OPTIMAL TRANSPORT IS NOT VOLUME-OPTIMAL FOR CONFORMAL PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Transport-based methods construct multivariate prediction regions as pullbacks of quantile regions of a simple reference distribution: any invertible map that transports the reference measure to the conditional law of the response yields regions with exact conditional coverage. This validity, however, says nothing about the volume of the resulting regions. Recent work has favored optimal-transport (Brenier) maps, implicitly treating the induced regions as canonical. We show that this choice is provably suboptimal. For a standard Gaussian reference, we show that, among elliptical targets with finite second moments, exactly the Gaussian and spherical targets yield prediction regions whose volumes equal those of the corresponding highest-density regions (HDR) at every coverage level; for every other target in this class, the Brenier-induced region has strictly greater Lebesgue volume over a nonempty open interval of coverage levels. For the compactly supported references used in center-outward multivariate ranks — the uniform ball and the spherical uniform — the same volume gap holds for every non-spherical elliptical target, without exception. We then factor admissible transports into a Brenier map and a measure-preserving rearrangement, thereby separating coverage from volume. We parameterize this class using smooth measure-preserving flows generated by skew-symmetric matrix fields with at most $2(d_y - 1)$ nonzero entries and derive a stochastic volume objective. Experiments on synthetic and real data show that measure-preserving rearrangements yield smaller prediction regions.

1 INTRODUCTION

For a response $Y \in \mathbb{R}^{d_y}$, two prediction regions can have the same probability and very different volumes. Let $U \sim \mu_U$, let $T_{x\#}\mu_U = \mu_{Y|X=x}$, let $\mathcal{B}_{d_y}(0, r_{1-\alpha})$ define a d_y-dimensional ball, centered at 0 with radius $r_{1-\alpha}$ and define

$$\mathcal{C}^T_{1-\alpha}(x) := T_x\big(\mathcal{B}_{d_y}(0, r_{1-\alpha})\big), \qquad \mu_U\big(\mathcal{B}_{d_y}(0, r_{1-\alpha})\big) = 1 - \alpha.$$

Then $\mathcal{C}^T_{1-\alpha}(x)$ has probability $1 - \alpha$ for every such T_x. Its volume, however, depends on T_x.

Exact transport gives the desired population coverage, and split conformal calibration corrects the radius when the transport is estimated. But calibration cannot repair a geometrically inefficient transport: it merely rescales the image of the same latent ball. Coverage is thus fixed by probability preservation, whereas volume depends on the representative chosen among the many transports with the same pushforward law.

Multivariate quantile methods often use the Brenier map $Q_x = \nabla\varphi_x$ (Carlier et al., 2016; 2017; Chernozhukov et al., 2017). Although this map minimizes the quadratic transport cost, this objective does not minimize the volume of $Q_x(\mathcal{B}_{d_y}(0, r))$. On the other hand, when a highest-density region exists, it has minimum volume among all sets with the same probability mass (Neyman & Pearson, 1933; Box & Tiao, 1973; Hyndman, 1996). Our main theoretical results show that, even in relatively simple settings, the Brenier map does not recover minimum-volume sets.

Figure 1 illustrates the discrepancy. For a standard Gaussian U, the map T_b of Appendix C has unit Jacobian determinant and maps every centered Gaussian ball onto a density superlevel set of its image distribution — a minimum-volume set at the corresponding probability level. Its Jacobian

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.

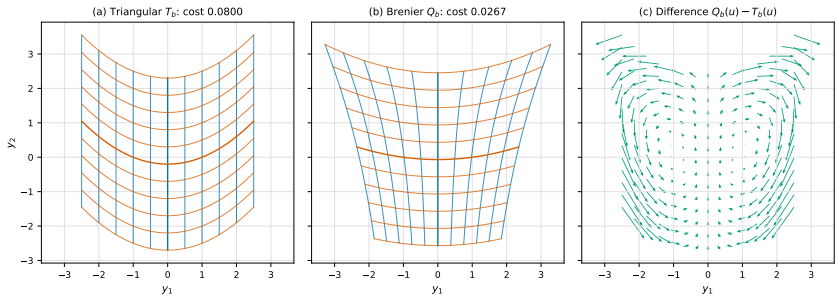


Figure 1: **Two transports for the banana distribution.** Left: The triangular map T_b sends the ball to the highest-density region. Middle: The Brenier map Q_b has lower quadratic cost but its image has larger volume. Right: The difference between the maps.

being non-symmetric, T_b is not the gradient of a convex potential, hence differs from the Brenier map Q_b.

To lower the volume of prediction regions induced by Brenier map, we use the non-uniqueness of the transport. Every sufficiently regular map from μ_U to $\mu_{Y|X=x}$ can be written as

$$Q_x \circ S_x, \qquad S_{x\#}\mu_U = \mu_U,$$

where the map S_x changes the image of the reference ball without changing its probability. We parameterize S_x as a flow map that preserves μ_U, and minimize the volume of $Q_x \circ S_x(\mathcal{B}_{d_y}(0, r_{1-\alpha}))$.

Our contributions are as follows:

- We show that the Brenier image of a reference ball need not have minimum volume among all sets of the same coverage, even for smooth, unimodal source and target densities (Section 3).
- We derive a new characterization of measure-preserving flows and provide a novel, sparse parameterization for a Gaussian reference measure (Section 4.1).
- We derive a volume objective and identify the population optimum over smooth measure-preserving maps (Section 4.2).
- We suggest a numerically stable surrogate function that shares the population optimum with the original volume objective (Section 4.2).

Related work. Monge–Kantorovich methods build multivariate quantiles from gradients of convex functions (Chernozhukov et al., 2017; Hallin et al., 2021; Carlier et al., 2016). Conformal versions obtain coverage by calibrating the radial score (Thurin et al., 2025; Kondratyev et al., 2026; Klein et al., 2026). Minimum-volume, level-set, and volume-optimizing conformal methods instead target small regions directly (Polonik, 1997; Scott & Nowak, 2006; Izbicki et al., 2022; Sampson & Chan, 2026; Gao et al., 2025; Braun et al., 2026a; Luo & Zhou, 2026). Our focus is the measure-preserving orbit of a fixed transport and the volume consequences of choosing its Brenier representative. Appendix A gives a more detailed comparison.

2 Constructing Multivariate Prediction Regions

We describe the transport construction and fix the coverage convention used throughout the paper. Let (X, Y) be a random variable that take values in $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}$. We write μ_X for the law of X and $\mu_{Y|X=x}$ for a regular conditional law of Y given $X = x$. Let U take values in $\mathcal{U} \subseteq \mathbb{R}^{d_y}$ with reference law μ_U; as is most common in practice, we adopt the standard Gaussian $\mu_U = \mathcal{N}(0, \mathrm{I}_{d_y})$. We often use the following assumptions.

A 1. *The reference law μ_U has a density f_U with respect to Lebesgue measure on $\mathbb{R}^{d_y}$, and its support is convex.*

A 2. *For μ_X-almost every $x \in \mathcal{X}$, the conditional law $\mu_{Y|X=x}$ has a density $f_{Y|X}(\cdot \mid x)$ with respect to Lebesgue measure on $\mathbb{R}^{d_y}$.*

A conditional transport is a measurable map $T\colon \mathcal{U} \times \mathcal{X} \to \mathcal{Y}$ such that $T(\cdot, x)_{\#}\mu_U = \mu_{Y|X=x}$ for μ_X-almost every x. We also assume that there exists a jointly measurable map $T^{-1}\colon \mathcal{Y} \times \mathcal{X} \to \mathcal{U}$

that is an a.e. inverse of $T(\cdot, x)$ for μ_X-almost every x; this fixes a common measurable version of the inverse. To construct a prediction region we choose $r_{1-\alpha}$ with $\mu_U\big(\mathcal{B}_{d_y}(0, r_{1-\alpha})\big) = 1-\alpha$ and define

$$\mathcal{C}^T_{1-\alpha}(x) := \big\{y \in \mathcal{Y}\colon \big\|T^{-1}(y,x)\big\|_2 \le r_{1-\alpha}\big\}, \qquad \text{so that} \quad \mu_{Y|X=x}\big(\mathcal{C}^T_{1-\alpha}(x)\big) = 1-\alpha$$

for μ_X-almost every x, by the pushforward property. Thus an exact conditional transport gives exact conditional coverage. In practice the transport is estimated, and split conformal calibration will later replace this statement with a finite-sample marginal guarantee.

Among all transports between μ_U and $\mu_{Y|X=x}$, a particularly important class consists of gradients of convex potentials, recently employed in conformal vector quantile regression (Thurin et al., 2025; Klein et al., 2026; Kondratyev et al., 2026).

Under **A** 1 and 2 and provided $\mathbb{E}\|Y\|_2 < \infty$, it was shown in (Carlier et al., 2016; 2017) (see also McCann, 1995, where no moment assumption is required) that there exists a measurable map $Q_{Y|X}\colon \mathcal{U} \times \mathcal{X} \to \mathcal{Y}$, unique $\mu_U \otimes \mu_X$-a.e., transporting μ_U to $\mu_{Y|X=x}$ for μ_X-a.e. x and equal to the gradient in u of a potential $\varphi(\cdot, x)$ that is convex for each x. The inverse map exists almost-everywhere and is given by $Q^{-1}_{Y|X} = \nabla\varphi^\star$, where $\varphi^\star$ is Legendre transform of φ (McCann, 1995). We refer to $Q_{Y|X} = \nabla_u\varphi$ as the conditional vector quantile map or the Brenier map.

Between two measures, there exist multiple valid transport maps. The next factorization makes the non-uniqueness of the transport explicit.

Theorem 1 (Transport factorization, Theorem. 1.53, Santambrogio (2015))**.** *Let μ_U and μ be absolutely continuous, and let $Q\colon \mathcal{U} \to \mathcal{Y}$ be an a.e.-invertible Brenier transport from μ_U to μ. A measurable map $G\colon \mathcal{U} \to \mathcal{Y}$ satisfies $G_\#\mu_U = \mu$ if and only if*

$$G = Q \circ S \qquad \mu_U\text{-almost everywhere}$$

for a measurable map $S\colon \mathcal{U} \to \mathcal{U}$ satisfying $S_\#\mu_U = \mu_U$. If G is also invertible up to null sets, then S is invertible up to null sets.

For later use, the factorization also has a conditional form: if Q and G are conditional transports and Q has a jointly measurable a.e. inverse, then $S(u,x) := Q^{-1}(G(u,x),x)$ is jointly measurable and satisfies $S(\cdot,x)_\#\mu_U = \mu_U$ for μ_X-almost every x; if G also has a jointly measurable a.e. inverse, so does S, namely $S^{-1}(u,x) = G^{-1}(Q(u,x),x)$. All conditional constructions below assume these measurable versions; the population theorems are stated for a fixed distribution and do not require them.

Let us now define the volume-optimal prediction regions. Suppose that there exists a threshold $\tau_{1-\alpha}(x) > 0$ such that

$$\mu_{Y|X=x}\big(\{y\colon f_{Y|X}(y \mid x) \ge \tau_{1-\alpha}(x)\}\big) = 1-\alpha.$$

The conditional highest-density region is

$$\mathrm{HDR}^{f_{Y|X}}_{1-\alpha}(x) := \big\{y \in \mathcal{Y}\colon f_{Y|X}(y \mid x) \ge \tau_{1-\alpha}(x)\big\}.$$

It is known that such sets provide prediction regions with the smallest Lebesgue volume at the required coverage level. See details in Appendix B.

3 Optimal transport is volume-suboptimal

We now show that, in a fairly general setting, the prediction regions induced by the Brenier map do not coincide with the conditional highest density regions $\mathrm{HDR}^{f_{Y|X}}_{1-\alpha}(x)$.

Since the statement below is conditional on $X = x$ and x plays no role, we suppress it from the notation. Let $d_y \ge 2$, and let μ_U be the standard Gaussian distribution. Let μ be a probability distribution on $\mathbb{R}^{d_y}$, absolutely continuous with respect to Lebesgue measure, with density

$$f(y) = C\, g\big((y-m)^\top \Sigma^{-1}(y-m)\big), \qquad y \in \mathbb{R}^{d_y}, \tag{1}$$

where $m \in \mathbb{R}^{d_y}$, $\Sigma \in \mathbb{R}^{d_y \times d_y}$ is symmetric positive definite, $C > 0$, and $g : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and strictly decreasing; assume in addition that $\int_{\mathbb{R}^{d_y}} \|y\|_2^2\, f(y)\, dy < \infty$. Let $Q = \nabla\varphi$

be the Brenier map transporting μ_U to μ. For $\alpha \in (0,1)$, let $r_{1-\alpha}$ be such that $\mu_U\big(\mathcal{B}_{d_y}(0, r_{1-\alpha})\big) = 1-\alpha$, and define the Brenier-induced prediction region

$$\mathcal{C}^Q_{1-\alpha} := \left\{y \in \mathbb{R}^{d_y} \colon \|Q^{-1}(y)\|_2 \le r_{1-\alpha}\right\}, \qquad \text{so that } \mu\big(\mathcal{C}^Q_{1-\alpha}\big) = 1-\alpha.$$

Let $\mathrm{HDR}^f_{1-\alpha}$ denote the highest density region of μ with probability content $1-\alpha$. It is the Mahalanobis ellipsoid

$$\mathrm{HDR}^f_{1-\alpha} = \{y \in \mathbb{R}^{d_y} \colon (y-m)^\top \Sigma^{-1}(y-m) \le q_{1-\alpha}\}$$

for the unique $q_{1-\alpha} > 0$ such that this set has μ-mass $1-\alpha$, and it is the Lebesgue-a.e. unique minimum-volume set of that mass. See the derivations in Appendix B.

The following theorem shows that, although the Brenier map is optimal for the quadratic transport cost, the family of prediction regions it induces is not volume-optimal for such a non-spherical target.

Theorem 2 (Brenier map does not generate the HDR family)**.** *Let $d_y \ge 2$, let μ_U be the standard Gaussian distribution, and let μ have density* (1) *with finite second moment, where Σ is symmetric positive definite with $\Sigma \notin \{\sigma \mathrm{I} \colon \sigma > 0\}$ and g is continuous, strictly decreasing, and not proportional to $s \mapsto \mathrm{e}^{-\beta s}$ for any $\beta > 0$. Let Q be the Brenier map from μ_U to μ. Then there exists a nonempty open interval $I \subseteq (0,1)$ such that, for every $\alpha \in I$,*

$$\mathrm{Vol}\Big(\mathcal{C}^Q_{1-\alpha}\Big) > \mathrm{Vol}\Big(\mathrm{HDR}^f_{1-\alpha}\Big).$$

In particular, the family $\{\mathcal{C}^Q_{1-\alpha}\}_{\alpha\in(0,1)}$ does not coincide, up to Lebesgue-null sets, with the family of HDRs of μ.

Proof outline. The complete proof is in Appendix E; it instantiates a rigidity lemma (Lemma 16, Appendix D) valid for every spherically symmetric reference. *(1)* Matching the radial distribution functions of $\|U\|_2$ and of the Mahalanobis radius $\rho_\Sigma(Y) := \|\Sigma^{-1/2}(Y-m)\|_2$ yields an increasing radial map a and the explicit transport $T_{\mathrm{HDR}}(u) := m + \Sigma^{1/2} a(\|u\|_2)\, u/\|u\|_2$, which pushes μ_U onto μ and maps *every* centered ball onto the HDR of the corresponding mass. *(2)* When Σ is non-spherical and a nonlinear, the Jacobian of T_{HDR} is non-symmetric on a set of positive measure, so T_{HDR} is not a Brenier map; a is linear precisely when g is exponential — the excluded Gaussian case, where T_{HDR} is the affine Brenier map itself. *(3)* If the Brenier regions matched the HDRs at almost every level, a radial–angular rigidity argument — in which the second moment enters through the correlation $\int u^\top y \,\mathrm{d}\pi$ — would force $Q = T_{\mathrm{HDR}}$ a.e., contradicting (2); a continuity argument then upgrades the discrepancy to an open interval of levels, on which strict HDR optimality (Lemma 12) yields the strict volume gap.

Remark 3 (Role of the assumptions)**.** *The second-moment condition makes the correlation $\int u^\top y \,\mathrm{d}\pi$ finite for every coupling π of (μ_U, μ) (by Cauchy–Schwarz) and is what allows us to characterize Q variationally (Knott & Smith, 1984; Brenier, 1991); whether it can be relaxed is open, and a first moment suffices for compactly supported references (Theorem 17, Appendix D). The exclusion of exponential profiles delineates a genuine exception: such a target is Gaussian, the Brenier map is then affine, and its regions coincide with the HDR ellipsoids at* every *level (Theorem 18 (ii)). Finally, the theorem cannot be strengthened to* all *$\alpha \in (0,1)$ without further arguments: our proof shows only that the set of levels with a strict gap has positive measure and is open.*

Remark 4 (Dependence on the reference measure)**.** *Multivariate ranks and quantiles are usually built from compactly supported references: the spherical uniform on the unit ball for center-outward ranks (Chernozhukov et al., 2017; Hallin et al., 2021; Thurin et al., 2025; Vedula et al., 2023b), or the uniform cube in vector quantile regression (Carlier et al., 2016). The suboptimality is not specific to the Gaussian reference: Theorem 17 in Appendix D establishes the same volume gap for both the uniform ball and the spherical uniform — there without any excluded target and under a first moment only — and the discussion following it details the differences with the Gaussian case; the cube, not being spherically symmetric, falls outside the scope of our analysis. We nevertheless adopt the Gaussian reference: it retains the distribution-freeness of transport-based ranks (Hallin et al., 2021) while being supported on all of $\mathbb{R}^{d_y}$, which is what the smooth-flow correction of Section 4 requires (Proposition 5); a compactly supported reference would force boundary conditions on the velocity field.*

4 RECOVERING VOLUME OPTIMALITY BY REARRANGEMENT

As in Section 3, the statements of this section are conditional on $X = x$: we write $\mu := \mu_{Y|X=x}$, work under **A**1 and 2, and suppress x from the notation. Let $Q\colon \mathcal{U} \to \mathcal{Y}$ be an a.e.-invertible Brenier transport with $Q_{\#}\mu_U = \mu$. If $S\colon \mathcal{U} \to \mathcal{U}$ is a.e. invertible and measure preserving, $S_{\#}\mu_U = \mu_U$, then

$$(Q \circ S)_{\#}\mu_U = \mu, \qquad \mathcal{C}_{1-\alpha}^{Q\circ S} = Q\big(S(\mathcal{B}_{d_y}(0, r_{1-\alpha}))\big),$$

the second identity holding up to μ-null sets. Conversely, by Theorem 1, every a.e.-invertible transport T with $T_{\#}\mu_U = \mu$ is of this form with $S = Q^{-1} \circ T$. The map S thus changes the image of the reference ball without changing its probability — the degree of freedom we exploit to reduce volume.

4.1 MEASURE-PRESERVING FLOWS

Write $\mu_U(du) = f_U(u)\,du$ and assume in this subsection that f_U is smooth and strictly positive on $\mathbb{R}^{d_y}$. Let S_t be the flow of a time-dependent vector field v:

$$\partial_t S_t(u) = v(S_t(u), t), \qquad S_0(u) = u.$$

We make the following assumption on $v\colon \mathbb{R}^{d_y} \times \mathbb{R}_+ \to \mathbb{R}^{d_y}$. Its two parts play distinct roles: part (i) guarantees well-posedness of the flow and suffices for the sufficiency direction of Proposition 5; the higher regularity of part (ii) is used only for the necessity direction.

A3. *(i) v and $D_u v$ are continuous on $\mathbb{R}^{d_y} \times \mathbb{R}_+$, and for every $\bar{t} > 0$ there exists $c(\bar{t}) < \infty$ such that $\|v(u,t)\|_2 \le c(\bar{t})\,(1 + \|u\|_2)$ for all $(u,t) \in \mathbb{R}^{d_y} \times [0, \bar{t}]$.*
(ii) $v(\cdot, t) \in C^{d_y - 1}(\mathbb{R}^{d_y}, \mathbb{R}^{d_y})$ for every $t \ge 0$.

Under **A**3(i), every S_t is a globally defined C^1-diffeomorphism of $\mathbb{R}^{d_y}$, and

$$(S_t)_{\#}\mu_U = \mu_U \ \text{ for all } t \ge 0 \quad \Longleftrightarrow \quad \nabla\cdot\big(f_U v(\cdot, t)\big) = 0 \ \text{ on } \mathbb{R}^{d_y} \text{ for all } t \ge 0.$$

See Appendix F for more details. For a matrix field $A\colon \mathbb{R}^{d_y} \times \mathbb{R}_+ \to \mathbb{R}^{d_y \times d_y}$ we write $(\operatorname{div} A)_i := \sum_{j=1}^{d_y} \partial_j A_{ij}$. Under the full **A**3, Proposition 5 gives the following parametrization.

Proposition 5 (Measure-preserving vector fields)**.** *Let $d_y \ge 2$, let f_U be smooth and strictly positive on $\mathbb{R}^{d_y}$, and let v satisfy **A**3(i).*

(a) (Sufficiency.) If, for every $t \ge 0$,

$$v(\cdot, t) = \operatorname{div} A(\cdot, t) + A(\cdot, t)\nabla \log f_U \tag{2}$$

for a skew-symmetric matrix field $A(\cdot, t)\colon \mathbb{R}^{d_y} \to \mathbb{R}^{d_y \times d_y}$ with C^1 entries, jointly measurable in (u, t), then the flow of v preserves μ_U.
*(b) (Necessity.) Conversely, if v also satisfies **A**3(ii) and the flow of v preserves μ_U, then v admits the representation (2) for every $t \ge 0$; moreover, $A(\cdot, t)$ can be chosen with at most $2(d_y - 1)$ nonzero entries.*

The following corollary makes this proposition concrete for the standard Gaussian distribution, which is often employed as a reference distribution when constructing transport maps.

Corollary 6 (Gaussian reference)**.** *Let $\mu_U = \mathcal{N}(0, \mathrm{I}_{d_y})$ and let v satisfy **A**3(i). If, for every $t \ge 0$,*

$$v(u, t) = \operatorname{div} A(u, t) - A(u, t)u \tag{3}$$

*for a skew-symmetric matrix field $A(\cdot, t)$ with C^1 entries, then the flow of v preserves μ_U. Conversely, if v also satisfies **A**3(ii) and its flow preserves μ_U, then v is of the form (3) for every $t \ge 0$, with $A(\cdot, t)$ having at most $2(d_y - 1)$ nonzero entries.*

Proofs of these results are provided in Appendix F.

We parameterize the band entries $a_k = a_k(u, t, x; \theta)$ by networks that are C^2 in u, jointly continuous in (u, t) together with their first two u-derivatives, and bounded together with these derivatives uniformly in u, locally uniformly in (t, x). The resulting v satisfies **A**3(i), the linear growth coming from the terms $a_i u_{i+1}$ in (3) together with the uniform bounds on the a_k and their first derivatives.

Note that the parameterized field is only C^1 in u, so **A**3(ii) may fail for $d_y \geq 4$; this is harmless, since only the sufficiency direction of Corollary 6 is invoked: because the band field A has C^1 entries, exact preservation of μ_U follows from (3), and **A**3(ii) is not needed. We define $S_\theta(\cdot, x)$ as the time-one flow map and obtain its inverse by integrating the reversed field $\tilde{v}(u, s) = -v(u, 1-s)$.

Exact measure preservation allows for efficient computation of the Jacobian determinant. The change-of-variables formula gives $f_U(S_t(u)) \det D_u S_t(u) = f_U(u)$, and in the case where the source distribution is chosen to be standard Gaussian, it simplifies to

$$\det D_u S_\theta(u, x) = \frac{f_U(u)}{f_U(S_\theta(u, x))} = \exp\left\{ \frac{\|S_\theta(u, x)\|^2 - \|u\|^2}{2} \right\}. \tag{4}$$

4.2 Volume objective

To apply measure-preserving rearrangement to the reconstruction of HDR sets, we propose minimizing the volume of prediction regions. Such a volume admits a useful representation, given in the following proposition.

Proposition 7 (Volume under a rearrangement)**.** *Let $\mu_U(du) = f_U(u)\, du$ with $f_U > 0$ Lebesgue-a.e. on $\mathcal{B}_{d_y}(0, r)$, and let μ have density f. Let T and S be everywhere differentiable and injective outside μ_U-null sets, with $T_\# \mu_U = \mu$ and $S_\# \mu_U = \mu_U$. Then*

$$\mathrm{Vol}\big((T \circ S)(\mathcal{B}_{d_y}(0, r))\big) = \int_{\mathcal{B}_{d_y}(0,r)} \frac{|\det DT(S(u))|}{f_U(S(u))}\, \mathrm{d}\mu_U(u) = \int_{\mathcal{B}_{d_y}(0,r)} \frac{\mathrm{d}\mu_U(u)}{f(T(S(u)))}. \tag{5}$$

Thus, minimizing the volume moves $T(S(\mathcal{B}_{d_y}(0, r)))$ toward high-density points. At the population level, optimizing over the unrestricted class of measure-preserving rearrangements recovers the highest density region exactly.

Proposition 8 (Attainability)**.** *Let $\alpha \in (0, 1)$, let μ_U satisfy **A** 1, and let T be a.e. invertible with $T_\# \mu_U = \mu$, where μ has density $f > 0$ Lebesgue-a.e. Suppose there exists $\tau_{1-\alpha} > 0$ with $\mu\big(\{f \geq \tau_{1-\alpha}\}\big) = 1 - \alpha$, and set $\mathrm{HDR}^f_{1-\alpha} := \{f \geq \tau_{1-\alpha}\}$. Then*

$$\inf_{S : S_\# \mu_U = \mu_U} \mathrm{Vol}\big(\mathcal{C}^{T \circ S}_{1-\alpha}\big) = \mathrm{Vol}(\mathrm{HDR}^f_{1-\alpha}),$$

where the infimum is over almost-everywhere invertible measure-preserving maps and is attained; S is a minimizer if and only if $\mu_U\big(\mathcal{B}_{d_y}(0, r_{1-\alpha}) \triangle S^{-1}(T^{-1}(\mathrm{HDR}^f_{1-\alpha}))\big) = 0$.

A direct Monte Carlo approximation of the volume objective can be numerically unstable, since the integrand $1/f(T(S(u)))$ may vary over several orders of magnitude, particularly in higher dimensions. We therefore optimize a log-domain surrogate.

The characterization given in Proposition 7 suggests replacing $1/f$ by another strictly decreasing function of the density. The following lemma justifies this change.

Lemma 9. *Let $(\mathcal{Y}, \mu_Y)$ be a probability space, let $f \colon \mathcal{Y} \to (0, \infty)$ be measurable, and let $p \in (0, 1)$. Assume that there exists a threshold $c_p > 0$ such that $\mu_Y(\{f \geq c_p\}) = p$, and that f has no plateau at c_p, i.e. $\mu_Y(\{f = c_p\}) = 0$. Let $\phi \colon (0, \infty) \to \mathbb{R}$ be strictly decreasing, and assume that the necessary integrals are finite. Then*

$$\operatorname*{argmin}_{A : \mu_Y(A) = p} \int_A \phi(f(y))\, \mathrm{d}\mu_Y(y) = \operatorname*{argmax}_{A : \mu_Y(A) = p} \int_A f(y)\, \mathrm{d}\mu_Y(y) = \big\{f \geq c_p\big\},$$

where equality is understood up to μ_Y-null sets.

Proofs for this section are provided in Appendix G; Appendix H gives a finite-sample oracle inequality for $\widehat{L}$ when T is known.

By Proposition 8, the unrestricted class of a.e.-invertible measure-preserving rearrangements can attain the HDR. Taking $\phi(x) = 1/x$ recovers the volume objective, whereas the numerically more convenient choice $\phi(x) = -\log x$ has the same population HDR minimizer by Lemma 9. This yields the following log-domain objective:

$$
\begin{aligned}
L(S) &= \frac{1}{1-\alpha} \int_{\mathcal{B}_{d_y}(0, r_{1-\alpha})} -\log f(T(S(u)))\, \mathrm{d}\mu_U(u) \\
&= \widetilde{C} + \frac{1}{1-\alpha} \int_{\mathcal{B}_{d_y}(0, r_{1-\alpha})} \left[\log|\det DT(S(u))| + \frac{\|S(u)\|_2^2}{2} \right] \mathrm{d}\mu_U(u),
\end{aligned} \tag{6}
$$

where $\widetilde{C}$ is a constant independent of S; the second equality uses the change of variables, the chain rule, and the closed-form Jacobian of equation (4).

This modified objective is numerically more stable, since the Jacobian determinants are computed in the log domain and the objective admits an unbiased Monte Carlo estimate. Numerically, we replace it by its sample-based estimate and sample the coverage level uniformly at random from $(0, 1)$. We also condition the rearrangement on the context variable. Hence, for each point (y_i, x_i), we sample a coverage level α_i, followed by m points $u_i^j \sim \mu_U\big(\cdot \mid U \in \mathcal{B}_{d_y}(0, r_{1-\alpha_i})\big)$, and average over a batch of samples, giving the final finite-sample loss:

$$
\widehat{L}(S) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\log|\det D_u T(S(u_i^j, x_i), x_i)| + \frac{\|S(u_i^j, x_i)\|_2^2}{2} \right].
$$

For a level-independent rearrangement the regions are automatically nested in the level; with the level as an input of S, nestedness is not enforced and region crossing — the analogue of quantile crossing — may occur. At deployment a single level, fixed before calibration, is used, so validity is unaffected.

Remark 10 (Relation between surrogate and original objective)**.** *With $\mu_U^\alpha := \mu_U\big(\cdot \mid U \in \mathcal{B}_{d_y}(0, r_{1-\alpha})\big)$ and $A = T(S(\mathcal{B}_{d_y}(0, r_{1-\alpha})))$, Jensen's inequality gives*

$$
\begin{aligned}
&\log \mathrm{Vol}(A) = \log(1-\alpha) + L(S) + \Delta_\alpha(S), \\
&\Delta_\alpha(S) := \log \mathbb{E}_{\mu_U^\alpha}\left[e^{-\log f(T(S(u)))} \right] - \mathbb{E}_{\mu_U^\alpha}[-\log f(T(S(u)))] \geq 0.
\end{aligned}
$$

The surrogate is thus a lower bound on the log-volume, and the gap $\Delta_\alpha(S)$ does not vanish at the optimum (f is not constant on an HDR). The two objectives are aligned at the level of minimizers*, not of values: this is precisely the content of Lemma 9, which applies to $\phi(x) = 1/x$ and $\phi(x) = -\log x$ alike. Away from the optimum, decreasing the surrogate need not decrease the volume, whereas the original objective orders rearrangements exactly by volume.*

Remark 11 (On the parametrization of the rearrangement)**.** *Proposition 8 concerns the unrestricted class of a.e.-invertible measure-preserving rearrangements. The flow parametrization is more restrictive: S and, in our setting, T are diffeomorphisms, so the region $T(S(\mathcal{B}_{d_y}(0, r)))$ is diffeomorphic to a ball, and HDRs that are disconnected or contain holes cannot be recovered exactly. The population optimum is then unattainable within the parametrized class, and the caveat of the preceding remark applies.*

Calibration. Marginal coverage is restored by split conformal calibration, regardless of how well T and S are fitted. Let the calibration pairs $(X_j, Y_j)_{j=1}^n$ and the test pair (X_{n+1}, Y_{n+1}) be exchangeable and independent of the data used to fit $\hat{T}$ and $\hat{S}$. We select $R_i = \left\| \hat{S}^{-1}\Big(\hat{T}^{-1}(y_i, x_i), x_i\Big) \right\|_2$ as a conformal score.

Let $R_{(1)} \leq \cdots \leq R_{(n)} \leq R_{(n+1)} := \infty$ denote the order statistics, set $k = \lceil (n+1)(1-\alpha) \rceil$, and $\rho_{1-\alpha} = R_{(k)}$. The conformal set

$$
\hat{\mathcal{C}}_{1-\alpha}^{\hat{T} \circ \hat{S}}(x) = \left\{ y \colon \left\| \hat{S}^{-1}\Big(\hat{T}^{-1}(y, x), x\Big) \right\|_2 \leq \rho_{1-\alpha} \right\}
$$

then guarantees

$$
\mathbb{P}_{(Y,X) \sim \mu_{YX}} \left(Y \in \hat{\mathcal{C}}_{1-\alpha}^{\hat{T} \circ \hat{S}}(X) \right) \geq 1 - \alpha.
$$

Validity only requires the score to be a fixed measurable function of (y, x); in particular, it is unaffected when $\hat{T}^{-1}$ and $\hat{S}^{-1}$ are numerical approximations of the exact inverses.

$d\backslash k$	1 (control)	2	4	8	16	32
2	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	-.01 [-.01, -.01]	-.01 [-.02, -.01]	-.02 [-.02, -.02]
4	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	-.01 [-.01, -.01]	-.03 [-.04, -.03]	-.07 [-.07, -.06]
8	.00 [.00, .00]	.00 [.00, .00]	-.06 [-.06, -.06]	-.11 [-.12, -.11]	-.22 [-.22, -.21]	-.34 [-.36, -.33]
16	.00 [.00, .00]	.00 [-.01, .00]	-.08 [-.11, -.08]	-.30 [-.38, -.29]	-.73 [-.74, -.63]	-1.16 [-1.21, -.95]
$\mathcal{N}(0, \Sigma(k))$	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]

Table 1: Coverage-matched $\mathrm{NVF}(\hat{T}, 0.1)$ of the estimated Brenier map across dimensions d and anisotropy parameters k. The closer the metric is to 0, the smaller the gap in volumes. Entries report the median, with 25%–75% quantiles in brackets. The column $k = 1$ is the spherical negative control.

5 Numerical Experiments

We first demonstrate the discrepancy between optimal transport and HDR-generating transports on synthetic data, then show on real data that the volume of the induced prediction regions is efficiently reduced by measure-preserving rearrangement.

Throughout our experiments, we use Neural Optimal Transport (Kondratyev et al., 2026) as the base model for the Brenier map and RealNVP (Dinh et al., 2016) as the base model for triangular flows. All experiments are repeated over 10 seeds. The training procedure, architectural choices, and detailed experimental setup are described in Appendix I.

Our theory is stated for Brenier maps, but the rearrangement applies to any transport with computable $\log|\det DT|$; we also apply it to normalizing flows, showing numerically that the gap extends beyond optimal transport.

Illustrating the Brenier gap. To test the gap in Theorem 2 directly, we estimate the Brenier map (PICNN, Appendix I.3) from the standard Gaussian distribution to a Student-t target with $\nu = 3$ and $\Sigma = \mathrm{diag}(k^{1-1/d}, k^{-1/d}, \ldots, k^{-1/d})$, and report the coverage-matched, dimension-normalized log-volume ratio $\mathrm{NVF}(\hat{T}, \alpha) := \frac{1}{d}\big(\log \mathrm{Vol}(\mathrm{HDR}_c) - \log \mathrm{Vol}(\mathcal{C}^{\hat{T}}_{1-\alpha})\big)$, where c is the realized coverage of $\mathcal{C}^{\hat{T}}_{1-\alpha}$ (see Table 1). The column $k = 1$ is a negative control: the target is spherical, the Brenier map is the closed-form radial map, and the true gap is zero. Thus, $\mathrm{NVF}(k{=}1)$ measures the estimation error of the pipeline. Net of this control, the gap grows with both the anisotropy k and the dimension d. Since NVF is normalized by d, even a constant NVF corresponds to a volume ratio that grows exponentially with dimension. We also include a control experiment, denoted by $\mathcal{N}(0, \Sigma(k))$, in which we vary the degree of anisotropy to verify that our chosen solver can recover the optimal solution in the simple setting of Gaussian-to-Gaussian transport. In this setting, the optimal transport map is known to be linear. Thus, verifying that the coverage gap arises from the sub-optimality of optimal transport only.

Controlled examples. We next evaluate HDR reconstruction on three synthetic datasets (see Appendix I), pushforwards of a standard Gaussian by unit-Jacobian maps in dimension $d \in \{2, 4, 8\}$, so that HDR volumes are analytic. We report, before and after rearrangement, the Volume Ratio $\mathrm{VR}(\hat{T}) := \mathrm{Vol}(\mathrm{HDR}^f_{1-\alpha})/\,\mathrm{Vol}(\mathcal{C}^{\hat{T}}_{1-\alpha}) \leq 1$, with $\mathrm{VR} = 1$ for perfect reconstruction. Table 2 and Figure 4a (see Appendix I) show that the rearrangement drives VR close to one at unchanged coverage.

Real datasets. We evaluate on standard multi-target regression benchmarks used in previous work on uncertainty estimation (Plassier et al., 2025; Dheur et al., 2025; Kondratyev et al., 2026): `scm20d`, `blog`, `onp` (Tsymbalov et al., 2019; Fernandes et al., 2015) and `bio` (Feldman et al., 2023). For the single-target datasets `blog` and `bio`, we follow (Feldman et al., 2023) and augment the response with one of the input features as a second target. For `onp` we use eight of the variables as targets (see Appendix I.2). We additionally consider the `qm9` dataset (Wu et al., 2018), which contains approximately 130k small organic molecules, with four quantum-chemical properties as targets: dipole moment, isotropic polarizability, and the HOMO and LUMO energies. Input features are standardized to zero mean and unit variance; responses are left unmodified. The response dimensions are thus 16 (`scm20d`), 8 (`onp`), 4 (`qm9`) and 2 (`blog`, `bio`).

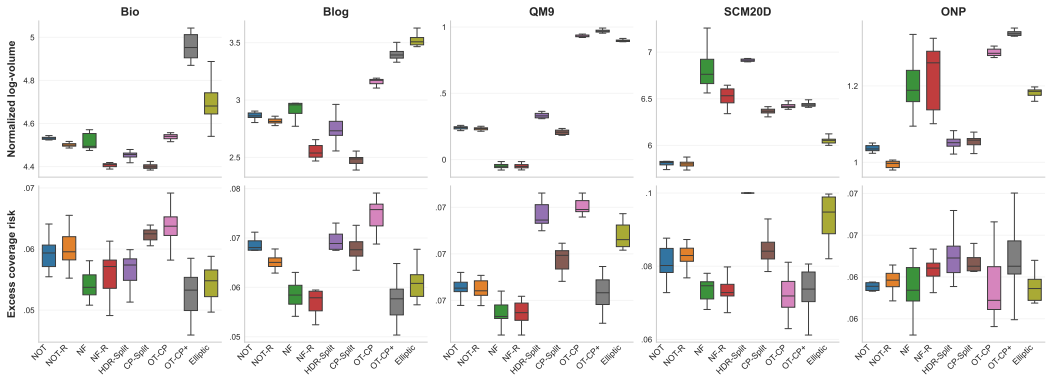


Figure 2: Normalized log-volume (lower is better) and Excess Coverage Risk (lower is better). NOT: neural optimal transport; NF: RealNVP flow with the latent-norm score, i.e. CONTRA (Fang et al., 2025); suffix -R: measure-preserving rearrangement of the base model.

Benchmark. In addition to analyzing the effect of rearrangement on the induced prediction regions, we also compare the resulting regions with current state-of-the-art methods. We consider optimal-transport-based rearrangements of residuals (OT-CP, OT-CP+ by Thurin et al. (2025)), along with elliptic calibrations (Ellipse from (Messoudi et al., 2022)). To calculate the residuals, we train a random forest regressor. We also add density-based (CP-split and HDR-split; Izbicki et al., 2022) and flow-based (CONTRA; Fang et al., 2025) models to our benchmark. Beyond optimal transport, we show that triangular normalizing flows benefit from the same rearrangement.

Metrics. Marginal coverage is attained by all methods on all datasets (Appendix I). We report the normalized log-volume and, to check that the rearrangement does not degrade conditional coverage, the excess coverage risk (ECR) (Braun et al., 2026b); worst-slab coverage (WSC) (Cauchois et al., 2021) and marginal coverage are tabulated in Appendix I. Figure 2 compares each base model with its rearrangement, together with the density-based baselines; See Appendix I for full benchmark.

Two patterns emerge from Figure 2. First, the rearrangement improves its own base transport in eight of the ten (dataset, base model) pairs while preserving conditional coverage (ECR) — the claim supported by our theory, since volume is a property of the representative chosen within the measure-preserving orbit; the two exceptions (RealNVP on `qm9` and `onp`) lie within interquartile ranges. Second, the comparison with density-based splits tracks the output dimension: CP-Split yields the smallest regions on the two-dimensional datasets, where conditional density estimation is easy, whereas rearranged transports dominate all baselines when $d_y \geq 4$. Transport-based regions are also natural when an invertible transport is needed anyway (ranks, quantiles, sampling), since CP-Split and HDR-Split do not allow sampling from the predicted region directly, only using costly methods such as rejection sampling.

6 Conclusion

We studied transport-based multivariate prediction regions through volume efficiency. For exact transports, the Brenier transport is strictly suboptimal on an open interval of levels for the non-spherical elliptical targets covered by our theorems, and the finite-second-moment Gaussian-reference exceptions are characterized exactly. Over measurable measure-preserving bijections the optimal population volume equals the HDR volume; sparse skew-symmetric fields generate smooth measure-preserving flows with a stochastic volume objective; split conformal calibration retains finite-sample marginal coverage for approximately fitted maps.

Limitations. Smooth flow images of a ball are diffeomorphic to a ball: disconnected HDRs can only be approximated, and the optimum is attained only over measurable rearrangements. The suboptimality theory covers elliptical targets with finite second moments; beyond marginal validity, conditional coverage is assessed empirically. The surrogate shares its population minimizer with the volume but is only a lower bound away from it; the oracle inequality of Appendix H covers a known transport only.

REPRODUCIBILITY STATEMENT

All theoretical results are stated with their assumptions and proved in Appendices B–H. Appendix I specifies the synthetic targets in closed form (with unit-Jacobian proofs), the real datasets with their sizes and splits, the architectures and every training hyperparameter of the base transports and of the rearrangement, the ODE solver settings, the volume estimators, and the definitions and estimators of all metrics; results are reported as medians with interquartile ranges over 10 seeds. Code reproducing all experiments is available at this anonymized repository URL.

ETHICS STATEMENT

This work is methodological. It uses publicly available benchmark datasets, involves no human subjects or personal data, and we foresee no direct negative societal impact. As with any uncertainty-quantification method, the marginal coverage guarantee should not be mistaken for a conditional one when prediction regions inform individual decisions.

AI USE STATEMENT

Large language models were used as an assistant for editing and proofreading the text, for checking the internal consistency of proofs and cross-references, and for reviewing experiment code. All technical content — theorems, proofs, algorithms, experiments and their interpretation — was produced and verified by the authors, who take full responsibility for it.

REFERENCES

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, 2nd edition, 2008.

Brandon Amos, Lei Xu, and J Zico Kolter. Input convex neural networks. In *International Conference on Machine Learning*, pp. 146–155. PMLR, 2017.

Brandon Amos et al. Tutorial on amortized optimization. *Foundations and Trends® in Machine Learning*, 16(5):592–732, 2023.

Anastasios N Angelopoulos, Stephen Bates, et al. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4):494–591, 2023.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. The limits of distribution-free conditional predictive inference. *Information and Inference*, 10(2):455–482, 2021.

Vladimir I. Bogachev. *Measure Theory*, volume 2. Springer, Berlin, 2007.

George E. P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison–Wesley, Reading, MA, 1973.

Axel Brando, Barcelona Supercomputing Center, José Rodríguez-Serrano, Jordi Vitrià, et al. Deep non-crossing quantiles through the partial derivative. In *International Conference on Artificial Intelligence and Statistics*, pp. 7902–7914. PMLR, 2022.

Sacha Braun, Liviu Aolaritei, Michael I Jordan, and Francis Bach. Minimum volume conformal sets for multivariate regression. *Journal of the American Statistical Association*, pp. 1–27, 2026a.

Sacha Braun, David Holzmüller, Michael I. Jordan, and Francis Bach. Conditional coverage diagnostics for conformal prediction. In *Forty-third International Conference on Machine Learning*, 2026b. URL https://openreview.net/forum?id=vaApZm6MKM.

Sacha Braun, Michael I. Jordan, and Francis Bach. Super-level-set regression: Conditional quantiles via volume minimization. *arXiv preprint arXiv:2605.06210*, 2026c.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

Charlotte Bunne, Andreas Krause, and Marco Cuturi. Supervised training of conditional monge maps. In *Advances in Neural Information Processing Systems*, volume 35, pp. 6859–6872, 2022.

Luis A. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992.

Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. Vector quantile regression: An optimal transport approach. *The Annals of Statistics*, 44(3):1165 – 1192, 2016.

Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. Vector quantile regression beyond the specified case. *Journal of Multivariate Analysis*, 161:96–102, 2017.

Guillaume Carlier, Victor Chernozhukov, Gwendoline De Bie, and Alfred Galichon. Vector quantile regression and optimal transport, from theory to numerics. *Empirical Economics*, 62(1):35–62, 2022.

Maxime Cauchois, Suyash Gupta, and John C Duchi. Knowing what you know: valid and validated confidence sets in multiclass and multilabel prediction. *Journal of Machine Learning Research*, 22(81):1–42, 2021.

Probal Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996. doi: 10.1080/01621459.1996.10476975.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge–kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223–256, 2017. doi: 10.1214/16-AOS1450.

Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118(48):e2107794118, 2021.

Nicolo Colombo. Normalizing flows for conformal regression. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024.

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

Victor Dheur, Matteo Fontana, Yorick Estievenart, Naomi Desobry, and Souhaib Ben Taieb. A unified comparative study with generalized conformity scores for multi-output conformal regression. In *Proceedings of the 42nd International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 2025.

Nathaniel Diamant, Ehsan Hajiramezanali, Tommaso Biancalani, and Gabriele Scalia. Conformalized deep splines for optimal and efficient prediction sets. In *International Conference on Artificial Intelligence and Statistics*, pp. 1657–1665. PMLR, 2024.

Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: Non-linear independent components estimation. In *International Conference on Learning Representations, Workshop Track*, 2015.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

John H. J. Einmahl and David M. Mason. Generalized quantile processes. *The Annals of Statistics*, 20(2):1062–1078, 1992.

Kai-Tai Fang, Samuel Kotz, and Kai Wang Ng. *Symmetric Multivariate and Related Distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1990.

Zhenhan Fang, Aixin Tan, and Jian Huang. Contra: Conformal prediction region via normalizing flow transformation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Shai Feldman, Stephen Bates, and Yaniv Romano. Calibrated multiple-output quantile regression with representation learning. *Journal of Machine Learning Research*, 24(24):1–48, 2023.

Kelwin Fernandes, Pedro Vinagre, and Paulo Cortez. A proactive intelligent decision support system for predicting the popularity of online news. In *Portuguese conference on artificial intelligence*, pp. 535–546. Springer, 2015.

Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, Princeton and Oxford, 2018.

Chao Gao, Liren Shan, Vaidehi Srinivas, and Aravindan Vijayaraghavan. Volume optimality in conformal prediction with structured prediction sets. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 18495–18527, 2025.

Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis Bach. Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Promit Ghosal and Bodhisattva Sen. Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *The Annals of Statistics*, 50(2):1012–1037, 2022.

Arnab Ghosh, Harkirat Behl, Emilien Dupont, Philip Torr, and Vinay Namboodiri. Steer: Simple temporal regularization for neural ode. *Advances in neural information processing systems*, 33: 14831–14843, 2020.

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. FFJORD: Free-form continuous dynamics for scalable reversible generative models. In *International Conference on Learning Representations*, 2019.

Marc Hallin and Dimitri Konen. Multivariate quantiles: Geometric and measure-transportation-based contours. In *Applications of Optimal Transport to Economics and Related Topics*, pp. 61–78. Springer, 2024.

Marc Hallin and Miroslav Šiman. Multiple-output quantile regression. In Roger Koenker, Victor Chernozhukov, Xin He, and Limin Peng (eds.), *Handbook of Quantile Regression*, pp. 185–207. Chapman & Hall/CRC, 2017.

Marc Hallin, Eustasio Del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139–1165, 2021.

Philip Hartman. *Ordinary Differential Equations*, volume 38 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 2nd edition, 2002.

David Holzmüller, Léo Grinsztajn, and Ingo Steinwart. Better by default: Strong pre-tuned mlps and boosted trees on tabular data. In *Neural Information Processing Systems*, 2024.

Chin-Wei Huang, Ricky TQ Chen, Christos Tsirigotis, and Aaron Courville. Convex potential flows: Universal probability distributions with optimal transport and convex optimization. In *International Conference on Learning Representations*, 2021.

Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166–1194, 2021.

Rob J Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50 (2):120–126, 1996.

Rafael Izbicki, Gilson Shimizu, and Rafael Stern. Flexible distribution-free conditional predictive bands using density estimators. 108:3068–3077, 26–28 Aug 2020. URL `https://proceedings.mlr.press/v108/izbicki20a.html`.

Rafael Izbicki, Gilson Shimizu, and Rafael B Stern. Cd-split and hpd-split: Efficient conformal regions in high dimensions. *Journal of Machine Learning Research*, 23(87):1–32, 2022.

Chancellor Johnstone and Bruce Cox. Conformal uncertainty sets for robust optimization. In *Conformal and Probabilistic Prediction and Applications (COPA)*, pp. 72–90, 2021.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

Patrick Kidger, Ricky TQ Chen, and Terry J Lyons. " hey, that's not an ode": Faster ode adjoints via seminorms. In *ICML*, pp. 5443–5452, 2021.

Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Michal Klein, Louis Béthune, Eugene Ndiaye, and Marco Cuturi. Multivariate conformal prediction using optimal transport. *Transactions on Machine Learning Research*, 2026.

Martin Knott and Cyril S Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43(1):39–49, 1984.

Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005. doi: 10.1017/CBO9780511754098.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International Conference on Machine Learning*, pp. 5361–5370, 2020.

Vladimir Kondratyev, Alexander Fishkov, Mahmoud Hegazy, Nikita Kotelevskii, Rémi Flamary, Maxim Panov, and Eric Moulines. Neural optimal transport meets multivariate conformal prediction. In *International Conference on Learning Representations*, 2026.

Alexander Korotin, Vage Egiazarian, Arip Asadulaev, Alexander Safin, and Evgeny Burnaev. Wasserstein-2 generative networks. In *International Conference on Learning Representations*, 2019.

Alexander Korotin, Lingxiao Li, Aude Genevay, Justin M Solomon, Alexander Filippov, and Evgeny Burnaev. Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark. In *Advances in Neural Information Processing Systems*, volume 34, pp. 14593–14605, 2021a.

Alexander Korotin, Lingxiao Li, Justin Solomon, and Evgeny Burnaev. Continuous wasserstein-2 barycenter estimation without minimax optimization. In *International Conference on Learning Representations*, 2021b.

Jing Lei, James Robins, and Larry Wasserman. Distribution-free prediction sets. *Journal of the American Statistical Association*, 108(501):278–287, 2013.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 2018.

Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2nd edition, 2001.

Rui Luo and Zhixin Zhou. Density-sorted prediction set: Efficient conformal prediction for multi-target regression. *Pattern Recognition*, 172:112513, 2026. doi: 10.1016/j.patcog.2025.112513.

Laura Lützow, Simone Garatti, Marco C. Campi, Lars Lindemann, and Matthias Althoff. Multivariable conformal prediction: Optimizing prediction sets without data splitting. *arXiv preprint arXiv:2605.12341*, 2026.

Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pp. 6672–6681. PMLR, 2020.

Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323, 1995.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, 120:108101, 2021.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Ellipsoidal conformal inference for multi-target regression. In *Conformal and Probabilistic Prediction and Applications (COPA)*, pp. 294–306, 2022.

Guillaume Morel, Lucas Drumetz, Simon Benaïchouche, Nicolas Courty, and François Rousseau. Turning normalizing flows into monge maps with geodesic gaussian preserving flows. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=2UQv8L1Cv9.

Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231:289–337, 1933.

Harris Papadopoulos, Kostas Proedrou, Vladimir Vovk, and Alexander Gammerman. Inductive confidence machines for regression. In *European Conference on Machine Learning*, pp. 345–356, 2002.

Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Vincent Plassier, Alexander Fishkov, Mohsen Guizani, Maxim Panov, and Eric Moulines. Probabilistic conformal prediction with approximate conditional validity. In *The Thirteenth International Conference on Learning Representations*, 2024.

Vincent Plassier, Alexander Fishkov, Victor Dheur, Mohsen Guizani, Souhaib Ben Taieb, Maxim Panov, and Eric Moulines. Rectifying conformity scores for better conditional coverage. In *Forty-second International Conference on Machine Learning*, 2025.

Wolfgang Polonik. Minimum volume sets and generalized quantile processes. *Stochastic Processes and their Applications*, 69(1):1–24, 1997.

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.

Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Aviv A Rosenberg, Sanketh Vedula, Yaniv Romano, and Alexander Bronstein. Fast nonlinear vector quantile regression. In *The Eleventh International Conference on Learning Representations*, 2023.

Max Sampson and Kung-Sik Chan. Highest probability density conformal regions. *Journal of Computational and Graphical Statistics*, 35(1):241–250, 2026. doi: 10.1080/10618600.2025.2520584.

Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, 2015.

Clayton D. Scott and Robert D. Nowak. Learning minimum volume sets. *Journal of Machine Learning Research*, 7:665–704, 2006.

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

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

Sophia Huiwen Sun and Rose Yu. Copula conformal prediction for multi-step time series prediction. In *The Twelfth International Conference on Learning Representations*, 2024.

Amirhossein Taghvaei and Amin Jalali. 2-wasserstein approximation via restricted convex potentials with application to improved training for gans. *arXiv preprint arXiv:1902.07197*, 2019.

Gauthier Thurin, Kimia Nadjahi, and Claire Boyer. Optimal transport-based conformal prediction. In *Forty-second International Conference on Machine Learning*, 2025.

Alexandre B. Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969, 1997.

Evgenii Tsymbalov, Sergei Makarychev, Alexander Shapeev, and Maxim Panov. Deeper connections between neural networks and gaussian processes speed-up active learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, IJCAI'19, pp. 3599–3605. AAAI Press, 2019. ISBN 9780999241141.

John W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1974*, volume 2, pp. 523–531, 1975.

Théo Uscidda and Marco Cuturi. The monge gap: A regularizer to learn all transport maps. In *International Conference on Machine Learning*, pp. 34709–34733. PMLR, 2023.

Sanketh Vedula, Irene Tallini, Aviv A Rosenberg, Marco Pegoraro, Emanuele Rodolà, Yaniv Romano, and Alexander Bronstein. Continuous vector quantile regression. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023a.

Sanketh Vedula, Irene Tallini, Aviv A. Rosenberg, Marco Pegoraro, Emanuele Rodolà, Yaniv Romano, and Alexander M. Bronstein. Continuous vector quantile regression. In *ICML 2023 Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023b.

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 2009.

Vladimir Vovk. Conditional validity of inductive conformal predictors. In *Asian Conference on Machine Learning*, pp. 475–490, 2012.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, New York, 2005.

Zhendong Wang, Ruijiang Gao, Mingzhang Yin, Mingyuan Zhou, and David M. Blei. Probabilistic conformal prediction using conditional random samples. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 8814–8836, 2023.

Zheyu Oliver Wang, Ricardo Baptista, Youssef Marzouk, Lars Ruthotto, and Deepanshu Verma. Efficient neural network approaches for conditional optimal transport with applications in bayesian inference. *SIAM Journal on Scientific Computing*, 47(4):C979–C1005, 2025.

Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

Ruiyao Zhang, Ping Zhou, and Tianyou Chai. Improved copula-based conformal prediction for uncertainty quantification of multi-output regression. *Journal of Process Control*, 129:103036, 2023.

Yanfei Zhou, Lars Lindemann, and Matteo Sesia. Conformalized adaptive forecasting of heterogeneous trajectories. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

A Extended related work

This appendix complements the discussion in the introduction with a broader account of the literature, organized along the components of our construction: conformal prediction, multivariate conformal regions, transport-based multivariate quantiles, minimum-volume sets, the estimation of optimal transport maps, and measure-preserving flows.

Conformal prediction. Conformal prediction originates in the work of Vovk and collaborators (Vovk et al., 2005; Shafer & Vovk, 2008); the split (inductive) variant used in this paper goes back to Papadopoulos et al. (2002), and its modern statistical treatment to Lei et al. (2018); see (Angelopoulos et al., 2023) for a recent survey. Much of the univariate literature concerns the design of the conformity score: residual-based scores, quantile-based scores (Romano et al., 2019), and rank- or distribution-based scores (Chernozhukov et al., 2021), the latter being the univariate shadow of the multivariate rank $\|T^{-1}(y, x)\|_2$ employed here. A parallel line delineates what is achievable: exact conditional coverage is impossible without distributional assumptions (Vovk, 2012; Barber et al., 2021). This impossibility is one motivation for our viewpoint. Under exchangeability, calibration gives marginal validity for any fixed score. The choice of score then determines the efficiency of the region, which we measure by Lebesgue volume.

Multivariate conformal prediction. Extending conformal prediction to $\mathbb{R}^{d_y}$ requires either a scalar score summarizing a multivariate residual or a genuinely set-valued construction. The first family includes norms of (standardized) residuals and Mahalanobis-type scores, which produce ellipsoidal regions (Johnstone & Cox, 2021; Messoudi et al., 2022), and copula-based aggregation of coordinatewise scores (Messoudi et al., 2021; Sun & Yu, 2024); see (Dheur et al., 2025) for a systematic empirical comparison of multi-output constructions. These methods impose a fixed geometry (ellipsoids, boxes, or copula level sets). They can be efficient when the conditional law matches that geometry and can lose efficiency otherwise. The second family is density-based: Lei et al. (2013) construct distribution-free prediction sets from density level sets and prove asymptotic volume optimality, and the CD-split and HPD-split procedures of Izbicki et al. (2020; 2022) calibrate highest-predictive-density regions built from conditional density estimates. These share our objective — the HDR benchmark — but reach it through explicit conditional density estimation, which is delicate in moderate dimension. Related procedures include highest-probability-density conformal regions (Sampson & Chan, 2026) and density-sorted prediction sets based on conditional normalizing flows (Luo & Zhou, 2026). A third family is generative: probabilistic conformal prediction (Wang et al., 2023) takes unions of balls around samples from a conditional generative model, and a rapidly growing body of work builds regions from normalizing flows and related generative transports (Fang et al., 2025; Colombo, 2024; Brando et al., 2022; Diamant et al., 2024; Plassier et al., 2024; Zhang et al., 2023; Zhou et al., 2024; Wang et al., 2025). Flow-based regions are, in our terminology, transport-based regions: the flow is a map T pushing a reference onto (an estimate of) $\mu_{Y|X=x}$, and the region is the image of a reference set. Many such transports are trained primarily by likelihood rather than directly for prediction-set volume; density-sorted prediction sets are an important exception because they explicitly threshold the learned conditional density. Our question is different: for a fixed transport, what can be achieved by moving within its measure-preserving rearrangement orbit? Pre-composition leaves population coverage unchanged, the measurable orbit contains an HDR minimizer, and improvement within the smooth flow class is an approximation question. Representation-learning approaches such as Feldman et al. (2023) calibrate spherical quantiles in a learned latent space and face the analogous issue: the latent geometry is a modeling choice, not a volume-optimal one. Finally, Gao et al. (2025) establish near-optimal volume within prescribed finite-VC set families, while Braun et al. (2026a) optimize flexible set parametrizations under conformal constraints. Concurrent work develops direct super-level-set regression (Braun et al., 2026c) and joint shape optimization and calibration without data splitting (Lützow et al., 2026). Our contribution differs in expressing invertible transport-based regions through measure-preserving rearrangements, in proving that the uncorrected optimal-transport choice is strictly suboptimal (see Theorems 2 and 18), and in establishing that the HDR volume is attained within the measurable rearrangement family (see Proposition 8).

Multivariate quantiles, ranks, and optimal transport. The absence of a canonical order on $\mathbb{R}^{d_y}$ has produced several notions of multivariate quantile: halfspace depth (Tukey, 1975), spa-

tial and geometric quantiles (Chaudhuri, 1996), and directional multiple-output quantile regression (Hallin & Šiman, 2017; Koenker, 2005). Monge–Kantorovich depth (Chernozhukov et al., 2017) and center-outward distribution functions (Hallin et al., 2021) define quantile regions as images of nested reference balls under an optimal transport map, with distribution-freeness and consistency established in Ghosal & Sen (2022); see Hallin & Konen (2024) for a recent account. On the regression side, vector quantile regression (Carlier et al., 2016; 2017; 2022) and its scalable variants (Rosenberg et al., 2023; Vedula et al., 2023a) estimate conditional Brenier maps, and the conformal instantiations (Thurin et al., 2025; Klein et al., 2026; Kondratyev et al., 2026) are the immediate context of this paper. All of these constructions fix the rearrangement to the identity, i.e. select the Brenier representative in the polar factorization. Theorems 2 and 18 show that this choice incurs a strict volume gap for broad classes of elliptical targets, and that the answer depends on the reference measure — a dependence already implicit in the literature on center-outward quantiles, where the spherical uniform reference is standard, but whose efficiency consequences had, to the best of our knowledge, not been isolated.

Minimum-volume sets and highest density regions. That density superlevel sets minimize volume at fixed probability content is classical: it is the Neyman–Pearson comparison (Neyman & Pearson, 1933), the highest-posterior-density construction of Box & Tiao (1973, Sec. 2.8), the bathtub principle (Lieb & Loss, 2001, Thm. 1.14), and the operational definition of HDRs in Hyndman (1996). The statistical literature studies the estimation of such sets directly: generalized quantile processes (Einmahl & Mason, 1992), minimum-volume set estimation (Polonik, 1997; Scott & Nowak, 2006), and density level-set estimation (Tsybakov, 1997). These estimators construct the set from a density estimate and inherit its curse of dimensionality; finite-sample coverage is not their objective. Our construction is complementary: the set is obtained as the image of a calibrated reference ball under a learned map, and marginal coverage is guaranteed by conformalization irrespective of estimation error. Restricted volume optimality within finite-complexity set classes (Gao et al., 2025) and direct volume-based conditional level-set regression (Braun et al., 2026c) pursue related efficiency targets under different admissible classes and statistical guarantees.

Optimal transport: theory and map estimation. The structural results underlying the paper are Brenier's theorem and polar factorization (Brenier, 1991; Knott & Smith, 1984), McCann's moment-free uniqueness (McCann, 1995), and Caffarelli's regularity theory (Caffarelli, 1992), which supports the differentiability assumptions of Section 4.2 under suitable conditions on the source and target densities; see Villani (2009); Santambrogio (2015); Galichon (2018) for comprehensive treatments. On the computational side, entropic regularization (Cuturi, 2013; Peyré et al., 2019) and stochastic dual solvers (Genevay et al., 2016; Seguy et al., 2018) underpin large-scale estimation, with statistical rates for transport-map estimation established by Hütter & Rigollet (2021) and, for entropic estimators, by Pooladian & Niles-Weed (2021). Neural estimators parameterize the Brenier potential by input-convex networks (Amos et al., 2017; Makkuva et al., 2020; Taghvaei & Jalali, 2019; Korotin et al., 2019; 2021b;a; Huang et al., 2021), possibly with Monge-gap regularization (Uscidda & Cuturi, 2023) or amortized conditioning (Bunne et al., 2022; Amos et al., 2023). Any of these estimators can serve as the frozen transport T of Section 4.2; the conformal step protects marginal coverage against their estimation error, and the rearrangement is trained to improve their region geometry.

Measure-preserving flows and volume control in generative models. Our parametrization of rearrangements as flows relies on the well-posedness theory of the continuity equation (Ambrosio et al., 2008, Sec. 8.1) and is implemented with neural ODEs (Chen et al., 2018). The contrast with continuous normalizing flows (Grathwohl et al., 2019) is direct: there, the map is unconstrained and the change of volume is *computed* by integrating a divergence along the trajectory; here, the map is constrained to preserve μ_U exactly, and the Jacobian determinant is available in closed form (see Section 4.1, equation (4)). Discrete-time analogues include volume-preserving couplings (Dinh et al., 2015) and invertible architectures with tractable log-determinants (Kingma & Dhariwal, 2018); divergence-free vector fields also appear in equivariant flows for symmetric densities (Köhler et al., 2020), where they preserve a group-invariant reference. We use exactly measure-preserving flows as a *post-hoc efficiency layer* for prediction regions, with the skew-potential representation of Proposition 5 and a band structure containing $O(d_y)$ nonzero matrix entries.

There is also work using measure-preserving flows in the context of constructing optimal transport solutions (Morel et al., 2023). They first project the distribution onto a cube and then parametrize a volume-preserving flow in this new domain. There are two main differences from our approach: (i) our solution is exact, up to numerical integration error, whereas the parametrization in the previous work relies on heuristics to satisfy the boundary conditions for the dynamics on the cube; and (ii) we provide a sparse parametrization, making the numerical implementation more efficient.

Summary of positioning. Relative to this literature, the paper contributes: (i) strict volume-suboptimality results for optimal-transport prediction regions, for both the spherical-uniform and the Gaussian reference, with a characterization, among finite-second-moment elliptical targets, of the exceptional cases for the Gaussian reference (Theorems 2 and 18); (ii) a factorization of admissible invertible corrections and a sparse skew-potential representation of smooth measure-preserving flow fields (Proposition 5); (iii) a proof that the HDR volume is attained over measurable measure-preserving a.e.-bijections (Proposition 8); and (iv) a modular framework — transport estimation, rearrangement learning, conformal calibration — in which validity and efficiency are handled in separate stages.

B Optimality of Lebesgue volume of HDR sets

Lemma 12 (Strict optimality of highest density regions)**.** *Let μ be a probability measure with density f with respect to the Lebesgue measure on $\mathbb{R}^{d_y}$. Fix $p \in (0,1)$ and suppose that, for some $\tau_p \in \mathbb{R}$, the set*

$$H_p := \{y \in \mathbb{R}^{d_y} : f(y) \geq \tau_p\}$$

satisfies $\mu(H_p) = p$. Then $\tau_p > 0$ and $\mathrm{Vol}(H_p) \leq p/\tau_p < \infty$, and for every measurable set $A \subseteq \mathbb{R}^{d_y}$ such that $\mu(A) \geq p$,

$$\mathrm{Vol}(A) \geq \mathrm{Vol}(H_p),$$

with equality if and only if $\mathrm{Vol}(A \triangle H_p) = 0$; in that case $\mu(A) = p$.

Proof. Since $f \geq 0$, if $\tau_p \leq 0$ then $H_p = \mathbb{R}^{d_y}$ and $\mu(H_p) = 1 > p$; hence $\tau_p > 0$. Markov's inequality gives $p = \int_{H_p} f \geq \tau_p \,\mathrm{Vol}(H_p)$, so $\mathrm{Vol}(H_p) \leq p/\tau_p < \infty$.

Let A be measurable with $\mu(A) \geq p$; we may assume $\mathrm{Vol}(A) < \infty$. Subtracting $\mu(A \cap H_p)$ from $\mu(A) \geq p = \mu(H_p)$ yields

$$\int_{A \setminus H_p} f(y)\, dy \;\geq\; \int_{H_p \setminus A} f(y)\, dy. \tag{7}$$

By definition of H_p, $f < \tau_p$ pointwise on $A \setminus H_p \subseteq H_p^{\mathrm{c}}$ and $f \geq \tau_p$ on $H_p \setminus A \subseteq H_p$. Therefore

$$\tau_p \,\mathrm{Vol}(A \setminus H_p) \;\geq\; \int_{A \setminus H_p} f \;\geq\; \int_{H_p \setminus A} f \;\geq\; \tau_p \,\mathrm{Vol}(H_p \setminus A). \tag{8}$$

Dividing by $\tau_p > 0$ and adding $\mathrm{Vol}(A \cap H_p) \leq \mathrm{Vol}(H_p) < \infty$ gives $\mathrm{Vol}(A) \geq \mathrm{Vol}(H_p)$.

Suppose now $\mathrm{Vol}(A) = \mathrm{Vol}(H_p)$. Both volumes being finite, $\mathrm{Vol}(A \setminus H_p) = \mathrm{Vol}(H_p \setminus A) =: V$, so the two outer terms of (8) are equal and every inequality in (8) is an equality. If $V > 0$, then, since $\tau_p - f > 0$ pointwise on $A \setminus H_p$ and $\mathrm{Vol}(A \setminus H_p) = V > 0$, the first inequality in (8) is strict, a contradiction. Hence $V = 0$, that is, $\mathrm{Vol}(A \triangle H_p) = 0$, and then $\mu(A) = \mu(H_p) = p$ because $\mu \ll \mathrm{Leb}$. The converse implication is immediate. □

Remark 13. *Lemma 12 is classical. The minimal-volume property of highest (posterior) density regions goes back to Box & Tiao (1973, Sec. 2.8) (see also Hyndman, 1996); the comparison argument in* (8) *is the one used in the Neyman–Pearson lemma (Neyman & Pearson, 1933). In measure-theoretic form the statement is the "bathtub principle" (Lieb & Loss, 2001, Thm. 1.14), whose uniqueness clause corresponds exactly to the hypothesis that the* non-strict *superlevel set H_p carries mass exactly p; no assumption on the level set $\{f = \tau_p\}$ is needed.*

Lemma 14 (HDRs of strictly decreasing elliptical laws)**.** *Let μ be a probability measure with density*

$$f(y) = C\, g\big((y-m)^\top \Sigma^{-1}(y-m)\big), \qquad y \in \mathbb{R}^{d_y},$$

where Σ is symmetric positive definite, $C > 0$, and $g : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and strictly decreasing. Then $g > 0$ on $\mathbb{R}_+$, and for every $p \in (0,1)$ there exists a unique $q_p > 0$ such that the ellipsoid

$$H_p := \{y \in \mathbb{R}^{d_y} : (y-m)^\top \Sigma^{-1}(y-m) \le q_p\}$$

satisfies $\mu(H_p) = p$. Moreover, with $\tau_p := Cg(q_p)$,

$$H_p = \{y : f(y) \ge \tau_p\}, \qquad \mathrm{Vol}\big(\{y : f(y) = \tau_p\}\big) = \mathrm{Vol}(\partial H_p) = 0.$$

In particular, for every $p \in (0,1)$ the exact-mass hypothesis of Lemma 12 is satisfied, and H_p is the Lebesgue-a.e. unique minimum-volume set of μ-mass p, i.e. the HDR of μ at level p.

Proof. If $g(s_0) = 0$ for some $s_0 \ge 0$, strict decrease would force $g(s) < 0$ for $s > s_0$, contradicting $g \ge 0$; hence $g > 0$ everywhere, and $f > 0$ on $\mathbb{R}^{d_y}$.

Write $A := \Sigma^{1/2}$ and $\rho_\Sigma(y) := \|A^{-1}(y-m)\|_2$, so that $H_p = \{\rho_\Sigma^2 \le q_p\}$. The change of variables $y = m + Az$ gives, for $t \ge 0$,

$$M(t) := \mu\big(\{\rho_\Sigma \le t\}\big) = d_y\, \omega_{d_y}\, C \det(A) \int_0^t s^{d_y - 1} g(s^2)\, ds,$$

where $\omega_{d_y} := \mathrm{Vol}(\mathcal{B}_{d_y}(0,1))$. Since the integrand is continuous and strictly positive for $s > 0$, M is continuous and strictly increasing on $[0,\infty)$, with $M(0) = 0$ and $\lim_{t\to\infty} M(t) = 1$. Hence for each $p \in (0,1)$ there is a unique $t_p > 0$ with $M(t_p) = p$; set $q_p := t_p^2$.

For the identification of H_p as a superlevel set, note that by monotonicity $\rho_\Sigma^2(y) \le q_p$ implies $f(y) = Cg(\rho_\Sigma^2(y)) \ge Cg(q_p) = \tau_p$; conversely $f(y) \ge \tau_p$ means $g(\rho_\Sigma^2(y)) \ge g(q_p)$, which by *strict* decrease of g forces $\rho_\Sigma^2(y) \le q_p$. Hence $H_p = \{f \ge \tau_p\}$ exactly, and $\mu(\{f \ge \tau_p\}) = p$. By injectivity of g, $\{f = \tau_p\} = \{\rho_\Sigma^2 = q_p\} = \partial H_p$, which is the image of the unit sphere under the affine map $z \mapsto m + t_p A z$ and therefore Lebesgue-null. Finally, since the map $\tau \mapsto \mu(\{f \ge \tau\})$ is, by the above, the strictly decreasing continuous bijection $\tau \mapsto M(g^{-1}(\tau/C)^{1/2})$ on the relevant range, τ_p is the unique threshold satisfying the exact-mass condition in the definition of the HDR in Section 2. The last assertion follows from Lemma 12. □

Remark 15. *For the density form and geometry of elliptical distributions see Fang et al. (1990, Ch. 2); the identification of the HDRs of unimodal elliptical laws with Mahalanobis ellipsoids is folklore (see, e.g., Hallin et al., 2021, App. C, for the same whitening in the measure-transport context), and a complete proof is given above for self-containedness.*

C THE BANANA EXAMPLE

C.1 THE TARGET AND ITS OPTIMAL REGIONS

Let $Z = (Z_1, Z_2) \sim \mathcal{N}(0, \mathrm{I}_2)$ and, for $b \in \mathbb{R}$, define

$$T_b(z_1, z_2) := \big(z_1, z_2 + b(z_1^2 - 1)\big), \qquad \mu_b := (T_b)_\# \mathcal{N}(0, \mathrm{I}_2).$$

The subtraction of one centers the second coordinate. Since

$$T_b^{-1}(y_1, y_2) = \big(y_1, y_2 - b(y_1^2 - 1)\big), \qquad \det DT_b = 1,$$

the density of μ_b is

$$f_b(y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[y_1^2 + \big(y_2 - b(y_1^2 - 1)\big)^2\right]\right\}. \tag{9}$$

Let $q_{1-\alpha}$ denote the $(1-\alpha)$ quantile of χ^2_2. Here

$$q_{1-\alpha} = -2 \log \alpha.$$

Because (9) is a decreasing function of $\|T_b^{-1}(y)\|^2$, its $(1-\alpha)$ highest-density region is exactly

$$\mathrm{HDR}_{1-\alpha}^{f_b} = T_b\big(\mathcal{B}_2(0, \sqrt{q_{1-\alpha}})\big). \tag{10}$$

The unit Jacobian gives

$$\mathrm{Vol}\big(\mathrm{HDR}_{1-\alpha}^{f_b}\big) = \pi q_{1-\alpha}. \tag{11}$$

Thus the nonlinear triangular map T_b sends every centered Gaussian ball to an optimal region, simultaneously for all α.

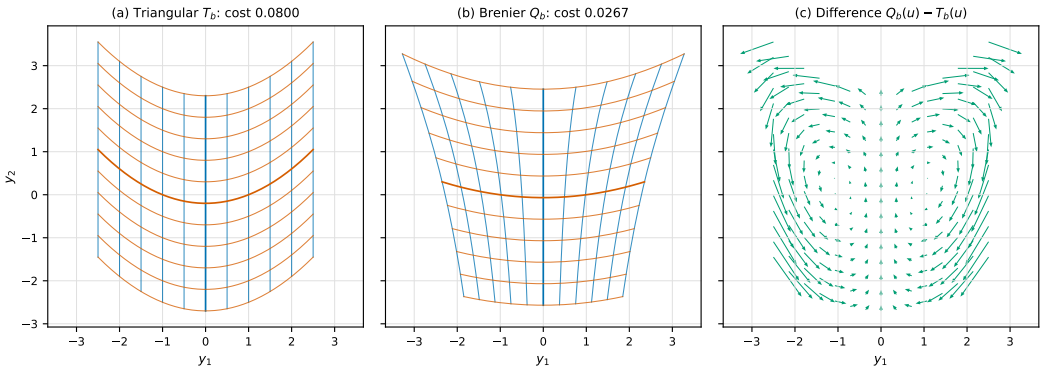


Figure 3: Images of a Cartesian grid under T_b and the numerical Brenier map Q_b for $b = 0.2$, followed by their difference. Both maps have pushforward μ_b, but only Q_b minimizes quadratic transport cost, while only T_b maps every centered ball to the corresponding highest-density region.

C.2 The Brenier map

Let $Q_b = \nabla\varphi_b$ be the Brenier map from $\mathcal{N}(0, \mathrm{I}_2)$ to μ_b. Although T_b has the correct pushforward, it is not a gradient when $b \neq 0$, because its Jacobian is not symmetric. It therefore differs from Q_b. Under the smooth small-b expansion stated below, the transport costs already separate the two maps:

$$\mathbb{E}\|T_b(Z) - Z\|^2 = 2b^2, \qquad \mathbb{E}\|Q_b(Z) - Z\|^2 = \frac{2}{3}b^2 + o(b^2).$$

The Brenier potential solves the Gaussian Monge–Ampère equation

$$f_b(\nabla\varphi_b(u)) \det D^2\varphi_b(u) = \frac{1}{2\pi}e^{-\|u\|^2/2}.$$

The following calculation is formal: we do not prove here the required smooth dependence of the Brenier potential on b. Conditional on that regularity, the remainders below are understood on each fixed bounded set, and the expansion is used only as an analytic illustration; the main theorems do not depend on it. Expanding

$$\varphi_b(u) = \frac{1}{2}\|u\|^2 + b\psi(u) + b^2\chi(u) + O(b^3)$$

and matching powers of b gives

$$\psi(u) = \frac{1}{3}(u_1^2 - 1)u_2,$$

and

$$\chi(u) = -\frac{u_1^4}{18} + \frac{5u_1^2u_2^2}{18} - \frac{u_1^2}{18} + \frac{u_2^2}{6}.$$

Consequently,

$$\begin{aligned} Q_{b,1}(u) &= u_1 + \frac{2b}{3}u_1u_2 + \frac{b^2}{9}u_1(-2u_1^2 + 5u_2^2 - 1) + O(b^3), \\ Q_{b,2}(u) &= u_2 + \frac{b}{3}(u_1^2 - 1) + \frac{b^2}{9}u_2(5u_1^2 + 3) + O(b^3). \end{aligned} \tag{12}$$

In particular,

$$\det DQ_b(u) = 1 + \frac{2b}{3}u_2 + \frac{b^2}{9}(2 - 5u_1^2 + 5u_2^2) + O(b^3).$$

D A RIGIDITY LEMMA AND THE UNIFORM-BALL REFERENCE

Throughout this appendix, "a.e." refers to the Lebesgue measure, and $\triangle$ denotes the symmetric difference.

The proof of Theorem 2 — restated as part (i) of Theorem 18 in Appendix E — and the proof of the uniform-ball counterpart stated below (Theorem 17) share a common core, which we isolate as a lemma: a rigidity statement for spherically symmetric references and elliptical targets. We first describe its setting.

Setting. Let $d_y \geq 2$ and $\bar{R} \in (0, \infty]$. Let μ_U be a spherically symmetric probability measure on $\mathbb{R}^{d_y}$ with $\mathbb{E}\|U\|_2 < \infty$ for $U \sim \mu_U$, whose radial distribution function

$$G(r) := \mathbb{P}\big(\|U\|_2 \leq r\big)$$

satisfies $G(0) = 0$ and $\lim_{r \to \bar{R}} G(r) = 1$ and is continuously differentiable with $G' > 0$ on $(0, \bar{R})$. Then μ_U is absolutely continuous, with Lebesgue density $G'(\|u\|_2)/(d_y \omega_{d_y} \|u\|_2^{d_y - 1})$ on $\mathcal{B}_{d_y}(0, \bar{R}) \setminus \{0\}$, where $\omega_{d_y} := \mathrm{Vol}(\mathcal{B}_{d_y}(0, 1))$: this follows by polar disintegration together with $G(r) = \int_0^r G'(s)\, ds$, which holds since G is C^1 on $(0, \bar{R})$ with $G(0) = 0$. In particular μ_U and the Lebesgue measure restricted to $\mathcal{B}_{d_y}(0, \bar{R})$ have the same null sets. Let μ have density (1), where Σ is symmetric positive definite and g is continuous and strictly decreasing, and assume $\int \|y\|_2 f(y)\, dy < \infty$. Let $A := \Sigma^{1/2}$ be the symmetric positive definite square root of Σ, and let $Z := A^{-1}(Y - m)$ for $Y \sim \mu$. The change of variables $y = m + Az$ shows that Z has the spherical density $\bar{f}(z) = \bar{C}\, g(\|z\|_2^2)$ with $\bar{C} := C \det(A)$. If g vanished at some s_0, strict decrease would force $g < 0$ beyond s_0; hence $g > 0$ everywhere, so $f > 0$ on $\mathbb{R}^{d_y}$ and μ and the Lebesgue measure have the same null sets. Define the radial distribution function

$$F_\nu(t) := \mathbb{P}(\|Z\|_2 \leq t) = d_y\, \omega_{d_y}\, \bar{C} \int_0^t s^{d_y - 1} g(s^2)\, ds, \qquad t \geq 0.$$

Then $F_\nu \in C^1([0, \infty))$ with $F_\nu'(t) = d_y \omega_{d_y} \bar{C}\, t^{d_y - 1} g(t^2) > 0$ for $t > 0$, $F_\nu(0) = 0$ and $F_\nu(t) \to 1$. Consequently

$$a := F_\nu^{-1} \circ G$$

restricts to an increasing C^1-diffeomorphism of $(0, \bar{R})$ onto $(0, \infty)$, with $a(0+) = 0$. Let $\rho_\Sigma(y) := \|A^{-1}(y - m)\|_2$ denote the Mahalanobis radius and

$$H_t := \{y : \rho_\Sigma(y) \leq t\}, \qquad t \geq 0,$$

so that $\mu(H_t) = F_\nu(t)$. Finally, let $Q = \nabla\varphi$ be a Brenier map with $Q_{\#}\mu_U = \mu$, and let $Q^{-1} := \nabla\varphi^*$ denote its almost-everywhere inverse, where φ^* is the convex conjugate of φ, so that $Q^{-1} \circ Q = \mathrm{id}$ μ_U-a.e. and $(Q^{-1})_{\#}\mu = \mu_U$ (McCann, 1995); see also Santambrogio (2015, Sec. 1.3) for the finite-second-moment case. Write

$$h(y) := \|\nabla\varphi^*(y)\|_2, \qquad C_r := \{y : h(y) \leq r\},$$

so that $\mu(C_r) = \mu_U\big(\mathcal{B}_{d_y}(0, r)\big) = G(r)$ for every $r > 0$ by the inverse pushforward.

Lemma 16 (Radial–angular rigidity)**.** *In the setting above, assume in addition:*

(a) $\int |u^\top y|\, d\pi(u, y) < \infty$ *for every coupling* $\pi \in \Pi(\mu_U, \mu)$ *of* μ_U *and* μ*;*

(b) $\Sigma \notin \{\sigma \mathrm{I} : \sigma > 0\}$*;*

(c) there is no $c > 0$ *with* $a(r) = cr$ *for all* $r \in (0, \bar{R})$*.*

Then, for every $r \in (0, \bar{R})$, $\mu(C_r) = \mu(H_{a(r)}) = G(r)$ *and* $H_{a(r)} = \mathrm{HDR}^f_{G(r)}$*. Moreover,*

$$B_{\mathrm{bad}} := \big\{r \in (0, \bar{R}) : \mathrm{Vol}\big(C_r \triangle H_{a(r)}\big) > 0\big\}$$

contains a nonempty open interval, and $\mathrm{Vol}(C_r) > \mathrm{Vol}\big(H_{a(r)}\big)$ *for every* $r \in B_{\mathrm{bad}}$*.*

Taking the lemma for granted, results for specific reference measures follow by instantiation. The Gaussian instantiation, which proves Theorem 2, is carried out in Appendix E. We record here the instantiation for the two compactly supported references used in the literature on transport-based multivariate ranks and quantile regions: the uniform distribution on a ball, and the *spherical uniform* distribution — the law of $R'\Theta$ with $R' \sim \mathrm{Unif}(0, R)$ and Θ uniform on the unit sphere, independent — which underlies the center-outward distribution function of Chernozhukov et al. (2017) and Hallin et al. (2021). The two laws must not be confused: the former has constant density on the ball, whereas the latter has density $\big(R\, d_y \omega_{d_y} \|u\|_2^{d_y-1}\big)^{-1}$ on the punctured ball, which is *unbounded at the origin* when $d_y \geq 2$. The rigidity lemma nevertheless covers both, because its hypotheses bear only on the radial distribution function G, not on the density.

Theorem 17 (Compactly supported spherical references)**.** *Let $d_y \geq 2$ and let μ_U be either*

(i) the uniform distribution on $\mathcal{B}_{d_y}(0, R)$, in which case set $r_{1-\alpha} := R(1-\alpha)^{1/d_y}$; or
(ii) the spherical uniform distribution on $\mathcal{B}_{d_y}(0, R)$, in which case set $r_{1-\alpha} := R(1-\alpha)$;

in both cases $\mu_U\big(\mathcal{B}_{d_y}(0, r_{1-\alpha})\big) = 1-\alpha$. Let μ have density (1) with $\int \|y\|_2 f(y)\, dy < \infty$, where Σ is symmetric positive definite with $\Sigma \notin \{\sigma \mathrm{I} : \sigma > 0\}$ and g is continuous and strictly decreasing. Let Q be the Brenier map from μ_U to μ and $\mathcal{C}^Q_{1-\alpha} := \{y : \|Q^{-1}(y)\|_2 \leq r_{1-\alpha}\}$. Then there exists a nonempty open interval $I \subseteq (0,1)$ such that, for every $\alpha \in I$,

$$\mathrm{Vol}\big(\mathcal{C}^Q_{1-\alpha}\big) > \mathrm{Vol}\big(\mathrm{HDR}^f_{1-\alpha}\big).$$

Proof. In case (i), $\bar{R} = R$ and $G(r) = (r/R)^{d_y}$ on $[0, R]$, continuously differentiable with $G'(r) = d_y r^{d_y-1}/R^{d_y} > 0$ on $(0, R)$; in case (ii), $\bar{R} = R$ and $G(r) = r/R$ on $[0, R]$, with $G' \equiv 1/R > 0$. In both cases $G(0) = 0$, $G(R) = 1$ and $\mathbb{E}\|U\|_2 \leq R < \infty$, so the setting of Lemma 16 applies. Hypothesis (a) of the lemma holds in both cases because $\|u\|_2 \leq R$ on the support of μ_U and $\int \|y\|_2 f(y)\, dy < \infty$, so $\int |u^\top y|\, d\pi \leq R \int \|y\|_2\, d\mu < \infty$ for every coupling; (b) is assumed.

For (c), suppose $a(r) = cr$ on $(0, R)$ for some $c > 0$, i.e. $F_\nu(cr) = G(r)$ on $(0, R)$. Differentiating and inserting the formula $F'_\nu(t) = d_y \omega_{d_y} \bar{C}\, t^{d_y-1} g(t^2)$ gives, in case (i),

$$g(c^2 r^2) = \big(\bar{C}\, \omega_{d_y} c^{d_y} R^{d_y}\big)^{-1} \qquad \text{for all } r \in (0, R):$$

g would be *constant*, contradicting its strict decrease. In case (ii) the same computation gives

$$g(c^2 r^2) = \big(R\, d_y \omega_{d_y} \bar{C}\, c^{d_y}\big)^{-1} r^{-(d_y-1)} \qquad \text{for all } r \in (0, R):$$

since $d_y \geq 2$, g would be *unbounded* as $r \downarrow 0$, contradicting its continuity at 0. In both cases (c) holds.

Lemma 16 therefore applies and yields a nonempty open interval $J \subseteq (0, R)$ of radii with $\mathrm{Vol}(C_r) > \mathrm{Vol}(H_{a(r)})$, together with the identities $\mu(C_r) = G(r)$ and $H_{a(r)} = \mathrm{HDR}^f_{G(r)}$ for every $r \in (0, R)$. Since $G(r_{1-\alpha}) = 1-\alpha$, these identities give $\mathcal{C}^Q_{1-\alpha} = C_{r_{1-\alpha}}$ and $\mathrm{HDR}^f_{1-\alpha} = H_{a(r_{1-\alpha})}$. In both cases $r \mapsto 1 - G(r)$ is a C^1-diffeomorphism of $(0, R)$ onto $(0, 1)$, so the image I of J under this map is a nonempty open interval of levels with $\mathrm{Vol}(\mathcal{C}^Q_{1-\alpha}) > \mathrm{Vol}(\mathrm{HDR}^f_{1-\alpha})$ for every $\alpha \in I$, proving the theorem. □

Comparison with the Gaussian reference. Theorem 17 and Theorem 2 follow from the same rigidity lemma, but they differ in three respects; we spell these out because each difference is easy to overlook.

(1) Moment assumptions. For a compactly supported reference, hypothesis (a) of the lemma — finiteness of $\int |u^\top y|\, d\pi$ for *every* coupling π, not merely the optimal one — is automatic under a first moment on the target, since $|u^\top y| \leq R\|y\|_2$ on the support. For the Gaussian reference, u is unbounded and the bound is obtained from the Cauchy–Schwarz inequality, which consumes a *second* moment of the target; whether a first moment suffices in that case is open.

(2) Exceptional targets. Hypothesis (c) — the radial map a is not linear — fails exactly when the target is an affine image of the reference; then T_{HDR} is affine, hence a convex gradient, so it *is* the Brenier map and its regions coincide with the HDRs at every level. Whether such targets exist *inside*

the class (1) depends on the reference. For the Gaussian reference, the affine images are the Gaussian laws, which do belong to the class: the exception is genuine, and this is why Theorem 2 must exclude exponential profiles. For the uniform ball, the affine image is the uniform law on an ellipsoid, whose profile is constant where positive: it violates the *strict decrease* of g. For the spherical uniform, the affine image has density proportional to $\rho_\Sigma(y)^{-(d_y-1)}$ on an ellipsoid: its profile is strictly decreasing where positive, and what fails is the *continuity of g at zero* — the density is unbounded at the mode — together with positivity beyond the ellipsoid. The same hypothesis is thus secured by three different clauses of the target class; for the two compactly supported references the exceptional set is empty, and no exclusion needs to be stated in Theorem 17.

(3) Regularity of the reference, and the correction step. The uniform ball has a density that is discontinuous on $\mathbb{R}^{d_y}$ (it vanishes outside its support), and the spherical uniform has a density that is unbounded at the origin. Neither is covered by the smooth-flow machinery of Section 4, which requires a smooth, strictly positive reference density on all of $\mathbb{R}^{d_y}$; the rigidity lemma, by contrast, only requires the radial distribution function to be C^1 with positive derivative, which all three references satisfy. The suboptimality phenomenon therefore covers the references of the center-outward literature, while the correction mechanism is developed for the Gaussian reference (cf. Remark 4). A last, cosmetic difference is the radius calibration: $r_{1-\alpha} = R(1-\alpha)$ for the spherical uniform — the linearity expressing the distribution-freeness of center-outward ranks — $R(1-\alpha)^{1/d_y}$ for the uniform ball, and the χ_{d_y} quantile $G_{d_y}^{-1}(1-\alpha)$ for the Gaussian.

Proof of Lemma 16. The mass identities are immediate: $\mu(C_r) = G(r)$ was noted in the setting, and $\mu(H_{a(r)}) = F_\nu(a(r)) = G(r)$. For the HDR identification, fix $r \in (0, \bar{R})$ and set $p := G(r) \in (0,1)$. By Lemma 14 there is a unique $q_p > 0$ such that the ellipsoid $\{y : (y-m)^\top \Sigma^{-1}(y-m) \le q_p\} = H_{\sqrt{q_p}}$ has μ-mass p and equals HDR_p^f; since $\mu(H_{a(r)}) = p$ and $t \mapsto \mu(H_t)$ is strictly increasing, $\sqrt{q_p} = a(r)$, i.e. $H_{a(r)} = \mathrm{HDR}_{G(r)}^f$. The remainder of the proof establishes the statement about B_{bad}, in five steps.

Step 1: a transport generating the HDR family. Define $T_{\mathrm{HDR}}(0) := m$ and

$$T_{\mathrm{HDR}}(u) := m + A\, a(\|u\|_2)\, \frac{u}{\|u\|_2}, \qquad u \in \mathcal{B}_{d_y}(0,\bar{R}) \setminus \{0\}.$$

Write $U = \rho\,\Theta$ with $\rho := \|U\|_2$ and $\Theta := U/\rho$. Since μ_U is spherically symmetric and $\mathbb{P}(\rho = 0) = G(0) = 0$, the direction Θ is uniform on the unit sphere and independent of ρ, and $\mathbb{P}(\rho \le r) = G(r)$. Hence, for $t \ge 0$, $\mathbb{P}(a(\rho) \le t) = \mathbb{P}\big(\rho \le a^{-1}(t)\big) = F_\nu(t)$, i.e. $a(\rho)$ is distributed as $\|Z\|_2$. Because Z has a spherical density, $Z = \|Z\|_2\, W'$ with $W' := Z/\|Z\|_2$ uniform on the sphere and independent of $\|Z\|_2$; therefore $a(\rho)\Theta$ has the law of Z, and $T_{\mathrm{HDR}}(U) = m + A\,a(\rho)\Theta$ has the law of Y. Thus $(T_{\mathrm{HDR}})_{\#}\mu_U = \mu$, and by construction

$$\rho_\Sigma\big(T_{\mathrm{HDR}}(u)\big) = a(\|u\|_2) \qquad \text{for all } u \in \mathcal{B}_{d_y}(0,\bar{R}) \setminus \{0\}. \tag{13}$$

Step 2: T_{HDR} is not a.e. equal to the gradient of a convex function. Let $\varrho(r) := a(r)/r$ for $r \in (0,\bar{R})$; then $\varrho \in C^1((0,\bar{R}))$ is positive, and $T_{\mathrm{HDR}}(u) = m + \varrho(\|u\|_2)\,Au$ is C^1 on the punctured ball, with

$$DT_{\mathrm{HDR}}(u) = \varrho(\|u\|_2)\,A + \frac{\varrho'(\|u\|_2)}{\|u\|_2}\,Auu^\top.$$

The first term is symmetric. For the second, $Auu^\top$ is symmetric if and only if $Auu^\top = uu^\top A$, which (applying both sides to u) holds if and only if Au is collinear with u, i.e. u is an eigenvector of A. Since $\Sigma \notin \{\sigma \mathrm{I}\}$ by (b), the symmetric matrix A has at least two distinct eigenvalues, so the union of its eigenspaces is a finite union of proper linear subspaces and is Lebesgue-null.

Next, ϱ' does not vanish identically: otherwise $\varrho \equiv c > 0$ on $(0,\bar{R})$, i.e. $a(r) = cr$, which is excluded by (c). Since ϱ' is continuous, $G_0 := \{r \in (0,\bar{R}) : \varrho'(r) \neq 0\}$ is open and nonempty, and

$$N := \big\{u : \|u\|_2 \in G_0\big\} \setminus \{u : Au \text{ collinear with } u\}$$

has positive Lebesgue measure, with $DT_{\mathrm{HDR}}(u)$ non-symmetric for every $u \in N$.

Suppose now, towards a contradiction with later steps, that $T_{\mathrm{HDR}} = \nabla\psi$ a.e. on $\mathcal{B}_{d_y}(0,\bar{R})$ for some convex ψ. Then ψ is finite on a dense subset of the open ball; since every point of the ball lies in a simplex with vertices in that dense set and $\operatorname{dom}\psi$ is convex, ψ is finite (and continuous) on all of it. By Alexandrov's theorem (Villani, 2009, Thm. 14.25), $\nabla\psi$ is approximately differentiable a.e. with *symmetric* approximate differential. Approximate differentials depend only on the Lebesgue-equivalence class of the map, and wherever T_{HDR} is classically differentiable its approximate differential is DT_{HDR}. Hence $DT_{\mathrm{HDR}}(u)$ would be symmetric for a.e. u, contradicting $\mathrm{Vol}(N) > 0$. Therefore no convex ψ with $T_{\mathrm{HDR}} = \nabla\psi$ a.e. exists.

Step 3: radial rigidity. Suppose, for contradiction, that

$$\mathrm{Vol}\big(C_r \triangle H_{a(r)}\big) = 0 \qquad \text{for Lebesgue-a.e. } r \in (0,\bar{R}). \tag{14}$$

Since $\mathrm{Vol}(C_r \triangle H_{a(r)}) = \int_{\mathbb{R}^{d_y}} \mathbb{1}\{\mathbb{1}\{h(y) \le r\} \neq \mathbb{1}\{\rho_\Sigma(y) \le a(r)\}\}\,\mathrm{d}y$, Tonelli's theorem applied on $\mathbb{R}^{d_y} \times (0,\bar{R})$ shows that for a.e. $y \in \mathbb{R}^{d_y}$,

$$\mathbb{1}\{r \ge h(y)\} = \mathbb{1}\{\rho_\Sigma(y) \le a(r)\} = \mathbb{1}\{r \ge a^{-1}(\rho_\Sigma(y))\} \qquad \text{for a.e. } r \in (0,\bar{R}),$$

where we used that a is an increasing bijection of $(0,\bar{R})$ onto $(0,\infty)$. Moreover $0 < h(y) < \bar{R}$ and $\rho_\Sigma(y) > 0$ for μ-a.e. (hence a.e.) y: indeed $\mu(\{h = 0\}) = \mu_U(\{0\}) = 0$ and $\mu(\{h \ge \bar{R}\}) = \mu_U(\{u : \|u\|_2 \ge \bar{R}\}) = 0$ by $(\nabla\varphi^*)_{\#}\mu = \mu_U$, while $\{\rho_\Sigma = 0\} = \{m\}$. For each such y, the two indicator functions of r above are right-continuous step functions on $(0,\bar{R})$ with a single jump, at $h(y)$ and at $a^{-1}(\rho_\Sigma(y))$ respectively; agreeing for a.e. r, they must have the same jump location. Hence

$$\rho_\Sigma(y) = a\big(h(y)\big) \qquad \text{for a.e. } y \in \mathbb{R}^{d_y}.$$

Since μ and Lebesgue share null sets, this identity holds μ-a.e.; pulling it back through Q (using $Q_{\#}\mu_U = \mu$ and $h(Q(u)) = \|\nabla\varphi^*(\nabla\varphi(u))\|_2 = \|u\|_2$ for μ_U-a.e. u) yields

$$\rho_\Sigma\big(Q(u)\big) = a(\|u\|_2) \qquad \text{for } \mu_U\text{-a.e. } u. \tag{15}$$

Step 4: angular rigidity. For $\pi \in \Pi(\mu_U,\mu)$ set $J(\pi) := \int u^\top y\, d\pi(u,y)$, which is finite by (a). Since Q is the gradient of a convex function, the coupling $\pi_Q := (\mathrm{id}, Q)_{\#}\mu_U$ maximizes J over $\Pi(\mu_U,\mu)$ (Knott & Smith, 1984; Brenier, 1991). (The cited results are stated for finite second moments, which we do not assume; under (a) the claim follows from Fenchel–Young: $\varphi(u) + \varphi^*(Q(u)) = u^\top Q(u)$ for μ_U-a.e. u, and since φ and φ^* admit affine minorants integrable against μ_U and μ respectively ($\mathbb{E}\|U\|_2 < \infty$ and $\int \|y\|_2\, d\mu < \infty$), while the sum of their integrals equals $J(\pi_Q) < \infty$, both $\int \varphi\, d\mu_U$ and $\int \varphi^*\, d\mu$ are finite; hence for every $\pi \in \Pi(\mu_U,\mu)$, $J(\pi) \le \int(\varphi(u) + \varphi^*(y))\, d\pi = \int \varphi\, d\mu_U + \int \varphi^*\, d\mu = J(\pi_Q)$.) Let

$$\Pi_{\mathrm{rad}} := \big\{\pi \in \Pi(\mu_U,\mu) : \|A^{-1}(y-m)\|_2 = a(\|u\|_2)\ \pi\text{-a.s.}\big\};$$

by (15), $\pi_Q \in \Pi_{\mathrm{rad}}$, and therefore π_Q also maximizes J over Π_{rad}, whose supremum equals the unconstrained one.

We first record that $\mathbb{E}[\rho\, a(\rho)] < \infty$: the coupling $\pi_{\mathrm{HDR}} := (\mathrm{id}, T_{\mathrm{HDR}})_{\#}\mu_U$ belongs to $\Pi(\mu_U,\mu)$ by Step 1, so (a) gives $\mathbb{E}\big|U^\top T_{\mathrm{HDR}}(U)\big| < \infty$; since $U^\top T_{\mathrm{HDR}}(U) = U^\top m + \rho\, a(\rho)\, \Theta^\top A\Theta$ with $\mathbb{E}\|U\|_2 < \infty$ and $\Theta^\top A\Theta \ge \lambda_{\min}(A) > 0$, the claim follows.

Let $(U,Y) \sim \pi \in \Pi_{\mathrm{rad}}$ and set $\rho := \|U\|_2$, $\Theta := U/\rho$, $V := A^{-1}(Y-m)$ and $W := V/\|V\|_2$; these are a.s. well defined since $\rho > 0$ a.s. and $\|V\|_2 = a(\rho) > 0$ a.s. As in Step 1, Θ is uniform on the sphere and independent of ρ, and W is uniform on the sphere and independent of $\|V\|_2$. Since a is strictly increasing, $\rho = a^{-1}(\|V\|_2)$, so W is also independent of ρ. Using $\mathbb{E}[U] = 0$ (spherical symmetry) and $Y = m + AV$,

$$J(\pi) = \mathbb{E}\big[U^\top AV\big] = \mathbb{E}\big[\rho\, a(\rho)\, \Theta^\top AW\big] = \mathbb{E}\big[\rho\, a(\rho)\ \mathrm{Tr}(AM_\rho)\big], \qquad M_\rho := \mathbb{E}\big[W\Theta^\top \mid \rho\big].$$

The conditional expectation is needed: W and Θ are each independent of ρ, but their coupling may depend on ρ. Conditionally on ρ, both Θ and W are uniform on the sphere, so for all unit vectors v, w, by Cauchy–Schwarz,

$$\big|v^\top M_\rho w\big| = \big|\mathbb{E}[(v^\top W)(w^\top\Theta) \mid \rho]\big| \le \big(\mathbb{E}[(v^\top W)^2 \mid \rho]\big)^{1/2}\big(\mathbb{E}[(w^\top\Theta)^2 \mid \rho]\big)^{1/2} = \frac{1}{d_y}.$$

Diagonalizing $A = \sum_{i=1}^{d_y} \lambda_i e_i e_i^\top$ with $\lambda_i > 0$ gives $\mathrm{Tr}(AM_\rho) = \sum_i \lambda_i\, e_i^\top M_\rho e_i \le \mathrm{Tr}(A)/d_y$, and hence

$$J(\pi) \;\le\; \mathbb{E}[\rho\, a(\rho)]\, \frac{\mathrm{Tr}(A)}{d_y} \;<\; \infty,$$

with equality for $\pi_{\mathrm{HDR}} \in \Pi_{\mathrm{rad}}$, for which $W = \Theta$ a.s. Since π_Q attains the supremum of J over Π_{rad}, we get, under π_Q, $\mathbb{E}\big[\rho\, a(\rho)\big(\mathrm{Tr}(A)/d_y - \mathrm{Tr}(AM_\rho)\big)\big] = 0$ with a nonnegative integrand and $\rho\, a(\rho) > 0$ a.s.; hence $\mathrm{Tr}(AM_\rho) = \mathrm{Tr}(A)/d_y$ for a.e. ρ. Because every λ_i is positive, this forces $e_i^\top M_\rho e_i = 1/d_y$ for every i; since both conditional second moments equal $1/d_y$, this gives $\mathbb{E}[(e_i^\top W - e_i^\top \Theta)^2 \mid \rho] = 2/d_y - 2/d_y = 0$ a.s., i.e. $e_i^\top W = e_i^\top \Theta$ a.s. for each i. Hence $W = \Theta$ a.s. under π_Q, i.e.

$$Q(u) = m + A\, a(\|u\|_2)\, \frac{u}{\|u\|_2} = T_{\mathrm{HDR}}(u) \qquad \text{for } \mu_U\text{-a.e. } u.$$

Since $Q = \nabla\varphi$ μ_U-a.e. with φ convex, and μ_U has the same null sets as Lebesgue on $\mathcal{B}_{d_y}(0, \bar{R})$, T_{HDR} would coincide a.e. with a convex gradient, contradicting Step 2. Therefore, hypothesis (14) fails:

$$B_{\mathrm{bad}} = \big\{r \in (0, \bar{R}) : \mathrm{Vol}\big(C_r \triangle H_{a(r)}\big) > 0\big\} \quad \text{has positive Lebesgue measure.} \tag{16}$$

Step 5: an open interval of radii, and conclusion. We first show that B_{bad} is open. Fix $k \in \mathbb{N}$ and let $K := \overline{\mathcal{B}}_{d_y}(0, k)$. For every $r \in (0, \bar{R})$, $(\nabla\varphi^*)_{\#}\mu = \mu_U$ gives $\mu(\{h = r\}) = \mu_U(\{u : \|u\|_2 = r\}) = 0$ because G is continuous; equivalence of μ and Lebesgue measure therefore gives $\mathrm{Vol}(\{h = r\}) = 0$. Also, $\mathrm{Vol}(\{\rho_\Sigma = t\}) = 0$ for every t, since $\{\rho_\Sigma = t\}$ is an affine image of a sphere; by monotone convergence along the nested families $(C_r)_r$ and $(H_t)_t$, the maps $r \mapsto \mathrm{Vol}(C_r \cap K)$ and $r \mapsto \mathrm{Vol}(H_{a(r)} \cap K)$ are continuous on $(0, \bar{R})$. Using $\mathbb{1}_{C \triangle H} \le \mathbb{1}_{C \triangle C'} + \mathbb{1}_{C' \triangle H'} + \mathbb{1}_{H' \triangle H}$ and nestedness,

$$\begin{aligned}&\big|\mathrm{Vol}\big((C_{r'} \triangle H_{a(r')}) \cap K\big) - \mathrm{Vol}\big((C_r \triangle H_{a(r)}) \cap K\big)\big| \\ &\qquad \le \big|\mathrm{Vol}(C_{r'} \cap K) - \mathrm{Vol}(C_r \cap K)\big| + \big|\mathrm{Vol}(H_{a(r')} \cap K) - \mathrm{Vol}(H_{a(r)} \cap K)\big|,\end{aligned}$$

so $r \mapsto \mathrm{Vol}\big((C_r \triangle H_{a(r)}) \cap K\big)$ is continuous for each k, and $B_{\mathrm{bad}} = \bigcup_{k \ge 1} \big\{r : \mathrm{Vol}\big((C_r \triangle H_{a(r)}) \cap \overline{\mathcal{B}}_{d_y}(0, k)\big) > 0\big\}$ is open. By (16) it is nonempty, hence contains a nonempty open interval.

Finally, fix $r \in B_{\mathrm{bad}}$ and set $p := G(r)$. The sets C_r and $H_{a(r)} = \mathrm{HDR}_p^f$ have the same μ-mass p, and by Lemma 14 the exact-mass hypothesis of Lemma 12 holds at level p. Since $\mathrm{Vol}\big(C_r \triangle H_{a(r)}\big) > 0$, the equality case of Lemma 12 is excluded, whence $\mathrm{Vol}(C_r) > \mathrm{Vol}\big(H_{a(r)}\big)$. □

E THE GAUSSIAN-REFERENCE CASE

Throughout this appendix, $\mu_U = \mathcal{N}(0, \mathrm{I}_{d_y})$ is the standard Gaussian reference and

$$G_{d_y}(r) := \mathbb{P}(\|U\|_2 \le r), \qquad U \sim \mu_U.$$

The density $G'_{d_y}(r) = 2^{1-d_y/2} r^{d_y - 1} \mathrm{e}^{-r^2/2} / \Gamma(d_y/2)$ is positive on $(0, \infty)$. For $\alpha \in (0, 1)$, let $r_{1-\alpha} := G_{d_y}^{-1}(1 - \alpha)$ and

$$\mathcal{C}_{1-\alpha}^Q := \{y \in \mathbb{R}^{d_y} : \|Q^{-1}(y)\|_2 \le r_{1-\alpha}\}.$$

This appendix first gives the proof of Theorem 2, and then completes it into a dichotomy (Theorem 18 below): within the elliptical class (1) with finite second moment, the targets excluded from Theorem 2 — Gaussian targets and spherical targets — are exactly those for which the Brenier construction *does* generate the HDR family.

Proof of Theorem 2. We verify the hypotheses of Lemma 16 in Section D, with $\bar{R} = \infty$ and $G = G_{d_y}$. The radial distribution function G_{d_y} satisfies $G_{d_y}(0) = 0$ and $G_{d_y}(r) \to 1$ and is continuously

differentiable with $G'_{d_y} > 0$ on $(0,\infty)$, and $\mathbb{E}\|U\|_2 < \infty$; the second-moment assumption implies $\int \|y\|_2 f(y)\,\mathrm{d}y < \infty$. Hypothesis (a) holds by Cauchy–Schwarz: for every coupling π of μ_U and μ,

$$\int |u^\top y|\,\mathrm{d}\pi(u,y) \le \Big(\mathbb{E}\|U\|_2^2\Big)^{1/2}\Big(\int \|y\|_2^2 f(y)\,\mathrm{d}y\Big)^{1/2} < \infty.$$

Hypothesis (b) is assumed. For (c), suppose $a(r) = cr$ for every $r > 0$ and some $c > 0$, where $a = F_\nu^{-1} \circ G_{d_y}$ is the radial map of the lemma. Differentiating $F_\nu(cr) = G_{d_y}(r)$ and cancelling r^{d_y-1} gives

$$g(c^2 r^2) = K\mathrm{e}^{-r^2/2}, \qquad r > 0,$$

for a constant $K > 0$. Continuity at zero then gives $g(s) = g(0)\mathrm{e}^{-s/(2c^2)}$ for every $s \ge 0$, contrary to the hypothesis that g is not proportional to an exponential.

Lemma 16 therefore yields a nonempty open interval of radii r on which $\mu(C_r) = \mu(H_{a(r)}) = G_{d_y}(r)$, $H_{a(r)} = \mathrm{HDR}^f_{G_{d_y}(r)}$, and $\mathrm{Vol}(C_r) > \mathrm{Vol}(H_{a(r)})$, where $h(y) = \|\nabla\varphi^*(y)\|_2$, $C_r = \{y : h(y) \le r\}$, $\rho_\Sigma(y) = \|A^{-1}(y-m)\|_2$ and $H_t = \{y : \rho_\Sigma(y) \le t\}$ are as in the lemma. Since $\mathcal{C}^Q_{1-\alpha} = C_{r_{1-\alpha}}$ with $r_{1-\alpha} = G_{d_y}^{-1}(1-\alpha)$, and $r \mapsto 1 - G_{d_y}(r)$ is a C^1-diffeomorphism of $(0,\infty)$ onto $(0,1)$, the image of this interval is a nonempty open interval of levels α with $\mathrm{Vol}(\mathcal{C}^Q_{1-\alpha}) > \mathrm{Vol}(\mathrm{HDR}^f_{1-\alpha})$. □

Theorem 18 (Gaussian reference: a dichotomy for elliptical targets)**.** *Let $d_y \ge 2$, let $\mu_U = \mathcal{N}(0, \mathrm{I}_{d_y})$, and let μ have density* (1)*, where Σ is symmetric positive definite and $g : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and strictly decreasing. Let $Q = \nabla\varphi$ be the Brenier map transporting μ_U to μ and $Q^{-1} = \nabla\varphi^*$ its a.e. inverse.*

(i) Assume $\int \|y\|_2^2 f(y)\,\mathrm{d}y < \infty$, that $\Sigma \notin \{\sigma\mathrm{I} : \sigma > 0\}$, and that g is not proportional to $s \mapsto \mathrm{e}^{-\beta s}$ for any $\beta > 0$. Then there is a nonempty open interval $I \subseteq (0,1)$ such that, for every $\alpha \in I$,

$$\mathrm{Vol}(\mathcal{C}^Q_{1-\alpha}) > \mathrm{Vol}(\mathrm{HDR}^f_{1-\alpha}).$$

(ii) Conversely, if μ is Gaussian, or if $\Sigma \in \{\sigma\mathrm{I} : \sigma > 0\}$, then $\mathcal{C}^Q_{1-\alpha} = \mathrm{HDR}^f_{1-\alpha}$ up to Lebesgue-null sets for every $\alpha \in (0,1)$. No moment assumption is required.

Part (i) restates Theorem 2 and was proved above. Part (ii) shows that the excluded targets are genuine exceptions; we now prove it.

Proof of (ii). Suppose first that $\mu = \mathcal{N}(m', \Sigma')$ is Gaussian. The strict decrease of g makes m the unique mode of (1), so $m' = m$. Fix a unit vector e. Along $y = m + te$, comparison of the two density formulas gives

$$\log g\big(t^2 e^\top \Sigma^{-1} e\big) = \text{constant} - \frac{t^2}{2} e^\top (\Sigma')^{-1} e.$$

Thus $g(s) = g(0)\mathrm{e}^{-\beta s}$ for some $\beta > 0$, and consistency over all directions gives $\Sigma' = \Sigma/(2\beta)$. Let $L := (\Sigma')^{1/2}$. The map $u \mapsto m + Lu$ is the gradient of a convex function and pushes μ_U to μ. Uniqueness of the Brenier map (McCann, 1995) yields $Q(u) = m + Lu$ μ_U-almost everywhere. Hence $\mathcal{C}^Q_{1-\alpha}$ is the Mahalanobis ellipsoid

$$\{y : (y-m)^\top (\Sigma')^{-1}(y-m) \le r^2_{1-\alpha}\}.$$

It has μ-mass $1-\alpha$ and is the corresponding HDR by Lemma 14.

Suppose next that $\Sigma = \sigma\mathrm{I}$ for some $\sigma > 0$. For $Y \sim \mu$, let $Z := \sigma^{-1/2}(Y - m)$ and define

$$F_\nu(t) := \mathbb{P}(\|Z\|_2 \le t), \qquad a(r) := F_\nu^{-1}(G_{d_y}(r)).$$

The radial density is continuous and strictly positive on $(0,\infty)$, so a is continuous and strictly increasing. Define $T(0) := m$ and, for $u \ne 0$,

$$T(u) := m + \sigma^{1/2} a(\|u\|_2)\frac{u}{\|u\|_2}.$$

Writing $U = \rho\Theta$, the direction Θ is uniform on the unit sphere and independent of ρ, while $\mathbb{P}(a(\rho) \le t) = F_\nu(t)$. Thus $T_\# \mu_U = \mu$. Moreover, $T = \nabla\Phi$ for

$$\Phi(u) := m^\top u + \sigma^{1/2} \int_0^{\|u\|_2} a(s)\,\mathrm{d}s.$$

The radial part of Φ is nondecreasing and convex, so Φ is convex. Uniqueness gives $Q = T$ μ_U-almost everywhere. Therefore

$$\mathcal{C}^Q_{1-\alpha} = \{y : \sigma^{-1/2}\|y - m\|_2 \le a(r_{1-\alpha})\}$$

up to a null set. This set has mass $1 - \alpha$ and is the HDR by Lemma 14. □

Remark 19 (Sharpness and moment condition)**.** *Within the elliptical family* (1) *with finite second moment, the two parts are exhaustive: with a Gaussian reference, the Brenier construction is volume-optimal at every level precisely for Gaussian or spherical targets. (A target with $\Sigma \notin \{\sigma \mathrm{I}\}$ and exponential g is Gaussian, hence covered by part (ii).) The second-moment assumption in part (i) makes the correlation functional finite for every coupling; part (ii) uses only uniqueness of the Brenier map. Whether part (i) remains true under a first-moment assumption is open; non-Gaussian, non-spherical targets with infinite second moment are not covered.*

F MEASURE-PRESERVING FLOWS

Throughout this appendix, f_U is a strictly positive, smooth probability density on $\mathbb{R}^{d_y}$, $v : \mathbb{R}^{d_y} \times \mathbb{R}_+ \to \mathbb{R}^{d_y}$ satisfies **A**3(i) stated in Section 4.1 — the higher regularity **A**3(ii) is invoked only where the necessity direction requires it, in Step 4 — and $m(u,t) := f_U(u)\, v(u,t)$ denotes the associated momentum field. We write $\partial_i := \partial/\partial u_i$, use the row-wise divergence $(\operatorname{div} A)_i = \sum_j \partial_j A_{ij}$ of Section 4.1, the convention $b_0 = b_{d_y} \equiv 0$ for the band potentials introduced in Step 4, and abbreviate $d := d_y$.

Proof of Proposition 5. The proof proceeds in four steps: well-posedness of the flow; reduction of measure preservation to the stationary continuity equation; sufficiency of the skew-potential representation; and its necessity, established constructively together with the sparsity claim.

Step 1: the flow is a well-defined family of C^1-diffeomorphisms. Under **A** 3(i), v is continuous on $\mathbb{R}^d \times \mathbb{R}_+$ and, since $(u,t) \mapsto D_u v(u,t)$ is continuous, $v(\cdot,t)$ is Lipschitz on compact subsets of $\mathbb{R}^d$, locally uniformly in t. By the Picard–Lindelöf theorem, the initial value problem $\frac{d}{dt}\Psi(u,t) = v(\Psi(u,t),t)$, $\Psi(u,0) = u$, has a unique maximal solution for every u; the linear-growth bound in **A** 3(i) and Grönwall's inequality give $\|\Psi(u,t)\|_2 \le \big(\|u\|_2 + tc(\bar t)\big)e^{tc(\bar t)}$ for $t \le \bar t$, so no trajectory explodes in finite time and the flow is global. Continuous differentiability of the flow map in u follows from the continuity of $D_u v$: the matrix $J(u,t) := D_u\Psi(u,t)$ solves the variational equation $\dot J = D_u v(\Psi(u,t),t)\, J$, $J(u,0) = \mathrm{I}$, and by Liouville's formula $\det J(u,t) = \exp\big(\int_0^t \operatorname{Tr} D_u v(\Psi(u,s),s)\,ds\big) > 0$. Finally, for fixed t the map $\Psi(\cdot,t)$ is a bijection of $\mathbb{R}^d$: its inverse is the time-t flow of the reversed field $\tilde v(u,s) := -v(u,t-s)$, which also satisfies **A**3(i) on $[0,t]$. Hence each $\Psi(\cdot,t)$ is a C^1-diffeomorphism of $\mathbb{R}^d$; see, e.g., Hartman (2002, Ch. II and V).

Step 2: preservation is equivalent to the weighted divergence condition. The Jacobian determinant in Step 1 is positive. Differentiating the change-of-variables density along a trajectory gives

$$\begin{aligned}\frac{d}{dt}\big[f_U\big(\Psi(u,t)\big)\det D_u\Psi(u,t)\big] &= \det D_u\Psi(u,t)\Big[\nabla f_U\big(\Psi(u,t)\big)^\top v\big(\Psi(u,t),t\big)\\ &\qquad + f_U\big(\Psi(u,t)\big)\operatorname{Tr} D_u v\big(\Psi(u,t),t\big)\Big]\\ &= \det D_u\Psi(u,t)\,\nabla\cdot\big(f_U v(\cdot,t)\big)\big(\Psi(u,t)\big).\end{aligned}$$

If the weighted divergence vanishes for almost every t, integration in time yields

$$f_U\big(\Psi(u,t)\big)\det D_u\Psi(u,t) = f_U(u).$$

The change-of-variables formula then gives $\Psi(\cdot,t)_{\#}\mu_U = \mu_U$. Conversely, if the flow preserves μ_U, the same change-of-variables formula gives this identity for almost every u. Both sides are continuous in u, so it holds for every u. Differentiation in t, followed by the fact that $\Psi(\cdot,t)$ is onto and has positive Jacobian determinant, gives $\nabla\cdot\big(f_U v(\cdot,t)\big) = 0$. This proves the required equivalence without an additional moment assumption on μ_U.

It therefore remains to prove: *for a fixed time t (suppressed from the notation), a C^1 field v — with $v \in C^{d-1}$ for the necessity direction, as provided by **A** 3(ii) — satisfies $\nabla\cdot(f_U v) = 0$ if and only if $v = \operatorname{div} A + A\nabla \log f_U$ for some skew-symmetric A, which can be taken band-sparse.* By the product rule, componentwise,

$$\operatorname{div}(f_U A) = f_U\big(\operatorname{div} A + A\nabla \log f_U\big), \tag{17}$$

so, writing $\tilde{A} := f_U A$ (skew-symmetric if and only if A is, since $f_U > 0$), the representation is equivalent to $m = \operatorname{div}\tilde{A}$ for the momentum field $m = f_U v$.

Step 3: sufficiency. Let $\tilde{A}$ be skew-symmetric with entries in C^1, and let $m = \operatorname{div}\tilde{A}$. For every $\phi \in C_c^\infty(\mathbb{R}^d)$, integrating by parts,

$$\int m^\top \nabla\phi \, du = \sum_{i,j}\int \partial_j \tilde{A}_{ij}\,\partial_i\phi \, du = -\sum_{i,j}\int \tilde{A}_{ij}\,\partial_j\partial_i\phi \, du = 0,$$

because the skew-symmetric matrix $\tilde{A}(u)$ is paired against the symmetric Hessian $\nabla^2\phi(u)$ at every point. Hence $\nabla\cdot m = 0$ in the sense of distributions, and pointwise when $m \in C^1$. Note that this direction requires no more than C^1 regularity of A and no sparsity.

Step 4: necessity and sparsity. Let $m \in C^{d-1}(\mathbb{R}^d;\mathbb{R}^d)$ — this is where the additional regularity $v(\cdot,t) \in C^{d-1}$, hence $m(\cdot,t) = f_U v(\cdot,t) \in C^{d-1}$, is used — and assume $\sum_{i=1}^d \partial_i m_i = 0$. We construct scalar fields $b_1,\dots,b_{d-1}$ such that the skew-symmetric band matrix

$$\tilde{A}_{k,k+1} := b_k, \qquad \tilde{A}_{k+1,k} := -b_k, \qquad k = 1,\dots,d-1,$$

with all other entries zero, satisfies $\operatorname{div}\tilde{A} = m$, i.e., componentwise, with $b_0 = b_d = 0$,

$$\begin{aligned} \partial_2 b_1 &= m_1, \\ \partial_{i+1}b_i - \partial_{i-1}b_{i-1} &= m_i, \qquad i = 2,\dots,d-1 \\ -\partial_{d-1}b_{d-1} &= m_d. \end{aligned} \tag{18}$$

This is a system of d equations in $d-1$ unknowns; the divergence constraint provides exactly the compatibility that makes it solvable.

Recursive construction. Define, for $i = 1,\dots,d-1$,

$$b_i(u) := \int_0^{u_{i+1}} \big(m_i + \partial_{i-1}b_{i-1}\big)(u_1,\dots,u_i,s,u_{i+2},\dots,u_d)\,ds, \tag{19}$$

so that $\partial_{i+1}b_i = m_i + \partial_{i-1}b_{i-1}$ by construction: the first $d-1$ equations of (18) hold. Since one integration is performed and one derivative is taken at each stage, $b_i \in C^{d-i}$; in particular every b_i with $i \le d-2$ is at least C^2, which legitimates the interchanges of derivatives below, and $b_{d-1} \in C^1$.

A telescoping identity. We claim that, for $1 \le j \le d-1$,

$$\partial_j\,\partial_{j+1}b_j = \sum_{i=1}^{j}\partial_i m_i \qquad \text{on } \mathbb{R}^d. \tag{20}$$

For $j = 1$, differentiate $\partial_2 b_1 = m_1$ with respect to u_1. Assume (20) for $j-1$. Differentiating the j-th recursion $\partial_{j+1}b_j = m_j + \partial_{j-1}b_{j-1}$ with respect to u_j and using Schwarz's theorem on $b_{j-1} \in C^2$,

$$\partial_j\partial_{j+1}b_j = \partial_j m_j + \partial_j\partial_{j-1}b_{j-1} = \partial_j m_j + \partial_{j-1}\big(\partial_j b_{j-1}\big),$$

and $\partial_j b_{j-1} = \partial_{(j-1)+1}b_{j-1}$, so the last term is $\partial_{j-1}\partial_j b_{j-1} = \sum_{i\le j-1}\partial_i m_i$ by the induction hypothesis. This proves (20). In particular, for $j = d-1$, the divergence constraint gives

$$\partial_{d-1}\,\partial_d b_{d-1} = \sum_{i=1}^{d-1}\partial_i m_i = -\,\partial_d m_d. \tag{21}$$

Boundary corrector. The last equation of (18), $-\partial_{d-1}b_{d-1} = m_d$, does not yet hold: by (19), b_{d-1} vanishes on the hyperplane $\{u_d = 0\}$, and so does its tangential derivative $\partial_{d-1}b_{d-1}$, whereas m_d need not. We therefore replace b_{d-1} by

$$\bar{b}_{d-1}(u) := b_{d-1}(u) + c(u_1, \dots, u_{d-1}),$$
$$c(u_1, \dots, u_{d-1}) := -\int_0^{u_{d-1}} m_d(u_1, \dots, u_{d-2}, s, 0)\, ds.$$

Since c does not depend on u_d, the $(d-1)$-th equation $\partial_d \bar{b}_{d-1} = m_{d-1} + \partial_{d-2}b_{d-2}$ is unaffected. For the d-th equation, set $F(u) := \partial_{d-1}\bar{b}_{d-1}(u) + m_d(u)$. On $\{u_d = 0\}$ we have $\partial_{d-1}b_{d-1} = 0$ and $\partial_{d-1}c = -m_d(\cdot, 0)$, so $F(\cdot, 0) \equiv 0$. Since b_{d-1} is only C^1, Schwarz's theorem does not apply directly; instead, write $g := m_{d-1} + \partial_{d-2}b_{d-2} \in C^1$, so that $\partial_d b_{d-1} = g$ by (19) and, differentiating under the integral sign in (19) ($\partial_{d-1}g$ being continuous), $\partial_{d-1}b_{d-1}(u) = \int_0^{u_d} \partial_{d-1}g\, ds$. Hence both mixed partials $\partial_d\partial_{d-1}b_{d-1}$ and $\partial_{d-1}\partial_d b_{d-1}$ exist, are continuous, and equal $\partial_{d-1}g$. Combining with (21), and since the corrector term $\partial_d\partial_{d-1}c$ vanishes (c does not depend on u_d),

$$\partial_d F = \partial_d\partial_{d-1}b_{d-1} + \partial_d m_d = \partial_{d-1}\partial_d b_{d-1} + \partial_d m_d = -\partial_d m_d + \partial_d m_d = 0.$$

Hence $F \equiv 0$ on $\mathbb{R}^d$: the d-th equation holds, and $\operatorname{div} \tilde{A} = m$ everywhere, where $\tilde{A}$ is the band matrix built from $b_1, \dots, b_{d-2}, \bar{b}_{d-1}$.

Conclusion of Step 4. The matrix $A := \tilde{A}/f_U$ is skew-symmetric, and its only possible nonzero entries are $\pm b_k/f_U$ (with b_{d-1} replaced by $\bar{b}_{d-1}$) on the first off-diagonals. It therefore has at most $2(d-1)$ nonzero entries. By (17), $\operatorname{div} A + A\nabla \log f_U = \operatorname{div} \tilde{A}/f_U = m/f_U = v$. The recursion uses only spatial differentiation and integration over finite intervals. Under **A**3(i), v is jointly continuous and its spatial derivatives are pointwise limits of difference quotients, hence jointly measurable in (u, t); consequently each $b_k(u, t)$, the boundary corrector, and the resulting $A(u, t)$ are jointly measurable in (u, t). Since Step 2 gives $\nabla\cdot(f_U v(\cdot, t)) = 0$ at every t and **A**3(ii) provides $v(\cdot, t) \in C^{d-1}$ at every t, this proves the necessity direction, together with the sparsity claim, for every t. *(Only the last off-diagonal pair requires the corrector; when $m_d(\cdot, 0) \equiv 0$ — e.g. for fields vanishing on the hyperplane — the raw recursion already suffices.)* □

Remarks. *(i) Regularity.* Each stage of (19) trades one derivative of m for one potential, so the construction consumes $d-2$ derivatives before the final C^2/C^1 requirements; this is the origin of the hypothesis $v(\cdot, t) \in C^{d-1}$ (**A** 3(ii)) in the necessity part of Proposition 5. The hypothesis is not optimal: for $v(\cdot, t)$ merely C^1, the same statement holds with $A(\cdot, t)$ understood in the sense of distributions (the pairing argument of Step 3 is already distributional), but we do not pursue this, as the fields used in practice are smooth. *(ii) Gauge freedom.* The potential is far from unique: any skew field B with $\operatorname{div}(f_U B) = 0$ may be added to A without changing v; the band construction is one convenient gauge, whose $d-1$ scalar degrees of freedom match the pointwise codimension of the constraint $\nabla\cdot(f_U v) = 0$, cf. the discussion in Section 4.1. In the language of differential forms, Step 4 is a quantitative Poincaré lemma: the $(d-1)$-form dual to the divergence-free momentum m is closed, hence exact on $\mathbb{R}^d$, and the band potential is a choice of primitive. *(iii) The Gaussian case.* For $\mu_U = \mathcal{N}(0, \mathrm{I}_d)$, $\nabla \log f_U(u) = -u$ and (17) reduces to the componentwise formula of Corollary 6: with the band entries $a_k := b_k/f_U$, $v_i = \partial_{i+1}a_i - \partial_{i-1}a_{i-1} - a_i u_{i+1} + a_{i-1}u_{i-1}$, matching (3).

G VOLUME-OBJECTIVE

Proof of Proposition 7. Everywhere differentiable maps satisfy Luzin's property (N), and $T \circ S$ is injective outside a μ_U-null set, which is Lebesgue-null on $\mathcal{B}_{d_y}(0, r)$ because $f_U > 0$ a.e. there. The area formula therefore gives

$$\operatorname{Vol}\big((T \circ S)(\mathcal{B}_{d_y}(0, r))\big) = \int_{\mathcal{B}_{d_y}(0,r)} |\det DT(S(u))|\, |\det DS(u)|\, du.$$

For an everywhere differentiable map Φ that is injective outside a μ_U-null set and pushes μ_U to a law with density g, the area formula applied on a Borel set of full μ_U-measure where Φ is injective, combined with the Luzin–Suslin theorem to identify the densities, gives $f_U(w) = g(\Phi(w))\, |\det D\Phi(w)|$ for μ_U-a.e. w. Applied to S with $g = f_U$, this yields

$f_U(S(u))\,|\det DS(u)| = f_U(u)$ μ_U-a.e. (equation (4) is the special case $S = S_\theta$), so the right-hand side equals $\int_{\mathcal{B}_{d_y}(0,r)} |\det DT(S(u))|\, f_U(S(u))^{-1}\, d\mu_U(u)$. Applied to T with $g = f$, it yields $f_U(w) = f(T(w))\,|\det DT(w)|$ for μ_U-a.e. w, hence at $w = S(u)$ for μ_U-a.e. u because $S_\#\mu_U = \mu_U$; moreover $f_U(S(u)) > 0$ and $f(T(S(u))) > 0$ hold μ_U-a.e. Substituting yields equation (5). □

Proof of Proposition 8. Let $B := \mathcal{B}_{d_y}(0, r_{1-\alpha})$ and $H := \mathrm{HDR}^f_{1-\alpha}$; the radius $r_{1-\alpha}$ exists because, under **A** 1, $r \mapsto \mu_U(\mathcal{B}_{d_y}(0,r))$ is continuous and increases from 0 to 1. Since $f > 0$ Lebesgue-a.e., μ and Lebesgue measure have the same null sets. For every measure-preserving a.e.-bijection S, the map $T \circ S$ has a.e. inverse $S^{-1} \circ T^{-1}$, and

$$\mu\big(\mathcal{C}^{T\circ S}_{1-\alpha}\big) = \mu_U\big(\{u : S^{-1}(u) \in B\}\big) = \mu_U\big(S^{-1}(\{u : S^{-1}(u) \in B\})\big) = \mu_U(B),$$

which equals $1-\alpha$; the three equalities use, respectively, $T^{-1} \circ T = \mathrm{id}$ μ_U-a.e. together with $T_\#\mu_U = \mu$, then $S_\#\mu_U = \mu_U$, then $S^{-1} \circ S = \mathrm{id}$ μ_U-a.e. It follows from Lemma 12 that

$$\mathrm{Vol}\big(\mathcal{C}^{T\circ S}_{1-\alpha}\big) \geq \mathrm{Vol}(H),$$

with equality if and only if $\mathcal{C}^{T\circ S}_{1-\alpha} = H$ up to Lebesgue-null sets. Pulling back by T — which satisfies $\mu(E) = \mu_U(T^{-1}(E))$ for every Borel E and $T^{-1}(T(u)) = u$ for μ_U-a.e. u — equality is equivalent to $\mu_U\big(\{u : S^{-1}(u) \in B\} \triangle T^{-1}(H)\big) = 0$, and, pulling back further by S (using $S_\#\mu_U = \mu_U$ and $S^{-1}(S(u)) = u$ μ_U-a.e.), to $\mu_U\big(B \triangle S^{-1}(T^{-1}(H))\big) = 0$. This is the characterization stated in the proposition.

It remains to construct a rearrangement with this property. The sets B and $T^{-1}(H)$ have the same μ_U-mass $1-\alpha \in (0,1)$, and their complements have mass $\alpha \in (0,1)$. Replacing $T^{-1}(H)$ by a Borel set equal to it up to a μ_U-null set if necessary, and normalizing the restrictions of μ_U to B and $T^{-1}(H)$ by $1-\alpha$ and to their complements by α, we obtain atomless standard probability spaces. The isomorphism theorem (Bogachev, 2007, Thm. 9.2.2) then gives measure-preserving Borel bijections modulo null sets

$$S_1 : B \to T^{-1}(H) \quad \text{and} \quad S_2 : B^{\mathrm{c}} \to \big(T^{-1}(H)\big)^{\mathrm{c}}.$$

Define $S^\star(u) = S_1(u)$ for $u \in B$ and $S^\star(u) = S_2(u)$ for $u \in B^{\mathrm{c}}$. Then $S^\star$ is a measure-preserving a.e.-bijection with $\mu_U\big(B \triangle (S^\star)^{-1}(T^{-1}(H))\big) = 0$, so $\mathcal{C}^{T\circ S^\star}_{1-\alpha} = H$ up to Lebesgue-null sets. This proves attainment and the characterization of all minimizers. □

Proof of Lemma 9. Set

$$H_p := \{y \in \mathcal{Y} : f(y) \geq c_p\}.$$

By assumption,

$$\mu_Y(H_p) = p.$$

Let $A \subseteq \mathcal{Y}$ be any measurable set satisfying

$$\mu_Y(A) = p.$$

Since $\mu_Y(A) = \mu_Y(H_p)$, we have

$$\mu_Y(A \setminus H_p) = \mu_Y(H_p \setminus A).$$

Moreover, since

$$\mu_Y(\{f = c_p\}) = 0,$$

we have, up to μ_Y-null sets,

$$f < c_p \text{ on } A \setminus H_p \quad \text{and} \quad f > c_p \text{ on } H_p \setminus A.$$

Because ϕ is strictly decreasing,

$$\phi(f) > \phi(c_p) \text{ on } A \setminus H_p \quad \text{and} \quad \phi(f) < \phi(c_p) \text{ on } H_p \setminus A.$$

Therefore,

$$\begin{aligned}\int_A \phi(f)\,\mathrm{d}\mu_Y - \int_{H_p} \phi(f)\,\mathrm{d}\mu_Y &= \int_{A\setminus H_p} \phi(f)\,\mathrm{d}\mu_Y - \int_{H_p\setminus A} \phi(f)\,\mathrm{d}\mu_Y \\ &\geq \phi(c_p)\,(\mu_Y(A \setminus H_p) - \mu_Y(H_p \setminus A)) = 0.\end{aligned}$$

Hence H_p minimizes

$$A \longmapsto \int_A \phi(f(y))\,\mathrm{d}\mu_Y(y)$$

over all measurable sets satisfying $\mu_Y(A) = p$. Similarly,

$$\begin{aligned}\int_{H_p} f\,\mathrm{d}\mu_Y - \int_A f\,\mathrm{d}\mu_Y &= \int_{H_p\setminus A} f\,\mathrm{d}\mu_Y - \int_{A\setminus H_p} f\,\mathrm{d}\mu_Y \\ &\geq c_p\mu_Y(H_p\setminus A) - c_p\mu_Y(A\setminus H_p) = 0.\end{aligned}$$

Thus H_p also maximizes

$$A \longmapsto \int_A f(y)\,\mathrm{d}\mu_Y(y)$$

over all measurable sets satisfying $\mu_Y(A) = p$. Finally, if

$$\mu_Y(A\triangle H_p) > 0,$$

then

$$\mu_Y(A\setminus H_p) = \mu_Y(H_p\setminus A) > 0.$$

Since $\mu_Y(\{f = c_p\}) = 0$, the inequalities

$$f < c_p \quad \text{on } A\setminus H_p, \qquad f > c_p \quad \text{on } H_p\setminus A$$

are strict almost everywhere, the two inequalities above are also strict. Hence A cannot be an optimizer.

Therefore H_p is the unique optimizer up to μ_Y-null sets for both problems, which proves

$$\operatorname*{argmin}_{A:\mu_Y(A)=p}\int_A \phi(f(y))\,\mathrm{d}\mu_Y(y) = \operatorname*{argmax}_{A:\mu_Y(A)=p}\int_A f(y)\,\mathrm{d}\mu_Y(y).$$

□

H AN ORACLE INEQUALITY FOR THE SURROGATE LOSS

This appendix gives a finite-sample guarantee for the surrogate loss of Section 4.2 when the conditional transport T is fixed and known. All probabilities are conditional on T; the only statistical randomness comes from the covariates, the sampled coverage levels, and the reference points. Because the surrogate is an average of logarithms — rather than the logarithm of an inner Monte Carlo average — its empirical version is an *unbiased* estimator of the population risk, and the analysis reduces to a single uniform concentration step: no inner-sample bias term appears, and the constants enter linearly rather than exponentially through the envelope.

H.1 SETTING AND OBJECTIVES

Let $\Theta \subseteq \mathbb{R}^{d_\theta}$ be a parameter set of Euclidean diameter at most $D_\Theta < \infty$, and for each $\theta \in \Theta$ let $S_\theta(\cdot, x, \alpha)$ be the level-conditioned rearrangement of Section 4.1, satisfying $S_\theta(\cdot, x, \alpha)_{\#}\mu_U = \mu_U$ for μ_X-almost every x and every level α. Fix a level floor $\alpha_- \in (0, 1)$ and let the level be drawn uniformly on $[\alpha_-, 1)$, so that the sampled radius satisfies $r_{1-\alpha} \leq \bar{r} := r_{1-\alpha_-} < \infty$. (The practical sampler of Section 4.2 uses $\alpha \sim \mathrm{Unif}(0, 1)$; the floor is what makes the uniform envelope below finite, since $r_{1-\alpha} \to \infty$ as $\alpha \downarrow 0$. See Remark 23 for the unbounded case.) Write

$$\lambda_\alpha := \mu_U\big(\cdot \mid \|U\|_2 \leq r_{1-\alpha}\big), \qquad g_\theta(u, x, \alpha) := \log\big|\det D_z T\big(S_\theta(u, x, \alpha), x\big)\big| + \tfrac{1}{2}\big\|S_\theta(u, x, \alpha)\big\|_2^2.$$

The population and empirical risks are

$$\overline{L}(\theta) := \mathbb{E}\Big[\int g_\theta(u, X, A)\,\lambda_A(\mathrm{d}u)\Big], \qquad \widehat{L}_{n,m}(\theta) := \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} g_\theta\big(U_{ij}, X_i, A_i\big), \tag{22}$$

where $(X_i, A_i)_{i\leq n}$ are i.i.d. with $X_i \sim \mu_X$ and $A_i \sim \mathrm{Unif}([\alpha_-, 1))$ independent, and, conditionally on (X_i, A_i), the inner points $U_{i1}, \ldots, U_{im}$ are i.i.d. with law λ_{A_i}, independently across i. By

construction $\mathbb{E}[\widehat{L}_{n,m}(\theta)] = \overline{L}(\theta)$ for every θ: the estimator is unbiased, in contrast with the log-of-average objective, whose empirical version carries a Jensen bias of order $1/m$. For each fixed (x, α), the inner integral in (22) equals the per-level surrogate $L(S_\theta)$ of Section 4.2 up to an additive constant that does not depend on θ; minimizers are therefore unchanged.

Given $\varepsilon_{\mathrm{opt}} \geq 0$, an $\varepsilon_{\mathrm{opt}}$-empirical minimizer is any measurable $\widehat{\theta} \in \Theta$ with

$$\widehat{L}_{n,m}(\widehat{\theta}) \leq \inf_{\theta \in \Theta} \widehat{L}_{n,m}(\theta) + \varepsilon_{\mathrm{opt}}. \tag{23}$$

H.2 Assumption and main result

A4 (Surrogate envelope)**.** *The map $(\theta, u, x, \alpha) \mapsto g_\theta(u, x, \alpha)$ is jointly measurable, and there exist constants $H < \infty$ and $J < \infty$ and a measurable set of full measure for the sampling law above on which, simultaneously for all $\theta, \theta' \in \Theta$,*

$$|g_\theta(u, x, \alpha)| \leq H, \qquad |g_\theta(u, x, \alpha) - g_{\theta'}(u, x, \alpha)| \leq J \, \|\theta - \theta'\|_2.$$

Primitive sufficient conditions mirror those of the flow parametrization. Suppose the band potentials and their first u-derivatives are bounded uniformly in u, locally uniformly in (t, x, α, θ), as in Section 4.1; then Grönwall's inequality gives, for u in the sampled ball $\mathcal{B}_{d_y}(0, \bar{r})$, a finite reachable radius $\|S_\theta(u, x, \alpha)\|_2 \leq R_S$, and a parameter-stability bound $\|S_\theta(u, x, \alpha) - S_{\theta'}(u, x, \alpha)\|_2 \leq C_S \|\theta - \theta'\|_2$ with $C_S = K_\theta(\mathrm{e}^{K_z} - 1)/K_z$, where K_z and K_θ bound $\|D_z v_\theta\|$ and $\|D_\theta v_\theta\|$ on the relevant domain. If, on $\mathcal{B}_{d_y}(0, R_S)$, $\big|\log|\det D_z T(z, x)|\big| \leq H_T$ and $\|\nabla_z \log|\det D_z T(z, x)|\|_2 \leq J_T$, then **A**4 holds with

$$H = H_T + \tfrac{1}{2} R_S^2, \qquad J = (J_T + R_S)\, C_S. \tag{24}$$

Theorem 20 (Oracle inequality for the surrogate loss)**.** *Let **A** 4 hold, let the data be sampled as above, and let $\widehat{\theta}$ satisfy* (23)*. Then, for every $\eta \in (0, 1)$ and $\delta > 0$, with probability at least $1 - \eta$,*

$$\sup_{\theta \in \Theta} \big|\widehat{L}_{n,m}(\theta) - \overline{L}(\theta)\big| \leq \Delta_n(\delta, \eta) := 2J\delta + H\sqrt{\frac{2}{n}\Big[\log\frac{2}{\eta} + d_\theta \log\Big(1 + \frac{2D_\Theta}{\delta}\Big)\Big]}, \tag{25}$$

and, on the same event,

$$\overline{L}(\widehat{\theta}) - \inf_{\theta \in \Theta} \overline{L}(\theta) \leq 2\,\Delta_n(\delta, \eta) + \varepsilon_{\mathrm{opt}}. \tag{26}$$

In particular, for $J > 0$ the choice $\delta = \min\{D_\Theta,\, H/(J\sqrt{n})\}$ makes the discretization term of order $n^{-1/2}$, and the bound reads, up to logarithmic factors, $\overline{L}(\widehat{\theta}) - \inf_\Theta \overline{L} \lesssim H\sqrt{d_\theta/n} + \varepsilon_{\mathrm{opt}}$.

Proof. Define the block averages $\bar{g}_\theta(i) := m^{-1}\sum_{j=1}^m g_\theta(U_{ij}, X_i, A_i)$, so that $\widehat{L}_{n,m}(\theta) = n^{-1}\sum_{i=1}^n \bar{g}_\theta(i)$. The triples $(X_i, A_i, U_{i,1:m})$ are i.i.d. across i, hence so are the blocks $\bar{g}_\theta(i)$, which take values in $[-H, H]$ by **A**4 and have mean $\overline{L}(\theta)$. Fix a δ-net $\Theta_\delta \subseteq \Theta$ with $|\Theta_\delta| \leq (1{+}2D_\Theta/\delta)^{d_\theta}$, which exists because Θ has diameter at most D_Θ. For each $\theta_0 \in \Theta_\delta$, Hoeffding's inequality gives $\mathbb{P}\big(|\widehat{L}_{n,m}(\theta_0) - \overline{L}(\theta_0)| > t\big) \leq 2\exp\{-nt^2/(2H^2)\}$; a union bound over the net yields, with probability at least $1 - \eta$,

$$\max_{\theta_0 \in \Theta_\delta} \big|\widehat{L}_{n,m}(\theta_0) - \overline{L}(\theta_0)\big| \leq H\sqrt{\frac{2}{n}\log\frac{2|\Theta_\delta|}{\eta}}.$$

For arbitrary $\theta \in \Theta$, pick $\theta_0 \in \Theta_\delta$ with $\|\theta - \theta_0\|_2 \leq \delta$; the J-Lipschitz property of g_θ in θ transfers to both $\widehat{L}_{n,m}$ and $\overline{L}$, adding at most $2J\delta$. This proves (25). For (26), let $\varepsilon > 0$ and $\theta_\varepsilon \in \Theta$ with $\overline{L}(\theta_\varepsilon) \leq \inf_\Theta \overline{L} + \varepsilon$; on the event of (25),

$$\overline{L}(\widehat{\theta}) - \overline{L}(\theta_\varepsilon) = \big[\overline{L} - \widehat{L}_{n,m}\big](\widehat{\theta}) + \big[\widehat{L}_{n,m}(\widehat{\theta}) - \widehat{L}_{n,m}(\theta_\varepsilon)\big] + \big[\widehat{L}_{n,m} - \overline{L}\big](\theta_\varepsilon) \leq 2\Delta_n(\delta, \eta) + \varepsilon_{\mathrm{opt}},$$

using (23) for the middle term. Let $\varepsilon \downarrow 0$. □

H.3 REMARKS

Remark 21 (Comparison with the log-of-average objective)**.** *An objective of the form* $\log\big(m^{-1}\sum_j \exp h_j\big)$ *— the direct Monte Carlo version of the log-*volume *— is biased at finite* m *by Jensen's inequality, with a bias bound of order* $(\mathrm{e}^{2H'}-1)^2/(8m)$ *for an envelope* H' *on the log-weights, and its uniform analysis requires controlling exponentiated envelopes. The surrogate loss removes both effects: the estimator is unbiased for every* (n, m)*, and* H *enters* (25) *linearly. This is the statistical counterpart of the numerical weight-degeneracy argument for adopting the log-domain objective.*

Remark 22 (Role of the inner sample size)**.** *Hoeffding's inequality uses only the boundedness of the block averages, so* m *does not appear in* (25)*. A Bernstein refinement replaces* H^2/n *by* σ^2/n *up to lower-order terms, with block variance*

$$\sigma^2 = \mathrm{Var}\Big(\int g_\theta \,\mathrm{d}\lambda_A(\cdot, X, A)\Big) + \frac{1}{m}\,\mathbb{E}\big[\mathrm{Var}\big(g_\theta(U, X, A) \mid X, A\big)\big] :$$

increasing m *shrinks only the inner component, and the rate remains* $n^{-1/2}$ *in the number of covariate–level draws. The inner sample size is thus a compute–variance trade-off, not a rate parameter.*

Remark 23 (Level floor and unbounded levels)**.** *The floor* α_- *makes the sampled ball bounded and hence the envelope* H *finite; without it,* $\sup_u \|S_\theta(u, x, \alpha)\|_2^2$ *is infinite and* ***A*** *4 fails. Since* $\frac{1}{2}\|S_\theta(U)\|_2^2$ *inherits sub-exponential tails from the Gaussian reference under the stability bounds above, a Bernstein–Orlicz argument would extend Theorem 20 to* $\alpha \sim \mathrm{Unif}(0, 1)$ *with* H *replaced by sub-exponential norms; we do not pursue this, as training with a small level floor is indistinguishable in practice.*

Remark 24 (From surrogate risk to volume)**.** *The guarantee* (26) *controls the excess* surrogate *risk within the class* Θ*. By the discussion in Section 4.2, the surrogate is aligned with the volume at the level of population minimizers over the unrestricted rearrangement class (Lemma 9), but it is only a lower bound on the log-volume away from the optimum. Accordingly, the decomposition*

$$\overline{L}(\widehat{\theta}) - \inf_{S:\, S_\# \mu_U = \mu_U} \overline{L}(S) = \Big[\overline{L}(\widehat{\theta}) - \inf_{\theta\in\Theta} \overline{L}(\theta)\Big] + \mathfrak{A}_\Theta, \qquad \mathfrak{A}_\Theta := \inf_{\theta\in\Theta} \overline{L}(\theta) - \inf_{S:\, S_\# \mu_U = \mu_U} \overline{L}(S) \;\geq\; 0,$$

separates the statistical and optimization errors, controlled by Theorem 20, from the approximation error $\mathfrak{A}_\Theta$ *of the smooth band-flow class, which is not controlled by increasing* n *or* m *and contains, in particular, the connectedness obstruction of Section 4.2.*

Remark 25 (Scope)**.** *The theorem treats* T *as deterministic. If* T *is estimated on an independent split, the result applies conditionally on the fitted transport whenever the realized constants* (H, J) *in* (24) *are controlled; if* T *and* S_θ *are fitted on the same observations, the complexity of the joint class must be analyzed and the present bound does not apply. Conformal validity (Section 4) is unaffected in either case.*

I DETAILED EXPERIMENTAL SECTION

I.1 PARAMETRIZATION OF MEASURE-PRESERVING FLOWS

The rearrangement $S_\theta(\cdot, x)$ is defined as the time-one flow map of the Gaussian-preserving vector field introduced in Corollary 6:

$$v_\theta(u, t, x) = \mathrm{div}\, A_\theta(u, t, x) - A_\theta(u, t, x)\, u,$$

where A_θ is a skew-symmetric band matrix constructed from the potentials $a_1, \ldots, a_{d_y-1}$. Componentwise,

$$(v_\theta)_i = \partial_{i+1} a_i - \partial_{i-1} a_{i-1} - a_i u_{i+1} + a_{i-1} u_{i-1}, \qquad a_0 = a_{d_y} \equiv 0.$$

To parametrize the skew-symmetric matrix, we first parametrize a time and context dependent function

$$\widetilde{a}_\theta : \mathbb{R}^{d_y} \times \mathbb{R}^+ \times \mathcal{X} \to \mathbb{R}^{d_y - 1},$$

and then compute the partial derivatives required to construct the velocity field at each integration step. Rather than relying on PyTorch's automatic differentiation machinery to compute these derivatives, we evaluate the required first-order partial derivatives explicitly using the chain rule. This substantially reduces the computational cost of evaluating the velocity field.

The network $\widetilde{a}_\theta$ is a multilayer perceptron with two hidden layers of width 256. We use SiLU after the first hidden layer and $\tanh$ after the second, applying one last linear layer to produce the output of size $d-1$. The covariate x and time t are provided as additional inputs to the network. The flow is integrated using the adaptive-step *dopri5* solver with $\mathrm{rtol} = 10^{-5}$ and $\mathrm{atol} = 10^{-4}$. At inference time, the flow is integrated from $t = 0$ to $t = 1$. During training, following Ghosh et al. (2020), we randomize the terminal integration time in a small neighborhood of one, sampling

$$t_{\mathrm{end}} \sim \mathrm{Uniform}(0.95, 1.05).$$

The inverse map S_θ^{-1} is obtained by integrating the reversed vector field

$$\widetilde{v}_\theta(u, s, x) = -v_\theta(u, 1-s, x)$$

from $s = 0$ to $s = 1$ using the same integration scheme. The Jacobian determinant does not need to be estimated: by exact preservation of the Gaussian reference measure, it is available in closed form from Equation (4).

To backpropagate through the ODE solution, we differentiate directly through the solver whenever memory permits. When memory is limiting, we instead use the adjoint method of Chen et al. (2018). In the latter case, we employ the seminorm modification proposed by Kidger et al. (2021).

I.2 Description of datasets

Synthetic datasets. All three synthetic datasets are pushforwards of the standard Gaussian $\mathcal{N}(0, \mathrm{I}_2)$ by a C^1-diffeomorphism with unit Jacobian determinant. The following elementary lemma explains why this construction gives analytic HDRs, and hence an exact value for the denominator of the Volume Ratio.

Lemma 26 (Unit-Jacobian transports have analytic HDRs)**.** *Let $T : \mathbb{R}^{d_y} \to \mathbb{R}^{d_y}$ be a C^1-diffeomorphism with $|\det DT| \equiv 1$ and let $\mu := T_{\#}\mathcal{N}(0, \mathrm{I}_{d_y})$. Then μ has density $f(y) = (2\pi)^{-d_y/2} \exp\{-\|T^{-1}(y)\|_2^2/2\}$, which is a strictly decreasing function of $\|T^{-1}(y)\|_2$. Consequently, for every $\alpha \in (0,1)$,*

$$\mathrm{HDR}^f_{1-\alpha} = T\big(\mathcal{B}_{d_y}(0, \sqrt{q_{1-\alpha}})\big), \qquad \mathrm{Vol}\big(\mathrm{HDR}^f_{1-\alpha}\big) = \omega_{d_y}\, q_{1-\alpha}^{d_y/2},$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the $\chi^2_{d_y}$ distribution and $\omega_{d_y} = \mathrm{Vol}(\mathcal{B}_{d_y}(0,1))$.

Proof. The density formula is the change of variables with unit Jacobian. The superlevel sets of f are therefore the sets $\{y : \|T^{-1}(y)\|_2^2 \le q\} = T(\mathcal{B}_{d_y}(0, \sqrt{q}))$, whose μ-mass is $\mathbb{P}(\chi^2_{d_y} \le q)$, and whose Lebesgue volume equals that of the ball because $|\det DT| \equiv 1$. The identification with the HDR follows as in Lemma 14. □

For $d_y = 2$ the following maps are chosen:

- **Banana:**

$$T(u_1, u_2) = \left(2u_1,\ \frac{u_2}{2} + u_1^2 + 8\right).$$

  The Jacobian is triangular and satisfies

$$\det DT(u) = 2 \cdot \frac{1}{2} = 1.$$

  This is a rescaled and shifted variant of the map T_b of Appendix C; the rescaling preserves the unit Jacobian.
- **Sinusoidal:**

$$T(u_1, u_2) = \left(\frac{u_1}{s},\ s u_2 + a \sin(\omega u_1 + \varphi)\right),$$

	NOT		
Dataset	$d = 2$	$d = 4$	$d = 8$
Banana	0.43 [0.38, 0.50]	0.41 [0.40, 0.45]	0.43 [0.37, 0.47]
Sinusoidal	0.89 [0.88, 0.90]	0.88 [0.87, 0.90]	0.89 [0.87, 0.91]
Star-shaped	0.94 [0.94, 0.95]	0.93 [0.91, 0.94]	0.94 [0.90, 0.95]
	NOT-R		
Dataset	$d = 2$	$d = 4$	$d = 8$
Banana	0.96 [0.93, 0.97]	0.91 [0.89, 0.93]	0.89 [0.86, 0.90]
Sinusoidal	0.98 [0.97, 0.98]	0.96 [0.96, 0.97]	0.95 [0.93, 0.96]
Star-shaped	0.98 [0.97, 0.99]	0.97 [0.97, 0.98]	0.97 [0.96, 0.98]

Table 2: Volume ratio VR for Neural OT (NOT) and its rearrangement (NOT-R) on the synthetic datasets in dimension $d \in \{2, 4, 8\}$, at coverage level 90%; median over 10 seeds, interquartile range in brackets. VR closer to 1 means better HDR reconstruction.

where $s > 0$ is the vertical scale, a is the amplitude, ω is the frequency, and φ is the phase. Its Jacobian is

$$DT(u) = \begin{pmatrix} 1/s & 0 \\ a\omega\cos(\omega u_1 + \varphi) & s \end{pmatrix},$$

and hence

$$\det DT(u) = 1.$$

- **Star-Shaped:** writing

$$U = (r\cos\theta, r\sin\theta),$$

let a denote the petal amplitude and define

$$c = \sqrt{1 + \frac{a^2}{2}}, \qquad R(\phi) = \frac{1 + a\cos(3\phi)}{c}.$$

The target angle ϕ is defined implicitly as the unique solution of

$$\theta = \phi + \frac{2a}{3c^2}\sin(3\phi) + \frac{a^2}{12c^2}\sin(6\phi),$$

and the radial coordinate is transformed according to

$$r' = rR(\phi).$$

Thus,

$$T(U) = (r'\cos\phi, r'\sin\phi).$$

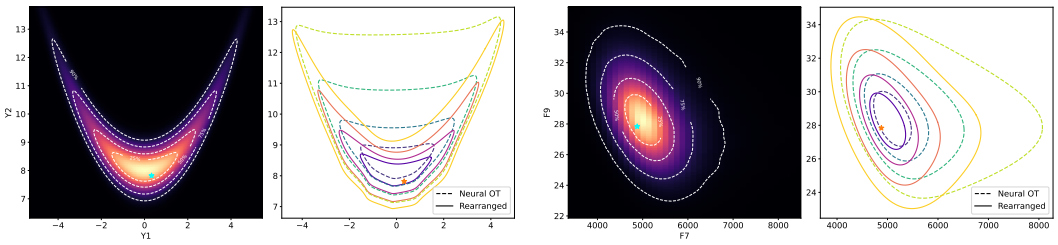
Since

$$\frac{d\theta}{d\phi} = R(\phi)^2,$$

the Cartesian Jacobian determinant satisfies

$$|\det DT| = \frac{r'}{r}\frac{\partial r'}{\partial r}\frac{d\phi}{d\theta} = R(\phi)^2\frac{1}{R(\phi)^2} = 1.$$

For each dataset, we draw 100000 / 1000 / 1000 points for training, calibration, and evaluation. The synthetic experiments are unconditional ($d_x = 0$). To reach dimension $d \in \{4, 8\}$, the planar map is extended by the identity on $d - 2$ additional independent standard Gaussian coordinates; the extended map still has unit Jacobian, so its HDRs remain analytic by Lemma 26. Table 2 reports the volume ratios discussed in Section 5. All synthetic experiments use the coverage level 90%; marginal coverage holds on all datasets and is not reported.



(a) Banana dataset.

(b) Bio dataset.

Figure 4: Effect of the rearrangement on the prediction regions. Within each panel: (estimated) probability density (left); prediction regions induced by optimal transport (dashed) and by its rearrangement (solid) (right).

Real datasets. Table 3 summarizes the benchmarks. `scm20d` is the supply-chain forecasting task from the Mulan multi-target repository; `blog` is BlogFeedback (UCI); `bio` is Physicochemical Properties of Protein Tertiary Structure (UCI); `qm9` is the quantum-chemistry benchmark of Wu et al. (2018). For the single-target datasets `blog` and `bio`, the response is augmented with one input feature used as a second output, following Feldman et al. (2023).

We adapt the Online News Popularity (ONP) dataset to obtain a multivariate regression task with an eight-dimensional continuous response. Specifically, we define the response as the log-transformed number of article shares, global subjectivity, global sentiment polarity, average positive and negative polarity, title subjectivity, title sentiment polarity, and average token length. To prevent target leakage, we remove from the covariates all response variables, deterministic transformations of the responses (e.g., absolute title subjectivity and polarity), other closely related sentiment/polarity statistics, and features derived from historical keyword and self-reference share counts. We additionally discard the URL and time-delta fields. Continuous covariates and each response coordinate are standardized independently using statistics computed on the training set only, while binary indicator variables are left unchanged; the same transformations are subsequently applied to the calibration and test sets.

Table 3: Summary of the real datasets.

Dataset	n	d_x	d_y	train/cal/test	source
`bio`	45730	8	2	60% / 20% / 20%	UCI
`blog`	52396	279	2	60% / 20% / 20%	UCI
`qm9`	130830	100	4	60% / 20% / 20%	Wu et al. (2018)
`scm20d`	8965	61	16	60% / 20% / 20%	Mulan
`onp`	39641	29	8	60% / 20% / 20%	Tsymbalov et al. (2019); Fernandes et al. (2015)

I.3 Description of numerical setup

Base transports. The Brenier map is estimated with Neural Optimal Transport (Kondratyev et al., 2026). We adapt their architecture, using Partial ICNN's (Amos et al., 2017) with 8 layers of size 64 for all our experiments, with Softplus activations.

The triangular baseline is a conditional RealNVP (Dinh et al., 2016) with 4 flow layers, each consisting of 4 layers of size 64.

In both cases the transport is trained on the training split and then *frozen*; the rearrangement never modifies it.

Training the rearrangement. The rearrangement is trained using stochastic gradient descent on the finite-sample loss $\widehat{L}$ of Section 4.2: for each covariate x_i in the minibatch, we sample a coverage level $\alpha_i \sim \mathrm{Unif}(0,1)$, then m reference points $u_i^j \sim \mu_U(\cdot \mid U \in \mathcal{B}_{d_y}(0, r_{1-\alpha_i}))$ by drawing a uniform direction on the unit sphere and an independent radius $G_{d_y}^{-1}(V)$ with $V \sim \mathrm{Uniform}(0, 1-\alpha_i)$, where G_{d_y} is the radial distribution function of the standard Gaussian.

We then average $\log|\det D_u T(S_\theta(u_i^j, x_i), x_i)| + \frac{1}{2}\|S_\theta(u_i^j, x_i)\|_2^2$. We use the Adam optimizer, with default parameters, setting learning rate to $1e-3$ and weight decay to $1e-4$. We also clip gradient norms at 10.

We use minibatch of size 1024, with $m = 16$ samples for volume estimates, for 250 epochs.

Student-t sweep (Table 1). The reference is the standard Gaussian distribution, so that $r_{1-\alpha}$ is the $(1-\alpha)$-quantile of the χ_d distribution. The target has density (1) with

$$g(s) = (1 + s/\nu)^{-(\nu+d)/2}, \quad \nu = 3, \quad m = 0$$

and $\Sigma = \mathrm{diag}(k^{1-1/d}, k^{-1/d}, \ldots, k^{-1/d})$, whose determinant is 1 for every k, so that volumes are comparable across anisotropies. Since $\rho_\Sigma(Y)^2/d \sim F(d,\nu)$, the HDR of mass c is the Mahalanobis ellipsoid with $q_c = d\, F_{d,\nu}^{-1}(c)$ and $\mathrm{Vol}(\mathrm{HDR}_c) = \omega_d\, q_c^{d/2}$.

Estimation of the Brenier map. For $k \neq 1$ no closed form exists (Section D); the map is estimated by a PICNN trained with Neural OT (Kondratyev et al., 2026), with the architecture and training settings of the benchmark above, and its inverse $\hat{Q}^{-1}$ is computed by solving the convex-conjugate problem numerically. The estimated map is the gradient of a convex potential, hence the Brenier map of *some* target rather than of the Student-t law: its regions do not have coverage exactly $1-\alpha$ under μ.

Coverage matching. For each cell (d,k) we estimate the realized coverage $c = \mu(\mathcal{C}_{1-\alpha}^{\hat{Q}})$ from $N_c = 10^5$ Student-t samples and compare $\mathrm{Vol}(\mathcal{C}_{1-\alpha}^{\hat{Q}})$ with $\mathrm{Vol}(\mathrm{HDR}_c)$, the HDR of the same mass; without this matching, a region of mass below $1-\alpha$ can be smaller than $\mathrm{HDR}_{1-\alpha}$ and the ratio is not bounded below by one.

Volume estimation. The volume

$$\mathrm{Vol}(\mathcal{C}_{1-\alpha}^{\hat{Q}}) = \mathrm{Vol}(\mathcal{B}_d(0, r_{1-\alpha}))\, \mathbb{E}_{u\sim\mathrm{Unif}(\mathcal{B}_d(0,r_{1-\alpha}))} |\det D\hat{Q}(u)|$$

is estimated with $M = 10^5$ uniform points by log-sum-exp.

Negative control. For $k = 1$ the target is spherical and the Brenier map is the closed-form radial map $u \mapsto a(\|u\|_2)\, u/\|u\|_2$ with $a = F_\nu^{-1} \circ G_d$, where G_d is the distribution function of the χ_d law and $F_\nu(t) = \mathbb{P}(F(d,\nu) \le t^2/d)$ (Theorem 18(ii)), so the true gap is exactly zero. Running the same estimation pipeline at $k = 1$ therefore measures its total error per dimension, and the quantity comparable with the population gap is $\mathrm{NVF}(k) - \mathrm{NVF}(k{=}1)$.

Calibration and evaluation. Split conformal calibration uses the score

$$R_i = \left\| \hat{S}^{-1}\big(\hat{T}^{-1}(y_i, x_i), x_i\big) \right\|_2$$

on the calibration split, as in Section 4, with a coverage level of $1-\alpha = 0.9$.

For transport-based models, we compute the volume as

$$\mathrm{Vol}(C_{1-\alpha}^F) = \int_{\mathcal{B}(0,r_{1-\alpha})} |\det DF(u)|\, du,$$

where F denotes either a transport map or its composition with a measure-preserving rearrangement. We estimate this integral by Monte Carlo using $k = 10{,}000$ samples

$$u_i \sim \mathrm{Uniform}\big(\mathcal{B}(0, r_{1-\alpha})\big).$$

In particular, we report the log-volume estimate

$$\log \widehat{\mathrm{Vol}}(C_{1-\alpha}^F) = \log \mathrm{Vol}\big(\mathcal{B}(0, r_{1-\alpha})\big) + \mathrm{logsumexp}_{i=1,\ldots,k} \left(\log|\det DF(u_i)|\right) - \log k.$$

For ellipsoid-based methods, the prediction region at x is of the form

$$C_{1-\alpha}(x) = \left\{ y \in \mathbb{R}^{d_y} : (y - \mu(x))^\top \Sigma(x)^{-1} (y - \mu(x)) \le r_{1-\alpha}^2 \right\}.$$

Its volume is available in closed form:

$$\mathrm{Vol}(C_{1-\alpha}(x)) = \frac{\pi^{d_y/2}}{\Gamma(d_y/2+1)} r_{1-\alpha}^{d_y} \det\big(\Sigma(x)^{-1}\big)^{-1/2}.$$

For numerical stability, we compute it in the log domain as

$$\log \mathrm{Vol}(C_{1-\alpha}(x)) = \frac{d_y}{2}\log\pi - \log\Gamma\left(\frac{d_y}{2}+1\right) + d_y \log r_{1-\alpha} - \frac{1}{2}\log\det\big(\Sigma(x)^{-1}\big).$$

Finally, when neither of the previous approaches is available, we estimate the volume by Monte Carlo, assuming that membership in the prediction region $\widetilde{C}_{1-\alpha}(x)$ can be evaluated. We first construct a global bounding box by computing the coordinatewise minimum $y^{\min}$ and maximum $y^{\max}$ prediction values over the training set. We then draw $k = 10{,}000$ samples uniformly from this box and estimate

$$\widehat{\mathrm{Vol}}\big(\widetilde{C}_{1-\alpha}(x)\big) = \mathrm{Vol}(B)\frac{1}{k}\sum_{i=1}^{k} \mathbb{I}\left[y_i \in \widetilde{C}_{1-\alpha}(x)\right], \qquad y_i \sim \mathrm{Uniform}(B),$$

where

$$B = \prod_{j=1}^{d_y} [y_j^{\min}, y_j^{\max}], \qquad \mathrm{Vol}(B) = \prod_{j=1}^{d_y} (y_j^{\max} - y_j^{\min}).$$

The HDR volumes of the synthetic target distributions are available analytically by Lemma 26.

Compute. All experiments were run on a single NVIDIA Blackwell GPU (96 GB).

I.4 Metrics

Normalized Log-Volume. Instead of reporting volume directly, we report the dimension-normalized log-volume, defined as

$$\mathrm{LV}(C) = \log\left(\mathrm{Vol}(C)^{1/d_y}\right) = \frac{\log \mathrm{Vol}(C)}{d_y}. \tag{27}$$

In our benchmarks, we suppress the dependence on C and refer to this metric simply as LV.

Marginal Coverage. To compute marginal coverage, we report

$$\text{Marginal Coverage} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left[y_i \in C_{1-\alpha}(x_i)\right]$$

over held-out test set, where $C_{1-\alpha}$ represents the prediction region.

Worst-Slab Coverage. We compute the worst-slab coverage metric (WSC) of Cauchois et al. (2021) over $10{,}000$ randomly sampled directions.

Excess Coverage Risk. We report the Excess Coverage Risk (ECR) of Braun et al. (2026b), using the ℓ_2 base norm and five-fold cross-validation. The metric is defined as

$$\mathrm{ECR} = \mathbb{E}_X\left[\left\|(1-\alpha) - \mathbb{P}\big(Y \in C(X) \mid X\big)\right\|_2\right].$$

To estimate the conditional coverage probability, we train a LightGBM model (Ke et al., 2017) using the default configuration of Holzmüller et al. (2024).

I.5 Extended benchmark results

In this section, we report the full comparison introduced in Section 5. For completeness, we include residual-based methods: OT-CP, OT-CP+ (Thurin et al., 2025), and ellipsoidal calibration (Elliptic; Messoudi et al., 2022). To compute the residuals, we first fit a random forest regressor and then calibrate the resulting residuals on a held-out calibration set. For CONTRA (Fang et al., 2025), we train a RealNVP model (Dinh et al., 2016) using the hyperparameters specified above; in all tables, CONTRA is denoted NF and its rearrangement NF-R.

Table 4: Normalized log-volume (LV) across datasets (lower is better). We report the median with the interquartile range (25%–75% quantiles) in brackets.

Method	Bio	Blog	QM9	SCM20D	ONP
NOT	4.53 [4.53, 4.54]	2.86 [2.85, 2.89]	.24 [.23, .25]	5.82 [5.79, 5.83]	1.04 [1.03, 1.04]
NOT-R	4.50 [4.50, 4.51]	2.81 [2.80, 2.83]	.23 [.22, .24]	5.80 [5.78, 5.82]	1.00 [0.99, 1.00]
NF	4.49 [4.49, 4.55]	2.96 [2.88, 2.97]	−.05 [−.06, −.04]	6.76 [6.66, 6.92]	1.19 [1.16, 1.24]
NF-R	4.41 [4.40, 4.41]	2.54 [2.50, 2.60]	−.05 [−.06, −.04]	6.53 [6.46, 6.61]	1.26 [1.14, 1.29]
HDR-Split	4.46 [4.44, 4.46]	2.73 [2.69, 2.81]	.33 [.31, .35]	6.91 [6.90, 6.93]	1.05 [1.04, 1.06]
CP-Split	4.40 [4.39, 4.41]	2.49 [2.45, 2.50]	.21 [.19, .22]	6.37 [6.35, 6.39]	1.06 [1.05, 1.06]
OT-CP	4.54 [4.53, 4.55]	3.17 [3.14, 3.18]	.93 [.93, .94]	6.42 [6.41, 6.44]	1.29 [1.28, 1.30]
OT-CP+	4.95 [4.91, 5.01]	3.39 [3.37, 3.42]	.97 [.96, .98]	6.43 [6.42, 6.45]	1.34 [1.33, 1.34]
Elliptic	4.68 [4.64, 4.74]	3.51 [3.48, 3.54]	.90 [.89, .90]	6.05 [6.03, 6.07]	1.19 [1.18, 1.19]

Table 5: Excess coverage risk across datasets. We report the median with the interquartile range (25%–75% quantiles) in brackets.

Method	Bio	Blog	QM9	SCM20D	ONP
NOT	.06 [.06, .06]	.07 [.07, .07]	.07 [.07, .07]	.08 [.08, .08]	.06 [.06, .06]
NOT-R	.06 [.06, .06]	.07 [.06, .07]	.07 [.07, .07]	.08 [.08, .08]	.06 [.06, .06]
NF	.05 [.05, .06]	.06 [.06, .06]	.06 [.06, .06]	.07 [.07, .08]	.06 [.06, .06]
NF-R	.06 [.05, .06]	.06 [.06, .06]	.06 [.06, .06]	.07 [.07, .08]	.06 [.06, .06]
HDR-Split	.06 [.05, .06]	.07 [.07, .07]	.07 [.07, .08]	.10 [.10, .10]	.06 [.06, .06]
CP-Split	.06 [.06, .06]	.07 [.07, .07]	.07 [.07, .07]	.08 [.08, .09]	.06 [.06, .06]
OT-CP	.06 [.06, .07]	.08 [.07, .08]	.07 [.07, .08]	.07 [.07, .08]	.06 [.06, .06]
OT-CP+	.05 [.05, .06]	.06 [.05, .06]	.07 [.06, .07]	.07 [.07, .08]	.06 [.06, .06]
Elliptic	.05 [.05, .06]	.06 [.06, .06]	.07 [.07, .07]	.09 [.09, .10]	.06 [.06, .06]

Table 6: Worst-slab coverage (WSC) across datasets. We report the median with the interquartile range (25%–75% quantiles) in brackets.

Method	Bio	Blog	QM9	SCM20D	ONP
NOT	.87 [.86, .88]	.78 [.76, .78]	.86 [.85, .87]	.90 [.87, .91]	.90 [.88, .90]
NOT-R	.87 [.85, .88]	.78 [.76, .80]	.86 [.85, .87]	.88 [.87, .92]	.89 [.88, .89]
NF	.89 [.88, .89]	.81 [.80, .82]	.88 [.87, .88]	.89 [.88, .91]	.90 [.89, .91]
NF-R	.89 [.88, .90]	.81 [.80, .82]	.88 [.87, .88]	.91 [.89, .91]	.88 [.86, .89]
HDR-Split	.90 [.89, .92]	.85 [.83, .86]	.90 [.89, .91]	1.00 [1.00, 1.00]	.90 [.89, .91]
CP-Split	.79 [.78, .81]	.75 [.75, .76]	.80 [.79, .80]	.86 [.83, .87]	.79 [.78, .81]
OT-CP	.68 [.67, .70]	.73 [.72, .75]	.77 [.77, .78]	.87 [.85, .88]	.80 [.78, .82]
OT-CP+	.90 [.89, .90]	.88 [.88, .89]	.87 [.87, .89]	.90 [.89, .93]	.90 [.89, .91]
Elliptic	.81 [.80, .83]	.87 [.85, .87]	.82 [.81, .82]	.86 [.85, .90]	.87 [.84, .88]

Table 7: Marginal coverage across datasets. We report the median with the interquartile range (25%–75% quantiles) in brackets.

Method	Bio	Blog	QM9	SCM20D	ONP
NOT	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .91]	.90 [.90, .90]
NOT-R	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .91]	.90 [.90, .90]
NF	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .90]	.90 [.89, .91]	.90 [.90, .90]
NF-R	.90 [.90, .90]	.90 [.89, .90]	.90 [.90, .90]	.90 [.90, .91]	.90 [.90, .90]
HDR-Split	.91 [.90, .91]	.92 [.92, .93]	.92 [.92, .92]	1.00 [1.00, 1.00]	.90 [.90, .91]
CP-Split	.90 [.90, .90]	.89 [.89, .90]	.90 [.90, .90]	.90 [.89, .90]	.90 [.90, .90]
OT-CP	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .90]	.91 [.89, .92]	.90 [.89, .90]
OT-CP+	.91 [.90, .91]	.90 [.90, .91]	.90 [.90, .91]	.91 [.90, .92]	.90 [.90, .91]
Elliptic	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .90]	.90 [.90, .91]	.90 [.90, .90]