Robust image recovery via total-variation minimization

Rachel Ward

University of Texas at Austin

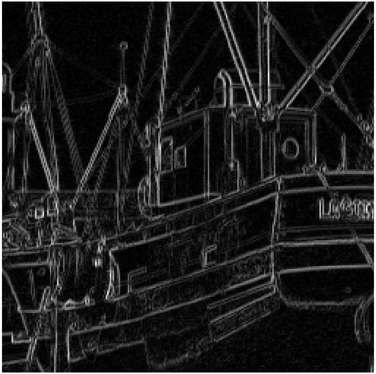
(Joint work with Deanna Needell, Claremont McKenna College)

February 16, 2012

Images are compressible

256 $\times$ 256 “Boats” image

Images are compressible
in discrete gradient



Images are compressible
in discrete gradient

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

$$X_x : \mathbb{R}^{N\times N} \to \mathbb{R}^{(N-1)\times N}, \qquad (X_x)_{j,k} = X_{j,k} - X_{j-1,k},$$
$$X_y : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times(N-1)}, \qquad (X_y)_{j,k} = X_{j,k} - X_{j,k-1},$$

the discrete gradient operator is

$$\left[TV[X]\right]_{j,k} = (X_x)_{j,k} + i(X_y)_{j,k}$$

Images are compressible
in discrete gradient


$\|X\|_p := \left(\sum_{j=1}^{N} \sum_{k=1}^{N} |X_{j,k}|^p \right)^{1/p}$

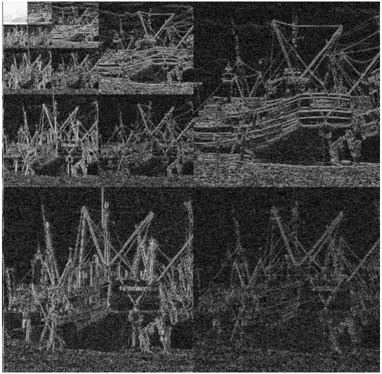
X is s-sparse if $\|X\|_0 := \{\#(j,k) : X_{j,k} \neq 0\} \leq s$

X_s is the best s-sparse approximation to X

$\sigma_s(X)_p = \|X - X_s\|_p$ is the best s-term approximation error in ℓ_p.

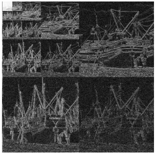
"Phantom": $\|TV[X]\|_0 = .03N^2$, "Boats": $\sigma_s(TV[X])_2$ decays quickly in s

Images are compressible in Wavelet bases



Two-dimensional Haar Wavelet Transform of “Boats”

Images are compressible
in Wavelet bases

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

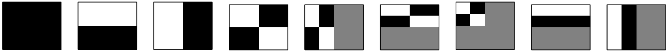


Figure: Haar basis functions

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

Images are compressible in Wavelet bases

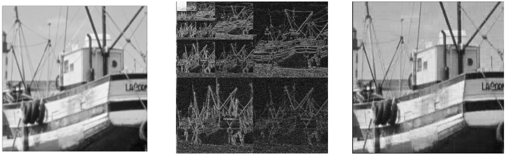


Figure: Boats image, 2D Haar transform, and compression using 10% Haar coefficients

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

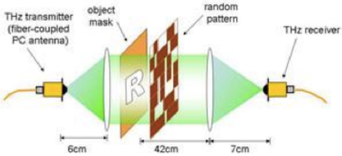
X is s-sparse (in Haar basis) if $\|c\|_0 \leq s$

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

$$\sigma_s^w(X)_p = \|X - X_s^w\|_p$$

Imaging via Compressed Sensing

Imaging via compressed sensing



Instead of storing all N^2 pixels of $X \in \mathbb{R}^{N \times N}$ and then compressing,

acquire information about X through $m \ll N^2$ nonadaptive linear measurements of the form

$$y_\ell = \langle A_\ell, X \rangle = \mathrm{trace}(A_\ell^* X)$$

or concisely $y = \mathcal{A}(X)$

Imaging via compressed sensing

More realistically, measurements are noisy,

$$y_\ell = \langle A_\ell, X \rangle + \xi_\ell,$$

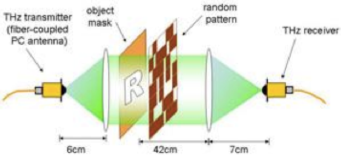
concisely $y = \mathcal{A}(X) + \xi$.

The goal is to use measurements A_ℓ and reconstruction algorithm such that $X \in \mathbb{R}^{N \times N}$ is reconstructed from $y \in \mathbb{R}^m$ efficiently and robustly

Robust: Reconstruction error $\|\hat{X} - X\|_2$ comparable to both noise level $\varepsilon = \|\xi\|_2$ and best s-term approximation error in (discrete gradient, wavelet basis) with $s \lesssim m/\log(N)$.

Efficient: Using a polynomial-time algorithm

Imaging via compressed sensing



Results in compressed sensing [CRT '06, etc ...] imply:

- if $X \in \mathbb{R}^{N \times N}$ is s-sparse in an orthonormal basis B
- if we use $m \gtrsim s \log(N)$ measurements $y_\ell = \langle A_\ell, X \rangle$ where A_ℓ are i.i.d. Gaussian random matrices

then with high probability,

$$X = \operatorname*{argmin}_{Z \in \mathbb{R}^{N \times N}} \|BZ\|_1 \quad \text{subject to } \mathcal{A}(Z) = y$$

Imaging via compressed sensing

Moreover,

- if $X \in \mathbb{R}^{N\times N}$ is approximately s-sparse in orthonormal basis B
- if we use $m \gtrsim s\log(N)$ noisy measurements $y_\ell = \langle A_\ell, X\rangle + \eta_\ell$ with A_ℓ i.i.d. Gaussian
- If $\hat{X} = \operatorname{argmin} \|BZ\|_1 \quad \text{subject to } \|\mathcal{A}(Z) - y\|_2 \le \varepsilon$,

then

$$\|X - \hat{X}\|_2 \lesssim \|X - X_s^B\|_1/\sqrt{s} + \varepsilon$$

This implies a strategy for reconstructing images up to their best s-term Haar approximation using $m \gtrsim s\log(N)$ measurements.

Imaging via compressed sensing

Moreover,

- if $X \in \mathbb{R}^{N\times N}$ is approximately s-sparse in orthonormal basis B
- if we use $m \gtrsim s\log(N)$ noisy measurements $y_\ell = \langle A_\ell, X\rangle + \eta_\ell$ with A_ℓ i.i.d. Gaussian
- If $\hat{X} = \mathrm{argmin}\,\|BZ\|_1 \quad \text{subject to } \|\mathcal{A}(Z) - y\|_2 \leq \varepsilon,$

then

$$\|X - \hat{X}\|_2 \lesssim \|X - X_s^B\|_1/\sqrt{s} + \varepsilon$$

This implies a strategy for reconstructing images up to their best s-term Haar approximation using $m \gtrsim s\log(N)$ measurements.

Imaging via compressed sensing

Let's compare two compressed sensing reconstruction algorithms

$$\hat{X}_{Haar} = \operatorname{argmin} \|H(Z)\|_1 \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon \qquad (L_1)$$

and

$$\hat{X}_{TV} = \operatorname{argmin} \|TV[Z]\|_1 \quad \text{subject to} \quad \|\mathcal{A}Z - y\|_2 \leq \varepsilon. \quad (TV)$$

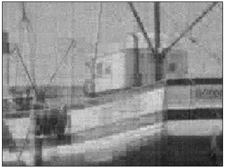
$\|Z\|_{TV} = \|TV[Z]\|_1$.

The mapping $Z \rightarrow TV[Z]$ is not orthonormal (inverse norm grows with N), stable image recovery via (TV) is not immediately justified.

Imaging via compressed sensing

(a) Original

(b) TV



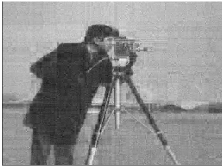
(c) L_1

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

Imaging via compressed sensing

(a) Original

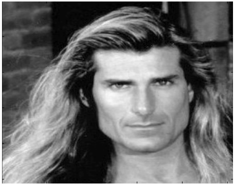
(b) TV



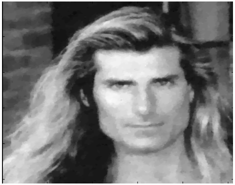
(c) L_1

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

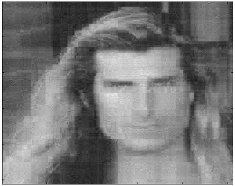
Imaging via compressed sensing



(a) Original



(b) TV



(c) L_1

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

Stable signal recovery using total-variation minimization

Our main result:

Theorem
There are choices of $m \gtrsim s \log(N)$ *measurements of the form*

$$\mathcal{A}(X) = (\langle X, A_\ell \rangle)_{\ell=1}^{m}$$

such that given $y = \mathcal{A}(X) + \xi$ *and*

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

with high probability

$$\|X - \hat{X}\|_2 \lesssim \log(\log(N)) \cdot \left[\frac{\sigma(TV[X])_1}{\sqrt{s}} + \varepsilon\right]$$

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

Stable signal recovery using total-variation minimization

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

$$\Longrightarrow \|X - \hat{X}\|_2 \lesssim \log(\log(N)) \cdot \left[\frac{\sigma(TV[X])_1}{\sqrt{s}} + \varepsilon \right]$$

Method of proof:

1. First prove stable *gradient* recovery
2. Translate stable *gradient* recovery to stable *signal* recovery using the following (nontrivial) relationship between total variation and decay of Haar wavelet coefficients:

Theorem (Cohen, DeVore, Petrushev, Xu, 1999)

Let $c_{(1)} \geq c_{(2)} \geq \dots c_{(N^2)}$ be the bivariate Haar coefficients of an image $Z \in \mathbb{R}^{N \times N}$, arranged in decreasing order of magnitude. Then

$$|c_{(k)}| \leq 10^5 \frac{\|Z\|_{TV}}{k}$$

II. Stable signal recovery from stable gradient recovery

$\mathcal{A}(Z) = \langle A_\ell, Z \rangle$, A_ℓ are i.i.d. Gaussian,

$$\hat{X} = \text{argmin} \, \|Z\|_{TV} \quad \mathcal{A}(Z - X) = y$$

1. [CDPX '99] Let $D = X - \hat{X}$. If $c_{(k)}$ are the Haar coefficients HD in decreasing arrangement, then

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

so $c = HD$ is compressible.

2. Gaussian random matrices are rotation-invariant, and $\mathcal{A}(D) = 0$ implies $c = HD$ is in the null space of an $(m \times N^2)$ Gaussian matrix. Then $c = HD$ must also be flat. (Null space property)

Together these imply that $\|D\|_2 = \|HD\|_2 \leq \log(N)\|TV[D]\|_2$

Summary

We use the (nontrivial) relationship between the total variation norm and compressibility of Haar wavelet coefficients to prove near-optimal robust image recovery via total-variation minimization

Images are sparser in discrete gradient than in Wavelet bases, so our results are in line with numerical studies

Open questions

1. The relationship between Haar compressibiity and total variation norm doesn't hold in one-dimension. What about stable (1D) signal recovery?
2. Do our stability results generalize to more practical compressed sensing measurement ensembles (e.g. partial random Fourier measurements?) (We have sub-optimal results)
3. [Patel, Maleh, Gilbert, Chellappa '11] Images are even sparser in individual directional derivatives X_x, X_y. If we minimize separately over directional derivatives, can we still prove stable recovery?