Compressed Sensing - beyond the Shannon Paradigm

Wolfgang Dahmen

Institut für Geometrie und Praktische Mathematik
RWTH Aachen

joint work with Albert Cohen and Ron DeVore

ISWCS-11, Aachen, Nov.8, 2011

igpm

Contents

Contents

Contents

Contents

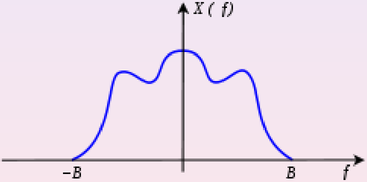
Contents

Contents

The Nyquist-Shannon Theorem

A bandlimited "signal" $f(t)$, $t \in \mathbb{R}$, with bandwidth $2B$ can be reconstructed exactly from the samples $f(k\tau)$, $\tau < 1/2B$, ...

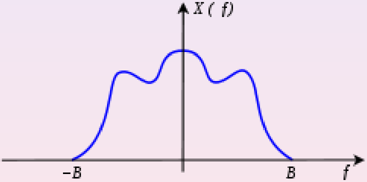
$$X(f)(\omega) := \int_{\mathbb{R}} f(t)e^{-i2\pi t\omega}dt = 0, \quad |\omega| > B, \quad \rightsquigarrow$$



The Nyquist-Shannon Theorem

A bandlimited "signal" $f(t)$, $t \in \mathbb{R}$, with bandwidth $2B$ can be reconstructed exactly from the samples $f(k\tau)$, $\tau < 1/2B$, ...

$$X(f)(\omega) := \int_{\mathbb{R}} f(t) e^{-i2\pi t\omega} dt = 0, \quad |\omega| > B, \quad \rightsquigarrow$$

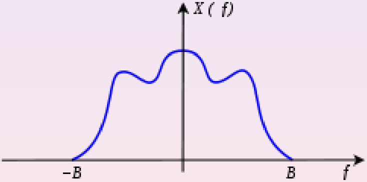


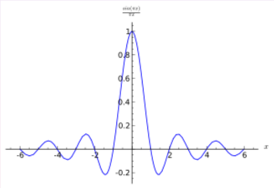
$$f(t) = \sum_{k=-\infty}^{\infty} f(k\tau) \frac{\sin \pi(\tau^{-1} t - k)}{\pi(\tau^{-1} t - k)}$$

The Nyquist-Shannon Theorem

A bandlimited “signal” $f(t)$, $t \in \mathbb{R}$, with bandwidth $2B$ can be reconstructed exactly from the samples $f(k\tau)$, $\tau < 1/2B$, ...

$$X(f)(\omega) := \int_{\mathbb{R}} f(t) e^{-i2\pi t\omega} dt = 0, \quad |\omega| > B, \quad \rightsquigarrow$$





$$f(t) = \sum_{k=-\infty}^{\infty} f(k\tau) \frac{\sin \pi(\tau^{-1} t - k)}{\pi(\tau^{-1} t - k)}$$

A Dilemma...

Common features of many signals:

A Dilemma...

Common features of many signals:

- ...they are far from bandlimited

A Dilemma...

Common features of many signals:

- ...they are far from bandlimited
- ...however, they are sparse, i.e. their information content is small compared with their size

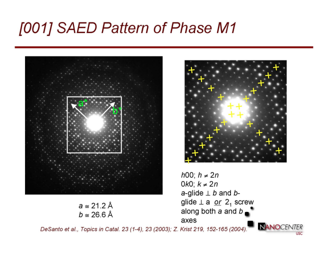
A Dilemma...

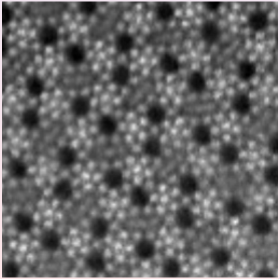
Common features of many signals:

- ...they are far from bandlimited
- ...however, they are sparse, i.e. their information content is small compared with their size
- ...measurements are difficult/expensive/harmful

(Information-)Sparse Signals

...are those sparse?





Attempt of a “Shortcut”

- employ more general measurements than classical sampling

Attempt of a “Shortcut”

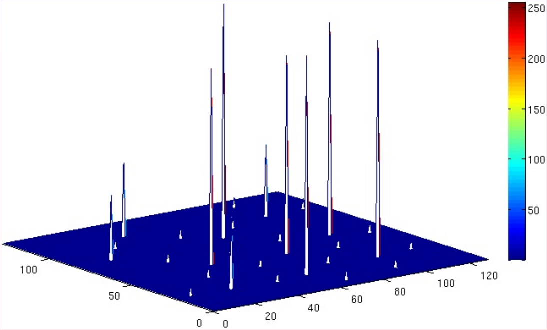
- employ more general measurements than classical sampling
- Question: can one sample at a rate comparable to information content rather than signal size?

Attempt of a “Shortcut”

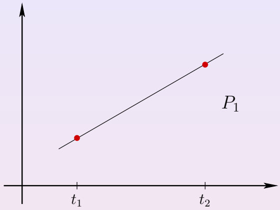
- employ more general measurements than classical sampling
- Question: can one sample at a rate comparable to information content rather than signal size?
- ...what exactly would one like to know?

Attempt of a "Shortcut"

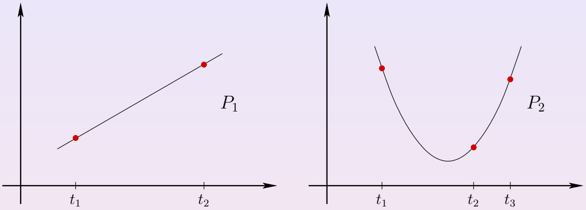
- employ more general measurements than classical sampling
- Question: can one sample at a rate comparable to information content rather than signal size?
- ...what exactly would one like to know?



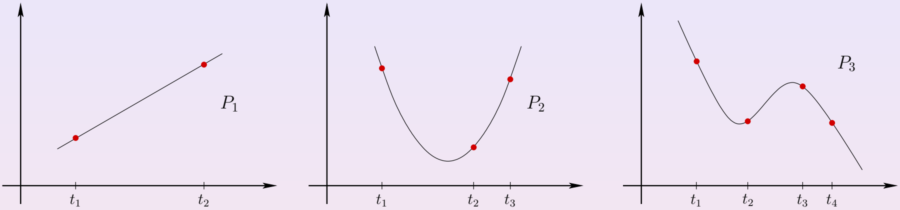
An Instructive Example: Polynomials



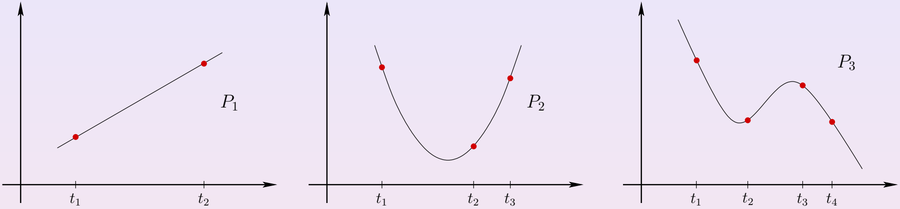
An Instructive Example: Polynomials



An Instructive Example: Polynomials

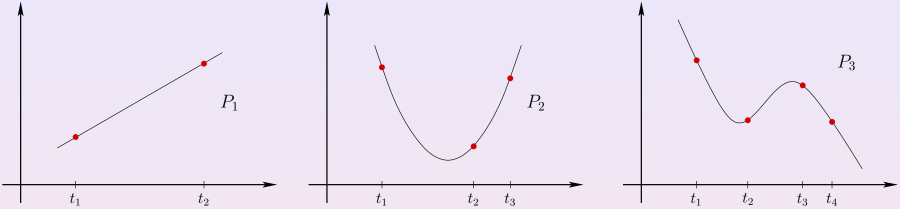


An Instructive Example: Polynomials



$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

An Instructive Example: Polynomials



$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

$$y_i = P_{N-1}(t_i),\ i = 1, \ldots, N, \quad \Rightarrow \quad x_0 \ldots, x_N$$

“Sparse” Polynomials

$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

“Sparse” Polynomials

$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

$$y_i = P_N(t_i) = \sum_{j=1}^{N} \underbrace{(t_i^{j-1})}_{=:\phi_{i,j}} x_j \qquad , \ i = 1, \ldots, N,$$

"Sparse" Polynomials

$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

$$y_i = P_N(t_i) = \sum_{j=1}^{N} \underbrace{(t_i^{j-1})}_{=:\phi_{i,j}} x_j =: \phi_i \cdot x, \ \ i = 1, \ldots, N,$$

“Sparse” Polynomials

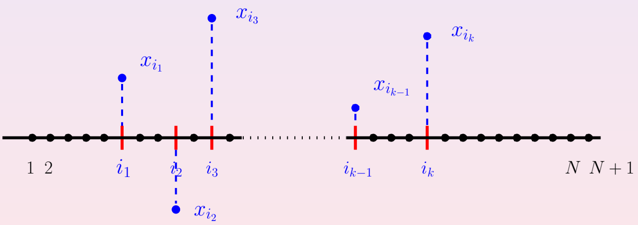
$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

$$y_i = P_N(t_i) = \sum_{j=1}^{N} \underbrace{(t_i^{j-1})}_{=:\phi_{i,j}} x_j =: \phi_i \cdot x, \; i = 1, \ldots, N, \quad \Rightarrow \quad x_0 \ldots, x_N$$

"Sparse" Polynomials

$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

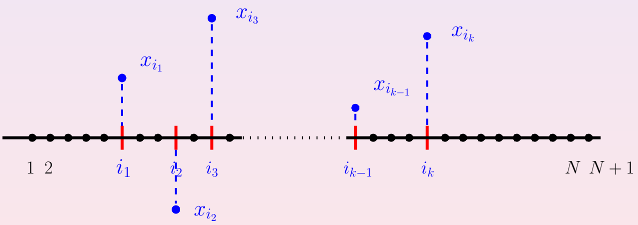
$$y_i = P_N(t_i) = \sum_{j=1}^{N} \underbrace{(t_i^{j-1})}_{=:\phi_{i,j}} x_j =: \phi_i \cdot x, \; i = 1, \ldots, N, \quad \Rightarrow \quad x_0 \ldots, x_N$$



"Sparse" Polynomials

$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

$$y_i = P_N(t_i) = \sum_{j=1}^{N} \underbrace{(t_i^{j-1})}_{=:\phi_{i,j}} x_j =: \phi_i \cdot x, \; i = 1, \ldots, N, \quad \Rightarrow \quad x_0 \ldots, x_N$$



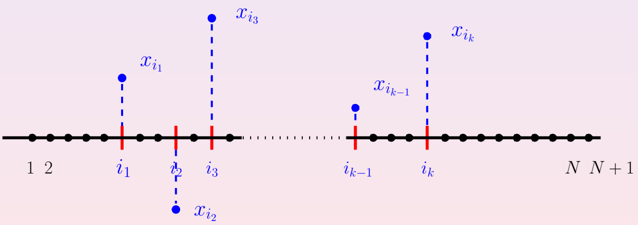
$$y_i = P_N(t_i), \; i = 0, \ldots, \qquad n \qquad \overset{?}{\Longrightarrow} \quad x_{i_1} \ldots, x_{i_k}$$

igpm

“Sparse” Polynomials

$$P_{N-1}(t) = x_1 t^0 + x_2 t^1 + x_3 t^2 + \cdots + x_{N-1} t^{N-2} + x_N t^{N-1}$$

$$y_i = P_N(t_i) = \sum_{j=1}^{N} \underbrace{(t_i^{j-1})}_{=:\phi_{i,j}} x_j =: \phi_i \cdot x, \; i = 1, \ldots, N, \quad \Rightarrow \quad x_0 \ldots, x_N$$



$$y_i = P_N(t_i), \; i = 0, \ldots, ? = n < N \quad \overset{?}{\Longrightarrow} \quad x_{i_1} \ldots, x_{i_k}$$

igpm

The Mathematical Model

- $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ "large" signal

Candés/Romberg/Tao, Donoho, Gilbert/Strauss/Tropp, Tanner, Rice group, Needell/Vershynin, CDD...

The Mathematical Model

- $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ "large" signal
- "Measurements" $x \to \Phi_i \cdot x$

$$\phi_i \cdot x = \sum_{j=1}^{N} \phi_{i,j} x_j, \quad i = 1, \dots, n \ll N$$

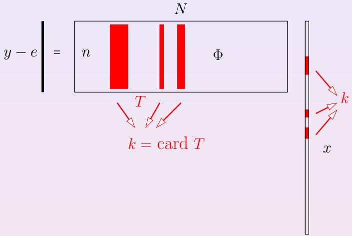
$\rightsquigarrow$ sensing matrix: $\Phi = (\phi_{i,j})_{i,j=1}^{n,N}$

Candés/Romberg/Tao, Donoho, Gilbert/Strauss/Tropp, Tanner, Rice group, Needell/Vershynin, CDD...

The Mathematical Model

- $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ "large" signal
- "Measurements" $x \to \Phi_i \cdot x$

$$\phi_i \cdot x = \sum_{j=1}^{N} \phi_{i,j} x_j, \quad i = 1, \ldots, n \ll N$$

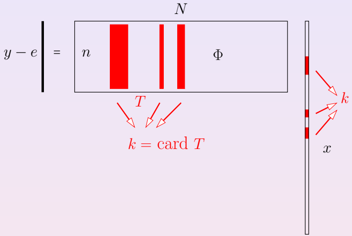


$\rightsquigarrow$ sensing matrix: $\Phi = (\phi_{i,j})_{i,j=1}^{n,N}$

Candés/Romberg/Tao, Donoho, Gilbert/Strauss/Tropp, Tanner, Rice group, Needell/Vershynin, CDD...

The Mathematical Model

- $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ "large" signal
- "Measurements" $x \to \Phi_i \cdot x$



$$\phi_i \cdot x = \sum_{j=1}^{N} \phi_{i,j} x_j, \quad i = 1, \ldots, n \ll N$$

$\rightsquigarrow$ sensing matrix: $\Phi = (\phi_{i,j})_{i,j=1}^{n,N}$

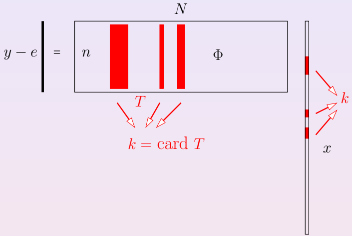
- k-sparse vektors: $\Sigma_k = \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \le k\}$

Candés/Romberg/Tao, Donoho, Gilbert/Strauss/Tropp, Tanner, Rice group, Needell/Vershynin, CDD...

The Mathematical Model

- $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ "large" signal
- "Measurements" $x \to \Phi_i \cdot x$

$$\phi_i \cdot x = \sum_{j=1}^{N} \phi_{i,j} x_j, \quad i = 1, \ldots, n \ll N$$



$\rightsquigarrow$ sensing matrix: $\Phi = (\phi_{i,j})_{i,j=1}^{n,N}$

- k-sparse vektors: $\Sigma_k = \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \le k\}$

  Question: given $y = \Phi x$ for an $x \in \Sigma_k$,

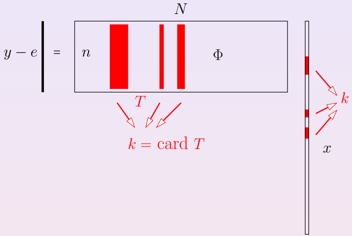
Candés/Romberg/Tao, Donoho, Gilbert/Strauss/Tropp, Tanner, Rice group, Needell/Vershynin, CDD...

igpm

The Mathematical Model

- $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ "large" signal
- "Measurements" $x \to \Phi_i \cdot x$

$$\phi_i \cdot x = \sum_{j=1}^{N} \phi_{i,j} x_j, \quad i = 1, \dots, n \ll N$$



$\rightsquigarrow$ sensing matrix: $\Phi = (\phi_{i,j})_{i,j=1}^{n,N}$

- k-sparse vektors: $\Sigma_k = \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \le k\}$

Question: given $y = \Phi x$ for an $x \in \Sigma_k$, does there exist a decoder

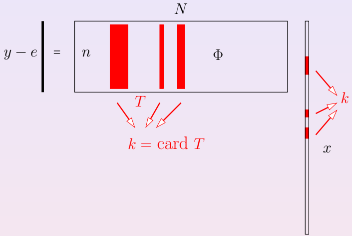
$$\Delta : \mathbb{R}^n \to \mathbb{R}^N, \quad \text{such that } x = \Delta(y), \qquad ?$$

Candés/Romberg/Tao, Donoho, Gilbert/Strauss/Tropp, Tanner, Rice group, Needell/Vershynin, CDD...

The Mathematical Model

- $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ "large" signal
- "Measurements" $x \to \Phi_i \cdot x$

$$\phi_i \cdot x = \sum_{j=1}^{N} \phi_{i,j} x_j, \quad i = 1, \ldots, n \ll N$$



$\rightsquigarrow$ sensing matrix: $\Phi = (\phi_{i,j})_{i,j=1}^{n,N}$

- k-sparse vektors: $\Sigma_k = \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \le k\}$

Question: given $y = \Phi x$ for an $x \in \Sigma_k$, does there exist a decoder

$$\Delta : \mathbb{R}^n \to \mathbb{R}^N, \quad \text{such that } x = \Delta(y), \text{ for } n \sim k \ll N?$$

Candés/Romberg/Tao, Donoho, Gilbert/Strauss/Tropp, Tanner, Rice group, Needell/Vershynin, CDD...

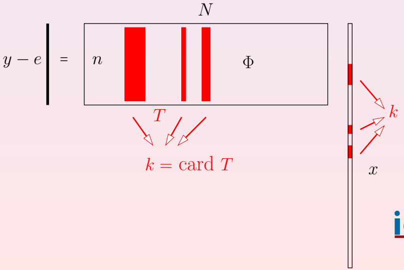
igpm

Back to Polynomials: $P_{N-1}(t_i) = \sum_{j=1}^{N} \phi_{i,j} x_j,\ i \le n \ll N$

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}
=
\begin{pmatrix}
t_1^0 & \cdots & t_1^{i_1} & \cdots & t_1^{i_k} & \cdots & t_1^{N-1} \\
t_2^0 & \cdots & t_2^{i_1} & \cdots & t_2^{i_k} & \cdots & t_2^{N-1} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
t_{n-1}^0 & \cdots & t_{n-1}^{i_1} & \cdots & t_{n-1}^{i_k} & \cdots & t_{n-1}^{N-1} \\
t_n^0 & \cdots & t_n^{i_1} & \cdots & t_n^{i_k} & \cdots & t_n^{N-1}
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{pmatrix}
$$

Back to Polynomials: $P_{N-1}(t_i) = \sum_{j=1}^{N} \phi_{i,j} x_j,\ i \le n \ll N$

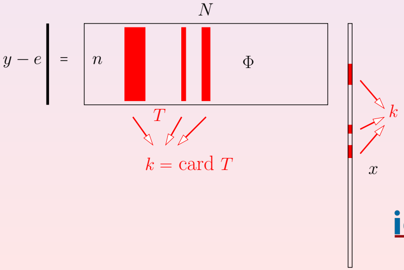
$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} t_1^0 & \cdots & t_1^{i_1} & \cdots & t_1^{i_k} & \cdots & t_1^{N-1} \\ t_2^0 & \cdots & t_2^{i_1} & \cdots & t_2^{i_k} & \cdots & t_2^{N-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ t_{n-1}^0 & \cdots & t_{n-1}^{i_1} & \cdots & t_{n-1}^{i_k} & \cdots & t_{n-1}^{N-1} \\ t_n^0 & \cdots & t_n^{i_1} & \cdots & t_n^{i_k} & \cdots & t_n^{N-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{pmatrix}$$



Back to Polynomials: $P_{N-1}(t_i) = \sum_{j=1}^{N} \phi_{i,j} x_j,\ i \le n \ll N$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} t_1^0 & \cdots & t_1^{i_1} & \cdots & t_1^{i_k} & \cdots & t_1^{N-1} \\ t_2^0 & \cdots & t_2^{i_1} & \cdots & t_2^{i_k} & \cdots & t_2^{N-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ t_{n-1}^0 & \cdots & t_{n-1}^{i_1} & \cdots & t_{n-1}^{i_k} & \cdots & t_{n-1}^{N-1} \\ t_n^0 & \cdots & t_n^{i_1} & \cdots & t_n^{i_k} & \cdots & t_n^{N-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{pmatrix}$$

$$y = \Phi x$$

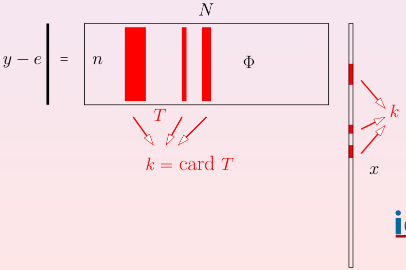


Back to Polynomials: $P_{N-1}(t_i) = \sum_{j=1}^{N} \phi_{i,j} x_j,\ i \le n \ll N$

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}
=
\begin{pmatrix}
t_1^0 & \cdots & t_1^{i_1} & \cdots & t_1^{i_k} & \cdots & t_1^{N-1} \\
t_2^0 & \cdots & t_2^{i_1} & \cdots & t_2^{i_k} & \cdots & t_2^{N-1} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
t_{n-1}^0 & \cdots & t_{n-1}^{i_1} & \cdots & t_{n-1}^{i_k} & \cdots & t_{n-1}^{N-1} \\
t_n^0 & \cdots & t_n^{i_1} & \cdots & t_n^{i_k} & \cdots & t_n^{N-1}
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{pmatrix}
$$

$$y = \Phi x$$

$$\Delta : y \xrightarrow{?} x$$



Are $n \geq 2k$ Measurements Sufficient?...

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

Are $n \geq 2k$ Measurements Sufficient?...

$$\Sigma_k := \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \leq k\},$$

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

(i) $\exists\ \Delta$ so that $\Delta(\Phi x) = x$, for all $x \in \Sigma_k$,

Are $n \geq 2k$ Measurements Sufficient?...

$$\Sigma_k := \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \leq k\},$$

$$\ker \Phi := \{z \in \mathbb{R}^N : \Phi z = 0\}$$

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

(i) $\exists\ \Delta$ so that $\Delta(\Phi x) = x$, for all $x \in \Sigma_k$,

(ii) $\Sigma_{2k} \cap \ker \Phi = \{0\}$,

Are $n \geq 2k$ Measurements Sufficient?...

$$\Sigma_k := \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \leq k\},$$

$$\ker \Phi := \{z \in \mathbb{R}^N : \Phi z = 0\}$$

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

(i) $\exists \;\; \Delta$ so that $\Delta(\Phi x) = x$, for all $x \in \Sigma_k$,

(ii) $\Sigma_{2k} \cap \ker \Phi = \{0\}$,

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

Are $n \geq 2k$ Measurements Sufficient?...

$$\Sigma_k := \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \leq k\},$$

$$\ker \Phi := \{z \in \mathbb{R}^N : \Phi z = 0\}$$

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

(i) $\exists\ \Delta$ so that $\Delta(\Phi x) = x$, for all $x \in \Sigma_k$,

(ii) $\Sigma_{2k} \cap \ker \Phi = \{0\}$,

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

Proof of: (i)⇒(ii): Suppose $x \in \Sigma_{2k} \cap \ker \Phi$

Are $n \geq 2k$ Measurements Sufficient?...

$$\Sigma_k := \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \leq k\},$$

$$\ker \Phi := \{z \in \mathbb{R}^N : \Phi z = 0\}$$

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

(i) $\exists\ \Delta$ so that $\Delta(\Phi x) = x$, for all $x \in \Sigma_k$,

(ii) $\Sigma_{2k} \cap \ker \Phi = \{0\}$,

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

Proof of: (i)$\Rightarrow$(ii): Suppose $x \in \Sigma_{2k} \cap \ker \Phi \ \rightsquigarrow x = x_0 - x_1$ where $x_0, x_1 \in \Sigma_k$.

Are $n \geq 2k$ Measurements Sufficient?...

$\Sigma_k := \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \leq k\}$,

$\ker \Phi := \{z \in \mathbb{R}^N : \Phi z = 0\}$

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

(i) $\exists\ \Delta$ so that $\Delta(\Phi x) = x$, for all $x \in \Sigma_k$,

(ii) $\Sigma_{2k} \cap \ker \Phi = \{0\}$,

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

Proof of: (i)⇒(ii): Suppose $x \in \Sigma_{2k} \cap \ker \Phi \rightsquigarrow x = x_0 - x_1$ where $x_0, x_1 \in \Sigma_k$. Since $\Phi x_0 = \Phi x_1$, (i) $\Longrightarrow\ x_0 = x_1$

Are $n \geq 2k$ Measurements Sufficient?...

$$\Sigma_k := \{z \in \mathbb{R}^N : \#\mathrm{supp}\,(z) \leq k\},$$

$$\ker \Phi := \{z \in \mathbb{R}^N : \Phi z = 0\}$$

LEMMA:

Let $\Phi \in \mathbb{R}^{n \times N}$ and $2k \leq n$. Then the following are equivalent:

(i) $\exists\ \Delta$ so that $\Delta(\Phi x) = x$, for all $x \in \Sigma_k$,

(ii) $\Sigma_{2k} \cap \ker \Phi = \{0\}$,

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

Proof of: (i)⇒(ii): Suppose $x \in \Sigma_{2k} \cap \ker \Phi \ \rightsquigarrow x = x_0 - x_1$ where $x_0, x_1 \in \Sigma_k$. Since $\Phi x_0 = \Phi x_1$, (i) $\Longrightarrow\ x_0 = x_1 \ \Longrightarrow\ x = x_0 - x_1 = 0$.

igpm

Polynomials Form a Descarte System

Are $n = 2k$ measurements sufficient for k-sparse polynomials?

Polynomials Form a Descarte System

Are $n = 2k$ measurements sufficient for k-sparse polynomials?

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_{2k} \end{pmatrix} = \begin{pmatrix} t_1^0 & \cdots & t_1^{i_1} & \cdots & t_1^{i_k} & \cdots & t_1^{N-1} \\ t_2^0 & \cdots & t_2^{i_1} & \cdots & t_2^{i_k} & \cdots & t_2^{N-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ t_{2k-1}^0 & \cdots & t_{2k-1}^{i_1} & \cdots & t_{2k-1}^{i_k} & \cdots & t_{2k-1}^{N-1} \\ t_{2k}^0 & \cdots & t_{2k}^{i_1} & \cdots & t_{2k}^{i_k} & \cdots & t_{2k}^{N-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{pmatrix}$$

Polynomials Form a Descarte System

Are $n = 2k$ measurements sufficient for k-sparse polynomials?

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_{2k} \end{pmatrix} = \begin{pmatrix} t_1^0 & \cdots & t_1^{i_1} & \cdots & t_1^{i_k} & \cdots & t_1^{N-1} \\ t_2^0 & \cdots & t_2^{i_1} & \cdots & t_2^{i_k} & \cdots & t_2^{N-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ t_{2k-1}^0 & \cdots & t_{2k-1}^{i_1} & \cdots & t_{2k-1}^{i_k} & \cdots & t_{2k-1}^{N-1} \\ t_{2k}^0 & \cdots & t_{2k}^{i_1} & \cdots & t_{2k}^{i_k} & \cdots & t_{2k}^{N-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{pmatrix}$$

...yes, all submatrices with $2k$ columns have full rank

Polynomials Form a Descarte System

Are $n = 2k$ measurements sufficient for k-sparse polynomials?

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_{2k} \end{pmatrix} = \begin{pmatrix} t_1^0 & \cdots & t_1^{i_1} & \cdots & t_1^{i_k} & \cdots & t_1^{N-1} \\ t_2^0 & \cdots & t_2^{i_1} & \cdots & t_2^{i_k} & \cdots & t_2^{N-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ t_{2k-1}^0 & \cdots & t_{2k-1}^{i_1} & \cdots & t_{2k-1}^{i_k} & \cdots & t_{2k-1}^{N-1} \\ t_{2k}^0 & \cdots & t_{2k}^{i_1} & \cdots & t_{2k}^{i_k} & \cdots & t_{2k}^{N-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{pmatrix}$$

...yes, all submatrices with $2k$ columns have full rank

BUT...

A Few Wrinkles...

- how to compute the sparse solution? (decoder)

A Few Wrinkles...

- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)

A Few Wrinkles...

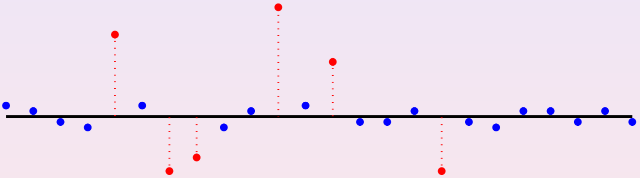
- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)
- Is the reconstruction stable? (quality)

A Few Wrinkles...

- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)
- Is the reconstruction stable? (quality)
- BUT... reale signals are not exactly sparse!

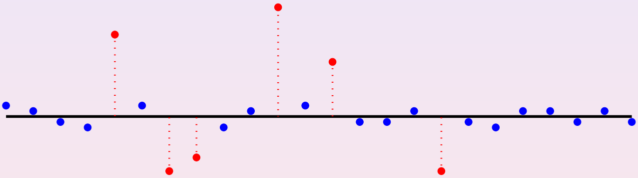
A Few Wrinkles...

- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)
- Is the reconstruction stable? (quality)
- BUT... reale signals are not exactly sparse!



A Few Wrinkles...

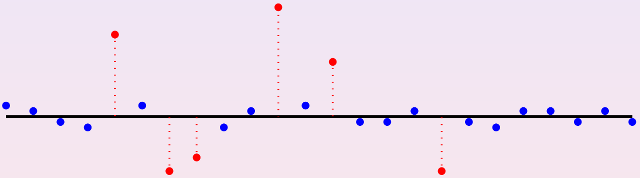
- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)
- Is the reconstruction stable? (quality)
- BUT... reale signals are not exactly sparse!



Best k-term approximation accuracy would be great:

A Few Wrinkles...

- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)
- Is the reconstruction stable? (quality)
- BUT... reale signals are not exactly sparse!

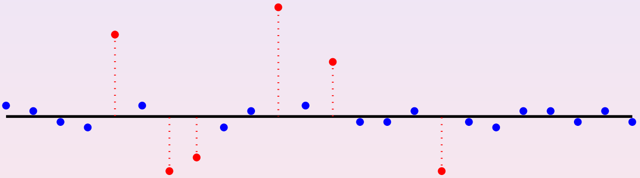


Best k-term approximation accuracy would be great:

$$\|x - z\|_{\ell_p}, \qquad \|x\|_{\ell_p} := \left(\sum_{j=1}^{N} |x_j|^p \right)^{1/p}$$

A Few Wrinkles...

- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)
- Is the reconstruction stable? (quality)
- BUT... reale signals are not exactly sparse!

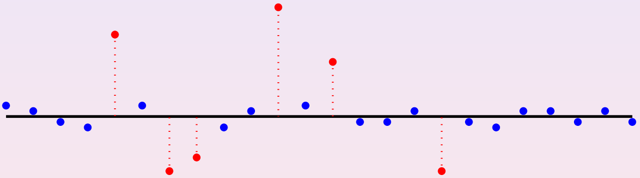


Best k-term approximation accuracy would be great:

$$\inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}, \qquad \|x\|_{\ell_p} := \left(\sum_{j=1}^{N} |x_j|^p \right)^{1/p}$$

A Few Wrinkles...

- how to compute the sparse solution? (decoder)
- that seems to be NP-hard (practical feasibility)
- Is the reconstruction stable? (quality)
- BUT... reale signals are not exactly sparse!

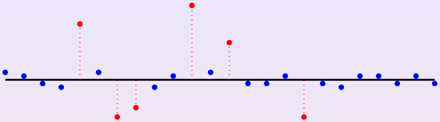


Best k-term approximation accuracy would be great:

$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}, \qquad \|x\|_{\ell_p} := \left(\sum_{j=1}^{N} |x_j|^p \right)^{1/p}$$

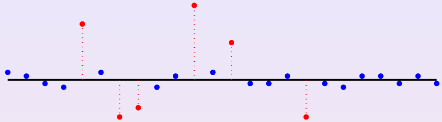
igpm

A Better Benchmark than Exact Sparse Recovery...



$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}$$

A Better Benchmark than Exact Sparse Recovery...

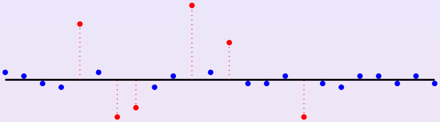


$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}$$

k-Term optimality:

Given N, n, how large can k become, so that there exists a encoder/decoder - pair (Φ, Δ), so that

A Better Benchmark than Exact Sparse Recovery...



$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}$$

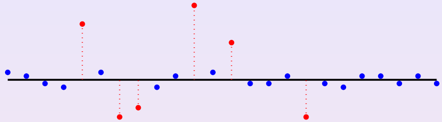
k-Term optimality:

Given N, n, how large can k become, so that there exists a encoder/decoder - pair (Φ, Δ), so that

$$\|x - \Delta(\Phi x)\|_{\ell_2} \leq C_0 \sigma_k(x)_{\ell_2}, \quad \forall\, x \in \mathbb{R}^N \quad (\mathrm{IO}(k)) \tag{1}$$

holds

A Better Benchmark than Exact Sparse Recovery...



$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}$$

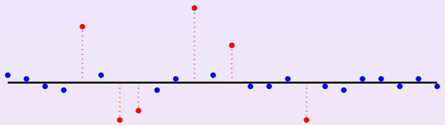
k-Term optimality:

Given N, n, how large can k become, so that there exists a encoder/decoder - pair (Φ, Δ), so that

$$\|x - \Delta(\Phi x)\|_{\ell_2} \leq C_0 \sigma_k(x)_{\ell_2}, \quad \forall\, x \in \mathbb{R}^N \quad (\mathrm{IO}(k)) \tag{1}$$

holds – ? k, Φ, Δ ?

A Better Benchmark than Exact Sparse Recovery...



$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}$$

k-Term optimality:

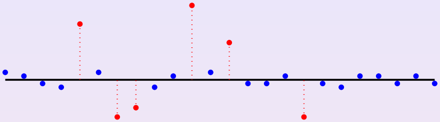
Given N, n, how large can k become, so that there exists a encoder/decoder - pair (Φ, Δ), so that

$$\|x - \Delta(\Phi x)\|_{\ell_2} \le C_0 \sigma_k(x)_{\ell_2}, \quad \forall\, x \in \mathbb{R}^N \quad (\mathrm{IO}(k)) \tag{1}$$

holds – ? k, Φ, Δ ?

Note: $x \in \Sigma_k$,

A Better Benchmark than Exact Sparse Recovery...



$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}$$

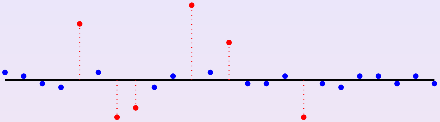
k-Term optimality:

Given N, n, how large can k become, so that there exists a encoder/decoder - pair (Φ, Δ), so that

$$\|x - \Delta(\Phi x)\|_{\ell_2} \le C_0 \sigma_k(x)_{\ell_2}, \quad \forall\, x \in \mathbb{R}^N \quad (\mathrm{IO}(k)) \tag{1}$$

holds – ? k, Φ, Δ ?

Note: $x \in \Sigma_k$, $\quad \mathrm{IO}(k) \Longrightarrow \Delta(\Phi x) = x$



$$\sigma_k(x)_{\ell_p} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p}$$

k-Term optimality:

Given N, n, how large can k become, so that there exists a encoder/decoder - pair (Φ, Δ), so that

$$\|x - \Delta(\Phi x)\|_{\ell_2} \leq C_0 \sigma_k(x)_{\ell_2}, \quad \forall\, x \in \mathbb{R}^N \quad (\mathrm{IO}(k)) \tag{1}$$

holds – ? k, Φ, Δ ?

Note: $x \in \Sigma_k$, $\mathrm{IO}(k) \Longrightarrow \Delta(\Phi x) = x$ exact reconstruction!

igpm

Guiding Questions

- What is the maximal optimality range k?

Guiding Questions

- What is the maximal optimality range k?
- Which sensing matrices Φ achieve the maximal range?

Guiding Questions

- What is the maximal optimality range k?
- Which sensing matrices Φ achieve the maximal range?
- How to develop k-term optimal decoders?

Maximal Optimality Range k? - back to the 70's

THEOREM:

$$\|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \qquad \sigma_k(x)_{\ell_2}$$

Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

$$\max_{x \in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x \in K} \sigma_k(x)_{\ell_2}$$

Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

$$\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

$$\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

$$\Rightarrow \quad k \leq c_0 n / \log(N/n)$$

Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

$$\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

$$\Rightarrow \quad k \leq c_0 n / \log(N/n)$$

$\rightsquigarrow$ Maximal optimality range $k \leq c_0 n / \log(N/n)$ (instead of $k \leq n/2$)

Maximal Optimality Range k? - back to the 70's

THEOREM:

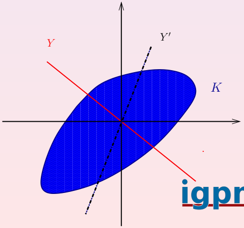
Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

$$\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

$$\Rightarrow \quad k \leq c_0 n/\log(N/n)$$

$\rightsquigarrow$ Maximal optimality range $k \leq c_0 n/\log(N/n)$ (instead of $k \leq n/2$)

Gelfand widths: (Kashin, Gluskin/Garnaev)



Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

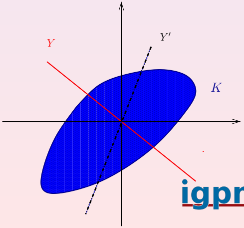
$$\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

$$\Rightarrow \quad k \leq c_0 n / \log(N/n)$$

$\rightsquigarrow$ Maximal optimality range $k \leq c_0 n/\log(N/n)$ (instead of $k \leq n/2$)

Gelfand widths: (Kashin, Gluskin/Garnaev)

$$\sup\{\|x\|_X \, ; \, x \in K \cap Y\}$$



THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

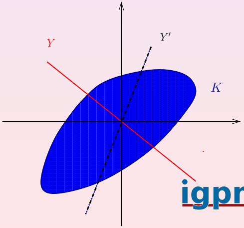
$$\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \stackrel{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

$$\Rightarrow \quad k \leq c_0 n / \log(N/n)$$

$\rightsquigarrow$ Maximal optimality range $k \leq c_0 n / \log(N/n)$ (instead of $k \leq n/2$)

Gelfand widths: (Kashin, Gluskin/Garnaev)

$$\sup\{\|x\|_X \,;\, x \in K \cap Y\}$$



Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

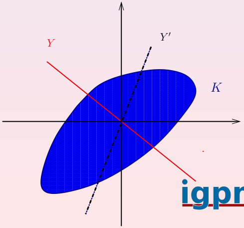
$$\inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

$$\Rightarrow \quad k \leq c_0 n/\log(N/n)$$

$\rightsquigarrow$ Maximal optimality range $k \leq c_0 n/\log(N/n)$ (instead of $k \leq n/2$)

Gelfand widths: (Kashin, Gluskin/Garnaev)

$$d^n(K)_X := \inf_{Y:\mathrm{codim}\, Y\leq n} \sup\{\|x\|_X \, ; \, x \in K \cap Y\}$$



Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

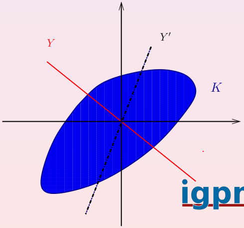
$$\sqrt{\frac{\log(N/n)+1}{n}} \sim \inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2}$$

$$\Rightarrow \quad k \leq c_0 n / \log(N/n)$$

$\rightsquigarrow$ Maximal optimality range $k \leq c_0 n/\log(N/n)$ (instead of $k \leq n/2$)

Gelfand widths: (Kashin, Gluskin/Garnaev)

$$d^n(K)_X := \inf_{Y:\text{codim } Y \leq n} \sup\{\|x\|_X \,;\, x \in K \cap Y\}$$



Maximal Optimality Range k? - back to the 70's

THEOREM:

Let $K := U(\ell_1^N) = \{x \in \mathbb{R}^N : \|x\|_{\ell_1} \leq 1\}$

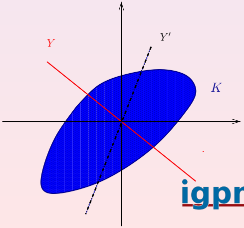
$$\sqrt{\frac{\log(N/n)+1}{n}} \sim \inf_{(\Phi,\Delta)\in\mathcal{A}_{n,N}} \max_{x\in K} \|x - \Delta(\Phi x)\|_{\ell_2} \overset{!}{\leq} C_0 \max_{x\in K} \sigma_k(x)_{\ell_2} \leq C_0 k^{-1/2}$$

$$\Rightarrow \quad k \leq c_0 n/\log(N/n)$$

$\rightsquigarrow$ Maximal optimality range $k \leq c_0 n/\log(N/n)$ (instead of $k \leq n/2$)

Gelfand widths: (Kashin, Gluskin/Garnaev)

$$d^n(K)_X := \inf_{Y:\mathrm{codim}\, Y \leq n} \sup\{\|x\|_X \,;\, x \in K \cap Y\}$$



A Breakthrough - ℓ_1-Minimimization - Φ, Δ?

What are good sensing matrices?

A Breakthrough - ℓ_1-Minimimization - Φ, Δ?

What are good sensing matrices?

Restricted Isometry Property - RIP(k, δ)

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

A Breakthrough - ℓ_1-Minimimization - Φ, Δ?

What are good sensing matrices?

Restricted Isometry Property - RIP(k, δ)

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

THEOREM [CT]: Praktikable decoder– ℓ_1-minimization

Let Φ satisfy RIP$(3k, \delta)$, decoder: $\Delta(y) := \operatorname{argmin}\{\|z\|_{\ell_1} : \Phi z = y\}$

A Breakthrough - ℓ_1-Minimimization - Φ, Δ?

What are good sensing matrices?

Restricted Isometry Property - RIP(k, δ)

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

THEOREM [CT]: Prakticable decoder– ℓ_1-minimization

Let Φ satisfy RIP$(3k, \delta)$, decoder: $\Delta(y) := \operatorname{argmin}\{\|z\|_{\ell_1} : \Phi z = y\}$

$$\Longrightarrow \|x - \Delta(\Phi x)\|_{\ell_2} \lesssim \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}} \quad k\text{-sparse exact}$$

A Breakthrough - ℓ_1-Minimimization - Φ, Δ?

What are good sensing matrices?

Restricted Isometry Property - RIP(k, δ)

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

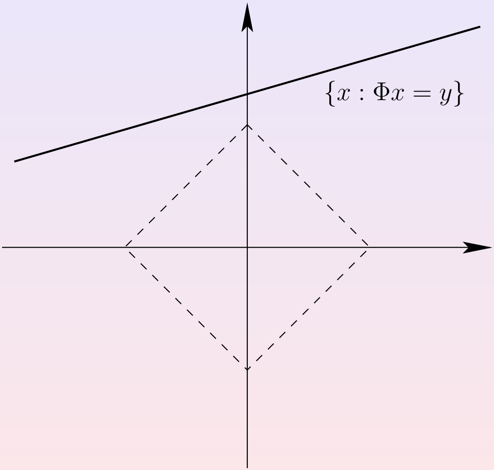
THEOREM [CT]: Praktikable decoder– ℓ_1-minimization

Let Φ satisfy RIP$(3k, \delta)$, decoder: $\Delta(y) := \operatorname{argmin}\{\|z\|_{\ell_1} : \Phi z = y\}$

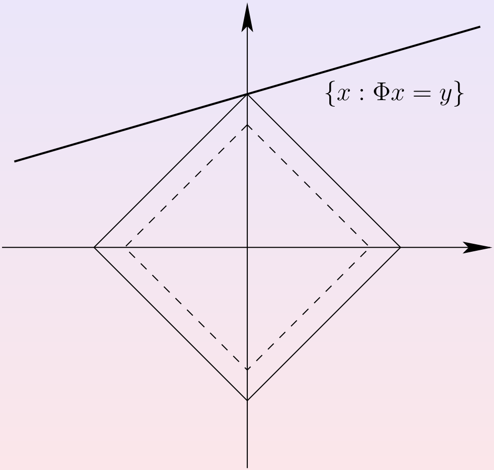
$$\Longrightarrow \|x - \Delta(\Phi x)\|_{\ell_2} \lesssim \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}} \quad k\text{-sparse exact}$$

Convex relaxation: $\min \|\cdot\|_{\ell_0} \rightarrow \min \|\cdot\|_{\ell_1}$

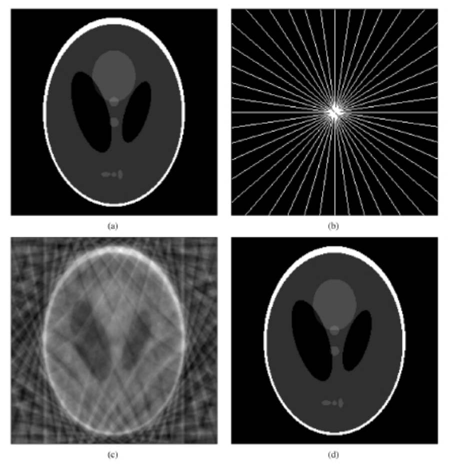
A Geometric Explanation - $n = 1, N = 2$



A Geometric Explanation - $n = 1, N = 2$



Closely Related: TV-Minimization... - Δ?



(Candes/Romberg/Tao)

How to Realize RIP? - Randomness Helps.. - Φ?

The “Rice - One Pixel - Camera”

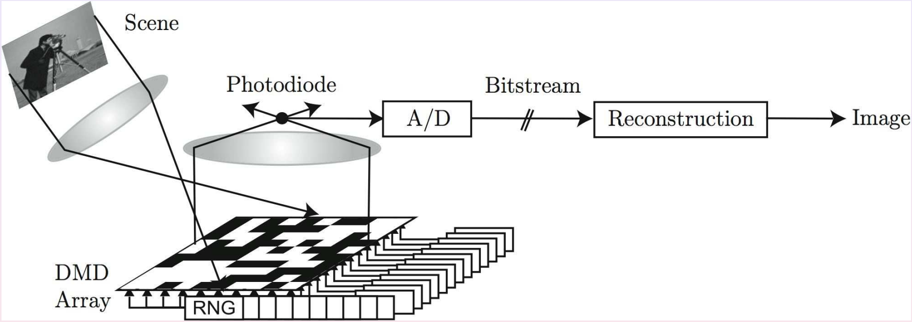


Figure: (R. Baraniuk, Rice University)

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

- **CoMP** : For each $x \in \mathbb{R}^N$, $\delta \in (0, 1]$,

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

- **CoMP** : For each $x \in \mathbb{R}^N$, $\delta \in (0, 1]$,

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

- **CoMP** : For each $x \in \mathbb{R}^N$, $\delta \in (0, 1]$,

$$\left| \|\Phi(\omega)x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2 \right| \leq \delta \|x\|_{\ell_2}^2$$

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

- **CoMP** : For each $x \in \mathbb{R}^N$, $\delta \in (0, 1]$,

$$\rho\Big\{\omega : \big|\|\Phi(\omega)x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2\big| \le \delta\|x\|_{\ell_2}^2\Big\} \ge 1 - be^{-cn\delta^2}$$

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

- **CoMP** : For each $x \in \mathbb{R}^N$, $\delta \in (0, 1]$,

$$\rho\Big\{\omega : \big|\|\Phi(\omega)x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2\big| \leq \delta\|x\|_{\ell_2}^2\Big\} \geq 1 - be^{-cn\delta^2}$$

THEOREM:

CoMP implies RIP(k, δ) for each $\delta \in (0, 1]$ with high probability when

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

- **CoMP** : For each $x \in \mathbb{R}^N$, $\delta \in (0, 1]$,

$$\rho\Big\{\omega : \big|\|\Phi(\omega)x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2\big| \le \delta\|x\|_{\ell_2}^2\Big\} \ge 1 - be^{-cn\delta^2}$$

THEOREM:

CoMP implies RIP(k, δ) for each $\delta \in (0, 1]$ with high probability when

$$k \le c_0 n / \log(N/n)$$

Randomness Helps...Concentration of Measure Property

Gaussian matrices: $\phi_{i,j} \in \mathcal{N}(0, 1/n)$

Bernoulli matricen: $\phi_{i,j} = \pm 1/\sqrt{n}$ with equal probability

In general: $\Phi = \Phi(\omega)$: $n \times N$ over a probability space (Ω, ρ)

- **CoMP** : For each $x \in \mathbb{R}^N$, $\delta \in (0, 1]$,

$$\rho\Big\{\omega : \big|\|\Phi(\omega)x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2\big| \leq \delta\|x\|_{\ell_2}^2\Big\} \geq 1 - be^{-cn\delta^2}$$

THEOREM:

CoMP implies RIP(k, δ) for each $\delta \in (0, 1]$ with high probability when

$$k \leq c_0 n / \log(N/n)$$

Gauß-, Bernoulli-, uniformly distributed points on the $(n-1)$-sphere satisfy **CoMP**

The Norm Matters...

THEOREM: $\|x\|_{\ell_1} = \sum_{j=1}^{N} |x_j|$:

Assume that Φ satisfies RIP$(3k, \delta)$ and $\Delta(y) := \mathrm{argmin}_{\Phi z = y} \|z\|_{\ell_1}$

The Norm Matters...

THEOREM: $\|x\|_{\ell_1} = \sum_{j=1}^N |x_j|$:

Assume that Φ satisfies RIP$(3k, \delta)$ and $\Delta(y) := \mathrm{argmin}_{\Phi z = y} \|z\|_{\ell_1}$
$\Longrightarrow$

$$\|x - \Delta(\Phi x)\|_{\ell_1} \leq C(\delta)\sigma_k(x)_{\ell_1}$$

The Norm Matters...

THEOREM: $\|x\|_{\ell_1} = \sum_{j=1}^N |x_j|$:

Assume that Φ satisfies RIP$(3k, \delta)$ and $\Delta(y) := \mathrm{argmin}_{\Phi z = y} \|z\|_{\ell_1}$ $\Longrightarrow$

$$\|x - \Delta(\Phi x)\|_{\ell_1} \leq C(\delta)\sigma_k(x)_{\ell_1}$$

THEOREM: $\|x\|_{\ell_2} = \left(\sum_{j=1}^N |x_j|^2\right)^{1/2}$:

- For each (Φ, Δ) : 1-term optimality $\forall\, x \in \mathbb{R}^N \Longrightarrow n \geq aN$

The Norm Matters...

THEOREM: $\|x\|_{\ell_1} = \sum_{j=1}^N |x_j|$:

Assume that Φ satisfies RIP$(3k,\delta)$ and $\Delta(y) := \mathrm{argmin}_{\Phi z = y} \|z\|_{\ell_1}$ $\Longrightarrow$

$$\|x - \Delta(\Phi x)\|_{\ell_1} \le C(\delta)\sigma_k(x)_{\ell_1}$$

THEOREM: $\|x\|_{\ell_2} = \left(\sum_{j=1}^N |x_j|^2\right)^{1/2}$:

- For each (Φ,Δ) : 1-term optimality $\forall\, x \in \mathbb{R}^N \Longrightarrow n \ge aN$

- Assume that $\{\Phi(\omega)\}$ satisfies CoMP. Then $\exists\ \Delta$ s.t. for each $x \in \mathbb{R}^N$ with high probability (with respect to .Φ) one has

$$\|x - \Delta(\Phi x)\|_{\ell_2} \le C_0\sigma_k(x)_{\ell_2}, \quad k \lesssim n/\log(N/n)$$

An Optimal Decoder

Decoder:

(i) Set $r^0 := y$, $x^0 := 0$, $j = 0$, $\Lambda_0 = \bar{\Lambda}_0 = \emptyset$

An Optimal Decoder

Decoder:

(i) Set $r^0 := y$, $x^0 := 0$, $j = 0$, $\Lambda_0 = \bar{\Lambda}_0 = \emptyset$

(ii) compute $r^j := y - \Phi x^j$ and set

$$\Gamma_{j+1} := \left\{ i \in \{1, \dots, N\} : |\langle r^j, \phi_i \rangle| \geq \frac{\delta \|r^j\|_{\ell_2}}{\sqrt{k}} \right\}$$

An Optimal Decoder

Decoder:

(i) Set $r^0 := y$, $x^0 := 0$, $j = 0$, $\Lambda_0 = \bar{\Lambda}_0 = \emptyset$

(ii) compute $r^j := y - \Phi x^j$ and set

$$\Gamma_{j+1} := \left\{ i \in \{1, \dots, N\} : |\langle r^j, \phi_i \rangle| \geq \frac{\delta \|r^j\|_{\ell_2}}{\sqrt{k}} \right\}$$

If $\Gamma_{j+1} = \emptyset$ or $j = k$, stop; output $\Lambda^* = \Lambda_j$ and $x^* := x^j$.
else $\bar{\Lambda}_{j+1} := \Lambda_j \cup \Gamma_{j+1}$;

An Optimal Decoder

Decoder:

(i) Set $r^0 := y$, $x^0 := 0$, $j = 0$, $\Lambda_0 = \bar{\Lambda}_0 = \emptyset$

(ii) compute $r^j := y - \Phi x^j$ and set

$$\Gamma_{j+1} := \left\{ i \in \{1, \ldots, N\} : |\langle r^j, \phi_i \rangle| \geq \frac{\delta \|r^j\|_{\ell_2}}{\sqrt{k}} \right\}$$

If $\Gamma_{j+1} = \emptyset$ or $j = k$, stop; output $\Lambda^* = \Lambda_j$ and $x^* := x^j$.
else $\bar{\Lambda}_{j+1} := \Lambda_j \cup \Gamma_{j+1}$;

(iii) compute

$$\hat{x} := \mathrm{argmin}_{\mathrm{supp} z = \bar{\Lambda}_{j+1}} \|\Phi z - y\|_{\ell_2}$$

An Optimal Decoder

Decoder:

(i) Set $r^0 := y$, $x^0 := 0$, $j = 0$, $\Lambda_0 = \bar{\Lambda}_0 = \emptyset$

(ii) compute $r^j := y - \Phi x^j$ and set

$$\Gamma_{j+1} := \left\{ i \in \{1, \dots, N\} : |\langle r^j, \phi_i \rangle| \geq \frac{\delta \|r^j\|_{\ell_2}}{\sqrt{k}} \right\}$$

If $\Gamma_{j+1} = \emptyset$ or $j = k$, stop; output $\Lambda^* = \Lambda_j$ and $x^* := x^j$.
else $\bar{\Lambda}_{j+1} := \Lambda_j \cup \Gamma_{j+1}$;

(iii) compute

$$\hat{x} := \mathrm{argmin}_{\mathrm{supp} z = \bar{\Lambda}_{j+1}} \|\Phi z - y\|_{\ell_2}$$

take $x^{j+1} =$ best k-term approximation of $\hat{x}$ und $\Lambda_{j+1} := \mathrm{supp}\, x^{j+1}$, set $j + 1 \to j$, and go to (ii).

Optimality

- k-Term optimality for ℓ_2 with high probability for the maximal range of k

Optimality

- k-Term optimality for ℓ_2 with high probability for the maximal range of k
- Robust with respect to noise

Optimality

- k-Term optimality for ℓ_2 with high probability for the maximal range of k
- Robust with respect to noise

THEOREM:

Let $\delta \leq 1/16$ and let $\{\Phi = \Phi(\omega)\}$ satisfy **CoMP**.

- k-Term optimality for ℓ_2 with high probability for the maximal range of k
- Robust with respect to noise

THEOREM:

Let $\delta \leq 1/16$ and let $\{\Phi = \Phi(\omega)\}$ satisfy **CoMP**. If $y = \Phi x + e$, then

$$\|x - x^*\|_{\ell_2} \leq C(\sigma_k(x)_{\ell_2} + \|e\|_{\ell_2}), \quad k \leq an/\log(N/n)$$

for all draws from $\Omega^* \subset \Omega$ of measure at least $1 - N^{-\beta}$, where $a \searrow$ when $\beta \nearrow$.

Optimality

- k-Term optimality for ℓ_2 with high probability for the maximal range of k
- Robust with respect to noise

THEOREM:

Let $\delta \le 1/16$ and let $\{\Phi = \Phi(\omega)\}$ satisfy **CoMP**. If $y = \Phi x + e$, then

$$\|x - x^*\|_{\ell_2} \le C(\sigma_k(x)_{\ell_2} + \|e\|_{\ell_2}), \quad k \le an/\log(N/n)$$

for all draws from $\Omega^* \subset \Omega$ of measure at least $1 - N^{-\beta}$, where $a \searrow$ when $\beta \nearrow$.

- Thresholding is more efficient, but requires smaller δ

Optimality

- k-Term optimality for ℓ_2 with high probability for the maximal range of k
- Robust with respect to noise

THEOREM:

Let $\delta \leq 1/16$ and let $\{\Phi = \Phi(\omega)\}$ satisfy **CoMP**. If $y = \Phi x + e$, then

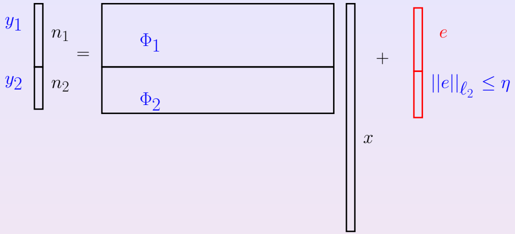
$$\|x - x^*\|_{\ell_2} \leq C(\sigma_k(x)_{\ell_2} + \|e\|_{\ell_2}), \quad k \leq an/\log(N/n)$$

for all draws from $\Omega^* \subset \Omega$ of measure at least $1 - N^{-\beta}$, where $a \searrow$ when $\beta \nearrow$.

- Thresholding is more efficient, but requires smaller δ
- ℓ_1-minimimization/regularization is more robust but more expensive...

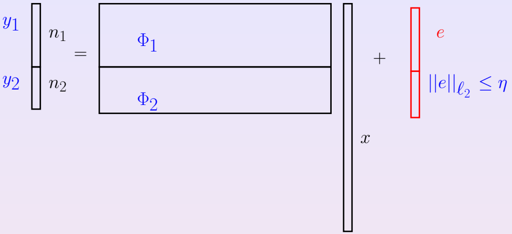
igpm

ℓ_1-Regularization



$$\begin{aligned} y_1 &= \Phi_1 x + e_1 \\ y_2 &= \Phi_2 x + e_2 \end{aligned}$$

ℓ_1-Regularization



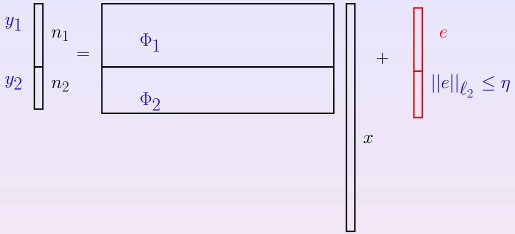
$$
\begin{aligned}
y_1 &= \Phi_1 x + e_1 \\
y_2 &= \Phi_2 x + e_2
\end{aligned}
$$

Decoder:

1. For $j = 1, \cdots, J$ set $\epsilon_j = 2^{2-j}\|y\|_{\ell_2}$, and solve

$$z_j := \operatorname{Argmin}_{\|\Phi_1 z - y_1\|_{\ell_2} \leq \epsilon_j} \|z\|_{\ell_1}$$

ℓ_1-Regularization



$$y_1 = \Phi_1 x + e_1$$
$$y_2 = \Phi_2 x + e_2$$

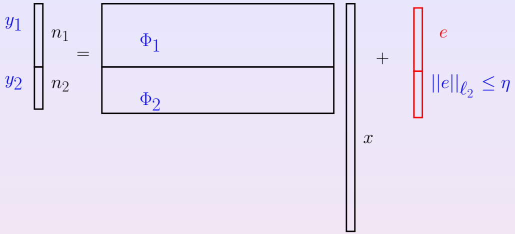
Decoder:

1. For $j = 1, \cdots, J$ set $\epsilon_j = 2^{2-j}\|y\|_{\ell_2}$, and solve

$$z_j := \mathrm{Argmin}_{\|\Phi_1 z - y_1\|_{\ell_2} \leq \epsilon_j} \|z\|_{\ell_1}$$

2. Define $j^* = \mathrm{Argmin}_{j=1,\cdots,J}\|\Phi_2 z_j - y_2\|_{\ell_2}$

ℓ_1-Regularization



$$y_1 = \Phi_1 x + e_1$$
$$y_2 = \Phi_2 x + e_2$$

Decoder:

1. For $j = 1, \cdots, J$ set $\epsilon_j = 2^{2-j}\|y\|_{\ell_2}$, and solve

$$z_j := \operatorname{Argmin}_{\|\Phi_1 z - y_1\|_{\ell_2} \leq \epsilon_j} \|z\|_{\ell_1}$$

2. Define $j^* = \operatorname{Argmin}_{j=1,\cdots,J} \|\Phi_2 z_j - y_2\|_{\ell_2}$ and set

$$x^* = \Delta(y) = z_{j^*}.$$

Near Optimal Performance

THEOREM:

For $k \lesssim n/\log(\frac{N}{n})$ one has

Near Optimal Performance

THEOREM:

For $k \lesssim n/\log(\frac{N}{n})$ one has

$$\|x - x^*\|_{\ell_2} \leq C(\sigma_k(x)_{\ell_2} + 2^{-J}\|x\|_{\ell_2} + \eta),$$

Near Optimal Performance

THEOREM:

For $k \lesssim n/\log(\frac{N}{n})$ one has

$$\|x - x^*\|_{\ell_2} \leq C(\sigma_k(x)_{\ell_2} + 2^{-J}\|x\|_{\ell_2} + \eta),$$

with probability larger than $1 - t(n_1, n_2)$

Near Optimal Performance

THEOREM:

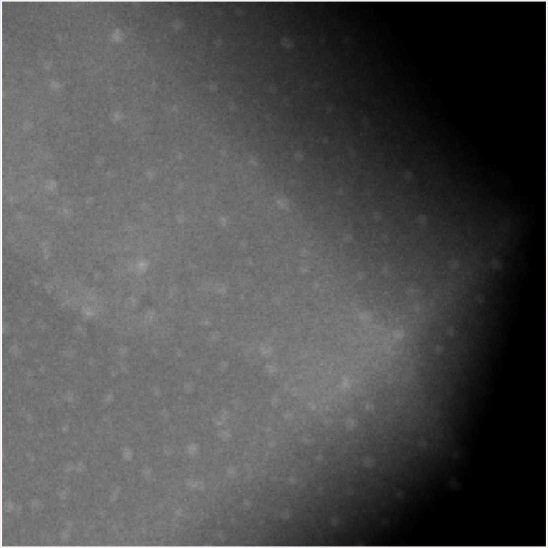
For $k \lesssim n/\log(\frac{N}{n})$ one has

$$\|x - x^*\|_{\ell_2} \le C(\sigma_k(x)_{\ell_2} + 2^{-J}\|x\|_{\ell_2} + \eta),$$

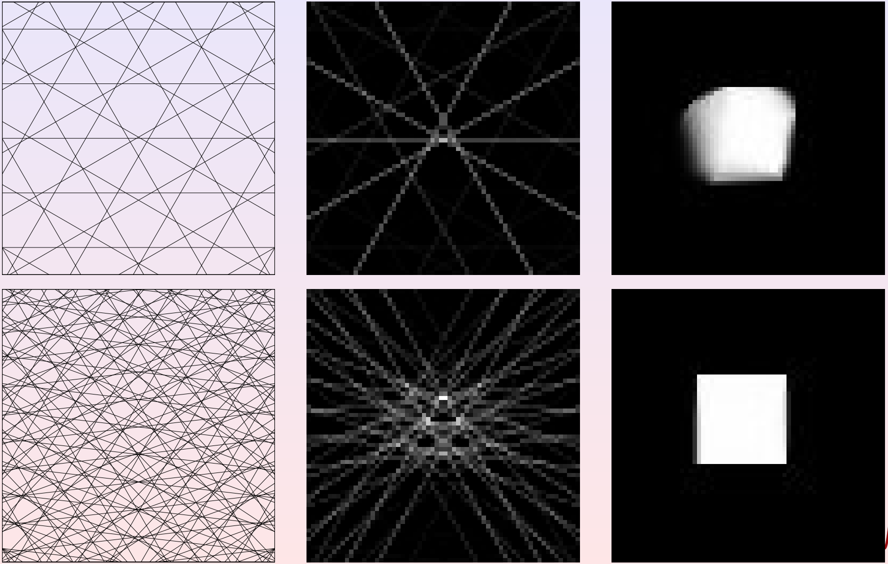
with probability larger than $1 - t(n_1, n_2)$ where

$$t(n_1, n_2) = Jb_2 e^{-c_2 n_2/2} + b_1(e^{-n_1 \frac{c_1}{32}} + e^{-c_1 n_1} + e^{-c_1 n_1/2})$$

Electron-Tomography



Tests



Tests