Compressed Sensing and Matrix Problems

In-Jae Kim

in-jae.kim@mnsu.edu

Department of Mathematics and Statistics

Minnesota State University, Mankato

August 13, 2009

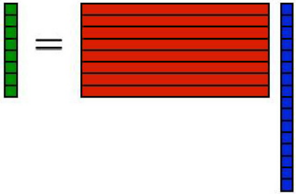
1. Introduction
   - How to understand compressed sensing as a linear algebraist
   - Recovering k-sparse signals
2. Recovering compressible signals
   - Compressible signals
   - Performance over K
3. Instance-Optimality
   - What is instance-optimality?
   - Null Space Property
   - Restricted Isometry Property
   - Good Sensing Matrices
4. Deterministic Construction of Sensing Matrices

Compressed Sensing in signal processing

- Compressed sensing is a recent concept in signal processing where one seeks to reconstruct efficiently a sparse (or compressible) signal from a minimal number of linear and non-adaptive measurements.

Compressed Sensing to Me

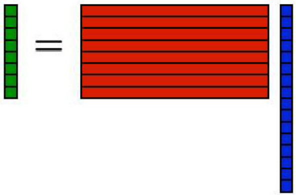
- Construct a underdetermined linear system $y = \Phi x$ with coefficient matrix Φ having many columns and relatively small number of rows, and solve this linear system



- sparse signal = column vector in $\mathbb{R}^N$ with at most k (significant) nonzero entries ($k << N$)
- n measurements = n inner products ($n << N$)

Compressed Sensing to Me

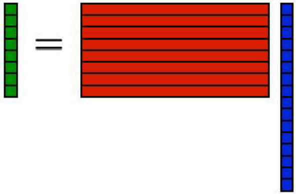
- Construct a underdetermined linear system $y = \Phi x$ with coefficient matrix Φ having many columns and relatively small number of rows, and solve this linear system



- sparse signal = column vector in $\mathbb{R}^N$ with at most k (significant) nonzero entries ($k << N$)
- n measurements = n inner products ($n << N$)

Compressed Sensing to Me

- Construct a underdetermined linear system $y = \Phi x$ with coefficient matrix Φ having many columns and relatively small number of rows, and solve this linear system



- sparse signal = column vector in $\mathbb{R}^N$ with at most k (significant) nonzero entries ($k << N$)
- n measurements = n inner products ($n << N$)

Signals with at most k nonzero entries

- $\Sigma_k = \{x \in \mathbb{R}^N \mid \#\text{supp}(x) \leq k\}$
- $\mathcal{F}(y) = \{x \in \mathbb{R}^N \mid \Phi x = y\} = x_0 + \mathcal{N}(\Phi)$

Theorem

If Φ is an $n \times N$ real matrix, then the following are equivalent:

(i) *$\mathcal{F}(y)$ contains at most one element in Σ_k for each $y \in \mathbb{R}^n$.*

(ii) $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$

(iii) *For any set T with $\#T = 2k$, the matrix Φ_T has rank $2k$.*

(iv) *For any set T with $\#T = 2k$, the $2k \times 2k$ matrix $\Phi_T^t \Phi_T$ is positive definite.*

Signals with at most k nonzero entries

- $\Sigma_k = \{x \in \mathbb{R}^N \mid \#\mathrm{supp}(x) \leq k\}$
- $\mathcal{F}(y) = \{x \in \mathbb{R}^N \mid \Phi x = y\} = x_0 + \mathcal{N}(\Phi)$

Theorem

If Φ is an $n \times N$ real matrix, then the following are equivalent:

(i) *$\mathcal{F}(y)$ contains at most one element in Σ_k for each $y \in \mathbb{R}^n$.*

(ii) $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$

(iii) *For any set T with $\#T = 2k$, the matrix Φ_T has rank $2k$.*

(iv) *For any set T with $\#T = 2k$, the $2k \times 2k$ matrix $\Phi_T^t \Phi_T$ is positive definite.*

Signals with at most k nonzero entries

- $\Sigma_k = \{x \in \mathbb{R}^N \mid \#\mathrm{supp}(x) \le k\}$
- $\mathcal{F}(y) = \{x \in \mathbb{R}^N \mid \Phi x = y\} = x_0 + \mathcal{N}(\Phi)$

Theorem

If Φ is an $n \times N$ real matrix, then the following are equivalent:

(i) *$\mathcal{F}(y)$ contains at most one element in Σ_k for each $y \in \mathbb{R}^n$.*

(ii) $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$

(iii) *For any set T with $\#T = 2k$, the matrix Φ_T has rank $2k$.*

(iv) *For any set T with $\#T = 2k$, the $2k \times 2k$ matrix $\Phi_T^t \Phi_T$ is positive definite.*

Signals with at most k nonzero entries

Theorem

If Φ is an $n \times N$ real matrix, then the following are equivalent:

(i) $\mathcal{F}(y)$ *contains at most one element in* Σ_k *for each* $y \in \mathbb{R}^n$.

(ii) $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$

Proof.

- Assume (i). To prove (ii), let $\eta \in \mathcal{N}(\Phi) \cap \Sigma_{2k}$. We can write $\eta = \eta_0 - \eta_1$ with $\eta_0, \eta_1 \in \Sigma_k$. Then $\Phi\eta_0 = \Phi\eta_1$ and hence, by (i), $\eta_0 = \eta_1$. This implies that $\eta = 0$.
- Assume (ii) and let $y \in \mathbb{R}^n$. Suppose that there are two vectors $x_1, x_2 \in \Sigma_k$ with $\Phi x_1 = \Phi x_2 = y$. Then $\eta = x_1 - x_2 \in \mathcal{N}(\Phi) \cap \Sigma_{2k}$ and by (ii), $\eta = 0$. Thus, $x_1 = x_2$. □

Signals with at most k nonzero entries

Theorem

If Φ is an $n \times N$ real matrix, then the following are equivalent:

(i) $\mathcal{F}(y)$ *contains at most one element in* Σ_k *for each* $y \in \mathbb{R}^n$.

(ii) $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$

Proof.

- Assume (i). To prove (ii), let $\eta \in \mathcal{N}(\Phi) \cap \Sigma_{2k}$. We can write $\eta = \eta_0 - \eta_1$ with $\eta_0, \eta_1 \in \Sigma_k$. Then $\Phi\eta_0 = \Phi\eta_1$ and hence, by (i), $\eta_0 = \eta_1$. This implies that $\eta = 0$.
- Assume (ii) and let $y \in \mathbb{R}^n$. Suppose that there are two vectors $x_1, x_2 \in \Sigma_k$ with $\Phi x_1 = \Phi x_2 = y$. Then $\eta = x_1 - x_2 \in \mathcal{N}(\Phi) \cap \Sigma_{2k}$ and by (ii), $\eta = 0$. Thus, $x_1 = x_2$. □

Signals with at most k nonzero entries

Theorem

If Φ is an $n \times N$ real matrix, then the following are equivalent:

(i) $\mathcal{F}(y)$ *contains at most one element in* Σ_k *for each* $y \in \mathbb{R}^n$.

(ii) $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$

Proof.

- Assume (i). To prove (ii), let $\eta \in \mathcal{N}(\Phi) \cap \Sigma_{2k}$. We can write $\eta = \eta_0 - \eta_1$ with $\eta_0, \eta_1 \in \Sigma_k$. Then $\Phi\eta_0 = \Phi\eta_1$ and hence, by (i), $\eta_0 = \eta_1$. This implies that $\eta = 0$.
- Assume (ii) and let $y \in \mathbb{R}^n$. Suppose that there are two vectors $x_1, x_2 \in \Sigma_k$ with $\Phi x_1 = \Phi x_2 = y$. Then $\eta = x_1 - x_2 \in \mathcal{N}(\Phi) \cap \Sigma_{2k}$ and by (ii), $\eta = 0$. Thus, $x_1 = x_2$. $\square$

Signals with at most k nonzero entries

- For given k, how many measurements do we need to take in order to recover every k-sparse vector?
- In another word what would be the smallest value of n for which we can construct a matrix such that any $2k$ of its N columns are linearly independent?
- $n = 2k$
- We can use a Vandermonde matrix Φ with $a_1 < a_2 < \cdots < a_N$

$$\Phi_{2k\times N} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_N \\ \vdots & \vdots & \vdots & \vdots \\ a_1^{2k-1} & a_2^{2k-1} & \cdots & a_N^{2k-1} \end{bmatrix}$$

Signals with at most k nonzero entries

- For given k, how many measurements do we need to take in order to recover every k-sparse vector?
- In another word what would be the smallest value of n for which we can construct a matrix such that any $2k$ of its N columns are linearly independent?
- $n = 2k$
- We can use a Vandermonde matrix Φ with $a_1 < a_2 < \cdots < a_N$

$$\Phi_{2k \times N} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_N \\ \vdots & \vdots & \vdots & \vdots \\ a_1^{2k-1} & a_2^{2k-1} & \cdots & a_N^{2k-1} \end{bmatrix}$$

Signals with at most k nonzero entries

- For given k, how many measurements do we need to take in order to recover every k-sparse vector?
- In another word what would be the smallest value of n for which we can construct a matrix such that any $2k$ of its N columns are linearly independent?
- $n = 2k$
- We can use a Vandermonde matrix Φ with $a_1 < a_2 < \cdots < a_N$

$$\Phi_{2k\times N} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_N \\ \vdots & \vdots & \vdots & \vdots \\ a_1^{2k-1} & a_2^{2k-1} & \cdots & a_N^{2k-1} \end{bmatrix}$$

Signals with at most k nonzero entries

- For given k, how many measurements do we need to take in order to recover every k-sparse vector?
- In another word what would be the smallest value of n for which we can construct a matrix such that any $2k$ of its N columns are linearly independent?
- $n = 2k$
- We can use a Vandermonde matrix Φ with $a_1 < a_2 < \cdots < a_N$

$$\Phi_{2k\times N} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_N \\ \vdots & \vdots & \vdots & \vdots \\ a_1^{2k-1} & a_2^{2k-1} & \cdots & a_N^{2k-1} \end{bmatrix}$$

Decoder

Goal: $\Delta(y) = \text{Argmin}_{z \in \Sigma_k} ||y - \Phi z||_{\ell_2^n}$

- $X_T = \{z \mid \text{supp}(z) \subseteq T\}$
- $x_T = \text{Argmin}_{z \in X_T} ||y - \Phi z||_{\ell_2^n}$, equivalently, $x_T = [\Phi_T^t \Phi_T]^{-1} \Phi_T^t y$ (decouple all k-dimensional subspace and solve the minimum problem)
- $T^* = \text{Argmin}_{|T|=k} ||y - \Phi x_T||_{\ell_2^n}$
- $\Delta(y) = x_{T^*}$ (k-sparse)

Decoder

Goal: $\Delta(y) = \text{Argmin}_{z \in \Sigma_k} ||y - \Phi z||_{\ell_2^n}$

- $X_T = \{z \mid \text{supp}(z) \subseteq T\}$
- $x_T = \text{Argmin}_{z \in X_T} ||y - \Phi z||_{\ell_2^n}$, equivalently, $x_T = [\Phi_T^t \Phi_T]^{-1} \Phi_T^t y$ (decouple all k-dimensional subspace and solve the minimum problem)
- $T^* = \text{Argmin}_{|T|=k} ||y - \Phi x_T||_{\ell_2^n}$
- $\Delta(y) = x_{T^*}$ (k-sparse)

Decoder

Goal: $\Delta(y) = \text{Argmin}_{z \in \Sigma_k} ||y - \Phi z||_{\ell_2^n}$

- $X_T = \{z \mid \text{supp}(z) \subseteq T\}$
- $x_T = \text{Argmin}_{z \in X_T} ||y - \Phi z||_{\ell_2^n}$, equivalently, $x_T = [\Phi_T^t \Phi_T]^{-1} \Phi_T^t y$ (decouple all k-dimensional subspace and solve the minimum problem)
- $T^* = \text{Argmin}_{|T|=k} ||y - \Phi x_T||_{\ell_2^n}$
- $\Delta(y) = x_{T^*}$ (k-sparse)

Decoder

Goal: $\Delta(y) = \text{Argmin}_{z \in \Sigma_k} ||y - \Phi z||_{\ell_2^n}$

- $X_T = \{z \mid \text{supp}(z) \subseteq T\}$
- $x_T = \text{Argmin}_{z \in X_T} ||y - \Phi z||_{\ell_2^n}$, equivalently, $x_T = [\Phi_T^t \Phi_T]^{-1} \Phi_T^t y$ (decouple all k-dimensional subspace and solve the minimum problem)
- $T^* = \text{Argmin}_{|T|=k} ||y - \Phi x_T||_{\ell_2^n}$
- $\Delta(y) = x_{T^*}$ (k-sparse)

Decoder

Goal: $\Delta(y) = \text{Argmin}_{z \in \Sigma_k} ||y - \Phi z||_{\ell_2^n}$

- $X_T = \{z \mid \text{supp}(z) \subseteq T\}$
- $x_T = \text{Argmin}_{z \in X_T} ||y - \Phi z||_{\ell_2^n}$, equivalently, $x_T = [\Phi_T^t \Phi_T]^{-1} \Phi_T^t y$ (decouple all k-dimensional subspace and solve the minimum problem)
- $T^* = \text{Argmin}_{|T|=k} ||y - \Phi x_T||_{\ell_2^n}$
- $\Delta(y) = x_{T^*}$ (k-sparse)

Signals with at most k nonzero entries

- Are we done?
- We need to decode y (Solving $\Phi x = y$) efficiently.
- Are we done?
- We usually do not have the priori information that the signal x is k-sparse.

Signals with at most k nonzero entries

- Are we done?
- We need to decode y (Solving $\Phi x = y$) efficiently.
- Are we done?
- We usually do not have the priori information that the signal x is k-sparse.

Signals with at most k nonzero entries

- Are we done?
- We need to decode y (Solving $\Phi x = y$) efficiently.
- Are we done?
- We usually do not have the priori information that the signal x is k-sparse.

Signals with at most k nonzero entries

- Are we done?
- We need to decode y (Solving $\Phi x = y$) efficiently.
- Are we done?
- We usually do not have the priori information that the signal x is k-sparse.

Compressible Signals

- We want to recover signals that are close to be k-sparse (there are around k spikes and other entries are small in magnitude)
- We do not assume that the signal x is k-sparse but somehow can be approximated by a k-sparse vector.
- The signal x has small best k-term approximation error

$$\sigma_k(x)_X = \inf_{z \in \Sigma_k} ||x - z||_X,$$

where $X = \ell_q^N$.

- $\mathcal{A}_X^s$ = the set of signals x for which $\sigma_k(x)_X \leq M \cdot k^{-s}$
- $x \in w\ell_p$ if $|\Lambda_\epsilon(x)| \leq M^p \epsilon^{-p}$
- If $X = \ell_q^N$, then $\mathcal{A}_X^s = w\ell_p^N$ with $p = (s + \frac{1}{q})^{-1}$.

Compressible Signals

- We want to recover signals that are close to be k-sparse (there are around k spikes and other entries are small in magnitude)
- We do not assume that the signal x is k-sparse but somehow can be approximated by a k-sparse vector.
- The signal x has small best k-term approximation error

$$\sigma_k(x)_X = \inf_{z \in \Sigma_k} ||x - z||_X,$$

  where $X = \ell_q^N$.
- $\mathcal{A}_X^s$ = the set of signals x for which $\sigma_k(x)_X \leq M \cdot k^{-s}$
- $x \in w\ell_p$ if $|\Lambda_\epsilon(x)| \leq M^p \epsilon^{-p}$
- If $X = \ell_q^N$, then $\mathcal{A}_X^s = w\ell_p^N$ with $p = (s + \frac{1}{q})^{-1}$.

Compressible Signals

- We want to recover signals that are close to be k-sparse (there are around k spikes and other entries are small in magnitude)
- We do not assume that the signal x is k-sparse but somehow can be approximated by a k-sparse vector.
- The signal x has small best k-term approximation error

$$\sigma_k(x)_X = \inf_{z \in \Sigma_k} ||x - z||_X,$$

where $X = \ell_q^N$.

- $\mathcal{A}_X^s$ = the set of signals x for which $\sigma_k(x)_X \leq M \cdot k^{-s}$
- $x \in w\ell_p$ if $|\Lambda_\epsilon(x)| \leq M^p \epsilon^{-p}$
- If $X = \ell_q^N$, then $\mathcal{A}_X^s = w\ell_p^N$ with $p = (s + \frac{1}{q})^{-1}$.

Compressible Signals

- We want to recover signals that are close to be k-sparse (there are around k spikes and other entries are small in magnitude)
- We do not assume that the signal x is k-sparse but somehow can be approximated by a k-sparse vector.
- The signal x has small best k-term approximation error

$$\sigma_k(x)_X = \inf_{z \in \Sigma_k} ||x - z||_X,$$

  where $X = \ell_q^N$.
- $\mathcal{A}_X^s =$ the set of signals x for which $\sigma_k(x)_X \leq M \cdot k^{-s}$
- $x \in w\ell_p$ if $|\Lambda_\epsilon(x)| \leq M^p \epsilon^{-p}$
- If $X = \ell_q^N$, then $\mathcal{A}_X^s = w\ell_p^N$ with $p = (s + \frac{1}{q})^{-1}$.

Compressible Signals

- We want to recover signals that are close to be k-sparse (there are around k spikes and other entries are small in magnitude)
- We do not assume that the signal x is k-sparse but somehow can be approximated by a k-sparse vector.
- The signal x has small best k-term approximation error

$$\sigma_k(x)_X = \inf_{z \in \Sigma_k} ||x - z||_X,$$

where $X = \ell_q^N$.

- $\mathcal{A}_X^s$ = the set of signals x for which $\sigma_k(x)_X \leq M \cdot k^{-s}$
- $x \in w\ell_p$ if $|\Lambda_\epsilon(x)| \leq M^p \epsilon^{-p}$
- If $X = \ell_q^N$, then $\mathcal{A}_X^s = w\ell_p^N$ with $p = (s + \frac{1}{q})^{-1}$.

Compressible Signals

- We want to recover signals that are close to be k-sparse (there are around k spikes and other entries are small in magnitude)
- We do not assume that the signal x is k-sparse but somehow can be approximated by a k-sparse vector.
- The signal x has small best k-term approximation error

$$\sigma_k(x)_X = \inf_{z \in \Sigma_k} ||x - z||_X,$$

  where $X = \ell_q^N$.
- $\mathcal{A}_X^s =$ the set of signals x for which $\sigma_k(x)_X \leq M \cdot k^{-s}$
- $x \in w\ell_p$ if $|\Lambda_\epsilon(x)| \leq M^p \epsilon^{-p}$
- If $X = \ell_q^N$, then $\mathcal{A}_X^s = w\ell_p^N$ with $p = (s + \frac{1}{q})^{-1}$.

Performance over a class K of compressible signals

- Let $\mathcal{A}_{n,N} = \{(\Phi, \Delta) \mid \Phi \text{ is } n \times N\}$ be the set of pairs of an encoder (Φ, linear) and a decoder (Δ, usually nonlinear).
- The best performance on a compact set K in $X = \ell_q^N$ is

$$E_n(K)_X = \inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \sup_{x\in K} ||x - \Delta(\Phi x)||_X.$$

- $E_n(K)_X$ is closed related to Gelfand width of K with order n

$$d^n(K)_X = \inf_{\text{codim}(Y)\leq n} \sup_{x\in K\cap Y} ||x||_X.$$

Theorem

If $K = -K$ and $K + K \subseteq CK$, then

$$d^n(K)_X \leq E_n(K)_X \leq C \cdot d^n(K)_X.$$

$d^n(K)_X$ is equivalent to $E_n(K)_X$

Note that for an $n \times N$ matrix Φ, the null space $Y = \mathcal{N}(\Phi)$ is of codimension at most n. Conversely, for any subspace Y of $\mathbb{R}^N$ of codimension n, we can associate its orthogonal complement $Y^\perp$ of codimension n to the $n \times N$ matrix Φ whose rows are formed by any basis for $Y^\perp$. Hence, we have

$$d^n(K)_X = \inf_{\Phi_{n\times N}} \sup\{||\eta||_X \mid \eta \in \mathcal{N}(\Phi) \cap K\}.$$

(Goal: $d^n(K)_X \le E_n(K)_X$)
If (Φ, Δ) is any encoder-decoder pair and $z = \Delta(0)$, then for any $\eta \in \mathcal{N}(\Phi)$, we also have $-\eta \in \mathcal{N}(\Phi)$. It follows that

$$(||-\eta||_X =)||\eta||_X \le \max(||\eta - z||_X, ||-\eta - z||_X).$$

Since $K = -K$, we have

$$(\inf) d^n(K)_X \le \sup_{\eta\in\mathcal{N}(\Phi)\cap K} ||\eta||_X \le \sup_{\eta\in\mathcal{N}(\Phi)\cap K} ||\eta - \Delta(\Phi\eta)||_X \le (\inf) \sup_{x\in K} ||x - \Delta(\Phi x)||_X.$$

Proof Continued

(Goal: $E_n(K)_X \leq C \cdot d^n(K)_X$)
Let Y be a subspace of $\mathbb{R}^N$, having codimension n, and Φ be the associated matrix. Choose a decoder Δ so that $\Delta(y)$ is in $K \cap \mathcal{F}(y)$. Now if $x \in K$, then $x - \Delta(\Phi x) \in CK$. It is also in $\mathcal{N}(\Phi) = Y$. Hence, $C^{-1}(x - \Delta(\Phi x))$ is in $K \cap Y$. Thus,

$$C^{-1}E_n(K)_X \leq C^{-1} \sup_{x \in K} ||x - \Delta(\Phi x)||_X = C^{-1} \sup_{\eta \in C(K \cap Y)} ||\eta||_X.$$

This implies that

$$C^{-1}E_n(K)_X \leq C^{-1} \cdot C \inf_{\mathrm{codim}(Y)=n} \sup_{\eta \in (K \cap Y)} ||\eta||_X = d^n(K)_X. \quad \blacksquare$$

Performance over K

- Let $1 \leq n \leq N$ and $1 < p < q \leq 2$. If $K = U(\ell_p^N)$, then
$$C_1 \Psi(n, N, p, q) \leq d^n(K)_{\ell_p} \leq C_2 \Psi(n, N, p, q),$$
where C_1 and C_2 only depend on p and q.
- $\Psi(n, N, 1, 2) = \min \left\{ 1, \sqrt{\frac{\log(N/n)}{n}} \right\}$
- $$C_1 \sqrt{\frac{\log(N/n)}{n}} \leq E_n(U(\ell_1^N))_{\ell_2^N} \leq C_2' \sqrt{\frac{\log(N/n)}{n}}.$$
- For $p = 1$ and $q = 2$, if we take n so that $n = k\log(N/n)$, then we have
$$E_n(U(\ell_1^N))_{\ell_2^N} \leq C' \sqrt{\frac{\log(N/n)}{n}} = C' \frac{1}{\sqrt{k}}.$$

Performance over K

- Let $1 \leq n \leq N$ and $1 < p < q \leq 2$. If $K = U(\ell_p^N)$, then
$$C_1 \Psi(n, N, p, q) \leq d^n(K)_{\ell_p} \leq C_2 \Psi(n, N, p, q),$$
where C_1 and C_2 only depend on p and q.
- $\Psi(n, N, 1, 2) = \min \left\{ 1, \sqrt{\frac{\log(N/n)}{n}} \right\}$
- $$C_1 \sqrt{\frac{\log(N/n)}{n}} \leq E_n(U(\ell_1^N))_{\ell_2^N} \leq C_2' \sqrt{\frac{\log(N/n)}{n}}.$$
- For $p = 1$ and $q = 2$, if we take n so that $n = k \log(N/n)$, then we have
$$E_n(U(\ell_1^N))_{\ell_2^N} \leq C' \sqrt{\frac{\log(N/n)}{n}} = C' \frac{1}{\sqrt{k}}.$$

Performance over K

- Let $1 \leq n \leq N$ and $1 < p < q \leq 2$. If $K = U(\ell_p^N)$, then
$$C_1 \Psi(n, N, p, q) \leq d^n(K)_{\ell_p} \leq C_2 \Psi(n, N, p, q),$$
where C_1 and C_2 only depend on p and q.
- $\Psi(n, N, 1, 2) = \min \left\{ 1, \sqrt{\frac{\log(N/n)}{n}} \right\}$
- $$C_1 \sqrt{\frac{\log(N/n)}{n}} \leq E_n(U(\ell_1^N))_{\ell_2^N} \leq C_2' \sqrt{\frac{\log(N/n)}{n}}.$$
- For $p = 1$ and $q = 2$, if we take n so that $n = k \log(N/n)$, then we have
$$E_n(U(\ell_1^N))_{\ell_2^N} \leq C' \sqrt{\frac{\log(N/n)}{n}} = C' \frac{1}{\sqrt{k}}.$$

Performance over K

- Let $1 \leq n \leq N$ and $1 < p < q \leq 2$. If $K = U(\ell_p^N)$, then

$$C_1 \Psi(n, N, p, q) \leq d^n(K)_{\ell_p} \leq C_2 \Psi(n, N, p, q),$$

  where C_1 and C_2 only depend on p and q.
- $\Psi(n, N, 1, 2) = \min\left\{1, \sqrt{\frac{\log(N/n)}{n}}\right\}$
-
$$C_1 \sqrt{\frac{\log(N/n)}{n}} \leq E_n(U(\ell_1^N))_{\ell_2^N} \leq C_2' \sqrt{\frac{\log(N/n)}{n}}.$$
- For $p = 1$ and $q = 2$, if we take n so that $n = k \log(N/n)$, then we have

$$E_n(U(\ell_1^N))_{\ell_2^N} \leq C' \sqrt{\frac{\log(N/n)}{n}} = C' \frac{1}{\sqrt{k}}.$$

Performance over K

- It can be shown that

$$\frac{1}{2\sqrt{k}} \leq \sigma_k(U(\ell_1^N))_{\ell_2^N} \leq \frac{1}{\sqrt{k}}.$$

- Hence,

$$E_n(U(\ell_1^N))_{\ell_2^N} \leq C_0 \sigma_k(U(\ell_1^N))_{\ell_2^N}.$$

- Using n (with $n = k \log(N/n)$) measurements, we get the same performance as if we took the N measurements (of the coordinates of x and then retained the largest k of them).

Performance over K

- It can be shown that

$$\frac{1}{2\sqrt{k}} \leq \sigma_k(U(\ell_1^N))_{\ell_2^N} \leq \frac{1}{\sqrt{k}}.$$

- Hence,

$$E_n(U(\ell_1^N))_{\ell_2^N} \leq C_0 \sigma_k(U(\ell_1^N))_{\ell_2^N}.$$

- Using n (with $n = k\log(N/n)$) measurements, we get the same performance as if we took the N measurements (of the coordinates of x and then retained the largest k of them).

Performance over K

- It can be shown that
$$\frac{1}{2\sqrt{k}} \leq \sigma_k(U(\ell_1^N))_{\ell_2^N} \leq \frac{1}{\sqrt{k}}.$$
- Hence,
$$E_n(U(\ell_1^N))_{\ell_2^N} \leq C_0 \sigma_k(U(\ell_1^N))_{\ell_2^N}.$$
- Using n (with $n = k\log(N/n)$) measurements, we get the same performance as if we took the N measurements (of the coordinates of x and then retained the largest k of them).

Instance Optimality: Individual vectors

Instance-Optimality

We say (Φ, Δ) is instance-optimal of order k for X if for an absolute constant $C_0 > 0$ (independent of k, n, N)

$$||x - \Delta(\Phi x)||_X \leq C_0 \cdot \sigma_k(x)_X \quad \forall x \in X.$$

Instance-Optimality & NSP

- We say that Φ has the null space property of order k in X if

$$||\eta||_X \leq C_1 \cdot \sigma_k(\eta)_X, \quad \forall \eta \in \mathcal{N}(\Phi).$$

Equivalently, for every set T with $\#T = k$,

$$(||\eta_T||_X \leq)||\eta||_X \leq C_1||\eta_{T^c}||_X, \quad \forall \eta \in \mathcal{N}(\Phi).$$

Instance-Optimality and NSP

Given an $n \times N$ matrix Φ, a norm $||\cdot||_X$ and a value of k, a sufficient condition that there exists a decoder Δ with instance-optimality of order k with constant C_0 is that Φ has the null space property of order $2k$ with constant $C_1 = \frac{C_0}{2}$, and a necessary condition is that Φ has the null space property of order $2k$ with constant $C_1 = C_0$.

Instance-Optimality & NSP

- We say that Φ has the null space property of order k in X if

$$||\eta||_X \leq C_1 \cdot \sigma_k(\eta)_X, \quad \forall \eta \in \mathcal{N}(\Phi).$$

Equivalently, for every set T with $\#T = k$,

$$(||\eta_T||_X \leq)||\eta||_X \leq C_1||\eta_{T^c}||_X, \quad \forall \eta \in \mathcal{N}(\Phi).$$

Instance-Optimality and NSP

Given an $n \times N$ matrix Φ, a norm $||\cdot||_X$ and a value of k, a sufficient condition that there exists a decoder Δ with instance-optimality of order k with constant C_0 is that Φ has the null space property of order $2k$ with constant $C_1 = \frac{C_0}{2}$, and a necessary condition is that Φ has the null space property of order $2k$ with constant $C_1 = C_0$.

RIP & NSP

RIP

An $n \times N$ matrix Φ has the restricted isometry property of order k if there exists a $\delta_k \in (0,1)$ such that

$$(1-\delta_k)||x||_{\ell_2} \leq ||\Phi x||_{\ell_2} \leq (1+\delta_k)||x||_{\ell_2} \quad \forall x \in \Sigma_k.$$

Theorem

If Φ has the RIP of order $3k$ with constant $\delta_{3k} \in (0,1)$, then for each $\eta \in \mathcal{N}(\Phi)$,

$$||\eta||_{\ell_1} \leq \left(1 + \frac{(1+\delta_{3k})\sqrt{2}}{(1-\delta_{3k})}\right) ||\eta_{T^c}||_{\ell_1}$$

for all T with $|T| = 2k$.

RIP=Good

Corollary

If Φ has the RIP for $3k$ with constant $\delta_{3k} \in (0,1)$, then there is an instance-optimal pair (Φ, Δ) of order k such that

$$||x - \Delta(\Phi x)||_{\ell_1} \leq 2 \cdot \left(1 + \frac{(1+\delta_{3k})\sqrt{2}}{(1-\delta_{3k})}\right) \sigma_k(x)_{\ell_1} \quad \forall x \in X.$$

- The good matrices for compressed sensing are the ones with the RIP of order $k \leq c_0 \frac{n}{\log(N/n)}$ (c_0 is a constant which when made small will make δ_k small as well).
- Matrices built using random entries from certain probability distributions will have the RIP with high probability.
- Random Matrices = Good with high probability

RIP=Good

Corollary

If Φ has the RIP for $3k$ with constant $\delta_{3k} \in (0,1)$, then there is an instance-optimal pair (Φ, Δ) of order k such that

$$||x - \Delta(\Phi x)||_{\ell_1} \leq 2 \cdot \left(1 + \frac{(1+\delta_{3k})\sqrt{2}}{(1-\delta_{3k})}\right) \sigma_k(x)_{\ell_1} \quad \forall x \in X.$$

- The good matrices for compressed sensing are the ones with the RIP of order $k \leq c_0 \frac{n}{\log(N/n)}$ (c_0 is a constant which when made small will make δ_k small as well).
- Matrices built using random entries from certain probability distributions will have the RIP with high probability.
- Random Matrices = Good with high probability

RIP=Good

Corollary

If Φ has the RIP for $3k$ with constant $\delta_{3k} \in (0,1)$, then there is an instance-optimal pair (Φ, Δ) of order k such that

$$||x - \Delta(\Phi x)||_{\ell_1} \leq 2 \cdot \left(1 + \frac{(1+\delta_{3k})\sqrt{2}}{(1-\delta_{3k})}\right) \sigma_k(x)_{\ell_1} \quad \forall x \in X.$$

- The good matrices for compressed sensing are the ones with the RIP of order $k \leq c_0 \frac{n}{\log(N/n)}$ (c_0 is a constant which when made small will make δ_k small as well).
- Matrices built using random entries from certain probability distributions will have the RIP with high probability.
- Random Matrices = Good with high probability

$(0,1)$-sensing matrices

- Consider a $(1,-1)$-random matrix $\Phi(\omega)$ from Bernoulli distribution with mean 0. Then $\frac{1}{2}(J+\Phi(\omega))$ is a $(0,1)$-matrix, where each entry of J is 1.
- We have two equivalent sensing matrices, having almost same null space.

$(0,1)$-sensing matrices

- Consider a $(1,-1)$-random matrix $\Phi(\omega)$ from Bernoulli distribution with mean 0. Then $\frac{1}{2}(J+\Phi(\omega))$ is a $(0,1)$-matrix, where each entry of J is 1.
- We have two equivalent sensing matrices, having almost same null space.

Deterministic $(0,1)$-sensing matrices

- Checking the RIP is not so easy, so we want to create a $(0,1)$-matrix Φ_0 for which we can check the RIP (after normalization).
- To create positions of ones, we use $F = \mathbb{Z}_p$ (p prime).
- The rows of Φ_0 are labeled by points $(x,y) \in F \times F$ lexicographically (there are $n = p^2$ rows).
- The columns v_P of Φ_0 are labeled by polynomials $P(x)$ in $F[x]$ of degree at most r (there are $N = p^{r+1}$ columns).
- The column v_P of Φ_0 has ones in the positions $(x, P(x))$ (there are exactly p ones in each column).

Deterministic $(0,1)$-sensing matrices

- Checking the RIP is not so easy, so we want to create a $(0,1)$-matrix Φ_0 for which we can check the RIP (after normalization).
- To create positions of ones, we use $F = \mathbb{Z}_p$ (p prime).
- The rows of Φ_0 are labeled by points $(x, y) \in F \times F$ lexicographically (there are $n = p^2$ rows).
- The columns v_P of Φ_0 are labeled by polynomials $P(x)$ in $F[x]$ of degree at most r (there are $N = p^{r+1}$ columns).
- The column v_P of Φ_0 has ones in the positions $(x, P(x))$ (there are exactly p ones in each column).

Deterministic $(0,1)$-sensing matrices

- Checking the RIP is not so easy, so we want to create a $(0,1)$-matrix Φ_0 for which we can check the RIP (after normalization).
- To create positions of ones, we use $F = \mathbb{Z}_p$ (p prime).
- The rows of Φ_0 are labeled by points $(x, y) \in F \times F$ lexicographically (there are $n = p^2$ rows).
- The columns v_P of Φ_0 are labeled by polynomials $P(x)$ in $F[x]$ of degree at most r (there are $N = p^{r+1}$ columns).
- The column v_P of Φ_0 has ones in the positions $(x, P(x))$ (there are exactly p ones in each column).

Deterministic $(0,1)$-sensing matrices

- Checking the RIP is not so easy, so we want to create a $(0,1)$-matrix Φ_0 for which we can check the RIP (after normalization).
- To create positions of ones, we use $F = \mathbb{Z}_p$ (p prime).
- The rows of Φ_0 are labeled by points $(x, y) \in F \times F$ lexicographically (there are $n = p^2$ rows).
- The columns v_P of Φ_0 are labeled by polynomials $P(x)$ in $F[x]$ of degree at most r (there are $N = p^{r+1}$ columns).
- The column v_P of Φ_0 has ones in the positions $(x, P(x))$ (there are exactly p ones in each column).

Deterministic $(0,1)$-sensing matrices

- Checking the RIP is not so easy, so we want to create a $(0,1)$-matrix Φ_0 for which we can check the RIP (after normalization).
- To create positions of ones, we use $F = \mathbb{Z}_p$ (p prime).
- The rows of Φ_0 are labeled by points $(x, y) \in F \times F$ lexicographically (there are $n = p^2$ rows).
- The columns v_P of Φ_0 are labeled by polynomials $P(x)$ in $F[x]$ of degree at most r (there are $N = p^{r+1}$ columns).
- The column v_P of Φ_0 has ones in the positions $(x, P(x))$ (there are exactly p ones in each column).

$k < \frac{\sqrt{n}\log n}{2\log(N/n)} + 1$

RIP

$$(1-\delta)||x||^2_{\ell_2^N} \le ||\Phi x||^2_{\ell_2^n} \le (1+\delta)||x||^2_{\ell_2^N} \quad \forall x \in \Sigma_k$$

Theorem

The matrix $\Phi = \frac{1}{\sqrt{p}}\Phi_0$ *satisfies the RIP with* $\delta = \frac{(k-1)r}{p}$ *for any* $k < \frac{p}{r} + 1$.

k in n, N

$$k < \frac{\sqrt{n}\log n}{2\log(N/n)} + 1$$

Proof

- $T \subseteq \{1, \ldots, N\}$ with $\# T = k$
- Investigate the range for the eigenvalues of the $k \times k$ Grammian matrix $G_T = \Phi_T^T \Phi_T$
- All the diagonal entries of G_T are 1
- All off-diagonal entries are nonnegative and $\leq \frac{r}{p}$
- Can write G_T as $I + A$ where $A = [a_{ij}]$ with $a_{ii} = 0$ and $|a_{ij}| \leq \frac{r}{p}$ $(i \neq j)$
- Eigenvalues of A are in $[-\frac{(k-1)r}{p}, \frac{(k-1)r}{p}](= [-\delta, \delta])$
- Eigenvalues of G_T are in $[1 - \delta, 1 + \delta]$
- As long as $\frac{(k-1)r}{p} < 1$, i.e., $k < \frac{p}{r} + 1$, the matrix Φ has the RIP of order k.

Proof

- $T \subseteq \{1, \dots, N\}$ with $\#T = k$
- Investigate the range for the eigenvalues of the $k \times k$ Grammian matrix $G_T = \Phi_T^T \Phi_T$
- All the diagonal entries of G_T are 1
- All off-diagonal entries are nonnegative and $\leq \frac{r}{p}$
- Can write G_T as $I + A$ where $A = [a_{ij}]$ with $a_{ii} = 0$ and $|a_{ij}| \leq \frac{r}{p}$ $(i \neq j)$
- Eigenvalues of A are in $[-\frac{(k-1)r}{p}, \frac{(k-1)r}{p}] (= [-\delta, \delta])$
- Eigenvalues of G_T are in $[1-\delta, 1+\delta]$
- As long as $\frac{(k-1)r}{p} < 1$, i.e., $k < \frac{p}{r} + 1$, the matrix Φ has the RIP of order k.

Proof

- $T \subseteq \{1, \ldots, N\}$ with $\#T = k$
- Investigate the range for the eigenvalues of the $k \times k$ Grammian matrix $G_T = \Phi_T^T \Phi_T$
- All the diagonal entries of G_T are 1
- All off-diagonal entries are nonnegative and $\leq \frac{r}{p}$
- Can write G_T as $I + A$ where $A = [a_{ij}]$ with $a_{ii} = 0$ and $|a_{ij}| \leq \frac{r}{p}$ $(i \neq j)$
- Eigenvalues of A are in $[-\frac{(k-1)r}{p}, \frac{(k-1)r}{p}] (= [-\delta, \delta])$
- Eigenvalues of G_T are in $[1 - \delta, 1 + \delta]$
- As long as $\frac{(k-1)r}{p} < 1$, i.e., $k < \frac{p}{r} + 1$, the matrix Φ has the RIP of order k.

Proof

- $T \subseteq \{1, \ldots, N\}$ with $\#T = k$
- Investigate the range for the eigenvalues of the $k \times k$ Grammian matrix $G_T = \Phi_T^T \Phi_T$
- All the diagonal entries of G_T are 1
- All off-diagonal entries are nonnegative and $\leq \frac{r}{p}$
- Can write G_T as $I + A$ where $A = [a_{ij}]$ with $a_{ii} = 0$ and $|a_{ij}| \leq \frac{r}{p}$ $(i \neq j)$
- Eigenvalues of A are in $[-\frac{(k-1)r}{p}, \frac{(k-1)r}{p}](= [-\delta, \delta])$
- Eigenvalues of G_T are in $[1 - \delta, 1 + \delta]$
- As long as $\frac{(k-1)r}{p} < 1$, i.e., $k < \frac{p}{r} + 1$, the matrix Φ has the RIP of order k.

Proof

- $T \subseteq \{1, \ldots, N\}$ with $\# T = k$
- Investigate the range for the eigenvalues of the $k \times k$ Grammian matrix $G_T = \Phi_T^T \Phi_T$
- All the diagonal entries of G_T are 1
- All off-diagonal entries are nonnegative and $\leq \frac{r}{p}$
- Can write G_T as $I + A$ where $A = [a_{ij}]$ with $a_{ii} = 0$ and $|a_{ij}| \leq \frac{r}{p}$ $(i \neq j)$
- Eigenvalues of A are in $[-\frac{(k-1)r}{p}, \frac{(k-1)r}{p}](= [-\delta, \delta])$
- Eigenvalues of G_T are in $[1 - \delta, 1 + \delta]$
- As long as $\frac{(k-1)r}{p} < 1$, i.e., $k < \frac{p}{r} + 1$, the matrix Φ has the RIP of order k.

Proof

- $T \subseteq \{1, \ldots, N\}$ with $\# T = k$
- Investigate the range for the eigenvalues of the $k \times k$ Grammian matrix $G_T = \Phi_T^T \Phi_T$
- All the diagonal entries of G_T are 1
- All off-diagonal entries are nonnegative and $\leq \frac{r}{p}$
- Can write G_T as $I + A$ where $A = [a_{ij}]$ with $a_{ii} = 0$ and $|a_{ij}| \leq \frac{r}{p}$ $(i \neq j)$
- Eigenvalues of A are in $[-\frac{(k-1)r}{p}, \frac{(k-1)r}{p}](= [-\delta, \delta])$
- Eigenvalues of G_T are in $[1 - \delta, 1 + \delta]$
- As long as $\frac{(k-1)r}{p} < 1$, i.e., $k < \frac{p}{r} + 1$, the matrix Φ has the RIP of order k.

Proof

- $T \subseteq \{1, \ldots, N\}$ with $\# T = k$
- Investigate the range for the eigenvalues of the $k \times k$ Grammian matrix $G_T = \Phi_T^T \Phi_T$
- All the diagonal entries of G_T are 1
- All off-diagonal entries are nonnegative and $\leq \frac{r}{p}$
- Can write G_T as $I + A$ where $A = [a_{ij}]$ with $a_{ii} = 0$ and $|a_{ij}| \leq \frac{r}{p}$ $(i \neq j)$
- Eigenvalues of A are in $[-\frac{(k-1)r}{p}, \frac{(k-1)r}{p}](= [-\delta, \delta])$
- Eigenvalues of G_T are in $[1 - \delta, 1 + \delta]$
- As long as $\frac{(k-1)r}{p} < 1$, i.e., $k < \frac{p}{r} + 1$, the matrix Φ has the RIP of order k.

Remarks

- The order of the RIP of the deterministic matrix falls far short of the range $\frac{n}{\log(N/n)}$ known for probabilistic constructions.

Question

Does there exist a deterministic construction for $n \times N$ matrices Φ that have the RIP of order $k \approx \frac{n^a}{\log(N/n)}$ for $\frac{1}{2} < a \leq 1$?

Remarks

- The order of the RIP of the deterministic matrix falls far short of the range $\frac{n}{\log(N/n)}$ known for probabilistic constructions.

Question

Does there exist a deterministic construction for $n \times N$ matrices Φ that have the RIP of order $k \approx \frac{n^a}{\log(N/n)}$ for $\frac{1}{2} < a \leq 1$?

References

Richard Baraniuk, Mark Davenport, Ronald DeVore, Michael Wakin.
A simple proof of the restricted isometry property for random matrices,
Constructive Approximation 28 (2008), 253–263.

Albert Cohen, Wolfgan Dahmen, Ronald DeVore.
Compressed sensing and best k-term approximation,
Journal of the American Mathematical Society 22 (2009), 211–231.

Ronald DeVore.
Deterministic construction of compressed sensing matrices,
Journal of Complexity 23 (2007), 918–925.