Compressed sensing and optimal denoising of sparse monotone signals

Eftychios A. Pnevmatikakis

Flatiron Institute, Simons Foundation, New York, NY

Summary

We are interested in signals $\mathbf{x_0} \in \mathbb{R}^N$ that are

- monotone: $\mathbf{x_0}(i+1) \geq \mathbf{x_0}(i)$
- sparsely varying: $\mathbf{x_0}(i+1) > \mathbf{x_0}(i)$ only for a small number k of indices i

We consider the following two problems:

- Compressed Sensing:

$$\min_{\mathbf{x}} f(\mathbf{x}), \text{ subject to } A\mathbf{x} = A\mathbf{x_0}, \quad \text{(CS)}$$

with $A \in \mathbb{R}^{m \times N}$ random i.i.d. Gaussian

- Optimal Denoising:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \lambda f(\mathbf{x}), \quad \text{(DN)}$$

where $f : \mathbb{R}^N \mapsto \mathbb{R} \cup \{\infty\}$ is a structure inducing function for the space of monotone sparsely varying signals:

$$f(\mathbf{x}) = \begin{cases} \mathbf{x}(N) - \mathbf{x}(1), & \mathbf{x}(i+1) \geq \mathbf{x}(i), i \in [N-1] \\ \infty, & \text{otherwise.} \end{cases} \quad \text{(RTV)}$$

Results

For the (CS) problem:

- Closed form expression for the number of measurements m required for successful reconstruction with high probability.
- We show that m depends not only on the number of changing points k but also on the location.
- Characterize best, worst, and average cases.
- Compare with the case of non-negative sparse signals.

For the (DN) problem:

- Characterize minimax cost and its dependence on the set of changing points.
- Calculate optimal value for regularizer λ.

Basic tools [1]

Definition (Descent cones)

The descent cone of a convex function $f : \mathbb{R}^N \mapsto \mathbb{R}$ at a point $\mathbf{x} \in \mathbb{R}^N$ is defined as the set of all non-increasing directions, i.e.,

$$\mathcal{D}(f, \mathbf{x}) = \bigcup_{\tau > 0} \{\mathbf{y} \in \mathbb{R}^N : f(\mathbf{x} + \tau \mathbf{y}) \leq f(\mathbf{x})\}.$$

Example:

$$f(x) = |x| \Rightarrow \mathcal{D}(f, x) = \begin{cases} \mathbb{R}_-, & x > 0 \\ \mathbb{R}_+, & x < 0 \end{cases}.$$

Definition (Statistical dimension)

The statistical dimension (SD) of a convex cone $\mathcal{C} \in \mathbb{R}^N$ is defined as

$$\delta(\mathcal{C}) = \mathbb{E}_{\mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_N)} \|\Pi_{\mathcal{C}}(\mathbf{g})\|^2,$$

where $\mathbf{g}$ is a standard Gaussian vector, and $\Pi_{\mathcal{C}}$ is the projection onto $\mathcal{C}$. Example: $\delta(\mathbb{R}^n_+) = n/2$.

Theorem (Phase transitions (Amelunxen et al.))

For an i.i.d. standard random Gaussian matrix $A \in \mathbb{R}^{m \times N}$ the convex problem (CS) succeeds with probability at least $1 - \exp(-t^2/4)$ if

$$m \geq \delta(\mathcal{D}(f, \mathbf{x_0})) + t\sqrt{N},$$

and fails with probability at least $1 - \exp(-t^2/4)$ if

$$m \leq \delta(\mathcal{D}(f, \mathbf{x_0})) - t\sqrt{N}.$$

Phase transitions for sparsely varying monotone signals

Lemma (Descent cones for monotone sparsely varying signals)

Let $\Omega = \{i \in \{2, \dots, N\} : \mathbf{x_0}(i) > \mathbf{x_0}(i-1)\}$ and define $i_1 < i_2 < \dots < i_k$ the elements of Ω in increasing order. The descent cone of the norm f of (RTV) at $\mathbf{x_0}$ is given by

$$\mathcal{D}(f, \mathbf{x_0}) = \left\{ \mathbf{y} \in \mathbb{R}^N : \begin{array}{c} \mathbf{y}(i_1) \leq \mathbf{y}(i_1+1) \leq \dots \leq \mathbf{y}(i_2 - 1) \\ \vdots \\ \mathbf{y}(i_{k-1}) \leq \mathbf{y}(i_{k-1}+1) \leq \dots \leq \mathbf{y}(i_k - 1) \\ \mathbf{y}(i_k) \leq \mathbf{y}(i_k+1) \leq \dots \leq \mathbf{y}(N) \leq \mathbf{y}(1) \leq \dots \leq \mathbf{y}(i_1 - 1) \end{array} \right\}.$$

The descent cone $\mathcal{D}(f, \mathbf{x_0})$ decomposes as the product of simpler "monotone" cones.

Fact (Statistical Dimension of "monotone" cones (Amelunxen et al.))

Let the cones

$$\mathcal{C}_1^N = \{\mathbf{x} \in \mathbb{R}^N : \mathbf{x}(1) \leq \mathbf{x}(2) \leq \dots \leq \mathbf{x}(N)\}$$
$$\mathcal{C}_2^N = \{\mathbf{x} \in \mathbb{R}^N : 0 \leq \mathbf{x}(1) \leq \mathbf{x}(2) \leq \dots \leq \mathbf{x}(N)\}.$$

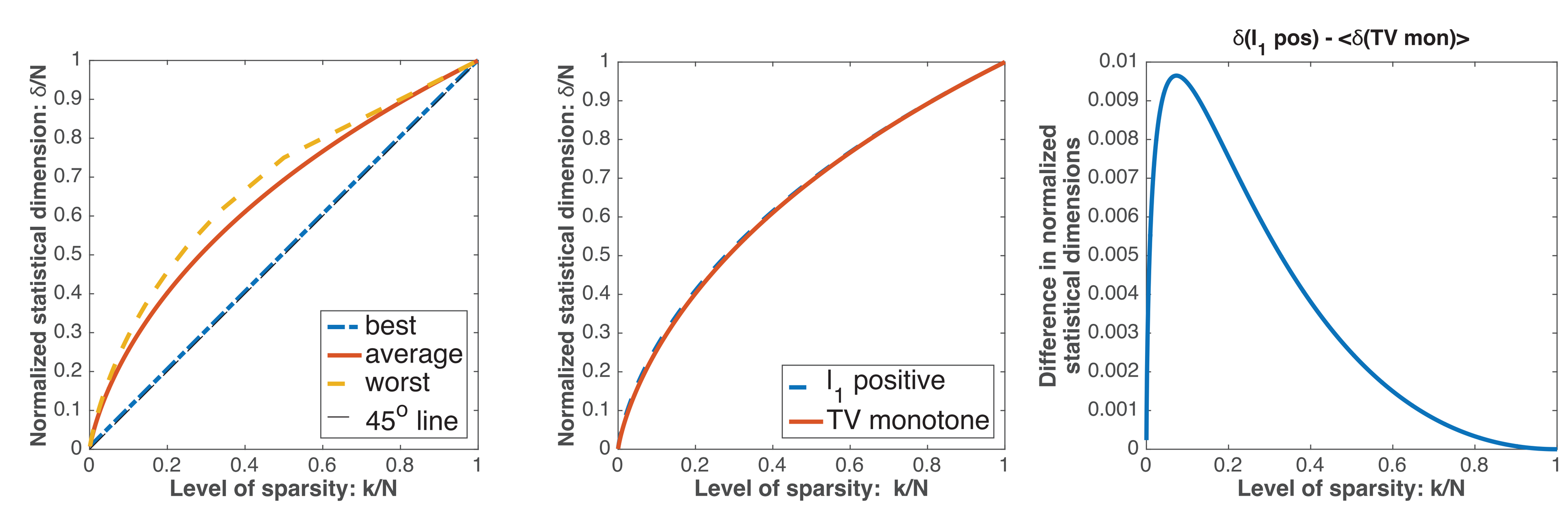
Then we have $\delta(\mathcal{C}_1^N) = H_N$, and $\delta(\mathcal{C}_2^N) = \frac{1}{2}H_N$, where $H_N = \sum_{i=1}^N \frac{1}{i}$, denotes the N-th harmonic number.

Theorem (Phase transition curves)

Let $\Omega = \{i \in \{2, \dots, N\} : \mathbf{x_0}(i) > \mathbf{x_0}(i-1)\}$ and define $i_1 < i_2 < \dots < i_k$ the elements of Ω in increasing order. The SD of the descent cone at $\mathbf{x_0}$ equals

$$\delta(\mathcal{D}(f, \mathbf{x_0})) = \sum_{j=2}^{k} H_{i_j - i_{j-1}} + H_{N + i_1 - i_k}.$$

Dependence on number and location of change points



- Best case: All changing points are occurring simultaneously.

$$\delta(\mathcal{D}(f, \mathbf{x_0})) = (k-1) + H_{N+1-k}$$

- Worst case: All changing points are occurring periodically every N/k steps. With $r_{N,k} = \mathrm{mod}(N, k)$,

$$\delta(\mathcal{D}(f, \mathbf{x_0})) = (k - r_{N,k})H_{\lfloor N/k \rfloor} + r_{N,k}H_{\lfloor N/k \rfloor + 1} \to k(\log(N/k) + \gamma),$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

- Average case: The k change points are chosen uniformly at random. For $N, k \to \infty$ with $k/N = \varepsilon$, $0 < \varepsilon < 1$

$$\delta(\mathcal{D}(f, \mathbf{x_0})) = \frac{k \log(1/\varepsilon)}{1 - \varepsilon}.$$

Relationship with CS of non-negative sparse signals using the l_1 norm

- The phase transition curve for the recovery of k-sparse of non-negative signals using l_1-norm minimization can be computed analytically [2].
- This curve is very close but not identical to the average PTC computed above (middle panel).
- The difference between the two different curves (right panel) attains a maximum of ≈ 0.0096 for $k/N \approx 0.0731$.

Optimal Denoising

Theorem (Minimax risk (Oymak and Hassibi))

Let $\mathbf{x}^*(\lambda)$ the solution of the denoising problem (PDN) with regularizer weight λ and let

$$\eta_f(\mathbf{x_0}) = \min_{\lambda \geq 0} \max_{\sigma > 0} \frac{\mathbb{E}\|\mathbf{x}^*(\lambda) - \mathbf{x_0}\|^2}{\sigma^2},$$

the minimax risk for $\mathbf{x_0}$ over all possible σ. Then:

$$\eta_f(\mathbf{x_0}) = \min_{\tau \geq 0} \mathbb{E}_{\mathbf{g} \sim \mathcal{N}(\mathbf{0}, I)}[\mathrm{dist}(\mathbf{g}, \tau \partial f(\mathbf{x_0}))^2], \quad \text{(MN)}$$

where $\mathbf{g}$ is a standard normal vector. Moreover the risk is maximized for $\sigma \to 0$ and if τ^* is the value that minimizes (MN), then $\lambda^* = \tau^* \sigma$ is the optimal choice as $\sigma \to 0$.

Theorem (Relationship with Statistical Dimension (Amelunxen et al.))

$$\delta(\mathcal{D}(f, \mathbf{x_0})) \leq \eta_f(\mathbf{x_0}) \leq \delta(\mathcal{D}(f, \mathbf{x_0})) + 2\frac{\sup_{\mathbf{w} \in \partial f(\mathbf{x}_i)} \|\mathbf{w}\|}{f(\mathbf{x_0}/\|\mathbf{x_0}\|)}.$$

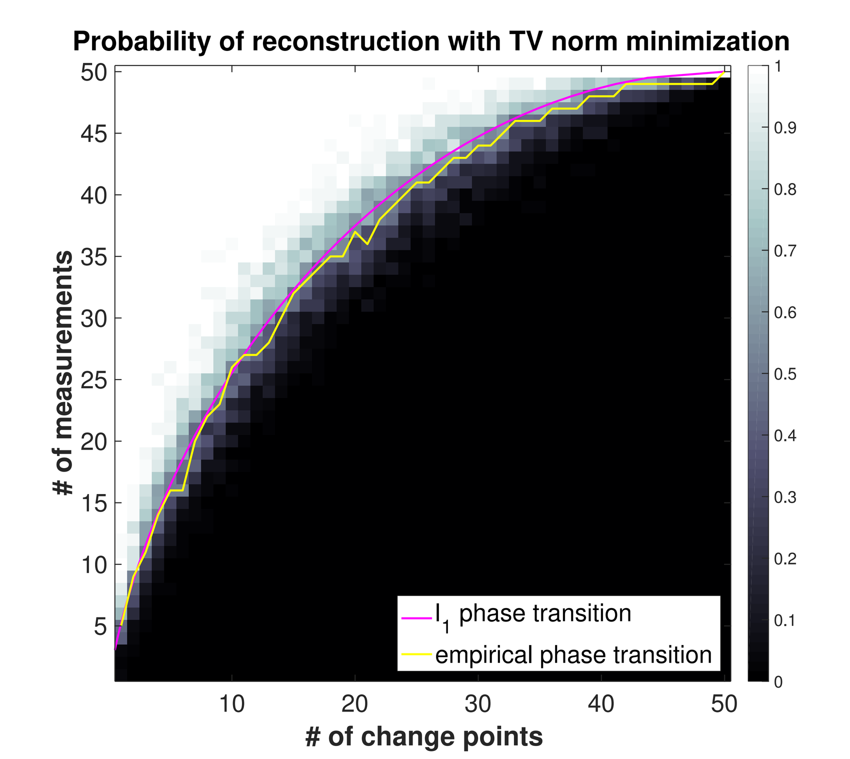
Theorem (Optimal Regularizer)

Let i_k denote the last element of Ω. Then the optimal τ^* for (MN) is given by

$$\tau^* = \max\left(\max_{j = i_k, \dots, N} \left\{\sum_{n=j}^{N} \mathbf{g}(n)\right\}, 0\right),$$

i.e., $\tau^* = M(N - i_k + 1)$, where $M(n)$ is the expected value of the maximum of a standard Gaussian random walk of n steps. It holds that $M(n) \leq \sqrt{\frac{2n}{\pi}}$.

Recovering sparsely varying signals with the TV norm



Empirical calculation of reconstruction probability for sparsely varying signals. 50-dimensional piecewise constant signals were constructed with variable number of change points k and locations chosen uniformly at random. For each signal a random Gaussian sensing matrix was constructed with variable number of rows (measurements) m. Reconstruction was attempted by minimizing the total variation (TV) norm subject to the measurements. The probability of success (color coded in the background) undergoes a phase transition. The empirical 50% success line (yellow) lies very close to the PTC for sparse signals (magenta) as is theoretically computed in [2].

Theoretical results are harder because the descent cone of the TV norm has more complex structure.

Acknowledgements

We thank L. Paninski, J. Tropp, and M. McCoy for useful discussions.

References

1. Amelunxen, D., M. Lotz, M. B. McCoy, and J. A. Tropp (2014). **Living on the edge: Phase transitions in convex programs with random data**. *Information and Inference*, iau005.
2. Donoho, D., A. Maleki, and A. Montanari (2009). **Message-passing algorithms for compressed sensing**. *Proceedings of the National Academy of Sciences* 106(45), 18914.
3. Oymak, S. and B. Hassibi (2012). **On a relation between the minimax risk and the phase transitions of compressed recovery**. *In Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on, pp. 1018?1025. IEEE.*

email: epnevmatikakis@simonsfoundation.org