Compressive Sensing

CS 763

Ajit Rajwade

Outline of the Lectures

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

INTRODUCTION + MOTIVATION

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

$$F(u) = \sqrt{N} \sum_{n=0}^{N-1} f(n) e^{-i2\pi un/N}$$

u-th Fourier coefficient

Fourier coefficient

$$f(n) = \frac{1}{\sqrt{N}} \sum_{u=0}^{N-1} F(u) e^{i2\pi un/N}$$

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

(complex) sinusoidal bases

$$\mathbf{f} = \mathbf{HF}$$

Vector of *N* Fourier coefficients

N x *N* orthonormal matrix (Fourier basis matrix)

Fourier Basis Matrix

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

$$\mathbf{F} = \mathbf{H}^{\mathrm{T}}\mathbf{f}$$

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

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

n varies across rows (constant over all entries in a given row). *u* varies across columns (constant across all entries in a given column)

2D Fourier Transform

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

F and ***f*** are vectors of length *MN*

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

Vector of size *MN* x 1

$$\mathbf{f} = \mathbf{HF}$$

Matrix of size *MN* x *MN*

Discrete Cosine Transform (DCT) in 1D

$$F(u) = \sum_{n=0}^{N-1} f(n) a_N^{un}$$

$$f(n) = \sum_{u=0}^{N-1} F(u) \tilde{a}_N^{un}$$

$$DCT : a_N^{un} = \sqrt{\frac{1}{N}}, u = 0$$

$$a_N^{un} = \sqrt{\frac{2}{N}} \cos\left(\frac{\pi(2n+1)u}{2N}\right), u = 1 \ldots N-1$$

$$\tilde{a}_N^{un} = a_N^{un}$$

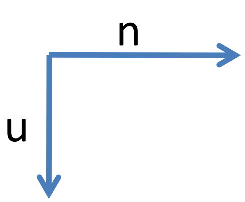
$$DFT : a_N^{un} = e^{-j2\pi\frac{un}{N}}$$

$$\tilde{a}_N^{un} = a^{*un}_{\ N} \text{(complex conjugate)}$$

Discrete Cosine Transform (DCT) in 1D

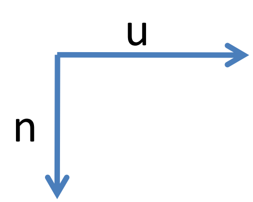
$$\begin{pmatrix} F(0) \\ \cdot \\ \cdot \\ \cdot \\ F(N-1) \end{pmatrix} = \begin{pmatrix} a_N^{0,0} & \cdot & \cdot & \cdot & a_N^{0,N-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_N^{N-1,0} & \cdot & \cdot & \cdot & a_N^{N-1,N-1} \end{pmatrix} \begin{pmatrix} f(0) \\ \cdot \\ \cdot \\ \cdot \\ f(N-1) \end{pmatrix}$$

$$\begin{pmatrix} f(0) \\ \cdot \\ \cdot \\ \cdot \\ f(N-1) \end{pmatrix} = \begin{pmatrix} a_N^{0,0} & \cdot & \cdot & \cdot & a_N^{0,N-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_N^{N-1,0} & \cdot & \cdot & \cdot & a_N^{N-1,N-1} \end{pmatrix} \begin{pmatrix} F(0) \\ \cdot \\ \cdot \\ \cdot \\ F(N-1) \end{pmatrix}$$



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

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



$$\tilde{\mathbf{A}} \in R^{N\times N} \ (DCT \text{ Basis Matrix})$$

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I}$$

$$DCT : \tilde{\mathbf{A}} = \mathbf{A}^T$$

$$DFT : \tilde{\mathbf{A}} = \mathbf{A}^{*T} \text{ (conjugate transpose)}$$

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.

$$\boxed{\begin{pmatrix} f(0) \\ \cdot \\ \cdot \\ \cdot \\ f(N-1) \end{pmatrix} = \begin{pmatrix} a_N^{0,0} & \cdot & \cdot & \cdot & a_N^{0,N-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_N^{N-1,0} & \cdot & \cdot & \cdot & a_N^{N-1,N-1} \end{pmatrix} \begin{pmatrix} F(0) \\ \cdot \\ \cdot \\ \cdot \\ F(N-1) \end{pmatrix}}$$

u

n

$$\tilde{\mathbf{A}} \in R^{N \times N} (DCT \text{ Basis Matrix})$$

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I}$$

- 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

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

$$f(n,m)=\sum_{u=0}^{N-1}F(u,v)\tilde{a}_{NM}^{unvm}$$

The DCT matrix is this case will have size *MN* x *MN*, and it will be the Kronecker product of two DCT matrices – one of size *M* x *M*, the other of size *N* x *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.

$$DCT:$$

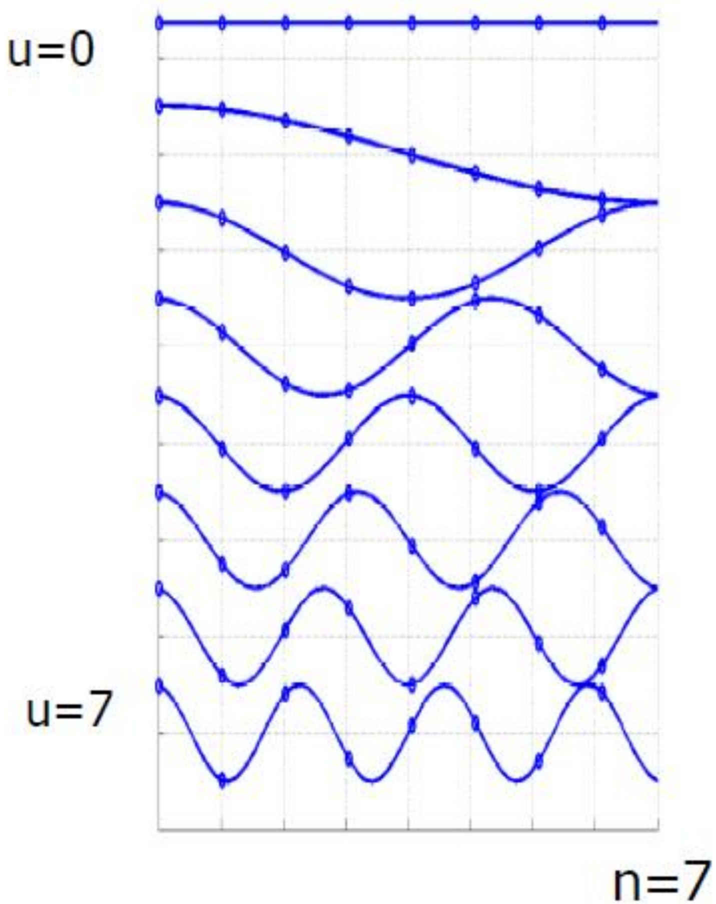
$$a_{NM}^{unvm}=\alpha(u)\alpha(v)\cos\left(\frac{\pi(2n+1)u}{2N}\right)\cos\left(\frac{\pi(2m+1)v}{2M}\right), u=0\ldots N-1, v=0\ldots M-1$$

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

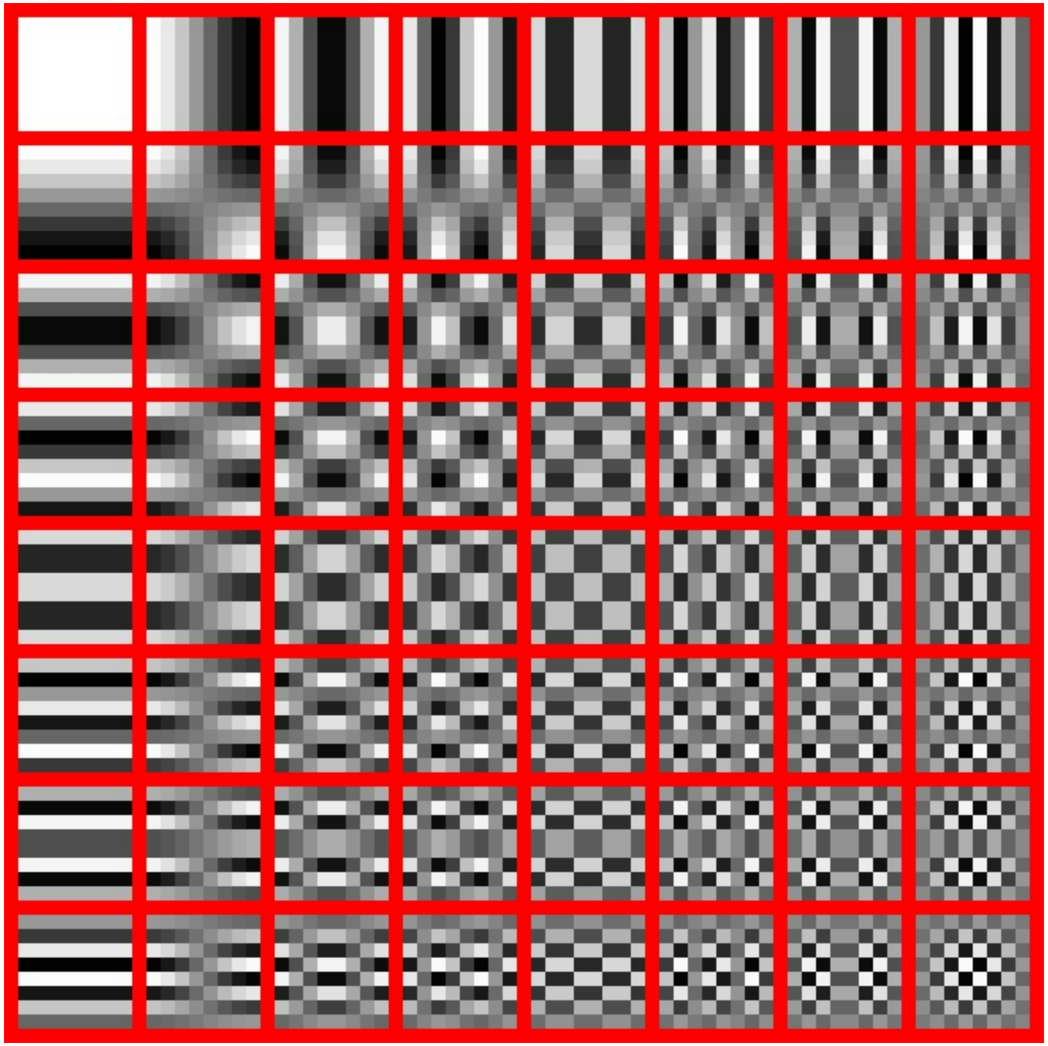
$$\alpha(v)=\sqrt{1/M}\ (v=0), \text{else } \alpha(v)=\sqrt{2/M}$$

$$\tilde{a}_{NM}^{unvm}=a_{NM}^{unvm}$$

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



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

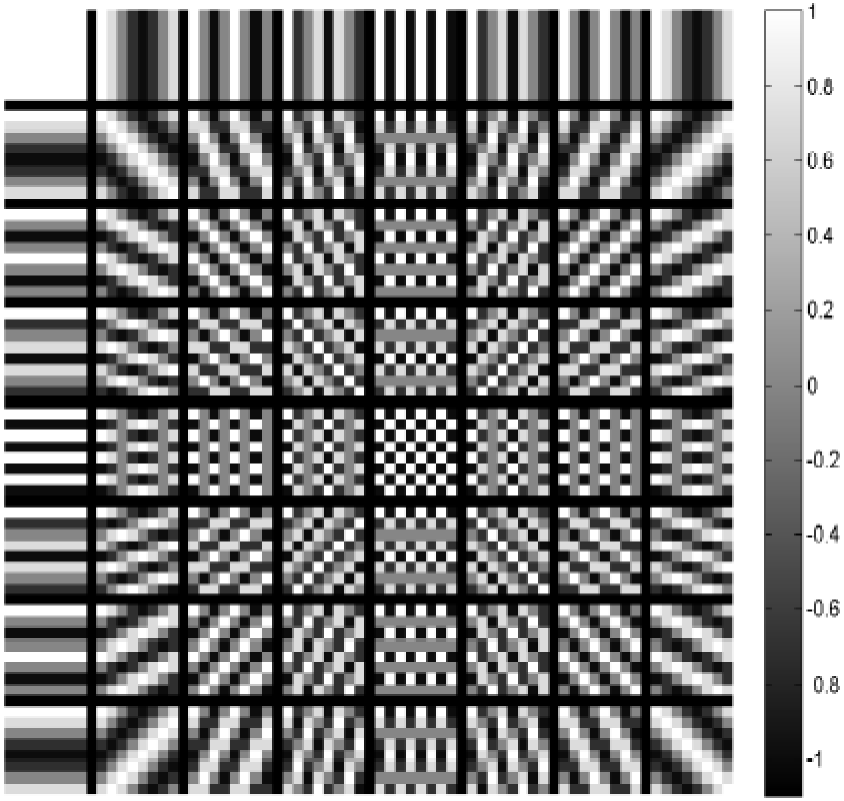


The DCT transforms an 8×8 block of input values to a linear combination of these 64 patterns. The patterns are referred to as the two-dimensional DCT *basis functions*, and the output values are referred to as *transform coefficients*.

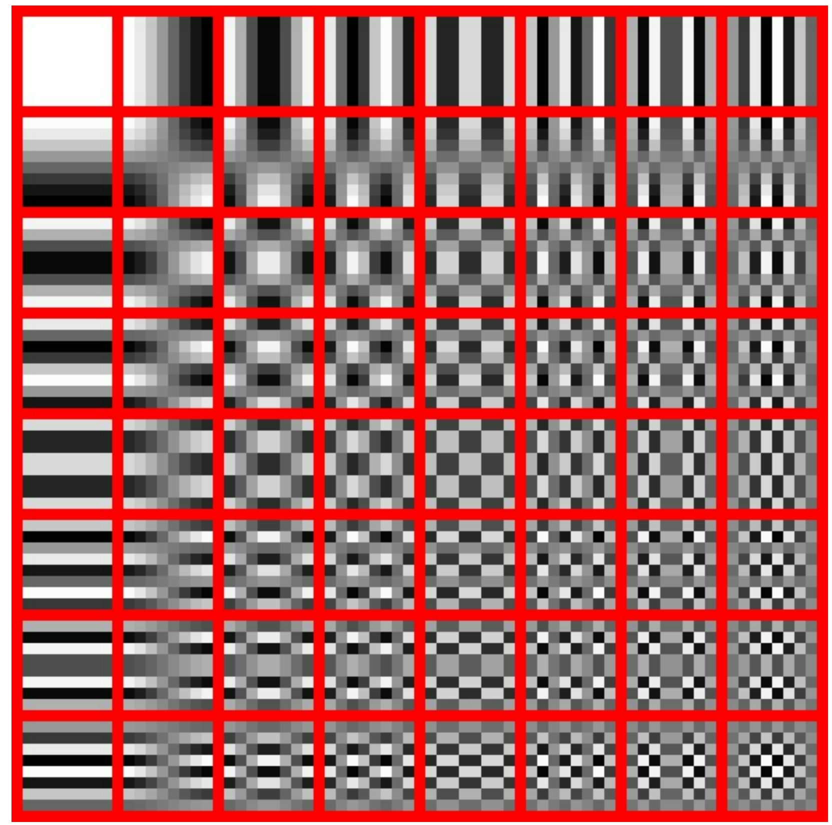
Each image here is obtained from the 8 x 8 outer product of a pair of DCT basis vectors. Each image is stretched between 0 and 255 – on a common scale.

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, 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?

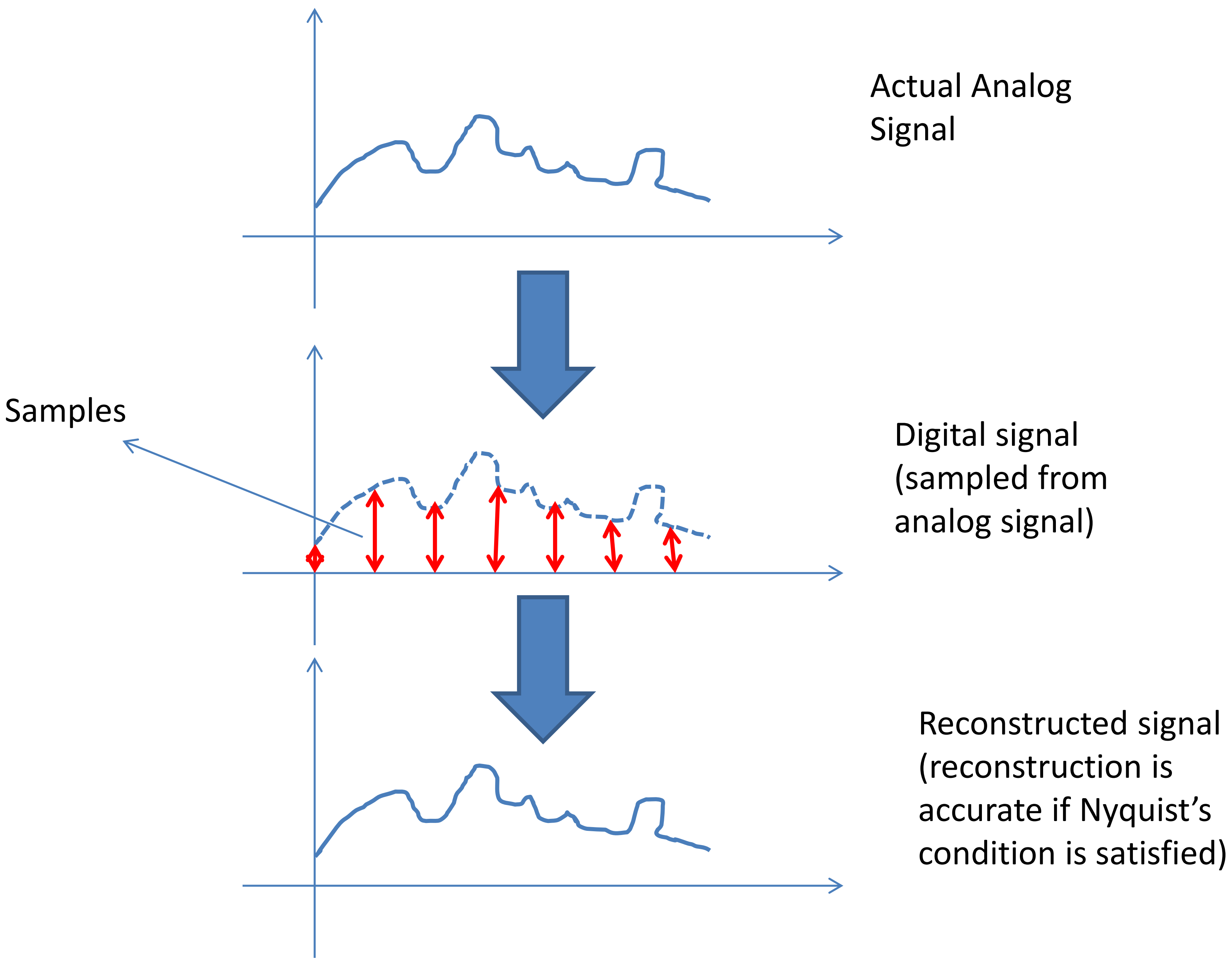
- It has the potential to dramatically improve acquisition speed for MRI, CT, hyper-spectral data and other modalities.

- 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.

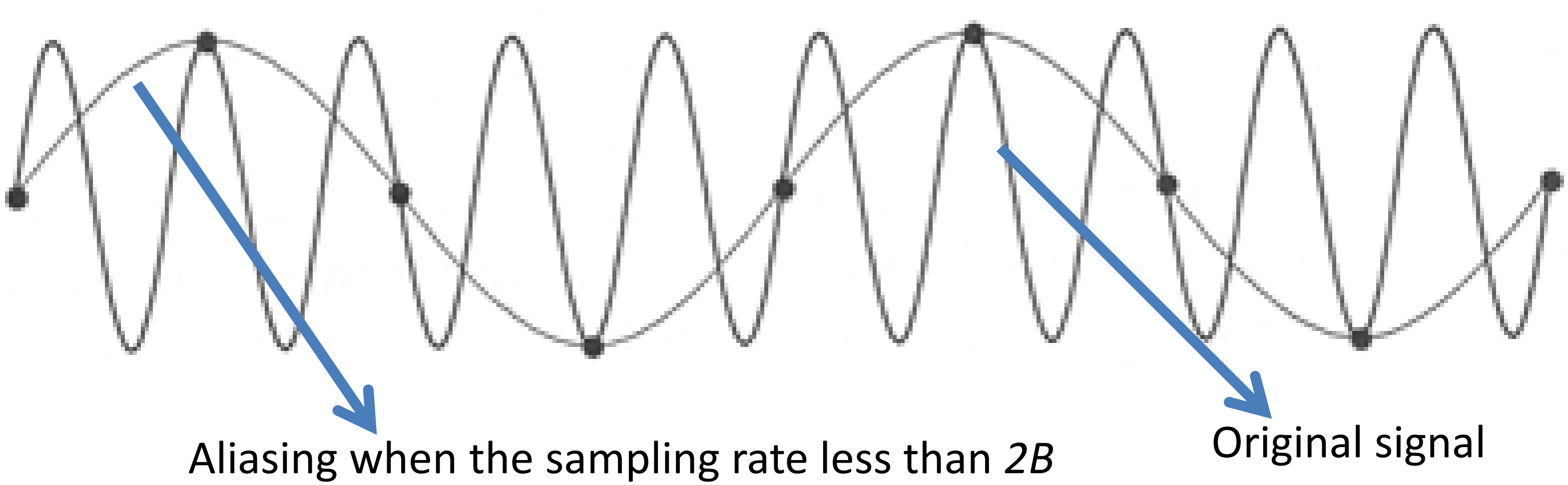

Actual Analog Signal
Samples
Digital signal (sampled from analog signal)
Reconstructed signal (reconstruction is accurate if Nyquist's condition is satisfied)

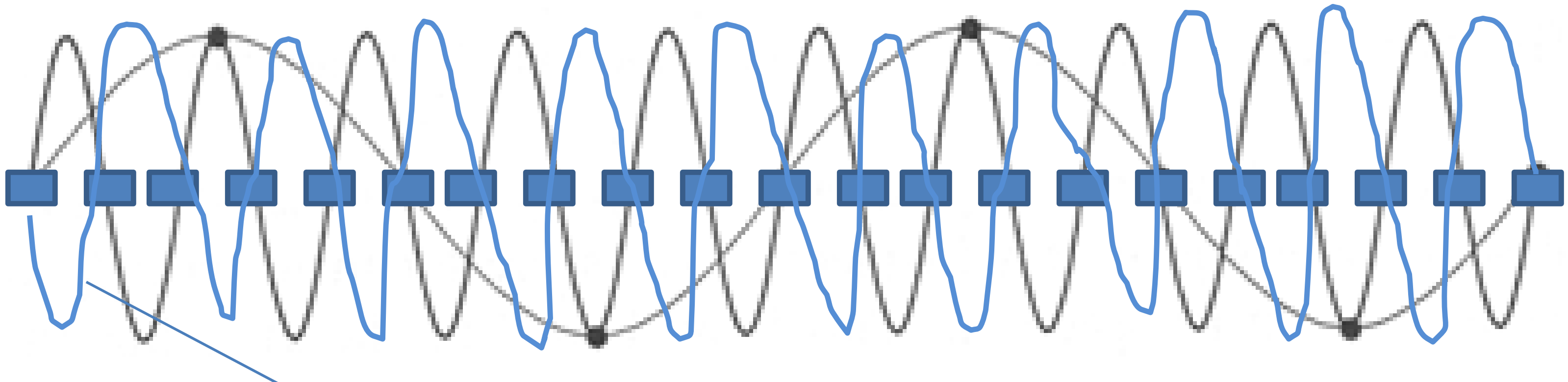
Band-limited signal

- 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).

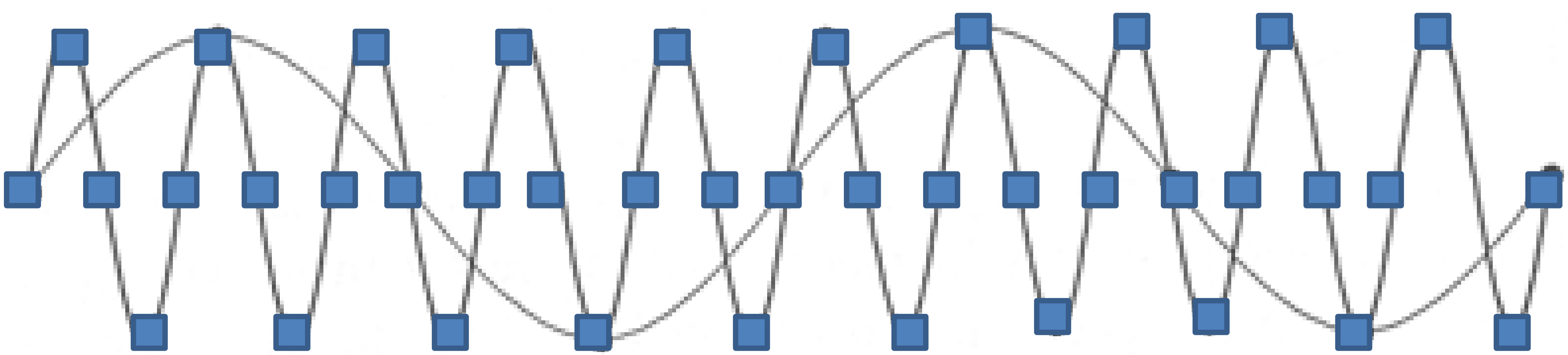
Aliasing

- 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

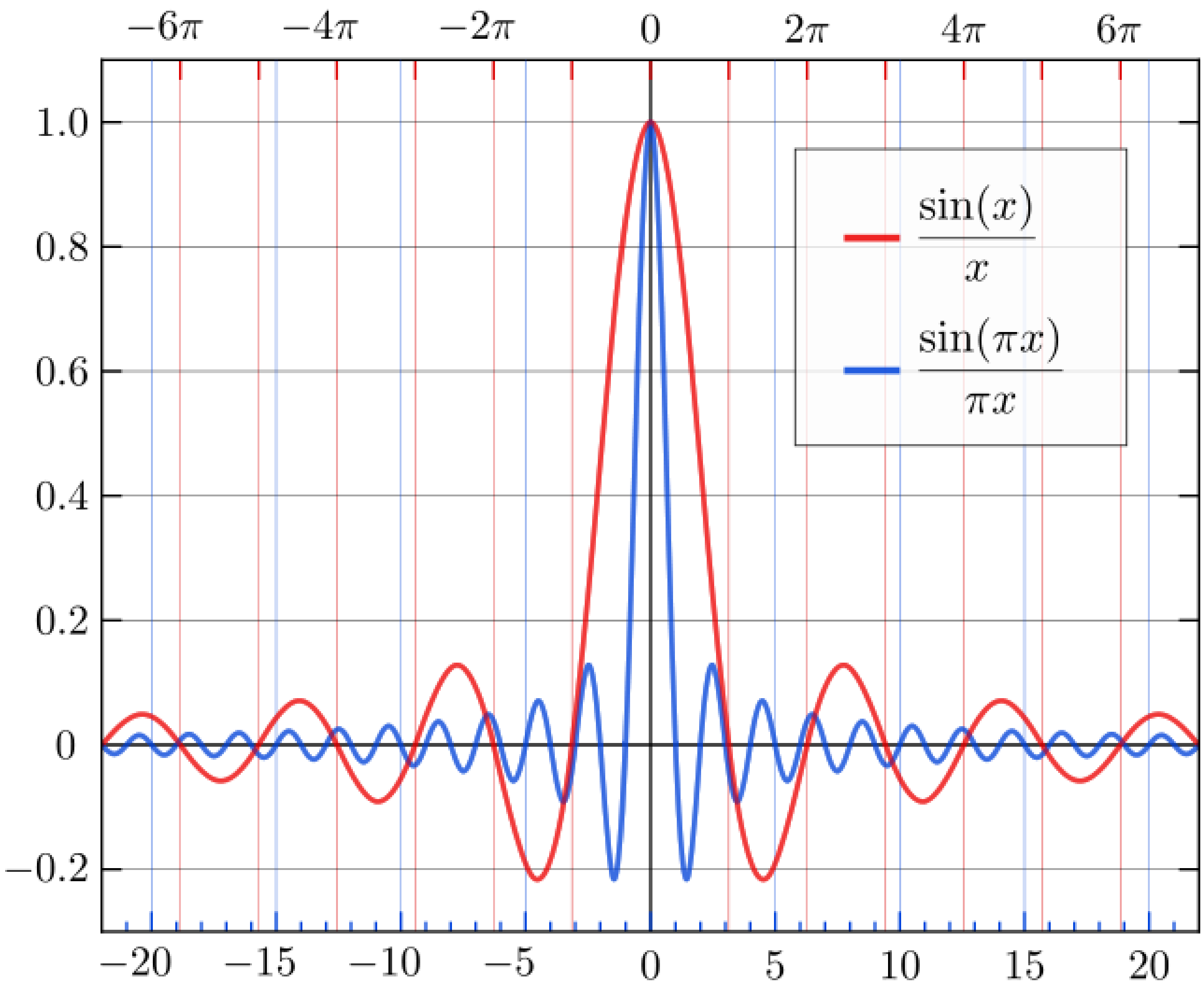
- The optimal reconstruction for band-limited signals from their digital samples proceeds using the **sinc interpolant**. It yields a time-domain (or space-domain signal of infinite extent with values that are very small outside a certain window).
- The associated formula is called Whittaker-Shannon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

Sampling period =1/Sampling rate

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

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).

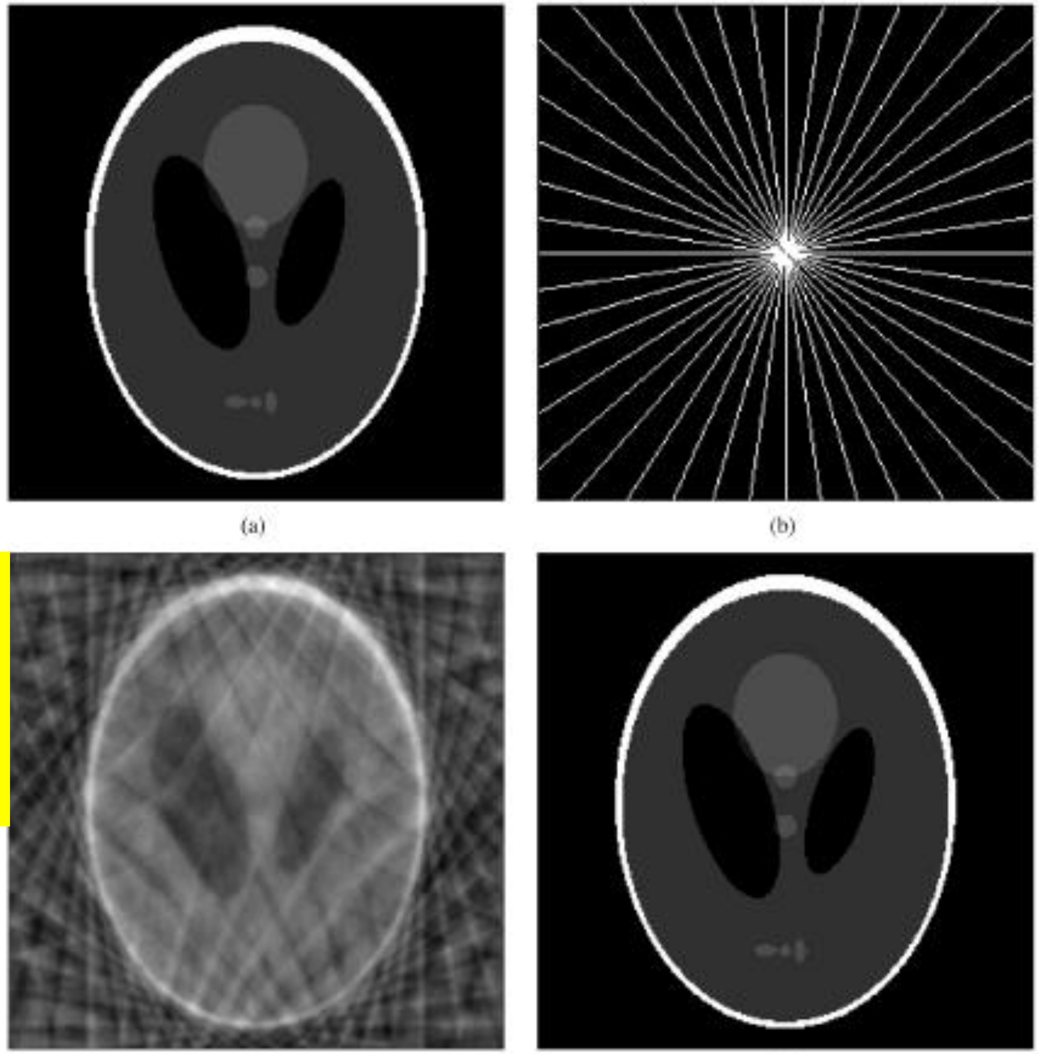
Motivation for CS: Candes' puzzling experiment (Circa 2004)

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

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

such that

$$\forall (u,v)\in \boldsymbol{C}, F(u,v)=G(u,v)$$



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

Fig. 1. Example of a simple recovery problem. (a) The Logan–Shepp phantom test image. (b) Sampling domain Ω in the frequency plane; Fourier coefficients are sampled along 22 approximately radial lines. (c) Minimum energy reconstruction obtained by setting unobserved Fourier coefficients to zero. (d) Reconstruction obtained by minimizing the total variation, as in (1.1). The reconstruction is an exact replica of the image in (a).

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 occurring naturally 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 $\mathbf{y} = \Phi\mathbf{f}$.
- 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 **f** be a vector with *n* elements.

- Since the measurements need to be compressive, $\mathbf{\Phi}$ must have fewer rows (*m*) than columns (*n*) to produce a measurement vector **y** with *m* elements.

- We know **y** and $\mathbf{\Phi}$, and we wish to estimate **f**.

CS: Theory

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

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

CS: Theory

- But CS theory states, that in certain cases, this system does have a unique solution.

- Conditions to be satisfied:

1. Vector **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.

Actually, the representation is compressible, i.e. most of the coefficients are close to zero, but not exactly zero. We will clarify this issue very soon.

Signal Sparsity

- Many signals have sparse* representations in standard orthonormal bases (denoted here as $\mathbf{\Psi}$).
- Example:

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

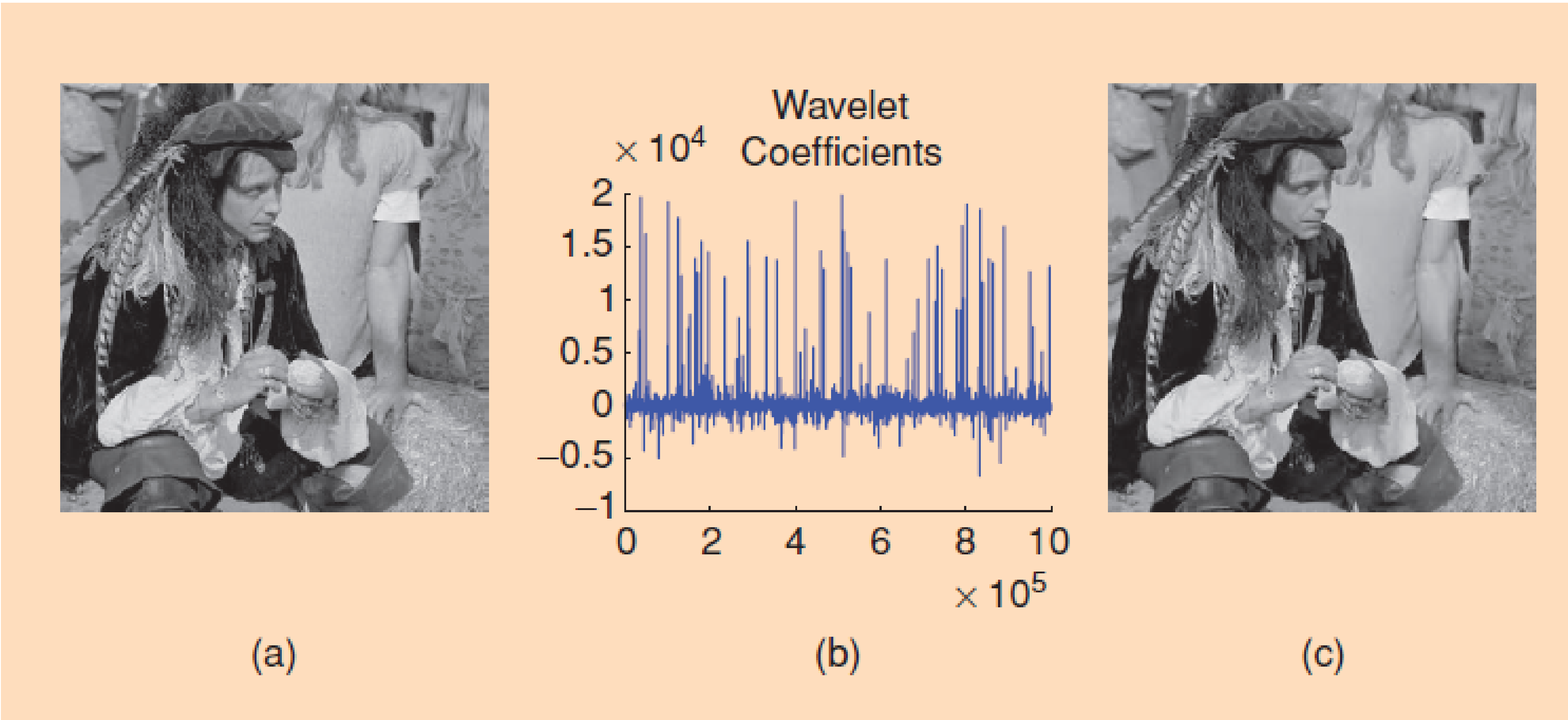
$$\mathbf{f} \in R^n, \boldsymbol{\theta} \in R^n, \mathbf{\Psi} \in R^{n \times n}, \mathbf{\Psi}^T \mathbf{\Psi} = \mathbf{I},$$

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

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

Image source: Candes and Wakin, “An introduction to compressive samplig”, IEEE Signal processing magazine

Signal Sparsity



[FIG1] (a) Original megapixel image with pixel values in the range [0,255] and (b) its wavelet transform coefficients (arranged in random order for enhanced visibility). Relatively few wavelet coefficients capture most of the signal energy; many such images are highly compressible. (c) The reconstruction obtained by zeroing out all the coefficients in the wavelet expansion but the 25,000 largest (pixel values are thresholded to the range [0,255]). The difference with the original picture is hardly noticeable. As we describe in “Undersampling and Sparse Signal Recovery,” this image can be perfectly recovered from just 96,000 incoherent measurements.

Incoherent Sampling

- Consider a sensing matrix Φ of size *m* by *n*, where *m* is much less than *n* (if *m* >= *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 $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is defined as:

$$\mu(\mathbf{\Psi},\mathbf{\Phi}) = \sqrt{n}\max_{\substack{1\le j\le m,\\ 1\le i\le n}} \left| < \mathbf{\Phi}^{(j)}, \mathbf{\Psi}_i > \right|$$

'j'-th row of $\mathbf{\Phi}$, i-th column of $\mathbf{\Psi}$

- 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 $\mathbf{\Psi}$ be the Fourier basis. A sampling basis $\mathbf{\Phi}$ that is incoherent with it is the standard spike (Dirac) basis: $\mathbf{\Phi}_{\mathbf{k}}(t) = \delta(t-k)$ which corresponds to the simplest and most conventional sampling basis in space or time (i.e. $\mathbf{\Phi} =$ 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:

$$\mathbf{y} = \mathbf{\Phi f} = \mathbf{\Phi \Psi \theta},$$

$$\mathbf{y} \in R^m, \mathbf{\Phi} \in R^{m \times n}, \mathbf{f} \in R^n, m << n$$

- The coefficients of the signal **f** in $\mathbf{\Phi}$ – denoted as θ – and hence the signal **f** itself - can be recovered by solving the following constrained minimization problem:

$$\text{Problem P0} : \min \|\boldsymbol{\theta}\|_0 \text{ such that } \mathbf{y} = \mathbf{\Phi \Psi \theta}$$

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

Signal Reconstruction

- P0 is a very difficult optimization problem to solve – it is NP-hard.
- Hence, a softer version (known as Basis Pursuit) is solved:

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

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 $\mathrm{P0}$ 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 2*S* columns of an *m* x *n* matrix **A** are linearly independent. Then any *S*-sparse signal **f** can be uniquely reconstructed from measurements **y** = **Af**. See proof on next slide.

Signal Reconstruction: uniqueness issues

- **Proof by contradiction:**
- Suppose we have S-sparse signals **f** and **f'** such that **y** = **Af** = **Af'**.
- Then **A**(**f**-**f'**) = **0** and hence **f**-**f'** lies in the null-space of **A**.
- But **f**-**f'** is a 2S-sparse signal.
- Hence **f**-**f'** cannot lie in the null-space of **A**, as any 2S-columns of **A** are linearly independent. Hence contradiction!

Signal Reconstruction: uniqueness issues

- But are these conditions on **A** valid?
- The answer is that we need measurement matrices such that **A** will obey this property.
- In fact, we will see something called the “Restricted isometry property” which **A** needs to satisfy to guarantee uniqueness.

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

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

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{f} = \boldsymbol{\Phi\Psi\theta},$$

$$\mathbf{y} \in R^m, \boldsymbol{\Phi} \in R^{m \times n}, \mathbf{f} \in R^n, m << n$$

Comments on Theorem 1

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

Comments on Theorem 1

- The algorithm to reconstruct ***f*** given ***y*** makes **<u>NO ASSUMPTIONS</u>** 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 L1 norm (and not L0 norm), we are still getting a reconstruction that is the sparsest (besides being accurate).

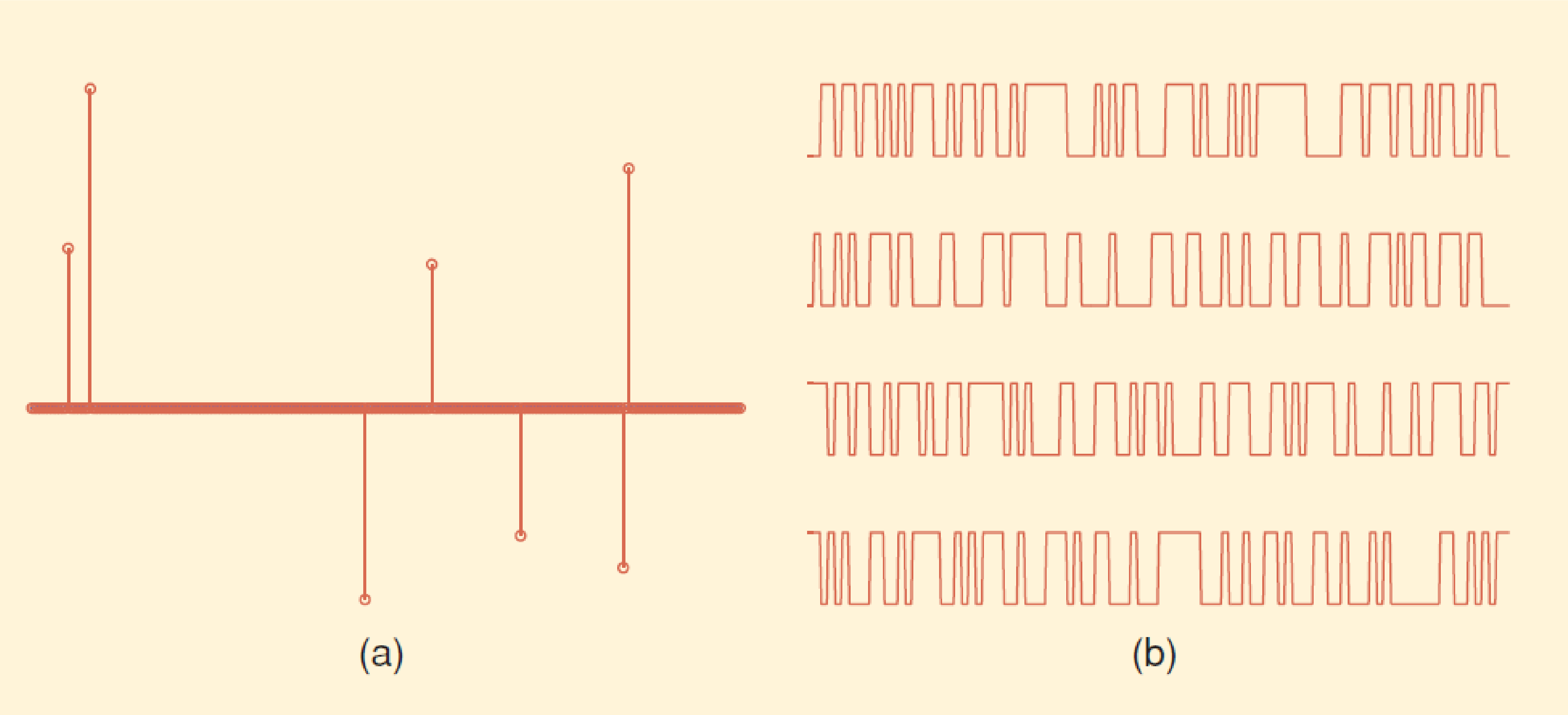
- Although L1-norm has been used in sparse recovery for a long time, the optimality proof was established only in 2005/2006 by Donoho, and by Candes, Romberg and Tao.

Comments on Theorem 1: A New Sampling Theorem?

- Let the signal basis $\mathbf{\Psi}$ be the Fourier basis. A sampling basis $\mathbf{\Phi}$ that is incoherent with it is the standard spike (Dirac) basis:
$$\mathbf{\Phi}_{\mathrm{k}}(t) = \delta(t-k)$$
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* Fourier coefficients of **f**, where
$$m \geq C \log(n/\delta) \|\boldsymbol{\theta}\|_0 \mu^2(\mathbf{\Psi}, \mathbf{\Phi}), \text{ i.e. } m \geq C \log(n/\delta) \|\boldsymbol{\theta}\|_0$$
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.

Intuition behind incoherence



[FIG1] Sampling a sparse vector. (a) An example of a very sparse vector. If we sample this vector directly with no knowledge of which components are active, we will see nothing most of the time. (b) Examples of pseudorandom, incoherent test vectors ϕ_k. With each inner product of a test vector from (b), we pick up a little bit of information about (a).

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

Intuition behind incoherence

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

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

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

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!

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

$$y_i = \left(\sum_{l=1}^{n} \mathbf{\Psi}_{\boldsymbol{l}} \alpha_l\right)^t \left(\sum_{k=1}^{n} \mathbf{\Psi}_{\mathbf{k}} \theta_k\right) = \sum_{k,l} \alpha_l \theta_k \mathbf{\Psi}_{\boldsymbol{l}}^{\mathbf{t}} \mathbf{\Psi}_{\mathbf{k}} = \sum_{k} \alpha_k \theta_k$$

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.

(*) Without loss of generality, we will assume that the rows of **A** have unit magnitude

New concept: Restricted Isometry Property (RIP)

- For integer S = 1,2,…,n, the restricted isometry constant (RIC) δ_S of a matrix **A** = $\mathbf{\Phi\Psi}$ (*) of size m by n is the smallest number such that for any S-sparse vector $\boldsymbol{\theta}$, i.e.

$$(1-\delta_S)\|\boldsymbol{\theta}\|^2 \le \|\mathbf{A}\boldsymbol{\theta}\|^2 \le (1+\delta_s)\|\boldsymbol{\theta}\|^2,$$

- We say that **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 **A** obeys RIP of order S, no S-sparse signal can lie in the null-space of **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 **x** that is *S*-sparse in some orthonormal basis in which it has representation **θ**, i.e. $\|\boldsymbol{\theta}\|_0 = S << n.$
- Then the following is *undesirable* for **A**

$$\mathbf{A}\boldsymbol{\theta}^{(1)} = \mathbf{A}\boldsymbol{\theta}^{(2)}, \text{for } \boldsymbol{\theta}^{(1)} \neq \boldsymbol{\theta}^{(2)}$$

- One way to ensure that this doesn't happen is to design **A** such that:

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

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

Restricted Isometry Property (RIP)

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

$$(1-\delta_{2S})\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^2 \le \|\mathbf{A}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\|^2 \le (1+\delta_{2S})\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^2,$$

$$\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \text{for which } \|\boldsymbol{\theta}_1\|_0 \le S, \|\boldsymbol{\theta}_2\|_0 \le S, \text{ where } \delta_{2S} << 1$$

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

Theorem 2

- Suppose the matrix $\mathbf{A} = \boldsymbol{\Phi\Psi}$ of size m by n (where sensing matrix $\boldsymbol{\Phi}$ has size m by n, and basis matrix $\boldsymbol{\Psi}$ 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 $\mathbf{f} = \boldsymbol{\Psi\theta}$, measurement vector $\mathbf{y} = \boldsymbol{\Phi\Psi\theta}$):

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

Then we have:

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|_2 \le \frac{C_0}{\sqrt{S}} \|\boldsymbol{\theta} - \boldsymbol{\theta}_S\|_1,$$

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|_1 \le C_0 \|\boldsymbol{\theta} - \boldsymbol{\theta}_S\|_1$$

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

Comments on Theorem 2

- This theorem says that the reconstruction for S-sparse signals is **always** exact if **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 of the signal **f**.

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_0 is independent of *n*, and it is a function of just δ_{2S}.

Compressive Sensing under Noise

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

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

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

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

  Then we have:

$\boldsymbol{\theta}_s$ is created by retaining the S largest magnitude elements of $\boldsymbol{\theta}$, 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 **f**, and (2) a term proportional to the noise variance.
- The constants C_0 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 functions of just δ_{2S}.

Randomness is super-cool!☺

- Consider any fixed orthonormal basis Ψ.
- For a sensing matrix Φ constructed in any of the following ways, the matrix **A** = $\Phi\Psi \in \mathcal{R}^{m\mathrm{x}n}$, 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 <u>any</u> orthonormal basis Ψ with very high probability.
- As such, we do not need to tune Φ for a given Ψ.

L_1 norm and L_0 norm

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

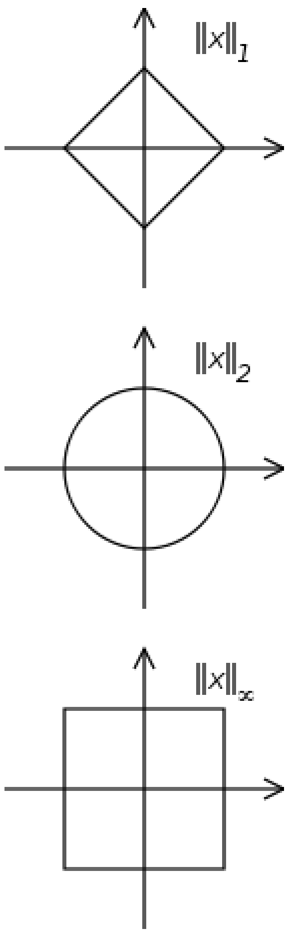
$$\begin{aligned}
&\mathbf{y} = \mathbf{A}\boldsymbol{\theta} \\
&\mathbf{y} \in R^n, \mathbf{A} \in R^{m \times n} \\
&\boldsymbol{\theta} \in R^n (m << n) \\
&\boldsymbol{\theta}^* = \min \|\boldsymbol{\theta}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\boldsymbol{\theta}
\end{aligned}$$

$$\begin{aligned}
&\mathbf{y} = \mathbf{A}\boldsymbol{\theta} \\
&\mathbf{y} \in R^n, \mathbf{A} \in R^{m \times n} \\
&\boldsymbol{\theta} \in R^n (m << n) \\
&\boldsymbol{\theta}^* = \min \|\boldsymbol{\theta}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\boldsymbol{\theta}
\end{aligned}$$

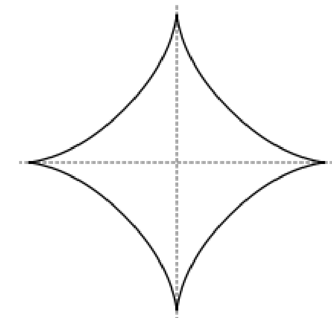
The L_1 norm is a “softer” version of the L_0 norm. Other Lp-norms where $0 < p < 1$ are possible and impose a stronger form of sparsity, but they lead to non-convex problems. Hence L_1 is preferred.

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

L_1 is good, L_2 is bad. Why?



<u>Unit Lp-circle</u> (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 = infinity (square with sides parallel to the axes); **right side**, p = 2/3 (astroid).



$$(|x|^p + |y|^p)^{\frac{1}{p}} = 1$$

In 2-D

$$\left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}} = 1$$

In d-dimensions

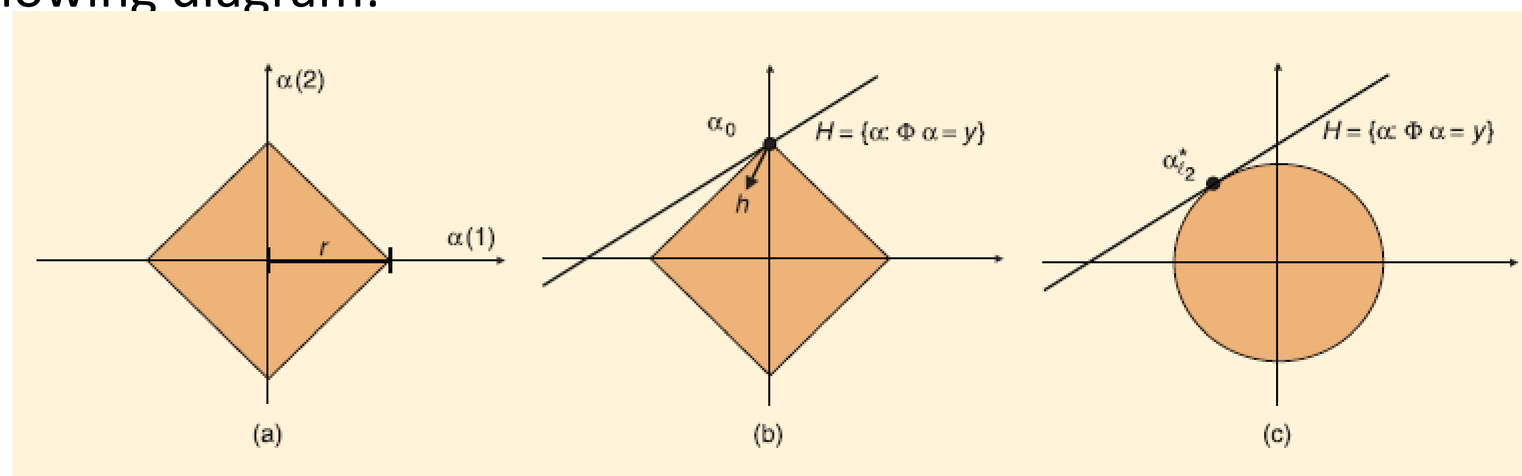
L_1 is good, L_2 is bad. Why?

- Look at the basic problem we are trying to solve:

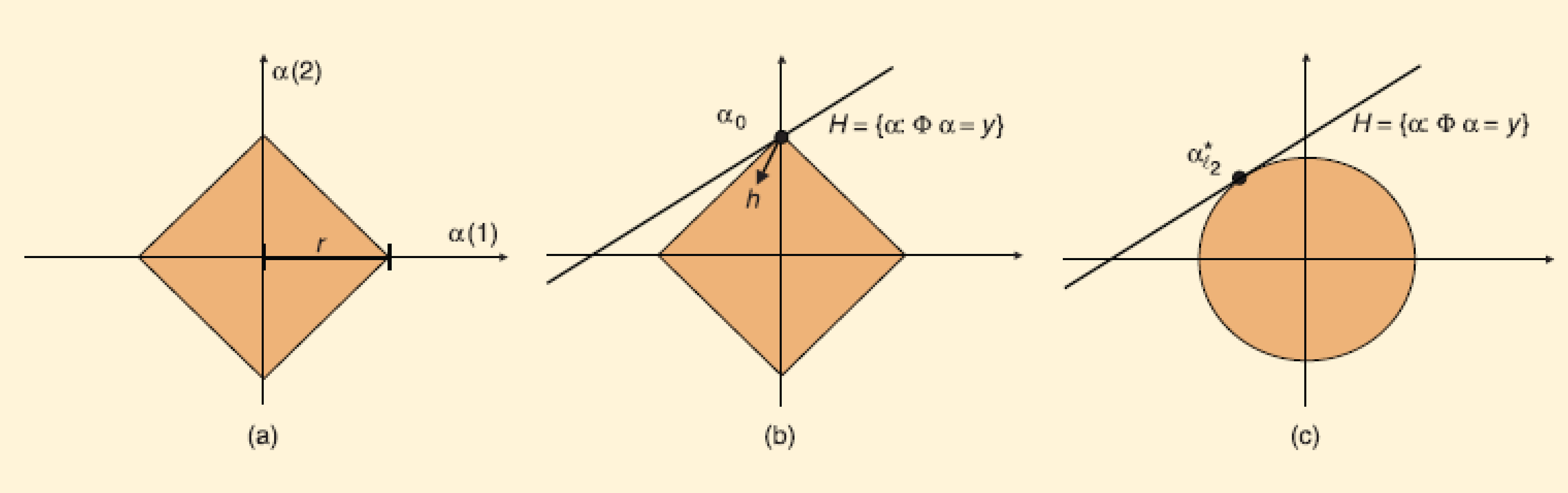
$$\min \|\alpha\|_p \text{ such that } y = \Phi\alpha;\ p \in \{1,2\}$$

- In 2D (i.e. α is a 2x1 vector, y is a scalar, Φ is a 1 x 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



[FIG3] Geometry of ℓ_1 recovery. (a) ℓ_1 ball of radius r; the orange region contains all $\alpha \in \mathbb{R}^2$ such that $|\alpha(1)| + |\alpha(2)| \le r$. (b) Solving the $\min -\ell_1$ problem (8) allows us to recover a sparse α_0 from $y = \Phi\alpha_0$, as the anisotropy of the ℓ_1 ball favors sparse vectors. Note that the descent vectors h pointing into the ℓ_1 ball from α_0 will be concentrated on the support of α_0. (c) Minimizing the ℓ_2 norm does not recover α_0. Since the ℓ_2 ball is isotropic, the $\min -\ell_2$ solution $\alpha^*_{\ell_2}$ will in general not be sparse at all.



[FIG3] **Geometry of ℓ_1 recovery. (a) ℓ_1 ball of radius r; the orange region contains all $\alpha \in \mathbb{R}^2$ such that $|\alpha(1)| + |\alpha(2)| \leq r$. (b) Solving the $\min -\ell_1$ problem (8) allows us to recover a sparse α_0 from $y = \Phi\alpha_0$, as the anisotropy of the ℓ_1 ball favors sparse vectors. Note that the descent vectors h pointing into the ℓ_1 ball from α_0 will be concentrated on the support of α_0. (c) Minimizing the ℓ_2 norm does not recover α_0. Since the ℓ_2 ball is isotropic, the $\min -\ell_2$ solution $\alpha^*_{\ell_2}$ will in general not be sparse at all.**

$$\min \|\alpha\|_p \text{ such that } y = \Phi\alpha$$

- Note that the line **y = Φα** represents a HARD constraint, i.e. any vector **α** must satisfy this constraint.
- We want to find the vector **α** with minimum Lp-norm which satisfies this constraint.
- So we grow Lp-circles of increasing radius (denoted as *r* and defined in the sense of the Lp-norm) starting from the origin, until the Lp-circle touches the line **y = Φα.**
- The very first time the Lp-circle touches the line, gives us the solution (the second time, the Lp-circle will have a greater radius and hence the solution vector **α** will have higher Lp-norm).
- For the L1-norm, the solution vector **α** will be sparse (see middle figure). For the L2-norm, the solution vector will not be sparse (rightmost figure).

L_1 is good, L_2 is bad. Why?

- The L1-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 L1-norm is a softened version of the L0-norm!

Compressive Sensing: Toy Example with images

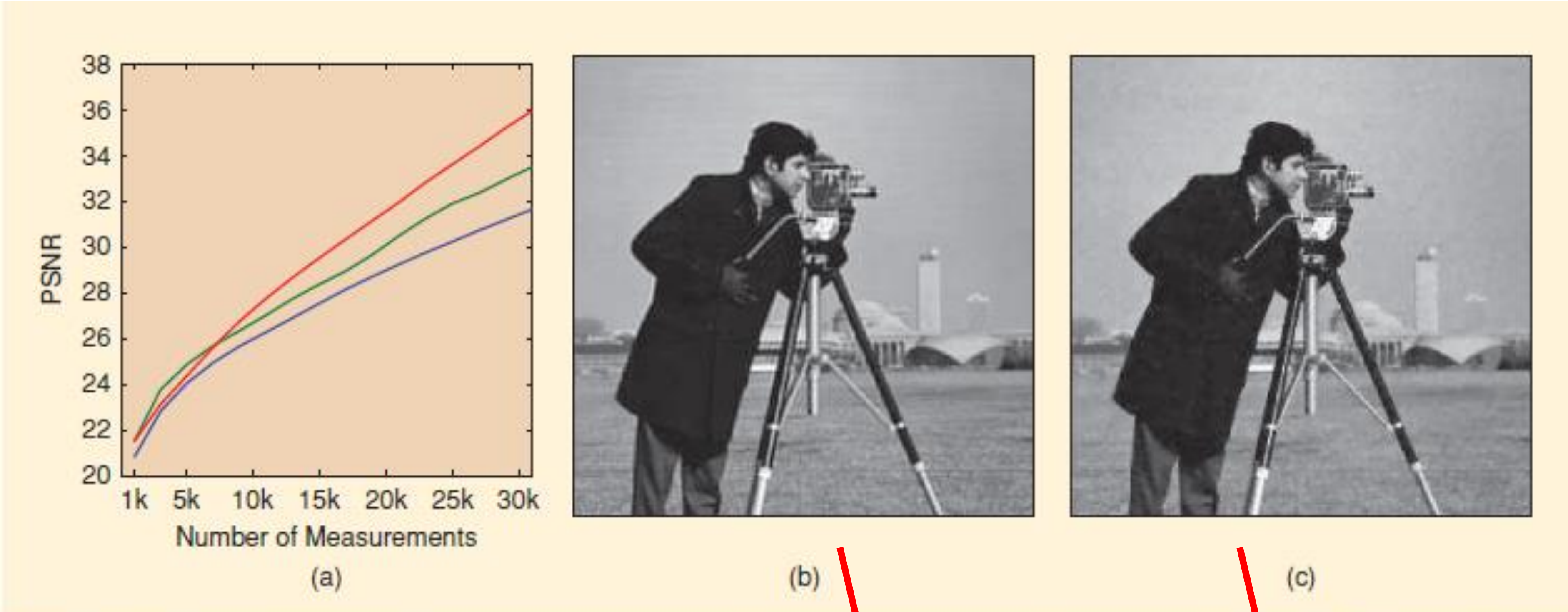
- We will compare image reconstruction by two different methods.
- We won't do true compressive sensing. Instead we'll apply different types of sensing matrices Φ synthetically (in software) to an image loaded into memory, merely for example sake.
- In the first method, we compute the DCT coefficients of an image (size N = 512 x 512), 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 the first 1000 largest DCT coefficients + random linear combinations of all DCT coefficients, yielding a total of *m* coefficients, where *m* is again varied from 1000 to 30000.
- The sensing matrix $\mathbf{\Phi}$ is appropriately assembled.
- In this case, the reconstruction proceeds by solving the following optimization problem:

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

Compressive Sensing: Toy Example with images



[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: $\mathrm{PSNR} = 20\log_{10}(255\cdot 256/\|X-\tilde{X}\|_2)$. (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)

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):

$$\min TV(f) \text{ such that } y = \Phi \Psi^T f,$$
$$TV(f) \leftarrow \iint \sqrt{f_x^2 + f_y^2}\, dxdy$$
$$= \text{L1} - \text{norm of the 2D gradient vector}$$

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.

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

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 $\mathbf{\Phi}$ yielding a vector **y** of only *m* << *n* measurements.

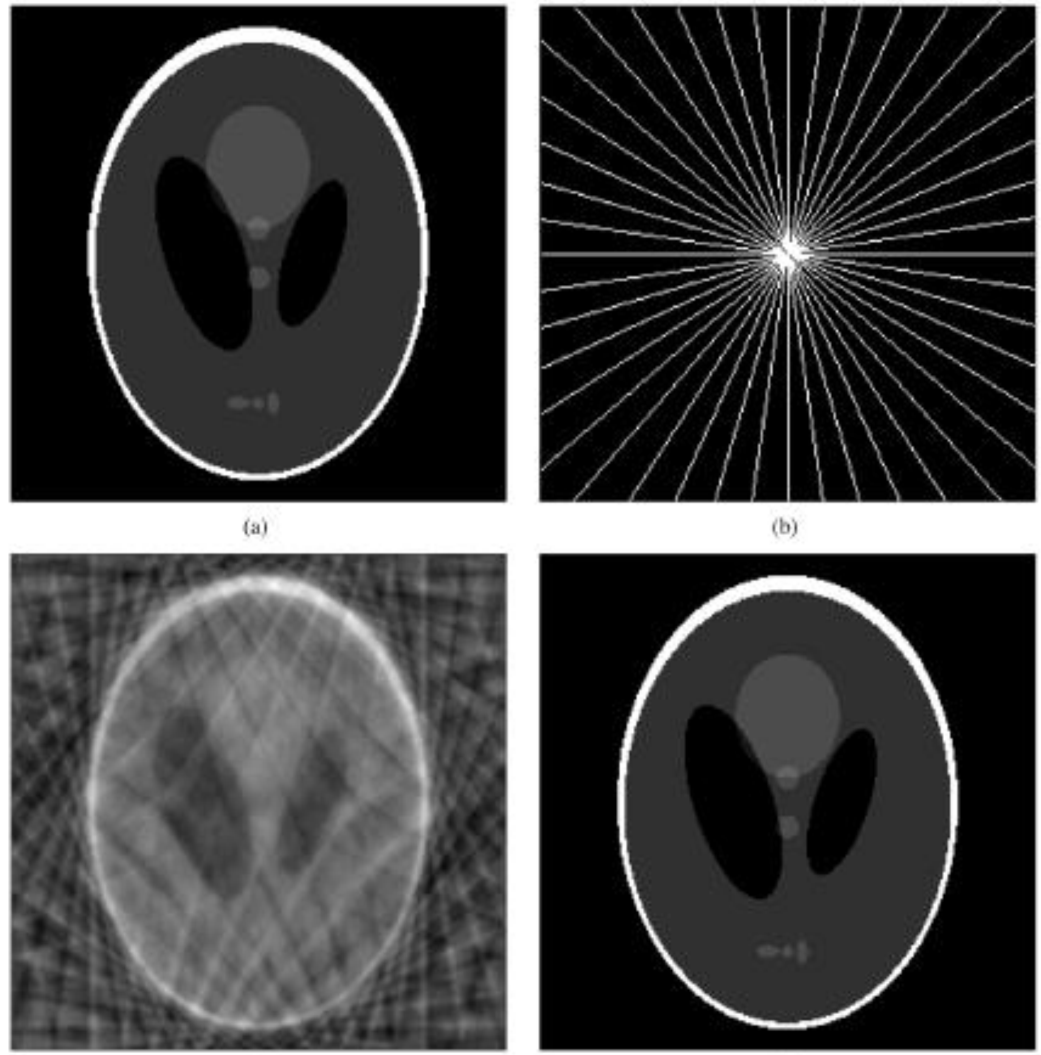
  If $m \geq C \log(n/\delta) \|\theta\|_0 \mu^2(\Psi, \Phi)$ (*C* is some constant) then the solution to the following is **exact** with probability $1-\delta$:

$$\min_f TV(f) \text{ such that } y = \Phi \Psi^T f$$

$$y \in R^m, \Phi \in R^{m \times n}, f \in R^n, m << n$$

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.



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

Fig. 1. Example of a simple recovery problem. (a) The Logan–Shepp phantom test image. (b) Sampling domain Ω in the frequency plane; Fourier coefficients are sampled along 22 approximately radial lines. (c) Minimum energy reconstruction obtained by setting unobserved Fourier coefficients to zero. (d) Reconstruction obtained by minimizing the total variation, as in (1.1). The reconstruction is an exact replica of the image in (a).

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 **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:

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

$$\mathbf{G} = \mathbf{I}_C\mathbf{D}\mathbf{H} = \mathbf{I}_C\mathbf{D}\mathbf{\Psi}^T\mathbf{h} = \mathbf{A}\mathbf{h}$$

where

$\mathbf{\Psi}^T\mathbf{h}$ = Fourier coefficients of $\mathbf{h}$

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

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

Experiment by Candes, Romberg and Tao

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

$$h(i) = \sqrt{(f(x,y) - f(x-1,y))^2 + (f(x,y) - f(x,y-1))^2}$$
$$= |(f(x,y) - f(x-1,y)) + j(f(x,y) - f(x,y-1))|$$

where pixel (x, y) maps on to index i

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:

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

- In the former case, we are measuring the image/signal (**spatial domain**) through $\mathbf{\Phi}$.
- In the second case, we are making measurements in the **frequency domain** (Fourier coefficient magnitudes) through $\mathbf{\Phi}$.

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

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*.
- Consider the following reconstruction criterion:

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

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.
- This was a very early application of L1-norm optimization for signal reconstruction.