Recovering overcomplete sparse representations from structured sensing

Deanna Needell

Claremont McKenna College

Feb. 2015

Support: Alfred P. Sloan Foundation and NSF CAREER #1348721.

Joint work with
Felix Krahmer (Univ. of Göttingen) and Rachel Ward (UT Austin)

Outline

Outline

Imaging via Fourier measurements

- Magnetic Resonance Imaging (MRI):
  Imaging method for medical diagnostics
- *Mathematical model:* For an image $f \in L_2([0,1]^2)$,
  MRI measures 2D-Fourier series coefficients

$$\mathcal{F}f(\omega_1, \omega_2) = \iint f(x,y) \exp(2\pi i(\omega_1 x + \omega_2 y)) dxdy.$$

- Discretize to obtain expansion of a *discrete image* $f \in \mathbb{C}^{N \times N}$ in the *discrete Fourier basis* consisting of the vectors

$$\varphi_{\omega_1,\omega_2}(t_1, t_2) = \frac{1}{N} e^{i2\pi(t_1\omega_1 + t_2\omega_2)/N}, \quad -N/2+1 \leq t_1, t_2 \leq N/2.$$

- Model can also be used for other applications.

Compressed sensing for discrete images

- *Model assumption:* the image $x \in \mathbb{C}^{N^2}$, is approximately s-sparse in a representation system $\{b_i\}$, i.e., $x \approx \sum_{j=1}^{s} x_{k_j} b_{k_j}$.
- *Suitable systems:* Wavelets, shearlets, ...
- *Intuition:* Low dimensionality due to image structure, but nonlinear.

Compressed sensing for discrete images

- *Model assumption:* the image $x \in \mathbb{C}^{N^2}$, is approximately s-sparse in a representation system $\{b_i\}$, i.e., $x \approx \sum_{j=1}^{s} x_{k_j} b_{k_j}$.
- *Suitable systems:* Wavelets, shearlets, ...
- *Intuition:* Low dimensionality due to image structure, but nonlinear.
- *Goal:* Reconstruction of x from $m \ll N^2$ linear measurements, that is, from $y = Ax$, where $A \in \mathbb{C}^{m \times N^2}$.
- Underdetermined system $\Rightarrow$ Many solutions.
- Sufficient condition for robust recovery via convex optimization:
  - $m \gtrsim s \log^{\alpha} N$ random measurements
  - Incoherence between measurements and basis elements.

Important tool: The Restricted Isometry Property

Definition (Candès-Romberg-Tao (2006))

A matrix $A \in \mathbb{C}^{m \times n}$ has the *Restricted Isometry Property* of order s and level $\delta \in (0,1)$ (in short, (s,δ)-RIP), if one has

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \qquad \text{for all } s\text{-sparse } x \in \mathbb{C}^n \qquad (1)$$

The Restricted Isometry Constant $\delta_s(A)$ is the smallest δ satisfying (1).

- **Idea:** Any submatrix of s columns is well-conditioned.
- **Typical result:** If $A \in \mathbb{C}^{m \times n}$ has the $(2k, \delta)$-RIP with $\delta \leq \frac{1}{\sqrt{2}}$ and the equation $y = Ax$ has a k-sparse solution $x^{\#}$, then one has $x^{\#} = \underset{Az=y}{\operatorname{argmin}} \|z\|_1$ (e.g., Cai et al. (2014)).
- Guarantee only works for basis representations

Compressed Sensing MRI - the basis case

Donoho-Lustig-Pauly (2007):
Use compressed sensing to reduce number of MRI measurements needed.

Main issue to address:

- Lack of incoherence between Fourier measurements and good bases for image representation

In this talk: address this issue, provide reconstruction guarantees

Outline

Dictionaries

Definition

A set $\mathbf{D} = \{d_1, \ldots, d_N\} \subset \mathbb{C}^n$ is called a *tight frame* if there exists $A > 0$ such that for all $x \in \mathbb{C}^n$

$$\sum_{i=1}^{N} |\langle d_i, x\rangle|^2 = A\|x\|_2^2$$

- Interpretation: Dictionary,
  the d_i's represent different features of the signal.
- Only few features active: Sparsity
- Redundancy $N > n$ allows for sparser representations.

Analysis vs. Synthesis Sparsity

- A signal $x \in \mathbb{C}^n$ is *synthesis s-sparse* with respect to a dictionary $\mathbf{D}$ if there exist $\{z_i\}_{i=1}^N$ such that $x = \sum_{i=1}^N z_i d_i$.
- A signal $x \in \mathbb{C}^n$ is *analysis s-sparse* with respect to a dictionary $\mathbf{D}$ if $\mathbf{D}^* x$ is s-sparse in $\mathbb{C}^N$.
- For many tight frames that appear in applications, synthesis sparse signals are also approximately analysis sparse.

Definition

For a dictionary $\mathbf{D} \in \mathbb{C}^{n \times N}$ and a sparsity level s, we define the *localization factor* as

$$\eta_{s,\mathbf{D}} = \eta \stackrel{\text{def}}{=} \sup_{\|\mathbf{D}\mathbf{z}\|_2 = 1, \|\mathbf{z}\|_0 \le s} \frac{\|\mathbf{D}^*\mathbf{D}\mathbf{z}\|_1}{\sqrt{s}}.$$

- *Assumption:* Localization factor not too large.

D-RIP

- If $\mathbf{D}$ is coherent, that is some d_i are very close, it may be impossible to recover $\mathbf{z}$ even from complete knowledge of $\mathbf{Dz}$.
- Consequently, RIP based guarantees cannot work.
- *Important observation:* No need to find $\mathbf{z}$, $\mathbf{x} = \mathbf{Dz}$ suffices.

Definition (Candès-Eldar-N-Randall (2010))

Fix a dictionary $\mathbf{D} \in \mathbb{C}^{n \times N}$ and matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$. The matrix $\mathbf{A}$ satisfies the $\mathbf{D}$-RIP with parameters δ and s if

$$(1-\delta)\|\mathbf{Dx}\|_2^2 \leq \|\mathbf{ADx}\|_2^2 \leq (1+\delta)\|\mathbf{Dx}\|_2^2$$

for all s-sparse vectors $\mathbf{x} \in \mathbb{C}^n$.

- *Examples of* $\mathbf{D}$*-RIP matrices:*
  Subgaussian matrices [Candès-Eldar-N-Randall (2010)],
  RIP matrices with random column signs [Krahmer-Ward (2011)]

D-RIP based recovery guarantees

Theorem (Candès-Eldar-N-Randall (2010))

For $\mathbf{A}$ *that has the* $\mathbf{D}$*-RIP and a measurement vector* $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{D}\mathbf{z}$ *consider the minimization problem*

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\tilde{\mathbf{x}} \in \mathbb{C}^n} \|\mathbf{D}^*\tilde{\mathbf{x}}\|_1 \quad \textit{subject to} \quad \mathbf{A}\tilde{\mathbf{x}} = \mathbf{y}.$$

Then the minimizer $\hat{\mathbf{x}}$ *satisfies*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C \frac{\|\mathbf{D}^*\mathbf{x} - (\mathbf{D}^*\mathbf{x})_s\|_1}{\sqrt{s}}$$

- Small localization factor $\Rightarrow$ small error bound
- In this talk: provide $\mathbf{D}$-RIP constructions closer to applications

Outline

Uniform sampling

The *mutual coherence* of two bases $\{\varphi_k\}$ and $\{b_j\}$ is defined to be

$$\mu = \sup_{j,k} |\langle b_j, \varphi_k \rangle|.$$

Theorem (Rudelson-Vershynin (2006), Rauhut (2007))

Consider the matrix $A = \Phi_\Omega B^* \in \mathbb{C}^{m \times N}$ *with entries*

$$A_{\ell,k} = \langle \varphi_{j_\ell}, b_k \rangle, \quad \ell \in [m], k \in [N], \tag{2}$$

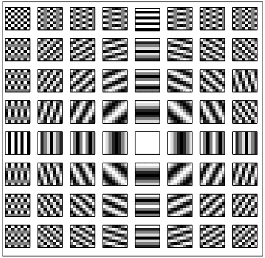
where the φ_{j_ℓ} *are independent samples drawn uniformly at random from an ONB* $\{\varphi_j\}_{j=1}^N$ *incoherent with the sparsity basis* $\{b_j\}$ *in the sense that* $\mu \leq KN^{-1/2}$*. Then once, for some* $s \gtrsim \log(N)$,

$$m \geq C\delta^{-2}K^2 s \log^3(s) \log(N), \tag{3}$$

with probability at least $1 - N^{-\gamma \log^3(s)}$*, the restricted isometry constant* δ_s *of* $\frac{1}{\sqrt{m}}A$ *satisfies* $\delta_s \leq \delta$*. The constants* $C, \gamma > 0$ *are universal.*

Variable density sampling

[Lustig-Donoho-Pauly (2007)]: "For a better performance with real images, one should be undersampling less near the k-space origin and more in the periphery of k-space. For example, one may choose samples randomly with sampling density scaling according to a power of distance from the origin."



- Idea by Puy-Vandergheynst-Wiaux (2011):
  - Variable density sampling can reduce coherence.
  - Strategy: Find optimal weights using convex optimization.
  - Work with problem specific discretization level.
  - No theoretical recovery guarantees.

Local coherence

- *Empirical observation of Puy et al.:*
  Often only few Fourier basis vectors have high coherence with the sparsity basis. Changing the weights can compensate for this inhomogeneity.
- We introduce the *local coherence* to address this issue.

Definition (Local coherence)

The *local coherence* of an ONB $\{\varphi_j\}_{j=1}^N$ of $\mathbb{C}^N$ with respect to another ONB $\{\psi_k\}_{k=1}^N$ of $\mathbb{C}^N$ is the function $\mu_{loc}(j) = \sup_{1\leq k\leq N} |\langle \varphi_j, \psi_k \rangle|$.

RIP for variable density subsampling

Theorem (Consequence of Rauhut-Ward '12)

Assume the local coherence of an ONB $\Phi = \{\varphi_j\}_{j=1}^N$ *with respect to an ONB* $\Psi = \{\psi_k\}_{k=1}^N$ *is pointwise bounded by the function* κ*, that is,* $\sup_{1 \le k \le N} |\langle \varphi_j, \psi_k \rangle| \le \kappa_j$*. Consider the matrix* $A \in \mathbb{C}^{m \times N}$ *with entries*

$$A_{\ell,k} = \langle \varphi_{j_\ell}, \psi_k \rangle, \quad j \in [m], k \in [N], \tag{4}$$

where the j_ℓ *are drawn independently according to* $\nu_\ell = \mathbb{P}(\ell_j = \ell) = \frac{\kappa_\ell^2}{\|\kappa\|_2^2}$*. Suppose that*

$$m \ge C\delta^{-2} \|\kappa\|_2^2 s \log^3(s) \log(N), \tag{5}$$

and let $D = diag(d_{j,j})$*, where* $d_{j,j} = \|\kappa\|_2 / \kappa_j$*. Then with probability at least* $1 - N^{-\gamma \log^3(s)}$*, the preconditioned matrix* $\frac{1}{\sqrt{m}} DA$ *has a restricted isometry constant* $\delta_s \le \delta$*. The constants* $C, \gamma > 0$ *are universal.*

Outline

Total variation

- Discrete image $\mathbf{x} = (x_{j,k}) \in \mathbb{C}^{n^2}$, $\quad (j,k) \in \{1,2,\dots,n\}^2 := [n]^2$
- Discrete directional derivatives

$$\begin{aligned}(x_u)_{j,k} &= x_{j,k+1} - x_{j,k},\\ (x_v)_{j,k} &= x_{j+1,k} - x_{j,k},\end{aligned}$$

- Discrete gradient $\nabla \mathbf{x} = (x_u, x_v)$ is very close to sparse.

- Not a basis representation, does not allow stable image reconstruction
- Total variation (TV): $\|\mathbf{x}\|_{TV} = \|\nabla \mathbf{x}\|_{\ell^1}$.

Compressed sensing via TV minimization

Proposition (N-Ward (2012))

Let $(\mathbf{a}_j)$ *be an orthonormal basis for* $\mathbb{C}^{n^2}$ *that is incoherent with the bivariate Haar basis* $(\mathbf{w}_j)$,

$$\sup_{j,k} |\langle \mathbf{a}_j, \mathbf{w}_k \rangle| \leq C/n.$$

Let $A_u : \mathbb{C}^{n^2} \to \mathbb{C}^m$ *consist of* $m \gtrsim s \log^6(n)$ *uniformly subsampled bases* $\mathbf{a}_j$ *as rows. Then with high probability, the following holds for all* $\mathbf{X} \in \mathbb{C}^{n^2}$*:*
If $y = A_u \mathbf{X} + \xi$ *with* $\|\xi\|_2 \leq \varepsilon$ *and*

$$\hat{\mathbf{X}} = \operatorname*{argmin}_{\mathbf{Z}} \|\mathbf{Z}\|_{TV} \quad \text{such that} \quad \|\mathcal{F}_u \mathbf{Z} - y\|_2 \leq \varepsilon$$

then

$$\|\hat{X} - X\|_2 \lesssim (\frac{\|\nabla X - (\nabla X)_s\|_1}{\sqrt{s}} + \varepsilon$$

Outline

Local incoherence between Fourier and Haar systems

Theorem (Krahmer-Ward (2014))

Fix $\delta \in (0, 1/3)$ *and integers* $N = 2^p$, m, *and* s *such that*

$$m \gtrsim s \log^3 s \log^5 N.$$

Select m *discrete frequencies* (Ω_1^j, Ω_2^j) *independently according to*

$$\mu(\omega_1, \omega_2) := \mathbb{P}\big[(\Omega_1^j, \Omega_2^j) = (\omega_1, \omega_2)\big] \propto \min\left(1, \frac{C'}{\omega_1^2 + \omega_2^2}\right), \quad -\frac{N}{2} + 1 \leq \omega_1, \omega_2 \leq \frac{N}{2},$$

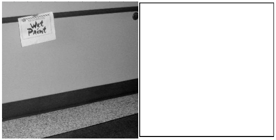
and let $\mathcal{F}_\Omega : \mathbb{C}^{N^2} \to \mathbb{C}^m$ *be the DFT matrix restricted to* $\{(\Omega_1^j, \Omega_2^j)\}$.
Then with high probability, the following holds for all $f \in \mathbb{C}^{N \times N}$.
Given measurements $y = \mathcal{F}_\Omega f$, *the TV-minimizer*

$$f^\# = \operatorname*{argmin}_{g \in \mathbb{C}^{N \times N}} \|g\|_{TV} \quad \text{such that} \quad \mathcal{F}_\Omega g = y,$$

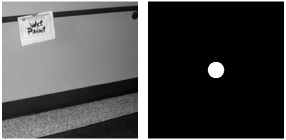
approximates f *to within the best* s*-term approximation error of* ∇f*:*

$$\|f - f^\#\|_2 \leq C \frac{\|\nabla f - (\nabla f)_s\|_1}{\sqrt{s}}.$$

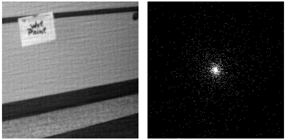
Numerical Simulations



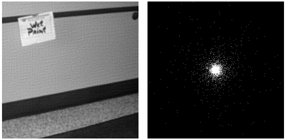
(a) Original image



(b) Lowest frequencies only



(c) Sample $\propto (k_1^2 + k_2^2)^{-1}$

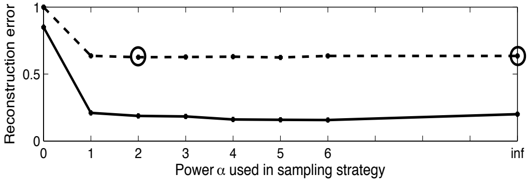


(d) Sample $\propto (k_1^2 + k_2^2)^{-3/2}$

Figure : Reconstruction using $m = 12,000$ noiseless partial DFT measurements with frequencies $\Omega = (k_1, k_2)$ sampled from various distributions.

- Relative reconstruction errors (b) .18, (c) .21, and (d) .19

Which sampling density to choose?



(a) Reconstruction errors by various power-law density sampling at low noise (filled line) and high noise (dashed line)

- Are all the reconstructions of comparable quality?

Really all the same?

(b) The *wet paint* reconstructions indicated by the circled errors on the error plot, zoomed in on the paint sign.

At high noise level, inverse quadratic-law sampling (left) still reconstructs fine details of the image better than low frequency-only sampling (right).

Theorem (Krahmer-N-Ward (2014))

Fix a sparsity level $s < N$, *and constant* $0 < \delta < 1$. *Let* $\mathbf{D} \in \mathbb{C}^{n \times N}$ *be a tight frame, let* $\mathbf{A} = \{a_1, \ldots, a_n\}$ *be an ONB of* $\mathbb{C}^n$, *and* $\kappa \in \mathbb{R}^n_+$ *an entrywise upper bound of the local coherence, that is,*

$$\mu_i^{loc}(\mathbf{A}, \mathbf{D}) = \sup_{j \in [N]} |\langle a_i, d_j \rangle| \leq \kappa_i.$$

Consider the localization energy η. *Construct* $\tilde{\mathbf{A}} \in \mathbb{C}^{m \times n}$ *by sampling vectors from* $\mathbf{A}$ *at random according to the probability distribution* ν *given by* $\nu(i) = \frac{\kappa_i^2}{\|\kappa\|_2^2}$ *and normalizing by* $\sqrt{n/m}$. *Then as long as*

$$\begin{aligned} m &\geq C\delta^{-2} s \|\kappa\|_2^2 (\eta)^2 \log^3(s\eta^2) \log(N), \quad \textit{and} \\ m &\geq C\delta^{-2} s \|\kappa\|_2^2 \eta^2 \log(1/\gamma) \end{aligned} \tag{6}$$

then with probability $1 - \gamma$, $\tilde{\mathbf{A}}$ *satisfies the* $\mathbf{D}$*-RIP with parameters* s *and* δ.

Consequences

- Recovery guarantees for Fourier measurements and Haar wavelet frames of redundancy 2 by previous local coherence analysis.
- Constant local coherence: Implies incoherence based guarantees (for example for oversampled Fourier dictionary).

Outline

Summary and open questions

- Compressive imaging via variable density sampling
- Uniform recovery guarantees for approximately Haar- and Gradient-sparse images
- First guarantees for Fourier measurements and dictionary sparsity
- *Goals:*
  - Basis-independent error bounds via continuous total variation.
  - Why is cubic decay better than quadratic?
  - Sharper error bounds using shearlets, curvelets
  - What is the right noise model?