Compressed sensing with coherent and redundant dictionaries

Deanna Needell

Stanford University
Joint work with E. J. Candès and Y. C. Eldar

ICNAAM Rhodes, Greece, Sept. 2010

Background to CS

The issue

Traditional data acquisition is wasteful.

The idea

Combine acquisition and compression.

The solution

Compressed sensing allows us to do this $\rightarrow$ reconstruct a signal from its (compressed) measurements.

Background to CS

The issue

Traditional data acquisition is wasteful.

The idea

Combine acquisition and compression.

The solution

Compressed sensing allows us to do this $\rightarrow$ reconstruct a signal from its (compressed) measurements.

Background to CS

The issue

Traditional data acquisition is wasteful.

The idea

Combine acquisition and compression.

The solution

Compressed sensing allows us to do this $\rightarrow$ reconstruct a signal from its (compressed) measurements.

Notation

1. f is an unknown signal in $\mathbb{R}^n$.
2. Measurement matrix $A : \mathbb{R}^n \to \mathbb{R}^m$.
3. Noisy measurements $y = Af + z$.

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

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

Notation

1. f is an unknown signal in $\mathbb{R}^n$.
2. Measurement matrix $A : \mathbb{R}^n \to \mathbb{R}^m$.
3. Noisy measurements $y = Af + z$.

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

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

Notation

1. f is an unknown signal in $\mathbb{R}^n$.
2. Measurement matrix $A : \mathbb{R}^n \to \mathbb{R}^m$.
3. Noisy measurements $y = Af + z$.

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

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

Notation

1. f is an unknown signal in $\mathbb{R}^n$.
2. Measurement matrix $A : \mathbb{R}^n \to \mathbb{R}^m$.
3. Noisy measurements $y = Af + z$.

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

4. Problem: Reconstruct signal f from measurements y

Is this possible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension n suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll n$.
- With respect to some other basis: $f = Dx$ where $\|x\|_0 \leq s \ll n$.

In practice, we encounter compressible signals.

Is this possible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension n suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll n$.
- With respect to some other basis: $f = Dx$ where $\|x\|_0 \leq s \ll n$.

In practice, we encounter compressible signals.

Is this possible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension n suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\mathrm{supp}(f)| \leq s \ll n$.
- With respect to some other basis: $f = Dx$ where $\|x\|_0 \leq s \ll n$.

In practice, we encounter compressible signals.

Is this possible?

Without further assumptions, this problem is ill-posed.

Why will this work?

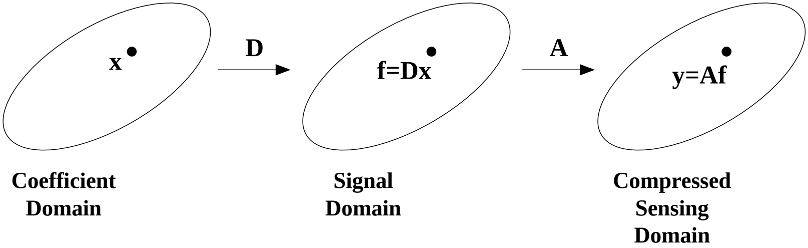
Most signals of interest contain far less information than their dimension n suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\mathrm{def}}{=} |\mathrm{supp}(f)| \leq s \ll n$.
- With respect to some other basis: $f = Dx$ where $\|x\|_0 \leq s \ll n$.

In practice, we encounter compressible signals.

The CS Process



Conditions on A?

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

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

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

Conditions on A?

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

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

- 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:

$$x^\star = \operatorname{argmin} ||w||_1 \quad \text{such that} \quad \|Aw - y\|_2 \leq \varepsilon,$$

L1-Minimization [Candès-Romberg-Tao]

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

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

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:

$$x^\star = \operatorname{argmin} ||w||_1 \quad \text{such that} \quad \|Aw - y\|_2 \leq \varepsilon,$$

L1-Minimization [Candès-Romberg-Tao]

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

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

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:

$$x^\star = \operatorname{argmin} ||w||_1 \quad \text{such that} \quad \|Aw - y\|_2 \leq \varepsilon,$$

L1-Minimization [Candès-Romberg-Tao]

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

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

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

The News

Good News

Many methods hold for signals f which are sparse in the coordinate basis or in some other orthonormal basis (ONB).

Bad News

There are many applications for which the signal f is sparse not in an ONB, but in some overcomplete dictionary! This means that $f = Dx$ where D is a redundant dictionary. When D is not an ONB, AD is not at all likely to satisfy the RIP (or be incoherent).

The News

Good News

Many methods hold for signals f which are sparse in the coordinate basis or in some other orthonormal basis (ONB).

Bad News

There are many applications for which the signal f is sparse not in an ONB, but in some overcomplete dictionary! This means that $f = Dx$ where D is a redundant dictionary. When D is not an ONB, AD is not at all likely to satisfy the RIP (or be incoherent).

The News

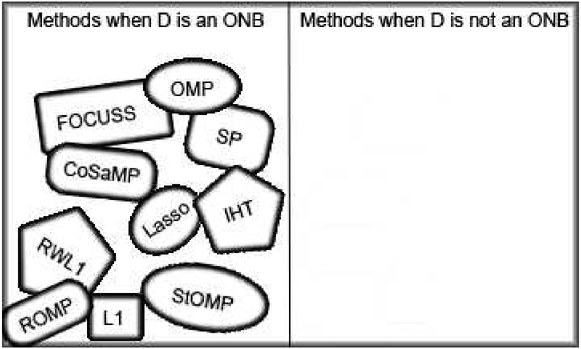
Good News

Many methods hold for signals f which are sparse in the coordinate basis or in some other orthonormal basis (ONB).

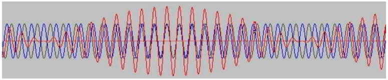
Bad News

There are many applications for which the signal f is sparse not in an ONB, but in some overcomplete dictionary! This means that $f = Dx$ where D is a redundant dictionary. When D is not an ONB, AD is not at all likely to satisfy the RIP (or be incoherent).

Big Picture

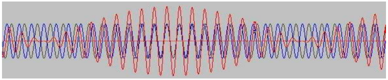

Methods when D is an ONB
Methods when D is not an ONB
OMP
FOCUSS
SP
CoSaMP
Lasso
IHT
RWL1
ROMP
L1
StOMP

Example: Oversampled DFT



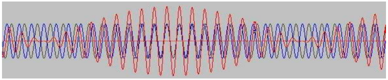
- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi i k t/n}$
- Sparse in the DFT = superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT: frequencies may be over even smaller intervals or intervals of varying length.
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ current CS does not apply.

Example: Oversampled DFT



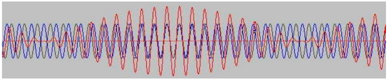
- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi i k t/n}$
- Sparse in the DFT $=$ superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT: frequencies may be over even smaller intervals or intervals of varying length.
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ current CS does not apply.

Example: Oversampled DFT



- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi ikt/n}$
- Sparse in the DFT $=$ superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT: frequencies may be over even smaller intervals or intervals of varying length.
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ current CS does not apply.

Example: Oversampled DFT



- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi i k t/n}$
- Sparse in the DFT $=$ superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT: frequencies may be over even smaller intervals or intervals of varying length.
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ current CS does not apply.

Example: Gabor frames

- Gabor frame: $G_k(t) = g(t - k_2 a)e^{2\pi i k_1 bt}$
- Radar, sonar, and imaging system applications use Gabor frames and wish to recover signals in this basis.
- Then D is an overcomplete frame with possibly highly coherent columns $\rightarrow$ current CS does not apply.

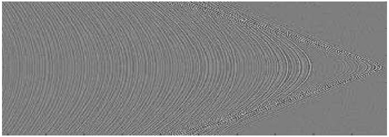
Example: Gabor frames

- Gabor frame: $G_k(t) = g(t - k_2 a)e^{2\pi i k_1 bt}$
- Radar, sonar, and imaging system applications use Gabor frames and wish to recover signals in this basis.
- Then D is an overcomplete frame with possibly highly coherent columns $\rightarrow$ current CS does not apply.

Example: Gabor frames

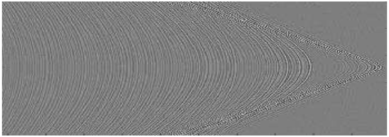
- Gabor frame: $G_k(t) = g(t - k_2 a)e^{2\pi i k_1 bt}$
- Radar, sonar, and imaging system applications use Gabor frames and wish to recover signals in this basis.
- Then D is an overcomplete frame with possibly highly coherent columns $\rightarrow$ current CS does not apply.

Example: Curvelet frames



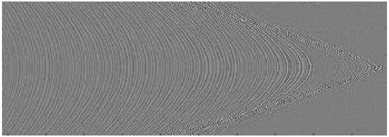
- A Curvelet frame has some properties of an ONB but is overcomplete.
- Curvelets approximate well the curved singularities in images and are thus used widely in image processing.
- Again, this means D is an overcomplete dictionary $\rightarrow$ current CS does not apply.

Example: Curvelet frames



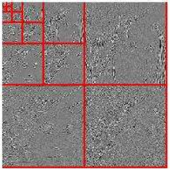
- A Curvelet frame has some properties of an ONB but is overcomplete.
- Curvelets approximate well the curved singularities in images and are thus used widely in image processing.
- Again, this means D is an overcomplete dictionary $\rightarrow$ current CS does not apply.

Example: Curvelet frames



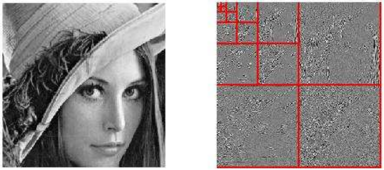
- A Curvelet frame has some properties of an ONB but is overcomplete.
- Curvelets approximate well the curved singularities in images and are thus used widely in image processing.
- Again, this means D is an overcomplete dictionary $\rightarrow$ current CS does not apply.

Example: UWT



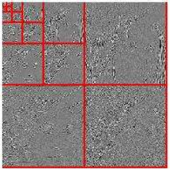
- The undecimated wavelet transform has a translation invariance property that is missing in the DWT.
- The UWT is overcomplete and this redundancy has been found to be helpful in image processing.
- Again, this means D is a redundant dictionary $\rightarrow$ current CS does not apply.

Example: UWT



- The undecimated wavelet transform has a translation invariance property that is missing in the DWT.
- The UWT is overcomplete and this redundancy has been found to be helpful in image processing.
- Again, this means D is a redundant dictionary $\rightarrow$ current CS does not apply.

Example: UWT



- The undecimated wavelet transform has a translation invariance property that is missing in the DWT.
- The UWT is overcomplete and this redundancy has been found to be helpful in image processing.
- Again, this means D is a redundant dictionary $\rightarrow$ current CS does not apply.

Example: Concatenations

$$\begin{bmatrix} I & F \end{bmatrix}$$

- In many applications, a signal may be sparse in several ONBs.
- Correlations between the bases mean current CS techniques do not apply.

Example: Concatenations

$$\begin{bmatrix} \\ I & \qquad F \\ \\ \end{bmatrix}$$

- In many applications, a signal may be sparse in several ONBs.
- Correlations between the bases mean current CS techniques do not apply.

ℓ_1-Analysis

Proposed Method

It has been observed (empirically) that ℓ_1-analysis often succeeds:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^n}{\operatorname{argmin}} \|D^*\tilde{f}\|_1 \quad \text{subject to} \quad \|A\tilde{f} - y\|_2 \leq \varepsilon.$$

Condition on A?

Let Σ_s be the union of all subspaces spanned by all subsets of s columns of D.

D-RIP

We say that the measurement matrix A obeys the *restricted isometry property adapted to D* (D-RIP) with constant δ_s if

$$(1-\delta_s)\|v\|_2^2 \leq \|Av\|_2^2 \leq (1+\delta_s)\|v\|_2^2$$

holds for all $v \in \Sigma_s$.

Similarly to the RIP, Guassian, subgaussian, and Bernoulli matrices satisfy the D-RIP with $m \approx s\log(d/s)$. Matrices with a fast multiply (DFT with random signs) also satisfy the D-RIP with m approximately of this order.

Condition on A?

Let Σ_s be the union of all subspaces spanned by all subsets of s columns of D.

D-RIP

We say that the measurement matrix A obeys the *restricted isometry property adapted to D* (D-RIP) with constant δ_s if

$$(1-\delta_s)\|v\|_2^2 \le \|Av\|_2^2 \le (1+\delta_s)\|v\|_2^2$$

holds for all $v \in \Sigma_s$.

Similarly to the RIP, Guassian, subgaussian, and Bernoulli matrices satisfy the D-RIP with $m \approx s\log(d/s)$. Matrices with a fast multiply (DFT with random signs) also satisfy the D-RIP with m approximately of this order.

Condition on A?

Let Σ_s be the union of all subspaces spanned by all subsets of s columns of D.

D-RIP

We say that the measurement matrix A obeys the *restricted isometry property adapted to D* (D-RIP) with constant δ_s if

$$(1-\delta_s)\|v\|_2^2 \le \|Av\|_2^2 \le (1+\delta_s)\|v\|_2^2$$

holds for all $v \in \Sigma_s$.

Similarly to the RIP, Guassian, subgaussian, and Bernoulli matrices satisfy the D-RIP with $m \approx s\log(d/s)$. Matrices with a fast multiply (DFT with random signs) also satisfy the D-RIP with m approximately of this order.

Main Result

Theorem

Let D be an arbitrary tight frame and let A be a measurement matrix satisfying D-RIP with $\delta_{2s} < 0.08$. Then the solution $\hat{f}$ to ℓ_1-analysis satisfies

$$\|\hat{f} - f\|_2 \leq C_0\varepsilon + C_1 \frac{\|D^*f - (D^*f)_s\|_1}{\sqrt{s}},$$

where the constants C_0 and C_1 may only depend on δ_{2s}.

Implications

In other words,

Our result says that ℓ_1-analysis is very accurate when D^*f has rapidly decaying coefficients. This is the case in applications using the Oversampled DFT, Gabor frames, Undecimated WT, and Curvelet frames (and many others).

This will not necessarily be the case when using concatenations of two ONBs $\rightarrow$ ℓ_1-analysis not the right method.

Implications

In other words,

Our result says that ℓ_1-analysis is very accurate when D^*f has rapidly decaying coefficients. This is the case in applications using the Oversampled DFT, Gabor frames, Undecimated WT, and Curvelet frames (and many others).

This will not necessarily be the case when using concatenations of two ONBs $\rightarrow$ ℓ_1-analysis not the right method.

Implications

In other words,

Our result says that ℓ_1-analysis is very accurate when D^*f has rapidly decaying coefficients. This is the case in applications using the Oversampled DFT, Gabor frames, Undecimated WT, and Curvelet frames (and many others).

This will not necessarily be the case when using concatenations of two ONBs $\rightarrow$ ℓ_1-analysis not the right method.

Experimental Setup

$n = 8192, m = 400, d = 491,520$
A: $m \times n$ Gaussian, D: $n \times d$ Gabor

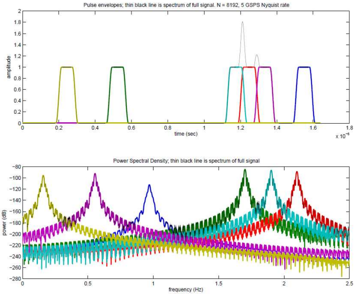


Figure: The signal is a superposition of 6 radar pulses, each of which being about 200 ns long, and with frequency carriers distributed between 50 MHz and 2.5 GHz (top plot). As can be seen, three of these pulses overlap in the time domain.

Experimental Results

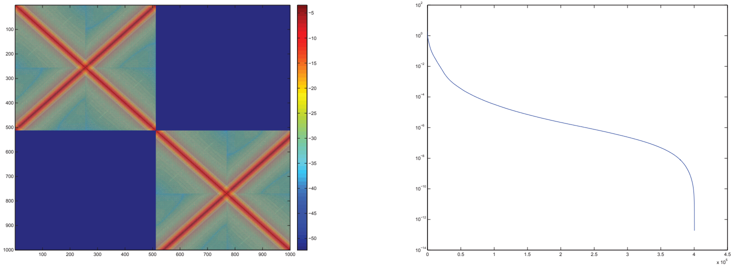


Figure: Portion of the matrix $D^* D$, in log-scale (left). Sorted analysis coefficients (in absolute value) of the signal from Figure 1 (right).

Experimental Results

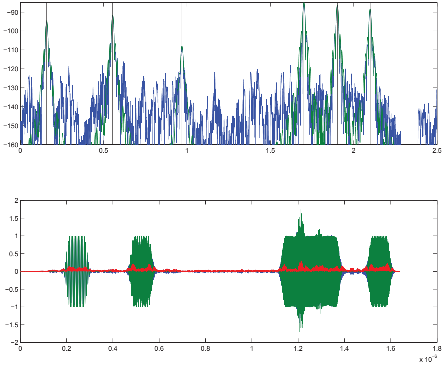


Figure: Recovery in both the time (below) and frequency (above) domains by ℓ_1-analysis. Blue denotes the recovered signal, green the actual signal, and red the difference between the two.

Experimental Results

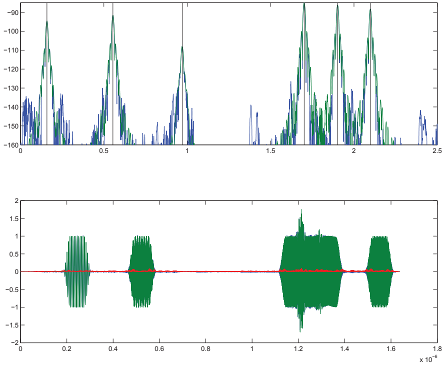


Figure: Recovery in both the time (below) and frequency (above) domains by ℓ_1-analysis after one reweighted iteration. Blue denotes the recovered signal, green the actual signal, and red the difference.

Experimental Results

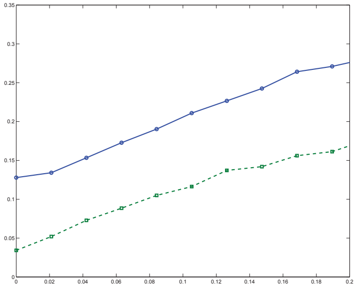


Figure: Relative recovery error of ℓ_1-analysis as a function of the (normalized) noise level, averaged over 5 trials. The solid line denotes standard ℓ_1-analysis, and the dashed line denotes ℓ_1-analysis with 3 reweighted iterations. The x-axis is the relative noise level $\sqrt{m}\sigma/\|Af\|_2$ while the y-axis is the relative error $\|\hat{f} - f\|_2/\|f\|_2$.

Experimental Results

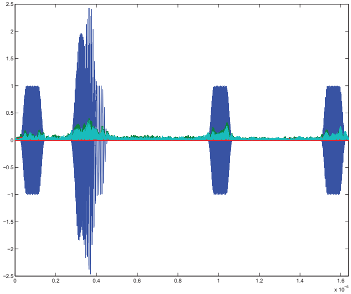


Figure: Relative error $\|\hat{f} - f\|_2/\|f\|_2$ of a compressible signal. Blue denotes the actual signal, while green, red, and cyan denote the recovery error from ℓ_1-analysis, reweighted ℓ_1-analysis, and ℓ_1-synthesis, respectively.

For more information

E-mail:

- dneedell@stanford.edu

Web:

- www-stat.stanford.edu/~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):12071223, 2006.
- E. J. Candès, Y. C. Eldar, and D. Needell. Compressed sensing with coherent and redundant dictionaries. Submitted for publication.