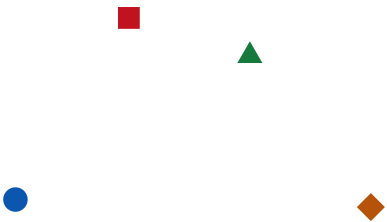
Optimal Average Distortion for Non-Expansive Embeddings of Euclidean Metrics into the Line

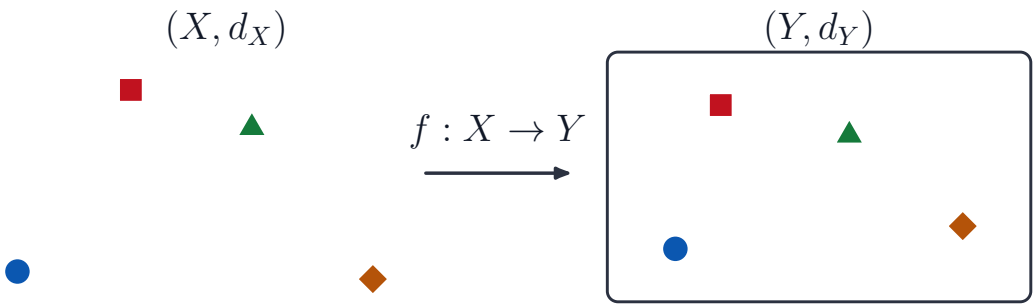
Jack Spalding-Jamieson

Metric embeddings

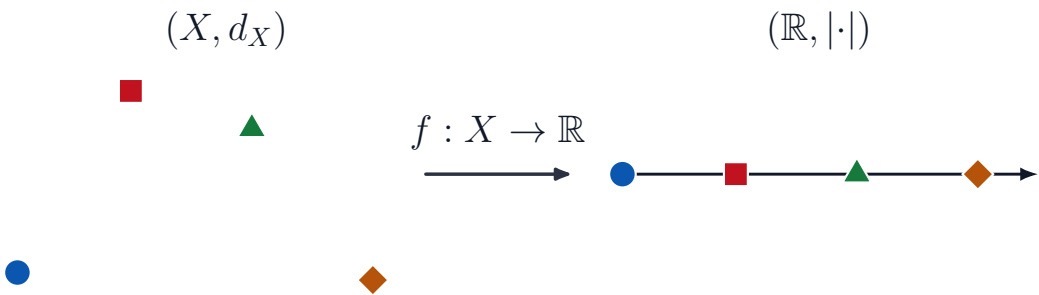
(X, d_X)



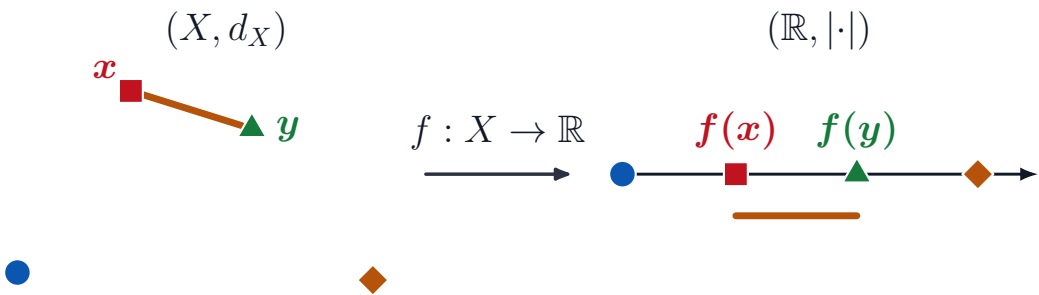
Metric embeddings



Metric embeddings into the line



Non-expansive metric embeddings into the line

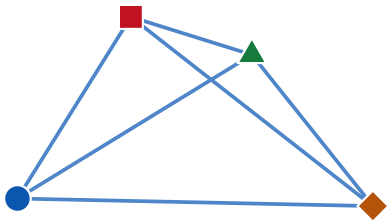


$$|f(x) - f(y)| \leq d_X(x, y) \quad \text{for all } x, y \in X$$

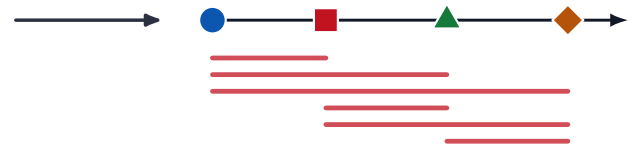
Non-expansive metric embeddings into the line with bounded average distortion

(X, d_X)

$(\mathbb{R}, |\cdot|)$



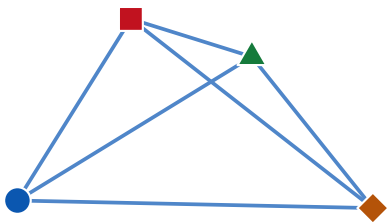
$f : X \to \mathbb{R}$



$$\mathrm{avd}(f) := \frac{\sum_{x,y \in X} d_X(x,y)}{\sum_{x,y \in X} |f(x) - f(y)|}$$

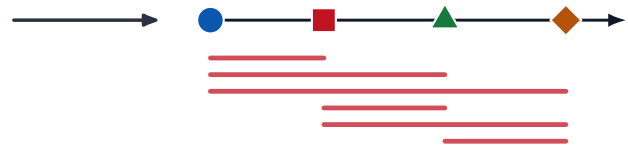
Non-expansive metric embeddings of Euclidean metrics into the line with bounded average distortion

$(X, \|\cdot\|)$



$f : X \to \mathbb{R}$

$(\mathbb{R}, |\cdot|)$



$X \subset \mathbb{R}^k, \ |X| = n$

$$\mathrm{avd}(f) = \frac{\sum_{x,y \in X} \|x - y\|}{\sum_{x,y \in X} |f(x) - f(y)|}$$

Known bounds by metric class

metric class	upper bound	lower bound
general	$O(\log n)$ Bourgain 1985	$\Omega(\log n)$ expanders

Known bounds by metric class

metric class	upper bound	lower bound
general	$O(\log n)$ Bourgain 1985	$\Omega(\log n)$ expanders
K_h-minor-free	$O(\log h)$ CF 2025	$\Omega(\log h)$ expanders
$K_h^{(1)}$-induced-minor-free	$O(\log h)$ DRSW 2027	$\Omega(\log h)$ expanders

Known bounds by metric class

metric class	upper bound	lower bound
general	$O(\log n)$ Bourgain 1985	$\Omega(\log n)$ expanders
K_h-minor-free	$O(\log h)$ CF 2025	$\Omega(\log h)$ expanders
$K_h^{(1)}$-induced-minor-free	$O(\log h)$ DRSW 2027	$\Omega(\log h)$ expanders
squared Euclidean	$O\left(\sqrt{\log n}\right)$ ARV 2004	$\Omega\left(\sqrt{\log n}\right)$ Hamming cube

Known bounds by metric class

metric class	upper bound	lower bound
general	$O(\log n)$ Bourgain 1985	$\Omega(\log n)$ expanders
K_h-minor-free	$O(\log h)$ CF 2025	$\Omega(\log h)$ expanders
$K_h^{(1)}$-induced-minor-free	$O(\log h)$ DRSW 2027	$\Omega(\log h)$ expanders
squared Euclidean	$O(\sqrt{\log n})$ ARV 2004	$\Omega(\sqrt{\log n})$ Hamming cube
Euclidean	$O(\sqrt{\log n})$ MR 2001	$\Omega(\sqrt{\log n / \log\log n})$ MR 2001

Main result

Theorem

For every integer $k \geq 1$, every n-point $X \subset \mathbb{R}^k$ admits a non-expansive $f : X \to \mathbb{R}$ with

$$\operatorname{avd}(f) = O\left(\sqrt{\frac{\log n}{\log \log n}}\right).$$

Clustered or well-spread

$$\Delta := \frac{1}{n^2} \sum_{x,y \in X} \|x - y\|$$

Clustered or well-spread

$$\Delta := \frac{1}{n^2} \sum_{x,y \in X} \|x - y\|$$

Clustered

$|B(x_*, \Delta/4)| \geq n/64$
for some $x_* \in X$



$f(x) := d(x, B(x_*, \Delta/4))$
$\mathrm{avd}(f) \leq 256$

Clustered or well-spread

$$\Delta := \frac{1}{n^2} \sum_{x,y \in X} \|x - y\|$$

Clustered

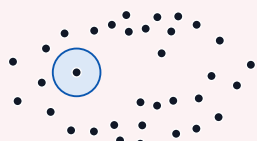
$|B(x_*, \Delta/4)| \geq n/64$
for some $x_* \in X$



$f(x) := d(x, B(x_*, \Delta/4))$
$\mathrm{avd}(f) \leq 256$

Well-spread

$|B(x_*, \Delta/4)| < n/64$
for every $x_* \in X$



Kantorovich-Rubinstein duality

Disjoint $X^-, X^+ \subset X$, $|X^-| = |X^+|$

Kantorovich-Rubinstein duality

Disjoint $X^-, X^+ \subset X, \quad |X^-| = |X^+|$

$$\max_{f \text{ non-expansive}} \left(\sum_{y \in X^+} f(y) - \sum_{x \in X^-} f(x) \right)$$

Kantorovich-Rubinstein duality

Disjoint $X^-, X^+ \subset X$, $|X^-| = |X^+|$

$$\max_{f \text{ non-expansive}} \left(\sum_{y \in X^+} f(y) - \sum_{x \in X^-} f(x) \right) = \min_M w(M)$$

M: perfect matching between X^- and X^+, $w(M) := \sum_{uv \in M} \|u - v\|$

Kantorovich-Rubinstein duality

Disjoint $X^-, X^+ \subset X$, $|X^-| = |X^+|$

$$\max_{f \text{ non-expansive}} \left(\sum_{y \in X^+} f(y) - \sum_{x \in X^-} f(x) \right) = \min_M w(M)$$

M: perfect matching between X^- and X^+, $w(M) := \sum_{uv \in M} \|u - v\|$

$$\operatorname{avd}(f) = \frac{n^2 \Delta}{\sum_{x,y \in X} |f(x) - f(y)|}$$

Kantorovich-Rubinstein duality

Disjoint $X^-, X^+ \subset X$, $|X^-| = |X^+|$

$$\max_{f \text{ non-expansive}} \left(\sum_{y \in X^+} f(y) - \sum_{x \in X^-} f(x) \right) = \min_M w(M)$$

M: perfect matching between X^- and X^+, $w(M) := \sum_{uv \in M} \|u - v\|$

$$\mathrm{avd}(f) = \frac{n^2 \Delta}{\sum_{x,y \in X} |f(x) - f(y)|} \leq \frac{n\Delta}{\sum_{y \in X^+} f(y) - \sum_{x \in X^-} f(x)}$$

Kantorovich-Rubinstein duality

Disjoint $X^-, X^+ \subset X$, $|X^-| = |X^+|$

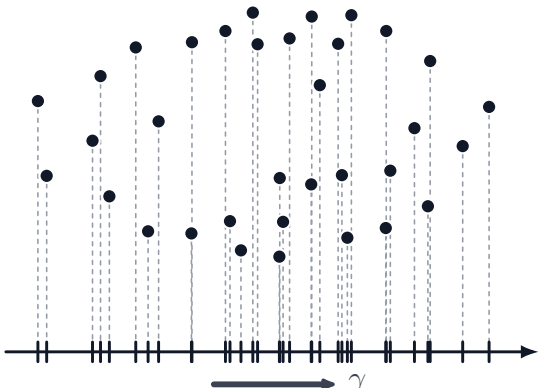
$$\max_{f \text{ non-expansive}} \left(\sum_{y \in X^+} f(y) - \sum_{x \in X^-} f(x) \right) = \min_M w(M)$$

M: perfect matching between X^- and X^+, $w(M) := \sum_{uv \in M} \|u - v\|$

$$\mathrm{avd}(f) = \frac{n^2 \Delta}{\sum_{x,y \in X} |f(x) - f(y)|} \leq \frac{n\Delta}{\sum_{y \in X^+} f(y) - \sum_{x \in X^-} f(x)}$$

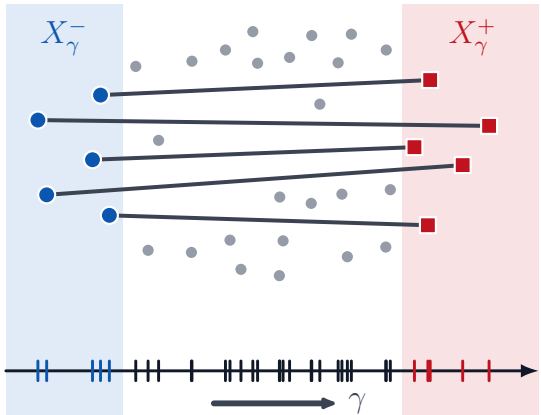
Goal: $|X^{\pm}| = \lceil n/64 \rceil$ and $\min_M w(M) = \Omega\big(n\Delta\sqrt{\log\log n / \log n}\big)$

Random projection



$$\gamma \sim \mathcal{N}(0,1)^k$$
$$x \mapsto \langle \gamma, x \rangle$$

Random projection



$\gamma \sim \mathcal{N}(0,1)^k$

$x \mapsto \langle \gamma, x \rangle$

X_γ^-, X_γ^+: lowest, highest $\lceil n/64 \rceil$ projected points

A_γ: min-weight matching between X_γ^- and X_γ^+

Technical steps

Lemma

For each $g \in \mathbb{R}^k$, let K_g be a symmetric, nonnegative matrix on X with row sums 1. Take independent $x_0 \sim \mathrm{Unif}(X)$ and $\gamma \sim \mathcal{N}(0,1)^k$. Sample $x_1 \sim K_\gamma(x_0, \cdot)$.

$$(\mathbb{E}\,|\langle \gamma, x_1 - x_0 \rangle|)^2 = O\left(\mathbb{E}\|x_1 - x_0\|^2 \cdot \frac{\log n}{\log\log n}\right).$$

Technical steps

Lemma

For each $g \in \mathbb{R}^k$, let K_g be a symmetric, nonnegative matrix on X with row sums 1. Take independent $x_0 \sim \mathrm{Unif}(X)$ and $\gamma \sim \mathcal{N}(0,1)^k$. Sample $x_1 \sim K_\gamma(x_0, \cdot)$.

$$(\mathbb{E}\,|\langle \gamma, x_1 - x_0 \rangle|)^2 = O\left(\mathbb{E}\|x_1 - x_0\|^2 \cdot \frac{\log n}{\log\log n}\right).$$

Corollary

If X is well-spread, then

$$\mathbb{E}_\gamma\, w(A_\gamma) = \Omega\left(n\Delta\sqrt{\frac{\log\log n}{\log n}}\right).$$

Conclusion

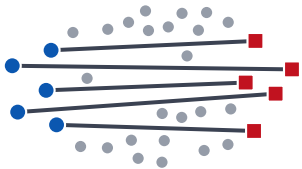
Euclidean metrics: $\mathrm{avd} = O\left(\sqrt{\frac{\log n}{\log \log n}}\right)$

Conclusion

Euclidean metrics: $\mathrm{avd} = O\left(\sqrt{\frac{\log n}{\log\log n}}\right)$



clustered core
distance to the core



well-spread
needs a long matching

Maybe future work:

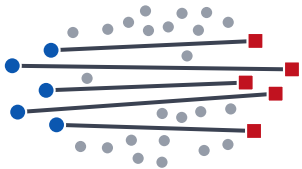
Uniform ARV L_1 embedding?

Conclusion

Euclidean metrics: $\mathrm{avd} = O\left(\sqrt{\frac{\log n}{\log\log n}}\right)$



clustered core
distance to the core



well-spread
needs a long matching

Maybe future work:

Uniform ARV L_1 embedding?

Fin