Price-aware transport subspaces for Wasserstein DRO

Martin Betz[a,*], Heinrich Daßer[b], Florian Roesel[a]

[a]Department of Mathematics, Friedrich-Alexander-Universität Erlangen–Nürnberg, Nürnberger Straße 74, 91052 Erlangen, Germany
[b]Computational Science and Mathematical Methods, Scientific Computing Center, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Abstract

We quantify the robust value lost by restricting Wasserstein transport to a common low-dimensional subspace. Affine transport sets and outer moment sets yield exact bounds for quadratic losses with nonsingular nominal covariance. A spectral rule augments a prescribed base space and attains a sharp, base-dependent approximation factor for the aggregate first-order restriction loss. Eigenvalue tails yield dimension and finite-radius decision certificates. Gradient principal component analysis admits no bounded approximation factor. Exact examples and synthetic reduced decision problems illustrate the guarantees.

Keywords: Distributionally robust optimization, ambiguity-set inclusion, Wasserstein distance, subspace approximation, decision regret

1. Introduction

For a loss $\ell_x : \mathbb{R}^d \to \mathbb{R}$ and a nominal law Q_0, consider

$$\Phi_\varepsilon(x) = \sup_{Q\in\mathcal{B}_\varepsilon} \mathbb{E}_Q \ell_x, \qquad \mathcal{B}_\varepsilon = \{Q \in \mathcal{P}_2(\mathbb{R}^d) : W_2(Q, Q_0) \le \varepsilon\}, \tag{1}$$

where W_2 uses the Euclidean ground norm and $\mathcal{P}_2$ denotes laws with finite second moment. Restricting admissible transports gives a lower approximation, while retaining only mean and covariance gives a Gelbrich outer approximation [1, 2]. The associated objective gaps measure the accuracy of these inner and outer approximations.

This letter connects those inclusions to the design of a common reduced ambiguity set. Such a set describes the movements of the nominal observations by a few transport directions shared by a finite list of candidate decisions. The question is how much robust value this restriction costs and how to choose the directions. For a space that does not change with the decision, the restricted problem has the same form as the robust problem with only a few transport directions, and its value bounds the robust optimum from below. We first distinguish exactness for quadratic losses from a persistent outer-relaxation gap for general smooth losses. This contrast motivates approximating from the inside, since the gap of a transport restriction vanishes with the radius and has a computable first-order price. We then ask how to add m transport directions to a prescribed base space so as to preserve worst-case values across a finite list of decisions. The main result is a sharp, base-dependent approximation factor $2/(1 + \rho)$ for the lost first-order value. The base may retain translations, affine fields, or other movements required by the specific problem setting. The guarantee improves as the base captures more of every gradient. Since ρ is a minimum over the candidates, a single poorly represented candidate can lead to a factor close to two. The resulting value brackets also screen a finite set of decisions and, when a global lower bound is available, bound regret over the full feasible decision set.

The Wasserstein-ball first-order expansion is established in [3, 4]. Namkoong [5, Proposition 2] expresses sensitivity under a prescribed transport restriction through the projected gradient. Our local expansion uses the same Taylor argument, with an explicit finite-radius remainder. Our contribution is to choose additional transport directions jointly for several decisions and to obtain a sharp approximation factor that accounts for the prescribed base space. The spectral construction also provides an initialization for direct objective refinement. Its relation to R_1-PCA is discussed after Theorem 6. PCA-based DRO reductions also appear in moment and mixed moment–Wasserstein models [6, 7]. Our reduced coordinates parameterize displacement fields over nominal observations. We compare the spectral rule with ordinary gradient PCA on the same field space and budget.

2. Subset bounds: exactness and a persistent gap

Before designing a reduced set, we ask from which side it should approximate the full ball. This section shows that outer moment sets are exact only in special cases and can leave a gap that persists as the radius tends to zero. Section 3 shows that the gap of an inner transport restriction vanishes with the radius.

The next two propositions recall the set inclusions and quadratic exactness results that follow from classical coupling bounds and the Gelbrich–Bures affine construction [1, 2]. Fix $Q_0 = \sum_{j=1}^k q_j \delta_{z_j}$, with distinct $z_j \in \mathbb{R}^d$, $q_j > 0$, and $\sum_j q_j = 1$. Let μ_0 and Σ_0 be its mean and covariance. Fields $v = (v_j)_{j=1}^k$ form the Hilbert space $L^2(Q_0) = (\mathbb{R}^d)^k$ with

$$\langle u, v\rangle = \sum_j q_j u_j^\top v_j, \qquad \|v\|^2 = \sum_j q_j \|v_j\|_2^2.$$

*Corresponding author.
Email address: martin.betz@fau.de (Martin Betz)

Its transported law is $Q_v = \sum_j q_j \delta_{z_j + v_j}$.

Definition 1 (Inner and outer sets)**.** For a linear field space $V \subseteq L^2(Q_0)$, put

$$\mathcal{B}_\varepsilon^V = \{Q_v : v \in V,\ \|v\| \le \varepsilon\}, \quad \Phi_\varepsilon^V(x) = \sup_{Q \in \mathcal{B}_\varepsilon^V} \mathbb{E}_Q \ell_x.$$

The translation and affine field spaces are

$$V_{\mathrm{tr}} = \{(t)_j : t \in \mathbb{R}^d\},$$
$$V_{\mathrm{aff}} = \{(t + B(z_j - \mu_0))_j : t \in \mathbb{R}^d,\ B \in \mathbb{R}^{d \times d}\}.$$

For $Q \in \mathcal{P}_2$ with mean μ and covariance Σ, define

$$G(Q)^2 = \|\mu - \mu_0\|_2^2 + \mathrm{tr}(\Sigma + \Sigma_0 - 2(\Sigma_0^{1/2}\Sigma\Sigma_0^{1/2})^{1/2}),$$
$$\mathcal{G}_\varepsilon = \{Q \in \mathcal{P}_2 : G(Q) \le \varepsilon\},$$
$$\mathcal{I}_\varepsilon = \{Q \in \mathcal{P}_2 : \|\mu - \mu_0\|_2^2 + \mathrm{tr}(\Sigma + \Sigma_0) \le \varepsilon^2\}.$$

Write $\Phi_\varepsilon^{\mathrm{G}}(x) = \sup_{Q \in \mathcal{G}_\varepsilon} \mathbb{E}_Q \ell_x$. We call $\mathcal{B}_\varepsilon^V$ a transport inner set, $\mathcal{G}_\varepsilon$ the Gelbrich outer set and $\mathcal{I}_\varepsilon$ the moment inner set. For a single loss we drop the index x, and for a candidate x_i we write $\Phi_\varepsilon^V(x_i)$ and $\Phi_\varepsilon^{\mathrm{G}}(x_i)$.

Proposition 2 (The subset chain)**.** *For $V_{\mathrm{tr}} \subseteq V \subseteq L^2(Q_0)$,*

$$\mathcal{B}_\varepsilon^{V_{\mathrm{tr}}} \subseteq \mathcal{B}_\varepsilon^V \subseteq \mathcal{B}_\varepsilon^{L^2(Q_0)} \subseteq \mathcal{B}_\varepsilon \subseteq \mathcal{G}_\varepsilon, \tag{2}$$

and $\mathcal{I}_\varepsilon \subseteq \mathcal{B}_\varepsilon$. The moment inner set is empty precisely when $\varepsilon^2 < \mathrm{tr}\,\Sigma_0$. For $\ell(z) = c + a^\top z$,

$$\Phi_\varepsilon^{V_{\mathrm{tr}}} = \Phi_\varepsilon = \Phi_\varepsilon^{\mathrm{G}} = a^\top \mu_0 + c + \varepsilon \|a\|_2. \tag{3}$$

Proof. The natural and independent couplings, respectively, give

$$W_2(Q_v, Q_0)^2 \le \sum_j q_j \|v_j\|_2^2,$$
$$W_2(Q, Q_0)^2 \le \|\mu - \mu_0\|_2^2 + \mathrm{tr}(\Sigma + \Sigma_0).$$

The Gelbrich inequality [1] gives $G(Q) \le W_2(Q, Q_0)$. Since $G(Q) \ge \|\mu - \mu_0\|_2$,

$$\sup_{Q \in \mathcal{G}_\varepsilon} (a^\top \mu + c) \le a^\top \mu_0 + c + \varepsilon \|a\|_2 = \mathbb{E}_{Q_v} \ell, \quad v_j = \varepsilon a / \|a\|_2.$$

For $a = 0$ take $v = 0$. Nonemptiness of $\mathcal{I}_\varepsilon$ follows by testing $Q = \delta_{\mu_0}$, necessity follows from nonnegativity of its two variable terms. □

Every transport inner set contains Q_0 at every radius. The Bures affine map [1, 2] further yields exact quadratic values by matching feasible first and second moments.

Proposition 3 (Closure for quadratic losses)**.** *If $\Sigma_0 \succ 0$, then every feasible moment pair of $\mathcal{G}_\varepsilon$ is attained by a law in $\mathcal{B}_\varepsilon^{V_{\mathrm{aff}}}$. Consequently, for every symmetric H and $\ell(z) = c + a^\top z + \frac{1}{2} z^\top H z$,*

$$\Phi_\varepsilon^{V_{\mathrm{aff}}} = \Phi_\varepsilon = \Phi_\varepsilon^{\mathrm{G}}. \tag{4}$$

Proof. For a feasible pair (μ, Σ), set

$$A = \Sigma_0^{-1/2}(\Sigma_0^{1/2}\Sigma\Sigma_0^{1/2})^{1/2}\Sigma_0^{-1/2}, \quad v_j = \mu - \mu_0 + (A - I)(z_j - \mu_0).$$

Then $A\Sigma_0 A = \Sigma$, and

$$\begin{aligned} \mathbb{E}_{Q_v} z &= \mu, \qquad \mathrm{Cov}_{Q_v}(z) = \Sigma, \\ \|v\|^2 &= \|\mu - \mu_0\|_2^2 + \mathrm{tr}((A - I)\Sigma_0(A - I)) \\ &= G(Q)^2 \le \varepsilon^2, \\ \mathbb{E}_{Q_v} \ell &= c + a^\top \mu + \tfrac{1}{2}\mu^\top H \mu + \tfrac{1}{2}\,\mathrm{tr}(H\Sigma) = \mathbb{E}_Q \ell. \end{aligned}$$

Thus $\Phi_\varepsilon^{\mathrm{G}} \le \Phi_\varepsilon^{V_{\mathrm{aff}}} \le \Phi_\varepsilon \le \Phi_\varepsilon^{\mathrm{G}}$. □

For a nonquadratic loss, the moment outer set can have a positive value gap already at radius zero. Consider the smooth convex loss and nominal law

$$\ell(z) = \sqrt{1 + z^2}, \qquad Q_0 = \tfrac{1}{2}\delta_0 + \tfrac{1}{4}\delta_{\sqrt{2}} + \tfrac{1}{4}\delta_{-\sqrt{2}}. \tag{5}$$

Here $\mu_0 = 0$, $\Sigma_0 = 1$, and at radius zero, with the radius as subscript,

$$\Phi_0 = (1 + \sqrt{3})/2, \qquad \Phi_0^{\mathrm{G}} = \sqrt{2}, \tag{6}$$
$$\Phi_0^{\mathrm{G}} - \Phi_0 = \sqrt{2} - (1 + \sqrt{3})/2 > 0.$$

Indeed $\mathbb{E}_Q \sqrt{1 + z^2} \le \sqrt{1 + \mathbb{E}_Q z^2} = \sqrt{2}$ for all $Q \in \mathcal{G}_0$, with equality at $(\delta_{-1} + \delta_1)/2$. This positive gap persists at small positive radii. Outer moment bounds can therefore remain loose at every small radius. The accuracy of transport inner approximations is governed by the loss gradients.

3. From subset inclusion to lost robust value

This section quantifies the cost of a fixed transport restriction. Up to a curvature remainder of order ε^2, the lost value equals ε times a gradient-projection price, which becomes the design objective in Section 4.

Write $g_x = (\nabla \ell_x(z_j))_j$, $\sigma_x = \|g_x\|$, and P_V for orthogonal projection in $L^2(Q_0)$. Assume throughout that $\nabla \ell_x$ is globally M_x-Lipschitz. For twice continuously differentiable losses, M_x can be obtained from a global Hessian bound $M_x \ge \sup_z \|\nabla^2 \ell_x(z)\|_{\mathrm{op}}$. For a quadratic loss with Hessian H_x, this gives $M_x = \|H_x\|_{\mathrm{op}}$. Evaluating curvature only at the nominal observations does not establish such a bound.

Proposition 4 (Finite-radius gap)**.** *With $\Phi_0(x) = \mathbb{E}_{Q_0} \ell_x$, for every V and $\varepsilon \ge 0$,*

$$|\Phi_\varepsilon(x) - \Phi_0(x) - \varepsilon\sigma_x| \le M_x \varepsilon^2/2, \tag{7}$$
$$|\Phi_\varepsilon^V(x) - \Phi_0(x) - \varepsilon\|P_V g_x\|| \le M_x \varepsilon^2/2. \tag{8}$$

In particular, with $\delta_x(V) = \sigma_x - \|P_V g_x\|$ and $\Delta_{\varepsilon,x}(V) = \Phi_\varepsilon(x) - \Phi_\varepsilon^V(x)$,

$$\left|\Delta_{\varepsilon,x}(V) - \varepsilon\delta_x(V)\right| \le M_x \varepsilon^2. \tag{9}$$

Proof. Taylor's inequality and its expectation give

$$|\ell_x(z+u) - \ell_x(z) - \nabla\ell_x(z)^\top u| \le \tfrac{1}{2}M_x\|u\|_2^2,$$
$$|\mathbb{E}_{Q_v}\ell_x - \Phi_0(x) - \langle g_x, v\rangle| \le \tfrac{1}{2}M_x\|v\|^2,$$
$$\sup_{v\in V, \|v\|\le\varepsilon} \langle g_x, v\rangle = \varepsilon\|P_V g_x\|.$$

The last supremum is attained by $v = \varepsilon P_V g_x/\|P_V g_x\|$, or $v = 0$ if the projection vanishes, proving (8). For an optimal coupling (Z, Y) of $Q_0, Q \in \mathcal{B}_\varepsilon$,

$$\begin{aligned}\mathbb{E}_Q\ell_x - \Phi_0(x) &\le \mathbb{E}[\nabla\ell_x(Z)^\top(Y-Z)] + \tfrac{1}{2}M_x\mathbb{E}\|Y-Z\|_2^2\\ &\le \sigma_x(\mathbb{E}\|Y-Z\|_2^2)^{1/2} + \tfrac{1}{2}M_x\mathbb{E}\|Y-Z\|_2^2\\ &\le \varepsilon\sigma_x + \tfrac{1}{2}M_x\varepsilon^2.\end{aligned}$$

The reverse bound uses the feasible field $v = \varepsilon g_x/\sigma_x$ (zero if $\sigma_x = 0$). Subtracting (7) and (8) proves (9). □

The proof also explains why local bounds cannot replace the global assumption in the full-ball estimate. For deterministic transports, $\|v\| \le \varepsilon$ implies $\|v_j\|_2 \le \varepsilon/\sqrt{q_j}$, so bounds on the balls $B(z_j, \varepsilon/\sqrt{q_j})$ suffice for (8). A Wasserstein coupling can instead split the mass at z_j and send a small part arbitrarily far. Bounds on fixed neighborhoods therefore do not control the Taylor remainder in (7).

For (5), $M = 1$, $\sigma = 1/\sqrt{3}$. Since $\mathcal{G}_0 \subseteq \mathcal{G}_\varepsilon$, the persistent-gap statement also has an explicit positive-radius bound:

$$\Phi_\varepsilon^{\mathrm{G}} - \Phi_\varepsilon \ge \sqrt{2} - (1+\sqrt{3})/2 - \varepsilon/\sqrt{3} - \varepsilon^2/2.$$

Equation (9) gives a vanishing error for every fixed transport restriction and identifies its first-order price.

4. A sharp bound on the price of approximation

We now choose the restriction. The problem setting prescribes a base space of mandatory movements, such as translations, and a budget of additional directions. We choose these directions to approximately minimize the aggregate price over a finite list of candidates, with a sharp guarantee, and derive decision certificates from the same spectral quantities.

Let $x_1, \dots, x_N$ be fixed candidates with weights $\omega_i > 0$, $\sum_i \omega_i = 1$. Put $g_i = g_{x_i}$, $\sigma_i = \|g_i\|$, and $\overline{M} = \sum_i \omega_i M_i$. Fix a base space S with $V_{\mathrm{tr}} \subseteq S \subseteq L^2(Q_0)$. Admissible spaces have the form

$$V = S \oplus W, \qquad W \subseteq S^\perp, \quad \dim W \le m.$$

Thus $\mathcal{B}_\varepsilon^S \subseteq \mathcal{B}_\varepsilon^V \subseteq \mathcal{B}_\varepsilon$. The base encodes common translations or affine movements. The budget m counts *additional* displacement patterns over the observations. The construction below uses only the gradients g_i. The curvature constants M_i enter when we translate its first-order guarantee into finite-radius value bounds.

Definition 5 (Restriction prices and spectral rule)**.** Set $\beta_i = \|P_S g_i\|$, $h_i = (I - P_S)g_i$, and

$$\delta(V) = \sum_i \omega_i(\sigma_i - \|P_V g_i\|),$$
$$\Delta_\varepsilon(V) = \sum_i \omega_i(\Phi_\varepsilon(x_i) - \Phi_\varepsilon^V(x_i)).$$

Identify fields with $\widehat{v} = (\sqrt{q_j}v_j)_j \in \mathbb{R}^{dk}$. Define

$$C = \sum_{i:h_i\neq 0} \frac{\omega_i}{\sigma_i+\beta_i}\widehat{h_i}\widehat{h_i}^\top, \qquad \rho = \min_{i:h_i\neq 0}\frac{\beta_i}{\sigma_i}. \tag{10}$$

Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ be its eigenvalues, $r_m = \sum_{j>m}\lambda_j$, and $V_m = S \oplus W_m$, where W_m spans up to m leading positive-eigenvalue directions. If all $h_i = 0$, set $C = 0, \rho = 1, W_m = \{0\}$.

Theorem 6 (Sharp base-dependent approximation)**.** *Let $\alpha = 2/(1+\rho)$. With infima over admissible V,*

$$\frac{r_m}{\alpha} \le \inf_V \delta(V) \le \delta(V_m) \le r_m, \tag{11}$$
$$\delta(V_m) \le \alpha \inf_V \delta(V), \tag{12}$$
$$\Delta_\varepsilon(V_m) \le \alpha \inf_V \Delta_\varepsilon(V) + (1+\alpha)\overline{M}\varepsilon^2. \tag{13}$$

For every fixed $0 \le \rho < 1$, the factor α is sharp, already for convex quadratic losses with translations as the base.

Proof. For $h_i \neq 0$, Pythagoras and rationalizing the difference give

$$\sigma_i - \|P_V g_i\| = \frac{\|(I-P_W)h_i\|^2}{\sigma_i + \|P_V g_i\|}, \tag{14}$$
$$\beta_i \le \|P_V g_i\| \le \sigma_i \implies \frac{\sigma_i+\beta_i}{2\sigma_i} \le \frac{\sigma_i+\beta_i}{\sigma_i+\|P_V g_i\|} \le 1.$$

Consequently, with $\overline{\delta}(V) = \mathrm{tr}((I-P_W)C)$,

$$\alpha^{-1}\overline{\delta}(V) \le \delta(V) \le \overline{\delta}(V), \qquad \inf_V \overline{\delta}(V) = \overline{\delta}(V_m) = r_m. \tag{15}$$

The second equality is Ky Fan's maximum principle [8]. Thus $\frac{r_m}{\alpha} \le \inf_V \delta(V) \le \delta(V_m) \le r_m$ and $\delta(V_m) \le \overline{\delta}(V_m) \le \overline{\delta}(V) \le \alpha\delta(V)$. Furthermore, (9) implies uniformly in V that $|\Delta_\varepsilon(V) - \varepsilon\delta(V)| \le \overline{M}\varepsilon^2$, so for every admissible V,

$$\begin{aligned}\Delta_\varepsilon(V_m) &\le \varepsilon\delta(V_m) + \overline{M}\varepsilon^2\\ &\le \alpha\varepsilon\delta(V) + \overline{M}\varepsilon^2\\ &\le \alpha\Delta_\varepsilon(V) + (1+\alpha)\overline{M}\varepsilon^2.\end{aligned}$$

For sharpness fix $d \ge 2$, take Q_0 uniform on $\{-1,1\}^d$, equal candidate weights, $S = V_{\mathrm{tr}}$, $m = d-1$, and

$$\ell_i(z) = s_i(\rho z_i + \tfrac{1}{2}\sqrt{1-\rho^2}\,z_i^2), \quad s_i = 1\ (i<d), \quad s_d = 1 - 1/d.$$

Then $\sigma_i = s_i, \beta_i = \rho s_i$, and the normalized centered fields are orthonormal. The eigenvalues of C are $(1-\rho)s_i/d$, so

$$\delta(V_m) = \frac{(1-\rho)(1-1/d)}{d}.$$

Let W be the hyperplane in their span perpendicular to their normalized sum. Each centered field retains the fraction $1-1/d$ of its squared norm, hence

$$\delta(S\oplus W) = (1-d^{-2})\left(1 - \sqrt{1-(1-\rho^2)/d}\right).$$

Therefore

$$\frac{\delta(V_m)}{\inf_V \delta(V)} \geq \frac{\delta(V_m)}{\delta(S \oplus W)} \longrightarrow \frac{2}{1+\rho} = \alpha.$$

□

The normalization in (10) is the smallest coefficient giving a separable quadratic upper bound on each candidate's price for every added space. Indeed (14) at $W = \{0\}$ gives $\frac{\sigma_i - \beta_i}{\|h_i\|^2} = \frac{1}{\sigma_i + \beta_i}$, and (14) proves validity for every W. This normalization accounts for the gradient already represented in the base. The factor α equals two at $\rho = 0$ and tends to one as $\rho \uparrow 1$. Since ρ is a minimum over the candidates, translations alone can leave α close to two. On the balanced instances of Table 3, the mean factor is 1.812 with translations, and 1.549 and 1.275 for fields affine in one and three coordinates. The usefulness of (13) depends on the size of the curvature remainder relative to the leading restriction price. For $r_m > 0$, their balance is measured by $\theta_\varepsilon = \overline{M}\varepsilon / r_m$. The bound $\inf_V \Delta_\varepsilon(V) \geq \varepsilon r_m(1/\alpha - \theta_\varepsilon)$ is positive only when $\theta_\varepsilon < 1/\alpha$. To obtain a finite-radius comparison close to the first-order factor α, this ratio must be small relative to $1/\alpha$. The finite-radius guarantee therefore does not imply that the spectral rule outperforms PCA or direct refinement at every radius. Figure 1(a) places S and V_m in the subset chain. Panels (b,c) illustrate the continuous-decision example and its finite set of decisions in Section 6.

Direct refinement and the relation to R_1-PCA.

The spectral space is a starting point. A local refinement can lower the price further while keeping the guarantee. In an orthonormal basis of $\operatorname{span}\{h_i\}$, write the added space as a matrix U with $U^\top U = I$ and define

$$f(U) = \sum_i \omega_i \sqrt{\beta_i^2 + \|U^\top h_i\|_2^2},$$
$$\delta(S \oplus \operatorname{range} U) = \sum_i \omega_i \sigma_i - f(U).$$

Ding et al. [9] introduced rotationally invariant residual-based R_1-PCA. When the base projections vanish, maximizing f is precisely the projected-norm variant in [10, Eq. (1.3)]. A retention guarantee with factor $\kappa < 1$ yields only

$$\delta(V_{\text{alg}}) \leq (1-\kappa) \sum_i \omega_i \sigma_i + \kappa \inf_V \delta(V),$$

which leaves an additive term even when the optimal lost value tends to zero. Theorem 6 gives a sharp multiplicative guarantee for that *loss*, incorporating the mandatory base. A standard local polar refinement directly increases the retained-norm objective:

$$p_i = \sqrt{\beta_i^2 + \|U^\top h_i\|_2^2}, \qquad D = \sum_{i:p_i>0} \frac{\omega_i}{p_i} h_i h_i^\top U, \tag{16}$$
$$U^+ \in \arg\max_{Z^\top Z = I} \operatorname{tr}(Z^\top D).$$

The second line is a thin polar factor, any missing columns are completed within the residual span. Terms with $p_i = 0$, which occur only if $\beta_i = 0$ and $U^\top h_i = 0$, contribute the zero subgradient. Convexity of f as a function of the matrix U gives

$$f(U^+) \geq f(U) + \operatorname{tr}(D^\top (U^+ - U)) \geq f(U).$$

Thus each local refinement step decreases or preserves δ. Initialized at V_m, the iterates retain (12) and its finite-radius consequence (13). Section 6 compares this polar refinement with its spectral initialization. The eigenvalue tail provides a common lower bound on the optimal price.

The spectral quantities also certify the use of V_m in decisions: the required dimension, attainable tolerances, and screening and regret bounds for the candidates.

Corollary 7 (Dimension, screening, and regret)**.** *The smallest budget permitting $\Phi_\varepsilon(x_i) - \Phi_\varepsilon^V(x_i) = o(\varepsilon)$ for every candidate is* $\operatorname{rank} C$. *For $\varepsilon > 0$,*

$$\varepsilon r_m + \overline{M}\varepsilon^2 \leq \eta \implies \Delta_\varepsilon(V_m) \leq \eta,$$
$$\varepsilon r_m/\alpha - \overline{M}\varepsilon^2 > \eta \implies \Delta_\varepsilon(V) > \eta \quad \text{for all admissible } V.$$

Let $\underline{\Phi}_i$ be the value of the explicit field $\varepsilon P_{V_m} g_i / \|P_{V_m} g_i\|$ (zero if its denominator vanishes), and put

$$\overline{\Phi}_\varepsilon^{\mathrm{T}}(x_i) = \Phi_0(x_i) + \varepsilon\sigma_i + M_i\varepsilon^2/2,$$
$$\underline{\Phi}_\varepsilon^{\mathrm{T}}(x_i) = \Phi_0(x_i) + \varepsilon\sigma_i - M_i\varepsilon^2/2,$$
$$\overline{\Phi}_i = \min\{\Phi_\varepsilon^{\mathrm{G}}(x_i),\ \overline{\Phi}_\varepsilon^{\mathrm{T}}(x_i)\},$$
$$\underline{\Phi}_i^+ = \max\{\underline{\Phi}_i,\ \underline{\Phi}_\varepsilon^{\mathrm{T}}(x_i)\}.$$

Every Wasserstein-ball minimizer survives $\underline{\Phi}_i^+ \leq \min_j \overline{\Phi}_j$. If $\hat{\imath} \in \arg\min_i \underline{\Phi}_i$, its full-ball regret $R_{\hat{\imath}} = \Phi_\varepsilon(x_{\hat{\imath}}) - \min_i \Phi_\varepsilon(x_i)$ satisfies

$$R_{\hat{\imath}} \leq \overline{\Phi}_{\hat{\imath}} - \min_i \underline{\Phi}_i^+ \leq \overline{\Phi}_{\hat{\imath}} - \underline{\Phi}_{\hat{\imath}}, \tag{17}$$
$$\overline{\Phi}_{\hat{\imath}} - \underline{\Phi}_{\hat{\imath}} \leq \varepsilon\delta_{\hat{\imath}}(V_m) + M_{\hat{\imath}}\varepsilon^2 \leq \frac{\varepsilon r_m}{\omega_{\hat{\imath}}} + M_{\hat{\imath}}\varepsilon^2.$$

If $\overline{\Phi}_{\hat{\imath}} < \min_{i \neq \hat{\imath}} \underline{\Phi}_i^+$, then $\hat{\imath}$ is the unique Wasserstein-ball minimizer.

Proof. Positive weights give $r_m = 0 \iff m \geq \operatorname{rank} C$ and $\operatorname{range} C = \operatorname{span}\{h_i\}_i$. For V_m containing this range, (9) gives all gaps $O(\varepsilon^2)$. If $m < \operatorname{rank} C$, (11) gives $\delta(V) \geq r_m/\alpha > 0$, and (9) precludes all gaps being $o(\varepsilon)$. The two tolerance claims follow from (11) and (9). For a full-ball minimizer b,

$$\underline{\Phi}_b^+ \leq \Phi_\varepsilon(x_b) \leq \Phi_\varepsilon(x_j) \leq \overline{\Phi}_j \quad (\forall j), \qquad \min_i \underline{\Phi}_i^+ \geq \min_i \underline{\Phi}_i = \underline{\Phi}_{\hat{\imath}}.$$

Finally, Taylor's lower bound and (11) yield

$$\overline{\Phi}_{\hat{\imath}} - \underline{\Phi}_{\hat{\imath}} \leq \varepsilon(\sigma_{\hat{\imath}} - \|P_{V_m} g_{\hat{\imath}}\|) + M_{\hat{\imath}}\varepsilon^2,$$
$$\omega_{\hat{\imath}}(\sigma_{\hat{\imath}} - \|P_{V_m} g_{\hat{\imath}}\|) \leq \delta(V_m) \leq r_m.$$

□

Corollary 7 concerns a prescribed menu. When the restricted problem can be minimized over the full feasible set, the same inclusion bounds the regret relative to the global optimum. Here the common space matters: if each $\ell_x(z)$ is convex in x, then $x \mapsto \Phi_\varepsilon^V(x)$ is a supremum of convex functions for fixed V, so $\underline{v}_\varepsilon^V$ is a bound from a convex problem whose inner maximization runs over only $\dim V$ directions.

Proposition 8 (Global decision certificate)**.** *Let $\mathcal{X}$ be a nonempty feasible decision set, let $x_1, \dots, x_N \in \mathcal{X}$, and use one fixed field space V for all $x \in \mathcal{X}$. Suppose $\underline{v}_\varepsilon^V \le v_\varepsilon^V$ is a global lower bound and $\Phi_\varepsilon(x_i) \le \overline{\Phi}_i$ for each candidate. With $v_\varepsilon = \inf_{x\in\mathcal{X}} \Phi_\varepsilon(x)$ and $\overline{v}_\varepsilon = \min_{1\le i\le N} \overline{\Phi}_i$,*

$$\begin{aligned} \underline{v}_\varepsilon^V &\le v_\varepsilon \le \overline{v}_\varepsilon, \\ \Phi_\varepsilon(x_a) - v_\varepsilon &\le \overline{v}_\varepsilon - \underline{v}_\varepsilon^V \end{aligned} \tag{18}$$

for any index a attaining $\overline{v}_\varepsilon$. The same bound holds for an exact minimizer of Φ_ε over the candidates retained by the screening rule of Corollary 7. For any $x \in \mathcal{X}$, define the curvature upper bound $\overline{\Phi}_\varepsilon^{\mathrm{T}}(x) = \mathbb{E}_{Q_0}\ell_x + \varepsilon\|g_x\| + M_x\varepsilon^2/2$. If x_ε^V globally minimizes $\Phi_\varepsilon^V(x)$ over $\mathcal{X}$, then

$$\begin{aligned} \Phi_\varepsilon(x_\varepsilon^V) - v_\varepsilon &\le \overline{\Phi}_\varepsilon^{\mathrm{T}}(x_\varepsilon^V) - \Phi_\varepsilon^V(x_\varepsilon^V) \\ &\le \varepsilon\delta_{x_\varepsilon^V}(V) + M_{x_\varepsilon^V}\varepsilon^2. \end{aligned}$$

Proof. The inclusion $\mathcal{B}_\varepsilon^V \subseteq \mathcal{B}_\varepsilon$ gives $\Phi_\varepsilon^V(x) \le \Phi_\varepsilon(x)$ for every $x \in \mathcal{X}$, hence $\underline{v}_\varepsilon^V \le v_\varepsilon$. A candidate minimizing $\overline{\Phi}_i$ has full-ball value at most $\overline{v}_\varepsilon$. The selected candidate minimizes Φ_ε over the candidate list and has value no larger than that candidate. For x_ε^V, use $\Phi_\varepsilon^V(x_\varepsilon^V) = v_\varepsilon^V$ and Proposition 4. □

For finite $\mathcal{X}$, minimizing candidate lower bounds gives $\underline{v}_\varepsilon^V$. For continuous $\mathcal{X}$, convexity in x alone does not suffice: the restricted inner problem and its subgradients must also be computed globally. The second bound involves the price at x_ε^V, which Corollary 7 bounds by r_m/ω_a when x_ε^V is a candidate x_a. This motivates building V_m from candidates near the expected solution.

The candidate-specific certificate in (17) uses the projected gradients directly. The final eigenvalue bound converts aggregate price into an individual bound through $1/\omega_{\hat{\imath}}$. For convex losses, $\Phi_0(x_i) + \varepsilon\sigma_i$ can replace the Taylor lower bound in $\underline{\Phi}_i^+$. Indeed, evaluating the supporting-plane inequality at the feasible field $v = \varepsilon g_i/\sigma_i$ gives $\Phi_\varepsilon(x_i) \ge \Phi_0(x_i) + \varepsilon\sigma_i$ (take $v = 0$ if $\sigma_i = 0$). We use this sharper lower bound in the convex quadratic comparison in Section 6. The Taylor lower bound remains valid without convexity, and the curvature upper bound gives an alternative to $\Phi_\varepsilon^{\mathrm{G}}(x_i)$. For nonquadratic losses, $\Phi_\varepsilon^{\mathrm{G}}(x_i)$ may be intractable and using only the curvature branch preserves the certificates but can widen them. Selecting $\hat{\imath}$ by the explicit values $\underline{\Phi}_i$ needs no further solves. The first bound in (17) certifies any other selection in the same way. These candidate intervals can be evaluated independently, while a common space specifies the movements shared by all policies.

The nonzero eigenpairs of C follow from the $N \times N$ Gram matrix of $\sqrt{\omega_i/(\sigma_i + \beta_i)}\,\widehat{h_i}$. Dense construction costs $O(N^2kd)$ and an eigensolve $O(N^3)$, in addition to gradient evaluations and base projections. The certificates use explicit fields $\underline{\Phi}_i$, evaluating exact restricted values $\Phi_\varepsilon^V(x_i)$ can require a reduced nonlinear maximization.

5. Comparison with ordinary gradient PCA

The spectral rule weights each residual gradient by its contribution to lost robust value. To assess this weighting, we compare it with the standard choice that omits it.

For the same base and budget, ordinary gradient PCA takes the leading directions of $C_{\mathrm{PCA}} = \sum_i \omega_i \widehat{h_i}\widehat{h_i}^{\mathsf{T}}$. It minimizes squared residual energy, while (10) weights residuals by their contribution to lost robust value. The matrices are proportional when all $\sigma_i + \beta_i$ coincide.

Proposition 9 (Unbounded separation from PCA)**.** *There are two equally weighted convex quadratic losses, a four-point nominal distribution in $\mathbb{R}^2$, and budget $m = 1$ for which*

$$\frac{\delta(V_{\mathrm{PCA}})}{\delta(V_1)} \longrightarrow \infty, \qquad \frac{\Delta_\varepsilon(V_{\mathrm{PCA}})}{\Delta_\varepsilon(V_1)} \longrightarrow \infty \quad \text{for every fixed } \varepsilon > 0.$$

Proof. Let $\sigma_A > 5/2$, Q_0 uniform on $\{-1, 1\}^2$, $S = V_{\mathrm{tr}}$, and

$$\ell_A(z) = \sqrt{\sigma_A^2 - 4}\, z_1 + z_1^2, \qquad \ell_B(z) = \tfrac{1}{2} z_2^2.$$

The centered fields are orthogonal, with

$$(\sigma_A, \beta_A, \|h_A\|) = (\sigma_A, \sqrt{\sigma_A^2 - 4}, 2), \qquad (\sigma_B, \beta_B, \|h_B\|) = (1, 0, 1).$$

The two eigenvalues on their respective centered directions are

$$C_{\mathrm{PCA}} : (2, 1/2), \qquad C : \left(\frac{2}{\sigma_A + \sqrt{\sigma_A^2 - 4}}, \frac{1}{2} \right).$$

Thus PCA retains h_A and the price-aware rule retains h_B, giving

$$\begin{aligned} \delta(V_{\mathrm{PCA}}) = \tfrac{1}{2}, \quad \delta(V_1) = \tfrac{1}{2}(\sigma_A - \sqrt{\sigma_A^2 - 4}), \\ \frac{\delta(V_{\mathrm{PCA}})}{\delta(V_1)} = \frac{\sigma_A + \sqrt{\sigma_A^2 - 4}}{4} \to \infty. \end{aligned}$$

For $\ell(z) = c + az_s + \frac{H}{2} z_s^2$, $H \ge 0$, direct maximization gives

$$\begin{aligned} \Phi_\varepsilon &= c + H/2 + \varepsilon\sqrt{a^2 + H^2} + H\varepsilon^2/2, \\ \Phi_\varepsilon^{V_{\mathrm{tr}}} &= c + H/2 + \varepsilon|a| + H\varepsilon^2/2. \end{aligned} \tag{19}$$

The first value is attained by $v_{j,s} = \varepsilon(a + Hz_{j,s})/\sqrt{a^2 + H^2}$ (zero field if $a = H = 0$) and equals the upper bound (7), the second uses $v_{j,s} = \varepsilon\,\mathrm{sgn}(a)$, choosing either sign if $a = 0$. With the relevant centered direction included, Φ_ε is attained. Without it, only the coordinate's constant displacement affects the loss. Therefore both aggregate gaps equal $\varepsilon\delta$, proving the second limit. □

Table 1: Exact four-scenario example. The selected policy minimizes the displayed worst-case costs. Regret uses full-ball values.

Evaluation	Cost A	Cost B	Selected	Regret
Full ball	30.6200	30.7200	A	0
Gradient PCA	30.6200	30.5200	B	0.1000
Price-aware	30.5796	30.7200	A	0

Combined with (12), Proposition 9 shows that gradient PCA admits no bounded approximation factor for δ. This is a worst-case statement, on individual instances the ordering can reverse. PCA achieves a smaller restriction price for the following three equally weighted convex losses under the same four-point nominal law:

$$\ell_1 = 9z_1 + z_1^2 + z_2^2, \quad \ell_2 = z_1 + \tfrac{3}{2}z_2^2, \quad \ell_3 = \tfrac{1}{2}z_1^2.$$

In the orthonormal centered-field basis $(z_1 e_1, z_2 e_2)$, their coordinates are $(2, 2)$, $(0, 3)$, and $(1, 0)$, with base norms $9, 1, 0$. Direct diagonalization of the two 2×2 matrices yields $\delta(V_{\mathrm{PCA}}) \approx 0.29846 < \delta(V_1) \approx 0.33547$.

6. Decision example and reproducible comparison

This section applies the bounds to decisions. An exact example shows that the choice of space can change the selected policy. The screening rule and the global certificate turn value brackets into decision guarantees and a synthetic study compares the spectral rule with PCA on prices and on reduced decision problems.

For an exact decision example, use Proposition 9 with $T = 10$, $\varepsilon = 0.2$, and add constants $c_A = 27.58$, $c_B = 30$. They calibrate a close robust ranking and do not affect either subspace. Formula (19) gives the exact values in Table 1. The aggregate restriction gaps are 0.1000 for PCA and 0.0202041 for the price-aware rule. The latter selects the robustly preferable policy. Proposition 3 also identifies both full-ball values with their outer moment bounds.

Screening.

For a prescribed finite menu, compute the bounds of Corollary 7, put $u_* = \min_{1 \le i \le N} \overline{\Phi}_i$, and discard candidates with $\underline{\Phi}_i^+ > u_*$. If one candidate remains, it minimizes the full-ball value over the decision set. Otherwise, evaluate the full-ball values only for the retained candidates and select a minimizer. Proposition 8 extends the certificate to the full feasible decision set when a global lower bound is available. Screening alone does not need a common space, since each bracket may use its own feasible field. With V_m, the brackets reuse the directions that also define the reduced problem of Proposition 8.

A continuous-decision example.

Use the four-point nominal law of Proposition 9, $\varepsilon = 0.2$, and the common feasible set $\mathcal{X} = [0, 1]$. For $x \in \mathcal{X}$, put

$$\ell_x(z) = (x - \tfrac{1}{2})^2 + x(\sqrt{96}\, z_1 + z_1^2 - 3).$$

Equation (19) gives $\Phi_{0.2}(x) = (x - \frac{1}{2})^2 + 0.04x$, with global minimizer $x_\varepsilon = 0.48$ and value $v_\varepsilon = 0.0196$. Minimization over the set $\{0, \frac{1}{2}, 1\}$ gives $\widehat{x} = \frac{1}{2}$, whose value is 0.0200 and global regret is 0.0004. Globally minimizing the translation restriction gives $B_{\mathrm{tr}} = -0.00020414$. The one-direction spectral space built from this set contains its sole centered gradient direction, so $B_1 = \min_{x \in \mathcal{X}} \Phi_\varepsilon^{V_1}(x) = v_\varepsilon$. The curvature upper bound at $\widehat{x}$ is $U(\widehat{x}) = 0.0200$. Proposition 8 therefore gives regret bounds $U(\widehat{x}) - B_{\mathrm{tr}} = 0.02020414$ and $U(\widehat{x}) - B_1 = 0.0004$, although x_ε is absent from the menu. The translation bound is conservative, while the spectral bound equals the actual regret. Computing either global lower bound requires optimization over $\mathcal{X}$. Finite-set screening alone gives no guarantee outside the menu, and the synthetic comparisons below concern finite-set values.

Synthetic comparison.

The examples above are constructed. The following study checks on generated instances how the factor of Theorem 6 behaves and how the spectral prices compare with PCA. Claude and GPT assisted with implementing and checking the numerical methods and plotting scripts. All reported results were computed using the stated instance generators, random seeds and solvers. For the numerical comparison, we use 100 convex quadratic instances: 40 heterogeneous instances and 20 each with balanced gradients, weak linear terms, and heavy-tailed observations. The first three designs have $(d, k, N) = (6, 32, 12)$; the last has $(10, 48, 24)$. All use equal weights and translations as the base. We compare gradient PCA, the spectral rule, and the local refinement (16) with five initial spaces: the spectral and PCA spaces and three seeded random spaces. For a smooth nonquadratic comparison, we also use 20 log-sum-exp instances at each $(d, k, N) = (4, 16, 8), (6, 32, 12), (8, 48, 16)$, with

$$\ell_i(z) = \log \sum_{r=1}^{3} e^{a_{ir}^\top z + b_{ir}} + u_i^\top z.$$

The code specifies the Gaussian generator and log-uniform scales. Let A_i have rows $a_{ir}^\top$, $r = 1, 2, 3$. The softmax Hessian gives $M_i = \|A_i\|_2^2/2$. At $\varepsilon = 0.2$, a pairwise-slope curvature bound certifies concavity of the transport-dual inner maximizations. Table 2 reports the mean leading price and finite-radius gap for both families with $m = 3$ and $\varepsilon = 0.2$. Across all seven designs, the reduction in mean leading price is accompanied by a reduction in the mean finite-radius gap at $\varepsilon = 0.2$. For $m = 3$, Figure 2, panel (a), compares the relative reductions in δ and $\Delta_{0.2}$ for each instance. All 100 points lie in the positive quadrant, and the two reductions differ by at most 2.74 percentage points. Table 2 also reports the curvature-to-tail ratios that determine how closely the finite-radius guarantee follows the sharp first-order factor. For these instances, the uniform curvature remainder does not certify a multiplicative factor close to α. The numerical comparisons show that the spectral rule can nevertheless perform well in this regime.

Direct refinement further reduces δ by an average of 0.79% over both budgets $m \in \{1, 3\}$, showing that the spectral construction already provides a competitive starting point for the tested local optimization. At $\varepsilon = 0.8$, the spectral space yields a smaller

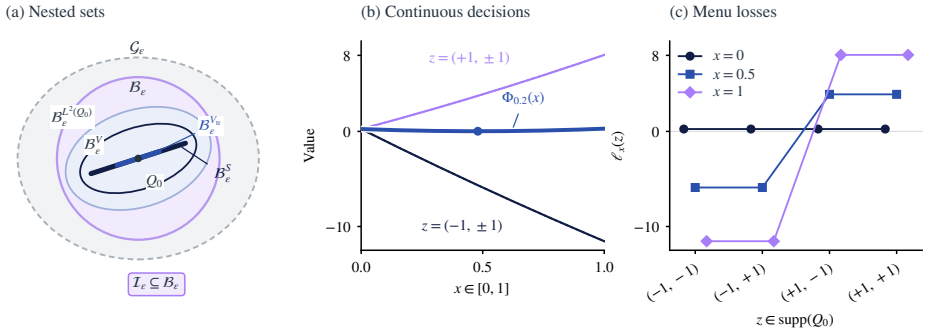


Figure 1: Subset geometry and continuous decisions. (a) Schematic sets of Proposition 2, with $V = V_m = S \oplus W_m$, $\mathcal{I}_\varepsilon$ is separate. (b) Four nominal-scenario losses and $\Phi_{0.2}(x)$ for Proposition 9, pairs with the same z_1 coincide, and $x_\varepsilon = 0.48$. (c) Scenario losses for the set $\{0, \frac{1}{2}, 1\}$, which omits x_ε.

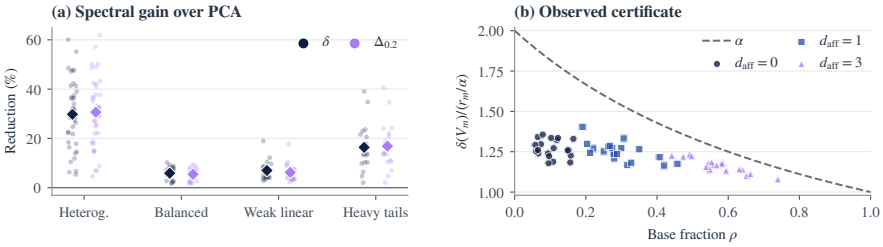


Figure 2: Spectral versus PCA. (a) Paired gains in δ and $\Delta_{0.2}$, diamonds are design means. (b) Factor $\delta(V_m)/(r_m/\alpha)$ and bound α.

Table 2: Mean leading and finite-radius prices, $m = 3$ and $\varepsilon = 0.2$; Direct denotes five-start refinement. The last column gives the mean ratio $\theta_{0.2}$. At radii 0.05 and 0.8, the ratios are one quarter and four times as large.

Design	Metric	PCA	Spectral	Direct	$\theta_{0.2}$
Heterogeneous	δ	0.27183	0.19529	0.19230	3.93
	$\Delta_{0.2}$	0.06033	0.04292	0.04248	
Balanced	δ	0.51850	0.48780	0.48368	1.02
	$\Delta_{0.2}$	0.11688	0.11050	0.10983	
Weak linear	δ	0.59022	0.55115	0.54763	0.86
	$\Delta_{0.2}$	0.13209	0.12436	0.12384	
Heavy tails	δ	0.53068	0.44313	0.43811	1.49
	$\Delta_{0.2}$	0.11611	0.09639	0.09553	
LSE, $d = 4$	δ	0.01457	0.01210	0.01208	58.46
	$\Delta_{0.2}$	0.00330	0.00278	0.00278	
LSE, $d = 6$	δ	0.03739	0.03166	0.03158	16.18
	$\Delta_{0.2}$	0.00804	0.00680	0.00680	
LSE, $d = 8$	δ	0.05655	0.05189	0.05169	8.16
	$\Delta_{0.2}$	0.01201	0.01104	0.01102	

finite-radius gap than its refinement in 145 of 200 paired comparisons across $m \in \{1, 3\}$. Refinement optimizes the leading price, so its additional improvement need not persist when curvature becomes important.

On the 20 balanced instances, we enlarge the base to fields affine in the first $d_{\mathrm{aff}} \in \{0, 1, 3, 6\}$ coordinates, with $d_{\mathrm{aff}} = 0$ denoting translations, and add up to three residual directions. Table 3 reports the mean base fraction, factor, and eigenvalue tail, together with the mean instancewise certified ratio $\delta(V)/(r_m/\alpha)$ for the spectral space and its direct refinement. Direct refinement gives only a small additional reduction. For $r_m > 0$, the largest spectral ratio is 1.404; Figure 2, panel (b), plots the instancewise ratios against ρ. Since $r_m/\alpha \le \inf_V \delta(V)$, these ratios bound the unknown approximation ratios from above. Total dimension

Table 3: Base-space sweep on 20 balanced instances, $m = 3$. Entries are instancewise means; the last two columns give $\delta(V)/(r_m/\alpha)$ when $r_m > 0$. Direct uses five local starts.

d_{aff}	ρ	α	r_m	Spectral	Direct
0	0.105	1.812	0.702	1.269	1.259
1	0.294	1.549	0.460	1.250	1.242
3	0.571	1.275	0.186	1.164	1.161
6	1	1	0	—	—

increases with d_{aff}, so the comparison does not isolate the effect of ρ. Full affine fields recover the full quadratic values.

Reduced decision problems.

Proposition 8 is the setting in which a common space is needed, and we test it on the simplex $\mathcal{X}$ of portfolio weights. The quadratic family uses $\ell_x(z) = -x^\top z + \frac{\gamma}{2}(x^\top z)^2$ with $\gamma = 15$, eight assets and 40 factor-model scenarios. The nonquadratic family uses $\ell_x(z) = \tau \log \sum_{r=1}^3 e^{(z^\top A_r x + b_r^\top x)/\tau}$ with $\sum_r A_r = 0$, $\tau = 0.5$, six features and 32 standard normal scenarios, so Proposition 3 does not apply. Both losses are convex in x, and each family has 20 instances. Spaces with translations as the base and $m = 3$ are built from 12 candidates, either random portfolios or perturbations of the nominal minimizer. Cutting planes certify $\underline{v}_\varepsilon^V \le v_\varepsilon^V$. A cut stays valid when the inner maximization stops at a local maximum, since $x \mapsto \mathbb{E}_{Q_0} \ell_x(z + v)$ is convex and bounded by $\Phi_\varepsilon^V(x)$ for every fixed feasible field v. Table 4 reports the share $(\underline{v}_\varepsilon^V - v_0)/(v_\varepsilon - v_0)$ of the robustness premium, where v_0 is the nominal optimum. For the nonquadratic family, v_ε is enclosed by the full-field bound and a Wasserstein dual bound with strongly concave inner problems. The enclosure is exact up to 10^{-5} of the premium at all radii. With random

Table 4: Share of the robustness premium, $(\underline{v}^V_\varepsilon - v_0)/(v_\varepsilon - v_0)$, means over 20 instances, $m = 3$, translations as base.

Loss	Candidates	ε	Transl.	PCA	Spectral
Quadratic	random	0.02	0.333	0.504	0.604
		0.1	0.527	0.610	0.672
	local	0.02	0.333	0.968	0.975
		0.1	0.527	0.875	0.880
Log-sum-exp	random	0.05	0.365	0.790	0.830
		0.4	0.448	0.786	0.826
	local	0.05	0.365	0.995	0.996
		0.4	0.448	0.977	0.978

candidates, the spectral space certifies more of the premium than PCA in all quadratic instances and in 17 of 20 nonquadratic ones at every radius. With candidates near the nominal minimizer, both spaces certify at least 87%, and translations alone at most 53%. The full problems remain tractable at these sizes, so the comparison measures how much robust value a three-direction common space retains and makes no claim about computational savings. In the quadratic family, the reduced minimizers are not systematically better decisions than the nominal minimizer, so the bound is the relevant output.

7. Conclusions

We asked how much robust value a common low-dimensional transport restriction loses and how to choose it. The subset sandwich closes for affine losses using translations and for quadratic losses using affine fields when the nominal covariance is nonsingular. For smooth losses, gradient projections quantify the value lost by a transport restriction. The spectral rule approximates the best aggregate leading price around a prescribed base within the sharp factor $2/(1+\rho)$ and supplies dimension and decision certificates. Gradient PCA, which preserves gradient energy, admits no such factor: its lost robust value can be arbitrarily larger, although it can be smaller on individual instances. The same value bounds screen finite candidate menus. A fixed common space turns the robust decision problem into a reduced problem whose value bounds the robust optimum from below. Together with an upper bound at the reduced minimizer, it bounds the regret even when the optimizer is absent from the candidate menu. In synthetic portfolio problems with quadratic and log-sum-exp losses, three spectral directions certified on average 60 to 83 percent of the robustness premium from random candidates and at least 88 percent from candidates near the nominal minimizer, more than PCA and well above translations alone. Finite-radius accuracy is governed by the curvature remainder, and a computational benefit of reduced spaces for problems whose full dual is intractable remains to be shown.

Additional information.

Contributions (CRediT). Conceptualization: Betz, Roesel; methodology, formal analysis, validation: Betz, Daßer; coding, writing (review/editing): all authors; visualization: Betz; supervision: Roesel. *Competing interests:* none. *Data and code:* Code and synthetic instance generators for the numerical comparisons are archived at `https://doi.org/10.5281/zenodo.23964954` under the MIT license.

Acknowledgements.

Doctoral support: M. Betz—Federal Ministry for Economic Affairs and Energy (BMWE), KoPF, grant 20Q2221B; H. Daßer—Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Priority Programme "Theoretical Foundations of Deep Learning" (SPP 2298). The funders have no active role in this research.

Declaration of generative AI use

Claude (Anthropic) and GPT (OpenAI) assisted with literature searches, theoretical verification, code, figures and formulations. The authors reviewed the outputs and take full responsibility for the article.

References

[1] M. Gelbrich, On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces, Mathematische Nachrichten 147 (1) (1990) 185–203. `doi:10.1002/mana.19901470121`.

[2] D. Kuhn, P. Mohajerin Esfahani, V. A. Nguyen, S. Shafieezadeh-Abadeh, Wasserstein distributionally robust optimization: Theory and applications in machine learning, in: Operations Research & Management Science in the Age of Analytics, INFORMS TutORials in Operations Research, INFORMS, 2019, pp. 130–166. `doi:10.1287/educ.2019.0198`.

[3] R. Gao, Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality, Operations Research 71 (6) (2023) 2291–2306. `doi:10.1287/opre.2022.2326`.

[4] D. Bartl, S. Drapeau, J. Obłój, J. Wiesel, Sensitivity analysis of Wasserstein distributionally robust optimization problems, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 477 (2256) (2021) 20210176. `doi:10.1098/rspa.2021.0176`.

[5] H. Namkoong, Local sensitivity under transport restrictions, preprint (2026). `arXiv:2606.04276`, `doi:10.48550/arXiv.2606.04276`. URL `https://arxiv.org/abs/2606.04276`

[6] J. Cheng, R. L.-Y. Chen, H. N. Najm, A. Pinar, C. Safta, J.-P. Watson, Distributionally robust optimization with principal component analysis, SIAM Journal on Optimization 28 (2) (2018) 1817–1841. `doi:10.1137/16M1075910`.

[7] M. Cheramin, J. Cheng, R. Jiang, K. Pan, Computationally efficient approximations for distributionally robust optimization under moment and Wasserstein ambiguity, INFORMS Journal on Computing 34 (3) (2022) 1768–1794. `doi:10.1287/ijoc.2021.1123`.

[8] K. Fan, On a theorem of Weyl concerning eigenvalues of linear transformations I, Proceedings of the National Academy of Sciences 35 (11) (1949) 652–655. `doi:10.1073/pnas.35.11.652`.

[9] C. Ding, D. Zhou, X. He, H. Zha, R_1-PCA: Rotational invariant L_1-norm principal component analysis for robust subspace factorization, in: Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 281–288. `doi:10.1145/1143844.1143880`.

[10] A. Naor, O. Regev, T. Vidick, Efficient rounding for the noncommutative Grothendieck inequality, Theory of Computing 10 (11) (2014) 257–295. `doi:10.4086/toc.2014.v010a011`.