Sparse recovery with minimal incoherence

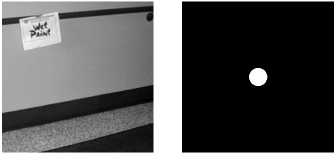
Rachel Ward,
University of Texas at Austin

Duke University, 2013

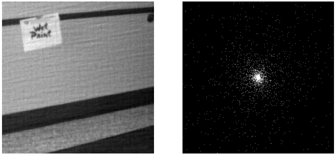
Research supported in part by the ONR, Air Force, and Sloan Foundation

1024 by 1024 pixel Image to be recovered from 1% subsampled frequency measurements

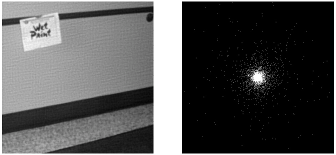
Image to be recovered from 1% subsampled frequency measurements



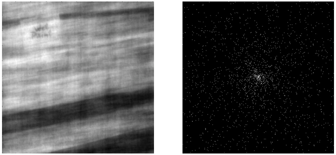
Lowest frequencies only



$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-1}$



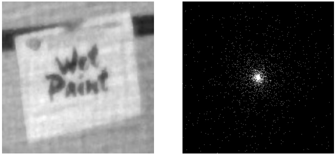
$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-3/2}$



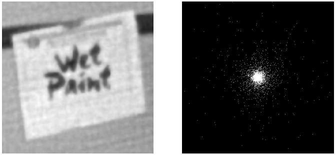
$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-1/2}$

Image to be recovered from 1% subsampled frequency measurements

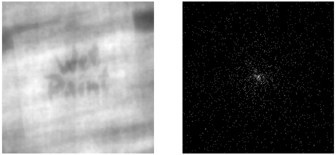
Lowest frequencies only



$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-1}$



$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-3/2}$



$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-1/2}$

Incoherent sampling

1. N-dimensional signal of interest $\mathbf{x} = (x_j)_{j=1}^N$ with assumed sparsity w.r.t. orthonormal basis Ψ: $\mathbf{x} = \Psi\mathbf{b}$ and $\mathbf{b}$ is sparse
2. $m \times N$ measurement matrix A ($m \ll N$) as a subset of rows from an alternative orthonormal basis Φ^*
3. Derive sampling strategies to recover $\mathbf{x}$ from compressed measurements $\mathbf{y} = A\mathbf{x}$.

Orthonormal bases Φ and Ψ are said to be *mutually incoherent* if

$$\mu(\Phi, \Psi) = \max_{1 \leq k,j \leq N} |\langle \phi_j, \psi_k \rangle| = \frac{\kappa}{\sqrt{N}}$$

Example: DFT and identity matrix

Incoherent sampling

1. N-dimensional signal of interest $\mathbf{x} = (x_j)_{j=1}^N$ with assumed sparsity w.r.t. orthonormal basis Ψ: $\mathbf{x} = \Psi\mathbf{b}$ and $\mathbf{b}$ is sparse
2. $m \times N$ measurement matrix A ($m \ll N$) as a subset of rows from an alternative orthonormal basis Φ^*
3. Derive sampling strategies to recover $\mathbf{x}$ from compressed measurements $\mathbf{y} = A\mathbf{x}$.

Orthonormal bases Φ and Ψ are said to be *mutually incoherent* if

$$\mu(\Phi, \Psi) = \max_{1 \le k, j \le N} |\langle \phi_j, \psi_k \rangle| = \frac{\kappa}{\sqrt{N}}$$

Example: DFT and identity matrix

If Φ and Ψ are incoherent, $\max_{1 \leq k,j \leq N} |\langle \phi_j, \psi_k \rangle| \leq \kappa/\sqrt{N}$:

Sampling strategy: $A = P_\Omega \Phi^*$ where P_Ω selects m rows of Φ^* i.i.d. from the uniform distribution

Reconstruction strategy from $\mathbf{y} = A\mathbf{x}$: solve the ℓ_1 minimization program

$$\mathbf{x}^\# = \arg\min_{\mathbf{u} \in \mathbb{R}^N} \|\Psi^* \mathbf{u}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{u}$$

Incoherent sampling:[1] If $m \gtrsim C s \kappa^2 \log^4(N)$, then w.h.p.,

1. $\mathbf{x} = \mathbf{x}^\#$ if $\Psi^* \mathbf{x}$ is s-sparse
2. $\|\mathbf{x} - \mathbf{x}^\#\|_2 \lesssim \frac{1}{\sqrt{s}} \|\Psi^* \mathbf{x} - (\Psi^* \mathbf{x})_s\|_1$ for arbitrary $\mathbf{x}$ (stability)

[1] Candès, Romberg, Tao 06, Donoho 06, Rudelson Vershynin 08, Rauhut 09

If Φ and Ψ are incoherent, $\max_{1 \le k,j \le N} | \langle \phi_j, \psi_k \rangle | \le \kappa/\sqrt{N}$:

Sampling strategy: $A = P_\Omega \Phi^*$ where P_Ω selects m rows of Φ^* i.i.d. from the uniform distribution

Reconstruction strategy from $\mathbf{y} = A\mathbf{x}$: solve the ℓ_1 minimization program

$$\mathbf{x}^\# = \arg \min_{\mathbf{u} \in \mathbb{R}^N} \|\Psi^* \mathbf{u}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{u}$$

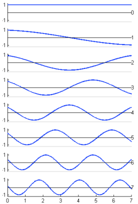
Incoherent sampling:[1] If $m \gtrsim Cs\kappa^2 \log^4(N)$, then w.h.p.,

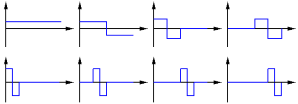
1. $\mathbf{x} = \mathbf{x}^\#$ if $\Psi^* \mathbf{x}$ is s-sparse
2. $\|\mathbf{x} - \mathbf{x}^\#\|_2 \lesssim \frac{1}{\sqrt{s}} \|\Psi^* \mathbf{x} - (\Psi^* \mathbf{x})_s\|_1$ for arbitrary $\mathbf{x}$ (stability)

[1] Candès, Romberg, Tao 06, Donoho 06, Rudelson Vershynin 08, Rauhut 09

But in important applications, sensing and sparsity bases are not incoherent. General sampling strategy?

Example: Imaging with Fourier measurements



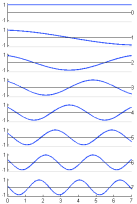


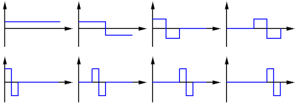
Natural images are compressible in wavelets. In applications (e.g. MRI) measurements are in Fourier domain. Fourier and wavelets have worst possible mutual coherence!

Closer look at coherence structure (Krahmer, W., 2012):

- 1D: If ϕ_ℓ is ℓth row of DFT and Ψ is Haar wavelet basis, $\max_j |\langle \phi_\ell, \psi_j \rangle| \lesssim \frac{1}{\ell^{1/2}}$
- 2D: $\max_\lambda |\langle \phi_{\ell_1,\ell_2}, \psi_\lambda \rangle| \lesssim \left(\ell_1^2 + \ell_2^2\right)^{-1/2}$

Example: Imaging with Fourier measurements



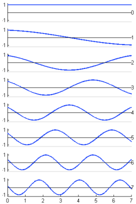


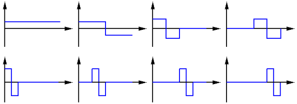
Natural images are compressible in wavelets. In applications (e.g. MRI) measurements are in Fourier domain. Fourier and wavelets have worst possible mutual coherence!

Closer look at coherence structure (Krahmer, W., 2012):

- 1D: If ϕ_ℓ is ℓth row of DFT and Ψ is Haar wavelet basis, $\max_j |\langle \phi_\ell, \psi_j \rangle| \lesssim \frac{1}{\ell^{1/2}}$
- 2D: $\max_\lambda |\langle \phi_{\ell_1,\ell_2}, \psi_\lambda \rangle| \lesssim \left(\ell_1^2 + \ell_2^2\right)^{-1/2}$

Example: Imaging with Fourier measurements





Natural images are compressible in wavelets. In applications (e.g. MRI) measurements are in Fourier domain. Fourier and wavelets have worst possible mutual coherence!

Closer look at coherence structure (Krahmer, W., 2012):

- 1D: If ϕ_ℓ is ℓth row of DFT and Ψ is Haar wavelet basis, $\max_j | \langle \phi_\ell, \psi_j \rangle | \lesssim \frac{1}{\ell^{1/2}}$
- 2D: $\max_\lambda | \langle \phi_{\ell_1,\ell_2}, \psi_\lambda \rangle | \lesssim \left(\ell_1^2 + \ell_2^2\right)^{-1/2}$

Coherence-based sampling

Define the *local* coherences from basis Φ onto basis Ψ by

$$\mu_\ell = \max_j |\langle \phi_\ell, \psi_j \rangle| \leq \kappa_\ell / \sqrt{N}, \quad \ell = 1, \ldots, N$$

Proposed sampling strategy: Select m rows of Φ^* i.i.d. from variable density with weights κ_ℓ^2 to form sensing matrix A.

$\mathbf{x}^\# = \arg\min_{\mathbf{u} \in \mathbb{R}^N} \|\Psi^* \mathbf{u}\|_1$ subject to $\mathbf{y} = A\mathbf{u}$

Theorem: (Krahmer, Rauhut, W '12): *If* $m \gtrsim s\left(\frac{1}{N}\sum_{\ell=1}^N \kappa_\ell^2\right) \log^4(N)$, *then w.h.p.,*

1. $\mathbf{x}^\# = \mathbf{x}$ *if* $\Psi^*\mathbf{x}$ *is s-sparse*
2. $\|\mathbf{x} - \mathbf{x}^\#\|_2 \lesssim \frac{1}{\sqrt{s}} \|\Psi^*\mathbf{x} - (\Psi^*\mathbf{x})_s\|_1$ *for any* $\mathbf{x}$.

Coherence-based sampling

Define the *local* coherences from basis Φ onto basis Ψ by

$$\mu_\ell = \max_j |\langle \phi_\ell, \psi_j \rangle| \leq \kappa_\ell / \sqrt{N}, \quad \ell = 1, \ldots, N$$

Proposed sampling strategy: Select m rows of Φ^* i.i.d. from variable density with weights κ_ℓ^2 to form sensing matrix A.

$\mathbf{x}^\# = \arg\min_{\mathbf{u} \in \mathbb{R}^N} \|\Psi^* \mathbf{u}\|_1$ subject to $\mathbf{y} = A\mathbf{u}$

Theorem: (Krahmer, Rauhut, W '12): *If* $m \gtrsim s\big(\frac{1}{N}\sum_{\ell=1}^N \kappa_\ell^2\big) \log^4(N)$, *then w.h.p.,*

1. $\mathbf{x}^\# = \mathbf{x}$ *if* $\Psi^*\mathbf{x}$ *is s-sparse*
2. $\|\mathbf{x} - \mathbf{x}^\#\|_2 \lesssim \frac{1}{\sqrt{s}} \|\Psi^*\mathbf{x} - (\Psi^*\mathbf{x})_s\|_1$ *for any* $\mathbf{x}$.

Define the *local* coherences from basis Φ onto basis Ψ by

$$\mu_\ell = \max_j | \langle \phi_\ell, \psi_j \rangle | \leq \kappa_\ell / \sqrt{N}, \quad \ell = 1, \dots, N$$

Theorem: (Krahmer, Rauhut, W '12): *If* $m \gtrsim s \left(\frac{1}{N} \sum_{\ell=1}^{N} \kappa_\ell^2 \right) \log^4(N)$, *then w.h.p.,*

1. $\mathbf{x}^\# = \mathbf{x}$ *if* $\Psi^* \mathbf{x}$ *is s-sparse*
2. $\|\mathbf{x} - \mathbf{x}^\#\|_2 \lesssim \frac{1}{\sqrt{s}} \|\Psi^* \mathbf{x} - (\Psi^* \mathbf{x})_s\|_1$ *for any* $\mathbf{x}$.

2D Fourier $\rightarrow$ Haar:

1. $\kappa_{\max}^2 = N$, but $\frac{1}{N} \sum_{\ell_1, \ell_2}^{N} \kappa_{\ell_1, \ell_2}^2 \leq \log(N)$, so optimal sparse recovery results up to factor of $\log(N)$.
2. Variable density sampling srategy: $(\ell_1, \ell_2) \sim \frac{1}{\ell_1^2 + \ell_2^2}$

Proof: Use theory of bounded orthonormal systems (Rauhut '09) and do "change of measure" to derive a restricted isometry property for matrix $DA\Psi$ with diagonal preconditioner D

A more technical result: consider for reconstruction instead

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{u}\in\mathbb{R}^N} \|\nabla \mathbf{u}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{u}$$

Theorem: (Needell, W '12): *If $m \gtrsim s\log^5(N)\log^3(s)$ frequencies are drawn from the variable density $(\ell_1^2 + \ell_2^2)^{-1}$ to form A, then w.h.p. for all* $\mathbf{x}$,

$$\|\mathbf{x} - \mathbf{x}\|_2 \lesssim \frac{1}{\sqrt{s}}\|\nabla \mathbf{x} - (\nabla \mathbf{x})_s\|_1.$$

Proof uses Bernstein-type estimates of Cohen-Devore-Petrushev-Xu (2000) which bound the compressibility of 2D wavelet coefficients by the total variation.

Proof: Use theory of bounded orthonormal systems (Rauhut '09) and do "change of measure" to derive a restricted isometry property for matrix $DA\Psi$ with diagonal preconditioner D

A more technical result: consider for reconstruction instead

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{u} \in \mathbb{R}^N} \|\nabla \mathbf{u}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{u}$$

Theorem: (Needell, W '12): *If* $m \gtrsim s \log^5(N) \log^3(s)$ *frequencies are drawn from the variable density* $(\ell_1^2 + \ell_2^2)^{-1}$ *to form* A*, then w.h.p. for all* $\mathbf{x}$*,*

$$\|\mathbf{x} - \mathbf{x}\|_2 \lesssim \frac{1}{\sqrt{s}} \|\nabla \mathbf{x} - (\nabla \mathbf{x})_s\|_1.$$

Proof uses Bernstein-type estimates of Cohen-Devore-Petrushev-Xu (2000) which bound the compressibility of 2D wavelet coefficients by the total variation.

Proof: Use theory of bounded orthonormal systems (Rauhut '09) and do "change of measure" to derive a restricted isometry property for matrix $DA\Psi$ with diagonal preconditioner D

A more technical result: consider for reconstruction instead

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{u} \in \mathbb{R}^N} \|\nabla \mathbf{u}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{u}$$

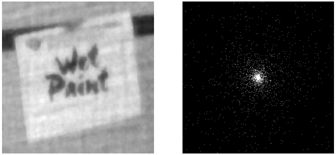
Theorem: (Needell, W '12): *If $m \gtrsim s \log^5(N) \log^3(s)$ frequencies are drawn from the variable density $(\ell_1^2 + \ell_2^2)^{-1}$ to form A, then w.h.p. for all $\mathbf{x}$,*

$$\|\mathbf{x} - \mathbf{x}\|_2 \lesssim \frac{1}{\sqrt{s}} \|\nabla \mathbf{x} - (\nabla \mathbf{x})_s\|_1.$$

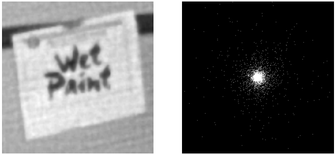
Proof uses Bernstein-type estimates of Cohen-Devore-Petrushev-Xu (2000) which bound the compressibility of 2D wavelet coefficients by the total variation.

Image to be recovered from 1% subsampled frequency measurements

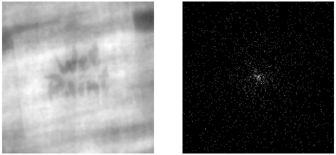
Lowest frequencies only



$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-1}$



$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-3/2}$



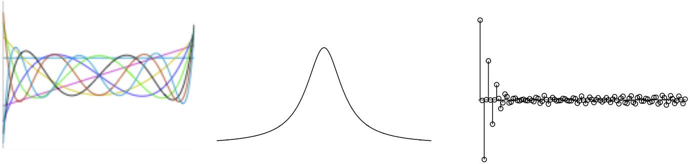
$(k_1, k_2) \sim (k_1^2 + k_2^2)^{-1/2}$

Related work

- **Lustig, Donoho, Pauly 2007, Lustig, Donoho, Santos, Pauly 2008:** Compressed sensing MRI. Empirical studies suggest to sample K-space according to densities scaling inversely to power of the distance to the origin.
- **G. Puy, P. Vandergheynst, and Y. Wiaux, 2011**: Proposed to use what we call "local coherences" to derive sampling strategies for Fourier/wavelet systems.
- **Adcock, Hansen, Poon, Roman 2013:** Asymptotic incoherence and asymptotic sparsity for Fourier / wavelet systems. Multilevel random subsampling. Take into account structured sparsity in wavelets

Coherence-based sampling in polynomial interpolation

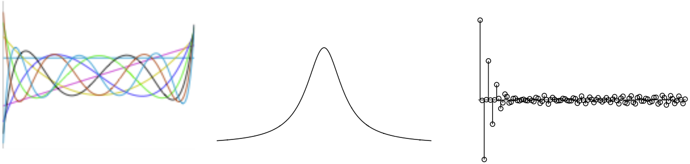
Polynomial interpolation



The Legendre polynomials, a smooth function, and its Legendre series coefficients.

- The Legendre polynomials $(L_j)_{j\geq 0}$ form an orthonormal basis for $L_2([-1,1])$, $\int_{-1}^{1} L_j(x)L_k(x)dx = \delta_{jk}$.
- Consider f with sparse Legendre expansion $f(x) \approx \sum_{j=0}^{N} c_j L_j(x)$ and $\|c\|_0 \leq s$.
- Sampling / reconstruction strategy for interpolating f from sampling points $f(x_1), f(x_2), \ldots, f(x_m)$?

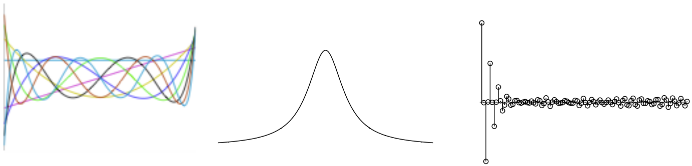
Polynomial interpolation



The Legendre polynomials, a smooth function, and its Legendre series coefficients.

- The Legendre polynomials $(L_j)_{j\geq 0}$ form an orthonormal basis for $L_2([-1,1])$, $\int_{-1}^{1} L_j(x)L_k(x)dx = \delta_{jk}$.
- Consider f with sparse Legendre expansion $f(x) \approx \sum_{j=0}^{N} c_j L_j(x)$ and $\|c\|_0 \leq s$.
- Sampling / reconstruction strategy for interpolating f from sampling points $f(x_1), f(x_2), \dots, f(x_m)$?

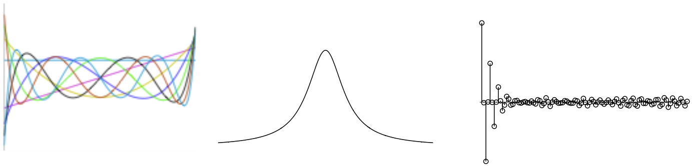
Polynomial interpolation



The Legendre polynomials, a smooth function, and its Legendre series coefficients.

- Legendre polynomial uniform bound: $\|L_j\|_\infty = \sqrt{2j+1}$.
- Finer measure of growth: $|L_j(x)| \leq \kappa(x) = \frac{2}{\sqrt{\pi}}\left(\frac{1}{1-x^2}\right)^{1/4}$
- $\int_{-1}^{1} \kappa^2(x)dx \leq 3$ - Bounded average local coherence
- Coherence-based sampling strategy [Rauhut, W, '10]: $x_1, x_2, \ldots, x_m \sim \frac{1}{\pi(1-x^2)^{1/2}}dx$ from Chebyshev measure. Aligns with classical results on Lagrange interpolation.

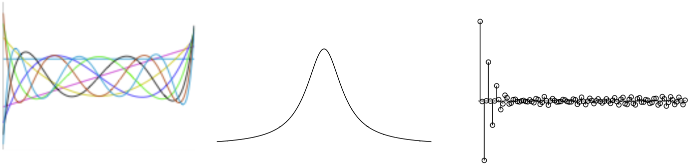
Polynomial interpolation



The Legendre polynomials, a smooth function, and its Legendre series coefficients.

- Legendre polynomial uniform bound: $\|L_j\|_\infty = \sqrt{2j+1}$.
- Finer measure of growth: $|L_j(x)| \leq \kappa(x) = \frac{2}{\sqrt{\pi}}\left(\frac{1}{1-x^2}\right)^{1/4}$
- $\int_{-1}^{1} \kappa^2(x)dx \leq 3$ - Bounded average local coherence
- Coherence-based sampling strategy [Rauhut, W, '10]: $x_1, x_2, \ldots, x_m \sim \frac{1}{\pi(1-x^2)^{1/2}}dx$ from Chebyshev measure. Aligns with classical results on Lagrange interpolation.

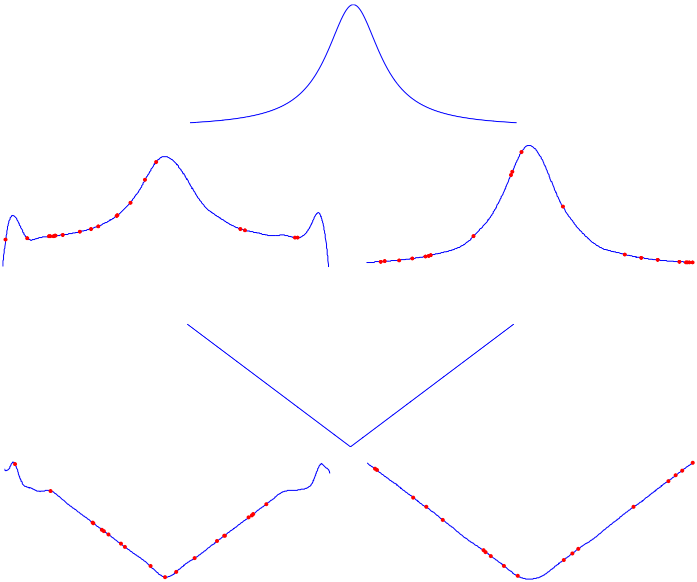
Polynomial interpolation



The Legendre polynomials, a smooth function, and its Legendre series coefficients.

- Legendre polynomial uniform bound: $\|L_j\|_\infty = \sqrt{2j+1}$.
- Finer measure of growth: $|L_j(x)| \le \kappa(x) = \frac{2}{\sqrt{\pi}}\left(\frac{1}{1-x^2}\right)^{1/4}$
- $\int_{-1}^{1} \kappa^2(x)dx \le 3$ - Bounded average local coherence
- Coherence-based sampling strategy [Rauhut, W, '10]: $x_1, x_2, \dots, x_m \sim \frac{1}{\pi(1-x^2)^{1/2}}dx$ from Chebyshev measure. Aligns with classical results on Lagrange interpolation.

uniformly distributed (left) and Chebyshev-distributed (right)



Coherence-based sampling for low-rank matrix completion

Coherent matrix completion

For rank-r matrix $M = U\Sigma V^* \in \mathbb{R}^{n\times n}$, let $u_i \in \mathbb{R}^r$ be ith row of U, $v_j \in \mathbb{R}^r$ is jth row of V.

Probability of observing entry $M_{i,j}$ is $w_{ij} \propto \|u_i\|^2 + \|v_j\|^2$.

$\|u_i\|^2$ and $\|v_j\|^2$ are called the statistical leverage scores of M. Used in sampling strategies for matrix column subset selection[2].

Theorem (Bhojanapalli, Chen, Sanghavi, W '2013:) *Suppose $m = Cn^{1.5} r \log^2(n)$ entries of M are revealed, sampled according to weights w_{ij}. Then with probability at least $1 - n^{-2}$, M is the unique solution to the nuclear norm minimization problem*

$$\text{Minimize } \|Z\|_* \quad \text{subject to} \quad Z_{i,j} = M_{i,j}, \quad (i,j) \in \Omega,$$

[2] Mahoney 2011

Coherent matrix completion

For rank-r matrix $M = U\Sigma V^* \in \mathbb{R}^{n\times n}$, let $u_i \in \mathbb{R}^r$ be ith row of U, $v_j \in \mathbb{R}^r$ is jth row of V.

Probability of observing entry $M_{i,j}$ is $w_{ij} \propto \|u_i\|^2 + \|v_j\|^2$.

$\|u_i\|^2$ and $\|v_j\|^2$ are called the statistical leverage scores of M. Used in sampling strategies for matrix column subset selection[2].

Theorem (Bhojanapalli, Chen, Sanghavi, W '2013:) *Suppose $m = Cn^{1.5} r \log^2(n)$ entries of M are revealed, sampled according to weights w_{ij}. Then with probability at least $1 - n^{-2}$, M is the unique solution to the nuclear norm minimization problem*

$$\textit{Minimize } \|Z\|_* \quad \textit{subject to} \quad Z_{i,j} = M_{i,j}, \quad (i,j) \in \Omega,$$

[2] Mahoney 2011

Summary

One only needs bounded average coherence between sensing and sparsity bases, rather than bounded maximal coherence, to achieve instance-optimal recovery rates, as long as one samples measurements the right way.

Many open questions:

- Incorporate structured sparsity
- Measurement error
- Improved bounds for coherent matrix completion (we require $m \sim n^{3/2}$ but should be $m \sim n$ samples)
- Adaptive sampling: learn local coherence structure as you sample
- Results for Fourer/wavelet sampling in continuous setting

That's all, thanks!

Summary

One only needs bounded average coherence between sensing and sparsity bases, rather than bounded maximal coherence, to achieve instance-optimal recovery rates, as long as one samples measurements the right way.

Many open questions:

- Incorporate structured sparsity
- Measurement error
- Improved bounds for coherent matrix completion (we require $m \sim n^{3/2}$ but should be $m \sim n$ samples)
- Adaptive sampling: learn local coherence structure as you sample
- Results for Fourer/wavelet sampling in continuous setting

That's all, thanks!