Some topics on Sparsity

Sangnam Nam

August 12, 2010
Why Sparsity?: randomly generated image
Why Sparsity?: randomly generated image
Why Sparsity?: randomly generated image
Why Sparsity?: randomly generated image
Why Sparsity?: randomly generated image
Why Sparsity?: randomly generated image
Why Sparsity?: DCT of randomly generated image
Why Sparsity?: DCT of randomly generated image
Compression: Image
Compression: Image
Compression: Audio
Denoising
Denoising
Denoising: Simple example

signal $x$

dct transform of $x$
Denoising: Simple example

signal + noise  
dct transform of (signal + noise)
Denoising: Simple example

Denoising with hard thresholding at 0.07
Denoising: Simple example

Denoising with hard thresholding at 0.18
Denoising: Simple example

With augmented dictionary:

Transform coefficients

Denoised Result
Inpainting: Original Image

original image 512x512 = 262144
Inpainting: letter

(Reuters) - Eating bacon, sausage, hot dogs and other processed meats such as salami and sausages, has been linked to increased risk of heart attacks and diabetes, according to a new study.

"To lower risk of heart attacks and diabetes, people should eat less processed meats,"

"Processed meats such as bacon, sausages, hot dogs and other processed meats, have been linked to increased risk of heart attacks and diabetes, according to a new study."

Based on her findings, she said people who eat one serving of processed meat daily have a 12% greater risk of heart disease and a 17% greater risk of diabetes.

The American Meat Institute objected to the findings, saying the study did not take into account the benefits of consuming lean cuts of meat.

"At best, this hypothesis merits further study. It is certainly not a reason to ignore the many benefits of consuming lean cuts of meat," Institute spokesman John Althouse said.

Most dietary guidelines recommend eating less meat, but studies rarely look for differences in risk between different types of meat.

She and colleagues did a systematic review of nearly 100,000 people from 14 studies.

They defined processed meat as any meat preserved by smoking, curing, salting or adding other preservatives.
Inpainting: letter

filled-in image: snr=25.2543
Inpainting: letter

reconstructed image: snr=34.3195
Inpainting: lines

Given image $512 \times 512 = 262144$
Inpainting: lines

filled-in image: snr=23.5627
Inpainting: lines

reconstructed image: snr=30.8017
Inpainting: random

Given image $512 \times 512$, 262144, 77.8% missing
Inpainting: random

filled-in image: snr=16.6297
Inpainting: random

reconstructed image: snr=21.263
Compressed Sensing

original image 256x256 = 65536
Compressed Sensing

Sample locations: 3032
Compressed Sensing

reconstructed image
Compressed Sensing
Compressed Sensing

sample locations: 7112
Compressed Sensing

reconstructed image
Separation
Separation
Super-resolution
Super-resolution
Notation/Terminology

- $y \in \mathbb{R}^d$: signal
- $\phi \in \Phi$: a column of $\Phi \in \mathbb{R}^{d \times K}$
- $\{\phi : \phi \in \Phi\}$ spans $\mathbb{R}^d \Rightarrow \Phi$: a dictionary for $\mathbb{R}^d$
- $\phi \in \Phi$: an atom
- $y = \Phi x \Rightarrow x$: a coefficient vector / representation
Sparse Model

signal  dictionary  representation
\( \ell_0 \)-minimization

To find the sparsest representation of \( y \): we solve

\[
\min_x \|x\|_0 \quad \text{subject to} \quad y = Mx
\]
$\ell_0$-minimization

To find the sparsest representation of $y$: we solve

$$\min_x \|x\|_0 \quad \text{subject to } y = Mx$$

How do we know we have the solution?
Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \# \{ i : x_i \neq 0 \}$. 
Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \# \{ i : x_i \neq 0 \}$.
- Let $\text{Spark of } \Phi := \min_{x \in \ker(\Phi), x \neq 0} \|x\|_0$.
Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \#\{i : x_i \neq 0\}$.
- Let $\text{Spark of } \Phi := \min_{x \in \ker(\Phi), x \neq 0} \|x\|_0$.
- Let $y = Mx$, $\|x\|_0 \leq (\text{Spark of } \Phi)/2$.
Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \#\{i : x_i \neq 0\}$.
- Let $\text{Spark of } \Phi := \min_{x \in \ker(\Phi), x \neq 0} \|x\|_0$.
- Let $y = Mx$, $\|x\|_0 \leq (\text{Spark of } \Phi)/2$

$\Rightarrow x$ is unique solution of $\min_z \|z\|_0$ subject to $y = Mz$. 
Good, we (kind of) know when we have the solution.
\( \ell_0 \)-minimization

Good, we (kind of) know when we have the solution. Unfortunately, solving \( \ell_0 \)-minimization is NP-hard.
$\ell_0$-minimization

Good, we (kind of) know when we have the solution. Unfortunately, solving $\ell_0$-minimization is NP-hard. What to do?
Alternatives

- Greedy Algorithm/Matching Pursuit
- Convex Optimization/Basis Pursuit
Matching Pursuit

Matching Pursuit [Mallat & Zhang]

1. Given signal $y$, dictionary $\Phi$.
2. Set $k = 0$, $\hat{y}_0 = 0$, $r_0 = y$.
3. Find $\phi := \arg \min_{\phi \in \Phi} \langle \phi, r_k \rangle$.
4. Update $\hat{y}_{k+1} := \hat{y}_k + \langle r_k, \phi \rangle \phi$, $r_{k+1} := r_k - \langle r_k, \phi \rangle \phi$.
5. Increment $k$.
6. If $\hat{y}_k$ uses a specified number of atoms or $r_k$ is smaller than a specified error, stop. Otherwise, go to step 3.
Energy Preservation

Energy Preservation (Pythagoras theorem)

\[ \| r_k \|_2^2 = \| r_{k+1} \|_2^2 + |\langle r_k, \phi \rangle|^2 \]
Drawback of Matching Pursuit?

MP can select an atom more than once.
Orthogonal Matching Pursuit

Orthogonal Matching Pursuit (OMP)

1. Given signal $y$, dictionary $\Phi$.
2. Set $k = 0$, $\Lambda = \emptyset$, $\hat{y}_0 = 0$, $r_0 = y$.
3. Find $i_k := \arg\min_i \langle \phi_i, r_k \rangle$ and set $\Lambda := \Lambda \cup \{i_k\}$.
4. Compute $\Delta \hat{y} := \text{(best } \ell_2\text{-approximation from span of } \Phi_\Lambda \text{ to } r_k\text{)}$, update $\hat{y}_{k+1} := \hat{y}_k + \Delta \hat{y}$, and $r_{k+1} := r_k - \Delta \hat{y}$.
5. Increment $k$.
6. If $\hat{y}_k$ uses a specified number of atoms or $r_k$ is smaller than a specified error, stop. Otherwise, go to step 3.
Variety of Matching Pursuits

- Morphological Component Analysis [MCA, Bobin et al]
- Stagewise OMP [Donoho et al]
- CoSAMP [Needell & Tropp]
- ROMP [Needell & Vershynin]
- Iterative Hard Thresholding [Blumensath & Davies]
To find good approximation of $f : [0, 5] \rightarrow \mathbb{R}$, one may sample the function values

1. Uniformly at $N$ locations for some large $N$, or
2. Adaptively sample—more locations near 5 in the figure— at some large number $N$ locations, then, use linear interpolation, cubic spline approximation, etc.
What if it is a polynomial of degree 6?

What we know it is a polynomial of degree 6? One needs only 7 different samples! Approximation is perfect!
What if it is a sum of 6 monomials of degree less than 200? How many samples do we need?
Think about Economy!

Digital Camera

1. High resolution photo (e.g. size 1024x1024) requires large number of samples (e.g. 1024x1024 samples).
2. Image is (immediately) compressed to smaller and lossy JPEG file.
3. Not a big deal for digital cameras?
Shannon Sampling / Nyquist Rate

Nyquist rate $\Rightarrow$ 44.1 KHz sampling rate
Shannon Sampling / Nyquist Rate

Nyquist rate $\Rightarrow$ 44.1 KHz sampling rate
$\Rightarrow$ large audio file
Shannon Sampling / Nyquist Rate

Nyquist rate ⇒ 44.1 KHz sampling rate
⇒ large audio file
⇒ Compress to lossy MP3 file.
Fewer Samples

- Can we sample less to begin with?
Fewer Samples

- Can we sample less to begin with?
- How do we recover/approximate the original data?
Fewer Samples

- Can we sample less to begin with?
- How do we recover/approximate the original data?
- Key idea to exploit: Sparsity
\( \ell_1 \)-minimization

Relax

\[
\min_x \|x\|_0 \quad \text{subject to } y = \Phi x
\]

Solve

\[
\min_x \|x\|_1 \quad \text{subject to } y = \Phi x
\]
$\ell_1$-minimization

Relax

$$\min_x \|x\|_0 \quad \text{subject to} \quad y = \Phi x$$

Solve

$$\min_x \|x\|_1 \quad \text{subject to} \quad y = \Phi x$$

Convex problem. Can be recast into Linear Programming.
Why does $\ell_1$-minimization work?

Feasible solutions
Why does $\ell_1$-minimization work?

$\ell_2$-balls

$\ell_1$-balls

$\ell_p$-balls, $0 < p < 1$
Why does $\ell_1$-minimization work?

$\ell_2$-minimizer  $\ell_1$-minimizer  $\ell_p$-minimizer, $0 < p < 1$
Why does $\ell_1$-minimization work?

When can we guarantee successful recovery?
Null Space Property

Notation:
\[ N(\Phi) := \text{(Kernel / Null Space of } \Phi) \]
\[ \Lambda := \text{(Support of } x^*) \]
\[ \bar{\Lambda} := \Lambda^c \]
Null Space Property

\[
\min_{x} \|x\|_1 \quad \text{s.t.} \quad \Phi x = \Phi x^* \quad (A)
\]

\(\ell_1\)-minimization (A) recovers \(x^*\) if and only if

\[
|\langle z, \text{sign}(x^*) \rangle| < \|z_{\Lambda}\|_1
\]

for all \(z \in N(\Phi)\).
Null Space Property

\[
\min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = \Phi x^* \tag{A}
\]

\(\ell_1\)-minimization (A) recovers \(x^*\) if and only if

\[|\langle z, \text{sign}(x^*) \rangle| < \|z_{\overline{\Lambda}}\|_1\]

\(\ell_1\)-minimization (A) recovers every \(x^*\) with support \(\Lambda\) if

\[\|z_{\Lambda}\|_1 < \|z_{\overline{\Lambda}}\|_1\]

for all \(z \in \mathcal{N}(\Phi)\).
Null Space Property

\[
\min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = \Phi x^*
\] (A)

\ell_1\text{-minimization (A) recovers } x^* \text{ if and only if }

\[|\langle z, \text{sign}(x^*) \rangle| < \|z_\Lambda\|_1\]

\ell_1\text{-minimization (A) recovers every } x^* \text{ with support } \Lambda \text{ if }

\[\|z_\Lambda\|_1 < \|z_{\bar{\Lambda}}\|_1\]

\ell_1\text{-minimization (A) recovers every } k\text{-sparse } x^* \text{ if }

\[\|z_\Lambda\|_1 < \|z_{\bar{\Lambda}}\|_1\]

for all \( z \in \mathcal{N}(\Phi) \) and \( \Lambda \) with \( \#\Lambda \leq k \).
Uniqueness for $\ell_1$ using NSP

If $x^*$ is $k$-sparse, then

$$\|x^* + z\|_1 \geq \|x^*\|_1 - \|z_\Lambda\|_1 + \|z_{\bar{\Lambda}}\|_1 \geq \|x^*\|_1.$$ 

I.e., $x^*$ is the unique $\ell_1$-minimizer.
Coherence

Coherence $M(\Phi) := \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$. 

Theorem

If $\|x^*\|_0 \leq \frac{1}{2}(1 + M(\Phi))$ is a solution of $y = Mx$, then $x^*$ is the unique minimizer of the $\ell_1$-problem. $x^*$ is also the unique minimizer of the $\ell_0$-problem!
Coherence

Coherence $M(\Phi) := \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$.

[Gribonval,Nielsen]

**Theorem**

If $\|x^*\|_0 \leq \frac{1}{2} \left( 1 + \frac{1}{M(\Phi)} \right)$ is a solution of $y = Mx$, then $x^*$ is the unique minimizer of the $\ell_1$-problem. $x^*$ is also the unique minimizer of the $\ell_0$-problem!
Derivation of $\ell_1$-guaranteeing sparsity level

For $z \in N(\Phi)$,

$$\phi_j z_j = - \sum_{i \neq j} \phi_i z_i.$$
Derivation of $\ell_1$-guaranteeing sparsity level

For $z \in \mathcal{N}(\Phi)$,

$$\phi_j z_j = - \sum_{i \neq j} \phi_i z_i.$$ 

$$|z_j| \leq M(\Phi) \sum_{i \neq j} |z_i|$$

$$\|z\|_1 \leq (\#\Lambda)M(\Phi)\|z\|_1 \Rightarrow \|z\|_1 \leq 2\|z\|_1$$

if $\#\Lambda \leq 1/(1 + M(\Phi))$. 

Derivation of $\ell_1$-guaranteeing sparsity level

For $z \in N(\Phi)$,

$$\phi_j z_j = -\sum_{i \neq j} \phi_i z_i.$$ 

$$|z_j| \leq M(\Phi) \sum_{i \neq j} |z_i|$$

$$(1 + M(\Phi)) |z_j| \leq M(\Phi) \|z\|_1$$
Derivation of $\ell_1$-guaranteeing sparsity level

For $z \in N(\Phi)$,

\[ \phi_j z_j = - \sum_{i \neq j} \phi_i z_i. \]

\[ |z_j| \leq M(\Phi) \sum_{i \neq j} |z_i| \]

\[ (1 + M(\Phi))|z_j| \leq M(\Phi) \|z\|_1 \]

\[ (1 + M(\Phi)) \|z_\Lambda\|_1 \leq (#\Lambda)M(\Phi) \|z\|_1 \]
Derivation of \( \ell_1 \)-guaranteeing sparsity level

For \( z \in N(\Phi) \),

\[
\phi_j z_j = - \sum_{i \neq j} \phi_i z_i.
\]

\[
|z_j| \leq M(\Phi) \sum_{i \neq j} |z_i|
\]

\[
(1 + M(\Phi))|z_j| \leq M(\Phi) \|z\|_1
\]

\[
(1 + M(\Phi)) \|z_\Lambda\|_1 \leq (#\Lambda) M(\Phi) \|z\|_1
\]

\[
\Rightarrow \|z_\Lambda\|_1 \leq \frac{\|z\|_1}{2} \text{ if } #\Lambda \leq \frac{1}{2}(1 + \frac{1}{M(\Phi)}).
\]
Coherence gives suboptimal estimates

For $\Phi \in \mathbb{R}^{d \times K}$, we may suppose $M(\Phi) = O\left(\frac{1}{\sqrt{d}}\right)$.
Coherence gives suboptimal estimates

For $\Phi \in \mathbb{R}^{d \times K}$, we may suppose $M(\Phi) = O\left(\frac{1}{\sqrt{d}}\right)$.

We can recover $x^*$ if $x^*$ is $\frac{1}{2} \left(1 + \sqrt{O(d)}\right) = O(\sqrt{d})$ sparse.
Coherence gives suboptimal estimates

For $\Phi \in \mathbb{R}^{d \times K}$, we may suppose $M(\Phi) = O\left(\frac{1}{\sqrt{d}}\right)$.

We can recover $x^*$ if $x^*$ is $\frac{1}{2} \left(1 + \sqrt{O(d)}\right) = O(\sqrt{d})$ sparse.

To put it differently, in order to recover $k$-sparse $x^*$, $d = O(k^2)$ measurements are desired.
Recovery of Logan-Shepp phantom using Fourier Transform Samples

An Image to recover:

original image512x512=262144
Recovery of Logan-Shepp phantom using Fourier Transform Samples

Available samples in Fourier Domain (2.7% (4.3%)):

sample locations: 7112
Recovery of Logan-Shepp phantom using Fourier Transform Samples

Reconstructed image—perfect!
How many samples do we need?

To recover

\[ k \leq m \leq N \]

Need to identify \( k \) nonzero positions out of \( N \) locations

\[ \log_2 \left( \frac{N}{k} \right) \approx k \log N \] measurements
How many samples do we need?

To recover

we (obviously) need \( k \leq m \leq N \) samples.
How many samples do we need?

To recover

we (obviously) need \( k \leq m \leq N \) samples.

Need to identify \( k \) nonzero positions out of \( N \) locations \( \Rightarrow \)

\[
\log_2 \left( \binom{N}{k} \right) \approx k \log \frac{N}{k}
\]

measurements
Random Sampling of Fourier Transform

[Candès, Romberg, Tao]

Theorem

Let \( N = n^2 \), \( f \) be an \( n \times n \) real-valued image supported on \( T \). Let \( \Omega \) be a subset of \( \{1, \ldots, n\}^2 \) chosen uniformly at random. If

\[
\#\Omega \geq C(\# T) \log N,
\]

then with high probability \((1 - O(N^{-M}))\), the minimizer to

\[
\min_g \|g\|_1 \quad s.t. \quad \hat{g}|_{\Omega} = \hat{f}|_{\Omega}
\]

is unique and is equal to \( f \).
For $s \in \mathbb{N}$, a matrix $\Phi$ satisfies the Restricted Isometry Property with the isometry constant $\delta_s$ if

$$(1 - \delta_s) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2$$

for all $s$-sparse signal $x$. 
Perfect Recovery condition in terms of RIP constant

[Candès]

**Theorem**

Assume that $\delta_{2s} < \sqrt{2} - 1$. Then, the solution $x^*$ to

$$
\min_{\tilde{x}} \|\tilde{x}\|_1 \quad s.t. \quad M\tilde{x} = Mx
$$

satisfies

$$
\|x^* - x\|_2 \leq C_0 s^{-1/2} \|x - x_s\|_1
$$

where $C_0$ is a constant.

**Remark:** The sensing matrix $M$ is non-adaptive, but the $\ell_1$-minimization with $M$ recovers every $s$-sparse signal.
[Candès]

**Theorem**

*If $h$ is in the nullspace of $\Phi$, then*

$$
\|h_T\|_1 \leq \rho \|h_{T^c}\|_1, \quad \rho := \sqrt{2}\delta_{2s}(1 - \delta_{2s})^{-1}
$$

*for every $T$ with $\#T = s$.***
Examples of Measurement Matrices with good RIP

- Random matrices with i.i.d. entries. If entries of $\Phi \in \mathbb{R}^{m \times d}$ are drawn from i.i.d. Gaussian with mean 0 and variance $1/m$, then with overwhelming probability the RIC $\delta_s$ is less than $\sqrt{2} - 1$ when $s \leq Cm / \log(d/m)$.
- Fourier ensemble. $\Phi \in \mathbb{R}^{m \times d}$ is constructed by sampling random $m$ rows of the discrete Fourier transform.
- General orthogonal measurement ensembles.
Questions

1. ‘Random dictionaries’ are good for compressed sensing. Are there any deterministic dictionaries that are as good as random ones?
2. Any alternative to RIP?
3. What happens when the model is not exactly sparse?
4. What happens when there is noise?
Stable and Robust Recovery

We observe

\[ y = \Phi x + n. \]

We reconstruct \( x \) as the solution to

\[
\min_{\tilde{x}} \| \tilde{x} \|_1 \quad \text{s.t.} \quad \| y - \Phi \tilde{x} \|_2 \leq \epsilon.
\]
Stable and Robust Recovery

[Candès]

**Theorem**

Assume that \( \delta_{2s} < \sqrt{2} - 1 \) and \( \|n\|_2 \leq \epsilon \). Then, the solution \( x^* \) satisfies

\[
\|x^* - x\|_2 \leq C_0 s^{-1/2} \|x - x_s\|_1 + C_1 \epsilon
\]

where \( C_0 \) and \( C_1 \) are some constants.
Global Optimization Approach for Noisy Problem

We solve

$$\min_x \frac{1}{2} \| Mx - y \|_2^2 + \lambda \| x \|_1$$

for some $\lambda > 0$. 
Global Optimization Approach for Noisy Problem

We solve

$$\min_x \frac{1}{2} \| Mx - y \|^2_2 + \lambda \| x \|_1$$

for some $\lambda > 0$.

$\lambda \to 0$?
Global Optimization Approach for Noisy Problem

We solve

$$\min_x \frac{1}{2} \|Mx - y\|_2^2 + \lambda \|x\|_1$$

for some $\lambda > 0$.

$\lambda \to 0$?

$\lambda \to \infty$?
Compressed Sensing

Paper: Compressed Sensing by David Donoho

$$\|\theta - \theta_N\|_2 \leq C(N + 1)^{1/2 - 1/p}, \quad N = 0, 1, \ldots$$
We Need Good Dictionaries

Techniques/Algorithms of sparse modeling begin with some dictionary $\Phi$ that provides sparse representations of the class of signals of interest. What are those $\Phi$’s?
Good Dictionaries

- Smooth functions $\Rightarrow$ Fourier transform
- Smooth functions with point singularities $\Rightarrow$ Wavelets
- Singularities along smooth curves $\Rightarrow$ Curvelets, Shearlets, etc.
- …
PCA / SVD?

We want to design/learn dictionaries from data.
We want to design/learn dictionaries from data.

Principal Component Analysis ⇒ Not suitable for non-Gaussian mixtures
We want to design/learn dictionaries from data.

Principal Component Analysis $\Rightarrow$ Not suitable for non-Gaussian mixtures

[Olshausen & Field]
Used Sparsity Promoting regularization to learn Dictionary
$\Rightarrow$ Localized, oriented, bandpass receptive fields emerge
Dictionary Learning via $\ell_1$-criterion

If $\Phi$ were a good dictionary for a signal $x$, then we can recover $x$ by solving

$$\min_{\tilde{x}} \|\tilde{x}\|_1 \quad \text{s.t.} \quad y = \Phi \tilde{x}.$$
Dictionary Learning via $\ell_1$-criterion

If $\Phi$ were a good dictionary for a signal $x$, then we can recover $x$ by solving

$$\min_{\tilde{x}} \|\tilde{x}\|_1 \quad \text{s.t.} \quad y = \Phi \tilde{x}.$$ 

$\Rightarrow$ Given a data set $Y := [y_1, \ldots, y_N]$, we may find a good dictionary $\Phi$ by solving

$$\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X.$$
Some issues

- We can decrease $\|X\|_1$ as small as we wish by scaling because
  \[ \alpha \Phi \alpha^{-1} X = \Phi X \]
  for any $\alpha \in \mathbb{R}$. 

The problem
\[
\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X
\]
is not convex.
Some issues

- We can decrease $\|X\|_1$ as small as we wish by scaling because
  
  $$\alpha \Phi \alpha^{-1} X = \Phi X$$

  for any $\alpha \in \mathbb{R}$.

- The problem
  
  $$\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X$$

  is not convex.
Some issues

- We can decrease $\|X\|_1$ as small as we wish by scaling because

$$\alpha \Phi \alpha^{-1} X = \Phi X$$

for any $\alpha \in \mathbb{R}$.

- The problem

$$\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X$$

is not convex.

The first issue can be dealt with by requiring that each column of $\Phi$ to be of unit length. We will write $\Phi \in \mathcal{U}(d, K)$ to mean that $\Phi \in \mathbb{R}^{d \times K}$ is such a dictionary.
Dictionary Identification

General Question: Given a data set $Y$, we learned a dictionary $\Phi$ via some learning method. Is $\Phi$ the ‘right’ one?
Suppose that the training data $Y$ is generated by

$$Y = \Phi_0 X_0.$$ 

Under what condition on $X_0$ (and $\Phi_0$), can we be sure that the minimization problem

$$\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X$$

will recover the pair $\Phi_0$ and $X_0$?
Necessary and Sufficient Condition

[Gribonval & Schnass]

Theorem

Suppose that $K \geq d$. $(\Phi_0, X_0)$ is a strict local minimum of the problem

$$\min_{\phi, X} \|X\|_1 \quad s.t. \quad \Phi X = \Phi_0 X_0, \Phi \in \mathcal{U}(d, K)$$

if and only if

$$|\langle CX_0 + V, \text{sign}(X_0) \rangle| < \| (CX_0 + V) \overline{\Lambda} \|_1$$

for every $CX_0 + V \neq 0$ with $\text{diag}(\Phi_0^* \Phi_0 C) = 0$ and $V \in N(\Phi_0)$. 
Sufficient Condition for Basis

Notation:

Coefficient matrix $X \ 0$
Sufficient Condition for Basis

**Theorem**

$(\Phi_0, X_0)$ is a strict local minimum of the problem

$$\min_{\Phi, X} \|X\|_1 \quad s.t. \quad \Phi X = \Phi_0 X_0, \Phi \in U(d, d)$$

if for every $k = 1, \ldots, d$, there exists $d_k \in \mathbb{R}^d$ with $\|d_k\|_\infty < 1$ such that

$$\bar{X}_k d_k = X_k s_k - \text{diag} \left( (\|x^i\|)_i \right) m_k$$

where $m_k$ is the $k$-th column of $\Phi_0^* \Phi_0$. 


Geometric Interpretation

Local minimum

Not local minimum

\[ u_k := X_k s_k - \text{diag} \left( \left\langle \| x^i \| \right\rangle \right) m_k, \quad Q_k := [-1, 1]^{r_k}. \]
$X_0$ follows the Bernoulli-Gaussian Model with parameter $0 < p < 1$ if each entry of $X_0$ is the product of independent Bernoulli($p$) random variable $\xi_{ij}$ and Normal random variable $g_{ij}$, i.e.,

$$x_{ij} = \xi_{ij}g_{ij}.$$
If $X_0$ follows the Bernoulli-Gaussian distribution with parameter $p$, then with high probability $\bar{X}_k Q_k$ contains the $\ell_2$-ball of radius

$$\alpha \approx Np(1-p)\sqrt{\frac{2}{\pi}},$$
Concentration of Measure Phenomena

If $X_0$ follows the Bernoulli-Gaussian distribution with parameter $p$, then with high probability $\bar{X}_k Q_k$ contains the $\ell_2$-ball of radius

$$\alpha \approx Np(1 - p) \sqrt{\frac{2}{\pi}},$$

$X_k s_k$ is contained in the $\ell_2$-ball of radius

$$\beta \approx \sqrt{NKp},$$
Concentration of Measure Phenomena

If $X_0$ follows the Bernoulli-Gaussian distribution with parameter $p$, then with high probability $\bar{X}_k Q_k$ contains the $\ell_2$-ball of radius

$$\alpha \approx Np(1 - p) \sqrt{\frac{2}{\pi}},$$

$X_k s_k$ is contained in the $\ell_2$-ball of radius

$$\beta \approx \sqrt{NK} p,$$

and $\text{diag} \left( (\|x^i\|_i) \right) m_k$ is contained in the $\ell_2$-ball of radius

$$\gamma \approx Np \sqrt{\frac{2}{\pi}}.$$
Sufficient Condition under Bernoulli-Gaussian Model

If the coherence $M(\Phi_0)$ of $\Phi_0$ is less than $1 - p$ then for large enough $N$, $\Phi_0$ is locally identifiable.
Sufficient Condition under Bernoulli-Gaussian Model

If the coherence \( M(\Phi_0) \) of \( \Phi_0 \) is less than \( 1 - p \) then for large enough \( N \), \( \Phi_0 \) is locally identifiable.

**Surprise:** One needs only

\[
N \geq CK \log K.
\]
Some dictionary learning methods

- K-SVD [Elad et al.]
- ISI [Gowreesunker]
- ...