Robust image recovery via total-variation minimization

Deanna Needell

Claremont McKenna College

Joint work with Rachel Ward
Level Set Seminar, UCLA, July 2012

Collaborator

Joint work with Rachel Ward [Univ. of Texas, Austin]

D. Needell and R. Ward. Stable image reconstruction using total variation minimization, Submitted, Mar. 2012.

Outline

- Compressed Sensing (CS)
  - Applications
  - Mathematical Formulation & Methods
- Extensions to other dictionaries
  - Applications
  - Mathematical Formulation & Methods
- Imaging with CS
  - Theoretical possibilities
  - Total Variation
  - Empirical observations
  - New Results
  - Proof Sketch

The mathematical problem (notation)

1. Signal of interest $f \in \mathbb{C}^d (= \mathbb{C}^{N \times N})$
2. Measurement operator $\mathcal{A} : \mathbb{C}^d \to \mathbb{C}^m$.
3. Measurements $y = \mathcal{A}f + \xi$.

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

4. Problem: Reconstruct signal f from measurements y

The mathematical problem (notation)

1. Signal of interest $f \in \mathbb{C}^d (= \mathbb{C}^{N \times N})$
2. Measurement operator $\mathcal{A} : \mathbb{C}^d \to \mathbb{C}^m$.
3. Measurements $y = \mathcal{A}f + \xi$.

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

4. Problem: Reconstruct signal f from measurements y

The mathematical problem (notation)

1. Signal of interest $f \in \mathbb{C}^d (= \mathbb{C}^{N \times N})$
2. Measurement operator $\mathcal{A} : \mathbb{C}^d \to \mathbb{C}^m$.
3. Measurements $y = \mathcal{A}f + \xi$.

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

4. Problem: Reconstruct signal f from measurements y

The mathematical problem (notation)

1. Signal of interest $f \in \mathbb{C}^d (= \mathbb{C}^{N\times N})$
2. Measurement operator $\mathcal{A} : \mathbb{C}^d \to \mathbb{C}^m$.
3. Measurements $y = \mathcal{A}f + \xi$.

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

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

Sparsity

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

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

Assume f is sparse:

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

In practice, we encounter compressible signals.

Sparsity

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

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

Assume f is sparse:

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

In practice, we encounter compressible signals.

Sparsity

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

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

Assume f is sparse:

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

In practice, we encounter compressible signals.

Sparsity

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

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

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \le s \ll d$
- 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.

Sparsity

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

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

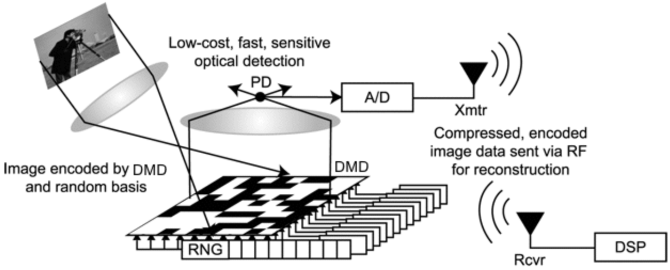
Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \le s \ll d$
- 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.

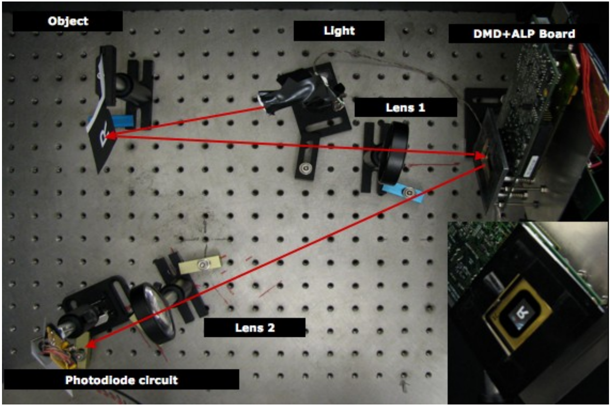
Digital Cameras

Save your pennies to buy the new digital camera?



Digital Cameras

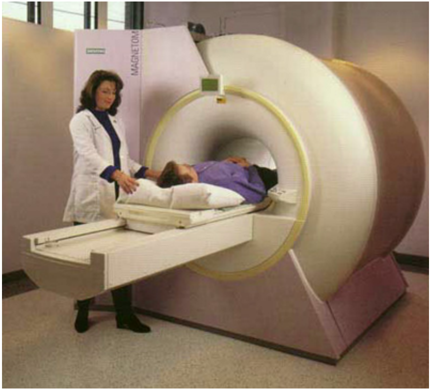
Save your pennies to buy the new digital camera?



MRI

Feeling claustrophobic?

It'll only last a quick 45 minutes...



MRI

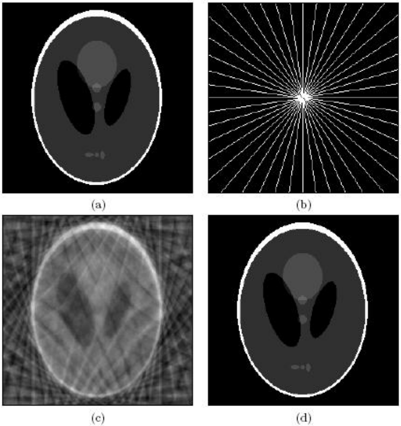
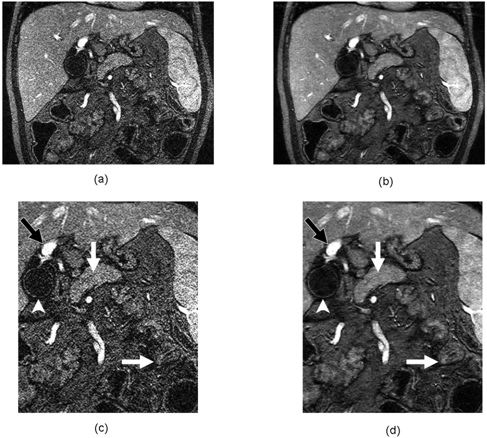


Figure 1: Example of a simple recovery problem. (a) The Logan-Shepp phantom test image. (b) Sampling domain Ω in the frequency plane; Fourier coefficients are sampled along 22 approximately radial lines. (c) Minimum energy reconstruction obtained by setting unobserved Fourier coefficients to zero. (d) Reconstruction obtained by minimizing the total variation, as in (1.1). The reconstruction is an exact replica of the image in (a).

Pediatric MRI



(a–d) Submillimeter near-isotropic–resolution contrast-enhanced T1-weighted MR images in 8-year-old boy. (a, c) Standard and (b, d) compressed sensing reconstruction images. (c, d) Zoomed images show improved delineation of the pancreatic duct (vertical arrow), bowel (horizontal arrow), and gallbladder wall (arrowhead), and equivalent definition of portal vein (black arrow) with L1 SPIR-iT reconstruction.

(Caffey Award : Faster Pediatric MRI Via Compressed Sensing - Shreyas Vasanawala et.al. (Stanford University))

Many more...

- Radar
- Error Correction
- Computational Biology (DNA Microarrays)
- Geophysical Data Analysis
- Data Mining, classification
- Neuroscience
- Imaging
- ...

Sparsity...

Sparsity in coordinate basis: f=x

Reconstructing the signal f from measurements y

ℓ_1-minimization [Candès-Romberg-Tao]

Let A satisfy the *Restricted Isometry Property* and set:

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

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.

Reconstructing the signal f from measurements y

ℓ_1-minimization [Candès-Romberg-Tao]

Let A satisfy the *Restricted Isometry Property* and set:

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

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.

Reconstructing the signal f from measurements y

ℓ_1-minimization [Candès-Romberg-Tao]

Let A satisfy the *Restricted Isometry Property* and set:

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

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.

Restricted Isometry Property

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

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

- 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: $m \gtrsim s \log^4 d$.

Restricted Isometry Property

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

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

- 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: $m \gtrsim s \log^4 d$.

Restricted Isometry Property

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

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

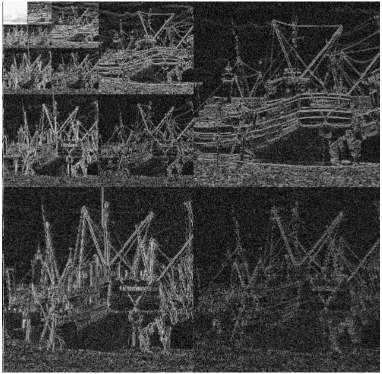
- Gaussian or Bernoulli measurement matrices satisfy the RIP with high probability when

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

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

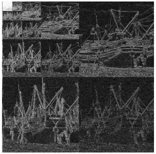
Sparsity...

In orthonormal basis: $f = Bx$



Natural Images

Images are compressible in *Wavelet bases*.



$$X = \sum_{j,k=1}^{N} c_{j,k} H_{j,k}, \qquad c_{j,k} = \langle X, H_{j,k} \rangle, \qquad \|X\|_2 = \|c\|_2,$$

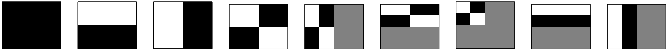


Figure: Haar basis functions

Wavelet transform is orthonormal and multi-scale. Sparsity level of image is higher on detail coefficients.

Natural images

Images are compressible in *Wavelet bases*.

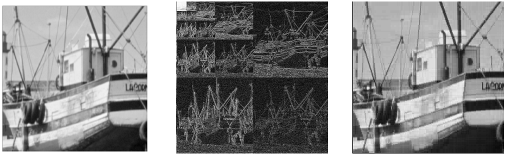


Figure: Boats image, 2D Haar transform, and compression from 10% of Haar coefficients

$$X = H^{-1}H(X) = \sum_{j,k=1}^{N} c_{j,k} H_{j,k}$$

X is s-sparse (in Haar basis) if $\|c\|_0 \leq s$

X_s^w is the best s-term approximation to X in Haar basis

Image compression: $X \rightarrow X_s^w$

Natural images

Images are compressible in *Wavelet bases*.

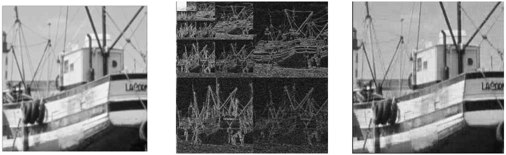


Figure: Boats image, 2D Haar transform, and compression from 10% of Haar coefficients

$$X = H^{-1}H(X) = \sum_{j,k=1}^{N} c_{j,k} H_{j,k}$$

X is s-sparse (in Haar basis) if $\|c\|_0 \leq s$

X_s^w is the best s-term approximation to X in Haar basis

Image compression: $X \to X_s^w$

Natural images

Images are compressible in *Wavelet bases*.

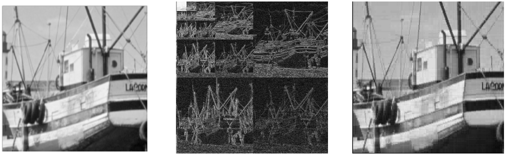


Figure: Boats image, 2D Haar transform, and compression from 10% of Haar coefficients

$$X = H^{-1}H(X) = \sum_{j,k=1}^{N} c_{j,k} H_{j,k}$$

X is s-sparse (in Haar basis) if $\|c\|_0 \le s$

X_s^w is the best s-term approximation to X in Haar basis

Image compression: $X \to X_s^w$

Natural images

Images are compressible in *Wavelet bases*.

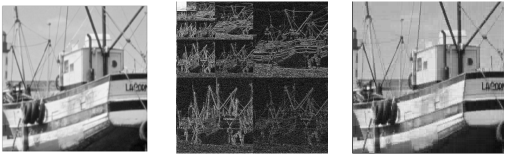


Figure: Boats image, 2D Haar transform, and compression from 10% of Haar coefficients

$$X = H^{-1}H(X) = \sum_{j,k=1}^{N} c_{j,k} H_{j,k}$$

X is s-sparse (in Haar basis) if $\|c\|_0 \leq s$

X_s^w is the best s-term approximation to X in Haar basis

Image compression: $X \to X_s^w$

Sparsity in orthonormal basis B

L1-minimization Method

For orthonormal basis B, $f = Bx$ with x sparse, one may solve the ℓ_1-minimization program:

$$\hat{f} = \operatorname*{argmin}_{\tilde{f} \in \mathbb{C}^n} \|B^{-1}\tilde{f}\|_1 \quad \text{subject to} \quad \|\mathcal{A}\tilde{f} - y\|_2 \leq \varepsilon.$$

Same results hold.

Sparsity in orthonormal basis B

L1-minimization Method

For orthonormal basis B, $f = Bx$ with x sparse, one may solve the ℓ_1-minimization program:

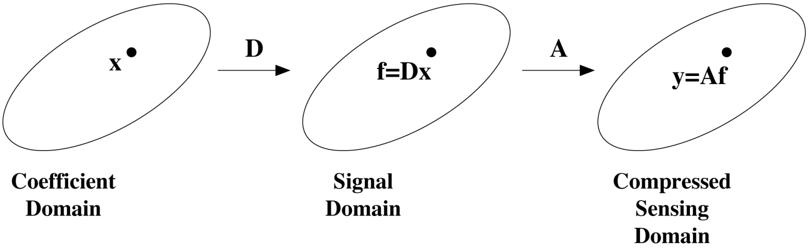
$$\hat{f} = \operatorname*{argmin}_{\tilde{f} \in \mathbb{C}^n} \|B^{-1}\tilde{f}\|_1 \quad \text{subject to} \quad \|\mathcal{A}\tilde{f} - y\|_2 \leq \varepsilon.$$

Same results hold.

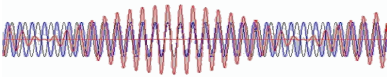
Sparsity...

In arbitrary dictionary: $f = Dx$

The CS Process

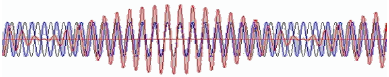


Example: Oversampled DFT



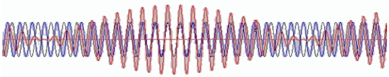
- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi ikt/n}$
- Sparse in the DFT $\rightarrow$ superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT:
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ conventional CS does not apply.

Example: Oversampled DFT



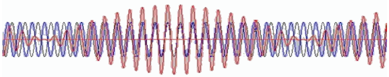
- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi ikt/n}$
- Sparse in the DFT $\rightarrow$ superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT:
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ conventional CS does not apply.

Example: Oversampled DFT



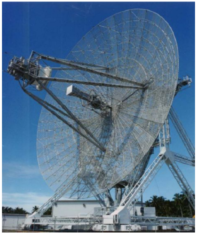
- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi ikt/n}$
- Sparse in the DFT $\rightarrow$ superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT:
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ conventional CS does not apply.

Example: Oversampled DFT



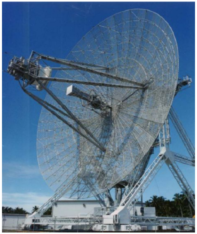
- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi ikt/n}$
- Sparse in the DFT $\rightarrow$ superpositions of sinusoids with frequencies in the lattice.
- Instead, use the oversampled DFT:
- Then D is an overcomplete frame with highly coherent columns $\rightarrow$ conventional CS does not apply.

Example: Gabor frames



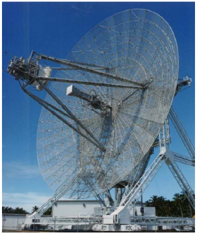
- Gabor frame: $G_k(t) = g(t - k_2 a)e^{2\pi i k_1 b t}$
- Radar, sonar, and imaging system applications use Gabor frames and wish to recover signals in this basis.
- Then D is an overcomplete frame with possibly highly coherent columns → conventional CS does not apply.

Example: Gabor frames



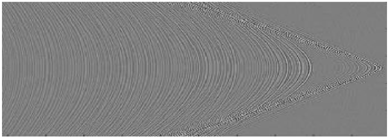
- Gabor frame: $G_k(t) = g(t - k_2 a)e^{2\pi i k_1 b t}$
- Radar, sonar, and imaging system applications use Gabor frames and wish to recover signals in this basis.
- Then D is an overcomplete frame with possibly highly coherent columns $\rightarrow$ conventional CS does not apply.

Example: Gabor frames



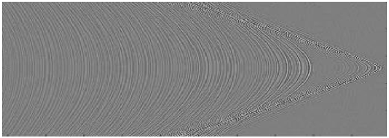
- Gabor frame: $G_k(t) = g(t - k_2 a)e^{2\pi i k_1 b t}$
- Radar, sonar, and imaging system applications use Gabor frames and wish to recover signals in this basis.
- Then D is an overcomplete frame with possibly highly coherent columns $\rightarrow$ conventional CS does not apply.

Example: Curvelet frames



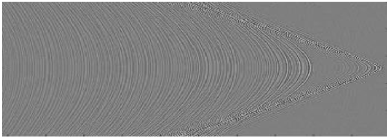
- A Curvelet frame has some properties of an ONB but is overcomplete.
- Curvelets approximate well the curved singularities in images and are thus used widely in image processing.
- Again, this means D is an overcomplete dictionary $\rightarrow$ conventional CS does not apply.

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 $\rightarrow$ conventional CS does not apply.

Example: Curvelet frames

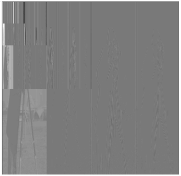


- A Curvelet frame has some properties of an ONB but is overcomplete.
- Curvelets approximate well the curved singularities in images and are thus used widely in image processing.
- Again, this means D is an overcomplete dictionary $\rightarrow$ conventional CS does not apply.

Example: UWT

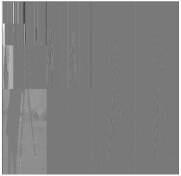
- The undecimated wavelet transform has a translation invariance property that is missing in the DWT.
- The UWT is overcomplete and this redundancy has been found to be helpful in image processing.
- Again, this means D is a redundant dictionary $\rightarrow$ conventional CS does not apply.

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 $\rightarrow$ conventional 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 $\rightarrow$ conventional CS does not apply.

ℓ_1-Synthesis Method

For arbitrary tight frame D, one may solve the ℓ_1-synthesis program:

$$\hat{f} = D\left(\operatorname*{argmin}_{\tilde{x}\in\mathbb{C}^n} \|\tilde{x}\|_1 \quad \text{subject to} \quad \|\mathcal{A}D\tilde{x} - y\|_2 \le \varepsilon\right).$$

Some work on this method [Candès et.al., Rauhut et.al., Elad et.al.,...]

ℓ_1-Synthesis Method

For arbitrary tight frame D, one may solve the ℓ_1-synthesis program:

$$\hat{f} = D\left(\underset{\tilde{x}\in\mathbb{C}^n}{\operatorname{argmin}} \|\tilde{x}\|_1 \quad \text{subject to} \quad \|\mathcal{A}D\tilde{x} - y\|_2 \leq \varepsilon\right).$$

Some work on this method [Candès et.al., Rauhut et.al., Elad et.al.,...]

ℓ_1-Analysis Method

For arbitrary tight frame D, one may solve the ℓ_1-analysis program:

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

Condition on A?

D-RIP

We say that the measurement matrix $\mathcal{A}$ obeys the *restricted isometry property adapted to D* (D-RIP) if there is $\delta < c$ such that

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

holds for all s-sparse x.

Similarly to the RIP, many classes of random matrices satisfy the D-RIP with $m \approx s\log(d/s)$. Randomly perturbed RIP matrices satisfy D-RIP [Krahmer-Ward '11], or similar property, Weibull matrices [Foucart '12], ...

Condition on A?

D-RIP

We say that the measurement matrix $\mathcal{A}$ obeys the *restricted isometry property adapted to D* (D-RIP) if there is $\delta < c$ such that

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

holds for all s-sparse x.

Similarly to the RIP, many classes of random matrices satisfy the D-RIP with $m \approx s\log(d/s)$. Randomly perturbed RIP matrices satisfy D-RIP [Krahmer-Ward '11], or similar property, Weibull matrices [Foucart '12], ...

Condition on A?

D-RIP

We say that the measurement matrix $\mathcal{A}$ obeys the *restricted isometry property adapted to D* (D-RIP) if there is $\delta < c$ such that

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

holds for all s-sparse x.

Similarly to the RIP, many classes of random matrices satisfy the D-RIP with $m \approx s\log(d/s)$. Randomly perturbed RIP matrices satisfy D-RIP [Krahmer-Ward '11], or similar property, Weibull matrices [Foucart '12], ...

CS with tight frame dictionaries

Theorem [Candès-Eldar-N-Randall]

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

$$\|\hat{f} - f\|_2 \lesssim \varepsilon + \frac{\|D^*f - (D^*f)_s\|_1}{\sqrt{s}}.$$

Implications

In other words,

This result says that ℓ_1-analysis is very accurate when $D^* f$ has rapidly decaying coefficients and D is a tight frame. This is the case in for example applications using the Oversampled DFT, Gabor frames, Undecimated WT, and Curvelet frames.

Implications

In other words,

This result says that ℓ_1-analysis is very accurate when D^*f has rapidly decaying coefficients and D is a tight frame. This is the case in for example applications using the Oversampled DFT, Gabor frames, Undecimated WT, and Curvelet frames.

Experimental Setup

$n = 8192, m = 400, d = 491,520$

A: $m \times n$ Gaussian, D: $n \times d$ Gabor

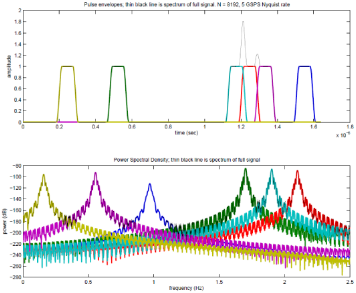


Figure: The signal is a superposition of 6 radar pulses, each of which being about 200 ns long, and with frequency carriers distributed between 50 MHz and 2.5 GHz (top plot). As can be seen, three of these pulses overlap in the time domain.

Experimental Results

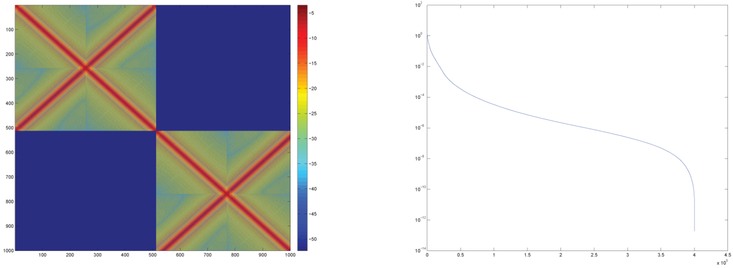


Figure: Portion of the matrix $D^* D$, in log-scale (left). Sorted analysis coefficients (in absolute value) of the signal from Figure 3 (right).

Experimental Results

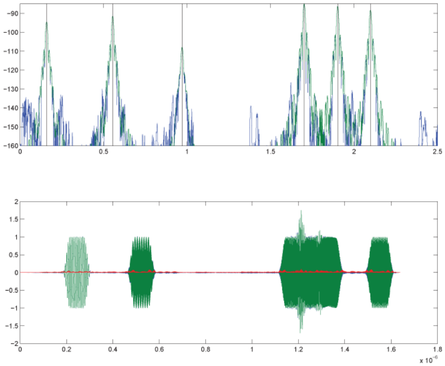


Figure: Recovery in both the time (below) and frequency (above) domains by ℓ_1-analysis. Blue denotes the recovered signal, green the actual signal, and red the difference between the two.

Implications

In other words,

This result says that ℓ_1-analysis is very accurate when D^*f has rapidly decaying coefficients and D is a tight frame. This is the case in for example applications using the Oversampled DFT, Gabor frames, Undecimated WT, and Curvelet frames.

Other sparsifying transforms for images?

Implications

In other words,

This result says that ℓ_1-analysis is very accurate when D^*f has rapidly decaying coefficients and D is a tight frame. This is the case in for example applications using the Oversampled DFT, Gabor frames, Undecimated WT, and Curvelet frames.

Other sparsifying transforms for images?

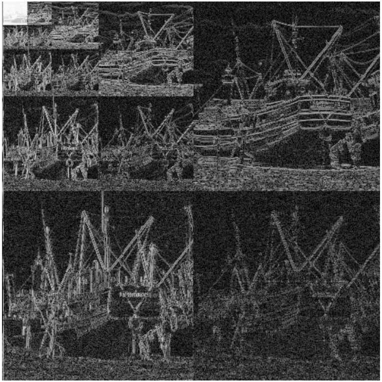
Natural images

Sparse...

256×256 “Boats” image

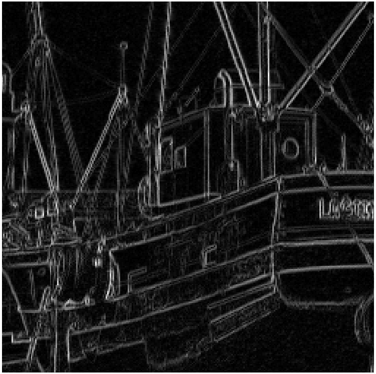
Natural images

Sparse wavelet representation...



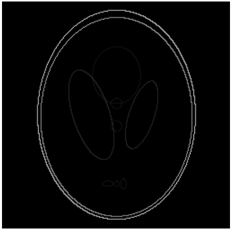
Natural images

Images are compressible in *discrete gradient*.



Natural images

Images are compressible in *discrete gradient*.



The discrete directional derivatives of an image $f \in \mathbb{C}^{N \times N}$ are

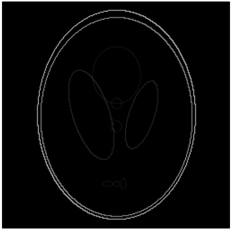
$$f_x : \mathbb{C}^{N \times N} \to \mathbb{C}^{(N-1) \times N}, \qquad (f_x)_{j,k} = f_{j,k} - f_{j-1,k},$$
$$f_y : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times (N-1)}, \qquad (f_y)_{j,k} = f_{j,k} - f_{j,k-1},$$

the discrete gradient operator is

$$\nabla[f] = (f_x, f_y)$$

Natural images

Images are compressible in *discrete gradient*.



The discrete directional derivatives of an image $f \in \mathbb{C}^{N\times N}$ are

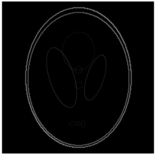
$$f_x : \mathbb{C}^{N\times N} \to \mathbb{C}^{(N-1)\times N}, \qquad (f_x)_{j,k} = f_{j,k} - f_{j-1,k},$$
$$f_y : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times (N-1)}, \qquad (f_y)_{j,k} = f_{j,k} - f_{j,k-1},$$

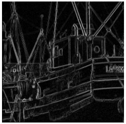
the discrete gradient operator is

$$\nabla[f] = (f_x, f_y)$$

Natural Notation

Images are compressible in *discrete gradient*.



- $\|f\|_p := \left(\sum_{j=1}^N \sum_{k=1}^N |f_{j,k}|^p\right)^{1/p}$
- f is s-sparse if $\|f\|_0 := \{\#(j,k) : f_{j,k} \neq 0\} \leq s$
- f_s is the best s-sparse approximation to f
- "Phantom": $\|\nabla[f]\|_0 = .03N^2$
- "Boats": $\|\nabla[f] - \nabla[f]_s\|_2$ decays quickly in s

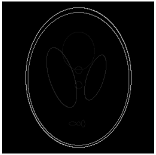
Natural Notation

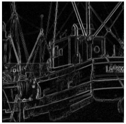
Images are compressible in *discrete gradient*.

- $\|f\|_p := \left(\sum_{j=1}^{N} \sum_{k=1}^{N} |f_{j,k}|^p \right)^{1/p}$
- f is s-sparse if $\|f\|_0 := \{\#(j,k) : f_{j,k} \neq 0\} \leq s$
- f_s is the best s-sparse approximation to f
- “Phantom”: $\|\nabla[f]\|_0 = .03N^2$
- “Boats”: $\|\nabla[f] - \nabla[f]_s\|_2$ decays quickly in s

Natural Notation

Images are compressible in *discrete gradient*.

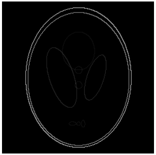



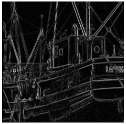


- $\|f\|_p := \left(\sum_{j=1}^N \sum_{k=1}^N |f_{j,k}|^p\right)^{1/p}$
- f is s-sparse if $\|f\|_0 := \{\#(j,k) : f_{j,k} \neq 0\} \leq s$
- f_s is the best s-sparse approximation to f
- "Phantom": $\|\nabla[f]\|_0 = .03N^2$
- "Boats": $\|\nabla[f] - \nabla[f]_s\|_2$ decays quickly in s

Natural Notation

Images are compressible in *discrete gradient*.



- $\|f\|_p := \left(\sum_{j=1}^N \sum_{k=1}^N |f_{j,k}|^p\right)^{1/p}$
- f is s-sparse if $\|f\|_0 := \{\#(j,k) : f_{j,k} \neq 0\} \leq s$
- f_s is the best s-sparse approximation to f
- "Phantom": $\|\nabla[f]\|_0 = .03N^2$
- "Boats": $\|\nabla[f] - \nabla[f]_s\|_2$ decays quickly in s

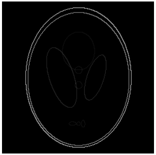
Natural Notation

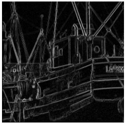
Images are compressible in *discrete gradient*.

- $\|f\|_p := \left(\sum_{j=1}^N \sum_{k=1}^N |f_{j,k}|^p \right)^{1/p}$
- f is s-sparse if $\|f\|_0 := \{\#(j,k) : f_{j,k} \neq 0\} \leq s$
- f_s is the best s-sparse approximation to f
- "Phantom": $\|\nabla[f]\|_0 = .03N^2$
- "Boats": $\|\nabla[f] - \nabla[f]_s\|_2$ decays quickly in s

Natural Notation

Images are compressible in *discrete gradient*.



- $\|f\|_p := \left(\sum_{j=1}^{N} \sum_{k=1}^{N} |f_{j,k}|^p \right)^{1/p}$
- f is s-sparse if $\|f\|_0 := \{\#(j,k) : f_{j,k} \neq 0\} \leq s$
- f_s is the best s-sparse approximation to f
- "Phantom": $\|\nabla[f]\|_0 = .03N^2$
- "Boats": $\|\nabla[f] - \nabla[f]_s\|_2$ decays quickly in s

Sparsity in gradient

CS Theory

The gradient operator ∇ is not an orthonormal basis or a tight frame.

Comparison of two compressed sensing reconstruction algorithms

Haar-minimization (L_1-Haar)

$$\hat{f}_{Haar} = \text{argmin} \|H(Z)\|_1 \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon$$

Total Variation minimization (TV)

$$\hat{f}_{TV} = \text{argmin} \|\nabla[Z]\|_1 \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon, \text{ where}$$

$\|Z\|_{TV} = \|\nabla[Z]\|_1$ is the *total-variation norm.*

Comparison of two compressed sensing reconstruction algorithms

Haar-minimization (L_1-Haar)

$$\hat{f}_{Haar} = \text{argmin} \, \|H(Z)\|_1 \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon$$

Total Variation minimization (TV)

$$\hat{f}_{TV} = \text{argmin} \, \|\nabla[Z]\|_1 \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon,$$ where

$\|Z\|_{TV} = \|\nabla[Z]\|_1$ is the *total-variation norm*.

Imaging via compressed sensing

(a) Original

(b) TV

(c) L_1-Haar

Figure: Reconstruction using $m = .2N^2$

Imaging via compressed sensing

(a) Original

(b) TV

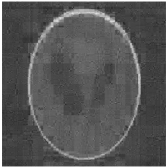
(c) L_1-Haar

Figure: Reconstruction using $m = .2N^2$ measurements

Imaging via compressed sensing

(a) Original

(b) TV



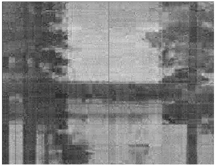
(c) L_1-Haar

Figure: Reconstruction using $m = .2N^2$ measurements.

Imaging via compressed sensing

(a) (Quantization)

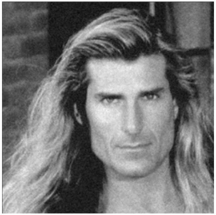
(b) TV



(c) L_1-Haar

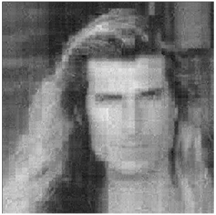
Figure: Reconstruction using $m = .2N^2$ measurements

Imaging via compressed sensing



(a) (Gaussian)

(b) TV



(c) L_1-Haar

Imaging via compressed sensing

InView (Austin TX)

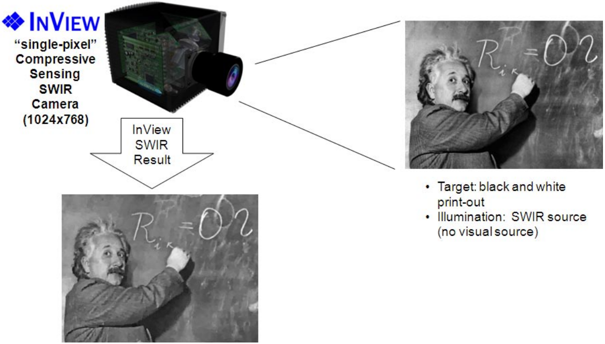


Figure: SWIR Reconstruction using $m = .5N^2$ measurements

Imaging via compressed sensing

InView (Austin TX)

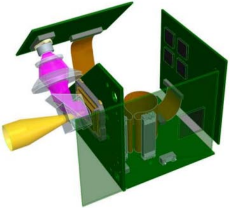


Figure: InView SWIR camera

Empirical $\rightarrow$ Theoretical?

TV Works

Empirically, it has been well known that

$$\hat{f}_{TV} = \operatorname{argmin} \|Z\|_{TV} \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon, \qquad (TV)$$

provides quality, stable image recovery.

No provable stability guarantees.

Empirical → Theoretical?

TV Works

Empirically, it has been well known that

$$\hat{f}_{TV} = \text{argmin} \|Z\|_{TV} \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon, \qquad (TV)$$

provides quality, stable image recovery.

No provable stability guarantees.

Stable signal recovery using total-variation minimization

Theorem (N-Ward '12)

From $m \gtrsim s\log(N)$ linear RIP measurements, for any $f \in \mathbb{C}^{N\times N}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \textit{such that} \quad \|\mathcal{A}(Z) - y\|_2 \le \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N)$ factor

Stable signal recovery using total-variation minimization

Theorem (N-Ward '12)

From $m \gtrsim s\log(N)$ linear RIP measurements, for any $f \in \mathbb{C}^{N\times N}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \textit{such that} \quad \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N)$ factor

Stable signal recovery using total-variation minimization

Theorem (N-Ward '12)

From $m \gtrsim s\log(N)$ linear RIP measurements, for any $f \in \mathbb{C}^{N\times N}$,

$$\hat{f} = \text{argmin}\, \|Z\|_{TV} \quad \textit{such that} \quad \|\mathcal{A}(Z) - y\|_2 \le \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N)$ factor

Stable signal recovery using total-variation minimization

Theorem (N-Ward '12)

From $m \gtrsim s\log(N)$ linear RIP measurements, for any $f \in \mathbb{C}^{N\times N}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \text{such that} \quad \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N)$ factor

Stable signal recovery using total-variation minimization

Theorem (N-Ward '12)

From $m \gtrsim s\log(N)$ linear RIP measurements, for any $f \in \mathbb{C}^{N\times N}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \text{such that} \quad \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N)$ factor

Higher dimensional objects

Movies, higher dimensional objects?

Theorem (N-Ward '12, In preparation.)

From $m \gtrsim s\log(N^d)$ linear RIP measurements, for any $f \in \mathbb{C}^{N^d}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \textit{such that} \quad \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N^d/s) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N^d/s)$ factor

Higher dimensional objects

Movies, higher dimensional objects?

Theorem (N-Ward '12, In preparation.)

From $m \gtrsim s\log(N^d)$ *linear RIP measurements, for any* $f \in \mathbb{C}^{N^d}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \text{such that} \quad \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N^d/s) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N^d/s)$ factor

Higher dimensional objects

Movies, higher dimensional objects?

Theorem (N-Ward '12, In preparation.)

From $m \gtrsim s\log(N^d)$ *linear RIP measurements, for any* $f \in \mathbb{C}^{N^d}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \text{such that} \quad \|\mathcal{A}(Z) - y\|_2 \le \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N^d/s) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N^d/s)$ factor

Higher dimensional objects

Movies, higher dimensional objects?

Theorem (N-Ward '12, In preparation.)

From $m \gtrsim s\log(N^d)$ linear RIP measurements, for any $f \in \mathbb{C}^{N^d}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \textit{such that} \quad \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N^d/s) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N^d/s)$ factor

Higher dimensional objects

Movies, higher dimensional objects?

Theorem (N-Ward '12, In preparation.)

From $m \gtrsim s \log(N^d)$ linear RIP measurements, for any $f \in \mathbb{C}^{N^d}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \text{such that} \quad \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N^d/s) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N^d/s)$ factor

Higher dimensional objects

Movies, higher dimensional objects?

Theorem (N-Ward '12, In preparation.)

From $m \gtrsim s\log(N^d)$ *linear RIP measurements, for any* $f \in \mathbb{C}^{N^d}$,

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \textit{such that} \quad \|\mathcal{A}(Z) - y\|_2 \le \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N^d/s) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N^d/s)$ factor

Proof Sketch

Strengthened Sobolev inequalities for random subspaces

Discrete Sobolev inequalities

Proposition (Sobolev inequality for discrete images)

Let $X \in \mathbb{R}^{N \times N}$ be mean-zero. Then

$$\|X\|_2 \leq \|X\|_{TV}$$

Proposition (New: Strengthed Sobolev inequality)

With probability $\geq 1 - e^{-cm}$, the following holds for all images $X \in \mathbb{R}^{N \times N}$ in the null space of an $m \times N^2$ random Gaussian matrix

$$\|X\|_2 \lesssim \frac{[\log(N)]^{3/2}}{\sqrt{m}} \|X\|_{TV}$$

Discrete Sobolev inequalities

Proposition (Sobolev inequality for discrete images)

Let $X \in \mathbb{R}^{N \times N}$ be mean-zero. Then

$$\|X\|_2 \leq \|X\|_{TV}$$

Proposition (New: Strengthed Sobolev inequality)

With probability $\geq 1 - e^{-cm}$, the following holds for all images $X \in \mathbb{R}^{N \times N}$ in the null space of an $m \times N^2$ random Gaussian matrix

$$\|X\|_2 \lesssim \frac{[\log(N)]^{3/2}}{\sqrt{m}} \|X\|_{TV}$$

Strengthened Sobolev inequalities

Proof ingredients:

1. [CDPX 99:] Denote the bivariate Haar wavelet coefficients of $X \in \mathbb{R}^{N \times N}$ by $c_{(1)} \geq c_{(2)} \geq \cdots \geq c_{(N^2)}$. Then

$$|c_{(k)}| \lesssim \frac{\|X\|_{TV}}{k}$$

   That is, the sequence is in weak-ℓ_1.

2. If $\Phi : \mathbb{R}^d \to \mathbb{R}^m$ has (properly normalized) i.i.d. Gaussian entries then with probability exceeding $1 - e^{-cm}$, Φ has the RIP of order $s \sim \frac{m}{\log d}$:

$$\frac{3}{4}\|f\|_2 \leq \|\Phi f\|_2 \leq \frac{5}{4}\|f\|_2 \quad \text{for all } s\text{-sparse } f.$$

Strengthened Sobolev inequalities: proof

Let $\Phi : \mathbb{R}^d \to \mathbb{R}^m$ be a Gaussian matrix and $\Phi X = 0$.

Suppose that $\Psi = \Phi \mathcal{H}^* : \mathbb{R}^d \to \mathbb{R}^m$ has the RIP of order $2s$.

Decompose $c = \mathcal{H}X$ into s-sparse blocks $c = c_{S_0} + c_{S_1} + c_{S_2} + \dots$

Then $\Psi c = \Phi \mathcal{H}^* \mathcal{H} X = \Phi X = 0$ and

$$\begin{aligned} 0 &\geq \|\Psi(c_{S_0} + c_{S_1})\|_2 - \sum_{j \geq 2} \|\Psi c_{S_j}\|_2 \\ \text{(RIP of } \Psi) \quad &\geq \frac{3}{4}\|c_{S_0} + c_{S_1}\|_2 - \frac{5}{4}\sum_{j \geq 2} \|c_{S_j}\|_2 \\ \text{(block trick)} &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\sum_{j \geq 1} \|c_{S_j}\|_1 \\ (c \text{ in weak } \ell_1) &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\|X\|_{TV} \log(d/s) \end{aligned}$$

Strengthened Sobolev inequalities: proof

Let $\Phi : \mathbb{R}^d \to \mathbb{R}^m$ be a Gaussian matrix and $\Phi X = 0$.

Suppose that $\Psi = \Phi \mathcal{H}^* : \mathbb{R}^d \to \mathbb{R}^m$ has the RIP of order $2s$.

Decompose $c = \mathcal{H}X$ into s-sparse blocks $c = c_{S_0} + c_{S_1} + c_{S_2} + \dots$

Then $\Psi c = \Phi \mathcal{H}^* \mathcal{H} X = \Phi X = 0$ and

$$
\begin{aligned}
0 &\geq \|\Psi(c_{S_0} + c_{S_1})\|_2 - \sum_{j \geq 2} \|\Psi c_{S_j}\|_2 \\
\text{(RIP of } \Psi) \quad &\geq \frac{3}{4}\|c_{S_0} + c_{S_1}\|_2 - \frac{5}{4}\sum_{j \geq 2} \|c_{S_j}\|_2 \\
\text{(block trick)} &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\sum_{j \geq 1} \|c_{S_j}\|_1 \\
(c \text{ in weak } \ell_1) &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\|X\|_{TV} \log(d/s)
\end{aligned}
$$

Strengthened Sobolev inequalities: proof

Let $\Phi : \mathbb{R}^d \to \mathbb{R}^m$ be a Gaussian matrix and $\Phi X = 0$.

Suppose that $\Psi = \Phi \mathcal{H}^* : \mathbb{R}^d \to \mathbb{R}^m$ has the RIP of order $2s$.

Decompose $c = \mathcal{H}X$ into s-sparse blocks $c = c_{S_0} + c_{S_1} + c_{S_2} + \dots$

Then $\Psi c = \Phi \mathcal{H}^* \mathcal{H} X = \Phi X = 0$ and

$$
\begin{aligned}
0 &\geq \|\Psi(c_{S_0} + c_{S_1})\|_2 - \sum_{j \geq 2} \|\Psi c_{S_j}\|_2 \\
\text{(RIP of } \Psi) \quad &\geq \frac{3}{4}\|c_{S_0} + c_{S_1}\|_2 - \frac{5}{4} \sum_{j \geq 2} \|c_{S_j}\|_2 \\
\text{(block trick)} &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}} \sum_{j \geq 1} \|c_{S_j}\|_1 \\
(c \text{ in weak } \ell_1) &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}} \|X\|_{TV} \log(d/s)
\end{aligned}
$$

Strengthened Sobolev inequalities: proof

Let $\Phi : \mathbb{R}^d \to \mathbb{R}^m$ be a Gaussian matrix and $\Phi X = 0$.

Suppose that $\Psi = \Phi \mathcal{H}^* : \mathbb{R}^d \to \mathbb{R}^m$ has the RIP of order $2s$.

Decompose $c = \mathcal{H}X$ into s-sparse blocks $c = c_{S_0} + c_{S_1} + c_{S_2} + \dots$

Then $\Psi c = \Phi \mathcal{H}^* \mathcal{H} X = \Phi X = 0$ and

$$
\begin{aligned}
0 &\geq \|\Psi(c_{S_0} + c_{S_1})\|_2 - \sum_{j \geq 2} \|\Psi c_{S_j}\|_2 \\
(\text{RIP of } \Psi) \quad &\geq \frac{3}{4}\|c_{S_0} + c_{S_1}\|_2 - \frac{5}{4}\sum_{j \geq 2} \|c_{S_j}\|_2 \\
(\text{block trick}) &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\sum_{j \geq 1} \|c_{S_j}\|_1 \\
(c \text{ in weak } \ell_1) &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\|X\|_{TV} \log(d/s)
\end{aligned}
$$

Strengthened Sobolev inequalities: proof

Let $\Phi : \mathbb{R}^d \to \mathbb{R}^m$ be a Gaussian matrix and $\Phi X = 0$.

Suppose that $\Psi = \Phi\mathcal{H}^* : \mathbb{R}^d \to \mathbb{R}^m$ has the RIP of order $2s$.

Decompose $c = \mathcal{H}X$ into s-sparse blocks $c = c_{S_0} + c_{S_1} + c_{S_2} + \dots$

Then $\Psi c = \Phi\mathcal{H}^*\mathcal{H}X = \Phi X = 0$ and

$$\begin{aligned}
0 &\geq \|\Psi(c_{S_0} + c_{S_1})\|_2 - \sum_{j\geq 2} \|\Psi c_{S_j}\|_2 \\
\text{(RIP of } \Psi) \quad &\geq \frac{3}{4}\|c_{S_0} + c_{S_1}\|_2 - \frac{5}{4}\sum_{j\geq 2}\|c_{S_j}\|_2 \\
\text{(block trick)} &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\sum_{j\geq 1}\|c_{S_j}\|_1 \\
(c \text{ in weak } \ell_1) &\geq \frac{3}{4}\|c_{S_0}\|_2 - \frac{5/4}{\sqrt{s}}\|X\|_{TV}\log(d/s)
\end{aligned}$$

Strengthened Sobolev inequalities: proof

So

1. $\|c_{S_0}\|_2 \leq \frac{1}{\sqrt{s}}\|X\|_{TV} \log(d/s)$,
2. $\|c - c_{S_0}\|_2 \leq \frac{1}{\sqrt{s}}\|X\|_{TV}$ (c is in weak ℓ_1)

Then

$$\begin{aligned} \|X\|_2 &= \|c\|_2 \leq \|c_{S_0}\|_2 + \|c - c_{S_0}\|_2 \\ &\leq \frac{\log(d/s)}{\sqrt{s}}\|X\|_{TV} \end{aligned}$$

Proof is complete, because with probability $1 - \varepsilon^{-cm}$, RIP of $\Phi\mathcal{H}^*$ holds with $s \sim m/\log(d)$.

Stable signal recovery using total-variation minimization

Method of proof:

1. First prove stable *gradient* recovery
2. Translate stable *gradient* recovery to stable *signal* recovery using the strengthened Sobolev inequality.

Stable signal recovery using total-variation minimization

Method of proof:

1. First prove stable *gradient* recovery
2. Translate stable *gradient* recovery to stable *signal* recovery using the strengthened Sobolev inequality.

Open questions

1. Remove the log factor?
2. Our results do not immediately hold for vectors. What about stable (1D) signal recovery?
3. [Patel, Maleh, Gilbert, Chellappa '11] Images are even sparser in individual directional derivatives f_x, f_y. If we minimize separately over directional derivatives, can we still prove stable recovery?
4. Other optimization problems

Open questions

1. Remove the log factor?
2. Our results do not immediately hold for vectors. What about stable (1D) signal recovery?
3. [Patel, Maleh, Gilbert, Chellappa '11] Images are even sparser in individual directional derivatives f_x, f_y. If we minimize separately over directional derivatives, can we still prove stable recovery?
4. Other optimization problems

Open questions

1. Remove the log factor?
2. Our results do not immediately hold for vectors. What about stable (1D) signal recovery?
3. [Patel, Maleh, Gilbert, Chellappa '11] Images are even sparser in individual directional derivatives f_x, f_y. If we minimize separately over directional derivatives, can we still prove stable recovery?
4. Other optimization problems

Open questions

1. Remove the log factor?
2. Our results do not immediately hold for vectors. What about stable (1D) signal recovery?
3. [Patel, Maleh, Gilbert, Chellappa '11] Images are even sparser in individual directional derivatives f_x, f_y. If we minimize separately over directional derivatives, can we still prove stable recovery?
4. Other optimization problems

Movies are very sparse!

- Movies are very sparse in all three dimensions
- Silicon Retina (Institute of Neuroinformatics) is the design of a camera that mimics retinas
- Dynamic Vision Sensor (DVS) from Silicon Retina, Institute of Neuroinformatics, Zurich

Movies are very sparse!

- Movies are very sparse in all three dimensions
- Silicon Retina (Institute of Neuroinformatics) is the design of a camera that mimics retinas
- Dynamic Vision Sensor (DVS) from Silicon Retina, Institute of Neuroinformatics, Zurich

Movies are very sparse!

- Movies are very sparse in all three dimensions
- Silicon Retina (Institute of Neuroinformatics) is the design of a camera that mimics retinas
- Dynamic Vision Sensor (DVS) from Silicon Retina, Institute of Neuroinformatics, Zurich

Fast vision in bad lighting

Figure: ("RoboGoalie", Silicon Retina, Institute of Neuroinformatics)

Fluid Particle Tracking Velocimetry

Figure: ("PTV", Silicon Retina, Institute of Neuroinformatics)

Mobile Robotics

Figure: (“Robotic Driver”, Silicon Retina, Institute of Neuroinformatics)

Sleep disorder research

Figure: ("Sleeping Mouse", Silicon Retina, Institute of Neuroinformatics)

Thank you!

E-mail:

- dneedell@cmc.edu

Web:

- www.cmc.edu/pages/faculty/DNeedell

References:

- A. Cohen, R. DeVore, P. Petrushev and H. Xu, Nonlinear approximation and the space $BV(\mathbb{R}^2)$, Am. J. of Math, 121:587-628, 1999.
- 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, D. Needell and P. Randall. Compressed sensing with coherent and redundant dictionaries. Applied and Computational Harmonic Analysis, 31(1):59-73, 2010.
- D. Needell and R. Ward. Stable image reconstruction using total variation minimization. Submitted.