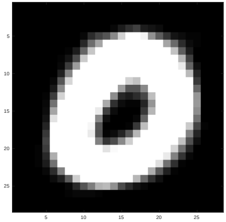
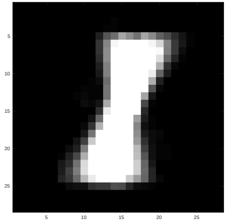
Rethinking Sketching as Sampling: Efficient Approximate Solution to Linear Inverse Problems

Fernando Gama, A. G. Marques, G. Mateos & A. Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
fgama@seas.upenn.edu

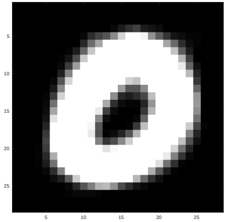
GlobalSIP, December 9, 2016

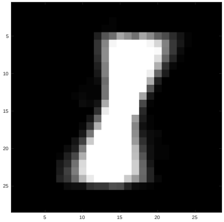
PCA Classification



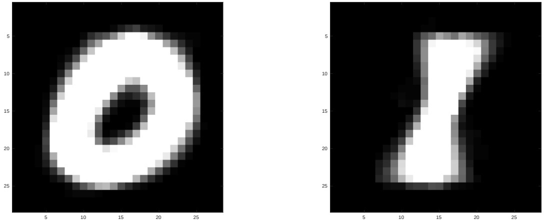


PCA Classification

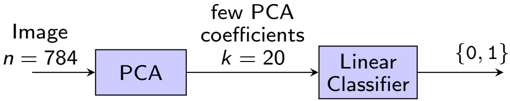




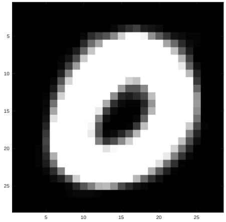
- Classify images according to the digits handwritten on them

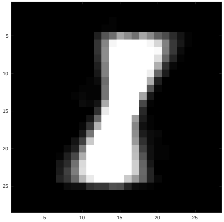


- Classify images according to the digits handwritten on them
- Perform PCA $\Rightarrow$ Keep first few coefficients $\Rightarrow$ Apply linear classifier



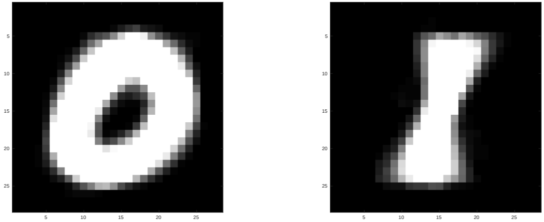
PCA Classification



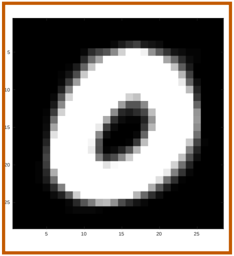


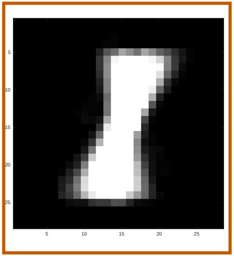
- Few PCA coefficients $\Rightarrow$ Problem is inherently lower-dimensional
- Improves classification task $\Rightarrow$ Low-pass filter to remove noise
- Lower-dimensional representation can also save computational cost

Computational Cost

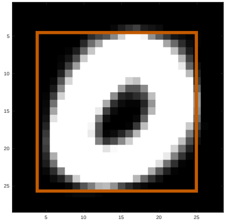


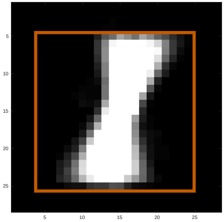
- Note that in performing PCA we need the complete image
- However, there are pixels that do not contribute to classification
  - ⇒ Pixels on the border of the image, for example
- And there are pixels that are more important for classification
  - ⇒ Pixels that are white in one image but black in the other



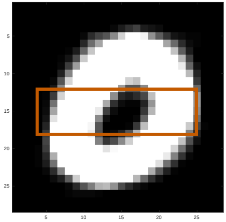


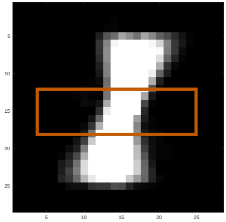
- Note that in performing PCA we need the complete image
- However, there are pixels that do not contribute to classification
  - ⇒ Pixels on the border of the image, for example
- And there are pixels that are more important for classification
  - ⇒ Pixels that are white in one image but black in the other





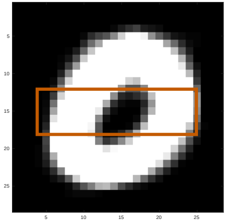
- Note that in performing PCA we need the complete image
- However, there are pixels that do not contribute to classification
  - ⇒ Pixels on the border of the image, for example
- And there are pixels that are more important for classification
  - ⇒ Pixels that are white in one image but black in the other

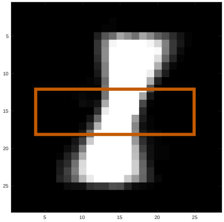




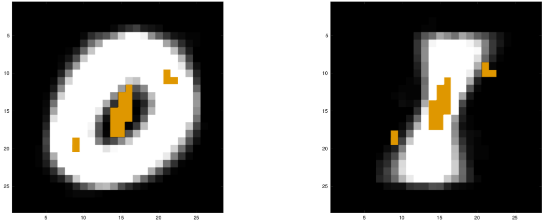
- Note that in performing PCA we need the complete image
- However, there are pixels that do not contribute to classification
  - ⇒ Pixels on the border of the image, for example
- And there are pixels that are more important for classification
  - ⇒ Pixels that are white in one image but black in the other

Sampling

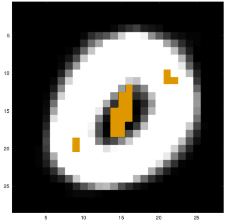


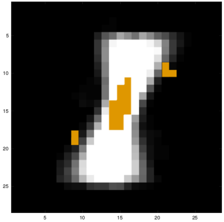


- Few nonzero PCA coefficients $\Rightarrow$ Bandlimited signal $\Rightarrow$ Sampling
- Subspace representation on covariance graph (not all pixels are useful)
  $\Rightarrow$ Linear combination of a few eigenvectors weighted by PCA coeff.
- Extend to arbitrary graphs $\Rightarrow$ Sampling of bandlimited graph signals



- Few nonzero PCA coefficients $\Rightarrow$ Bandlimited signal $\Rightarrow$ Sampling
- Subspace representation on covariance graph (not all pixels are useful)
  $\Rightarrow$ Linear combination of a few eigenvectors weighted by PCA coeff.
- Extend to arbitrary graphs $\Rightarrow$ Sampling of bandlimited graph signals





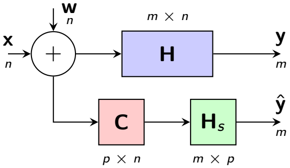
- Few nonzero PCA coefficients ⇒ Bandlimited signal ⇒ Sampling
- Subspace representation on covariance graph (not all pixels are useful)
  ⇒ Linear combination of a few eigenvectors weighted by PCA coeff.
- Extend to arbitrary graphs ⇒ Sampling of bandlimited graph signals
- Design a classifier to operate on the samples ⇒ Reduce dimensionality

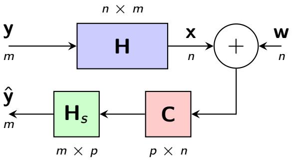
Rethinking Sketching as Sampling

- Sketching $\Rightarrow$ Reduce dimensionality of linear transformations
- Projection on a lower-dimensional subspace $\Rightarrow$ Smaller size matrix
  $\Rightarrow$ Matrix sketch retains the most outstanding characteristics
- Smaller matrix operates on smaller vector to compute the result
  $\Rightarrow$ Project vector on a lower-dimensional subspace $\Rightarrow$ Sampling

Rethinking Sketching as Sampling

- Sketching $\Rightarrow$ Reduce dimensionality of linear transformations
- Projection on a lower-dimensional subspace $\Rightarrow$ Smaller size matrix
  $\Rightarrow$ Matrix sketch retains the most outstanding characteristics
- Smaller matrix operates on smaller vector to compute the result
  $\Rightarrow$ Project vector on a lower-dimensional subspace $\Rightarrow$ Sampling
- Jointly design sampling of signal and sketching of linear transform
  $\Rightarrow$ Obtain approximate solution by operating only on few samples





Sampling of Graph Signals

- Graph signals defined on top of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with n nodes
- Irregular support captured by normal graph shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$
- Define the graph Fourier transform (GFT) $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$

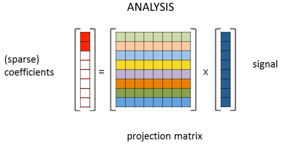
  $\Rightarrow$ Linear combination weighted by GFT coefficients $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$ (iGFT)

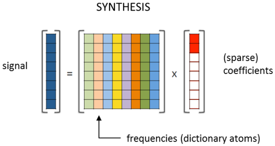
Sampling of Graph Signals

- Graph signals defined on top of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with n nodes
- Irregular support captured by normal graph shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$
- Define the graph Fourier transform (GFT) $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$

  $\Rightarrow$ Linear combination weighted by GFT coefficients $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$ (iGFT)
- Bandlimited graph signal $\Rightarrow \tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_k; \mathbf{0}_{n-k}]$ with $k \ll n$ $\Rightarrow \mathbf{x} = \mathbf{V}_k\tilde{\mathbf{x}}_k$

  $\Rightarrow$ Active eigenbasis of vectors $\mathbf{V}_k = [\mathbf{V}_k, \mathbf{0}_{n\times(n-k)}]$
- Signal as a linear combination of few elements in $\mathbf{V}_k$ $\Rightarrow$ Sampling





Sketching

- Estimate the input to a linear transform by measuring the output
  - $\Rightarrow$ The model is $\mathbf{x} = \mathbf{H}\mathbf{y}$, with $\mathbf{H} \in \mathbb{R}^{n \times m}$ and where $n \gg m$
  - $\Rightarrow$ LS solution $\Rightarrow$ Computationally costly (pseudo-)inverse

Sketching

- Estimate the input to a linear transform by measuring the output
  - $\Rightarrow$ The model is $\mathbf{x} = \mathbf{H}\mathbf{y}$, with $\mathbf{H} \in \mathbb{R}^{n \times m}$ and where $n \gg m$
  - $\Rightarrow$ LS solution $\Rightarrow$ Computationally costly (pseudo-)inverse

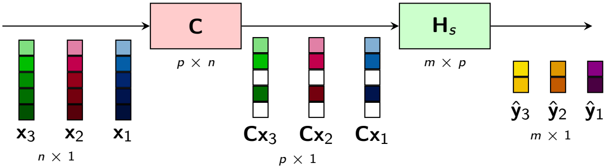
- Traditional sketching $\Rightarrow$ Reduce dimension of the linear problem
- Compress $\mathbf{H}$ and $\mathbf{x}$ $\Rightarrow$ $\mathbf{KH}$ and $\mathbf{Kx}$, $\mathbf{K} \in \mathbb{R}^{p \times n}$ random, $p \ll n$
  - $\Rightarrow$ Random projection on a lower-dimensional subspace
  - $\Rightarrow$ Solution of smaller problem $\min_{\mathbf{y}} \|(\mathbf{KH})\mathbf{y} - (\mathbf{Kx})\|_2^2$ $\Rightarrow$ Faster
- Design $\mathbf{K}$ such that $\mathbf{KH}$ and $\mathbf{Kx}$ retains important traits of the problem
  - $\Rightarrow$ Then, solving for $(\mathbf{KH}, \mathbf{Kx})$ yields a good approximation

- We consider a deterministic design to obtain a smaller matrix sketch

Operating Conditions

- Sequence of signals to be processed by the same linear transform
  $\Rightarrow$ Matrix $\mathbf{H}$ is *big* $\Rightarrow$ Computationally intensive to operate with
- Realizations of a bandlimited graph random process $\Rightarrow$ $\mathbf{R}_x$ singular
- Enough computational power available prior to processing of signals
- Process sequence of signals fast $\Rightarrow$ Apply smaller matrix to samples

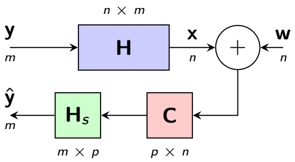
- Traditional sampling $\Rightarrow$ Ignores further processing on the signal
- Traditional sketching $\Rightarrow$ Recomputes sketch for each realization $\mathbf{x}$

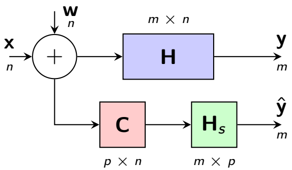


Design $\mathbf{C}, \mathbf{H}_s$ based on $\mathbf{H}$ and statistics of signal $\mathbf{R}_x$ and noise $\mathbf{R}_w$

Problem Statement

- Design a sampling matrix $\mathbf{C}$ that selects $k \leq p \ll n$ samples
- Design a deterministic sketch $\mathbf{H}_s$ to be directly applied to samples
- Joint design of sketching and sampling prior to start of sequence
  - $\Rightarrow$ Minimize the MSE relative to using full $\mathbf{H}$ on the full signal $\mathbf{x}$
- Processing of signals reduces to sampling and matrix multiplication
- The computational cost of processing is reduced by a factor of p/n





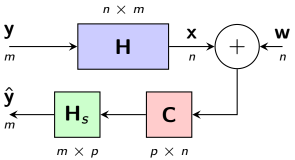
Inverse Linear Problem

- Use noisy output $(\mathbf{x} + \mathbf{w}) \in \mathbb{R}^n$ to estimate input $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} = \mathbf{H}\mathbf{y}$
- Linear model $\mathbf{H} \in \mathbb{R}^{n \times m}$ tall matrix with $m \ll n$ and full rank
- Output signal $\mathbf{x} \in \mathbb{R}^n$ is k-bandlimited with known $\mathbf{R}_x \succeq \mathbf{0}$ (singular)
- Input noise $\mathbf{w}$, indep. of $\mathbf{x}$ with known covariance matrix $\mathbf{R}_w \succ \mathbf{0}$

- Design sketch $\mathbf{H}_s^* \in \mathbb{R}^{m \times p}$ and a selection matrix $\mathbf{C}^* \in \mathbb{R}^{p \times n}$

$$\{\mathbf{C}^*, \mathbf{H}_s^*\} := \underset{C \in \mathcal{C}_{pn}, \mathbf{H}_s}{\operatorname{argmin}} \mathbb{E}\left[\|\mathbf{H}\mathbf{H}_s\mathbf{C}(\mathbf{x}+\mathbf{w}) - \mathbf{x}\|_2^2\right]$$

- Solve this problem before processing the sequence of signals



Inverse Linear Problem: Noisy case

- Two-stage optimization to solve $\min \mathbb{E}\left[\|\mathbf{H}\mathbf{H}_s\mathbf{C}(\mathbf{x}+\mathbf{w})-\mathbf{x}\|_2^2\right]$

1. Design matrix sketch $\mathbf{H}_s^* = \mathbf{H}_s^*(\mathbf{C})$ then replace on objective function

$$\mathbf{H}_s^*(\mathbf{C}) = \mathbf{A}_{LS}\mathbf{R}_x\mathbf{C}^T\left(\mathbf{C}(\mathbf{R}_x+\mathbf{R}_w)\mathbf{C}^T\right)^{-1}$$

$\Rightarrow$ This is the LS solution with a preprocessing to deal with the noise

2. Define auxiliary matrix $\mathbf{G} = \mathbf{H}\mathbf{A}_{LS}$ and obtain $\mathbf{C}^*$ by solving

$$\min_{\mathbf{C}\in\mathcal{C}_{pn}} \operatorname{tr}\left[\mathbf{R}_x - \mathbf{G}\mathbf{R}_x\mathbf{C}^T\left(\mathbf{C}(\mathbf{R}_x+\mathbf{R}_w)\mathbf{C}^T\right)^{-1}\mathbf{C}\mathbf{R}_x\mathbf{G}^T\right]$$

$\Rightarrow$ Tradeoff between output energy and noise of the selected samples

$\Rightarrow$ This is a binary optimization problem over selection matrix $\mathbf{C}$

$\Rightarrow$ There are $\binom{n}{p}$ possible solutions $\Rightarrow$ Prohibitive to test all of them

Inverse Linear Problem: Noisy case

- Binary constraints are inherent to the selection problem
- Equivalent problem with linear objective function and LMIs
  - $\Rightarrow$ It would be an SDP except for binary constraint
- Observe that $\mathbf{C}^T\mathbf{C} = \text{diag}(\mathbf{c})$ $\Rightarrow$ Sampling vector $\mathbf{c} \in \{0,1\}^n$
- Define $\bar{\mathbf{C}}_\alpha = \text{diag}(\mathbf{c})/\alpha$, $\alpha > 0$ and $\bar{\mathbf{R}}_\alpha = \mathbf{R}_x + \mathbf{R}_w - \alpha \mathbf{I}_n$
- Problem over $\mathbf{C}$ can be posed as an equivalent problem over $\mathbf{c}$

$$\min_{\mathbf{c}\in\{0,1,\}^n, \mathbf{Y}, \bar{\mathbf{C}}_\alpha} \text{tr}[\mathbf{Y}]$$

$$\text{s. t. } \bar{\mathbf{C}}_\alpha = \alpha^{-1}\text{diag}(\mathbf{c}) \ , \ \mathbf{c}^T\mathbf{1}_n = p$$

$$\begin{bmatrix} \mathbf{Y} - \mathbf{R}_x + \mathbf{G}\mathbf{R}_x\bar{\mathbf{C}}_\alpha\mathbf{R}_x\mathbf{G}^T & \mathbf{G}\mathbf{R}_x\bar{\mathbf{C}}_\alpha \\ \bar{\mathbf{C}}_\alpha\mathbf{R}_x\mathbf{G}^T & \bar{\mathbf{R}}_\alpha^{-1} + \bar{\mathbf{C}}_\alpha \end{bmatrix} \succeq \mathbf{0}$$

- This is also a complicated problem but slightly more tractable

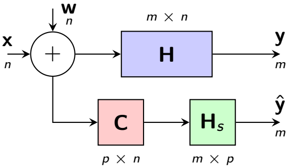
Direct Linear Problem

- Given noisy input $(\mathbf{x} + \mathbf{w}) \in \mathbb{R}^n$ estimate output $\mathbf{y} = \mathbf{H}\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$
- Design sketch $\mathbf{H}_s \in \mathbb{R}^{m \times p}$ and $p \times n$ selection matrix $\mathbf{C}$

$$\{\mathbf{C}^*, \mathbf{H}_s^*\} := \underset{C \in \mathcal{C}_{pn}, \mathbf{H}_s}{\text{argmin}} \ \mathbb{E}\left[\|\mathbf{H}_s\mathbf{C}(\mathbf{x}+\mathbf{w}) - \mathbf{H}\mathbf{x}\|_2^2\right]$$

- Two stage optimization $\Rightarrow$ Matrix sketch $\mathbf{H}_s$ and sampling scheme $\mathbf{C}$
- Can be reformulated as an equivalent problem over selection vector $\mathbf{c}$

$\Rightarrow$ Linear objective function, LMIs constraints, binary constraint



Sampling Heuristics

- Solving sampling problems might be intractable $\Rightarrow$ Heuristic solutions
- Convex relaxation $[\mathcal{O}((n+m)^{3.5})]$ $\Rightarrow \mathbf{c} \in [0,1]^n$ $\Rightarrow$ SDPs

  $\Rightarrow$ Tresholding $\Rightarrow$ Set the p highest values to 1 and the rest to 0

  $\Rightarrow$ Random $\Rightarrow$ Use relaxed solution as distribution to select nodes
- Noise-blind Heuristic $[\mathcal{O}(n \log n)]$ $\Rightarrow$ p rows of $\mathbf{R}_x \mathbf{G}^T$ with largest $\| \cdot \|_2$

$$\min_{\mathbf{C} \in \mathcal{C}_{pn}} \operatorname{tr} \left[\mathbf{R}_x - \mathbf{G}\mathbf{R}_x \mathbf{C}^T \left(\mathbf{C} (\mathbf{R}_x + \mathbf{R}_w) \mathbf{C}^T \right)^{-1} \mathbf{C} \mathbf{R}_x \mathbf{G}^T \right]$$

- Greedy approach $[\mathcal{O}(np(mnp + p^3))]$ $\Rightarrow$ Select best node iteratively

Example: Computing the Graph Fourier Transform

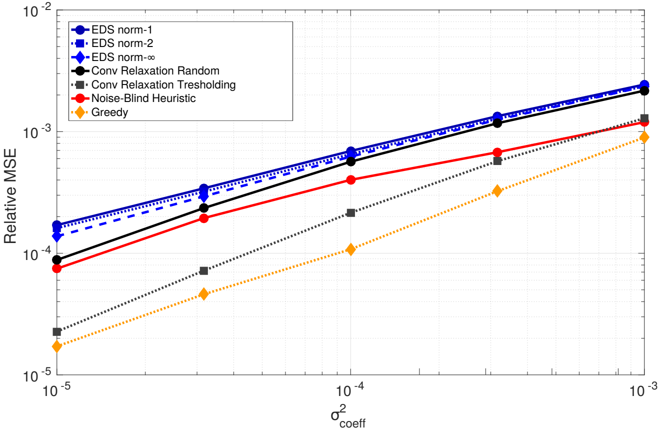
- Consider a bandlimited graph signal $\mathbf{x} = \mathbf{V}_k \tilde{\mathbf{x}}_k$ $\Rightarrow$ $\tilde{\mathbf{x}}_k$: freq. coeff.
  $\Rightarrow$ Inverse linear model $\Rightarrow$ $\mathbf{x} = \mathbf{V}_k \tilde{\mathbf{x}}_k$ $\Rightarrow$ Transform $\mathbf{H} = \mathbf{V}_k$
- Sequence of noisy signals $(\mathbf{x} + \mathbf{w})$ $\Rightarrow$ Fast computation of the GFT
  $\Rightarrow$ $\mathbf{w}$: white gaussian zero-mean noise of power prop. to energy of $\mathbf{x}$

Example: Computing the Graph Fourier Transform

- Consider a bandlimited graph signal $\mathbf{x} = \mathbf{V}_k \tilde{\mathbf{x}}_k \Rightarrow \tilde{\mathbf{x}}_k$: freq. coeff.
  - $\Rightarrow$ Inverse linear model $\Rightarrow \mathbf{x} = \mathbf{V}_k \tilde{\mathbf{x}}_k \Rightarrow$ Transform $\mathbf{H} = \mathbf{V}_k$
- Sequence of noisy signals $(\mathbf{x} + \mathbf{w}) \Rightarrow$ Fast computation of the GFT
  - $\Rightarrow \mathbf{w}$: white gaussian zero-mean noise of power prop. to energy of $\mathbf{x}$
- Compare between different heuristics proposed for the joint design
- Compare with other traditional sampling schemes for reconstruction
  - $\Rightarrow$ Experimentally Design Sampling (EDS) technique
  - $\Rightarrow$ Assign to each node the norm of the rows of $\mathbf{V}_k$
  - $\Rightarrow$ Sample with replacement with a distribution prop. to this norm

Example: Approximating the GFT

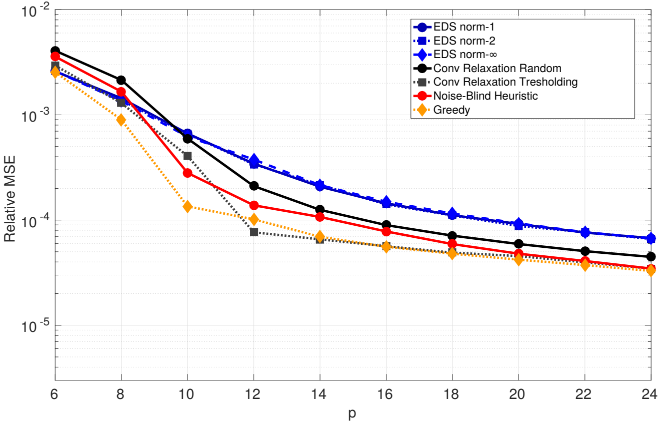
- Erdős-Rényi graph of size $n = 100$ with probablity 0.2
- Signal bandlimited with $k = 10$ freq. coeff. $\Rightarrow p = k = 10$



- Error of $2 \cdot 10^{-5}$ reducing computational complexity by 10

Example: Approximating the GFT

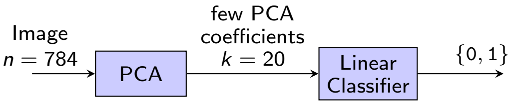
- Erdős-Rényi graph of size $n = 100$ with probablity 0.2
- Signal bandlimited with $k = 10$ freq. coeff. $\Rightarrow \sigma^2_{\text{coeff}} = 10^{-4}$



- Error of 10^{-4} reducing computational complexity by $100/24 = 4.167$

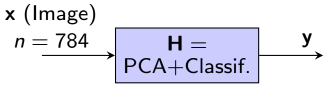
Classification of handwritten digits

- Classify images of handwritten digits of the MNIST database
- Linear classifier in the PCA domain $\Rightarrow$ Expensive linear operation
  $\Rightarrow$ Subsume PCA and classifier in one linear operator



- Classify images of handwritten digits of the MNIST database
- Linear classifier in the PCA domain $\Rightarrow$ Expensive linear operation

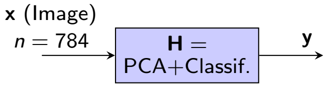
  $\Rightarrow$ Subsume PCA and classifier in one linear operator



Classification of handwritten digits

- Classify images of handwritten digits of the MNIST database
- Linear classifier in the PCA domain $\Rightarrow$ Expensive linear operation

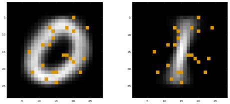
  $\Rightarrow$ Subsume PCA and classifier in one linear operator



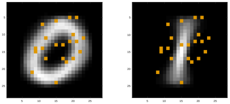
- Classify images by operating directly on a subset of pixels
- Images of size $n = 784$ pixels $\Rightarrow$ Use only $p = 20$ pixels

  $\Rightarrow$ Processing costs reduced by 39.2 for each image

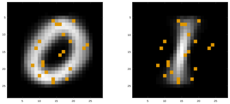
Classification of handwritten digits



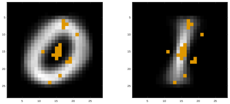
(a) EDS norm-1



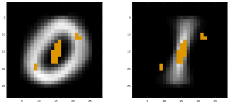
(b) EDS norm-2



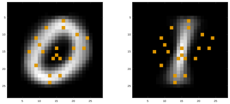
(c) EDS norm-∞



(d) Tresholding



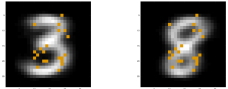
(e) Noise-Blind



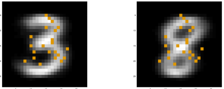
(f) Greedy

- Sketching and sampling techniques achieve perfect classification

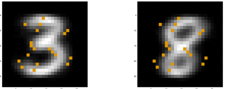
Classification of handwritten digits



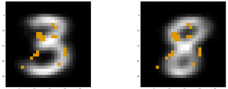
(a) EDS norm-1



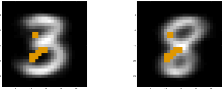
(b) EDS norm-2



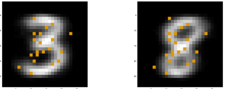
(c) EDS norm-∞



(d) Tresholding



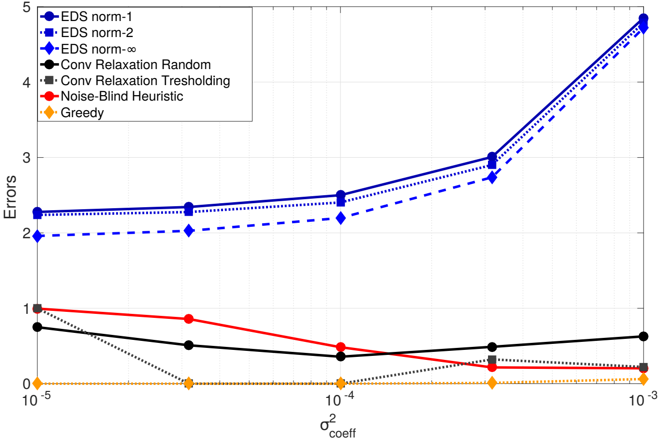
(e) Noise-Blind



(f) Greedy

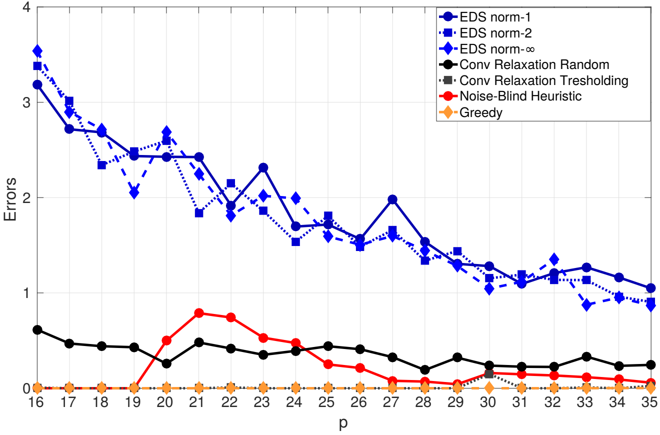
- Error rate using full image: 4.00%
  - $\Rightarrow$ Greedy approach using 20 pixels: 4.53%

- 200 image classification as a function of noise for $p = 20$ pixels



Classification of handwritten digits

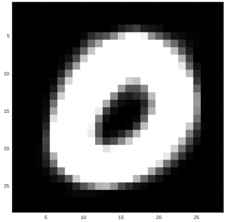
- 200 image classification as a function of the number of pixels

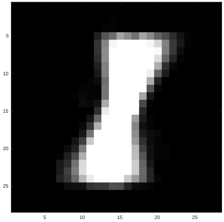


Conclusions

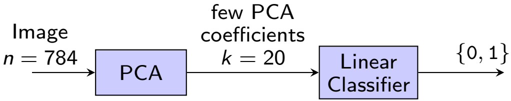
- Optimal sketch and sampling for processing bandlimited graph signals
  - $\Rightarrow$ Obtain approximate solution by operating only on a few samples
  - $\Rightarrow$ Accelerate processing of a sequence of bandlimited signals
- Joint design of matrix sketch and sampling scheme (prior to processing)
  - $\Rightarrow$ Two-stage optimization $\Rightarrow$ Heuristic solutions for sampling problem
- Fast computation of GFT of a bandlimited graph signal
  - $\Rightarrow$ Errors in the order of 10^{-5} reducing the cost 10 times
- Classification of images of size 784 pixels of handwritten digits
  - $\Rightarrow$ Using as few as 20 pixels $\Rightarrow$ 40 times less computational cost
- Journal version available on arXiv: `arxiv.org/abs/1611.00119`

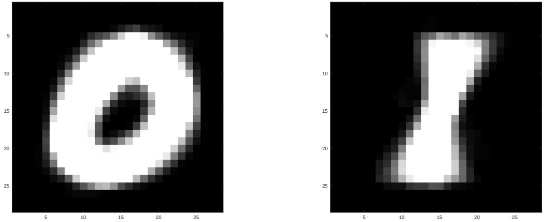
PCA Classification





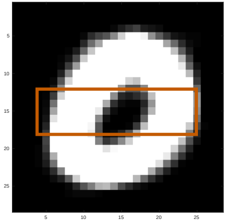
- Classify images according to the digits handwritten on them
- Perform PCA $\Rightarrow$ Keep first few coefficients $\Rightarrow$ Apply linear classifier

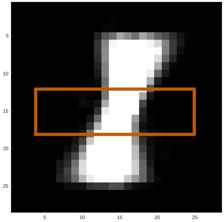




- Few PCA coefficients $\Rightarrow$ Problem is inherently lower-dimensional
- Improves classification task $\Rightarrow$ Low-pass filter to remove noise
- Lower-dimensional representation can also save computational cost

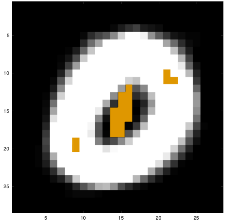
Computational Cost

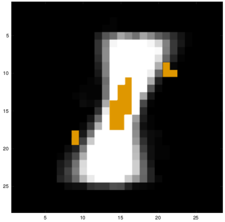




- Note that in performing PCA we need the complete image
- However, there are pixels that do not contribute to classification
  - ⇒ Pixels on the border of the image, for example
- And there are pixels that are more important for classification
  - ⇒ Pixels that are white in one image but black in the other

Sampling

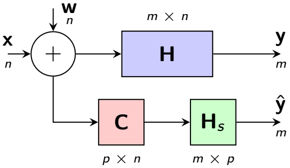


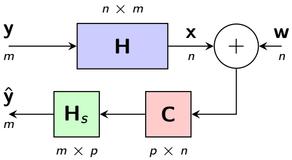


- Few nonzero PCA coefficients $\Rightarrow$ Bandlimited signal $\Rightarrow$ Sampling
- Subspace representation on covariance graph (not all pixels are useful)
  $\Rightarrow$ Linear combination of a few eigenvectors weighted by PCA coeff.
- Extend to arbitrary graphs $\Rightarrow$ Sampling of bandlimited graph signals
- Design a classifier to operate on the samples $\Rightarrow$ Reduce dimensionality

Rethinking Sketching as Sampling

- Sketching ⇒ Reduce dimensionality of linear transformations
- Projection on a lower-dimensional subspace ⇒ Smaller size matrix
  ⇒ Matrix sketch retains the most outstanding characteristics
- Smaller matrix operates on smaller vector to compute the result
  ⇒ Project vector on a lower-dimensional subspace ⇒ Sampling
- Jointly design sampling of signal and sketching of linear transform
  ⇒ Obtain approximate solution by operating only on few samples





- Graph signals defined on top of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with n nodes
- Irregular support captured by normal graph shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$
- Define the graph Fourier transform (GFT) $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$

  $\Rightarrow$ Linear combination weighted by GFT coefficients $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$ (iGFT)

- Bandlimited graph signal $\Rightarrow \tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_k; \mathbf{0}_{n-k}]$ with $k \ll n$ $\Rightarrow \mathbf{x} = \mathbf{V}_k\tilde{\mathbf{x}}_k$

  $\Rightarrow$ Active eigenbasis of vectors $\mathbf{V}_k = [\mathbf{V}_k, \mathbf{0}_{n\times(n-k)}]$

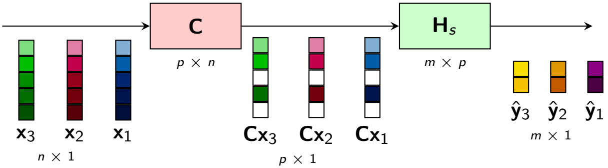
- Signal as a linear combination of few elements in $\mathbf{V}_k$ $\Rightarrow$ Sampling

Sketching

- Estimate the input to a linear transform by measuring the output
  - $\Rightarrow$ The model is $\mathbf{x} = \mathbf{H}\mathbf{y}$, with $\mathbf{H} \in \mathbb{R}^{n \times m}$ and where $n \gg m$
  - $\Rightarrow$ LS solution $\Rightarrow$ Computationally costly (pseudo-)inverse
- Traditional sketching $\Rightarrow$ Reduce dimension of the linear problem
- Compress $\mathbf{H}$ and $\mathbf{x}$ $\Rightarrow$ $\mathbf{KH}$ and $\mathbf{Kx}$, $\mathbf{K} \in \mathbb{R}^{p \times n}$ random, $p \ll n$
  - $\Rightarrow$ Random projection on a lower-dimensional subspace
  - $\Rightarrow$ Solution of smaller problem $\min_{\mathbf{y}} \|(\mathbf{KH})\mathbf{y} - (\mathbf{Kx})\|_2^2$ $\Rightarrow$ Faster
- Design $\mathbf{K}$ such that $\mathbf{KH}$ and $\mathbf{Kx}$ retains important traits of the problem
  - $\Rightarrow$ Then, solving for $(\mathbf{KH}, \mathbf{Kx})$ yields a good approximation
- We consider a deterministic design to obtain a smaller matrix sketch

Operating Conditions

- Sequence of signals to be processed by the same linear transform
  $\Rightarrow$ Matrix $\mathbf{H}$ is *big* $\Rightarrow$ Computationally intensive to operate with
- Realizations of a bandlimited graph random process $\Rightarrow$ $\mathbf{R}_x$ singular
- Enough computational power available prior to processing of signals
- Process sequence of signals fast $\Rightarrow$ Apply smaller matrix to samples
- Traditional sampling $\Rightarrow$ Ignores further processing on the signal
- Traditional sketching $\Rightarrow$ Recomputes sketch for each realization $\mathbf{x}$

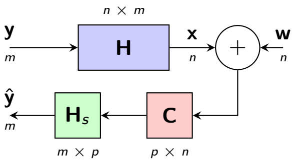


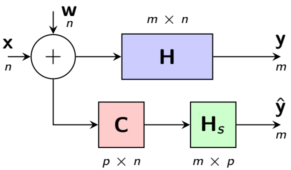
Design $\mathbf{C}, \mathbf{H}_s$ based on $\mathbf{H}$ and statistics of signal $\mathbf{R}_x$ and noise $\mathbf{R}_w$

Problem Statement

- Design a sampling matrix $\mathbf{C}$ that selects $k \leq p \ll n$ samples
- Design a deterministic sketch $\mathbf{H}_s$ to be directly applied to samples
- Joint design of sketching and sampling prior to start of sequence

  $\Rightarrow$ Minimize the MSE relative to using full $\mathbf{H}$ on the full signal $\mathbf{x}$
- Processing of signals reduces to sampling and matrix multiplication
- The computational cost of processing is reduced by a factor of p/n





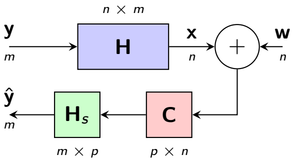
Inverse Linear Problem

- Use noisy output $(\mathbf{x} + \mathbf{w}) \in \mathbb{R}^n$ to estimate input $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} = \mathbf{H}\mathbf{y}$
- Linear model $\mathbf{H} \in \mathbb{R}^{n \times m}$ tall matrix with $m \ll n$ and full rank
- Output signal $\mathbf{x} \in \mathbb{R}^n$ is k-bandlimited with known $\mathbf{R}_x \succeq \mathbf{0}$ (singular)
- Input noise $\mathbf{w}$, indep. of $\mathbf{x}$ with known covariance matrix $\mathbf{R}_w \succ \mathbf{0}$

- Design sketch $\mathbf{H}_s^* \in \mathbb{R}^{m \times p}$ and a selection matrix $\mathbf{C}^* \in \mathbb{R}^{p \times n}$

$$\{\mathbf{C}^*, \mathbf{H}_s^*\} := \underset{C \in \mathcal{C}_{pn}, \mathbf{H}_s}{\operatorname{argmin}} \mathbb{E}\left[\|\mathbf{H}\mathbf{H}_s\mathbf{C}(\mathbf{x} + \mathbf{w}) - \mathbf{x}\|_2^2\right]$$

- Solve this problem before processing the sequence of signals



Inverse Linear Problem: Noisy case

- Two-stage optimization to solve $\min \mathbb{E}\left[\|\mathbf{H}\mathbf{H}_s\mathbf{C}(\mathbf{x}+\mathbf{w})-\mathbf{x}\|_2^2\right]$

1. Design matrix sketch $\mathbf{H}_s^* = \mathbf{H}_s^*(\mathbf{C})$ then replace on objective function

$$\mathbf{H}_s^*(\mathbf{C}) = \mathbf{A}_{LS}\mathbf{R}_x\mathbf{C}^T\left(\mathbf{C}(\mathbf{R}_x+\mathbf{R}_w)\mathbf{C}^T\right)^{-1}$$

$\Rightarrow$ This is the LS solution with a preprocessing to deal with the noise

2. Define auxiliary matrix $\mathbf{G} = \mathbf{H}\mathbf{A}_{LS}$ and obtain $\mathbf{C}^*$ by solving

$$\min_{\mathbf{C}\in\mathcal{C}_{pn}} \operatorname{tr}\left[\mathbf{R}_x - \mathbf{G}\mathbf{R}_x\mathbf{C}^T\left(\mathbf{C}(\mathbf{R}_x+\mathbf{R}_w)\mathbf{C}^T\right)^{-1}\mathbf{C}\mathbf{R}_x\mathbf{G}^T\right]$$

$\Rightarrow$ Tradeoff between output energy and noise of the selected samples

$\Rightarrow$ This is a binary optimization problem over selection matrix $\mathbf{C}$

$\Rightarrow$ There are $\binom{n}{p}$ possible solutions $\Rightarrow$ Prohibitive to test all of them

Inverse Linear Problem: Noisy case

- Binary constraints are inherent to the selection problem
- Equivalent problem with linear objective function and LMIs

  $\Rightarrow$ It would be an SDP except for binary constraint
- Observe that $\mathbf{C}^T\mathbf{C} = \text{diag}(\mathbf{c})$ $\Rightarrow$ Sampling vector $\mathbf{c} \in \{0,1\}^n$
- Define $\bar{\mathbf{C}}_\alpha = \text{diag}(\mathbf{c})/\alpha$, $\alpha > 0$ and $\bar{\mathbf{R}}_\alpha = \mathbf{R}_x + \mathbf{R}_w - \alpha \mathbf{I}_n$
- Problem over $\mathbf{C}$ can be posed as an equivalent problem over $\mathbf{c}$

$$\min_{\mathbf{c}\in\{0,1,\}^n, \mathbf{Y}, \bar{\mathbf{C}}_\alpha} \text{tr}[\mathbf{Y}]$$

$$\text{s. t. } \bar{\mathbf{C}}_\alpha = \alpha^{-1}\text{diag}(\mathbf{c}) \ , \ \mathbf{c}^T\mathbf{1}_n = p$$

$$\begin{bmatrix} \mathbf{Y} - \mathbf{R}_x + \mathbf{G}\mathbf{R}_x\bar{\mathbf{C}}_\alpha\mathbf{R}_x\mathbf{G}^T & \mathbf{G}\mathbf{R}_x\bar{\mathbf{C}}_\alpha \\ \bar{\mathbf{C}}_\alpha\mathbf{R}_x\mathbf{G}^T & \bar{\mathbf{R}}_\alpha^{-1} + \bar{\mathbf{C}}_\alpha \end{bmatrix} \succeq \mathbf{0}$$

- This is also a complicated problem but slightly more tractable

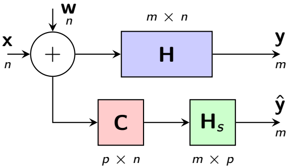
Direct Linear Problem

- Given noisy input $(\mathbf{x}+\mathbf{w}) \in \mathbb{R}^n$ estimate output $\mathbf{y} = \mathbf{H}\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$
- Design sketch $\mathbf{H}_s \in \mathbb{R}^{m\times p}$ and $p \times n$ selection matrix $\mathbf{C}$

$$\{\mathbf{C}^*, \mathbf{H}_s^*\} := \underset{C\in\mathcal{C}_{pn},\mathbf{H}_s}{\operatorname{argmin}} \mathbb{E}\left[\|\mathbf{H}_s\mathbf{C}(\mathbf{x}+\mathbf{w}) - \mathbf{H}\mathbf{x}\|_2^2\right]$$

- Two stage optimization $\Rightarrow$ Matrix sketch $\mathbf{H}_s$ and sampling scheme $\mathbf{C}$
- Can be reformulated as an equivalent problem over selection vector $\mathbf{c}$

  $\Rightarrow$ Linear objective function, LMIs constraints, binary constraint



Sampling Heuristics

- Solving sampling problems might be intractable $\Rightarrow$ Heuristic solutions
- Convex relaxation $[\mathcal{O}((n+m)^{3.5})]$ $\Rightarrow \mathbf{c} \in [0,1]^n$ $\Rightarrow$ SDPs

  $\Rightarrow$ Tresholding $\Rightarrow$ Set the p highest values to 1 and the rest to 0

  $\Rightarrow$ Random $\Rightarrow$ Use relaxed solution as distribution to select nodes
- Noise-blind Heuristic $[\mathcal{O}(n \log n)]$ $\Rightarrow$ p rows of $\mathbf{R}_x \mathbf{G}^T$ with largest $\|\cdot\|_2$

$$\min_{\mathbf{C} \in \mathcal{C}_{pn}} \operatorname{tr}\left[\mathbf{R}_x - \mathbf{G}\mathbf{R}_x\mathbf{C}^T\left(\mathbf{C}(\mathbf{R}_x + \mathbf{R}_w)\mathbf{C}^T\right)^{-1}\mathbf{C}\mathbf{R}_x\mathbf{G}^T\right]$$

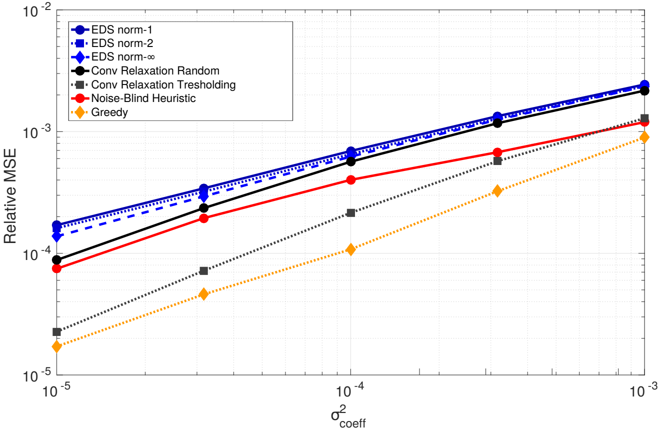
- Greedy approach $[\mathcal{O}(np(mnp + p^3))]$ $\Rightarrow$ Select best node iteratively

Example: Computing the Graph Fourier Transform

- Consider a bandlimited graph signal $\mathbf{x} = \mathbf{V}_k \tilde{\mathbf{x}}_k$ $\Rightarrow \tilde{\mathbf{x}}_k$: freq. coeff.
  - $\Rightarrow$ Inverse linear model $\Rightarrow \mathbf{x} = \mathbf{V}_k \tilde{\mathbf{x}}_k$ $\Rightarrow$ Transform $\mathbf{H} = \mathbf{V}_k$
- Sequence of noisy signals $(\mathbf{x} + \mathbf{w})$ $\Rightarrow$ Fast computation of the GFT
  - $\Rightarrow \mathbf{w}$: white gaussian zero-mean noise of power prop. to energy of $\mathbf{x}$
- Compare between different heuristics proposed for the joint design
- Compare with other traditional sampling schemes for reconstruction
  - $\Rightarrow$ Experimentally Design Sampling (EDS) technique
  - $\Rightarrow$ Assign to each node the norm of the rows of $\mathbf{V}_k$
  - $\Rightarrow$ Sample with replacement with a distribution prop. to this norm

Example: Approximating the GFT

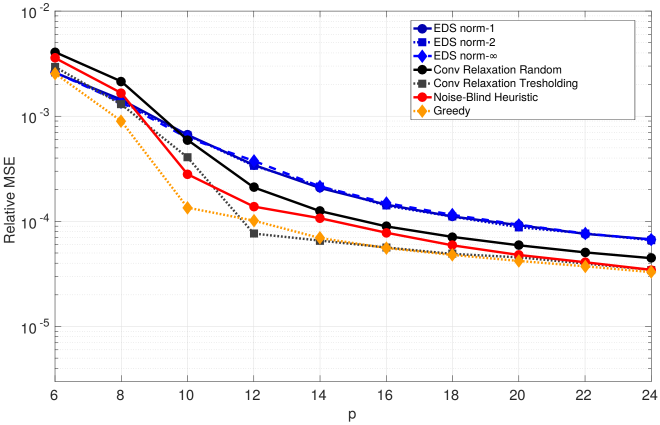
- Erdős-Rényi graph of size $n = 100$ with probablity 0.2
- Signal bandlimited with $k = 10$ freq. coeff. $\Rightarrow p = k = 10$



- Error of $2 \cdot 10^{-5}$ reducing computational complexity by 10

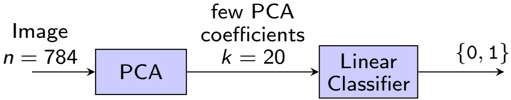
Example: Approximating the GFT

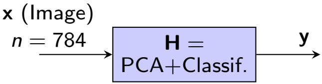
- Erdős-Rényi graph of size $n = 100$ with probablity 0.2
- Signal bandlimited with $k = 10$ freq. coeff. $\Rightarrow \sigma^2_{\text{coeff}} = 10^{-4}$



- Error of 10^{-4} reducing computational complexity by $100/24 = 4.167$

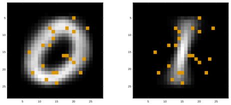
- Classify images of handwritten digits of the MNIST database
- Linear classifier in the PCA domain $\Rightarrow$ Expensive linear operation
  $\Rightarrow$ Subsume PCA and classifier in one linear operator



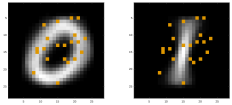


- Classify images by operating directly on a subset of pixels
- Images of size $n = 784$ pixels $\Rightarrow$ Use only $p = 20$ pixels
  $\Rightarrow$ Processing costs reduced by 39.2 for each image

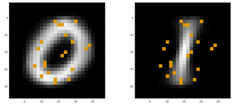
Classification of handwritten digits



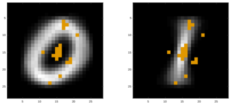
(a) EDS norm-1



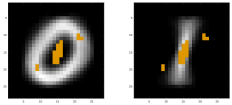
(b) EDS norm-2



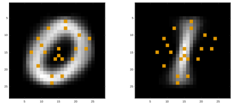
(c) EDS norm-∞



(d) Tresholding



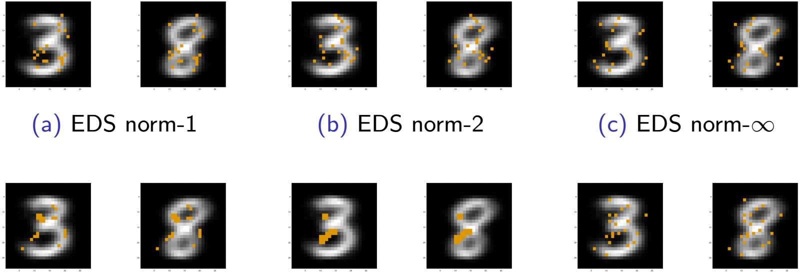
(e) Noise-Blind



(f) Greedy

- Sketching and sampling techniques achieve perfect classification

Classification of handwritten digits



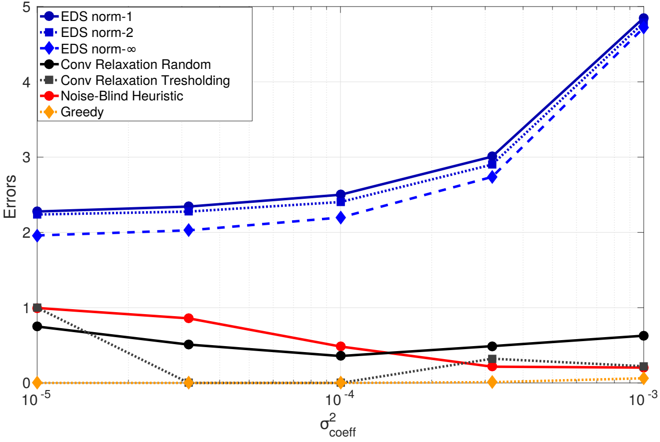
(a) EDS norm-1 (b) EDS norm-2 (c) EDS norm-∞

(d) Tresholding (e) Noise-Blind (f) Greedy

- Error rate using full image: 4.00%
  - $\Rightarrow$ Greedy approach using 20 pixels: 4.53%

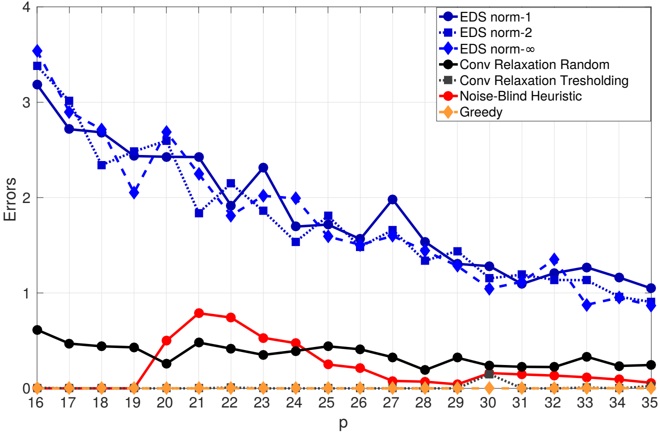
Classification of handwritten digits

- 200 image classification as a function of noise for $p = 20$ pixels



Classification of handwritten digits

- 200 image classification as a function of the number of pixels



Conclusions

- Optimal sketch and sampling for processing bandlimited graph signals
  - $\Rightarrow$ Obtain approximate solution by operating only on a few samples
  - $\Rightarrow$ Accelerate processing of a sequence of bandlimited signals
- Joint design of matrix sketch and sampling scheme (prior to processing)
  - $\Rightarrow$ Two-stage optimization $\Rightarrow$ Heuristic solutions for sampling problem
- Fast computation of GFT of a bandlimited graph signal
  - $\Rightarrow$ Errors in the order of 10^{-5} reducing the cost 10 times
- Classification of images of size 784 pixels of handwritten digits
  - $\Rightarrow$ Using as few as 20 pixels $\Rightarrow$ 40 times less computational cost
- Journal version available on arXiv: `arxiv.org/abs/1611.00119`