Optimality of the Johnson-Lindenstrauss Transform for Average Distortion Measures

Yair Bartal
Hebrew University

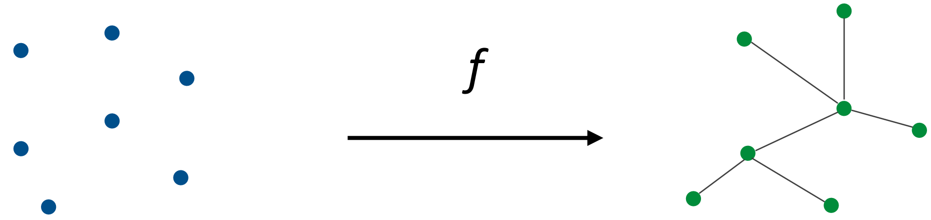
Ora N. Fandina
Århus University

Kasper Green Larsen
Århus University

June 7-10, 2022

SoCG, Berlin

metric embedding



For an input space X find a "nice" Y and $\boldsymbol{f}: \boldsymbol{X} \to \boldsymbol{Y}$ with a small distance error

(worst-case) *distortion* is the smallest stretch $\alpha \geq 1$ s.t. **for all** pairs $u, v \in X$

$$\mathrm{d}_{\mathrm{X}}(\mathrm{u}, \mathrm{v}) \leq c \cdot d_Y\big(f(u), f(v)\big) \leq \alpha \cdot d_X(u, v) \quad \text{(for some const. c>0)}$$

- worst-case guarantee is too strong for practical applications

- preserving **most** of the distances will usually suffice

Average distortion
multiplicative

$$dist_f(u,v) = \max\{ \frac{d_Y(u,v)}{d_X(u,v)}, \frac{d_X(u,v)}{d_Y(u,v)} \}$$

$$\ell_q^{(\Pi)}\text{-}dist(f) = \left(E_{u,v \sim \Pi}\left[dist_f(u,v)^q\right]\right)^{1/q}$$

Average distortion
additive

$$error(u,v) = |d_Y(u,v) - d_X(u,v)|$$

$$Stress_q(f) = \left(\frac{E_\Pi[(error(u,v))^q]}{E_\Pi[(d_X(u,v))^q]}\right)^{1/q}$$

Many heuristics optimizing these notions

Few works with rigorous analysis

Goal Given a finite Euclidean X and a target dim. $k \geq 1$

compute $f: X \to \ell_2^k$ with the smallest worst-case/average distortion

[JL84] Projecting onto a random subspace of dim. $\mathrm{k} = O(\log n / \epsilon^2)$

with const. prob. $1 + \epsilon$ worst-case distortion

[tight, LN17]

[Mat90] Projecting onto $\mathrm{k} \geq 1$ dim. gives worst-case distortion $\tilde{O}\left(n^{\frac{2}{k}}\right)$

[almost tight, Mat90]

[BFN19] Projecting onto a random subspace

$1+\epsilon$ multipl. average distortion $k = O\left(\max\{\frac{1}{\epsilon^2}, \frac{q}{\epsilon}\}\right)$

$O(\epsilon)$ additive average distortion $k = O(q/\epsilon^2)$

$k = \Omega(1/\epsilon)$

This work

The JL transform is optimal for average distortions

- multiplicative $k = \Omega\left(\max\{\frac{1}{\epsilon^2}, \frac{q}{\epsilon}\}\right)$
- additive $k = \Omega(q/\epsilon^2)$

Theorem
q=1

There is a Euclidean space I such that any $f: I \to \ell_2^k$ with $1+\epsilon$ multipl. average dist. must have $k = \Omega(1/\epsilon^2)$

Proof. encoding/volume argument [LN17]

- construct a (large) family P of different pointsets
- If each $I \in P$ embeds with $1+\epsilon$ average dist. into k dims

  using f uniquely encode each I with (few) at most $L(\epsilon, k)$ bits

volume argument: $|P| \leq 2^{L(\epsilon,k)}$

$\log |P| \leq L(\epsilon, k)$

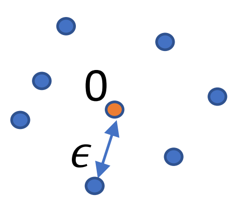
constructing the family P

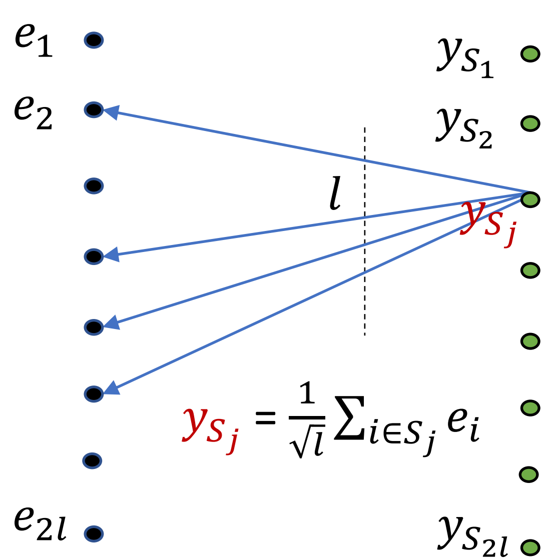
$l := 1/c\epsilon^2$

Each instance **I** in P 6l points in the unit ball in R^{2l}

$|\mathrm{P}| = \binom{2l}{l}^{2l} \gtrsim 2^{2l^2}$

$|O| = 2l$ | E standard basis | $Y\ [S_1, \dots, S_{2l}]$





y_{S_j} can be recovered by

$$< y_{S_j}, \mathrm{e}_i > = \begin{cases} 0, & if\ e_i \notin S_j \\ \frac{1}{\sqrt{l}} = c\epsilon & othrws \end{cases}$$

$\log |P| \leq L(\epsilon, k)$

$2\mathrm{l}^2 \leq L(\epsilon, k)$

encode I with few bits

assume $f: I \to \ell_2^k$ with $1 + \epsilon$ average distortion

Ideally

- f preserves **all** distances up to $1 \pm \epsilon$

all $< f(y_S), f(e_i) > = < y_S, e_i > \pm \Theta(\epsilon)$

- images have $0/1$ coordinates

k bits for a point in I

$6lk$ bits for I

Volume argument $1.8\, l^2 + \Theta\left(l \cdot k\, log\left(\frac{1}{\epsilon} \cdot \frac{1}{\sqrt{k}}\right)\right) \geq 2l^2$

$k \gtrsim l = \Omega(1/\epsilon^2)$

Reality

- f preserves a **special subset** of inner products

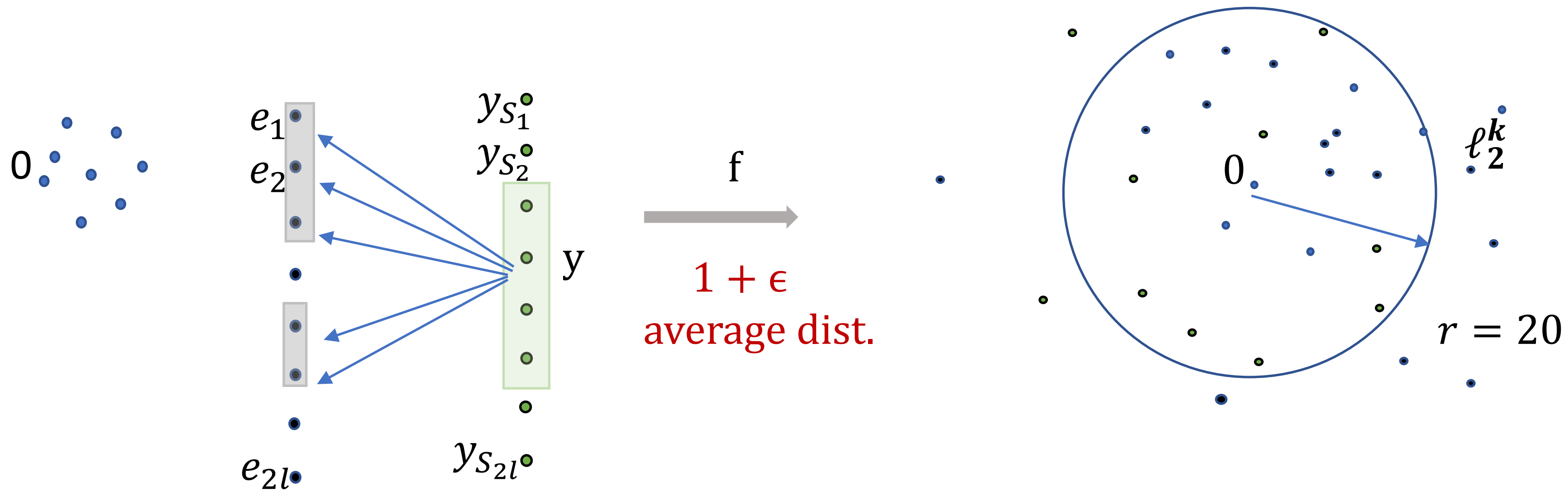
enables to recover entire I

- images can be encoded with

$\approx k\, log\left(\frac{1}{\epsilon} \cdot \frac{1}{\sqrt{k}}\right)$ bits for a point in I

$1.8\, l^2 + \Theta\left(l \cdot k\, log\left(\frac{1}{\epsilon} \cdot \frac{1}{\sqrt{k}}\right)\right)$ bits for I

structural lemma



a special subset

$f(\hat{Y})$ and all $f(E_y) \subset B^k(20)$

a large (const. frac.) $\hat{Y} \subseteq Y$

for all y $\in \hat{Y}$ there is a large (const. frac) $E_y \subseteq E$

$\forall\, e \in E_y \;\; < f(y), f(e) > = < y, e > \pm\, \Theta(\epsilon)$

random shift to ϵ- grid [AK17]

- inner products up to $\pm\Theta(\epsilon)$
- grid size $\cong \left(\frac{\text{const}}{\epsilon\sqrt{k}}\right)^k$

$\Theta\left(k \log\left(\frac{1}{\epsilon} \cdot \frac{1}{\sqrt{k}}\right)\right)$ bits for image

conclusion

The JL transform is also optimal for average distortions

Thanks! Questions?