ETH zürich

Chapter 10

Signal Approximation and Compressed Sensing

Content

- **Signals and Sparse Representation**
  - Orthogonal Bases
  - Approximation in Orthogonal Bases
  - Reconstruction in Overcomplete Bases
- **Random Projection and Approximation**
  - Johnson-Lindenstrauss Approximation
  - Compressed Sensing
- **Equivalence between l0 and l1 Recovery**
  - Restricted Nullspace Property and Geometric Intuition
  - Restricted Isometry Property

ETH zürich

1.1. Orthogonal Bases

- (general) Signal:

e.g. data such as sea water levels, audio recordings, photographic images, video data, and financial data

represent the signal by a vector $\theta^* \in R^p$

1.1. Orthogonal Bases

- Orthonormal basis: A basis with finite dimension whose vectors are all unit vectors and orthogonal to each other
- $\{\psi_j\}_{j=1}^{p}$ orthonormal basis of $R^p \longrightarrow \Psi := [\psi_1 \; \psi_2 \ldots \psi_p]$ is a p x p matrix with $\Psi^T \Psi = I_{p \times p}$

1.1. Orthogonal Bases

- Basis coefficient $\beta^* = \Psi^T \theta^* = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \cdots \\ \psi_p \end{pmatrix} \theta^* = \begin{pmatrix} \sum_{i=1}^p \theta_i^* \psi_{i1} \\ \sum_{i=1}^p \theta_i^* \psi_{i2} \\ \cdots \\ \sum_{i=1}^p \theta_i^* \psi_{ip} \end{pmatrix}$ of a signal

$\theta^* \in R^p$ with $\theta^* := \sum_{j=1}^p \beta_j^* \psi_{ij}$

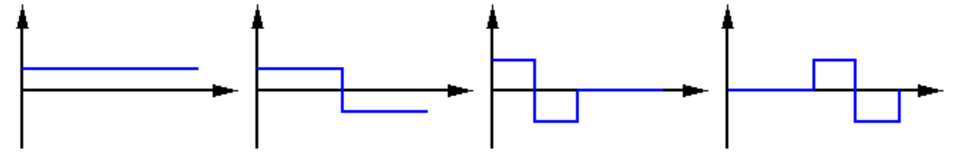
- the j-th basis coefficient can be written as a inner product

$$\beta_j^* := < \theta^*, \psi_j > = \sum_{i=1}^p \theta_j^* \psi_{ij}$$

1.1. Orthogonal Bases

- Example (wavelet transform)

$$\Psi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{-1}{2} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{-1}{2} & 0 & \frac{-1}{\sqrt{2}} \end{pmatrix}$$



http://fourier.eng.hmc.edu/e161/lectures/Haar/index.html

as orthonormal basis matrix

4 column vectors = 4 bases for Haar transform (signal length is 4)

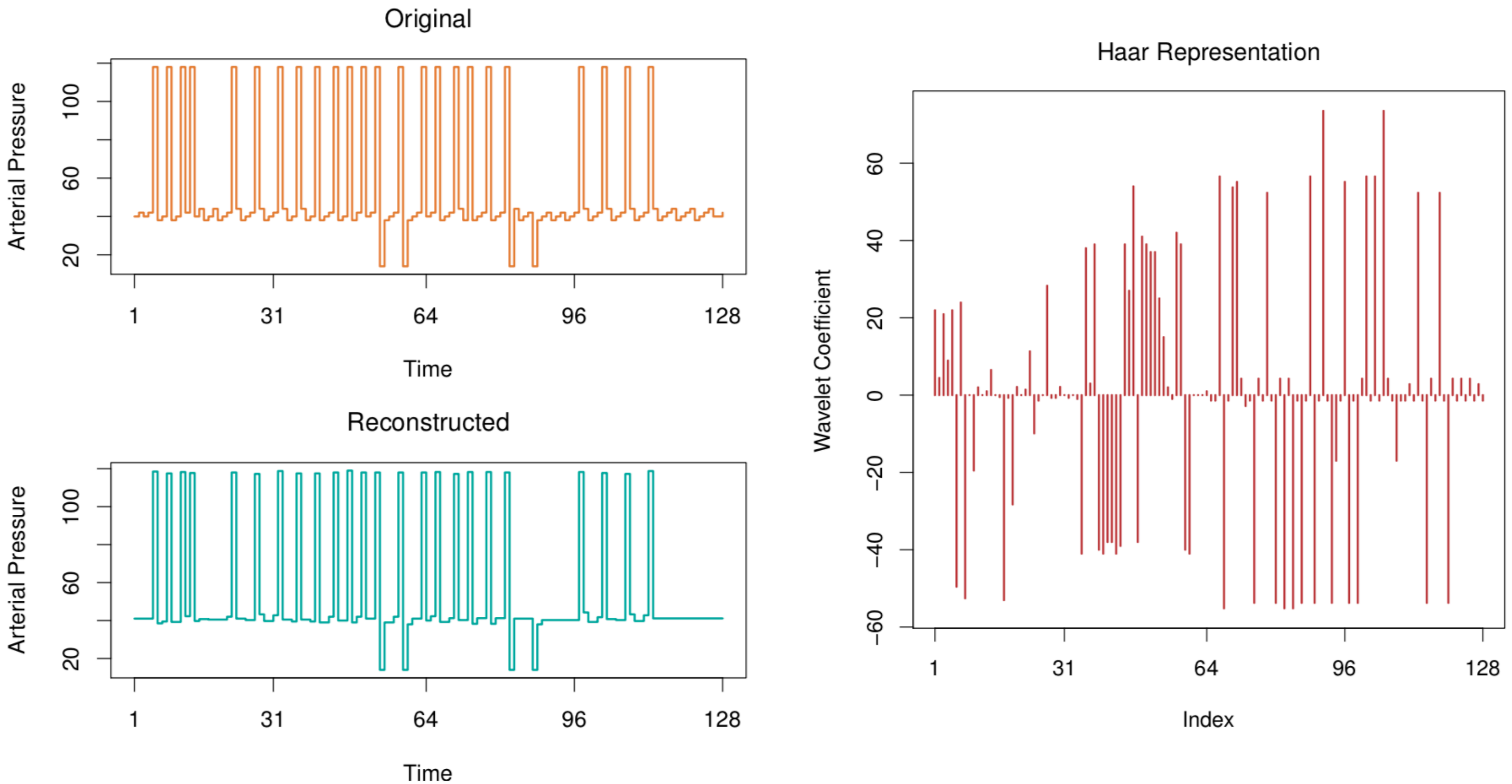


Figure 10.1 *Illustration of sparsity in time series data. Left, top panel: Signal $\theta^* \in \mathbb{R}^p$ of arterial pressure versus time over $p = 128$ points. Left, bottom panel: Reconstruction $\widehat{\theta}^{128}$ based on retaining the largest (in absolute amplitude) $k = 64$ coefficients from the Haar basis. Right: Haar basis coefficients $\beta^* = \mathbf{\Psi}^T \theta^*$ of the signal.*

Statistical Learning with Sparsity the Lasso and Generalizations, Trevor Hastie, Robert Tibshirani, Martin Wainwright, CRC Press, p. 270

1.2. Approximation in Orthogonal Bases

- Goal of signal compression: represent signal $\theta^* \in R^p$ using k<<p coefficients
- How?
- Use only sparse subset of the orthogonal vectors $\{\psi_j\}_{j=1}^p$ for $k \in \{1,\ldots,p\}$: consider reconstruction

$$\Psi\beta = \sum_{j=1}^p \beta_j \psi_j, \ \|\beta\|_0 := \sum_{j=1}^p I[\beta_j \neq 0] \leq k$$

1.2. Approximation in Orthogonal Bases

- Optimal k-sparse approximation:

- Compute: $\widehat{\beta^k} \in argmin_{\beta \in R^p} ||\theta* - \Psi^T \beta||_2^2 \; so\, that \; ||\beta||_0 \leq k$

- Reconstruction: $\theta^k := \sum_{j=1}^{p} \widehat{\beta}_j^k \psi_j$

⟶ best least-squares approximation to the signal $\theta* \in R^p$ based on k terms

⟶ non-convex and combinatorial problem

1.2. Approximation in Orthogonal Bases

- solve by taking first k coefficients with largest absolute values:

1. Order vector $\beta^* \in R^p$ of basis coefficients: $|\beta^*_{(1)}| \geq |\beta^*_{(2)}| \geq \ldots \geq |\beta^*_{(p)}|$

2. For given $k \in \{1,\ldots,p\}$ choose the first k terms: $\widehat{\theta}_k := \sum_{j=1}^k \beta^*_{(j)} \psi_{\sigma(j)}$

1.2. Approximation in Orthogonal Bases

- Complexity of the procedure

1. Compute basis coefficients $\beta_j^* := < \theta^*, \psi_j > = \sum_{i=1}^{p} \theta_j^* \psi_{ij} \longrightarrow O(p^2)$
2. Sort coefficients in terms of absolute values $\longrightarrow O(p \, log p)$
3. Extract first k coefficients
4. Compute the best k-term approximation $\widehat{\theta}_k := \sum_{j=1}^{k} \beta_{(j)}^* \psi_{\sigma(j)}$

(a) (b)

Figure 10.3 *Illustration of image compression based on wavelet thresholding. (a) Zoomed portion of the original "Boats" image from Figure 10.2(a). (b) Reconstruction based on retaining 5% of the wavelet coefficients largest in absolute magnitude. Note that the distortion is quite small, and concentrated mainly on the fine-scale features of the image.*

Statistical Learning with Sparsity the Lasso and Generalizations, Trevor Hastie, Robert Tibshirani, Martin Wainwright, CRC Press, p. 272

1.3. Reconstruction in Overcomplete Bases

- Problem: Only limited class of signals has sparse representations in ANY orthonormal bases

- Solution: Combine different orthonormal bases
  $\longrightarrow$ use subsets of vectors from both bases simultaneously

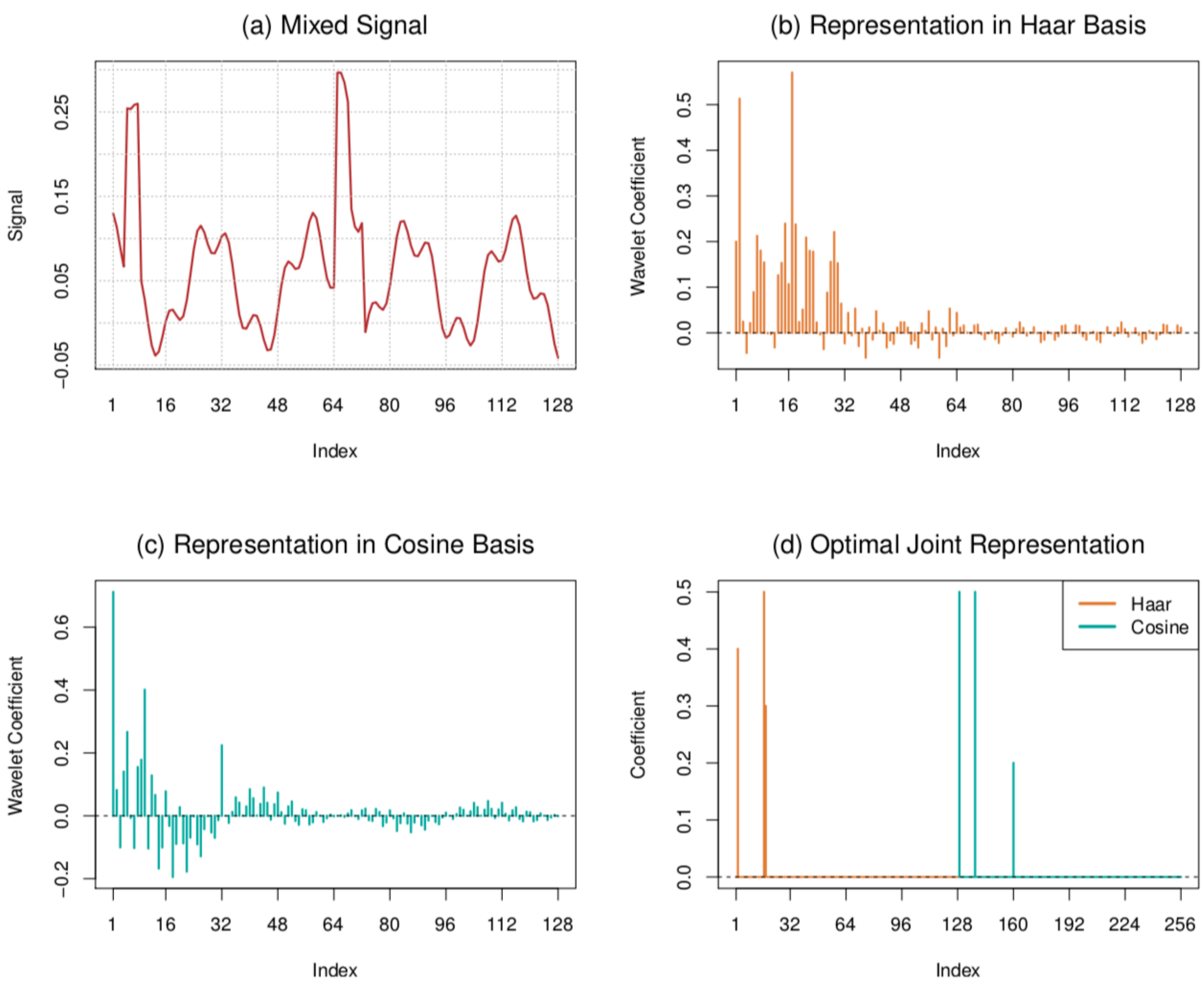


Figure 10.4 *(a) Original signal $\theta^* \in \mathbb{R}^p$ with $p = 128$. (b) Representation $\boldsymbol{\Psi}^T\theta^*$ in the Haar basis. (c) Representation $\boldsymbol{\Phi}^T\theta^*$ in the discrete cosine basis. (d) Coefficients $(\widehat{\alpha}, \widehat{\beta}) \in \mathbb{R}^p \times \mathbb{R}^p$ of the optimally sparse joint representation obtained by solving basis pursuit linear program* (10.11).

Statistical Learning with Sparsity the Lasso and Generalizations, Trevor Hastie, Robert Tibshirani, Martin Wainwright, CRC Press, p. 275

ETH zürich

1.3. Reconstruction in Overcomplete Bases

- two pairs of orthonormal bases: $\{\psi_j\}_{j=1}^{p}$, $\{\phi_j\}_{j=1}^{p}$
- reconstruction of the form:

$$\sum_{j=1}^{p} \alpha_j \phi_j + \sum_{j=1}^{p} \beta_j \psi_j \; so\, that \; \|\alpha\|_0 + \|\beta\|_0 \leq k$$

- optimisation problem:

$$minimize_{(\alpha,\beta)\in R^p \times R^p} \|\theta^* - \Phi\alpha - \Psi\beta\|_2^2 \; so\, that \; \|\alpha\|_0 + \|\beta\|_0 \leq k$$

1.3. Reconstruction in Overcomplete Bases

- difficult to solve (non convex)

- consider convex program:

$$minimize_{(\alpha,\beta)\in R^p \times R^p} \|\theta^* - \Phi\alpha - \Psi\beta\|_2^2 \; so\,that \; \|\alpha\|_1 + \|\beta\|_1 \leq R$$

⟶ constrained version of the lasso program/ basis-pursuit program

- consider even simpler problem:

$$minimize_{(\alpha,\beta)\in R^p \times R^p} \|\alpha\|_1 + \|\beta\|_1 \; so\,that \; \theta^* = [\Phi \; \Psi]\begin{bmatrix}\alpha \\ \beta\end{bmatrix}$$

⟶ linear program (LP)/ basis-pursuit linear program

1.3. Reconstruction in Overcomplete Bases

- Question: When is basis-pursuit linear program equivalent to l0 constraint problem in general?

- Answer: It depends on the degree of coherence between two bases. (s. 3.2)

2.1. Johnson-Lindenstrauss Approximation

$For\, any\, 0 < \epsilon < 1\, and\, any\, integer\, n, let\, k\, be\, a\, postive\, integer\, such\, that$

$$k \geq 4(\epsilon^2/2 - \epsilon^3/3)^{-1} ln\, n\, .$$

$Then\, for, any\, set\, V\, of\, n\, points\, in\, R^d, there\, is\, a\, map\, f\, :\, R^d\, \rightarrow\, R^k\, such\, that\, for\, all\, u, v \in\, V,$

$$(1-\epsilon)\|u-v\|_2^2 \leq \|F(u)-F(v))\|_2^2 \leq (1+\epsilon)\|u-x\|_2^2.$$

$Further\, this\, map\, can\, be\, found\, in\, polynomial\, .$

Taken from Sanjoy Dasgupta, Anupam Gupta: An elementary proof of the Johnson Lindenstrauss Lemma

2.1. Johnson-Lindenstrauss Approximation

- establish the existence of distance-preserving dimension reduction projection
- provide an explicit bound on the dimension required for approximate distance preserving
- provide an explicit construction of the random projection

2.1. Johnson-Lindenstrauss Approximation

- Definition of Random Projection:

A random projection of a signal $\theta*$ is a measurement of the form

$$y_i = < z_i, \theta* > = \sum_{j=1}^{p} z_{ij}\theta_j^*$$

with $z_i \in R^p$ random vector.

2.1. Johnson-Lindenstrauss Approximation

- PART 1
- Given: n data points (d dimension), relative error ε
- Compute: minimum dimension k with

$$k = 4(\epsilon^2/2 - \epsilon^3/3)^{-1} ln\, n$$

- larger ε smaller k
- k only depends on number of points n, not on dimension of points d

2.1. Johnson-Lindenstrauss Approximation

- PART 2
- Given: input matrix X of n points in d-dimensional space
- Compute: map F which preserves distance approximately

- F can be a simple matrix A constructed by sampling k x d i.i.d. draws from a Gaussian matrix with mean 0 and variance 1/k

$$F(u) := \frac{1}{\sqrt{k}} Zu,\ Z \in R^{k \times d}\ random\ matrix\ with\ each\ Z_{ij} \sim N(0,1)\ i.i.d$$

Compressed Sensing

- Motivation: In previous algorithm we discard most β_j^*'s, so do we really need to calculate them all?

- Oracle technique: We know which subset of k coefficients will be retained for sparse approximation $\rightarrow$ only need to compute this subset of basis coefficients

- Compressed sensing:
  - Instead of precomputing all coefficients $\beta^* = \Psi^T \theta^*$, we compute N random projections y=Zθ (with N<<p)
  - Mimics behaviour of the oracle technique with only little computational overhead.

Compressed Sensing

- Setup of our problem:
  - Given: $y \in \mathbb{R}^N$: Vector of random projections of signal θ^*
    $Z \in \mathbb{R}^{N \times p}$: Design matrix used to compute random projections
  - Goal: Recover signal $\theta^* \in \mathbb{R}^p$
  - Problem: $y = Z\theta$ is highly underdetermined as $N << p$

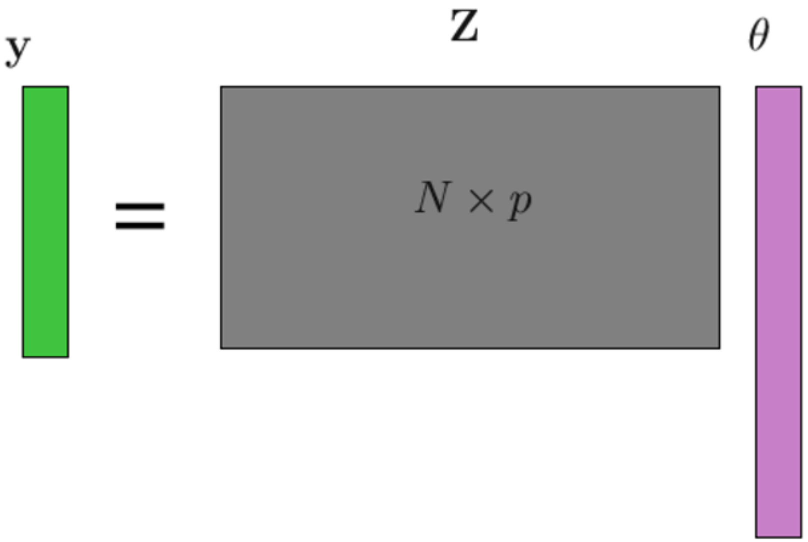
Example :

$y = x_1 + x_2$ if y=1:

$x_1 = 1$ and $x_2 = 0$

$x_1 = 0.5$ and $x_2 = 0.5$

etc...

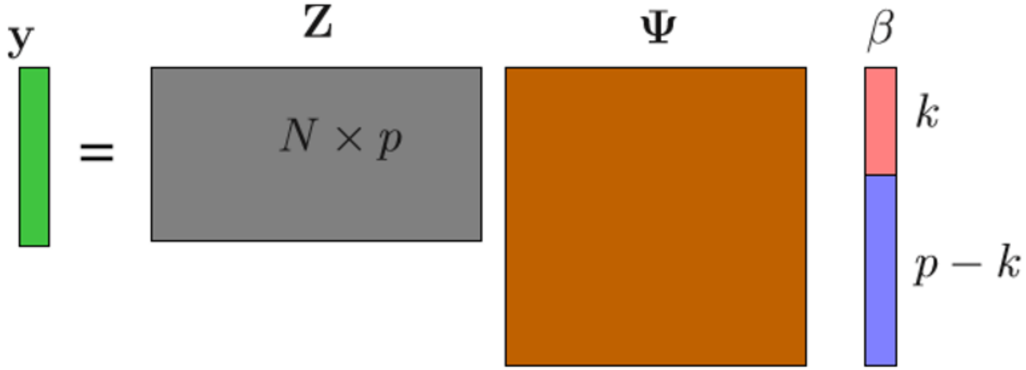


(a)

$$\underset{\theta \in \mathbb{R}^p}{\text{minimize}} \left\| \Psi^T \theta \right\|_0 \quad \text{such that } y = Z\theta$$

$$\Big\downarrow \ \ell_1 - \text{relaxation}$$

$$\underset{\theta \in \mathbb{R}^p}{\text{minimize}} \left\| \Psi^T \theta \right\|_1 \quad \text{such that } y = Z\theta$$



(b)

Equivalently we can write:

$$\underset{\beta \in R^p}{\text{minimize}} \left\| \beta \right\|_1 \quad \text{such that } y = \tilde{Z}\beta \text{ where } \tilde{Z} = Z\Psi$$

Picture: Trevor Hastie, Robert Tibshirani, Martin Wainwright: Statistical Learning with Sparsity the Lasso and Generalizations, CRC Press, 2015

Compressed Sensing

1. For given sample size N, compute random projections $y_i = \left\langle z_i, \theta^* \right\rangle$

   (Or ideally measure the random projections y instead of full signal θ^*)

2. Estimate θ^* by solving linear program:

   $$\underset{\theta \in \mathbb{R}^p}{\text{minimize}} \left\| \Psi^T \theta \right\|_1 \text{ such that } y = Z\theta, \text{ to obtain } \hat{\theta}.$$

- Family of procedures depending on random projection vectors

  simplest choice: $z_{ij} \sim N(0, \frac{1}{N})$

- Combination of random projection and l_1 – relaxation

Compressed Sensing

- Before: $\underset{\theta \in \mathbb{R}^p}{\text{minimize}} \left\| \Psi^T \theta \right\|_1$ such that $y = Z\theta$
- In reality we often have noisy data: $y = Z\theta^* + \varepsilon$ with $\left\| \varepsilon \right\|_2 \leq e$.

$\rightarrow$ We introduce a bound e > 0 on the noise level :

$$\underset{\theta \in \mathbb{R}^p}{\text{minimize}} \left\| \Psi^T \theta \right\|_1 \text{ such that } \left\| y - Z\theta \right\|_2 \leq e.$$

Sensing matrices in Compressed Sensing

Random Sensing Matrices $Z \in \mathbb{R}^{N \times p}$:

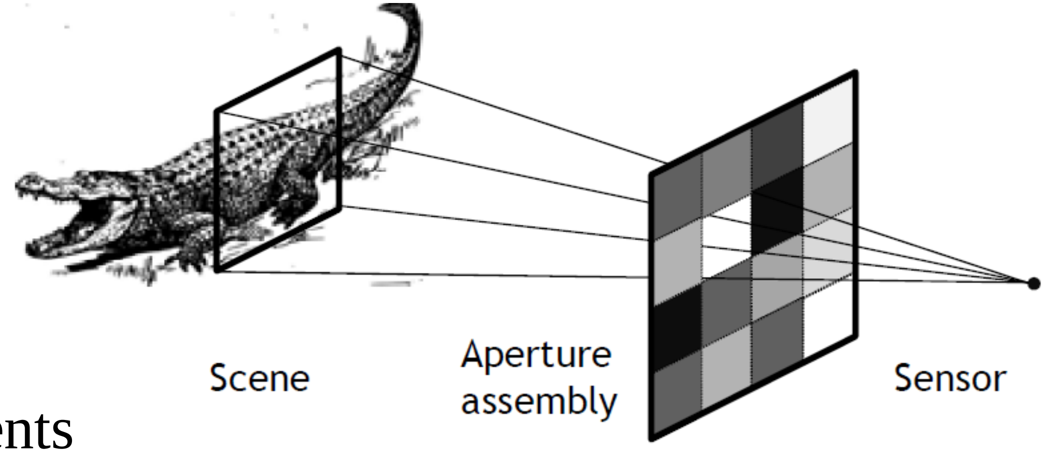
- Gaussian Random Matrix: $z_{i,j} \sim N(0,1/N)$
- Bernoulli Random Matrix: $z_{i,j} = \begin{cases} \frac{1}{\sqrt{N}} \\ -\frac{1}{\sqrt{N}} \end{cases}$
- Satisfy RIP with almost optimal order $\rightarrow$ allows sparse recovery using l_1-minimization.
- No fast matrix multiplication that may speed up the algorithm is available.
- Storing an unstructured matrix may be difficult

$\Bigg\}$ large scale problems not practicable with Gaussian/Bernoulli matrices

$\rightarrow$ Deterministic Sensing Matrices

Application: Lensless Imaging

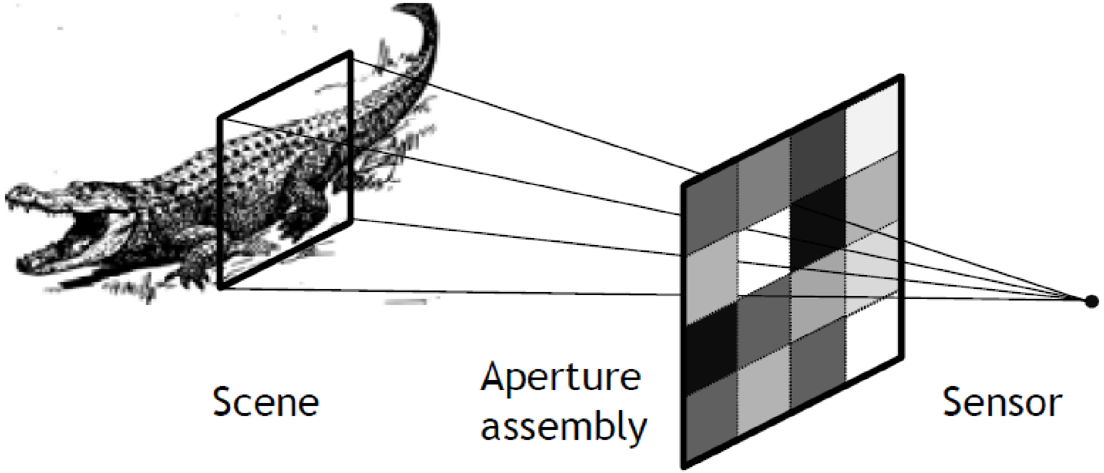
Picture and Information from:
G. Huang, H Jiang, K. Matthews, P. Wilford: Lensless Imaging by Compressive Sensing, arXiv, 2013.



Architecture:

- Aperture assembly:
  - Consists of 2 dimensional array of aperture elements
  - The transmittance of each aperture element is independently controllable.
- Sensor:
  - Used for taking compressive measurements
- Compressive sensing matrix
  - Implemented by adjusting transmittance of individual aperture elements according to the values of the sensing matrix.

Application: Lensless Imaging



How it works:

- One Element of aperture together with sensor

$\rightarrow$ cone of rays.

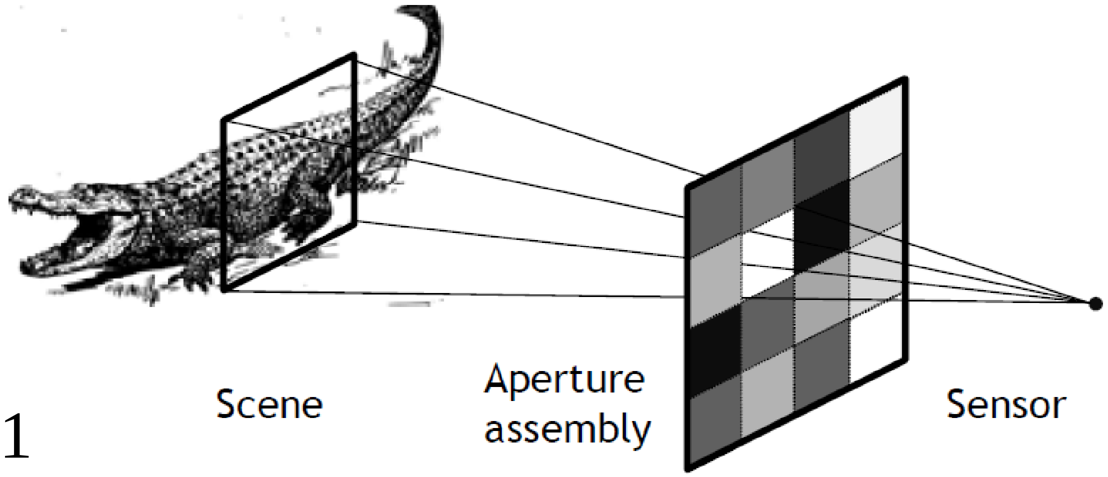
- Integration of rays in this cone is defined as a pixel value of the image

Example: "traditional way of representing an image, pixel by pixel"

1. Open one aperture element and close all others
2. Take a measurement, get one pixel value for that aperture element
3. Repeat that for all aperture elements

$\rightarrow$ We take as many measurements as number of pixels

Compressive Meassurements



- Goal: #Measurements << #Pixels

1. Define a sensing matrix:
   - Entries = random numbers between 0 and 1
   - Each row defines a pattern for the aperture elements
   - #Colums = #Elements in aperture assembly

   $\rightarrow$ each element in a row corresponds to one aperture element
2. Sensor takes one measurement for each pattern

   (integrates all rays after their intensity is modulated by transmittances)

$\rightarrow$ measurement = projection of image onto row of sensing matrix

- We take less measurements than pixels and try to reconstruct the image using CS

Compressive Meassurements

Picture from: G. Huang, H Jiang, K. Matthews, P. Wilford: Lensless Imaging by Compressive Sensing, arXiv, 2013.

Figure 5. Reconstructed images of "Soccer", 12.5%.

- Goal: #Measurements << #Pixels

1. Define a sensing matrix:
   - Entries = random numbers between 0 and 1
   - Each row defines a pattern for the aperture elements
   - #Colums = #Elements in aperture assembly

   → each element in a row corresponds to one aperture element

2. Sensor takes one measurement for each pattern

   (integrates all rays after their intensity is modulated by transmittances)

→ measurement = projection of image onto row of sensing matrix

- We take less measurements than pixels and try to reconstruct the image

Other Applications

- X-Ray (to reduce radiation exposure)
- Single Pixel Camera
- Magnetics Resonance Imaging
- Radar
- Machine learning

⋮

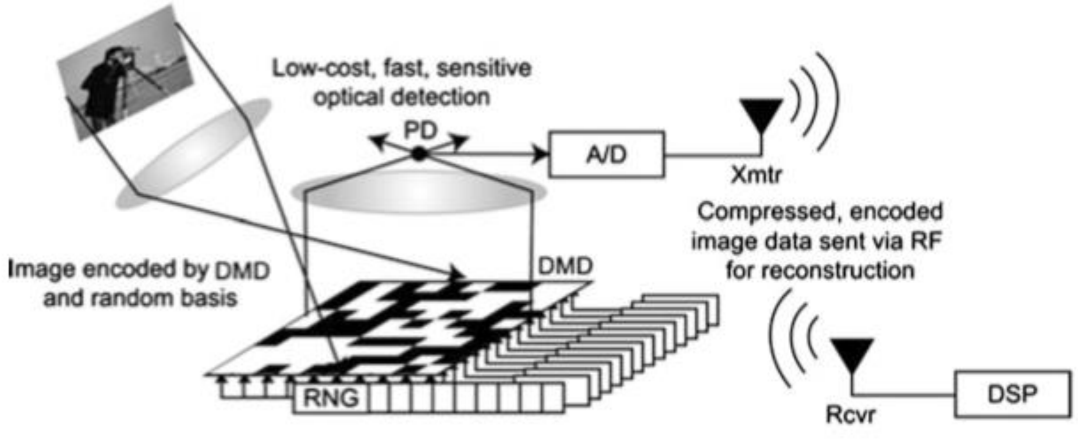


Fig. 1.5 Schematic representation of a single-pixel camera (Image courtesy of Rice University)

https://www.google.ch/search?q=single+pixel+camera&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjWzob_-I3fAhXWFsAKHQlfDvwQ_AUIDigB&biw=1536&bih=747#imgrc=-4OBKTuGUIgQpM:

Figure 3. Prototype device. Top: lab setup. Bottom left: the LCD screen as the aperture assembly. Bottom right: the sensor board with two sensors, indicated by the red circle.

Picture:
G. Huang, H Jiang, K. Matthews, P. Wilford: Lensless Imaging by Compressive Sensing, arXiv, 2013.

Equivalence between l_0 and l_1 Recovery

Question

When is solving the l_1-relaxation equivalent to solving the original l_0-problem?

Equivalence between l_0 and l_1 Recovery

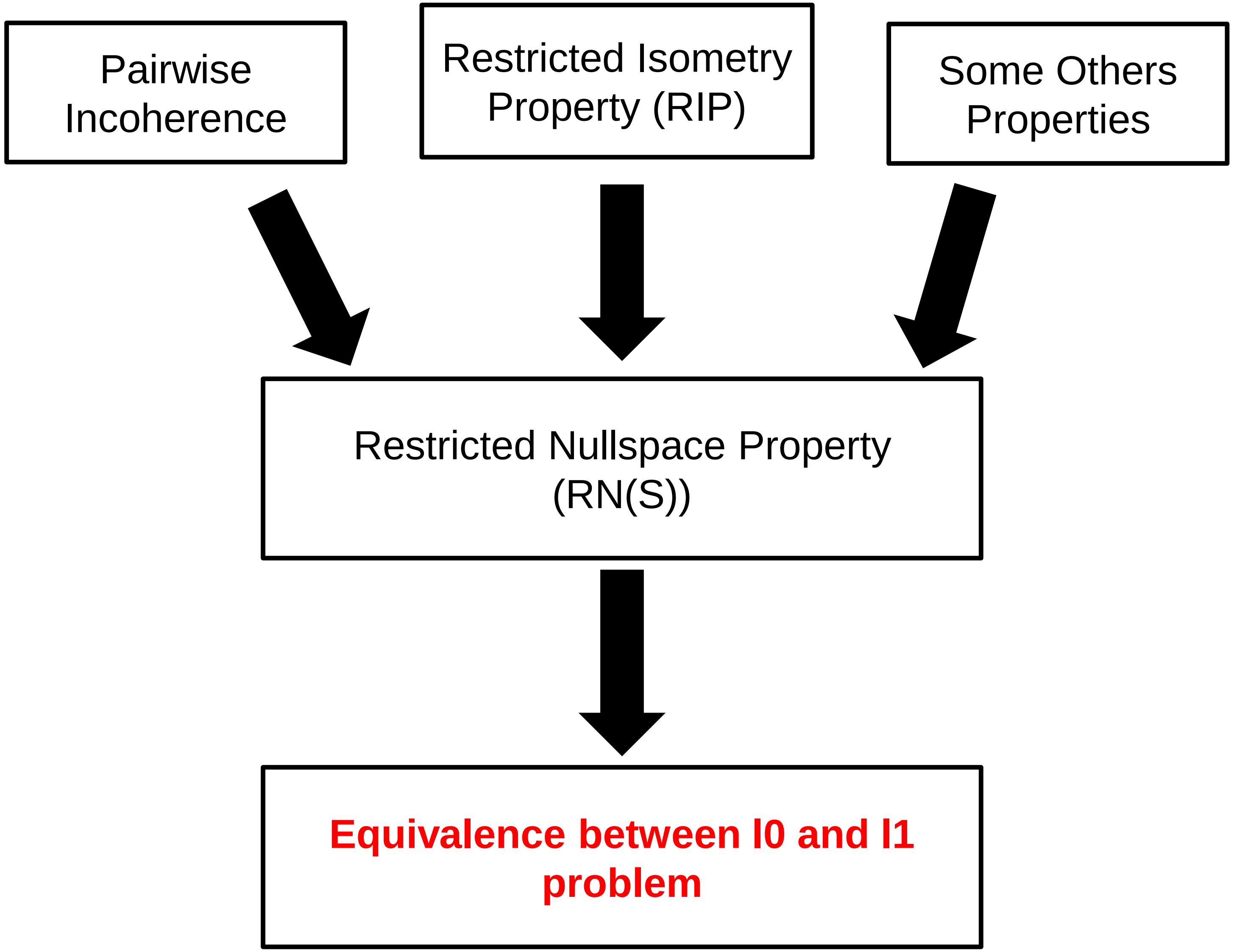
Given: $y \in \mathbb{R}^N$ and $X \in \mathbb{R}^{Nxp}$

l_0 problem: $\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \|\beta\|_0$ such that $X\beta = y$

l_1 problem: $\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \|\beta\|_1$ such that $X\beta = y$

Example: Compressed Sensing

$X = \tilde{Z} = Z\Psi$


Pairwise Incoherence
Restricted Isometry Property (RIP)
Some Others Properties
Restricted Nullspace Property (RN(S))
Equivalence between l0 and l1 problem

Restricted Nullspace Property

An $N \times p$ matrix X satisfies the *restricted nullspace property* for a set $S \subseteq \{1, \ldots, p\}$ if

$$\|\beta_S\|_1 < \|\beta_{S^C}\|_1 \text{ for all } \beta \in \ker(X) \setminus \{0\}$$

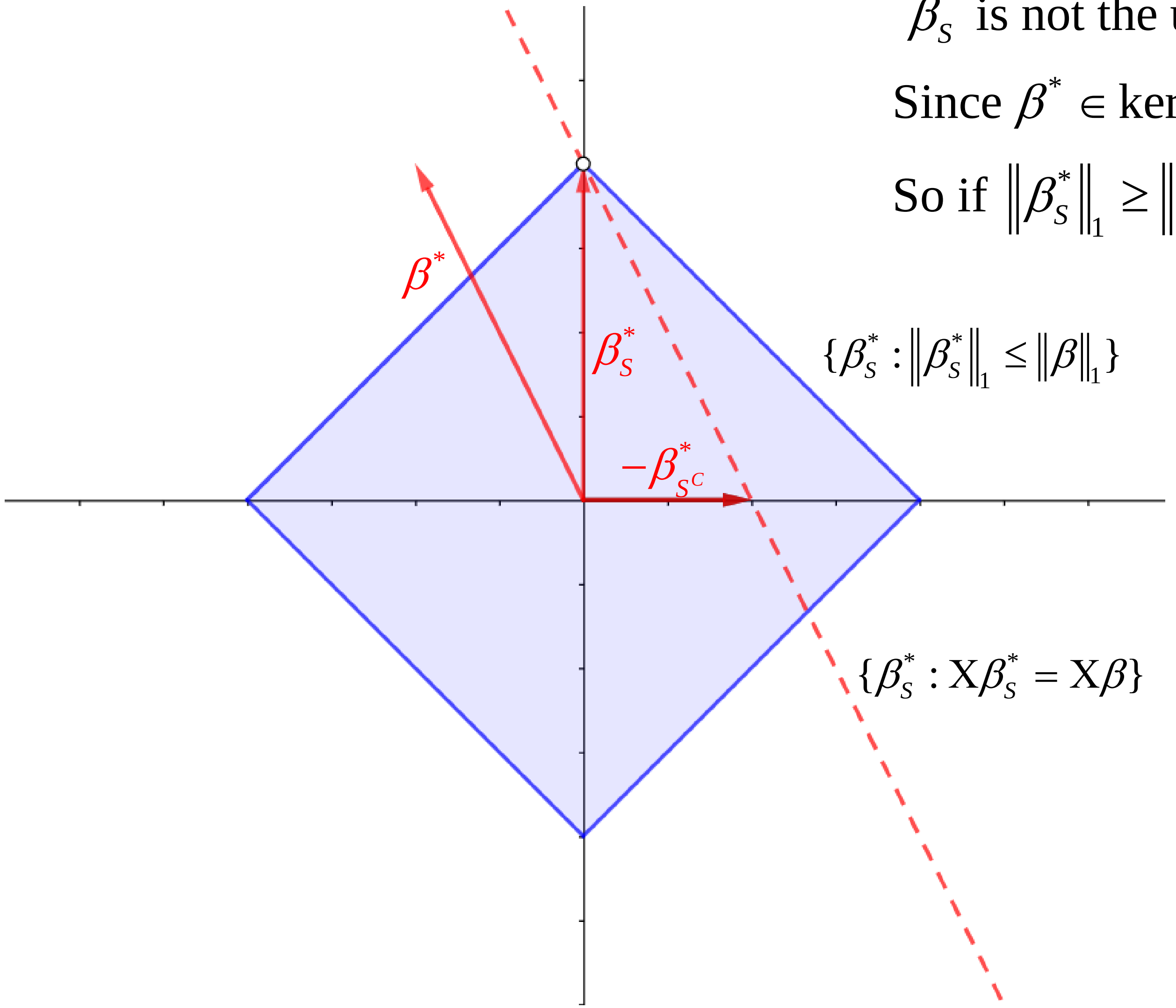
It is said to satisfy the null space property of order k if it satisfies the Nullspace property for any set S with $\mathrm{card}(S) \leq k$.

Theorem : Equivalence between l_0 and l_1 Recovery

If a given matrix $\mathrm{X} \in \mathbb{R}^{N \times p}$ satisfies the null space property for a set S, every vector $\beta^* \in \mathbb{R}^p$ supported on this set S is the unique solution of the l_1 – problem with $y = \mathrm{X}\beta^*$.

- l_1 problem: $\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \|\beta\|_1$ such that $y = \mathrm{X}\beta$

What goes wrong if $\|\beta_S\|_1 \geq \|\beta_{S^C}\|_1$?



Chose $\beta^* \in \ker(X)\backslash\{0\}$ and show:

β_S^* is not the unique minimizer of $\|\beta\|_1$ subject to $X\beta = X\beta_S^*$

Since $\beta^* \in \ker(X)$ we have: $X(\beta_S^* + \beta_{S^C}^*) = 0 \rightarrow X\beta_S^* = -X\beta_{S^C}^*$

So if $\|\beta_S^*\|_1 \geq \|\beta_{S^C}^*\|$ then β_S^* is not a unique solution anymore.

Picture selfmade, inspired by: https://dustingmixon.wordpress.com/2013/10/28/a-geometric-intuition-for-the-null-space-property/

Varifying the restricted nullspace property

Restricted Isometry Property (RIP) :

For tolerance $\delta \in (0,1)$ and $2k \in \{1,...,p\}$ we say that $RIP(2k,\delta)$ holds if

$$\frac{\|X_S u\|_2^2}{\|u\|_2^2} \in [1-\delta, 1+\delta] \text{ for all } u \in \mathbb{R}^k \setminus \{0\}$$

for all subsets $S \subset \{1,...,p\}$ of kardinality $2k$.

- Intuition:
  - RIP holds if X_S changes length of vectors very little, Eigenvalues close to 1
  - Every set of columns of size at most $2k$ approximatly behaves like orthonormal system

RIP implies restricted nullspace property

$\text{RIP}(2k,\delta)$ holds with $\delta < \frac{1}{3}$ $\rightarrow$ Nullspace property of order k holds

$\rightarrow$ the l_1 – relaxation is exact for all vectors supported on at most k elements.

- Advantage: RIP constant δ does not depend on k.
- Problem: constraint on a huge number of submatrices, $\binom{p}{2k}$ in total.
- Various choices of random projection matrix X satisfy RIP with high probability as long as $N \succsim k \log \frac{ep}{k}$. (only small multiplicative overhead compared to oracle)

Proof of Theorem:

- Suppose X satisfies RN(S) property.
- β_0, β_1 : optimal solution to $l_0, l_1 -$ problem.
- Define error vector $\Delta := \beta_0 - \beta_1$
- Goal: Show that $\Delta=0$, for that we try to show: $\Delta \in \ker(X)$ and $\|\Delta_S\|_1 \geq \|\Delta_{S^C}\|_1$
- We have: $X\beta_0 = y = X\beta_1 \rightarrow X(\beta_o - \beta_1) = 0 \rightarrow \Delta \in \ker(X)$
- Writing $\beta_0 = \beta_1 + \Delta,$ we get:

$$\|\beta_{0,S}\|_1 \geq \|\beta_1\|_1 = \|\beta_{0,S} + \Delta_S\|_1 + \|\Delta_{S^C}\|_1 \geq \|\beta_{0,S}\|_1 - \|\Delta_S\|_1 + \|\Delta_{S^C}\|_1$$

- Rearranging terms $\rightarrow \|\Delta_S\|_1 \geq \|\Delta_{S^C}\|_1$

Summary

- Sparse Representation of Signals
  - Sparsity of a Signal depends on the Choice of the Basis.
  - Reconstruction in Overcomplete Bases
- Random Projection
  - Johnson-Lindenstrauss Approximation
  - R-Code showing distance preserving property
- Compressed Sensing
  - R-Code showing recovery of an Image
  - Applications
- Equivalence between l0 and l1 Recovery
  - Restricted Nullspace Property and Geometric Intuition
  - Restricted Isometry Proeprty

Reference

- Restricted Nullspace Property: https://dustingmixon.wordpress.com/2013/10/28/a-geometric-intuition-for-the-null-space-property/
- G. Huang, H Jiang, K. Matthews, P. Wilford: Lensless Imaging by Compressive Sensing, arXiv, 2013.
- C. Hegde, R. Baraniuk, M. A. Davenport, M. F. Duarte: An Introduction to Compressive Sensing. OpenStax CNX. 27. Aug. 2014 http://cnx.org/contents/f70b6ba0-b9f0-460f-8828-e8fc6179e65f@5.12
- Trevor Hastie, Robert Tibshirani, Martin Wainwright: Statistical Learning with Sparsity the Lasso and Generalizations, CRC Press, 2015
- Sanjoy Dasgupta, Anupam Gupta: An elementary proof of the Johnson Lindenstrauss Lemma
- E. Candès, J. Romberg, and T. Tao: Signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. vol. 59, no. 8, 2005, pp. 1207-1223.
- E. Candès: Compressive sampling, Proceedings of the International Congress of Mathematicians, Madrid, August 22-30, 2006: invited lectures. 2006.