

CO 769. Jan 6, 2009

1

## RECENT ADVANCES IN COMPRESSIVE SENSING & MINIMIZATION

Compressive (= compressed) sensing has emerged as a major topic in signal processing, statistics, computer science and optimization during the past 5 years. This course covers recent papers and background.

Compressive sensing draws upon

- optimization + convexity
- prob. + meas. theory
- Analysis
  - signal processing
  - Algorithms + complexity.

Thm 1: Let  $A \in \mathbb{R}^{m \times n}$  be a random matrix:

Choose each entry independently from the normal distribution with mean = 0, variance = 1. Then with high prob.  $A$  has the following properties:

① For every vector  $\underline{x} \in \mathbb{R}^n$  such that  $\underline{x}$  has at most  $f(m, n)$  non-zero entries.

$\underline{x}$  is the unique solution to the linear equations

$$A \underline{\xi} = \underline{b} \quad \text{where} \quad \underline{b} = A \underline{x} \quad \text{such that} \quad \underline{\xi} \text{ has}$$

at most  $f(m, n)$  non-zero entries.   
 Divided by a modest factor more later.

at most  $\leq f(m, n)$  non-zero entries.

② For every vector  $\underline{x} \in \mathbb{R}^n$ , st.  $\underline{x}$  has at most  $f(m, n)$  non-zero entries,  $\underline{x}$  is the unique soln to the optimization problem

$$\min \|\underline{\xi}\|_1, \leftarrow L_1\text{-norm} \approx 1\text{-norm of } \underline{\xi}$$

$$\text{st. } A \underline{\xi} = \underline{b} = A \underline{x}$$

means

$$|\xi_1| + \dots + |\xi_n|$$

- what is the use?
- why is it unexpected?
- How to prove?
- How to extend it?

What is the use?

suppose  $A$  signal is being monitored and it's mostly 0's.

Say total signal length is  $n$ . and at most  $k$  non-zeros

Now to compress

obvious approach. If signal is  $x_1, x_2, \dots, x_n$

Then compress it as  $(i_1, x_{i_1}), \dots, (i_k, x_{i_k})$  where

$i_1, \dots, i_k \in \{1, \dots, n\}$  are positions of non-zero entries.

Problem: must acquire whole signal first to apply this compression.

Instead, can take  $m$  inner products of  $x$  with random vectors. (I.e. rows of  $A$ ), and store these inner products.

Theorem 1 says signal can then be recovered.

Recovery procedure is to solve  $\ell_1$  optimization problem.

This type of sampling is non-adaptive.

More general setting.

The signal  $x \in \mathbb{R}^n$  is not sparse, but becomes sparse after

application of a fixed orthogonal transformation.

E.g. 2D wavelet transformation. is applied to an image.

especially a cartoon image. the resulting vector will be very sparse.

Theorem 2: Recall  $\Phi \in \mathbb{R}^{m \times n}$  is orthogonal. if  $\Phi^T \Phi = I$

equivalently, cols of  $\Phi$  form an orthonormal basis of  $\mathbb{R}^m$  identity

Theorem 2: Let  $\Phi$  be orthogonal, let  $A$  be as above.

Then with high probability,

(a) For every vector  $x$  such that  $\Phi x$  has at most  $f(m, n)$  non-zero entries.  $x$  is the unique solution to the equation

$$A \xi = \underline{b}, \text{ where } \underline{b} = Ax \text{ and } \Phi \xi \text{ has } \leq f(m, n) \text{ non zeros.}$$

(b) for every  $x$  st  $\|\Phi x\|_0 \leq f(m, n)$

$x$  is the unique soln. to the zero norm, # of nonzero entries.

$$\text{with } \|\Phi \xi\|_1$$

$$\text{st. } A \xi = \underline{b}$$

PF follows from theorem 1. Apply theorem 1 to  $A\Phi^{-1}$  and define  $y = \Phi x$

Need following fact:

Distribution of  $m \times n$  matrices defined in theorem 1 is invariant under mult. by an ortho. mtr.

Signal recovery Algorithm:

minimize  $\|x\|_1$  It is a special case of Linear Programming

st.  $Ax = b$

Standard Equality Form Linear programming SEFLP

Not the same  $\min c^T x$  — Linear objective func.

st.  $Ax = b$  — Linear equality constraints

$x \geq 0$  — Linear Inequality special case sign constraints

$A \in \mathbb{R}^{m \times n}$  usually  $m < n$   $b \in \mathbb{R}^m$

most general form of LP

$\min c^T x$

st.  $A_1 x \geq b_1$

$A_2 x \geq b_2$

$x_i \geq 0$  for  $i \in I \cup \{1, \dots, n\}$

Fact: general form of LP can be reduced to S.E.F.

$\min \|x\|_1$  st.  $Ax = b$

Equivalent to

$\min t_1 + \dots + t_n$

st.  $t_1 \geq x_1$   
 $t_1 \geq -x_1$   
 $\vdots$

$t_n \geq x_n$

$t_n \geq -x_n$

$Ax = b$

the more obvious approach of constraining

$t_1 = |x_i|$  won't work because cannot

be expressed as linear constraints.

However, thanks to the form of the objective function.

$t_1 \geq |x_i|$  suffices.

There are algorithms for LP.

simplex. requires typically  $O(mn^2)$  operations.

interior-point and ellipsoid methods known to be polynomial time.

specialized methods exist for  $L_1$ -minimization.

even more specialized algorithms exist for compressive sensing.

— Why is theorem 1 unexpected?

① Classical theory of signal processing says that to accurately represent a signal requires a number of samples proportional to  $1/\text{max-frequency}$ .

② The general problem

$$\min \|x\|_0$$

$$\text{st. } Ax = b$$

is NP-hard. means (roughly speaking)

probably no polynomial time algorithm exists)

Yet we solve this easily easily in theorem 1.

Two key ingredients allow easy soln.

① Randomized constraints

② Advance knowledge that a sparse soln exists.

Convexity

Probability meas. theory

Signal processing

complexity

Jan 8 2009

Let  $G(n, k) = \{ \text{set of subspaces of } \mathbb{R}^n \text{ of dim } k \}$   
( $\hookrightarrow$  Grassmann manifold)

Let  $\Phi \in \mathbb{R}^{n \times n}$  be orthogonal

Suppose  $S \subset G(n, k)$ , let  $K \in S$

Define  $\Phi K$  by  $\Phi x \in \Phi K \Leftrightarrow x \in K$

FACT: If  $K$  is a subspace of dim  $k$ , so is  $\Phi K$ .

Define  $\Phi S$  by

$$\Phi K \in \Phi S \Leftrightarrow K \in S$$

A probability measure on  $G(n, k)$  is orthogonally invariant if for any  $S \subset G(n, k)$  st.  $\text{prob}(S)$  is defined and for all orthogonal matrices,

$$\text{Prob}(\Phi S) = \text{Prob}(S)$$

Fact: There is only one orthogonally invariant probability measure on  $G(n, k)$

FACT: Can sample uniformly from this distribution by generating an  $(m \times n)$  MTR by choosing each entry indept. from normal dist. mean 0 variance 1.

And taking  $K = \text{null}(A)$

Norms: The  $p$ -norm or  $L_p$ -norm of  $x \in \mathbb{R}^m$  is defined by

$$\|x\|_p = (\sum |x_i|^p)^{1/p} \quad (p > 0)$$

Special cases:  $p=2$  Euclidean norm  $\sqrt{x_1^2 + \dots + x_n^2}$

$p=1$   $L_1$ -norm  $|x_1| + \dots + |x_n|$

$$\lim_{p \rightarrow \infty} \|x\|_p = \max\{|x_1|, \dots, |x_n|\} =: \|x\|_\infty$$

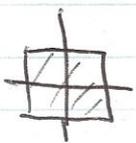
For  $0 < p < 1$   $\|x\|_p$  NOT Norm,  $\rightarrow$  Triangle ineq. violated.

eg  $\| \begin{pmatrix} 1 \\ 0 \end{pmatrix} \|_{1/2} = (1^{1/2} + 0^{1/2})^2 = 1$

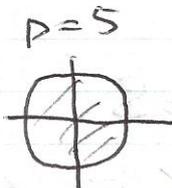
$$\| \begin{pmatrix} 0 \\ 1 \end{pmatrix} \|_{1/2} = 1 \quad \| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \|_{1/2} = 4$$

Convention:  $\|x\|_0$  denotes the # of non-zero entries of  $x$

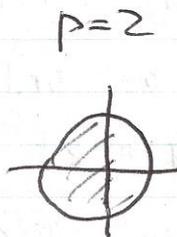
Unit Balls of  $p$ -norms  $n=2$



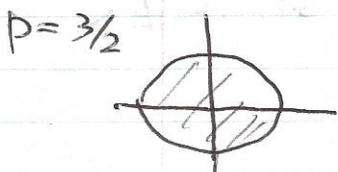
$p=0$



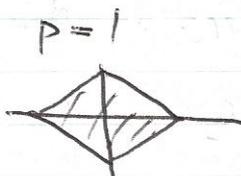
$p=5$



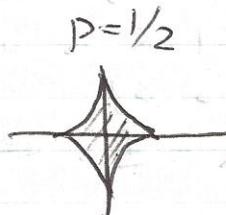
$p=2$



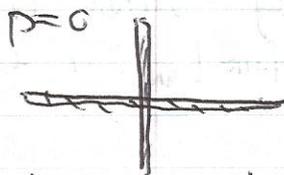
$p=3/2$



$p=1$



$p=1/2$



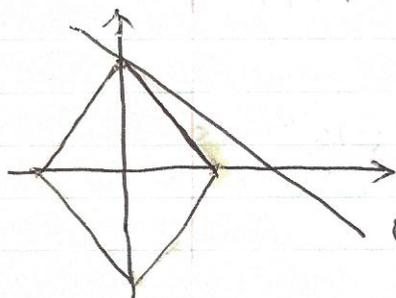
$p=0$

consider  $\min \|x\|_p$  st.  $Ax=b$

For  $0 \leq p \leq 1$  sparsity enhanced in solution

For  $p > 1$  solution in general is dense

$$\min \|x\|_1 \quad \text{st. } a^T x = b$$



$a^T x = b$  Proof:

Simple relationships b/w  $L_1$  and  $L_2$ -norm

$$\forall x \in \mathbb{R}^n, \|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2$$

$$\textcircled{a} \|x\|_2^2 = x_1^2 + x_2^2 + \dots + x_n^2$$

$$\|x\|_1^2 = (|x_1| + \dots + |x_n|)^2$$

$$= x_1^2 + \dots + x_n^2 + 2|x_1||x_2| + \dots + 2|x_{n-1}||x_n|$$

$\textcircled{a}$  is tight, Take  $x = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$l_1$ -norm of  $x$  equals  $\sum x^T$  for some  $\Sigma$  of the form  $\geq c$

$$\Sigma = \begin{pmatrix} \pm 1 \\ \vdots \\ \pm 1 \end{pmatrix}$$

inner product

Reall Cauchy-Schwarz inequality

$$\underline{a}^T \underline{b} \leq \|\underline{a}\|_2 \|\underline{b}\|_2$$

$$\Rightarrow \|\underline{x}\|_1 \leq \|\underline{1}\|_2 \|\underline{x}\|_2 = \sqrt{n} \|\underline{x}\|_2$$

ⓑ is tight achieved by  $\underline{x} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

Fact for a random  $\underline{x}$ ,

$$\|\underline{x}\|_1 \sim \sqrt{n} \|\underline{x}\|_2 \quad \text{so case ⓑ typical}$$

LET  $m, n$  be positive integers st.  $m < n$ .

Let  $V$  be an  $n-m$  dimensional subspace of  $\mathbb{R}^n$  say

that  $V$  has the  $f(m, n)$  - Approximate spherical section property if

$$\forall \underline{v} \in V - \{0\} \cdot \frac{\|\underline{v}\|_1}{\|\underline{v}\|_2} \geq f(m, n)$$

Dvoretzky 1961

Any convex body has a fairly high dimensional slice that is approximately spherical.

Thm: (ZHANG Attributes Koshin, Garnaev, GLUSKIN (KGG))

If  $V \in G(n, n-m)$ , is chosen at random according to above distribution.

then it has the  $f_0(m, n)$  - Approximately spherical section property with probability  $\geq 1 - e^{-C_1(n-m)}$

What is  $f(m, n) = \frac{C_2 \sqrt{m}}{\sqrt{1 + \log(n/m)}}$

eg if  $m = \text{constant}$ ,  $n$ . Then  $f(m, n) = C_2 \sqrt{m}$

Hence for almost all  $V$ ,  $\forall \underline{v} \in V \setminus \{0\}$

$$\text{const } \sqrt{m} \leq \frac{\|\underline{v}\|_1}{\|\underline{v}\|_2} \leq \sqrt{n}$$

ineq ⓑ above

Lemma:  $\forall \underline{x} \in \mathbb{R}^n$ ,  $\|\underline{x}\|_1 \leq \sqrt{\|\underline{x}\|_0} \|\underline{x}\|_2$

PF: use C-S ineq again

$$\|\underline{x}\|_1 = \underline{x}^T \underline{1} \quad \text{where } \underline{1}_i = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{if } x_i < 0 \\ 1 & \text{if } x_i = 0 \end{cases}$$

OBS:  $\|s\|_2 = \sqrt{\|x\|_0}$

proof of part (a) of Theorem 1

Choose  $V \in \mathbb{G}(n, n-m)$ , wish to show that with high prob. null space of  $A$  in Theorem 1.

$V$  has the following property

$1 - e^{-c_1(n-m)}$

(a)  $\forall x \neq 0$ ,  $\|x\|_0 \leq f(m, n)$ , No other  $\sum_{i \in \text{supp}(x)} \in \mathbb{R}^n$

satisfies  $\sum -x \in V$  and  $\|\xi\|_0 \leq f(m, n)$   
 previously stated as

PF: require that  $V$  has  $A\xi = b$  where  $b = Ax \Rightarrow$

$f_0(m, n)$  - a approximately Spherical section property  
 $\xi - x \in \text{null}(A)$

We will prove contrapositive, suppose

$\xi \neq x$  and  $\xi - x \in V$ . Then  $\|\xi\|_0 > f(m, n)$

OBS:  $\|\xi - x\|_0 \leq \|\xi\|_0 + \|x\|_0 \leq f_0(m, n) + \|\xi\|_0$   
 $\geq \left( \frac{\|\xi - x\|_1}{\|\xi - x\|_2} \right)^2$  by the lemma

Assumption  $\frac{\|\xi - x\|_1}{\|\xi - x\|_2} \geq f_0(m, n)$  by assumption about  $V$ .

Combine  $\|\xi\|_0 + f(m, n) \geq f_0(m, n)^2$

i.e.  $\|\xi\|_0 \geq f_0(m, n)^2 - f(m, n)$

choose  $f(m, n) \leq \frac{1}{2} f_0(m, n)^2$

To ensure that  $\|\xi\|_0 > \frac{1}{2} f_0(m, n)^2 > f(m, n)$

Note this is a simplification of Zhang's proof.

proof of (b) of Theorem 1:

$$f(m, n) = \text{const.} \frac{m}{1 + \log(n/m)}$$

$\forall x (\neq 0)$  such that  $\|x\|_0 \leq f(m, n)$

$x$  is the unique solution to  $\min \|\xi\|_1$  st  $\xi - x \in V$  (\*)

Additional assumptions  $f(m, n) < n/2$

prove contrapositive

Assume  $\xi$  is distinct from  $x$  and is minimizer of

(\*) And conclude  $\|x\|_0 > f(m, n)$

Let  $\text{Supp}(x) = \{i \in \{1, \dots, n\}, x_i \neq 0\}$

Let  $S = \text{Supp}(x)$  and  $\bar{S}$  its complement

$$\rightarrow \text{OBS } |S| = \|x\|_0$$

$$\|x\|_1 \geq \|\xi\|_1 = \|\xi_S\|_1 + \|\xi_{\bar{S}}\|_1 = \|\xi_S\|_1 + \|\xi_{\bar{S}} - x_{\bar{S}}\|_1$$

$\xi$  is the minimizer by hypothesis

$$\geq \|x_S\|_1 - \|x_S - \xi_S\|_1 + \|\xi_{\bar{S}} - x_{\bar{S}}\|_1$$

$$= \|x\|_1 - \|\xi_S - x_S\|_1 + \|\xi_{\bar{S}} - x_{\bar{S}}\|_1$$

$$\Rightarrow \|\xi_{\bar{S}} - x_{\bar{S}}\|_1 \geq \|\xi_S - x_S\|_1$$

Jan B. 2009.  $\in \mathbb{R}^n$

subspace  $V$  satisfies the approximate spherical section property

DIM is  $m$

w.t.p. (wants to prove)  $\forall x$  st  $\|x\|_0 \leq f(m, n)$

$x$  is the unique soln. to  $\min \|\xi\|_1$  st  $\xi - x \in V$ .

additional Assumption  $f(m, n) < n/2$  we will prove the

Let ~~S~~ contrapositive Assume  $\xi$  is a minimizer Distinct from  $x$  and conclude that  $\|x\|_0 > f(m, n)$

Let  $S = \text{supp}(x)$  From last time

$$\|x\|_1 \geq \|\xi\|_1 \dots \geq \|x\|_1 - \|\xi_S - x_S\|_1 + \|\xi_{\bar{S}} - x_{\bar{S}}\|_1$$

$$\Rightarrow \|\sum_S - x_S\|_1 \geq \|\sum_S - x_S\|$$

Question: What is the max value of  $\frac{\|\sum_S - x\|_1}{\|\sum_S - x\|_2}$

$$\text{st. } \|\sum_S - x_S\|_1 \geq \|\sum_S - x_S\|_1 \quad ?$$

Lemma  $\max \left\{ \frac{\|z\|_1}{\|z\|_2} : \|z_S\|_1 \geq \|z_{\bar{S}}\|_1, z \neq 0 \right\}$

$$\leq 2 \sqrt{\frac{|S||\bar{S}|}{|S|+|\bar{S}|}} \text{ where } S \subseteq \{1, \dots, n\} \text{ satisfies}$$

$$|S| < n/2$$

proof: OBS: Objective  $\frac{\|z\|_1}{\|z\|_2}$  and constraint

$\|z_S\|_1 \geq \|z_{\bar{S}}\|_1$  Invariant under Rescaling and under flipping signs.

Thus WLOG, Assume  $z_1, \dots, z_n \geq 0$ , and  $z_1^2 + \dots + z_n^2 = 1$

problem becomes  $\max z_1 + \dots + z_n$  call this

$$\text{st. } z_1^2 + \dots + z_n^2 = 1 \quad \left. \begin{array}{l} \text{**} \\ \downarrow \end{array} \right\}$$

$$\sum_{i \in S} z_i \geq \sum_{i \in \bar{S}} z_i$$

$$z_1 \geq 0 \dots z_n \geq 0$$

OK to Relax scaling constraints to  $z_1^2 + \dots + z_n^2 \leq 1$  (\*\*)

The problem is convex opt. hence the KKT suffices for global optimality. conditions

claim: optimal soln. is

$$z_i^* = \begin{cases} a/|S| & \text{for } i \in S \\ a/|\bar{S}| & \text{for } i \in \bar{S} \end{cases}$$

check constraint (\*\*)

where  $a = \sqrt{\frac{|S||\bar{S}|}{|S|+|\bar{S}|}}$

$$|S| \cdot \frac{a^2}{|S|^2} + |\bar{S}| \cdot \frac{a^2}{|\bar{S}|^2} = \frac{a^2}{|S|} + \frac{a^2}{|\bar{S}|} = 1$$

so (\*\*) satisfies as equality.

So KKT conds satisfied if  $\exists \lambda_1, \lambda_2 \geq 0$  st.

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ \vdots \\ -1 \end{pmatrix} + 2\lambda_2 \begin{pmatrix} a/|s| \\ a/|s| \\ \vdots \\ a/|s| \end{pmatrix}$$

Gradient evaluated at  $z^*$ .

$\left. \begin{matrix} s \\ \vdots \\ s \end{matrix} \right\} \begin{matrix} s \text{ entries} \\ 3 \text{ entries} \end{matrix}$

i.e.  $\begin{cases} \lambda_1 + 2\lambda_2 a/|s| = 1 \\ -\lambda_1 + 2\lambda_2 a/|s| = 1 \end{cases}$

$$\Rightarrow \lambda_2 = \left( \frac{1}{2a(\frac{1}{|s|} + \frac{1}{|s|})} \right) \geq 0 \Rightarrow \lambda_1 = \frac{\frac{1}{|s|} - \frac{1}{|s|}}{\frac{1}{|s|} + \frac{1}{|s|}} \geq 0 \text{ by assumption}$$

OBJ: Func. val. at  $z^*$  is

$$\sum_{i=1}^n z_i^* = |s| \frac{a}{|s|} + |s| \frac{a}{|s|} = 2a$$

$$= 2 \sqrt{\frac{|s||s|}{|s|+|s|}} \text{ proves the lemma } \square$$

Back to part of (b). Have concluded that

$$\frac{\|\underline{z} - \underline{x}\|_1}{\|\underline{z} - \underline{x}\|_2} \leq 2 \sqrt{\frac{|s||s|}{|s|+|s|}} \leq 2\sqrt{|s|} = 2\sqrt{\|\underline{x}\|_0}$$

Div. Num. and denominator by  $|s|$

on the other hand  $\frac{\|\underline{z} - \underline{x}\|_1}{\|\underline{z} - \underline{x}\|_2} \geq f_0(m,n)$  (4 approx. sph sec. property of  $v$ )

$$\Rightarrow 2\sqrt{\|\underline{x}\|_0} \geq f_0(m,n)$$

$$\Rightarrow \|\underline{x}\|_0 \geq \frac{f_0(m,n)^2}{4} \Rightarrow \|\underline{x}\|_0 \geq f(m,n)$$

$$\text{provided } f(m,n) < \frac{f_0(m,n)^2}{4}$$

This result comes from

Donoho, For most large underdetermined systems of linear equations.

The minimal  $L_1$ -norm solution is also the sparsest solution.

Simplification: By

Zhang, on theory of compressive sensing. via  $L_1$ -minimization

simple derivations + extensions

And By S.V:

Extensions of this thm:

① Other ways to select A

② errors Two sources of error in compressive sensing

- What if signal  $x$  to be recovered is not sparse but only approximately sparse.

- What if measurements are approximate?

Several papers in the field have focused on partial Fourier information.

Supp.  $x \in \mathbb{C}^N$ , define the discrete Fourier transform (DFT)

of  $x$  to be  $y \in \mathbb{C}^N$ .

$$\text{st. } y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{-2\pi i j k / N} x_j$$

Note:  $y \in \mathbb{C}^N$  via  $0, \dots, N-1$

Note that this is linear. I.e.

$$y = \Phi x \quad \text{where } \Phi = \mathbb{C}^{N \times N} \text{ given by}$$

$$\Phi_{jk} = e^{-2\pi i j k / N} / \sqrt{N}$$

Thm:  $\Phi$  is unitary.  $\Phi^H \Phi = I$ .

"Unitary" is a generalization of orthogonal.

$\Phi^H$  means conjugate transpose of  $\Phi$ . Also denoted  $\Phi^*$ .

$$\text{PF: } (\Phi^H)_{kl} = e^{2\pi i k l / N} / \sqrt{N}$$

Thus if  $y = \Phi x$  and  $z = \Phi^H y$

Then

$$z_l = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i k l / N} y_k = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i k l / N} \sum_{j=0}^{N-1} e^{-2\pi i j k / N} x_j$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} \left[ \sum_{l=0}^{N-1} e^{-2\pi i j k l / N} e^{2\pi i k l / N} \right] x_j$$

$$(X) = \sum_{k=0}^{N-1} e^{2\pi i k (l-j) / N}$$

Recall  $\sum_{k=0}^{N-1} f^k = \frac{f^N - 1}{f - 1}$  ( $f \neq 1$ )

$$= \frac{(e^{2\pi i (l-j) / N})^N - 1}{e^{2\pi i (l-j) / N} - 1} = 0 \text{ provided } (l \neq j)$$

$$(X) = \sum_{k=0}^{N-1} 1 = N$$

thus  $Z_l = \frac{1}{N} \sum_{j=0}^{N-1} \begin{cases} 0 & \text{if } (l \neq j) \\ N & \text{if } (l = j) \end{cases} x_j = x_l$

so  $Z = X$  so  $\Phi^H \Phi = I$

Note:  $\Phi^H$  called (inverse DFT)

Some authors put the (+) in DFT and (-) in IDFT.

... 1 in front of DFT and  $\frac{1}{N}$  in IDFT.

often write  $\Phi X$  as  $\hat{X}$

often  $\hat{X}$  subscript with  $w$  often written  $\hat{X}(w)$

suppose  $\hat{X} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$  pos  $w^*$  implies that

pure tone signal  $x_j = \frac{1}{N} e^{2\pi i j w^* / N}$  ze. sampling the func.  $\frac{1}{N} e^{2\pi i w^* t}$

at evenly spaced t's.

This  $X$  is called a "pure tone".

Jan 15, 2009.

FAST FOURIER Transform COOLEY-TUKEY

Given  $X \in \mathbb{C}^N$ ,  $N$  is power of 2.

say  $N=2^P$ , can compute  $\hat{X}$  IN  $O(N \log N)$ . Arithmetic ops.

Obvious algorithm of multiplying  $\Phi X$  would require  $O(N^2)$  ops.

① RECURSIVELY Apply FFT to  $x_0, x_2, \dots, x_{N-2}$

To get  $\hat{\Phi} \begin{pmatrix} x_0 \\ \vdots \\ x_{N-2} \end{pmatrix}$  DFT in  $\mathbb{C}^{N/2 \times N/2}$  - call the result  $v_0, v_2, \dots, v_{N-2}$

② recursively apply to  $x_1, x_3, \dots, x_{N-1}$

OBS:  $v_{2k} = \frac{1}{\sqrt{N/2}} \sum_{j=0}^{N/2-1} e^{-2\pi i j k / (N/2)} x_{2j}$  - call the result  $w_1, \dots, w_{N-1}$

$$w_{2k+1} = \frac{1}{\sqrt{N/2}} \sum_{j=0}^{N/2-1} e^{-2\pi i j k / (N/2)} x_{2j+1}$$

OBS:  $y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{-2\pi i j k / N} x_j$

$$= \frac{1}{\sqrt{N}} \sum_{\substack{j=0 \\ \hat{2j}=j}}^{N/2-1} e^{-4\pi i \hat{j} k / N} x_{\hat{2j}} + \frac{1}{\sqrt{N}} \sum_{\substack{j=0 \\ \hat{2j+1}=j}}^{N/2-1} e^{-4\pi i \hat{j} k / N - 2\pi i k / N} x_{\hat{2j+1}}$$

$$= \begin{cases} \frac{1}{\sqrt{2}} v_{2k} + \frac{1}{\sqrt{2}} w_{2k+1} e^{-2\pi i k / N} & \text{if } k < N/2 \\ \frac{1}{\sqrt{2}} v_{2(k-N/2)} + \frac{1}{\sqrt{2}} w_{2(k-N/2)+1} e^{-2\pi i k / N} & \text{if } k \geq N/2 \end{cases}$$

so we need two recursive calls to FFT, each with

$N/2$  ops. plus  $O(N)$  ops to recover  $y$  from  $v$  and  $w$

so let  $T(N) =$  Total # of operations for FFT on length- $N$  vector

Obtain inequality

$$T(N) \leq 2T(N/2) + \text{const. } N \quad \text{Has solution}$$

$$T(N) \leq \text{const} + N \log N.$$

In compressive sensing, consider  $A$  chosen to be "partial Fourier information" i.e. a submatrix consisting of rows of  $\hat{\Phi}$ .

Two groups with this same idea.  
GILBERT et al.

eg.  $m$  non zero entries

suppose  $\hat{x} \in \mathbb{R}^n$  is sparse, can we determine  $\hat{x}$  using Fourier transform of  $x$ .

$m$   $\rightarrow$  smallest-factor. samples of  $x$ ?

(samples chosen at random)

Candès's Ramore, Tao, posed dual question

Given  $x \in \mathbb{R}^n$ ,  $x$  sparse. Can you determine  $x$  knowing some (random selection) of entries of  $\hat{x}$

In Gilbert et al work,

$A$  is a randomly selected set of rows of  $\Phi^H$

In C-R-T.  $A$  is randomly selected rows of  $\Phi$ .

In both cases: we want a sparsest solution to

$$A \hat{x} = b$$

The difference btw  $\Phi$  vs.  $\Phi^H$  is inconsequential mathematically.

But other aspects of the papers are quite different.

① Gilbert. for every sparse  $\hat{x}$ , there is a high probability over choices of randomization in the algorithm (which rows of  $\Phi^H$  to select) can recover  $\hat{x}$ . C.R.T. similar. But followed by

② Candès + Tao.

if a random subset of rows of  $\Phi$  is selected with high prob.

you can recover it from  $Ax$ , for every sparse  $x$ .

C.R.T use  $L_1$ -minimization. (once for all)

Gilbert uses a specialized algorithm seemingly unconnected to optimization

Gilbert algorithm very fast.

Gilbert, Strauss, Tropp.

Fast Fourier Sampling A tutorial.

Improved time bounds for

Gilbert, Muthukrishnan, Strauss. near optimal sparse Fourier representation

Gilbert, Indyk,

Muthukrishnan, Strauss.

Near optimal sparse

Fourier representation via sampling.

Key idea: Apply DFT to a short subsequence of evenly spaced entries of  $x$  with randomly chosen spacing  $\sigma$ .

Suppose  $\hat{x}$  is sparse, say contains only  $m$  frequencies  $\omega_1, \dots, \omega_m$ .

So  $\hat{x}(\omega_k) \neq 0$ ,  $k=1, \dots, m$   $\hat{x}(\omega) = 0$  else.

Assume  $N$  is a power of 2. Choose  $b$  and  $K$  deterministically (more later) choose  $\sigma \in \{1, 3, 5, \dots, N-1\}$  uniformly at random

Let  $y_l = x_{b+l\sigma \pmod N}$  for  $l=0, \dots, K-1$ .

idea  $m < K \ll N$

compute DFT of  $y$ , call it  $\hat{y} \in \mathbb{C}^K$ . (use FFT)

OBS.  $x_j = \frac{1}{N} \sum_{\nu=1}^m \hat{x}(\omega_\nu) e^{2\pi i \omega_\nu j / N}$  (IDFT, zero terms dropped)

Hence  $y_l = \frac{1}{N} \sum_{\nu=1}^m \hat{x}(\omega_\nu) e^{2\pi i \omega_\nu (b+l\sigma) / N}$  ( $l=0, \dots, K-1$ ).

Hence  $\hat{y}(\psi) = \frac{1}{\sqrt{KN}} \sum_{l=0}^{K-1} e^{-2\pi i \psi l / K} y_l$

$$= \frac{1}{\sqrt{KN}} \sum_{\nu=1}^m \left[ \sum_{l=0}^{K-1} e^{-2\pi i \psi l / K} e^{2\pi i \omega_\nu (b+l\sigma) / N} \right] \hat{x}(\omega_\nu)$$

(\*)

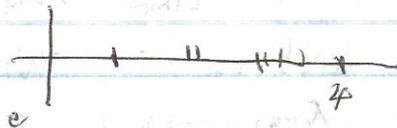
$$(*) = \sum_{l=0}^{K-1} e^{2\pi i \omega_\nu b / N} e^{2\pi i (\sigma \omega_\nu / N - \psi / K) l}$$

$$= e^{2\pi i \omega_\nu b / N} \frac{e^{2\pi i (\sigma \omega_\nu / N - \psi / K) K} - 1}{e^{2\pi i (\sigma \omega_\nu / N - \psi / K)} - 1}$$

\* Denominator close to zero if

$$\sigma \omega_\nu / N \approx \psi / K \pmod{1}$$

suppose  $\frac{\omega_1}{N}, \dots, \frac{\omega_m}{N}$  plots like this



after a multiplication by  $\sigma$  and then taking mod 1

Clusters will be broken up and  $\sigma w_u/N$  values will be well separated.

look for  $\frac{\psi}{k}$ 's (lie in  $(0,1)$ ) that come close to a  $\frac{\sigma w_u}{N}$ ; can detect these because  $\sum (\psi)$  will blow up.

try to recover  $w_1, \dots, w_m$  from this information.

Second idea from Gilbert et al is Bit testing.

any integer in the range  $0, \dots, 2^p - 1$  may be written as a sequence of  $p$  binary digits (called bits)

Fix  $u$ , write  $w_u$  as  $= d_p \dots d_0$

To determine  $d_0$ , choose  $\psi$  st.  $\sum (\psi)$  Large.   
 bit representation of  $w_u$

From above derivation.

$$\sum (\psi) = \frac{1}{jkN} \sum_{k=0}^{j-1} x^{jk w_u} \underbrace{\text{same } \psi}_{\text{say } k=0 \text{ to be recovered.}}$$
$$\frac{e^{2\pi i (k \sigma w_u/N - \psi)} - 1}{e^{2\pi i (k \sigma w_u/N - \psi/k)} - 1} = e^{2\pi i w_u b/N}$$

Now repeat with

$$\sum_{k=0}^{j-1} x^{k \psi} = x^{b \psi/2} + L_0 \pmod{N} \text{ for } k=0, \dots, k-1$$

By same analysis,

$$\sum (\psi) = \{ \text{all the same} \} e^{2\pi i w_u b/N} \cdot e^{2\pi i N/2 \cdot w_u/N}$$

the algorithm checks whether  $\sum (\psi) \approx \sum (\psi)$  or  $\sum (\psi) \approx -\sum (\psi)$

$$\sum (\psi) = \begin{cases} 1 & w_u \text{ even} \\ -1 & w_u \text{ odd} \end{cases}$$

This determines the trailing bits of  $w_u$

This will fail if  $\sum (\psi)$  is large because  $\psi/k$  is close to

two or more values of  $\frac{\sigma w_u}{N} \pmod{1}$ .  
Repeat the same trick with  $h$ .

$b + N/4, b + N/8, \dots$  to recover all bits of all  $w_u$ 's.

Hilroy

AER <sup>max</sup>

Jan 20, 2009.

Gilbert Algorithm  $\mathbb{C}^N$  / Fourier transform

Given a signal  $\underline{x}$  st.  $\underline{x}$  is sparse (or approximately) say  $\|\hat{\underline{x}}\|_0 = m$   $m \ll N$

Wish to sample a small number of entries of  $\underline{x}$  to determine  $\underline{x}$ .  
select a subsequence of equally spaced values of  $\underline{x}$ , say

$$x_{(b+kL) \bmod N} \quad k=0, \dots, k-1.$$

Fill this subsequence  $\sum_0, \dots, \sum_{k-1}$  Compute  $\sum_0^1$  Look for  $\psi$ 's in  $0, 1, \dots, k-1$  st.  $\sum_0^1(\psi)$  is large, these  $\psi$ 's are st.  $\frac{\psi}{k} \approx \frac{\sigma W_\psi}{N}$  where  $W_\psi$  is one of the  $N$  positions in which  $\hat{x}$  is non zero.

using bit-testing (repeat this calculation for different values of  $b$ ) can determine bits of  $W_\psi$ .

this procedure fails if choice of  $\sigma$  unlucky.

$$\frac{\sigma W_{\psi_1}}{N} \approx \frac{\sigma W_{\psi_2}}{N} \pmod{1} \quad \text{— in this case, bit testing fails.}$$

Therefore, procedure must be repeated to increase prob. of success.

Once  $W_\psi, \psi=1, \dots, m$  Found, need additional procedure to determine  $\hat{x}(W_1) \dots \hat{x}(W_m)$

Gilbert et al running time estimate.

suppose that there exists a good approximation  $R$  to  $\hat{x}$  st.

$$\|R\|_0 \leq m \quad \frac{\|\hat{x}\|_2}{\|\hat{x}-R\|_2} \leq M$$

suppose acceptable prob. of failure is  $\delta$ .

supp. seek  $R'$  st.

$$\|R'\|_0 \leq m, \quad \|\hat{x}-R'\|_2 \leq (1+\epsilon) \|\hat{x}-R\|_2$$

$$\text{Then \# OPS} \leq m \cdot \left(\log \frac{1}{\delta}\right)^{c_1} \cdot (\log N)^{c_2} \cdot (\log m)^{c_3} \cdot \left(\frac{1}{\epsilon}\right)^{c_4} \cdot (\log M)^{c_5}$$

$\uparrow$   
 $\log_2 N$  bits  
in each  $W_\psi$

$\underbrace{\hspace{10em}}_{\text{check } \delta}$

Strong compressive sensing You select a sampling operator once, almost any choice works with high probability. Then the measurements are successfully recovered exactly sparse signal.

Weak compressive sensing: For each sparse signal, almost every measurement operator recovers it. andes, Tomer, Tao.

Thm: Let  $V$  be a randomly chosen  $k$ -Gilbert et al. subspace of  $\mathbb{R}^n$ . Then with prob. (over choice of  $V$ ) of  $1 - P_0(k, n)$ ,  $V$  has the  $f_1(k, n)$  approximate spherical section property sth small for  $\forall v \in V, \|v\|_1 \geq f_1(k, n) \|v\|_2$  determined by pf.

PF: — Donoho, PISZER, S.V. (Stephen Vavasis)

choose  $v = \text{range}(B)$  where  $B \in \mathbb{R}^{n \times k}$  with indep. normally distributed  $(\mathcal{N}(0, 1))$  entries.

Equiv. to: choosing  $v = \text{null}(B')$  where  $B' \in \mathbb{R}^{(n-k) \times n}$ ,  $\mathcal{N}(0, 1)$  indep. entries.

Recall; A random variable  $X$  has pdf.  $f(x)$

if  $f(x)$  is nonnegative, integrable,  $\int_{-\infty}^{\infty} f(x) dx = 1$

and  $\text{Prob}(x \in [a, b]) = \int_a^b f(x) dx$ .

most famous pdf. is for  $\mathcal{N}(\mu, \sigma^2)$

Normal of Gaussian Distribution with mean  $\mu$  and variance  $\sigma^2$ .

has pdf.  $\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

special case  $\mathcal{N}(0, 1)$  has pdf  $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  for standard normal.

Start with a single  $x \in \mathbb{R}^k$ , will show that

$$\|Bx\|_1 \geq f_1(k, n) \|Bx\|_2$$

in  $V$  since  $v \in \text{range}(B)$

$B \in \mathbb{R}^{n \times k}$  as above  
 $B_{ij} \sim \mathcal{N}(0, 1)$

$X$  not random,  $B$  is

WLOG.  $\|X\|_2 = 1$

Claim 1: There is exponentially small prob. that

$$\|BX\|_2 \geq \sqrt{c_1 n}$$

proof of claim 1: A typical entry of  $BX$  is  $B(i, :) \cdot X = \sum_{j=1}^n B(i, j) X_j$

Fact: A sum of  $k$  indep. Gaussians with  $\mu_1, \dots, \mu_k$  and variances  $\sigma_1^2, \dots, \sigma_k^2$  is also gaussian with mean  $\mu_1 + \dots + \mu_k$  and variance  $\sigma_1^2 + \dots + \sigma_k^2$

Hence  $B(i, :) \cdot X$  is Gaussian with mean zero, and variance

$$X(1)^2 + \dots + X(n)^2 = 1 \text{ by assumption } (\|X\|_2 = 1)$$

Thus,  $B(i, :) \cdot X \sim \mathcal{N}(0, 1)$  Thus

$$\|BX\|_2^2 = (B(1, :) \cdot X)^2 + \dots + (B(n, :) \cdot X)^2$$

This is  $\chi_n^2$  distribution with freedom  $n$ .

It has pdf  $\frac{x^{n/2-1} \cdot e^{-x/2}}{\Gamma(n/2) 2^{n/2}}$

Recall  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$  if  $z$  is pos. integer.

$$\Gamma(z) = (z-1)!$$

Assume  $n$  Even.

$$\text{This prob } (\|BX\|_2^2 > M) = \frac{1}{\Gamma(n/2) 2^{n/2}} \int_M^\infty x^{n/2-1} e^{-x/2} dx$$

Change variable  $\xi = x - M$

$$= \frac{1}{\Gamma(n/2) 2^{n/2}} \int_0^\infty (\xi + M)^{n/2-1} e^{-(\xi+M)/2} d\xi$$

$$= \frac{e^{-M/2}}{\Gamma(n/2) 2^{n/2}} \int_0^\infty \sum_{i=0}^{n/2-1} \binom{n/2-1}{i} M^{n/2-1-i} \xi^i e^{-\xi/2} d\xi$$

$$= \frac{e^{-M/2}}{\Gamma(n/2) 2^{n/2}} \sum_{i=0}^{n/2-1} \binom{n/2-1}{i} M^{n/2-1-i} \int_0^\infty \xi^i e^{-\xi/2} d\xi = \int_0^\infty (2\psi)^i e^{-\psi} d\psi$$

$$= \frac{e^{-M/2}}{\Gamma(n/2) 2^{n/2}} \sum_{i=0}^{n/2-1} M^{n/2-1-i} \binom{n/2-1}{i} i! 2^{i+1} \quad d\psi = d\xi/2 = 2^{i+1} \cdot i!$$

$$= \frac{e^{-M/2}}{\Gamma(\frac{n}{2}) 2^{n/2}} \sum_{i=0}^{n/2-1} M^{\frac{n}{2}-i} \cdot 2^{2i+1} \cdot \underbrace{\left(\frac{n}{2}-1\right) \dots \left(\frac{n}{2}-i\right)}_{M^{\frac{n}{2}-1} \cdot 2^i \cdot M^{\frac{n}{2}-2} \cdot 2^2 \cdot \left(\frac{n}{2}-1\right) + M^{\frac{n}{2}-3} \cdot 2^3 \cdot \left(\frac{n}{2}-1\right) \left(\frac{n}{2}-2\right)}$$

The ratio of second to 1st term

$$\frac{M^{\frac{n}{2}-2} \cdot 2^2 \left(\frac{n}{2}-1\right)}{M^{\frac{n}{2}-1} \cdot 2} = \frac{2\left(\frac{n}{2}-1\right)}{M}$$

ratio 3rd to 2nd =  $\frac{2\left(\frac{n}{2}-2\right)}{M}$

If  $M \geq 2n$  Then all these ratios are  $\leq \frac{1}{2}$

∴ in this case, sum is bounded by twice first term.

$$\leq \frac{e^{-M/2}}{\Gamma(\frac{n}{2}) 2^{n/2}} 4 \cdot M^{\frac{n}{2}-1}$$

take  $M = 8n$  Then  $P(\|BX\|_2^2 \geq M) \leq \frac{e^{-4n}}{\Gamma(\frac{n}{2})! 2^{n/2}} \cdot 4(8n)^{\frac{n}{2}-1}$

LOG of RHS =  $-4n - \log(\frac{n}{2}-1)! - \frac{n}{2} \log 2 + \log 4 + (\frac{n}{2}-1)(\log n + \log 8)$

$$[\log(\frac{n}{2}-1)! = \frac{n}{2} \log \frac{n}{2} - \frac{n}{2} + O(\log n) \text{ Stirling}]$$

$$= -4n - \left(\frac{n}{2} \log \frac{n}{2}\right) + \frac{n}{2} + O(\log n) - \frac{n}{2} \ln 2 + \log 4 + \frac{n}{2} \log n + \frac{n}{2} \log 8 - \log n - \ln 8$$

$$= -4n + \frac{n}{2} \log 2 - \frac{n}{2} - \frac{n}{2} \log 2 + \frac{n}{2} \log 8 + O(\log n)$$

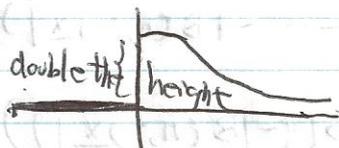
$$= -\text{const} \cdot n + O(\log n)$$

Thus  $\text{prob}(\|BX\|_2^2 \geq 8n) \leq e^{-c_1 n}$  Concludes claim 1.

claim 2: Exponentially small prob. that

$$\|BX\|_1 \leq C_2 n$$

Proof:  $B(i, : )X \sim \mathcal{N}(0, 1)$   $|B(i, : )X|$  follows the half normal distribution



Jan 21, 2009

Facts: mean  $\sqrt{2/\pi}$  variance  $\sigma^2 = 1 - \frac{2}{\pi}$

Thm: (Alternative version of "KKG" theorem)

$V \subset \mathbb{R}^n$   $k$ -dim. chosen at random.

Then with prob  $\geq 1 - P_1(k, n) \quad \forall v \in V$

$$\|v\|_1 \geq f_1(k, n) \|v\|_2$$

proof: choose  $B \subset \mathbb{R}^{n \times k}$  matrix at random  $V = \text{range}(B)$

Let  $x \in \mathbb{R}^n \quad \|x\|_2 = 1$  (Deterministic)

claim 1.  $\text{Prob}(\|Bx\|_2^2 \geq \delta n) \leq e^{-2n}$   
random

return to proof: suppose take  $M = \varrho n \quad \varrho \geq 2$ .

Then obtain for suff. large  $n$ ,

claim 1q  $\text{Prob}(\|Bx\|_2^2 \geq \varrho n) \leq e^{-\frac{1}{2}n(\log \varrho - \varrho)}$

(more general)

Recall strategy of proof: to show that very unlikely that

$\|Bx\|_2$  is large

$\|Bx\|_1$  is small

claim 2: Exponentially small probability, that  $\|Bx\|_1 \leq \varrho n$

$B(i, :)x$  follows the half-normal distribution.

has PDF  $f(x) = \begin{cases} \frac{2}{\sqrt{\pi}} e^{-x^2/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$$\text{Prob}(\|Bx\|_1 \leq \varrho n) = \text{Prob}(|B(1, :)x| + \dots + |B(n, :)x| \leq \varrho n)$$

Let  $\varphi$  be  $\varphi(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$

$$= \text{Prob}(\varrho n - |B(1, :)x| - \dots - |B(n, :)x| \geq 0)$$

$$= \mathbb{E}(\varphi(\varrho n - |B(1, :)x| - \dots - |B(n, :)x|))$$

Bernstein  $\varphi(x) \leq e^{hx} \quad \forall x \quad \forall h > 0$

$$\leq \mathbb{E}(e^{h(\varrho n - |B(1, :)x| - \dots - |B(n, :)x|)})$$

$$= \mathbb{E}(e^{h\varrho n} e^{-h(|B(1, :)x| + \dots + |B(n, :)x|)})$$

FAT  $E(XY) = E(X)E(Y)$  ZF  $X \perp Y$

$$\mathbb{E} \left[ \exp(h(q - |B(v):)X|) \right]$$

select one factor

$$E(\exp(h(q - |B(v):)X|))$$

~ Think of this as a random variable following half normal distribution

$$= \int_0^\infty \exp(h(q-t)) H(t) dt \quad H(t) = \begin{cases} \frac{\sqrt{2}}{\pi} e^{-t^2/2} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

$$= \int_0^\infty \frac{\sqrt{2}}{\pi} \exp(hq) \cdot \underbrace{\exp(h(-t)) e^{-t^2/2}}_{\text{focus on this}} dt$$

$$\int_0^\infty \exp(-ht - t^2/2) dt$$

Aside: for  $a \geq 0$   $\text{erfc}(a) \equiv \int_a^\infty e^{-x^2/2} dx$  Change of variables  $v = x-a$

$$\text{erfc}(a) \leq \frac{\exp(-a^2)}{a \cdot \text{const}}$$

$$= \int_0^\infty e^{-(v+a)^2/2} dv$$

$$= \int_0^\infty e^{-v^2/2 - 2va - a^2/2} dv$$

$$= e^{-a^2/2} \int_0^\infty e^{-v^2/2 - 2va} dv$$

Let  $t = v\sqrt{2}$ ,  $\frac{t^2}{2} = v^2$ ,  $dv = dt/\sqrt{2}$

$$\text{erfc}(a) = \frac{e^{-a^2}}{\sqrt{2}} \int_0^\infty e^{-t^2/2 - \sqrt{2}at} dt$$

This  $\frac{e^{-a^2}}{\sqrt{2}} \int_0^\infty e^{-t^2/2 - \sqrt{2}at} dt \leq \frac{e^{-a^2}}{a \cdot \text{const} \sqrt{2}}$

End of aside

hence,  $E(\exp(h(q - |B(v):)X|)) \leq \exp(hq) \frac{\text{const}}{h}$

Take  $h \sim \frac{1}{q}$  such that  $\exp(hq) = \frac{e}{\text{const}}$

Thus  $E(\dots) \leq \text{const} \cdot q$

Thus  $P(\|BX\|_1 \leq qn) \leq (\text{const} \cdot q)^n$

Claim 3:  $\text{prob}(\|BX\|_1 \geq qn) \text{ exponentially small}$

Using Bernstein analysis.

$$\text{Prob}(\|BX\|_1 \geq qn) \leq \alpha^n \quad \text{where}$$

$$\alpha = \int_{\frac{h}{2}}^{\infty} \frac{1}{\sqrt{\pi}} \exp(-ht) \int_0^{\infty} \exp(ht - \frac{t^2}{2}) dt$$

$$\int_0^{\infty} \exp(ht - \frac{t^2}{2}) dt \leq \int_{-\infty}^{\infty} \exp(ht - \frac{t^2}{2}) dt$$

$$u = t - h \Rightarrow du = dt$$

$$ht - \frac{t^2}{2} = -\frac{u^2 + h^2}{2}$$

$$= \int_{-\infty}^{\infty} e^{-h^2/2} \cdot e^{-u^2/2} du = \text{const.} \cdot e^{-h^2/2}$$

$$\text{Thus } \alpha \leq \text{const.} \cdot \exp\left(\frac{h^2}{2} - hq\right)$$

Take  $h=q$

$$\leq \text{const} \exp(-qn^2/2)$$

$$\text{Thus } \text{prob}(\|BX\|_1 \geq qn) \leq e^{-qn^2/2} \cdot \text{const.}$$

Let  $S^{k-1}$  be the  $k$ -dim 2-norm unit sphere lying in  $\mathbb{R}^k$ :

$$\text{ie. } \{x \in \mathbb{R}^k : \|x\|_2 = 1\}$$

A  $\delta$ -net  $W$  lying in  $\mathbb{R}^k$  is a set of points

$$\{x_1, \dots, x_n\} \text{ st. } \forall y \in S^{k-1}, \exists x \in W \text{ st. } \|y - x\|_2 \leq \delta$$

Claim 4: There exists a  $\delta$ -net in  $S^{k-1}$  with at most  $(1 + \frac{2}{\delta})^k$  points.

Proof: Build net  $W$  by starting with  $W = \emptyset$  and adding points one at a time.

add  $x$  if  $\|x - x_i\|_2 > \delta \quad \forall x_i \text{ already in } W$ .

Stop when no such  $x$  can be found.

When procedure stops,  $W$  is a  $\delta$ -net. If not,  $\exists y \in S^{k-1}$  st.

$\|y - x_i\|_2 > \delta$  for all  $x_i \in W$ , then  $y$  should be added to  $W$

Upon completion, if  $x_i, x_j$  are distinct points in  $W$ .

Then  $\|x_i - x_j\| > \delta$ , Thus  $B(x_i, \delta/2), \dots, B(x_N, \delta/2)$

are disjoint, where  $W = \{x_1, \dots, x_N\}$

and  $B(x, r)$  means <sup>Euclidean</sup> ball of radius  $r$  about  $x$ .

But these  $N$  balls all lie inside  $B(0, H \frac{\delta}{2})$  since

$$\|x_i\| = 1 \text{ for } i=1, \dots, N$$

Facts  $\text{Vol}(B(x, r)) = c_k \cdot r^k$   
 $x \in \mathbb{R}^k$

$$\text{Thus } \text{Vol}(B(x_1, \delta/2)) + \dots + \text{Vol}(B(x_N, \delta/2)) \leq \text{Vol}(B(0, H \frac{\delta}{2}))$$

$$\Rightarrow N c_k \left(\frac{\delta}{2}\right)^k \leq c_k \left(H \frac{\delta}{2}\right)^k$$

$$\Rightarrow N \leq \left(H \frac{2}{\delta}\right)^k \dots \text{OBS: proof does not require Euclidean,}$$

Claim 5, Apply union bound to claim 1, 2, 3.

Let  $x_1, \dots, x_N \in S^{k-1}$

a)  $\text{prob}(\|B x_i\|_2 \geq \epsilon_n) \leq N e^{-n(\log 2 - \epsilon)}$   
 for some  $i$

b)  $\text{prob}(\|B x_i\|_1 \leq \epsilon_n) \text{ for some } x_i \leq N (\text{const} \cdot \epsilon)^n$   
 $\downarrow$   
 $B$  random here

c)  $\text{prob}(\|B x_i\|_1 \geq \epsilon_n) \text{ for some } i \leq N e^{-(\frac{\epsilon^2}{2} + \epsilon \cdot \text{const})n}$

Claim 6: <sup>Prister</sup> suppose  $\{x_1, \dots, x_N\}$  form a  $\delta$ -net, then

$\forall x \in S^{k-1}$ , For any  $B$ ,  $\|B x\|_{\square} \leq \frac{1}{1-\delta} \max_{i=1, \dots, N} \|B x_i\|_{\square}$  for 2-norm.

proof: write  $x$  as an infinite summation

Find  $i$  st.  $\|x - x_i\|_2 \leq \delta$ . Define  $y_0 = x_i$

Find  $i'$  st.  $\left\| \frac{x - y_0}{\|x - y_0\|_2} \right\|_{\square} \leq \delta$ . Let  $y_1 = x_{i'}$ .  $\|x - y_0\|_2$

Hilroy

Jan 27, 20

Thm.  $\forall$  A random  $k$ -Dim subspace of  $\mathbb{R}^n$ .

Then with prob.  $\geq 1 - n^{-c(k,n)}$   $\forall v \in V$

$$\|v\|_1 \geq f_1(k,n) \|v\|_2$$

Claim 5 (Union bound) Let  $x_1, \dots, x_N$  lie in  $S^k$ . Then

(a) Prob  $\{ \|Bx_i\|_2 \geq \sqrt{n} \text{ for some } i \} \leq$

$$N \exp(-(\log 2 - \epsilon)n)$$

(b) Similar for prob  $\{ \|Bx_i\|_1 \leq \epsilon n \}$

(c) - - - -  $\|Bx_i\|_1 \geq \epsilon n$

Claim 6. Supp  $\{x_1, \dots, x_N\}$  form a  $\delta$ -net in 2-norm.

Then  $\forall B, \forall x \in S^{k-1}$

$$\|Bx\|_{\square} \leq \frac{1}{1-\delta} \max_{i=1, \dots, N} \|Bx_i\|_{\square}$$

PF: write  $x$  as infinite sum as follows

Find  $i$ , st.  $\|x_i - x\|_2 \leq \delta$ , Let  $y_0 = x_i$

Let  $\tilde{x} = \frac{x - y_0}{\|x - y_0\|_2}$  Find  $i'$  st  $\|\tilde{x} - x_{i'}\| \leq \delta$  Let  $y_1 = x_{i'}$

Let  $\tilde{\tilde{x}} = \frac{\tilde{x} - y_1}{\|\tilde{x} - y_1\|} \parallel x - y_0 \|_2 = \delta$   
 $\parallel \tilde{\tilde{x}} - y_1 \parallel \parallel x - y_0 \|_2$  etc. Each step.

obtain a unit vector, approx. it, subtract off approx.

normalize the residue/remainder

can check  $\|x - y_0\| \leq \delta$   
 $\|x - y_0 - y_1\| \leq \delta^2$   
 $\vdots$   
 And  $\|y_j\| \leq \delta^j$  for  $j=1, 2, \dots$

For each  $j$ ,  $\frac{\|y_j\|}{\|y_1\|_2}$  is in  $\delta$ -net

To prove the claim:

$$\|Bx\|_{\square} = \|B \sum_{j=0}^{\infty} y_j\|_{\square} \leq \sum_{j=0}^{\infty} \|By_j\|_{\square}$$

$$= \sum_{j=0}^{\infty} \left\| \frac{By_j}{\|y_j\|_2} \right\|_{\square} \cdot \|y_j\|_2$$

$$\leq \sum_{i=0}^{\infty} \left( \max_{i=1, \dots, N} \|BX_i\|_{\square} \right) \cdot \|y_i\|_2$$

$$\leq \max_{i=1, \dots, N} \|BX_i\|_{\square} \left( 1 + \delta + \delta^2 + \dots \right) = \frac{1}{1-\delta}$$

Claim 7: Suppose  $\{x_1, \dots, x_N\}$  form a  $\delta$ -net. then  
 $\forall x \in S^{k-1}, \|Bx\|_{\square} \geq \min_{i=1, \dots, N} \|Bx_i\|_{\square} - \frac{\delta}{1-\delta} \max_{i=1, \dots, N} \|Bx_i\|_{\square}$

PF: Construct the same sequence of  $y_i$ 's as in claim 6.

Then

$$\begin{aligned} \|Bx\|_{\square} &= \left\| B \sum_{i=0}^{\infty} y_i \right\|_{\square} \geq \|By_0\|_{\square} - \|By_1\|_{\square} - \dots \\ &\geq \min_{i=1, \dots, N} \|Bx_i\|_{\square} - \max_{i=1, \dots, N} \|Bx_i\|_{\square} \left[ \delta + \delta^2 + \dots \right] \\ &= \frac{\delta}{1-\delta} \end{aligned}$$

proves the claim.

Combine claims to prove Thm.

prob(  $\|Bx\|_1 \leq \epsilon n$  for ~~some~~ any  $x \in S^{k-1}$  )  
 by claim 7.

Thus with prob.  $\geq 1 - N(c\epsilon_1)^n - N \exp(-\frac{q_2^2}{2} n)$   
 at least

$$\begin{aligned} \min_{i=1, \dots, N} \|Bx_i\|_1 &\geq \epsilon_1 n \\ \max_{i=1, \dots, N} \|Bx_i\|_1 &\leq q_2 n \end{aligned}$$

so by claim 7

$$\min_{x \in S^{k-1}} \|Bx\|_1 \geq \epsilon_1 n - \frac{\delta}{1-\delta} q_2 n \text{ with same prob.}$$

So must select  $\epsilon_1, q_2, \delta$  to make this positive.

So i.e. select  $\epsilon_1 = .1, q_2 = 10, \delta = 1/200$ .

$$\text{In this case } N = \left( 1 + \frac{2}{\delta} \right)^k = \text{const }^k$$

Aside  $\|Bx\|_1 \geq \min_{i=1, \dots, N} \|Bx_i\|_1 - \frac{\delta}{1-\delta} \max_{i=1, \dots, N} \|Bx_i\|_1$

By claim 5 with prob.  $\geq N \exp(-c\epsilon)^n$

$$\begin{aligned} \min_{i=1, \dots, N} \|Bx_i\|_1 &\geq \epsilon n \\ \text{prob} &\geq N \exp(-\frac{\epsilon^2}{2} n) \\ \max_{i=1, \dots, N} \|Bx_i\|_1 &\leq \epsilon n \end{aligned}$$

so provided  $n \geq (\text{const. factor})^k$ , The decreasing factors

$(1 - q_1)^n \exp(-q_2^2 n/2)$  decrease faster than  $n$  increases.

so exponentially small prob that

$\min_{x \in S^{k-1}} \|Bx\|_1$  smaller than  $\text{const} \cdot n$

for the same reason exponentially small prob. that

$\max_{x \in S^k} \|Bx\|_2$  exceeds  $\text{const} \cdot \sqrt{n}$ .

use union bound last time

provided  $n \geq \text{const factor } k$ .

Exponentially small prob ( $\exp(\text{const} \cdot n + \text{const} \cdot k)$ )

That  $\frac{\|Bx\|_1}{\|Bx\|_2} \leq \text{const} \cdot \sqrt{n}$  for  $\forall x \in S^{k-1}$

(compare to version of KGA Thm in Zhang's paper with prob at least  $1 - e^{-ck}$ )

$\frac{\|Bv\|_1}{\|Bv\|_2} \geq \text{const} \frac{\sqrt{n-k}}{\sqrt{k \log(1 + \frac{k}{n-k})}}$  Before  $m = n - k$

For  $k =$  for small  $\text{const } n$

Result's about same (except for const's)

For  $k$  very small new thm better

As  $k$  approaches  $n$ , new thm loses strength.

### Candes and Tao near optimal signal Recovery

Given orthonormal basis  $\varphi_1, \dots, \varphi_N$  of  $\mathbb{R}^N$ , say that

$f \in \mathbb{R}^N$ , lies in the weak  $\ell_p$ -ball of radius  $R$

written as  $f \in W_p(R)$  if denoting by  $\theta_1, \dots, \theta_N$

the values of  $\varphi_1^T f, \dots, \varphi_N^T f$

Express  $f$  in basis  $\varphi_1, \dots, \varphi_N$

sorted into decreasing order by magnitude.  
 The inequality  $|\theta_i| \leq R i^{-1/p}$  holds for all  $i$   
 (called a power law)

FACT:  $f \in wlp(R)$   $\Rightarrow$   $\|\Phi^T f\|_p \leq (\ln(NH))^{1/p} R$  ( $p$  can be less than 1 or zero.)

$$\Phi^T = \begin{bmatrix} \phi_1^T \\ \vdots \\ \phi_n^T \end{bmatrix}$$

proof:  $\|\Phi^T f\|_p := \left( \sum_{i=1}^N |\phi_i^T f|^p \right)^{1/p} = \left( \sum_{i=1}^N |\theta_i|^p \right)^{1/p}$   
 $\leq \left( \sum_{i=1}^N (R i^{-1/p})^p \right)^{1/p} = R \left( \sum_{i=1}^N i^{-1} \right)^{1/p} \leq R (\ln(NH))^{1/p}$

partial converse:

$$\|\Phi^T f\|_p \leq R \Leftrightarrow$$

$$\sum_{i=1}^N |\theta_i|^p \leq R^p$$

For given  $\lambda > 0$  count # of  $i$  st  $|\theta_i| \geq \lambda$   
 same as #  $i$  st  $|\theta_i|^p \geq \lambda^p$  by above ineq.

This number is  $\leq \frac{R^p}{\lambda^p}$

Implies that  $|\theta_{\lfloor R^p/\lambda^p \rfloor + 1}| < \lambda$

Let  $i = \lfloor R^p/\lambda^p \rfloor + 1 \approx R^p/\lambda^p$

So  $\lambda \approx R i^{-1/p}$

And prev. ineq says

$$|\theta_i| \leq R i^{-1/p}$$

In practice, for correct  $\Phi$  (Fourier wavelet), commonly occurring signals lie in  $wlp(R)$  for a  $p \in (0, 1)$

Membership in  $wlp(R)$  implies that for any  $k=1, \dots, N$ , there is a good  $k$ -term estimate of  $\Phi^T f$ , namely,  $(\theta_1, \dots, \theta_k, 0, \dots, 0)$  permuted to correct position. (call this  $\theta_k(f)$ )

Aside for  $p < 2$   $\int_0^{\infty} k e^{-2t} dt = \frac{k}{1-2/p}$

How good is this approximation?

$$\| \Phi^T f - \Theta_k(f) \|_2 = \left( \sum_{j=k+1}^N \theta_j^2 \right)^{1/2}$$

$$\leq \left( \sum_{j=k+1}^N R^2 j^{-2/p} \right)^{1/2} \leq R k^{\frac{1}{2} - \frac{1}{p}} \cdot \text{const} \cdot C_p$$

since  $\Phi$  orthogonal.

$$\| f - \Phi \Theta_k(f) \| \leq R k^{\frac{1}{2} - \frac{1}{p}} \cdot C_p$$

If  $\Phi^T f$  has  $\leq k$  non-zero entries, then

$k$ -term approx. exact  $\forall k \geq R$

Jan 29, 2009. CO 769.

more in a minute

Thm (1.1 - C, T) Let  $W$  be a random  $k \times N$  matrix, ( $k < N$ )

Let  $f$  be in  $W L_p(\mathbb{R})$  for some  $0 < p < 1$ ,  $R > 0$ .

Basis  $\Phi$  or for  $p=1$ ,  $\|f\|_1 \leq R$

Let  $f^*$  solve  $\min \| \Phi g \|_1$  s.t.  $Wg = Wf$

then with prob  $\geq 1 - C_{\alpha, p} N^{-p/\alpha}$

$\alpha$  suff small arbitrary constant,  $p$  is fixed const  
prob. over choice of  $W$ , since  $W$  works for all  $f$

$$\| f - f^* \|_2 \leq C_{p, \alpha} R k^{\frac{1}{2} - \frac{1}{p}} (\log N)^{\frac{1}{p} - \frac{1}{2}}$$

Recall.  $\| f - \Phi \Theta_k(f) \| \leq R k^{\frac{1}{2} - \frac{1}{p}} C_p$

thus. Compressive sensing returns answer almost as good as if you knew

$R, p$  and all the leading entries of  $\Phi^T f$ .

Required properties of choice of random matrix.

$$= \Delta(N, p)$$

Property 1: UUP or Uniform uncertainty principle.

there is a  $\lambda$  st.

with prob. at least  $1 - C_{\alpha, p} N^{-p/\alpha}$  for all  $\alpha \in (0, \lambda)$   $W$  chosen

at random has the property that for all subsets  $T \subset \{1, \dots, N\}$

$$\text{st. } |T| \leq \alpha k / \lambda, \quad \frac{1}{2} \frac{k}{N} \leq \sigma_{|T|}^2 \left( \frac{W(T, \cdot)}{W(\cdot, T)} \right) \leq \sigma_1^2 \left( \frac{W(T, \cdot)}{W(\cdot, T)} \right) \leq \frac{3}{2} \frac{k}{N}$$

↑ Smallest singular values
 ↓ largest

Here  $\sigma_i(A)$  denotes the  $i$ th singular value of matrix  $A$ .  
Singular values are the <sup>sqrt of</sup> eigen values of  $A^T A$  listed in non-increasing order.

Some facts about singular values.

① The every  $A \in \mathbb{R}^{m \times n}$  has a SVD. If  $m \leq n$ , written as

$$A = Q \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_m & \\ & & & 0 \end{pmatrix} Z^T \quad \begin{matrix} m \times n \text{ ortho.} \\ \text{singular values.} \\ n \times n \text{ ortho.} \end{matrix}$$

Singular value  $\sigma_i(A)$  is a norm, denoted  $\|A\|$  or  $\|A\|_2$

FACT.  $\|A\|_2 = \sup \{ \|Ax\|_2 \mid \|x\|_2 = 1 \}$

If  $A \in \mathbb{R}^{m \times n}$ ,  $m \leq n$ , condition number of  $A$  is defined to be  $\sigma_1(A) / \sigma_m(A)$

UUP implies that every submatrix of  $w$  of appropriate size is well conditioned.

Property 2: ERP, Exact Reconstruction principle

$w$  has ERP with factor  $\lambda$  if for all suff. small  $\alpha > 0$  for every  $T \subset \{1, \dots, N\}$   $|T| \leq \alpha k / \lambda$

for every choice of  $\{s_i, i \in T\}$  s.t.  $s_i = \pm 1$ , with high prob.  $1 - C_{\alpha, p} N^{-p}$

over choice of  $w$  and  $T$ , there exists a  $\underline{w} \in \mathbb{R}^k$  s.t.

$W^T \underline{w}$  ( $\in \mathbb{R}^N$ ) equals  $\underline{s}$  for all  $i \in T$ , and ABS. val of entries of  $W^T \underline{w}$  indexed by  $\{1, \dots, N\} - T$  is  $\leq 1/2$

Zhang' Don't need UUP or ERP, just need/use spherical section property.

UUP is NOT preserved if  $w$  replaced by  $Gw$ , where  $G \in \mathbb{R}^{k \times k}$  non-singular. However, replacing  $w$  by  $Gw$  has no effect on compressive sensing since compressive sensing only  $\text{Null}(W)$  uses

Thm: (C-T)

① IF  $W_{i,j} \sim \mathcal{N}(0, 1/N)$  iid. Then  $w$  has UUP and ERP with  $\lambda = \log N$

*Alroy*

② If  $W$  has entries iid.  $W_{ij} = \begin{cases} 1/\sqrt{N} & \text{with prob } 1/2 \\ -1/\sqrt{N} & \dots 1/2 \end{cases}$   
has UUP and ERP.

③ For  $W$  is randomly chosen set of rows of DFT MTR,  $W$  has UUP and ERP with  $\lambda = (\log N)^b$ .

Thm(1.2) Restatement of 1.1)  
 $W$  chosen so that the conditions of UUP and ERP hold  
Factor  $\lambda_1$       Factor  $\lambda_2$

And  $f \in \mathbb{R}^N$  as above ( $f \in Wp(\mathbb{R}), 0 < p < 1$  or  $\|\Phi^T f\| \leq R$ )

Then  $\|f - f^*\|_2 \leq C_{p,\alpha} R (k/\lambda)^{\frac{1}{2}-\frac{p}{\alpha}}$

where  $f^*$  solves  $\min \| \Phi^T g \|_1$  st.  $Wg = wf$   
and  $\lambda = \max(\lambda_1, \lambda_2)$ .

From now on, take  $\Phi = I$ .

Lemma (2.1) Suppose that  $W$  has ERP, and supp  $f = f_0 + h$

st.  $\text{support}(f) = T \quad |T| \leq \alpha k/\lambda$

then with prob.  $1 - C_{\alpha,p} N^{-p/\alpha}$  solution  $f^*$  to  $\min \|g\|_1$  st.

$\|Wg\| = Wg = wf$  satisfies

$$\sum_{i \in T} |f_i^*| \leq 4 \|h\|_1$$

proof: Assume  $T$  valid for ERP

can find  $\underline{w} \in \mathbb{R}^k$  st. if

$P = W^T \underline{w}$  Then  $P_i = \text{sgn}((f_0)_i) \quad \forall i \in T \quad \|P_i\| \leq 1/2$

$$(f^*)^T P = (f^*)^T W^T \underline{w} = f^T W^T \underline{w} = (f_0 + h)^T P$$

since  $Wf^* = wf$

so  $(f^*)^T P \geq \|f_0\|_1 - \frac{1}{2} \|h\|_1$   
 $= f_0^T P$

on the other hand

$$|(f^*)^T P| \leq \sum_{i \in T} |f_i^* P_i| + \sum_{i \notin T} |f_i^* P_i|$$

$$\leq \sum_{i \in T} |f_i^*| + \frac{1}{2} \sum_{i \notin T} |f_i^*| - \sum_{i \notin T} |f_i^*|$$

$$= \underbrace{\sum_{i \in T} |f_i^*|}_{\|f^*\|_1} - \frac{1}{2} \sum_{i \notin T} |f_i^*|$$

put together  $\|f^*\|_1 - \frac{1}{2} \sum_{i \notin T} |f_i^*| \geq \|f\|_1 - \|h\|_1$

But also optimality of  $f^*$  in Opt prob.

$$\Rightarrow \|f^*\|_1 \leq \|f\|_1 \leq \|f_0\|_1 + \|h\|_1$$

$$\|f_0\|_1 + \|h\|_1 - \frac{1}{2} \sum_{i \notin T} |f_i^*| \geq \|f_0\|_1 - \|h\|_1$$

$$\Rightarrow \sum_{i \notin T} |f_i^*| \leq 4 \|h\|_1$$

Lemma says that if  $f$  small on a certain set of entries, so is  $f^*$ .

Claim If  $f \in W_p(R)$ , then  $f^*$  has a similar property.

Suppose  $f \in W_p(R)$ ,  $T$  indexes the  $|T|$  largest entries of  $f$

write  $f = f_0 + h$   
 $f_0$  supported on  $T$        $h$  supported on  $\bar{T}$

Feb 3, 2007, CO 769  
 Lemma  $W$  has ERP,  $f = f_0 + h$  st  $\text{supp}(f_0) = T$

$|T| \leq \alpha k / \lambda$ , then with high prob.  $1 - CN^{-p/\alpha}$

$\min \|g\|_1$ , st.  $Wg = wf$  satisfies

$$\sum_{i \in \bar{T}} |(f^*)_i| \leq 4 \|h\|_1$$

complement of  $T$ .

IF  $f \in W_p(R)$  and  $T$  indexes the  $|T|$  largest entries (in magnitude) of  $f$ , then

$$f = f_0 + h$$

supported on  $T$       supported on  $\bar{T}$

ESTIMATE:  $\|h\|_1 \leq R|T|^{1-1/p} C_p$

Corollary (2.2)  $|(f^*)|_m| \leq R|T|^{-1/p} C_p$  For  $m > 2|T|$

chrt largest entries of  $f^*$  Almost says  $f^* \in W_p(\mathbb{R})$

Pf: By Lemma,  $\|f^*(\bar{T})\|_1 \leq 4\|f(\bar{T})\|_1$

Like  $h$  is to lemma

Let  $E_m$  index the  $m$ th largest entries of  $f^*$

Say  $m > 2|T|$ , OBS.  $|E_m \cap \bar{T}| > m - |T|$

So  $\|f^*(E_m \cap \bar{T})\|_1 \geq (m - |T|) |(f^*)|_m| \geq |T| |(f^*)|_m|$

$\|f^*(E_m \cap \bar{T})\|_1 \leq \|f^*(\bar{T})\|_1 \leq 4R|T|^{1-1/p} C_p$

Combine inequalities  $|(f^*)|_m| \leq 4R|T|^{1-1/p} C_p$

Thm (3.1) suppose  $W$  satisfies the LUP inequalities.

Then  $\forall T, S, |T| \leq \alpha k/\lambda$ , and  $\forall f \in \mathbb{R}^T$

There exist  $f^{\text{ext}} \in \mathbb{R}^N, s.t.$

①  $f^{\text{ext}}(T) = f$

②  $\exists W \in \mathbb{R}^{k \times k}$  st.  $W^T W = f^{\text{ext}}$

③ for any  $E \subset \{1, \dots, N\}, \|f^{\text{ext}}(E)\|_2 \leq C \left( \frac{|E|}{\alpha k/\lambda} \right)^{1/2} \|f\|_2$

Aside: Facts ① Singular values of  $A^T A$  are squares of singular values of  $A$ .

② If  $A$  is square invertible, then singular values of  $A^{-1}$  are reciprocals of  $A$ .

Take  $f^{\text{ext}} = \underbrace{W^T}_{W(\cdot, T)} (W(\cdot, T))^T f$

② is obvious

To prove ①,

$$f^{\text{ext}}(T) = \underbrace{W(\cdot, T)^T W(\cdot, T)}_{I} (W(\cdot, T))^T W(\cdot, T) f = f$$

Proof of (3)

First supp.  $|E| \leq \alpha k/\lambda$

Then  
→

$$\|W(:, T)\| \leq \sqrt{3k/2N} \quad (\text{By UVP})$$

$$\|W(:, E)\| \leq \sqrt{3k/2N}$$

so

$$\begin{aligned} \|f^{\text{ext}}(E)\|_2 &= \|W(:, E)^T W(:, T) (W(:, T)^T W(:, T))^{-1} f\|_2 \\ &\leq \|W(:, E)\| \|W(:, T)\| \|(W(:, T)^T W(:, T))^{-1}\| \|f\| \\ &\leq \sqrt{3k/2N} \sqrt{3k/2N} \frac{2N}{k} \|f\| \\ &\leq 3\|f\| \end{aligned}$$

Consider

Larger E, supp. e.g.

$$E = E_1 \cup E_2 \cup \dots \cup E_L \quad \text{where } |E_i| \leq \alpha k/\lambda \quad \text{for } i=1, \dots, L$$

Disjoint partition

$$|E_i| \leq \alpha k/\lambda$$

$$\|f^{\text{ext}}(E)\|_2^2 = \|f^{\text{ext}}(E_1)\|_2^2 + \dots + \|f^{\text{ext}}(E_L)\|_2^2$$

$$\leq 9\|f\|^2 + \dots + 9\|f\|^2$$

$$= 9L\|f\|_2^2 \quad \text{so}$$

$$\|f^{\text{ext}}(E)\|_2 \leq 3\sqrt{L} \|f\|_2$$

$$\sqrt{L} = \sqrt{|E| / (\alpha k/\lambda)} \leq \sqrt{|E| / (\alpha k/\lambda)}$$

proves (3.1).

Proof of Thm (1.2), w.t.p.  $\|f - f^*\|_2 \leq C_{p,\alpha} \cdot R(k/\lambda)^{\frac{1}{2} - \frac{1}{p}}$

Let  $T_0$  be the indices of the  $\alpha k/\lambda$  largest entries of  $f$  and  $T_1$  same for  $f^*$ . And let  $T = T_0 \cup T_1$

so  $\alpha k/\lambda \leq |T| \leq 2 \alpha k/\lambda$

$$\|f(\bar{T}) - f^*(\bar{T})\|_1 \leq \|f(\bar{T})\|_1 + \|f^*(\bar{T})\|_1$$

$$\bar{T} \subset \bar{T}_0 \leq \|f(\bar{T}_0)\|_1 + \|f^*(\bar{T}_1)\|_1$$

use  $\in \text{wlp}(R)$   
 $\leq R|T|^{r/p} C_p$

use Cor. 2.2.

Apply to the first  $|T_1|$  over 2 entries of  $T_1$

Fact  $\|X\|_2 \leq \sqrt{\|X\|_1 \cdot \|X\|_\infty}$       Cpf:  $\sqrt{x_1^2 + \dots + x_n^2} \leq \sqrt{|x_1| \max |x_i| + \dots + |x_n| \max |x_i|}$

Also  $\|f(\bar{T}) - f^*(\bar{T})\|_\infty \leq \|f(\bar{T}_0)\|_\infty + \|f^*(\bar{T})\|_\infty \leq R|\bar{T}|^{-1/p} C_P \leq \sqrt{\|X\|_1 \cdot \|X\|_\infty}$

$\Rightarrow \|f(\bar{T}) - f^*(\bar{T})\|_2 \leq C|\bar{T}|^{\frac{1}{2} - \frac{1}{p}}$

Now we must consider

$\|f(\bar{T}) - f^*(\bar{T})\|_2$  case

Apply Thm (3.1) Find  $g = W^T W$  st.

for  $i \in \bar{T}$ ,  $g_i = f_i - f_i^*$  and

$\sum_{i \in E} g_i^2 \leq C \sum_{i \in E} (f_i - f_i^*)^2 \dots (*)$

for all  $E \subset \bar{T}$ , st.  $|E| \leq \alpha k/\lambda$

OBS:  $(f - f^*)^T g = (f - f^*) W^T W = 0$

Recall  $W f^* = W f$

So  $\sum_{i \in \bar{T}} (f_i - f_i^*) g_i = - \sum_{i \in \bar{T}} (f_i - f_i^*) g_i \quad (**)$

$= \|f(\bar{T}) - f^*(\bar{T})\|_2^2$   
by choice of  $g$ .

Partition  $\bar{T}$  into blocks of size  $|\bar{T}|$ , say  $|\bar{T}| = B_0 \cup B_1 \dots B_L$

$|B_0| = \dots = |B_{L-1}| = |\bar{T}|, \quad |B_L| \leq |\bar{T}|$

$B_0$  contains indices  $i \in \bar{T}$  where  $|f_i - f_i^*|$  largest

$B_1$  next largest, etc.

For  $\mu = 0, \dots, L$

$\sum_{i \in B_\mu} (f_i - f_i^*) g_i \leq \|f(B_\mu) - f^*(B_\mu)\|_2 \|g(B_\mu)\|_2$  / Cauchy-Schwarz

$\leq C \|f - f^*\|_2$

$\|f(B_\mu) - f^*(B_\mu)\|_2$

call this  $I_\mu$

Recall  $\|X\|_2 \leq \sqrt{\|X\|_1} \|X\|_\infty$

$$I_0 \leq \|f(B_0) - f^*(B_0)\|_{\infty} \sqrt{|T|} \leq C |T|^{\frac{1}{2} - \frac{1}{p}} \quad \cdot 37$$

Same reason as in 1<sup>st</sup> part of proof — large entries excluded.

For  $M \geq 1$

$$I_M \leq \sqrt{|T|} \cdot \|f(B_M) - f^*(B_M)\|_{\infty} \leq \sqrt{|T|} \cdot \text{Smallest entry of } f(B_{M-1}) - f^*(B_{M-1}) \\ \leq \sqrt{|T|} \cdot \frac{\|f(B_{M-1}) - f^*(B_{M-1})\|_1}{|T|}$$

The sum of  $i \in T$

$$\sum_{i \in T} (f_i - f_i^*) g_i = \sum_{\mu} \sum_{i \in B_{\mu}} (f_i - f_i^*) g_i \\ \leq \sum_{\mu} \|f - f^*\|_2 \cdot I_{\mu} = \|f - f^*\|_2 \cdot I_0 \\ + \sum_{\mu=1}^L I_{\mu} \\ \sum_{\mu=1}^L I_{\mu} \leq \frac{1}{\sqrt{|T|}} \|f(\bar{T}) - f^*(\bar{T})\|_1 \leq \frac{1}{\sqrt{|T|}} \cdot C \cdot |T|^{1 - \frac{1}{p}} \\ = C |T|^{\frac{1}{2} - \frac{1}{p}} \quad \downarrow \text{first part of proof.}$$

$$\Rightarrow \sum_{i \in T} (f_i - f_i^*) g_i \leq C |T|^{\frac{1}{2} - \frac{1}{p}} \|f - f^*\|_2$$

plug into (++)

$$\|f(T) - f^*(T)\|_2^2 \leq C |T|^{\frac{1}{2} - \frac{1}{p}} \|f - f^*\|_2$$

$$\|f(T) - f^*(T)\|_2 \leq C |T|^{\frac{1}{2} - \frac{1}{p}} \quad \text{Divide by this}$$

Feb 5, 2009.

Explicit (Deterministic) Construction of Compressive Sensing Matrix.

Why explicit construction. For use in a device, random choice might not be feasible, is easy way to verify LUP or K&G properties.

BURUSWAMI, LOE, RAZBOROV. Almost Euclidean Subspaces of

$L_N^1$  via Expander codes

Main theorem

For all  $N \geq \frac{1}{\eta}$  ( $\eta \rightarrow 0$  as  $N \rightarrow \infty$ )  
 st. one can construct a space  $V \subset \mathbb{R}^N$  st.  
 $\dim(V) = (1-\eta)N$  and for  $\forall v \in V - \{0\}$ ,

$$\frac{\|v\|_1}{\|v\|_2} \geq \frac{\sqrt{N} \eta^{-\log \log N}}{\log \log N^{C \log \log N}}$$

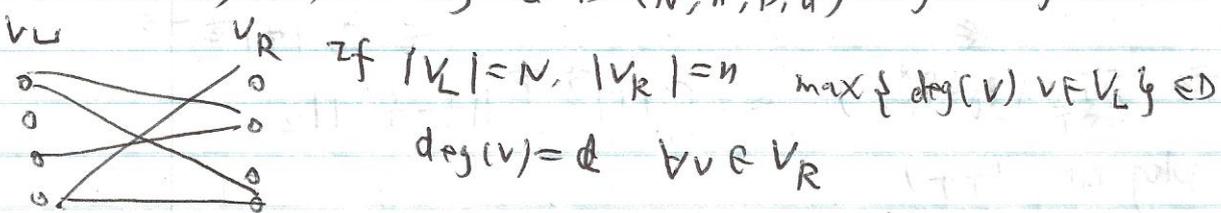
The best possible denominator

RGG. is  $\sqrt{N \log N}$

Note  $\log \log N^{C \log \log N}$  grows faster than  $\log N$ , but more slowly than  $N^\epsilon$  for any  $\epsilon > 0$

Bipartite graph

$G = (V_L, V_R, E)$  say  $G$  is  $(N, n, D, d)$  right regular.



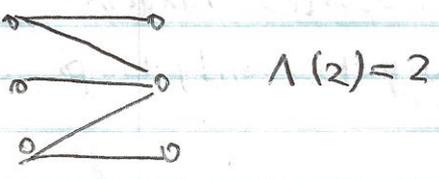
Edges have one end in  $V_L$ , and the other in  $V_R$

For a graph in which  $V$  is the vertices.

If  $S \subset V$ , write  $\Gamma(S) = \{ \text{neighbors of } S \}$

The expansion profile of  $G = (V_L, V_R, E)$  is defined by

$$\Lambda(m) = \min \{ |\Gamma(S)| : S \subset V_L, |S| \geq m \}$$



Lemma (2-3) Given a left-regular bipartite graph  $G$  st.  $N$  nodes in  $V_L$ ,  $n$  nodes in  $V_R$ ,  $D \leq \deg \leq D$  on left,  $D \leq \deg \leq d$  on right.

Can construct  $(N, n, D', \hat{d})$  right-regular graph  $G'$ .

st.  $D' \leq 2D$ ,  $n' \leq 2n$

PR: Let  $d = \lceil \text{Avg degree on right} \rceil = \lceil \frac{\sum_{v \in V_R} \deg(v)}{|V_R|} \rceil$   
(arbitrary deterministic)

Split each node in  $V_R$  into multiple nodes each with degree exactly  $d$ , except possibly the last one

Now insert edges on right to raise the nodes ( $\leq n$  of them) of degree  $< d$  on the right to degree exactly  $d$ .

Total # of new nodes added to  $V_R$

$$\# \text{ new nodes} \leq \sum_{v \in V_R} (\lceil \deg(v)/d \rceil - 1)$$

$$\leq \sum_{v \in V_R} \deg(v)/d$$

$$= \sum_{v \in V_R} \deg(v) / \lceil \frac{\sum_{v \in V_R} \deg(v)}{|V_R|} \rceil \leq |V_R|$$

Shows  $n' \leq 2n$ .

At most  $n$  <sup>new</sup> nodes have  $\deg < d$ , so # edges to be added to make new  $V_R$  regular  $\leq (d-1)n$

$$(d-1)n \leq \sum_{v \in V_R} \frac{\deg(v)}{|V_R|} \cdot n = \sum_{v \in V_R} \deg(v) \equiv \# \text{ original edges}$$

so # of original edges =  $D \cdot N$

so distribute the end points of new edges deterministically but arbitrarily so that each node in  $V_L$  get at most  $\leq D$  new edges, shows  $D' \leq 2D$ .

Thm (2.7) (Number theoretic Result) (construction of expander with a good profile for small  $m$ )

There exists universal  $\epsilon_0 > 0$  st. Following true.

Let  $p$  be a prime, consider  $G_p = (V_L, V_R, E)$  defined as follows

$$V_L = \{0, 1, \dots, p-1\}^3 \quad V_R = \{1, 2, 3, 4\} \times \{0, 1, \dots, p-1\}$$

$$|V_L| = p^3$$

$$|V_R| = 4p$$

$E$  connects  $(a, b, c)$  to  $(1, a)$   $(2, b)$   $(3, c)$  and  $(4, (abc) \bmod p)$

for all  $(a, b, c) \in \{0, \dots, p-1\}^3$

Then  $\lambda(m) \geq \min(p^{2/3}, m^{1/3 + \epsilon_0})$

Thm (2.8) There is an explicit  $(N, n, 8, \text{const } N^{2/3})$  right regular graph with  $\lambda(m) \geq \min(n^{2/3}, m^{1/3 + \epsilon_0})$

Aside: construction in Lemma 2.3 does not decrease  $\lambda(m)$  for any  $m$ .

Pf. let  $p$  be the smallest prime st.  $p^3 \geq N$  (check)

Note: number theory fact.

$$N^{1/3} \leq p \leq 2N^{1/3} \text{ construct op as in Thm 2.7.}$$

Delete an arbitrary set of  $p^3 - N$  nodes on left yielding  $H$

so that  $H$  has  $N$  vertices on left,  $4p$  on right.

left-regular of deg. 4.

Apply Lemma 2.3 to obtain  $(N, n, 8, d)$  right regular graph

deg.  $d \approx N^{2/3}$  about equal to  $n^2$  same (or better)

$\wedge$

Thm 2.6. (More complicated number theoretic construction)

Better for larger value  $m$ )

There exists explicit  $(N, n, 4, \text{const. } d)$  Expander for every  $d \geq 5$  and  $N \geq d$  st.

$$\lambda(m) \geq \min\left(\frac{m}{2d}, \sqrt{2Nm}\right)$$

Distortion of a subspace  $V \subseteq \mathbb{R}^n$  is defined by

$$\Delta(V) = \frac{1}{\inf_{v \in V, \|v\|_1 = 1} \|v\|_2}$$

K&G says for random  $V$  of Dim  $n \times m$ .  
distortion of  $V$ .

$$\Delta(V) \leq \sqrt{\frac{n}{m} \log\left(\frac{n}{m}\right)}$$

GLR seeks  $V$  with low  $\Delta(V)$

spread: say  $V \in \mathbb{R}^{n \times m}$  has a spread of  $(t, \epsilon)$  if for every  $S \subset \{1, \dots, n\}$  st.  $|S| \leq t$ , for every  $x \in V$

$$\|x\|_2$$

$$\|x(\bar{S})\|_2 \geq \epsilon \|x\|_2$$

i.e. Large entries of  $x$  can not be concentrated in a small # of positions.

Thm (2.11) (a) If  $V$  has spread  $(t, \epsilon)$  Then

$$\Delta(V) \leq \sqrt{\frac{n}{t}} \epsilon^{-2}$$

(b) If  $V$  has distortion  $\Delta$ , Then it has a spread of

$$\left(\frac{n}{\Delta^2}, \frac{1}{4\Delta}\right)$$

PR: maybe homework.

say:  $V$  is  $(t, T, \epsilon)$  -spread  $(t, \epsilon)$  if

$$\min_{\substack{S \subset \{1, \dots, n\} \\ |S| \leq T}} \|x(\bar{S})\|_2 \geq \epsilon \min_{\substack{S \subset \{1, \dots, n\} \\ |S| \leq t}} \|x(\bar{S})\|_2$$

$$(t, \epsilon) \text{-spread} \Leftrightarrow (0, t\epsilon) \text{-spread}$$

$$\Leftrightarrow \left(\frac{t}{2}, t, \epsilon\right) \text{-spread} \quad \text{— allow real numbers.}$$

Lemma (2.15) If  $X_1 \subset \mathbb{R}^n$  with spread

$$(t_0, t_1, \epsilon_1) \text{ and } X_2 \subset \mathbb{R}^n \text{ with spread } (t_1, t_2, \epsilon_2)$$

Then  $X_1 \cap X_2$  is  $(t_0, t_2, \epsilon_1, \epsilon_2)$  -spread.

Feb 10, 2009. 0769 Compressive Sensing

Corrections on KT Theorem

$$\sum_{i \in E} g_i \leq \sum_{i \in E} (f - f_i^*)$$

$$\sum_{i \in B_U} (f_i - f_i^*) g_i \leq \|f(B_U) - f^*(B_U)\|_2 \cdot \|g(B_U)\|_2$$

$$\leq I_U \cdot \|f(T) - f^*(T)\|_2$$

In subsequent derivations  $\|f - f^*\|_2$  should be  $\|f(T) - f^*(T)\|_2$

At start of proof:

- upper bound on  $\|f(\bar{T})\|_1$  by Assumption
- ...  $\|f^*(\bar{T})\|_1$  by Lemma 2.1
- ...  $\|f^*(\bar{T})\|_\infty$  by Assumption
- ...  $\|f^*(T)\|_\infty$  by Corollary 2.2

Thm 3.2 (Coding Theory) For all pos. integers st.  $k$  is a power of 4, and  $R \leq d \leq k^2/2$ , there exists an explicit  $k \times d$  matrix  $A$  st.

- ① Each entry of  $A$  is  $\pm 1/\sqrt{k}$ , so that  $\|A(:, j)\|_2 = 1 \quad \forall j$
- ② For all  $j, j'$  st  $j \neq j'$ ,  $|A(:, j)^T A(:, j')| \leq 1/\sqrt{k}$
- ③  $\|A\| \leq \sqrt{d/k}$

$D_{\max}(A)$   
 Lemma B-1) suppose  $A \in \mathbb{R}^{k \times d}$  satisfies

- ①  $|A(:, j)^T A(:, j')| \leq \tau \quad \forall j, j' \text{ st. } j \neq j'$
- ②  $\|A(:, j)\|_2 = 1 \quad \forall j$

Then  $\text{null}(A)$  is  $(\frac{1}{2\tau}, \frac{1}{2\|A\|})$ -spread.

proof: Supp  $S \subset \{1, \dots, d\} \quad |S| \leq \frac{1}{2\tau}$

Let  $t = |S|$  consider  $\mathcal{Q} = A(:, S)^T A(:, S)$

so  $\mathcal{Q}$  is an  $t \times t$  symmetric semi-definite

Further  $|\Phi(j, j)| = 1 \quad \forall j = 0, \dots, t$  and  $|\Phi(j, j')| \leq \tau$   
 $\forall j, j'$  st.  $j \neq j'$  so  
 $\Phi = I + \Psi$  and  $|\Psi(j, j')| \leq \tau \quad \forall j, j'$

$\Rightarrow \|\Psi\|_F \leq t\tau \Rightarrow \|\Psi\|_2 \leq t\tau \Rightarrow$  All eigen values of  $\Psi$   
 lie in  $[-t\tau, t\tau] \Rightarrow$  All eigen values of  $\Phi$  lie in  
 $[1-t\tau, 1+t\tau] \Rightarrow$  singular values of  $A(\cdot, s)$  lie in  
 $[\sqrt{1-t\tau}, \sqrt{1+t\tau}]$ . (Compare to UUP)

$$\Rightarrow \forall y \in \mathbb{R}^t \quad \|A(\cdot, s)y\|_2 \geq \sqrt{1-t\tau} \|y\|_2$$

By choice of  $t \quad \sqrt{1-t\tau} \geq \sqrt{1/2}$

$$|s| \leq \frac{1}{2t}$$

Choose  $x \in \text{Null}(A) - \{0\}$

$$Ax = 0 \Rightarrow A(\cdot, s)x(s) = -A(\cdot, \bar{s})x(\bar{s})$$

Combine with prev. ineq  $\Rightarrow$

$$\begin{aligned} \|A(\cdot, \bar{s})x(\bar{s})\|_2 &\geq \frac{1}{\sqrt{2}} \|x(\bar{s})\|_2 \\ &\leq \|A(\cdot, \bar{s})\|_2 \|x(\bar{s})\|_2 \\ &\leq \|A\|_2 \|x(\bar{s})\|_2 \end{aligned}$$

combine  $\|x(\bar{s})\|_2 \leq \sqrt{2} \|A\|_2 \|x(\bar{s})\|_2$

$$\Rightarrow \|x(\bar{s})\|_2^2 \leq 2 \|A\|_2^2 \|x(\bar{s})\|_2^2$$

Add  $\|x(\bar{s})\|_2^2$  to both sides we get

$$\|x\|_2^2 \leq (2 \|A\|_2^2 + 1) \|x(\bar{s})\|_2^2$$

$$\text{spread} \geq \frac{1}{\sqrt{2 \|A\|_2^2 + 1}} \geq \frac{1}{\sqrt{2 \|A\|_2^2 + 2 \|A\|_2^2}} = \frac{1}{2 \|A\|_2}$$

concl. of the prev. results: For every  $k$  a power of 4, every  $d$   
 s.t.  $k \leq d \leq k^2/2$ ,  $\exists$  explicit subspace of  $\mathbb{R}^d$  of dimension  
 $d-k$  whose spread is  $(\frac{\sqrt{k}}{2}, \frac{1}{4} \sqrt{\frac{k}{d}})$

Thm.

Defn: (4.1) Given a right regular bipartite graph  $G$ ,

$$|V_L| = N \quad |V_R| = n \quad \deg(v) = d \quad \forall v \in V_R$$

and also given subspace  $L \subset \mathbb{R}^d$

$$\text{Let } X(G, L) = \{ \underline{x} \in \mathbb{R}^N, \underline{x}(\Gamma(i)) \in L \quad \forall i \in V_R \}$$

claim:  $X(G, L)$  is a subspace

$$\dim(X(G, L)) \geq N - (d - \dim(L))n$$

Thm(4.2) suppose  $G$  is an  $(N, n, D, d)$  right-regular bipartite

graph with expansion profile  $\Delta(\cdot)$  suppose  $L$  is subspace of

$\mathbb{R}^d$  has spread  $(t, \epsilon)$

Then  $\forall T_0 \in (0, N]$

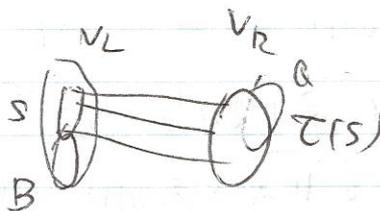
$$X(G, L) \text{ has spread } (T_0, \frac{t}{D} \Delta(T_0), \frac{\epsilon}{2D})$$

pf: consider  $\underline{x} \in X(G, L)$  st.  $\|\underline{x}\|_2 = 1$

Take  $S \subset \{1, \dots, N\}$  st.  $|S| \leq T$

$$T = \frac{t}{D} \Delta(T_0) \text{ - MUST show}$$

$$\|\underline{x}(S)\| \geq \frac{\epsilon}{\sqrt{2D}} \cdot \min_{|B| \leq T_0} \|\underline{x}(B)\|_2$$



$$\text{Let } Q = \{ j \in \Gamma(S) : |\Gamma(j) \cap S| > t \}$$

$$\text{Let } B = \{ i \in S : \Gamma(i) \subset Q \}$$

OBS:  $|Q| \leq \# \text{ edges btw } S \text{ and } \Gamma(S)$

because each entry of  $Q$  "uses up"  $> t$  edges connecting  $S$  to  $\Gamma(S)$

$$\leq D |S| \quad \text{max-deg over } V_L \text{ is } D$$

$$\leq DT \quad |S| \leq T$$

Thus  $|Q| < \frac{DT}{t} \leq \Delta(T_0)$  by def. of  $T$ .

Also  $|Q| \geq |C(B)| \Rightarrow |C(B)| < \Lambda(T_0)$

By defn of  $\Lambda \Rightarrow |B| < T_0$

thus B a candidate in 'nu', so suff to prove

$$\|X(\bar{s})\|_2 \geq \frac{\epsilon}{\sqrt{2D}} \|X(\bar{B})\|_2$$

$$\|X(\bar{B})\|_2^2 = \|X(\bar{s})\|_2^2 + \|X(s-B)\|_2^2$$

so every vertex  $s-B$  has at least 1 neighbour in  $C(s)-Q$

$$\Rightarrow \sum_{j \in C(s)-Q} \|X(\tau(j))\|_2^2 \geq \|X(s-B)\|_2^2$$

combine  $\|X(\bar{B})\|_2^2 \leq \|X(\bar{s})\|_2^2 + \sum_{j \in C(s)-Q} \|X(\tau(j))\|_2^2$

$\forall j \in C(s)-Q. |C(j) \cap S| \leq t$  By defn of Q.

$$\Rightarrow \sum_{j \in C(s)-Q} \|X(\underbrace{C(j)-S}_{=C(j)-(C(j) \cap S)} )\|_2^2 \geq \epsilon^2 \sum_{j \in C(s)-Q} \|X(C(j))\|_2^2$$

size=d                      size  $\leq t$

by construction of  $\mathcal{L}$

This follows from the fact that  $X(C(j)) \in \mathcal{L}$  and spread of  $\mathcal{L}$ .

combine:

$$\|X(\bar{B})\|_2^2 \leq \|X(\bar{s})\|_2^2 + \frac{1}{\epsilon^2} \sum_{j \in C(s)-Q} \|X(C(j)-S)\|_2^2$$

Next  $\sum_{j \in C(s)-Q} \|X(C(j)-S)\|_2^2 \leq \sum_{j=1}^n \|X(C(j)-S)\|_2^2 \leq D \|X(\bar{s})\|_2^2$

combine:   
 $\|X(\bar{B})\|_2^2 \leq \|X(\bar{s})\|_2^2 + \frac{1}{\epsilon^2} D \|X(\bar{s})\|_2^2$    
 only elements of  $\bar{s}$  appear here can appear at most D times.

$$\Rightarrow \|X(\bar{s})\|_2 \geq \frac{1}{\sqrt{4 + D/\epsilon^2}} \|X(\bar{B})\|_2$$

$$\geq \frac{1}{\sqrt{D/\epsilon^2 + D/\epsilon^2}} = \frac{\epsilon}{\sqrt{2D}}$$

Proposed schedule:

Mar 3	Zhou	Apr 2	SZKORA
Mar 5	ONYANG		
10	AMES	No lectures in March 17, 19	
12	Cheung	No lectures reading week	
24	KARZMI		
26	ALAMPAR-YA ZDZ		
31	Zhang		

Algorithm for CS signal recovery

Problem given by  $\min \|x\|_1$   
st.  $Ax = b$

- ① General purpose LP solver
- ② First order methods for convex problems (Nesterov (2003))
- ③ Orthogonal matching  
Applicable only to CS

Tropp-Gilbert signal recovery from random measurements via orthogonal matching pursuit 2007

Given  $A, b$  as above, say  $A \in \mathbb{R}^{m \times n}$   
s.t.  $x$  known to be sparse,  $\|x\|_0 = k$   
 $k < m < n$   
 $\text{supp}(x)$

Orthogonal Matching Algorithm / subset of  $\{1, \dots, n\}$

- ① initialize  $\Lambda = \emptyset, r = b$
- ② For  $j = 1, \dots, k$ 
  - ⓐ Find  $u \in \{1, \dots, n\}$  st.  $[A(:, u)]^T r$  is max
  - ⓑ  $\Lambda = \Lambda \cup \{u\}$
  - ⓒ solve linear least squares problem  
 $\min \|A(:, \Lambda)z - r\|_2$   
 $z \in \mathbb{R}^{|\Lambda|}$   
Let  $z^*$  be optimum
  - ⓓ  $r = r - A(:, \Lambda)z^*$

③ Output  $x$  where  $x(\Lambda)$  chosen to minimize  $\|A(:, \Lambda)x(\Lambda) - b\|_2$   
 $x(\bar{\Lambda}) = 0$

Note Minimizer  $x(1)$  can be recovered from sequences  $r^s, p^s$

$$\begin{aligned} \text{SUPP } k=2 \\ r^{(0)} &= b \\ r^{(1)} &= b - A(:, u^{(1)}) z^{*(1)} \\ r^{(2)} &= r^{(1)} - A(:, \{u^{(1)}, u^{(2)}\}) z^{*(2)} \\ &= b - A(:, u^{(1)}) z^{*(1)} - A(:, \{u^{(1)}, u^{(2)}\}) z^{*(2)} \\ &= b - A(:, \{u^{(1)}, u^{(2)}\}) \begin{bmatrix} z^{*(1)} \\ 0 \end{bmatrix} + z^{*(2)} \end{aligned}$$

Important fact about solving Linear Linear squares  
 SUPP MUST solve sequence of LLS problems in which  $j^{\text{th}}$  coeff. matrix is equal to  $(j+1)^{\text{st}}$  with one more column.  
 Then the cost of  $j^{\text{th}}$  is  $O(m-j+1^2)$  operations  
 (As opposed to  $O(mj^2)$ )

Compute QR factorization of coeff. matrix on  $(j+1)^{\text{st}}$  step by updating QR factorization from  $j^{\text{th}}$  step.  
 With this trick, LLS solving is low order work compared to step @ which is  $mn$  per iteration, so  $O(mnk)$  ops total.

- Assumptions on A
- ① Cols of A chosen independently. (In prob. sense)
  - ② Expected 2-norm of each col. is 1.
  - ③ Let V be an arbitrary  $m \times k$  matrix st.

$$\|V(:, j)\|_2 \leq 1 \quad \forall j=1, \dots, k. \quad \text{Then } \forall \epsilon=1, \dots, k$$

$$\text{Prob}(\|V^T A(:, \Lambda)\|_\infty \leq \epsilon) \geq 1 - 2k \exp(-\epsilon^2 m)$$

Q4 try Chebyshev ineq

- ④ For any  $\Lambda \subset \{1, \dots, n\}$  st.  $|\Lambda| = k$ ,  
 $\text{Prob}(\sigma_k(A(:, \Lambda)) \geq \frac{1}{2}) \geq 1 - e^{-cm}$

Reminiscent of UUP

Note spherical Section (KGG) Assumption is not sufficient <sup>on A</sup> for OMP

Thus: ① Gaussian Matrix scaled by  $1/\sqrt{m}$  has properties ①-④

② Bernoulli matrix "you select each entry to be  $\pm 1/\sqrt{m}$  with prob  $1/2, 1/2$  has properties ①-④

Main Thm. 6

Fix  $\delta \in (0, 0.36)$  supp.

$m \geq \text{const. } k \cdot \log(n/\delta)$  (This constant depends on constants in ①-④) supp.  $x$  is a sparse signal.  $\|x\|_0 \leq k$  for random choice of  $A$  from a distribution satisfying above assumptions, OMP applied to DATA  $(A, b)$  (where  $b = Ax$ ) will correctly recover  $x$  with prob  $\geq 1 - \delta$ .

Note: This is "weak" compressive sensing.

PF: Let  $\Lambda^*$  denote supp( $x$ )

Define  $r \in \mathbb{R}^m$

$$P(r) = \|A(:, \Lambda^*)^T r\|_2$$

OMP Guaranteed to select a column of  $\Lambda_j^*$  on step  $j$  provided

$$P(r^{(j)}) < 1$$

Imagine Applying OMP to  $A(:, \Lambda^*), b$

Let  $q^{(1)} \dots q^{(k)}$  be a sequence of residuals in imaginary case

Claim by induction.  $q^{(j)} = r^{(j)}$   $\forall j = 1, \dots, k$  with high probability

True provided

$$P(q^{(j)}) < 1 \quad \forall j = 1, \dots, k$$

Ensures that  $\Lambda^{(j)} \in \Lambda^*$

So must evaluate

$$P(P(q_j) < 1, \forall j = 1, \dots, k)$$

Define  $\hat{q}_j = \frac{q_j \cdot \frac{1}{2}}{\|A(:, \Lambda^*)^T q_j\|_2}$   $\forall j = 1, \dots, k$

OBS:  $\|q_j\|_2 = \frac{1}{2} \cdot \frac{\|q_j\|_2}{\|A(:, \Lambda^*)^T q_j\|_2} \leq \frac{1}{2} \cdot \frac{1}{\sigma_{\min}(A(:, \Lambda^*))} \stackrel{\text{assume}}{\leq} 2$

Assum  $\Rightarrow 2$

If it fails, add failure prob to overall prob of failure of OMP

$$\begin{aligned}
 P(q_j) &= \frac{\|A(:, \bar{\lambda}^*)^T q_j\|_\infty}{\|A(:, \bar{\lambda}^*)^T q_j\|_\infty} \\
 &\leq \frac{\sqrt{k} \|A(:, \bar{\lambda}^*)^T q_j\|_\infty}{\|A(:, \bar{\lambda}^*)^T q_j\|_2} \quad \text{by (*)} \\
 &= \sqrt{k} \|A(:, \bar{\lambda}^*)^T \hat{q}_j\|_\infty \frac{\|A(:, \bar{\lambda}^*)^T q_j\|_2}{\|A(:, \bar{\lambda}^*)^T \hat{q}_j\|_2} \frac{1}{\sqrt{2}} \\
 &= 2\sqrt{k} \|A(:, \bar{\lambda}^*)^T \hat{q}_j\|_\infty
 \end{aligned}$$

$$\begin{aligned}
 \text{Prob} ( P(q_j) < 1 \quad \forall j=1, \dots, k ) &\geq \text{Prob} ( 2\sqrt{k} \|A(:, \bar{\lambda}^*)^T \hat{q}_j\|_\infty < 1 \\
 &\quad \forall j=1, \dots, k ) \\
 &= \text{prob} ( 2\sqrt{k} \|A(:, \mu)^T \hat{q}_j\|_\infty < 1 \quad \forall j=1, \dots, k ) \\
 &\quad \forall \mu \in \bar{\lambda}^*
 \end{aligned}$$

Introduce  $V = [\hat{q}_1, \dots, \hat{q}_k]$   $m \times k$  matrix each col has

$$\begin{aligned}
 &= \text{Prob} ( 2\sqrt{k} \|A(:, \mu)^T V\|_\infty < 1 \quad \forall \mu \in \bar{\lambda}^* ) \quad \text{2-norm} \leq 1 \\
 &= \prod_{\mu \in \bar{\lambda}^*} \text{prob} ( 2\sqrt{k} \|A(:, \mu)^T V\|_\infty \leq 1 ) \quad \text{by independence of cols of } A
 \end{aligned}$$

Apply property ③ Note cols of  $V$  depend only on  $A(:, \bar{\lambda}^*)$  and hence may be considered Fixed, arbitrary, deterministic w.r.t

$$\begin{aligned}
 \epsilon &= 1/2\sqrt{k} \quad \text{cols in } \bar{\lambda}^* \\
 &\geq \prod_{\mu \in \bar{\lambda}^*} ( 1 - 2k \exp(-\frac{C_m}{4k}) ) = ( 1 - 2k \exp(-\frac{C_m}{4k}) )^{n-k}
 \end{aligned}$$

$$\begin{aligned}
 \text{Fact: for } p \geq 1 \quad (1-x)^p &\geq 1 - px \\
 x &\leq 1 \\
 &\quad (n-k) k \leq n^2/4 \\
 &\geq 1 - 2(n-k) k \exp(-\frac{C_m}{4k}) \geq 1 - n^2/2 \exp(-\frac{C_m}{4k})
 \end{aligned}$$

choose  $k \leq \text{const} \cdot m \cdot \log(n/\delta)$

ensures failure prob  $\delta$

Feb. 24, 2009.

Recall Basic Algorithmic problem of C-S is

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

Orthogonal matching pursuit works with high prob. (over choice of  $A$  and  $x$ )

Running time is  $O(mk)$  where  $A \in \mathbb{R}^{m \times n}$  and  $k$  has  $\|x\|_0$

Other algorithms: simplex, interior point

First-order proximate

reformulate as S.E.F Linear programming

$$\min e^T x^+ + e^T x^- \quad e = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\text{s.t. } Ax^+ - Ax^- = b$$

$$x^+ \geq 0, x^- \geq 0$$

$$x = x^+ - x^-$$

$$\text{Dual: } \max b^T y \text{ s.t. } A^T y \leq e$$

$$\text{i.e. } \max b^T y \text{ s.t. } \|A^T y\|_\infty \leq 1$$

weak duality Applied here

if  $x, y$  feasible for primal, dual respectively,

$$b^T y = x^T A^T y \leq \|x\|_1 \|A^T y\|_\infty \leq \|x\|_1$$

Complementary slackness

At optimality,

$$x_i^+ (1 - a_i^T y) = 0 \quad \text{for } i=1, \dots, n$$

$$x_i^- (1 + a_i^T y) = 0$$

The simplex algorithm maintains an  $m \times n$  submatrix of  $[A, -A]$  called the basis

On each iteration, must solve  $m \times m$  linear system because basis maintained in inverse or factored form requires  $O(m^2)$  ops. Scanning for new non basic variable to enter basis require  $O(mn)$  ops.

worst case # of operations is exponential.

$\binom{n}{m}$  (obvious upper bound) But in practice, # of iterations is  $\sim 3n$  if use Heuristic bound rather than worst-case bound, # op is  $O(mn^2)$  since  $n \gg k$ , This is worse than OMP.

simplex, ZPT, 1st order proximal may be preferable to omp for two reasons:

- ① No probabilistic assumptions.
- ② work even if not compressive sensing ( $X$  not sparse)

Issue with simplex for C-S, optimal sol is degenerate.

In nondegenerate SEF, LP,  $\|x^*\|_0 = m$  But for C-S.

$\|x^*\|_0 = k \ll m$ , Degeneracy may create problems.

Interior point Method.

on each iteration, maintain primal and dual feasible iterates  $x, y$  st.

$$\|XSe - Me\| \leq \text{const } M.$$

where  $X$  is a variable in the algorithm that

driven to zero.  $e = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$   $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$   $s = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}$

where  $s_i$  - slack in the dual constraint

For this problem, there are  $2n$  vars.

So  $x = \begin{pmatrix} x_1^+ \dots x_n^+ \\ x_1^- \dots x_n^- \end{pmatrix}$   $s = \begin{pmatrix} s_1^+ \dots s_n^+ \\ s_1^- \dots s_n^- \end{pmatrix}$

$$s = \begin{pmatrix} 1 - a_1^T y \\ \vdots \\ 1 - a_n^T y \end{pmatrix}$$

On each iteration, approximately solve non-linear eqs.

$$A \bar{x} = b$$

and  $\bar{x}_i \cdot \bar{s}_i = \bar{m} \quad i=1, \dots, n.$

$m$  th nonlinear eqs. For  $m$  th variables  
( $x, y$ ),  $\bar{m}$  is less than current  $m$

Apply one step of Newton's method to solve linear eqs.

Solution of resulting linear eqs requires  $O(m^2 n)$  OPS

Number of iterations required Theoretically is

$$O(\sqrt{n} L) \text{ where } L = \# \text{ bits required to write } A, b.$$

In practice, # iterations is 30.

with this heuristics, run time is  $O(m^2 n)$

$V + Y \in$  has an interior pt method whose running time depends on  $(XA)$  but indep. of  $b$ .

Perhaps preferable for CS because  $A$  is fixed in advance, but  $b$  is data dependent.

$$\bar{X}_A = \max \{ \|A_B^{-1} A\| \mid A_B \text{ is invertible max submat of } A \}$$

perhaps can be proved that  $\bar{X}_A$

is modest under CS assumptions.

first-order proximal-type methods

(Ivesterov)

$$\min f(x) + P(x)$$

where  $P(x)$  is convex, "simple" non smooth

$f(x)$  is convex, smooth

Example: Reformulate CS as

$$\min \|Ax - b\|_2$$

$$\text{st. } \|x\|_1 \leq \delta$$

equiv. to  $\min \|x\|_1$ , st. ~~HA~~  $Ax = b$  provided  $\delta = \|x^*\|_1$

if  $\delta$  not known in advance, could find it via bisection  
To use above form, Take

$$f(x) = \frac{1}{2} \|Ax - b\|_2^2$$

$$P(x) = \begin{cases} 0 & \text{if } \|x\|_1 \leq \delta \\ \infty & \text{if } \|x\|_1 > \delta \end{cases}$$

Proximal method requires two ops

- ① Compute  $Ax - b$ , ie.  $\nabla f(x)$   $mn$  ops.
- ② Solve subproblems of the form

$$\min c^T x$$

$$\text{st. } \|x\|_1 \leq \delta$$

both can be done in  $O(n)$  ops.

and

$$\min \|x - x_0\|_2$$

$$\text{st. } \|x\|_1 \leq \delta$$

Running time is  $O(L/\epsilon)$  iterations  
 where  $L = \text{Lipschitz const of } \nabla f = \sup_{x \neq y} \frac{\|\nabla f(x) - \nabla f(y)\|_2}{\|x - y\|_2}$

In this context.

$$= \sup_{x \neq y} \frac{\|A(x - y)\|_2}{\|x - y\|_2} = \sigma_{\max}(A)$$

in  $c-s$  context, this is promising.

$\epsilon =$  Desired accuracy of solutions.

Tues Lec. in mc 5136

Complexity = Classifying computational problems as hard or easy  
 A problem means a description of function to be computed.

The function maps inputs called instances to outputs

E.g. Linear programming is a problem

Instance:  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c \in \mathbb{R}^n$

output  $x^*$  solving  $\min c^T x$  st.  $Ax = b$ ,  $x \geq 0$

An instance of LP

$\min \begin{bmatrix} 3 \\ 2 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  st.  $x_1 + x_2 = 4$ ,  $x_1 \geq 0$ ,  $x_2 \geq 0$

A decision problem: The output is either "Yes" or "No"

posing LP as a decision problem

Instance  $A, b, c$  as above,  $\lambda \in \mathbb{R}$

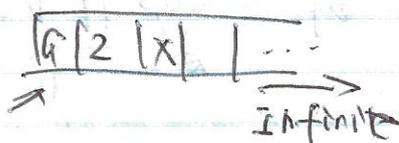
output = "Yes" if  $\exists x^* Ax^* = b$ ,  $x^* \geq 0$ ,  $(c^T x^* \leq \lambda)$

"No" Else

Algorithm to solve a problem Compute the function is a "Turing machine"

A very simple model here of a computer

Tape divided into cells



each cell contains a symbol

But any specific

computation uses only a finite # of

Initially, the input instance follows by blanks

Tape head reads cell and writes according

to a program of finite length statement like

step 1: if current cell is 'a', change to 'b'

move the head left, jump to step 13

TM: can carry out all the instructions of an ordinary computer.

Real numbers NOT representable on a TM.

T.M. Typically represents reals with rationals.

Polynomial time: A problem is solvable in polynomial time if there is a TM to solve it whose running time is bounded above by a polynomial function of length of the input.

Example 1 PSI. Given  $a, b \in \mathbb{Z}^+$ , (~~find their GCD~~)

Are they relatively prime?

EUCLID'S GCD Alg. Can show that either  $a$  or  $b$  decreases by at least  $1/2$  per iteration.

So # iterations  $\leq \underbrace{(\lceil \log_2 a \rceil + \lceil \log_2 b \rceil)}_{\text{length of input}}$

each iteration requires integer algorithm poly  $(\lceil \log_2 a \rceil + \lceil \log_2 b \rceil)$

3 Linear programming KHACHYAN 1979 showed

ellipsoid algorithm requires  $O(n^2 L)$  iterations to find  $x^*$ .

$n$  = Dimension of  $x^*$ ,  $L$  = # bits in  $(A, b, c)$

Decision problem  $\Pi$  lies in  $\boxed{NP}$ .

If there exists a Turing Machine  $M$ , (called "certificate checker") as follows:

$M$  takes as input order pairs of strings  $(x, y)$

where  $x$  is an instance of  $\Pi$  and  $y$  is a candidate certificate

1  $M$  is poly time in length of  $(x, y)$

2 If  $x$  is a Yes-instance of  $\Pi$ , then there exists

a  $\gamma$  st. length  $(\gamma \leq \text{poly}(\text{length}(x)))$  and  $M$  outputs "Yes" for  $(x, \gamma)$

③ If  $x$  is a No-instance of  $\Pi$ , then  $M$  outputs "No" for all pairs  $(x, \gamma)$

① Linear programming.

If  $x$  = instance  $(A, b, c, \lambda)$  as above, take  $y = x^*$

I.e. the vector st.  $Ax^* = b$   $c^T x^* \leq \lambda$

If a Yes-instance.  $x^* \geq 0$

If  $M$  given  $(x, y)$  then it merely substitute  $x^*$  into the constraints, outputs "Yes" if all satisfied, else it outputs "No"

Attention: Must show  $\text{length}(x^*) \leq \text{poly}(\text{length}(A, b, c, \lambda))$

subtlety

② Satisfiability: Given a boolean formula

$[(x_1 \wedge x_2) \vee (\bar{x}_1 \wedge \bar{x}_2)] \wedge \bar{x}_1$  is there a setting of vars to either true or false to make the overall formula true?

This is in NP. A Yes-instance is certified by providing the setting of vars to make the formula true.

Above is satisfied by "FF"

A polynomial time reduction from decision problem  $\Pi_1$  to  $\Pi_2$  is a mapping  $\phi$  from instance of  $\Pi_1$  to instance of  $\Pi_2$  st.

①  $\phi$  computable in poly time.

②  $x$  A Yes-instance of  $\Pi_1 \Leftrightarrow \phi(x)$  is a Yes-instance of  $\Pi_2$

Example: 3-SAT special case of satisfiability.

Formula must be a conjunction of clauses.

$C_1 \wedge C_2 \wedge C_3 \wedge \dots \wedge C_m$

each clause is a disjunction of 3 variables or complemented vars.



because of omitted edges, clique can contain at most  $m$  literal and  $m$  clause vertices.

OBS: If we select literal vertices corresponding to a satisfying assignment, and if we select clause vertex for each clause corresponding its true literals, then all edges will present and hence a clique is found. Conversely, if  $(m+1)$ -vertices form a clique, we can deduce the satisfying assignment.

Say that A (Decision) problem  $\Pi$  is  $\boxed{NP\text{-hard Complete}}$

If ①  $\Pi \in NP$ .

② For every  $\Pi' \in NP$ , there is a poly reduction from  $\Pi'$  to  $\Pi$  (the hardest problem in  $NP$ )

Thm (Cook 1971) 3SAT is  $NP$ -complete.

PF: Already show 3-SAT  $\in NP$ .

suppose  $\Pi' \in NP$ . it has a certificate checker  $M$ .

write down a huge but poly-sized boolean formula  $\varphi$  in which vars.

entries of  $\varphi$  (ex.  $x_i$  stands for

"1st symbol of  $y$  is '0'")

vars for contents of each tape cell at each time step of a certificate checker.

write down any clauses to express consistency and correct operations of certificate checker. Clause that states that on the last step,  $M$  outputs "yes"!

Cor. clique is  $NP$ -complete.

PF: clique  $\in NP$  (certificate is the list of vertices of the clique)

if  $\Pi'$  is in NP, has reduce of  $\Pi$  to 3SAT and 'reduction of' 3SAT to clique, Hence reduction of  $\Pi'$  to clique. (composition)

thus to show that  $\Pi$  is an NP-complete problem, Must first show it is in NP, then find a poly-time reduction of known NP-complete problem to  $\Pi$ .

Say that  $\Pi$  is NP-hard if an NP-complete problem can be reduced to  $\Pi$ .  
polynomially

Three dimensional matching (3DM) — NP-hard

Given three sets of vertices,  $X, Y, Z$ .

given a list of ordered triples,  $L = \{x_1y_1z_1, \dots, x_ny_nz_n\}$

is there a subset  $L'$  of  $L$  such that each entry of  $X \cup Y \cup Z$  occurs exactly once in  $L'$ ?

Thm: (KARP 1972) 3DM is NP-complete.

Thm: Sparsest vector problem Given  $A \in \mathbb{Z}^{m \times n}$ ,  $b \in \mathbb{Z}^m$  and  $k$ , is there an  $x$  st.  $Ax = b$   $\|x\|_0 \leq k$ ?

Sparsest vector : NP-complete. (reductions from 3DM)