Overview of Compressed Sensing

(An Imaging Perspective)

Justin Romberg, Georgia Tech

Image Compression

- Modern imaging applications: 10s of millions to billions of pixels
- Compression algorithms make this huge amount of data manageable
  Compression ratio somewhere around $\mathbf{100 : 1}$
- "Turn huge data set into significantly smaller one without losing much"
- Compressive Sampling *avoids the large data set* altogether

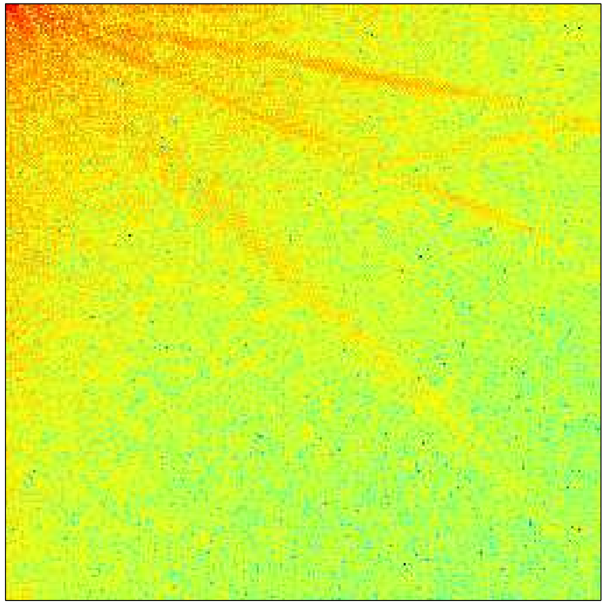
Transform Coding

- *Transform* image into an appropriate domain
  - energy compaction
  - sparsity/compressibility
  - accurate low-order approximations
- *Quantize* the important coefficients
- Effectiveness closely tied to transform

Classical Image Compression (JPEG)

- Discrete Cosine Transform (DCT)
  Basically a real-valued Fourier transform (sinusoids)

- Model: most of the energy is at low frequencies

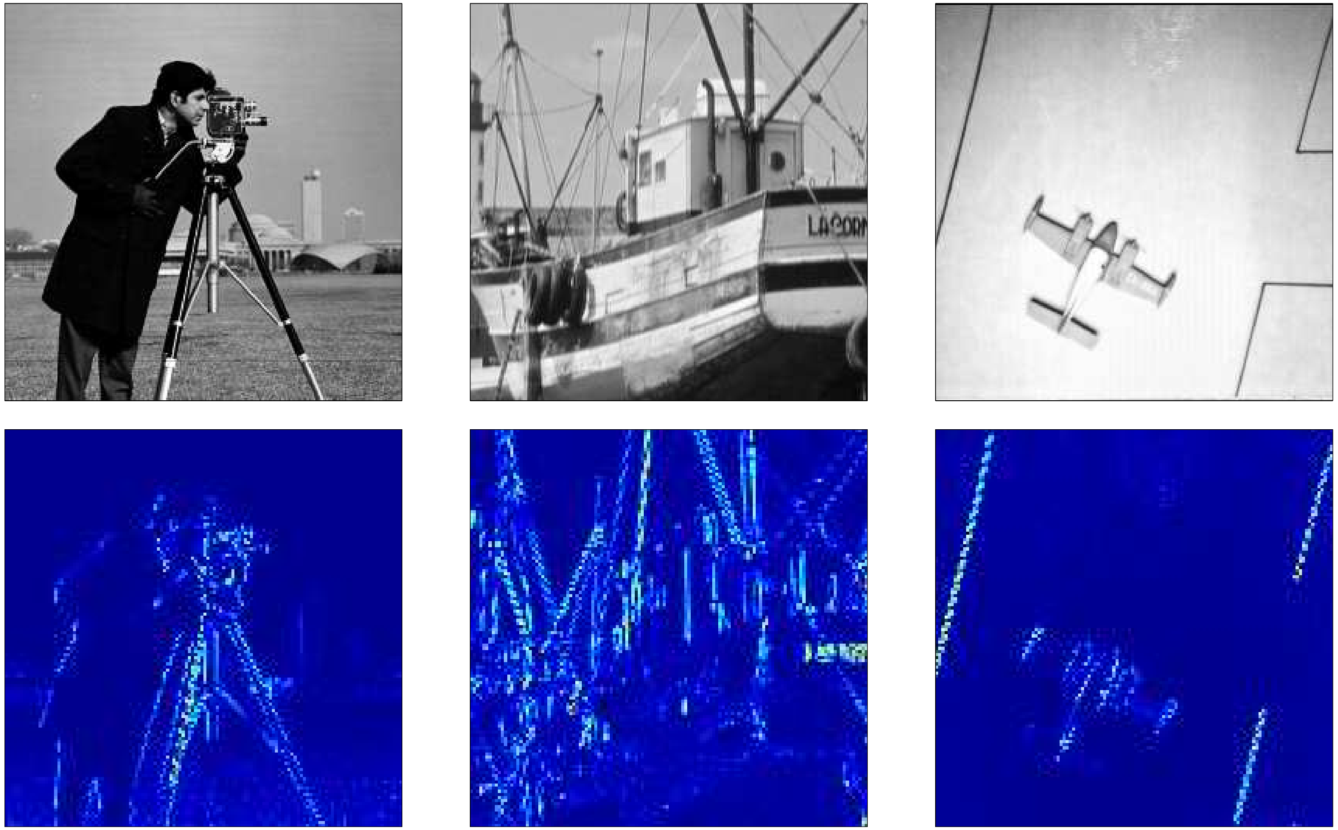


- Use a *fixed* quantization map
  (keep low frequencies, zero-out high-frequencies)

- Performance $\approx$ linear approximation in DCT basis

Modern Image Compression

- Based on the wavelet transform
- *Locations* of important wavelet coefficients vary



- Coder observes which wavelets are significant and keeps them (and codes their location)

Modern Image Compression

- Based on the wavelet transform
- *Locations* of important wavelet coefficients vary
- Coder observes which wavelets are significant and keeps them (and codes their location)
- *Adaptation* to edges is what gives wavelets the advantage
- Performance $\approx$ nonlinear approximation in wavelet basis

Coded Imaging

- Use transform in the acquisition process
- Instead of samples, measure against *codes* or *test vectors*

$$y_1 = \langle f, \phi_1 \rangle, \quad y_2 = \langle f, \phi_2 \rangle, \quad \ldots, y_m = \langle f, \phi_m \rangle$$

$$y = \Phi f$$

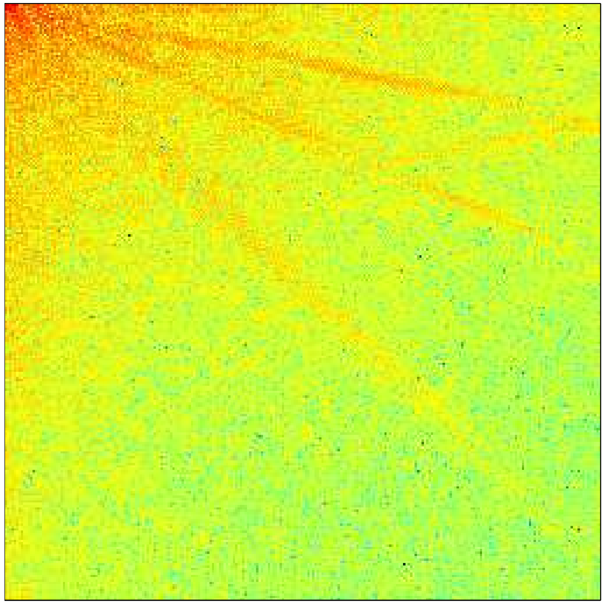
- Examples:
  - MRI, the ϕ_k are sinusoids, measuring Fourier coefficints
  - Tomography, the ϕ_k are ridge functions, measuring line integrals
  - Digital camera, the ϕ_k are small indicator functions, measuring pixels
- Sometimes we are stuck with what we can measure,
  we will broaden are perspective here and see what we *should* measure
- General coded imaging devices are being built
  Single-pixel camera at Rice
  Hyperspectral imager at Yale
  Analog imager at Georgia Tech

DCT Imaging

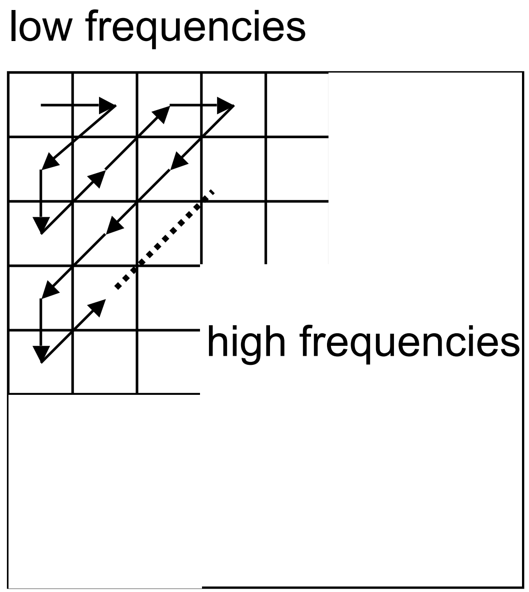
- How should we choose the measurements?
- Idea 1: Match the measurements to image structure
  Inspiration: JPEG
- Most of the energy is at low frequencies, so we start there

image

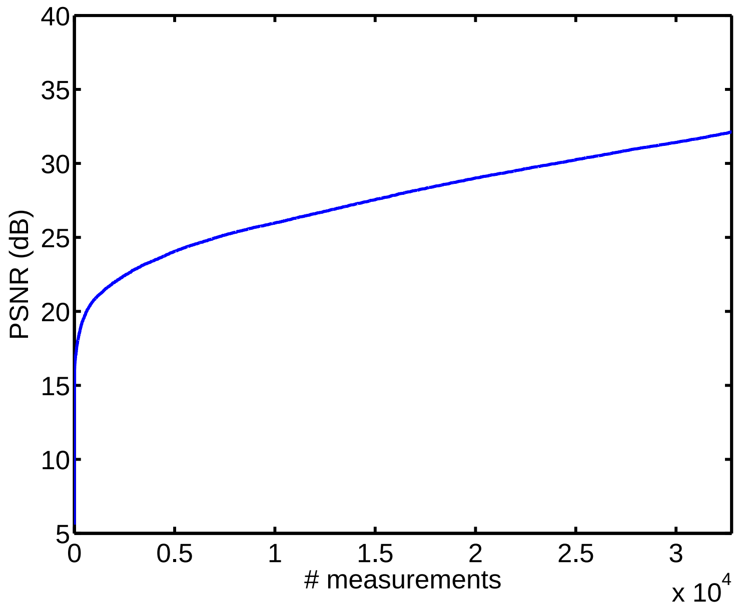
DCT



acquisition order



Stylized Performance

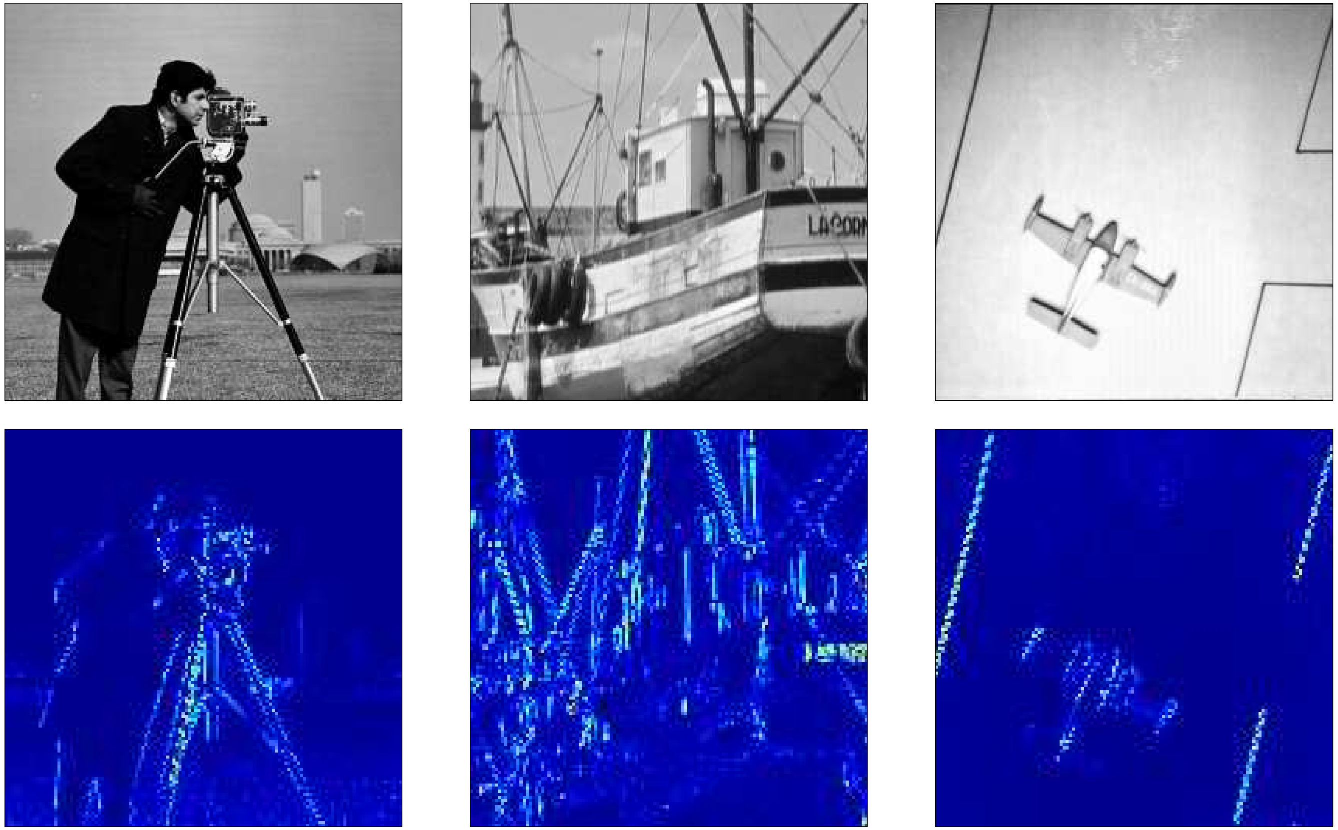


1000 term approx

- We get a “rough sketch” very quickly
- $\approx \mathbf{95\%}$ of the energy in $\mathbf{1000}$ terms
- The details come in much more slowly
  (we are after $\approx \mathbf{99.95\%}$ of the energy)

Wavelet Imaging?

- Want to measure wavelets, but which ones?



The Big Question

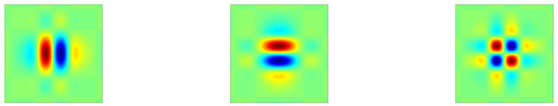
Can we get adaptive approximation performance from a fixed set of measurements?

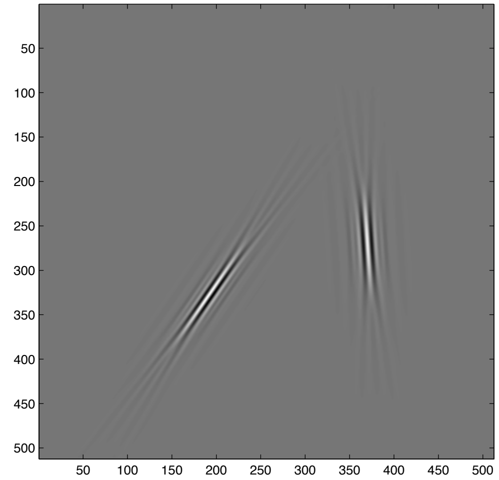
- Surprisingly: yes.
- More surprising: measurements should *not* match image structure at all
- The measurements should look like random noise

Representation vs. Measurements

- Image structure: *local, coherent*

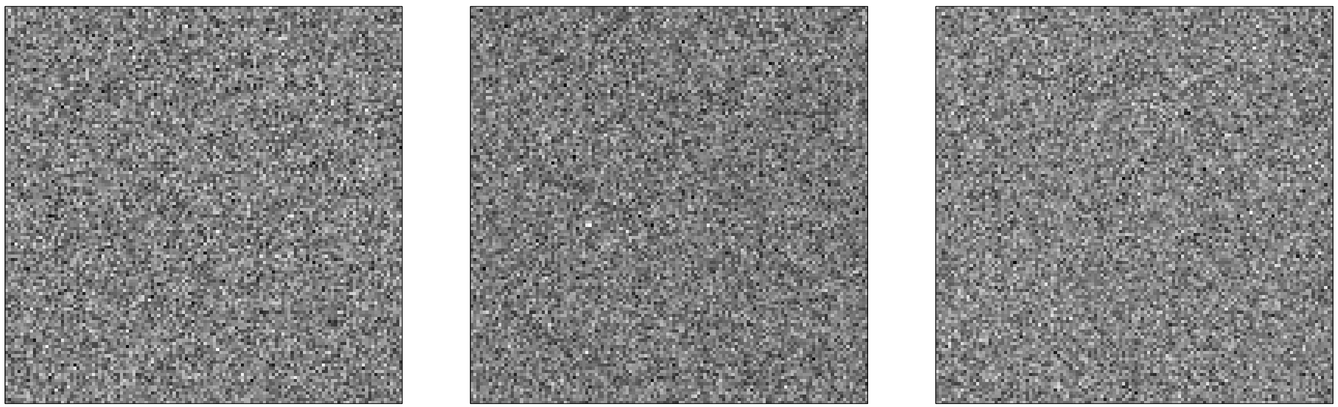
  Good basis functions:



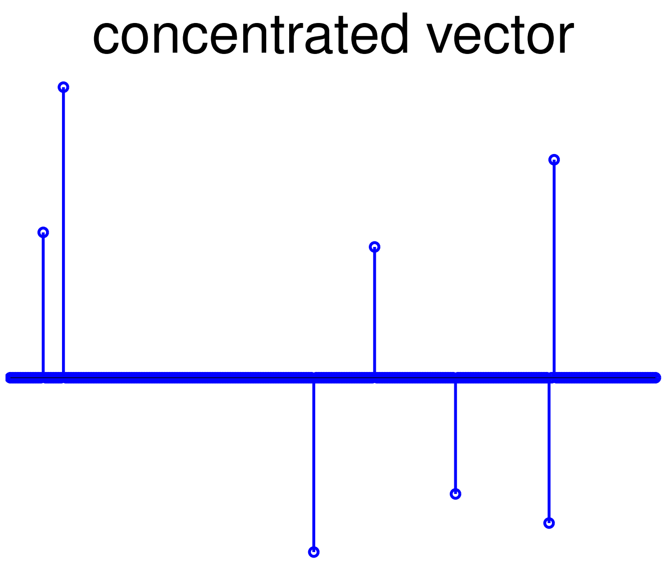


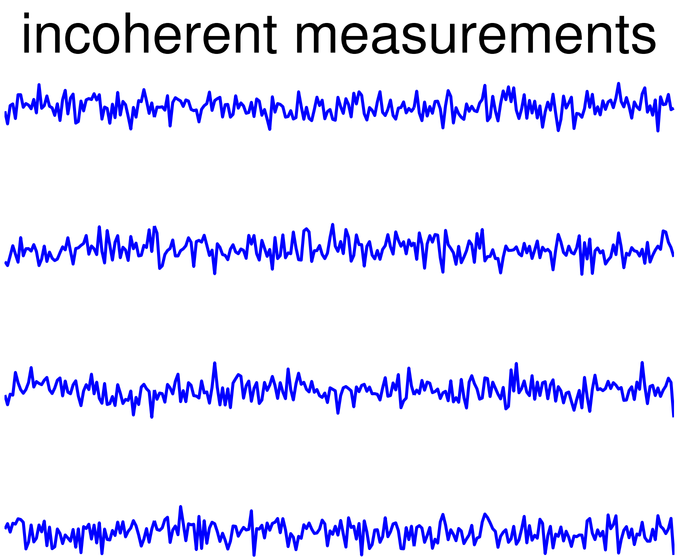
- Measurements: *global, incoherent*

  Good test functions:



Motivation: Sampling a Sparse Vector





- Signal is local, measurements are global
- Each measurement picks up a little information about each component
- Triangulate significant components from measurements
- Formalization: Relies on uncertainty principles between sparsity basis and measurement system

The Uniform Uncertainty Principle

- Φ obeys a UUP for sets of size S if

$$0.8 \cdot \frac{m}{n} \cdot \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq 1.2 \cdot \frac{m}{n} \cdot \|x\|_2^2$$

  for every S-sparse vector x

- Examples: Φ obeys UUP for $S \lesssim m/\log n$ when
  - ϕ_k = random Gaussian
  - ϕ_k = random binary
  - ϕ_k = randomly selected Fourier samples (extra $\log$ factors apply)
- We call these types of measurements *incoherent*

UUP and Sparse Recovery

- UUP $\Rightarrow$ there can be only one sparse explanation for measurements (more or less automatic)
- Say x_0 is S-sparse, and we measure $y = \Phi x_0$
  If we search for the sparsest vector that explains y, we will find x_0:

$$\min_x \#\{t : x(t) \neq 0\} \quad \text{subject to} \quad \Phi x = y$$

- This is nice, but impossible (combinatorial)
- But, we can use the ℓ_1 norm as a *proxy* for sparsity

Sparse Recovery via ℓ_1 Minimization

- Say x_0 is S-sparse, Φ obeys UUP for sets of size $4S$
- Measure $y = \Phi x_0$
- Then solving

$$\min_x \|x\|_{\ell_1} \quad \text{subject to} \quad \Phi x = y$$

  will recover x_0 exactly
- We can recover x_0 from

$$m \gtrsim S \cdot \log n$$

  incoherent measurements by solving a *tractable* program
- *Number of measurements* $\approx$ *number of active components*

Transform Domain Recovery

- Sparsity basis $\boldsymbol{\Psi}$ (e.g. wavelets)
- Reconstruct by solving

$$\min_{\boldsymbol{\alpha}} \ \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{y}$$

- Need measurement to be incoherent in the $\boldsymbol{\Psi}$ domain
  - Random Gaussian: still incoherent (exactly the same)
  - Random binary: still incoherent
  - General rule: just make $\boldsymbol{\Phi}$ unstructured wrt $\boldsymbol{\Psi}$

General Recovery

- Sparsity basis Ψ (orthonormal, $\Psi^*\Psi = I$)
  (signal/image model)
- Measurement basis M (orthogonal, $M^*M = nI$)
  (acquisition system)
- Select measurements by sampling at random in M domain:
  $\Phi = M_\Omega$, Ω = sample index set
- Recover solving

$$\min \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \Phi'\alpha = y, \qquad \Phi' = \Phi\Psi$$

- Exact recovery when

$$m \gtrsim \mu^2(\Phi, \Psi) \cdot S \cdot \log n$$

  for vast majority of signals and sample sets
- Recovery conditions based on different *weak uncertainty principles*

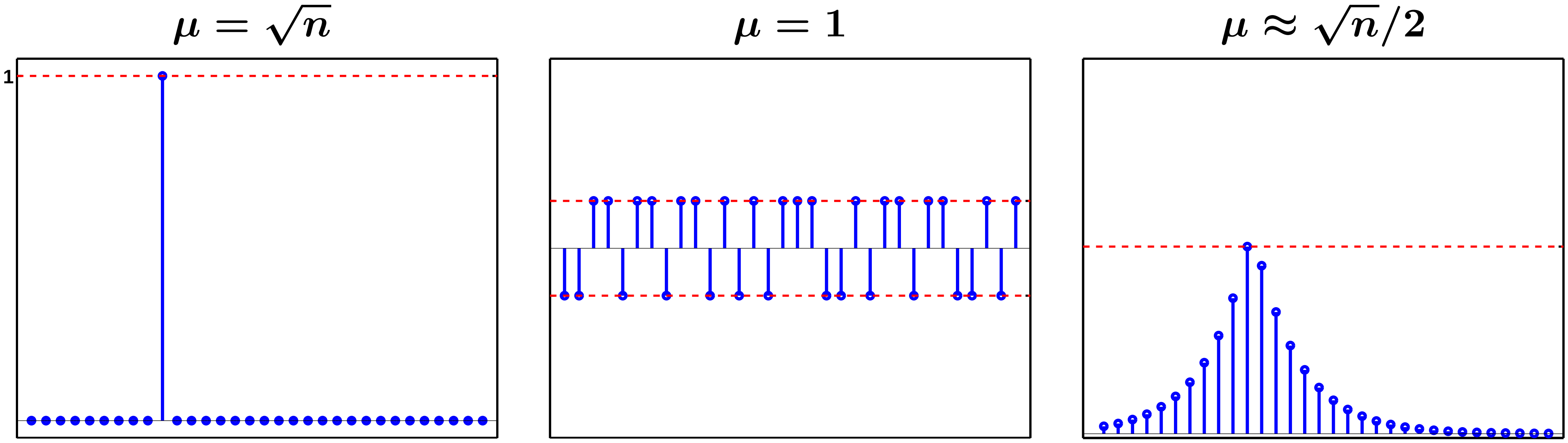
What is $\mu(\Phi, \Psi)$?

$$m \gtrsim \mu^2(\Phi, \Psi) \cdot S \cdot \log n$$

- *Mutual coherence*

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{j,k} |\langle \phi_k, \psi_j \rangle| = \sqrt{n} \cdot \max_{j,k} |(M\Psi)_{j,k}|$$

- Rows of Φ':



- Maximum value in M (not relationship between columns)

Motivation for Structured Measurements

- Properties of system: $\mathbf{\Psi}$ fixed by image model, M fixed by physics
- Algorithmic:
  - solving $\mathbf{min}\ \boldsymbol{\ell_1}$ requires solving (large) systems of equations

$$\boldsymbol{M_\Omega \Psi \Sigma_i \Psi^T M_\Omega^T z = w}$$

  iterative methods need *fast* implicit algorithms for $\mathbf{\Psi}, \boldsymbol{M}$

Fast Measurements

- Say we want to take $\mathbf{20,000}$ measurements of a $\mathbf{512 \times 512}$ image ($\boldsymbol{n = 262,144}$)
- If $\boldsymbol{\Phi}$ is Gaussian, with each entry a float, it would take more than an entire DVD just to hold $\boldsymbol{\Phi}$
- Need fast, implicit, noise-like measurement systems to make recovery feasible
- Noiselet system (Coifman,Geshwind,Meyer ’01)
  - perfectly incoherent with Haar system
  - performs the same as Gaussian (in numerical experiments) for recovering spikes, sparse wavelets, sparse Fourier signals
  - $\boldsymbol{O(n \log n)}$

Large Scale Example

- $n = 1024^2 \approx 10^6$, $m = 100,000$
- Perfectly sparse image (in wavelet domain), $S = 25,000$
- Recovered to 4 digits in 50 iterations
  (5 digits in 52 iterations, 6 digits in 54 iterations,$\ldots$)
- Recovery time ≈ 30 minutes on a desktop (Matlab code)
- ≈ 5000–$10,000$ applications of Φ'

Stability

- Real images are not exactly sparse
- For Φ' obeying UUP for sets of size $4S$, and *general* α, recovery obeys

$$\|\alpha_0 - \alpha^*\|_2 \lesssim \frac{\|\alpha_0 - \alpha_{0,S}\|}{\sqrt{S}}$$

  $\alpha_{0,S}$ = best S-term approximation

- Compressible: if transform coefficients decay

$$|\alpha|_{0(k)} \lesssim k^{-r}, \quad r > 1$$

  $|\alpha|_{0(k)} = k$th largest coefficient, then

$$\begin{aligned} \|\alpha_0 - \alpha_{0,S}\|_2 &\lesssim S^{-r+1/2} \\ \|\alpha_0 - \alpha^*\|_2 &\lesssim S^{-r+1/2} \end{aligned}$$

- *Recovery error $\sim$ adaptive approximation error*

Stability

- What if the measurements are noisy?

$$y = \Phi' \alpha_0 + e, \qquad \|e\|_2 \le \epsilon$$

- *Relax* the recovery program; solve

$$\min_{\alpha} \ \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \|\Phi'\alpha - y\|_2 \le \epsilon$$

- The recovery error obeys

$$\|\alpha_0 - \alpha^*\|_2 \quad \lesssim \quad \sqrt{\frac{n}{m}} \cdot \epsilon \ + \ \frac{\|\alpha_0 - \alpha_{0,S}\|_{\ell_1}}{\sqrt{S}}$$

measurement error $+$ *approximation error*

Total-variation Recovery

- Recover by solving

$$\min \ \mathbf{TV}(x) \quad \text{such that} \quad \|\Phi x - y\|_2 \le \epsilon$$

where

$$\mathbf{TV}(x) = \sum_{i,j} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}$$

(sum of the magnitude of the gradient at each pixel)

- Alternate image model, popularized by Rudin-Osher-Fatemi (1992)
- Find an image with *sparse gradient* that matches observations
- Second-order cone program ($\approx$ linear program)
- Exact recovery from clean measurements on piecewise-constant images

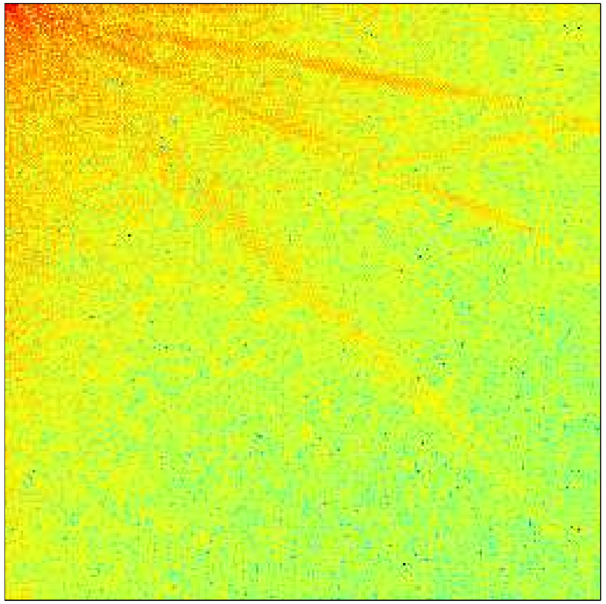
Back to our imaging simulation . . .

DCT Imaging

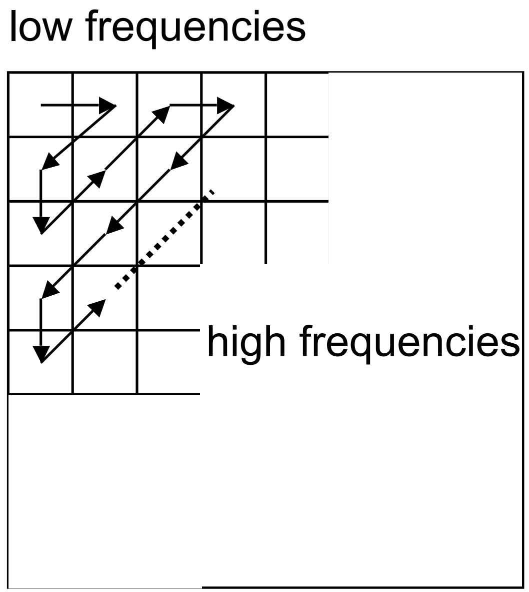
- How should we choose the measurements?
- Idea 1: Match the measurements to image structure
  Inspiration: JPEG
- Most of the energy is at low frequencies, so we start there

image

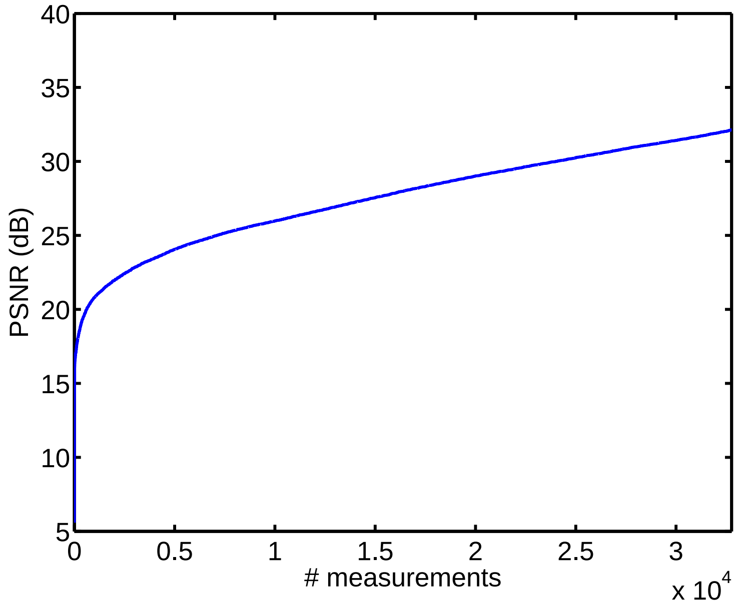
DCT



acquisition order



Stylized Performance



1000 term approx

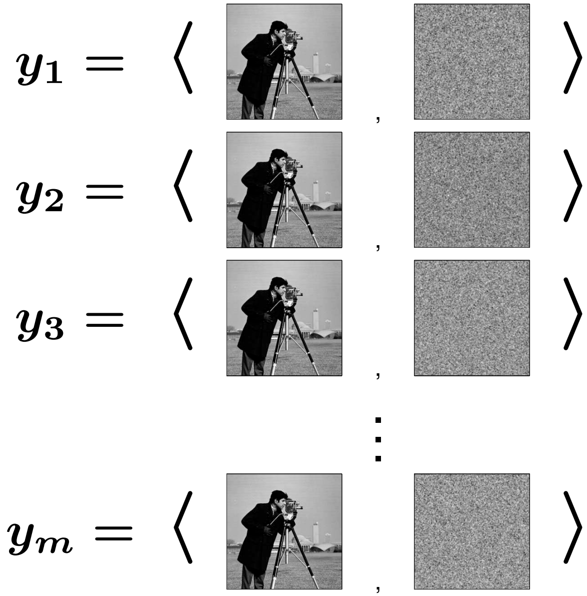
- We get a "rough sketch" very quickly
- $\approx \mathbf{95\%}$ of the energy in $\mathbf{1000}$ terms
- The details come in much more slowly
  (we are after $\approx \mathbf{99.95\%}$ of the energy)

Compressive Imaging

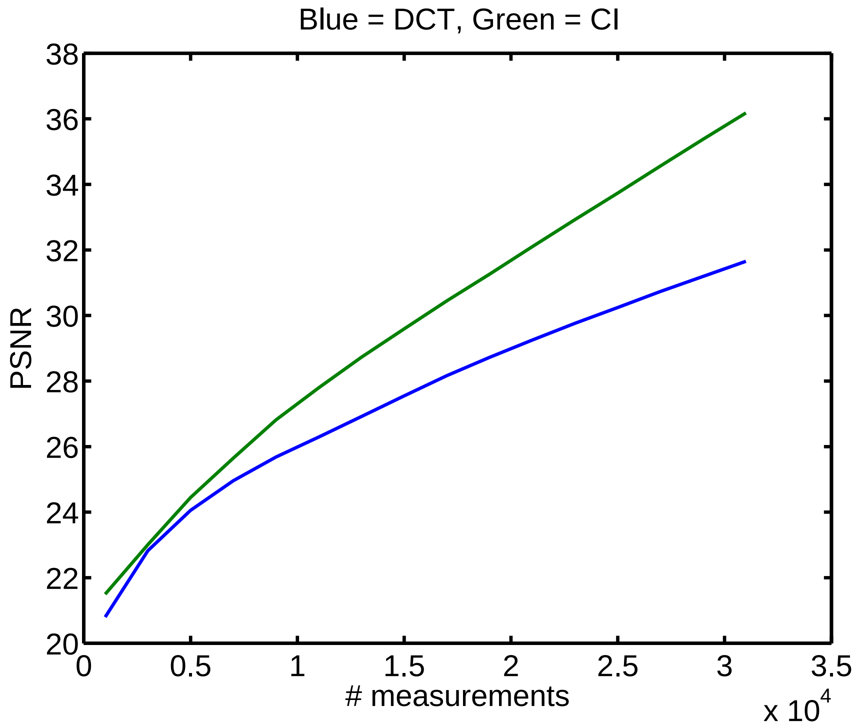
- Idea 2: Take first 1000 DCT coefficients,
  then switch to random measurements, recover using TV minimization

image

measurements



Compressive Imaging: Stylized Performance



- Not only is Compressive Imaging better, it is getting better faster!
- CI is much better at “filling in the details”

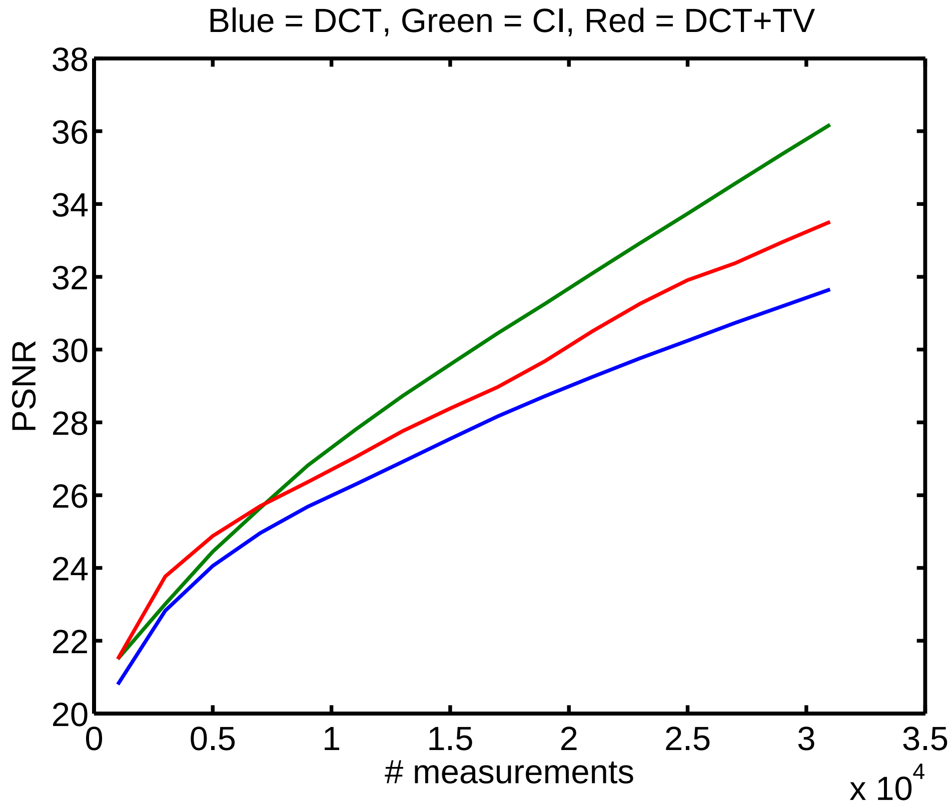
Compressive Imaging: Example

linear, 21k meas

CI, 21k meas

Compressive Imaging: Stylized Performance

- Is the performance gain just coming from the modeling?
- Take low frequency DCT measurement, recover using TV



- Random measurements are really helping!

Examples of Immediate Applications

- Fast MRI
- Fast Tomography
- Low-power Sensing
- High-speed analog-to-digital conversion (ADC)