Compressed Sensing

presented by

Anamitra Makur
School of Electrical & Electronic Engineering
Nanyang Technological University
Singapore

SPCOM 2016
June 12
Bangalore

Sections

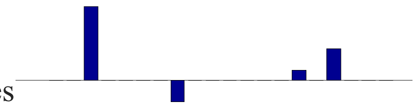
- Introduction
- Analysis of l_1 minimization
- Measurement matrix
- Reconstruction algorithms
- Applications
- Conclusion

Section 1. INTRODUCTION

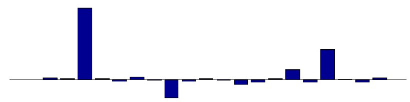
- Introduction to sparse signal
- Compressed sensing versus conventional sampling
- The idea of compressed sensing
- Reconstruction by l_0 minimization
- Reconstruction by l_1 minimization

Sparse signal

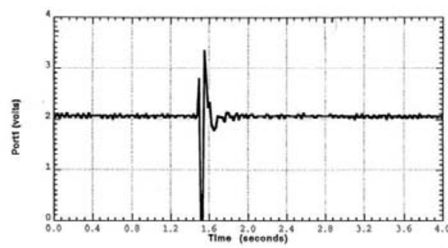
Sparse = mostly zero
Sparsity = no of non-zero values



Approximately/nearly sparse
= mostly insignificant values



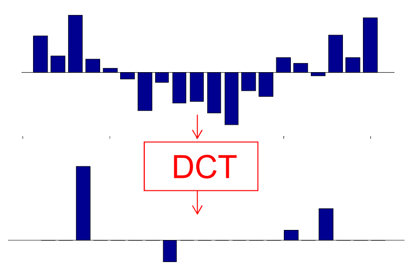
Real-life example:
seismograph (we are lucky
earthquakes are sparse)



Sparse signal

Most real-life signals
do not look sparse

They may be sparse
under appropriate basis

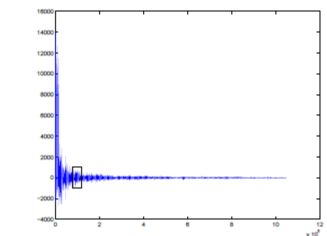


These are implicitly sparse signals

Sparse signal

Real-life example: image

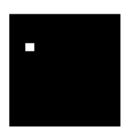
wavelet



CS versus sampling

CS = compressed/compressive sensing/sampling

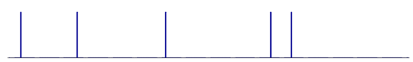
Conventional sampling: local/sparse

image sample/pixel = ⟨  ,  ⟩

Assume each measurement has a cost
→ Use minimum number of measurements
Is local measurement a good choice for sparse signals?

CS versus sampling

Real-life example: US army during WW-II, syphilis test
(0 = negative, 1 = positive)

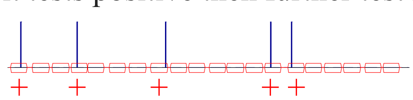


Syphilis blood test was expensive.
Random individual test:

→ all negative!

CS versus sampling

US army used group testing, pooling blood from a few soldiers.
If it tests positive then further test may be done.

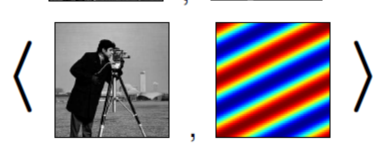


→ captures positive cases

Such global measurements exist for image:

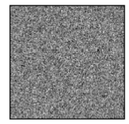
In tomography, measurement = (line integral)

In MRI, measurement = (sinusoids)



CS versus sampling

For sparse signals, global/dense measurements are better, as long as the signal may be reconstructed back from its measurements

In CS, measurement = (random) $\Big\langle$  ,  $\Big\rangle$

CS guarantees that a sparse signal may be reconstructed back from its measurements

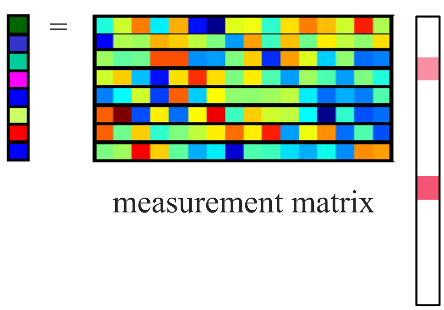
Further, CS requires less measurements/samples than conventional sampling for sparse signals

Compressed sensing

x = explicitly/implicitly sparse signal of dimension d

x has sparsity s

Take m random measurements y (m measurements)



$$y_{m\times 1} = A_{m\times d} \, . \, x_{d\times 1}$$

Compressed sensing

Few measurements: $m << d$
Can x be reconstructed from y?

Example: $m = 4$, $d = 6$

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 2 \\ 0 \end{bmatrix}$$

To find x, consider solving

$$x_0 + x_2 + x_5 = 0$$
$$x_1 + x_2 + x_3 = 2$$
$$\vdots$$

More unknowns than equations
→ Infinite solutions!
Typical of inverse problems

Compressed sensing

Some solutions:

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \\ 4/3 \\ 1/3 \\ -2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 2 \\ 0 \end{bmatrix}$$

l_0 norm	l_1 norm	l_2 norm
1	2	2
3	6	3.5
4	4	2
5	10	4.5
6	3.3	1.6

$$\|x\|_0 = \sum_i x_i^0 \qquad \|x\|_1 = \sum_i |x_i|$$

$$\|x\|_2 = \sqrt{\sum_i x_i^2}$$

In order to choose from infinite solutions, use some criterion such as the l_0 norm (sparsity)

Compressed sensing

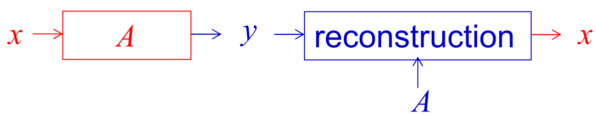
Under certain conditions, an explicitly sparse signal x of sparsity s and dimension d may be reconstructed from m measurements

Sensing: given signal x and measurement matrix A,

$$y = A.x$$

Reconstruction: given measurement y and measurement matrix A,

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



Compressed sensing

For nearly sparse u, let x be the most significant s-sparse part of u

Sensing: given signal u and measurement matrix A,

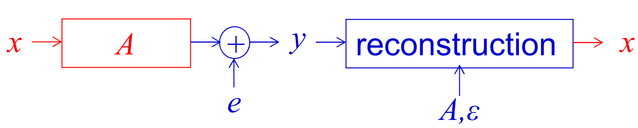
$$y = A.u = A.x + A(u - x) = A.x + e$$

where let $\|e\|_2 \leq \varepsilon$

Also for noisy measurement y, sensing is $A.x$ but $y = A.x + e$

Reconstruction: given y, A, and ε

$$\min_x \|x\|_0 \quad \text{subject to} \quad \|A.x - y\|_2 \leq \varepsilon$$



Compressed sensing

For implicitly sparse x in a basis Ψ such that $x = \Psi.u$ where u is sparse:

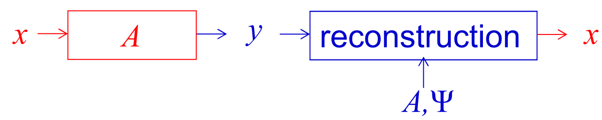
Sensing: $y = A.x$

Reconstruction: $\min_x \|\Psi^{-1}.x\|_0$ subject to $A.x = y$

or, use the explicit reconstruction with a modified measurement matrix,

$$\min_u \|u\|_0 \quad \text{subject to} \quad A\Psi.u = y$$

such that the solution is $x = \Psi.u$



16 Anamitra Makur NANYANG TECHNOLOGICAL UNIVERSITY

l_0 minimization

Recovery using l_0 minimization:

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

Example:
Find x where
$\|x\|_0 = 2$

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 3 \\ 5 \end{bmatrix} \qquad x = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ 3 \\ 0 \end{bmatrix}$$

(a sophisticated version of US army group testing)

17 Anamitra Makur NANYANG TECHNOLOGICAL UNIVERSITY

l_0 minimization

Can there be solutions other than x?

Same A:
Find x where
$\|x\|_0 = 2$

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} \qquad x = \begin{bmatrix} 3 \\ 3 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \ldots$$

l_0 minimization

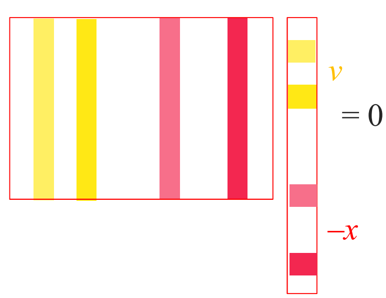
Theorem: Unique recovery using l_0 minimization is possible if spark of $A > 2s$.

spark of A = smallest number of linearly dependent columns of A

Proof: Let there be another solution v such that

$A.v = y, \quad \|v\|_0 \leq s$

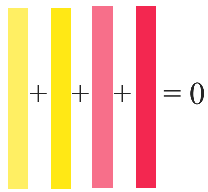
Then $A.v - A.x = A.(v - x) = 0$



s = sparsity

l_0 minimization

Or weighted sum of $\|v-x\|_0$ columns of A is zero



Or $\|v-x\|_0$ columns of A are linearly dependent

But $\|v\|_0 \le s, \|x\|_0 = s$

So even if v and x has entirely different non-zero locations, $\|v-x\|_0 \le 2s$ which is not possible, since spark of $A > 2s$ means any $2s$ or less columns are linearly independent

l_0 minimization

Check our example:

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}, s = 2$$

Some set of $2s = 4$ columns are not linearly independent, such as

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = 0$$

l_0 minimization

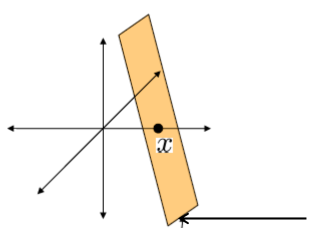
Good news: recovery is guaranteed using l_0 minimization

Bad news: l_0 minimization is essentially combinatorial
(try out all possible combinations)
Therefore complexity is high (NP-hard)

l_1 minimization

Convex relaxation: relax highly non-convex l_0 minimization to convex l_1 minimization

$$\min_x \|x\|_0 \quad \text{subject to} \quad A.x = y$$
$$\downarrow$$
$$\min_x \|x\|_1 \quad \text{subject to} \quad A.x = y$$

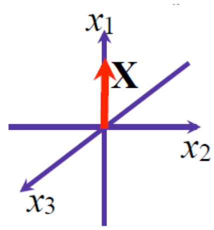


All solutions lie in a $d-m$ dimensional null space plane $\{v : A.v = y\}$

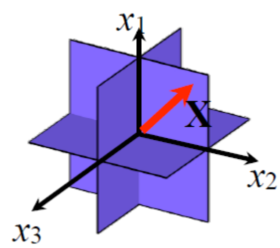
solution = where the null space plane intersects the l_* ball
(region with l_* norm = constant)

l_1 minimization

l_0 'ball' is infinite, looks like

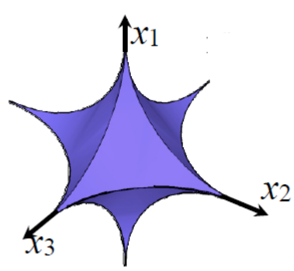


l_0 norm = 1
highly non-convex

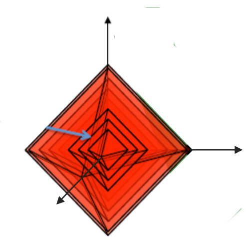


l_0 norm = 2
highly non-convex

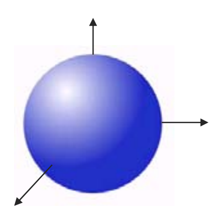
l_1 minimization



l_p ball for $p < 1$
non-convex

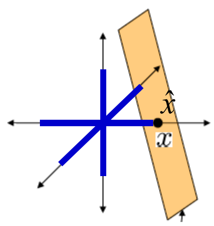


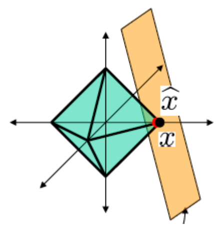
l_1 ball
convex

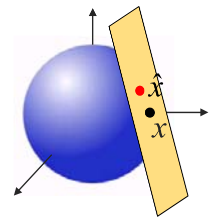


l_2 ball
convex

l_1 minimization







l_0 'ball:'
correct solution
if $m \geq 2s$

l_1 ball:
correct solution
under certain
conditions

l_2 ball:
wrong solution,
typically not
sparse

m = measurements, s = sparsity

Section 2. ANALYSIS OF l_1 MINIMIZATION

- l_1 minimization: theorem and sketch of proof
- Proof part 1: tail probabilities
- Proof part 2: number of measurements
- Proof part 3: RIP constant
- Proof part 4: null space property

Analysis of l_1 minimization

Recovery using l_1 minimization:

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

Theorem: For a random Gaussian matrix A, choosing $m \geq c_5 s \log(d/s)$, recovery using l_1 minimization occurs with probability $1 - e^{-c_3 H.d}$.
(c_i = constants > 0, $H = H(2s/d)$ = entropy)

(Proofs using JL lemma show with probability $1 - e^{-c_6 m}$)

d = dimension, m = measurements, s = sparsity

Analysis of l_1 minimization

Non-uniform recovery result: For a given x, choose A at random. Show recovery in probability for a constraint on m.

Easier to show. For one x, some A may not work. For another x, some other A doesn't work. It is possible that all A are such that they don't work for some x.

Uniform recovery result: For any x, choose A at random. Show recovery in probability for a constraint on m.

For all x, a random A works with certain probability.

Analysis of l_1 minimization

Sketch of proof: random measurement matrix A

probabilistic

Proof 1	tail probabilities of singular values of sub-matrix of A to RIP constant

$$p\left\{\sqrt{1+\delta_{2s}} \geq 1+\lambda\right\} \leq e^{-c_3 H.d}$$

$$\sqrt{1+\delta_{2s}} < 1+\lambda$$

deterministic

Proof 2	required m so that $\lambda \leq (2-\sqrt{3})/\sqrt{3}$

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

Analysis of l_1 minimization

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

d e t e r m i n i s t i c

Proof 3	RIP constant to null space property

A satisfies null space property

Proof 4	null space property means recovery

recovery

Analysis of l_1 minimization: Proof 1

Proof 1: random measurement matrix A

Proof 1	tail probabilities of singular values of sub-matrix of A to RIP constant

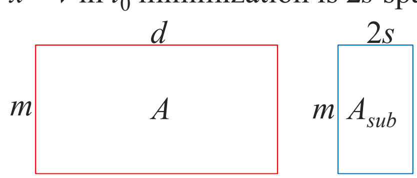
$$p\left\{\sqrt{1+\delta_{2s}} \ge 1+\lambda\right\} \le e^{-c_3 H.d}$$

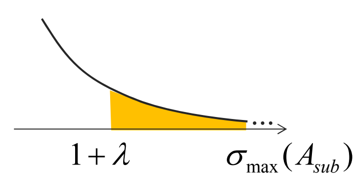
This part finds upper bound on the probability that the RIP constant is much larger than 1.

Analysis of l_1 minimization: Proof 1

Random Gaussian matrix = elements of A iid Gaussian

A_{sub} = some column sub-matrix of A made of $2s$ columns (recall $x - v$ in l_0 minimization is $2s$ sparse), singular values ≈ 1





Applying concentration inequality for Gaussian measures,

$$p\left\{\sigma_{\max}(A_{sub}) > 1 + \underbrace{\sqrt{\frac{2s}{m}} + t + ..}_{\lambda}\right\} \le e^{-mt^2/2}$$

Analysis of l_1 minimization: Proof 1

Let $\lambda = \sqrt{\frac{2s}{m}} + t = \sqrt{\frac{2s}{m}} + \sqrt{(2+2c_1)\frac{H.d}{m}}$

where $H = H\left(\frac{2s}{d}\right) = -\frac{2s}{d}\log\frac{2s}{d} - \frac{d-2s}{d}\log\frac{d-2s}{d}$

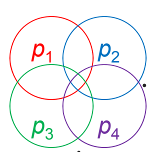
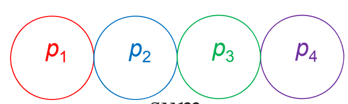
then $p\{\sigma_{\max}(A_{sub}) > 1+\lambda\} \le e^{-mt^2/2} = e^{-(1+c_1)H.d}$

Let $\overline{\sigma} = \max_{A_{sub}} \sigma_{\max}(A_{sub})$, the maximum singular value among all possible A_{sub}

Even if any one $\sigma_{\max}(A_{sub}) > 1+\lambda$, then $\overline{\sigma} > 1+\lambda$

So $p\{\overline{\sigma} > 1+\lambda\} = p\{\text{union of all } \sigma_{\max}(A_{sub}) > 1+\lambda\}$

Analysis of l_1 minimization: Proof 1

$p\{\overline{\sigma} > 1+\lambda\} =$  ... $\le$  ...

There are $\binom{d}{2s}$ different A_{sub}, all having identical probability

$$p\{\overline{\sigma} > 1+\lambda\} \le \binom{d}{2s} p\{\sigma_{\max}(A_{sub}) > 1+\lambda\}$$
$$\le \binom{d}{2s} e^{-(1+c_1)H.d}$$

Analysis of l_1 minimization: Proof 1

Now $\log\binom{d}{2s} = \log d! - \log 2s! - \log(d-2s)!$

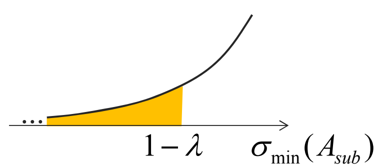
Using Stirling approximation
$\log d! \approx d\log d - d + ..$

$$= -2s\log\frac{2s}{d} - (d-2s)\log\frac{d-2s}{d} + ..$$
$$= H.d + ..$$

$$p\{\overline{\sigma} > 1+\lambda\} \le e^{H.d+..}.e^{-(1+c_1)H.d}$$
$$= e^{-c_1 H.d+..}$$
$$\le e^{-c_2 H.d}$$

Analysis of l_1 minimization: Proof 1

Similarly, for minimum singular value of A_{sub}



$$\underline{\sigma} = \min_{A_{sub}} \sigma_{\min}(A_{sub})$$

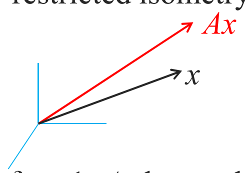
It may be shown that $p\{\underline{\sigma} < 1-\lambda\} \le e^{-c_2 H.d}$

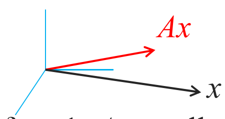
Analysis of l_1 minimization: Proof 1

Restricted isometry property (RIP) of order s:
For a measurement matrix A, the l_2 norm of Ax for all s-sparse signal x is bounded by

$$(1-\delta_s)\,\|x\|_2^2 \;\le\; \|Ax\|_2^2 \;\le\; (1+\delta_s)\,\|x\|_2^2$$

δ_s = restricted isometry constant, or simply RIP constant





$1 + \delta_s > 1$, Ax larger than x $\quad$ $1 - \delta_s < 1$, Ax smaller than x

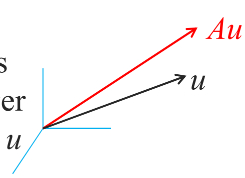
Small δ_s means Ax is nearly equal to x (in l_2 norm sense)

Analysis of l_1 minimization: Proof 1

Consider RIP for $2s$-sparse signal u

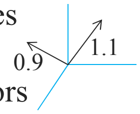
$$(1-\delta_{2s})\,\|u\|_2^2 \;\le\; \|Au\|_2^2 \;\le\; (1+\delta_{2s})\,\|u\|_2^2$$

$1+\delta_{2s}$ measures how much larger can Au be than u

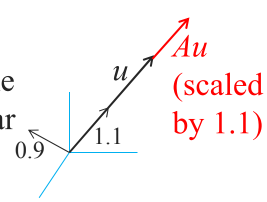


u in which direction results in largest Au ?

singular values and right-singular vectors of A (≈ 1)



direction of the largest singular value



Analysis of l_1 minimization: Proof 1

$\overline{\sigma}$ is the maximum singular value among all possible A_{sub}

$$(1-\delta_{2s})\,\|u_{sub}\|_2^2 \;\le\; \|A_{sub}u_{sub}\|_2^2 \;\le\; (1+\delta_{2s})\,\|u_{sub}\|_2^2$$

$$\text{Since } \|A_{sub}u_{sub}\|_2^2 \le \overline{\sigma}^2\,\|u_{sub}\|_2^2\,,\quad \sqrt{1+\delta_{2s}} \overset{?}{=} \overline{\sigma}$$

We have to check lower bound, too

$$\text{Since } \underline{\sigma}^2\,\|u_{sub}\|_2^2 \le \|A_{sub}u_{sub}\|_2^2\,,\quad \sqrt{1-\delta_{2s}} \overset{?}{=} \underline{\sigma}$$

$$\text{or}\quad \sqrt{1+\delta_{2s}} \overset{?}{=} \sqrt{2-\underline{\sigma}^2}$$

Combining, $\sqrt{1+\delta_{2s}} = \max\left\{\overline{\sigma}, \sqrt{2-\underline{\sigma}^2}\right\}$

Analysis of l_1 minimization: Proof 1

If either $\overline{\sigma} > 1+\lambda$ or $\sqrt{2-\underline{\sigma}^2} > 1+\lambda$, then $\sqrt{1+\delta_{2s}} > 1+\lambda$

$p\left\{\sqrt{1+\delta_{2s}} > 1+\lambda\right\} =$ $\overline{\sigma}$ $\sqrt{2-\underline{\sigma}^2}$ $\le$ $\overline{\sigma}$ $\sqrt{2-\underline{\sigma}^2}$

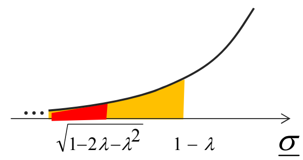
union sum

So $p\left\{\sqrt{1+\delta_{2s}} > 1+\lambda\right\} \le p\{\overline{\sigma} > 1+\lambda\} + p\left\{\sqrt{2-\underline{\sigma}^2} > 1+\lambda\right\}$

Analysis of l_1 minimization: Proof 1

Now $p\left\{\sqrt{2-\underline{\sigma}^2} > 1+\lambda\right\}$

$$= p\left\{\underline{\sigma} < \sqrt{1-2\lambda-\lambda^2}\right\}$$
$$< p\left\{\underline{\sigma} < \sqrt{1-2\lambda+\lambda^2}\right\}$$
$$= p\left\{\underline{\sigma} < 1-\lambda\right\}$$



So $p\left\{\sqrt{1+\delta_{2s}} > 1+\lambda\right\} < p\left\{\overline{\sigma} > 1+\lambda\right\} + p\left\{\underline{\sigma} < 1-\lambda\right\}$

$$\le 2e^{-c_2 H.d}$$
$$\le e^{-c_3 H.d}$$

Analysis of l_1 minimization: Proof 2

Proof 2: $\sqrt{1+\delta_{2s}} < 1+\lambda$

Proof 2 required m so that $\lambda \le (2-\sqrt{3})/\sqrt{3}$

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

This part finds the required m such that the RIP constant of A is small.

Analysis of l_1 minimization: Proof 2

As we see later in proof part 3, recovery requires $\delta_{2s} < ⅓$

Or, probability that $\delta_{2s} \geq ⅓$ should be very small

$$\text{If } \lambda \leq \frac{2-\sqrt{3}}{\sqrt{3}} \text{ then } p\left\{\delta_{2s} \geq \frac{1}{3}\right\} = p\left\{\sqrt{1+\delta_{2s}} \geq \frac{2}{\sqrt{3}}\right\}$$
$$\leq p\left\{\sqrt{1+\delta_{2s}} \geq 1+\lambda\right\}$$
$$\leq e^{-c_3 H.d}$$
$$\text{or } \quad p\left\{\delta_{2s} < \frac{1}{3}\right\} \geq 1 - e^{-c_3 H.d}$$

Analysis of l_1 minimization: Proof 2

From proof part 1, $\lambda = \sqrt{\frac{2s}{m}} + \sqrt{(2+2c_1)\frac{H.d}{m}}$

$$\lambda^2 m = 2s + (2+2c_1)H.d + ..$$
$$\approx 2s + (2+2c_1)H.d$$
$$= 2s + (2+2c_1)\left(-2s\log\frac{2s}{d} - \underset{\downarrow}{(d-2s)\log\frac{d-2s}{d}}\right)$$

since $2s << d$, this term is smaller

$$\leq 2s + c_4\left(-2s\log\frac{2s}{d}\right)$$
$$= 2s + c_4\left(2s\log\frac{d}{2s}\right)$$

Analysis of l_1 minimization: Proof 2

$$= 2c_4 s \log\frac{d}{s} - (2c_4 \log 2 - 2)s \quad \le 2c_4 s \log\frac{d}{s}$$

Choose $m \ge c_5 s \log\frac{d}{s}$ such that $\lambda^2 m \le 2c_4 s \log\frac{d}{s} \le \frac{2c_4}{c_5} m$

or $\lambda \le \sqrt{\frac{2c_4}{c_5}}$

Recall that c_4 depends on c_1 .
Choose c_5 large enough to make $\sqrt{\frac{2c_4}{c_5}} \le \frac{2-\sqrt{3}}{\sqrt{3}}$

Therefore $\delta_{2s} < \frac{1}{3}$ with very high probability

Anamitra Makur NANYANG TECHNOLOGICAL UNIVERSITY

Analysis of l_1 minimization: Proof 3

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

Proof 3	RIP constant to null space property

A satisfies null space property

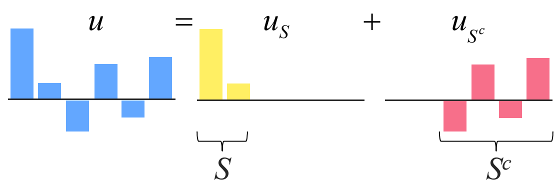
This part shows that a small enough RIP constant implies that null space vectors of A are not sparse (null space property).

Anamitra Makur NANYANG TECHNOLOGICAL UNIVERSITY

Analysis of l_1 minimization: Proof 3

Null space property of order s:

For all null space vectors u (such that $A.u = 0$),
for all subset S consisting of s components of u,
split u into parts u_S and u_{S^c}:

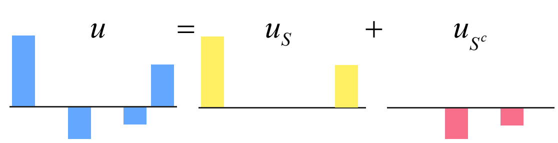


Then A satisfies null space property of order s if $\|u_S\|_1 < \|u_{S^c}\|_1$

Analysis of l_1 minimization: Proof 3

Null space property implies that u is not sparse

In the earlier example, if u is sparse, then S may be chosen such that $\|u_S\|_1 > \|u_{S^c}\|_1$:



The worst case S is choosing s (absolute-)largest components of u
Let this case be $u_S = u_0$

Analysis of l_1 minimization: Proof 3

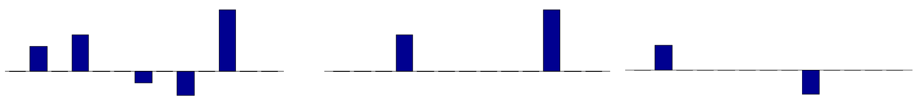
Split u into s-sparse parts from large to small components

$$u = u_0 + u_1 + u_2 + \cdots$$

s (absolute-)largest components

s+1 to 2s (absolute-)largest components

s=2 example: u = u_0 + u_1 ...



Analysis of l_1 minimization: Proof 3

How does the l_1 norm of a s-sparse vector compare to its l_2 norm?

Example: $v = [\ldots \underbrace{a..a\ldots a}_{s\ \text{'}a\text{' terms}} \ldots]$

l_1 norm = $s|a|$ but l_2 norm = $\sqrt{sa^2}$

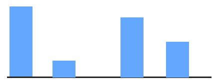
$\rightarrow$ l_1 norm is larger than l_2 norm

Scaling the l_2 norm by $\sqrt{s}$, $\|v\|_1 = \sqrt{s}\,\|v\|_2$

The next slide formally shows that in general, $\|v\|_1 \le \sqrt{s}\,\|v\|_2$

Analysis of l_1 minimization: Proof 3

For any s-sparse vector v with components $\{v\}_i$, take the inner product of $|v|$ =



with $w =$ [1 0 1 0 0 1 0 1 0]

Note that $<|v|, w> = \|v\|_1$, the l_1 norm of v

Also $<|v|, w>^2 = \left(\sum_{nonzero} |\{v\}_i|.1\right)^2 \le \left(\sum_{nonzero} \{v\}_i^2\right)\left(\sum_{nonzero} 1^2\right)$ from Cauchy-Schwartz inequality

l_2 norm of v

or $\|v\|_1^2 \le \|v\|_2^2\, s$ Therefore $\|u_0\|_1 \le \|u_0\|_2 \sqrt{s}$

Analysis of l_1 minimization: Proof 3

Since u_0 is s-sparse, using RIP result $(1-\delta_s)\|u_0\|_2^2 \le \|Au_0\|_2^2$

$$\|u_0\|_2^2 \sqrt{s} \le \frac{\sqrt{s}}{1-\delta_s}\|Au_0\|_2^2$$

$$= \frac{\sqrt{s}}{1-\delta_s}\langle Au_0, Au_0\rangle$$

Since $A(u_0 + u_1 + u_2 + \cdots) = 0$,

$$Au_0 = -Au_1 - Au_2 - \cdots$$

$$= \frac{\sqrt{s}}{1-\delta_s}\langle Au_0, -Au_1 - Au_2 - \cdots\rangle$$

$$= \frac{\sqrt{s}}{1-\delta_s}\sum_{k\ge 1}\langle Au_0, -Au_k\rangle$$

Analysis of l_1 minimization: Proof 3

We'll use the following results of RIP constant:

1) Let u, v be sparse vectors with support S_u , S_v such that

$$S_u \cap S_v = \phi, \quad |S_u| + |S_v| = 2s$$

Then $$|\langle Au, Av\rangle| \le \delta_{2s} \|u\|_2 \|v\|_2$$

2) if $s < t$ then $\delta_s \le \delta_t$

Anamitra Makur

Analysis of l_1 minimization: Proof 3

so $$\frac{\sqrt{s}}{1-\delta_s}\sum_{k\ge1}\langle Au_0, -Au_k\rangle \le \frac{\sqrt{s}}{1-\delta_s}\sum_{k\ge1}\delta_{2s}\|u_0\|_2\|u_k\|_2 \quad \text{using result 1)}$$

$$\le \frac{\sqrt{s}\delta_{2s}}{1-\delta_{2s}}\|u_0\|_2\sum_{k\ge1}\|u_k\|_2 \quad \text{using result 2)}$$

Therefore $$\|u_0\|_2^2\sqrt{s} \le \frac{\sqrt{s}\delta_{2s}}{1-\delta_{2s}}\|u_0\|_2\sum_{k\ge1}\|u_k\|_2$$

Or $$\|u_0\|_2\sqrt{s} \le \frac{\sqrt{s}\delta_{2s}}{1-\delta_{2s}}\sum_{k\ge1}\|u_k\|_2$$

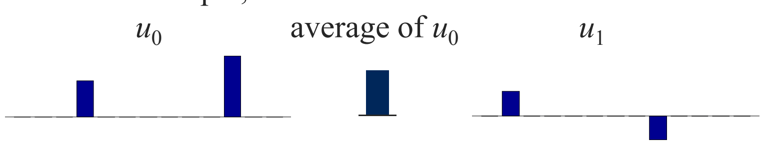
Therefore $$\|u_0\|_1 \le \frac{\sqrt{s}\delta_{2s}}{1-\delta_{2s}}\sum_{k\ge1}\|u_k\|_2$$

Anamitra Makur

Analysis of l_1 minimization: Proof 3

Any component in u_k is smaller than average absolute component in $u_{k\text{-}1}$

In earlier example,



$$\text{so} \quad |\{u_k\}_i| \leq \frac{\|u_{k-1}\|_1}{s}$$

Analysis of l_1 minimization: Proof 3

$$\{u_k\}_i^2 \leq \frac{\|u_{k-1}\|_1^2}{s^2}$$

$$\sum_{nonzero} \{u_k\}_i^2 \leq \frac{\|u_{k-1}\|_1^2}{s}$$

$$\sqrt{\sum_{nonzero} \{u_k\}_i^2} \leq \frac{\|u_{k-1}\|_1}{\sqrt{s}}$$

$$\|u_k\|_2 \leq \frac{\|u_{k-1}\|_1}{\sqrt{s}}$$

$$\frac{\sqrt{s}\delta_{2s}}{1-\delta_{2s}} \sum_{k\geq 1} \|u_k\|_2 \leq \frac{\sqrt{s}\delta_{2s}}{1-\delta_{2s}} \sum_{k\geq 1} \frac{\|u_{k-1}\|_1}{\sqrt{s}}$$

Analysis of l_1 minimization: Proof 3

Therefore
$$\|u_0\|_1 \le \frac{\delta_{2s}}{1-\delta_{2s}} \sum_{k\ge 1} \|u_{k-1}\|_1$$
$$\le \frac{\delta_{2s}}{1-\delta_{2s}} (\|u_0\|_1 + \underbrace{\|u_1\|_1 + \|u_2\|_1 + \cdots}_{\|u_{0^c}\|_1})$$

Since $\delta_{2s} < \frac{1}{3}$, $\frac{\delta_{2s}}{1-\delta_{2s}} < \frac{1}{2}$, or $\|u_0\|_1 < \frac{1}{2}(\|u_0\|_1 + \|u_{0^c}\|_1)$

Therefore $\|u_0\|_1 < \|u_{0^c}\|_1$ or A satisfies null space property of order s

Analysis of l_1 minimization: Proof 4

Proof 4: A satisfies null space property

Proof 4 | null space property means recovery

recovery

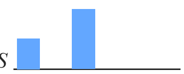
Apart from the sparse solution, consider any other solution.
The difference between these two solutions (a null space vector) is not sparse.
Thus the other solution is not sparse, and has a larger l_1 norm.
The sparse solution indeed has minimum l_1 norm.

Analysis of l_1 minimization: Proof 4

Let x be the s-sparse solution with support S x 

Let there be another solution v such that $A.x = A.v$ v 

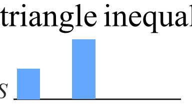
Split v into v_S and v_{S^c} v_S v_{S^c} 

$x = (x - v_S) + v_S$ $x - v_S$  v_S 

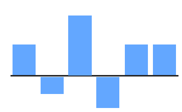
Anamitra Makur NANYANG TECHNOLOGICAL UNIVERSITY

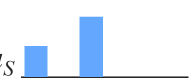
Analysis of l_1 minimization: Proof 4

$\|x\|_1 \le \|x - v_S\|_1 + \|v_S\|_1$ from triangle inequality

$x - v_S$  v_S 

$= \|x_S - v_S\|_1 + \|v_S\|_1$ since the support of x is S

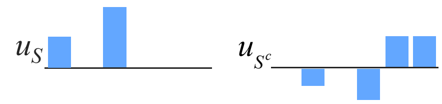
Let $u = x - v$ u 

$= \|u_S\|_1 + \|v_S\|_1$ u_S  v_S 

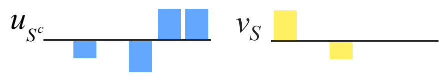
Anamitra Makur NANYANG TECHNOLOGICAL UNIVERSITY

Analysis of l_1 minimization: Proof 4

since $A.u$=0, from null space property, $\|u_S\|_1 < \|u_{S^c}\|_1$

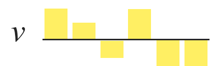


$< \|u_{S^c}\|_1 + \|v_S\|_1$



$= \|-v_{S^c}\|_1 + \|v_S\|_1$ since $u_{S^c} = x_{S^c} - v_{S^c} = -v_{S^c}$

$= \|v\|_1$



Analysis of l_1 minimization: Proof 4

Thus, for a sparse signal x, for any other solution v

$$\|x\|_1 < \|v\|_1$$

or the l_1 minimization will recover x successfully.

In fact, null space property is a necessary and sufficient condition for recovery using l_1 minimization (necessary not proved here).

In conclusion, from proof part 1, recovery using l_1 minimization occurs with probability $1 - e^{-c_3 H.d}$

Section 3. MEASUREMENT MATRIX

- Desirable properties of measurement matrices
- Random Gaussian and Bernoulli matrices
- Structured random/deterministic matrices
- Deterministic matrices
- Measurement matrices for implicitly sparse signals

Properties of measurement matrix

Desirable properties of the measurement matrix:

From l_1 minimization: restricted isometry property (RIP)

$$(1-\delta_{2s})\,\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta_{2s})\,\|x\|_2^2$$

Make δ_{2s} small so that recovery is possible

→ Difficult to measure, let alone design, δ_{2s}
Requires checking the eigenvalues of each sub-matrix

s = sparsity

Properties of measurement matrix

From l_0 minimization: make columns linearly independent
→ make columns incoherent (different from each other)

Coherence of A having columns $\mathbf{a}_j$

$$\mu = \max_{i \neq j} \frac{|< \mathbf{a}_i, \mathbf{a}_j >|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2}$$

→ Easy to measure, iterative design procedures available

Random measurement matrix

Random Gaussian matrix: $A = \{a_{ij}\}$ such that
elements a_{ij} are iid normal with mean 0, variance $\frac{1}{m}$

We have already seen that RIP of order $2s$ holds in probability for this matrix for sufficiently large m

Random Bernoulli/Rademacher matrix:
elements a_{ij} are iid binary with equiprobable values $\pm\frac{1}{\sqrt{m}}$

RIP of order $2s$ holds for Bernoulli matrix, too

d = dimension, m = measurements, s = sparsity

Structured random matrix

Partial random Toeplitz matrix:
Begin with a d x d random Toeplitz matrix

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{d-1} \\ a_{-1} & a_0 & a_1 & \cdots & a_{d-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{-d+1} & a_{-d+2} & a_{-d+3} & \cdots & a_0 \end{bmatrix}$$

Pick m random rows of the above matrix

Structured random matrix

Partial random circulant matrix:
Begin with a d x d random circulant matrix

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{d-1} \\ a_{d-1} & a_0 & a_1 & \cdots & a_{d-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix}$$

Pick m random rows of the above matrix

Structured deterministic matrix

Partial Fourier matrix:
Begin with a d x d Fourier matrix W

$$w_{pq} = \frac{1}{\sqrt{d}} e^{2\pi i \frac{pq}{d}}$$

Pick m random rows of the above matrix
RIP of order $2s$ holds for partial Fourier matrix
(works with other orthonormal bases)

Advantages of structured random/deterministic matrices:
- Less storage requirement than unstructured matrix
- Fast matrix-vector multiplication possible using FFT

Deterministic measurement matrix

Example: $d = p^{r+1}$ where p = prime, $0 < r < p$, $m = p^2$
Consider all polynomials of degree r in the finite field GF(p):

$$q(w) = a_0 + a_1 w + \cdots + a_r w^r, \quad a_i \in \mathrm{GF}(p)$$

Construct p x p matrix for $q(w)$:

$$\begin{array}{cc} \boxed{\begin{array}{ccc} 0\cdots 0 & \frac{1}{\sqrt{p}} & 0\cdots 0 \\ & & \\ & & \end{array}} & \leftarrow w \\ \uparrow & \\ q(w) & \end{array}$$

Rearrange to a column

There are p^{r+1} polynomials
Construct A from these columns

For any $k < \frac{p}{r} + 1$, A satisfies $\delta_k = \frac{(k-1)r}{p}$

Case of implicitly sparse signals

For implicitly sparse signals, the measurement matrix is $A\Psi$ where Ψ is the basis

A and Ψ should not be coherent
Example: random Gaussian matrix is incoherent with any basis with high probability
Let A = random Gaussian matrix, Ψ = DCT basis, then $A\Psi$ remains white Gaussian noise

Random Bernoulli matrix is also incoherent with any basis

Case of implicitly sparse signals

For some signals/applications, there may be multiple bases Ψ_1 and Ψ_2, which can be concatenated to obtain a joint basis if they are not coherent

Mutual coherence of Ψ_1 and Ψ_2 having columns $\mathbf{a}_j$ and $\mathbf{b}_j$

$$\mu = \max_{i,j} \frac{|<\mathbf{a}_i, \mathbf{b}_j>|}{\|\mathbf{a}_i\|_2 \|\mathbf{b}_j\|_2}$$

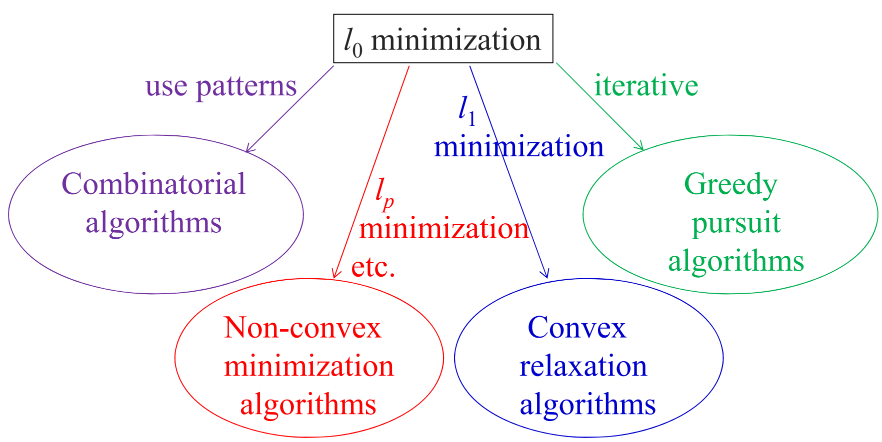
Example: $\Psi_1 = d \times d$ Fourier basis (localized in frequency)
$\Psi_2 = d \times d$ impulse basis or identity basis (localized in time)
$[\Psi_1\ \Psi_2]$ has mutual coherence $\mu = \frac{1}{\sqrt{d}}$, RIP constant $\delta_s \leq \frac{s-1}{\sqrt{d}}$

Haar wavelets and noiselets are also incoherent

Section 4. RECONSTRUCTION ALGORITHMS

- Combinatorial algorithms
- Non-convex minimization algorithms
- Convex relaxation algorithms
- BP
- Greedy pursuit algorithms
- OMP

Reconstruction algorithms



Combinatorial algorithms

Exploits specific patterns in the sparsity/measurements
Applicable only for applications where such patterns exist

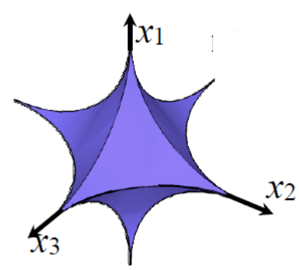
Examples:
HHS (heavy hitters on steroids)
CP (chaining pursuits)
FSA (Fourier sampling algorithm)
Sudocodes
etc.

computation = very low

Non-convex minimization algorithms

Relax l_0 minimization to a non-convex but favorable minimization problem, such as:

l_p minimization where $0 < p < 1$
(recall that l_p ball is not convex)

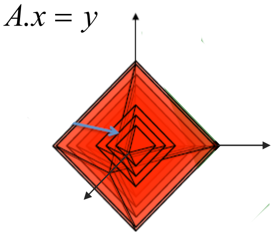


Examples:
IRL1 (iterative reweighted l_1)
ISD (iterative support detection)
SBL (sparse Bayesian learning)
etc.

Convex relaxation algorithms: BP

BP (basis pursuit) $\min_x \|x\|_1$ subject to $A.x = y$

Recall that l_1 ball is convex:



→ minimization may be solved using
- linear program for real values
- Second order cone program for complex values

polynomial time complexity
sparsity s not required
typically performs the best
several fast algorithms exist, such as sparselab/CVX solver

Convex relaxation algorithms

BPDN (basis pursuit de-noising)

$$\min_x \|x\|_1 \quad \text{subject to} \quad \|A.x - y\|_2 \le \varepsilon$$

LASSO (least absolute shrinkage and selection operator)

$$\min_x \|A.x - y\|_2 \quad \text{subject to} \quad \|x\|_1 \le \varepsilon$$

DS (Dantzig selector)

$$\min_x \|x\|_1 \quad \text{subject to} \quad \|A.x - y\|_\infty \le \varepsilon$$

and more such as LARS (least angle regression), etc.

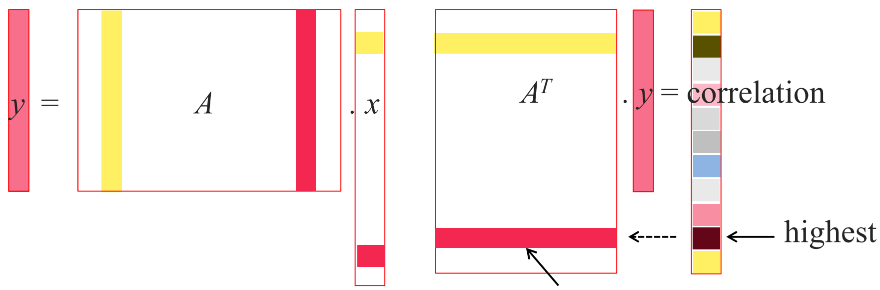
computation = very high

Greedy pursuit algorithms: OMP

OMP (orthogonal matching pursuit)

y = weighted sum of few columns of A

Find which column is most correlated with y (greedy!)



add this column to support set $\hat{A}$

Greedy pursuit algorithms: OMP

In some iteration, support set $\hat{A}$ has a few columns

Signal estimate for $\hat{A}$ = solution with minimum error in l_2 norm

$$\hat{x} = \hat{A}^{\dagger} y \text{ in } \hat{A}, \; 0 \text{ in other positions}$$

where $\hat{A}^{\dagger}$ = pseudo-inverse of $\hat{A}$ (found recursively, no inverse)

Why in l_2 norm?

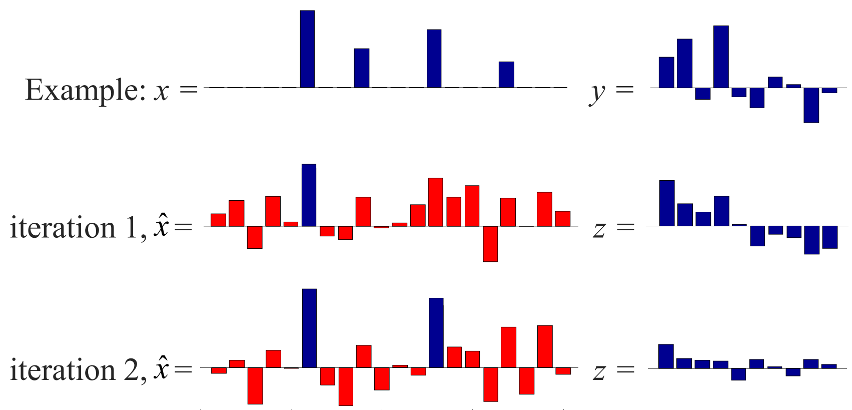
- Signal remains sparse since $\hat{A}$ has few columns
- Nice closed form solution
- If there is an exact solution, l_2 norm will find it

What is left? residue $z = y - A\hat{x}$

Find which column is most correlated with z

Continue for s iterations → sparsity s required

Greedy pursuit algorithms: OMP



and so on

Greedy pursuit algorithms

Greedy pursuits involve iterative estimation, greedy step in each iteration, most having convergence guarantee

Examples:
IHT/IST (iterative hard/soft thresholding)
AMP (approximate message passing)
MP (matching pursuit)
OMP (orthogonal matching pursuit)
SP (subspace pursuit)
CoSAMP (compressive sampling matching pursuit)
BAOMP (backtracking-based adaptive OMP)
etc.
computation = low

Section 5. APPLICATIONS

- Compression
- Denoising
- Classification/recognition
- Data acquisition

Compression

- Typically with a basis
- Typically with other refinements
- Universal encoder (basis not required)
- Asymmetrical: simple encoder, complex decoder

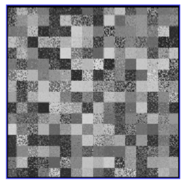
	H.264	CS based
video with small object size	30 sec	0.6 sec
video with large object size	18 sec	0.75 sec

Example: video compression, encoding time

Anamitra Makur NANYANG TECHNOLOGICAL UNIVERSITY

Compression

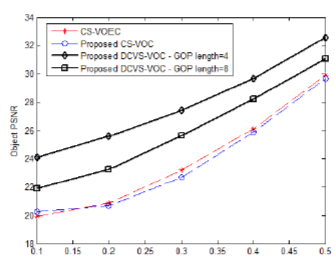
- Possible to compress after encryption



Example: encrypted image after compress.+reconstruction

Compression

- Distributed compressive sensing (independent encoder, joint decoder)



Example: video compression, PSNR results

Denoising

- Dictionary design
- Sparse representation of signal on the dictionary permits removing noise, filling up missing values, etc.


missing pixels

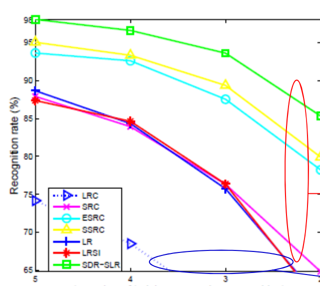

CS


noisy image


CS

Classification/recognition

- Robust



CS based

not CS based

Example: expression-robust face recognition

Data acquisition

Useful where measurements are limited by nature, such as
- Image acquisition
- MRI (magnetic resonance imaging)
- Radar

Section 6. CONCLUSION

Compressed sensing:

- Founded on theoretical guarantees
- General framework with several variations
- Many algorithms
- Applications in diverse fields

Thank you for your attention!