# THE PRICE OF REGULARIZING SQUARED DISTANCE IN HYPERBOLIC OPTIMAL TRANSPORT

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Changing a transport cost to obtain regularity also changes the task being solved. We quantify this price for squared distance on hyperbolic space, whose negative curvature violates the Ma–Trudinger–Wang (MTW) condition. Among normalized radial costs in dimension $d \geq 3$, critical log-cosh minimizes uniform deviation from squared distance among full-A3w costs, uniquely on each bounded distance interval. A prescribed transverse MTW floor admits an explicit sharp envelope with interval rigidity. Its weakest full margin and mixed-Hessian condition number obey $m_D^\star(\mathsf{K}_D^\star)^2 = \Lambda$, exposing the sensitivity cost of this optimal repair. Positive full margins require additional distortion: finite directional candidates and a sixth-order obstruction sharpen analytic bounds, while calibrated log-cosh supplies a feasible family-wise frontier. The critical repair's distortion and conditioning penalties vanish in the flat limit. In a fixed three-target OT problem, new-seed experiments show that repair changes the optimal transport regions at a measurable original-objective cost. Separate inversion experiments recover the predicted local sensitivity. Together, the theory and experiments distinguish the geometric benefit of regularization from its task and numerical prices.

## 1 INTRODUCTION

An optimal transport cost specifies which assignments are preferable. Replacing that cost to obtain a regular transport map therefore changes the problem, even when the data and solver stay fixed. This matters for hyperbolic representations of hierarchies (Nickel & Kiela, 2017; 2018; Sala et al., 2018): aligning two distributions naturally starts from the metric displacement cost

$$c_0(x, y) = \tfrac{1}{2} d_\kappa(x, y)^2, \tag{1}$$

where the ambient curvature is $-\kappa^2$. Yet this cost violates the Ma–Trudinger–Wang (MTW) condition, a necessary condition for continuity guarantees over arbitrary smooth positive data (Loeper, 2009). Smooth densities alone cannot remove this geometric obstruction.

Log-cosh is already known to satisfy A3w (Lee & Li, 2012) and is used in hyperbolic OT (Alvarez-Melis et al., 2020; Ng & Gu, 2025). Its admissibility leaves the minimum repair price undetermined. *How much must the transport task change to remove the curvature obstruction, and what numerical sensitivity does that change create?* We seek the regular cost that best preserves the task and quantify the sensitivity of its transport equation.

Figure 1 illustrates the issue on a fixed three-target problem. Squared distance sends the origin away from the central target; a log-cosh repair sends it to that target. Both solutions satisfy the same mass constraints. The repair changes an actual optimal assignment, and evaluating it under the original squared objective reveals its price. This is the behavior our theory is designed to quantify, rather than an assumption that a regular cost improves every matching task.

We fix the cost scale by $\ell''(0) = 1$ and measure the largest deviation from $r^2/2$ on a distance range $[0, D]$. This criterion controls original-objective regret for every pair of probability distributions supported within that range. The main results are:

- **Sharp repair prices.** For $d \geq 3$, we identify the closest full-A3w repair on each distance interval and the rigid envelope for every nonnegative transverse floor.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

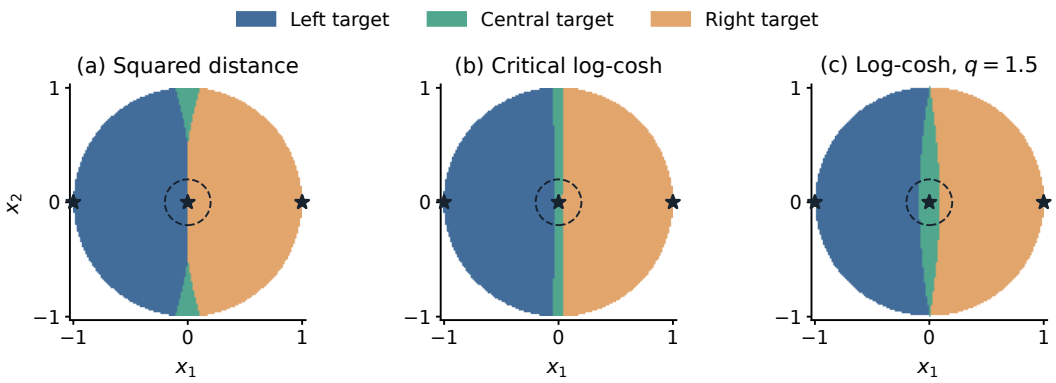


Figure 1: **Repairing the cost changes the optimal transport.** Fixed source points, targets (stars), and target masses; only the cost changes. Colors show target cells defined by optimal discrete dual potentials on the $x_3 = 0$ section of a three-dimensional problem ($\kappa = 1$, $n = 2049$, seed 100). The dashed circle marks $r = 0.2$. The critical repair brings the origin into the central cell; its original-objective regret is $0.00103$. This section does not establish three-dimensional disconnectedness or smooth-map regularity.

- **The numerical price of the optimum.** The envelope's weakest full margin and mixed-Hessian condition number satisfy $m_D^\star(\mathsf{K}_D^\star)^2 = \Lambda$. The same geometry that limits distortion therefore determines loss of full margin and increased inverse sensitivity.
- **What an everywhere-positive margin costs.** Full directional minimization reduces to finitely many candidates for $d \geq 2$. A sixth-order obstruction strengthens finite-cap bounds after fourth-order saturation. Calibrated log-cosh has an exact family-wise frontier; the unrestricted positive full-margin infimum remains open.

The critical repair's prices vanish as $\kappa \downarrow 0$, isolating negative curvature's contribution. We test transport-region changes, original-objective regret, and local inverse sensitivity separately. All application results are retained, including matching experiments without a confirmed benefit.

## 2 Related Work

**Regularity for a fixed cost.** The MTW condition (Ma et al., 2005) connects cost derivatives to map regularity. Loeper (2009) established its necessity for continuity guarantees over arbitrary smooth positive data; Kim & McCann (2010) gave its geometric formulation. Regularity theorems also impose density and domain hypotheses (Figalli et al., 2013; Liu et al., 2010; Trudinger & Wang, 2009); see De Philippis & Figalli (2014). Partial regularity can hold without MTW (De Philippis & Figalli, 2015). Our target is the cost obstruction to the stronger continuity guarantee.

**Regular hyperbolic costs.** Lee & Li (2012) establish scalar MTW criteria and show that log-cosh satisfies A3w. We use these criteria to prove its optimal fidelity among normalized radial repairs in $d \geq 3$, with equality forcing the profile on the distance interval. The positive-budget envelope and its conditioning law extend this extremal analysis. Cost constructions (Nguyen, 2026), regional certificates (Shivakumar et al., 2026), and log-cosh solvers (Ng & Gu, 2025) address complementary questions.

**Learning or modifying the cost.** Ground metric learning fits a cost to labeled data (Cuturi & Avis, 2014); inverse OT infers costs from observed couplings (Stuart & Wolfram, 2020; Chiu et al., 2022). Here the original task cost is fixed, and the admissible modification is constrained by MTW geometry. Entropic regularization instead penalizes the coupling while retaining the chosen ground cost (Cuturi, 2013; Peyré & Cuturi, 2019). Its approximation bias (Weed, 2018) and numerical stabilization (Schmitzer, 2019) are separate from the mixed-Hessian sensitivity studied here.

**Hyperbolic representations and learned maps.** Hyperbolic networks exploit hierarchy (Ganea et al., 2018; Chami et al., 2019), and OT aligns such representations (Alvarez-Melis et al., 2020). Embedding precision tradeoffs (Sala et al., 2018) concern representation error; ours concern the cost on fixed geometry. Neural and statistical map estimation (Seguy et al., 2018; Makkuva et al., 2020; Manole et al., 2024) require separate approximation assumptions. Our direct-map bounds do not rule out weaker-norm estimation of discontinuous maps (Pooladian et al., 2023).

## 3 Setup and the cost of changing the task

Let $\mathbb{H}^d_\kappa$ have constant sectional curvature $-\kappa^2$, with $d \geq 2$ and $\kappa > 0$. It is complete, simply connected, and has no cut locus. We study $c_\ell(x, y) = \ell(d_\kappa(x, y))$: radial costs preserve all ambient isometries, so only the distance penalty changes.

**Definition 1** (Admissible profiles)**.** The class $\mathcal{H}_\kappa$ consists of profiles with an even $C^4$ extension to $\mathbb{R}$, with $\ell(0) = 0$, $\ell''(0) = 1$, and $\ell'(r), \ell''(r) > 0$ for $r > 0$.

Evenness gives $\ell'(0) = \ell^{(3)}(0) = 0$ and $C^4$ regularity across the diagonal. Positivity ensures twist and nondegeneracy. The normalization removes the additive constant and fixes the scale of the original objective. Positive rescaling changes fidelity and absolute inverse sensitivity, but not a spectral condition number.

Identify covectors with vectors via the Riemannian metric. In the Lee–Li convention,

$$\mathrm{MTW}_x(u, p, w) = -\frac{3}{2}\,\frac{\partial^4}{\partial s^2 \partial t^2} c\big(\gamma(t), c\text{-}\exp_x(p + sw)\big)\Big|_{s=t=0}, \tag{2}$$

where $\gamma$ is a geodesic with $\gamma(0) = x$, $\dot\gamma(0) = u$, $\langle u, w\rangle = 0$, and $c\text{-}\exp_x$ inverts $y \mapsto -D_x c(x, y)$. A3w requires nonnegative MTW; A3s requires strict positivity on nonzero orthogonal directions. A stated margin uses unit directions. Map regularity additionally requires the density, domain, and $c$-convexity hypotheses of the cited PDE results. All norms are Riemannian; endpoint tangent spaces are compared by parallel transport.

**Definition 2** (Fidelity, margin, and conditioning)**.** For $D > 0$, with $\operatorname{cond} M = \sigma_{\max}(M)/\sigma_{\min}(M)$ in orthonormal frames, let

$$\Delta(\ell; D) = \sup_{0 \leq r \leq D} \big[r^2/2 - \ell(r)\big], \tag{3}$$

$$m(\ell; D) = \inf_{\substack{d_\kappa(x,y) \leq D,\ \|u\| = \|w\| = 1 \\ \langle u, w\rangle = 0}} \mathrm{MTW}_x(u, -D_x c_\ell(x, y), w), \tag{4}$$

$$\mathsf{K}(\ell; D) = \sup_{d_\kappa(x,y) \leq D} \operatorname{cond}(-D^2_{xy} c_\ell). \tag{5}$$

The envelope below proves $\ell \leq r^2/2$ on its feasible class, so $\Delta$ there equals uniform absolute cost deviation. If exact minimizers $\pi_c, \pi_0$ share the same feasible coupling set and $0 \leq c_0 - c \leq \Delta$, then their original-objective regret obeys $0 \leq \int c_0 \,\mathrm{d}(\pi_c - \pi_0) \leq \Delta$ (Proposition 19). This motivates fidelity without assuming that a small objective change forces a small change of plan.

**Two constraints.** Put $z = \ell'(r)$ and $g = (\ell')^{-1}$. The Lee–Li functions are

$$A(z) = 1/g'(z), \qquad B(z) = \kappa z \coth(\kappa g(z)), \qquad \delta(z) = B'(z)/z. \tag{6}$$

For $d \geq 3$, two orthogonal directions tangent to a distance sphere have MTW value $-3\delta/2$. Define

$$\Lambda_\perp(\ell; D) = \inf_{0 \leq z \leq \ell'(D)} -3\delta(z)/2. \tag{7}$$

The full margin satisfies $m \leq \Lambda_\perp$. Thus transverse and full constraints lead to different design problems:

$$V_\perp(\Lambda, D) = \inf_{\ell \in \mathcal{H}_\kappa:\, \Lambda_\perp(\ell; D) \geq \Lambda} \Delta(\ell; D), \qquad d \geq 3, \tag{8}$$

$$V_{\mathrm{full}}(\Gamma, D) = \inf_{\ell \in \mathcal{H}_\kappa:\, m(\ell; D) \geq \Gamma} \Delta(\ell; D), \qquad d \geq 2. \tag{9}$$

We solve $V_\perp$ and, in $d \geq 3$, $V_{\mathrm{full}}(0, D)$. Positive full margins have analytic bounds. Throughout, $D$ bounds *all* admissible source–target distances; a distance quantile cannot replace it.

**All directions can be checked explicitly.** Let $\alpha, \beta, \gamma, \delta$ be the scalar coefficients in (54). For orthogonal unit directions with radial energies $x = |u_0|^2, y = |w_0|^2$, their MTW value is $-3Q_z(x, y)/2$, where

$$Q_z(x, y) = \delta + (\beta - \delta)x + (\gamma - \delta)y + (\alpha - \beta - \gamma + \delta)xy. \tag{10}$$

**Proposition 3** (Finite directional candidates)**.** *Fix $z \geq 0$, set $s = \alpha - \beta - \gamma + \delta$ and $h = \alpha - 2\gamma + \delta$, and let*

$$\mathcal{C}_2 = \{\beta, \gamma\} \cup \begin{cases} \{\gamma + h^2/(4s)\}, & s > 0,\ 0 \leq h \leq 2s, \\ \varnothing, & \textit{otherwise}. \end{cases} \tag{11}$$

*Set $\mathcal{C}_d = \mathcal{C}_2 \cup \{\delta\}$ for $d \geq 3$. Then*

$$\max_{u \perp w,\, \|u\| = \|w\| = 1} Q_z = \max \mathcal{C}_d(z), \qquad m(\ell; D) = \inf_{0 \leq z \leq \ell'(D)} \left[ -\tfrac{3}{2} \max \mathcal{C}_d(z) \right]. \tag{12}$$

In $d = 2$, orthogonality forces $x + y = 1$; in $d \geq 3$, it gives the triangle $x, y \geq 0$, $x + y \leq 1$. Maximizing the quadratic on its edges yields the candidates. The remaining full-margin optimization is over profiles and distances.

## 4 The local price of MTW regularity

The first question is whether repair can be arbitrarily small near $x = y$. It cannot: the fourth derivative of $\ell$ must compensate for the negative ambient curvature.

**Definition 4** (Local prices)**.** For unit $u \perp w$, define

$$\mathsf{R}_0 = \mathrm{MTW}_x(u, 0, w), \quad \mathsf{F}_0 = \lim_{r \downarrow 0} \frac{12[r^2/2 - \ell(r)]}{r^4}, \quad \mathsf{N}_0 = \lim_{r \downarrow 0} \frac{2[1 - \ell''(r)]}{r^2}. \tag{13}$$

Isotropy makes $\mathsf{R}_0$ independent of the chosen unit pair.

**Lemma 5** (Diagonal MTW identity)**.** *For every $\ell \in \mathcal{H}_\kappa$,*

$$\mathsf{R}_0(\ell) = -\kappa^2 - \frac{1}{2}\ell^{(4)}(0). \tag{14}$$

The proof in Section A expands the inverse momentum and Lee–Li coefficients. It fixes the factor $-3/2$ in (2); an alternative convention requires rescaling every prescribed margin consistently.

**Theorem 6** (Local repair identity and jet rigidity)**.** *For every $\ell \in \mathcal{H}_\kappa$,*

$$\mathsf{F}_0(\ell) = \kappa^2 + \mathsf{R}_0(\ell) = \frac{1}{2}\mathsf{N}_0(\ell). \tag{15}$$

*Consequently, if $\mathsf{R}_0(\ell) \geq \Gamma \geq 0$, then*

$$\mathsf{F}_0(\ell) \geq \kappa^2 + \Gamma, \qquad \mathsf{N}_0(\ell) \geq 2(\kappa^2 + \Gamma). \tag{16}$$

*Both constants are attained by the calibrated family*

$$\ell_q(r) = \frac{\log\cosh(q\kappa r)}{q^2\kappa^2}, \qquad q > 0, \tag{17}$$

*with*

$$q_\Gamma = \sqrt{1 + \Gamma/\kappa^2}. \tag{18}$$

*Equality in either inequality of* (16) *holds if and only if*

$$\ell(r) = \frac{r^2}{2} - \frac{\kappa^2 + \Gamma}{12} r^4 + o(r^4), \tag{19}$$

*so the fourth-order jet agrees with that of $\ell_{q_\Gamma}$.*

In words, even removing the negative sign ($\Gamma = 0$) costs at least $\kappa^2 r^4/12$ to leading order. The same coefficient weakens the radial mixed Hessian. Here a *jet* means the Taylor coefficients at zero: matching the fourth-order price need not determine the cost at finite distances.

## 5 THE SHARP TRANSVERSE REPAIR

We now ask how far below squared distance every repair must lie on $[0, D]$. An *envelope* is the largest feasible profile at each distance. Since feasible costs lie below $r^2/2$, this upper envelope gives the smallest distortion.

**Theorem 7** (Sharp transverse fidelity envelope and interval rigidity)**.** *Let $d \geq 3$, $D > 0$, $\Lambda \geq 0$, and $a = \Lambda/3$. Suppose $\ell \in \mathcal{H}_\kappa$ satisfies*

$$-\frac{3}{2}\delta(z) \geq \Lambda \qquad \text{for } 0 \leq z \leq \ell'(D). \tag{20}$$

*Define*

$$z_a(r) := \frac{2\tanh(\kappa r)}{\kappa + \sqrt{\kappa^2 + 4a\tanh^2(\kappa r)}}, \tag{21}$$

$$\ell_a^\star(r) := \int_0^r z_a(t)\,\mathrm{d}t. \tag{22}$$

*Then, for every $0 \leq r \leq D$,*

$$\ell'(r) \leq z_a(r), \qquad \ell(r) \leq \ell_a^\star(r) \leq \frac{r^2}{2}. \tag{23}$$

*The profile $\ell_a^\star$ belongs to $\mathcal{H}_\kappa$ and attains $-3\delta/2 = \Lambda$ at every finite distance. If equality $\ell(R) = \ell_a^\star(R)$ holds for some $R \in (0, D]$, then $\ell = \ell_a^\star$ on $[0, R]$. Hence $\ell_a^\star$ is the unique pointwise maximal feasible profile on every distance cap.*

The proof integrates $B'(z) \leq -2az$ from $B(0) = 1$ to obtain $B(z) \leq 1 - az^2$. Equality determines $z_a$; integrating once more determines $\ell_a^\star$. Hence, for every $D > 0$,

$$\Delta(\ell; D) \geq \Delta(\ell_a^\star; D) = \frac{D^2}{2} - \int_0^D z_a(t)\,\mathrm{d}t. \tag{24}$$

At $\Lambda = 0$, the envelope is critical log-cosh and is itself full A3w. The necessary transverse constraint therefore yields an exact full-A3w optimum, as follows.

**Corollary 8** (Least-distorting A3w repair and cap rigidity)**.** *Let $d \geq 3$ and $D > 0$. Among all $\ell \in \mathcal{H}_\kappa$ for which $c_\ell$ satisfies A3w at every source–target pair of distance at most $D$, the critical log-cosh profile*

$$\ell_1(r) = \frac{1}{\kappa^2}\log\cosh(\kappa r) \tag{25}$$

*minimizes $\Delta(\ell; D)$. The minimum is*

$$\Delta(\ell_1; D) = \frac{D^2}{2} - \frac{1}{\kappa^2}\log\cosh(\kappa D). \tag{26}$$

*Every feasible profile attaining this value agrees with $\ell_1$ on $[0, D]$.*

**Corollary 9** (Flat-limit collapse of the curvature repair)**.** *Fix $D < \infty$. For the critical repair $\ell_1^\kappa(r) = \kappa^{-2}\log\cosh(\kappa r)$, uniformly for $0 \leq r \leq D$ as $\kappa \downarrow 0$,*

$$\ell_1^\kappa(r) = \frac{r^2}{2} - \frac{\kappa^2 r^4}{12} + O(\kappa^4 D^6), \tag{27}$$

$$\Delta(\ell_1^\kappa; D) = \frac{\kappa^2 D^4}{12} - \frac{\kappa^4 D^6}{45} + O(\kappa^6 D^8), \tag{28}$$

$$\mathsf{K}(\ell_1^\kappa; D) = \cosh(\kappa D) = 1 + \frac{\kappa^2 D^2}{2} + O(\kappa^4 D^4). \tag{29}$$

*Hence $\ell_1^\kappa \to r^2/2$, the least A3w fidelity price tends to zero, and the condition number tends to one. In the Euclidean limit, squared distance itself has zero MTW tensor, so the A3w repair problem collapses.*

## 6 THE SENSITIVITY COST OF THE OPTIMAL TRANSVERSE REPAIR

The closest repair need not be the easiest to invert. The mixed Hessian $M = -D^2_{xy}c$ converts endpoint changes into momentum changes; $M^{-1}$ gives the local inverse response. Its condition number measures directional imbalance, while $1/\sigma_{\min}(M)$ measures the largest absolute amplification.

**Theorem 10** (Full MTW margin and conditioning of the envelope extremizer)**.** *Let $d \geq 2$, $a > 0$, and $c^\star_a = \ell^\star_a \circ d_\kappa$. For $z = z_a(r)$, set*

$$H(z) = 1 + az^2, \qquad N(z) = (1 - az^2)^2 - \kappa^2 z^2. \tag{30}$$

*Then $N(z_a(r)) > 0$ at every finite $r$, the MTW tensor of $c^\star_a$ is strictly positive at every finite source–target pair, and $c^\star_a$ satisfies A3s on every bounded distance cap. Its mixed-Hessian singular values are*

$$\rho^\star_a(r) = \frac{N(z)}{H(z)}, \qquad \tau^\star_a(r) = \sqrt{N(z)} \quad \textit{with multiplicity } d-1. \tag{31}$$

*The weakest full MTW direction at distance $r$ is radial–transverse and has value*

$$m^\star_a(r) = \frac{3aN(z)}{H(z)^2}. \tag{32}$$

*Consequently, for every $D > 0$,*

$$m(\ell^\star_a; D) = \frac{3aN(z_a(D))}{H(z_a(D))^2}, \qquad \mathsf{K}(\ell^\star_a; D) = \frac{H(z_a(D))}{\sqrt{N(z_a(D))}}, \tag{33}$$

*and the exact law*

$$\boxed{m(\ell^\star_a; D)\, \mathsf{K}(\ell^\star_a; D)^2 = 3a = \Lambda.} \tag{34}$$

*holds. Both quantities in* (33) *are extrema over the full distance cap.*

The weakest direction in Theorem 10 is a radial–transverse pair; the smallest mixed-Hessian singular value is a single radial direction. The two notions are distinct. Although the envelope matches the fourth-order coefficient of $q^2\kappa^2 = \kappa^2 + \Lambda$, its full margin on $D > 0$ is $\Lambda/(\mathsf{K}^\star_D)^2 < \Lambda$. Its maximum absolute inverse gain is $\mathsf{G}(\ell^\star_a; D) = H(z_a(D))/N(z_a(D))$. Distance selection should check both $\mathsf{K}$ and $\mathsf{G}$; evaluate $N$ using the stable identity (145).

**Corollary 11** (Unattainable lower bound under a full MTW margin)**.** *Let $d \geq 3$, $D > 0$, and $\Gamma \geq 0$. If $\ell \in \mathcal{H}_\kappa$ satisfies*

$$m(\ell; D) \geq \Gamma, \tag{35}$$

*then*

$$\Delta(\ell; D) \geq \Delta(\ell^\star_{\Gamma/3}; D). \tag{36}$$

*For $\Gamma = 0$, equality is attained by critical log-cosh. For $\Gamma > 0$, every feasible profile has strictly larger distortion than the right-hand side of* (36)*. This is a pointwise statement over profiles; it does not imply strict separation of their infimum from the displayed lower bound.*

## 7 PAYING FOR A UNIFORM FULL MARGIN

The envelope's full margin decreases with distance. To keep a prescribed positive margin everywhere, we return to the log-cosh family in (17). Its extra cost deficit gives an explicit feasible comparison for the harder repair problem.

**Theorem 12** (Exact MTW phase diagram of the calibrated log-cosh family)**.** *For $q > 0$, let $c_q(x, y) = \ell_q(d_\kappa(x, y))$ with* (17)*. Then:*

1. *if $0 < q < 1$, $c_q$ violates A3w at the diagonal;*
2. *if $q = 1$, the MTW tensor vanishes identically, so $c_1$ satisfies A3w;*
3. *if $q > 1$, then for every finite source–target pair and every $u, w$ with $\langle u, w\rangle = 0$,*

$$\mathrm{MTW}_x(u, -D_x c_q(x, y), w) \geq \Gamma_q \|u\|^2 \|w\|^2, \qquad \Gamma_q := \kappa^2(q^2 - 1). \tag{37}$$

*Equality holds at the diagonal. Thus $\Gamma_q$ is the exact full-space MTW margin of $c_q$.*

The critical case $q = 1$ removes the curvature obstruction exactly. Increasing $q$ makes the cost flatter than squared distance and provides a positive full margin at every distance. The proof in Section C controls all directional coefficients through a positive-coefficient series.

For finite-radius conditioning, let $D$ bound all source–target distances. The mixed derivative $-D^2_{xy}c_q$ has one radial and $d-1$ transverse singular values,

$$\rho_q(r) = \operatorname{sech}^2(q\kappa r), \qquad \tau_q(r) = \frac{\tanh(q\kappa r)}{q\sinh(\kappa r)}. \tag{38}$$

For $q \geq 1$, $\rho_q(r) \leq \tau_q(r)$. Define

$$\Delta_q(D) := \sup_{0\leq r\leq D}\left[\frac{r^2}{2} - \ell_q(r)\right] = \frac{D^2}{2} - \frac{\log\cosh(q\kappa D)}{q^2\kappa^2}, \tag{39}$$

$$\mathsf{K}_q(D) := \sup_{0\leq r\leq D} \operatorname{cond}(-D^2_{xy}c_q) = \frac{\sinh(2q\kappa D)}{2q\sinh(\kappa D)}, \tag{40}$$

$$\mathsf{G}_q(D) := \sup_{0\leq r\leq D} \|D(c\text{-}\exp^q_x)\|_{\mathrm{op}} = \cosh^2(q\kappa D). \tag{41}$$

The inverse gain $\mathsf{G}_q$ differs from the condition number $\mathsf{K}_q$ and the Lee–Li function $A(z)$.

**Theorem 13** (Optimal log-cosh member at a prescribed full margin)**.** *Fix $D > 0$. The exact margin $\Gamma_q = \kappa^2(q^2-1)$ and all three prices in* (39)–(41) *are strictly increasing for $q \geq 1$. Hence, for a prescribed $\Gamma \geq 0$, $q_\Gamma = \sqrt{1+\Gamma/\kappa^2}$ is the unique family member with $\Gamma_q \geq \Gamma$ that simultaneously minimizes $\Delta_q(D)$, $\mathsf{K}_q(D)$, and $\mathsf{G}_q(D)$.*

Theorem 13 is a unique optimum within the calibrated $q$-family. The universal extremal statement is instead Theorem 7: if $\Lambda = \Gamma_q$, then $c_q$ is feasible for its transverse constraint and

$$\ell_q(r) \leq \ell^\star_{\Lambda/3}(r), \qquad r \geq 0, \tag{42}$$

with strict inequality for $q > 1$ and $r > 0$. The additional distortion in $c_q$ buys the distance-independent full margin (37); the least-distorting path instead follows the exact decay law (34).

**Corollary 14** (Analytic bracket for the full-margin problem)**.** *Let $d \geq 3$, $D > 0$, and $\Gamma \geq 0$, and set $q_\Gamma = \sqrt{1+\Gamma/\kappa^2}$. Then*

$$\Delta(\ell^\star_{\Gamma/3}; D) \leq V_{\mathrm{full}}(\Gamma, D) \leq \Delta_{q_\Gamma}(D). \tag{43}$$

*For $\Gamma > 0$, the lower endpoint is unattainable and, as $D \downarrow 0$,*

$$\Delta_{q_\Gamma}(D) - \Delta(\ell^\star_{\Gamma/3}; D) = \frac{\Gamma(3\kappa^2+4\Gamma)}{270}D^6 + O(D^8). \tag{44}$$

*Thus the two bounds agree through fourth order; their first difference is proportional to $D^6$. This local agreement need not persist at large distances.*

The remaining question is whether a full-margin repair can match the envelope beyond fourth order. If the fourth-order price is minimal, the next theorem forces an additional sixth-order deficit, including in $d = 2$. Let $\mathcal{H}^{(6)}_\kappa \subset \mathcal{H}_\kappa$ contain the profiles with an even $C^6$ extension and write $c_6(\ell) = \ell^{(6)}(0)/720$. Define

$$c_{6,\perp}(\Gamma) := \frac{\kappa^4}{45} + \frac{\Gamma\kappa^2}{18} + \frac{\Gamma^2}{27}, \qquad \mathcal{P}_6(\ell;\Gamma) := c_{6,\perp}(\Gamma) - c_6(\ell), \tag{45}$$

where $c_{6,\perp}(\Gamma)$ is the sixth-order coefficient of the matched transverse envelope $\ell^\star_{\Gamma/3}$.

**Theorem 15** (Sixth-order full-MTW jet price)**.** *Let $d \geq 2$, $D > 0$, and $\Gamma > 0$. If $\ell \in \mathcal{H}^{(6)}_\kappa$ satisfies*

$$m(\ell; D) \geq \Gamma, \qquad \mathsf{F}_0(\ell) = \kappa^2 + \Gamma, \tag{46}$$

*then*

$$c_6(\ell) \leq \frac{\kappa^4}{45} + \frac{5\Gamma\kappa^2}{108} + \frac{2\Gamma^2}{81}, \qquad \mathcal{P}_6(\ell;\Gamma) \geq \frac{\Gamma(3\kappa^2+4\Gamma)}{324}. \tag{47}$$

*Let $P^\star_6(\Gamma, D)$ be the infimum of $\mathcal{P}_6$ over the class in* (46)*. The feasible profile $\ell_{q_\Gamma}$ gives*

$$\frac{\Gamma(3\kappa^2+4\Gamma)}{324} \leq P^\star_6(\Gamma, D) \leq \frac{\Gamma(3\kappa^2+4\Gamma)}{270}. \tag{48}$$

*Its width is $\Gamma(3\kappa^2+4\Gamma)/1620$, one sixth of the full envelope-to-log-cosh sixth-order gap.*

The proof in Section C uses the radial–transverse candidate $\beta$: fourth-order saturation forces its $r^2$ coefficient to be nonpositive. This is a jet constraint, not a uniform finite-cap gap without uniform control of higher-order remainders.

## 8 EXPERIMENTS: TASK CHANGE AND NUMERICAL PRICE

We test three consequences of the design question: whether repair changes an actual optimal coupling, how much original-objective value it sacrifices, and whether inversion exhibits the predicted sensitivity. The new three-target experiment includes a complete CPU implementation and disjoint pilot and evaluation seeds. The earlier M0–M3 exports supply broader geometry, inverse, and matching checks. Appendix H reports the full evidence, including null application results.

**Evaluation under the original task.** For exact minimizers $\pi_c, \pi_0$ over the same coupling set, the cost-deficit bound $0 \le c_0 - c \le \Delta(c; D)$ implies

$$0 \le R_0(\pi_c) := \int c_0 \, \mathrm{d}\pi_c - \int c_0 \, \mathrm{d}\pi_0 \le \Delta(c; D). \tag{49}$$

This is an objective guarantee, not a bound on assignment changes. We use unregularized LP solutions for this comparison. Sinkhorn results are reported separately because entropy changes the optimization problem (Cuturi, 2013; Weed, 2018).

**A fixed task with a visible structural change.** In $\mathbb{H}^3_\kappa$, sources are the exponential image of a unit ball in normal coordinates. The three targets are $\exp_o(-e_1), o, \exp_o(e_1)$, with fixed masses $(15/32, 1/16, 15/32)$. Scrambled Sobol points are reflected across $x_1 = 0$, and the origin is appended as a probe. Source weights are $1/n$; the LP may split mass. We freeze the construction after pilot seeds 0–9 and evaluate seeds 100–109 at $n \in \{513, 1025, 2049\}$ and $\kappa \in \{0, 0.5, 1\}$. Squared distance, critical log-cosh, and $q = 1.5$ log-cosh give 270 LP solves, all successful. Costs share $\ell''(0) = 1$; no embedding or registration is trained.

At $\kappa = 1$, squared distance sends no origin mass to the central target, whereas both repairs send all of it there, at every seed and resolution. Figure 1 displays cells induced by the optimal dual of the first evaluation seed. The effect extends beyond the single probe: at $n = 2049$, the mean central-target share among sources with $r \le 0.2$ rises from zero to $27.8\%$ for the critical repair (Table 1). The larger $q$ produces a further change and a larger original-task regret. These are different margin levels, not a comparison at the same positive full margin.

Table 1: **Transport changes at a measurable task cost.** Three-target OT at $\kappa = 1$, $n = 2049$, ten evaluation seeds. Origin share is identical across seeds; neighborhood share is a mean; regret is mean $\pm$ standard deviation. All costs use the same points and marginals.

| Cost | Origin $\to$ center | $r \le 0.2 \to$ center | $10^3 R_0$ |
|---|---|---|---|
| Squared distance | $0\%$ | $0\%$ | $0$ |
| Critical log-cosh | $100\%$ | $27.80\%$ | $1.017 \pm 0.162$ |
| Log-cosh, $q = 1.5$ | $100\%$ | $54.49\%$ | $3.253 \pm 0.269$ |

The maximum marginal residual is $2.49 \times 10^{-14}$ and maximum primal–dual gap is $1.22 \times 10^{-15}$, far below the observed differences. An independent sorting solution of this symmetric LP agrees with all 270 results to approximately $10^{-15}$. At $\kappa = 0$, all three costs and solutions coincide. At $\kappa = 0.5$, squared distance also sends the origin to the center. Thus the structural change is reproducible in the declared geometry, not universal over all negative curvatures. Atomic targets and a two-dimensional cell section do not test smooth Monge-map regularity.

**Measured inverse sensitivity.** The earlier inversion experiment uses $d = 3$, $\kappa = 1$, $D \in \{0.5, 1, 2, 3\}$, and $\Lambda \in \{0.1, 0.25, 0.5, 1\}$. It compares squared distance, critical log-cosh, $\ell^\star_{\Lambda/3}$, and $q = \sqrt{1 + \Lambda/\kappa^2}$. The latter two match their fourth-order coefficients. For $M = -D^2_{xy} c$, a unit momentum perturbation $w$ has predicted gain $\|M^{-1} w\|$, with maximum $1/\sigma_{\min}(M)$. Across

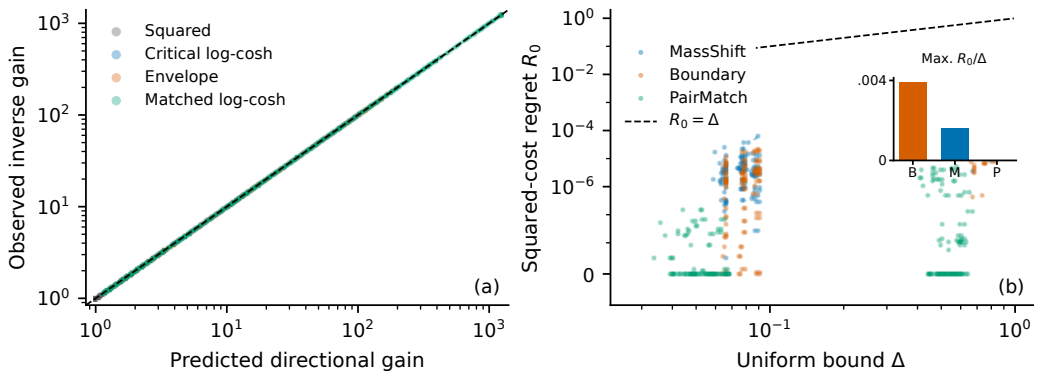


Figure 2: **Numerical sensitivity and objective fidelity.** Left: predicted and measured directional inverse gain (25,000 comparisons). Right: exact-OT regret versus the uniform bound (720 evaluations on 240 instances). Dashed lines indicate equality. Inset: maximum $R_0/\Delta$ for Boundary, MassShift, and PairMatch. Small regret can coexist with changed assignments; these experiments test different consequences of the cost.

25,000 exported comparisons, measured $c$-exponential inversion gives median, 95th-percentile, and maximum relative errors $6.32 \times 10^{-5}$, $1.14 \times 10^{-3}$, and $1.60 \times 10^{-2}$ (Figure 2). This measures local inverse sensitivity; it does not identify $\mathsf{K}$ with a Sinkhorn convergence rate.

**Choosing a usable distance range.** The geometry also provides an explicit design check. For critical log-cosh, budgets $\mathsf{K} \le K_{\max}$ and $\mathsf{G} \le G_{\max}$, with both thresholds at least one, require

$$D \le \kappa^{-1} \min\{\operatorname{arcosh} K_{\max}, \operatorname{arcosh} \sqrt{G_{\max}}\}. \tag{50}$$

The cap must bound every admissible source–target distance; an empirical distance percentile cannot replace it.

**Broader transport and matching checks.** Exact OT on 256-point MassShift, Boundary, and PairMatch clouds crosses $D \in \{1, 2\}$, $\Lambda \in \{0.25, 0.5\}$, and 20 instances per cell. All 720 repaired-cost evaluations satisfy (49) within tolerance; the maximum $R_0/\Delta$ is $0.00394$, so the bound is conservative on these data. Across 80 paired PairMatch instances, P@1 is $55.752\%$ for squared distance and $55.747\%$ for critical log-cosh; its paired difference interval includes zero. Flat-limit, finite-candidate, and sixth-order audits appear in the appendix. WordNet and OAEI Anatomy matching do not establish a robust downstream benefit. Their low recovery, registration behavior, and numerical exclusions are reported alongside the other experiments; MTW repair is not an accuracy guarantee.

## 9 Discussion and Conclusion

The cost of obtaining regularity has two distinct parts: changing the objective and increasing inverse sensitivity. In $d \ge 3$, critical log-cosh is the closest normalized radial full-A3w repair of squared distance. The transverse envelope extends this optimum to positive transverse floors, with an exact relation between full margin and conditioning. Finite directional candidates and a sixth-order obstruction narrow the positive full-margin problem without identifying its finite-cap infimum.

The three-target experiment makes the task change visible in an optimal coupling, while perturbation and exact-OT experiments quantify inverse sensitivity and original-objective regret. These conclusions concern normalized radial costs and the stated distance caps. The sharp envelope requires $d \ge 3$; the sixth-order bound assumes fourth-order saturation. Map regularity needs the additional PDE hypotheses, and matching accuracy is a separate empirical endpoint. The application runs reported in the appendix do not establish a downstream advantage. The resulting theory provides a principled cost comparison: a regular repair has a quantifiable price, which must be assessed against the task it is meant to preserve.

### AI USE STATEMENT

Generative AI tools assisted with literature search, mathematical formulation, proof drafting and checking, experimental design, implementation, interpretation of results, figures, and language editing. The authors take responsibility for the mathematical claims, experimental records, and all AI-assisted material in the paper and supplement.

### ETHICS STATEMENT

The work uses no human-subject or private data. Empirical evaluation uses public lexical and ontology benchmarks under their licenses; numerical audits and discrete matching results are not treated as proofs about continuous Monge maps.

### REPRODUCIBILITY STATEMENT

The appendices give complete proofs, experimental settings, all reported comparison families, and the scope of the empirical conclusions. The companion archive includes exported records and figure-generation scripts for the original experiments, together with a complete CPU implementation, seeds, and numerical outputs for the three-target experiment. The latter can be rerun from its source generator. The original embedding and registration pipelines are represented by their exported results; their training checkpoints and complete execution logs are not included. Numerical exclusions and unsuccessful application endpoints are retained.

## REFERENCES

David Alvarez-Melis, Youssef Mroueh, and Tommi Jaakkola. Unsupervised hierarchy matching with optimal transport over hyperbolic spaces. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1606–1617, 2020.

Ines Chami, Rex Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Wei-Ting Chiu, Pei Wang, and Patrick Shafto. Discrete probabilistic inverse optimal transport. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 3925–3946, 2022. URL https://proceedings.mlr.press/v162/chiu22b.html.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Marco Cuturi and David Avis. Ground metric learning. *Journal of Machine Learning Research*, 15 (17):533–564, 2014. URL https://jmlr.org/papers/v15/cuturi14a.html.

Guido De Philippis and Alessio Figalli. The Monge–Ampère equation and its link to optimal transportation. *Bulletin of the American Mathematical Society*, 51(4):527–580, 2014. URL https://arxiv.org/abs/1310.6167.

Guido De Philippis and Alessio Figalli. Partial regularity for optimal transport maps. *Publications Mathématiques de l'IHÉS*, 121:81–112, 2015. URL https://arxiv.org/abs/1209.5640.

Zlatan Dragisic, Valentina Ivanova, Huanyu Li, and Patrick Lambrix. Experiences from the anatomy track in the ontology alignment evaluation initiative. *Journal of Biomedical Semantics*, 8(1):56, 2017. doi: 10.1186/s13326-017-0166-5.

Alessio Figalli, Young-Heon Kim, and Robert J. McCann. Hölder continuity and injectivity of optimal maps. *Archive for Rational Mechanics and Analysis*, 209(3):747–795, 2013. doi: 10.1007/s00205-013-0629-5.

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Young-Heon Kim and Robert J. McCann. Continuity, curvature, and the general covariance of optimal transportation. *Journal of the European Mathematical Society*, 12(4):1009–1040, 2010. doi: 10.4171/JEMS/221.

Paul W. Y. Lee and Jiayong Li. New examples satisfying Ma–Trudinger–Wang conditions. *SIAM Journal on Mathematical Analysis*, 44(1):61–73, 2012. doi: 10.1137/110820543.

Jiakun Liu, Neil S. Trudinger, and Xu-Jia Wang. Interior $C^{2,\alpha}$ regularity for potential functions in optimal transportation. *Communications in Partial Differential Equations*, 35(1):165–184, 2010. doi: 10.1080/03605300903236609.

Grégoire Loeper. On the regularity of solutions of optimal transportation problems. *Acta Mathematica*, 202(2):241–283, 2009. doi: 10.1007/s11511-009-0037-8.

Xi-Nan Ma, Neil S. Trudinger, and Xu-Jia Wang. Regularity of potential functions of the optimal transportation problem. *Archive for Rational Mechanics and Analysis*, 177(2):151–183, 2005. doi: 10.1007/s00205-005-0362-9.

Ashok Vardhan Makkuva, Amirhossein Taghvaei, Jason D. Lee, and Sewoong Oh. Optimal transport mapping via input convex neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6672–6681, 2020.

Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *The Annals of Statistics*, 52(3):966–998, 2024. doi: 10.1214/24-AOS2379.

George A. Miller. WordNet: A lexical database for english. *Communications of the ACM*, 38(11): 39–41, 1995. doi: 10.1145/219717.219748.

Yan Bin Ng and Xianfeng Gu. Hyperbolic optimal transport. *Mathematics, Computation and Geometry of Data*, 4(2):75–139, 2025. doi: 10.4310/MCGD.250905234225.

Du Nguyen. Families of costs with zero and nonnegative MTW tensor in optimal transport and the $c$-divergences. *Information Geometry*, 9:31–80, 2026. doi: 10.1007/s41884-025-00186-z.

Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Maximilian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3779–3788, 2018.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019. doi: 10.1561/2200000073.

Aram-Alexandre Pooladian, Vincent Divol, and Jonathan Niles-Weed. Minimax estimation of discontinuous optimal transport maps: The semi-discrete case. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 28128–28150, 2023.

Frederic Sala, Chris De Sa, Albert Gu, and Christopher Ré. Representation tradeoffs for hyperbolic embeddings. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4460–4469, 2018. URL `https://proceedings.mlr.press/v80/sala18a.html`.

Bernhard Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, 2019. doi: 10.1137/16M1106018.

Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel. Large-scale optimal transport and mapping estimation. In *International Conference on Learning Representations*, 2018.

Sachin Shivakumar, Georgiy A. Bondar, Gabriel Khan, and Abhishek Halder. Sum-of-squares programming for Ma–Trudinger–Wang regularity of optimal transport maps. *Journal of Optimization Theory and Applications*, 209(3):70, 2026. doi: 10.1007/s10957-026-03015-0.

Andrew M. Stuart and Marie-Therese Wolfram. Inverse optimal transport. *SIAM Journal on Applied Mathematics*, 80(1):599–619, 2020. doi: 10.1137/19M1261122.

Neil S. Trudinger and Xu-Jia Wang. On the second boundary value problem for monge–ampère type equations and optimal transportation. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*, 8(1):143–174, 2009.

Jonathan Weed. An explicit analysis of the entropic penalty in linear programming. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 1841–1855, 2018. URL https://proceedings.mlr.press/v75/weed18a.html.

## A PROOF OF THE DIAGONAL IDENTITY AND LOCAL TRILEMMA

We give a self-contained reduction to the scalar quantities in Lee & Li (2012). Let $g = (\ell')^{-1}$ and let $z = \ell'(r)$. On curvature $-\kappa^2$, define

$$A(z) = \frac{1}{g'(z)}, \qquad B(z) = \kappa z \coth(\kappa g(z)). \tag{51}$$

For $z > 0$, introduce

$$\alpha(z) = \frac{z^2 A''(z) + 6(A(z) - B(z)) - 4z(A'(z) - B'(z))}{z^2}, \tag{52}$$

$$\beta(z) = \frac{zA'(z) - 2(A(z) - B(z))}{z^2}, \tag{53}$$

$$\gamma(z) = B''(z), \qquad \delta(z) = \frac{B'(z)}{z}. \tag{54}$$

The continuous values at zero are defined by limits. For orthogonal $u, w$, the Lee–Li formula is

$$\mathrm{MTW} = -\frac{3}{2}\big[\alpha|u_0|^2|w_0|^2 + \beta|u_0|^2|w_1|^2 + \gamma|u_1|^2|w_0|^2 + \delta|u_1|^2|w_1|^2\big], \tag{55}$$

where the subscripts denote components parallel and orthogonal to the radial covector.

*Proof of Proposition 3.* Orthogonality gives $x + y \le 1$. When $d = 2$, an orthonormal pair is a basis, so Parseval's identity for the radial unit vector gives $x + y = 1$. Hence

$$Q_z(x, 1 - x) = \gamma + hx - sx^2, \qquad 0 \le x \le 1. \tag{56}$$

The endpoints give $\gamma$ and $\beta$. An additional maximum occurs only when $s > 0$ and the vertex $x = h/(2s)$ belongs to $[0, 1]$; its value is $\gamma + h^2/(4s)$. This proves the $d = 2$ candidate set.

For $d \ge 3$, every point of the triangle

$$\mathcal{T} = \{(x, y) : x \ge 0,\ y \ge 0,\ x + y \le 1\} \tag{57}$$

is realized by a unit orthogonal pair. Indeed, for $x < 1$ and an orthonormal frame $e_0, e_1, e_2$ with $e_0$ radial, one may take

$$u = \sqrt{x}\, e_0 + \sqrt{1 - x}\, e_1, \tag{58}$$

$$w = \sqrt{y}\, e_0 - \sqrt{\frac{xy}{1 - x}}\, e_1 + \sqrt{\frac{1 - x - y}{1 - x}}\, e_2, \tag{59}$$

with the remaining boundary point obtained by continuity. Substitution in (55) gives (10). For fixed $x$, $Q_z(x, y)$ is affine in $0 \le y \le 1 - x$, so its maximum occurs on $y = 0$ or $y = 1 - x$. The first edge contributes $\delta$ and $\beta$; the hypotenuse is (56) and contributes $\gamma$, $\beta$, and the same possible vertex. This proves (12). Since $\mathrm{MTW} = -3Q/2$, minimizing MTW reverses the directional maximum and yields the remaining identities. The statement at $z = 0$ follows by continuity. □

*Proof of Lemma 5.* Write $a = \ell^{(4)}(0)/24$. Evenness and normalization give

$$\ell(r) = \frac{r^2}{2} + ar^4 + o(r^4), \qquad \ell'(r) = r + 4ar^3 + o(r^3). \tag{60}$$

Series inversion yields

$$g(z) = z - 4az^3 + o(z^3), \qquad A(z) = \frac{1}{g'(z)} = 1 + 12az^2 + o(z^2). \tag{61}$$

Using $t \coth t = 1 + t^2/3 + o(t^2)$,

$$B(z) = \frac{z}{g(z)}\left[\kappa g(z) \coth(\kappa g(z))\right] \tag{62}$$

$$= 1 + \left(4a + \frac{\kappa^2}{3}\right) z^2 + o(z^2) \tag{63}$$

$$= 1 + \frac{\ell^{(4)}(0) + 2\kappa^2}{6} z^2 + o(z^2). \tag{64}$$

Substitution in (54) gives the common limit

$$\alpha(0) = \beta(0) = \gamma(0) = \delta(0) = \frac{\ell^{(4)}(0) + 2\kappa^2}{3}. \tag{65}$$

At the diagonal, isotropy and $u \perp w$ make the bracket in (55) equal to this common coefficient for unit $u, w$. Therefore

$$\mathsf{R}_0(\ell) = -\frac{3}{2}\frac{\ell^{(4)}(0) + 2\kappa^2}{3} = -\kappa^2 - \frac{1}{2}\ell^{(4)}(0). \tag{66}$$

□

*Proof of Theorem 6.* Taylor expansion gives

$$\mathsf{F}_0(\ell) = \lim_{r\downarrow 0} \frac{12[-\ell^{(4)}(0)r^4/24 + o(r^4)]}{r^4} = -\frac{1}{2}\ell^{(4)}(0), \tag{67}$$

$$\mathsf{N}_0(\ell) = \lim_{r\downarrow 0} \frac{2[-\ell^{(4)}(0)r^2/2 + o(r^2)]}{r^2} = -\ell^{(4)}(0). \tag{68}$$

Combining these identities with Lemma 5 proves (15) and (16). For (17),

$$\ell_q(r) = \frac{r^2}{2} - \frac{q^2\kappa^2}{12}r^4 + O(r^6), \tag{69}$$

so $\ell_q^{(4)}(0) = -2q^2\kappa^2$. Taking $q = q_\Gamma$ attains equality. Conversely, equality fixes $\ell^{(4)}(0) = -2(\kappa^2 + \Gamma)$, which is equivalent to (19). □

## B PROOFS FOR THE NONLOCAL EXTREMIZER

*Proof of Theorem 7.* The constraint (20) is $B'(z)/z \le -2a$. The normalization of $\ell$ gives $g(z) = z + o(z)$ and hence $B(0) = 1$. For $0 \le z \le \ell'(D)$, integration from zero yields

$$B(z) \le 1 - az^2. \tag{70}$$

For $0 \le r \le D$, set $z = \ell'(r)$. The identity $g(z) = r$ turns (70) into

$$az^2 \tanh(\kappa r) + \kappa z - \tanh(\kappa r) \le 0. \tag{71}$$

The left-hand side is strictly increasing for $z \ge 0$. Its positive root is exactly $z_a(r)$ in (21), proving the derivative comparison. Integration gives $\ell \le \ell_a^\star$. Since $z_a(r) \le \tanh(\kappa r)/\kappa \le r$, the squared-cost upper bound in (23) follows as well.

To verify attainability, define on its maximal positive interval

$$g_a(z) = \frac{1}{\kappa}\operatorname{artanh}\left(\frac{\kappa z}{1 - az^2}\right). \tag{72}$$

The interval ends at the positive solution of $1 - az^2 = \kappa z$. The function $g_a$ is odd, analytic, strictly increasing, and maps this interval onto $\mathbb{R}$. Direct differentiation gives

$$g_a'(z) = \frac{1 + az^2}{(1 - az^2)^2 - \kappa^2 z^2} > 0. \tag{73}$$

Its inverse is the odd function $z_a(r)$, so integrating $z_a$ produces an even smooth profile with $\ell_a^\star(0) = 0$, $(\ell_a^\star)''(0) = 1$, and positive first and second derivatives for $r > 0$. Moreover,

$$B_a^\star(z) = \kappa z \coth(\kappa g_a(z)) = 1 - az^2, \tag{74}$$

so $-3\delta/2 = 3a = \Lambda$ identically.

If $\ell(R) = \ell_a^\star(R)$, the continuous nonnegative function $z_a - \ell'$ has zero integral on $[0, R]$ and therefore vanishes there. The common value at zero then gives interval rigidity. □

*Proof of Corollary 8.* In dimension $d \geq 3$, the Lee–Li criterion includes $\delta \leq 0$ as a necessary condition for A3w. Apply Theorem 7 with $\Lambda = a = 0$. Its extremizer is $\ell_1(r) = \kappa^{-2} \log \cosh(\kappa r)$, whose MTW tensor vanishes identically. It is therefore feasible. The distortion formula and uniqueness on $[0, D]$ follow from (24) and interval rigidity. □

*Proof of Theorem 10.* For brevity write $H = 1 + az^2$ and $N = (1 - az^2)^2 - \kappa^2 z^2$. Equations (73) and (51) give

$$A(z) = \frac{N}{H}, \qquad B(z) = 1 - az^2. \tag{75}$$

At every finite $r$, the argument of $\operatorname{artanh}$ in (72) is strictly below one, hence $N > 0$. Substitution of (75) into (54) yields

$$\gamma(z) = \delta(z) = -2a, \tag{76}$$

$$\beta(z) = -\frac{2aN}{H^2}, \tag{77}$$

$$\alpha(z) + \delta(z) = 2\beta(z) - \frac{2a(4a + \kappa^2)z^2(1 + 5az^2)}{H^3}. \tag{78}$$

All inequalities in the Lee–Li criterion are strict when $a > 0$ and $z$ is finite. The explicit positive cap margin below is uniform for $0 \leq r \leq D$, proving A3s on every bounded distance cap. Notice that $0 < N/H^2 < 1$ away from the diagonal, so $\delta = \gamma < \beta < 0$ and $\alpha + \delta < 2\beta < 0 < 2\sqrt{\beta\gamma}$. At the diagonal, the common coefficient is $-2a < 0$, so the strict condition also holds there.

We next identify the weakest direction, not only the sign. For unit orthogonal $u, w$, set $x = |u_0|^2$ and $y = |w_0|^2$. Cauchy–Schwarz and $\langle u, w \rangle = 0$ give $xy \leq (1 - x)(1 - y)$, hence $x + y \leq 1$. The bracket $Q$ in (55) can be written, using $\gamma = \delta$, as

$$Q = \delta + (\beta - \delta)x + (\alpha - \beta)xy. \tag{79}$$

Let $p = \beta - \delta \geq 0$. Equation (78) gives $\alpha - \beta \leq p$, and therefore

$$Q \leq \delta + p(x + xy) \leq \delta + p\,x(2 - x) \leq \delta + p = \beta. \tag{80}$$

Equality is attained by a radial $u$ and a transverse $w$. Thus the pointwise full MTW margin is $-3\beta/2$, proving (32).

The general mixed-spectrum formula in Lemma 18 and (75) give the radial singular value $N/H$. To compute the transverse value, use

$$\tanh(\kappa r) = \frac{\kappa z}{1 - az^2}, \qquad \sinh(\kappa r) = \frac{\kappa z}{\sqrt{N}}, \tag{81}$$

which yields $\tau_a^\star = \sqrt{N}$. Since $H^2 - N = (4a + \kappa^2)z^2 \geq 0$, the radial singular value is the smaller one and the pointwise condition number is $H/\sqrt{N}$.

It remains to place both extrema at $r = D$. With $t = z^2$,

$$\frac{N}{H^2} = 1 - \frac{(4a + \kappa^2)t}{(1 + at)^2}. \tag{82}$$

The function $t/(1+at)^2$ is strictly increasing while $at < 1$. The latter follows from $1 - az^2 > \kappa z > 0$. Since $z_a(r)$ is strictly increasing, $N/H^2$ decreases with $r$. Hence the full margin is minimized and the condition number is maximized at $D$. Multiplying the two expressions in (33) proves (34). □

*Proof of Corollary 11.* When $d \geq 3$, choose $u$ and $w$ to be orthogonal unit vectors tangent to the distance sphere. Their MTW value is $-3\delta(z)/2$. Hence $m(\ell; D) \geq \Gamma$ implies the transverse constraint with parameter $\Gamma$ on $0 \leq z \leq \ell'(D)$. Applying Theorem 7 gives

$$\Delta(\ell; D) \geq \frac{D^2}{2} - \ell(D) \geq \frac{D^2}{2} - \ell^\star_{\Gamma/3}(D) = \Delta(\ell^\star_{\Gamma/3}; D). \tag{83}$$

Suppose $\Gamma > 0$ and equality holds. Both inequalities above must then be equalities, so $\ell(D) = \ell^\star_{\Gamma/3}(D)$. Interval rigidity gives $\ell = \ell^\star_{\Gamma/3}$ on $[0, D]$. With $a = \Gamma/3$, Theorem 10 then gives

$$m(\ell; D) = \frac{\Gamma}{\mathsf{K}(\ell^\star_a; D)^2} < \Gamma, \tag{84}$$

because $D > 0$ and $H(z_a(D))^2 - N(z_a(D)) = (4a + \kappa^2)z_a(D)^2 > 0$. This contradicts the assumed full margin. □

*Proofs of Corollaries 21 and 22.* From (72),

$$\tanh(\kappa r) = \frac{\kappa z}{1 - az^2}, \qquad N(z) = (1 - az^2)^2 \operatorname{sech}^2(\kappa r). \tag{85}$$

As $r \to \infty$, $z_a(r) \to \bar{z}_a$, where $1 - a\bar{z}_a^2 = \kappa \bar{z}_a$. Substitution in (33), followed by $\cosh(\kappa D) \sim e^{\kappa D}/2$ and $\operatorname{sech}^2(\kappa D) \sim 4e^{-2\kappa D}$, proves (136). Equation (137) follows directly from (40), (41), and the standard asymptotics of $\sinh$ and $\cosh$. □

## C Proof of the complete $q$-phase diagram

Write $\bar{r} = \kappa r$ and $\bar{\ell}(\bar{r}) = \kappa^2 \ell(\bar{r}/\kappa)$. This identifies $\mathbb{H}^d_\kappa$ with the curvature-$-1$ model up to the metric homothety $g_\kappa = \kappa^{-2} g_1$. For unit vectors in $g_\kappa$, the normalized MTW value is $\kappa^2$ times its curvature-$-1$ counterpart. It is therefore enough to prove the coefficient inequalities at $\kappa = 1$ and restore this factor at the end.

**Lemma 16** (Positive-coefficient representation)**.** *Let* $0 < a < 1$ *and*

$$B_a(s) = as \coth(a \operatorname{artanh} s), \qquad |s| < 1. \tag{86}$$

*Then*

$$B_a(s) = 1 - \sum_{n=1}^{\infty} c_n(a) s^{2n}, \tag{87}$$

*where*

$$c_1(a) = \frac{1 - a^2}{3}, \tag{88}$$

$$(2n+1)c_n(a) = (2n-3)c_{n-1}(a) + \sum_{j=1}^{n-1} c_j(a) c_{n-j}(a), \qquad n \geq 2. \tag{89}$$

*In particular,* $c_n(a) > 0$ *for every* $n \geq 1$.

*Proof.* Set $B = B_a$ and $u(s) = \tanh(a \operatorname{artanh} s) = as/B(s)$. Differentiation gives

$$u'(s) = \frac{a(1 - u(s)^2)}{1 - s^2}. \tag{90}$$

Eliminating $u$ yields

$$(B - sB')(1 - s^2) = B^2 - a^2 s^2. \tag{91}$$

Write $B = 1 - C$ and $C = \sum_{n \geq 1} c_n s^{2n}$. Equation (91) is equivalent to

$$s(1 - s^2)C' = C^2 - C - Cs^2 + (1 - a^2)s^2. \tag{92}$$

Equating the $s^2$ coefficient gives $c_1 = (1 - a^2)/3$; equating $s^{2n}$ for $n \geq 2$ gives (89). Induction proves positivity. To justify convergence, use the principal branch $\operatorname{artanh} s = \frac{1}{2} \operatorname{Log}((1+s)/(1-s))$. For $|s| < 1$,

$$\operatorname{Re} \frac{1+s}{1-s} = \frac{1 - |s|^2}{|1 - s|^2} > 0, \qquad |\operatorname{Im}(\operatorname{artanh} s)| < \frac{\pi}{4}.$$

Thus $a \operatorname{artanh} s$ cannot meet a nonzero pole $i\pi k$ of $\coth$. At $s = 0$, the prefactor $as$ cancels the pole and $B_a(0) = 1$. The even function $B_a$ is therefore analytic in the unit disk, so its Taylor series converges normally on compact subsets. Termwise differentiation on $(0, 1)$ is justified. □

*Proof of Theorem 12.* The diagonal identity and $\ell_q^{(4)}(0) = -2q^2\kappa^2$ give

$$\mathsf{R}_0(\ell_q) = \kappa^2(q^2 - 1), \tag{93}$$

which proves the first claim.

Set $\kappa = 1$ for the remaining sign calculation. Since

$$z = \ell_q'(r) = \frac{\tanh(qr)}{q}, \qquad s = qz, \qquad a = 1/q, \tag{94}$$

the inverse momentum is $g(z) = a \operatorname{artanh} s$. The functions in (51) become

$$A(z) = 1 - s^2, \qquad B(z) = B_a(s). \tag{95}$$

Using $\mathrm{d}/\mathrm{d}z = a^{-1}\mathrm{d}/\mathrm{d}s$, direct substitution in (54) gives

$$\beta = \frac{2(B_a - 1)}{a^2 s^2}, \qquad \gamma = \frac{B_a''}{a^2}, \qquad \delta = \frac{B_a'}{a^2 s}, \tag{96}$$

$$\alpha + \delta = \frac{6(1 - B_a) + 5sB_a'}{a^2 s^2}. \tag{97}$$

If $q = 1$, then $a = 1$, $B_a \equiv 1$, and all four coefficients vanish. Suppose $q > 1$. Termwise differentiation of Lemma 16 gives

$$\beta = -\frac{2}{a^2} \sum_{n \geq 1} c_n s^{2n-2}, \tag{98}$$

$$\delta = -\frac{1}{a^2} \sum_{n \geq 1} 2n c_n s^{2n-2}, \tag{99}$$

$$\gamma = -\frac{1}{a^2} \sum_{n \geq 1} 2n(2n-1) c_n s^{2n-2}, \tag{100}$$

$$\alpha + \delta = \frac{1}{a^2} \sum_{n \geq 1} (6 - 10n) c_n s^{2n-2}. \tag{101}$$

Set

$$h = \frac{2c_1(a)}{a^2} = \frac{2(q^2 - 1)}{3}. \tag{102}$$

Positivity of every $c_n$ and the coefficient inequalities $2n \geq 2$, $2n(2n-1) \geq 2$, and $6 - 10n \leq -4$ imply

$$\beta, \gamma, \delta \leq -h, \qquad \alpha + \delta \leq -2h. \tag{103}$$

For unit orthogonal $u, w$, let $x = |u_0|^2$ and $y = |w_0|^2$. As in the proof of Theorem 10, orthogonality gives $x + y \leq 1$. Rearranging the Lee–Li bracket gives

$$Q = (\alpha + \delta)xy + \beta x(1 - y) + \gamma(1 - x)y \tag{104}$$

$$+ \delta(1 - x - y). \tag{105}$$

All four weights are nonnegative. Using (103),

$$Q \leq -h\,[2xy + x(1 - y) + (1 - x)y + 1 - x - y] = -h. \tag{106}$$

Therefore $\mathrm{MTW} = -3Q/2 \geq q^2 - 1$ when $\kappa = 1$. At $s = 0$, all four coefficients equal $-h$, so equality holds. Restoring the curvature scale multiplies the MTW tensor by $\kappa^2$, which proves (37) and the exactness of $\Gamma_q$. □

**Strict separation of the two continuations.** Fix $q > 1$ and set $a_\star = \Gamma_q/3$. The full bound (37) implies the transverse constraint with budget $\Gamma_q$, so Theorem 7 gives $\ell_q \leq \ell^\star_{a_\star}$. The quadratic coefficient of the $B$-series is exactly $a_\star$, whereas Lemma 16 has $c_n > 0$ for every $n \geq 2$. Consequently $B_q(z) < 1 - a_\star z^2$ for $z > 0$. Interval rigidity in Theorem 7 rules out equality of the profiles at any positive radius, proving (42).

**Corollary 17** (Sixth-order separation of matched repairs)**.** *Let $a > 0$ and choose $q > 1$ so that $q^2\kappa^2 = \kappa^2 + 3a$. Then*

$$\ell^\star_a(r) - \ell_q(r) = \frac{a(\kappa^2 + 4a)}{30} r^6 + O(r^8) \qquad (r \downarrow 0). \tag{107}$$

*Thus the matched paths agree through order four and separate at order six.*

*Proof of Corollary 17.* Expanding (21) at zero and integrating gives

$$\ell^\star_a(r) = \frac{r^2}{2} - \frac{\kappa^2 + 3a}{12} r^4 + \left( \frac{\kappa^4}{45} + \frac{a\kappa^2}{6} + \frac{a^2}{3} \right) r^6 + O(r^8). \tag{108}$$

The Taylor series of the matched log-cosh profile is

$$\ell_q(r) = \frac{r^2}{2} - \frac{\kappa^2 + 3a}{12} r^4 + \frac{(\kappa^2 + 3a)^2}{45} r^6 + O(r^8). \tag{109}$$

Subtracting the two expressions gives (107). □

*Proof of Corollary 14.* The lower bound and its nonattainment for $\Gamma > 0$ are Corollary 11. By Theorem 12, the calibrated profile $\ell_{q_\Gamma}$ has full-space margin exactly $\Gamma$, so it is feasible for (9) and supplies the upper bound. With $a = \Gamma/3$, the matching relation $q_\Gamma^2 \kappa^2 = \kappa^2 + 3a$ holds. Hence Corollary 17 gives

$$\Delta_{q_\Gamma}(D) - \Delta(\ell^\star_{\Gamma/3}; D) = \ell^\star_{\Gamma/3}(D) - \ell_{q_\Gamma}(D) = \frac{\Gamma(3\kappa^2 + 4\Gamma)}{270} D^6 + O(D^8), \tag{110}$$

which proves (44). □

*Proof of Theorem 15.* Put $a = \Gamma/3$ and write

$$\ell(r) = \frac{r^2}{2} + c_4 r^4 + c_6 r^6 + o(r^6). \tag{111}$$

The equality in (46), together with Theorem 6, gives $c_4 = -(\kappa^2 + 3a)/12$. The coefficient $\beta$ is a member of the finite candidate set in Proposition 3 for every $d \geq 2$. Hence $m(\ell; D) \geq 3a$ implies $-3\beta(\ell'(r))/2 \geq 3a$, or $\beta(\ell'(r)) \leq -2a$, for all sufficiently small $r \geq 0$.

The chain rule makes the relevant coefficient explicit, without series inversion. Since $A(\ell'(r)) = \ell''(r)$ and $A'(\ell'(r)) = \ell^{(3)}(r)/\ell''(r)$, (54) gives

$$\beta(\ell'(r)) = \frac{\ell^{(3)}(r)}{\ell'(r)\ell''(r)} - \frac{2\ell''(r)}{\ell'(r)^2} + \frac{2\kappa \coth(\kappa r)}{\ell'(r)}. \tag{112}$$

An even $C^6$ profile has $\ell'(r) = r + 4c_4 r^3 + 6c_6 r^5 + o(r^5)$, $\ell''(r) = 1 + 12c_4 r^2 + 30c_6 r^4 + o(r^4)$, and $\ell^{(3)}(r) = 24c_4 r + 120c_6 r^3 + o(r^3)$. Using $\kappa \coth(\kappa r) = r^{-1} + \kappa^2 r/3 - \kappa^4 r^3/45 + O(r^5)$, the $r^{-2}$ terms in (112) cancel and leave

$$\begin{aligned} \beta(\ell'(r)) = 8c_4 + \frac{2\kappa^2}{3} \\ + \left(72c_6 - 256c_4^2 - \frac{8\kappa^2 c_4}{3} - \frac{2\kappa^4}{45}\right) r^2 + o(r^2). \end{aligned} \tag{113}$$

Substituting the saturated fourth-order coefficient yields

$$\beta(\ell'(r)) = -2a + \left(72c_6 - \frac{8}{5}\kappa^4 - 10a\kappa^2 - 16a^2\right) r^2 + o(r^2). \tag{114}$$

The displayed coefficient of $r^2$ cannot be positive. Therefore

$$c_6 \leq \frac{\kappa^4}{45} + \frac{5a\kappa^2}{36} + \frac{2a^2}{9} = \frac{\kappa^4}{45} + \frac{5\Gamma\kappa^2}{108} + \frac{2\Gamma^2}{81}. \tag{115}$$

The expansion in the proof of Corollary 17 gives

$$c_{6,\perp}(\Gamma) = \frac{\kappa^4}{45} + \frac{a\kappa^2}{6} + \frac{a^2}{3}. \tag{116}$$

Subtracting the preceding upper bound proves (47).

By Theorem 12, $\ell_{q_\Gamma}$ has full-space margin $\Gamma$; its Taylor coefficient is $c_6(\ell_{q_\Gamma}) = (\kappa^2 + \Gamma)^2/45$. Substitution in (45) gives $\mathcal{P}_6(\ell_{q_\Gamma}; \Gamma) = \Gamma(3\kappa^2 + 4\Gamma)/270$. The two bounds in (48) follow, and their difference is the claimed width. □

## D MIXED-HESSIAN SPECTRUM AND THE FINITE-RADIUS FRONTIER

**Lemma 18** (Mixed-Hessian singular values)**.** *Let $c_\ell = \ell \circ d_\kappa$ with $\ell', \ell'' > 0$. At a pair $(x, y)$ of distance $r > 0$, the singular values of $-D^2_{xy} c_\ell$ are*

$$\ell''(r) \quad \textit{in the radial direction}, \qquad \frac{\kappa \ell'(r)}{\sinh(\kappa r)} \quad \textit{with multiplicity } d-1. \tag{117}$$

*Both extend continuously to 1 at $r = 0$ under our normalization.*

*Proof.* Let $y = \exp_x(re)$ with $\|e\| = 1$. The covector $-D_x c_\ell$ identifies with $\ell'(r)e$. Varying $y$ radially differentiates its magnitude and gives $\ell''(r)$. For a transverse endpoint variation $v \perp e$, the Jacobi field along the geodesic with $J(0) = 0$ and $J(r) = v$ is

$$J(t) = \frac{\sinh(\kappa t)}{\sinh(\kappa r)} P_{r \to t} v. \tag{118}$$

Thus the initial direction changes by $\kappa v / \sinh(\kappa r)$ after parallel identification, giving the transverse singular value. The limits follow from $\ell'(r) = r + o(r)$ and $\sinh(\kappa r) = \kappa r + o(r)$. □

*Proof of Corollary 9.* The first two expansions follow from $\log\cosh t = t^2/2 - t^4/12 + t^6/45 + O(t^8)$ and the monotonicity of $r^2/2 - \ell_1^\kappa(r)$ on $[0, D]$. By Lemma 18, the radial and transverse singular values are $\operatorname{sech}^2(\kappa r)$ and $\operatorname{sech}(\kappa r)$, respectively. Their ratio is $\cosh(\kappa r)$ and is maximal at $D$. Finally, Euclidean squared distance has identity mixed Hessian and zero MTW tensor. □

*Proof of Theorem 13.* Substituting $\ell_q'$ and $\ell_q''$ into Lemma 18 gives (38). With $x = \kappa r$,

$$\frac{\tau_q(r)}{\rho_q(r)} = \frac{\sinh(2qx)}{2q \sinh x}. \tag{119}$$

The function $t \mapsto t \coth t$ is strictly increasing on $(0, \infty)$ because

$$\frac{\mathrm{d}}{\mathrm{d}t}(t \coth t) = \frac{\sinh t \cosh t - t}{\sinh^2 t} > 0. \tag{120}$$

It follows that the ratio is at least one and strictly increasing in $r > 0$. This proves (40). The smallest singular value is $\rho_q(r)$, yielding (41) by inversion. Both expressions are strictly increasing in $q$ because

$$\frac{\partial}{\partial q} \log \frac{\sinh(2qx)}{q} = 2x \coth(2qx) - \frac{1}{q} > 0. \tag{121}$$

For the fidelity term, $r^2/2 - \ell_q(r)$ is strictly increasing in $r > 0$ since

$$r - \ell_q'(r) = r - \frac{\tanh(q\kappa r)}{q\kappa} > 0. \tag{122}$$

This proves (39). For fixed $r > 0$, write $y = q\kappa r$ and observe that

$$\frac{\mathrm{d}}{\mathrm{d}y} \frac{\log\cosh y}{y^2} = \frac{y \tanh y - 2 \log\cosh y}{y^3} < 0, \tag{123}$$

because $2\log\cosh y - y\tanh y$ vanishes at zero and has derivative $\tanh y - y \operatorname{sech}^2 y > 0$. Hence $\Delta_q(D)$ is strictly increasing in $q$. The margin $\Gamma_q$ is also strictly increasing, so its constraint $\Gamma_q \geq \Gamma$ is equivalent to $q \geq q_\Gamma$. Strict monotonicity gives uniqueness and the displayed formulas follow by substitution. □

## E REGRET AND APPROXIMATION PROOFS

Let $\mu, \nu$ be probability measures and $\Pi(\mu, \nu)$ their coupling set. For a repaired cost $c$, write $\pi_c \in \arg\min_{\pi \in \Pi(\mu,\nu)} \int c \, \mathrm{d}\pi$ and define $\pi_0$ analogously.

**Proposition 19** (Squared-cost regret)**.** *Assume* $0 \le c_0 - c \le \Delta_c(D)$ *on* $\operatorname{supp}\mu \times \operatorname{supp}\nu$*. Then*

$$0 \le \int c_0 \,\mathrm{d}\pi_c - \int c_0 \,\mathrm{d}\pi_0 \le \Delta_c(D). \tag{124}$$

*If the squared-cost objective also satisfies*

$$\int c_0 \,\mathrm{d}\pi - \int c_0 \,\mathrm{d}\pi_0 \ge \omega(d_\Pi(\pi, \pi_0)) \tag{125}$$

*for an increasing modulus* $\omega$*, then* $d_\Pi(\pi_c, \pi_0) \le \omega^{-1}(\Delta_c(D))$.

**Proposition 20** (Continuous and fixed-Lipschitz obstacles)**.** *Suppose a map* $T$ *has one-sided limits* $y^+, y^-$ *at* $x_0$ *with* $d_\kappa(y^+, y^-) = J > 0$*. Every continuous direct map* $F$ *satisfies*

$$\sup_x d_\kappa(F(x), T(x)) \ge J/2. \tag{126}$$

*For the squared-risk statement, let* $\Sigma_0$ *be a compact* $C^1$ *interface patch with a normal tube* $\Phi(u,t)$*,* $u \in \Sigma_0$*,* $|t| < r_0$*. Assume that the tube Jacobian is at least* $j_0 > 0$*, the source density is at least* $m_0 > 0$*, and there are traces* $y^\pm(u)$ *such that, for* $0 < t < r_0$*,*

$$d_\kappa(y^+(u), y^-(u)) \ge J, \qquad d_\kappa(T(\Phi(u, \pm t)), y^\pm(u)) \le J/8. \tag{127}$$

*Then every* $L$*-Lipschitz direct map satisfies*

$$\int d_\kappa(F, T)^2 \,\mathrm{d}\mu \ge \frac{m_0 j_0 \operatorname{area}(\Sigma_0) J^2}{8} \min\left\{r_0, \frac{J}{8L}\right\}, \tag{128}$$

*with* $J/(8L) = \infty$ *when* $L = 0$*.*

*Proof of Proposition 19.* Let $d_c = c_0 - c$. By assumption, $0 \le d_c \le \Delta_c(D)$ on the support product. Optimality of $\pi_c$ for $c$ gives

$$\int c_0 \mathrm{d}\pi_c - \int c_0 \mathrm{d}\pi_0 = \left(\int c \mathrm{d}\pi_c - \int c \mathrm{d}\pi_0\right) + \left(\int d_c \mathrm{d}\pi_c - \int d_c \mathrm{d}\pi_0\right) \tag{129}$$

$$\le \Delta_c(D). \tag{130}$$

The lower bound follows from optimality of $\pi_0$ for $c_0$. Applying the assumed growth modulus at $\pi = \pi_c$ and then (124) proves the stability statement. □

*Proof of Proposition 20.* Choose sequences $x_n^\pm \to x_0$ from the two sides with $T(x_n^\pm) \to y^\pm$. Continuity gives $F(x_n^\pm) \to F(x_0)$. If the uniform error is $\varepsilon$, then

$$J \le d_\kappa(y^+, F(x_0)) + d_\kappa(F(x_0), y^-) \le 2\varepsilon. \tag{131}$$

For the tube bound, fix $u$ and $t \in (0, \tau)$ with

$$\tau = \min\{r_0, J/(8L)\}. \tag{132}$$

Let $e_\pm(t) = d_\kappa(F(\Phi(u, \pm t)), T(\Phi(u, \pm t)))$. The trace assumptions and the Lipschitz property imply

$$e_+(t) + e_-(t) \ge J - \frac{J}{8} - \frac{J}{8} - d_\kappa(F(\Phi(u,t)), F(\Phi(u,-t))) \tag{133}$$

$$\ge \frac{3J}{4} - 2Lt \ge \frac{J}{2}. \tag{134}$$

Therefore $e_+(t)^2 + e_-(t)^2 \ge (e_+(t) + e_-(t))^2/2 \ge J^2/8$. Integrating in $t$ and $u$, then using the density and tube-Jacobian lower bounds, gives (128). □

## F Exact relation to the Lee–Li criteria

For completeness, the criterion used above is as follows. In dimension two, A3w is equivalent to

$$\beta \le 0, \qquad \gamma \le 0, \qquad \alpha + \delta \le 2\sqrt{\beta\gamma}; \tag{135}$$

in dimensions at least three, one adds $\delta \le 0$. Replacing all inequalities by strict inequalities characterizes A3s. The functions are defined in (54). This paper uses the criterion only after verifying that the radial profile is even, $C^4$, and strictly convex, so A0–A2 hold on hyperbolic space (Lee & Li, 2012). For every $d \ge 2$, Proposition 3 is the quantitative version of this sign criterion: A3w at a fixed distance is equivalent to $\max \mathcal{C}_d(z) \le 0$, while $-3 \max \mathcal{C}_d(z)/2$ gives the exact pointwise margin. The sole dimensional change is the additional purely transverse candidate $\delta$ when $d \ge 3$.

**Convention audit.** In our notation the Lee–Li quartic bracket is $Q$ and $\mathrm{MTW} = -3Q/2$, as fixed in (2). Thus the coefficient conditions above are written as $Q \le 0$, while every numerical margin in the paper is measured after multiplication by $-3/2$. Conventions omitting this factor require consistent rescaling of $\Lambda$, $\Gamma$, and $m$. After this recalibration, the feasible profiles and rigidity conclusions are unchanged; the numerical budgets cannot be held fixed across conventions.

## G Additional conditioning and OT consequences

**Corollary 21** (Noncompact asymptotics of the envelope extremizer)**.** *Let $a > 0$, $\bar{z}_a = 2/(\kappa + \sqrt{\kappa^2 + 4a})$, and $\bar{H}_a = 1 + a\bar{z}_a^2$. As $D \to \infty$,*

$$\mathsf{K}(\ell_a^\star; D) \sim \frac{\bar{H}_a}{2\kappa \bar{z}_a} e^{\kappa D}, \qquad m(\ell_a^\star; D) \sim \frac{12a\kappa^2 \bar{z}_a^2}{\bar{H}_a^2} e^{-2\kappa D}. \tag{136}$$

*Thus the weakest full margin decays at the inverse square of the condition number.*

**Corollary 22** (Noncompact conditioning of the log-cosh path)**.** *For $q > 1$, as $D \to \infty$,*

$$\mathsf{K}_q(D) \sim \frac{1}{2q} e^{(2q-1)\kappa D}, \qquad \mathsf{G}_q(D) \sim \frac{1}{4} e^{2q\kappa D}. \tag{137}$$

*Compared with Corollary 21, the log-cosh path keeps its full MTW margin fixed while its condition number grows more rapidly.*

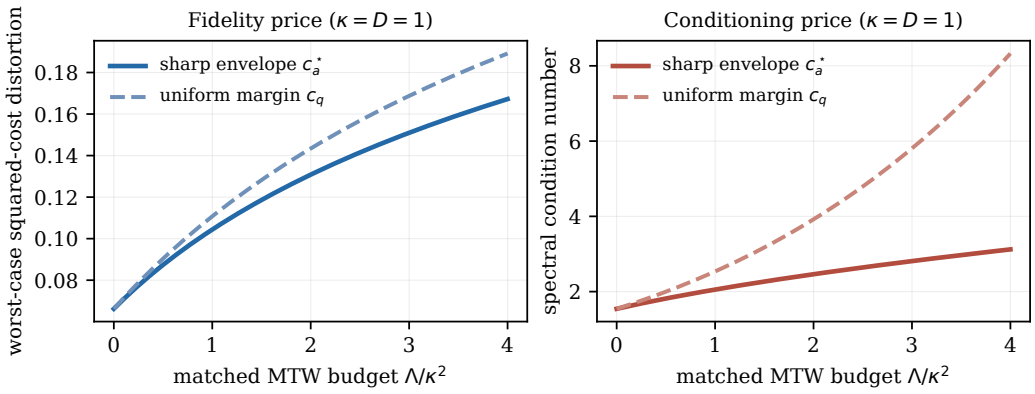


Figure 3: Two continuations of the same diagonal jet for $\kappa = D = 1$. The extremal path minimizes distortion for a prescribed transverse floor; the log-cosh path preserves the corresponding full margin on the entire space. All curves are generated from the displayed formulas.

The map-level implications use existing theory. When $q < 1$, Theorem 12 and Loeper's necessity result imply the existence of smooth positive data for which the $c_q$-optimal map is discontinuous. For $q \ge 1$ and for $c_a^\star$ with $a > 0$, the regularity conclusions of Figalli et al. (2013) and Liu et al. (2010) apply only when their density, domain, and $c$-convexity assumptions hold. MTW positivity alone does not supply those assumptions.

If $\pi_c$ and $\pi_0$ minimize a repaired cost $c$ and $c_0$, respectively, and $0 \le c_0 - c \le \Delta_c(D)$ on the admissible pairs, then Proposition 19 gives

$$0 \le \int c_0 \, \mathrm{d}\pi_c - \int c_0 \, \mathrm{d}\pi_0 \le \Delta_c(D). \tag{138}$$

This controls objective regret rather than distance between plans; the latter requires an additional growth condition. The distance cap in every theorem is the hard cap

$$D_{\mathrm{hard}} = \sup\{d_\kappa(x, y) : (x, y) \text{ is an admissible source–target pair}\}. \tag{139}$$

Quantiles $D_{95}$ and $D_{99}$ describe typical geometry but cannot replace $D_{\mathrm{hard}}$ in a uniform theorem. Distance selection should account for both anisotropy and absolute inverse gain: given budgets $K_{\max}$

and $G_{\max}$, require $\mathsf{K}(D) \le K_{\max}$ and $\mathsf{G}(D) \le G_{\max}$. Here $\mathsf{G}(\ell_a^\star; D) = H(z_a(D))/N(z_a(D))$, while $\mathsf{G}_q(D) = \cosh^2(q\kappa D)$. These are sensitivity screens, not solver-convergence guarantees.

The empirical evaluation tests mechanisms suggested by the theory. A curvature sweep checks the flat-limit collapse; controlled and real tasks ask whether singular-value and fidelity differences appear in inverse sensitivity, coupling stability, regret, and hierarchy alignment. Sinkhorn iterations are not identified with mixed-Hessian conditioning. The direct-map result in Proposition 20 is architecture-dependent and does not apply to discontinuous potential-gradient representations or unrestricted $L^2$ estimators (Makkuva et al., 2020; Pooladian et al., 2023).

# H Experimental details and additional results

## H.1 Evidence and numerical conventions

The evaluation separates three questions: whether independent calculations agree with the formulas, whether their local sensitivity prediction is observable, and whether changing the cost benefits a matching task. Agreement on the first two does not imply the third. The original M0–M3 plots are regenerated from exported numerical records. The three-target supplement includes both new LP runs and their complete implementation. The archive preserves the source numerical values. Summary-level quantities are distinguished below from instance-level observations.

The implementation uses float64 with nominal solver tolerance $10^{-8}$. The high-precision jet audit uses 336-bit arithmetic, with 160- and 240-bit checks. Costs are normalized by $\ell''(0) = 1$. Native-scale comparisons keep the support and admissible pairs fixed. Positive cost rescaling leaves exact-OT minimizers and $\mathsf{K}$ unchanged, but changes absolute inverse gain. In a Sinkhorn comparison (Cuturi, 2013; Peyré & Cuturi, 2019), scaling both cost and entropy parameter by the same positive factor leaves the mathematical regularized problem unchanged. We do not infer a property of the continuous Monge map from a discrete solver's convergence.

| Experiment | Numerical setting | Theoretical target |
|---|---|---|
| MTW directions (M0) | $d = 2, 3, 5, 10$; $\kappa = 0.5, 1, 2$; $D = 0.5, 1, 2, 3$ | Proposition 3 and Theorem 6 |
| Sixth-order jet (M0J) | $d = 2, 3, 5$; $\Gamma/\kappa^2 = 0.1, 0.5, 1, 2$; 160/240/336 bits | Theorem 15 |
| Flat limit (M0C) | $\kappa D = 0, 0.05, 0.1, 0.25, 0.5, 1, 1.5, 2$ | Corollary 9 |
| Local inverse (M1) | $d = 3$, $\kappa = 1$; perturbations $10^{-6}, 10^{-5}, 10^{-4}$ | Theorem 10 and Lemma 18 |
| Exact OT (M2) | $n = 256$; $D = 1, 2$; $\Lambda = 0.25, 0.5$; 20 instances per cell | Proposition 19 |
| Alignment (M3) | WordNet 3.0; OAEI Anatomy 2025; seeds 0–9 | Task utility, not an MTW theorem |

Table 2: Settings recorded in the experiment archive. A configuration cell, a perturbation direction, and an independent point-cloud instance are different analysis units.

## H.2 Three-target transport: complete construction and checks

**Supports, measure, and solver.** For each scrambled Sobol seed, generate $2^e$ points $u \in (0,1)^4$, with $e = 8, 9, 10$. Transform the first three coordinates by the inverse standard-normal CDF and normalize the resulting vector; multiply its radius by $u_4^{1/3}$. Reflect the resulting cloud across $x_1 = 0$ and append the origin, giving $n = 2^{e+1} + 1$. Apply $\exp_o$ to these normal-coordinate vectors. This samples the exponential pushforward of Euclidean unit-ball volume, not hyperbolic-volume-uniform measure. The explicit origin has weight $1/n$, which decreases with resolution. The separate neighborhood endpoint measures the fraction of sampled mass within $r \le 0.2$ sent to the central target.

The targets are $\exp_o(-e_1), o, \exp_o(e_1)$, their masses are $(15/32, 1/16, 15/32)$, and every source has mass $1/n$. Thus the feasible set is identical across costs within a curvature–seed–resolution cell. Its hard distance cap is at most two. We solve the $n \times 3$ unregularized transportation LP with SciPy's

HiGHS backend and nonnegative variables; the recorded primal and dual feasibility tolerances are $10^{-9}$. Because $n/16$ is nonintegral, mass splitting is allowed and necessary. Entropic regularization is absent.

Pilot seeds 0–9 determine the construction; the supplied freeze record specifies seeds 100–109 for the subsequent evaluation. We retain all three curvatures, all three resolutions, and all evaluation seeds. At zero curvature, both log-cosh profiles are assigned their continuous limit $r^2/2$. The source hash agrees across costs within every cell. Evaluation ran on CPU under Linux with Python 3.11.9, NumPy 2.4.3, and SciPy 1.17.1, according to the supplied environment record. The reported total is 270 LP solves in 31.4 seconds, excluding manuscript plotting; this is a run log, not a comparative speed benchmark.

**Why the center changes.** For $x = \exp_o(re)$, write $t = \langle e, e_1 \rangle$. Hyperbolic geometry gives

$$\cosh(\kappa d_\pm) = \cosh(\kappa r)\cosh\kappa \mp \sinh(\kappa r)\sinh\kappa\, t. \tag{140}$$

The reflected support and equal lateral masses allow a symmetric optimal target dual. Normalize its lateral entries to zero and write its central entry as $v_0$. Define

$$T_\ell(x) = \ell(r) + \ell(1) - \min\{\ell(d_-), \ell(d_+)\}, \qquad \lambda = v_0 + \ell(1). \tag{141}$$

The central dual cell is $T_\ell(x) \le \lambda$. For critical log-cosh,

$$T_{\ell_1}(x) = -\kappa^{-2}\log\bigl(1 - \tanh(\kappa r)\tanh\kappa\,|t|\bigr) \ge 0. \tag{142}$$

At the origin $T_{\ell_1} = 0$; a positive $\lambda$ puts an open origin neighborhood in the central cell. For squared distance at $\kappa = r = 1$ and $t = 0$,

$$T_{c_0} = 1 - \tfrac{1}{2}\operatorname{arcosh}(\cosh^2 1)^2 = -0.14515044\ldots, \tag{143}$$

while $T_{c_0}(o) = 0$. A negative central threshold can therefore collect off-axis sources while excluding the origin. This cost-comparison calculation explains the observed cell geometry. It uses the same negative-curvature obstruction studied by MTW theory; it is not a new proof of a regularity theorem.

At $\kappa = 1, n = 2049$, the ten squared-cost thresholds lie in $[-0.04213, -0.03032]$, the critical thresholds in $[0.02614, 0.02925]$, and the $q = 1.5$ thresholds in $[0.05505, 0.05778]$. Their signs are separated from zero by many orders of magnitude more than solver error. At $\kappa = 0.5$ the origin belongs to the central cell for all three costs, a retained control that limits the structural claim to the tested regime.

Table 3: Three-target evaluation across curvature and resolution. Ten scrambled-Sobol seeds per row and cost. Critical: $q = 1$; strict: $q = 1.5$. Regret is measured under the original squared objective. No resolution is treated as an additional independent seed.

| $\kappa$ | $n$ | Mean near-origin central share (%) | | | Mean $10^3 R_0$ | |
|---|---|---|---|---|---|---|
| | | Squared | Critical | Strict | Critical | Strict |
| 0 | 513 | 32.00 | 32.00 | 32.00 | 0.000 | 0.000 |
| 0 | 1025 | 30.51 | 30.51 | 30.51 | 0.000 | 0.000 |
| 0 | 2049 | 30.03 | 30.03 | 30.03 | 0.000 | 0.000 |
| 0.5 | 513 | 26.13 | 32.00 | 36.00 | 0.064 | 0.353 |
| 0.5 | 1025 | 22.99 | 30.51 | 40.57 | 0.081 | 0.386 |
| 0.5 | 2049 | 18.08 | 29.47 | 39.02 | 0.073 | 0.337 |
| 1 | 513 | 0.00 | 32.00 | 60.00 | 0.860 | 3.121 |
| 1 | 1025 | 0.00 | 29.46 | 56.77 | 0.934 | 3.312 |
| 1 | 2049 | 0.00 | 27.80 | 54.49 | 1.017 | 3.253 |

**Independent solution check.** For each source $x_i$, let $e_i = \min\{c(x_i, y_-), c(x_i, y_+)\}$ and $h_i = c(x_i, o) - e_i$. Starting from the cheapest lateral assignment, sending mass to the center changes the objective by $h_i$ per unit mass. Sort $h_i$ and allocate the lowest increments until the central budget $1/16$ is exhausted, splitting the last source if necessary. Reflection symmetrizes this allocation to meet both lateral budgets without changing the objective. This yields an independent solution of

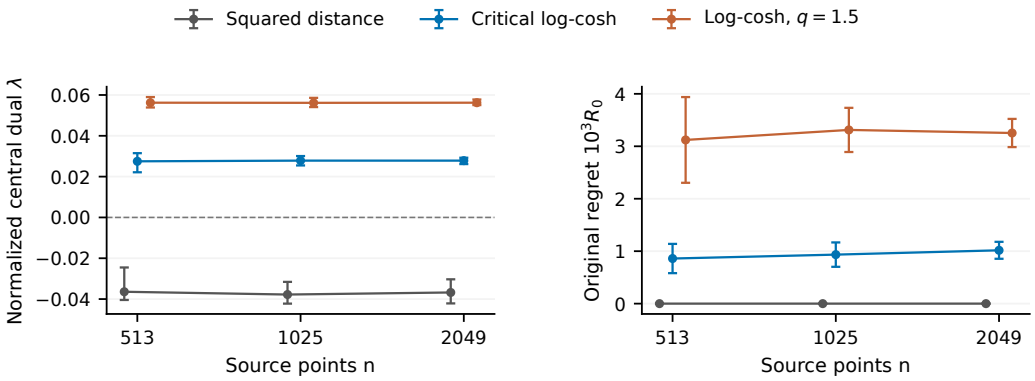


Figure 4: Three-target resolution check at $\kappa = 1$. Left: mean normalized central dual with min–max bars across ten seeds. Right: mean original-objective regret with one-standard-deviation bars. The same seeds are reused across nested resolutions; these are descriptive stability checks, not thirty independent samples.

the symmetric LP by sorting, without an LP solver. Its original squared objective is computed using the same allocation and the squared lateral costs; equality of mirror-pair costs makes boundary tie handling immaterial to that objective.

This calculation agrees with all 270 evaluation rows: maximum absolute discrepancies are $1.67 \times 10^{-16}$ in squared objective, $1.86 \times 10^{-16}$ in regret, and $9.37 \times 10^{-16}$ in normalized central threshold. Origin shares agree exactly. A separate rerun of the three costs at seed 100, $\kappa = 1, n = 2049$ also reproduces the structural endpoints and objectives. The original pilot reproduction differs only in roundoff-scale primal–dual gaps, with maximum change $1.67 \times 10^{-16}$; this is numerical reproduction within tolerance.

**Interpretation of the plots.** Figure 1 uses seed 100, the first evaluation seed, and displays the $x_3 = 0$ section of cells induced by an optimal discrete dual. It is not a separately solved two-dimensional OT problem. It does not establish connectivity or disconnectedness of a three-dimensional cell. All targets are atomic, so the experiment does not meet smooth-density hypotheses for Monge-map regularity. Neighborhood shares need not equal one: the observed change concerns the origin and a subset of its neighborhood. The $q = 1.5$ and critical profiles have different full margins; their regret comparison is specific to this task and is not a universal regret-optimality claim.

### H.3 Finite candidates, jets, and flat limits

The M0 coefficient grid uses $\Lambda/\kappa^2 \in \{0, 0.1, 0.25, 0.5, 1, 2\}$ and $q \in \{0.8, 0.95, 1, 1.1, 1.25, 1.5, 2\}$. The recorded direction-search resolutions are 1,441 angles in dimension two and 181 grid points per direction parameter in the three-dimensional audit. The overall finite-candidate summary records maximum absolute discrepancy $7.01 \times 10^{-9}$ against a $10^{-8}$ tolerance. This is an implementation check over sampled profiles and distances, not a feasibility certificate between samples.

The sixth-order audit fixes the fourth-order coefficient and perturbs the sixth-order boundary using $\eta \in \{-10^{-2}, -10^{-3}, 0, 10^{-3}, 10^{-2}\}$. Nonzero perturbations test the predicted sign of the coefficient in (114). At $\eta = 0$, the exported sign label is positive although the asymptotic coefficient is zero; the fitted residuals range from approximately $6.3 \times 10^{-10}$ to $2.38 \times 10^{-4}$. We retain these values and treat the zero-perturbation case as a boundary residual check, not a successful positive-sign test or exact equality. Finite-distance remainders and coefficient recovery must be separated before drawing a sharper conclusion at the boundary.

For critical log-cosh, set $x = \kappa D$. The dimensionless quantities are

$$f(x) = \frac{x^2/2 - \log\cosh x}{x^4}, \qquad k(x) = \frac{\cosh x - 1}{x^2}. \tag{144}$$

They approach $1/12$ and $1/2$. The values at $x = 0$ in the archive are analytic continuous extensions, not measured ratios of zeros. Figure 5 shows the exported evaluations alongside the exact functions. Varying curvature at fixed distance and varying distance at fixed curvature give the same dimensionless curves.

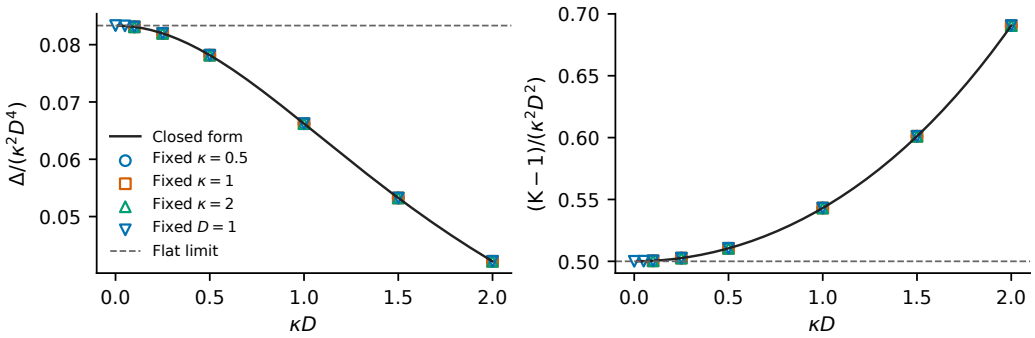


Figure 5: Flat-limit collapse of the critical repair. Lines evaluate the closed forms; markers are the supplied numerical evaluations. The dashed levels are the limiting coefficients $1/12$ and $1/2$. The endpoint at zero uses the analytic extension. This is a curvature-scaling check, not an independent task-performance experiment.

### H.4 Spectrum and inversion

For $y = \exp_x(re)$, momentum is $p = \ell'(r)e$ and $c\text{-}\exp_x(p) = \exp_x(g(\|p\|)p/\|p\|)$, where $g = (\ell')^{-1}$. The singular values in Lemma 18 therefore give both direction-dependent inverse gain and its worst-case value. The M1 design includes 128 geometry directions, random perturbations and singular-vector perturbations, and perturbation magnitudes from $10^{-6}$ to $10^{-4}$. We report the 25,000 comparisons present in the inverse-gain export, without treating them as 25,000 independent data sets. Their log-scale predicted–observed correlation is $0.9999998325$. Relative-error quantiles, rather than correlation alone, are reported in the main text because gain spans a wide dynamic range.

A separate extended-geometry summary records 544 spectral cells, a maximum relative AD–SVD discrepancy of $1.70 \times 10^{-8}$ among valid cells, and maximum small-step inverse-gain discrepancy $1.22 \times 10^{-4}$. These are distinct scopes and error definitions; they are not pooled with the main directional sample. In particular, the minimum singular vector is not an MTW direction pair.

At large distance, evaluating $N = (1 - az^2)^2 - \kappa^2 z^2$ by subtraction can lose significant digits. The exact identity

$$N(z_a(r)) = (1 - az_a(r)^2)^2 \operatorname{sech}^2(\kappa r) > 0 \tag{145}$$

avoids this cancellation. The application ledger flags 74 envelope-spectrum evaluations as invalid. They are excluded from spectrum-based interpretation, not interpreted as finite-distance degeneracy of the analytic profile. No aggregate failure-rate claim is made from these flags.

### H.5 Exact transport and scale controls

| Cost | P@1 (%) | Mean $R_0$ | Paired ΔP@1 (pp), 95% CI |
|---|---|---|---|
| Squared distance | 55.752 | 0 | Reference |
| Critical log-cosh | 55.747 | $2.11 \times 10^{-7}$ | $-0.0049$ $[-0.0439, 0.0293]$ |
| Envelope | 55.737 | $3.74 \times 10^{-7}$ | $-0.0146$ $[-0.0537, 0.0195]$ |
| Matched log-cosh | 55.737 | $3.76 \times 10^{-7}$ | $-0.0146$ $[-0.0537, 0.0195]$ |

Table 4: Exact PairMatch transport: 80 paired instances per cost, no recorded failures. Confidence intervals use 10,000 paired-bootstrap resamples; pp denotes percentage points. Small changes do not prove noninferiority.

The controlled design has three point-cloud families, named MassShift, Boundary, and PairMatch in the archive. They probe different support geometries; PairMatch additionally spans easy, medium, hard, and near-critical matching conditions. All comparisons use the same marginals and supports within an instance. The actual hard cap is $D_{\mathrm{hard}} = \max_{i,j} d_\kappa(x_i, y_j)$, not $D_{95}$ or $D_{99}$. The nominal grid contains 240 independent instances. Three repaired costs give 720 exact-regret records, and the paired-accuracy table has 80 instances per cost. Source records use $n = 256$. The smallest recorded regret is $-3.47 \times 10^{-18}$, a roundoff-scale deviation from nonnegativity; the largest is $3.24 \times 10^{-3}$. All upper bounds hold. The maximum ratio to the uniform bound is $0.00394$, so these data do not demonstrate sharpness of the distribution-free bound.

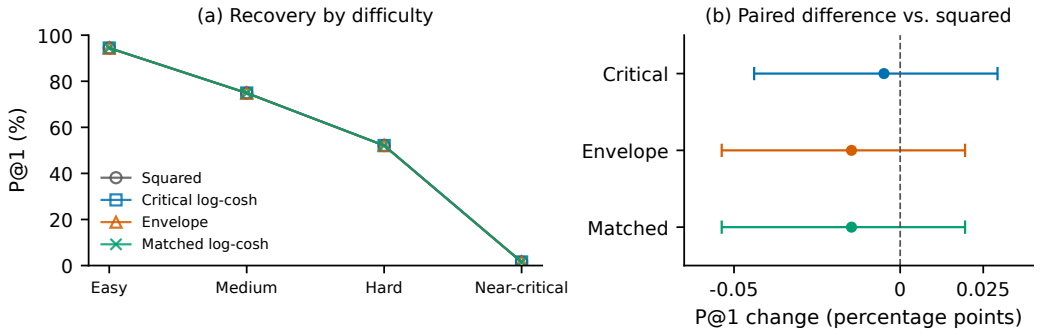


Figure 6: Exact-OT PairMatch recovery. Left: separately reconstructed means for each cost and difficulty. Right: paired P@1 differences from squared distance with the exported 95% bootstrap intervals (80 instances). Near-coincident curves do not establish equivalence.

The exported paired intervals use 10,000 bootstrap replicates at confidence level 95%, with recorded bootstrap seed 20270918. We report these exported intervals and make no new significance claim. An interval crossing zero is not a test of equivalence. Likewise, a low objective regret need not bound plan distance without the growth condition in Proposition 19.

The near-coincidence in Figure 6 is present in the instance-level export: critical log-cosh differs from squared distance in P@1 on 7 of 80 instances, and each matched repair on 9 of 80. The two matched repairs have equal P@1 on all 80 instances, but their squared-cost regret differs by more than $10^{-12}$ on 1 and coupling-TV diagnostic on 4. Each curve is computed from its own cost-labeled rows; the paired-difference panel resolves the small changes hidden by the common accuracy scale.

The Sinkhorn configuration uses $n = 2048$, at most 20,000 iterations, tolerance $10^{-8}$, and entropy ratios $\{0.005, 0.01, 0.02, 0.05\}$. Native, gradient-matched, and jointly cost/entropy-scaled controls are recorded. The native-scale PairMatch summary at entropy ratio $0.02$ has 80 instances per cost and no failures in that subset. The full aggregate ledger, however, records 33 solver failures. The available export does not provide a uniquely linked all-runs denominator for that aggregate count; we therefore report the count, do not infer an overall rate, and do not use Sinkhorn convergence as a headline result. The exact-OT result is separate from the regularized objective diagnostic.

### H.6 Hierarchy alignment: null results and numerical limitations

We retain the application experiments because they delimit the empirical scope of the theory. WordNet 3.0 (Miller, 1995) supplies a same-graph controlled comparison; OAEI Anatomy 2025 uses the mouse–human benchmark described by Dragisic et al. (2017) for cross-ontology matching. Both use structure-only hyperbolic representations. The configuration records dimension 10, embedding training for 200 epochs, 50 negative samples, and a two-layer registration network of hidden width 128. The audited equal-time retraining budget is 180 seconds, reduced from the planned 3,600 seconds, with seeds 0–9 retained. These are budget-limited pipeline evaluations, not comparisons against a tuned state-of-the-art ontology matcher.

The WordNet export contains 400 matching evaluation cells (80 clean and 320 perturbed), plus separate inverse and exact-OT records. Clean P@1 is approximately $9.9 \times 10^{-5}$ and P@10 approxi-

mately $10^{-3}$, with little difference between costs. These values are of the random-ranking order for a 10,000-candidate problem. We do not infer an accuracy or stability advantage.

| OAEI cost | Clean Hits@10 | Robust Hits@1 AUC |
|---|---|---|
| Squared distance | 0.002891 | 0.000300 |
| Critical log-cosh | 0.003557 | 0.000266 |
| Envelope | 0.003590 | 0.000266 |
| Matched log-cosh | 0.003590 | 0.000266 |

Table 5: Means of the ten exported OAEI seed endpoints. Absolute recovery is low. The robust superiority endpoint fails for all three repairs.

The equal envelope and matched-log-cosh Hits summaries in Table 5 also occur before rounding: their four seed-level endpoints coincide for all ten seeds. Other exported quantities distinguish the costs. Their MRR values differ on all 80 matched primary evaluation rows (maximum absolute difference $2.40 \times 10^{-5}$), and their optimized cost values differ on all 40 exact-audit rows. Thus the table faithfully reports the supplied endpoints; identical coarse recovery scores do not imply identical costs or transportation plans. The available exports do not independently verify the original prediction pipeline.

For critical log-cosh versus squared distance, the robust-Hits@1-AUC difference is $-3.37\times10^{-5}$ with paired 95% interval $[-2.34\times10^{-4}, 1.99\times10^{-4}]$. The clean-Hits@10 difference is $6.66\times10^{-4}$ with interval $[1.34\times10^{-4}, 1.13\times10^{-3}]$. Although the latter passes the prespecified absolute noninferiority margin of $-0.01$, that margin is larger than the baseline score itself. It is not informative about practical preservation at this accuracy scale. The joint application criterion is unmet because robust superiority fails. We do not replace that criterion after inspecting the results.

The registration audit selects step zero throughout the cost-specific and dropout retraining stages. A bounded-geometry intervention projects 98.68% of points on average, so it cannot isolate a pure change of cost. Sparse OT completes numerically but disagrees with dense OT in top-1 prediction by 99.62%, 76.31%, and 71.74% at neighborhood sizes $64$, $256$, and $1024$, respectively. It is not used as a faithful substitute for the dense reference. Exact-regret checks on these supports remain objective-level diagnostics and cannot rescue an application claim.

### H.7 Additional resolution, precision, and solver controls

The supplementary exports retain several views of the same observations in separate plotting panels. We summarize each relevant panel separately and do not count repeated views as new independent trials.

The direction-resolution check evaluates six grids, from 31 to 961 points, at the single recorded distance $r = 0.994167$. All six give zero recorded discrepancy from the coefficient-based reference. This corroborates that particular calculation; it is not a uniform error guarantee over profiles or distances. The broader finite-candidate comparison above remains the appropriate aggregate audit.

The jet-window export contains 2,160 coefficient fits per plotting panel, covering 12, 16, 20, and 24 fit points and 160-, 240-, and 336-bit precision. In the sign panel, all 1,728 nonzero-$\eta$ fits have the predicted sign. The 432 zero-$\eta$ fits retain the small positive slopes described above; the largest absolute fitted intercept is $1.66 \times 10^{-12}$. Increasing arithmetic precision alone therefore does not remove the finite-window boundary remainder. The flat-remainder export contains 26 settings shown in four panels, supplementing the dimensionless collapse in Figure 5.

The inverse-step export has 30,000 entries in each of four panels at perturbation magnitudes $10^{-6}$, $10^{-5}$, and $10^{-4}$. Its extrapolated gain panel gives median, 95th-percentile, and maximum relative discrepancies $6.41 \times 10^{-5}$, $1.14 \times 10^{-3}$, and $1.54 \times 10^{-2}$. The residual-versus-step panel has maximum Newton residual $9.97 \times 10^{-14}$; the iteration panel records one to four Newton steps. These are complementary views of local inversion and are not added to the 25,000 main-figure comparisons as independent samples.

The Sinkhorn diagnostic export contains 2,016 entries per panel across native, gradient-matched, and jointly scaled cost/entropy settings. The residual panel has median $8.57 \times 10^{-9}$, 95th percentile

$9.96 \times 10^{-9}$, and maximum $5.99 \times 10^{-7}$; 31 entries exceed $10^{-8}$. The iteration panel also has 31 entries at the 20,000-iteration cap. These panel counts do not reconcile the separate aggregate ledger's 33 failures, whose full run identifiers are unavailable. We preserve both scopes rather than report a misleading pooled success rate. Iteration medians across scale modes are not treated as paired speed comparisons: the exported panel groups have different configuration counts.

All fourteen original main and supplementary plot-data files and all original numerical result tables are included in the companion data archive. Source summary labels are retained for traceability; the scope and interpretation of each result follow the definitions in this paper.

### H.8 Reproduction scope

The companion archive includes the exported plot-level observations, paired summaries, parameter table, failure counts, and a deterministic script for the figures and calculations reported here. It reproduces the presentation from supplied records. For the original M0–M3 pipelines, those records are not a complete raw execution archive: the original point clouds, checkpoints, full solver logs, and complete training/transport implementation are not included. Consequently, rerunning those original pipelines requires additional artifacts. In contrast, the three-target generator and LP implementation are included and support a complete rerun. We do not claim that a successful plot rebuild verifies the training runs or constitutes an independent replication.

## I Boundary cases and scope

| Design problem | Search class | Result in this paper |
|---|---|---|
| Full A3w, $\Gamma = 0$ | All normalized radial costs, $d \geq 3$ | Critical log-cosh: exact optimum and interval rigidity |
| Transverse floor $\Lambda \geq 0$ | Same class, $d \geq 3$ | Sharp, attained, rigid envelope |
| Full margin $\Gamma > 0$ | All normalized radial costs, $d \geq 2$ | Exact directions and jet bounds; finite-cap bounds for $d \geq 3$; functional infimum open |
| Uniform full margin | Calibrated log-cosh family | Exact margin and unique family-wise frontier |

Table 6: The full A3w optimum is solved; the positive full-margin optimum remains open. The last row is restricted to the calibrated family.

**Dimension.** In $d = 2$, Proposition 3 still gives the exact full-MTW directional minimum, and the canonical families and jet bounds apply. The sharp transverse envelope requires two orthogonal directions tangent to a distance sphere, hence $d \geq 3$. No two-dimensional fidelity optimum is asserted.

**Curvature dependence.** Negative curvature enters the diagonal obstruction $-\kappa^2$, the coefficient $\kappa z \coth(\kappa g)$, and the transverse spectral factor $\kappa / \sinh(\kappa r)$. The critical repair and its prices converge to the quadratic Euclidean values as $\kappa \downarrow 0$ (Corollary 9).

**Optimality classes.** The transverse envelope is sharp over normalized radial profiles and has interval rigidity. Full A3w is solved in $d \geq 3$. Positive full margins have an unattainable lower bound and a feasible upper bound, but neither the exact infimum nor a positive gap is established. The calibrated log-cosh frontier is family-wise (Table 6).

**Sixth-order scope.** Theorem 15 bounds the next even coefficient for $C^6$ profiles with a full margin and fourth-order saturation, in every $d \geq 2$. It does not establish jet optimality of log-cosh or improve the finite-cap lower bound without uniform higher-order remainder control.

**Positive margins and sensitivity.** Increasing $q > 1$ raises the full margin but decreases $\rho_q(D) = \operatorname{sech}^2(q\kappa D)$. A stronger MTW condition thus need not improve numerical sensitivity; the experiments measure these separately.