Greedy Algorithms and Super-resolution

Deanna Needell

CLAREMONT McKENNA COLLEGE

Thanks to the Simons Foundation grant #274305, AIM SQuaRE Collaboration grant, and the Alfred P. Sloan Fellowship.

The Data Deluge

The Data Deluge

How can we handle all this data?

- Build hardware that can store and trasmit more data.
  - We need the resources.
  - There are fundamental limitiations to data storage.

- Design more efficient compression methods.
  - Enter the world of: *Compressive Sensing* (CS)
  - CS gives us efficient compression techniques: “Compressive”
  - More surprisingly, we can acquire the compression without ever having to acquire the entire object!: “Sensing”

A mathematical problem

1. Signal of interest $f \in \mathbb{C}^d (= \mathbb{C}^{N\times N})$

2. Measurement operator $\mathscr{A} : \mathbb{C}^d \to \mathbb{C}^m$.

3. Measurements $y = \mathscr{A} f + \xi$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad & \mathscr{A} & \quad \end{bmatrix} \begin{bmatrix} f \\ \\ \\ \end{bmatrix} + \begin{bmatrix} \xi \end{bmatrix}$$

4. *Problem:* Reconstruct signal f from measurements y

Sparsity

Measurements $y = \mathscr{A} f + \xi$.

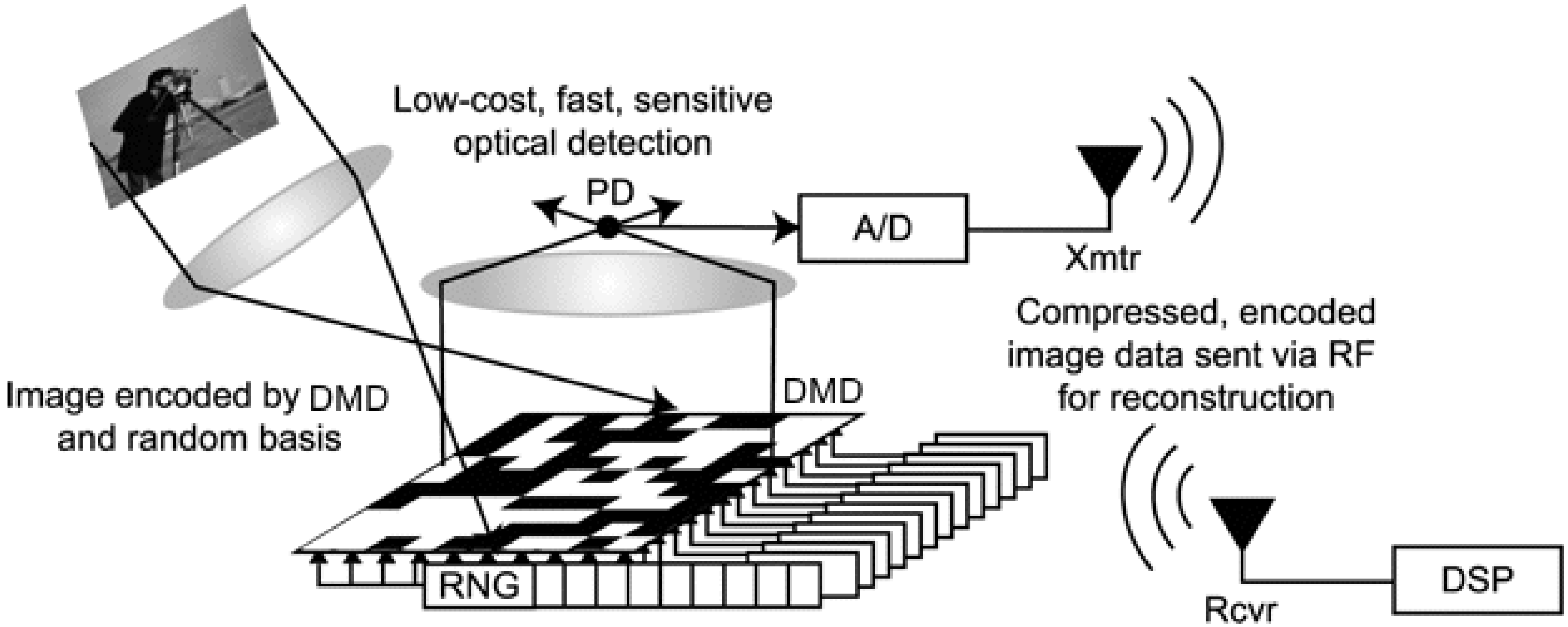
$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & \mathscr{A} & \end{bmatrix} \begin{bmatrix} f \\ \\ \\ \end{bmatrix} + \begin{bmatrix} \xi \end{bmatrix}$$

Assume f is *sparse*:

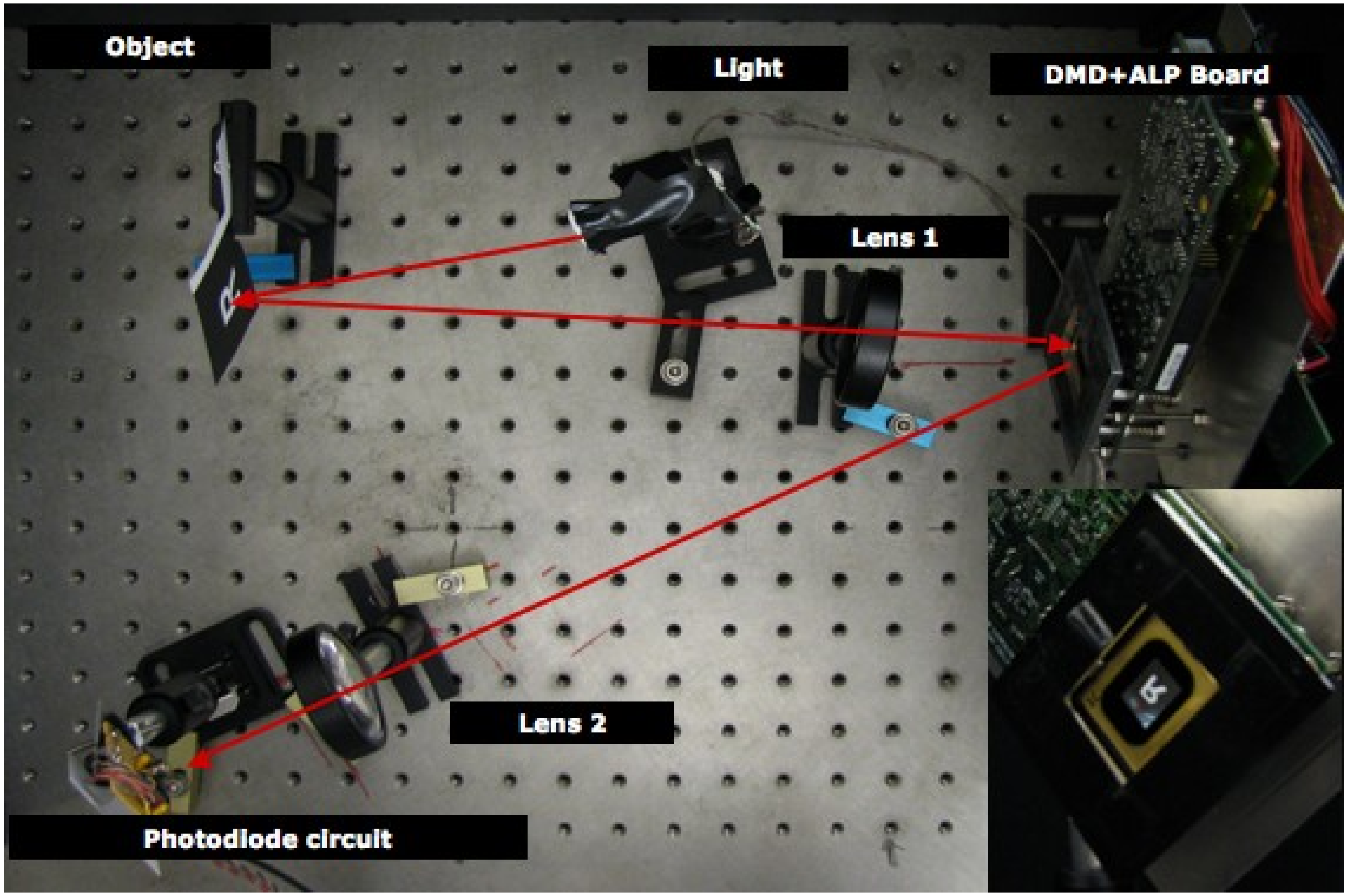
- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\operatorname{supp}(f)| \le s \ll d$
- In orthonormal basis: $f = Bx$ where $\|x\|_0 \le s \ll d$
- In other dictionary: $f = Dx$ where $\|x\|_0 \le s \ll d$

In practice, we encounter *compressible* signals.

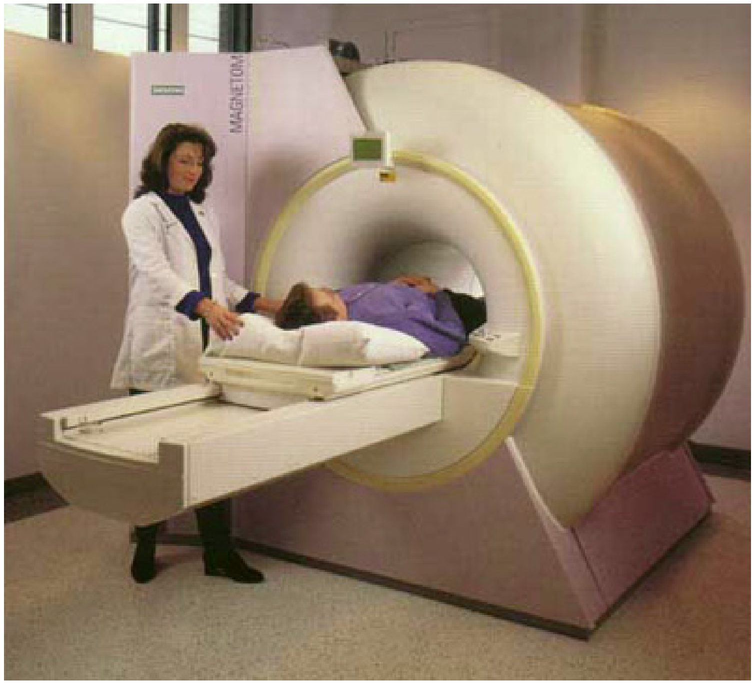
Digital Cameras



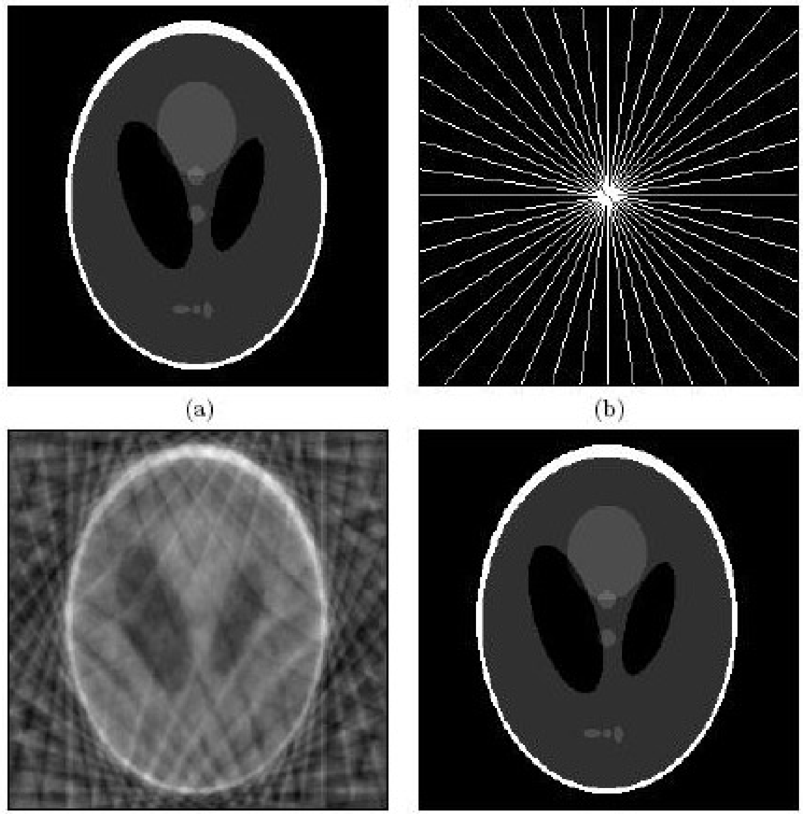
Digital Cameras



MRI



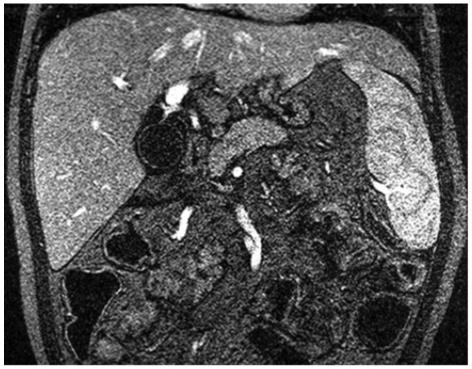
MRI



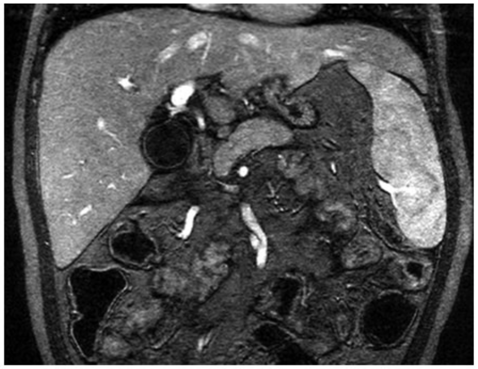
(a) (b) (c) (d)

(Candès et.al.)

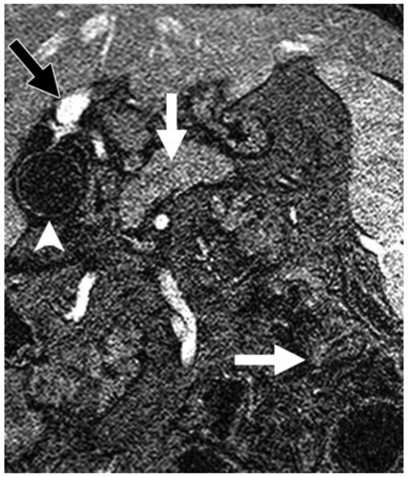
Pediatric MRI



(a)



(b)



(c)

(d)

Many more...

- Radar, Error Correction
- Computational Biology, Geophysical Data Analysis
- Data Mining, classification
- Neuroscience
- Imaging
- Sparse channel estimation, sparse initial state estimation
- Topology identification of interconnected systems
- ...

Background: Restricted Isometry Property

- $\mathscr{A}$ satisfies the Restricted Isometry Property (RIP) when there is $\delta < c$ such that

$$(1-\delta)\|f\|_2 \le \|\mathscr{A} f\|_2 \le (1+\delta)\|f\|_2 \quad \text{whenever } \|f\|_0 \le s.$$

- Gaussian or Bernoulli measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property: $m \gtrsim s \log^4 d$.

Reconstructing the signal f from measurements y

- ℓ_1-minimization [Candès-Romberg-Tao]

Let $\mathscr{A}$ satisfy the *Restricted Isometry Property* and set:

$$\hat{f} = \underset{g}{\operatorname{argmin}} \|g\|_1 \quad \text{such that} \quad \|\mathscr{A} g - y\|_2 \leq \varepsilon,$$

where $\|\xi\|_2 \leq \varepsilon$. Then we can stably recover the signal f:

$$\|f - \hat{f}\|_2 \lesssim \varepsilon + \frac{\|x - x_s\|_1}{\sqrt{s}}.$$

This error bound is optimal. Speed is polynomial (linear programming).

CoSaMP

COSAMP (N-Tropp)

input: Sampling operator A, measurements y, sparsity level s
initialize: Set $x^0 = 0,\ i = 0$.
repeat
signal proxy: Set $p = A^*(y - Ax^i),\ \Omega = \text{supp}(p_{2s}),\ T = \Omega \cup \text{supp}(x^i)$.
signal estimation: Using least-squares, set $b|_T = A_T^\dagger y$ and $b|_{T^c} = 0$.
prune and update: Increment i and to obtain the next approximation, set $x^i = b_s$.
output: s-sparse reconstructed vector $\widehat{x} = x^i$

Same guarantees under RIP as ℓ_1-minimization.

Super-resolution

✦ Goal: Produce high-resolution image from low-resolution samples

✦ Challenge: Model becomes $y = Ax + e$ where A is a (non-random) partial DFT. Goal: identify (support T of) sparse x.

Super-resolution

- Idea: Partial DFT has *translation invariance*: any restriction of a column a_k to $s \leq m$ consecutive elements gives rise to the same sequence, up to an overall scalar
- Moral: A is not an arbitrary dictionary, it has structure we should not ignore!

Super-resolution

✦ Idea: Partial DFT has *translation invariance*: any restriction of a column a_k to $s \leq m$ consecutive elements gives rise to the same sequence, up to an overall scalar

✦ Moral: A is not an arbitrary dictionary, it has structure we should not ignore!

Idea: Pick a number $1 < L < m$ and juxtaposes translated copies of y into the Hankel matrix $Y = \text{Hankel}(y)$, defined as

$$Y = \begin{pmatrix} y_0 & y_1 & \cdots & y_{m-L-1} \\ y_1 & y_2 & \cdots & y_{m-L} \\ \vdots & \vdots & \vdots & \vdots \\ y_{L-1} & y_L & \cdots & y_m \end{pmatrix}.$$

✦ *Wonderful fact*: Without noise, $\text{Ran}\, Y = \text{Ran}\, A_T^L$

Super-resolution

✦ Recovery using this idea: Loop over all atoms a_k and select those for which

$$\angle(a_k^L, \text{Ran}\, Y) = 0.$$

From this set T, recovery by solving

$$A_T \hat{x}_T = y, \qquad \hat{x}_{T^c} = 0.$$

✦ **Theorem [Demanet - N - Nguyen]** : If $m > 2|T|$ and $y = Ax$, then $\hat{x} = x$.

Super-resolution : Noise?

✦ With noise, we no longer have Ran $Y =$ Ran A_T^L

✦ **Theorem [Demanet - N - Nguyen]** : Let $y = Ax + e$ with $e \sim N(0, \sigma^2 I_m)$. Then with high probability,

$$\sin \angle(a_k^L, \text{Ran } Y) \leq c\,\varepsilon_1$$

for all indices k in the support set (and $c\varepsilon_1$ is explicitly computed).

✦ Extension: Choose atoms with small enough angles.

Super-resolution : Experimental Results

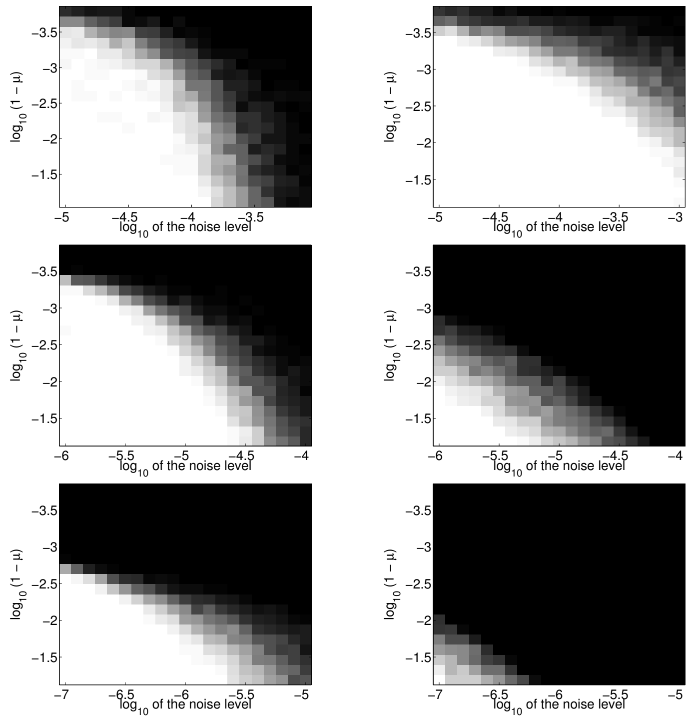


Figure 1: Probability of recovery, from 1 (white) to 0 (black) for the superset method (left column) and the matrix pencil method (right column). Top row: 2-sparse signal. Middle row: 3-sparse signal. Bottom row: 4-sparse signal. The plots show recovery as a function of the noise level (x-axis, $\log_{10}\sigma$) and the coherence (y-axis, $\log_{10}(1-\mu)$).

Partial Inversion

Can we adapt a method like CoSaMP to super-resolution?

CoSaMP (N-Tropp)

---

input: Sampling operator A, measurements y, sparsity level s
initialize: Set $x^0 = 0$, $i = 0$.
repeat
signal proxy: Set $p = A^*(y - Ax^i)$, $\Omega = \text{supp}(p_{2s})$, $T = \Omega \cup \text{supp}(x^i)$.
signal estimation: Using least-squares, set $b|_T = A_T^\dagger y$ and $b|_{T^c} = 0$.
prune and update: Increment i and to obtain the next approximation, set $x^i = b_s$.
output: s-sparse reconstructed vector $\widehat{x} = x^i$

Partial Inversion

General model:

- S: downsampling matrix
- H: filtering (antialiasing) operation
- Ψ: sparsifying basis
- $A = SH\Psi$: sampling operator
- $y = Ax + e$

Partial Inversion

256×256 example:

- SH: by shifting the filter kernel $h = \{0.1, 0.2, 0.4, 0.2, 0.1\}$ by two from one row to the next
- Ψ: Haar wavelet basis
- $A = SH\Psi$: sampling operator
- $y = Ax + e$

Absolute values of $A^* A$.

PARTIAL INVERSION (Chen-Divekar-N)

---

input: $y = Ax$, return best s-sparse approximation $\hat{x}$

initialize: $\hat{x} \leftarrow A^* y; I \leftarrow$ indices of the L-largest magnitudes of $\hat{x}$

repeat

signal proxy: $\hat{x}_I \leftarrow A_I^\dagger y$

$r \leftarrow y - A_I \hat{x}_I$

$J \leftarrow I^c$

signal estimation: $\hat{x}_J \leftarrow A_J^* r$

prune and update: $I \leftarrow$ indices of the L-largest magnitude components of $\hat{x}$

Partial Inversion

- PartInv:

$$\hat{x}_I = A_I^{\dagger} y = A_I^{\dagger}(A_I x_I + A_{I^c} x_{I^c})$$
$$= x_I + (A_I^* A_I)^{-1} A_I^* A_{I^c} x_{I^c}.$$

- CoSaMP:

$$\hat{x}_I = A_I^* y$$
$$= A_I^* A_I x_I + A_I^* A_{I^c} x_{I^c}$$
$$= x_I + (A_I^* A_I - I) x_I + A_I^* A_{I^c} x_{I^c}$$

Partial Inversion

✦ **Theorem [Chen- Divekar - N]** Let $x \in \mathbb{C}^N$ be a s-sparse vector with support set T satisfying

$$|x_i| \geq 3\varepsilon \|x\|_2, \quad \forall i \in T, \tag{1}$$

for some fixed constant $0 < \varepsilon \leq \frac{1}{3\sqrt{s}}$. Assume that the dictionary A satisfies the following properties:

$$\|A^*_{T_1} A_{T_1} x_{T_1}\|_2 \geq (1-\varepsilon)^2 \|x_{T_1}\|_2, \qquad \forall\ T_1 \subseteq T \tag{2}$$

$$\|A_I\| \leq C, \qquad \forall\ |I| \leq L \tag{3}$$

$$\|A_I^\dagger\| \leq C, \qquad \forall\ |I| \leq L \tag{4}$$

$$\|A_I A_I^\dagger A_{I^c \cap T}\| \leq \varepsilon / C, \qquad \forall\ |I| \leq L. \tag{5}$$

Then PartInv reconstructs the signal, $\hat{x} = x$ in at most s iterations.

PartInv : Experimental Results

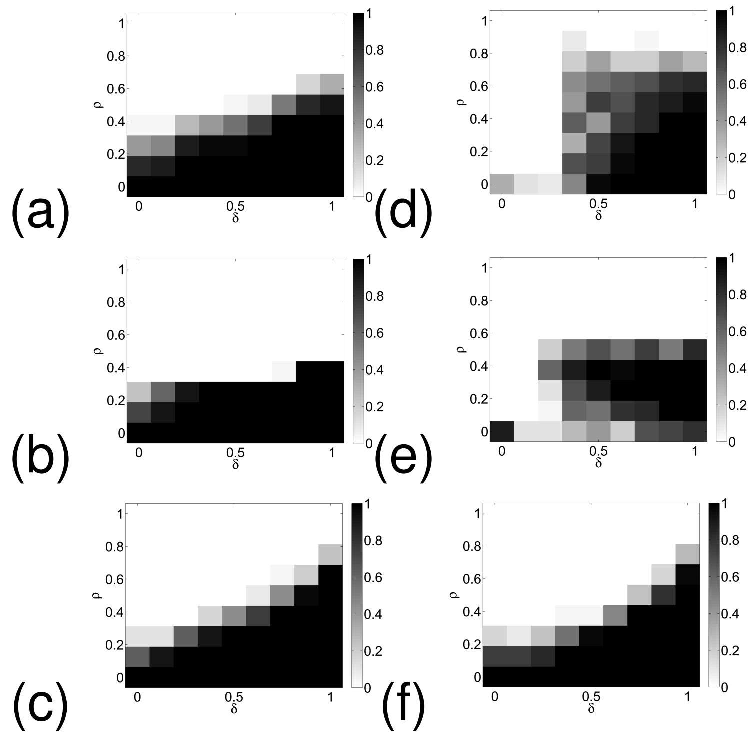


Figure 2: *Proportion of successes on Gaussian matrices using (a) PartInv, (b) CoSaMP and (c) ℓ_1-minimization, and proportion of successes on correlated column subset matrices using (d) PartInv, (e) CoSaMP and (f) ℓ_1-minimization for various values of $\delta = \frac{m}{n} \in (0, 1)$ (horizontal axis) and $\rho = \frac{s}{m} \in (0, 1)$ (vertical axis).*

Even newer braches of CS

1-bit compressive sensing

- Measurements: $y = \text{sign}(Af)$ (extreme quantization)
- Noise: Random or adversarial bit flips
- Assumption: signal f lies in some (convex) set K
- $\hat{f} = \max_x \langle y, Ax \rangle \quad \text{s.t.} \quad x \in K$
- (Plan-Vershynin): $\|\hat{f} - f\|_2 \lesssim w(K)/\sqrt{m}$
- Greedy methods for accurate recovery from optimal number of (e.g. Gaussian) measurements [Baraniuk et al.]

1-bit compressive sensing

✦ In general, results are of the form:

$$\|\hat{f} - f\|_2 \lesssim \lambda^{-c},$$

where $\lambda = \frac{m}{s\log(n/s)}$ is the oversampling factor.

✦ New results [Baraniuk-Foucart-N-Plan-Wootters]: Provide a reconstruction method to obtain

$$\|\hat{f} - f\|_2 \lesssim e^{-\lambda},$$

(in preparation).

1-bit compressive sensing

- To do:

- Optimal greedy methods for recovery (what is optimal?)

- Methods for recovery when sparsity is w.r.t. aribtrary dictionary D

- Mixed models of quantization – unified framework for all precision

Adaptive measurement schemes

- Design measurement operator on the fly
- Fundamental limitations on improved recovery [Candès-Davenport]
- However, improvements still possible (such as reduced number of measurements needed) [Aldroubi et al., Iwen-Tewfik, Indyk et al.]
- Adaptive measurement schemes for fixed sampling structures, total variation, sparsity in dictionaries, average case results, ...

Adaptive measurement schemes

- ✦ Sampling from constrained measurements
- ✧ Certain constrained settings don't afford improvements via adaptivity (Davenport-N)
- ✧ Identify geometric properties of constraints that offer adaptive improvements
- ✧ Design adaptive measurement schemes and recovery algorithms for those that do

Thank you!

E-mail:

✧ dneedell@cmc.edu

Web:

✧ www.cmc.edu/pages/faculty/DNeedell

References:

- E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 59(8):1207–1223, 2006.
- E. J. Candès, Y. C. Eldar, D. Needell and P. Randall. Compressed sensing with coherent and redundant dictionaries. Applied and Computational Harmonic Analysis, 31(1):59-73, 2010.
- M. A. Davenport, D. Needell and M. B. Wakin. Signal Space CoSaMP for Sparse Recovery with Redundant Dictionaries, IEEE Trans. Info. Theory, to appear.
- D. Needell and R. Ward. Stable image reconstruction using total variation minimization. SIAM J. Imaging Sciences, vol. 6, num. 2, pp. 1035-1058.
- D. Needell and R. Ward. Near-optimal compressed sensing guarantees for total variation minimization, IEEE Trans. Image Proc., iss. 99, 2013.
- L. Demanet, D. Needell and N. Nguyen. Super-resolution via superset selection and pruning. SAMPTA 2013.
- R. Giryes and D. Needell. Greedy Signal Space Methods for Incoherence and Beyond, Appl. Harmon. Analysis, to appear.