2^{p-1}}(1+o(1))$, and inverting β_* yields (17).

Proof of (c). Given any $\varepsilon>0$, choose $h\in(0,L/2]$ small enough that $\varepsilon_c(p,h)=h^p/\beta_*(p)<\varepsilon$, and embed three collinear points $F:=\{x_0,x_0+hv,x_0+2hv\}\subset X$ ($\|v\|_2=1$) along the segment. For any probability measures $\alpha,\beta\in\mathcal{P}(X)$ supported on F, every coupling $\pi\in\Pi(\alpha,\beta)$ is necessarily concentrated on $F\times F$, where $\int_{X\times X}c_p\,d\pi+\varepsilon\,\mathrm{KL}(\pi\|\alpha\otimes\beta)$ coincides identically with the finite-space objective (1) on F (and likewise for $\Pi(\alpha,\alpha)$ and $\Pi(\beta,\beta)$), so $S_\varepsilon^X(\alpha,\beta)=S_\varepsilon^F(\alpha,\beta)$. By Part (a) and Theorem 3.1(d), there exist probability measures α,β supported on F with $S_\varepsilon^X(\alpha,\beta)=S_\varepsilon^F(\alpha,\beta)<0$. □

5. Additive Positivity Semigroups and FitzGerald–Horn Re-Entrant Islands

In Theorem 4.2(c) and Theorem 4.3(a), the set of positive regularizations $\{\varepsilon>0:S_\varepsilon\ge 0\}$ was a single interval $(0,\varepsilon_c]$. It is tempting to conjecture that decreasing ε always preserves positivity of S_ε. We now show that this fails dramatically for $n\ge 4$: the inverse-regularization positivity set has a rich additive semigroup structure governed by Hadamard powers.

Theorem 5.1 (Additive positivity semigroup and FitzGerald–Horn threshold)**.** *Let $C=C^\top\in\mathbb{R}^{n\times n}$ ($n\ge 2$) satisfy $C_{ii}=0<C_{ij}$ for all $i\ne j$, and define the inverse-regularization positivity set*

$$\mathcal{B}_+(C)\;:=\;\big\{\beta>0:S_{1/\beta}(\alpha,\beta')\ge 0\;\forall\,\alpha,\beta'\in\Delta_n\big\}.$$

Then:

(a) $\mathcal{B}_+(C) = \{\beta > 0 : K_{1/\beta} \succeq 0\}$ *is a closed* additive subsemigroup *of* $(0,\infty)$*: if* $\beta_1, \beta_2 \in \mathcal{B}_+(C)$*, then* $\beta_1 + \beta_2 \in \mathcal{B}_+(C)$*.*

(b) *There exists a finite threshold* $\beta_{\mathrm{crit}}(C) > 0$ *such that* $[\beta_{\mathrm{crit}}(C), \infty) \subseteq \mathcal{B}_+(C)$*. For* $n = 2$*, every such* C *is CND and* $\mathcal{B}_+(C) = (0,\infty)$*. For* $n \geq 3$*, if* $\beta_0 \in \mathcal{B}_+(C)$*, then by FitzGerald–Horn's theorem* [3]*, every* $\beta \geq (n-2)\beta_0$ *lies in* $\mathcal{B}_+(C)$ *(equivalently,* $S_{\varepsilon_0} \geq 0 \implies S_\varepsilon \geq 0$ *for all* $\varepsilon \in (0, \varepsilon_0/(n-2))$*).*

Proof. By Theorem 3.1, $\mathcal{B}_+(C) = \{\beta > 0 : K_{1/\beta} \succeq 0\}$, which is closed in $(0,\infty)$ by continuity of $\beta \mapsto \lambda_{\min}(K_{1/\beta})$. Since $K_{1/(\beta_1+\beta_2)} = K_{1/\beta_1} \circ K_{1/\beta_2}$ (entrywise Hadamard product), the Schur product theorem immediately gives $\beta_1, \beta_2 \in \mathcal{B}_+(C) \implies \beta_1 + \beta_2 \in \mathcal{B}_+(C)$. Because $C_{ii} = 0$ and $c_{\min} := \min_{i \neq j} C_{ij} > 0$, Gershgorin's disk theorem gives $\lambda_{\min}(K_{1/\beta}) \geq 1 - (n-1)e^{-\beta c_{\min}} > 0$ for all $\beta > \frac{\log(n-1)}{c_{\min}}$, proving the existence of a finite $\beta_{\mathrm{crit}}(C) > 0$ with $[\beta_{\mathrm{crit}}(C), \infty) \subseteq \mathcal{B}_+(C)$. For $n = 2$, $K_{1/\beta} = \left(\begin{smallmatrix} 1 & e^{-\beta C_{12}} \\ e^{-\beta C_{12}} & 1 \end{smallmatrix}\right) \succ 0$ for every $\beta > 0$, so $\mathcal{B}_+(C) = (0,\infty)$. For $n \geq 3$, if $\beta_0 \in \mathcal{B}_+(C)$, then $A := K_{1/\beta_0}$ is an entrywise-positive $n \times n$ PSD matrix, so for any $\beta \geq (n-2)\beta_0$ (noting $n - 2 \geq 1$), the Hadamard power $K_{1/\beta} = A^{\circ(\beta/\beta_0)}$ is PSD by FitzGerald–Horn's critical exponent theorem [3, Theorem 2.2] (which proves $A^{\circ\gamma} \succeq 0$ for all $\gamma \geq n-2$). □

Theorem 5.2 (Existence of open re-entrant positivity islands for all $n \geq 4$)**.** *For every* $n \geq 4$*, there exists a symmetric non-negative zero-diagonal cost matrix* $C = C^\top \in \mathbb{R}^{n \times n}_{\geq 0}$ *(*$C_{ii} = 0 < C_{ij}$ *for* $i \neq j$*) whose Sinkhorn positivity set* $\{\varepsilon > 0 : S_\varepsilon > 0$ *off-diagonal*$\}$ *is* disconnected*: specifically, there exists a non-empty open re-entrant positivity island* $(\varepsilon_-, \varepsilon_+) \subset (\frac{2}{3}, \frac{4}{3})$ *containing* $\varepsilon = 1$ *such that*

- $\inf_{\alpha,\beta} S_\varepsilon(\alpha,\beta) < 0$ *for all* $\varepsilon \geq \frac{4}{3}$ *(high-*ε *indefinite regime),*
- $S_\varepsilon(\alpha,\beta) > 0$ *for all* $\alpha \neq \beta$ *whenever* $\varepsilon \in (\varepsilon_-, \varepsilon_+)$ *(open re-entrant positivity island containing* $\varepsilon = 1$*),*
- $\inf_{\alpha,\beta} S_\varepsilon(\alpha,\beta) < 0$ *on a non-empty open interval containing* $\varepsilon = \frac{2}{3}$ *(re-entrant negative gap strictly separating* $(\varepsilon_-, \varepsilon_+)$ *from the low-*ε *phase),*
- $S_\varepsilon(\alpha,\beta) > 0$ *for all* $\alpha \neq \beta$ *whenever* $\varepsilon \in (0, \frac{1}{2}]$ *(low-*ε *positive-definite phase).*

Proof. First take $n = 4$ and let $\theta = (0,\ 0.35,\ 0.75,\ 1.30) \in \mathbb{R}^4$. Since $|\theta_i - \theta_j| \leq 1.30 < \pi/2$, the 4×4 cosine Gram matrix $A_0 := (\cos(\theta_i - \theta_j))_{i,j=1}^4$ has strictly positive entries (with minimum entry $\cos(1.30) \approx 0.2675 > 0$), unit diagonal, and rank 2, since $A_0 = UU^\top$ where $U_i = (\cos\theta_i, \sin\theta_i) \in \mathbb{R}^2$; in particular, $A_0 \succeq 0$ with $\lambda_{\min}(A_0) = 0$. Direct evaluation of the 4×4 Hadamard powers $A_0^{\circ 3/2}$ and $A_0^{\circ 1/2}$ shows that both are indefinite, with $\lambda_{\min}(A_0^{\circ 3/2}) = -1.22233 \times 10^{-3} < 0$ and $\lambda_{\min}(A_0^{\circ 1/2}) = -3.55290 \times 10^{-2} < 0$ (for explicit elementary certificates independent of eigensolvers, the rational vectors $v_{3/2} := \frac{1}{100}(-32, 73, -58, 18)^\top$ and $v_{1/2} := \frac{1}{20}(-11, 7, 11, -10)^\top$ satisfy $v_{3/2}^\top A_0^{\circ 3/2} v_{3/2}/\|v_{3/2}\|^2 = -1.1773 \times 10^{-3} < 0$ and $v_{1/2}^\top A_0^{\circ 1/2} v_{1/2}/\|v_{1/2}\|^2 = -3.4751 \times 10^{-2} < 0$).

Now fix any $\delta \in (0,1)$ and define the perturbed matrix $A_\delta := (1-\delta)A_0 + \delta I_4$, whose entries lie in $(0,1]$ with unit diagonal, and set $C_\delta := -\log A_\delta \in \mathbb{R}^{4\times4}_{\geq 0}$ (entrywise). Because I_4 is supported solely on the diagonal and A_0 has unit diagonal, the Hadamard power of A_δ satisfies the *exact algebraic identity*

$$K_{1/\beta}(C_\delta) \;=\; A_\delta^{\circ\beta} \;=\; (1-\delta)^\beta\, A_0^{\circ\beta} + \big(1 - (1-\delta)^\beta\big) I_4 \qquad \forall\, \beta > 0. \tag{18}$$

Consequently, the minimum eigenvalue shifts linearly:

$$\lambda_{\min}\big(K_{1/\beta}(C_\delta)\big) \;=\; (1-\delta)^\beta\, \lambda_{\min}\big(A_0^{\circ\beta}\big) + \big(1 - (1-\delta)^\beta\big). \tag{19}$$

Choose $\delta := 2 \times 10^{-4}$ (any $0 < \delta < 8 \times 10^{-4}$ suffices). Evaluating (19) yields:

(1) At $\beta = 1$: $\lambda_{\min}(A_0) = 0$, so $\lambda_{\min}(K_1(C_\delta)) = \delta = +2 \times 10^{-4} > 0$.
(2) At $\beta = 3/2$: $1 - (1-\delta)^{3/2} \approx 3.0 \times 10^{-4} < 1.222 \times 10^{-3}(1-\delta)^{3/2}$, so $\lambda_{\min}(K_{2/3}(C_\delta)) = -9.22 \times 10^{-4} < 0$.
(3) For all $\beta \in (0, 3/4]$ (i.e. all $\varepsilon \geq 4/3$): we must have $\lambda_{\min}(K_{1/\beta}(C_\delta)) < 0$! Indeed, if some $\beta_0 \in (0, 3/4]$ belonged to $\mathcal{B}_+(C_\delta)$, then since $3/2 \geq 2\beta_0 = (4-2)\beta_0$, Theorem 5.1(b) (FitzGerald–Horn for $n = 4$) would force $3/2 \in \mathcal{B}_+(C_\delta)$, contradicting item (2).
(4) For all $\beta \geq 2$ (i.e. all $\varepsilon \in (0, 1/2]$): since $A_0 \succeq 0$ is 4×4 entrywise positive, FitzGerald–Horn's theorem [3, Theorem 2.2] guarantees $\lambda_{\min}(A_0^{\circ\beta}) \geq 0$, whence $\lambda_{\min}(K_{1/\beta}(C_\delta)) \geq 1 - (1-\delta)^\beta \geq 1 - (1-\delta)^2 = \delta(2-\delta) > 0$ for every $\beta \geq 2$.

By continuity of $\beta \mapsto \lambda_{\min}(K_{1/\beta}(C_\delta))$, letting $(\beta_-, \beta_+) \subset (\frac{3}{4}, \frac{3}{2})$ be the open connected component of $\{\beta > 0 : \lambda_{\min}(K_{1/\beta}(C_\delta)) > 0\}$ containing $\beta = 1$, and setting $\varepsilon_- := 1/\beta_+$ and $\varepsilon_+ := 1/\beta_-$, Theorem 3.1 establishes all four properties for $n = 4$ (numerically $\beta_- \approx 0.998259$, $\beta_+ \approx 1.060219$, and the low-ε threshold lies at $\beta_3 \approx 1.927538 < 2$).

For $n > 4$, extend C_δ to an $n \times n$ symmetric cost $C_\delta^{(n)}$ by setting $C_{\delta,ii}^{(n)} = 0$ and $C_{\delta,ij}^{(n)} = M$ whenever $\max(i, j) > 4$, where $M > 0$ is chosen large enough that $(n-1)e^{-M} < \frac{1}{2}\delta$ (which also ensures $(n-1)e^{-2M} < \frac{1}{2}\delta(2-\delta)$). Write $K_{1/\beta}(C_\delta^{(n)}) = \big(K_{1/\beta}(C_\delta) \oplus I_{n-4}\big) + E_M(\beta)$, where $E_M(\beta)$ has zero diagonal and off-diagonal entries bounded by $e^{-\beta M}$, hence $\|E_M(\beta)\|_{\mathrm{op}} \leq (n-1)e^{-\beta M}$ by Gershgorin's disk theorem. Therefore:

- For every $\beta \in (0, \frac{3}{4}] \cup \{\frac{3}{2}\}$, Cauchy interlacing with the leading 4×4 principal block $K_{1/\beta}(C_\delta)$ immediately yields $\lambda_{\min}(K_{1/\beta}(C_\delta^{(n)})) \leq \lambda_{\min}(K_{1/\beta}(C_\delta)) < 0$ (independently of M).
- At $\beta = 1$, Weyl's inequality gives $\lambda_{\min}(K_1(C_\delta^{(n)})) \geq \delta - (n-1)e^{-M} > \delta/2 > 0$.
- For every $\beta \geq 2$, $\lambda_{\min}(K_{1/\beta}(C_\delta) \oplus I_{n-4}) \geq 1 - (1-\delta)^\beta \geq \delta(2-\delta)$ and $\|E_M(\beta)\|_{\mathrm{op}} \leq (n-1)e^{-2M} < \frac{1}{2}\delta(2-\delta) \leq \frac{1}{2}(1-(1-\delta)^\beta)$, so Weyl's inequality gives $\lambda_{\min}(K_{1/\beta}(C_\delta^{(n)})) \geq \frac{1}{2}\big(1-(1-\delta)^\beta\big) > 0$.

Taking the open connected component of $\beta = 1$ inside $\{\beta > 0 : \lambda_{\min}(K_{1/\beta}(C_\delta^{(n)})) > 0\}$ completes the proof for every $n \geq 4$. □

Remark 5.3 (True metric spaces and re-entrancy)**.** Note that the cosine-log cost C_δ in Theorem 5.2 satisfies $C_{ij} = C_{ji} > 0$ $(i \neq j)$ and $C_{ii} = 0$, but violates the triangle inequality because $-\log\cos x \approx x^2/2$ is locally quadratic near $x = 0$. Whether a *true finite metric space* (satisfying the triangle inequality) can also exhibit a disconnected Sinkhorn positivity set $\mathcal{B}_+(d)$ remains an intriguing open question in metric geometry.

6. High-Precision Numerical Verification

Every theorem and formula above has been independently verified via our reproducible suite in `scripts/verify_sinkhorn.py` and `tests/test_sinkhorn.py`, combining 50-digit `mpmath` root-finding with log-domain Sinkhorn iterations converged to dual tolerance 10^{-14}.

Across 30 random 5×5 symmetric cost matrices and random strictly positive base measures $\mu \in \Delta_5^+$, the analytical diagonal Hessian (9) matches centered finite differences of $S_\varepsilon(\mu + t\eta, \mu - t\eta)$ $(t = 10^{-4})$ to maximum relative error 3.75×10^{-8}. On $K_{2,3}$ at $\varepsilon_c = 2/\log 2 = 2.8853900818\ldots$, both the interior affine zero-locus witness (13) (with $u = (1.2, 0.9, 1.1, 0.8, 1.3)$ and $t = 0.2$) and the disjoint-support boundary witness (μ_A, μ_B) yield $|S_{\varepsilon_c}| < 2.3 \times 10^{-16}$, while at $\varepsilon = 1.02\,\varepsilon_c$, letting w be a unit eigenvector of K_ε for $\lambda_{\min}(K_\varepsilon) < 0$ and setting $P_+ := \sum_{w_i > 0} w_i$, $P_- := \sum_{w_i < 0}(-w_i)$, $u_i := P_+$ if $w_i > 0$, P_- if $w_i < 0$, and 1 if $w_i = 0$ (normalized to $\tilde{u} := u/\sqrt{u^\top K_\varepsilon u}$, yielding $\mu_i := \tilde{u}_i(K_\varepsilon \tilde{u})_i \in \Delta_5^+$), unit negative eigenvector v of S_μ ($\lambda_{\min}(S_\mu) = -5.33153 \times 10^{-4}$), $\eta = D_\mu^{1/2} v$,

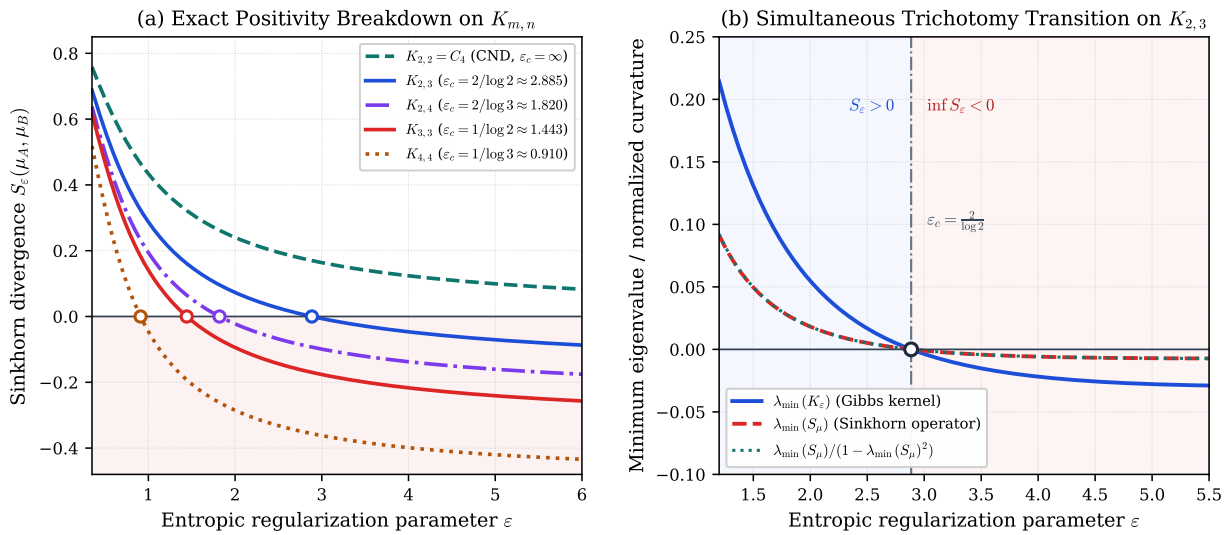


FIGURE 1. (a) Exact Sinkhorn divergence $S_\varepsilon(\mu_A,\mu_B)$ on complete bipartite graph metrics $K_{m,n}$, crossing zero at the exact threshold $\varepsilon_c(K_{m,n})=2/\log((m-1)(n-1))$ (Theorem 4.2(c)); $K_{2,2}=C_4$ is CND and remains positive for all $\varepsilon>0$. (b) Simultaneous zero crossing on $K_{2,3}$ of $\lambda_{\min}(K_\varepsilon)$, $\lambda_{\min}(S_\mu)$, and the normalized diagonal curvature $\lambda_{\min}(S_\mu)/(1-\lambda_{\min}(S_\mu)^2)$ at $\varepsilon_c=2/\log 2$.

TABLE 1. Exact critical regularization thresholds $\varepsilon_c(K_{m,n})=2/\log((m-1)(n-1))$, minimum Gibbs eigenvalue at ε_c, and negative Sinkhorn divergence $S_\varepsilon(\mu_A,\mu_B)$ at $\varepsilon=1.5\,\varepsilon_c$ and $\varepsilon\to\infty$.

Graph	$(m-1)(n-1)$	$\varepsilon_c(K_{m,n})$	$\lambda_{\min}(K_{\varepsilon_c})$	$S_{1.5\varepsilon_c}(\mu_A,\mu_B)$	$S_\infty(\mu_A,\mu_B)$
$K_{2,3}$	2	2.8853900818	-3.6×10^{-17}	-0.05577965	$-1/6\approx-0.16666667$
$K_{2,4}$	3	1.8204784533	$+7.6\times10^{-17}$	-0.08405854	$-1/4=-0.25000000$
$K_{2,5}$	4	1.4426950409	$+3.6\times10^{-16}$	-0.10118527	$-3/10=-0.30000000$
$K_{3,3}$	4	1.4426950409	-4.2×10^{-17}	-0.11303519	$-1/3\approx-0.33333333$
$K_{3,4}$	6	1.1162212531	$+9.7\times10^{-17}$	-0.14272597	$-5/12\approx-0.41666667$
$K_{4,4}$	9	0.9102392266	$+3.3\times10^{-16}$	-0.17363363	$-1/2=-0.50000000$
$K_{5,5}$	16	0.7213475204	$+2.8\times10^{-16}$	-0.21281830	$-3/5=-0.60000000$

and step $t=10^{-3}$ yields $S_\varepsilon(\mu+t\eta,\mu-t\eta)=-3.1382434\times10^{-9}$, matching the analytical prediction $2\varepsilon\frac{\lambda_{\min}(S_\mu)}{1-\lambda_{\min}(S_\mu)^2}t^2=-3.1382438\times10^{-9}$ to 7 digits.

Figure 1 and Table 1 verify the complete bipartite threshold (14). Figure 2(a) and Tables 2–3 confirm the exact super-quadratic phase curve (16)–(17) and continuum collapse on uniform N-point discretizations of $[0,1]$ at $p=3$. Finally, Figure 2(b) and Table 4 confirm all four phases of the FitzGerald–Horn re-entrant island via both $\lambda_{\min}(K_{1/\beta})$ and direct log-domain Sinkhorn evaluations.

REFERENCES

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

[2] J. Feydy, T. Séjourné, F.-X. Vialard, S.-I. Amari, A. Trouvé, and G. Peyré, *Interpolating between optimal transport and MMD using Sinkhorn divergences*, Proc. 22nd Int. Conf. Artif. Intell. Stat. (AISTATS), PMLR **89** (2019), 2681–2690.

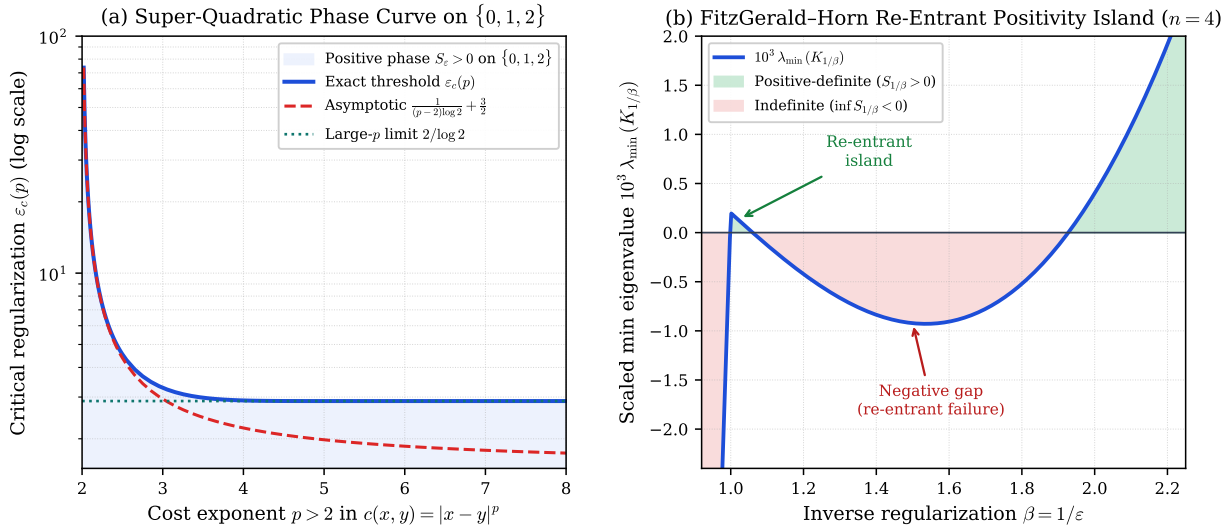


FIGURE 2. (a) Exact critical regularization curve $\varepsilon_c(p)$ for $|x-y|^p$ $(p > 2)$ on $\{0,1,2\}$ together with the two exact asymptotics $\frac{1}{(p-2)\log 2} + \frac{3}{2}$ $(p \to 2^+)$ and $2/\log 2$ $(p \to \infty)$ from Theorem 4.3. (b) FitzGerald–Horn re-entrant positivity island on $n = 4$ ($C_\delta = -\log((1-\delta)A_0 + \delta I_4)$, $\delta = 2 \times 10^{-4}$; Theorem 5.2): $S_{1/\beta}$ is strictly positive definite on the island $\beta \in (0.9983, 1.0602)$, becomes indefinite on the re-entrant negative gap $\beta \in (1.0602, 1.9275)$, and returns to strict positive definiteness for $\beta > 1.9275$.

TABLE 2. Exact 50-digit values of $\varepsilon_c(p)$ on $\{0,1,2\}$ for $|x-y|^p$ $(p > 2)$ compared with the near-quadratic expansion $\frac{1}{(p-2)\log 2} + \frac{3}{2}$ and the large-p limit $2/\log 2$.

p	$\varepsilon_c(p)$	$\frac{1}{(p-2)\log 2} + \frac{3}{2}$	Difference	$\varepsilon_c(p)/(2/\log 2)$
2.001	1444.19532978	1444.19504089	$+2.89 \times 10^{-4}$	500.519961
2.010	145.77240017	145.76950409	$+2.90 \times 10^{-3}$	50.520864
2.050	30.36853657	30.35390082	$+1.46 \times 10^{-2}$	10.524933
2.100	15.95659330	15.92695041	$+2.96 \times 10^{-2}$	5.530134
2.500	4.54566819	4.38539008	$+1.60 \times 10^{-1}$	1.575409
3.000	3.28203586	2.94269504	$+3.39 \times 10^{-1}$	1.137467
4.000	2.90224716	2.22134752	$+6.81 \times 10^{-1}$	1.005842
6.000	2.88539008	1.86067376	$+1.02 \times 10^{0}$	1.000000
10.000	2.88539008	1.68033688	$+1.21 \times 10^{0}$	1.000000

[3] C. H. FitzGerald and R. A. Horn, *On fractional Hadamard powers of positive definite matrices*, J. Math. Anal. Appl. **61** (1977), 633–642.
[4] A. Genevay, G. Peyré, and M. Cuturi, *Learning generative models with Sinkhorn divergences*, Proc. 21st Int. Conf. Artif. Intell. Stat. (AISTATS), PMLR **84** (2018), 1608–1617.
[5] H. Lavenant, J. Luckhardt, G. Mordant, B. Schmitzer, and L. Tamanini, *The Riemannian geometry of Sinkhorn divergences*, arXiv:2405.04987 (2024).
[6] G. Peyré and M. Cuturi, *Computational optimal transport: With applications to data science*, Found. Trends Mach. Learn. **11** (2019), 355–607.
[7] I. J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), 522–536.
[8] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist. **35** (1964), 876–879.

TABLE 3. Continuum collapse on uniform N-point grids of $[0,1]$ (spacing $h = (N-1)^{-1}$) for $c_3(x,y) = |x-y|^3$, compared against the 3-point upper bound $h^3\varepsilon_c(3)$ from Theorem 4.3(a),(c).

N	$h = (N-1)^{-1}$	$\varepsilon_c(\text{grid})$	$h^3\varepsilon_c(3)$	$\varepsilon_c(\text{grid})/(h^3\varepsilon_c(3))$
3	0.500000	$4.10254482 \times 10^{-1}$	$4.10254482 \times 10^{-1}$	1.000000
5	0.250000	$2.93117849 \times 10^{-2}$	$5.12818103 \times 10^{-2}$	0.571582
9	0.125000	$3.08787799 \times 10^{-3}$	$6.41022629 \times 10^{-3}$	0.481711
17	0.062500	$3.64968434 \times 10^{-4}$	$8.01278286 \times 10^{-4}$	0.455483
33	0.031250	$4.48524068 \times 10^{-5}$	$1.00159786 \times 10^{-4}$	0.447809

TABLE 4. Direct Sinkhorn verification across all four phases of the $n = 4$ FitzGerald–Horn cost C_δ ($\delta = 2 \times 10^{-4}$, Theorem 5.2), with numerical transition boundaries $\beta_1 = 0.998259$, $\beta_2 = 1.060219$, and $\beta_3 = 1.927538$. In indefinite regimes ($\beta \in \{0.90, 1.50\}$), the witness column reports $S_\varepsilon(\mu + 10^{-2}\eta, \mu - 10^{-2}\eta)$ along the negative eigenvector $\eta = D_\mu^{1/2} v$ of S_μ, where $\mu_i := \tilde{u}_i (K_{1/\beta}\tilde{u})_i$ is built from the unit negative eigenvector w of $K_{1/\beta}$ via the (P_+, P_-) scaling formula in Section 6; in positive regimes ($\beta \in \{1.00, 2.20\}$), it reports the minimum over 10 independent Dirichlet($\mathbf{1}$) pairs.

Regime	$\beta = 1/\varepsilon$	ε	$\lambda_{\min}(K_{1/\beta})$	Direct S_ε witness
High-ε indefinite	0.90	1.111111	-1.060626×10^{-2}	-7.496953×10^{-7}
Re-entrant island (> 0)	1.00	1.000000	$+2.000000 \times 10^{-4}$	$+3.034986 \times 10^{-3}$
Re-entrant negative gap	1.50	0.666667	-9.219788×10^{-4}	-3.877906×10^{-8}
Low-ε positive (> 0)	2.20	0.454545	$+1.915514 \times 10^{-3}$	$+4.700081 \times 10^{-3}$

INDEPENDENT RESEARCHER, AIRR SCIENCE
Email address: editor@airr.science