ECE 18-898G: Special Topics in Signal Processing: Sparsity, Structure, and Inference

Sparse Recovery via iterative (greedy) algorithms

Yuejie Chi

Department of Electrical and Computer Engineering

Carnegie Mellon University

Spring 2018

Outline

- One-step thresholding (OST)
- Orthogonal Matching Pursuit (OMP)
- Iterative Hard Thresholding (IHT)
- Compressive Sampling Matching Pursuit (CoSaMP)

iterative and simple, with explicit control of the sparsity level.

Greedy algorithms

Consider the noise-free measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

where

- $\boldsymbol{x} \in \mathbb{R}^p$ is k-sparse,
- $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p] \in \mathbb{R}^{n \times p}$ with unit-norm columns, i.e. $\|\boldsymbol{a}_i\|_2 = 1$.

Our goal is to estimate the *support* and *coefficients* of $\boldsymbol{x}$ from $\boldsymbol{y}$.

One-Step Thresholding

One-step thresholding

If $\boldsymbol{x}$ is 1-sparse as $\boldsymbol{x} = \boldsymbol{e}_i$ which is a basis vector in $\mathbb{R}^p$, then $\boldsymbol{y}$ is just $\boldsymbol{a}_i$, and a natural way to determine i is using *matched filter*:

$$i^* = \operatorname{argmax}_{1 \le i \le p} |\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle|$$

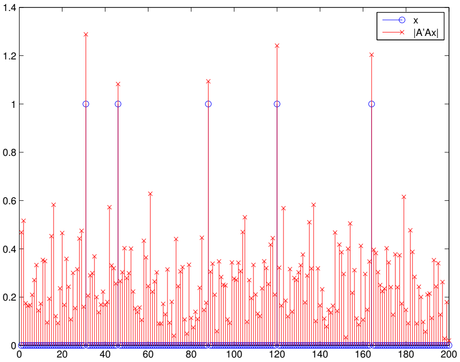
Algorithm 5.1 One-Step Thresholding (OST)

Input: Sparsity level k.

1. Compute:
$$\boldsymbol{z} = \boldsymbol{A}^{\mathsf{T}} \boldsymbol{y}$$
2. Find the support as the indices of the k largest entries of $|\boldsymbol{z}|$.

One-step thresholding

How well does it work?



If the interference from other nonzero entries of x is small enough, it is possible to read off the support of x from the largest entries of z.

Performance of OST

It is all about managing signal-to-interference ratio. Let the mutual coherence of $\boldsymbol{A}$ be

$$\mu = \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|.$$

Theorem 5.1 (OST)

Suppose that $\boldsymbol{x}$ be a k-sparse signal. OST recovers the support of $\boldsymbol{x}$ if

$$\frac{\min_i |x_i|}{\|\boldsymbol{x}\|_1} > \frac{2\mu}{(1+\mu)}.$$

- We can recover $\boldsymbol{x}$ if its smallest non-zero entry is not too small.
- If $|x_1| = \cdots = |x_k|$, the LHS becomes $1/k$ and for success support recovery we require

$$\frac{1}{k} \gtrsim \mu \sim \frac{1}{\sqrt{n}}$$

which requires $n \gtrsim k^2$.

Proof of Theorem 5.1

Note that

$$\boldsymbol{z} = \boldsymbol{A}^\top \boldsymbol{y} = \underbrace{\boldsymbol{A}^\top \boldsymbol{A}}_{\text{Gram matrix}} \boldsymbol{x}$$

- Wlog, assume $\boldsymbol{x}$ is k-sparse with the nonzero entries indexed by $\{1, \ldots, k\}$, in a descending order

$$|x_1| \geq |x_2| \geq \ldots \geq |x_k|.$$

- To guarantee the success of OST, we want to show

$$\min_{1 \leq i \leq k} |z_i| > \max_{k+1 \leq i \leq n} |z_i|.$$

Lower bound $\min_{1\le i\le k}|z_i|$

For $1 \le i \le k$,

$$\begin{aligned}
|z_i| &= |\boldsymbol{a}_i^\top \boldsymbol{A}\boldsymbol{x}| \\
&= |\boldsymbol{a}_i^\top(\boldsymbol{a}_i x_i + \sum_{j\neq i}\boldsymbol{a}_j x_j)| \\
&= |x_i + \sum_{j\neq i}\boldsymbol{a}_i^\top \boldsymbol{a}_j x_j| \\
&\ge |x_i| - \sum_{j\neq i}|\boldsymbol{a}_i^\top \boldsymbol{a}_j||x_j| \\
&\ge |x_i| - \mu(\|\boldsymbol{x}\|_1 - |x_i|) \\
&\ge (1+\mu)|x_i| - \mu\|\boldsymbol{x}\|_1,
\end{aligned}$$

therefore, $\min_{1\le i\le k}|z_i| \ge (1+\mu)\min_i |x_i| - \mu\|\boldsymbol{x}\|_1$.

Upper bound $\max_{k+1 \le i \le n} |z_i|$

For $k+1 \le i \le n$,

$$\begin{aligned} |z_i| &= |\boldsymbol{a}_i^\mathsf{T} \boldsymbol{A} \boldsymbol{x}| \\ &= |\boldsymbol{a}_i^\mathsf{T} \sum_{j=1}^{k} \boldsymbol{a}_j x_j| \\ &\le \sum_{j=1}^{k} |\boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j||x_j| \\ &\le \mu \|\boldsymbol{x}\|_1 \end{aligned}$$

Putting everything together

OST succeeds if

$$(1+\mu)\min_i |x_i| - \mu\|\boldsymbol{x}\|_1 > \mu\|\boldsymbol{x}\|_1$$

which yields

$$(1+\mu)\min_i |x_i| > 2\mu\|\boldsymbol{x}\|_1.$$

or equivalently

$$\frac{\min_i |x_i|}{\|\boldsymbol{x}\|_1} > \frac{2\mu}{(1+\mu)}.$$

Better strategies?

- False alarms and miss detections are possible when the signal is weak and interference is high.
- It is obvious better approaches exist, for example, by applying iterations.

The idea is through iterations, we can either iteratively identify new atoms in the sparse representation, or refine our earlier estimate.

- Orthogonal Matching Pursuit (OMP)
- Iterative Hard Thresholding (IHT)
- Compressive Sampling Matching Pursuit (CoSaMP)

Orthogonal Matching Pursuit (OMP)

Orthogonal Matching Pursuit

Idea: select one index at every iteration.

Algorithm 5.2 Orthogonal Matching Pursuit (OMP)

Input: Sparsity level k.

Initialization: Let $\boldsymbol{r}_0 = \boldsymbol{y}$, and $S_0 = \emptyset$.

for $t = 1, \cdots, k$**:**

1. Choose the atom that has the largest correlation with the residual:
$$i_t = \text{argmax}_j |\langle \boldsymbol{a}_j, \boldsymbol{r}_{i-1} \rangle|$$
2. Add i_t to the support set: $S_t = \{S_{t-1}, i_t\}$;
3. Update the residual as
$$\boldsymbol{r}_t = \left(\boldsymbol{I} - \boldsymbol{A}_{S_t} \boldsymbol{A}_{S_t}^{\dagger}\right) \boldsymbol{y}.$$

Properties of OMP

- It doesn't select the same atom twice. If $j \in S_{t-1}$ has been selected,

$$\begin{aligned} \langle \boldsymbol{a}_j, \boldsymbol{r}_{t-1} \rangle &= \langle \boldsymbol{a}_j, (\boldsymbol{I} - \boldsymbol{A}_{S_t} \boldsymbol{A}_{S_t}^{\dagger}) \boldsymbol{y} \rangle \\ &= \boldsymbol{y}^{\mathsf{T}} (\boldsymbol{I} - \boldsymbol{A}_{S_t} \boldsymbol{A}_{S_t}^{\dagger}) \boldsymbol{a}_j = 0, \end{aligned}$$

  therefore j won't be selected again by OMP.
- If in each step OMP selects a correct index in T, in k iterations it will select all indices in T and terminates.
- An alternative way to terminate OMP (without the knowledge of k) is to examine the norm of the residual $\|\boldsymbol{r}_j\|_2 < \epsilon$.

Exact Recovery Condition (ERC) for OMP

Theorem 5.2 (ERC, Tropp 2004)

Suppose that $\boldsymbol{x}$ be a k-sparse signal supported on T. OMP recovers $\boldsymbol{x}$ whenever

$$\max_{\boldsymbol{a} \in T^c} \|\boldsymbol{A}_T^{\dagger} \boldsymbol{a}\|_1 < 1$$

where $\dagger$ denotes pseudo-inverse.

- This condition also guarantees the success of Basis Pursuit (ℓ_1 minimization), see [Tropp 2004].
- Interestingly enough, this condition only depends on $\boldsymbol{A}$, not on the coefficients of $\boldsymbol{x}$ - much improved from OST.
- A natural question is when does this condition hold?

Exact Recovery Condition (ERC)

Theorem 5.3 (Tropp 2004)

ERC holds for every *superposition of* k *atoms from* $\boldsymbol{A}$ *whenever*

$$k < \frac{1}{2}(\mu^{-1} + 1)$$

- Same condition that guarantees unique sparse solution for ℓ_0/ℓ_1 minimization.
- Since $\mu = O(\frac{1}{\sqrt{n}})$, we recover sparsity level up to $k \lesssim O(1/\sqrt{n})$.

Proof for Theorem 5.2

- Proceed by induction.
- After t steps, assume OMP has already identified t correct indices in T. We would like to develop a condition that guarantees the next selected atom is also in T.
- Motivated by our earlier discussions with OST, we only need to examine if the ratio

$$\rho(\boldsymbol{r}_t) = \frac{\left\|\boldsymbol{A}_{T^c}^\mathsf{T}\boldsymbol{r}_t\right\|_\infty}{\left\|\boldsymbol{A}_T^\mathsf{T}\boldsymbol{r}_t\right\|_\infty} < 1.$$

Realizing that $\boldsymbol{r}_t \in \mathsf{Span}(\boldsymbol{A}_T)$, we write

$$\boldsymbol{r}_t = \boldsymbol{A}_T\boldsymbol{A}_T^\dagger\boldsymbol{r}_t = \boldsymbol{A}_T(\boldsymbol{A}_T^\mathsf{T}\boldsymbol{A}_T)^{-1}\boldsymbol{A}_T^\mathsf{T}\boldsymbol{r}_t = (\boldsymbol{A}_T^\dagger)^\mathsf{T}\boldsymbol{A}_T^\mathsf{T}\boldsymbol{r}_t.$$

Proof for Theorem 5.2

- This allows us to bound

$$\rho(\boldsymbol{r}_t) = \frac{\left\|\boldsymbol{A}_{T^c}^\mathsf{T}\boldsymbol{r}_t\right\|_\infty}{\left\|\boldsymbol{A}_T^\mathsf{T}\boldsymbol{r}_t\right\|_\infty} \le \frac{\left\|\boldsymbol{A}_{T^c}^\mathsf{T}(\boldsymbol{A}_T^\dagger)^\mathsf{T}\boldsymbol{A}_T^\mathsf{T}\boldsymbol{r}_t\right\|_\infty}{\left\|\boldsymbol{A}_T^\mathsf{T}\boldsymbol{r}_t\right\|_\infty} \le \left\|\boldsymbol{A}_{T^c}^\mathsf{T}(\boldsymbol{A}_T^\dagger)^\mathsf{T}\right\|_{\infty,\infty}$$

  where the last inequality follows from the definition of the matrix norm $\|\cdot\|_{p,p}$

$$\|\boldsymbol{R}\|_{p,p} := \max_{\boldsymbol{x}} \frac{\|\boldsymbol{R}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}.$$

  It is easy to check (by yourself) that
  - $\|\boldsymbol{R}\|_{\infty,\infty}$ equals the maximum absolute row sum of $\boldsymbol{R}$;
  - $\|\boldsymbol{R}\|_{1,1}$ equals the maximum absolute column sum of $\boldsymbol{R}$;
- We have

$$\rho(\boldsymbol{r}_t) \le \left\|\boldsymbol{A}_{T^c}^\mathsf{T}(\boldsymbol{A}_T^\dagger)^\mathsf{T}\right\|_{\infty,\infty} = \left\|\boldsymbol{A}_T^\dagger\boldsymbol{A}_{T^c}\right\|_{1,1} = \max_{i\in T^c}\left\|\boldsymbol{A}_T^\dagger\boldsymbol{a}_i\right\|_1.$$

Proof of Theorem 5.3

- Recall the ERC can be upper bounded as

$$\max_{i\in T^c}\left\|\boldsymbol{A}_T^{\dagger}\boldsymbol{a}_i\right\|_1 = \max_{i\in T^c}\left\|(\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{A}_T)^{-1}\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{a}_i\right\|_1$$
$$\leq \left\|(\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{A}_T)^{-1}\right\|_{1,1}\max_{i\in T^c}\left\|\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{a}_i\right\|_1, \qquad (*)$$

where the second term can be bounded by the Babel function

$$\max_{i\in T^c}\left\|\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{a}_i\right\|_1 = \max_{i\in T^c}\sum_{j\in T}|\langle\boldsymbol{a}_j,\boldsymbol{a}_i\rangle| \leq k\mu.$$

- For the first term, we set off to write $\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{A}_T$ as

$$\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{A}_T = \boldsymbol{I}+\boldsymbol{\Phi}$$

where $\phi_{ij} = \langle\boldsymbol{a}_{T_i},\boldsymbol{a}_{T_j}\rangle$, and

$$\|\boldsymbol{\Phi}\|_{1,1} = \max_{l}\sum_{j\neq l}|\langle\boldsymbol{a}_{T_l},\boldsymbol{a}_{T_j}\rangle| \leq \mu(k-1).$$

Proof of Theorem 5.3 continued

If $\|\mathbf{\Phi}\|_{1,1} < 1$, the *von Neumann series* $\sum_{k=0}^{\infty}(-\mathbf{\Phi})^k$ converges to $(\boldsymbol{I} + \mathbf{\Phi})^{-1}$, we can compute

$$\begin{aligned}\left\|(\boldsymbol{A}_T^\mathsf{T}\boldsymbol{A}_T)^{-1}\right\|_{1,1} &= \left\|(\boldsymbol{I} + \mathbf{\Phi})^{-1}\right\|_{1,1} \\ &= \left\|\sum_{k=0}^{\infty}(-\mathbf{\Phi})^k\right\|_{1,1} \\ &\le \sum_{k=0}^{\infty}\|(-\mathbf{\Phi})\|_{1,1}^k = \frac{1}{1 - \|\mathbf{\Phi}\|_{1,1}} \le \frac{1}{1 - \mu(k-1)}.\end{aligned}$$

Plugging this into (*), a sufficient condition to guarantee ERC is

$$\frac{\mu k}{1 - \mu(k-1)} < 1$$

which gives $k < \frac{1}{2}(1 + \mu^{-1})$.

OMP Performance via RIP

Theorem 5.4 (OMP via RIP, Davenport and Wakin, 2010)

Suppose that $\boldsymbol{A}$ satisfies the RIP of order $k+1$ with isometry constant $\delta_{k+1} < \frac{1}{3\sqrt{k}}$. Then for any k-sparse signal $\boldsymbol{x}$, OMP will recover it exactly from in k iterations.

- Under Gaussian design, we can guarantee RIP constant with $n \gtrsim k \log p / \delta_k^2 = O(k^2 \log p)$ measurements.

Iterative Hard Thresholding (IHT)

IHT as Proximal Gradient Descent

Consider the non-convex optimization problem directly:

$$\min \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_0 \le k.$$

Solve by proximal gradient descent:

$$\min \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + g(\boldsymbol{x}), \quad \text{where} \quad g(\boldsymbol{x}) = \begin{cases} 0, & \text{if } \|\boldsymbol{x}\|_0 \le k \\ \infty, & \text{else} \end{cases}$$

- **gradient descent:**

$$\boldsymbol{z}^t \leftarrow \boldsymbol{x}^t - \mu_t \underbrace{\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{y})}_{\text{gradient of } \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2}$$

- **projection:** keep only k largest (in magnitude) entries

Iterative hard thresholding (IHT)

Algorithm 5.3 Iterative Hard Thresholding (IHT)

Input: Sparsity level k.

for $t = 0, 1, \cdots$**:**

$$\boldsymbol{x}^{t+1} = \mathcal{P}_k\left(\boldsymbol{x}^t - \mu_t \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{y})\right)$$

where $\mathcal{P}_k(\boldsymbol{x}) := \arg\min_{\|\boldsymbol{z}\|_0 = k} \|\boldsymbol{z} - \boldsymbol{x}\|$ is best k-term approximation of $\boldsymbol{x}$.

- For appropriate step size, it converges to a *local minimum* of

$$\min \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_0 \le k.$$

- Every iteration produces a k-sparse solution.

Linear convergence of IHT under RIP

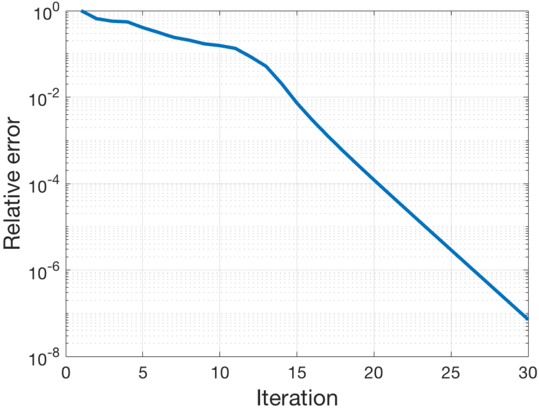
Theorem 5.5 (Blumensath & Davies '09)

Suppose $\boldsymbol{x}$ *is* k*-sparse, and RIP constant* $\delta_{3k} < 1/2$. *Then taking* $\mu_t \equiv 1$ *gives*

$$\left\| \boldsymbol{x}^t - \boldsymbol{x} \right\| \leq (2\delta_{3k})^t \left\| \boldsymbol{x}^0 - \boldsymbol{x} \right\|$$

- Under Gaussian design, need $n = O(k \log p)$ measurements.
- Under RIP, IHT attains ϵ-accuracy within $O\Big(\log \frac{1}{\epsilon}\Big)$ iterations
- Each iteration takes time proportional to a matrix-vector product

Numerical performance of IHT



Relative error $\frac{\|\boldsymbol{x}^t - \boldsymbol{x}\|}{\|\boldsymbol{x}\|}$ vs. iteration count t
$(n = 100,\ k = 5,\ p = 1000,\ A_{i,j} \sim \mathcal{N}(0, 1/n))$

Proof of Theorem 5.5

Let $\boldsymbol{z} := \boldsymbol{x}^t - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{y}) = \boldsymbol{x}^t - \boldsymbol{A}^\top\boldsymbol{A}(\boldsymbol{x}^t - \boldsymbol{x})$. By definition of $\mathcal{P}_k$,

$$\|\underbrace{\boldsymbol{x}}_{k\text{-sparse}} - \boldsymbol{z}\|^2 \geq \|\underbrace{\boldsymbol{x}^{t+1}}_{\text{best } k\text{-sparse}} - \boldsymbol{z}\|^2$$

$$= \|\boldsymbol{x}^{t+1} - \boldsymbol{x}\|^2 - 2\langle \boldsymbol{x}^{t+1} - \boldsymbol{x}, \boldsymbol{z} - \boldsymbol{x}\rangle + \|\boldsymbol{z} - \boldsymbol{x}\|^2$$

$$\begin{aligned}
\Longrightarrow \quad \|\boldsymbol{x}^{t+1} - \boldsymbol{x}\|^2 &\leq 2\langle \boldsymbol{x}^{t+1} - \boldsymbol{x},\ \boldsymbol{z} - \boldsymbol{x}\rangle \\
&= 2\Big\langle \boldsymbol{x}^{t+1} - \boldsymbol{x},\ (\boldsymbol{I} - \boldsymbol{A}^\top\boldsymbol{A})(\boldsymbol{x}^t - \boldsymbol{x})\Big\rangle \\
&\leq 2\delta_{3k}\|\boldsymbol{x}^{t+1} - \boldsymbol{x}\| \cdot \|\boldsymbol{x}^t - \boldsymbol{x}\| \qquad (5.1)
\end{aligned}$$

which gives

$$\|\boldsymbol{x}^{t+1} - \boldsymbol{x}\| \leq 2\delta_{3k}\|\boldsymbol{x}^t - \boldsymbol{x}\|$$

as claimed. Here, (5.1) follows from the following fact (homework)

$$|\langle \boldsymbol{u},\ (\boldsymbol{I} - \boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{v}\rangle| \leq \delta_s\|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\| \quad \text{with } s = |\text{supp}\,(\boldsymbol{u}) \cup \text{supp}\,(\boldsymbol{v})|$$

Compressive Sampling Matching Pursuit (CoSaMP)[1]

[1] See also *Subspace Pursuit* by Dai and Milenkovic

CoSaMP

Idea: add more to the support and then prune.

Algorithm 5.4 Compressive Sampling Matching Pursuit (CoSaMP)

Input: Sparsity level k.

Initialization: Let $\boldsymbol{r}_0 = \boldsymbol{y}$, $\boldsymbol{x}^0 = 0$, and $S = \emptyset$.

for $t = 1, 2, \cdots$**:**

1. Identify the support Ω_t of the $2k$ largest coefficients of
$$\boldsymbol{z}_t = \boldsymbol{A}^\top \boldsymbol{r}_{t-1};$$
2. Merge support: $S_t = \Omega_t \cup \text{supp}(\boldsymbol{x}^{t-1})$;
3. Least-squares estimation: $\boldsymbol{b}_S = \boldsymbol{A}_S^\dagger \boldsymbol{y}, \quad \boldsymbol{b}_{S^c} = 0$;
4. Prune: $\boldsymbol{x}^t = \boldsymbol{b}_k$ as the k-term approximation to $\boldsymbol{b}_S$;
5. Residual update: $\boldsymbol{r}_t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t$

Performance of CoSaMP via RIP

Theorem 5.6 (Needell and Tropp, 2008)

Assume $\boldsymbol{A}$ satisfies the RIP with $\delta_{2k} \leq 0.05$. For any k-sparse signal $\boldsymbol{x}$, the reconstruction in the tth iteration $\boldsymbol{x}^t$ is k-sparse, and satisfies

$$\left\|\boldsymbol{x}^{t+1} - \boldsymbol{x}\right\|_2 \leq 0.26 \cdot \|\boldsymbol{x}^t - \boldsymbol{x}\|_2.$$

Moreover, CoSaMP is exact after at most $6(k+1)$ iterations.

- Under Gaussian design, need $n = O(k \log p)$ measurements.
- Under RIP, CoSaMP attains ϵ-accuracy within $O\left(\log \frac{1}{\epsilon}\right)$ iterations
- Each iteration takes more time compared to IHT.

Iteration Count of CoSaMP

The number of iterations is at most $6(k+1)$, and could be as small as $\log k$.

It heavily relies on the coefficient profile.

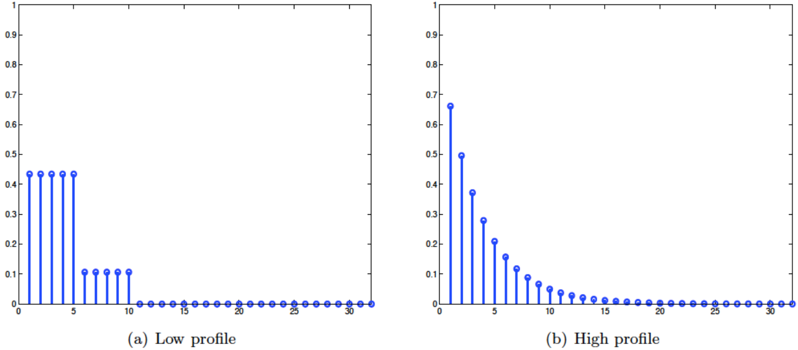


FIGURE 1. Illustration of two unit-norm signals with sharply different profiles.

Phase transition for inverse problems

Suppose $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ is i.i.d. Gaussian, and consider

$$\begin{aligned} \text{minimize}_{\boldsymbol{x} \in \mathbb{R}^p} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \end{aligned} \tag{5.2}$$

Key: using convex geometry

$$(5.2) \text{ succeeds}$$

$$\Updownarrow$$

$$\{\boldsymbol{h} : \boldsymbol{A}\boldsymbol{h} = \boldsymbol{0}\} \cap \mathcal{D}(f, \boldsymbol{x}) = \{\boldsymbol{0}\}$$

$$\Updownarrow$$

$$\underbrace{\text{stat-dim}(\{\boldsymbol{h} : \boldsymbol{A}\boldsymbol{h} = \boldsymbol{0}\})}_{= p-n} + \text{stat-dim}(\mathcal{D}(f, \boldsymbol{x})) \leq p \quad \text{(by Theorem ??)}$$

Phase transition for inverse problems

Suppose $\boldsymbol{A} \in \mathbb{R}^{n\times p}$ is i.i.d. Gaussian, and consider

$$\begin{array}{rl} \text{minimize}_{\boldsymbol{x}\in\mathbb{R}^p} & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \end{array} \tag{5.2}$$

Theorem 5.7 (Amelunxen, Lotz, McCoy & Tropp '13)

$$n > \textit{stat-dim}(\mathcal{D}(f,\boldsymbol{x})) + \Theta(\sqrt{p\log p})$$
$$\Longrightarrow \qquad (5.2) \text{ succeeds with high prob.}$$
$$n < \textit{stat-dim}(\mathcal{D}(f,\boldsymbol{x})) - \Theta(\sqrt{p\log p})$$
$$\Longrightarrow \qquad (5.2) \text{ fails with high prob.}$$

Statistical dimension:

$$\text{stat-dim}\Big(\mathcal{D}\left(\|\cdot\|_1, \boldsymbol{x}\right)\Big) = \inf_{\tau\geq 0}\left\{k\left(1+\tau^2\right) + (p-k)\sqrt{\frac{2}{\pi}}\int_{\tau}^{\infty}(z-\tau)^2 e^{-z^2}\mathrm{d}z\right\}$$

Numerical phase transition

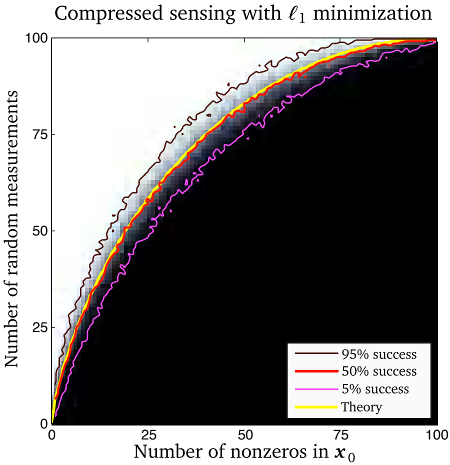


Figure credit: Amelunxen, Lotz, McCoy, & Tropp '13

Benchmark result by Donoho and Maleki

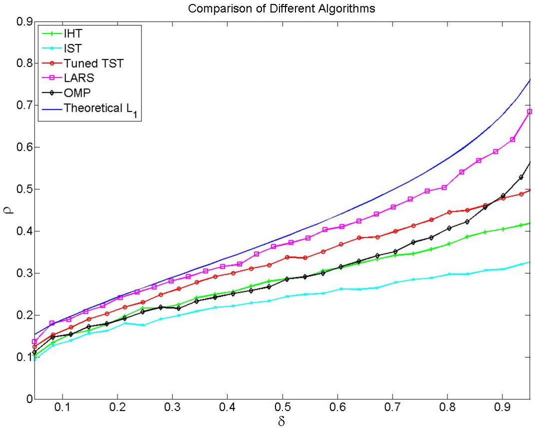


Figure 5.1: Phase Transitions of several algorithms at the standard suite. ρ is sparsity level and δ is subsampling ratio. [Donoho and Maleki, 2009].

References

[1] "*Greed is good: Algorithmic results for sparse approximation*," J. Tropp, *IEEE Transactions on Information Theory*, 2004.

[2] "*CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*," D. Needell and J. Tropp, *Applied and Computational Harmonic Analysis*, 2009.

[3] "*Orthogonal matching pursuit for sparse signal recovery with noise*," T. T. Cai and L. Wang, *IEEE Transactions on Information Theory*, 2011.

[4] "*Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property*," M. Davenport, M. Wakin, *IEEE Transactions on Information Theory*, 2010.

[5] "*Iterative hard thresholding for compressed sensing*," T. Blumensath, M. Davies, *Applied and computational harmonic analysis*, 2009.

[6] `https://sparselab.stanford.edu/`