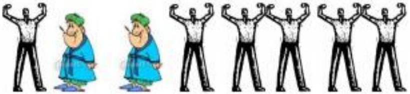
Sparsity-aware sampling theorems and applications

Rachel Ward
University of Texas at Austin

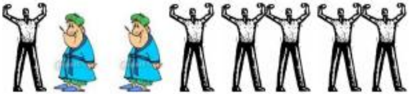
November, 2014

Sparsity-aware sampling: motivating example



Problem: $N = 100,000$ soldiers should be screened for syphilis.
Syphilis is rare (only about $s = 10$ expected out of $100,000$).
Doing a blood test is expensive. Do we need to take N blood tests?

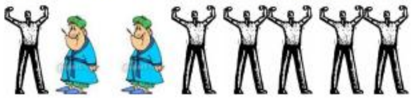
Sparsity-aware sampling: motivating example



Problem: $N = 100,000$ soldiers should be screened for syphilis. Syphilis is rare (only about $s = 10$ expected out of $100,000$). Doing a blood test is expensive. Do we need to take N blood tests?

Idea: Pool blood together. Test a combined blood sample to check if at least one soldier has syphilis.

Sparsity-aware sampling: motivating example



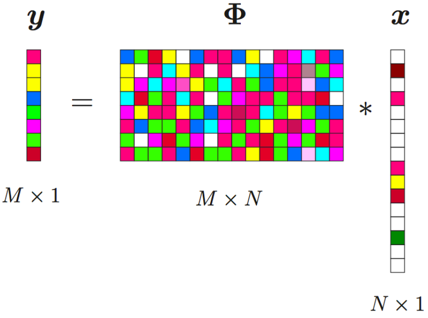
Problem: $N = 100,000$ soldiers should be screened for syphilis. Syphilis is rare (only about $s = 10$ expected out of $100,000$). Doing a blood test is expensive. Do we need to take N blood tests?

Idea: Pool blood together. Test a combined blood sample to check if at least one soldier has syphilis.

Only only need take $s \log N \ll N$ blood tests to identify infected soldiers. ("compressed" measurements).

Implemented by the U.S. Government during WWII

Compressive sensing



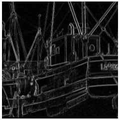
Main idea: Many natural signals / images of interest are *sparse* in some sense.

We say $\mathbf{x}$ is s-sparse if $\|\mathbf{x}\|_0 = \#\{j : |x_j| > 0\} \leq s$.

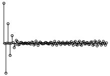
Theory: from only $m \approx s \log(N)$ *incoherent* linear measurements, can recover sparse signal as e.g. vector of minimal ℓ_1-norm satisfying $\mathbf{y} = \Phi \mathbf{x}$

Examples of sparsity:

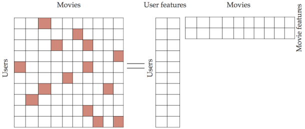
Natural images:



Smooth function interpolation



Low-rank matrices:



Incoherent sampling

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

Let (Φ, Ψ) is a pair of orthonormal bases of $\mathbb{R}^N$.

1. $\Phi = (\phi_j)$ is used for sensing: $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a subset of m rows of Φ
2. $\Psi = (\psi_k)$ is used to sparsely represent $\mathbf{x}$: $\mathbf{x} = \mathbf{\Psi}^* \mathbf{b}$, and $\mathbf{b}$ is assumed sparse

Definition

The coherence between Φ and Ψ is

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \le k,j \le N} | < \phi_j, \psi_k > |$$

Incoherent sampling

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

Let (Φ, Ψ) is a pair of orthonormal bases of $\mathbb{R}^N$.

1. $\Phi = (\phi_j)$ is used for sensing: $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a subset of m rows of Φ
2. $\Psi = (\psi_k)$ is used to sparsely represent $\mathbf{x}$: $\mathbf{x} = \mathbf{\Psi}^*\mathbf{b}$, and $\mathbf{b}$ is assumed sparse

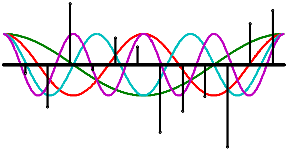
Definition

The coherence between Φ and Ψ is

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \leq k,j \leq N} | < \phi_j, \psi_k > |$$

If $\mu(\Phi, \Psi) = C$ a constant, then Φ and Ψ are called *incoherent*.

Incoherent sampling

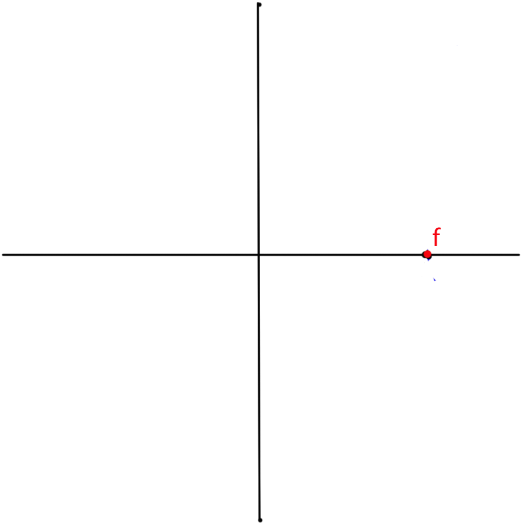


Example:

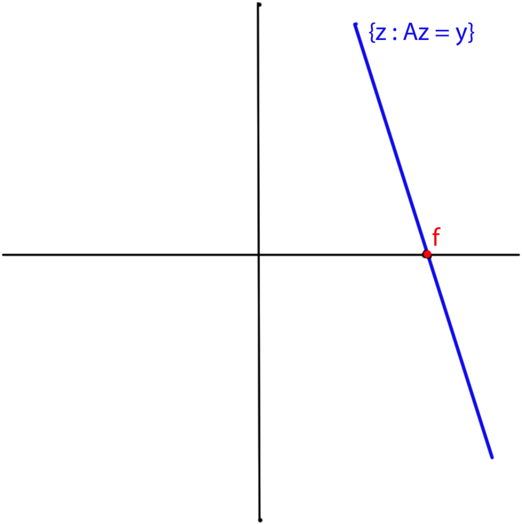
- Ψ = Identity. Signal is sparse in canonical/Kronecker basis
- Φ is discrete Fourier basis, $\phi_j = \left(\frac{1}{\sqrt{N}} e^{i2\pi jk/N}\right)_{k=0}^{N-1}$
- The Kronecker and Fourier bases are incoherent:

$$\mu(\Phi, \Psi) := \sqrt{N} \max_{j,k} | < \phi_j, \psi_k > | = 1.$$

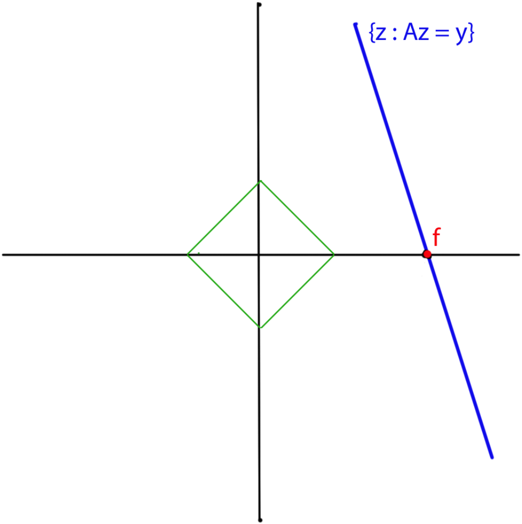
Why does ℓ_1 minimization work?



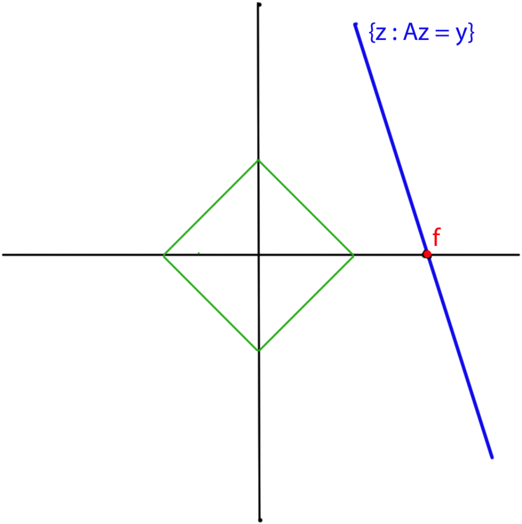
Why does ℓ_1 minimization work?



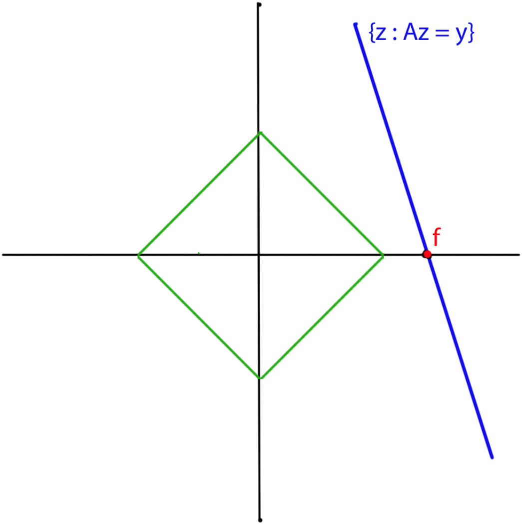
Why does ℓ_1 minimization work?



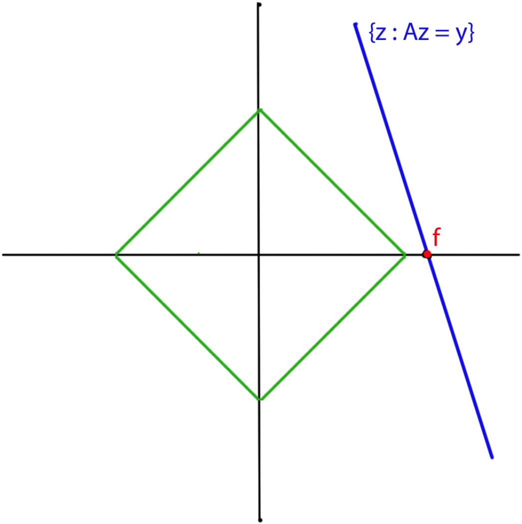
Why does ℓ_1 minimization work?



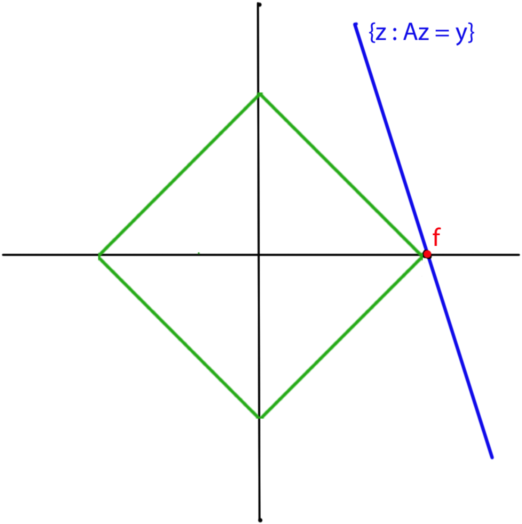
Why does ℓ_1 minimization work?



Why does ℓ_1 minimization work?



Why does ℓ_1 minimization work?



Reconstructing sparse signals

ℓ_1-minimization

$$\mathbf{x}^{\#} = \arg\min_{\mathbf{z}\in\mathbb{R}^N} \sum_{j=1}^{N} |z_j| \quad \text{such that} \quad \mathbf{Az} = \mathbf{Ax}.$$

or, if $\mathbf{x}$ is sparse with respect to basis Ψ,

$$\mathbf{x}^{\#} = \arg\min_{\mathbf{z}\in\mathbb{R}^N} \sum_{j=1}^{N} |(\Psi^* z)_j| \quad \text{such that} \quad \mathbf{Az} = \mathbf{Ax}.$$

Theorem (Sparse recovery via incoherent sampling[1])

Let (Φ, Ψ) *be a pair of incoherent orthonormal bases of* $\mathbb{R}^N$*.*

Select m *(possibly not distinct) rows of* Φ *i.i.d. uniformly to form* $\mathbf{A} : \mathbb{R}^N \to \mathbb{R}^m$*, where*

$$m \lesssim Cs \log(N).$$

With exceedingly high probability, the following holds: for all $\mathbf{x} \in \mathbb{R}^N$ *such that* $\Psi^* \mathbf{x}$ *is* s*-sparse,*

$$\mathbf{x} = arg \min_{\mathbf{z} \in \mathbb{R}^N} \sum_{j=1}^{N} |(\Psi^* z)_j| \quad such\ that \quad \mathbf{Az} = \mathbf{Ax}.$$

Such reconstruction is also stable to sparsity defects and robust to noise.

---

[1] Candès, Romberg, Tao '06, Rudelson Vershynin '08, ...

Theory is largely restricted to: incoherent measurement/sparsity bases, finite-dimensional spaces, and sparsity in orthonormal representations; not sufficient for key examples

Current research directions:

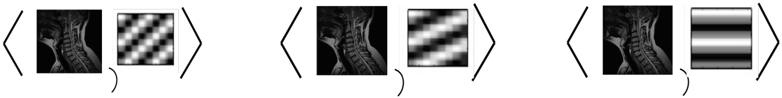
1. **Importance sampling** for compressive sensing applications
2. **Adaptive** sampling strategies
3. Extend theory from sparsity in orthonormal bases to sparsity in **redundant dictionaries**
4. Extend theory from finite-dimensional spaces to **infinite-dimensional** spaces

Compressive imaging

In MRI, one cannot observe the $N = n \times n$ pixel image directly; can only take samples from 2D (or 3D) discrete Fourier transform $\mathcal{F}$.

So we can acquire a number $m \ll N$ linear measurements of the form

$$y_{k_1,k_2} = (\mathcal{F}\mathbf{x})_{k_1,k_2} = \frac{1}{n}\sum_{j_1,j_2} x_{j_1,j_2} e^{2\pi i(k_1 j_1 + k_2 j_2)/n}, -n/2+1 \leq k_1, k_2, \leq n/2$$

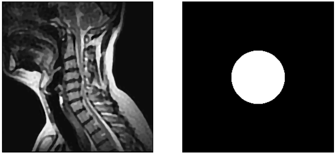


Smaller m means faster MRI scan! How to subsample in frequency domain?

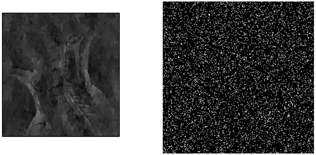
In the MRI setting ... random sampling fails

Reconstructions of an image from $m = .1N$ frequency measurements using *total variation minimization.*

Pixel space / Frequency space



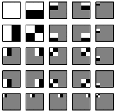
Reconstruction from lowest frequencies



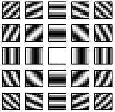
Reconstruction from uniformly subsampled frequencies

In the MRI setting ... random sampling fails

Image



Natural images are sparsely represented in 2D wavelet bases Ψ

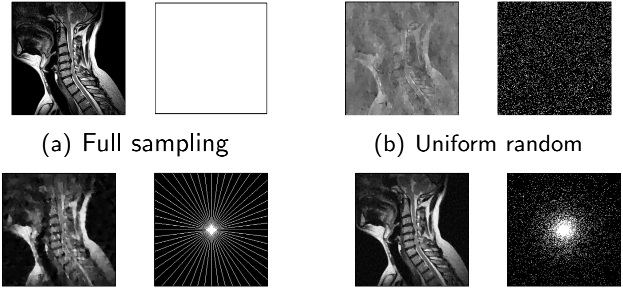


Possible sensing measurements are Fourier measurements Φ

This is because wavelet and Fourier bases are not incoherent

Importance sampling

Image domain / Fourier domain



(a) Full sampling (b) Uniform random

(c) Radial line sampling (d) Variable-density

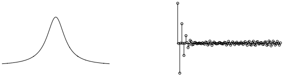
Used in MRI: radial-line sampling. New: "importance sampling": take random samples according to an inverse-square distance variable density: Draw frequency (k_1, k_2) with probability $\propto \frac{1}{k_1^2+k_2^2}$.

With variable density sampling, can extend compressed sensing results and prove that $m \gtrsim s \log(N)$ 2D DFT measurements suffice for recovering images with s-sparse wavelet expansions.

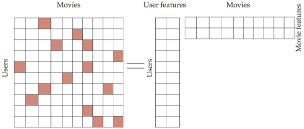
Examples of sparsity:

Natural images:

Smooth function interpolation

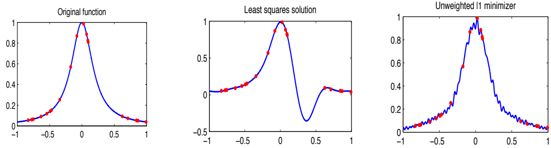


Low-rank matrices:



High-dimensional function interpolation

Given a function $f : \mathcal{D} \to \mathbb{C}$ on a d-dimensional domain $\mathcal{D}$, reconstruct or interpolate f from sample values $f(t_1), \dots, f(t_m)$.

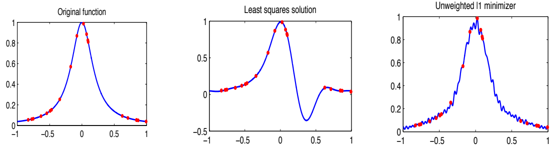


Assume the form $f(t) = \sum_{j \in \Gamma} x_j \psi_j(t)$ where $\mathbf{x}$ has assumed structure:

1. Sparsity: $\|\mathbf{x}\|_0 := \{\ell : x_\ell \neq 0\} \leq s$
2. Smoothness: coefficient decay $\sum_j j^r |x_j| < \infty$.

High-dimensional function interpolation

Given a function $f : \mathcal{D} \to \mathbb{C}$ on a d-dimensional domain $\mathcal{D}$, reconstruct or interpolate f from sample values $f(t_1), \ldots, f(t_m)$.



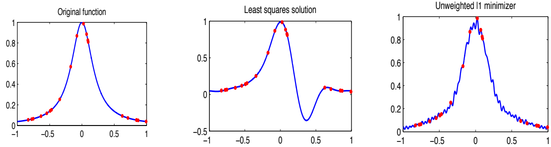
Assume the form $f(t) = \sum_{j \in \Gamma} x_j \psi_j(t)$ where $\mathbf{x}$ has assumed structure:

1. Sparsity: $\|\mathbf{x}\|_0 := \{\ell : x_\ell \neq 0\} \leq s$
2. Smoothness: coefficient decay $\sum_j j^r |x_j| < \infty$.

Smoothness assumption not strong enough to overcome *curse of dimensionality:* need $m \approx (\frac{1}{\varepsilon})^{d/r}$ sample values for accuracy ε.

High-dimensional function interpolation

Given a function $f : \mathcal{D} \to \mathbb{C}$ on a d-dimensional domain $\mathcal{D}$, reconstruct or interpolate f from sample values $f(t_1), \ldots, f(t_m)$.



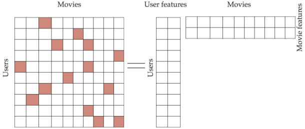
Assume the form $f(t) = \sum_{j \in \Gamma} x_j \psi_j(t)$ where $\mathbf{x}$ has assumed structure:

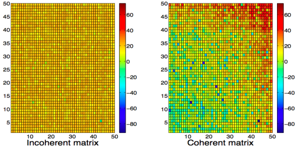
1. Sparsity: $\|\mathbf{x}\|_0 := \{\ell : x_\ell \neq 0\} \leq s$
2. Smoothness: coefficient decay $\sum_j j^r |x_j| < \infty$.

Smoothness assumption not strong enough to overcome *curse of dimensionality:* need $m \approx (\frac{1}{\varepsilon})^{d/r}$ sample values for accuracy ε.

Our work: combine smoothness + sparsity for *weighted* ℓ_1*-coefficient function spaces.* $m \approx (\frac{1}{\varepsilon}) s \log^3(s)$ samples sufficient to reconstruct such a function, *independent of dimension d*

Low-rank matrix completion / approximation





Previous results: a rank-r *incoherent* $n \times n$ matrix M may be completed (via convex optimization) from $m \approx nr \log^2(n)$ uniformly sampled entries

Our results: An *arbitrary* rank-r matrix M may be completed (via convex optimization) from $m \approx nr \log(n)$ entries, sampled according to a specific non-uniform distribution *adapted* to the matrix leverage scores.

Also: extensions to only approximately low-rank matrices, two-stage adaptive sampling

Summary

Compressed sensing and related optimization problems often assume incoherence between the sensing and sparsity bases to derive sparse recovery guarantees.

Incoherence is restrictive and not achievable in many problems of practical interest.

With small local coherence from one basis to another, one may derive sampling strategies and sparse recovery results for a wide range of new sensing problems (imaging, matrix completion, ...)

Also: weighted sparsity, measurement error, adaptive sampling ...