Compressed Sensing: Theory, Algorithms and Applications

Shihua Zhang

Fall 2019

Overview

Outline

Data: Increasingly Massive, High-dimensional...

Images:

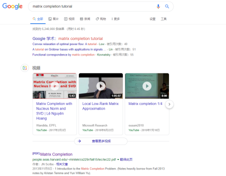
1M pixels
Compression
De-noising
Recognition...

Video:

1B voxels
Streaming
Tracking
Stabilization...

User data:

1B users
Clustering
Classification
Collaborative filtering...



Web data:

100B webpages
Indexing
Ranking
Search...

Data: Increasingly Massive, High-dimensional...

Images:

1M pixels

Compression

De-noising

Recognition...

Video:

1B voxels

Streaming

Tracking

Stabilization...

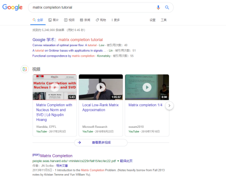
User data:

1B users

Clustering

Classification

Collaborative filtering...



Web data:

100B webpages

Indexing

Ranking

Search...

Generally, we think how to analyze such high-dimensional data.

Data: Increasingly Massive, High-dimensional...

Images:

1M pixels

Compression

De-noising

Recognition...

Video:

1B voxels

Streaming

Tracking

Stabilization...

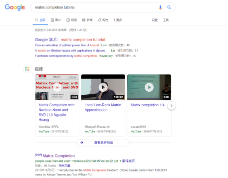
User data:

1B users

Clustering

Classification

Collaborative filtering...



Web data:

100B webpages

Indexing

Ranking

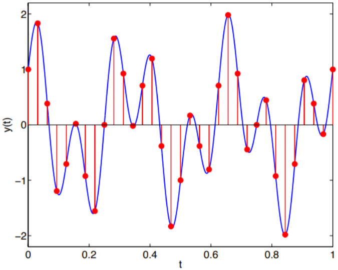
Search...

Generally, we think how to analyze such high-dimensional data.

How to measure such high-dimensional data?

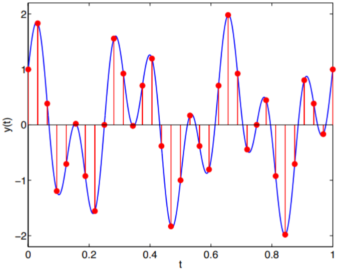
Data Acquisition

Foundation: Shannon-Nyquist Sampling Theorem:
"No information loss if we sample at 2*x* the bandwidth"



Data Acquisition

Foundation: Shannon-Nyquist Sampling Theorem:
"No information loss if we sample at 2*x* the bandwidth"



Success has many fathers...

Old-fashioned Thinking

Long-established paradigm for digital data acquisition.

- uniformly sample data at Nyquist rate.
- compress data.

$N \gg K$

$$x \longrightarrow \boxed{\text{sample}} \xrightarrow{N} \boxed{\textbf{compress}} \xrightarrow{K} \boxed{\text{transmit/store}}$$

JPEG
JPEG2000
...

Shannon sampling theorem

$$\boxed{\text{receive}} \xrightarrow{K} \boxed{\text{decompress}} \xrightarrow{N} \widehat{x}$$

Sparsity

Many real-world signals are compressible in the sense that they are well approximated by sparse signals.

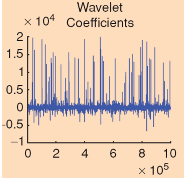


Figure: Left: original image. Right: reconstruction using 1% of the largest absolute wavelet coefficient, i.e., 99% of the coefficients are set to zeros

Sparsity

Many real-world signals are compressible in the sense that they are well approximated by sparse signals.

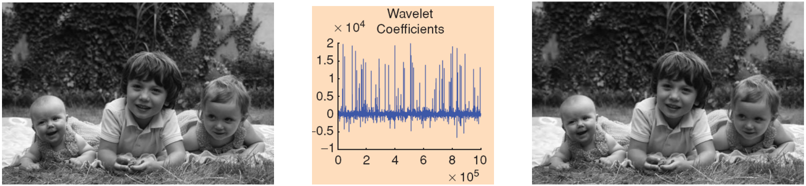


Figure: Left: original image. Right: reconstruction using 1% of the largest absolute wavelet coefficient, i.e., 99% of the coefficients are set to zeros

Throwing away most of the information won't affect the quality!

Motivation

- Why go to so much effort to acquire all the data when most of what we get will be thrown away?
- Cant we just directly measure the part that won't end up being thrown away?

Motivation

- Why go to so much effort to acquire all the data when most of what we get will be thrown away?
- Cant we just directly measure the part that won't end up being thrown away?

People involved

Emmanuel Candes
Caltech

Terence Tao
UCLA

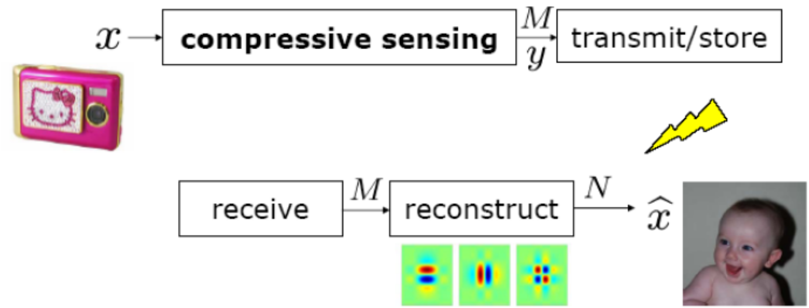
Justin Romberg
Georgia Tech

David Donoho
Stanford University

New Idea: Compressed Sensing (CS)

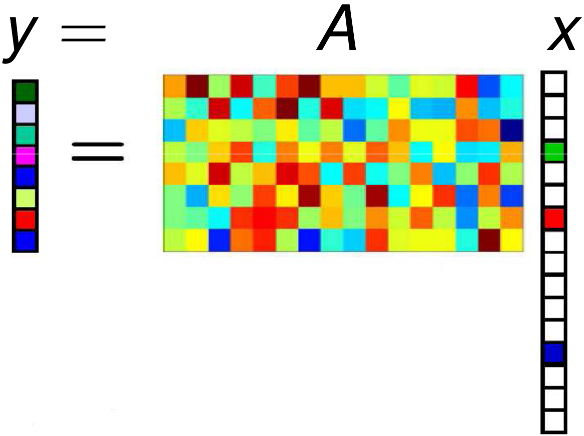
- Directly acquire “compressed” data.
- Replace samples by more general “measurements”.

$$K \approx \underline{M} \ll N$$



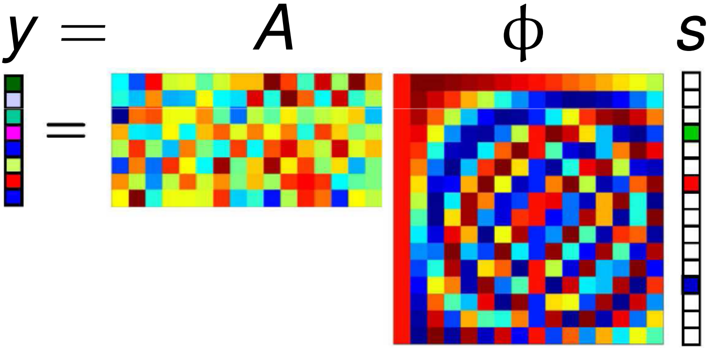
What is Compressed Sensing?

- Basic assumption: the signal x is sparse.
- Goal: given y, find x.
- The linear measurement process can be represented by a matrix $A \in \mathbb{C}^{m \times N}$.



What is Compressed Sensing?

- If x is sparse in a specific basis, i.e. $x = \phi s$

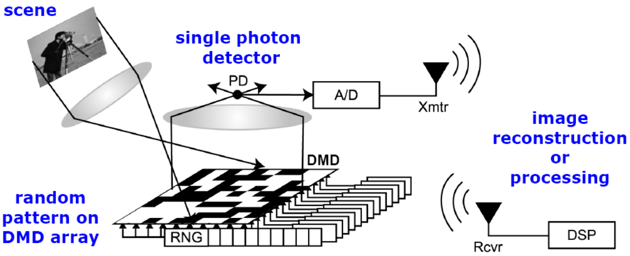


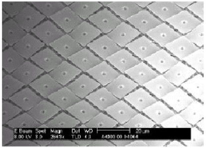
Outline

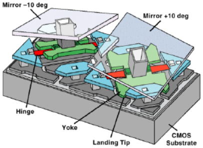
Outline

Single-Pixel Camera [Duarte et al., 2008]

- The crucial ingredient is a mirror array.
- Flip mirror array M times to acquire M measurements.



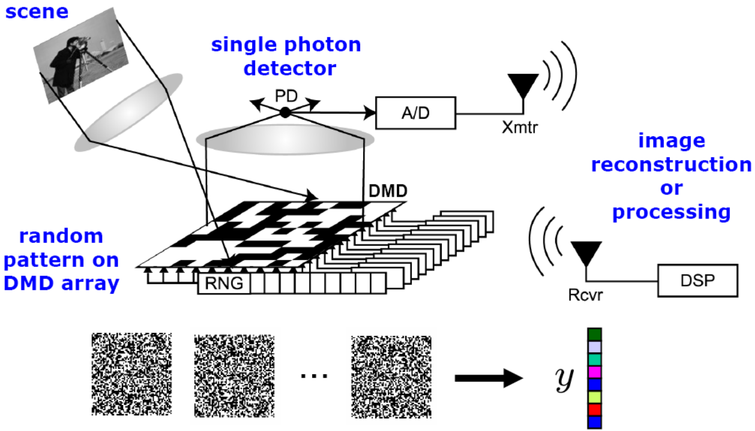




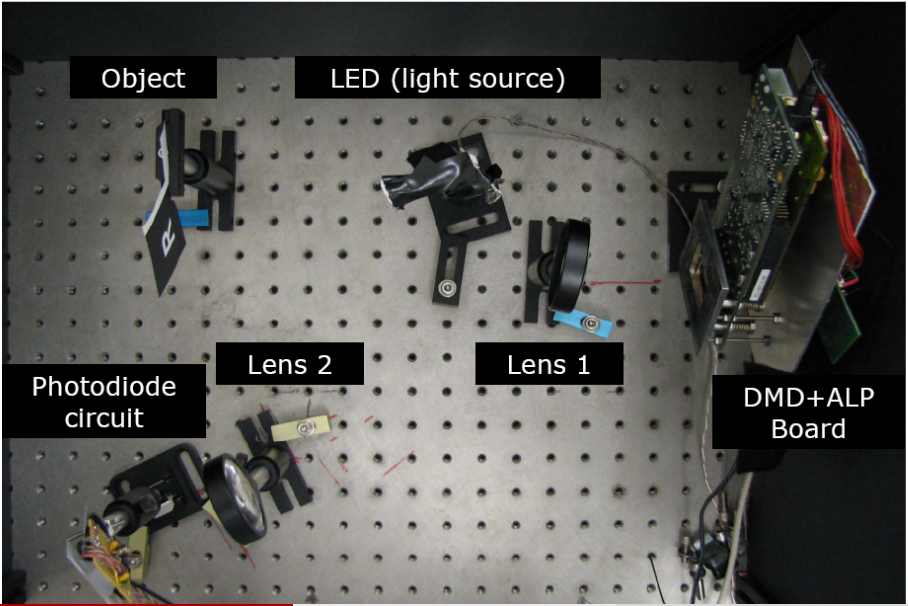
w/ Kevin Kelly

Single-Pixel Camera

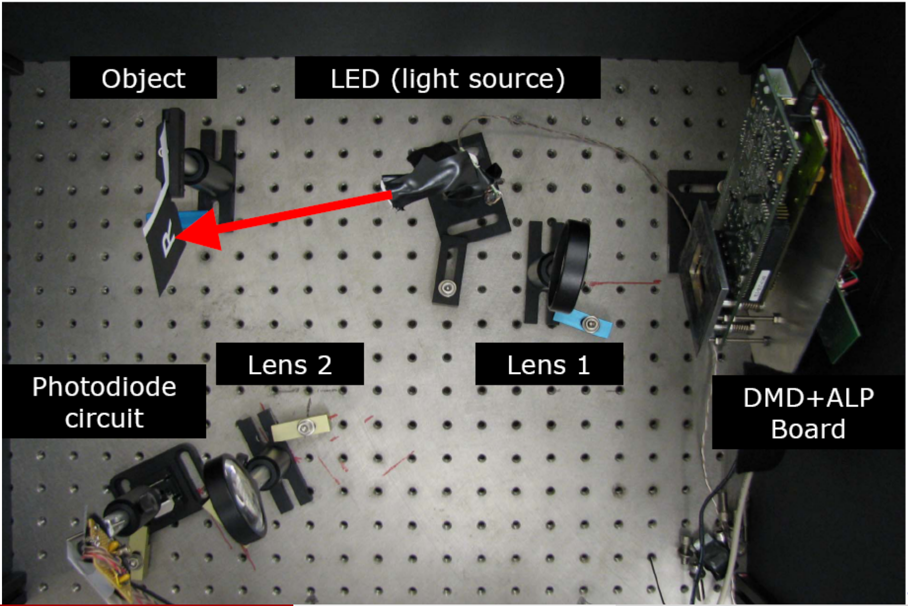
- Images are sparse in the basis of wavelet or discrete cosine transformation.



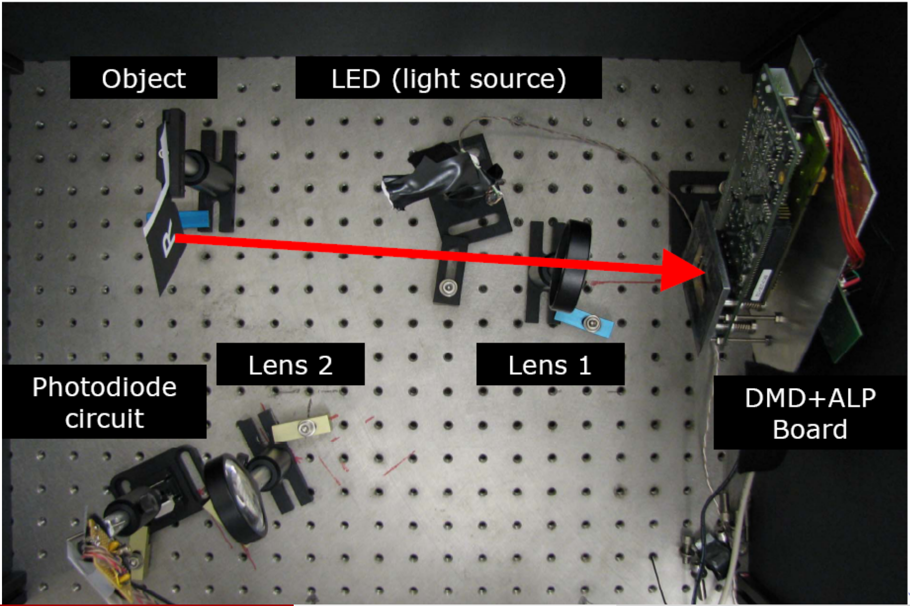
Single-Pixel Camera



Single-Pixel Camera



Single-Pixel Camera



Single-Pixel Camera

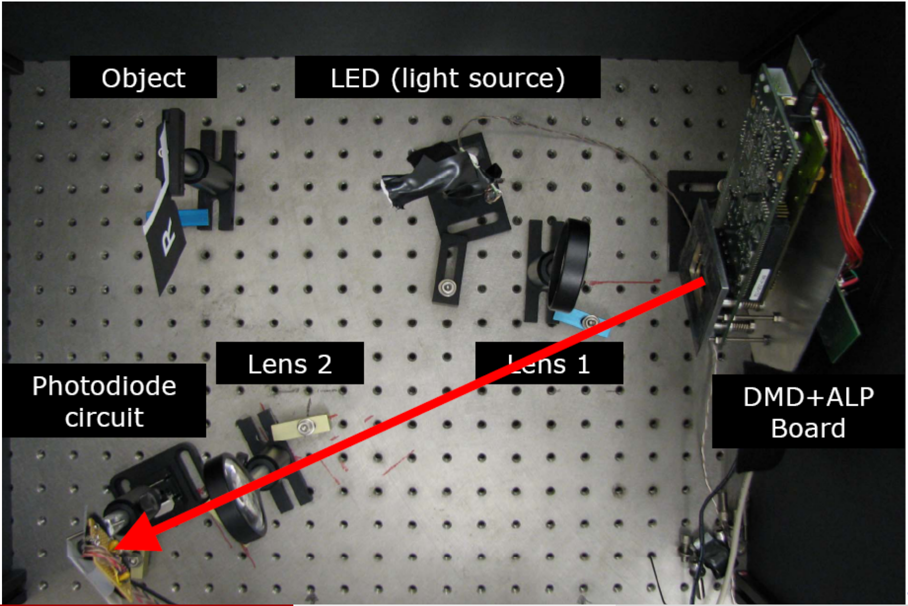
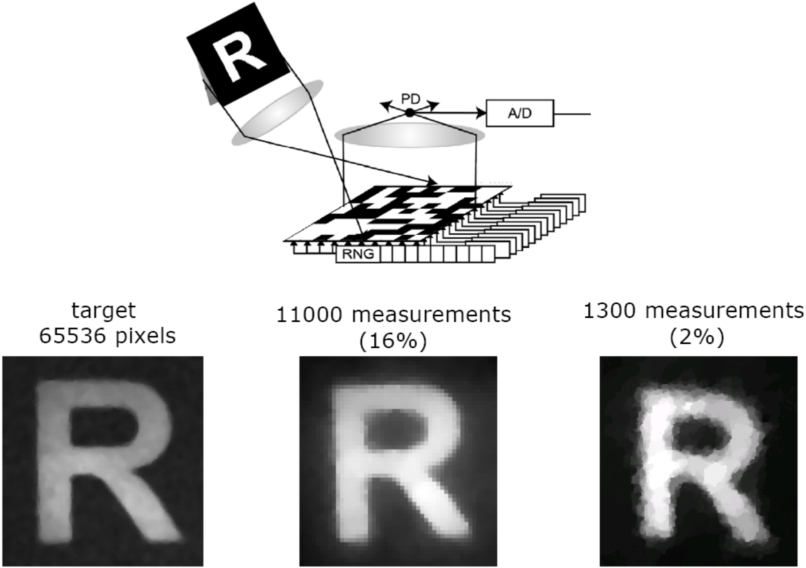


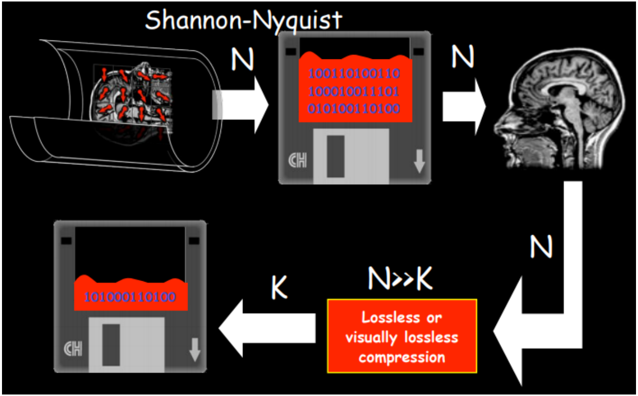
Image Acquisition



Outline

Magnetic Resonance Imaging [Haldar et al., 2010]

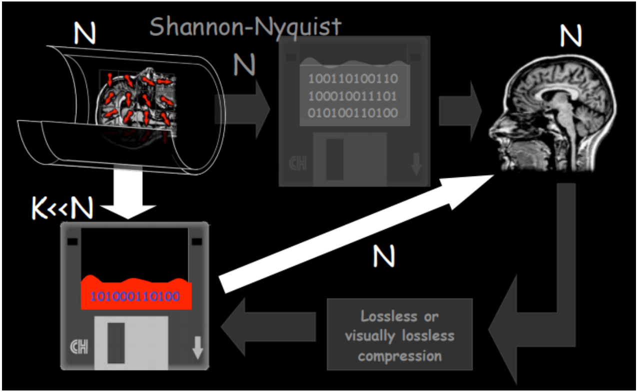
Standard approach: first collect, then compress.



More sampling effort!

Magnetic Resonance Imaging

Compressed Sensing: first compress, then reconstruct.



Less sampling time!

Benefit of CS

What will happen if MRI takes less time?

Benefit of CS

What will happen if MRI takes less time?

- Heart patients won't hold their breath for too long.
- Children won't lose patience.
- Less harmful radiation.
- ...

Magnetic Resonance Imaging

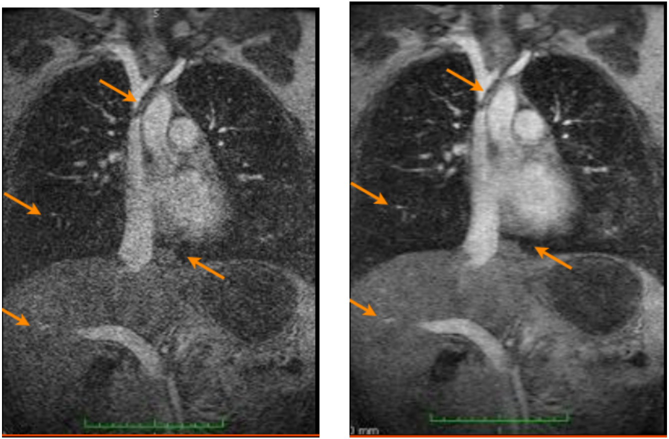


Figure: Left: Shannon-Nyquist. Right: Compressed Sensing

Magnetic Resonance Imaging

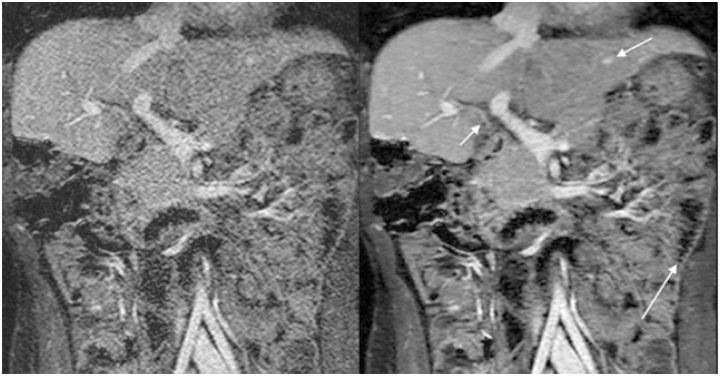


Figure: Left: Shannon-Nyquist. Right: Compressed Sensing

Better quality!

Outline

Outline

How to Implement CS?

Basic Questions

- How can one reconstruct x from $y = Ax$?
- How should one design the linear measurement process?

How to Implement CS?

Basic Questions

- How can one reconstruct x from $y = Ax$?
- How should one design the linear measurement process?

These two question can be separated though not entirely independent.

Optimization Methods

- Goal: given y, find x.
- Optimization formula:

$$\min \|x\|_0 \quad \text{s.t.} \quad Ax = y$$

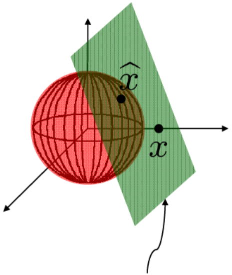
- Nonconvex and NP-hard, convex relaxation instead:

$$\min \|x\|_1 \quad \text{s.t.} \quad Ax = y$$

- The algorithm (problem) is called basis pursuit (BP) [Chen et al., 2001].

Geometry Interpretation

Why l_1 not l_2 works?

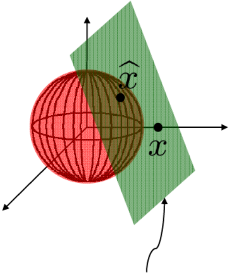


$$\{x : y = Ax\}$$

l_2-minimization is almost never sparse.

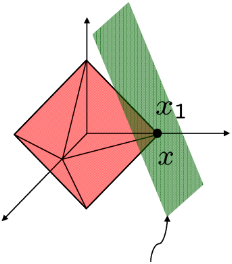
Geometry Interpretation

Why l_1 not l_2 works?



$$\{x : y = Ax\}$$

l_2-minimization is almost never sparse.



$$\{x : y = Ax\}$$

l_1-minimization = sparsest solution.

Quadratically Constrained Basis Pursuit

If the measurement vector y has noise:

$$\min \|x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leqslant \eta$$

The solution is related to the output of LASSO [Tibshirani, 1996], which is consistent in solving:

$$\min \|Ax - y\|_2 \quad \text{s.t.} \quad \|x\|_1 \leqslant \tau$$

Quadratically Constrained Basis Pursuit

If the measurement vector y has noise:

$$\min \|x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leqslant \eta$$

The solution is related to the output of LASSO [Tibshirani, 1996], which is consistent in solving:

$$\min \|Ax - y\|_2 \quad \text{s.t.} \quad \|x\|_1 \leqslant \tau$$

Proposition

If x is a unique minimizer of the quadratically constrained basis pursuit (QCBP) with $\eta \geqslant 0$, then there exists $\tau = \tau_x \geqslant 0$ such that x is a unique minimizer of the LASSO.

Conversion to Linear Programming

- Let $x = u - v$ where
  u keeps positive entities as they are, set negative entities to zeros.
  v keeps negative entities as they are, set positive entities to zeros.
- Let $z = \left[u^T, v^T\right]^T \in \mathbb{R}^{2N}$

$$\|x\|_1 = \mathbf{1}^T(u+v), Ax = A(u-v) = [A, -A]z$$

- Then, the optimization formula turn to:

$$\min_z \mathbf{1}^T z \quad \text{s.t.} \quad y = [A, -A]z, z \geqslant \mathbf{0}$$

How to Implement l_1-recovery

- Utilize various algorithms and packages for the convex optimization.
    - Algorithms: simplex method, interior point method, homotopy method, etc.
    - Packages: LAPACK, GPLK, L1-magic, CVX, L1-LS, Sparselab, etc.

- Computational cost is guaranteed to be $O(n^3)$.

l_p-recovery

- Relaxing to l_p-norm ($0 < p < 1$) often leads to good performance but nonconvex.
- For example, $L_{\frac{1}{2}}$ regularizer possesses sparsity, unbiasedness and Oracle property, and find its usage on **radar** [Xu et al., 2010].
- However, performing the optimization is a nontrivial task.

Iterated Reweighted Least Squares (IRLS) [Daubechies et al., 2010]

Key Idea: solving it with IRLS

- One can analytically solve the optimization of quadratic functions.

  $B = \mathrm{diag}\left(b_i\right), \left(B^{+}\right)_i = \begin{cases} b_i^{-1}, & \left(b_i \neq 0\right) \\ 0, & \left(b_i = 0\right) \end{cases}$

$$\min_x \sum_{i=1}^{N} \left(B^{+}\right)_i |x_i|^2 \quad \text{s.t} \quad y = Ax$$

$$x = BA^T \left(ABA^T\right)^{+} y$$

- Employ this formula for iterative optimization of l_p-minimization.

Iterated Reweighted Least Squares (IRLS) [Daubechies et al., 2010]

- x^k: Solution obtained at the kth iteration
  $X_k = \operatorname{diag}\left(\left|x_i^k\right|^{2-p}\right), \quad (X_k^+)_i = \begin{cases} \left|x_i^k\right|^{p-2}, & (x_i^k \neq 0) \\ 0, & (x_i^k = 0) \end{cases}$

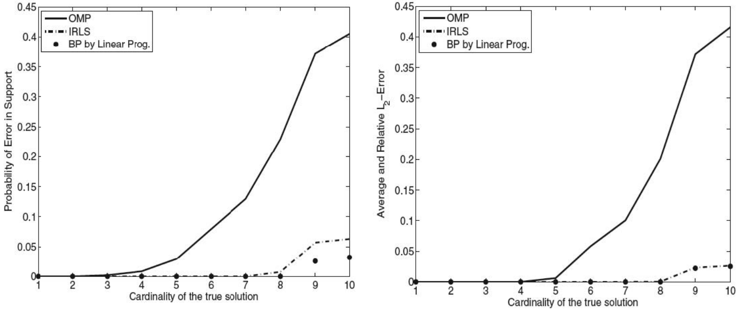
$$\min_x \sum_{i=1}^N \left(X_{k-1}^+\right)_i |x_i|^2 \quad \text{s.t} \quad y = Ax$$

$$x^k = X_{k-1}A^T\left(AX_{k-1}A^T\right)^+ y$$

- Converges to a local minimum.

Comparison

- Set $p = 1$.



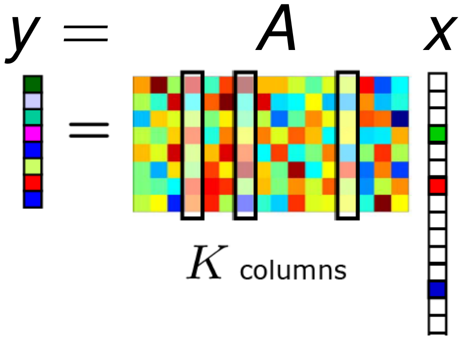
- Computational cost BP, IRLS $\gg$ OMP

Outline

Matching Pursuit (MP) [Mallat and Zhang, 1993]

Key Idea:

- Measurements y composed of sum of K columns of A.
- Identify which K columns sequentially according to size of contribution to y.



Matching Pursuit (MP)

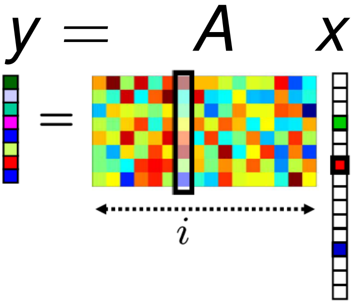
- For each column a_i compute

$$\hat{x}_i = \langle y, a_i \rangle$$

- Choose largest $\|\hat{x}_i\|$ (greedy).
- Update estimate $\hat{x}$ by adding in $\hat{x}_i$.
- Form residual measurement

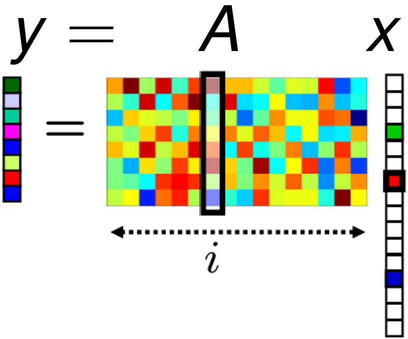
$$y' = y - x_i a_i$$

and iterate until convergence.



Orthogonal Matching Pursuit (OMP) [Tropp, 2004]

- Same procedure as Matching Pursuit.
- Except at each iteration: keep the residue orthogonal to selected columns.



Weakness: once an incorrect index selected, it remains in all the process.

Computational Cost

- Computational cost

s: final support size

OMP Exact enumeration

$$O(mns) \longleftarrow O(mn^s s^2)$$

- Drastic reduction (when $s \sim O(1)$)

Varieties of OMP

Variations in coefficient updates

- OMP minimizes over all the selected elements.
- MP minimizes over the most recently selected element.

Varieties of OMP

Variations in coefficient updates

- OMP minimizes over all the selected elements.
- MP minimizes over the most recently selected element.
- Gradient Pursuits (GP) employs gradient decent to update coefficients [Blumensath and Davies, 2008a], i.e.

$$\widehat{x}^{[i]}_{U^{[i]}} = x^{[i-1]}_{U^{[i]}} + a^{[i]} d^{[i]}_{U^{[i]}}$$

$$d^{[i]}_{U^{[i]}} := g^{[i]}_{U^{[i]}} = A^T_{U^{[i]}} \left(y - A_{U^{[i]}} \widehat{x}^{[i-1]}_{U^{[i]}} \right)$$

- Conjugate Gradient Pursuit (CGP) use conjugate gradient while selecting the update direction [Blumensath and Davies, 2008a].

Varieties of OMP

Variations in element selection

- StOMP to speed up, select multiple elements at a time [Donoho et al., 2006]:

$$U^{[i]} = U^{[i-1]} \cup \left\{ j : \left| g_j^{[i]} \right| / \|A_j\|_2 \geqslant \lambda^{[i]} \right\}$$

$$\lambda_{\text{stomp}}^{[i]} = t^{[i]} \left\| y - A_{U^{[i]}} \widehat{x}_{U^{[i]}}^{[i-1]} \right\|_2 / \sqrt{m}$$

$t^{[i]}$ usually take a value: $2 \leqslant t^{[i]} \leqslant 3$

Varieties of OMP

Variations in element selection

- StOMP to speed up, select multiple elements at a time [Donoho et al., 2006]:

$$U^{[i]} = U^{[i-1]} \cup \left\{ j : \left| g_j^{[i]} \right| / \|A_j\|_2 \geqslant \lambda^{[i]} \right\}$$

$$\lambda_{\text{stomp}}^{[i]} = t^{[i]} \left\| y - A_{U^{[i]}} \widehat{x}_{U^{[i]}}^{[i-1]} \right\|_2 / \sqrt{m}$$

$t^{[i]}$ usually take a value: $2 \leqslant t^{[i]} \leqslant 3$

Problem: the algorithm may terminates prematurely.

Varieties of OMP

Variations in element selection

- Stage weak element selection won't terminates too early.

$$\lambda_{weak}^{[i]} = \alpha \max_j \frac{\left| g_j^{[i]} \right|}{\|A_j\|_2}$$

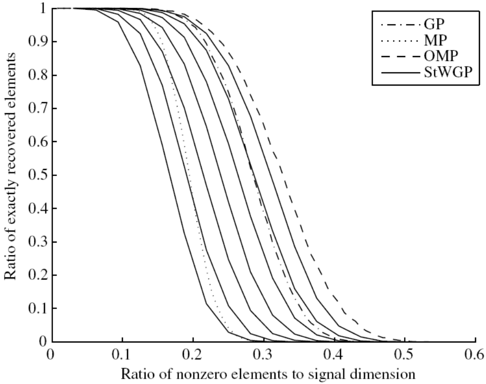
  A good combination is Stagewise Weak Conjugate Gradient Pursuit (StWGP) [Blumensath and Davies, 2009].

- Order Recursive Matching Pursuit (ORMP) Choose best-feat column combined with selected columns [Barron et al., 2008].
  - More accurate evaluation although cost increases.

Comparison of Cost

Algorithm	Computation cost	Storage cost
MP	$m + A + n$	$A + m + 2k + n$
OMP (QR)	$2mk + m + A + n$	$2(m+1)k + 0.5k(k+1) + A + n$
OMP (Chol)	$3A + 3k^2 + 2m + n$	$0.5k(i+1) + A + m + 2k + n$
GP	$2A + k + 3m + n$	$2m + A + 2k + n$
CGP	$2A + k + 3m + n$	$2m + A + 2k + n$
StWGP	$2A + k + 3m + 2n$	$2m + A + 2k + n$
StOMP (CG)	$(\nu + 2)A + k + 3m + n$	$2m + A + 2k + n$
ORMP	$2m(n-k) + 3m + A + n$	$2(m+1)k + 0.5k(k+1) + nm + n$

Comparison of Accuracy



The solid lines correspond to (from left to right) the parameter of conjugate gradient: $\alpha = 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0$

Outline

Thresholding-based Methods

Weakness of greedy algorithms: Bad selection remains.
Solution: Increase the number of iterations.

Hard thresholding operator $H_s(z)$

$$L_s(z) := \text{index set of } s \text{ largest absolute entries of } z \in \mathbb{C}^N,$$
$$H_s(z) := z_{L_s(z)}$$

Note: when the sparsity s is unknown, the hard thresholding operator gives ways to a soft thresholding operator.

Thresholding-based method

Basic thresholding (BT)

- First find active set U according to the contribution to A

$$U = L_s\left(A^* y\right)$$

- Then compute the least-square solution with U

$$x = \underset{z \in \mathbb{C}^N}{\operatorname{argmin}} \{\|y - Az\|_2, \operatorname{supp}(z) \subset U\}$$

Basic thresholding (BT) is fastest sacrificing recovery accuracy!

Compressive Sampling Matching Pursuit (CoSaMP) [Needell and Tropp, 2009]

- Define the active set U

$$U = \operatorname{supp}(x) \cup L_{2s}\left(A^*(y - Ax)\right)$$

- Find the least-square solution u

$$u = \underset{z \in \mathbb{C}^N}{\operatorname{argmin}} \{\|y - Az\|_2, \operatorname{supp}(z) \subset U\}$$

- Impose sparsity by hard thresholding operator

$$x = H_s(u)$$

- Iterate until converge.

Iterative Hard Thresholding (IHT) [Blumensath and Davies, 2008b]

- Transform the rectangular system $Ax = y$ to the square system

$$A^* Ax = A^* y$$

- Interpret it as the fixed-point equation

$$x = (I - A^* A)x + A^* y = x + A^*(y - Ax)$$

- Keep the s largest absolute entries

$$x^+ = H_s(x + A^*(y - Ax))$$

- Iterate until converge.
- Computational cost $O(mn)$/iteration.

Hard Thresholding Pursuit (HTP) [Foucart, 2011]

- As before, we can change

$$x^{+} = H_s(x + A^*(y - Ax))$$

into:

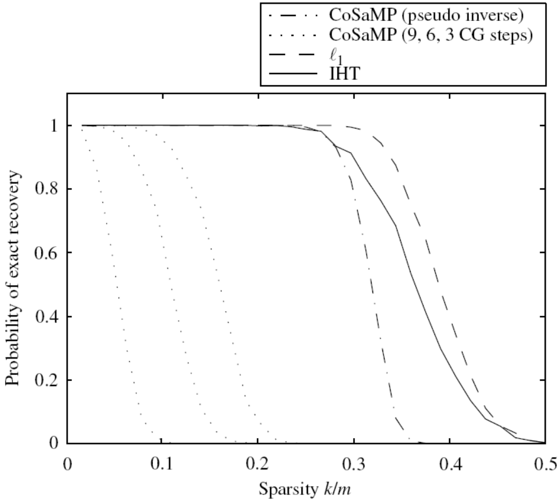
$$U = L_s\left(x + A^*(y - Ax)\right)$$

$$x^{+} = \underset{z\in\mathbb{C}^N}{\operatorname{argmin}}\{\|y - Az\|_2, \operatorname{supp}(z) \subset U\}$$

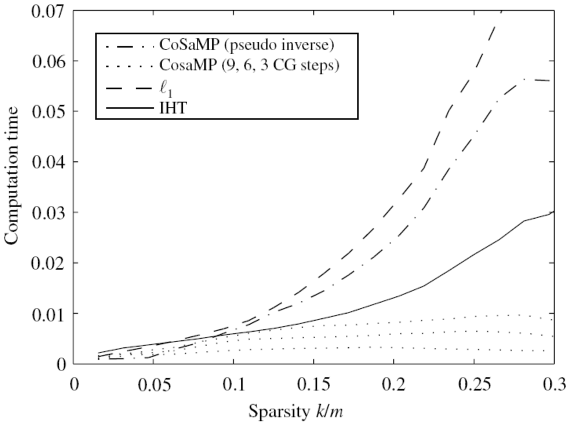
for better fits the measurements y.

- This algorithm called hard thresholding pursuit.

Comparison of Accuracy



Comparison of Cost



Which Algorithm Should One Choose?

Criterion 1: number of measurements

- Vary with each algorithm.
- A matter of numerical tests.

Criterion 2: speed

- Greedy algorithms and algorithms with orthogonal projection operator are sensitive to sparsity s.
- Basis pursuit and thresholding-based algorithms aren't influenced by the sparsity s.

Criterion 3: fast matrix–vector multiplication

- Complicated for orthogonal projection steps.
- Algorithms without orthogonal projection steps can be speed up.

Outline

Outline

What Matrix Can Guarantee Recovery?

Ideal: Solve...

$$\min_{z\in\mathbb{C}^N} \|x\|_0 \quad \text{s.t. } y = Ax \quad (P_0)$$

Convex Relaxation

$$\min_{z\in\mathbb{C}^N} \|x\|_1 \quad \text{s.t. } y = Ax \quad (P_1)$$

Sufficient Condition for '$l_0 = l_1$'?
Coherence!

Coherence

Definition

Let $A = (a_i)_{i=1}^N$ be an $m \times N$ matrix. Then its coherence $\mu(A)$ is

$$\mu(A) = \max_{i \neq j} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2} \in \left[\sqrt{\frac{N-m}{m(N-1)}}, 1 \right]$$

Theorem

[Gribonval and Nielsen, 2002][Donoho and Elad, 2003]

Let A be an $m \times N$ matrix, and let $x^0 \in \mathbb{R}^N \setminus 0$ has sparsity s

$$s < \frac{1}{2} \left(1 + \mu(A)^{-1} \right)$$

Then x^0 is the unique solution of

$\min \|x\|_0 \quad$ s.t. $y = Ax \quad$ and $\quad \min \|x\|_1 \quad$ s.t. $y = Ax$

Analysis of Coherence

- Choosing a matrix $A \in \mathbb{C}^{m \times N}$ with small coherence $\mu \leqslant c/\sqrt{m}$ the condition

  $$(2s-1)\mu < 1$$

  ensures recovery of s-sparse vectors , which means the number of measurements should satisfy:

  $$m \geqslant Cs^2$$

- For mildly large s, the estimate can be too pessimistic.
- RIP is better with less measurements.

Outline

Restricted Isometry Property [Candes and Tao, 2004]

Definition

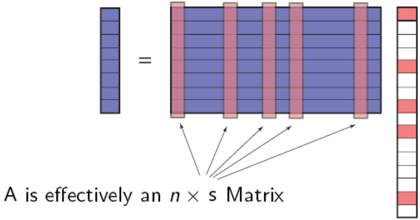
Let A be an $n \times N$ matrix. Then A has the Restricted Isometry Property (RIP) of order s, if their exists $\delta_s \in (0, 1)$ with

$$(1 - \delta_s) \|x\|_2^2 \leqslant \|Ax\|_2^2 \leqslant (1 + \delta_s) \|x\|_2^2 \quad \forall x \text{ with sparsity } s$$

Restricted Isometry Property

Key Idea 1: Sparsity

- Our signal is s-sparse:

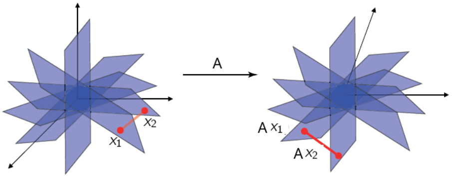


- Design A so that each of its $n \times s$ submatrices is full rank!

Restricted Isometry Property

Key Idea 2: Stable Embedding:

- *A* shall preserve the geometry of the set of sparse signals:



- Restricted Isometry Property:

$$\|x_1 - x_2\| \approx \|Ax_1 - Ax_2\|$$

But this is a combinatorial NP-hard design problem [Tillmann and Pfetsch, 2013]!

Recovery with RIP

Theorem [Cai et al., 2009]

Suppose that the 2sth restricted isometry constant of the matrix $A \in \mathbb{C}^{m \times N}$ satisfies

$$\delta_{2s} < \frac{1}{3}$$

Then every s-sparse vector $x \in \mathbb{C}^N$ is the unique solution of

$$\min_z \|z\|_1 \quad \text{s.t. } Az = Ax$$

Greedy and thresholding algorithms have similar results!

Recovery with RIP

Theorem [Candes et al., 2006]

Suppose that the $2s$th restricted isometry constant of the matrix $A \in \mathbb{C}^{m\times N}$ satisfies

$$\delta_{2s} < \frac{4}{\sqrt{41}} \approx 0.6246$$

Then, for any $x \in \mathbb{C}^N$ and $y \in \mathbb{C}^m$ with $\|Ax - y\|_2 \leqslant \eta$, a solution $x^\#$ of

$$\min_z \|z\|_1 \quad \text{s.t } \|Az - y\|_2 \leqslant \eta$$

approximates the vector x with errors

$$\left\|x - x^\#\right\|_1 \leqslant C\sigma_s(x)_1 + D\sqrt{s}\eta$$

$$\left\|x - x^\#\right\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta$$

where the constants $C, D > 0$ depend only on δ_{2s}

What Kind of Matrix Hold RIP?

Cons of deterministic matrix

- Problem lies in estimating the eigenvalues of submatrix of A.
- The basic tool, namely Gershgorin's disk theorem is not accurate enough.

Pros of random matrix

- A powerful set of tool of probability theory becomes available

Random Matrix

General Approach to RIP

- Design A to be a random matrix, e.g.
  - Gaussian iid
  - Bernoulli (± 1) iid
  - Subgaussian iid
  - ...
- Subgaussian is more general:

$$\mathbb{P}\left(|A_{j,k}| \geqslant t\right) \leqslant \beta e^{-\kappa t^2}$$

- Such matrices A have the Restricted Isometry Property with high probability, if

$$m = O\left(s \cdot \log\left(\frac{N}{s}\right)\right) << N$$

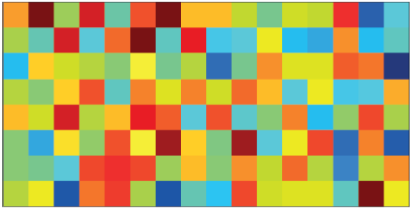
RIP for Subgaussian Random Matrix

Theorem [Mendelson et al., 2008]

Let A be an $m \times N$ subgaussian random matrix. Then there exists a constant $C > 0$ such that the restricted isometry constant of $\frac{1}{\sqrt{m}}A$ satisfies $\delta_s \leqslant \delta$ with probability at least $1 - \epsilon$ provided

$$m \geqslant C\delta^{-2}\left(s\ln(eN/s) + \ln\left(2\epsilon^{-1}\right)\right)$$

C depends only on the subgaussian parameters β, κ



Recovery with Subgaussian Random Matrix

Theorem

Let A be an $m \times N$ subgaussian random matrix. There exist constants $C_1, C_2 > 0$ only depending on the subgaussian parameters β, κ such that if, for $\epsilon \in (0, 1)$,

$$m \geqslant C_1 s \ln(eN/s) + C_2 \ln\left(2\epsilon^{-1}\right)$$

then with probability at least $1 - \epsilon$ every s-sparse vector $x \in \mathbb{C}^N$ is the unique solution of

$$\min_z \|z\|_1 \quad \text{s.t. } Az = Ax$$

Subgaussian Matrix $\Longrightarrow$ RIP $\Longrightarrow$ Perfect Recovery

Phase Transition

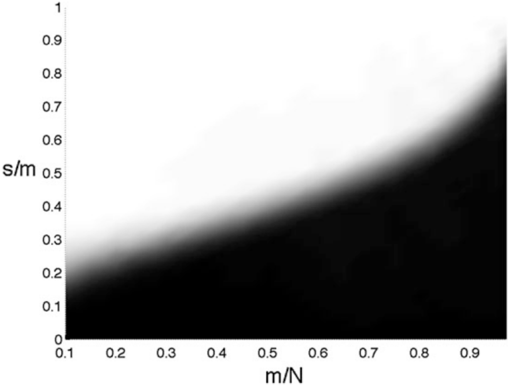


Figure: Black corresponds to 100% empirical success probability, white to 0% empirical success probability

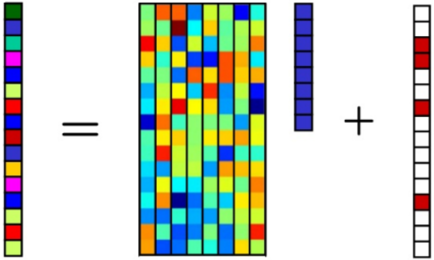
Outline

Outline

Error Correction [Candes and Tao, 2005]

Suppose errors happen occasionally while transmitting a signal...

$$y \quad = \quad B \quad x \quad + \quad e$$



How can we recover x from y?

Error Correction

- Construct a matrix A such that $AB = 0$.
- Let $y' = Ay$, then

$$y' = Ay = ABx + Ae = Ae$$

- Reconstruct e with compressed sensing.
- Acquire x by solving redundant equation:

$$y - e = Bx$$

Cryptology and Error Correction

Case 1

- Jerry had a secret x and sent Tom the cryptograph $y = Bx$.
- Tom understand y because he knows B. Then he took action.
- After a while, B was decoded according to y and Tom's action.

Case 2

- Jerry had a secret x and sent Tom the cryptograph $y = Bx + e$. e is a sparse error.
- Tom understand y because he knows B and compressed sensing. Then he took action.
- B is hard to decode because e is random.

Outline

From Vector to Matrix [Fazel, 2002]

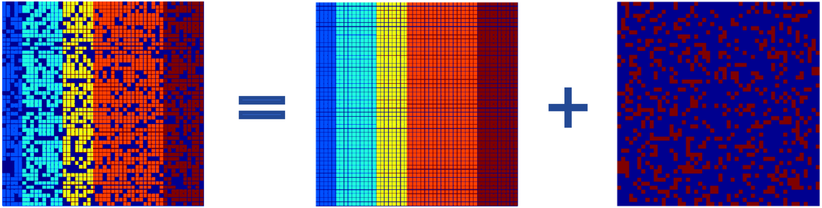
	Sparse Vector	*Low-Rank Matrix*
Low-dimensionality of	individual signal	correlated signals
Measure	L_0 norm $\|\|x\|\|_0$	$\mathrm{rank}(X)$
Convex Surrogate	L_1 norm $\|\|x\|\|_1$	Nuclear norm $\|\|X\|\|_*$
Compressed Sensing	$y = Ax$	$Y = A(X)$

In data science, matrix is more general than vector.

Robust PCA [Candès et al., 2011]

- Basic assumption:

X-observation $\qquad$ L-low rank $\qquad$ S-sparse



- Optimization formula:

$$\min \|L\|_* + \lambda \|S\|_1 \quad \text{s.t. } L + S = X$$

Semidefinite program, solvable in polynomial time.

Background Modeling from Video

Video X = Low-rank L + Sparse error S

Outline

Matrix Completion [Candès and Recht, 2009]

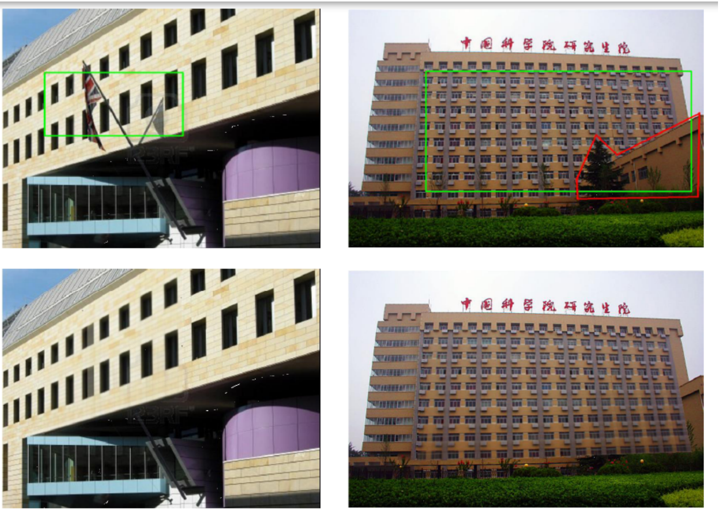
- Assumption: true data $\bar{X}$ is low-rank, but part of it is unknown.

- Optimization formula:

$$\min \|\bar{X}\|_* + \lambda\|E\|_1 \quad \text{s.t. } P_\Omega[\bar{X} + E] = X$$

$$X \subset \mathbb{R}^{m\times n} \quad \Omega \subset [m] \times [n]$$

Matrix Completion



Outline

Outline

Bayesian Compressed Sensing [Ji et al., 2008]

- Signal x is the sum of a sparse vector with large magnitude and a dense vector with small magnitude.

$$x = x_s + x_e$$

- Then

$$y = Ax = Ax_s + Ax_e = Ax_s + n$$

  - Central-Limit Theorem
  - n can be approximated by iid Gaussian
  - varaince σ^2 is unknown

- Then the Gaussian likelihood function is

$$p\left(y|x_s, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-K/2} \exp\left(-\frac{1}{2\sigma^2}\left\|y - Ax_s\right\|^2\right)$$

How to Impose Sparse on Signal x

- First define a zero-mean Gaussian prior on each element of x

$$p(x|\alpha) = \prod_{i=1}^{N} \mathcal{N}\left(x_i|0, \alpha_i^{-1}\right)$$

- Then a Gamma prior is considered over α

$$p(\alpha|a, b) = \prod_{i=1}^{N} \Gamma\left(\alpha_i|a, b\right)$$

- The overall prior on x is

$$p(x|a, b) = \prod_{i=1}^{N} \int_0^{\infty} \mathcal{N}\left(x_i|0, \alpha_i^{-1}\right) \Gamma\left(\alpha_i|a, b\right) d\alpha_i$$

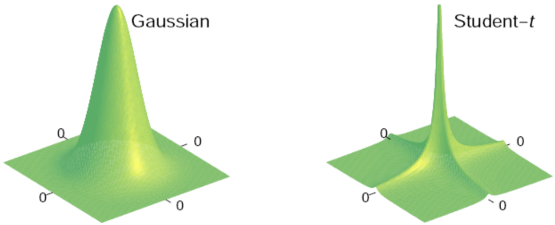
which can be evaluated analytically.

How to Impose Sparse on Signal x

Given a and b,

$$p(x_i) = \int p(x_i|\alpha_i)\, p(\alpha_i)\, d\alpha_i$$
$$= \frac{b^a \Gamma\left(a + \frac{1}{2}\right)}{(2\pi)^{\frac{1}{2}} \Gamma(a)} \left(b + x_i^2/2\right)^{-\left(a+\frac{1}{2}\right)}$$

which corresponds to the density of a Student-t distribution.



Encouraging sparsity!

Bayesian Compressed Sensing

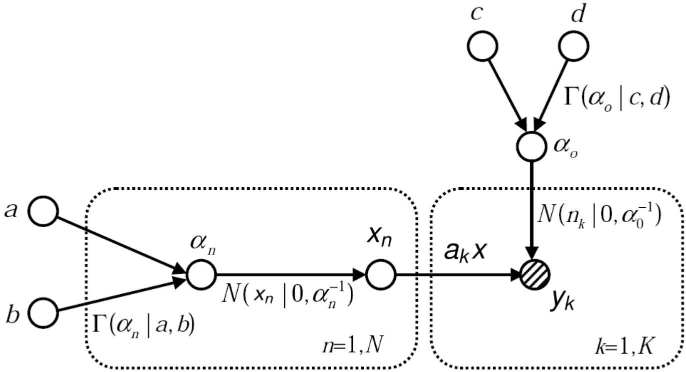


Figure: Graphical model of the Bayesian CS formulation

Bayesian Compressed Sensing

- Since the prior on x is conjugate, the posterior for x can be performed in a closed form.

$$
\begin{aligned}
\mu &= \alpha_0 \Sigma A^T y \\
\Sigma &= \left(\alpha_0 A^T A + W\right)^{-1} \\
W &= \operatorname{diag}\left(\alpha_1, \alpha_2, \ldots, \alpha_N\right)
\end{aligned}
$$

- α can be estimated via EM algorithm.
- One iterates two steps until a convergence criterion has been satisfied.

Conclusion

- Sparser than existing CS solutions.
- Computation time comparable to the state-of-the-art greedy algorithms (StOMP).
- BCS is an important complement to conventional CS formulation.

	BP	CFDR	CFAR	BCS
#Nonzeros	3840	1766	926	751
Time (secs)	162	10	28	15
Reconst. Error	0.1416	0.1826	0.1508	0.1498

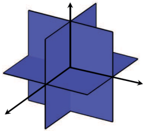
Table: Summary of the performance of Basis Pursuit, StOMP (CFDR and CFAR) and BCS

Outline

Deep Compressed Sensing [Wu et al., 2019]

Compressed Sensing

- Basic assumption: sparsity



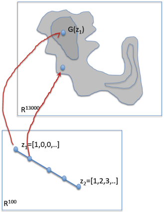
- optimization:

$$\min_x \|x\|_1 \quad \text{s.t. } Ax = y$$

- Convex

GAN

- Basic assumption: low-dim manifold



- Adversarial training
- Highly nonconvex

What will happen if we relax sparsity to low-dim manifold?

Deep Compressed Sensing

- Assume $x = G_\theta(z)$.
- The reconstruction problem similar to CS is :

$$\hat{z} = \arg\min_{z} E_\theta(m, z)$$

$$E_\theta = \|y - AG_\theta(z)\|_2^2$$

- Highly non-convex, optimization can be hard.
- Meta Learning can accelerate the optimization with only 3-5 gradient descent steps.

Deep Compressed Sensing

- If the generator maps all $G_\theta(z)$ into the null space of A, the reconstruction fail.
- The solution is to enforce the RIP via regularizing θ

$$\mathcal{L}_F = \mathbb{E}_{x_1,x_2}\left[\left(\|A(x_1 - x_2)\|_2 - \|x_1 - x_2\|_2\right)^2\right]$$

- Further, $A = A_\phi$ can be deep neural network to obtain better measurements.

Algorithm Scheme

Algorithm 2 Deep Compressed Sensing

Input: minibatchs of data $\{\boldsymbol{x}_i\}_{i=1}^N$, measurement function F_ϕ, generator G_θ, learning rate α, number of latent optimisation steps T

repeat

  Initialize generator parameters θ

  for $i = 1$ **to** N **do**

    Measure the signal $\boldsymbol{m}_i \leftarrow F_\phi(\boldsymbol{x}_i)$

    Sample $\hat{\boldsymbol{z}}_i \sim p_{\boldsymbol{z}}(\boldsymbol{z})$

    for $t = 1$ **to** T **do**

      Optimise $\hat{\boldsymbol{z}}_i \leftarrow \hat{\boldsymbol{z}}_i - \frac{\partial}{\partial \boldsymbol{z}} E_\theta(\boldsymbol{m}_i, \hat{\boldsymbol{z}}_i)$

    end for

  end for

  $\mathcal{L}_G = \frac{1}{N}\sum_{i=1}^N E_\theta(\boldsymbol{m}_i, \hat{\boldsymbol{z}}_i)$

  Compute $\mathcal{L}_F$ using eq. 12

  Option 1 : joint update $\theta \leftarrow \theta - \frac{\partial}{\partial\theta}(\mathcal{L}_G + \mathcal{L}_F)$

  Option 2 : alternating update

$$\theta \leftarrow \theta - \frac{\partial}{\partial\theta}\mathcal{L}_G \qquad \phi \leftarrow \phi - \frac{\partial}{\partial\phi}\mathcal{L}_F$$

until reaches the maximum training steps

Experiments

From sparsity to low-dimensional manifold,
does more knowledge mean less measurements needed?
Yes!!

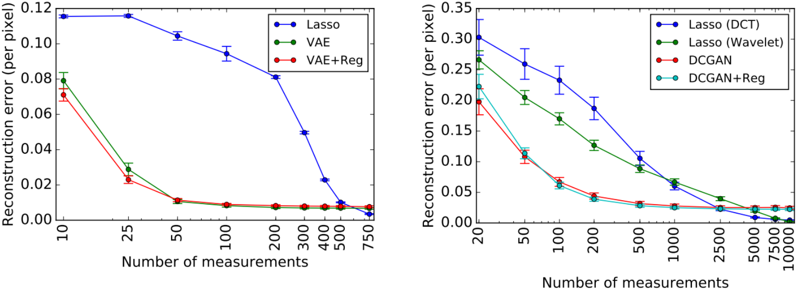


Figure: Left: Results on MNIST. Right: Results on celebA

Experiments

MODEL	10	25 MEASUREMENTS	STEPS
BASELINE	54.8	17.2	10×1000
LINEAR	10.8 ± 3.8	6.9 ± 2.7	3
NN	12.5 ± 2.2	10.2 ± 1.7	3
LINEAR(L)	6.5 ± 2.1	$4. \pm 1.4$	3
NN(L)	$\mathbf{5.3 \pm 1.9}$	$\mathbf{3.4 \pm 1.2}$	3

Figure: Reconstruction loss on MNIST test data using different measurement functions. (L) indicates learned measurement functions

MODEL	20	50 MEASUREMENTS	STEPS
BASELINE	156.8	82.3	2×500
LINEAR	34.7 ± 7.9	27.1 ± 6.1	3
NN	46.1 ± 8.9	41.2 ± 8.3	3
LINEAR(L)	26.2 ± 5.9	20.5 ± 4.3	3
NN(L)	$\mathbf{23.4 \pm 5.8}$	$\mathbf{18.5 \pm 4.3}$	3

Figure: Reconstruction loss on CelebA test data using different measurement functions.

Conclusion

- GAN can represent data distributions more concisely than standard sparsity models.
- DCS method can use 5-10x fewer measurements than conventional methods.
- After relatively few measurements, the signal reconstruction gets close to the optimal within the range of the generator.

References I

Barron, A. R., Cohen, A., Dahmen, W., DeVore, R. A., et al. (2008).
Approximation and learning by greedy algorithms.
The annals of statistics, 36(1):64–94.

Blumensath, T. and Davies, M. E. (2008a).
Gradient pursuits.
IEEE Transactions on Signal Processing, 56(6):2370–2382.

Blumensath, T. and Davies, M. E. (2008b).
Iterative thresholding for sparse approximations.
Journal of Fourier analysis and Applications, 14(5-6):629–654.

Blumensath, T. and Davies, M. E. (2009).
Stagewise weak gradient pursuits.
IEEE Transactions on Signal Processing, 57(11):4333–4346.

Cai, T. T., Wang, L., and Xu, G. (2009).
Shifting inequality and recovery of sparse signals.
IEEE Transactions on Signal Processing, 58(3):1300–1308.

Candes, E. and Tao, T. (2004).
Near optimal signal recovery from random projections: Universal encoding strategies?
arXiv preprint math/0410542.

References II

Candes, E. and Tao, T. (2005).
Decoding by linear programming.
arXiv preprint math/0502327.

Candès, E. J., Li, X., Ma, Y., and Wright, J. (2011).
Robust principal component analysis?
Journal of the ACM (JACM), 58(3):11.

Candès, E. J. and Recht, B. (2009).
Exact matrix completion via convex optimization.
Foundations of Computational mathematics, 9(6):717.

Candes, E. J., Romberg, J. K., and Tao, T. (2006).
Stable signal recovery from incomplete and inaccurate measurements.
Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 59(8):1207–1223.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001).
Atomic decomposition by basis pursuit.
SIAM review, 43(1):129–159.

References III

Daubechies, I., DeVore, R., Fornasier, M., and Güntürk, C. S. (2010).
Iteratively reweighted least squares minimization for sparse recovery.
Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 63(1):1–38.

Donoho, D. L., Drori, I., Tsaig, Y., and Starck, J.-L. (2006).
Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit.
Department of Statistics, Stanford University.

Donoho, D. L. and Elad, M. (2003).
Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization.
Proceedings of the National Academy of Sciences, 100(5):2197–2202.

Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., and Baraniuk, R. G. (2008).
Single-pixel imaging via compressive sampling.
IEEE signal processing magazine, 25(2):83–91.

Fazel, M. (2002).
Matrix rank minimization with applications.

Foucart, S. (2011).
Hard thresholding pursuit: an algorithm for compressive sensing.
SIAM Journal on Numerical Analysis, 49(6):2543–2563.

References IV

Gribonval, R. and Nielsen, M. (2002).
Sparse representations in unions of bases.

Haldar, J. P., Hernando, D., and Liang, Z.-P. (2010).
Compressed-sensing mri with random encoding.
IEEE transactions on Medical Imaging, 30(4):893–903.

Ji, S., Xue, Y., Carin, L., et al. (2008).
Bayesian compressive sensing.
IEEE Transactions on signal processing, 56(6):2346.

Mallat, S. G. and Zhang, Z. (1993).
Matching pursuits with time-frequency dictionaries.
IEEE Transactions on signal processing, 41(12):3397–3415.

Mendelson, S., Pajor, A., and Tomczak-Jaegermann, N. (2008).
Uniform uncertainty principle for bernoulli and subgaussian ensembles.
Constructive Approximation, 28(3):277–289.

Needell, D. and Tropp, J. A. (2009).
Cosamp: Iterative signal recovery from incomplete and inaccurate samples.
Applied and computational harmonic analysis, 26(3):301–321.

References V

Tibshirani, R. (1996).
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society: Series B (Methodological), 58(1):267–288.

Tillmann, A. M. and Pfetsch, M. E. (2013).
The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing.
IEEE Transactions on Information Theory, 60(2):1248–1259.

Tropp, J. A. (2004).
Greed is good: Algorithmic results for sparse approximation.
IEEE Transactions on Information theory, 50(10):2231–2242.

Wu, Y., Rosca, M., and Lillicrap, T. P. (2019).
Deep compressed sensing.
CoRR, abs/1905.06723.

Xu, Z., Zhang, H., Wang, Y., Chang, X., and Liang, Y. (2010).
L 1/2 regularization.
Science China Information Sciences, 53(6):1159–1169.