Mixed Operators in Compressed Sensing

Deanna Needell

Stanford University
Joint work with Matthew Herman, UCLA

AT13, San Antonio, March 2010

Notation

1. x is an unknown signal in $\mathbb{R}^d$.
2. Measurement matrix $A : \mathbb{R}^d \to \mathbb{R}^m$.
3. Noisy measurements $y = Ax + e$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad & A & \quad \end{bmatrix} \begin{bmatrix} x \end{bmatrix} + \begin{bmatrix} e \end{bmatrix}$$

4. Assume x is s-sparse: $\|x\|_0 \stackrel{\text{def}}{=} |\text{supp}(x)| \leq s \ll d$.
5. sparsity, measurements, dimension

Notation

1. x is an unknown signal in $\mathbb{R}^d$.
2. Measurement matrix $A : \mathbb{R}^d \to \mathbb{R}^m$.
3. Noisy measurements $y = Ax + e$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad A \quad \end{bmatrix} \begin{bmatrix} x \end{bmatrix} + \begin{bmatrix} e \end{bmatrix}$$

4. Assume x is s-sparse: $\|x\|_0 \stackrel{\text{def}}{=} |\text{supp}(x)| \le s \ll d$.
5. sparsity, measurements, dimension

Notation

1. x is an unknown signal in $\mathbb{R}^d$.
2. Measurement matrix $A : \mathbb{R}^d \to \mathbb{R}^m$.
3. Noisy measurements $y = Ax + e$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} x \\ \\ \end{bmatrix} + \begin{bmatrix} e \end{bmatrix}$$

4. Assume x is s-sparse: $\|x\|_0 \stackrel{\text{def}}{=} |\operatorname{supp}(x)| \leq s \ll d$.
5. sparsity, measurements, dimension

RIP

Restricted Isometry Property (RIP)

- A satisfies the restricted isometry property (RIP) with parameters (s, δ) (or with RIC δ_s) if

$$(1-\delta)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta)\|x\|_2^2 \quad \text{whenever } \|x\|_0 \le s.$$

- For Gaussian or Bernoulli measurement matrices, with high probability

$$\delta \le c < 1 \quad \text{when } m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property.

RIP

Restricted Isometry Property (RIP)

- A satisfies the restricted isometry property (RIP) with parameters (s, δ) (or with RIC δ_s) if

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \quad \text{whenever } \|x\|_0 \leq s.$$

- For Gaussian or Bernoulli measurement matrices, with high probability

$$\delta \leq c < 1 \quad \text{when } m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property.

Methods

The literature has provided us with many algorithms for recovery. One of these is ℓ_1-minimization:

$$x^\star = \operatorname{argmin} ||z||_1 \quad \text{such that} \quad \|Az - y\|_2 \leq \gamma,$$

where $\|e\|_2 \leq \gamma$.

L1-Minimization [Candès-Romberg-Tao]

Assume that the measurement matrix A satisfies the RIP with parameters $(3s, 0.2)$. Then the reconstructed signal $x^\star$ satisfies:

$$\|x^\star - x\|_2 \leq C \frac{\|x - x_s\|_1}{\sqrt{s}} + C\gamma.$$

The sharpest result is due to Foucart who shows the above holds with RIP parameters $(2s, 0.4652)$.

Methods

The literature has provided us with many algorithms for recovery. One of these is ℓ_1-minimization:

$$x^\star = \operatorname{argmin} ||z||_1 \quad \text{such that} \quad \|Az - y\|_2 \leq \gamma,$$

where $\|e\|_2 \leq \gamma$.

L1-Minimization [Candès-Romberg-Tao]

Assume that the measurement matrix A satisfies the RIP with parameters $(3s, 0.2)$. Then the reconstructed signal $x^\star$ satisfies:

$$\|x^\star - x\|_2 \leq C\frac{\|x - x_s\|_1}{\sqrt{s}} + C\gamma.$$

The sharpest result is due to Foucart who shows the above holds with RIP parameters $(2s, 0.4652)$.

Methods

The literature has provided us with many algorithms for recovery. One of these is ℓ_1-minimization:

$$x^\star = \operatorname{argmin} ||z||_1 \quad \text{such that} \quad \|Az - y\|_2 \leq \gamma,$$

where $\|e\|_2 \leq \gamma$.

L1-Minimization [Candès-Romberg-Tao]

Assume that the measurement matrix A satisfies the RIP with parameters $(3s, 0.2)$. Then the reconstructed signal $x^\star$ satisfies:

$$\|x^\star - x\|_2 \leq C\frac{\|x - x_s\|_1}{\sqrt{s}} + C\gamma.$$

The sharpest result is due to Foucart who shows the above holds with RIP parameters $(2s, 0.4652)$.

Methods

There are also greedy algorithms, which provide reconstruction guarantees and are often faster. For example, we have Compressive Sampling Matching Pursuit (N.-Tropp):

CoSaMP:

Initialize: $a = 0$, $v = y$

Signal Proxy: $u = A^*v$, $\Omega = \text{supp}(u_{2s})$, $T = \Omega \cup \text{supp}(a)$

Signal Estimation: $w|_T = A_T^\dagger y$, $w|_{T^c} = 0$

Prune: $a = w_s$

Sample Update: $v = y - Aa$

Methods

There are also greedy algorithms, which provide reconstruction guarantees and are often faster. For example, we have Compressive Sampling Matching Pursuit (N.-Tropp):

CoSaMP:

Initialize: $a = 0$, $v = y$

Signal Proxy: $u = A^* v$, $\Omega = \text{supp}\,(u_{2s})$, $T = \Omega \cup \text{supp}\,(a)$

Signal Estimation: $w|_T = A_T^{\dagger} y$, $w|_{T^c} = 0$

Prune: $a = w_s$

Sample Update: $v = y - Aa$

Guarantees

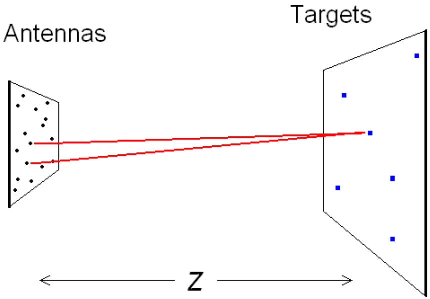
Theorem [N.-Tropp]:

For any measurement matrix satisfying the RIP with parameters $(2s, 0.1)$, the reconstructed signal $x^\sharp$ from its noisy measurements $y = Ax + e$ in at most $6s$ iterations:

$$\|x^\sharp - x\|_2 \leq C\left(\|e\|_2 + \frac{\|x - x_s\|_1}{\sqrt{s}}\right).$$

Applications of CS

Some of the many applications of compressed sensing physically implement the encoding matrix in a sensor. For example in remote sensing we have:



Applications of CS

The exact Green's function for the Helmholtz equation for monochromatic waves is

$$G^{ex}(\mathbf{a},\mathbf{r}) = \frac{\varepsilon^{i\omega|\mathbf{r}-\mathbf{a}|}}{4\pi|\mathbf{r}-\mathbf{a}|}, \qquad \mathbf{a} = (0,\xi,\eta), \quad \mathbf{r} = (z_0, x, y).$$

The paraxial approximation to Green's function is given by

$$G^{par}(\mathbf{a},\mathbf{r}) = \frac{\varepsilon^{i\omega z_0}}{4\pi z_0}\varepsilon^{i\omega|x-\xi|^2/(2z_0)}\varepsilon^{i\omega|y-\eta|^2/(2z_0)}.$$

Then the encoding matrix A is given by

$$A_{ij} = G^{ex}(\mathbf{a}_i,\mathbf{r}_j), \qquad i\text{th antenna}, \; j\text{th target location}.$$

and the decoding matrix,

$$\Phi_{ij} = G^{par}(\mathbf{a}_i,\mathbf{r}_j).$$

Applications of CS

The exact Green's function for the Helmholtz equation for monochromatic waves is

$$G^{\text{ex}}(\mathbf{a},\mathbf{r}) \;=\; \frac{\varepsilon^{\mathrm{i}\omega|\mathbf{r}-\mathbf{a}|}}{4\pi|\mathbf{r}-\mathbf{a}|}, \qquad \mathbf{a}=(0,\xi,\eta), \quad \mathbf{r}=(z_0,x,y).$$

The paraxial approximation to Green's function is given by

$$G^{\text{par}}(\mathbf{a},\mathbf{r}) \;=\; \frac{\varepsilon^{\mathrm{i}\omega z_0}}{4\pi z_0}\,\varepsilon^{\mathrm{i}\omega|x-\xi|^2/(2z_0)}\,\varepsilon^{\mathrm{i}\omega|y-\eta|^2/(2z_0)}.$$

Then the encoding matrix A is given by

$$A_{ij} \;=\; G^{\text{ex}}(\mathbf{a}_i,\mathbf{r}_j), \qquad i\text{th antenna}, \;\; j\text{th target location}.$$

and the decoding matrix,

$$\Phi_{ij} \;=\; G^{\text{par}}(\mathbf{a}_i,\mathbf{r}_j).$$

Applications of CS

The exact Green's function for the Helmholtz equation for monochromatic waves is

$$G^{\text{ex}}(\mathbf{a},\mathbf{r}) \;=\; \frac{\varepsilon^{i\omega|\mathbf{r}-\mathbf{a}|}}{4\pi|\mathbf{r}-\mathbf{a}|}, \qquad \mathbf{a}=(0,\xi,\eta), \quad \mathbf{r}=(z_0,x,y).$$

The paraxial approximation to Green's function is given by

$$G^{\text{par}}(\mathbf{a},\mathbf{r}) \;=\; \frac{\varepsilon^{i\omega z_0}}{4\pi z_0}\,\varepsilon^{i\omega|x-\xi|^2/(2z_0)}\,\varepsilon^{i\omega|y-\eta|^2/(2z_0)}.$$

Then the encoding matrix A is given by

$$A_{ij} \;=\; G^{\text{ex}}(\mathbf{a}_i,\mathbf{r}_j), \qquad i\text{th antenna}, \;\; j\text{th target location}.$$

and the decoding matrix,

$$\Phi_{ij} \;=\; G^{\text{par}}(\mathbf{a}_i,\mathbf{r}_j).$$

Applications of CS

The exact Green's function for the Helmholtz equation for monochromatic waves is

$$G^{\mathrm{ex}}(\mathbf{a},\mathbf{r}) = \frac{\varepsilon^{\mathrm{i}\omega|\mathbf{r}-\mathbf{a}|}}{4\pi|\mathbf{r}-\mathbf{a}|}, \qquad \mathbf{a} = (0,\xi,\eta), \quad \mathbf{r} = (z_0, x, y).$$

The paraxial approximation to Green's function is given by

$$G^{\mathrm{par}}(\mathbf{a},\mathbf{r}) = \frac{\varepsilon^{\mathrm{i}\omega z_0}}{4\pi z_0}\, \varepsilon^{\mathrm{i}\omega|x-\xi|^2/(2z_0)}\, \varepsilon^{\mathrm{i}\omega|y-\eta|^2/(2z_0)}.$$

Then the encoding matrix A is given by

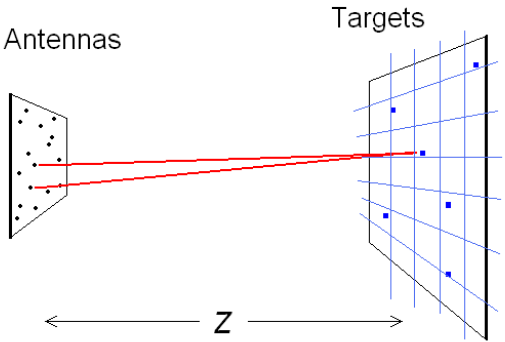
$$A_{ij} = G^{\mathrm{ex}}(\mathbf{a}_i,\mathbf{r}_j), \qquad i\text{th antenna}, \ \ j\text{th target location}.$$

and the decoding matrix,

$$\Phi_{ij} = G^{\mathrm{par}}(\mathbf{a}_i,\mathbf{r}_j).$$

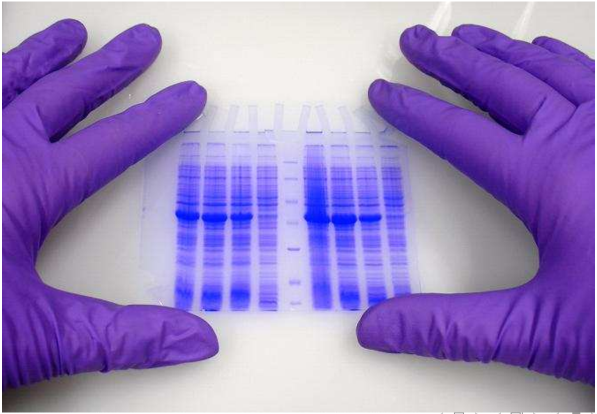
Applications of CS

We may also assume the targets like on some lattice, inducing error into the sensing matrix.



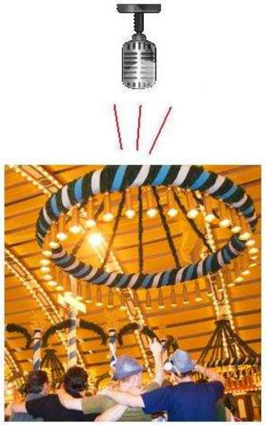
Applications of CS

Another example - Screening for genetic disorders using DNA samples. Error is introduced into the sensing matrix from human handling when pipetting the DNA samples.



Applications of CS

Another example - For source separation there are errors in estimating the mixing matrix.



Applications of CS

Another example - We may even encounter very small corruptions in the measurement matrix from its storage throughout time in memory.

Mixed Operators

Framework

We will now consider the framework in which we encode with one matrix A and decode with a possibly different matrix Φ. This yields a completely perturbed system that allows for *additive* error as well as *multiplicative* error.

Q: Why not simply treat the multiplicative noise in the same way as the additive noise?

A: These type of errors are *fundamentally* different. Increasing the strength of the signal will not reduce the signal to noise ratio in the multiplicative case.

Goals: How does this affect reconstruction error? How different can the two matrices be?

Mixed Operators

Framework

We will now consider the framework in which we encode with one matrix A and decode with a possibly different matrix Φ. This yields a completely perturbed system that allows for *additive* error as well as *multiplicative* error.

Q: Why not simply treat the multiplicative noise in the same way as the additive noise?

A: These type of errors are *fundamentally* different. Increasing the strength of the signal will not reduce the signal to noise ratio in the multiplicative case.

Goals: How does this affect reconstruction error? How different can the two matrices be?

Mixed Operators

Framework

We will now consider the framework in which we encode with one matrix A and decode with a possibly different matrix Φ. This yields a completely perturbed system that allows for *additive* error as well as *multiplicative* error.

Q: Why not simply treat the multiplicative noise in the same way as the additive noise?

A: These type of errors are *fundamentally* different. Increasing the strength of the signal will not reduce the signal to noise ratio in the multiplicative case.

Goals: How does this affect reconstruction error? How different can the two matrices be?

Mixed Operators

Framework

We will now consider the framework in which we encode with one matrix A and decode with a possibly different matrix Φ. This yields a completely perturbed system that allows for *additive* error as well as *multiplicative* error.

Q: Why not simply treat the multiplicative noise in the same way as the additive noise?

A: These type of errors are *fundamentally* different. Increasing the strength of the signal will not reduce the signal to noise ratio in the multiplicative case.

Goals: How does this affect reconstruction error? How different can the two matrices be?

Quantities & Assumptions

Quantities

- Sparsity: $\alpha_s = \frac{\|x-x_s\|_2}{\|x_s\|_2}$, $\beta_s = \frac{\|x-x_s\|_1}{\sqrt{s}\|x_s\|_2}$
- Perturbations: $\varepsilon_A^{(s)} = \frac{\|A-\Phi\|_2^{(s)}}{\|A\|_2^{(s)}}$, $\varepsilon_A = \frac{\|A-\Phi\|_2}{\|A\|_2}$, $\varepsilon = \|A - \Phi\|_2$
- RIP Ratios: $\kappa_A = \frac{\sqrt{1+\delta_s}}{\sqrt{1-\delta_s}}$, $\gamma_A = \frac{\|A\|_2}{\sqrt{1-\delta_s}}$

Assumptions

- RIP on A: $\delta_{2s} < \frac{\sqrt{2}}{\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2} - 1$
- Sparsity: $\alpha_s + \beta_s < \frac{1}{\kappa_{\mathbf{A}}^{(s)}}$

Quantities & Assumptions

Quantities

- Sparsity: $\alpha_s = \frac{\|x - x_s\|_2}{\|x_s\|_2}$, $\beta_s = \frac{\|x - x_s\|_1}{\sqrt{s}\|x_s\|_2}$
- Perturbations: $\varepsilon_A^{(s)} = \frac{\|A - \Phi\|_2^{(s)}}{\|A\|_2^{(s)}}$, $\varepsilon_A = \frac{\|A - \Phi\|_2}{\|A\|_2}$, $\varepsilon = \|A - \Phi\|_2$
- RIP Ratios: $\kappa_A = \frac{\sqrt{1+\delta_s}}{\sqrt{1-\delta_s}}$, $\gamma_A = \frac{\|A\|_2}{\sqrt{1-\delta_s}}$

Assumptions

- RIP on A: $\delta_{2s} < \frac{\sqrt{2}}{\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2} - 1$
- Sparsity: $\alpha_s + \beta_s < \frac{1}{\kappa_{\mathbf{A}}^{(s)}}$

Quantities & Assumptions

Quantities

- Sparsity: $\alpha_s = \frac{\|x - x_s\|_2}{\|x_s\|_2}$, $\beta_s = \frac{\|x - x_s\|_1}{\sqrt{s}\|x_s\|_2}$
- Perturbations: $\varepsilon_A^{(s)} = \frac{\|A - \Phi\|_2^{(s)}}{\|A\|_2^{(s)}}$, $\varepsilon_A = \frac{\|A - \Phi\|_2}{\|A\|_2}$, $\varepsilon = \|A - \Phi\|_2$
- RIP Ratios: $\kappa_A = \frac{\sqrt{1+\delta_s}}{\sqrt{1-\delta_s}}$, $\gamma_A = \frac{\|A\|_2}{\sqrt{1-\delta_s}}$

Assumptions

- RIP on A: $\delta_{2s} < \frac{\sqrt{2}}{\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2} - 1$
- Sparsity: $\alpha_s + \beta_s < \frac{1}{\kappa_{\mathbf{A}}^{(s)}}$

Quantities & Assumptions

Quantities

- Sparsity: $\alpha_s = \frac{\|x - x_s\|_2}{\|x_s\|_2}$, $\beta_s = \frac{\|x - x_s\|_1}{\sqrt{s}\|x_s\|_2}$
- Perturbations: $\varepsilon_A^{(s)} = \frac{\|A - \Phi\|_2^{(s)}}{\|A\|_2^{(s)}}$, $\varepsilon_A = \frac{\|A - \Phi\|_2}{\|A\|_2}$, $\varepsilon = \|A - \Phi\|_2$
- RIP Ratios: $\kappa_A = \frac{\sqrt{1+\delta_s}}{\sqrt{1-\delta_s}}$, $\gamma_A = \frac{\|A\|_2}{\sqrt{1-\delta_s}}$

Assumptions

- RIP on A: $\delta_{2s} < \frac{\sqrt{2}}{\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2} - 1$
- Sparsity: $\alpha_s + \beta_s < \frac{1}{\kappa_{\mathbf{A}}^{(s)}}$

Quantities & Assumptions

Quantities

- Sparsity: $\alpha_s = \frac{\|x-x_s\|_2}{\|x_s\|_2}$, $\beta_s = \frac{\|x-x_s\|_1}{\sqrt{s}\|x_s\|_2}$
- Perturbations: $\varepsilon_A^{(s)} = \frac{\|A-\Phi\|_2^{(s)}}{\|A\|_2^{(s)}}$, $\varepsilon_A = \frac{\|A-\Phi\|_2}{\|A\|_2}$, $\varepsilon = \|A - \Phi\|_2$
- RIP Ratios: $\kappa_A = \frac{\sqrt{1+\delta_s}}{\sqrt{1-\delta_s}}$, $\gamma_A = \frac{\|A\|_2}{\sqrt{1-\delta_s}}$

Assumptions

- RIP on A: $\delta_{2s} < \frac{\sqrt{2}}{\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2} - 1$
- Sparsity: $\alpha_s + \beta_s < \frac{1}{\kappa_{\mathbf{A}}^{(s)}}$

Results

Theorem [Herman-Strohmer]

Let x be an arbitrary signal with measurements $b = Ax$, corrupted with noise to form $y = Ax + e$. Set the total noise parameter

$$\varepsilon_{\mathbf{A},s,\mathbf{b}} := \left(\frac{\varepsilon_{\mathbf{A}}^{(s)} \kappa_{\mathbf{A}} + \varepsilon_{\mathbf{A}} \gamma_{\mathbf{A}} \alpha_s}{1 - \kappa_{\mathbf{A}}(\alpha_s + \beta_s)} \right) \|\mathbf{b}\|_2 + \|e\|_2.$$

Then under the above assumptions, the ℓ_1-reconstruction $x^\star$ using matrix Φ and noisy measurements $y = b + e$ satisfies

$$\|\mathbf{z}^\star - \mathbf{x}\|_2 \;\leq\; \frac{C_0}{\sqrt{s}} \|\mathbf{x} - \mathbf{x}_s\|_1 \;+\; C_1 \varepsilon_{\mathbf{A},s,\mathbf{b}}.$$

Numerical Results

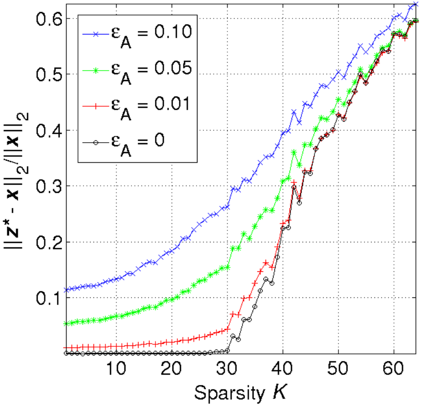


Figure: [“General Deviants: An Analysis of Perturbations in Compressed Sensing,” Herman, Strohmer ’09] (m=128, d=512)

Numerical Results

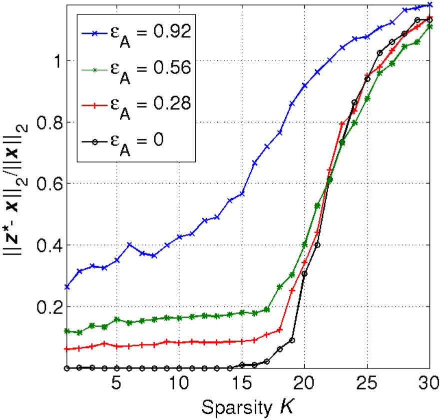


Figure: Simulation of remote sensing results.

Results

Theorem [Herman-N.]

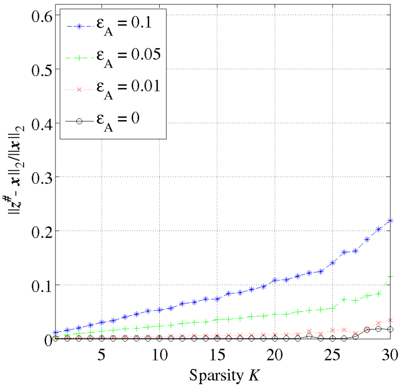
Let A be a measurement matrix with RIC

$$\delta_{4s} \leq \frac{1.1}{(1+\varepsilon_{\mathbf{A}}^{(4s)})^2} - 1.$$

Let x be an arbitrary signal with measurements $b = Ax$, corrupted with noise to form $y = Ax + e$. Then under similar assumptions, the reconstruction $x^\sharp$ using matrix Φ from CoSaMP satisfies

$$\|x^\sharp - x\|_2 \leq C \cdot \left(\|x - x_s\|_2 + \frac{\|x - x_s\|_1}{\sqrt{s}} + (\varepsilon \alpha_s + \varepsilon^{(s)})\|b\|_2 + \|e\|_2 \right).$$

Numerical Results



[“Mixed Operators in Compressed Sensing,” Herman, N. ’10] (m=128, d=512)

Summary

Conclusions

- Important to consider perturbations in the signal, measurements, and measurement matrices for applications of CS
- Stability of ℓ_1 and CoSaMP is a **linear** function of the perturbations $\|\mathbf{A} - \mathbf{\Phi}\|_2, \|\mathbf{e}\|_2$
- This type of analysis may lead to better strategies to minimize recovery error in particular applications

Summary

Conclusions

- Important to consider perturbations in the signal, measurements, and measurement matrices for applications of CS
- Stability of ℓ_1 and CoSaMP is a **linear** function of the perturbations $\|\mathbf{A} - \mathbf{\Phi}\|_2, \|\mathbf{e}\|_2$
- This type of analysis may lead to better strategies to minimize recovery error in particular applications

Summary

Conclusions

- Important to consider perturbations in the signal, measurements, and measurement matrices for applications of CS
- Stability of ℓ_1 and CoSaMP is a **linear** function of the perturbations $\|\mathbf{A} - \mathbf{\Phi}\|_2, \|\mathbf{e}\|_2$
- This type of analysis may lead to better strategies to minimize recovery error in particular applications

For more information

E-mail:

- dneedell@stanford.edu
- mattyh@math.ucla.edu

Web:

- www-stat.stanford.edu/~dneedell
- www.math.ucla.edu/~mattyh/

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):12071223, 2006.
- M. A. Herman and T. Strohmer. General Deviants: An analysis of perturbations in compressed sensing. IEEE Journal of Selected Topics in Sig. Proc.: Special Issue on Compressive Sensing, 4(2), Apr. 2010.
- M. A. Herman and D. Needell. Mixed Operators in Compressed Sensing. Proc. of CISS 2010.