Robust Intrinsic Scale Estimation on Metric Spaces

Jongmin Lee
(Joint work with Paromita Dubey and Sungkyu Jung)

Pusan National University

June 25, 2026 (ScoDA workshop)

What is scale (dispersion)?

Scale (dispersion) is one of the most basic statistical summaries. In Euclidean statistics:

- Variance measures spread around the mean;
- MAD (median absolute deviation) measures spread around the median;
- Many methods begin with center, then scale.

Question: Is this always the right order in metric spaces?

Data living on metric spaces

Examples include:

- shapes,
- networks,
- covariance matrices, and
- probability distributions.

Common structure: We may not have vector operations, but we do have distances.

Why this is problematic for random objects

In metric spaces, the center may not be reliable. The center may

- be non-unique;
- be non-existent;
- reflect geometric pathologies (smeariness, stickiness).

Takeaway: If the center is unstable, the center-based scale might inherit that instability.

Our viewpoint and goal

Viewpoint of this talk

- A center is already a statistical construction.
- Pairwise distances are primitive metric information.
- Scale should be definable before choosing a center.

Our goal is two-fold:

- construction of scale estimators based on **pairwise distances** (intrinsicity);

- construction of **robust** scale estimators.

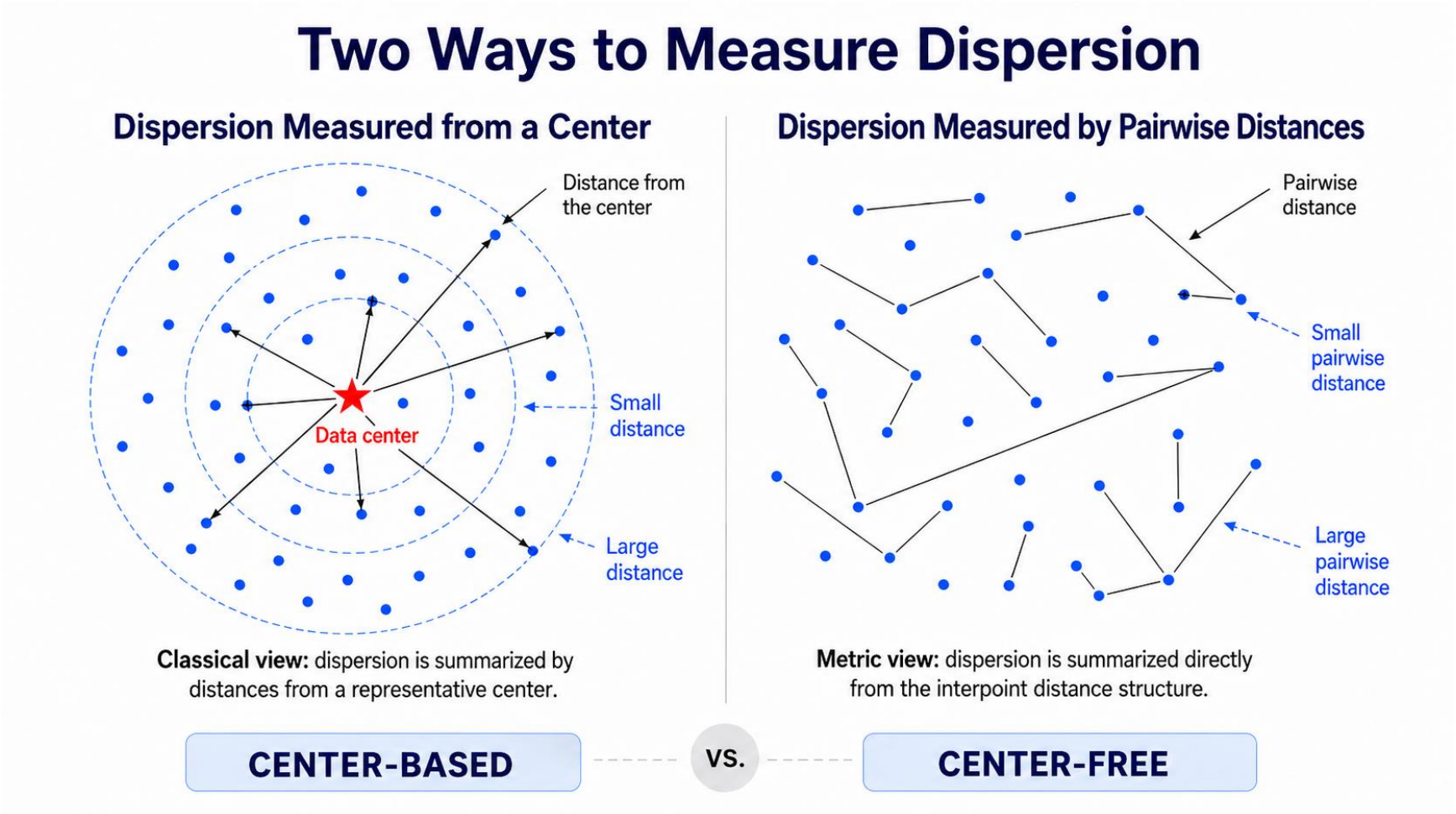
What intrinsicity does mean

- We use the term "intrinsic" in a stronger, center-free sense. A scale estimator (or functional) is called *intrinsic* if it is determined by the interpoint metric structure of the sample, rather than through an optimizing center.

- While scale in Euclidean statistics is often measured relative to a center, our viewpoint is that, in general metric spaces, scale should be definable prior to, and independently of, the choice of a center.

- Intrinsic does not mean: always better, always more efficient, a replacement for Fréchet-type methods. It means determined by the interpoint metric structure without defining centers.

Two ways to measure dispersion



Center-based: distances from a representative point

Center-free: distances among data points themselves

The second view uses only the metric configuration of the data.

Bickel-Lehmann (BL) axiom for spread

On the real line, if

- $F_X^{-1}(v) - F_X^{-1}(u) \leq F_Y^{-1}(v) - F_Y^{-1}(u)$ for all $0 < u < v < 1$
- $\Rightarrow S(X) \leq S(Y)$,

then S said to be a spread functional.

Question: How do we extend the BL axiom for spread into metric spaces?

From real-line spacings to metric spacings

- The real-line formulation uses: subtraction, quantiles, linear ordering.

- These are not available in a general metric space.
  Replacement idea

$$\text{quantile spacings} \quad \longrightarrow \quad \text{pairwise distances.}$$

Takeaway:
Pairwise distances are metric analogues of spacings.

Bickel-Lehmann (BL) spread axiom and its generalization

(BL axiom for spread) On the real line,

- $\dfrac{F_X^{-1}(v) - F_X^{-1}(u) \leq F_Y^{-1}(v) - F_Y^{-1}(u) \ (0 < u < v < 1)}{\Rightarrow S(X) \leq S(Y),}$

(Generalization)

- On the real line, for an independent uniform random variables U, V, when $V > U$ occurs

  $$(X - X' \overset{d}{=}) \ F_X^{-1}(V) - F_X^{-1}(U) \leq F_Y^{-1}(V) - F_Y^{-1}(U) \ (\overset{d}{=} Y - Y')$$

  by the probability integral transform.
- When $U \geq V$ occurs,

  $$(X' - X \overset{d}{=}) \ F_X^{-1}(V) - F_X^{-1}(U) \leq F_Y^{-1}(V) - F_Y^{-1}(U) \ (\overset{d}{=} Y' - Y)$$

How to generalize BL spread axiom into metric spaces?

Therefore,

$$\underline{F_X^{-1}(v) - F_X^{-1}(u) \leq F_Y^{-1}(v) - F_Y^{-1}(u) \text{ (for all } 0 < u < v < 1)} \tag{1}$$

implies

$$|X - X'| \leq_{\text{st}} |Y - Y'|, \tag{2}$$

where $Z \leq_{\text{st}} W$ implies $\mathbb{P}(Z > t) \leq \mathbb{P}(W > t)$ for all $t \in \mathbb{R}$.
The main idea is to replace (1) by (2). That is, we consider the following monotonicity ordering

$$|X - X'| \leq_{\text{st}} |Y - Y'| \Rightarrow S(X) \leq S(Y), \tag{3}$$

instead of BL spread axiom.

(3) can be written on metric spaces as follows:

$$d(X, X') \leq_{\text{st}} d(Y, Y') \Rightarrow S(X) \leq S(Y), \tag{4}$$

which we call pairwise-distance monotonicity axiom for ("nice" scale functional).

Question: What is the sample version of (4)?

- Denote the distribution functions of $d(X, X')$ and $d(Y, Y')$ by F_1 and F_2, respectively.

- $d(X, X') \leq_{\text{st}} d(Y, Y') \Leftrightarrow F_1(t) \geq F_2(t)\ (t > 0) \Leftrightarrow F_1^{-1}(u) \leq F_2^{-1}(u)\ (0 < u < 1)$

Pairwise-distance monotonicity (sample version)

Consider a set of pairwise distances of data points $\{d(x_i, x_j)\}_{i<j}$.
Let

$$r_{(1)}(X) \leq \cdots \leq r_{(N)}(X), \quad N = \frac{n(n-1)}{2},$$

be the ordered pairwise distances.
(Sample pairwise-distance monotonicity) If

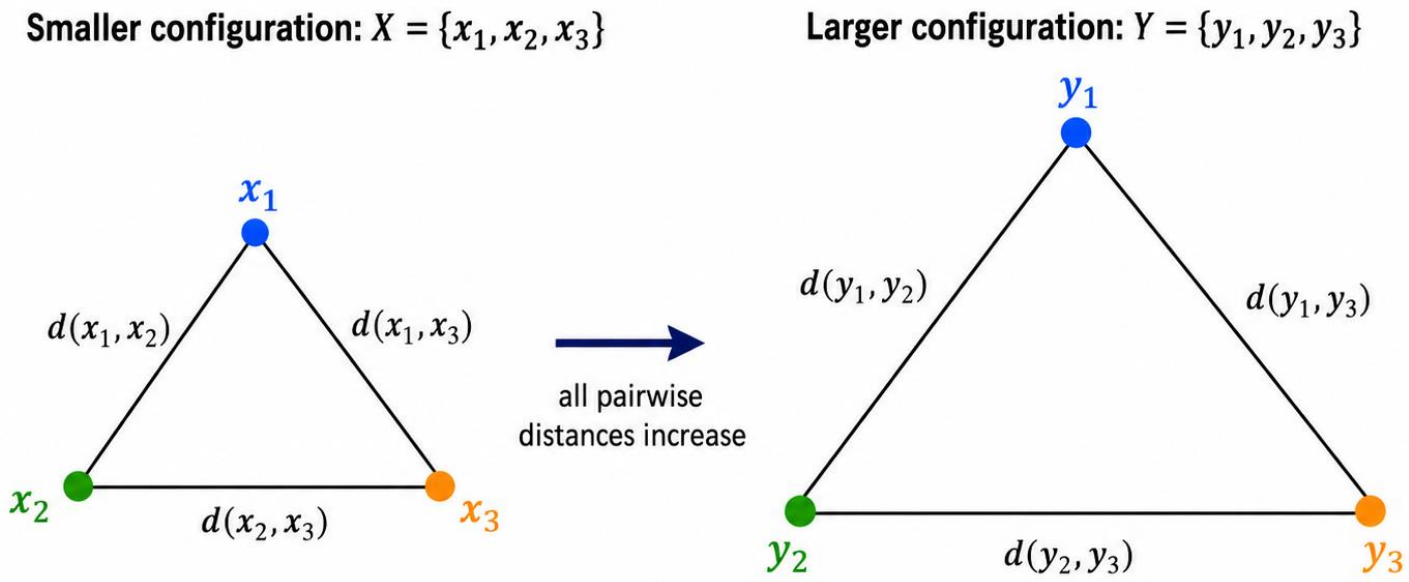
$$r_{(k)}(\mathbf{X}) \leq r_{(k)}(\mathbf{Y}) \quad \text{for all } k,$$

then

$$S_n(\mathbf{X}) \leq S_n(\mathbf{Y}).$$

Takeaway:
If all pairwise distances become larger, scale should not decrease.

Pairwise Distance Monotonicity



$$d(x_1, x_2) <= d(y_1, y_2),$$
$$d(x_1, x_3) <= d(y_1, y_3),$$
$$d(x_2, x_3) <= d(y_2, y_3).$$

Hence, X is no more dispersed than Y (pairwise distance monotonicity holds).

If every side length increases,

$$d(x_i, x_j) \leq d(y_i, y_j),$$

then the second configuration should not have smaller scale.

Axioms for intrinsic scale

A scale estimator is called *intrinsic* if it satisfies:

A1. Permutation invariance

A2. **Pairwise-distance monotonicity**

A3. Isometric invariance

A4. Strict positivity for distinct observations

A5. Metric-scaling equivariance

Takeaway:
Intrinsic scale depends only on the metric configuration and respects spread.

What intrinsicity does mean

- We use the term "intrinsic" in a stronger, center-free sense. A scale estimator (or functional) is called intrinsic if it is determined by the interpoint metric structure of the sample, rather than through an optimizing center.

- While scale in Euclidean statistics is often measured relative to a center, our viewpoint is that, in general metric spaces, scale should be definable prior to, and independently of, the choice of a center.

- Intrinsic does not mean: always better, always more efficient, a replacement for Fréchet-type methods. It means determined by the interpoint metric structure without defining centers.

The familiar benchmark: Fréchet SD

Empirical Fréchet standard deviation is defined as

$$\sigma_{F,n} := \{ \inf_{m \in M} \frac{1}{n} \sum_{i=1}^{n} d^2(x_i, m) \}^{1/2}.$$

It is natural, widely used, center-based.
Question: How robust is it?

Fréchet SD has a low breakdown point

Theorem

If (M, d) is unbounded, then

$$\varepsilon_n^*(\sigma_{F,n}, \mathbf{X}) = \frac{1}{n}.$$

A single outlier can make $\sigma_{F,n}$ explode.

Takeaway:

Fréchet SD is natural, but not high-breakdown robust.

Center-based robust alternatives

- Examples of Center-based robust scales include trimmed Fréchet SD, MAD from a Fréchet median.

- They can be robust. But, they still measure dispersion relative to a center.

Takeaway:
Robustness and center-freeness are different issues.

Weak intrinsicity

(A2-L) (Labelled pairwise-distance monotonicity) For given $\mathbf{X} = (x_1, \ldots, x_n)$ and $\mathbf{Y} = (y_1, \ldots, y_n)$, if, after relabeling,

$$d(x_{\pi(i)}, x_{\pi(j)}) \leq d(y_i, y_j) \quad \forall i < j,$$

then

$$S_n(\mathbf{X}) \leq S_n(\mathbf{Y}).$$

If a scale estimator satisfies (A1), (A2-L), (A3), (A4), and (A5), then we call it a weakly intrinsic scale.

Takeaway: Weak intrinsicity is a useful intermediate notion between center-based scales and strictly intrinsic scales.

The empirical Fréchet standard deviation satisfies (A2-L)

Theorem

The empirical Fréchet standard deviation $\sigma_{F,n}$ satisfies (A2-L) on Hadamard spaces with constant curvature

Proof.

Suppose that for given $\mathbf{X} = (x_1, \ldots, x_n)$, $\mathbf{Y} = (y_1, \ldots, y_n) \in M^n$, $d(x_i, x_j) \leq d(y_i, y_j)$ for all $i < j$. Let $S = \{y_1, \ldots, y_n\} \subset M$ and define a map $f : S \to M$ so that $f(y_k) = x_k$ for all $k = 1, \ldots n$. The map f is well-defined. Indeed, if $y_i = y_j$, then $d(x_i, x_j) \leq d(y_i, y_j) = 0$, implying that $x_i = x_j$. □

Proof.

Notably, f is 1-Lipschitz since $d(f(y_i), f(y_j)) = d(x_i, x_j) \leq d(y_i, y_j)$ holds for any i, j. By an application of the Kirszbraun-type extension theorem to the Hadamard space with constant curvature (M, d), there exists a 1-Lipschitz (nonexpansive) map $F : M \to M$ such that $F|_S = f$. Now denote the empirical Fréchet mean of $\mathbf{Y}$ by $\mu_{\mathbf{Y}}$, which exists and is unique. Then,

$$\sigma^2_{F,n}(\mathbf{X}) = \inf_{m \in M} \sum_{i=1}^{n} d^2(x_i, m) \leq \frac{1}{n} \sum_{i=1}^{n} d^2(x_i, F(\mu_Y)) \tag{5}$$

Since $x_i = F(y_i)$ for all i and F is 1-Lipschitz, $d(x_i, F(\mu_{\mathbf{Y}})) = d(F(y_i), F(\mu_{\mathbf{Y}})) \leq d(y_i, \mu_{\mathbf{Y}})$ for all i. By (5),

$$\sigma^2_{F,n}(\mathbf{X}) \leq \frac{1}{n} \sum_{i=1}^{n} d^2(x_i, F(\mu_Y)) \leq \frac{1}{n} \sum_{i=1}^{n} d^2(y_i, \mu_{\mathbf{Y}}) = \sigma^2_{F,n}(\mathbf{Y}),$$

Upper bound for robustness

Theorem

On a geodesically complete Hadamard space, for any weakly intrinsic scale S_n,

$$\varepsilon_n^*(S_n, \mathbf{X}) \leq \frac{\lfloor n/2 \rfloor}{n}.$$

No weakly intrinsic scale can beat one-half breakdown.

Half-mass radius

Half-mass radius

$$R_n(X) := \min_{1 \le i \le n} \inf \left\{ r \ge 0 : \frac{1}{n} \sum_{k=1}^{n} 1_{d(x_i, x_k) \le r} \ge \frac{1}{2} \right\}, .$$

Meaning: smallest radius of a ball, centered at an observation, containing at least half of the sample.

- R_n is a weakly intrinsic scale.
- R_n has asymptotically optimal robustness.

Robustness of $\mathcal{S}_n$ and $\mathcal{Q}_n$

Classical Rousseeuw-Croux scale estimators can be defined on metric spaces:
$\mathcal{S}_n(\mathbf{X}) := \text{med}_i \text{med}_j d(x_i, x_j)$
$\mathcal{Q}_n(\mathbf{X}) :=$ first quartile of $\{d(x_i, x_j)\}_{i<j}$

Theorem

Under distinct observations,

$$\varepsilon_n^*(\mathcal{S}_n, \mathbf{X}), \varepsilon_n^*(\mathcal{Q}_n, \mathbf{X}) \geq \frac{\lfloor n/2 \rfloor}{n}.$$

$\mathcal{S}_n$ and $\mathcal{Q}_n$ reach the optimal robustness level.

Comparison of estimators

Table: Comparison of scale estimators on metric spaces.

Scale estimators	Intrinsicity	Weak intrinsicity	Robustness
Empirical Fréchet SD	No	No	No
Trimmed Fréchet SD	No	No	Yes
MAD	No	No	Yes
Half-mass radius (R_n)	No	Yes	Yes
$\mathcal{S}_n$	No	Yes	Yes
$\mathcal{Q}_n$	Yes	Yes	Yes
Metric/distance SD	Yes	Yes	No

Notes. Robustness refers to high-breakdown robustness.The trimmed Fréchet standard deviation is robust when the trimming proportion is chosen close to $1/2$.

Center-freeness does not automatically imply robustness.

Contribution

This work proposes:

- a center-free axiomatic view of scale estimators on metric spaces;
- a monotonicity ordering based on pairwise distances;
- breakdown theory for various scale estimators;
- robust examples achieving the optimal breakdown level.

Takeaway:
The contribution is a framework for thinking about scale in metric spaces.

Take-home messages

Take-home messages

- Scale needs a monotonicity axiom to reflect dispersion.
- Pairwise distances monotonicity provide a center-free ordering of spread.
- Fréchet SD is natural, but has breakdown $1/n$ on unbounded spaces. R_n, $\mathcal{S}_n$, and $\mathcal{Q}_n$ provide robust center-free alternatives.
- Breakdown one-half is optimal under weak intrinsicity.

Final message:
Scale can be studied as a primitive metric object.

Further studies

- Asymptotic properties
- Outlier detection
- Real-data performances of scale estimators

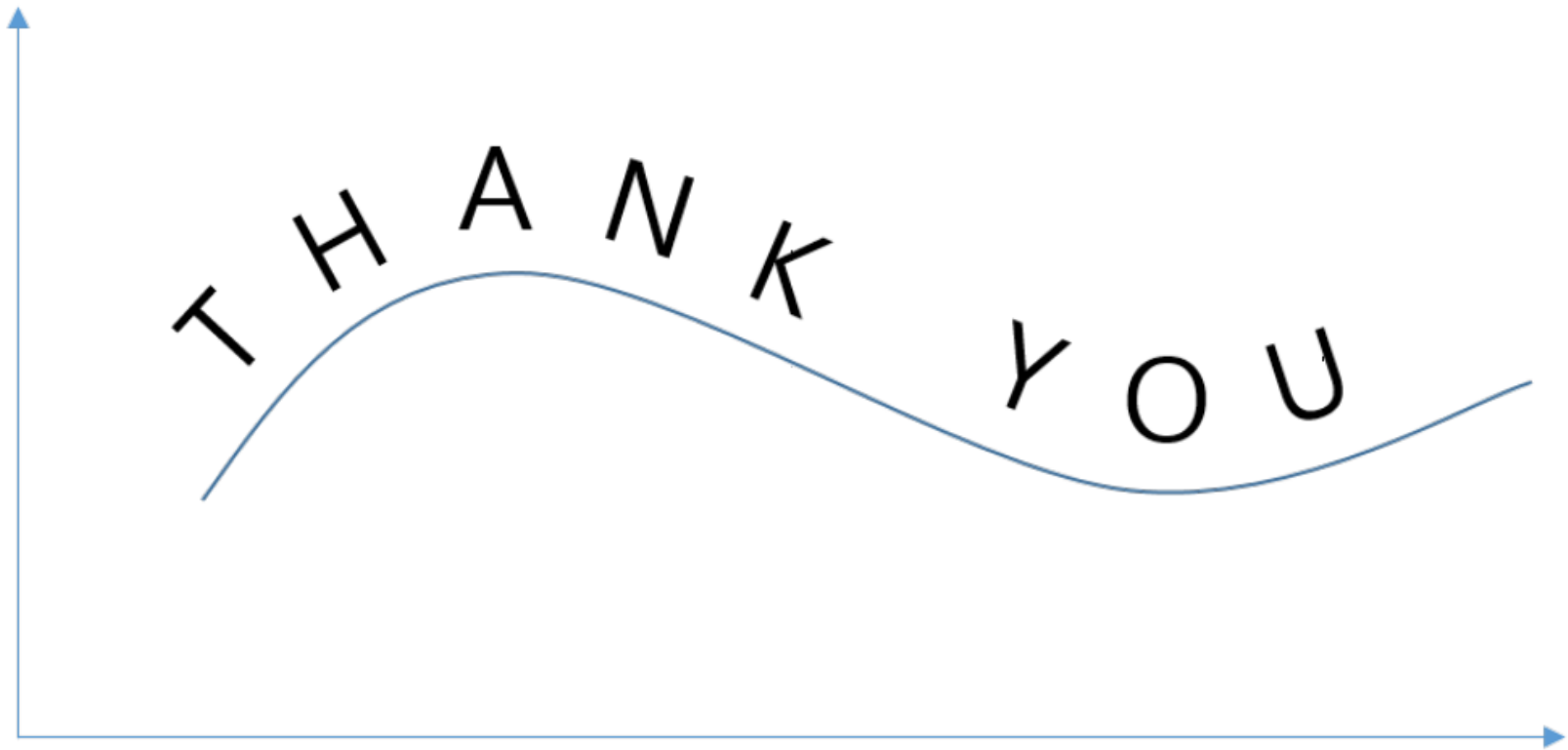
Appendix: A population counterpart of (A2-L)

We propose one based on a contraction order between probability measures on M. Specifically, for probability measures μ and ν on M, we write

$$\mu \preceq_{\mathrm{ctr}} \nu,$$

if there exists a nonexpansive (i.e., 1-Lipschitz) map $T : M \to M$ such that $T_{\#}\nu = \mu$. Thus, μ can be obtained from ν by a deterministic contraction.
For a scale functional S, a population analogue of Axiom (A2-L) is then

$$\mu \preceq_{\mathrm{ctr}} \nu \quad \Longrightarrow \quad S(\mu) \leq S(\nu).$$


THANK YOU