Magnitude Distance: A Geometric Measure of Dataset Similarity

Sahel Torkamani
School of Informatics
University of Edinburgh
sahel.torkamani@ed.ac.uk

Henry Gouk
School of Informatics
University of Edinburgh
henry.gouk@ed.ac.uk

Rik Sarkar
School of Informatics
University of Edinburgh
rsarkar@inf.ed.ac.uk

Abstract

Quantifying the distance between datasets is a fundamental question in mathematics and machine learning. We propose *magnitude distance*, a novel distance metric defined on datasets that is based on the notion of the *magnitude* of a metric space. It is an intuitive and outlier-robust geometric distance between two finite sets in $\mathbb{R}^D$. We prove various properties of magnitude distance, including aspects of the metric axioms and how it can be tuned to pay more attention to local versus global structures. An experimental example demonstrating the outlier robustness property of this approach is also given.

1 Introduction and Related Work

The problem of measuring the similarity or distance between two finite datasets plays an important role in generative modelling. The performance of generative models is often evaluated by the similarity of generated samples with the reference dataset, such as the Inception Score [14] or the Maximum Mean Discrepancy (MMD) [4]. Distance measures also play a key role in providing a learning signal during the optimization of model parameters. The choice of distance can be crucial in the viability of a prospective generative modelling approach. For example, Wasserstein Generative Adversarial Networks (WGANs) [2, 6] make use of the Wasserstein distance [16, 13, 11] in the loss function due to the issues identified with distribution divergence measures that were previously used for training GANs. Wasserstein distance and MMD have also recently been used for improving flow-based generative modelling [15, 19].

In this paper, we introduce the *magnitude distance*, a novel distance measuring the dissimilarity between finite sets $X, Y \in \mathbb{R}^D$, which captures geometric properties of the dataset. The distance is built on the magnitude of metric spaces introduced by [8]. Magnitude is a relatively new isometric invariant of metric spaces. Intuitively, magnitude can be seen as measuring the "effective size" of mathematical objects. It has been defined, adapted, and studied in many different contexts such as topology, finite metric spaces, entropy, compact metric spaces, graphs, and machine learning [8, 9, 3, 7, 5]. Magnitude distance inherits many properties of metric magnitude and is able to leverage the topological and geometric properties of datasets. A distinguishing property of the magnitude distance is that it has a scaling parameter, t, that controls the sensitivity to capturing local differences or global structures in data.

In summary, we define the magnitude distance d^t_{Mag} and its normalized variant for finite sets in $\mathbb{R}^D$. In the process of doing this, we extend existing results on the magnitude of metrics from non-redundant sets to all sets in Theorem 3. Furthermore, we analyse various properties of d^t_{Mag}, beginning with the metric axioms in Theorem 6. We prove non-negativity, and conditional identity of indiscernibles under a proposed notion of *magnitude equivalence*, and demonstrate the lack of a triangle inequality in general. We also study the behaviour of d^t_{Mag} over different values of the

EurIPS 2025 Workshop on Principles of Generative Modeling (PriGM).

scaling parameter in Theorem 7. Finally, in Theorem 8 we prove the global boundedness for outlier robustness.

2 Preliminaries

For a finite metric space, (X, d), we define the *similarity matrix* as $\zeta_X(x_i, x_j) := \exp(-d(x_i, x_j))$, for every $x_i, x_j \in X$. The concept of *magnitude*, defined in terms of a weighting, is given below.

Definition 1 (Metric Magnitude). A weighting of (X, d) is a function $w_X : X \to \mathbb{R}$ satisfying $\sum_{j \in X} \zeta_X(x_i, x_j)\, w_X(x_j) = 1$ for every $x_i \in X$, where $w_X(x_i)$ is called the *magnitude weight*. The magnitude of (X, d) is defined as

$$Mag(X, d) \coloneqq \sum_{x_i \in X} w_X(x_i). \tag{1}$$

In general, the existence of a suitable weighting, and therefore the magnitude, is not guaranteed. However, if X is a finite subset of $\mathbb{R}^D$, then ζ_X is a symmetric positive definite matrix [8, Theorem 2.5.3]. In particular, ζ_X is invertible so a unique weighting exists and the magnitude is well defined. In this case the weighting vector can be computed by inverting the similarity matrix ζ and summing all the entries; i.e., $w_X \coloneqq \zeta_X^{-1}\mathbb{1}$, where $\mathbb{1}$ is the $|X| \times 1$ column vector of all ones. Consequently, magnitude is the sum of all the entries of the weighting vector, i.e., $Mag(X) \coloneqq \mathbb{1}^* w_X = \mathbb{1}^* \zeta_X^{-1}\mathbb{1}$. When the distance measure, d, is understood from context, such as in $\mathbb{R}^D$, we often omit it in the notation.

One can introduce a parameter, $t \in \mathbb{R}_+$, to define the *scaled metric space* (tX, d_t), often denoted by tX. This is the metric with the same points as X and metric $d_t(x, y) \coloneqq t \cdot d(x, y)$. The *magnitude function* assigns each finite metric space X to a family of scaled metric spaces $\{tX\}_{t>0}$.

Definition 2 (Magnitude function). The *magnitude function* of a metric space X is given by

$$\mathrm{Mag}_X(t) = Mag(tX), \tag{2}$$

with the associated weighting vector denoted by $\mathbf{w}_X^t$.

3 Magnitude Distance

In the literature, a metric space is understood to be a *set* of distinct points—i.e., without duplicates. In the following we extend the notion of magnitude to finite *collections* of points that may contain duplicates. In this case, while the weighting vector may not be unique, the magnitude remains well-defined.

Theorem 3. *Let X be a finite collection of points in Euclidean space $\mathbb{R}^D$, and define the similarity matrix as before. Then:*

1. *If the elements of X are distinct, then the similarity matrix is symmetric positive definite. Therefore, the inverse exists, and the weighting vector and magnitude are uniquely well-defined. [8, Theorem 2.5.3].*
2. *If X contains duplicate points, the similarity matrix is symmetric positive semidefinite (but not definite). However, the magnitude of X is equal to that of X', where X' is the set of distinct elements in X. Also, while the weighting vector is not unique, all valid weightings for X can be defined by distributing each $\mathbf{w}_{X'}(i)$ among the duplicates of x_i (so that the coefficients sum to 1) where $\mathbf{w}_{X'}$ is the unique weighting on X'.*

Theorem 3 tells us magnitude is insensitive to redundancy. We therefore let $X, Y \subset \mathbb{R}^D$ be finite sets of points in Euclidean space, and the magnitude is computed disregarding sample redundancy within the sets. With this in place, we define the *magnitude distance*, which captures the dissimilarity between X and Y using magnitudes at a given scale, t.

Definition 4 (Magnitude Distance). For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d_{Mag}^t(X, Y) = 2\,\mathrm{Mag}_{X \cup Y}(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t), \tag{3}$$

and the normalized magnitude distance is defined as $\tilde{d}_{Mag}^t(X, Y) = \frac{d_{Mag}^t(X,Y)}{\mathrm{Mag}_{X \cup Y}(t)}$.

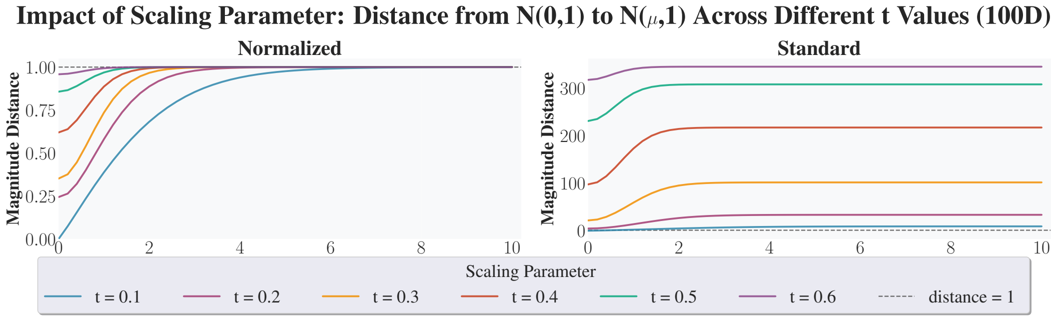

Figure 1: Magnitude distance from 200 samples of $N(0,1)$ to 200 samples of $N(x,1)$ with $t = 0.1$ (blue), 0.2 (purple), 0.3 (yellow), 0.4 (red), 0.5 (green), 0.6 (dark purple). The right plot displays the standard magnitude distance, while the left plot shows the normalized version. As the mean difference between samplings increases, both magnitude distances also increase. However, the standard magnitude distance converges toward different values depending on t. By Theorem 7, these limits are also bounded above by the cardinality of the symmetric difference of the samples, and the normalized magnitude distance consistently converges to 1, regardless of t.

The proposed magnitude distance depends on the scaling parameter, t, that controls the sensitivity of d^t_{Mag} to the separation between points. Intuitively, small t focuses on the structure of the entire data, while large t places more emphasis on the local differences and sample variability. Figure 1 illustrates the impact of the scaling parameter t in 100 dimensional space.

4 Properties

4.1 Metric Axioms

We now examine whether the magnitude distance satisfies the standard axioms of a metric. Before analyzing the axioms, we need to introduce the notion of *magnitude equivalency*.

Definition 5 (Magnitude Equivalence)**.** Let $X, Y \subset \mathbb{R}^D$ be finite sets with weighting vectors $\mathbf{w}^t_X$ and $\mathbf{w}^t_Y$ at scale t. The sets X and Y are *magnitude-equivalent* at scale t if they have the same support of non-zero weight entries,

$$X \underset{\mathrm{Mag}(t)}{=} Y \quad \iff \quad \{x \in X : \mathbf{w}^t_X(x) \neq 0\} = \{y \in Y : \mathbf{w}^t_Y(y) \neq 0\}. \tag{4}$$

Additional details on the properties of magnitude equivalence are provided in the Appendix B. Our main result regarding the metric properties of the magnitude distance is given below.

Theorem 6. *Magnitude distance d^t_{Mag} satisfies the following properties for $X, Y \subset \mathbb{R}^D$ and $t > 0$:*

- ***Symmetry:*** *$d^t_{Mag}(X,Y) = d^t_{Mag}(Y,X)$ by definition.*
- ***Non-negativity:*** *For any $t > 0$, we have $d^t_{Mag}(X,Y) \geq 0$.*
- ***Identity of indiscernibles:*** *$d^t_{Mag}(X,Y) = 0 \iff X \underset{\mathrm{Mag}(t)}{=} Y$.*
- ***Triangle inequality:*** *d^t_{Mag} does not satisfy the triangle inequality in $\mathbb{R}^D$ for $D > 1$.*

4.2 Magnitude Distance Over Different t Values

By analyzing the behavior of the magnitude distance across different values of the scaling parameter t, we show that it inherits similar properties of the magnitude function [8, Proposition 2.2.6].

Theorem 7. *For every finite metric sets X and Y, the magnitude distance $d^t_{Mag}(X)$:*

- *Converges to 0 as $t \to 0$.*
- *Converges to the cardinality of $X \Delta Y$ as $t \to \infty$.*

- *For $t \gg 0$, the magnitude distance $d^t_{Mag}(X)$ is increasing with respect to t.*

Technically, in Theorem 7, we show that as $t \to 0$, the magnitude distance converges to zero i.e., $d^t_{Mag}(X, Y) \to 0$. Moreover, the lower semicontinuity (with respect to the Gromov-Hausdorff distance) of the magnitude function on finite subsets of Euclidean space ensures that it is also lower semicontinuous [10, 8]. Consequently, for every two finite sets $X, Y \in \mathbb{R}^D$, there exists a sufficiently small value of t for which the magnitude distance is meaningful. This property ensures that d^t_{Mag} remains discriminative even in high-dimensional settings. In contrast, in high-dimensional spaces, classical distances such as Wasserstein distance are known to suffer from the *curse of dimensionality*, meaning that they concentrate around a common value in high-dimensional spaces [17, 18].

4.3 Robustness to Outliers

In this section we show magnitude distance is robust to outliers. Firstly, we prove that under natural assumptions, the distance is bounded.

Theorem 8. *Let $X, Y \subset \mathbb{R}^D$ be finite sets with nonnegative weighting vectors of X, Y, and $X \cup Y$. Then, we have:*

$$0 \leq d^t_{\text{Mag}}(X, Y) \leq 2(|X \cup Y|). \tag{5}$$

where $|X|$ and $|Y|$ denote the number of points in X and Y respectively.

Also, nonnegative weighting vectors are guaranteed in different finite metric spaces, such as all subsets of metric spaces when scaled up sufficiently, i.e., $t \gg 0$, and also $\mathbb{R}$ which this global boundedness exists for any scaling parameter. A direct consequence of Theorem 8 is that the distance's sensitivity to adding or adjusting samples is also bounded. In contrast, most distances, such as the Wasserstein distance, are extremely sensitive to outliers, as their sensitivity to adding noise is not bounded, and a single outlier can arbitrarily increase the distance.

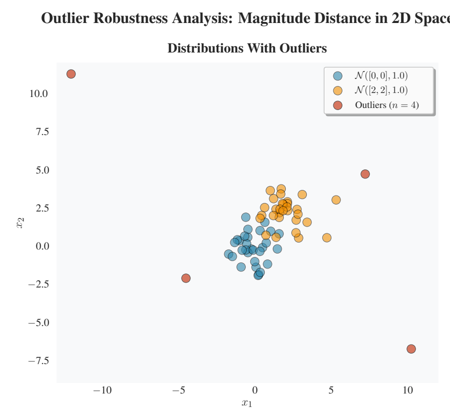

Figure 2: Outlier robustness in 2D with the baseline dataset $B \sim \mathcal{N}([0, 0], 1)$ (blue points), and set $Y \sim \mathcal{N}([2, 2], 1)$ (yellow points), with noisy variant Y^*, incorporating the outliers, Y' (red points).

To demonstrate this outlier robustness, we generate two datasets, B and Y, by sampling from normal distributions with different means. These are represented by blue points and yellow points, respectively. We also generate a third set of points, Y', with much higher dispersion. These datasets are shown in Figure 2. We consider the magnitude distance and Wasserstein distance for two cases: the distance between B and Y, and the distance between B and $Y^* = Y \cup Y'$. The relative change in magnitude distance with $t = 20$ and $t = 5$ are 6.85% and 10.29% respectively, compared to 17.61% for Wasserstein distance. Even for small values of t, where the magnitude distance is more sensitive to local changes, the relative change in magnitude distance due to the addition of outliers remains smaller than that in the Wasserstein distance. Moreover, this relative change decreases as t increases.

5 Conclusion

We have proposed a distance measuring the dissimilarity between finite sets in Euclidean space that we refer to as the magnitude distance. Our distance is based on metric space magnitude, a multi-scale invariant summarizing the geometrical characteristics of a dataset. The magnitude distance inherits these properties, capturing the geometry of data through its scaling parameter t, which controls its sensitivity to local or global data structural variations. Furthermore, we have shown that magnitude distance is robust to outliers. Magnitude distance thus gives a provably outlier-robust and flexible measure suitable for evaluating the performance of generative models and enabling model optimization. We anticipate that this measure will be useful for both training and evaluating generative models. A number of generative modelling frameworks, such as Simulation-Based Inference and Generative Adversarial Networks, require one to compute similarities between finite sets. The magnitude distance represents a flexible alternative to existing approaches like Wasserstein distance and MMD. Our future work will explore the practical application of magnitude distance in the context of improved generative modelling algorithms.

Acknowledgments and Disclosure of Funding

This project was supported by the Royal Academy of Engineering under the Research Fellowship programme.

References

[1] Rayna Andreeva, James Ward, Primoz Skraba, Jie Gao, and Rik Sarkar. Approximating metric magnitude of point sets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 15374–15381, 2025.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[3] Juan Antonio Barceló and Anthony Carbery. On the magnitudes of compact sets in euclidean spaces. *American Journal of Mathematics*, 140(2):449–494, 2018.

[4] Wacha Bounliphone, Eugene Belilovsky, Matthew B Blaschko, Ioannis Antonoglou, and Arthur Gretton. A test of relative similarity for model selection in generative models. *arXiv preprint arXiv:1511.04581*, 2015.

[5] Chad Giusti and Giuliamaria Menara. Eulerian magnitude homology: subgraph structure and random graphs. *arXiv preprint arXiv:2403.09248*, 2024.

[6] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

[7] Ryuki Kaneta and Masahiko Yoshinaga. Magnitude homology of metric spaces and order complexes. *Bulletin of the London Mathematical Society*, 53(3):893–905, 2021. doi: https://doi.org/10.1112/blms.12469. URL `https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/blms.12469`.

[8] Tom Leinster. The magnitude of metric spaces. *Documenta Mathematica*, 18:857–905, 2013.

[9] Tom Leinster. *Entropy and diversity: the axiomatic approach*. Cambridge university press, 2021.

[10] Tom Leinster and Simon Willerton. On the asymptotic magnitude of subsets of euclidean space. *Geometriae Dedicata*, 164(1):287–310, 2013.

[11] Victor M Panaretos and Yoav Zemel. Statistical aspects of wasserstein distances. *Annual review of statistics and its application*, 6(1):405–431, 2019.

[12] Ran Raz. On the complexity of matrix product. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 144–151, 2002.

[13] Ludger Rüschendorf. The wasserstein distance and approximation theorems. *Probability Theory and Related Fields*, 70(1):117–129, 1985.

[14] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

[15] Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

[16] SS Vallender. Calculation of the wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18(4):784–786, 1974.

[17] Michel Verleysen and Damien François. The curse of dimensionality in data mining and time series prediction. In *Proceedings of the 8th International Conference on Artificial Neural Networks: Computational Intelligence and Bioinspired Systems*, IWANN'05, page 758–770, Berlin, Heidelberg, 2005. Springer-Verlag. ISBN 3540262083. doi: 10.1007/11494669_93. URL https://doi.org/10.1007/11494669_93.

[18] Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.

[19] Linqi Zhou, Stefano Ermon, and Jiaming Song. Inductive moment matching. In *Forty-second International Conference on Machine Learning*, 2025.

A Proofs: Sets with Redundant Samples

Proposition (2.4.3, [8])**.** Let X be a positive definite finite metric space with similarity matrix ζ_X. Then

$$Mag(X) \;=\; \sup_{v\neq 0} \frac{\left(\sum_{x\in X} v(x)\right)^2}{v^*\zeta_X v}, \tag{6}$$

where the supremum is taken over all nonzero vectors $v \in \mathbb{R}^X$, and v^* denotes the transpose of v. Moreover, a vector v attains the supremum if and only if it is a nonzero scalar multiple of the unique weighting on X.

Theorem 3. *Let X be a finite collection of points in Euclidean space $\mathbb{R}^D$, and define the similarity matrix as before. Then:*

1. *If the elements of X are distinct, then the similarity matrix is symmetric positive definite. Therefore, the inverse exists, and the weighting vector and magnitude are uniquely well-defined. [8, Theorem 2.5.3].*

2. *If X contains duplicate points, the similarity matrix is symmetric positive semidefinite (but not definite). However, the magnitude of X is equal to that of X', where X' is the set of distinct elements in X. Also, while the weighting vector is not unique, all valid weightings for X can be defined by distributing each $\mathbf{w}_{X'}(i)$ among the duplicates of x_i (so that the coefficients sum to 1) where $\mathbf{w}_{X'}$ is the unique weighting on X'.*

Proof of Theorem 3. The first case follows directly from [8, Theorem 2.5.3], which shows that every finite subspace of Euclidean space is positive definite.

For the second case, let $X' = \{x_1, \cdots, x_n\}$ be the set of distinct elements in X, where each point x_i appears k_i times in X. Then, due to these duplications, the similarity matrix ζ_X is obtained from $\zeta_{X'}$ by repeating rows and columns corresponding to the multiplicities k_i. Therefore, ζ_X has linear dependencies between its rows (and columns) and $\operatorname{rank}(\zeta_X) = \operatorname{rank}(\zeta_{X'}) = n$. Thus, ζ_X is positive semidefinite and not definite.

Let $w_{X'}$ be the unique weighting on X'. We now show a valid weighting w_X for X by distributing each $\mathbf{w}_{X'}(j)$ arbitrarily among the k_j duplicates of x_j. For each x_j, choose $\alpha_{j,1}, \cdots, \alpha_{j,k_j}$ such that $\sum_{l=1}^{k_j} \alpha_{j,l} = 1$. Then, define the weights of the duplicates $x_j^{(l)} \in X$ by

$$w_X(x_j^{(l)}) \coloneqq \alpha_{j,l}\, w_{X'}(x_j), \qquad l = 1, \ldots, k_j$$

Now, we prove that w_X is a valid weighting. For every $y \in X$ with representative $x_i \in X'$ we have:

$$\sum_{y' \in X} \zeta_X(y, y')\, w_X(y) = \sum_{x_j \in X'} \sum_{l=1}^{k_j} \zeta_{X'}(x_i, x_j)\, \alpha_{j,l} w_{X'}(x_j)$$
$$= \sum_{j \in X'} \zeta_{X'}(x_i, x_j) w_{X'}(x_j) \underbrace{\sum_{l=1}^{k_j} \alpha_{j,l}}_{=1}$$
$$= \sum_{j \in X'} \zeta_{X'}(x_i, x_j) w_{X'}(x_j) \underset{(a)}{=} 1.$$

where in (a) we use the fact that $w_{X'}$ is a valid weighting for X'.

Moreover, we prove these are the only valid weightings for X by assuming the opposite. That is, assume the weights of the duplicates $x_j^{(l)} \in X$ can be written as

$$w_X(x_j^{(l)}) := \alpha_{j,l}\, w_{X'}(x_j), \qquad l = 1, \ldots, k_j$$

where there exists j such that $\sum_{l=1}^{k_j} \alpha_{j,l} \neq 1$. In this case, following the same approach, for every $y \in X$ with representative $x_i \in X'$ we have:

$$1 = \sum_{y' \in X} \zeta_X(y, y')\, w_X(y) = \sum_{x_j \in X'} \sum_{l=1}^{k_j} \zeta_{X'}(x_i, x_j)\, \alpha_{j,l} w_{X'}(x_j)$$
$$= \sum_{j \in X'} \zeta_{X'}(x_i, x_j) \underbrace{w_{X'}(x_j) \sum_{l=1}^{k_j} \alpha_{j,l}}_{(b)}$$

then, for every j, the coefficient $w_{X'}(x_j) \sum_{l=1}^{k_j} \alpha_{j,l}$ should be a valid weighting for X'. But this contradicts the uniqueness of $w_{X'}$. Therefore, all valid weightings for X can be defined by distributing each $\mathbf{w}_{X'}(i)$ among the duplicates of x_i, so that the coefficients sum to 1.

Finally, considering the magnitude defined as the sum of the entries of the weighting vector:

$$Mag(X) = \sum_{y \in X} w_X(y) = \sum_{x_j \in X'} \sum_{l=1}^{k_j} \alpha_{j,l} w_{X'}(x_j)$$
$$= \sum_{j \in X'} w_{X'}(x_j) \underbrace{\sum_{l=1}^{k_j} \alpha_{j,l}}_{=1}$$
$$= \sum_{j \in X'} w_{X'}(x_j) = Mag(X').$$

□

Corollary 9 (Non-Negative Weighting)**.** *Let X be a finite collection of points in Euclidean space $\mathbb{R}^D$, and let X' denote the set of distinct elements in X. If the unique weighting vector of X' is non-negative, X also has a non-negative weighting.*

B Proofs: Metric Axioms

Lemma 10. *For any $t > 0$, the magnitude distance satisfies the non-negativity axiom.*

Proof of Lemma 10. Let $X, Y \subset \mathbb{R}^D$ be finite sets, and let $Z = X \cup Y$. By[8, Corollary 2.4.4], the inclusion $X \subseteq Z$ implies $\mathrm{Mag}_Z(t) \geq \mathrm{Mag}_X(t)$, and similarly $\mathrm{Mag}_Z(t) \geq \mathrm{Mag}_Y(t)$. Therefore,

$$d^t_{Mag}(X, Y) = 2\,\mathrm{Mag}_Z(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t) \geq 0.$$

□

Corollary 11. *Let $X, Y \subset \mathbb{R}^D$ be finite sets with weighting vectors $\mathbf{w}_X^t$ and let t be a scaling parameter. The magnitude equivalency is an equivalence relation, satisfying*

- ***Reflexivity:*** $X \underset{\mathrm{Mag}(t)}{=} X$,
- ***Symmetry:*** *If* $X \underset{\mathrm{Mag}(t)}{=} Y$, *then* $Y \underset{\mathrm{Mag}(t)}{=} X$.
- ***Transitivity:*** *If* $X \underset{\mathrm{Mag}(t)}{=} Y$, *and* $Y \underset{\mathrm{Mag}(t)}{=} Z$, *then* $X \underset{\mathrm{Mag}(t)}{=} Z$.

Proof of Corollary 11. Let $\mathbf{w}_X^t$, $\mathbf{w}_Y^t$ and $\mathbf{w}_Z^t$ be weighting vectors at scale t for X, Y, and Z respectively. The first two properties are trivially true by definition. If $X \underset{\mathrm{Mag}(t)}{=} Y$, and $Y \underset{\mathrm{Mag}(t)}{=} Z$, then

$$\{x \in X : \mathbf{w}_X^t(x) \neq 0\} = \{y \in Y : \mathbf{w}_Y^t(y) \neq 0\},$$
$$\text{and } \{y \in Y : \mathbf{w}_Y^t(y) \neq 0\} = \{z \in Z : \mathbf{w}_Z^t(z) \neq 0\}.$$

Thus,

$$\{x \in X : \mathbf{w}_X^t(x) \neq 0\} = \{z \in Z : \mathbf{w}_Z^t(z) \neq 0\}.$$

□

Lemma 12. *Let $X, Y \subset \mathbb{R}^D$ be finite sets, and let $\mathbf{w}_X^t$, and $\mathbf{w}_Y^t$ denote the weighting vectors of X, and Y at scale t, respectively. Then,*

$$X \underset{\mathrm{Mag}(t)}{=} Y \quad \iff \begin{cases} \mathbf{w}_X^t(z) = \mathbf{w}_Y^t(z) & \text{if } z \in Q, \\ \mathbf{w}_X^t(z) = 0 & \text{if } z \in X \setminus Q, \\ \mathbf{w}_Y^t(z) = 0 & \text{if } z \in Y \setminus Q. \end{cases}$$

where $Q = X \cap Y$.

Proof of Lemma 12. We only prove the forward direction, as the converse direction is trivial. Let $Q = \{x \in X : \mathbf{w}_X^t(x) \neq 0\} = \{y \in Y : \mathbf{w}_Y^t(y) \neq 0\}$. Build vector $\mathbf{u}_X$ and $\mathbf{u}_Y$ such that for every $q \in Q$ we have $\mathbf{u}_X(q) = \mathbf{w}_X^t(q)$ and $\mathbf{u}_Y(q) = \mathbf{w}_Y^t(q)$. We know $\mathbf{w}_X^t(x) = 0$ for every $x \in X \setminus Q$ and by definition of weighting we have $1 = \zeta_{tX}\mathbf{w}_X^t = \zeta_{tQ}\mathbf{u}_X$ for $q \in Q$. Similarly, we have $1 = \zeta_{tQ}\mathbf{u}_Y$. Therefore, both $\mathbf{u}_X$ and $\mathbf{u}_Y$ are valid weighting for Q. By uniqueness of weightings, we have $\mathbf{u}_X = \mathbf{u}_Y$. Note that $Q \subseteq X \cap Y$ and weights outside Q vanish. □

Lemma 13. *Let $X, Z \subset \mathbb{R}^D$ be finite sets which $X \subseteq Z$. Then,* $\mathrm{Mag}_Z(t) = \mathrm{Mag}_X(t)$ *if and only if* $X \underset{\mathrm{Mag}(t)}{=} Z$.

Proof of Lemma 13. By [8, Proposition 2.4.3] we have:

$$Mag(X) = \sup_{v \neq 0} \frac{\left(\sum_{x \in X} v(x)\right)^2}{v^* \zeta_{tX} v},$$
$$Mag(Z) = \sup_{u \neq 0} \frac{\left(\sum_{z \in Z} u(z)\right)^2}{u^* \zeta_{tZ} u}. \tag{7}$$

First, we prove the forward direction and assume $Mag(X) = Mag(Z)$. Define a vector u on Z by extending the weighting $\mathbf{w}_X^t$ of X:

$$\mathbf{u} = \begin{cases} \mathbf{w}_X^t(x) & \text{if } z \in X, \\ 0 & \text{if } z \in Z \setminus X. \end{cases}$$

By definition of $\mathbf{u}$ and considering the magnitude as the sum of the weighting entries, we have:

$$\begin{aligned} \mathbf{u}^* \zeta_{tZ} \mathbf{u} &= \sum_{z_1, z_2 \in Z} \mathbf{u}(z_1) e^{-td(z_1, z_2)} \mathbf{u}(z_2) \\ &= \sum_{x_1, x_2 \in X} \mathbf{w}^t(x_1) e^{-td(x_1, x_2)} \mathbf{w}^t(x_2) \\ &= \mathbf{w}_X^{t\,*} \zeta_{tX} \mathbf{w}_X^t = \mathrm{Mag}_X(t), \end{aligned} \tag{8}$$

and

$$\sum_{z\in Z}\mathbf{u}(z)=\sum_{x\in X}\mathbf{w}_X^t(x)=\mathrm{Mag}_X\,(t)\,. \tag{9}$$

Therefore, $\mathbf{u}$ is the weighting on Z and by substituting the equation equation (7) with equation (8), equation (9), we have:

$$\frac{\left(\sum_{z\in Z}\mathbf{u}(z)\right)^2}{\mathbf{u}^*\zeta_{tZ}\mathbf{u}}=\frac{\mathrm{Mag}_X\,(t)^2}{\mathrm{Mag}_X\,(t)}=\mathrm{Mag}_X\,(t)\,. \tag{10}$$

Therefore, by considering equation (7), and equation (10) we have $\mathrm{Mag}_Z\,(t) = \sup_{u\neq 0}\frac{\left(\sum_{z\in Z}u(z)\right)^2}{u^*\zeta_{tZ}u}\geq \mathrm{Mag}_X\,(t)$. Then, based on our assumption i.e., $\mathrm{Mag}_X\,(t)=\mathrm{Mag}_Z\,(t)$, the supremum is achieved at $\mathbf{u}$. By [8, Proposition 2.4.3] the weighting vector $\mathbf{u}$ must be scalar multiple of the unique weighting $\mathbf{w}_Z^t$ on Z, i.e. for some $c\neq 0$ we have $\mathbf{u}=c\mathbf{w}_Z^t$. Then, by definition of $\mathbf{u}$ we have $c\mathbf{w}_Z^t(z)=\mathbf{u}(z)=0$ for $z\in Z\setminus X$ and $c\mathbf{w}_Z^t(z)=\mathbf{u}(z)$ for $z\in X$, implying $X \underset{\mathrm{Mag}(t)}{=} Z$.

While the converse direction is a direct conclusion of Lemma 12, in the following we explain another proof. For proving the converse direction, assume $X \underset{\mathrm{Mag}(t)}{=} Z$. Then for every $x\in Z\setminus X$, we have $\mathbf{w}_Z^t(z)=0$ where $\mathbf{w}_Z^t$ is the weighting vector of Z. Let $\mathbf{u}$ be a vector which $\mathbf{u}(x)=\mathbf{w}_Z^t(x)$ for every $x\in X$. Now by definition of weighting vector and $\mathbf{u}$ for any $x\in X$ we have $1=\zeta_{tZ}\mathbf{w}_Z^t=\zeta_{tX}\mathbf{u}$ implying that $\mathbf{u}$ is a valid weighting for X. This completes the proof as $\mathbf{w}_Z^t(z)=0$ for every $x\in Z\setminus X$, and

$$\mathrm{Mag}_X\,(t)=\sum_{x\in X}\mathbf{u}(x)=\sum_{x\in X}\mathbf{w}_Z^t(x)=\sum_{z\in Z}\mathbf{w}_Z^t(z)=\mathrm{Mag}_Z\,(t)\,.$$

□

Lemma 14. *Magnitude distance satisfies the identity of indiscernibles, with respect to the magnitude equivalency:*

$$d^t_{Mag}(X,Y)=0 \iff X \underset{\mathrm{Mag}(t)}{=} Y. \tag{11}$$

Proof of Lemma 14. By definition of the magnitude distance equation (3), we have $d^t_{Mag}(X,Y)=0$ if and only if $2\,\mathrm{Mag}_{X\cup Y}\,(t)=\mathrm{Mag}_X\,(t)+\mathrm{Mag}_Y\,(t)$. Let $Z=X\cup Y$. Then, by [8, Corollary 2.4.4], we know that $\mathrm{Mag}_Z\,(t)\geq\mathrm{Mag}_X\,(t)$ and $\mathrm{Mag}_Z\,(t)\geq\mathrm{Mag}_Y\,(t)$. Therefore, equality $2\,\mathrm{Mag}_{X\cup Y}\,(t)=\mathrm{Mag}_X\,(t)+\mathrm{Mag}_Y\,(t)$ holds if and only if both inequalities are tight i.e. $\mathrm{Mag}_Z\,(t)=\mathrm{Mag}_X\,(t)$ and $\mathrm{Mag}_Z\,(t)=\mathrm{Mag}_Y\,(t)$. By lemma 13, statements $\mathrm{Mag}_Z\,(t)=\mathrm{Mag}_X\,(t)$ and $X \underset{\mathrm{Mag}(t)}{=} Z$, and also statements $\mathrm{Mag}_Z\,(t)=\mathrm{Mag}_Y\,(t)$ and $Y \underset{\mathrm{Mag}(t)}{=} Z$ are equivalent. Thus,

$$d^t_{Mag}(X,Y)=0 \iff X \underset{\mathrm{Mag}(t)}{=} Z \text{ and } Y \underset{\mathrm{Mag}(t)}{=} Z.$$

Finally, Lemma 12 implies that whenever $X \underset{\mathrm{Mag}(t)}{=} Y$, then we have $X \underset{\mathrm{Mag}(t)}{=} Y \underset{\mathrm{Mag}(t)}{=} Z$. The converse direction also follows from the transitivity of magnitude equivalence. □

Lemma 15. *Magnitude distance does not satisfy the triangle inequality in general. Moreover, the triangle inequality for the magnitude distance is equivalent to the submodularity of magnitude.*

Proof of Theorem 15. Magnitude distance satisfies the triangle inequality if for every three finite sets of samples $X,Y,Z\subset\mathbb{R}^D$, if and only if we have:

$$d^t_{Mag}(X,Y)+d^t_{Mag}(Y,Z)\geq d^t_{Mag}(X,Z).$$

Substituting the definition of magnitude distance equation (3) in this inequality is equivalent to $\mathrm{Mag}_{X\cup Y}\,(t)+\mathrm{Mag}_{Y\cup Z}\,(t)-\mathrm{Mag}_Y\,(t)\geq\mathrm{Mag}_{X\cup Z}\,(t)$, which can be written as:

$$\mathrm{Mag}_{X\cup Y}\,(t)+\mathrm{Mag}_{Y\cup Z}\,(t)\geq\mathrm{Mag}_{X\cup Z}\,(t)+\mathrm{Mag}_Y\,(t)\,. \tag{12}$$

If we take $Y = X \cap Z$ then equation (12) becomes the submodularity inequality, i.e. $\mathrm{Mag}_X(t) + \mathrm{Mag}_Z(t) \geq \mathrm{Mag}_{X\cup Z}(t) + \mathrm{Mag}_{X\cap Z}(t)$. Conversely, define $X' = X \cup Y$ and $Z' = Z \cup Y$, then the submodularity equation for X' and Z' will be $\mathrm{Mag}_{X'}(t) + \mathrm{Mag}_{Z'}(t) \geq \mathrm{Mag}_{X'\cup Z'}(t) + \mathrm{Mag}_{X'\cap Z'}(t)$ which is

$$\mathrm{Mag}_{X\cup Y}(t) + \mathrm{Mag}_{Z\cup Y}(t) \geq \mathrm{Mag}_{(X\cup Y)\cup(Z\cup Y)}(t) + \mathrm{Mag}_{(X\cup Y)\cap(Z\cup Y)}(t).$$

Equivalently, $\mathrm{Mag}_{X\cup Y}(t) + \mathrm{Mag}_{Z\cup Y}(t) \geq \mathrm{Mag}_{X\cup Y\cup Z}(t) + \mathrm{Mag}_{Y\cup(X\cap Z)}(t)$ which by [8, Lemma 2.4.12] we have $\mathrm{Mag}_{X\cup Y\cup Z}(t) \geq \mathrm{Mag}_{X\cup Z}(t)$ and $\mathrm{Mag}_{Y\cup(X\cap Z)}(t) \geq \mathrm{Mag}_Y(t)$. This implies

$$\mathrm{Mag}_{X\cup Y}(t) + \mathrm{Mag}_{Z\cup Y}(t) \geq \mathrm{Mag}_{X\cup Y\cup Z}(t) + \mathrm{Mag}_{Y\cup(X\cap Z)}(t)$$

which is the equation (12). Therefore, the triangle inequality for the magnitude distance is equivalent to the submodularity of magnitude. Submodularity is known not to hold for magnitude in general [1]. Now, we show a counter-example that equation (12) does not hold in general using [1, Theorem 2]. Let $Y = \emptyset$ and $Z = \{0\}$. Then $\mathrm{Mag}_Y(t) = 0$, $\mathrm{Mag}_Z(t) = 1$, and $X \cup Y = X$. Then, X is build on $S = \{e_1, \cdots, e_D\}$ which is the standard basis vectors of $\mathbb{R}^D$ as $X = \{te_1, -te_1, \cdots, te_D, -te_D\}$. For these choices, equation (12) is:

$$\mathrm{Mag}_X(t) + 1 \geq \mathrm{Mag}_{X\cup Z}(t),$$

which this inequality does not hold with the appropriate choice of t and D. For instance, when $t = 5$ and $D = 500$, $\mathrm{Mag}_{X\cup Z}(5) - \mathrm{Mag}_X(5) \approx 7.18$. □

Proposition 16. *Magnitude distance satisfies the triangle inequality in* $\mathbb{R}$*.*

Proof of Proposition 16. By Lemma 15, the triangle inequality for the magnitude distance is equivalent to the submodularity of the magnitude. While submodularity is known not to hold for magnitude in general, it does hold in the one-dimensional case $\mathbb{R}$ [1, Theorem 3]. □

C Proofs: Magnitude Distance Over Different t Values

Proposition (2.2.6, [8])**.** Let X be a finite metric space. Then:

- For $t \gg 0$, the magnitude function of X is increasing with respect to t.
- For $t \gg 0$, there is a unique, positive, weighting on tX.
- Converges to the cardinality of X as $t \to \infty$.

Theorem 7. *For every finite metric sets X and Y, the magnitude distance $d^t_{Mag}(X)$:*

- *Converges to 0 as $t \to 0$.*
- *Converges to the cardinality of $X \Delta Y$ as $t \to \infty$.*
- *For $t \gg 0$, the magnitude distance $d^t_{Mag}(X)$ is increasing with respect to t.*

Proof of Theorem 7. The first two statements follow directly from Proposition . Now we prove the third one as follows. Let $M_{tX} = \zeta_{tX} - I$ be a matrix, where I is the identity matrix. Therefore, M_{tX} has zero diagonal as $\exp(-t\, d(x_i, x_i)) = 1$ for all $x_i \in X$. Also, off-diagonal entries are equal to similarity matrix and for $t \gg 0$, the off-diagonal entries of M_{tX} are exponentially small, i.e., $\exp(-t\, d(x_i, x_j)) = O(\exp(-t d_{\min}(X)))$ where $d_{\min}(X) = \min_{x_i \neq x_j} d(x_i, x_j)$.

By $t \to \infty$, matrix $\|M_{tX}\|$ converges to zero and for sufficiently large t, the the spectral radius of M_{tX} is less than 1. Then, ζ_{tX}^{-1} can be expanded using the converging Neumann series,

$$\zeta_{tX}^{-1} = (I + M_{tX})^{-1} = \sum_{k=0}^{\infty} (-1)^k M_{tX}^k.$$

Now recalling the definition of magnitude, using Neumann series, we have

$$\begin{aligned}\operatorname{Mag}_X(t) = \mathbf{1}^\top \zeta_{tX}^{-1}\mathbf{1} = \sum_{k=0}^{\infty}(-1)^k \mathbf{1}^\top M_{tX}^k \mathbf{1} \\ = |X| - \sum_{x_i \neq x_j, x_i, x_j \in X} e^{-t\,d(x_i,x_j)} + O(e^{-2td_{\min}(X)}).\end{aligned}$$

and the derivative is

$$\frac{d}{dt}\operatorname{Mag}_X(t) = 2\sum_{x_i \neq x_j, x_i, x_j \in X} d(x_i,x_j)e^{-t\,d(x_i,x_j)} + O(e^{-2td_{\min}(X)}).$$

Similarly, defining the magnitude of Y and $X \cup Y$ using Neumann series, the derivative of the magnitude distance can be written as

$$\begin{aligned}\frac{d}{dt}d^t_{\text{Mag}}(X,Y) &= 2\frac{d}{dt}\operatorname{Mag}_Z(t) - \frac{d}{dt}\operatorname{Mag}_X(t) - \frac{d}{dt}\operatorname{Mag}_Y(t) \\ &= 4\sum_{z_i \neq z_j, z_i, z_j \in X\cup Y} d(z_i,z_j)e^{-t\,d(z_i,z_j)} - 2\sum_{x_i \neq x_j, x_i, x_j \in X} d(x_i,x_j)e^{-t\,d(x_i,x_j)} \\ &\quad - 2\sum_{y_i \neq y_j, y_i, y_j \in Y} d(y_i,y_j)e^{-t\,d(y_i,y_j)} + O(e^{-2td_{\min}(X\cup Y)}).\end{aligned}$$

This can be simplified to

$$\begin{aligned}\frac{d}{dt}d^t_{\text{Mag}}(X,Y) &= 4\sum_{x_i \neq y_j, x_i\in X, y_j \in Y} d(x_i,y_j)e^{-t\,d(x_i,y_j)} + 2\sum_{x_i \neq x_j, x_i, x_j \in X} d(x_i,x_j)e^{-t\,d(x_i,x_j)} \\ &\quad + 2\sum_{y_i \neq y_j, y_i, y_j \in Y} d(y_i,y_j)e^{-t\,d(y_i,y_j)} + O(e^{-2td_{\min}(X\cup Y)}),\end{aligned}$$

where each term is nonnegative. This indicates, for sufficiently large t, magnitude distance is positive, up to exponentially small error. □

D Proofs: Outlier Robustness

Theorem 8. *Let $X, Y \subset \mathbb{R}^D$ be finite sets with nonnegative weighting vectors of X, Y, and $X \cup Y$. Then, we have:*

$$0 \leq d^t_{\text{Mag}}(X,Y) \leq 2(|X \cup Y|). \tag{5}$$

where $|X|$ and $|Y|$ denote the number of points in X and Y respectively.

Proof of Theorem8. The proof follows from [8, Lemma 2.4.12], which proves $\operatorname{Mag}_{X\cup Y}(t) < |X \cup Y|$, and from [8, Proposition 2.2.6], which guarantees $\operatorname{Mag}_X(t), \operatorname{Mag}_Y(t) \geq 0$, under the assumption of nonnegative weighting vectors for X, Y, and $X \cup Y$. Note that this assumption is satisfied for sufficiently large t, i.e., $t \gg 0$, by [8, Proposition 2.2.6]. □

E Computational Cost

Magnitude distance computational cost is in the order of magnitude computation in the size of the union of sets. By definition, magnitude computation in metric spaces requires consideration of all pairs of input points and inverting the similarity matrix. For a set of n points, the standard method of computing magnitude requires inverting an $n \times n$ matrix. The best known lower bound for matrix inversion is $\Omega(n^2 \log n)$ [12]. To address this challenge, [1] approximates magnitude by casting it as a convex optimization problem and proposes an iterative algorithm that achieves faster performance than matrix inversion. Further to avoid storing the entire similarity matrices, [1] introduces an approach that selects a smaller subset $S \subset X$ of representative points so that $\operatorname{Mag}_S(t)$ approximates $\operatorname{Mag}_X(t)$.