Efficient Point-to-Subspace Query in ℓ^1 with Application to Robust Face Recognition

Ju Sun, Yuqian Zhang, and John Wright

Department of Electrical Engineering, Columbia University

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Point-to-Subspace Query in ℓ^1

- **Problem Statement**: Given n linear subspaces $\mathcal{S}_1, \dots, \mathcal{S}_n$ of $\mathbb{R}^D$ of dimension r and a query point $\mathbf{q} \in \mathbb{R}^D$, determine the nearest $\mathcal{S}_i$ to $\mathbf{q}$ in ℓ^1 norm.
- **Motivation**:
  - Low-dimensional structures in visual data (e.g., lighting, poses)
  - Structure query as recognition, ℓ^1 for robustness (to, e.g., occlusions, shadows)
  - Efficiency: large D (e.g., # pixels) and large n (many subjects)

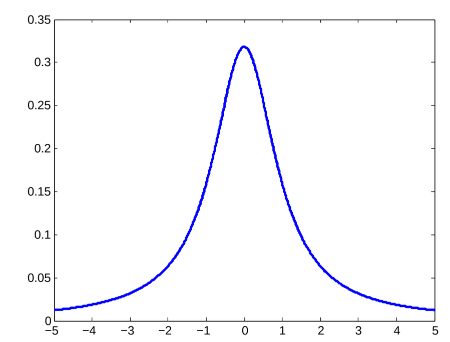
Existing Work

- **Efficiency**: Preprocessing and storage – $\mathrm{lowpoly}(D, r, n)$, Query – $\mathrm{lowpoly}(D, r, n^{o(1)})$

	$r = 0$	$r = 1$	$r \geq 2$
ℓ^2	✓ (e.g., LSH)	✓ (Andoni et al, SODA'09)	? Heuristics (Basri'11, Jain'10)
ℓ^1	✓ (Andoni'05)	? This work addresses D only	

  - ☹ Sublinear time algorithm for point-to-hyperplane in ℓ^2 (and ℓ^1) unlikely (Williams'05)
  - ☹ General low-distortion low-dimensional embedding for ℓ^1 impossible (Brinkman'05)
- **Precursor**☺: For a single subspace, $\ell^1_D \to \ell_1^{O(r \log r)}$ with distortion $O(r \log r)$ (Sohler and Woodruff, 11)
- **Error-Correction in** ℓ^1: dim. reduction determined by density of error e (Candes and Tao, 04)

Our Algorithm and Main Results



- Cauchy distribution: $p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$
- No finite mean or variance
- ℓ^1 Stable: for iid standard Cauchy RV's $\phi_1, \cdots, \phi_k$, $\sum_{i=1}^k \phi_k \sim \|\Phi\|_{\ell^1} x$, for x standard Cauchy.

- **Algorithm**: Generate a random matrix $\mathbf{P} \in \mathbb{R}^{d \times D}$ with iid Cauchy RV's ($d \ll D$)

  Preprocessing: Compute the projections $\mathbf{P}\mathcal{S}_1, \cdots, \mathcal{P}\mathcal{S}_n$
  Test: Compute the projection $\mathbf{Pq}$, and compute its ℓ^1 distance to each of $\mathbf{P}\mathcal{S}_i$

- **Theory**: In short, Cauchy projection with large enough d preserves the identity of nearest subspace with nontrivial probability. In full details

  Suppose we are given n linear subspaces $\{\mathcal{S}_1, \cdots, \mathcal{S}_n\}$ of dimension r in $\mathbb{R}^D$ and any query point $\mathbf{q}$, and that the ℓ^1 distances of $\mathbf{q}$ to each of $\{\mathcal{S}_1, \cdots, \mathcal{S}_n\}$ are $\xi_{1'} \leq \cdots \leq \xi_{n'}$ when arranged in ascending order, with $\xi_{2'}/\xi_{1'} \geq \eta > 1$. For any fixed $\alpha < 1 - 1/\eta$, there exists $d \sim O(r \log n)^{1/\alpha}$ (assuming $n > r$), if $\mathbf{P} \in \mathbb{R}^{d \times D}$ is iid Cauchy, we have

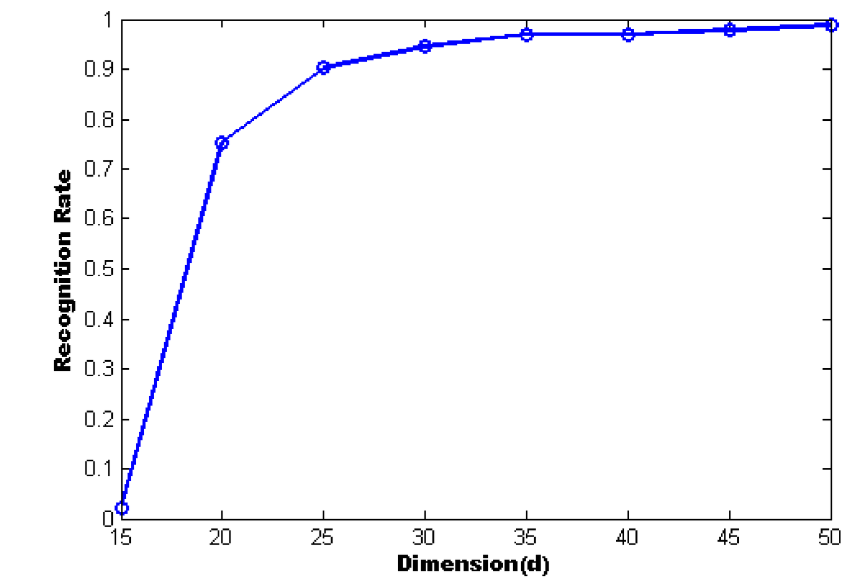
  $$\arg\min_{i \in [n]} d_{\ell^1}(\mathbf{Pq}, \mathbf{P}\mathcal{S}_i) = \arg\min_{i \in [n]} d_{\ell^1}(\mathbf{q}, \mathcal{S}_i)$$

  with (nonzero) constant probability.

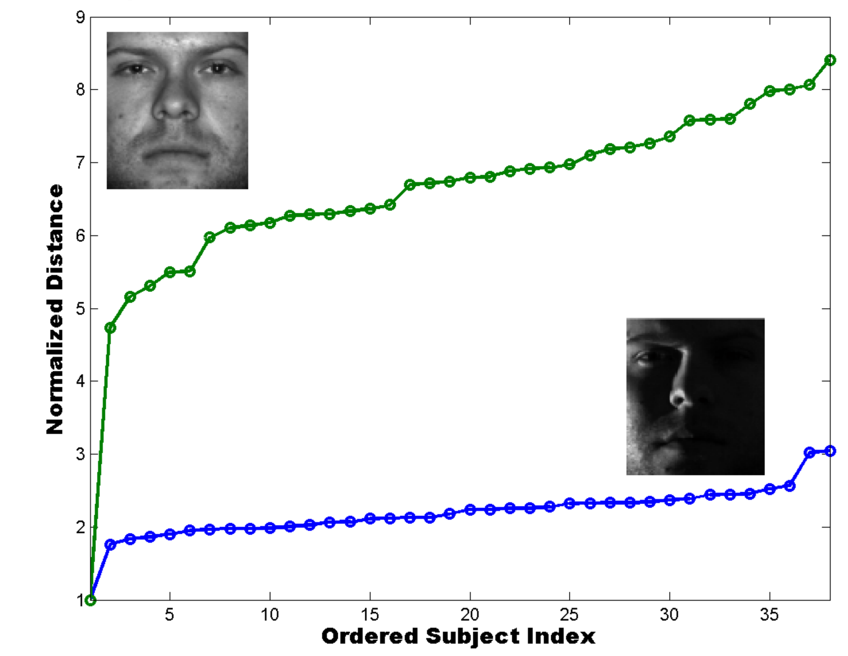
- **Implications**:
  - d depends on the relative gap η, and not on D.
  - d depends on $\log n$ – growing nicely wrt. # subspaces.
  - Independent trials can be taken to amplify the success probability. (Matter of low-dimensional ℓ^1 regressions!)
  - In case of ties, first k nearest neighbors can instead be considered.

Empirical Results

- Extended Yale B Face Data: $D \sim 30,000$, $n = 38$, $d = 9$. Subset under moderate lighting (single projection)



- Normalized distance gap for moderately/extremely illuminated samples.



Sketch of Analysis

Bounded expansion for the good subspace

- Distance after projection can easily be upper bounded:

$$d_{\ell^1}(\mathbf{Pq}, \mathbf{P}\mathcal{S}_\star) = \min_{\mathbf{h} \in \mathbf{P}\mathcal{S}_\star} \|\mathbf{Pq} - \mathbf{h}\|_1 \leq \|\mathbf{Pq} - \mathbf{Pv}_\star\|_1 = \|\mathbf{P}(\mathbf{q} - \mathbf{v}_\star)\|_1.$$

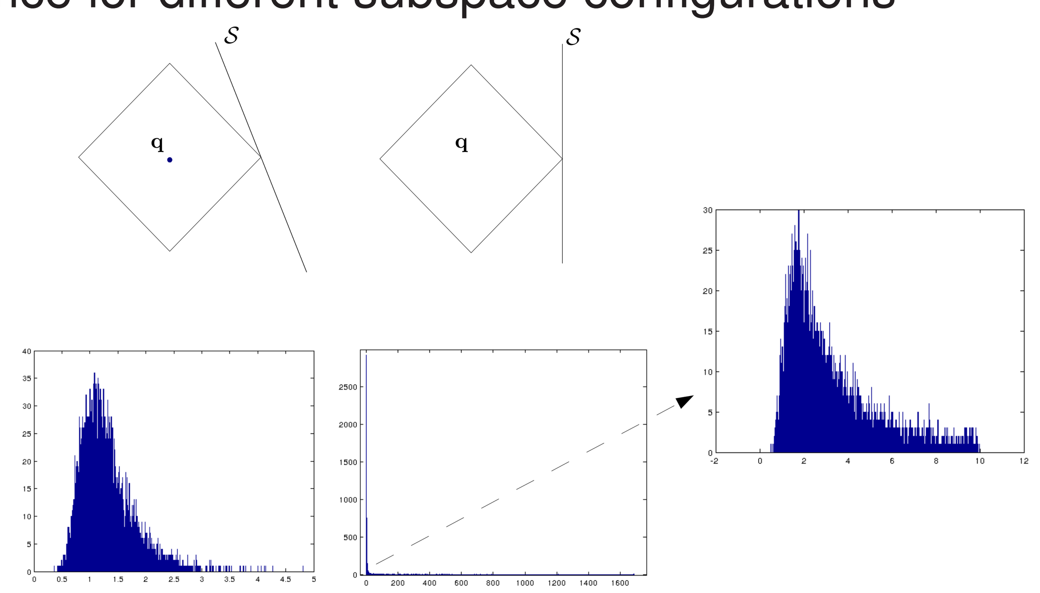
- Bounded expansion with nonzero constant probability:

  There exists numerical constant $c \in (0, 1)$ with the following property. If $\mathbf{w} \in \mathbb{R}^D$ be any fixed vector, and suppose that $\mathbf{P} \in \mathbb{R}^{d \times D}$ is a matrix with iid standard Cauchy entries. Then for any $\rho > 1$,

  $$\mathbb{P}\left[\|\mathbf{Pw}\|_1 > \rho \frac{2}{\pi} d \log d \|\mathbf{w}\|_1\right] < c + \frac{1-c}{\rho} < 1.$$

Idea of Proof

- Behavior of Cauchy projection on point-to-subspace ℓ^1 distance for different subspace configurations



- $\mathbf{P}$ does not increase the distance to "good" subspace too much, and
- $\mathbf{P}$ does not shrink the distances to "bad" subspace too much.

Bounded contraction for the bad subspaces

- Consider $\forall\ \mathbf{w} \in \mathcal{S}_i \oplus \mathbf{q}$, want to show $\|\mathbf{Pw}\|_1 \geq \gamma \|\mathbf{w}\|_1$ for some appropriate γ, then

$$d_{\ell^1}(\mathbf{Pq}, \mathbf{P}\mathcal{S}_i) = \min_{\mathbf{v} \in \mathcal{S}_i} \|\mathbf{Pq} - \mathbf{Pv}\|_1 \geq \min_{\mathbf{v} \in \mathcal{S}_i} \|\mathbf{P}(\mathbf{q} - \mathbf{v})\|_1 \geq \min_{\mathbf{v} \in \mathcal{S}_i} \gamma \|\mathbf{q} - \mathbf{v}\|_1 = \gamma\, d_{\ell^1}(\mathbf{q}, \mathcal{S}_i),$$

.

- Discretization argument on restricted unit ℓ^1 sphere onto the augmented subspaces $\Gamma = \{\mathbf{w} \mid \|\mathbf{w}\|_1 = 1\} \cap \tilde{\mathcal{S}}_i$. Points on the ε-net covered by the concentration results on lower tail.

  Let $\mathbf{P} \in \mathbb{R}^{d \times D}$ be an iid Cauchy matrix. Then for any fixed vector $\mathbf{w} \in \mathbb{R}^D$ and $\alpha, \delta \in (0, 1)$,

  $$\mathbb{P}\left[\|\mathbf{Pw}\|_1 < (1-\alpha)(1-\delta)\frac{2}{\pi} d \log d \|\mathbf{w}\|_1\right] < d^{1-\alpha} \exp\left(-\frac{\delta^2}{2\pi} d^\alpha\right).$$

- Points on Γ but off the ε-net covered by triangular inequalty, which is founded on well-conditioned basis for ℓ^1 subspaces.

Email: <jusun, yuqianzhang, johnwright>@ee.columbia.edu