A Convex Approach for Designing "Good" Linear Embeddings

Chinmay Hegde

Massachusetts
Institute of
Technology






Redundancy in Images

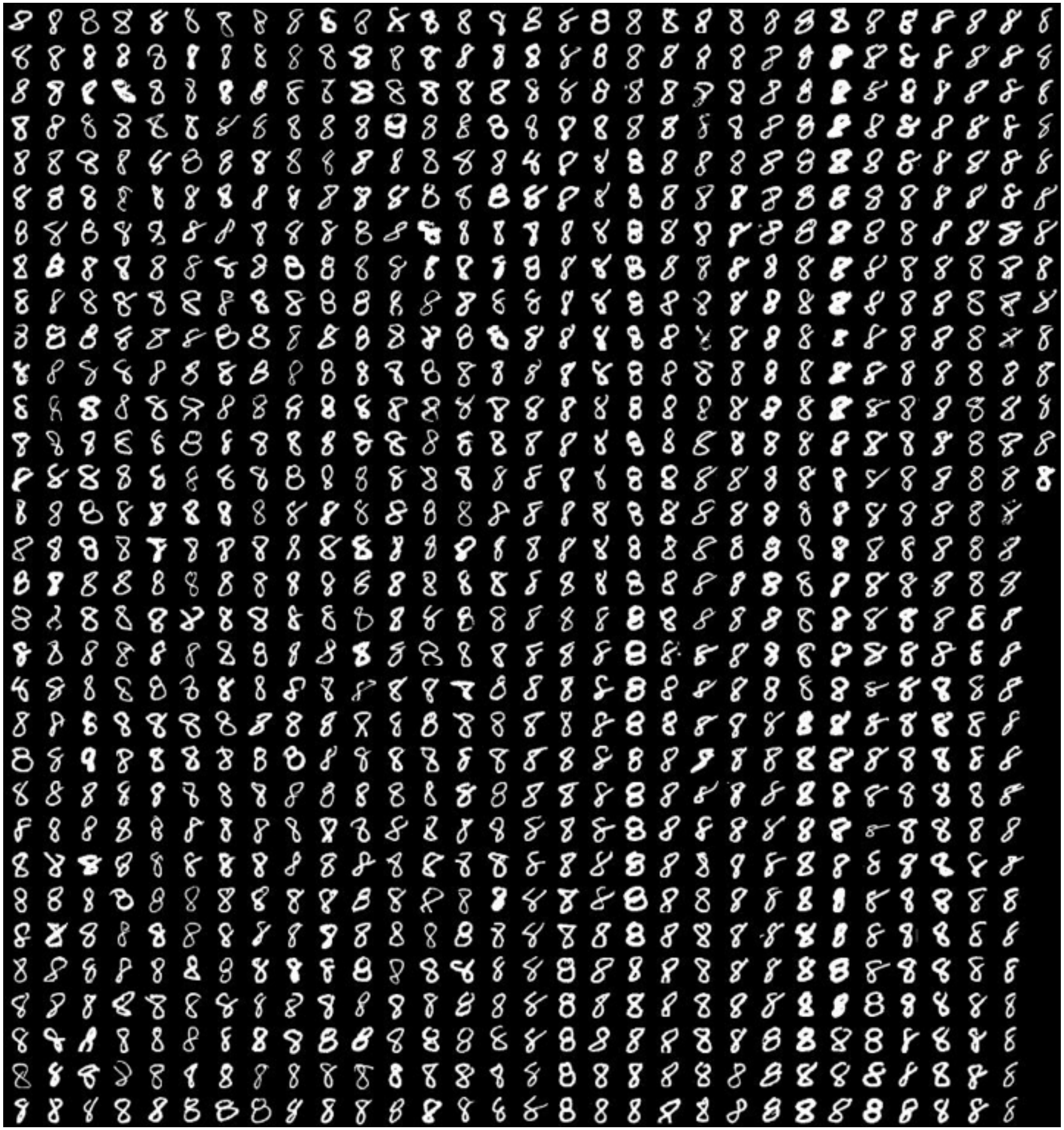
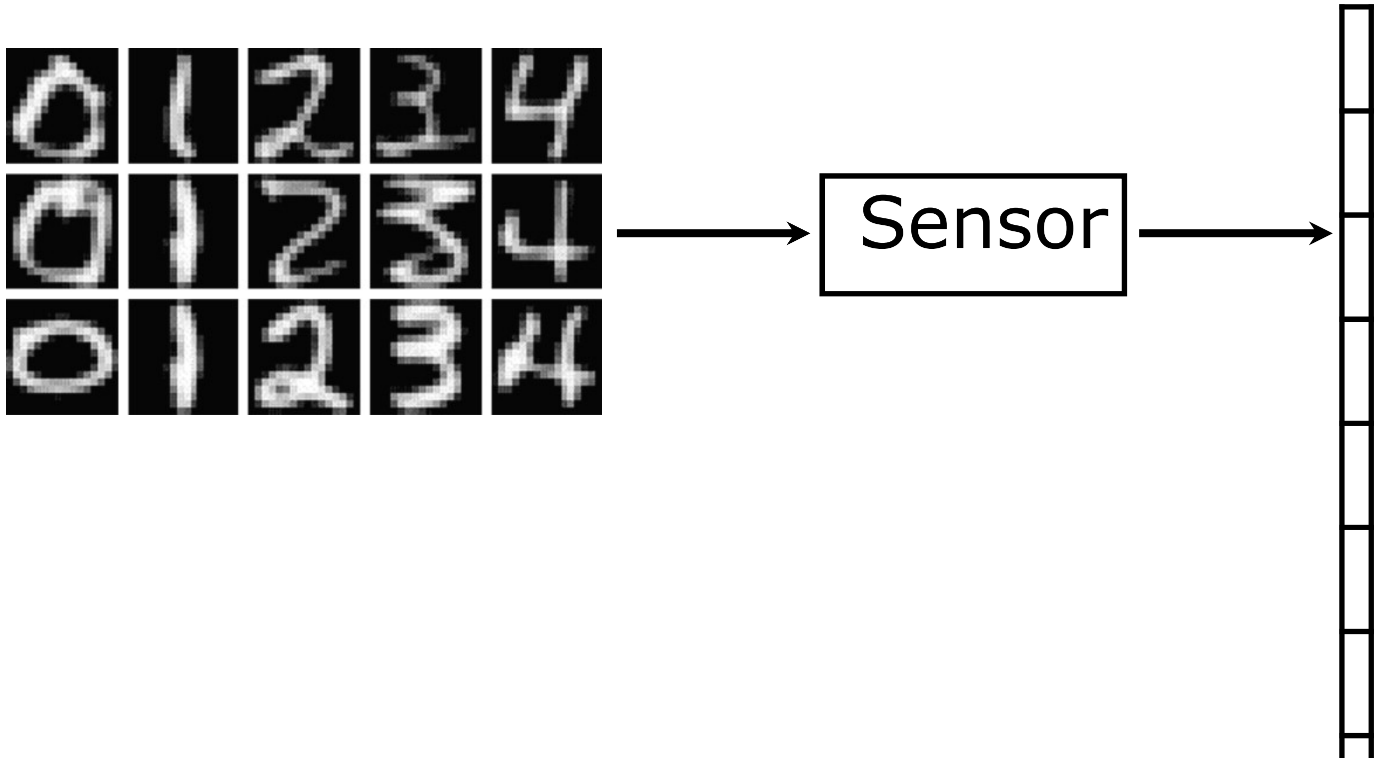


Image Acquisition



- Information content <<< size of signal
- Can we leverage this fact to build an acquisition system that is **specifically tailored** to such signals?

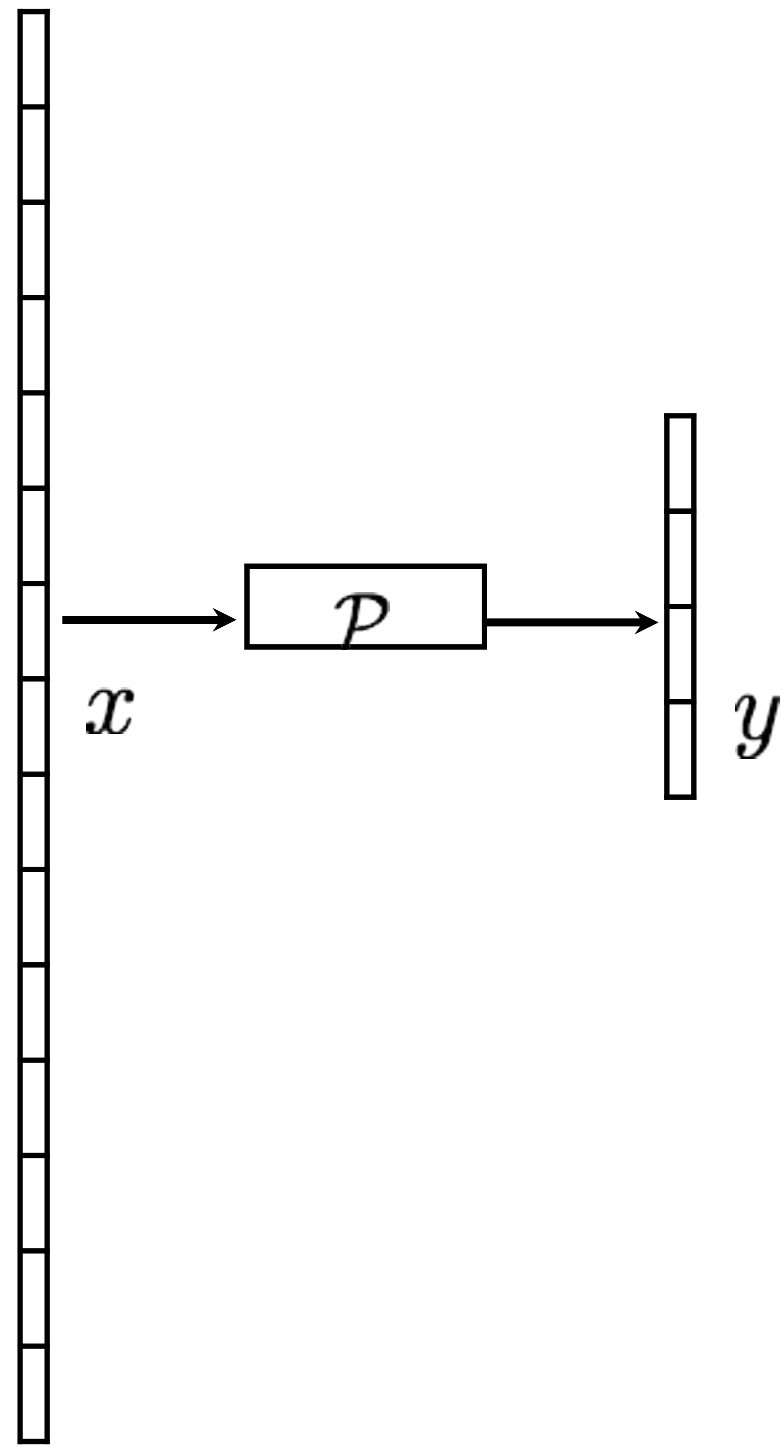
(Linear) Dim. Reduction

- Design criteria
  - ***Linear*** *mapping*
  - ***Reduce*** *dimensionality as much as possible*
  - *Ensure* ***information preservation***

$$\|\mathcal{P}x_1 - \mathcal{P}x_2\| \approx \|x_1 - x_2\|$$

"Near-isometric embedding"

- **Q. Is this possible?**



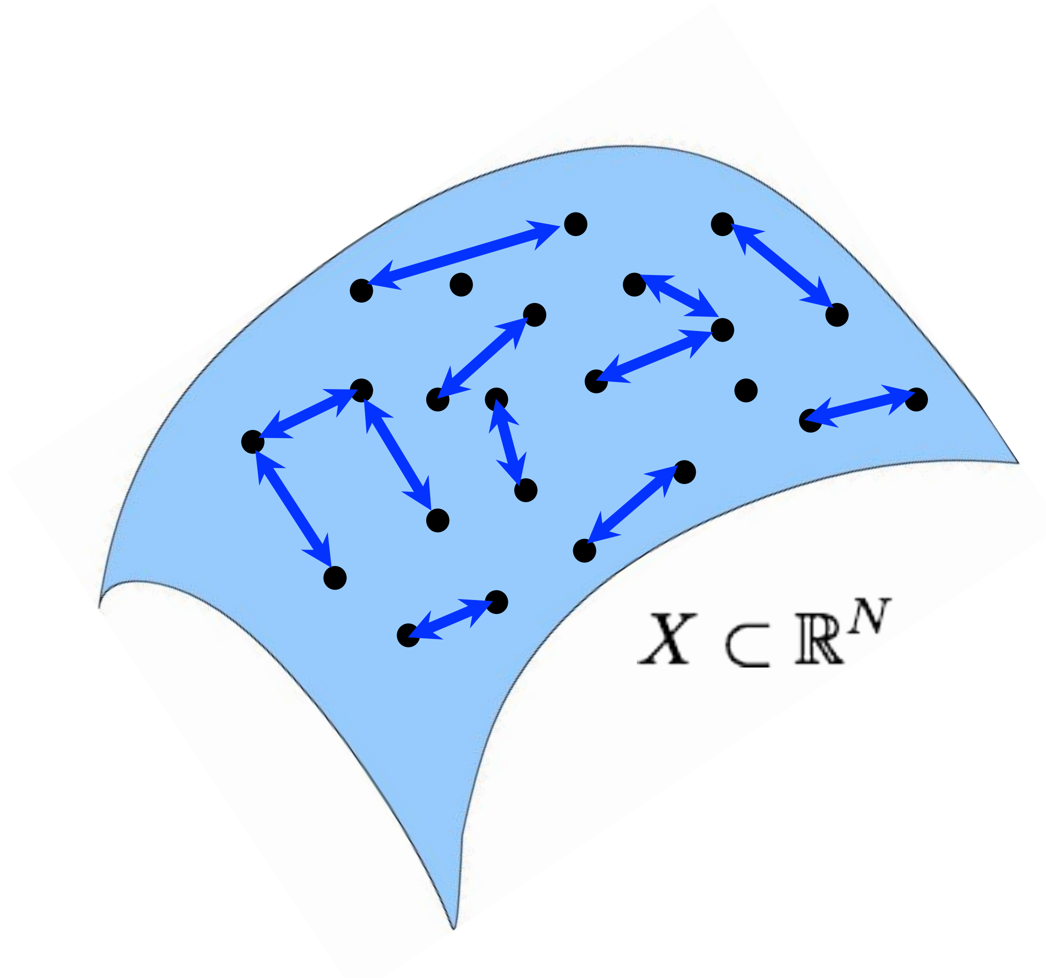
(Linear) Dim. Reduction

- **A. Yes**

[**JL**] Suppose $\#X = Q$

$$M = O\left(\frac{\log(Q)}{\delta^2}\right)$$

$$\Phi \sim \mathrm{subGauss}(M, N)$$



Then, w.h.p.,

$$1 - \delta \leq \frac{\|\Phi(x_1 - x_2)\|^2}{\|x_1 - x_2\|^2} \leq 1 + \delta$$

(Linear) Dim. Reduction

- **Q. Can we beat random projections?**

- **A**. …
  - On the one hand: Lower bounds for JL

$$M \geq \Omega\left(\frac{1}{\delta^2 \log(1/\delta)} \log Q\right)$$

[Alon '03]

(Linear) Dim. Reduction

- **Q. Can we beat random projections?**

- **A**. …
  - On the one hand: Lower bounds for JL

$$M \geq \Omega\left(\frac{1}{\delta^2 \log(1/\delta)} \log Q\right)$$

[Alon '03]

  - On the other hand: Carefully constructed linear projections can often do better

- **This Talk:** An ***optimization*** based approach for designing "good" linear embeddings

Designing a "Good" Linear Map

Given: (normalized) pairwise differences

$$v_1,\ v_2,\ \ldots\ ,v_Q, \quad \|v_i\|_2 = 1$$

Want: the "shortest" matrix Φ , such that

$$1 - \delta \leq \|\Phi v_i\|^2 \leq 1 + \delta$$
$$i = 1, 2, \ldots, Q$$

Designing a “Good” Linear Map

Given: (normalized) pairwise differences

$$v_1,\ v_2,\ \ldots\ , v_Q, \quad \|v_i\|_2 = 1$$

Want: the “shortest” matrix Φ , such that

$$1 - \delta \leq \|\Phi v_i\|^2 \leq 1 + \delta$$
$$i = 1, 2, \ldots, Q$$

$$\Updownarrow$$

$$\text{minimize } \#\text{rows}(\Phi)$$
$$\left| \|\Phi v_i\|_2^2 - 1 \right| \leq \delta$$
$$i = 1, 2, \ldots, Q$$

Designing a “Good” Linear Map

- Convert quadratic constraints in Φ into ***linear*** constraints in $P = \Phi^T \Phi$ (“Lifting trick”)

$$\mathcal{A} : P \mapsto \{v_i^T P v_i\}_{i=1}^{Q}$$

- Rank of *P* = number of rows in Φ

Designing a "Good" Linear Map

- Convert quadratic constraints in Φ into ***linear*** constraints in $P = \Phi^T \Phi$ ("Lifting trick")

$$\mathcal{A} : P \mapsto \{v_i^T P v_i\}_{i=1}^{Q}$$

- Use a ***nuclear-norm*** relaxation of the rank

- Simplified problem:

$$\begin{array}{c} \text{minimize } \#\text{rows}(\Phi) \\ \left| \|\Phi v_i\|_2^2 - 1 \right| \le \delta \\ i = 1, 2, \ldots, Q \end{array} \quad \Leftrightarrow \quad \begin{array}{c} \text{minimize } \|P\|_* \\ \|\mathcal{A}(P) - \mathbf{1}\|_\infty \le \delta \\ P \succ 0,\ P = P^T \end{array}$$

[HSYB12]

Semidefinite Formulation

$$\text{minimize } \|P\|_*$$
$$\|\mathcal{A}(P) - \mathbf{1}\|_\infty \leq \delta$$
$$P \succ 0,\ P = P^T$$

- Solvable via standard interior point techniques

- Rank of solution is controlled by δ

Semidefinite Formulation

$$\text{minimize } \|P\|_*$$
$$\|\mathcal{A}(P) - \mathbf{1}\|_\infty \leq \delta$$
$$P \succ 0,\ P = P^T$$

- Solvable via standard interior point techniques

- Rank of solution is controlled by δ

- Practical considerations: *N* large, *Q* very large!
  For a matrix *P* of size *N* x *N*, the computational costs *per iteration* scale as $O\left(Q^3 + QN^6 + Q^2N^4\right)$

Algorithm

$$\text{minimize } \|P\|_*$$
$$\|\mathcal{A}(P) - \mathbf{1}\|_\infty \leq \delta$$
$$P \succ 0,\ P = P^T$$

- Alternating Direction Method of Multipliers (**ADMM**)

$$\text{minimize } \|P\|_*$$
$$P = L,\ \mathcal{A}(L) = q,\ \|q - \mathbf{1}\|_\infty \leq \delta$$

[HSYB12]

Algorithm: “NuMax”

$$\begin{gathered}\text{minimize } \|P\|_* \\ \|\mathcal{A}(P) - \mathbf{1}\|_\infty \leq \delta \\ P \succ 0,\ P = P^T\end{gathered}$$

- Alternating Direction Method of Multipliers (**ADMM**)

$$\begin{gathered}\text{minimize } \|P\|_* \\ P = L,\ \mathcal{A}(L) = q,\ \|q - \mathbf{1}\|_\infty \leq \delta\end{gathered}$$

- solve for *P* using spectral thresholding
- solve for *L* using least-squares
- solve for *q* using “squishing”

Computational costs per iteration: $O\left(Q^2 N^2\right)$

[HSYB12]

Ext: Task Adaptivity

Can prune the secants according to the task at hand

$$\mathcal{A} : P \mapsto \{v_i^T P v_i\}_{i=1}^{Q}$$

- If goal is signal reconstruction, preserve **all** pairwise differences
- If goal is classification, preserve only **inter-class** pairwise differences

- Can preferentially weight the input vectors according to importance (connections with *boosting*)

Ext: Model Adaptivity

- Designing sensing matrices for redundant **dictionaries**

$$D = [d_1, d_2, \ \ldots, \ d_S]$$

- Desiderata: the “holographic basis” ΦD must be as ***incoherent*** as possible (in terms of mutual/ average coherence) for good reconstruction [Elad06]

Ext: Model Adaptivity

- Designing sensing matrices for redundant **dictionaries**

$$D = [d_1, d_2, \ \ldots, \ d_S]$$

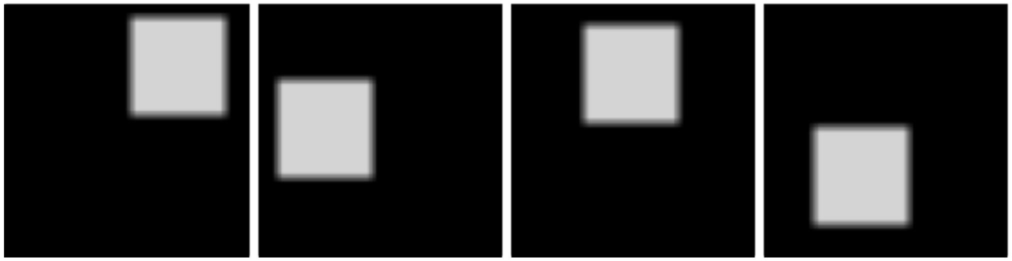
- Desiderata: the “holographic basis” ΦD must be as ***incoherent*** as possible (in terms of mutual/ average coherence) for good reconstruction [Elad06]

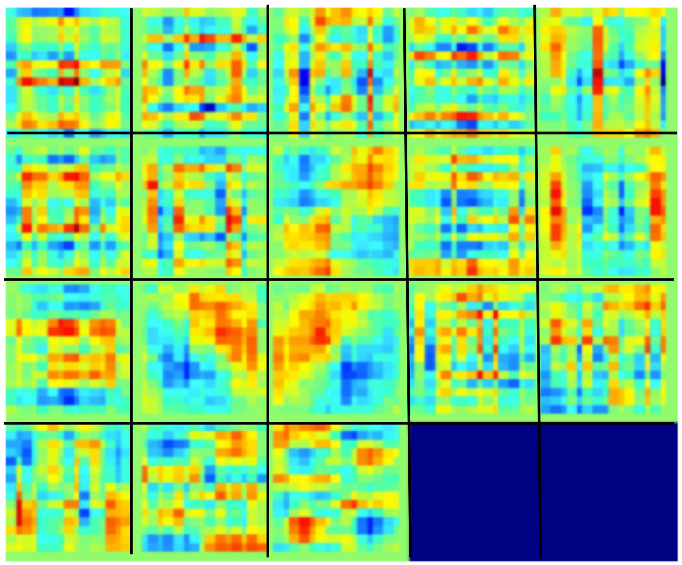
- **Upshot**: coherence → pairwise inner products → linear constraints: $\langle \Phi d_i, \Phi d_j \rangle = d_i^T P d_j$

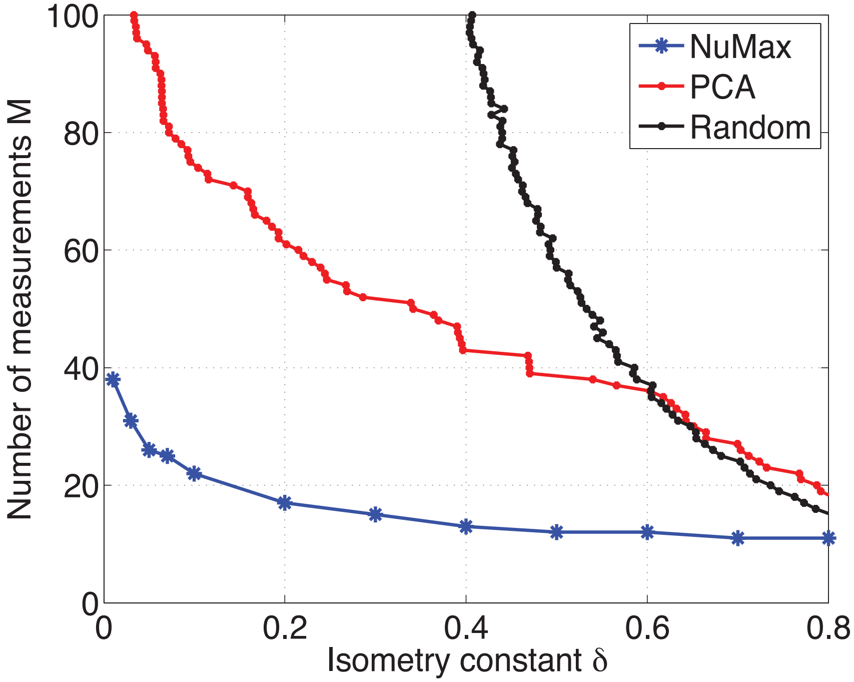
- Hence, simple variant of NuMax works

$$\mathcal{A} : P \mapsto \{d_i^T P d_j\}_{i,j}$$

Squares

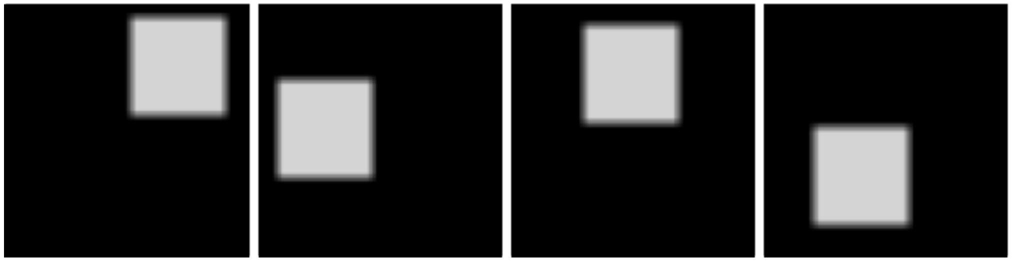


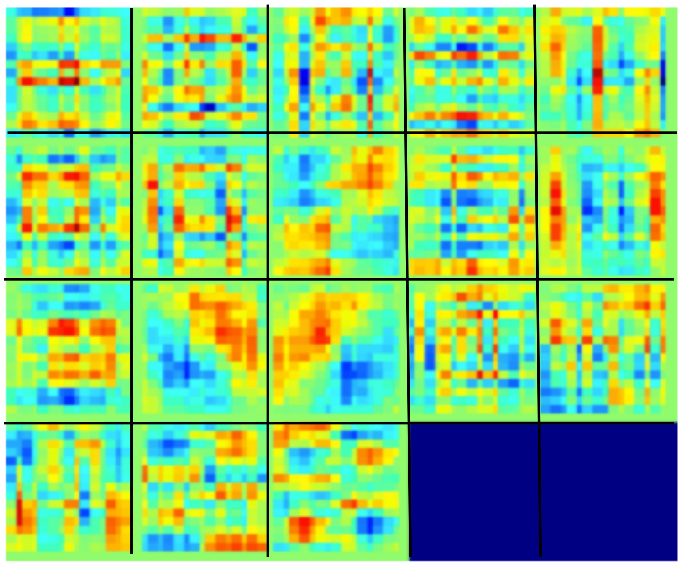


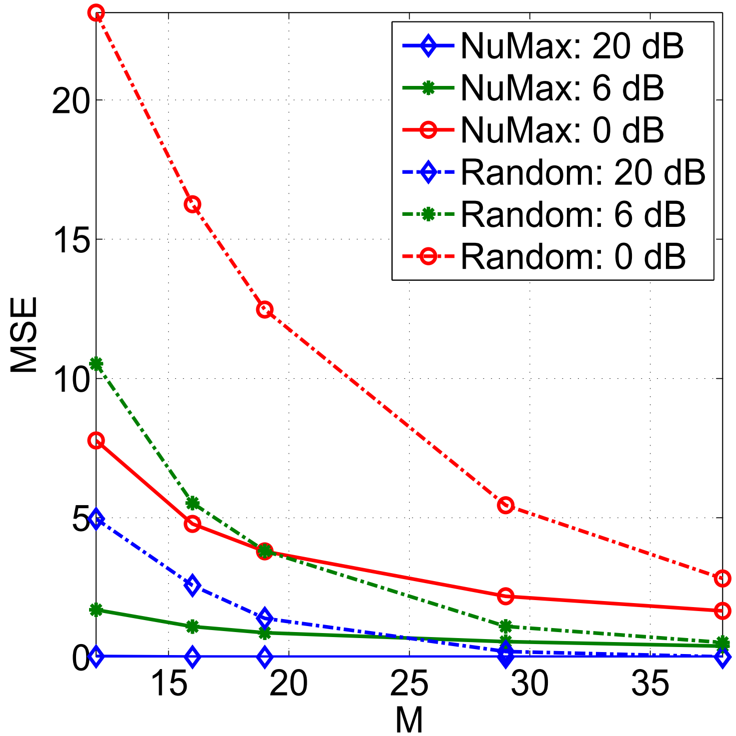


M=40 linear measurements enough to ensure isometry constant of 0.01

Squares: CS Reconstruction

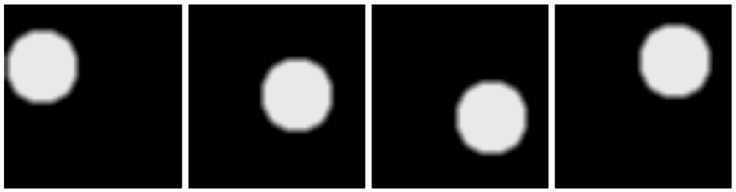


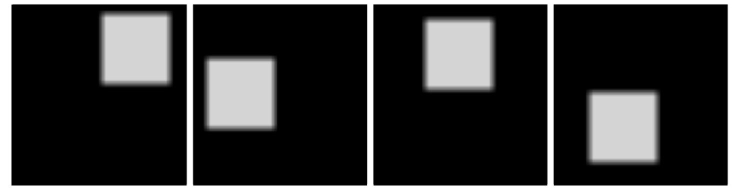


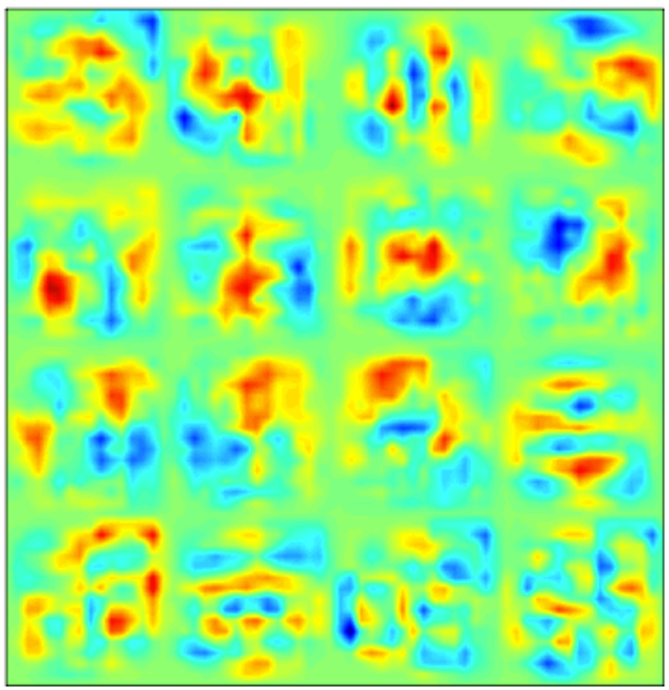


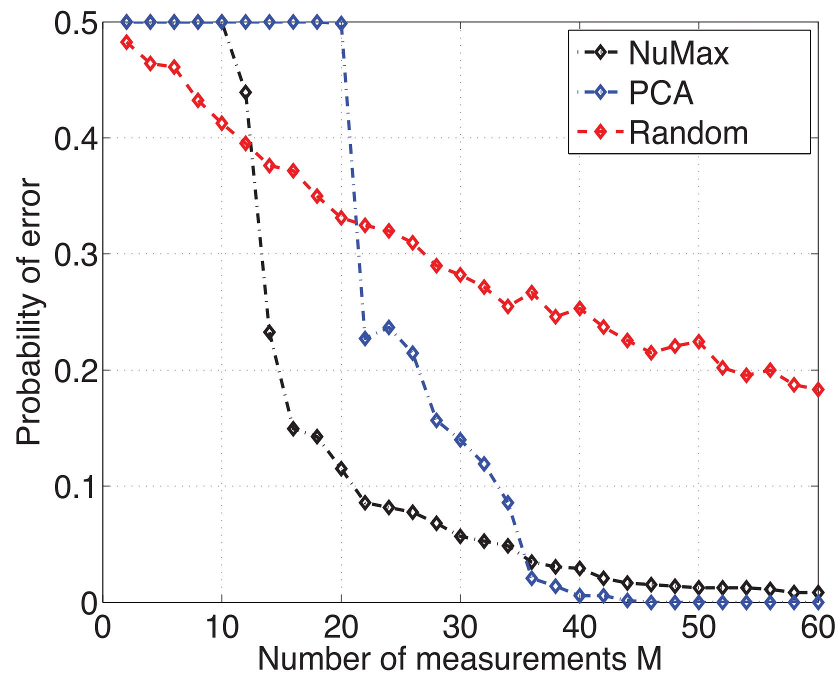
NuMax uniformly beats Random Projections

Circles vs. Squares



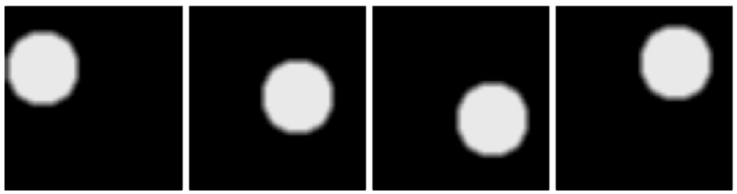


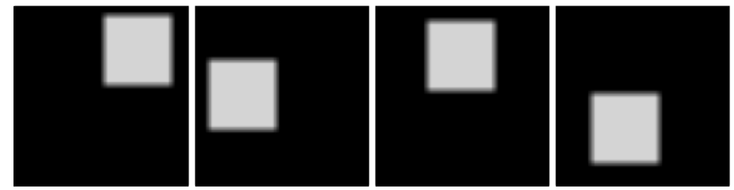


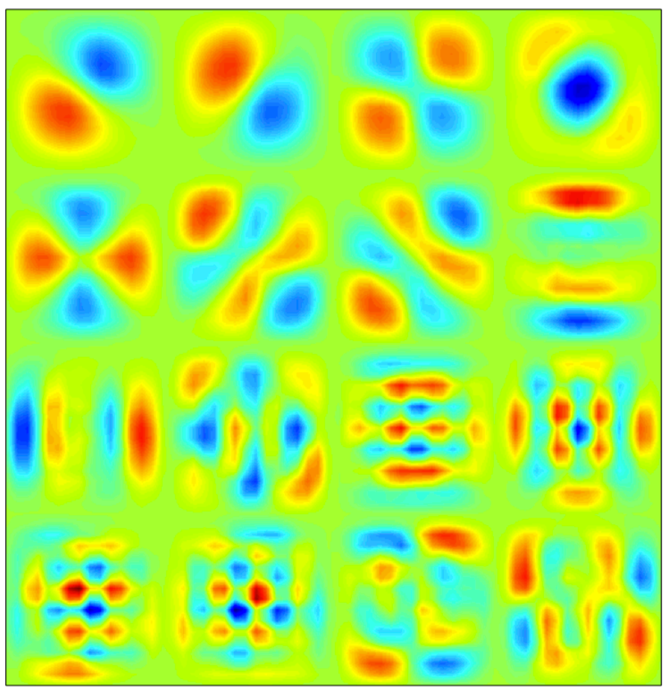


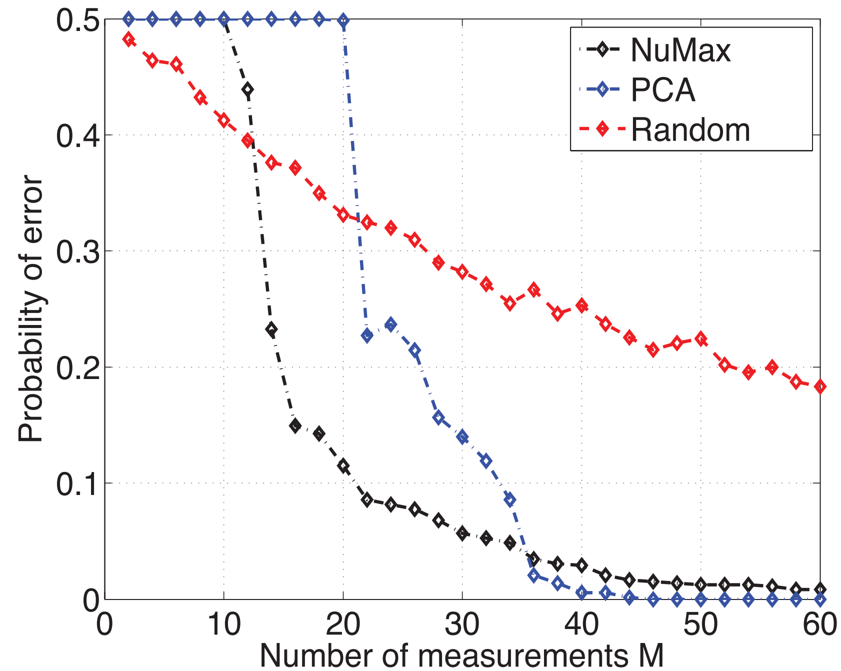
NuMax beats Random Projections, PCA

Circles vs. Squares



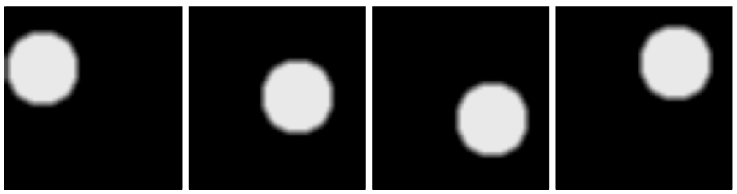


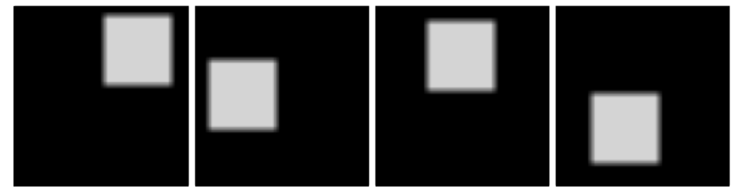


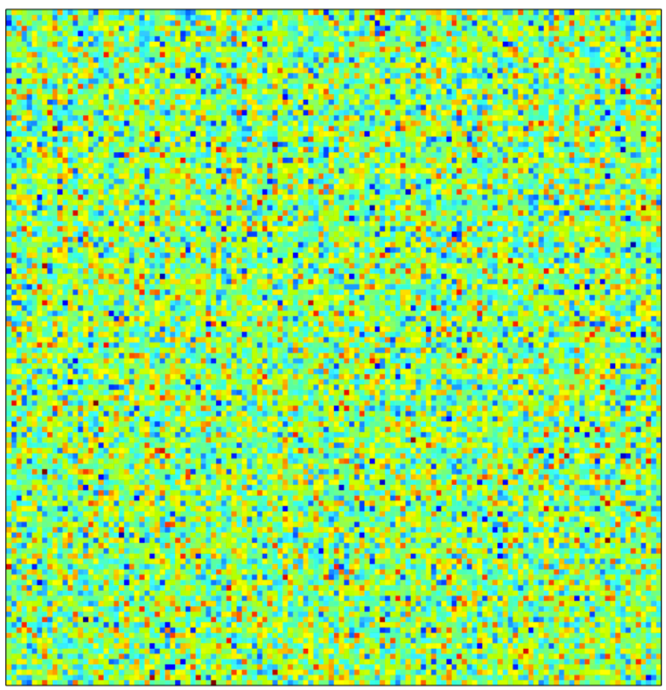


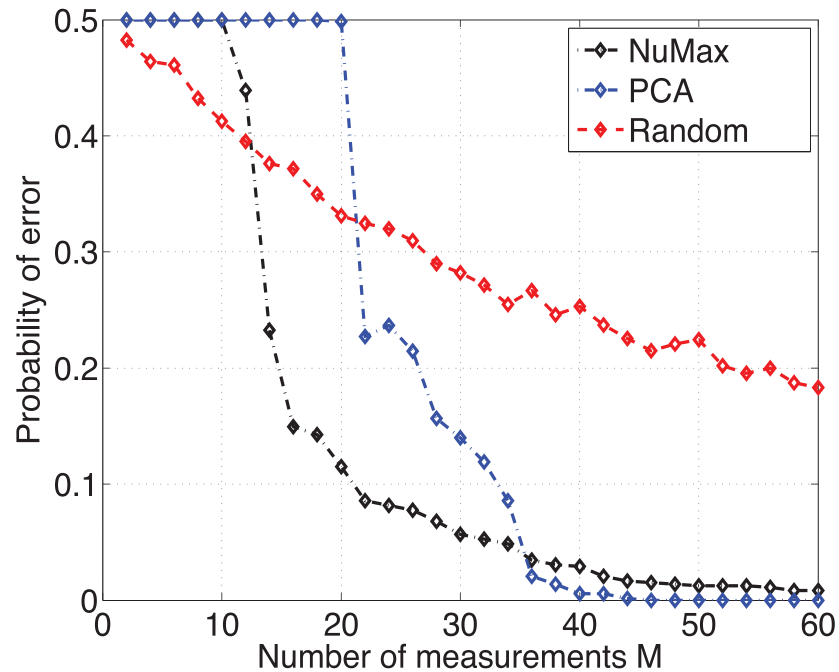
NuMax beats Random Projections, PCA

Circles vs. Squares



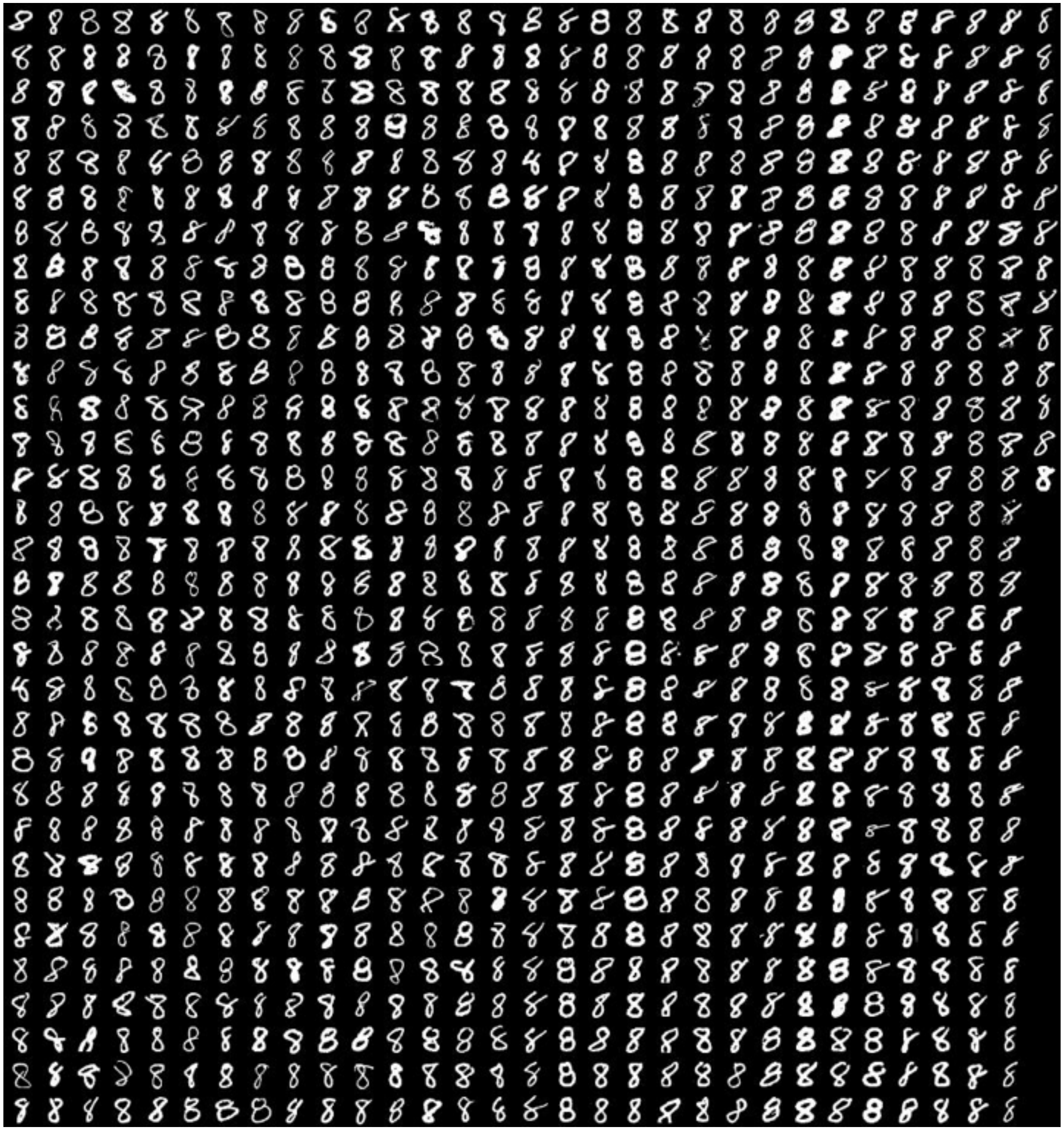


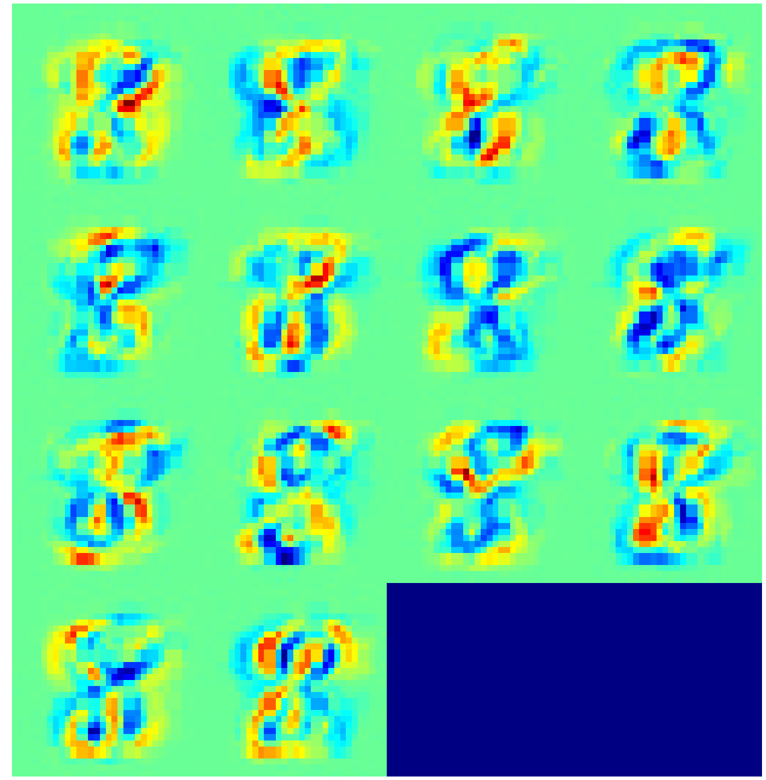




NuMax beats Random Projections, PCA

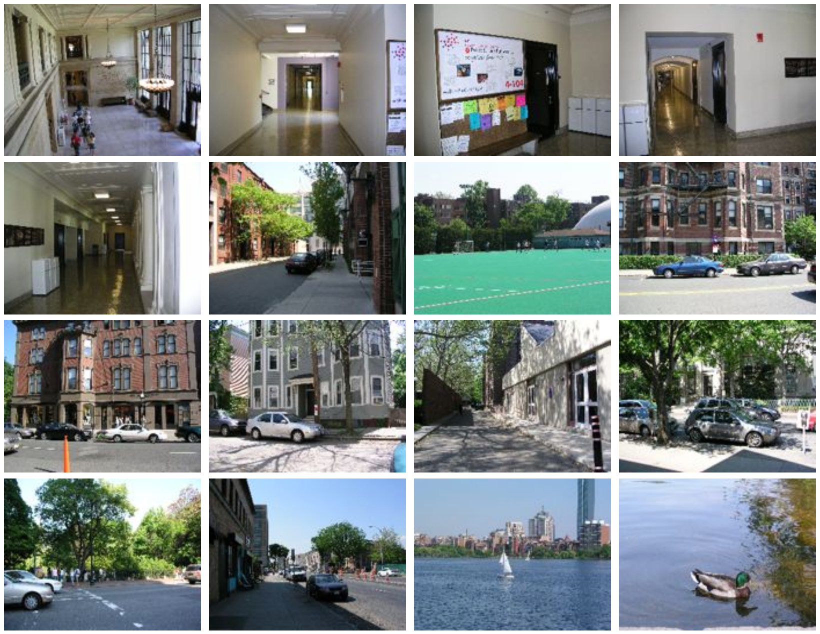
MNIST Dataset

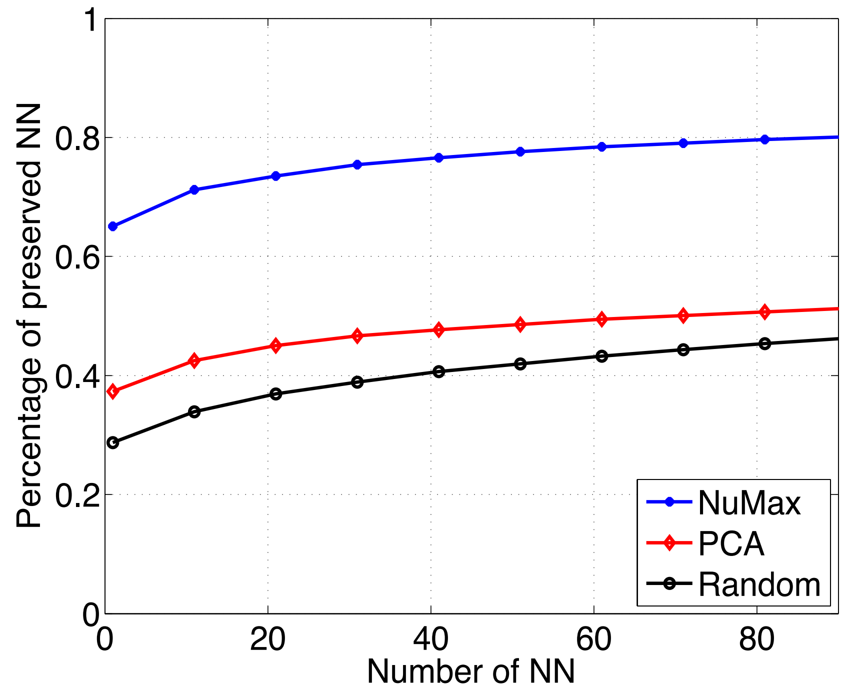




M = 10 basis functions suffice to achieve $\delta = 0.2$

LabelMe: Image Retrieval





- Goal: preserve *neighborhood structure*

- N = 512, Q = 4000, M = 45 suffices

- NuMax beats Random, PCA

(Some) Theory

$$\begin{array}{c} \text{minimize rank}(\Phi) \\ \left| \|\Phi v_i\|_2^2 - 1 \right| \leq \delta \\ i = 1, 2, \ldots, Q \end{array} \quad \Leftrightarrow \quad \begin{array}{c} \text{minimize } \|P\|_* \\ \|\mathcal{A}(P) - \mathbf{1}\|_\infty \leq \delta \\ P \succ 0, \; P = P^T \end{array}$$

- Can we predict (or bound) the rank of NuMax solution?

(Some) Theory

$$\begin{array}{c} \text{minimize rank}(\Phi) \\ \left| \|\Phi v_i\|_2^2 - 1 \right| \leq \delta \\ i = 1, 2, \ldots, Q \end{array} \quad \Leftrightarrow \quad \begin{array}{c} \text{minimize } \|P\|_* \\ \|\mathcal{A}(P) - \mathbf{1}\|_\infty \leq \delta \\ P \succ 0,\ P = P^T \end{array}$$

- Can we predict (or bound) the rank of NuMax solution?
- Result:

$$M^* \leq \left\lceil \frac{\sqrt{8Q+1} - 1}{2} \right\rceil$$

(easy: count the number of active constraints)

Summary

- Goal: develop a representation for data that is ***linear, isometric***

- Can be posed as a rank-minimization problem
  - Semi-definite program (SDP) achieves this
  - NuMax achieves this *very efficiently*

- **Applications**: Compressive sensing, classification, retrieval, ++

[HSYB12] “A Convex Approach for Learning Near-Isometric Linear Embeddings”, Nov 2012