Sparse Signal Processing

Lei Yu[1]

[1]School of Electronic and Information
Wuhan University

Master Course, 2017

Outline

Outline

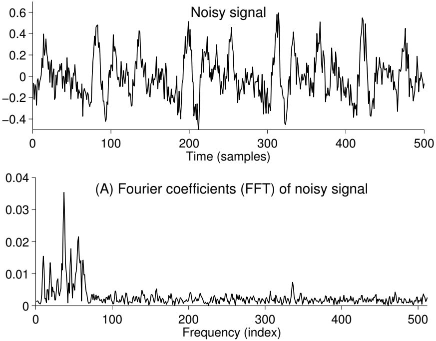
Motivation: Signal Denoising

Noisy speech signal y

$$y = s + w$$

s: noise-free speech signal

w: noise sequence



Motivation: Signal Denoising

Digital LTI filters are often used for noise reduction (denoising),

- lowpass filter
- highpass filter
- bandpass filter
- bandstop filter

but not applicable for

- the noise and signal overlap in the frequency domain
- the respective frequency bands are unknown

So, let's take a look at Sparsity!

Motivation: Signal Denoising

Assume the noise-freen speech signal s has a sparse set of Fourier coefficients:

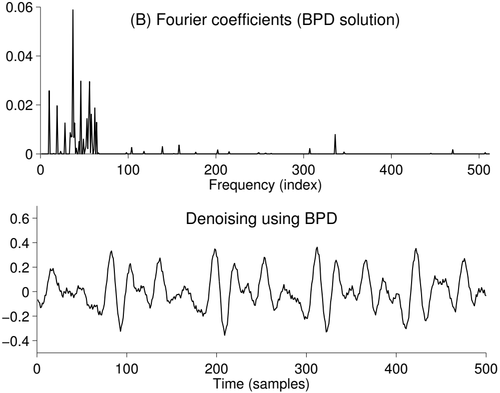
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{w}$$

$\boldsymbol{y}$: noisy speech signal, length M
$\boldsymbol{A}$: $M \times N$ DFT matrix
$\boldsymbol{c}$: sparse Fourier coefficients, length N
$\boldsymbol{w}$: noise, length M
Find estimation of $\boldsymbol{c}$ (BPD algorithm)

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \left\{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{c}\|_1 \right\}$$

Once $\hat{\boldsymbol{c}}$ is found, an estimate of the speech signal is given by $\hat{\boldsymbol{s}} = \boldsymbol{A}\hat{\boldsymbol{c}}$

Motivation: Signal Denoising



Motivation: Signal Deconvolution

If the signal of interest x is not only noisy but is also distorted by an LTI system with impulse response h, then the available data y is

$$y = h \circledast x + w \Longleftrightarrow \boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}$$

w is additive noise, h is known system function.
Applications include:

- echo cancellation
- direction of arrival estimation
- localization in GPS
- etc.

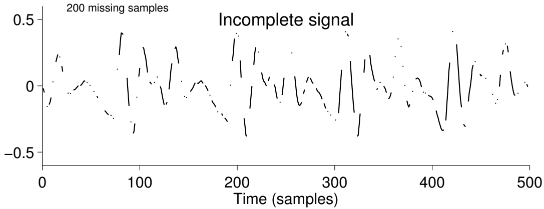
Motivation: Signal Inpainting

Due to data transmission/acquisition errors, some signal samples may be lost. Fill in missing values for error concealment.

Part of a signal or image may be intentionally deleted (image editing, etc). Convincingly fill in missing values according to the surrounding area to do inpainting.

$$y = Sx$$

S is the selection (sampling) operator



Motivation: Signal Separation

For a signal composed by two different type of data

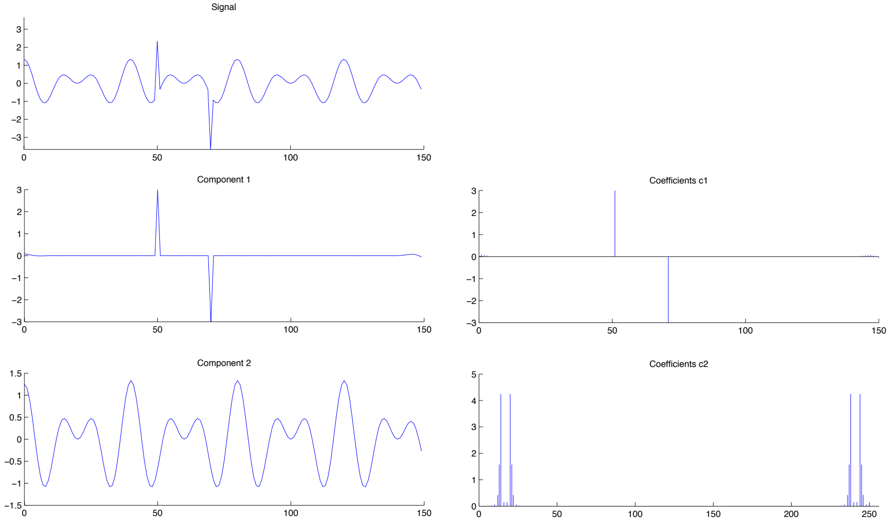
$$x = x_1 + x_2$$

x_1 is sparse under transform with operator A_1, x_2 is sparse under transform with operator A_2, then signal x can be separated by solving

$$\{\hat{c}_1, \hat{c}_2\} = \arg\min_{c_1, c_2} \left\{ \|x - A_1 c_1 - A_2 c_2\|_2^2 + \lambda_1 \|c_1\|_1 + \lambda_2 \|c_2\|_1 \right\}$$

Once get c_1, c_2, the two components can be estimated $\hat{x}_1 = A_1 \hat{c}_1, \hat{x}_2 = A_2 \hat{c}_2$.

Motivation: Signal Separation



Sparse Signal Processing

Canonical problem

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{n}$$

Find c via optimization

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \lambda \psi(\boldsymbol{c}) \}$$

Exploiting sparsity is good,

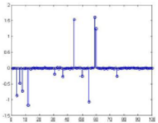
- normally with better performance than traditional method,
- linear measurement model,
- nonlinear, thus hard to solve

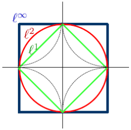
Outline

Outline

Sparsity

A vector $x \in \mathbb{R}^n$ is sparse if only a few entries are nonzero:





The number of nonzeros is called the ℓ^0-norm of x:

$$\|x\|_0 \triangleq \#\{i | x_i \neq 0\}$$

Denote Σ_k the set of all k-sparse signals. And geometrically

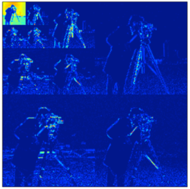
$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p} \Rightarrow \|x\|_0 = \lim_{p \to 0} \|x\|_p^p$$

Sparsity is Universal

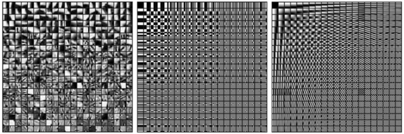
- Signal is itself not sparse at all, then *sparsify it*!!!

$$x = \Psi\alpha, \quad s.t. \quad \alpha \in \Sigma_k$$

  - Fixed dictionaries: Wavelet, DCT, etc.



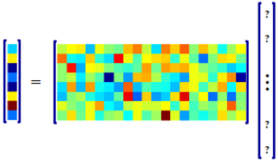
  - Learned dictionaries: K-SVD



The sparse solution

Underdetermined system

$$y = Ax$$



Look for the sparsest x that agrees with our observation:

$$\text{minimize} \quad \|x\|_0 \quad \text{subject to} \quad Ax = y.$$

Theorem 1 (Gorodnitsky+Rao '97) .
Suppose $y = Ax_0$, and let $k = \|x_0\|_0$. If $\mathrm{null}(A)$ contains no $2k$-sparse vectors, x_0 is the unique optimal solution to

$$\text{minimize} \; \|x\|_0 \quad \textit{subject to} \quad y = Ax.$$

The sparse solution

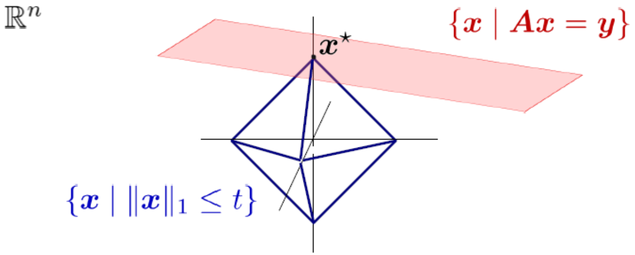
$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

NP-hard, hard to appx. [Natarjan '95], [Amaldi+Kann '97]

⬇⬇

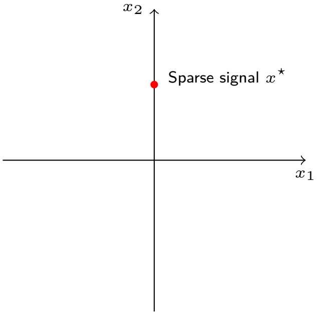
$$\text{minimize} \quad \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

Efficiently solvable



Linear Inverse Problem

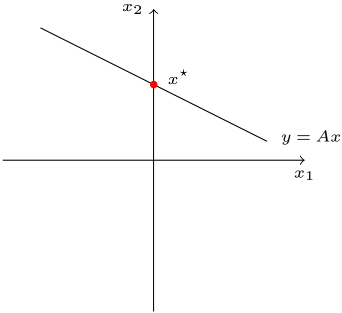
- Measurement fitness ($M \ll N$):
  $\hat{x} = \arg\min \|y - Ax\|_2^2$
  1. Infinite solutions;
  2. Over-fitting.
- ℓ_2 energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_2^2$
  1. Solution is not sparse.



Linear Inverse Problem

Measurement

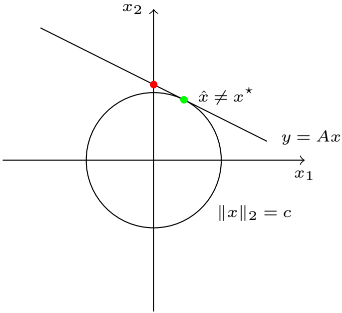
- Measurement fitness ($M \ll N$):
  $\hat{x} = \arg\min \|y - Ax\|_2^2$
  1. Infinite solutions;
  2. Over-fitting.
- ℓ_2 energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_2^2$
  1. Solution is not sparse.



Linear Inverse Problem

Measurement + ℓ_2 Energy

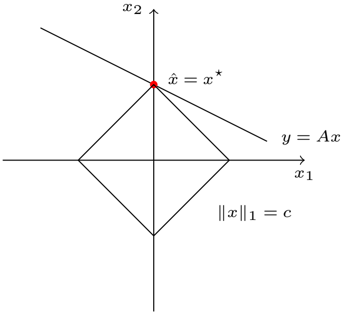
- Measurement fitness $(M \ll N)$:
  $\hat{x} = \arg\min \|y - Ax\|_2^2$
  1. Infinite solutions;
  2. Over-fitting.
- ℓ_2 energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_2^2$
  1. Solution is not sparse.



Linear Inverse Problem

Measurement + ℓ_1 Energy

- ℓ_1 energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_1$
  1. Unique sparse solution;
  2. Noise robustness.
- ℓ_p energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_p$
  1. Sharper, but non-convex.

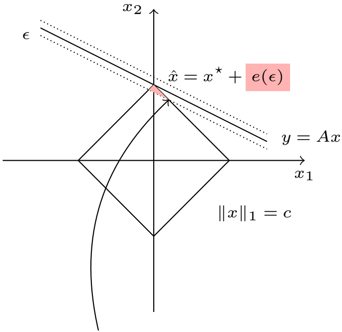


Linear Inverse Problem

Measurement + ℓ_1 Energy

- ℓ_1 energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_1$
  1. Unique sparse solution;
  2. Noise robustness.

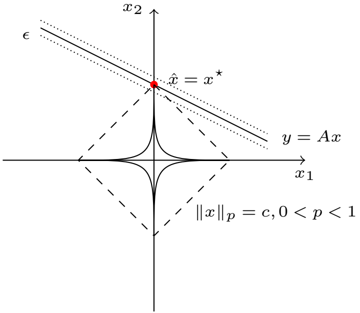
- ℓ_p energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_p$
  1. Sharper, but non-convex.



Linear Inverse Problem

Measurement + ℓ_1 Energy

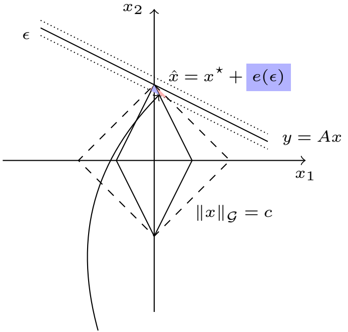
- ℓ_1 energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_1$
  1. Unique sparse solution;
  2. Noise robustness.
- ℓ_p energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_p$
  1. Sharper, but non-convex.



Linear Inverse Problem

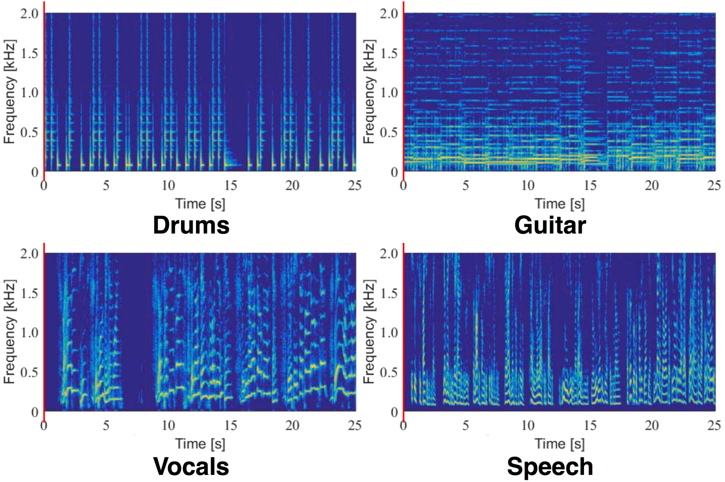
Measurement + ℓ_1 Energy + Structures

- Group energy limited:
  $\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda \|x\|_{\mathcal{G}}$
  where
  $$\|x\|_{\mathcal{G}} = \sum_{G \in \mathcal{G}} \left\{ \sum_{j \in G} d_j \cdot x_j^2 \right\}^{\frac{1}{2}}$$ with $\mathcal{G}$ the set of groups, d_j the weight.
  1. Sharper, still convex.

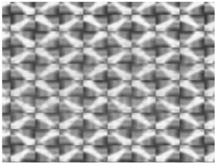


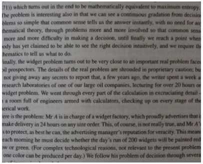
Outline

Low-Rank Model

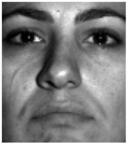


Low-Rank Model





71)) which turns out in the end to be mathematically equivalent to maximum entropy.
The problem is interesting also in that we can see a continuous gradation from decisio
blems so simple that common sense tells us the answer instantly, with no need for an
thematical theory, through problems more and more involved so that common sens
more and more difficulty in making a decision, until finally we reach a point when
ody has yet claimed to be able to see the right decision intuitively, and we require th
hematics to tell us what to do.
inally, the widget problem turns out to be very close to an important real problem face
il prospectors. The details of the real problem are shrouded in proprietary caution; bu
not giving away any secrets to report that, a few years ago, the writer spent a week a
research laboratories of one of our large oil companies, lecturing for over 20 hours o
widget problem. We went through every part of the calculation in excruciating detail
a room full of engineers armed with calculators, checking up on every stage of th
erical work.
ere is the problem: Mr A is in charge of a widget factory, which proudly advertises that i
make delivery in 24 hours on any size order. This, of course, is not really true, and Mr A'
s to protect, as best he can, the advertising manager's reputation for veracity. This mean
each morning he must decide whether the day's run of 200 widgets will be painted red
w or green. (For complex technological reasons, not relevant to the present problem
one color can be produced per day.) We follow his problem of decision through severa



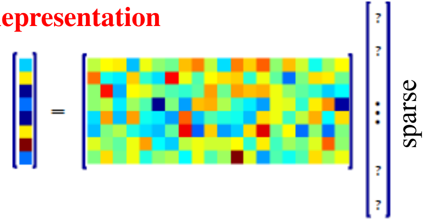
Visual data exhibit ***low-dimensional structures*** due to rich ***local*** regularities, ***global*** symmetries, ***repetitive*** patterns, or ***redundant*** sampling.

Low-Rank Model

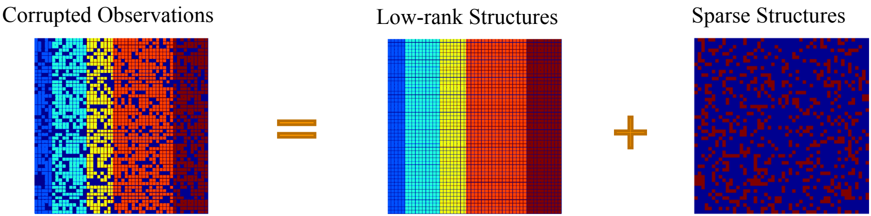
Sparse Representation

Underdetermined system

$$y = Ax$$



Robust PCA



Low-Rank Model

Low-rank model:

$$Y = X + E + N$$

Robust PCA (Non-tractable):

$$\min_{X,E} \frac{1}{\lambda} \|Y - X - E\|_F^2 + rank[X] + \frac{1}{n} \|E\|_0$$

Convex Relaxation:

$$\min_{X,E} \frac{1}{\lambda} \|Y - X - E\|_F^2 + \|X\|_* + \frac{1}{\sqrt{n}} \|E\|_1$$

Outline

Classical Sampling Formulation

The Shannon sampling theorem provides sufficient but not necessary conditions for perfect reconstruction.

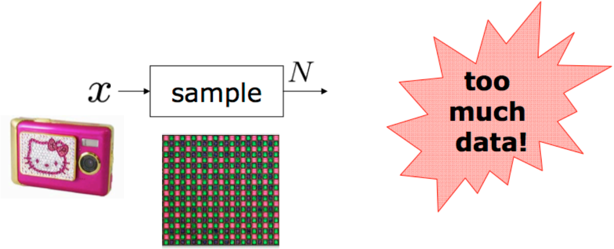
Moreover: How many real signals are bandlimited? How many realizable filters are ideal low-pass filters?

By the way, who discovered the sampling theorem? The list is long ;-)

- Whittaker 1915, 1935
- Kotelnikov 1933
- Nyquist 1928
- Raabe 1938
- Gabor 1946
- Shannon 1948
- Someya 1948

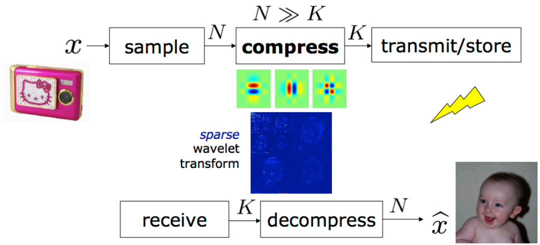
Recall of Sampling Theory

- Shannon's sampling theory: uniformly sample data at Nyquist rate (2 times of Fourier bandwidth)



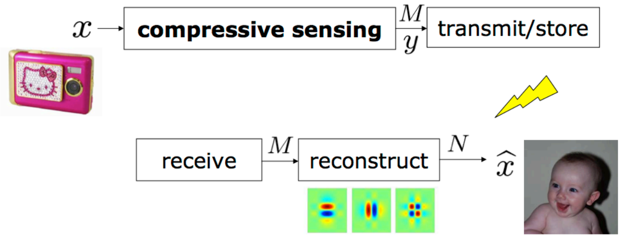
Procedure of sampling

- Traditional sampling theorem admits the following procedure:
  - Uniformly sample data at Nyquist rate
  - compress data
  - transmit and receive
  - decompress data



Compressive Sensing (CS)

- CS directly acquire “compressed” data

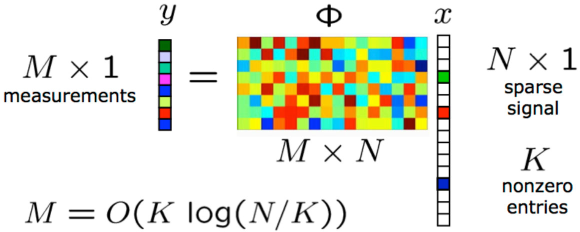


Compressive data acquisition

- When data is sparse, CS can directly acquire a compressed measurement with no information loss

$$y = Ax$$

- Random projection will work



CS v.s. Shannon's thoery

- Signal model
  - In CS, signals are sparse;
  - In Shannon's theory, signals are Fourier bandlimited
- Sampling procedure
  - In CS, acquire information via random projection
  - In Shannon's theory, acquire data via uniform sampling
- Recovery method
  - In CS, recover signal via nonlinear algorithm;
  - In Shannon's theory, recover signal via linear interpolation.

Compressive Sensing

Given a signal $x \in \mathbb{R}^n$, CS measurements are obtained by linear projection

$$y = Ax$$

with $A \in \mathbb{R}^{m \times n}$ the sensing matrix and $y \in \mathbb{R}^m$ the captured measurements.

Underdetermined

Notice that $m \ll n$, leading to a underdetermined linear system.

Questions:

1. ***How should we design the sensing matrix A?***
   $\rightarrow$ to preserve information
2. How can we recover the original signal x?
   $\rightarrow$ to recover information (Sparse Representation/Sparse Recovery)

Compressive Sensing

Given a signal $x \in \mathbb{R}^n$, CS measurements are obtained by linear projection

$$y = Ax$$

with $A \in \mathbb{R}^{m \times n}$ the sensing matrix and $y \in \mathbb{R}^m$ the captured measurements.

Underdetermined

Notice that $m \ll n$, leading to a underdetermined linear system.

Questions:

1. ***How should we design the sensing matrix A?***
   $\rightarrow$ to preserve information
2. How can we recover the original signal x?
   $\rightarrow$ to recover information (Sparse Representation/Sparse Recovery)

Compressive Sensing

Given a signal $x \in \mathbb{R}^n$, CS measurements are obtained by linear projection

$$y = Ax$$

with $A \in \mathbb{R}^{m \times n}$ the sensing matrix and $y \in \mathbb{R}^m$ the captured measurements.

Underdetermined

Notice that $m \ll n$, leading to a underdetermined linear system.

Questions:

1. ***How should we design the sensing matrix A?***
   $\rightarrow$ to preserve information
2. How can we recover the original signal x?
   $\rightarrow$ to recover information (Sparse Representation/Sparse Recovery)

Properties of Sensing Matrix

1. Spark
2. Null Space Property (NSP)
3. Restricted isometry Property (RIP)
4. Coherence

Spark

定义

The spark of a given matrix A is the smallest number of columns of A that are linearly dependent.

例

1. $A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$, *spark(A)=?*
2. $\forall A \in \mathbb{R}^{m \times n}$ with $m < n$, what is the maximum spark of A?

$$\forall A \in \mathbb{R}^{m \times n}, spark(A) \in [2, m+1]$$

Null space property

- **Null space**

$$\mathcal{N}(A) = \{z : Az = 0\}$$

- **Null space property (NSP)** A matrix A satisfies the NSP of order k if there exists a constant $C > 0$ such that,

$$\|h_\Lambda\|_2 \leq C \frac{\|h_{\Lambda^c}\|_1}{\sqrt{k}}$$

holds for all $h \in \mathcal{N}(A)$ and for all Λ with $|\Lambda| \leq k$.

Restricted isometry property

定义

A matrix A satisfies the Restricted Isometry Property (RIP) of order k if there exists a $\delta_k \in (0,1)$, such that (for all $x \in \Sigma_k$)

$$(1-\delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\|x\|_2^2$$

Links to singular values

$x \in \Sigma_k$, denote $\Gamma = supp\{x\}$, A_Γ the submatrix of A, the twoside inequalities is equivalent to

$$1-\delta_k \leq \frac{\|A_\Gamma x_\Gamma\|_2^2}{\|x_\Gamma\|_2^2} \leq 1+\delta_k$$

Note that $\frac{\|A_\Gamma x_\Gamma\|_2^2}{\|x_\Gamma\|_2^2}$ is bounded in $\left[\lambda_{\min}(A_\Gamma^T A_\Gamma), \lambda_{\max}(A_\Gamma^T A_\Gamma)\right]$.

Coherence

定义

The coherence of a matrix A, $\mu(A)$, is the largest absolute inner product between any two columns a_i, a_j of A

$$\mu(A) = \max_{1 \le i < j \le n} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}$$

Bounds of Coherence

$$\sqrt{\frac{n-m}{m(n-1)}} \le \mu(A) \le 1$$

if $m \ll n$, the lower bound is approximately $\frac{1}{\sqrt{m}}$.

Links between these properties

- **RIP $\Rightarrow$ NSP:** If $\delta_{2k} < \sqrt{2} - 1$, then A satisfies NSP of order $2k$, with constant
$$C = \frac{\sqrt{2}\delta_{2k}}{1 - (1 + \sqrt{2})\delta_{2k}}$$
- **Coherence $\Rightarrow$ RIP**
$$\delta_k = (k-1)\mu(A)$$
with $k < 1/\mu$.
- **Spark v.s. coherence**
$$spark(A) \geq 1 + \frac{1}{\mu(A)}$$

Information preserving

- **For sparse signals**, the CS measurements

$$y = Ax \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ with $m \ll n$, $x \in \Sigma_k$, and $y \in \mathbb{R}^m$.

- Information preserving $\Leftrightarrow$ uniqueness of solution.

Uniqueness of solution

定理

For any vector $y \in \mathbb{R}^m$, there exists at most one signal $x \in \Sigma_k$, such that $y = Ax$ if and only if $spark(A) > 2k$.

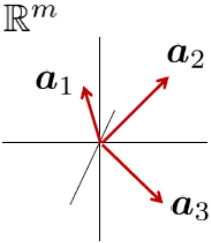
证明.

1. Necessity: Suppose $spark(A) \leq 2k$
   $\Rightarrow$ there exists $2k$ coloumns of A that are dependent
   $\Rightarrow \exists h \in \Sigma_{2k}$, s.t. $h \in \mathcal{N}(A) \Rightarrow \exists x_1, x_2 \in \Sigma_k$, s.t. $h = x_1 - x_2$
   $\Rightarrow Ax_1 = Ax_2$ (*Contradiction*)
2. Sufficiency: Suppose $\exists x_1, x_2 \in \Sigma_k$, s.t. $y = Ax_1 = Ax_2$
   $\Rightarrow h = x_1 - x_2 \in \Sigma_{2k}$, i.e. $Ah = 0$
   $\Rightarrow h = 0$, i.e. $x_1 = x_2$ (since $spark(A) > 2k$)

Intuition of Information Preserving

Suppose:

$$y = Ax = \sum_{i \in \mathsf{supp}(x)} a_i x_i$$



Intuition: Recovering x is "easier" if the a_i are not too similar ...
This is exactly the definition of coherence: (smaller the better)

$$\mu(A) = \max_{i \neq j} |\langle a_i, a_j \rangle|$$

Uniqueness of solution (other conditions)

Considering the sparse signals Σ_k, the uniqueness of solution

$$\forall x_1, x_2 \in \Sigma_k, x_1 \neq x_2 \Leftrightarrow Ax_1 \neq Ax_2$$

- Spark guarantee

$$spark(A) > 2k$$

- NSP guarantee

$$A \text{ satisfies NSP of order } 2k$$

- RIP guarantee

$$\delta_{2k} < 1$$

- Coherence guarantee

$$\mu(A) < \frac{1}{2k-1}$$

Limitations of Coherence

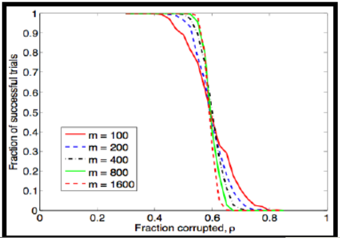
- For any $m \times n$ matrix A, its coherence

$$\mu(A) \geq \sqrt{\frac{n-m}{m(n-1)}}$$

- Thus sparsity level should satisfy

$$k < \frac{1}{2}(1+\mu(A)^{-1}) = O(\sqrt{m})$$

- Truth is often much better (Phase transition at $k = \alpha^* m$)



Necessary measurement number with RIP

定理 (Candes 2005, 2008)

Suppose $y = Ax_0$ *with RIP constant* $\delta_{2k} < \sqrt{2} - 1$, *then* x_0 *is the unique optimal solution to*

$$\min \|x\|_1, \quad s.t. y = Ax$$

定理

Let $A \in \mathbb{R}^{m \times n}$ *satisfies RIP of order* $2k$, *with constant* $\delta \in (0, \frac{1}{2}]$. *Then*

$$m \geq Ck \log\left(\frac{n}{k}\right)$$

where $C = 0.5\log(\sqrt{24}+1) \approx 0.28$.

$k \sim m$ when considering RIP of matrix A

Constructing sensing matrix

- Deterministic method
- Random method
  - For any random matrix,

    $$spark(A) = m + 1$$

    with probability 1.
  - For sub-Gaussian, if

    $$m = O\left(k/\delta_{2k}^2 \log\left(\frac{n}{k}\right)\right)$$

    then RIP of order $2k$ is fulfilled with probability at least $1 - 2\exp(-c_1\delta_{2k}^2 m)$.
  - For any zero-mean and finite variance distribution, it has

    $$\mu(A) = \sqrt{(2\log n)/m}$$

CS with Chaotic Sequence

- Logistic map

$$z_{n+1} = r z_n (1 - z_n)$$

- Constructing chaotic matrix

$$A = \sqrt{\frac{2}{m}} \begin{pmatrix} x_0 & \cdots & x_{m(n-1)} \\ x_1 & \cdots & x_{m(n-1)+1} \\ \vdots & \vdots & \vdots \\ x_{m-1} & \cdots & x_{mn-1} \end{pmatrix} \quad (2)$$

where $x_k = 1 - 2z_{n+kd}$ with z_{n+kd} the coefficient selected from generated chaotic set $Z(d, k, z_0) = \{z_n, z_{n+d}, \ldots, z_{n+kd}, \ldots\}$.

[1] L. Yu, etc., "Compressive Sensing With Chaotic Sequence," IEEE SPL, 2010.

CS with Chaotic Sequence

- Logistic map

$$z_{n+1} = r z_n (1 - z_n)$$

- Constructing chaotic matrix

$$A = \sqrt{\frac{2}{m}} \begin{pmatrix} x_0 & \cdots & x_{m(n-1)} \\ x_1 & \cdots & x_{m(n-1)+1} \\ \vdots & \vdots & \vdots \\ x_{m-1} & \cdots & x_{mn-1} \end{pmatrix} \tag{2}$$

where $x_k = 1 - 2z_{n+kd}$ with z_{n+kd} the coefficient selected from generated chaotic set $Z(d, k, z_0) = \{z_n, z_{n+d}, \ldots, z_{n+kd}, \ldots\}$.

[1] L. Yu, etc., "Compressive Sensing With Chaotic Sequence," IEEE SPL, 2010.

Statistical Independence

定理

Denote $Z = \{z_n, z_{n+1}, \cdots, z_{n+r}, \cdots\}$ *the sequence generated by Logistic map with intial state* $z_0 = \cos(2\pi x)$, *and integer* d *the sampling distance, then for any positive integer* $m_0, m_1 < 2^d$, *it has*

$$E(z_n^{m_0} z_{n+d}^{m_1}) = E(z_n^{m_0}) E(z_{n+d}^{m_1})$$

定理

Chaotic matrix $A \in \mathbb{R}^{m \times n}$ *constructed as* (2) *satisfies RIP of order* k *for constant* $\delta \in (0,1)$, *with overwhelming probability, providing that* $m \geq O(k \log(n/k))$.

Statistical Independence

定理

Denote $Z = \{z_n, z_{n+1}, \cdots, z_{n+r}, \cdots\}$ *the sequence generated by Logistic map with intial state* $z_0 = \cos(2\pi x)$, *and integer* d *the sampling distance, then for any positive integer* $m_0, m_1 < 2^d$, *it has*

$$E(z_n^{m_0} z_{n+d}^{m_1}) = E(z_n^{m_0}) E(z_{n+d}^{m_1})$$

定理

Chaotic matrix $A \in \mathbb{R}^{m \times n}$ *constructed as* (2) *satisfies RIP of order* k *for constant* $\delta \in (0,1)$, *with overwhelming probability, providing that* $m \geq O(k \log(n/k))$.

Performance: successful recovery rate

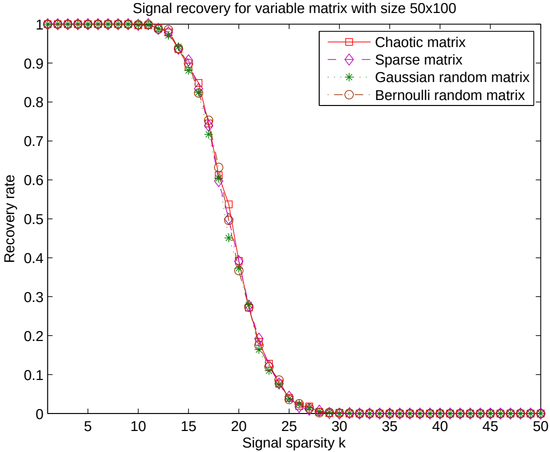


图: Signal recovery for different sensing matrix with size 50×100.

Outline

Sparse Representation/Sparse Recovery

1. Basis pursuit
2. Basis pursuit denoising
3. Matching pursuit
4. etc.
5. Bayesian approach
6. Deep learning
7. Analog approach

Basis pursuit

Basis pursuit (BP) problem:

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \|\boldsymbol{c}\|_1, \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{c}$$

- convex problem
- noise-free

Basis pursuit denoising

Basis pursuit denoising (BPD) problem

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \left\{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{c}\|_1 \right\}$$

- convex problem
- noisy, λ is a parameter balancing measurement fidelity and sparse prior.

Matching pursuit

Matching pursuit problem (approximately)

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2, \quad s.t. \quad \|\boldsymbol{c}\|_0 \leq K$$

- efficient
- approximately solve the L_0 problem.
- variations: orthogonal matching pursuit, CoSaMP, ...

Bayesian Interpretation

$$MAP: p(x|y) = \frac{p(y|x) \quad p(x)}{\int p(y|x)p(x)dx}$$

- Measurement Likelihood: Gaussian noise model

$$y - Ax \sim \mathcal{N}(0, \sigma_0)$$

- Energy Prior: sparse promoting model

$$\text{e.g.} \quad x \sim \text{Laplace}(0, b)$$

- How to introduce structures? ←-- Hierarchical Bayesian Model

Bayesian Interpretation

Measurement

$$MAP : p(x|y) = \frac{\boxed{p(y|x)}\, p(x)}{\int p(y|x)p(x)dx}$$

- Measurement Likelihood: Gaussian noise model

$$y - Ax \sim \mathcal{N}(0, \sigma_0)$$

- Energy Prior: sparse promoting model

$$\text{e.g.} \quad x \sim \text{Laplace}(0, b)$$

- How to introduce structures? ←-- Hierarchical Bayesian Model

Bayesian Interpretation

Measurement + ℓ_1 Energy

$$MAP : p(x|y) = \frac{p(y|x) \; p(x)}{\int p(y|x)p(x)dx}$$

- Measurement Likelihood: Gaussian noise model

$$y - Ax \sim \mathcal{N}(0, \sigma_0)$$

- Energy Prior: sparse promoting model

$$\text{e.g.} \quad x \sim \text{Laplace}(0, b)$$

- How to introduce structures? ←-- Hierarchical Bayesian Model

Bayesian Interpretation

Measurement + ℓ_1 Energy + Structures?

$$MAP: p(x|y) = \frac{p(y|x)\, p(x)}{\int p(y|x)p(x)dx}$$

- Measurement Likelihood: Gaussian noise model

$$y - Ax \sim \mathcal{N}(0, \sigma_0)$$

- Energy Prior: sparse promoting model

$$\text{e.g.} \quad x \sim \text{Laplace}(0, b)$$

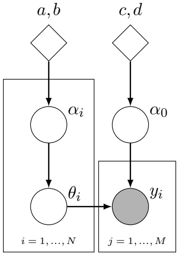
- How to introduce structures? ←-- Hierarchical Bayesian Model

Hierarchical Bayesian Model for CS

- Gamma-Gaussian Model $\rightsquigarrow$ Sparsity:

$$x \sim \mathcal{N}(0, \alpha^{-1})$$
$$\alpha_i \sim \Gamma(a, b)$$

- $x_i \sim \frac{1}{|x_i|}$, as $(a, b) \to (0, 0)$



Gamma-Gaussian Model for CS

- Noise tolerance, non-parametric;
- but no structure prior.

The 1st Proposed Model: CluSS-MCMC

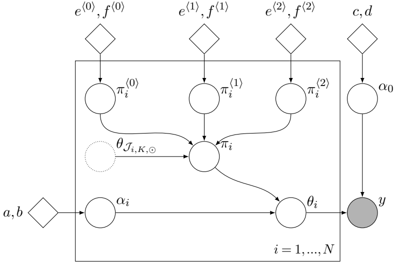
- Spike-and-slab model:

$$x_i \sim (1-\pi_i)\delta_0 + \pi_i \mathcal{N}(0, \alpha_i^{-1})$$

- Pattern selection:

$$\pi_i = \begin{cases} \pi_i^{\langle 0 \rangle}, & \text{if Pattern (a)} \\ \pi_i^{\langle 1 \rangle}, & \text{if Pattern (b)} \\ \pi_i^{\langle 2 \rangle}, & \text{if Pattern (c)} \end{cases}$$

1. Promote clusters, while eliminate isolates.



The 1st Proposed Model: CluSS-MCMC

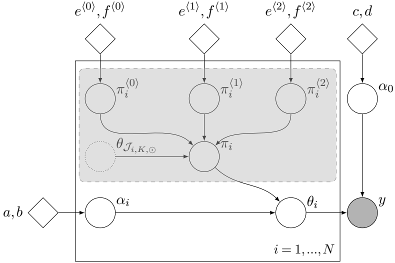
- Spike-and-slab model:

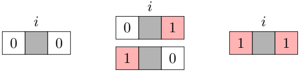
$$x_i \sim (1-\pi_i)\delta_0 + \pi_i \mathcal{N}(0, \alpha_i^{-1})$$

- Pattern selection:

$$\pi_i = \begin{cases} \pi_i^{\langle 0 \rangle}, & \text{if Pattern (a)} \\ \pi_i^{\langle 1 \rangle}, & \text{if Pattern (b)} \\ \pi_i^{\langle 2 \rangle}, & \text{if Pattern (c)} \end{cases}$$

1. Promote clusters, while eliminate isolates.





(a) (b) (c)

The 1st Proposed Model: CluSS-MCMC

It does introduce the cluster structure, while ...

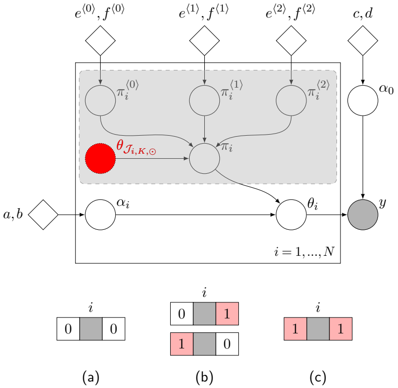
How to determine $x_i = 0$?

Threshold procedure

$|x_i| < t$: Ambiguous to determine t;

No explicit estimators

MCMC technique is exploited: slow.



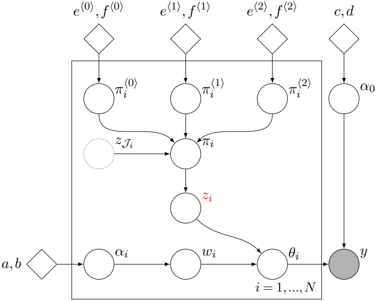
The 2nd Proposed Model: CluSS-VB

- Latent model:

$$x = w \circ z$$
$$w \sim \mathcal{N}(0, \alpha^{-1})$$
$$z_i \sim Bernoulli(\pi_i)$$

- Pattern selection:

$$\pi_i = \begin{cases} \pi_i^{\langle 0 \rangle}, & \text{if Pattern (a)} \\ \pi_i^{\langle 1 \rangle}, & \text{if Pattern (b)} \\ \pi_i^{\langle 2 \rangle}, & \text{if Pattern (c)} \end{cases}$$



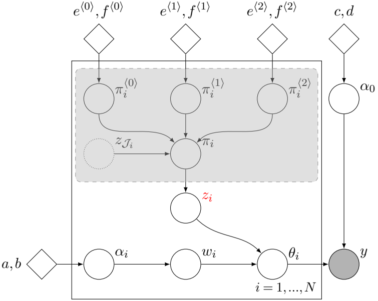
The 2nd Proposed Model: CluSS-VB

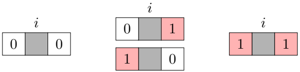
- Latent model:

$$x = w \circ z$$
$$w \sim \mathcal{N}(0, \alpha^{-1})$$
$$z_i \sim Bernoulli(\pi_i)$$

- Pattern selection:

$$\pi_i = \begin{cases} \pi_i^{\langle 0 \rangle}, & \text{if Pattern (a)} \\ \pi_i^{\langle 1 \rangle}, & \text{if Pattern (b)} \\ \pi_i^{\langle 2 \rangle}, & \text{if Pattern (c)} \end{cases}$$





(a) (b) (c)

The 2nd Proposed Model: CluSS-VB

It is faster and more robust, while not "elegant" ...

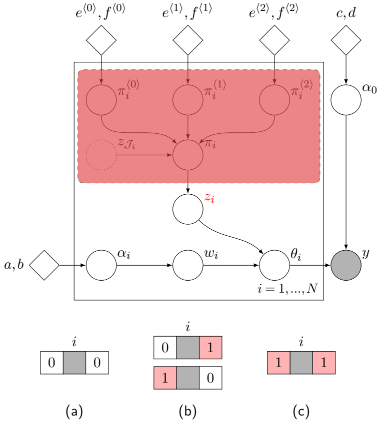
Pattern selection procedure:

Deterministic, hard decision, lots of parameters;

What's next?

"Pure" statistical model:

Statistical, soft decision, thus more robust.



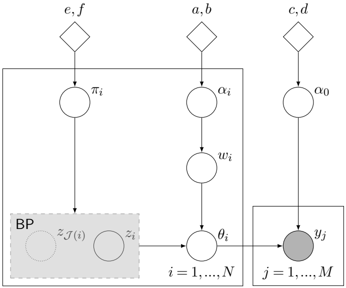
The 3rd Proposed Model: MBCS-LBP

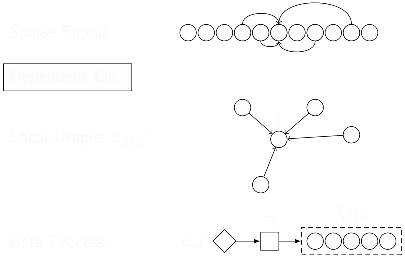
- Latent model:

$$x = w \circ z$$
$$w \sim \mathcal{N}(0, \alpha^{-1})$$

- Local beta process:

$$z_j \sim Bernoulli(\pi_i), \forall z_j \in z_{\mathcal{J}(i)}$$
$$\pi_i \sim Beta(e, f)$$





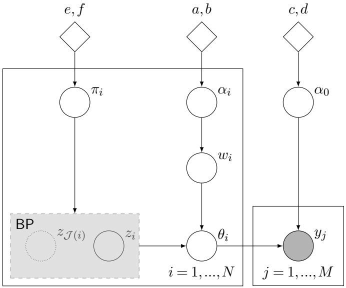
The 3rd Proposed Model: MBCS-LBP

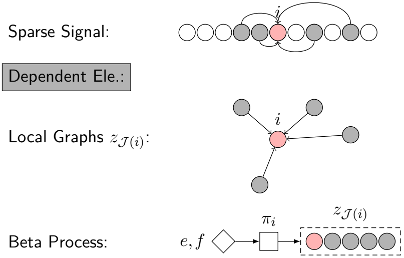
- Latent model:

$$x = w \circ z$$
$$w \sim \mathcal{N}(0, \alpha^{-1})$$

- Local beta process:

$$z_j \sim Bernoulli(\pi_i), \forall z_j \in z_{\mathcal{J}(i)}$$
$$\pi_i \sim Beta(e, f)$$





Experiments: Setting up

Default settings

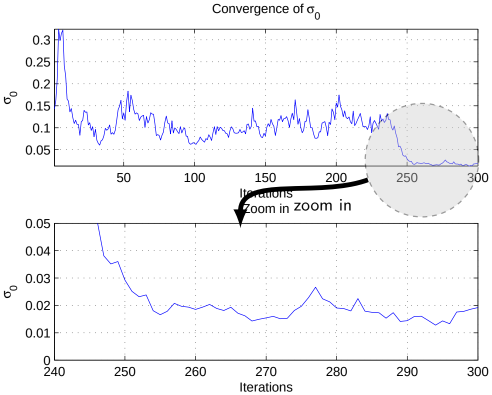
- Gaussian random sensing matrix A;
- Clustered ± 1(or Gaussian) spikes;
- Noise level $\sigma_0 = 0.02$.

Benchmark

- Basis Pursuit (BP);
- CoSaMP, Block CoSaMP;
- Bayesian CS.

Experiments: Convergence

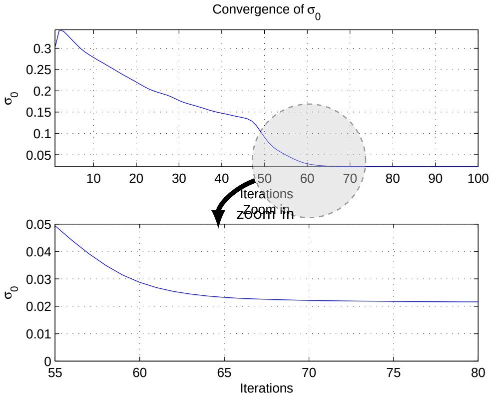
Evolution of variable of noise invariance $\sigma_0 = 0.02$:



CluSS-MCMC slower than CluSS-VB slower than MBCS-LBP

Experiments: Convergence

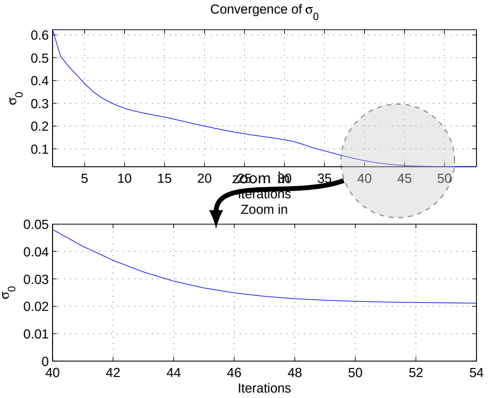
Evolution of variable of noise invariance $\sigma_0 = 0.02$:



CluSS-MCMC slower than CluSS-VB slower than MBCS-LBP

Experiments: Convergence

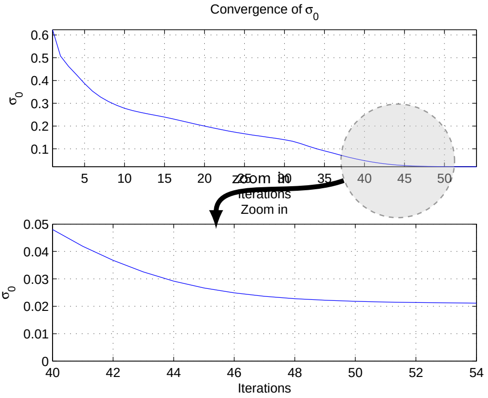
Evolution of variable of noise invariance $\sigma_0 = 0.02$:



CluSS-MCMC slower than CluSS-VB slower than MBCS-LBP

Experiments: Convergence

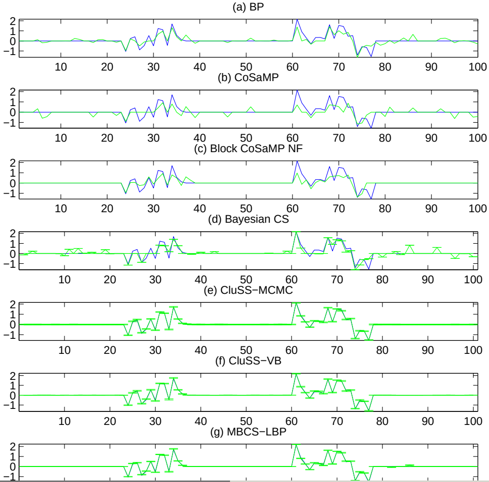
Evolution of variable of noise invariance $\sigma_0 = 0.02$:



CluSS-MCMC slower than CluSS-VB slower than MBCS-LBP

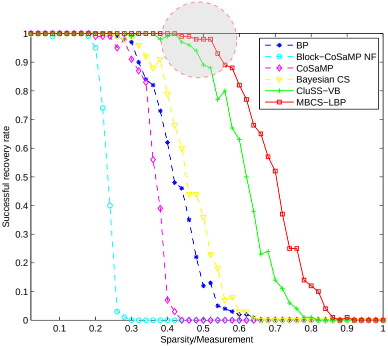
Experiments: General Comparison

- Clustered Gaussian Spikes:
  - Signal size $N = 100$;
  - Sparsity $s = 30$;
  - Clusters $C = 2$.
- Measurements $M = 50$.



Experiments: Successful Reconstruction Rate

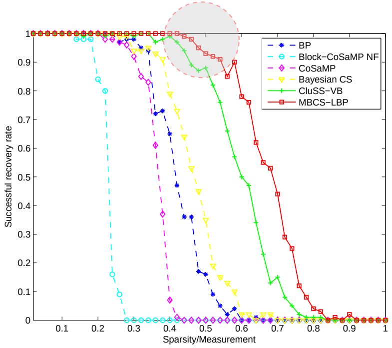
Successful reconstruction rate with clustered Gaussian spikes:



Number of Clusters: $C = 1$ better than $C = 2$ better than $C = 4$

Experiments: Successful Reconstruction Rate

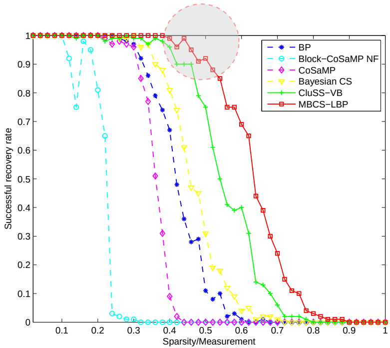
Successful reconstruction rate with clustered Gaussian spikes:



Number of Clusters: $C = 1$ better than $C = 2$ better than $C = 4$

Experiments: Successful Reconstruction Rate

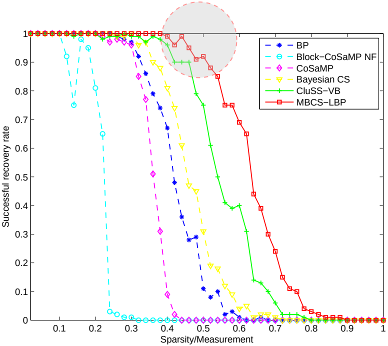
Successful reconstruction rate with clustered Gaussian spikes:



Number of Clusters: $C = 1$ better than $C = 2$ better than $C = 4$

Experiments: Successful Reconstruction Rate

Successful reconstruction rate with clustered Gaussian spikes:



Number of Clusters: $C = 1$ better than $C = 2$ better than $C = 4$

Experiments: Successful Reconstruction Rate

Successful reconstruction rate with clustered ± 1 spikes via CluSS-MCMC:

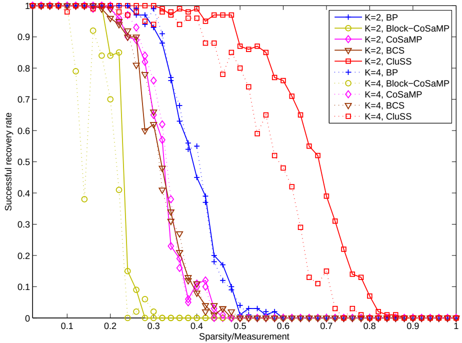
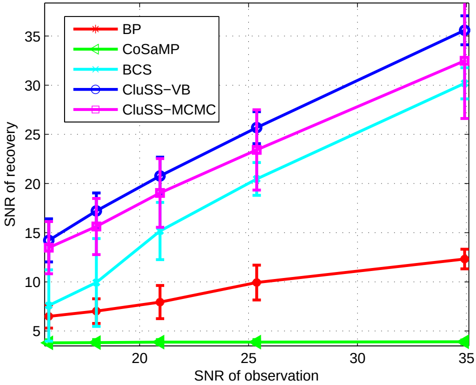


图: $K = 2$ means $C = 1$, while $K = 4$ means $C = 2$.

Experiments: Robust to Noise

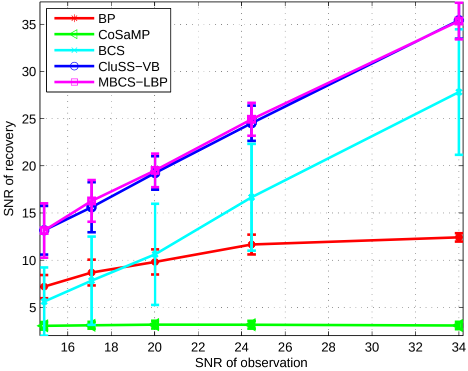
Ranging noise level $\sigma_0 = 0.01 \rightsquigarrow 0.09$:



CluSS-MCMC v.s. CluSS-VB v.s. MBCS-LBP

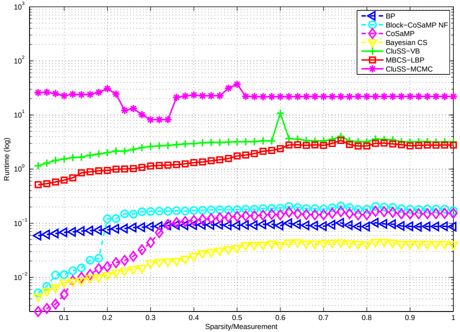
Experiments: Robust to Noise

Ranging noise level $\sigma_0 = 0.01 \rightsquigarrow 0.09$:



CluSS-MCMC v.s. CluSS-VB v.s. MBCS-LBP

Experiments: Complexity



Drawback

The proposed algorithms are much slower than the benchmark algorithms.

Experiments: 2D Images

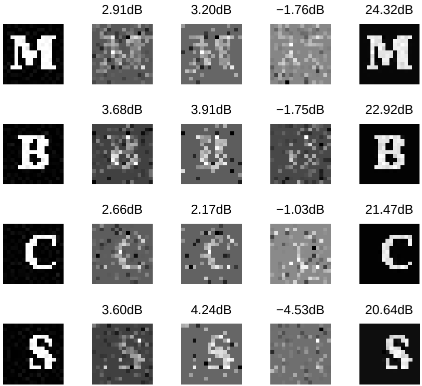


图: (1st col.) original signals, reconstructions via (2nd col.) BP, (3rd col.) CoSaMP, (4th col.) BCS and (5th col.) MBCS-LBP.

Dynamical Sparse Recovery with Finite-time Convergence

Recover Sparse Signals with Cucuits: A new dynamical system is constructed by introducing the parameter $\alpha \in (0,1]$, i.e.,

$$\begin{cases} \tau \dot{u}(t) = -\lceil u(t) + (\Phi^T\Phi - I)a(t) - \Phi^T y \rfloor^{\alpha} \\ \hat{x}(t) = a(t) \end{cases} \tag{3}$$

with $\lceil \cdot \rfloor^{\alpha}$ being a function defined as $\lceil \cdot \rfloor^{\alpha} = |\cdot|^{\alpha} \cdot \mathrm{sgn}(\cdot)$ where $|\cdot|, \cdot, \mathrm{sgn}$ are all element-wise operators, $\alpha \in \mathbb{R}_+$ denotes an exponential coefficient and

$$\mathrm{sgn}(\omega) \begin{cases} = 1, & \text{if } \omega > 0 \\ \in [-1,1], & \text{if } \omega = 0 \\ = -1, & \text{if } \omega < 0 \end{cases} .$$

L. Yu, G. Zheng and J. P. Barbot, "Dynamical Sparse Recovery With Finite-Time Convergence," in IEEE Transactions on Signal Processing, vol. 65, no. 23, pp. 6146-6157, Dec.1, 1 2017.

Dynamical Sparse Recovery

Optimization of sparse representation problem:

$$x^* = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda \psi(x) \tag{4}$$

and typically, the sparsity-inducing term $\psi(x) = \|x\|_1 \triangleq \sum_i |x_i|$ and $\lambda > 0$ is the balancing parameter.

Theorem

If sensing matrix satisfies RIP, the state $u(t)$ of (3) converges in finite time to its equilibrium point u^*, and $\hat{x}$(t) of (3) *converges in finite-time to x of* (4).

Dynamical Sparse Recovery

- Convergence Speed:

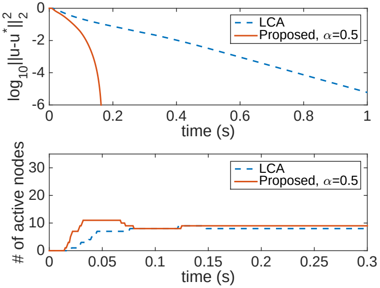


图: Evolutions of state error $\tilde{u}(t)$ and the number of active nodes with respect to time.

Dynamical Sparse Recovery

- Tracking ability:

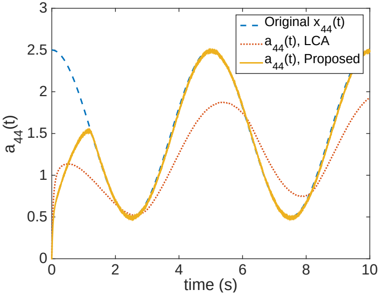
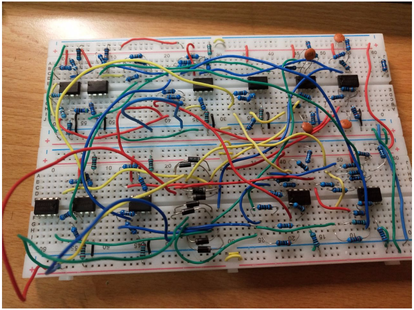


图: Estimation of time-varying sparse signals via LCA and the proposed system.

Dynamical Sparse Recovery



Working with Jiang Yulun.

Outline

Learning to Recover Sparse Signals

The canonical form of compressive sensing

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

- $\mathbf{y} \in \mathcal{R}^M$ is the measurement vector
- $\mathbf{A} \in \mathcal{R}^{M \times N}$ is a random sensing matrix with $M \ll N$ satisfying the so-called RIP
- $\mathbf{x} \in \mathcal{R}^N$ is the original sparse signal needed to be recovered with no more than K $(K < M)$ nonzero elements
- $\mathbf{e}$ is the error term consists of the possible noise and perturbations

Orthogonal Matching Pursuit

Algorithm 1 Orthogonal Matching Pursuit

Input: the sensing matrix $\mathbf{A} \in \mathcal{R}^{M \times N}$, the measurement $\mathbf{y} \in \mathcal{R}^{M}$, sparsity K

Output: the recovered sparse signal $\mathbf{x} \in \mathcal{R}^{N}$

1: Initialize $\mathbf{r}_0 = \mathbf{y}$, $\mathbf{\Lambda}_0 = \varnothing$, $\mathbf{A}_0 = \varnothing$, $t = 1$

2: $\boldsymbol{\lambda}_t = \arg \max_{j=1,2,\ldots,N} |< \mathbf{r}_{t-1}, \mathbf{a}_j >|$

3: $\mathbf{\Lambda}_t = \mathbf{\Lambda}_{t-1} \cup \boldsymbol{\lambda}_t$, $\mathbf{A}_t = \mathbf{A}_{t-1} \cup \mathbf{a}_{\boldsymbol{\lambda}_t}$

4: $\hat{\mathbf{x}}_t = \arg \min_{\mathbf{x}_t} \|\mathbf{y} - \mathbf{A}\mathbf{x}_t\|$

5: $\mathbf{r}_t = \mathbf{y} - \mathbf{A}_t \hat{\mathbf{x}}_t$

6: $t = t + 1$, if $t < K$ continue to 2, else goto 7

7: Output $\hat{\mathbf{x}}_t$

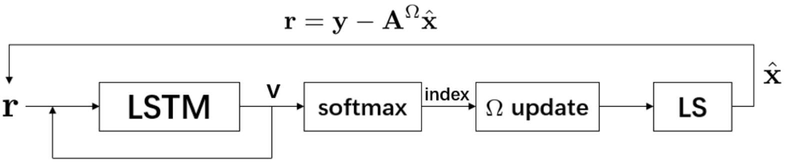
Structure Information

In reality, besides the sparsity property, the elements of sparse signals usually follow a certain structure which could be utilized to improve the recovery performance.

- block-sparse: block-based CoSaMP, block-sparse Bayesian learning
- tree-structure: TSW-CS
- uniform-sparse

Our Algorithm

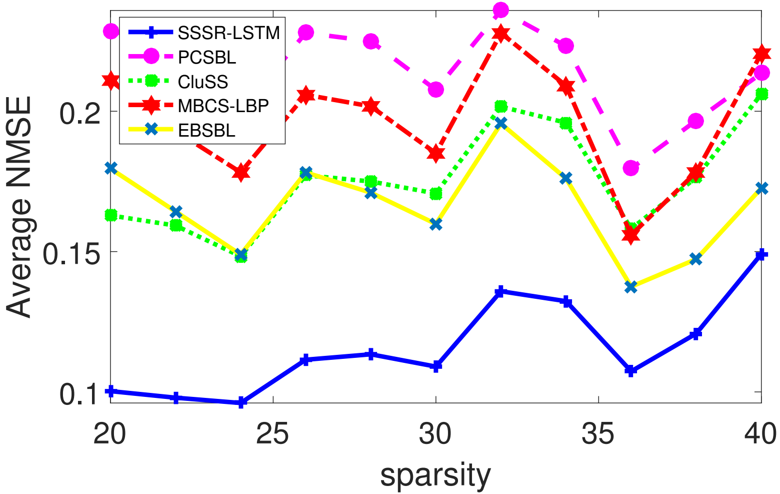
Instead of searching indices of nonzero elements by solving a maximization problem in OMP algorithm, we replace this step with learning approaches.



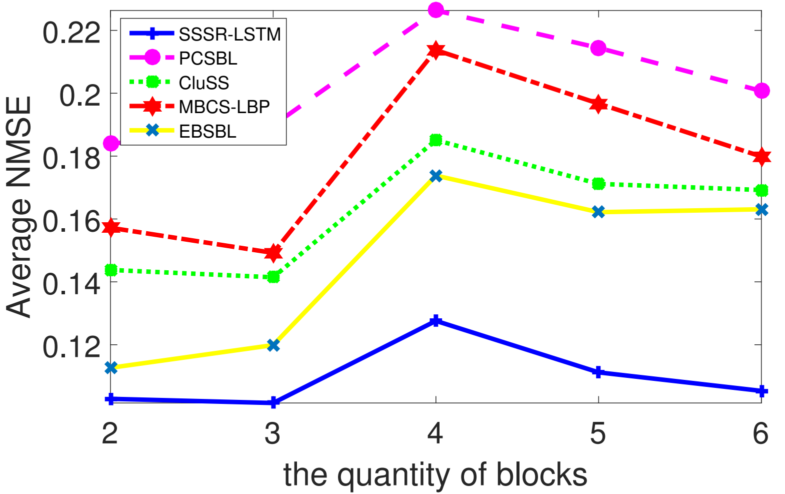
Flow diagram of the proposed algorithm

Working with Lv Chengcheng.

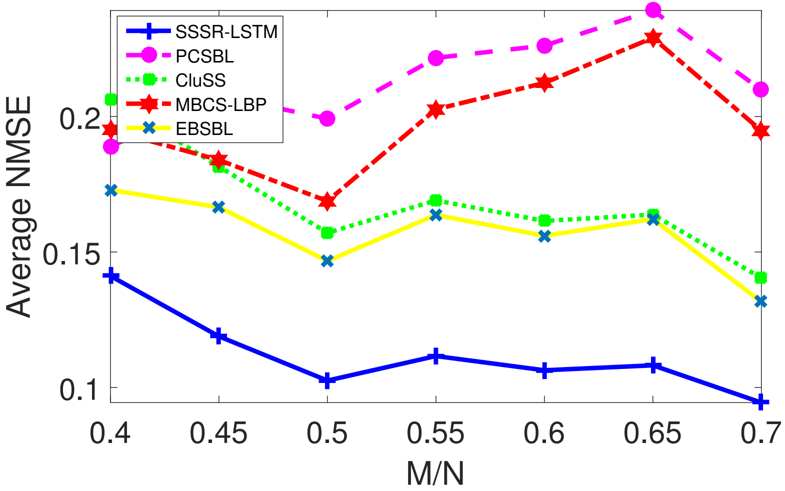
Block-Sparse Signals



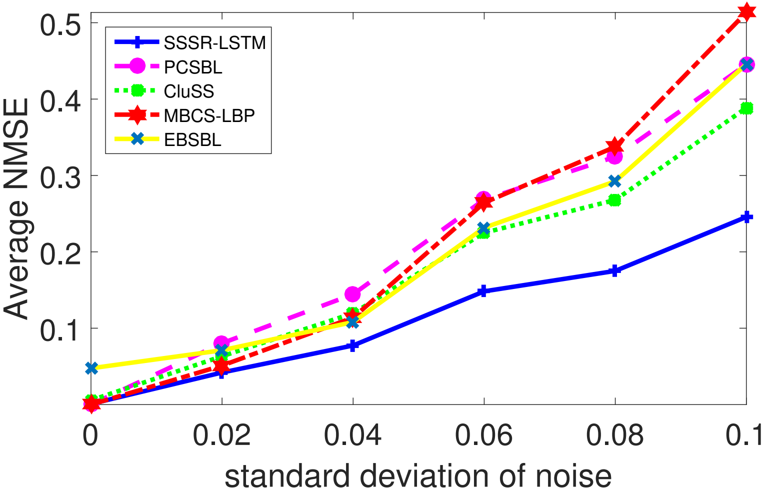
Block-Sparse Signals



Block-Sparse Signals



Block-Sparse Signals

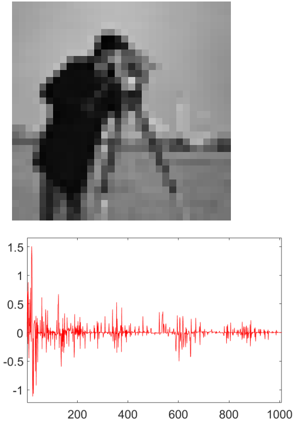


Block-Sparse Signals

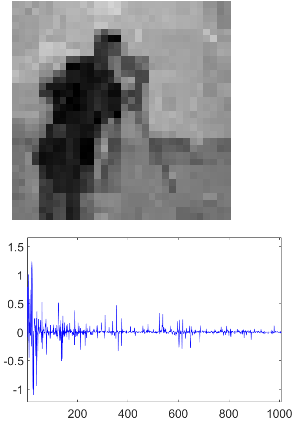
Average NMSE on different MNIST digit test images

digit	1	3	5	7	9
CluSS	0.288	0.250	0.254	0.295	0.341
MBCS-LBP	0.344	0.249	0.268	0.328	0.292
PCSBL	0.310	0.223	0.234	0.311	0.289
EBSBL	0.537	0.373	0.339	0.444	0.525
SSSR-LSTM	**0.150**	**0.194**	**0.186**	**0.190**	**0.161**

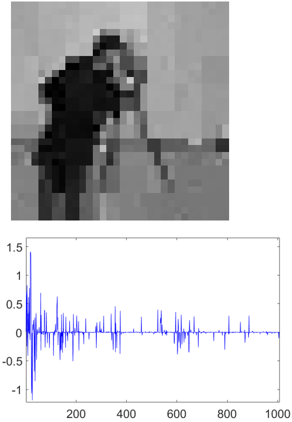
Tree-Struture Signals



Original

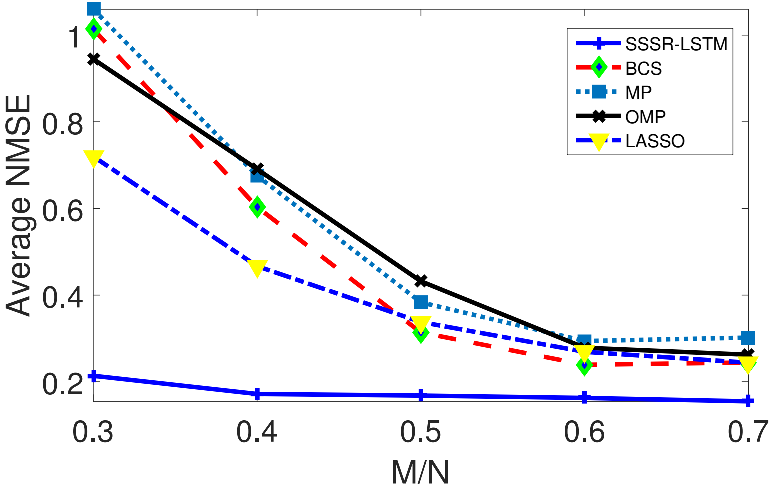


TSW-CS

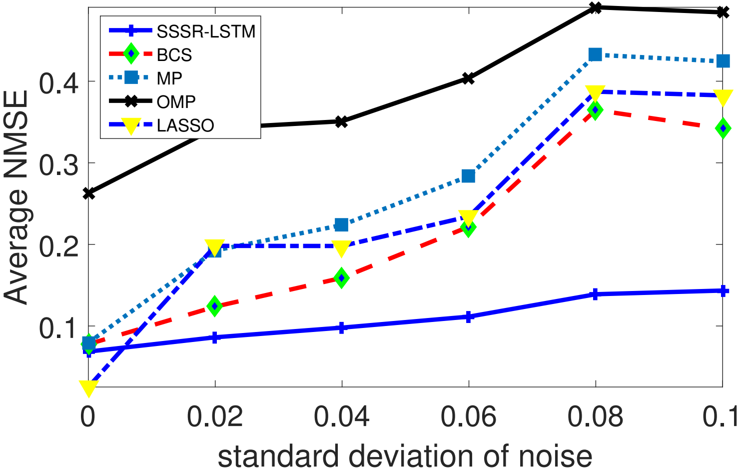


SSSR-LSTM

Uniform-Sparse Signals



Uniform-Sparse Signals

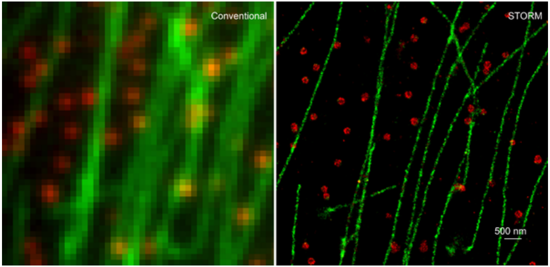


Outline

Outline

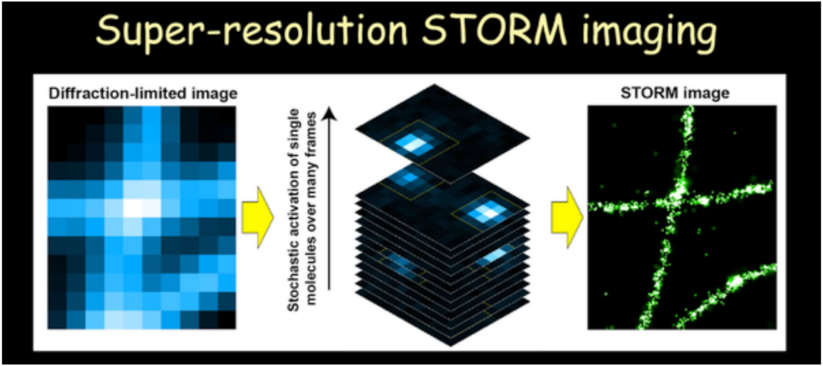
Biology

微管 (英语：Microtubule) 是细胞骨架的一个组成部分，可以在整个细胞质中找到。微管蛋白的这些管状聚合物可以增长长达 50 微米，具有 25 微米的平均长度，并且是高度动态的。微管的外径约为 24 纳米，而内直径为约 12 纳米。



M. Bates, B. Huang, G. T. Dempsey, X. Zhuang, Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes, Science 317 1749-1753 (2007)

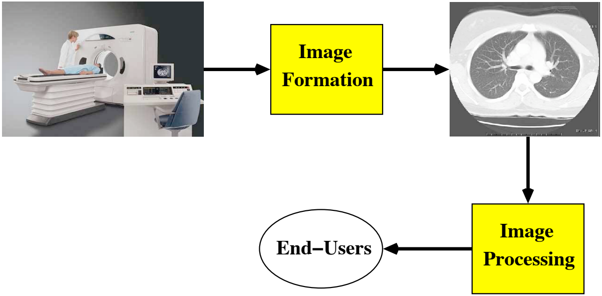
Biology



M. J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM) Nature Methods 3 793-795 (2006)

MRI

"In 2005, the U.S. spent 16% of its GDP on health care. It is projected that this will reach 20% by 2015." Goal: Individualized treatments based on low-cost and effective medical devices.



MRI

- MRI measurements is gathered from transform space (K-space):

$$\boldsymbol{b}_i = \int_x \gamma(x) \exp(-jk_i^T x) dx + n_i$$

$\boldsymbol{b}_i$ Samples in i-th channel

$\gamma(x)$ MRI image in time space (to be recovered)

n_i Noise in i-th channel

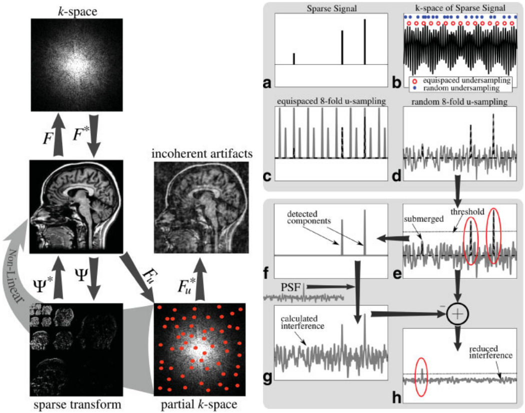
- Concise model

$$\boldsymbol{b} = \mathcal{A}(\gamma) + \boldsymbol{n}$$

- MRI reconstruction exploiting sparsity

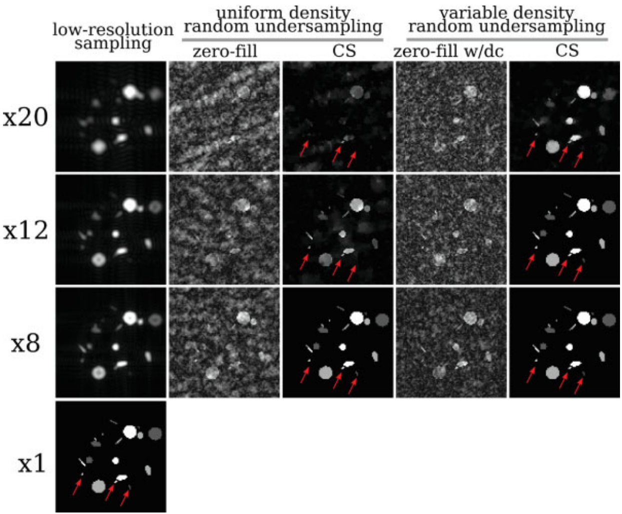
$$\gamma^* = \arg\min_{\gamma} \|\boldsymbol{b} - \mathcal{A}(\gamma)\|_2^2 + \lambda \|\Psi(\gamma)\|_1$$

Sparse MRI

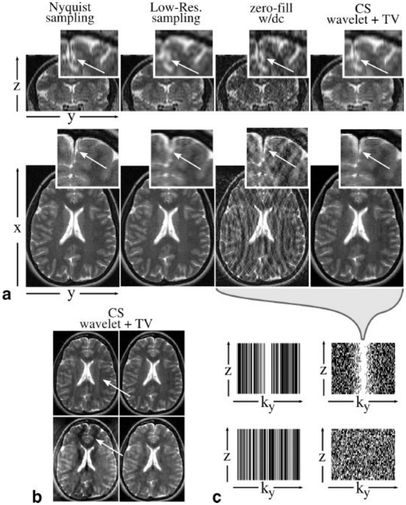


Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic resonance in medicine, 58(6), 1182-1195.

Sparse MRI

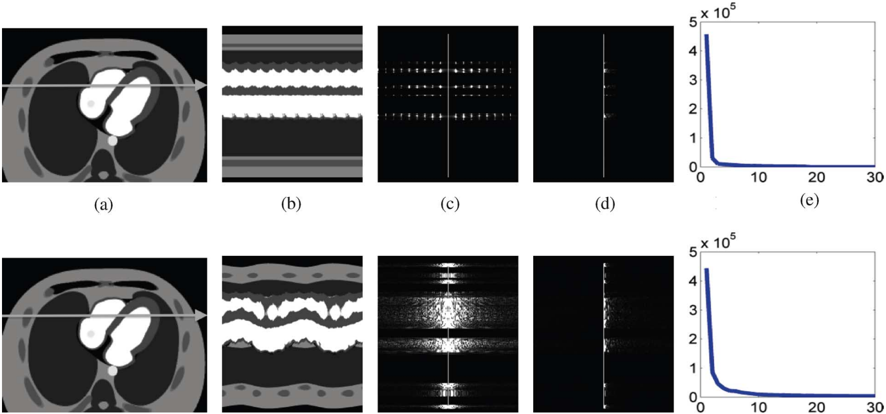

low-resolution sampling
uniform density random undersampling
zero-fill
CS
variable density random undersampling
zero-fill w/dc
CS
x20
x12
x8
x1

Sparse MRI


Nyquist sampling
Low-Res. sampling
zero-fill w/dc
CS wavelet + TV
z
y
x
a
CS wavelet + TV
b
c
k_y

Dynamic MRI exploiting Sparsity and Low-rank

- Lowrank in $x - t$ space



(a) (b) (c) (d) (e)

Dynamic MRI exploiting Sparsity and Low-rank

- MRI measurements in $k-t$ space

$$\boldsymbol{b}_i = \int_x \gamma(x,t)\exp(-jk_i^T x)dx + n_i$$

- Samples in $k-t$ space form a matrix
$\Gamma = [\gamma(x,t_0), \gamma(x,t_1), ..., \gamma(x,t_{n-1})]$

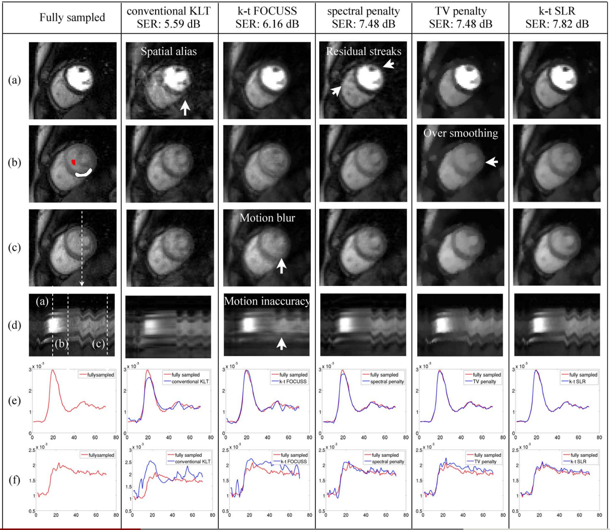
$$\boldsymbol{b} = \mathcal{A}(\Gamma) + \mathbf{n}$$

- Dynamic MRI exploiting Low-rank and Sparsity

$$\Gamma^* = \arg\min \|\mathcal{A}(\Gamma) - \boldsymbol{b}\|_2^2 + \lambda_1 \phi(\Gamma) + \lambda_2 \psi(\Gamma)$$

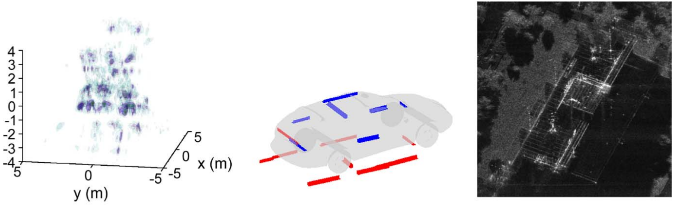
S. G. Lingala, Y. Hu, E. DiBella and M. Jacob, "Accelerated Dynamic MRI Exploiting Sparsity and Low-Rank Structure: k-t SLR," in IEEE Transactions on Medical Imaging, vol. 30, no. 5, pp. 1042-1054, May 2011.

Dynamic MRI exploiting Sparsity and Low-rank



Outline

Radar with Sparsity



(a) (b) (c)

Fig. 1. ***Radar images are compressible. (a) Matched filter three-dimensional image. (b) Nonlinear regression can yield a parsimonious representation of reflectors. (c) Radar image collected using MiniSAR demonstrating the compressibility of radar scenes.***

Radar with Sparsity

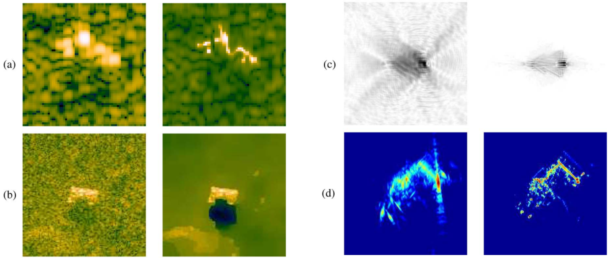


Fig. 2. ***SAR imaging examples. (Left) Conventional imaging and (right) ℓ_p-norm-based reconstruction. (a) MSTAR example with sparsity imposed on reflection coefficients [31]. (b) MSTAR example with sparsity imposed on reflectivity gradients [31]. (c) Passive radar imaging example [32]. (d) Backhoe data (see https://www.sdms.afrl.af.mil/main.php) example for wide-angle imaging aperture of 110°.***

Outline

Image Restoration

- Image restoration is one of the most important and basic areas in image processing.

Model

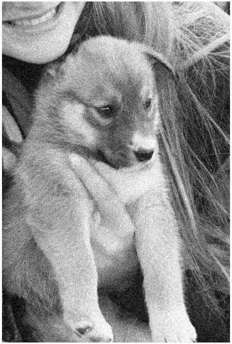
$$Y = HX + N$$

Y — Observed image
H — Degraded operator
X — Original image
N — Additive noise

- Different image restoration problem corresponds to different type of H.

Image Denoising

- When H is the identity matrix.

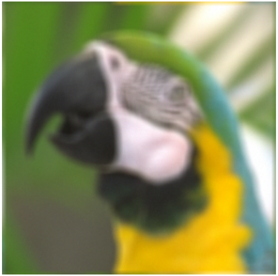


(a) Noisy

(b) Denoising Result

Image Deblurring

- When H is the convolution operator.



(a) Blurred Image

(b) Deblurred Result

Image Inpainting

- When H is the restriction operator.

(a) Miss 80 % pixels

(b) Inpainting Result

Image Inpainting

- When H is the restriction operator.

(a) Corrupted by Text

(b) Text Removal Result

Image Super-Resolution

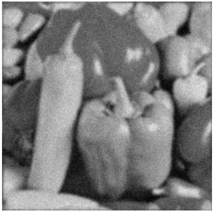
- When H is the downsampling operator.

(a) LR

(b) Interpolated by Bicubic

(c) Reconstruction based SR

Deblurring Results



(a) Blurred Image

(b) Reference 2, PSNR=26.36

(c) Reference 7, PSNR=25.32

(d) Reference 8, PSNR=28.65

[8] Dong W, Zhang L, Shi G, et al. Nonlocally centralized sparse representation for image restoration. IEEE Transactions on Image Processing, 2013, 22(4): 1620-1630.

Denoising Results

(a) Noisy Image

(b) Reference 9, PSNR=27.74

(c) Reference 8, PSNR=28.90

[9] Cai J F, Ji H, Shen Z, et al. Data-driven tight frame construction and image denoising. Applied and Computational Harmonic Analysis, 2014, 37(1): 89-105.

Super-Resolution Results

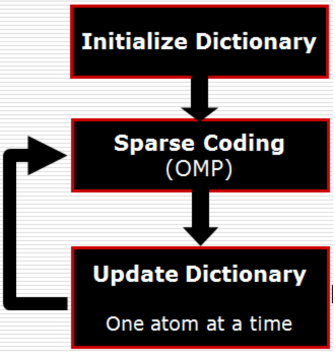

(a) LR


(b) Reference 8, PSNR=31.28


(c) Reference 6, PSNR=31.66

K-SVD

$$\min_{\boldsymbol{D},\boldsymbol{A}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{A}\|_F \quad s.t. \quad \|\boldsymbol{A}\|_0 \le K$$



Step 1. Update Sparse coefficients.

$$\min_{\boldsymbol{A}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{A}\|_F \quad s.t. \quad \|\boldsymbol{A}\|_0 \le K$$

OMP algorithm to solve above problem

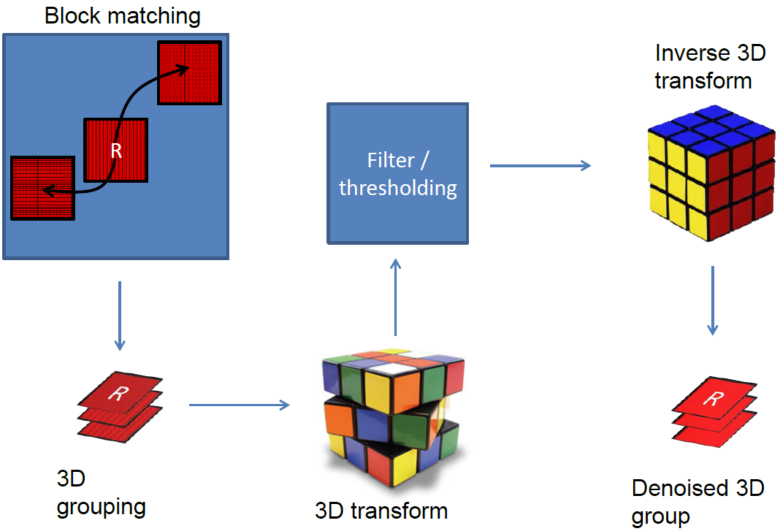
Step 2. Update Dictionary atoms.

$$\min_{\boldsymbol{D}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{A}\|_F$$

apply SVD to update one atom at a time

Alternate two steps until object function converges

Block Matching and 3-D filtering (BM3D)



Learned Simultaneous Sparse Coding (LSSC)

$$\min_{(\boldsymbol{A}_i)_{i=1}^n, \boldsymbol{D}\in C} \sum_{i=1}^{n} \frac{\|\boldsymbol{A}_i\|_{p,q}}{|S_i|^p} \quad s.t. \quad \forall i \sum_{j\in S_i} \left\|\boldsymbol{y}_j - \boldsymbol{D}\boldsymbol{\alpha}_{ij}\right\|_2^2 \leq \epsilon_i$$

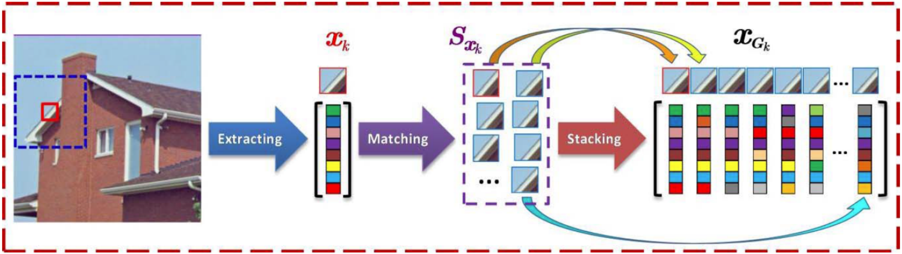
- **Step 1 :Patch Grouping**
  Stacking similar patches to obtain S_i.
- **Step 2 :Update Dictionary**
  Set $p=1, q=2$. Using Online dictionary learning to obtain $\boldsymbol{D}$.
- **Step 3 :Update Sparse Coefficients**
  Set $p=0, q=\infty$. Using OMP to update $\boldsymbol{A}_i$

Alternate above steps until object function converges

Group-Based Sparse Representation (GSR)

$$\min_{\boldsymbol{D}_x,\boldsymbol{\alpha}_{G_k}} \sum_{k=1}^{n} \|\boldsymbol{X}_{G_k} - \boldsymbol{D}_x \boldsymbol{\alpha}_{G_k}\|_2^2 + \lambda \sum_{k=1}^{n} \|\boldsymbol{\alpha}_{G_k}\|_0$$

- Construct 3D groups to stack similar patches. Meanwhile, dictionary and coefficients matrix are both 3D.
- Solving above problem using alternately update dictionary and coefficients matrix.



Results

(a) Noisy Image

(b) KSVD(PSNR=27.86)

(c) BM3D(PSNR=29.05)

(d) Blurred Image

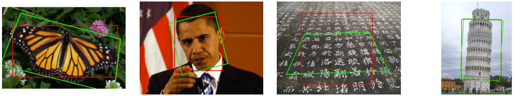
(e) BM3D(PSNR=27.66)

(f) GSR(PSNR=27.77)

Outline

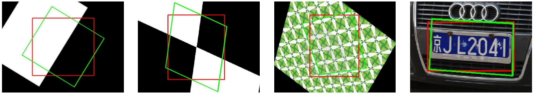
Image Rectification with Low-Rank and Sparsity

- Images with regular patterns have low-rank property

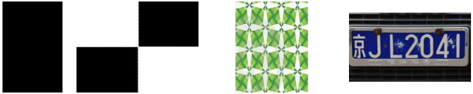


(a) Input ($r = 35$) (b) Input ($r = 15$) (c) Input ($r = 53$) (d) Input ($r = 13$)

(e) Output ($r = 14$) (f) Output ($r = 8$) (g) Output ($r = 19$) (h) Output ($r = 6$)



(a) Input($r = 11$) (b) Input($r = 16$) (c) Input($r = 10$) (d) Input($r = 24$)



(e) Output($r = 1$) (f) Output($r = 2$) (g) Output($r = 7$) (h) Output($r = 14$)

Image Rectification with Low-rank and Sparsity

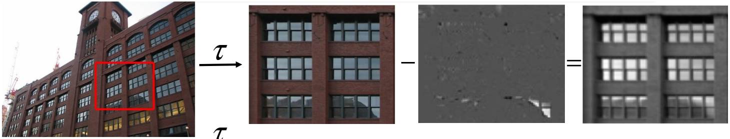
- TILT (Transform Invariant Low-rank Textures) model:

$$I \circ \tau = I^0 + E$$

I^0 Rectified low-rank image

E Sparse error

τ Image transform operator (nonlinear)



$$I \xrightarrow{\tau} I \circ \tau \quad - \quad E \quad = \quad I^0$$

- Rectifying images via optimization:

$$\min_{I^0, E, \Delta\tau} \|I^0\|_* + \lambda \|E\|_1, \quad s.t. \quad I \circ \tau + \nabla I \Delta\tau = I^0 + E$$

Image Rectification with Low-rank and Sparsity

Algorithm 1 (TILT via ALM)

Input: Initial rectangular window $I \in \mathbb{R}^{m\times n}$ in the input image, initial transformations τ in a certain group $\mathbb{G}$ (affine or projective), $\lambda > 0$.

While not converged **Do**

Step 1: normalize the image and compute the Jacobian w.r.t. transformation:

$$I \circ \tau \leftarrow \frac{I \circ \tau}{\|I \circ \tau\|_F}, \quad \nabla I \leftarrow \frac{\partial}{\partial \zeta}\left(\frac{I \circ \zeta}{\|I \circ \zeta\|_F}\right)\Big|_{\zeta=\tau};$$

Step 2: solve the linearized convex optimization (4):

$$\min_{I^0, E, \Delta\tau} \|I^0\|_* + \lambda\|E\|_1 \quad \text{subject to} \quad I \circ \tau + \nabla I \Delta\tau = I^0 + E,$$

with the initial conditions: $Y_0 = 0, E_0 = 0, \Delta\tau_0 = 0, \mu_0 > 0, \rho > 1, k = 0$:

While not converged **Do**

$$\begin{aligned}
(U_k, \Sigma_k, V_k) &\leftarrow \operatorname{svd}(I \circ \tau + \nabla I \Delta\tau_k - E_k + \mu_k^{-1} Y_k),\\
I^0_{k+1} &\leftarrow U_k \mathcal{S}_{\mu_k^{-1}}[\Sigma_k] V_k^T,\\
E_{k+1} &\leftarrow \mathcal{S}_{\lambda\mu_k^{-1}}[I \circ \tau + \nabla I \Delta\tau_k - I^0_{k+1} + \mu_k^{-1} Y_k],\\
\Delta\tau_{k+1} &\leftarrow (\nabla I^T \nabla I)^{-1} \nabla I^T (-I \circ \tau + I^0_{k+1} + E_{k+1} - \mu_k^{-1} Y_k),\\
Y_{k+1} &\leftarrow Y_k + \mu_k (I \circ \tau + \nabla I \Delta\tau_{k+1} - I^0_{k+1} - E_{k+1}),\\
\mu_{k+1} &\leftarrow \rho\mu_k,
\end{aligned}$$

End While

Step 3: update transformations: $\tau \leftarrow \tau + \Delta\tau_{k+1}$;

End While

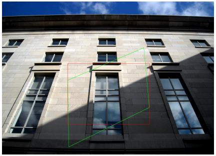
Output: I^0, E, τ.

[11] Zhang, Z., et al., TILT: Transform Invariant Low-Rank Textures. International Journal of Computer Vision, 2012. 99(1): p. 1-24.
[12] Ren, X. and Z. Lin, Linearized Alternating Direction Method with Adaptive Penalty and Warm Starts for Fast Solving Transform Invariant Low-Rank Textures. International Journal of Computer Vision, 2013. 104(1): p. 1-14.

Image Rectification with Low-rank and Sparsity

Image Rectification with Low-rank and Sparsity

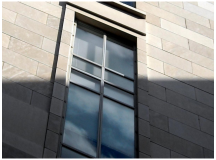
- Failure case:

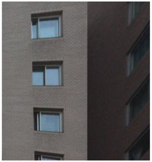


(a) high-rank structures

(b) two low-rank regions

(c) too much occlusion





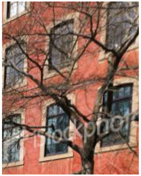


Image Rectification with Low-rank and Sparsity

本文提出一种基于稀疏贝叶斯框架[5, 6]的TILT方法详见附录B

$$\min_{I^0, E, \tau} \left\| I^0 \right\|_* + \xi \left\| E \right\|_1 \quad s.t. \;\; I \circ \tau + J \Delta\tau = I^0 + E$$

重写TILT原始数学模型为如下形式：

$$Y = I^0 + E + \Phi\Delta\tau + N; \;\; Y = I \circ \tau, \Phi = -J$$

对式子的每一项进行**贝叶斯建模**[5] ：Gaussian-Gamma模型诱导相关项的稀疏性和低秩性

待估计量	分布	方差
$I^0 = AB^T$	$A_{gk} \quad (0, \gamma_k^{-1}), B_{gk} \quad (0, \gamma_k^{-1})$	$\gamma_k :\ \Gamma(a_\gamma, b_\gamma)$
E	$E_{ij} \quad (0, \alpha_{ij}^{-1})$	$\alpha_{ij} :\ \Gamma(\mathrm{a}_\alpha, \mathrm{b}_\alpha)$
$\Delta\tau$	$S\Delta\tau \quad (0, \lambda^{-1})$	$\lambda :\ \Gamma(a_\lambda, b_\lambda)$
Y	$Y \quad (AB^T + E + \Phi\tau, \beta^{-1} I_{mn})$	$p(\beta) = \beta^{-1}$

然后利用**变分贝叶斯推导**[6]，给出每一个待估计量的更新公式

[13] Hu, S., et al., Sparse Bayesian learning for image rectification with transform invariant low-rank textures. Signal Processing, 2017. 137: p. 298-308.

Image Rectification with Low-rank and Sparsity

实验结果与分析[7, 8]——矫正成功率

实验说明：取一些典型的低秩图片，施加一个仿射变换，再给这些变换后的图片加入不同程度的随机污染，记录三种算法在不同程度下的随机污染能正确地恢复多少张图片。可以观察到，本文提出的BF-TILT对随机污染的**鲁棒性更高**。

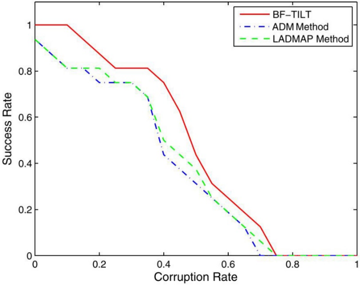


Image Rectification with Low-rank and Sparsity

实验结果与分析——恶劣情况下的矫正结果对比

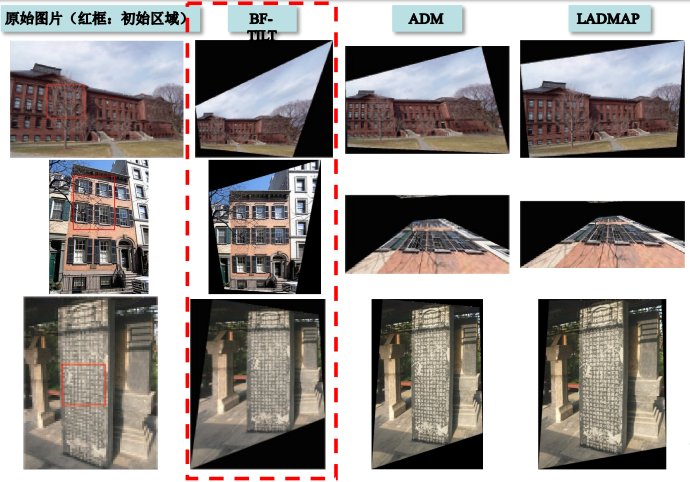
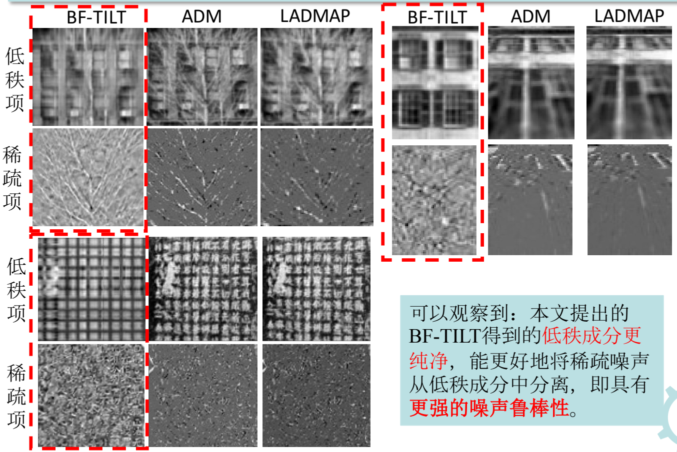


Image Rectification with Low-rank and Sparsity

实验结果——上一个实验恢复得到的低秩项与稀疏项对比

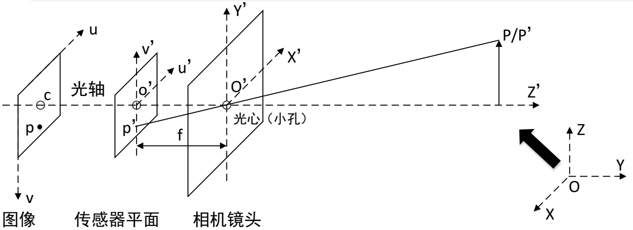


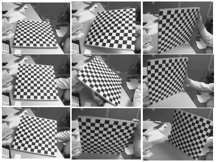
可以观察到：本文提出的BF-TILT得到的低秩成分更纯净，能更好地将稀疏噪声从低秩成分中分离，即具有**更强的噪声鲁棒性**。

Image Rectification with Low-rank and Sparsity

Camera Calibration with Low-rank and Sparsity

针孔摄像头：张正友标定法[1]





相机参数	标定结果
焦距	$[f_u、f_v]$ = [662.49480，664.67679] +/- [1.43392，1.54252]
主点位置	$[c_u、c_v]$ = [306.51221，241.75115] +/- [2.83469，2.60814]
扭曲	α = [0.00000] +/- [0.00000]

[14] Zhang, Z., A Flexible New Technique for Camera Calibration. IEEE Transactions on pattern analysis and machine intelligence, 2000. 22(11): p. 1330–1334.

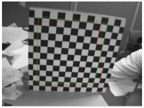
Camera Calibration with Low-rank and Sparsity

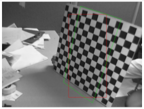
当变换τ为相机成像过程，TILT可用于相机标定

文献[5]解这样一个凸优化问题，原始的TILT数学模型经过变化

$$\min_{I^0,E,\tau^i,\tau^o} \left\|I^0\right\|_* + \xi\left\|E\right\|_1 \quad s.t. \quad I\circ\left(\tau^i,\tau^o\right)+J_{\tau^i}\Delta\tau^i ++J_{\tau^o}\Delta\tau^o = I^0+E$$

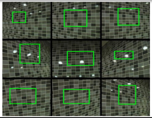
单张图片的标定方法——针孔摄像头

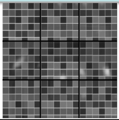




标定结果	
$[f_u, f_v]$	[662.49480, 664.67679]
f	677.2436
f	652.7898

多张图片的标定方法——针孔摄像头





标定结果

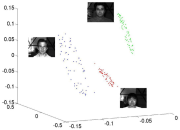
$$\begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1138.6 & 0 & 482.3 \\ 0 & 1127.8 & 267.7 \\ 0 & 0 & 1 \end{bmatrix}$$

[15] Zhang, Z., et al., Camera calibration with lens distortion from low-rank textures. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011: p. 2321-2328.

Outline

Face Recognition via Sparse Representation

1. Assume $\mathbf{y}$ belongs to Class i



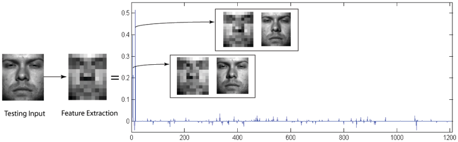
$$\begin{aligned}\mathbf{y} &= \alpha_{i,1}\mathbf{v}_{i,1} + \alpha_{i,2}\mathbf{v}_{i,2} + \cdots + \alpha_{i,n_1}\mathbf{v}_{i,n_i}, \\ &= A_i\alpha_i,\end{aligned}$$

where $A_i = [\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \cdots, \mathbf{v}_{i,n_i}]$.

2. Nevertheless, Class i is the **unknown** variable we need to solve:

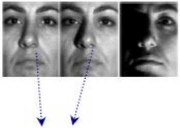
Sparse representation $\mathbf{y} = [A_1, A_2, \cdots, A_K]\begin{bmatrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K\end{bmatrix} = A\mathbf{x} \in \mathbb{R}^{3\cdot 640\cdot 480}$.

3. $\mathbf{x}_0 = [\,0 \cdots 0 \;\alpha_i^T\; 0 \cdots 0\,]^T \in \mathbb{R}^n$.

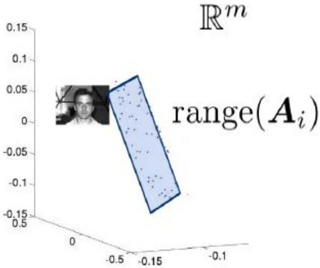


Sparse representation encodes membership!

Face Recognition



$$A_i = [\ |\ |\ \ldots\ ] \in \mathbb{R}^{m \times n_i}$$



y  $\approx\ x_{i,1}$  $+\ x_{i,2}$  $+ \ldots +\ x_{i,n}$  $= A_i x_i$

Face Recognition via Sparse Representation

Algorithm 1 (Recognition via Sparse Representation)

1: **Input:** a matrix of training images $A \in \mathbb{R}^{m \times n}$ for k subjects, a linear feature transform $R \in \mathbb{R}^{d \times m}$, a test image $\boldsymbol{y} \in \mathbb{R}^m$, and an error tolerance ϵ.

2: Compute features $\tilde{\boldsymbol{y}} = R\boldsymbol{y}$ and $\tilde{A} = RA$, and normalize $\tilde{\boldsymbol{y}}$ and columns of $\tilde{A}$ to unit length.

3: Solve the convex optimization problem (P_1'):

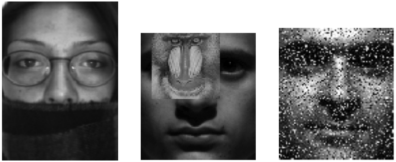
$$\min \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\tilde{\boldsymbol{y}} - \tilde{A}\boldsymbol{x}\|_2 \leq \epsilon.$$

4: Compute the residuals $r_i(\boldsymbol{y}) = \|\tilde{\boldsymbol{y}} - \tilde{A}\,\delta_i(\boldsymbol{x})\|_2$ for $i = 1, \ldots, k$.

5: **Output:** $\text{identity}(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$.

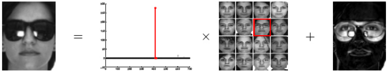
[10] John Wright, Allen Y. Yang, Arvind Ganesh, Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. To appear in PAMI, 2008.

Face Recognition via Sparse Representation



1. Sparse representation + sparse error

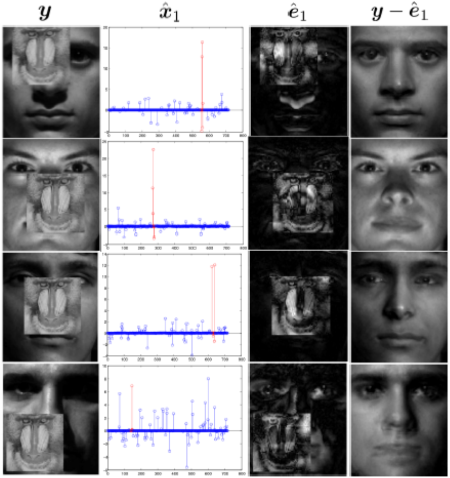
$$\mathbf{y} = A\mathbf{x} + \mathbf{e}$$

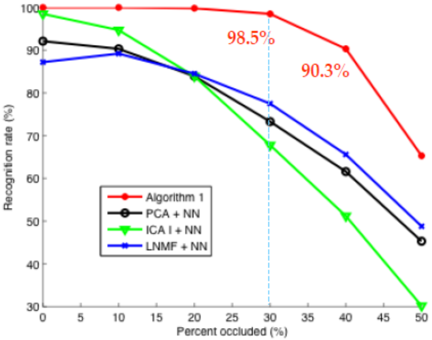


2. Occlusion compensation

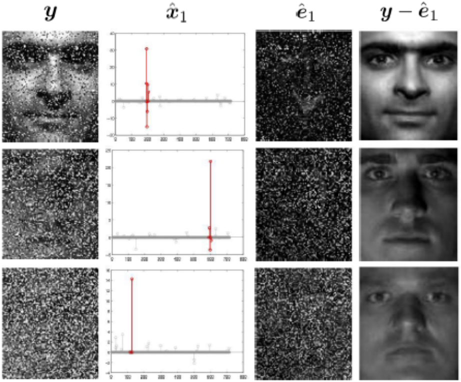
$$\mathbf{y} = \begin{pmatrix} A & \mid & I \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{e} \end{pmatrix} = B\mathbf{w}$$

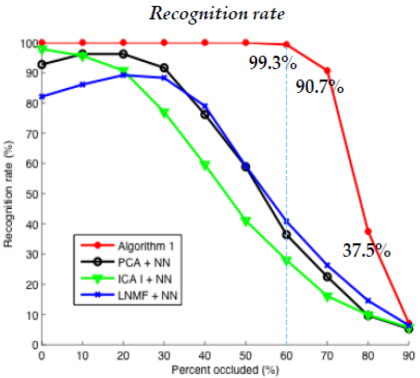
Face Recognition via Sparse Representation





Face Recognition via Sparse Representation





Q & A

$$g(x) = \lambda \|x\|_2^2$$

Energy

$$g(x) = \lambda \rho\{\mathbf{L}x\}$$

Robust Statistics

$$g(x) = \lambda \| |\nabla x| \|$$

Total Variation

$$g(x) = \lambda \|\mathbf{W}x\|_1$$

Wavelet Sparsity

$$g(x) = \lambda \|\alpha\|_0^0$$

$$\text{for } x = \mathbf{D}\alpha$$

Sparse Overcomplete

Structured

- Group sparsity
- Dynamic Group sparsity
- Graph Sparsity
- ......

Sparsity, Structured sparsity, Low-rank

What's next?