

OUT-OF-DISTRIBUTION WASSERSTEIN AMBIGUITY SET CONSTRUCTION WITH INCLUSIVE AND EXCLUSIVE EVIDENCE

BIAO HAN

ABSTRACT. Distributionally robust optimization hedges a decision against an ambiguity set of candidate distributions. In a cold-start setting the target distribution is unobserved and no sample from it is available, so only related source distributions provide evidence for constructing that set. We propose an inclusive-exclusive Wasserstein ball model that jointly learns a center and radius from inclusive sources believed compatible with the target and exclusive sources believed incompatible with it. Soft penalties balance enclosure and exclusion against ball size and margin. We identify parameter conditions for unboundedness and radius collapse, and show that small regularization permits an exact relaxation of the radius constraint. In this regime the Lagrangian dual involves a signed Wasserstein barycenter, and, assuming an exact signed-barycenter oracle, we propose a cutting-plane learning algorithm that constructs the ball and certifies bounds on the dual optimum and on the suboptimality of candidate balls. For quadratic transport under small regularization, we establish exact convex reformulations for univariate laws and for translations and scalings of a common base law. These yield finite-dimensional convex quadratic programs for univariate empirical sources and common-base families, with explicit recovery of a globally optimal ball. The learned ball then serves as the ambiguity set of the downstream robust out-of-distribution stochastic decision.

1. INTRODUCTION

Stochastic programming minimizes an expected cost under a probability law (Shapiro et al., 2021). When that law is uncertain, distributionally robust optimization (DRO) minimizes the worst-case expected cost over an *ambiguity set* of candidate laws (Kuhn et al., 2025). Data-driven Wasserstein DRO typically constructs this set around the empirical distribution of observations from the target law (Mohajerin Esfahani and Kuhn, 2018). In a cold-start decision, however, neither the target distribution nor a single sample from it is available. For example, stocking a new product must rely on demand records of other products. The central question is then how to construct an ambiguity set from related distributions alone.

Robust out-of-distribution stochastic optimization (RooDSO) (Li et al., 2026) addresses this setting by treating the sources and the unseen target as realizations of a meta-distribution and learning a support description from the sources. Kernel mean embeddings represent distributions as points in a reproducing kernel Hilbert space (Muandet et al., 2017), where support descriptions can be learned geometrically (Muandet and Schölkopf, 2013). Wasserstein geometry instead compares distributions directly through transport costs (Villani, 2009); enclosing-ball constructions in this space provide a way to learn both a center distribution and a radius (Han, 2026; Wang et al., 2025). These constructions supply a natural starting point, but a description based only on compatible sources does not explicitly use evidence about which distributions should be excluded.

In an era of superintelligence, the value judgment of a human decision maker is a scarce resource that deserves to be exploited to the full. Human judgment about the unknown rarely confines itself to what the unknown *is*; it usually extends to what the unknown is *not*. China’s classical *I Ching* (*Book of Changes*) (Wilhelm, 1967) likewise embodies a philosophy in which the ancient sages

Date: October 11, 2026.

2020 Mathematics Subject Classification. 90C15, 90C46, 62G35.

Key words and phrases. out-of-distribution stochastic optimization; Wasserstein ambiguity set; distributionally robust optimization; inclusive and exclusive evidence; signed Wasserstein barycenter; convex reformulation.

characterized the unknown through the complementary pair of *yin* and *yang*. Negative evidence is no less informative than positive evidence, and it too should inform the decision. Information from both sides is the most basic and most useful value judgment that human intelligence can supply.

Such negative information is often available alongside positive information. For a new product, some existing products may suggest plausible demand, while others may be known to follow incompatible demand patterns. We call the former *inclusive evidence* and the latter *exclusive evidence*. The two families play different roles: inclusive evidence guides the region of plausible target laws, whereas exclusive evidence helps delimit that region. It is not hard to see that exclusive evidence may calibrate the ambiguity set and so mitigate misspecification (the target law lying outside it) to some degree. Their use in ball learning is motivated by support vector data descriptions with negative examples (Tax and Duin, 2004; Mu and Nandi, 2009) and small-sphere–large-margin models (Liu et al., 2026). Here the objects to be enclosed or repelled are probability distributions, and the ball is constructed in Wasserstein space.

For $p \geq 1$, let $\mathcal{P}_p(\Xi)$ denote the Borel probability measures on an observation space $\Xi \subseteq \mathbb{R}^d$ with finite p th moments. The decision problem is as follows.

Problem 1 (Out-of-distribution decision with inclusive and exclusive evidence). *Let $\mathbb{P}^* \in \mathcal{P}_p(\Xi)$ be a target law from which no observations are available. The available data are finite samples from inclusive source laws $\{\mathbb{P}_i^+\}_{i=1}^{N^+}$ and exclusive source laws $\{\mathbb{P}_j^-\}_{j=1}^{N^-}$ in $\mathcal{P}_p(\Xi)$, labeled according to their believed compatibility or incompatibility with the target. Given a feasible decision set $\mathcal{X} \subseteq \mathbb{R}^n$ and a cost function $f: \mathcal{X} \times \Xi \rightarrow \mathbb{R}$, construct a Wasserstein ambiguity set $\mathcal{U} \subseteq \mathcal{P}_p(\Xi)$ from both evidence families and choose a decision by solving*

$$\min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbb{Q} \in \mathcal{U}} \mathbb{E}_{\xi \sim \mathbb{Q}} \{f(\mathbf{x}, \xi)\}. \quad (1.1)$$

The proposed construction learns a small Wasserstein ball while penalizing inclusive sources left outside and exclusive sources lying inside a surrounding margin. The evidence is soft: the labels guide the fitted geometry rather than impose exact membership constraints. Moving this construction from a Hilbert space to Wasserstein space requires a separate optimization analysis, since the inner-product identities underlying Hilbert-space reformulations are generally unavailable. Statistical coverage likewise requires assumptions relating the sources to the unseen target; compatibility labels alone do not ensure it.

The main contributions are as follows:

- (1) We formulate Wasserstein ball learning with inclusive and exclusive evidence, recovering inclusive-only support learning as a special case. We identify conditions for unboundedness and radius collapse, and establish an exact radius relaxation in the small-regularization regime.
- (2) We derive the Lagrangian dual of the relaxed model and express its inner problem as a signed Wasserstein barycenter. With an exact oracle, the resulting cutting-plane method provides dual optimality bounds and candidate-ball suboptimality bounds, distinguishing the two certificates without assuming strong duality.
- (3) For quadratic transport under small regularization, we prove exact convex reformulations for univariate laws and common-base translation and scaling families. We obtain finite-dimensional quadratic programs for univariate empirical sources and common-base families, and recover global optimizers of the original model.

Section 2 presents the model. Section 3 develops the dual and the learning algorithm. Section 4 establishes exact convex reformulations for quadratic transport under the stated geometric conditions. Section 5 concludes. Detailed proofs are collected in Appendix A.

2. METHODOLOGY

We construct the ambiguity set in Problem 1 by learning a ball from both evidence families. Inclusive sources are encouraged to lie inside the ball, while exclusive sources are encouraged to

remain beyond an outer margin. Both requirements allow violations, reflecting the uncertainty of the evidence labels.

2.1. Wasserstein Geometry. For $p \geq 1$, the p -Wasserstein transport cost is

$$\mathcal{W}_p^p(\mathbb{P}, \mathbb{Q}) = \inf_{\pi \in \Pi(\mathbb{P}, \mathbb{Q})} \int_{\Xi \times \Xi} \|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p d\pi(\boldsymbol{\xi}, \boldsymbol{\zeta}), \quad (2.1)$$

where $\Pi(\mathbb{P}, \mathbb{Q})$ is the set of couplings with marginals $\mathbb{P}$ and $\mathbb{Q}$. Throughout, $\|\cdot\|$ denotes the Euclidean norm and $\top$ denotes transposition. We write $\delta_{\boldsymbol{\xi}}$ for the Dirac law at $\boldsymbol{\xi}$. The p th root of the cost is the metric $\mathcal{W}_p$ (Villani, 2009). A center $\mathbb{P}_c \in \mathcal{P}_p(\Xi)$ and a radius $R \geq 0$ therefore describe the ball

$$\{\mathbb{Q} \in \mathcal{P}_p(\Xi) \mid \mathcal{W}_p(\mathbb{Q}, \mathbb{P}_c) \leq R\}.$$

We write $\rho := R^p$ and measure the margin in transport-cost units as well. The ball radius remains R , whereas R^p is a radius penalty, not a volume in the space of probability measures.

2.2. Inclusive-Exclusive Wasserstein Ball Learning. Let $N^+ \geq 1$, $N^- \geq 0$, and $N := N^+ + N^-$. Following the small-sphere-large-margin principle (Liu et al., 2026), we propose

$$\inf_{\substack{\mathbb{P}_c \in \mathcal{P}_p(\Xi) \\ \rho \geq 0, t \geq 0}} \nu \rho - \mu t + \frac{1}{N} \sum_{i=1}^{N^+} (\mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{P}_c) - \rho)_+ + \frac{b}{N} \sum_{j=1}^{N^-} (\rho + t - \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{P}_c))_+, \quad (2.2)$$

where $(z)_+ := \max\{z, 0\}$, $\nu > 0$, $b > 0$, and $\mu \geq 0$. The inclusive hinge penalizes a source outside radius $\rho^{1/p}$; the exclusive hinge penalizes a source inside the larger radius $(\rho + t)^{1/p}$. Thus t is the difference of radius powers, rather than the difference of radii.

The parameter ν penalizes the ball size, b weights exclusive violations relative to inclusive ones, and μ rewards a larger margin. The normalization by N assigns the same base weight to each source; the influence of each family also depends on its cardinality. These parameters govern the joint trade-off between enclosure, exclusion, and size.

Introducing nonnegative slacks gives the equivalent formulation

$$\begin{aligned} \inf_{\substack{\mathbb{P}_c \in \mathcal{P}_p(\Xi), \rho, t \geq 0 \\ \omega^+ \geq 0, \omega^- \geq 0}} \quad & \nu \rho - \mu t + \frac{1}{N} \sum_{i=1}^{N^+} \omega_i^+ + \frac{b}{N} \sum_{j=1}^{N^-} \omega_j^- \\ \text{s.t.} \quad & \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{P}_c) \leq \rho + \omega_i^+, \quad i = 1, \dots, N^+, \\ & \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{P}_c) \geq \rho + t - \omega_j^-, \quad j = 1, \dots, N^-. \end{aligned} \quad (2.3)$$

For a fixed center, both formulations are convex in their remaining variables. The exclusive constraints, however, generally destroy convexity in the center distribution.

Note that problems (2.2) and (2.3) are generally non-convex, because the exclusive side is not convex in the center. When $N^- = 0$ and $\mu = 0$, we may take $t = 0$. Dividing the objective by ν reduces (2.2) to

$$\inf_{\mathbb{P}_c \in \mathcal{P}_p(\Xi), \rho \geq 0} \rho + \frac{1}{N^+ + \nu} \sum_{i=1}^{N^+} (\mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{P}_c) - \rho)_+.$$

For $\nu \in (0, 1]$, this is the inclusive-only Wasserstein meta-distribution support learning model (Han, 2026). Its exclusion-level interpretation of ν does not automatically carry over to the two-sided model.

2.3. The Hyperparameter Regime. The margin reward must be controlled to prevent the objective from decreasing without bound. The following lemma shows the cases in which the objective is unbounded below.

Lemma 1 (Unboundedness). *For every $p \geq 1$, problems (2.2) and (2.3) are unbounded below if either (i) Ξ is nonempty and $\mu > bN^-/N$, or (ii) Ξ is unbounded and $\mu > N^+/N$. The proof is given in Appendix A.1.*

Hence, when the observation space Ξ is allowed to be unbounded, the condition that must be imposed to ensure a finite optimal value is

$$0 \leq \mu \leq \min \left\{ \frac{N^+}{N}, \frac{bN^-}{N} \right\}. \quad (2.4)$$

Write $F(\mathbb{Q}, \rho, t)$ for the objective in (2.2). The next result shows that below a threshold on the hyperparameters the radius constraint may be dropped without changing the infimum or the global minimizers.

Theorem 2 (Small-Regularization Regime). *Let $p \geq 1$, $\nu, b > 0$, $0 \leq \mu \leq bN^-/N$ (together with (2.5) below, this implies (2.4)), and suppose*

$$\nu + \mu < \frac{N^+}{N}. \quad (2.5)$$

Then (2.2) has the same infimum and global minimizers as

$$\inf_{\substack{\mathbb{Q} \in \mathcal{P}_p(\Xi) \\ \rho \in \mathbb{R}, t \geq 0}} F(\mathbb{Q}, \rho, t). \quad (2.6)$$

In particular, every global minimizer of (2.6) has $\rho \geq 0$, and every original global minimizer also solves the relaxed problem. The same conclusion holds for (2.3) when its radius constraint is relaxed. The proof is given in Appendix A.2.

When the hyperparameters exceed the threshold above, the model degenerates in the ways that the following theorem describes.

Theorem 3 (Large-Regularization Regime). *Let $p \geq 1$, $\nu, b > 0$, $0 \leq \mu \leq bN^-/N$, and assume*

$$\nu + \mu \geq \frac{N^+}{N}. \quad (2.7)$$

Write $\kappa = N^+/N$. Then the following conclusions hold.

(i) Zero-radius reduction. *Problem (2.2) has the same infimum as*

$$\inf_{\substack{\mathbb{Q} \in \mathcal{P}_p(\Xi) \\ t \geq 0}} \left\{ -\mu t + \frac{1}{N} \sum_i \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}) + \frac{b}{N} \sum_j (t - \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{Q}))_+ \right\}. \quad (2.8)$$

Every optimizer of this reduced problem gives an original optimizer with $\rho = 0$. Conversely, any original optimizer $(\mathbb{Q}^, \rho^*, t^*)$ gives a reduced optimizer $(\mathbb{Q}^*, \rho^* + t^*)$. If $\nu + \mu > \kappa$, all original optimizers have $\rho^* = 0$.*

(ii) No margin reward. *If $\mu = 0$, all original optimizers are precisely the triples $(\hat{\mathbb{Q}}, \rho, t)$ where $\hat{\mathbb{Q}}$ is an inclusive Wasserstein barycenter, that is, it minimizes $\sum_i \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q})$, and*

$$\begin{aligned} \rho, t &\geq 0, & (\nu - \kappa)\rho &= 0, \\ \rho &\leq \min_i \mathcal{W}_p^p(\mathbb{P}_i^+, \hat{\mathbb{Q}}), & \rho + t &\leq \min_j \mathcal{W}_p^p(\mathbb{P}_j^-, \hat{\mathbb{Q}}). \end{aligned} \quad (2.9)$$

The minimum over an empty exclusive family is $+\infty$.

The proof is given in Appendix A.3.

This paper mainly focuses on the non-degenerate case where the hyperparameters satisfy (2.5).

2.4. The Induced Ambiguity Set. An optimal solution of (2.2) supplies both a center and a radius, and the pair induces the ambiguity set that the decision is hedged against.

Definition 1 (Inclusive-Exclusive Wasserstein Ambiguity Set). *For an optimal solution $(\mathbb{P}_c^*, \rho^*, t^*)$ of (2.2), set $R^* := (\rho^*)^{1/p}$. The learned inclusive-exclusive Wasserstein ambiguity set is*

$$\mathcal{U}(\mathbb{P}_c^*, R^*) := \{\mathbb{Q} \in \mathcal{P}_p(\Xi) \mid \mathcal{W}_p(\mathbb{Q}, \mathbb{P}_c^*) \leq R^*\}. \quad (2.10)$$

Both evidence families determine the center and radius, but the margin t^* guides learning and does not enlarge the decision ambiguity set. Since violations are penalized rather than forbidden, some inclusive sources may remain outside and some exclusive sources may remain inside.

With finite source samples, the construction uses the empirical laws

$$\widehat{\mathbb{P}}_i^+ = \frac{1}{n_i^+} \sum_{k=1}^{n_i^+} \delta_{\xi_{ik}^+}, \quad \widehat{\mathbb{P}}_j^- = \frac{1}{n_j^-} \sum_{k=1}^{n_j^-} \delta_{\xi_{jk}^-},$$

in place of the source laws in (2.2). Substituting the resulting ball into (1.1) gives the robust decision problem.

Once the ball is fixed, (1.1) is a Wasserstein distributionally robust optimization problem. Under suitable assumptions on the loss function, decision set, and support, Wasserstein duality provides tractable reformulations (Mohajerin Esfahani and Kuhn, 2018; Zhao and Guan, 2018; Blanchet and Murthy, 2019; Gao and Kleywegt, 2023), including finite-dimensional convex programs for finitely supported nominal laws and appropriate problem structure (Kuhn et al., 2019, 2025). For the inclusive-only case of the model proposed here, that decision stage, together with its reformulation and the search over adverse scenarios, is developed in (Han, 2026). Since this paper concerns the construction of the ambiguity set, we do not repeat that treatment.

3. LEARNING STRATEGY

3.1. Dual Formulation. To learn the inclusive-exclusive Wasserstein ball proposed in Section 2, we present its dual in the following theorem.

Theorem 4 (Lagrangian Dual). *Suppose Ξ is nonempty, $p \geq 1$, $0 \leq \mu \leq bN^-/N$, and (2.5) holds. Define*

$$\mathcal{A} := \left\{ (\alpha^+, \alpha^-) \left| \begin{array}{l} 0 \leq \alpha_i^+ \leq \frac{1}{N}, \quad 0 \leq \alpha_j^- \leq \frac{b}{N}, \quad \forall i, j \\ \sum_i \alpha_i^+ - \sum_j \alpha_j^- = \nu, \quad \sum_j \alpha_j^- \geq \mu \end{array} \right. \right\}. \quad (3.1)$$

The Lagrangian dual of the radius-relaxed problem (2.6) has the equivalent form

$$\sup_{(\alpha^+, \alpha^-) \in \mathcal{A}} \inf_{\mathbb{Q} \in \mathcal{P}_p(\Xi)} \left\{ \sum_i \alpha_i^+ \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}) - \sum_j \alpha_j^- \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{Q}) \right\}. \quad (3.2)$$

By Theorem 2, (2.5) gives (2.6) the same infimum as (2.2). The proof is given in Appendix A.4.

The two evidence families enter (3.2) with opposite signs. Merging them into a single family indexed by $k = 1, \dots, N$, with

$$\mathbb{P}_k := \begin{cases} \mathbb{P}_k^+, & k \leq N^+, \\ \mathbb{P}_{k-N^+}^-, & k > N^+, \end{cases} \quad y_k := \begin{cases} +1, & k \leq N^+, \\ -1, & k > N^+, \end{cases} \quad c_k := \begin{cases} \frac{1}{N}, & k \leq N^+, \\ \frac{b}{N}, & k > N^+, \end{cases}$$

the label y_k records whether source k carries inclusive or exclusive evidence. Writing α_k for the multiplier of source k , the feasible set (3.1) becomes

$$\mathcal{A} = \left\{ \alpha \geq 0 \mid \alpha_k \leq c_k \quad \forall k, \quad \sum_{k=1}^N y_k \alpha_k = \nu, \quad \sum_{k: y_k = -1} \alpha_k \geq \mu \right\},$$

and (3.2) reads

$$\sup_{\alpha \in \mathcal{A}} \inf_{\mathbb{Q} \in \mathcal{P}_p(\Xi)} \sum_{k=1}^N y_k \alpha_k \mathcal{W}_p^p(\mathbb{P}_k, \mathbb{Q}). \quad (3.3)$$

The inner problem in (3.3) is a barycenter problem with weights of both signs. Since $\sum_k y_k \alpha_k = \nu > 0$, the scaled weights $a_k := y_k \alpha_k / \nu$ sum to one, so the inner minimization asks for a *generalized* (or *signed*) *Wasserstein barycenter* of the sources, which is a nascent subject: Gallouët et al. (2025) start from the extrapolation of a Wasserstein geodesic beyond its endpoints, with one positive and one negative weight; Tornabene et al. (2025) pose the quadratic-cost problem against a signed measure of unit total mass and settle its existence and uniqueness, Jacobs and Zhou (2026) extend it to general transport costs and to a dual in Kantorovich potentials, and Gallouët et al. (2026) obtain convex duals through Toland duality together with a Sinkhorn-based particle algorithm. When $y_k \equiv +1$ with nonnegative weights it is the classical Wasserstein barycenter (Agueh and Carlier, 2011) and recovers the inclusive-only model of (Han, 2026).

3.2. Signed Wasserstein Barycenter Oracle. We assume that the inner signed barycenter problem in (3.3) can be solved by an oracle; such an oracle may, for instance, be obtained by entropic or particle methods (Cuturi and Doucet, 2014; Gallouët et al., 2026). Writing $\mathbf{y} \odot \alpha := (y_1 \alpha_1, \dots, y_N \alpha_N)^\top$ for the signed weights, we define the oracle as follows.

Definition 2 (Signed-barycenter oracle). *A signed-barycenter oracle is a subroutine that, given the inclusive and exclusive source distributions $\{\mathbb{P}_k\}_{k=1}^N$ and the signed weights $\mathbf{y} \odot \alpha$, returns a law $\hat{\mathbb{Q}}$ and the value attained at it. The law solves the signed barycenter problem,*

$$\hat{\mathbb{Q}} \in \text{SBar}_p(\{\mathbb{P}_k\}_{k=1}^N; \mathbf{y} \odot \alpha) := \arg \min_{\mathbb{Q} \in \mathcal{P}_p(\Xi)} \sum_{k=1}^N y_k \alpha_k \mathcal{W}_p^p(\mathbb{P}_k, \mathbb{Q}), \quad (3.4)$$

so that $\hat{\mathbb{Q}}$ is a signed Wasserstein barycenter of the sources, and the value is

$$\Omega(\alpha) := \sum_{k=1}^N y_k \alpha_k \mathcal{W}_p^p(\mathbb{P}_k, \hat{\mathbb{Q}}). \quad (3.5)$$

Two structural facts drive what follows. Each transport cost is convex under linear mixtures of the center law $\mathbb{Q}$, so the signed objective in (3.4) is a difference of convex functions. The inner problem is generally nonconvex, and a minimizer need not be unique. Once $\mathbb{Q}$ is fixed, however, the objective in (3.5) is affine in α ; being the infimum over $\mathbb{Q}$ of such affine functions, Ω is concave on the polytope $\mathcal{A}$, and every signed barycenter supplies a supergradient, available in closed form:

$$\mathbf{g} := (y_k \mathcal{W}_p^p(\mathbb{P}_k, \hat{\mathbb{Q}}))_{k=1}^N \in \partial\Omega(\alpha), \quad \hat{\mathbb{Q}} \in \text{SBar}_p(\{\mathbb{P}_k\}_{k=1}^N; \mathbf{y} \odot \alpha), \quad (3.6)$$

where $\partial\Omega(\alpha)$ denotes the superdifferential of the concave function Ω at α .

3.3. Learning Algorithm. Under (2.4) and (2.5), we solve the dual by constraint generation, using the exact oracle of Section 3.2. Problem (3.3) has the equivalent semi-infinite formulation

$$\begin{aligned} \max_{\alpha \in \mathcal{A}, \theta \in \mathbb{R}} \quad & \theta \\ \text{s.t.} \quad & \theta \leq \sum_{k=1}^N y_k \alpha_k \mathcal{W}_p^p(\mathbb{P}_k, \mathbb{Q}), \quad \forall \mathbb{Q} \in \mathcal{P}_p(\Xi). \end{aligned} \quad (3.7)$$

For fixed weights, the oracle finds the most violated constraint. Each returned center adds one linear constraint to a finite master problem; no additional center variables are introduced into the master.

At iteration r , let the oracle return $\mathbb{Q}^r$ and $v_r = \Omega(\alpha^r)$. Set $d_k^r = \mathcal{W}_p^p(\mathbb{P}_k, \mathbb{Q}^r)$ and $\mathbf{g}^r = (y_k d_k^r)_{k=1}^N$, the supergradient from (3.6) at this iteration. Then

$$\Omega(\alpha) \leq v_r + \langle \mathbf{g}^r, \alpha - \alpha^r \rangle = \langle \mathbf{g}^r, \alpha \rangle.$$

Retaining all such cuts yields the linear master problem

$$\begin{aligned} U_r := \max_{\alpha \in \mathcal{A}, \theta \in \mathbb{R}} \quad & \theta \\ \text{s.t.} \quad & \theta \leq \langle \mathbf{g}^\ell, \alpha \rangle, \quad \ell = 0, \dots, r. \end{aligned} \quad (3.8)$$

Its optimal weights define α^{r+1} . If v_D denotes the dual optimal value, then

$$L_r := \max_{0 \leq \ell \leq r} v_\ell \leq v_D \leq U_r. \quad (3.9)$$

Thus $U_r - L_r \leq \varepsilon_D$ certifies an ε_D -optimal dual solution among the evaluated weights. The bounds L_r and U_r are respectively nondecreasing and nonincreasing.

Each center also yields a candidate ball by solving

$$J_r := \min_{\rho, t \geq 0} F(\mathbb{Q}^r, \rho, t). \quad (3.10)$$

This is a linear program in the slack formulation (2.3), since the distances d_k^r are fixed. Keeping the best candidate gives

$$B_r := \min_{0 \leq \ell \leq r} J_\ell, \quad L_r \leq v_D \leq v_P \leq B_r, \quad (3.11)$$

where v_P is the original optimal value. Hence $B_r - L_r$ bounds the suboptimality of the returned ball. Dual termination alone does not certify primal optimality, since a duality gap may remain.

Algorithm 1 reports the remaining gaps if the iteration limit is reached. Its certificates require exact oracle values; an approximate inner solver needs certified error bounds to retain valid dual lower bounds. Under the geometry of Section 4, the original problem can instead be solved directly through its exact convex reformulation.

4. EXACT CONVEX REFORMULATIONS FOR QUADRATIC TRANSPORT

For $p = 2$, additional Wasserstein geometry permits an exact convex reformulation in the small-regularization regime of Section 2.3.

We use either of the following geometric settings; their definitions are independent of the parameter regime.

- (i) *Univariate laws.* Let $\Xi = \mathbb{R}$ and let $\mathbb{P}_i^\pm \in \mathcal{P}_2(\mathbb{R})$ be arbitrary. Set $\mathcal{H} = L^2(0, 1)$, let $\mathcal{C}$ be the closed convex cone of functions admitting a nondecreasing representative, and let $z_i^\pm$ be the source quantile functions. A center coordinate $z \in \mathcal{C}$ represents $\text{Law}(z(U))$ for $U \sim \text{Unif}(0, 1)$.
- (ii) *Translations and scalings of a common base law.* Let $\Xi = \mathbb{R}^d$ and suppose

$$\mathbb{P}_i^\pm = \text{Law}(\mathbf{m}_i^\pm + a_i^\pm \mathbf{Z}), \quad a_i^\pm \geq 0, \quad \mathbb{E}\mathbf{Z} = 0, \quad \mathbb{E}\|\mathbf{Z}\|^2 = 1. \quad (4.1)$$

The same base law is used for all sources and need not be Gaussian. Set $\mathcal{H} = \mathbb{R}^{d+1}$, $\mathcal{C} = \mathbb{R}^d \times \mathbb{R}_+$, and $z_i^\pm = (\mathbf{m}_i^\pm, a_i^\pm)$. A center coordinate $z = (\mathbf{m}, a) \in \mathcal{C}$ represents $\text{Law}(\mathbf{m} + a\mathbf{Z})$.

Algorithm 1 Learning of the Inclusive-Exclusive Wasserstein Ambiguity Set

Input: Inclusive laws $\{\mathbb{P}_i^+\}_{i=1}^{N^+}$ and exclusive laws $\{\mathbb{P}_j^-\}_{j=1}^{N^-}$ in $\mathcal{P}_p(\Xi)$; transport order $p \geq 1$; parameters $\nu, b > 0$ and $\mu \geq 0$ satisfying (2.4) and (2.5); the exact oracle of Definition 2; dual tolerance $\varepsilon_D > 0$; maximum number of oracle calls $K \geq 1$.

Output: Best evaluated dual weights $\bar{\alpha}$; candidate center $\bar{\mathbb{Q}}$, radius $\bar{R}$, and margin $\bar{t}$; bounds L, U, B and termination status.

- 1: Use the source indexing, labels $\mathbf{y}$, and caps $\mathbf{c}$ of Section 3.1; choose $\alpha^0 \in \mathcal{A}$ with $\mathcal{A}$ given by (3.1).
- 2: Set $L = -\infty$, $U = B = +\infty$, and initialize an empty master cut set.
- 3: Set status to **iteration limit**.
- 4: **for** $r = 0, \dots, K - 1$ **do**
- 5: $\mathbb{Q}^r \leftarrow$ an element of $\text{SBar}_p(\{\mathbb{P}_k\}_{k=1}^N; \mathbf{y} \odot \alpha^r)$, returned by the oracle of Definition 2 according to (3.4).
- 6: Evaluate $d_k^r = \mathcal{W}_p^p(\mathbb{P}_k, \mathbb{Q}^r)$ for $k = 1, \dots, N$.
- 7: $v_r \leftarrow \Omega(\alpha^r)$, the oracle value $\sum_{k=1}^N y_k \alpha_k^r d_k^r$ in (3.5).
- 8: $\mathbf{g}^r \leftarrow (y_k d_k^r)_{k=1}^N$ by (3.6), evaluated at $(\alpha, \widehat{\mathbb{Q}}) = (\alpha^r, \mathbb{Q}^r)$.
- 9: **if** $v_r > L$ **then**
- 10: Update the dual lower bound $L = v_r$ and retain $\bar{\alpha} = \alpha^r$.
- 11: **end if**
- 12: Fix the center at $\mathbb{Q}^r$ and solve the linear program (2.3) using d_k^r , obtaining (ρ^r, t^r) and $J_r = F(\mathbb{Q}^r, \rho^r, t^r)$ as in (3.10).
- 13: **if** $J_r < B$ **then**
- 14: Update the primal upper bound $B = J_r$ and retain $(\bar{\mathbb{Q}}, \bar{\rho}, \bar{t}) = (\mathbb{Q}^r, \rho^r, t^r)$.
- 15: **end if**
- 16: Add the cut $\theta \leq \langle \mathbf{g}^r, \alpha \rangle$ to the master, retaining all previous cuts.
- 17: Solve (3.8); use its optimal weights as α^{r+1} and its value as the dual upper bound $U = U_r$.
- 18: **if** $U - L \leq \varepsilon_D$ **then**
- 19: Set status to **dual tolerance reached**; **break**.
- 20: **end if**
- 21: **end for**
- 22: Set $\bar{R} = \bar{\rho}^{1/p}$. Report $U - L$ as the dual error bound and $B - L$ as the candidate ball's suboptimality bound.
- 23: **return** $\bar{\alpha}$, $(\bar{\mathbb{Q}}, \bar{R}, \bar{t})$, (L, U, B) , and status.

In both settings, represented centers satisfy $\mathcal{W}_2^2(\mathbb{P}_i^\pm, \mathbb{Q}) = \|z_i^\pm - z\|_{\mathcal{H}}^2$.

Theorem 5 (Exact Convexification). *Let $p = 2$, $\nu, b > 0$, and suppose (2.4) and (2.5) hold. Use either geometric setting (i) or (ii) above, with its Hilbert space $\mathcal{H}$, closed convex set $\mathcal{C}$, and source coordinates $z_i^\pm$. In either setting, problem (2.3) has the same minimum as the convex quadratic optimization problem*

$$\begin{aligned}
 \min_{\substack{z \in \mathcal{C}, s \in \mathbb{R}, t \geq 0 \\ \omega^+, \omega^- \geq 0}} \quad & \nu \|z\|_{\mathcal{H}}^2 - \nu s - \mu t + \frac{1}{N} \sum_i \omega_i^+ + \frac{b}{N} \sum_j \omega_j^- \\
 \text{s.t.} \quad & \|z_i^+\|_{\mathcal{H}}^2 - 2\langle z_i^+, z \rangle_{\mathcal{H}} + s \leq \omega_i^+, \quad i = 1, \dots, N^+, \\
 & t - s - \|z_j^-\|_{\mathcal{H}}^2 + 2\langle z_j^-, z \rangle_{\mathcal{H}} \leq \omega_j^-, \quad j = 1, \dots, N^-.
 \end{aligned} \tag{4.2}$$

Every optimizer yields an original optimizer by setting

$$\rho^* = \|z^*\|_{\mathcal{H}}^2 - s^* \geq 0, \quad R^* = \sqrt{\rho^*}, \tag{4.3}$$

and retaining t^* and the slacks. In (i), the center is $\mathbb{P}_c^* = \text{Law}(z^*(U))$ with $U \sim \text{Unif}(0, 1)$. In (ii), writing $z^* = (\mathbf{m}^*, a^*)$, it is $\mathbb{P}_c^* = \text{Law}(\mathbf{m}^* + a^* \mathbf{Z})$. Conversely, every original optimizer has the corresponding center representation and solves (4.2) with $s^* = \|z^*\|_{\mathcal{H}}^2 - \rho^*$. The proof is given in Appendix A.5.

For empirical source laws in (i), (4.2) has an exact finite-dimensional formulation. Partition $(0, 1)$ at all source quantile breakpoints and replace z on each interval by its average. This preserves monotonicity and every $\langle z_i^\pm, z \rangle_{\mathcal{H}}$ while decreasing $\|z\|_{\mathcal{H}}^2$. The interval values therefore suffice, with linear monotonicity constraints and interval lengths as integration weights. In (ii), (4.2) is already finite-dimensional. In these finite-dimensional settings, the objective has a positive semidefinite Hessian and all constraints are linear, so a standard convex quadratic programming solver can jointly compute $(z^*, s^*, t^*, \omega^{+*}, \omega^{-*})$, for example by a primal–dual interior-point method. Feasibility residuals and the primal–dual gap provide numerical stopping criteria. The center is then reconstructed from z^* as above, and the radius follows from (4.3); no separate dual derivation or barycenter oracle is required. For general univariate source laws, numerical discretization of the quantile functions instead gives an approximation. In (ii), the common-base condition must hold for the source laws actually used in the optimization; independent empirical estimates need not preserve it even when the population laws satisfy it. If a common-base family is fitted first, the reformulation is exact for those fitted laws, not necessarily for the original empirical laws. These conclusions require the stated geometry; $p = 2$ alone does not justify this reformulation for arbitrary multivariate source laws.

5. CONCLUDING REMARKS

We have developed an inclusive-exclusive Wasserstein ball, learned from both inclusive and exclusive evidence when observations of the target distribution are unavailable. The model balances ball size, separation margin, and violations of the source labels. Its parameter analysis identifies an exact radius relaxation under small regularization, which yields a signed-barycenter dual and a cutting-plane learning algorithm with distinct dual and candidate-ball certificates. Under the same small-regularization condition, quadratic transport admits an exact convex reformulation for univariate laws and for common-base translation and scaling families, together with global optimizer recovery.

The general learning problem remains nonconvex, and the algorithm’s certificates depend on exact or certified inner solutions. Future work includes scalable signed-barycenter oracles with error bounds, extensions of the convex reformulation to broader multivariate families, and statistical calibration of the learned ball under explicit assumptions linking the sources to the unseen target.

REFERENCES

- Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- Jose Blanchet and Karthyek Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.
- Marco Cuturi and Arnaud Doucet. Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning*, pages 685–693. PMLR, 2014.
- Thomas O. Gallouët, Andrea Natale, and Gabriele Todeschi. Metric extrapolation in the Wasserstein space. *Calculus of Variations and Partial Differential Equations*, 64(5):147, 2025. doi: 10.1007/s00526-025-02996-6.
- Thomas O. Gallouët, Andrea Natale, and Gabriele Todeschi. Toland duality and particle approximations for signed Wasserstein barycenters. *arXiv preprint arXiv:2609.17073*, September 2026.

- Rui Gao and Anton J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655, 2023.
- Biao Han. Robust out-of-distribution stochastic optimization based on Wasserstein-metric meta-distribution support learning, October 2026. URL <https://optimization-online.org/?p=37187>.
- Matt Jacobs and Bohan Zhou. The signed Wasserstein barycenter problem. *arXiv preprint arXiv:2602.05976*, February 2026.
- Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*. INFORMS, 2019. doi: 10.1287/educ.2019.0198.
- Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. Distributionally robust optimization. *Acta Numerica*, 34:579–804, 2025.
- Xianyu Li, Huan Xu, Xiaolin Huang, and Chao Shang. Robust out-of-distribution stochastic optimization. *arXiv preprint arXiv:2604.20147*, April 2026.
- Hongying Liu, Hao Wang, Haoran Chu, and Yibo Wu. Towards convexity in anomaly detection: A new formulation of SSLM with unique optimal solutions. *Journal of Machine Learning Research*, 27(89):1–36, May 2026. ISSN 1533-7928.
- Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, 2018.
- Tingting Mu and Asoke K. Nandi. Multiclass Classification Based on Extended Support Vector Data Description. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, 39(5):1206–1216, October 2009. ISSN 1941-0492. doi: 10.1109/TSMCB.2009.2013962.
- Krikamol Muandet and Bernhard Schölkopf. One-class support measure machines for group anomaly detection. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 449–458, 2013.
- Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1-2):1–141, 2017.
- Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2021.
- David M.J. Tax and Robert P.W. Duin. Support vector data description. *Machine Learning*, 54(1): 45–66, January 2004. ISSN 1573-0565. doi: 10.1023/B:MACH.0000008084.60811.49.
- Francesco Tornabene, Marco Veneroni, and Giuseppe Savaré. Generalized Wasserstein barycenters. *Applied Mathematics & Optimization*, 92(3):68, 2025. doi: 10.1007/s00245-025-10351-6.
- Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.
- Yuntao Wang, Yuxuan Li, Qingyuan Yang, and Hu Ding. Finding Wasserstein ball center: Efficient algorithm and the applications in fairness. In *Proceedings of the Forty-second International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 63490–63515. PMLR, 2025. URL <https://proceedings.mlr.press/v267/wang25be.html>.
- Richard Wilhelm. *The I Ching, or, Book of Changes*, volume 19 of *Bollingen Series*. Princeton University Press, Princeton, NJ, 3rd edition, 1967. Rendered into English by Cary F. Baynes; foreword by C. G. Jung; preface to the third edition by Hellmut Wilhelm.
- Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.

APPENDIX A. PROOFS

A.1. Proof of Lemma 1. Write $F(\mathbb{Q}, \rho, t)$ for the objective in (2.2). If $\mu > bN^-/N$, fix any $\mathbb{Q} \in \mathcal{P}_p(\Xi)$ and set $\rho = 0$. For all sufficiently large t , every exclusive hinge is active, giving

$$F(\mathbb{Q}, 0, t) = \left(\frac{bN^-}{N} - \mu \right) t + \frac{1}{N} \sum_i \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}) - \frac{b}{N} \sum_j \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{Q}) \rightarrow -\infty.$$

This argument does not require Ξ to be unbounded.

If $\mu > N^+/N$ and Ξ is unbounded, choose $\xi_k \in \Xi$ with $\|\xi_k\| \rightarrow \infty$ and set $\mathbb{Q}_k = \delta_{\xi_k}$, $\rho_k = 0$, and $t_k = \|\xi_k\|^p$. For any source law $\mathbb{P}$, the reverse triangle inequality in L^p gives

$$|\mathcal{W}_p(\mathbb{P}, \mathbb{Q}_k) - \|\xi_k\|| \leq \left(\int_{\Xi} \|\xi\|^p d\mathbb{P}(\xi) \right)^{1/p}.$$

Consequently, $\mathcal{W}_p^p(\mathbb{P}, \mathbb{Q}_k)/t_k \rightarrow 1$ for every source, and

$$\frac{F(\mathbb{Q}_k, 0, t_k)}{t_k} = -\mu + \frac{1}{N} \sum_i \frac{\mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}_k)}{t_k} + \frac{b}{N} \sum_j \left(1 - \frac{\mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{Q}_k)}{t_k} \right)_+ \rightarrow \frac{N^+}{N} - \mu < 0.$$

Thus $F(\mathbb{Q}_k, 0, t_k) \rightarrow -\infty$. The same conclusions hold for (2.3) by taking the slacks equal to their hinge losses.

A.2. Proof of Theorem 2.

Proof. If $\rho < 0$, keep $\mathbb{Q}$ fixed and set $\rho' = 0$ and $t' = (\rho + t)_+$. Nonnegativity of all transport costs gives

$$F(\mathbb{Q}, \rho, t) - F(\mathbb{Q}, 0, t') = \left(\nu + \mu - \frac{N^+}{N} \right) \rho - \mu \min\{\rho + t, 0\} > 0.$$

Thus every relaxed point with negative radius power is strictly dominated by an original feasible point. The infima are equal, and the relaxed problem has no global minimizer with $\rho < 0$. Equality of the global minimizer sets follows immediately. Minimizing over the nonnegative slacks recovers F , giving the same conclusion for the slack formulation. $\square$

A.3. Proof of Theorem 3.

Proof. For (i), direct expansion using $(a)_+ = a + (-a)_+$ gives, for $\rho, t \geq 0$,

$$\begin{aligned} F(\mathbb{Q}, \rho, t) - F(\mathbb{Q}, 0, \rho + t) &= (\nu + \mu - \kappa)\rho \\ &\quad + \frac{1}{N} \sum_i (\rho - \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}))_+ \geq 0. \end{aligned} \tag{A.1}$$

Restricting to $\rho = 0$ therefore preserves the infimum and gives the stated optimizer correspondences. If $\nu + \mu > \kappa$ and $\rho > 0$, the inequality is strict, so no original optimizer has positive radius.

For (ii), $\mu = 0$ implies $\nu \geq \kappa$ and

$$F(\mathbb{Q}, \rho, t) \geq (\nu - \kappa)\rho + \frac{1}{N} \sum_i \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}) \geq \inf_{\mathbb{Q}'} \frac{1}{N} \sum_i \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}').$$

Each barycenter with $\rho = t = 0$ attains this lower bound. Equality throughout holds precisely when the center is a barycenter, the positive hinges equal their arguments, every negative hinge vanishes, and $(\nu - \kappa)\rho = 0$. These are exactly (2.9). $\square$

A.4. Proof of Theorem 4.

Proof. The primal to be dualized. By Theorem 2, (2.5) makes (2.2) share its infimum and global minimizers with the radius-relaxed problem (2.6), in which ρ ranges over all of $\mathbb{R}$; the same theorem relaxes the radius constraint of the slack form (2.3). We dualize that relaxed slack form: the variables are $\mathbb{Q} \in \mathcal{P}_p(\Xi)$, $\rho \in \mathbb{R}$, $t \geq 0$ and $\omega^\pm \geq 0$, the objective is the one in (2.3), and the constraints are its inclusive and exclusive transport constraints. Eliminating the slacks through $(z)_+ = \inf\{\omega \geq 0 \mid z \leq \omega\}$ returns (2.6), so the two carry the same value.

The Lagrangian. Attach a multiplier $\alpha_i^+ \geq 0$ to the inclusive constraint indexed by i and $\alpha_j^- \geq 0$ to the exclusive constraint indexed by j ; the restrictions $t \geq 0$ and $\omega^\pm \geq 0$ are retained rather than dualized. The Lagrangian is

$$\begin{aligned} \mathcal{L} = & \nu\rho - \mu t + \frac{1}{N} \sum_i \omega_i^+ + \frac{b}{N} \sum_j \omega_j^- \\ & + \sum_i \alpha_i^+ (\mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}) - \rho - \omega_i^+) \\ & + \sum_j \alpha_j^- (\rho + t - \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{Q}) - \omega_j^-), \end{aligned}$$

and collecting the terms in ρ , t , and the slack vectors gives

$$\begin{aligned} \mathcal{L} = & \sum_i \alpha_i^+ \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}) - \sum_j \alpha_j^- \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{Q}) \\ & + \left(\nu - \sum_i \alpha_i^+ + \sum_j \alpha_j^- \right) \rho + \left(\sum_j \alpha_j^- - \mu \right) t \\ & + \sum_i \left(\frac{1}{N} - \alpha_i^+ \right) \omega_i^+ + \sum_j \left(\frac{b}{N} - \alpha_j^- \right) \omega_j^-. \end{aligned}$$

The dual feasible set. A variable confined to the half-line $[0, \infty)$ leaves the inner infimum finite exactly when its coefficient is nonnegative; for the free variable ρ a nonzero coefficient lets $\rho \rightarrow \pm\infty$ drive $\mathcal{L} \rightarrow -\infty$, so its coefficient must vanish. The conditions are therefore

$$\alpha_i^+ \leq \frac{1}{N}, \quad \alpha_j^- \leq \frac{b}{N}, \quad \sum_j \alpha_j^- \geq \mu, \quad \sum_i \alpha_i^+ - \sum_j \alpha_j^- = \nu,$$

which together with $\alpha^+, \alpha^- \geq 0$ is exactly the set $\mathcal{A}$ of (3.1). The freedom of ρ is what makes the last condition an equality; had $\rho \geq 0$ been retained instead, it would read $\sum_i \alpha_i^+ - \sum_j \alpha_j^- \leq \nu$. Multipliers outside $\mathcal{A}$ send the inner infimum to $-\infty$ and can never attain the supremum, so the supremum in (3.2) is taken over $\mathcal{A}$ alone.

The dual function. For multipliers in $\mathcal{A}$ the coefficient of ρ vanishes identically and the remaining coefficients are nonnegative, so the infimum over ρ , t and the slacks is attained at $\rho = t = \omega^+ = \omega^- = 0$, where their contribution vanishes. What remains is the dual function Ω of (3.5), read in the split notation,

$$\Omega(\alpha^+, \alpha^-) = \inf_{\mathbb{Q} \in \mathcal{P}_p(\Xi)} \left\{ \sum_i \alpha_i^+ \mathcal{W}_p^p(\mathbb{P}_i^+, \mathbb{Q}) - \sum_j \alpha_j^- \mathcal{W}_p^p(\mathbb{P}_j^-, \mathbb{Q}) \right\}.$$

Infimizing $\mathcal{L}$ over $\mathbb{Q} \in \mathcal{P}_p(\Xi)$, $\rho \in \mathbb{R}$, $t \geq 0$, and $\omega^\pm \geq 0$, without imposing the dualized transport constraints, thus returns Ω on $\mathcal{A}$ and $-\infty$ elsewhere. Taking the supremum over the nonnegative multipliers gives (3.2). $\square$

A.5. Proof of Theorem 5.

Proof. Theorem 2 first allows $\rho \in \mathbb{R}$. In (i), the quantile identity gives $\mathcal{W}_2^2(\mathbb{P}_i^\pm, \mathbb{Q}) = \|z_i^\pm - z\|_{\mathcal{H}}^2$ for every center with quantile z .

In (ii), we must first justify restricting the center to the same family. Write an arbitrary center as $\mathbb{Q} = \text{Law}(\mathbf{m} + \mathbf{Y})$ with $\mathbb{E}\mathbf{Y} = 0$, and define

$$a := \sup_{\pi \in \Pi(\text{Law}(\mathbf{Z}), \text{Law}(\mathbf{Y}))} \mathbb{E}_\pi[\mathbf{Z}^\top \mathbf{Y}], \quad e := \mathbb{E}\|\mathbf{Y}\|^2 - a^2.$$

Independence and Cauchy–Schwarz imply $a \geq 0$ and $e \geq 0$. Expanding the transport cost, using $a_i^\pm \geq 0$, yields

$$\mathcal{W}_2^2(\mathbb{P}_i^\pm, \mathbb{Q}) = \|\mathbf{m}_i^\pm - \mathbf{m}\|^2 + (a_i^\pm - a)^2 + e.$$

For $\mathbb{Q}' = \text{Law}(\mathbf{m} + a\mathbf{Z})$, the same expansion gives $\mathcal{W}_2^2(\mathbb{Q}, \mathbb{Q}') = e$, and its source-to-center costs are $\|z_i^\pm - z\|_{\mathcal{H}}^2$. Replacing $(\mathbb{Q}, \rho)$ by $(\mathbb{Q}', \rho - e)$ preserves every hinge argument, so

$$F(\mathbb{Q}', \rho - e, t) = F(\mathbb{Q}, \rho, t) - \nu e.$$

Thus restricting the radius-relaxed problem to this family preserves its infimum. Since $\nu > 0$, every original optimizer, which also solves the relaxed problem, must have $e = 0$ and hence $\mathbb{Q} = \mathbb{Q}'$.

In both cases, set

$$s = \|z\|_{\mathcal{H}}^2 - \rho.$$

Expanding the squared distances gives (4.2). Its objective is convex and its displayed constraints are affine over the convex domain $\mathcal{C}$. Theorem 2 ensures that any optimizer has the nonnegative recovered radius in (4.3).

For completeness, the minimum is attained even though Ξ is unbounded. Choose weights within $0 \leq \alpha_i^+ \leq 1/N$ and $0 \leq \alpha_j^- \leq b/N$ whose sums are $\nu + \mu$ and μ , respectively; (2.5) and (2.4) guarantee their existence. Weighting the constraints shows that the objective is bounded below by $\nu\|z\|_{\mathcal{H}}^2 - 2\langle h, z \rangle_{\mathcal{H}} + c$, where $h = \sum_i \alpha_i^+ z_i^+ - \sum_j \alpha_j^- z_j^-$ and $c = \sum_i \alpha_i^+ \|z_i^+\|_{\mathcal{H}}^2 - \sum_j \alpha_j^- \|z_j^-\|_{\mathcal{H}}^2$. Thus a minimizing sequence has bounded z and uniformly bounded source-to-center costs, say by D . If $\rho < 0$, replace (ρ, t) by $(0, (\rho + t)_+)$ as in Theorem 2. For fixed center and $\rho, t \geq 0$, F has slope $\nu + bN^-/N > 0$ in ρ when $\rho > D$, and slope $bN^-/N - \mu \geq 0$ in t when $t > D$. Hence clipping ρ and t to $[0, D]$ cannot increase the objective. Taking the smallest feasible slacks also bounds them, and therefore s . A weakly convergent subsequence now exists; the closed convex feasible set and weak lower semicontinuity of the objective yield an optimizer. $\square$

Email address: hanb16@tsinghua.org.cn