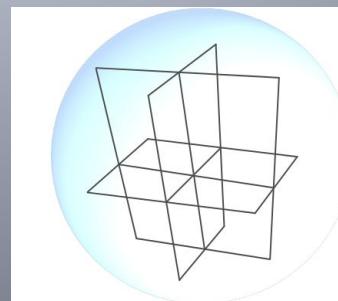


Compressive Sensing

CS 754
Ajit Rajwade



ViGIL
VISION, GRAPHICS
& IMAGING LABORATORY

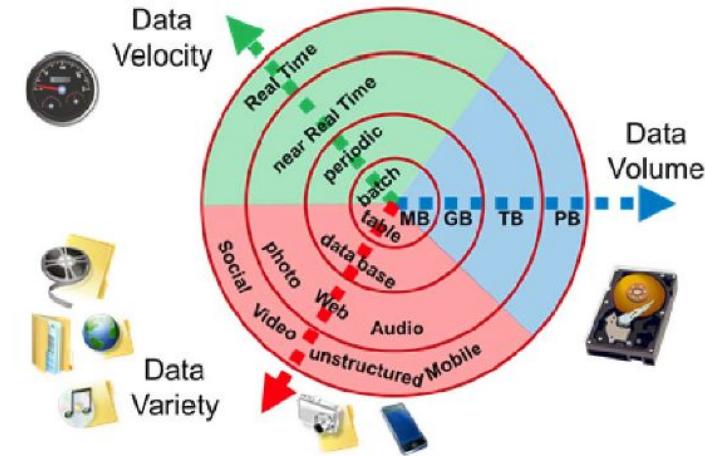
Outline of the Lectures

- Introduction and Motivation
- CS Theory: Key Theorems
- Practical Compressive Sensing Systems
- Reconstruction algorithms for Compressive Sensing Systems
- Proof of some of the key results

Introduction + MOTIVATION

The Era of Big Data

- Huge amounts of data being constantly acquired
- In many applications
 - ✓ E-commerce, stock markets
 - ✓ Weather forecasting
 - ✓ Sports (esp. cricket!)
 - ✓ News reports
 - ✓ Image data (**medical**, surveillance, satellite, etc.)



https://en.wikipedia.org/wiki/Big_data

CC image courtesy of Enderoo5 on
WikiCommons licensed under
CC-BY-SA-4.0

The Era of Big Data

- Do we **really need** to measure all these data?
- Can we **save resources** by **intelligent acquisition**?
- Resources:
 - Time (**MRI machine**)
 - Battery Power (**mobile phones**)
 - Radiation administered to patients (**CT machine**)
 - Chemical reagents (**Covid-19 RT-PCR testing**)
- Use **inherent properties** of (redundancy in) data

Conventional Sensing and Compression

- Before applying signal/image compression algorithms like MPEG, JPEG, JPEG-2000 etc., the measuring devices typically measure large amounts of data.
- The data is then converted to a transform domain where a majority of the transform coefficients turn out to have near-zero magnitude and can be discarded with small reconstruction error but great savings in storage.
- Example: a digital camera has a detector array where several thousand pixel values are first stored. This 2D array is fed to the JPEG algorithm which computes DCT coefficients of each block in the image, discarding smaller-valued DCT coefficients.

Converting to transform domain: DFT

- All discrete signals can be represented perfectly as linear combinations of **complex sinusoidal functions** of different frequencies.
- This is called as the Discrete Fourier transform (See next slide).
- What purpose does it serve? It enables determining the frequency content of the signal (eg: which frequencies were present in an audio signal and at what strength?)

Fourier Transform

$$F(u) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n) e^{-i 2\pi u n / N}$$

u -th Fourier coefficient; u = frequency index

$$f(n) = \frac{1}{\sqrt{N}} \sum_{u=0}^{N-1} F(u) e^{i 2\pi u n / N}$$

Value of signal f at location n (f is a vector of size N)

$i = \sqrt{-1}$

$$\mathbf{f} = \mathbf{H}\mathbf{F}$$

Vector of N Fourier coefficients

$$= \sum_{u=0}^{N-1} \mathbf{H}_u F(u)$$

$N \times N$ orthonormal matrix (Fourier basis matrix)

u -th column of \mathbf{H} (corresponding to Fourier coefficient $F(u)$)

Fourier Basis Matrix

$$\mathbf{F} = \mathbf{H}^T \mathbf{f} \quad \mathbf{H} \mathbf{H}^T = \mathbf{H}^T \mathbf{H} = \mathbf{I}$$

$$\begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ \vdots \\ f(N-1) \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{i2\pi(0)(0)/N} & e^{i2\pi(1)(0)/N} & e^{i2\pi(N-1)(0)/N} \\ e^{i2\pi(0)(1)/N} & e^{i2\pi(1)(1)/N} & e^{i2\pi(N-1)(1)/N} \\ & & \\ e^{i2\pi(0)(N-1)/N} & e^{i2\pi(1)(N-1)/N} & e^{i2\pi(N-1)(N-1)/N} \end{bmatrix} \begin{bmatrix} F(0) \\ F(1) \\ \vdots \\ \vdots \\ F(N-1) \end{bmatrix}$$

Time index n is constant over all entries in a given row.
Frequency index u is constant across all entries in a given column. Thus each column has a fixed frequency (of the complex exponential).

2D Fourier Transform

$$F(u, v) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-i2\pi(xu/M + yv/N)}$$

F and f are vectors of length MN

$$f(x, y) = \frac{1}{\sqrt{MN}} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{i2\pi(xu/M + yv/N)}$$

$$\mathbf{f} = \mathbf{H}\mathbf{F}$$

$$= \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) \mathbf{H}_{uv}$$

Vector of size $MN \times 1$. Note that though \mathbf{F} and \mathbf{f} are 2D signals, they are represented here in vectorized form.

Matrix of size $MN \times MN$, whose columns are indexed by a frequency pair u, v (note \mathbf{H}_{uv})

Discrete Cosine Transform (DCT) in 1D

$$F(u) = \sum_{n=0}^{N-1} f(n)(H^T)_{un} = \sum_{n=0}^{N-1} f(n)H_{nu}^* = (H^T f)_u$$

$$f(n) = \sum_{u=0}^{N-1} F(u)H_{nu} = (HF)_n$$

$$DCT : H_{nu} = \sqrt{\frac{1}{N}}, u = 0$$

For DFT, we would have $H_{nu} = e^{i2\pi \frac{un}{N}}$

$$H_{nu} = \sqrt{\frac{2}{N}} \cos\left(\frac{\pi(2n+1)u}{2N}\right), u = 1 \dots N-1$$

Note that by convention transpose here refers to the conjugate transpose (this is also a convention followed by MATLAB).

Discrete Cosine Transform (DCT) in 1D

$$\begin{pmatrix} f(0) \\ \vdots \\ f(N-1) \end{pmatrix} = \begin{pmatrix} H_{0,0} & \dots & \dots & H_{0,N-1} \\ \vdots & \ddots & \dots & \vdots \\ \vdots & \ddots & \dots & \vdots \\ \vdots & \ddots & \dots & \vdots \\ H_{N-1,0} & \dots & \dots & H_{N-1,N-1} \end{pmatrix} \begin{pmatrix} F(0) \\ \vdots \\ F(N-1) \end{pmatrix}$$

↓

$$\mathbf{H} \in R^{N \times N}$$

$$\mathbf{H}\mathbf{H}^T = \mathbf{I}$$

Across any column of \mathbf{H} , the frequency index u is constant and time (or spatial) index n varies. Across any row of \mathbf{H} , the frequency index varies and n is constant.

DCT

- Expresses a signal as a linear combination of **cosine** bases (as opposed to the complex exponentials as in the Fourier transform). The coefficients of this linear combination are called **DCT coefficients**.
- Is **real-valued** unlike the Fourier transform!
- Has better compaction properties for signals and images – as compared to DFT.

$$\begin{pmatrix} f(0) \\ \vdots \\ f(N-1) \end{pmatrix} = \begin{pmatrix} H_{0,0} & \dots & \dots & H_{0,N-1} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ H_{N-1,0} & \dots & \dots & H_{N-1,N-1} \end{pmatrix} \begin{pmatrix} F(0) \\ \vdots \\ F(N-1) \end{pmatrix}$$

$\mathbf{H} \in R^{N \times N}$ (DCT Basis Matrix)

$$\mathbf{H}\mathbf{H}^T = \mathbf{I}$$

- DCT basis matrix is orthonormal. The dot product of any row (or column) with itself is 1. The dot product of any two different rows (or two different columns) is 0. The inverse is equal to the transpose.
- Being orthonormal, it preserves the squared norm, i.e. $\|\mathbf{f}\|^2 = \|\mathbf{F}\|^2$
- DCT is NOT the real part of the Fourier!
- DCT basis matrix is **NOT** symmetric.
- Columns of the DCT matrix are called the **DCT basis vectors**.

DCT in 2D

$$F(u, v) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f(n, m) H_{nmuv}^*$$

$$f(n, m) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) H_{nmuv}$$

The DCT matrix in this case will have size $MN \times MN$, and it will be the Kronecker product of two DCT matrices – one of size $M \times M$, the other of size $N \times N$. The DCT matrix for the 2D case is also orthonormal, it is NOT symmetric and it is NOT the real part of the 2D DFT.

DCT :

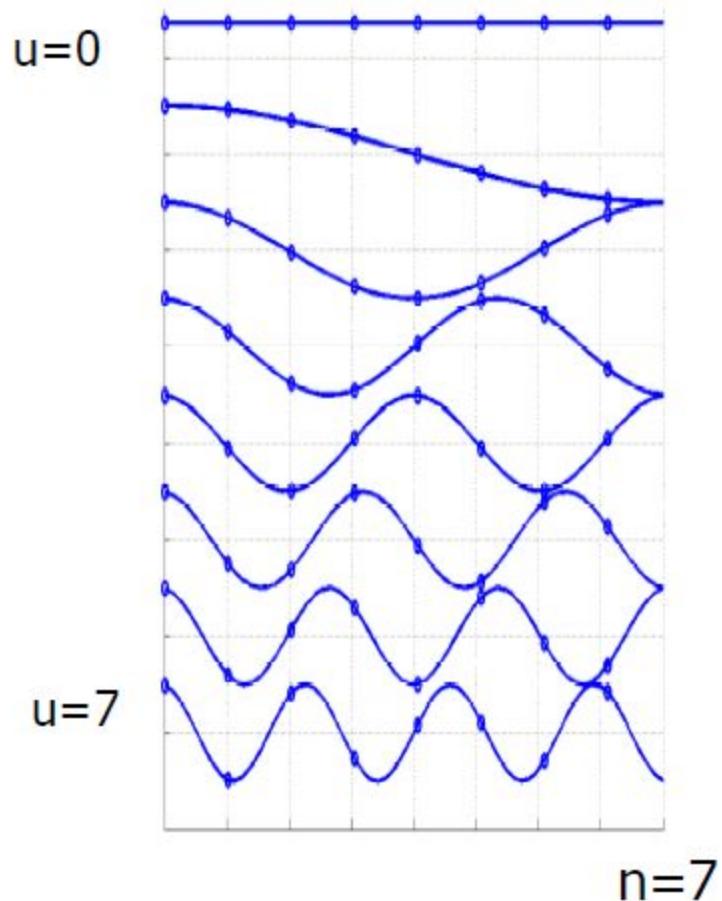
$$H_{nmuv} = \alpha(u)\alpha(v) \cos\left(\frac{\pi(2n+1)u}{2N}\right) \cos\left(\frac{\pi(2m+1)v}{2M}\right), u = 0 \dots N-1, v = 0 \dots M-1$$

$$\alpha(u) = \sqrt{1/N} \text{ (} u = 0 \text{), else } \alpha(u) = \sqrt{2/N}$$

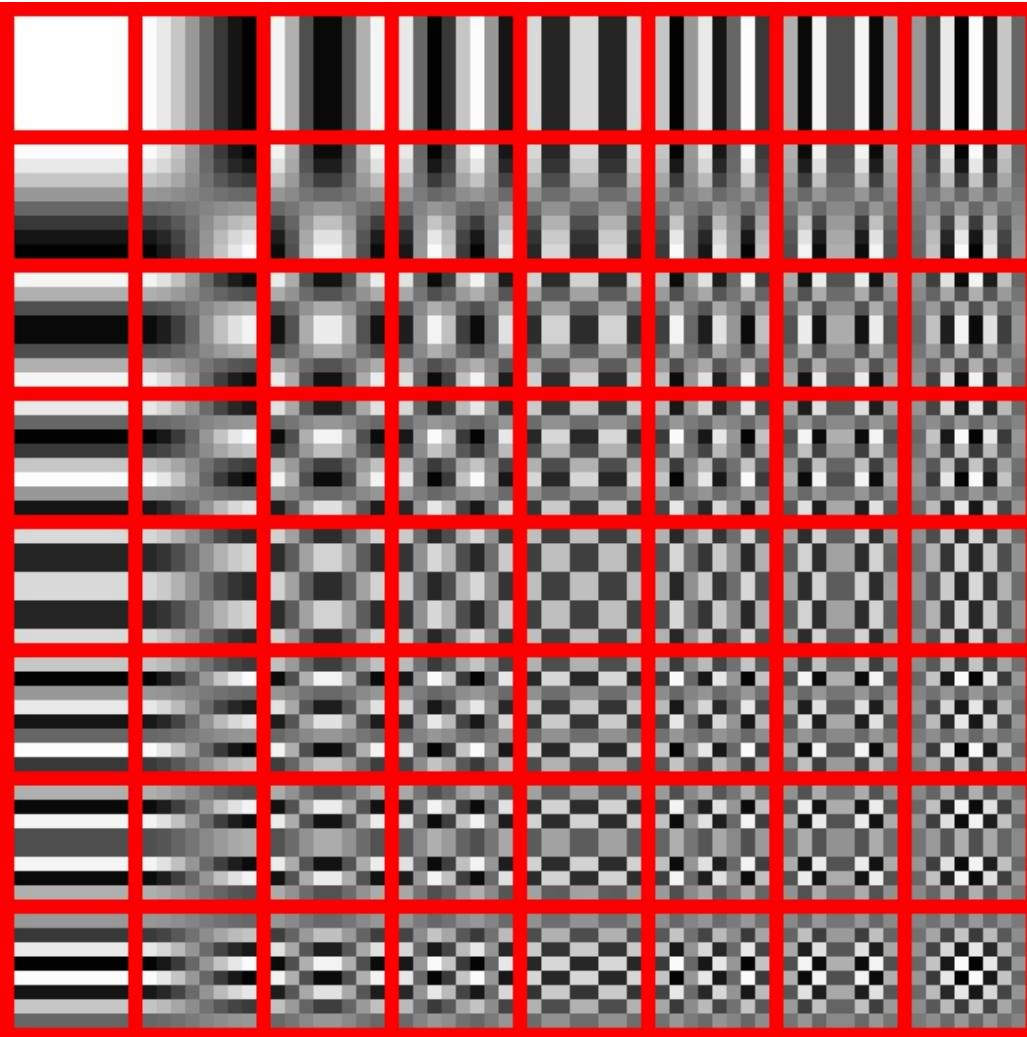
$$\alpha(v) = \sqrt{1/M} \text{ (} v = 0 \text{), else } \alpha(v) = \sqrt{2/M}$$

Do not interpret H_{nmuv} as a 4-D index, rather it is an element of the column of H indexed by frequency (u, v) . Which element? An element indexed by spatial coordinates (n, m) .

How do the DCT bases look like? (1D case)



How do the DCT bases look like? (2D-case)

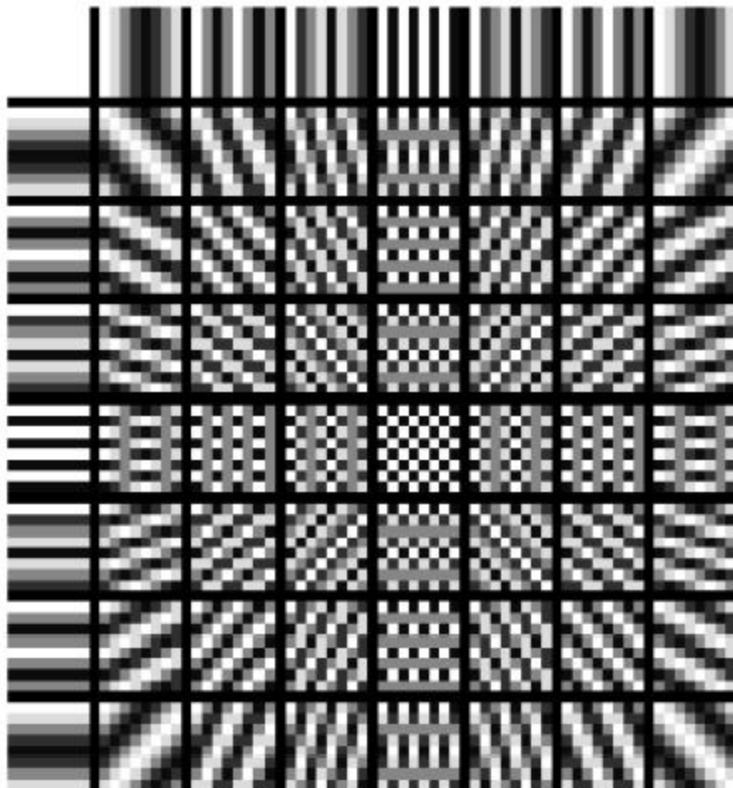


The DCT transforms an 8×8 block of input values to a linear combination of these 64 patterns. The patterns are referred to as the two-dimensional DCT *basis functions*, and the output values are referred to as *transform coefficients*.

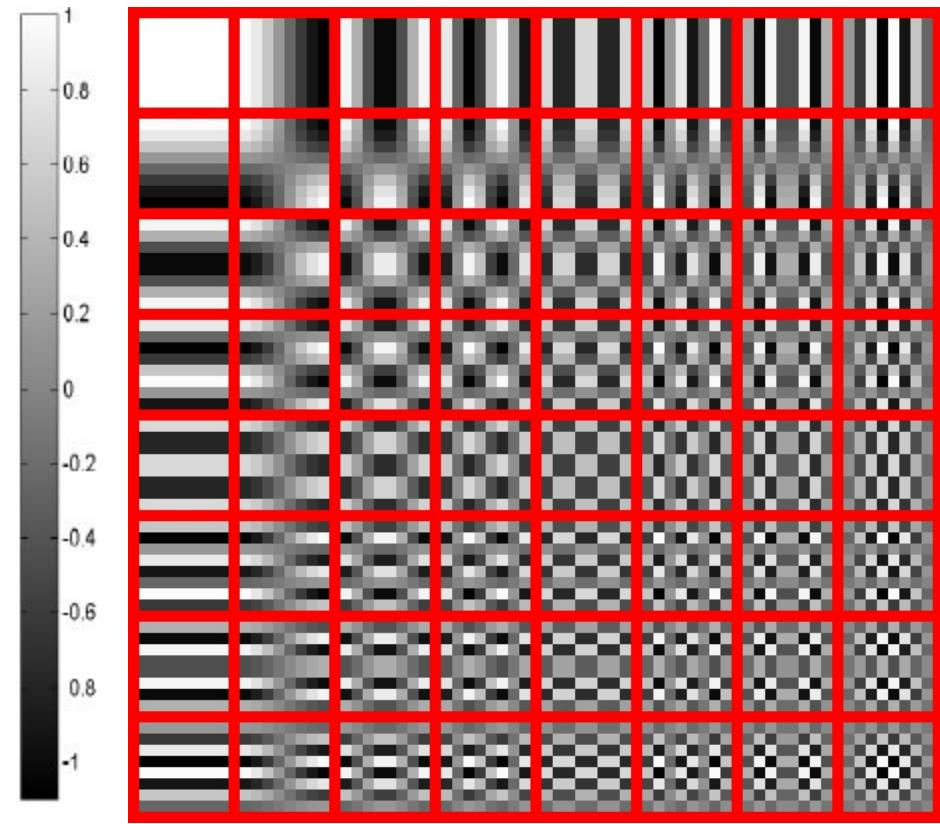
Each image here is obtained from the 8×8 outer product of a pair of DCT basis vectors. Each image is stretched between 0 and 255 – on a common scale.

<http://en.wikipedia.org/wiki/JPEG>

Again: DCT is NOT the real part of the DFT



Real part of DFT



DCT

<http://en.wikipedia.org/wiki/JPEG>

What is Compressive Sensing?

- Conventional sensing is a “measure and compress+throw” paradigm, which is wasteful!
- Especially for time-consuming acquisitions like MRI, CT, video photography, etc.
- Compressive sensing is a new technology where the data are acquired/measured in a compressed format!

What is Compressive Sensing?

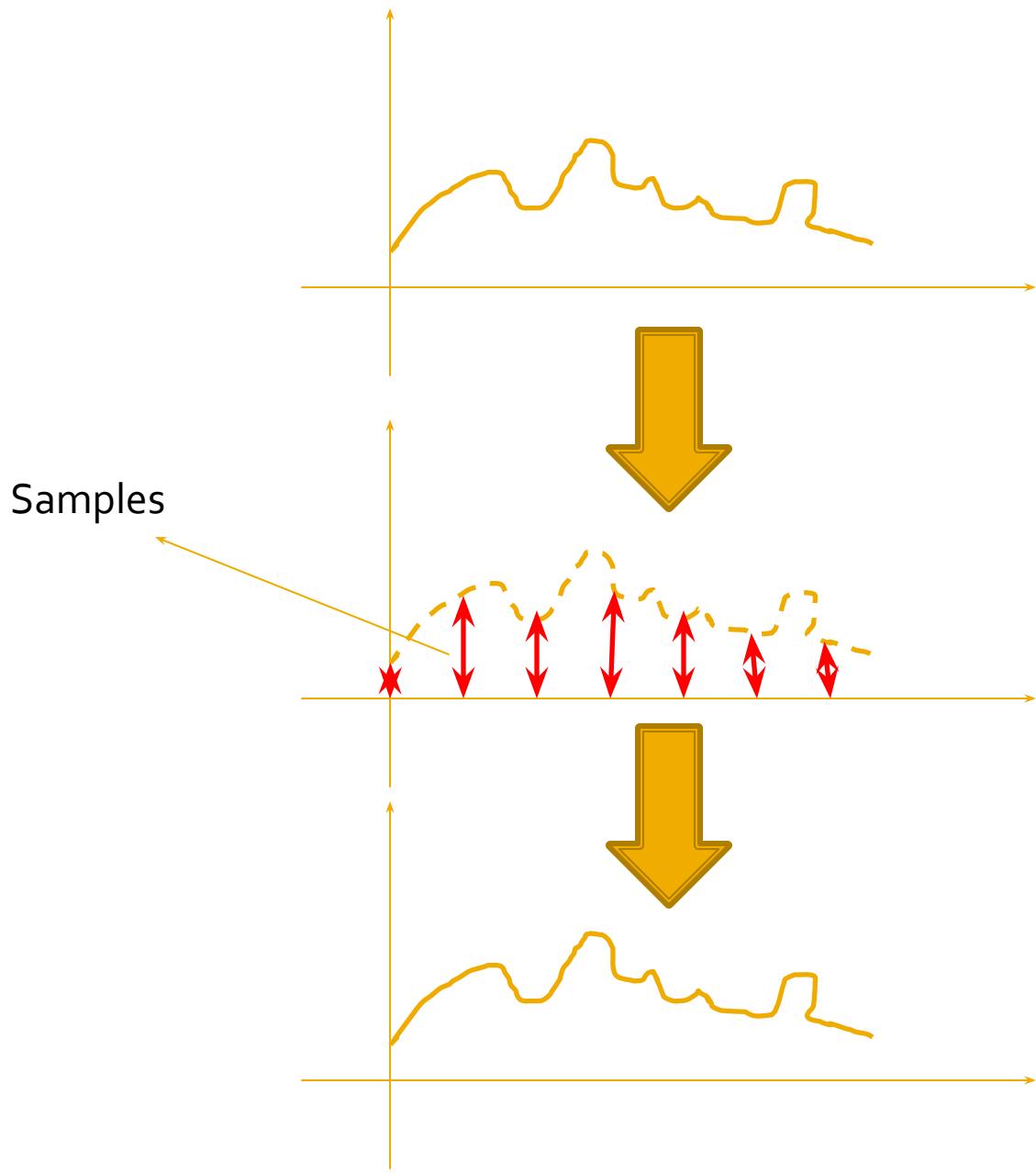
- These compressed measurements are then fed to some optimization algorithm (also called inversion algorithm) to produce the complete signal.
- This part is implemented in software.
- Under suitable conditions that the original signal and the measurement system must fulfill, the reconstruction is guaranteed to have very little or even zero error!

Why compressive sensing?

- It has the potential to dramatically improve acquisition speed for MRI, CT, hyper-spectral data and other modalities.
- Potential to dramatically improve video-camera frame rates without sacrificing spatial resolution.

Shannon's sampling theorem

- A band-limited signal with maximum frequency B can be **accurately** reconstructed from its uniformly spaced digital samples if the rate of sampling exceeds $2B$ (called **Nyquist rate**).
- Independently discovered by Shannon, Whitaker, Kotelnikov and Nyquist.



Actual Analog
Signal

Digital signal
(sampled from
analog signal)

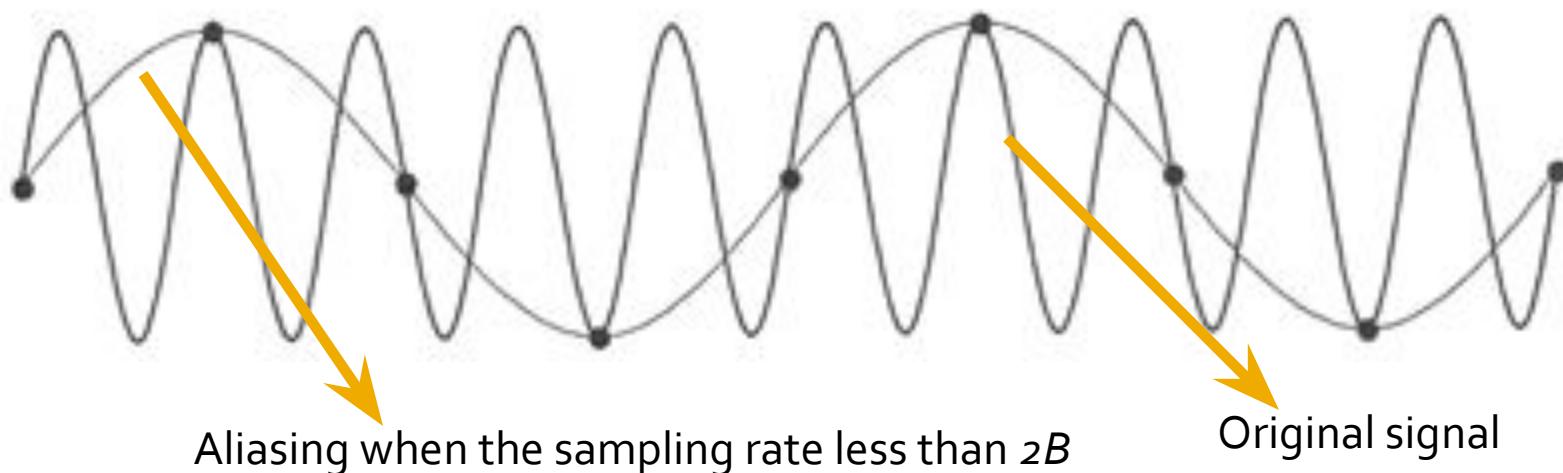
Reconstructed signal
(reconstruction is
accurate if Nyquist's
condition is satisfied)

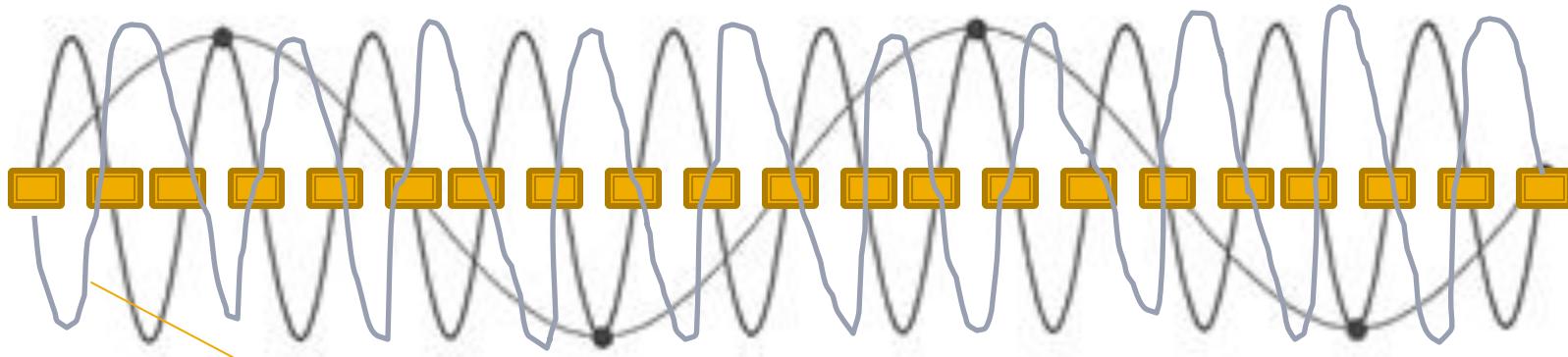
Band-limited signal

- A truly band-limited signal is an ideal concept and would take infinite time to transmit or infinite memory to store.
- Because a **band-limited signal can never be time-limited** (or space-limited).
- But many naturally occurring signals when measured by a sensing device are approximately band-limited (example: camera blurring function reduces high frequencies).

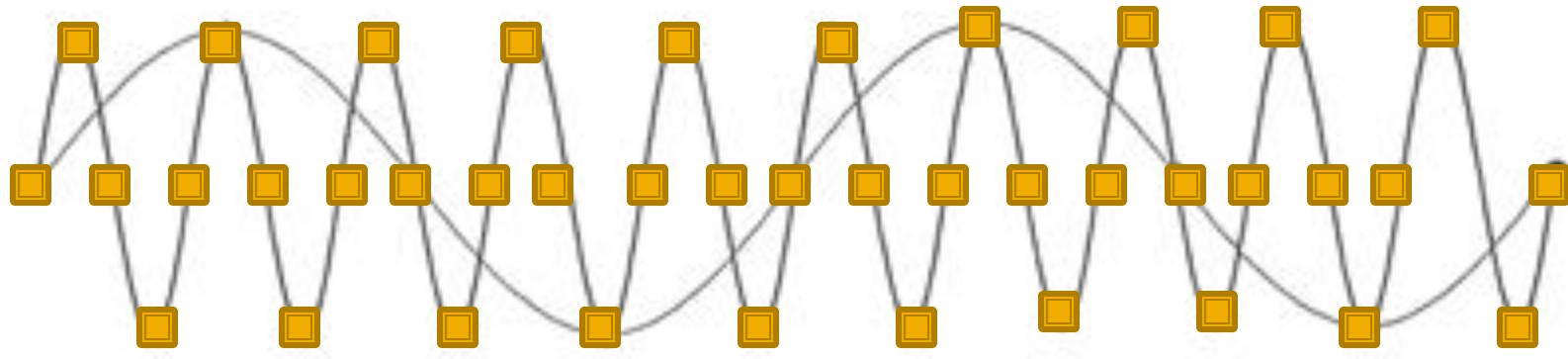
Aliasing

- Sampling an analog signal with maximum frequency B at a rate **less than or equal to $2B$** causes an artifact called **aliasing**.





Aliasing when the sampling rate is equal to $2B$



No aliasing when the sampling rate is more than $2B$

Whittaker-Shannon Interpolation Formula

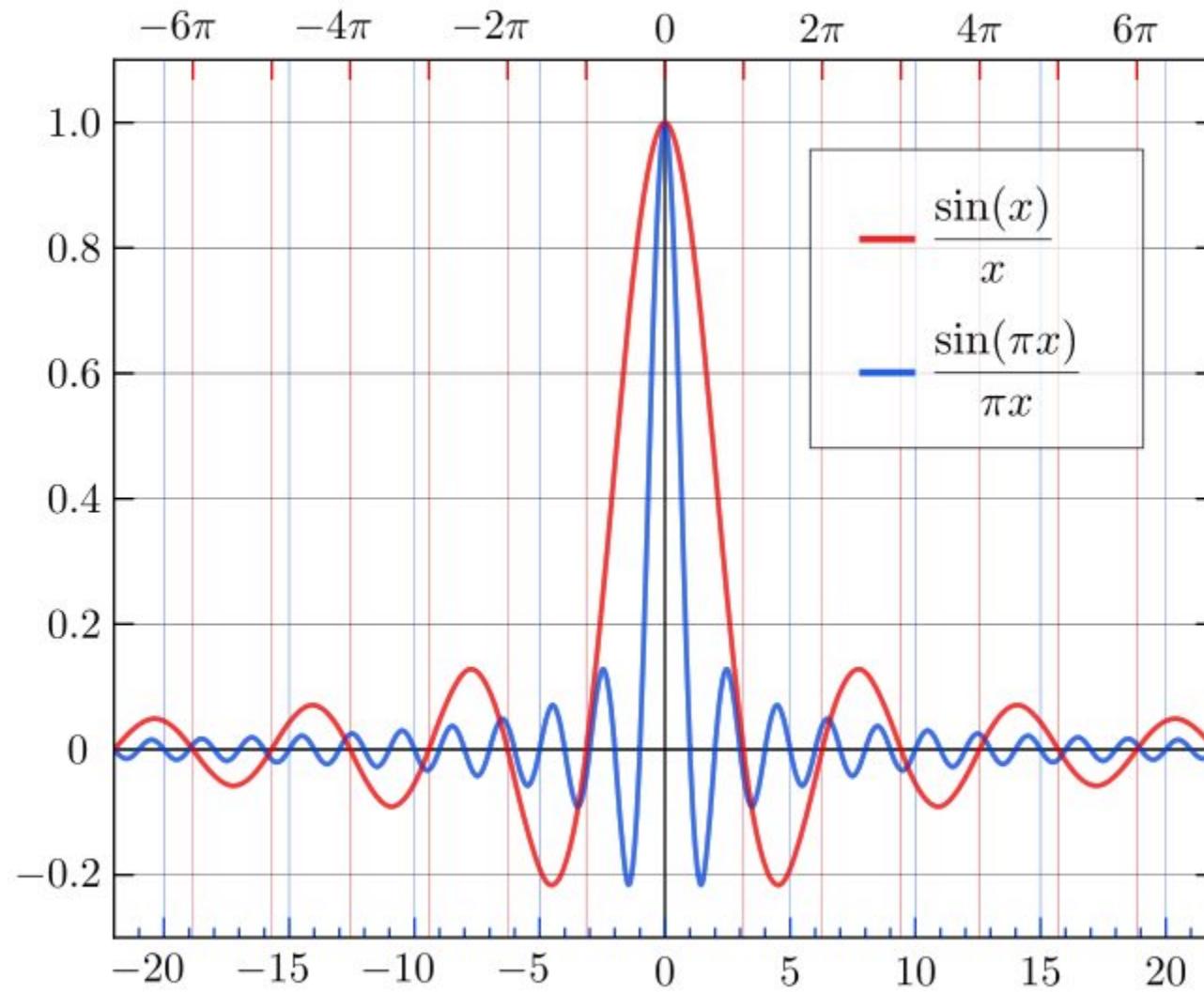
- The optimal reconstruction for band-limited signals from their digital samples proceeds using the **sinc interpolant**. It yields a time-domain (or space-domain signal of infinite extent with values that are very small outside a certain window).
- The associated formula is called Whittaker-Shannon interpolation formula.

$$f(t) = \sum_{n=-\infty}^{+\infty} f(n) \frac{\sin\left(\frac{t-nT}{T}\right)}{\left(\frac{t-nT}{T}\right)} = \sum_{n=-\infty}^{+\infty} f(n) \text{sinc}\left(\frac{t-nT}{T}\right) = \left(\sum_{n=-\infty}^{+\infty} f(n) \delta(t-nT) \right) * \text{sinc}\left(\frac{t}{T}\right)$$

Sampling period = 1/Sampling rate

Magnitude of the n-th sample

The sinc function (it is the Fourier transform of the rect function)



Some limitations of Shannon's theorem

- The samples need to be **uniformly spaced** (there are extensions for non-uniformly spaced samples, with the equivalent of Nyquist rate being an average sampling rate).
- The sampling rate needs to be **very high** if the original signal contains higher frequencies (to avoid aliasing).
- Does not account for several nice properties of naturally occurring signals (talks only about band-limitedness which is not perfectly realizable).

Motivation for CS: Candes' puzzling experiment (Circa 2004)

Ref: Candes, Romberg and Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", IEEE Transactions on Information Theory, Feb 2006.

$$\min_f \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \sqrt{f_x^2(x, y) + f_y^2(x, y)}$$

such that

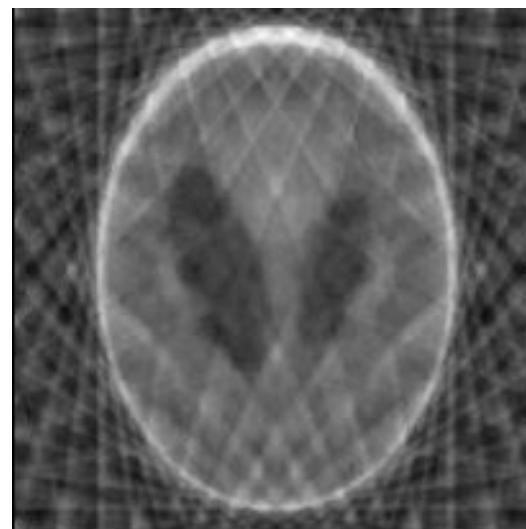
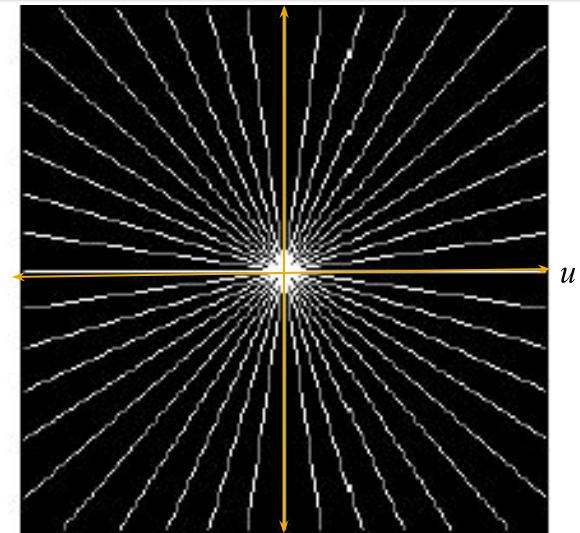
$$\forall (u, v) \in \mathbb{C}, [F(f)](u, v) = G(u, v)$$

$F \rightarrow$ Fourier operator

$f_x = x$ -derivative of f ;

$f_y = y$ -derivative of f

$\{G(u, v)\}$ = retained frequencies



Motivation for CS: Candes' puzzling experiment (Circa 2004)

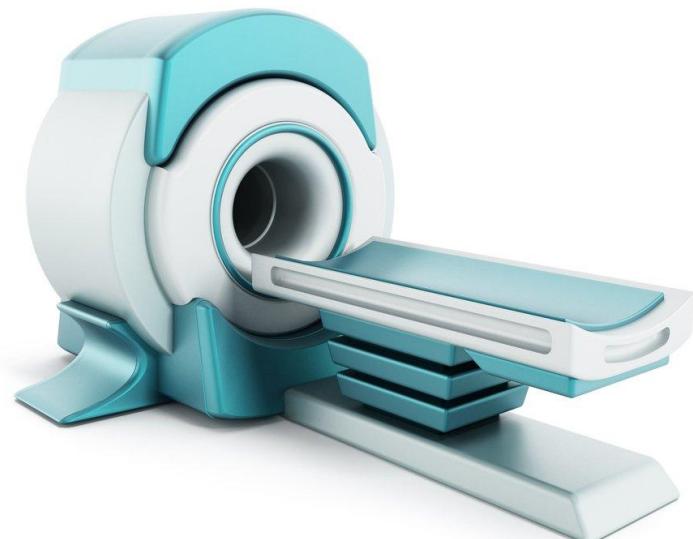
- The top-left image (previous slide) is a standard phantom used in medical imaging - called as Logan-Shepp phantom.
- The top-right image shows 22 radial directions with 512 samples along each, representing those Fourier frequencies (remember we are dealing with 2D frequencies!) which were measured.
- Bottom left: reconstruction obtained using inverse Fourier transform, assuming the rest of the Fourier coefficients were zero.
- Bottom right: image reconstructed by solving the constrained optimization problem in the yellow box on the previous slide. It gives a zero-error result!

THEORY OF COMPRESSIVE SENSING

CS: Theory

- Let a naturally occurring signal be denoted f .
- Let us denote the measurement of this signal (using some device) as y .
- Typically, the mathematical relationship between f and y can be expressed as a linear equation of the form $y = \Phi f$.
- Φ is called the measurement matrix (or the “sensing matrix” or the “forward model” of the measuring device).
- For standard digital camera, Φ may be approximated by a Gaussian blur.

Measuring devices can be represented as linear systems



CS: Theory

- Let \mathbf{f} be a vector with n elements.
- Since the measurements need to be compressive, Φ must have fewer rows (m) than columns (n) to produce a measurement vector \mathbf{y} with m elements.
- We know \mathbf{y} and Φ , and we wish to estimate \mathbf{f} .

CS: Theory

- We know \mathbf{y} and Φ , and we wish to estimate \mathbf{f} .
- We know that in *general*, this is an under-determined linear system, and hence there is no unique solution.
- Why?

$$\Phi \tilde{\mathbf{f}} = \Phi(\tilde{\mathbf{f}} + \mathbf{v}), \text{ where } \Phi \mathbf{v} = \mathbf{0}, \text{ i.e. } \mathbf{v} \in \text{Nullspace}(\Phi)$$

CS: Theory

- But CS theory states, that in certain cases, this system does have a unique solution.
- Conditions to be satisfied:
 1. Vector \mathbf{f} should be sparse (so not all vectors in R^n are potential solutions).
 2. Φ should be “very different from” (“incoherent with”) any row-subset of the identity matrix.

CS: Theory

- It turns out this is wonderful news for signal and image processing.
- Why?
 1. Natural signals/images have a sparse representation in some well-known orthonormal basis Ψ such as Fourier, DCT, Haar wavelet etc.
 2. The measurement matrix Φ can be designed to be incoherent (poorly correlated) with the signal basis matrix Ψ , i.e. the sensing waveforms (rows of Φ) have a very dense representation in the signal basis.

Signal Sparsity

- Many signals have sparse* representations in standard orthonormal bases (denoted here as Ψ).
- Example:

$$\mathbf{f} = \Psi \boldsymbol{\theta} = \sum_{k=1}^n \Psi_k \theta_k,$$

$$\mathbf{f} \in R^n, \boldsymbol{\theta} \in R^n, \Psi \in R^{n \times n}, \Psi^T \Psi = \mathbf{I},$$

$$\|\boldsymbol{\theta}\|_0 \ll n$$

This is the L₀ norm of a vector
= the number of non-zero elements in it

Signal Sparsity



The barbara image (left) of size 512×512 and its reconstruction from just the top 100,000 (out of 262144) DCT coefficients. The difference is indistinguishable and similar results hold for wavelet transforms as well

Incoherent Sampling

- Consider a sensing matrix Φ of size m by n , where m is much less than n (if $m \geq n$, there is no compression in the measurements!).
- We will assume (for convenience) that each row of Φ is unit-normalized.
- CS theory states that Φ and Ψ should be “incoherent” with each other.

Incoherent Sampling

- The coherence between Φ and Ψ is defined as:

$$\mu(\Psi, \Phi) = \sqrt{n} \max_{\substack{1 \leq j \leq m, \\ 1 \leq i \leq n}} \left| \langle \Phi^{(j)}, \Psi_i \rangle \right|$$


j'-th row of Φ , i-th column of Ψ

- We want this quantity μ to be as small as possible.
- Its value always lies in the range $(1, \sqrt{n})$.

Incoherent sampling: Example (1)

- Let the signal basis Ψ be the Fourier basis. A sampling basis Φ that is incoherent with it is the standard spike (Dirac) basis: $\Phi_k(t) = \delta(t - k)$ which corresponds to the simplest and most conventional sampling basis in space or time (i.e. Φ = identity matrix) .
- The associated coherence value is 1 (why?).

Incoherent sampling: Example (2) or Randomness is cool! 😊

- Sensing matrices whose entries are i.i.d. random draws from Gaussian or Bernoulli (+1/-1) distributions are incoherent with **any** given orthonormal basis Ψ with a very high probability.
- Implication: we want our sensing matrices to behave like noise! 😊

Signal Reconstruction

- Let the measured data be given as:
 $y = \Phi f = \Phi \Psi \theta,$
 $y \in R^m, \Phi \in R^{m \times n}, f \in R^n, m \ll n$
- The coefficients of the signal f in Ψ – denoted as θ – and hence the signal f itself - can be recovered by solving the following constrained minimization problem:

Problem P0 : $\min \|\theta\|_0$ such that $y = \Phi \Psi \theta$

This is the L₀ norm of a vector = the number of non-zero elements in it

Signal Reconstruction

- Po is a very difficult optimization problem to solve – it is NP-hard.
- Hence, a softer version (known as Basis Pursuit) is solved:

$$\text{Problem BP : } \min \|\theta\|_1 \text{ such that } \mathbf{y} = \Phi \Psi \theta$$

This is the L₁ norm of a vector = the sum total of the absolute values of all its elements

- This is a linear programming problem and can be solved with any LP solver (in Matlab, for example) or with packages like L₁-magic.

<http://users.ece.gatech.edu/~justin/l1magic/>

Signal Reconstruction: uniqueness issues

- Is P_0 guaranteed to have a unique solution at all? Why? (If answer is no, compressed sensing is not guaranteed to work!)
- Consider the case that any $2S$ columns of an $m \times n$ matrix Φ are linearly independent. Then any S -sparse signal f can be uniquely reconstructed from measurements $y = \Phi f$. See proof on next slide.

Signal Reconstruction: uniqueness issues

- **Proof by contradiction:**
- Suppose we have S -sparse signals \mathbf{f} and \mathbf{f}' such that $\mathbf{y} = \Phi\mathbf{f} = \Phi\mathbf{f}'$.
- Then $\Phi(\mathbf{f}-\mathbf{f}') = \mathbf{o}$ and hence $\mathbf{f}-\mathbf{f}'$ lies in the null-space of Φ .
- But $\mathbf{f}-\mathbf{f}'$ is a $2S$ -sparse signal. Let T be the set of **indices** of its $(2S)$ non-zero elements. Then $\Phi(\mathbf{f}-\mathbf{f}') = \Phi_T(\mathbf{f}-\mathbf{f}')_T = \mathbf{o}$ where Φ_T is a sub-matrix with column indices strictly from T and $(\mathbf{f}-\mathbf{f}')_T$ is a vector containing values of $\mathbf{f}-\mathbf{f}'$ strictly from T .
- Hence $(\mathbf{f}-\mathbf{f}')_T$ lies in the null-space of Φ_T , and hence there exist $2S$ columns of Φ that aren't linearly independent. Hence contradiction!

Signal Reconstruction: uniqueness issues

- But are these conditions on Φ valid?
- The answer is that we need measurement matrices such that Φ will obey this property.
- In fact, we will see something called the “Restricted isometry property”. If Φ satisfies this property, uniqueness is guaranteed.

Why is BP efficiently solvable?

- Because it can be expressed as a linear program, i.e. an optimization problem with linear objective functions and linear equality/inequality constraints.
- Linear programs can be solved in polynomial time.
- Why can it be cast as a linear program? See next slide.

Problem BP : $\min \|\boldsymbol{\theta}\|_1$ such that $\mathbf{y} = \mathbf{A}\boldsymbol{\theta}$

Define $\boldsymbol{\theta} = \boldsymbol{\alpha} - \boldsymbol{\beta}$, where $\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta} \in R^n$

$\boldsymbol{\alpha}$ is non - zero only at positive entries of $\boldsymbol{\theta}$

$\boldsymbol{\beta}$ is non - zero only at negative entries of $\boldsymbol{\theta}$

Define $\boldsymbol{\delta} = [\boldsymbol{\alpha}^t \ \boldsymbol{\beta}^t]^t \in R^{2n}$

$$\therefore \mathbf{A}\boldsymbol{\theta} = [\mathbf{A} \mid -\mathbf{A}]\boldsymbol{\delta}$$

$$\min \sum_{i=1}^{2n} \delta_i \text{ such that } \mathbf{y} = [\mathbf{A} \mid -\mathbf{A}]\boldsymbol{\delta}, \boldsymbol{\delta} \geq \mathbf{0}$$

This is an equivalent formulation of BP and it is a linear program. But why does the new formulation guarantee that the estimated $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ obey our constraint, i.e. the estimated $\boldsymbol{\alpha}$ is non-zero only at the positive entries of $\boldsymbol{\theta}$ and the estimated $\boldsymbol{\beta}$ is non-zero only at the negative entries of $\boldsymbol{\theta}$? See next slide for the answer.

Contradiction to our rescue!

Assume α and β obtained from the *optimal* δ have a non-disjoint support, i.e.

$$\exists i \text{ s.t. } \alpha_i > 0, \beta_i > 0.$$

WLOG assume $\alpha_i \geq \beta_i$.

Consider $\tilde{\alpha}_i = \alpha_i - \beta_i, \tilde{\beta}_i = 0$

Clearly $\tilde{\alpha}_i \geq 0, \tilde{\beta}_i \geq 0$ and $\mathbf{A}(\tilde{\alpha} - \tilde{\beta}) = \mathbf{A}(\alpha - \beta)$.

\therefore the constraint $\mathbf{y} = [\mathbf{A} \mid -\mathbf{A}] \tilde{\delta}$ is still satisfied where $\tilde{\delta} = (\tilde{\alpha} \mid \tilde{\beta})$.

But the objective function is now reduced by $2\beta_i$.

Hence our earlier solution could not have been optimal.

Contradiction!

Signal Reconstruction

- The first use of L₁-norm for signal reconstruction goes back to a PhD thesis in 1965 – by Logan.
- L₁-norm was used in geophysics way back in 1973 (ref: Claerbout and Muir, “Robust modeling with erratic data”, Geophysics, 1973).
- There is something stunning about the L₁-norm problem from the previous slide. We will see soon.
- Before we proceed: Note that there are other methods to (approximately) solve problem P0- eg: greedy approximation algorithms like orthogonal matching pursuit (OMP).

Ref: Candes, Romberg and Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", IEEE Transactions on Information Theory, Feb 2006.

Ref: Donoho, "Compressed Sensing", IEEE Transactions on Information Theory, April 2006.

Theorem 1

- Consider a signal vector \mathbf{f} (having length n) with a sparse representation in basis Ψ , i.e. $\mathbf{f} = \Psi\boldsymbol{\theta}$, where $\|\boldsymbol{\theta}\|_0 \ll n$.
- Suppose \mathbf{f} is measured through the measurement matrix Φ yielding a vector \mathbf{y} of only $m \ll n$ measurements.
- If $m \geq C \log(n/\delta) \|\boldsymbol{\theta}\|_0 \mu^2(\Psi, \Phi)$ (C is some constant) then the solution to the following is **exact** with probability $1-\delta$:

$$\min \|\boldsymbol{\theta}\|_1 \text{ such that } \mathbf{y} = \Phi\Psi\boldsymbol{\theta} \quad \mathbf{y} = \Phi\mathbf{f} = \Phi\Psi\boldsymbol{\theta},$$
$$\mathbf{y} \in R^m, \Phi \in R^{m \times n}, \mathbf{f} \in R^n, m \ll n$$

Comments on Theorem 1

- Look at the lower bound on the number of measurements m : $m \geq C \log(n/\delta) \|\boldsymbol{\theta}\|_0 \mu^2(\Psi, \Phi)$
- The **sparser** the signal \mathbf{f} in the chosen basis Ψ , the **fewer** the number of samples required for exact reconstruction.
- The lower the coherence (**greater the incoherence**) between Ψ and Φ , the **fewer** the number of samples required for exact reconstruction.
- When δ is smaller, the probability of exact reconstruction is higher. The number of samples required is higher, but the increase in number of samples **is an additive factor proportional to $\log \delta$** .

Comments on Theorem 1

- The algorithm to reconstruct f given y makes **NO ASSUMPTIONS** about the sparsity level of f in Ψ , or the **location/magnitude** of the non-zero coordinates of θ .
- The algorithm does not have any notion/consideration for “uniformly spaced samples” – unlike Nyquist’s theorem.
- Even though we are penalizing the L_1 norm (and not L_0 norm), we are still getting a reconstruction that is the sparsest (besides being accurate).
- Although L_1 -norm has been used in sparse recovery for a long time, the optimality proof was established only in 2005/2006 by Donoho, and by Candes, Romberg and Tao.

Comments on Theorem 1: A New Sampling Theorem?

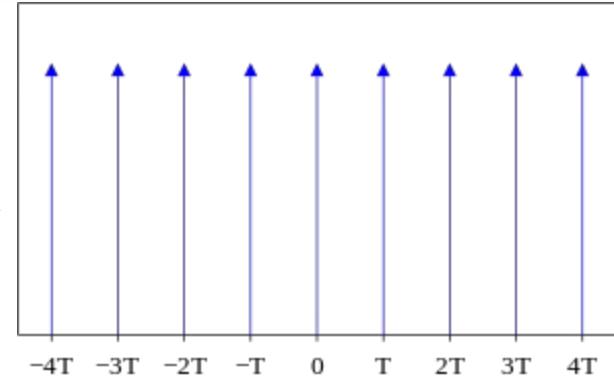
- Let the signal basis Ψ be the Fourier basis. A sampling basis Φ that is incoherent with it is the standard spike (Dirac) basis: $\Phi_k(t) = \delta(t - k)$ which corresponds to the simplest sampling basis in space or time.
- Let the signal f be sparse in this Fourier basis.
- Then given any (**not necessarily uniformly spaced**) m (time-domain) samples of f *chosen uniformly at random*, where $m \geq C \log(n/\delta) \|\Theta\|_0 \mu^2(\Psi, \Phi)$, i.e. $m \geq C \log(n/\delta) \|\Theta\|_0$ we can get an exact reconstruction of f with high probability.
- This acts like a new sampling theorem that does not require uniformly spaced time-samples or bandlimited signals.
- Unlike Shannon's sampling theorem, this one does not reconstruct an analog signal from the samples – just a discrete time signal is constructed.

Comments on Theorem 1: A New Sampling Theorem?

- The measurements are taken from a subset that is drawn uniformly at random from all possible subsets (of the same size).
- Hence there will be some combination of (measurement subset, set of non-zero indices of signal coefficients) for which the reconstruction will not succeed.
- Hence the element of probability of $1-\delta$ in the guarantee for signal reconstruction.

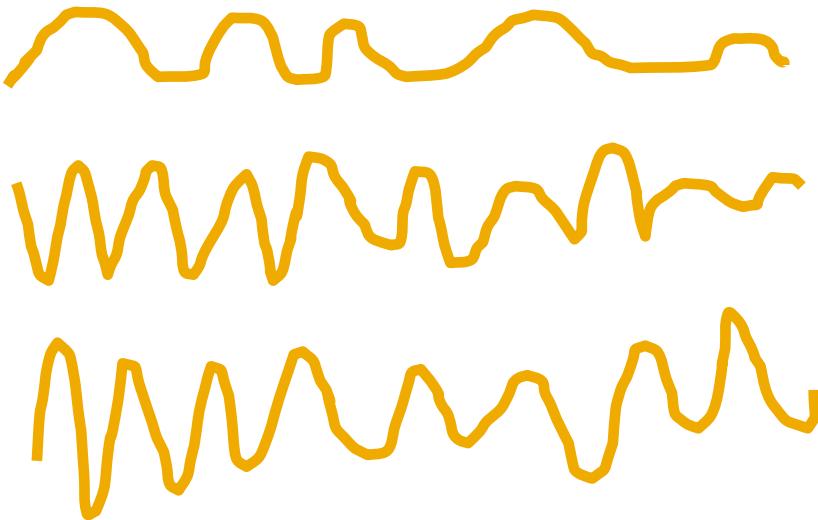
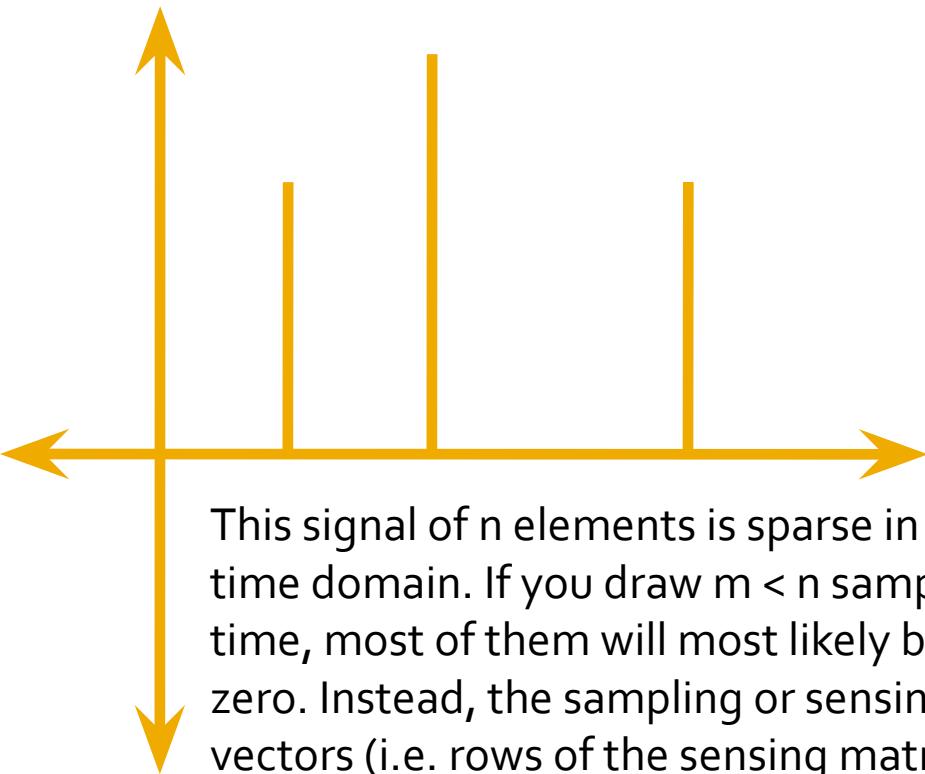
Comments on Theorem 1: A New Sampling Theorem?

- Here is one such example.
- Consider a 1D signal \mathbf{f} with N elements, consisting of uniformly spaced spikes of unit height.
- The spacing between spikes is \sqrt{N} .
- Such a signal \mathbf{f} is called a Dirac comb. It is sparse in time domain.
- The Fourier transform of such a signal is also a Dirac comb!
- Consider that the Fourier measurements (in \mathbf{y}) are at frequencies that are not multiples of \sqrt{N} .
- Then clearly, good signal reconstruction can never be guaranteed, as the observed frequencies and the frequency support of the signal are disjoint.
- However since the measurement frequencies are chosen uniformly at random, such “pathological” cases are rare!



“signal support” is a technical term = set of signal indices with non-zero values.

Intuition behind incoherence



Intuition behind incoherence

- Consider a signal that is sparse in (say) DCT basis – as in the left subfigure on the previous slide.
- Remember: we do not know in advance which DCT coefficients are non-zero and which aren't.
- The measurements are given as:

$$\mathbf{y} = \Phi \Psi \boldsymbol{\theta} = \Phi \sum_{k=1}^n \Psi_k \theta_k$$

$$y_i = \Phi^i \sum_{k=1}^n \Psi_k \theta_k$$

Intuition behind incoherence

- Now suppose our measurement functions (i.e. rows of Φ) were also sparse in the DCT basis.
- Then taking such measurements carries no information about the original signal – as most of these measurements will be zero!

$$\Phi^i = \left(\sum_{k=1}^n \Psi_k \alpha_k \right)^t, \alpha \text{ is a sparse vector}$$

$$y_i = \left(\sum_{l=1}^n \Psi_l \alpha_l \right)^t \left(\sum_{k=1}^n \Psi_k \theta_k \right) = \sum_{k,l} \alpha_l \theta_k \Psi_l^t \Psi_k = \sum_k \alpha_k \theta_k$$

Intuition behind incoherence

- Hence the measurement functions should be **non-sparse** linear combinations of the DCT basis vectors = having measurements which are **non-sparse** linear combinations of all the **DCT coefficients**.
- Thereby **each** measurement will have information about **all** the DCT coefficients.

New concept: Restricted Isometry Property (RIP)

- For integer $S = 1, 2, \dots, n$, the restricted isometry constant (RIC) δ_S of a matrix $\mathbf{A} = \Phi\Psi$ (*) of size m by n is the smallest number such that for any S -sparse vector $\boldsymbol{\theta}$, the following is true:
$$(1 - \delta_S) \|\boldsymbol{\theta}\|^2 \leq \|\mathbf{A}\boldsymbol{\theta}\|^2 \leq (1 + \delta_S) \|\boldsymbol{\theta}\|^2.$$
- We say that \mathbf{A} obeys the restricted isometry property (RIP) of order S if δ_S is not too close to 1 (rather as close to 0 as possible).
- If \mathbf{A} obeys RIP of order S , no S -sparse signal can lie in the null-space of \mathbf{A} (if it did, then we would have $\mathbf{A}\boldsymbol{\theta} = \mathbf{0}$ and that obviously does not preserve the squared magnitude of the vector $\boldsymbol{\theta}$).

Restricted Isometry Property

- Let's suppose we wanted to sense a signal \mathbf{x} that is S -sparse in some orthonormal basis in which it has representation $\boldsymbol{\theta}$, i.e. $\|\boldsymbol{\theta}\|_0 = S \ll n$.
- Then the following is *undesirable* for \mathbf{A}
 $\mathbf{A}\boldsymbol{\theta}^{(1)} = \mathbf{A}\boldsymbol{\theta}^{(2)}$, for $\boldsymbol{\theta}^{(1)} \neq \boldsymbol{\theta}^{(2)}$
- One way to ensure that this doesn't happen is to design \mathbf{A} such that:

$$\mathbf{A}\boldsymbol{\theta}^{(1)} \approx \mathbf{A}\boldsymbol{\theta}^{(2)} \leftrightarrow \boldsymbol{\theta}^{(1)} \approx \boldsymbol{\theta}^{(2)}, \text{i.e.}$$

$$\|\mathbf{A}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})\|^2 \approx \|(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})\|^2$$

Restricted Isometry Property (RIP)

- Note that the difference between two S -sparse vectors is $2S$ -sparse. Then, if \mathbf{A} obeys RIP of order $2S$, we have the following:

$$(1 - \delta_{2S})\|\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\|^2 \leq \|\mathbf{A}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})\|^2 \leq (1 + \delta_{2S})\|\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\|^2,$$

$\forall \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}$, for which $\|\boldsymbol{\theta}^{(1)}\|_0 \leq S, \|\boldsymbol{\theta}^{(2)}\|_0 \leq S$, where $\delta_{2S} \ll 1$

- Thus \mathbf{A} should **approximately preserve the squared differences between any two S -sparse vectors.**

Theorem 2

- Suppose the matrix $\mathbf{A} = \Phi\Psi$ of size m by n (where sensing matrix Φ has size m by n , and basis matrix Ψ has size n by n) has RIP property of order $2S$ where $\delta_{2S} < 0.41$. Let the solution of the following be denoted as θ^* , (for signal $\mathbf{f} = \Psi\theta$, measurement vector $\mathbf{y} = \Phi\Psi\theta$): $\min \|\theta\|_1$ such that $\mathbf{y} = \Phi\Psi\theta$

Then we have:

θ_s is created by retaining the S largest magnitude elements of θ in their correct locations, and setting the rest to 0.

$$\|\theta^* - \theta\|_2 \leq \frac{C_0}{\sqrt{S}} \|\theta - \theta_s\|_1,$$

$$\|\theta^* - \theta\|_1 \leq C_0 \|\theta - \theta_s\|_1$$

Comments on Theorem 2

- This theorem says that the reconstruction for S -sparse signals is **always** exact if \mathbf{A} satisfies the RIP.
- For signals that are not S -sparse, the reconstruction error is almost as good as what could be provided to us by an **oracle** which knew the S largest coefficients (and also their indices) of the signal vector \mathbf{f} in the basis Ψ .

Comments on Theorem 2

- We know that most signals are never exactly sparse in standard ortho-normal bases. But they are compressible, i.e. there is a sharp decay in the coefficient magnitude with many coefficients being nearly zero!
- Theorem 2 handles such compressible signals robustly. Theorem 1 did not apply to compressible signals! Here again, Theorem 2 is more powerful than Theorem 1.
- The constant C is independent of n , and it is an increasing function of just δ_{2S} in $[0,1]$.

Compressive Sensing under Noise

- So far we assumed that our measurements were exact, i.e. $\mathbf{y} = \Phi\mathbf{f} = \Phi\Psi\boldsymbol{\theta}$.
- But practical measurements are always noisy so that $\mathbf{y} = \Phi\mathbf{f} + \boldsymbol{\eta} = \Phi\Psi\boldsymbol{\theta} + \boldsymbol{\eta}$.
- Under the same assumption as before, we can estimate $\boldsymbol{\theta}$ by solving the following problem:

$$\min \|\boldsymbol{\theta}\|_1 \text{ such that } \|\mathbf{y} - \Phi\Psi\boldsymbol{\theta}\|_2^2 \leq \varepsilon$$

Convex problem, called as second-order cone program. Can be solved efficiently with standard packages including MATLAB and L1-MAGIC

Theorem 3

- Suppose the matrix $\mathbf{A} = \Phi\Psi$ of size m by n (where sensing matrix Φ has size m by n , and basis matrix Ψ has size n by n) has RIP property of order $2S$ where $\delta_{2S} < 0.41$. Let the solution of the following be denoted as $\boldsymbol{\theta}^*$, (for signal $\mathbf{f} = \Psi\boldsymbol{\theta}$, measurement vector $\mathbf{y} = \Phi\Psi\boldsymbol{\theta} + \boldsymbol{\eta}$):
$$\min \|\boldsymbol{\theta}\|_1 \text{ such that } \|\mathbf{y} - \Phi\Psi\boldsymbol{\theta}\|_2^2 \leq \varepsilon$$

Then we have:

$\boldsymbol{\theta}_S$ is created by retaining the S largest magnitude elements of $\boldsymbol{\theta}$ in their correct locations, and setting the rest to 0.

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|_2 \leq \frac{C_0}{\sqrt{S}} \|\boldsymbol{\theta} - \boldsymbol{\theta}_S\|_1 + C_1 \varepsilon$$

Comments on Theorem 3

- Theorem 3 is a direct extension of Theorem 2 for the case of noisy measurements.
- It states that the solution of the convex program (see previous slides) gives a reconstruction error which is the sum of two terms: (1) the error of an oracle solution where the oracle told us the S largest coefficients of the signal \mathbf{f} , and (2) a term proportional to the noise variance.
- The constants C_o and C_1 are very small (less than or equal to 5.5 and 6 respectively for $\delta_{2S} = 0.25$) , they are independent of n , and they are increasing functions of just δ_{2S} in $[0,1]$.

Comments on Theorem 3

- How do you choose ε in practice?
- It depends on the noise model.
- Note, if $(\|\eta\|_2)^2 \leq \kappa$, then one must choose $\varepsilon \geq \kappa$ (at least, with high probability).

Comments on Theorem 3

- Consider: $y = \Phi f + \eta = \Phi \Psi \theta + \eta$
- If for $i = 1$ to m , $\eta_i \sim \text{Uniform}(-r, r)$, i.e. uniform random noise with known r , then $\varepsilon \geq r^2 m$.
- This represents quantization noise.

Comments on Theorem 3

- Consider: $\mathbf{y} = \Phi\mathbf{f} + \boldsymbol{\eta} = \Phi\Psi\boldsymbol{\theta} + \boldsymbol{\eta}$
- If for each $i = 1$ to m , $\eta_i \sim N(0, \sigma^2)$, with known σ .
- Then, the squared magnitude of the vector $\boldsymbol{\eta}$ is a chi-square random variable.
- Hence with very high probability, the magnitude of $\boldsymbol{\eta}$ will lie within 3 standard deviations from the mean, i.e. set $\varepsilon \geq 9m\sigma^2$.
- In such a case, Theorem 3 will hold with high probability even when \mathbf{A} obeys RIP.

Randomness is super-cool! 😊

- Consider any fixed orthonormal basis Ψ .
- For a sensing matrix Φ constructed in any of the following ways, the matrix $\mathbf{A} = \Phi\Psi \in \mathbb{R}^{m \times n}$, will obey the RIP of order S with overwhelming probability provided that the number of rows $m \geq CS \log(n/S)$
 1. Φ contains entries from zero-mean Gaussian with variance $1/m$.
 2. Φ contains entries with values $\pm 1/\sqrt{m}$ with probability 0.5 (symmetric Bernoulli).
 3. Columns of Φ are sampled uniformly randomly from a unit sphere in m -dimensional space.

Randomness is super-super-cool! 😊

- These properties of random matrices hold true for **any** orthonormal basis Ψ with very high probability.
- As such, we do not need to tune Φ for a given Ψ .

Randomness is super-cool! 😊

- A sensing matrix Φ containing a random subset of rows of the Discrete Fourier Matrix also obeys the RIP.
- Provided the number of rows $m \geq O(S \log^4 n)$.
- Similar results hold for row-subsampled versions of other orthonormal matrices whose values are upper bounded by $O(1/n)$.

M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, Aug. 2008

L_1 norm and L_0 norm

- There is a special relationship between the following two problems :

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta}$$

$$\mathbf{y} \in R^n, \mathbf{A} \in R^{m \times n}$$

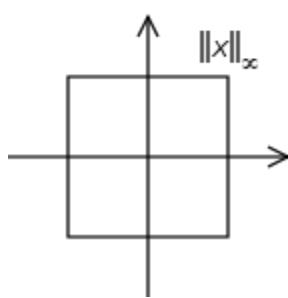
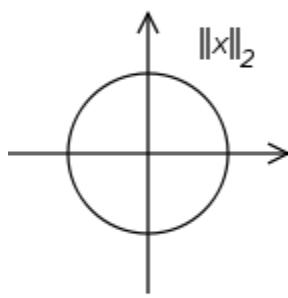
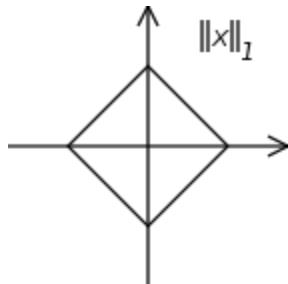
$$\boldsymbol{\theta} \in R^n (m \ll n)$$

$$\boldsymbol{\theta}^* = \min \|\boldsymbol{\theta}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\boldsymbol{\theta} \quad \boldsymbol{\theta}^* = \min \|\boldsymbol{\theta}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\boldsymbol{\theta}$$

The L_1 norm is a “softer” version of the L_0 norm. Other L_p -norms where $0 < p < 1$ are possible and impose a stronger form of sparsity, but they lead to non-convex problems. Hence L_1 is preferred.

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

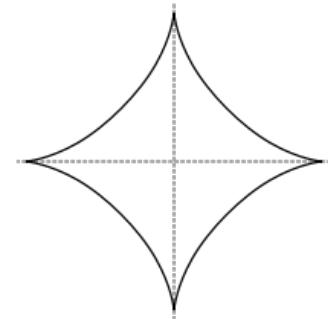
L_1 is good, L_2 is bad. Why?



Unit L_p -circle (defined below) for different values of p . **Left side**, top to bottom, $p = 1$ (square with diagonals coinciding with Cartesian axes), $p = 2$ (circle in the conventional sense), $p = \infty$ (square with sides parallel to the axes); **right side**, $p = 2/3$ (astroid).

$$(|x|^p + |y|^p)^{\frac{1}{p}} = 1 \quad \text{In 2-D}$$

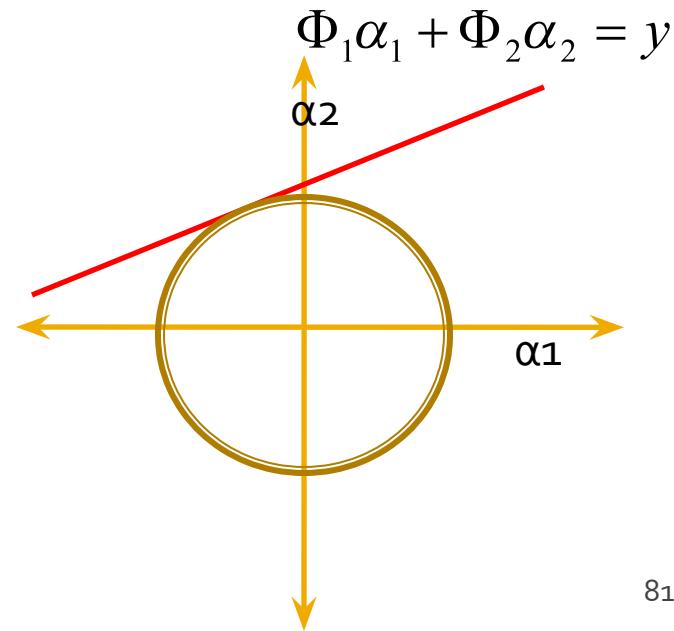
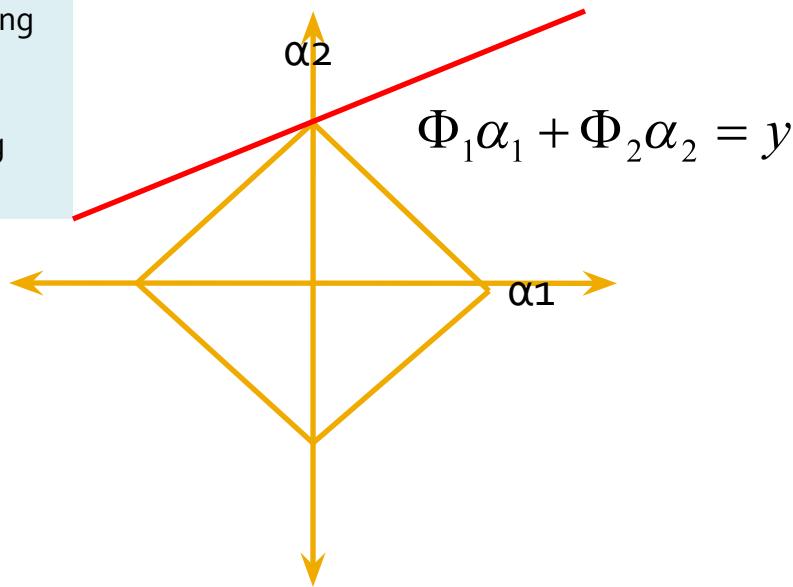
$$\left(\sum_{i=1}^d |x_i|^p \right)^{\frac{1}{p}} = 1 \quad \text{In d-dimensions}$$



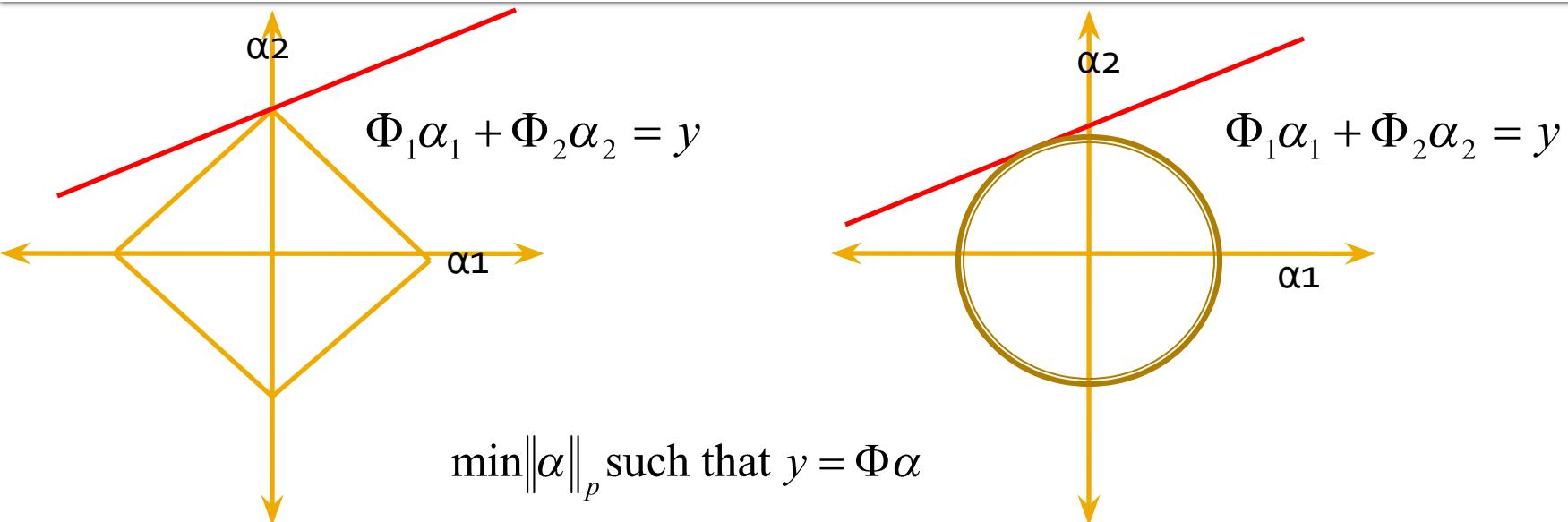
L_1 is good, L_2 is bad. Why?

- Look at the basic problem we are trying to solve:
 $\min\|\alpha\|_p$ such that $y = \Phi\alpha; p \in \{1,2\}$
- In 2D (i.e. α is a 2×1 vector, y is a scalar, Φ is a 1×2 matrix, and $y = \Phi\alpha$ being a line in 2D space), the problem can be represented as the following diagram:

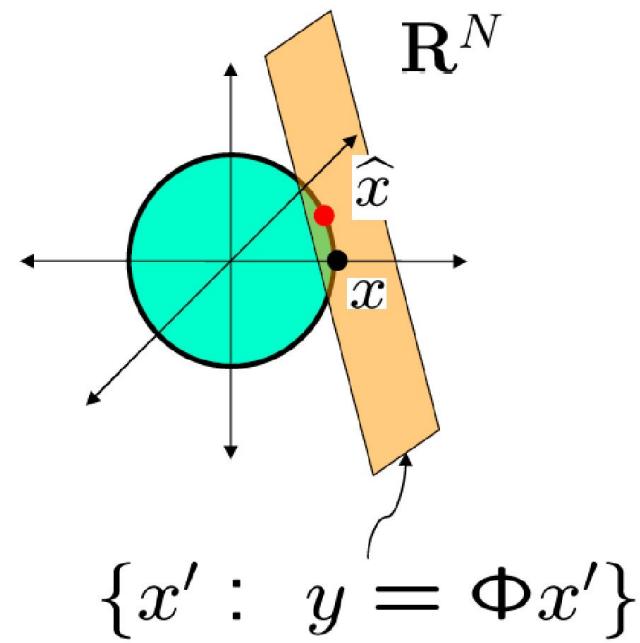
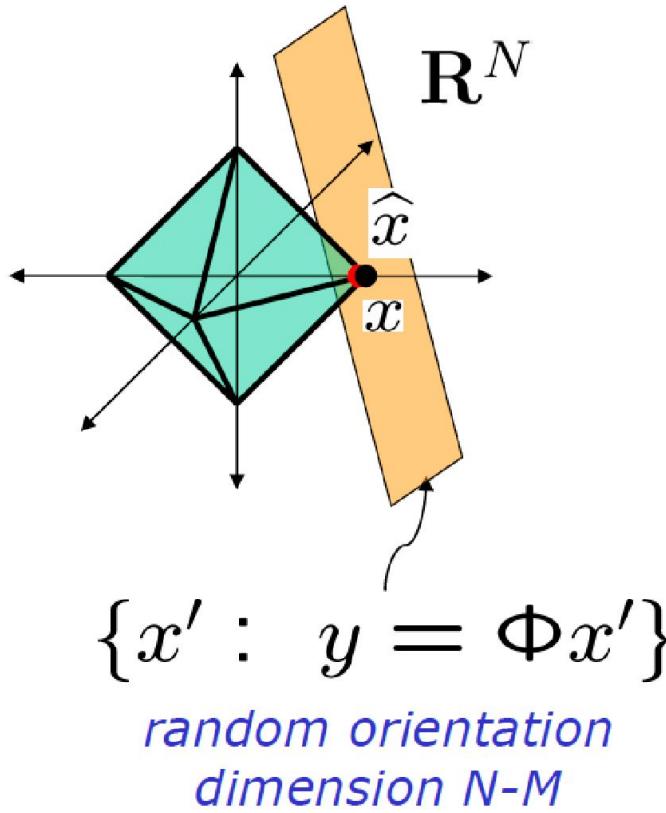
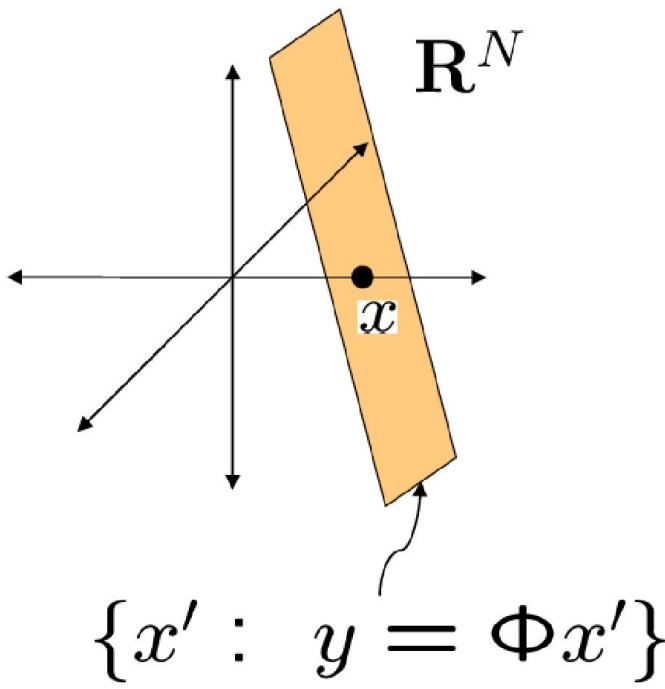
Image source:
Romberg, "Imaging
via compressive
sampling", IEEE
Signal Processing
Magazine



L_1 is good, L_2 is bad. Why?



- Note that the line $y = \Phi\alpha$ represents a HARD constraint, i.e. any vector α must satisfy this constraint.
- We want to find the vector α with minimum L_p -norm which satisfies this constraint.
- So we grow L_p -circles of increasing radius (denoted as r and defined in the sense of the L_p -norm) starting from the origin, until the L_p -circle touches the line $y = \Phi\alpha$.
- The very first time the L_p -circle touches the line, gives us the solution (the second time, the L_p -circle will have a greater radius and hence the solution vector α will have higher L_p -norm).
- For the L_1 -norm, the solution vector α will be sparse (see left figure). For the L_2 -norm, the solution vector will not be sparse (right figure).



Another example – see the L₁ ball versus the L₂ ball – both in 3D. In this case the constraint is a plane, given by $y = \Phi x$ where y is scalar, x is a vector of 3 elements, and Φ has size 1×3 .

Source of Images:

<http://web.yonsei.ac.kr/nipi/lectureNote/CompressedSensing%20by%20Romberg%20and%20Wakin.pdf>

L_1 is good, L_2 is bad. Why?

- The L_1 -norm minimization (subject to the chosen constraint) is guaranteed to give the sparsest possible solutions provided the conditions of theorems 1, 2 and 3 are met.
- Note that we get the sparsest possible solutions even though the L_1 -norm is a softened version of the L_0 -norm!

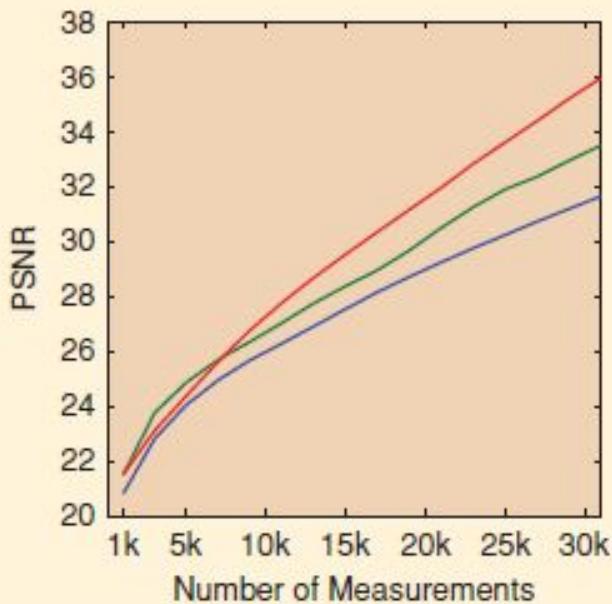
Compressive Sensing: Toy Example with images

- We will compare image reconstruction by two different methods.
- We won't do true compressive sensing. Instead we'll apply different types of sensing matrices Φ synthetically (in software) to an image loaded into memory, merely for example sake.
- In the first method, we compute the DCT coefficients of an image (size $N = 256 \times 256$), and reconstruct the image using the m largest absolute value DCT coefficients (and a simple matrix transpose). m is varied from 1000 to 30000.
- Note that this is equivalent to using a sensing matrix Φ of size m by N .

Compressive Sensing: Toy Example with images

- In the second method, we reconstruct the image using m random linear combinations of elements of signal \mathbf{x} , where m is again varied from 1000 to 30000.
- The sensing matrix Φ is appropriately assembled.
- In this case, the reconstruction proceeds by solving the following optimization problem:
$$\min \|\boldsymbol{\theta}\|_1 \text{ such that } \mathbf{y} = \Phi \Psi \boldsymbol{\theta}$$

Compressive Sensing: Toy Example with images



(a)



(b)



(c)



[FIG2] Coded imaging simulation. (a) Recovery error versus number of measurements for linear DCT acquisition (blue), compressive imaging (red), and DCT imaging augmented with total-variation minimization (green). The error is measured using the standard definition of peak signal-to-noise ratio: $\text{PSNR} = 20 \log_{10}(255 \cdot 256 / \|X - \hat{X}\|)$. (b) Image recovered using linear DCT acquisition with 21,000 measurements. (c) Image recovered using compressive imaging from 1,000 DCT and 20,000 noiselet measurements.

Better edge preservation with the randomly assembled Φ – which actually uses some high frequency information (unlike conventional DCT which discards higher frequencies). Higher PSNR values with the randomly assembled sensing matrix as compared to the linear DCT-based method.

Image source: Romberg, "Imaging via compressive sampling", IEEE Signal Processing Magazine

Compressed Sensing for Piecewise Constant Signals

- In the case of piecewise constant signals/images, the gradients are sparse (zero everywhere except along edges).
- For such images, we can solve the following problem (based on total variation or TV-norm):

$\min TV(f)$ such that $y = \Phi\Psi^T f$,

$$TV(f) \leftarrow \iint \sqrt{f_x^2 + f_y^2} dx dy$$

= L1 – norm of the 2D gradient vector

This optimization problem is called a second order cone program. It can be solved efficiently using packages like L1-magic. We will not go into more details, but the documentation of L1-magic is a good starting point for the curious readers.

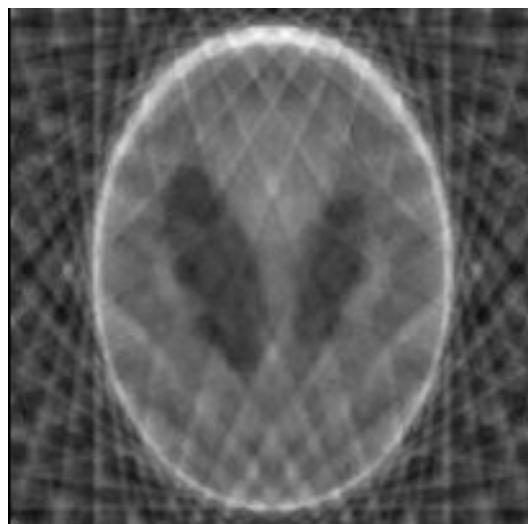
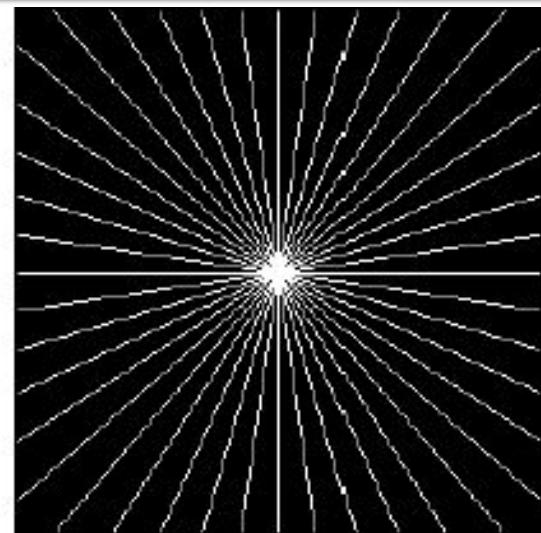
Theorem 4 (similar to Theorem 1)

- Consider a piecewise constant signal f (having length n). Suppose coefficients of f are measured through the measurement matrix Φ yielding a vector y of only $m \ll n$ measurements.
If $m \geq C \log(n/\delta) \|\theta\|_0 \mu^2(\Psi, \Phi)$ (C is some constant) then the solution to the following is **exact** with probability $1-\delta$:

$$\min_f TV(f) \text{ such that } y = \Phi \Psi^T f \quad y \in R^m, \Phi \in R^{m \times n}, f \in R^n, m \ll n$$

Experiment by Candes, Romberg and Tao

Ref: Candes, Romberg and Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", IEEE Transactions on Information Theory, Feb 2006.



Relevant theorem behind Candes' experiment

- Take a look at theorem 1 (scroll back!).
- Consider a 1D signal f (having N elements) that is piecewise constant – the gradients of such a signal form a sparse vector!
- Suppose we have measured a subset C of the Fourier coefficients of f . Let us call these measurements as $G(u)$, $u \in C$. These form a vector g .
- We wish to estimate f (exploiting the fact that it is piecewise constant) given g .
- How do we do this?

Relevant theorem behind Candes' experiment

- Define vector \mathbf{h} as follows:

$$h(x) = f(x) - f(x-1)$$

$$\therefore H(u) = (1 - e^{-j2\pi u/N})F(u)$$

- So we will do the following:

$$\mathbf{h}^* = \arg \min_{\mathbf{h}} \|\mathbf{h}\|_1 \text{ such that}$$

$$\mathbf{g} = \mathbf{I}_c \mathbf{D} \mathbf{H} = \mathbf{I}_c \mathbf{D} \Psi^* \mathbf{h} = \mathbf{A} \mathbf{h}$$

where

$\Psi^* \mathbf{h}$ = Fourier coefficients of \mathbf{h} (Note : Ψ^* stands for conjugate transpose of Ψ)

\mathbf{D} = diagonal matrix for which $D(u, u) = (1 - e^{-j2\pi u/N})^{-1}$

\mathbf{I}_c = matrix of size $|\mathbf{C}|$ by N which selects only those frequency coefficients from set \mathbf{C}

Experiment by Candes, Romberg and Tao

- Invoking Theorem 1 (scroll back), we can estimate \mathbf{h} accurately. Then we can estimate \mathbf{f} accurately (up to an additive constant).
- In case of images (2D signals) that are piecewise constant, the vector \mathbf{h} will contain the magnitudes of the gradients at all the pixels, i.e.

$$\begin{aligned} h(i) &= \sqrt{(f(x, y) - f(x-1, y))^2 + (f(x, y) - f(x, y-1))^2} \\ &= |(f(x, y) - f(x-1, y)) + j(f(x, y) - f(x, y-1))| \\ &\text{where pixel } (x, y) \text{ maps on to index } i, \\ &\text{and } j = \sqrt{-1} \end{aligned}$$

Experiment by Candes, Romberg and Tao

- Our measurement matrix in this case is a partial Fourier matrix. It is known to be highly incoherent with the spatial domain!
- The L₁-norm we are penalizing in this case turns out to be the total variation (**sum total of gradient magnitudes** over the entire image)!
- It was this experiment by Candes, Romberg and Tao (done around 2005) that inspired modern compressive sensing theory!

Experiment by Candes, Romberg and Tao

- Compare the following two problems:

$$\min \|\theta\|_1 \text{ such that } y = \Phi f = \Phi \Psi \theta \quad \min_f TV(f) \text{ such that } y = \Phi \Psi^* f$$

- In the former case, we are measuring the image/signal (**spatial domain**) through Φ .
- In the second case, we are making measurements in the **frequency domain** (Fourier coefficient values) through Φ .

Mutual coherence and Restricted Isometry Property

Concept of mutual coherence

- Consider sensing matrix Φ , representation matrix Ψ and matrix $\mathbf{A} = \Phi \Psi$.
- The mutual coherence of \mathbf{A} – denoted $\mu(\mathbf{A})$ - is defined as the maximum absolute normalized dot product of any column of \mathbf{A} with any other column of \mathbf{A} .
- In other words:

$$\mu(\mathbf{A}) = \max_{i,j, i \neq j} \frac{|\mathbf{A}_i^T \mathbf{A}_j|}{\|\mathbf{A}_i\|_2 \|\mathbf{A}_j\|_2}$$

Theorem 5

- Consider measurement $y = \Phi x + \eta = \Phi \Psi \theta + \eta = A\theta + \eta$.
- We want to estimate x via θ .
- Let θ_s be a sub-vector of θ with the S largest elements in correct locations.
- Then if $S < 0.5(1 + 1/\mu(A))$, the solution θ^* to the optimization problem P_2 yields the following error bounds:

$$\|\theta - \theta^*\|_2 \leq C_0(\varepsilon + \varepsilon') + C_1\|\theta - \theta_s\|_1, \|\eta\|_2 \leq \varepsilon$$

- See problem P_2 on next slide.

Theorem 5

- Problem P₂ is the following:
 $\min \|\boldsymbol{\theta}\|_1$ such that $\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2 \leq \varepsilon'$, $\varepsilon \leq \varepsilon'$
- The constants C_0 and C_1 depend upon S and $\mu(\mathbf{A})$.
- Note the similarity between this theorem and Theorem 3!
- This theorem handles noise as well as compressible signals but it uses the mutual coherence criterion and not the RIP (unlike theorem 3).

Theorem 3 versus Theorem 5

- The sufficient condition for Theorem 5 is easily verifiable in $O(mn^2)$ time.
- The RIC of a matrix (of arbitrary order S) cannot in general be computed efficiently.
- Why? Consider for S -sparse θ :

$$(1 - \delta_S) \|\theta\|^2 \leq \|A\theta\|^2 \leq (1 + \delta_s) \|\theta\|^2.$$

 RIC of A (of order S)

Theorem 3 versus Theorem 5

- Now consider the following relationship:

$$\delta_S = \max\{1 - \lambda_{\min}, \lambda_{\max} - 1\} \text{ where}$$

$$\lambda_{\max} = \max_{\boldsymbol{\theta}_\Gamma \in R^S, |\Gamma| \leq S} \frac{\|\mathbf{A}_\Gamma \boldsymbol{\theta}_\Gamma\|^2}{\|\boldsymbol{\theta}_\Gamma\|^2}, \lambda_{\min} = \min_{\boldsymbol{\theta}_\Gamma \in R^S, |\Gamma| \leq S} \frac{\|\mathbf{A}_\Gamma \boldsymbol{\theta}_\Gamma\|^2}{\|\boldsymbol{\theta}_\Gamma\|^2}$$

A subset (of maximum size S) of possible non-zero indices of vector $\boldsymbol{\theta}$

These indicate the maximum and minimum eigenvalues of the matrix $(\mathbf{A}_\Gamma)^T(\mathbf{A}_\Gamma)$ over all sets Γ . In general, the maximum eigenvalue of a symmetric matrix $\mathbf{B}^T \mathbf{B}$ is given by $\max_x \|\mathbf{B}\mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2$. The vector \mathbf{x} that realizes the maximum value is the corresponding eigenvector. Likewise for minimum eigenvalue.

Theorem 3 versus Theorem 5

- This therefore requires us to enumerate all the different subsets Γ of maximum size S , and compute the maximum and minimum eigenvalue of matrices $(\mathbf{A}_\Gamma)^\top(\mathbf{A}_\Gamma)$.
- Subset enumeration is a problem that requires $O(n^S)$ time.
- Since there is no upper bound on S , this is clearly not an efficient algorithm – or even a polynomial time algorithm.

Theorem 3 versus Theorem 5

- However the coherence based bounds are usually too conservative (why?), and the RIC-based bounds are considered tighter.
- We will see why when we do the proofs of these results, but we can have an intuitive explanation based on the famous Gershgorin's disc theorem.

Theorem 3 versus Theorem 5

- The Gershgorin's disc/circle theorem states that every eigenvalue of a square matrix \mathbf{B} lies in the union of the so-called **Gershgorin discs** $D(B_{ii}, r_i)$.
- Here r_i is the radius of the i -th Gershgorin disc and is defined as the sum of the absolute values of the off-diagonal elements of the i -th row.
- In other words, we have:

For every eigenvalue λ , $\exists i \in \{1, \dots, n\}$ such that $B_{ii} - r_i \leq \lambda \leq B_{ii} + r_i$

Theorem 3 versus Theorem 5

- Simple example of Gershgorin's theorem:

$$B = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \rightarrow (\lambda - 1)(\lambda - 4) - 4 = 0$$

$\therefore \lambda = 0$ or 5 , both of which lie in $D(1,2)$ or $D(4,2)$

Theorem 3 versus Theorem 5

- Applying this theorem to the matrix $\mathbf{A}^T \mathbf{A}$ and using the fact that the vector \mathbf{x} in the definition is S -sparse, we can conclude that $\delta_S \leq \mu(S - 1)$.
- Theorem 5 does not require the representation matrix Ψ to be orthonormal.
- And neither does Theorem 3.

Some history: Observations by Logan (1965)

- Consider a band-limited signal f (of maximum frequency B) corrupted by **impulse noise** (i.e. spikes) of arbitrary magnitude, yielding the corrupted signal: $y(t) = f(t) + \eta(t)$
- Let the support of the impulse noise be sparse in time (valid assumption), having length T . Support means the set of indices for which the impulse noise is non-zero.
- Consider the following reconstruction criterion:

$$\min_f \|y - f\|_1 \text{ subject to } \forall \omega > B, \hat{f}(\omega) = 0$$

Ref: Logan, "Properties of high-pass signals", PhD Thesis, Columbia University, 1965

Some history: Observations by Logan (1965)

- It turns out that this reconstruction yields **PERFECT** results **regardless of the noise magnitude**, provided that the following condition holds: $TB \leq \frac{\pi}{2}$
- This result exploits the band-limitedness of the signal f , but does not impose constraints on the sampling rate.
- Note: this result requires that if B is large, then T must be small.
- This was a very early application of L_1 -norm optimization for signal reconstruction.

Summary

- Key Theorems
- Concept of Incoherence
- Concept of RIP
- Concept of mutual coherence
- ℓ_0 , ℓ_1 and ℓ_2 norms for reconstruction
- Key references: within the slides, see also class web-page