Basic Algorithms

I-Liang Chern

September 22, 2016

Three categories of Algorithms in CS

The noiseless compressive sensing problem

$$\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_0 \text{ subject to } A\mathbf{z} = \mathbf{y}. \qquad \text{(P0)}$$

is a combinatorial optimization problem. One needs to search over all possible subindex set $S \subset [N]$. The numbers of steps are $O(2^N)$. This is NP-hard. We therefore look for other algorithms. There are three categories of algorithms:

- I. Optimization methods
- II. Greedy methods
- III. Thresholding-based methods

I. Optimization methods-1

- I. Noiseless Case: Instead of solving

$$\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_0 \text{ subject to } A\mathbf{z} = \mathbf{y}. \qquad \text{(P0)}$$

We solve the convex relaxation problem:

$$\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_1 \text{ subject to } A\mathbf{z} = \mathbf{y}. \qquad \text{(P1)}$$

This is called basis pursuit. It is a convex optimization problem. There are many algorithms available to solve this constrained optimization problem. We will discuss these algorithms in Chapter 15.

I. Optimization methods-2

- II. Noisy Case: Suppose the measurement has noise with noise level η. Instead of solving

$$\boxed{\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_0 \text{ subject to } \|A\mathbf{z}-\mathbf{y}\| \leq \eta}$$

we solve the quadratically constrained basis pursuit:

$$\boxed{\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_1 \text{ subject to } \|A\mathbf{z}-\mathbf{y}\|_2 \leq \eta.}$$

Equivalence-1

There are three equivalent versions of such constrained optimization optimization.

1. Quadratically constrained basis pursuit:

$$\boxed{\min \|\mathbf{z}\|_1 \text{ subject to } \|A\mathbf{z} = \mathbf{y}\|_2 \leq \eta;}$$

2. Basis pursuit denoising

$$\boxed{\min \lambda\|\mathbf{z}\|_1 + \|A\mathbf{z} - \mathbf{y}\|_2^2;}$$

3. LASSO:

$$\boxed{\min \|A\mathbf{z} - \mathbf{y}\|_2 \text{ subject to } \|\mathbf{z}\|_1 \leq \tau.}$$

Equivalence-2

(1) $\min \|\mathbf{z}\|_1$ subject to $\|A\mathbf{z} - \mathbf{y}\|_2 \leq \eta$;

(2) $\min \lambda\|\mathbf{z}\|_1 + \|A\mathbf{z} - \mathbf{y}\|_2^2$;

(3) $\min \|A\mathbf{z} - \mathbf{y}\|_2$ subject to $\|\mathbf{z}\|_1 \leq \tau$.

Proposition

- *(2) $\Rightarrow$ (1): If $\mathbf{x}$ is a minimizer of (2) with some $\lambda > 0$, then $\exists\, \eta_x$ such that $\mathbf{x}$ is the minimizer of (1).*
- *(3) $\Rightarrow$ (2): If $\mathbf{x}$ is a minimizer of (3) with some $\tau > 0$, then $\exists\, \lambda_x > 0$ such that $\mathbf{x}$ is a minimizer of (2)*
- *(1) $\Rightarrow$ (3): If $\mathbf{x}$ is the unique minimizer of (1) with some $\eta > 0$, then $\exists\, \tau_x$ such that $\mathbf{x}$ is a unique minimizer of (3).*

(2)⇒(1):

1. Set $\eta_x = \|A\mathbf{x} - \mathbf{y}\|$. Consider any $\mathbf{z}$ with $\|A\mathbf{z} - \mathbf{y}\| \leq \eta$.
2. $\lambda\|\mathbf{x}\|_1 + \|A\mathbf{x} - \mathbf{y}\|^2 \leq \lambda\|\mathbf{z}\|_1 + \|A\mathbf{z} - \mathbf{y}\|^2 \leq \lambda\|\mathbf{z}\|_1 + \|A\mathbf{x} - \mathbf{y}\|^2$. This leads to $\|\mathbf{x}\|_1 \leq \|\mathbf{z}\|_1$.

(3) ⇒ (2):

λ is the Lagrange multiplier for constrained optimization.

(1) ⇒ (3):

If $\mathbf{x}$ is the unique minimizer of (1), then set $\|x\|_1 = \tau$. For any $\|\mathbf{z}\|_1 \leq \tau$, since $\mathbf{x}$ is the unique minimizer satisfying the constraint, $\mathbf{z}$ cannot satisfy the constraint. Thus, $\|A\mathbf{z} - \mathbf{y}\|_2 > \eta \geq \|A\mathbf{x} - \mathbf{y}\|_2$. Thus, $\mathbf{x}$ is the unique minimizer of (3).

(3) ⇒ (1):

If $\mathbf{x}$ is the unique minimizer of (3), let us set $\|A\mathbf{x} - \mathbf{y}\|_2 = \eta$ and $\|x\|_1 = \tau$. For any $\mathbf{z}$ satisfying $\|A\mathbf{z} - \mathbf{y}\|_2 \leq \eta$, it cannot satisfy the constraint of (3), otherwise it violates the uniqueness of the minimizer of (3). Thus, $\|\mathbf{z}\|_1 > \tau = \|\mathbf{x}\|_1$. We obtain $\mathbf{x}$ is the minimizer of (1).

Optimization methods-3: Dantzig Selector

- In statistics, there is another kind model selector called Dantzig selector:

$$\boxed{\min \|\mathbf{z}\|_1 \text{ subject to } \|\mathbf{A}^*(\mathbf{A}\mathbf{z}-\mathbf{y})\|_\infty \leq \tau.}$$

- The residual $\mathbf{r} = \mathbf{A}\mathbf{z} - \mathbf{y}$ should have small correlation with all columns $\mathbf{a}_j$ of the matrix $\mathbf{A}$. Indeed,

$$\|\mathbf{A}^*(\mathbf{A}\mathbf{z}-\mathbf{y})\|_\infty = \max_j |\langle \mathbf{a}_j, \mathbf{r}\rangle|$$

- Dantzig selector is a convex optimization problem [Candes-Tao].
- The theory for basis pursuit is also valid for Dantzig selector.

II. Greedy Methods

- Suppose $\mathbf{x}$ an s-sparse vector and $\mathbf{A}$ an $m \times N$ matrix. We are given the measurement $\mathbf{y} = \mathbf{A}\mathbf{x}$. We want to recover $\mathbf{x}$.
- We introduce two greedy methods[1] [2]
  - Orthogonal Matching Pursuit
  - Compressive Sampling Matching Pursuit
- The goal is to find conditions on $\mathbf{A}$ such that these two methods can recover $\mathbf{x}$ exactly from $\mathbf{y}$.

---

[1] S. Mallat and Zhifeng Zhang, Matching Pursuirs with time-frequency dictionaries (1993)

[2] J. Tropp and Anna Gilbert, Signal recovery from random measurements via orthogonal matching pursuit (2007)

Matching Pursuits for Sparse Representation

Before introducing Orthogonal matching pursuit, we introduce matching pursuit for sparse representation.

- Goal: given signal $\mathbf{y}$ and an over-complete dictionary $D = \{\mathbf{a}_1, \mathbf{a}_2, \cdots\}$, where $\mathbf{a}_i$ with $\|\mathbf{a}_i\| = 1$ are called the atoms of D. The matching pursuit looks for a representation of $\mathbf{y}$ in terms of $\sum_i x_i \mathbf{a}_i$ that is sparse and approximates $\mathbf{y}$.
- Matching pursuit will first find the one atom that has the biggest inner product with the signal, then subtract the contribution due to that atom, and repeat the process until the signal is satisfactorily decomposed.

Matching Pursuit for orthonormal systems

- Input: $\mathbf{A}$ (orthonormal column vectors), $\mathbf{y}$;
- Output: $\mathbf{x}^{\#}$
- Initialization: $\mathbf{r}^0 = \mathbf{y}$, $\mathbf{x} \equiv 0$
- Iteration: stop when $\|\mathbf{r}^n\| \leq \epsilon$
  - $j_{n+1} := \text{argmax}_{j \in [N]} |\langle \mathbf{a}_j, \mathbf{r}^n \rangle|$
  - $x_{j_{n+1}} = \langle \mathbf{r}^n, \mathbf{a}_{j_{n+1}} \rangle$
  - $\mathbf{r}^{n+1} := \mathbf{r}^n - x_{j_{n+1}} \mathbf{a}_{j_{n+1}}$

Orthogonal Matching Pursuit

- Input: $\mathbf{A}$ (normalized column vectors $\|\mathbf{a}_j\|_2 = 1$) and $\mathbf{y}$.
- Initialization: $S^0 = \emptyset$, $\mathbf{x}_0 = 0$.
- Iteration: stop when $n = \bar{n}$
  - $S^{n+1} = S^n \cup \{j_{n+1}\}$, $j_{n+1} := \operatorname{argmax}_{j \in [N]} |\langle \mathbf{a}_j, \mathbf{y} - \mathbf{A}\mathbf{x}^n \rangle|$
  - $\mathbf{x}^{n+1} := \operatorname{argmin}_{\mathbf{z}} \{\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2, \operatorname{supp}(\mathbf{z}) \subset S^{n+1}\}$
- Output: $\mathbf{x}^{\#} = \mathbf{x}^{\bar{n}}$.

Remarks

- The step $\min\{\|\mathbf{Az}-\mathbf{y}\|_2, \text{supp}\,(\mathbf{z}) \subset S^{n+1}\}$ is computationally costly. There are fast method to speed up the computation: e.g. an efficient QR decomposition for A_{S^n} when a new column is added;
- The choice of the index j_{n+1} is dictated by a greedy strategy where one aims to reduce the L^2-norm of the residual $\mathbf{y}-\mathbf{Ax}^n$ as much as possible at each iteration.

Lemma

Given $S \subset [N]$, *if*

$$\mathbf{v} := \textit{argmin}\left\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2, \textit{supp}\,(\mathbf{z}) \subset S\right\}$$

then for any $j \in S$, $\langle \mathbf{a}_j, \mathbf{y} - \mathbf{A}\mathbf{v} \rangle = 0$

Proof. This is the Euler-Lagrange equation for the minimization problem. Indeed, for any $\mathbf{h}$ with $\text{supp}\,(\mathbf{h}) \subset S$,

$$\langle \mathbf{A}\mathbf{h}, \mathbf{y} - \mathbf{A}\mathbf{v} \rangle = 0.$$

Taking $\mathbf{h} = \mathbf{e}_j$ with $j \in S$, we get

$$\langle \mathbf{a}_j, \mathbf{y} - \mathbf{A}\mathbf{v} \rangle = \langle \mathbf{A}\mathbf{e}_j, \mathbf{y} - \mathbf{A}\mathbf{v} \rangle = 0.$$

Lemma

Let A be a matrix with normalized column vectors, $S \subset [N]$, $\mathbf{v} \in \mathbb{C}^N$ with support in S, and $j \in [N]$. If

$$\mathbf{w} := \textit{argmin}\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \,|\, \textit{supp}\,(\mathbf{z}) \subset S \cup \{j\}\}$$

then

$$\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 \leq \|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 - |\langle \mathbf{a}_j, \mathbf{y} - \mathbf{A}\mathbf{v}\rangle|^2.$$

Proof. (for real case)

$$\begin{aligned}
\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 &\leq \min_t \|\mathbf{y} - \mathbf{A}(\mathbf{v} + t\mathbf{e}_j)\|_2^2 \\
&= \min_t \left(\|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 - 2t\langle \mathbf{y} - \mathbf{A}\mathbf{v}, \mathbf{a}_j\rangle + t^2\|\mathbf{a}_j\|^2\right) \\
&= \|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 - |\langle \mathbf{y} - \mathbf{A}\mathbf{v}, \mathbf{a}_j\rangle|^2.
\end{aligned}$$

Proposition

Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ *and* $S \subset [N]$ *with* $|S| = s$. *Every* $0 \neq \mathbf{x}$ *supported on* S *is uniquely recovered from* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *after at most* s *iterations of OMP if and only if the matrix* $\mathbf{A}_S$ *is injective and*

$$\max_{j \in S} |(\mathbf{A}^* \mathbf{r})_j| > \max_{\ell \in \bar{S}} |(\mathbf{A}^* \mathbf{r})_\ell| \tag{0.1}$$

for all $\mathbf{r} = \mathbf{A}\mathbf{z}$ *with supp*$(\mathbf{z}) \subset S$.

Remark.

- $(\mathbf{A}^* \mathbf{r})_j = \langle \mathbf{a}_j, \mathbf{r} \rangle$ is the correlation between the residual and model vector $\mathbf{a}_j$.

Proof.

($\Rightarrow$)

1. $\mathbf{A}_S$ is injective: Suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ have the same support in S and $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$. Then through OMP, $\mathbf{x}_i$ can be recovered from $\mathbf{y}$ exactly. Thus, they are identical.
2. Since the index chosen at the first iteration always stays in the target support S, if $\mathbf{r} = \mathbf{A}\mathbf{z}$ for some $\mathbf{z}$ with support exact S, then an index $\ell \notin S$ cannot be chosen at the first iteration (otherwise, we cannot find $\mathbf{x}$ in s iteration). That is,

   $$|\langle \mathbf{a}_\ell, \mathbf{r} \rangle| < \max_{j \in S} |\langle \mathbf{a}_j, \mathbf{r} \rangle|$$

   This shows

   $$\max_{\ell \in \bar{S}} |\langle \mathbf{a}_\ell, \mathbf{r} \rangle| < \max_{j \in S} |\langle \mathbf{a}_j, \mathbf{r} \rangle|.$$

$(\Leftarrow)$:

1. Assuming $\mathbf{A}\mathbf{x}^n \neq \mathbf{y}$ for $n = 1, ..., s-1$, otherwise we have done. We want to show that $\mathbf{A}\mathbf{x}^s = \mathbf{y}$.
2. First, we claim that $S^n \subset S$ for $n = 1, ..., s$. This is proven inductively in n. If $1 \leq n \leq s-1$ and $S^n \subset S$, then

$$\mathbf{r}_n = \mathbf{y} - \mathbf{A}\mathbf{x}^n \in \{\mathbf{A}\mathbf{z} \mid \text{supp}\,(\mathbf{z}) \subset S\}.$$

   By our assumption, $j^{n+1} \subset S$. Hence $S^{n+1} = S^n \cup \{j_{n+1}\} \subset S$.

3 We claim that $|S^n| = n$ for $n = 1, ..., s$. We have seen that $\mathbf{r}^n := \mathbf{A}\mathbf{x}^n \perp \mathbf{a}_j$, for $j \in S^n$. If $j^{n+1} \subset S^n$, then

$$\max_{j \notin S^n} |\langle \mathbf{r}^n, \mathbf{a}_j \rangle| \leq \max_{j \in S^n} |\langle \mathbf{r}^n, \mathbf{a}_j \rangle| = 0.$$

This implies $\mathbf{r}^n = 0$. Thus, $|S^n| = n$.

4 Since $|S| = s$ and $|S^n| = n$, $S^n \subset S$. We conclude $S^s = S$.

Remark. The above proposition is equivalent to

$$\boxed{\|\mathbf{A}_S^\dagger \mathbf{A}_{\bar{S}}\|_{1\to 1} < 1.}$$

Proof.

1. $\mathbf{A}_S^\dagger := (\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*$ exists iff $\mathbf{A}_S$ is 1-1.
2. The condition (0.1) is equivalent to

$$\|\mathbf{A}_S^*\mathbf{A}_S\mathbf{u}\|_\infty > \|\mathbf{A}_{\bar{S}}^*\mathbf{A}_S\mathbf{u}\|_\infty \quad \forall\, 0 \neq \mathbf{u} \in \mathbb{C}^N.$$

3. Let $\mathbf{v} = \mathbf{A}_S^*\mathbf{A}_S\mathbf{u}$, then the above formula is

$$\|\mathbf{v}\|_\infty > \|\mathbf{A}_{\bar{S}}^*(\mathbf{A}_S^\dagger)^*\mathbf{v}\|_\infty$$

   This reads $\|\mathbf{A}_{\bar{S}}^*(\mathbf{A}_S^\dagger)^*\|_{\infty\to\infty} < 1$. This is equivalent to $\|\mathbf{A}_S^\dagger\mathbf{A}_{\bar{S}}\|_{1\to1} < 1$.

Weakness of OMP

1. You may not know s;
2. $\mathbf{A}$ may not satisfy (0.1):

$$\max_{j \in S} |(\mathbf{A}^*\mathbf{r})_j| > \max_{\ell \in \bar{S}} |(\mathbf{A}^*\mathbf{r})_\ell|$$

3. If you choose a wrong index j, it can not be removed.

An improved algorithm called Compressive Sampling Matching Pursuit.

Compressive Sampling Orthogonal Matching Pursuit (CoSaMP)-1

- Input: $\mathbf{A}, \mathbf{y}$ and s.
- Initialization: $\mathbf{x}^0 = 0$
- Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:
  - $U^{n+1} = \text{supp}\,(\mathbf{x}^n) \cup L_{2s}(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))$
  - $\mathbf{u}^{n+1} = \text{argmin}_{\mathbf{z}}\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \text{supp}\,(\mathbf{z}) \subset U^{n+1}\}$.
  - $\mathbf{x}^{n+1} = H_s(\mathbf{u}^{n+1})$.
- Output: $\mathbf{x}^{\#} = \mathbf{x}^{\bar{n}}$.

3

[3]Needel and Troop

CoSaMP-2

In CoSaMP, we use a notation called hard thresholding defined as the follows.

- Given $\mathbf{x} \in \mathbb{C}^N$, we define its best s sparse approximation $H_s(\mathbf{x})$ to be
  - $L_s(\mathbf{x}) :=$ index set of the s largest $|x_i|$.
  - $H_s(\mathbf{x}) := \mathbf{x}_{L_s(\mathbf{x})}$.
- The operator H_s is called the hard thresholding operator.

[4]

[4] For review, see monograph [472] by Temlyakov, Tropp in [476], Mallat, and Zhang in [145]

III. Thresholding based methods

- Basic thresholding method
- Iterative thresholding Method
- Hard thresholding pursuit

Basic Thresholding Method

1. Input: $\mathbf{A}$, $\mathbf{y}$ and s
2. Procedure:
   - $S^{\#} = L_s(\mathbf{A}^*\mathbf{y})$;
   - $\mathbf{x}^{\#} = \min\{\|\mathbf{y} - \mathbf{Az}\|^2 \mid \text{supp}\,(\mathbf{z}) \subset S^{\#}\}$
3. Output: $\mathbf{x}^{\#}$.

Proposition

A vector $\mathbf{x} \in \mathbb{C}^N$ *supported on a set* S *is recovered from* $\mathbf{y} = \mathbf{Ax}$ *via basic thresholding if and only if*

$$\min_{j\in S} |\langle \mathbf{a}_j, \mathbf{y}\rangle| > \max_{\ell\in\bar{S}} |\langle \mathbf{a}_\ell, \mathbf{y}\rangle|$$

Proof. $S^{\#} = S$ means the above inequality.

Iterative Thresholding Method

1. Input: $\mathbf{A}$, $\mathbf{y}$ and s
2. Initialization: $\mathbf{x}^0 = 0$
3. Repeat until a stopping criterion is met at $n = \bar{n}$

$$\mathbf{x}^{n+1} = H_s(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))$$

4. Output: $\mathbf{x}^{\#} = \mathbf{x}^{\bar{n}}$.

Remark:

- $\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x})$ is the negative gradient of $\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2$.
- We can also replace $\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)$ by $\tau^n \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)$ so that it is the greatest decent.

Hard Thresholding Pursuit

1. Input: $\mathbf{A}$, $\mathbf{y}$ and s
2. Initialization: $\mathbf{x}^0 = 0$
3. Repeat until a stopping criterion is met at $n = \bar{n}$:
   - $S^{n+1} = L_s(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))$
   - $\mathbf{x}^{n+1} = \text{argmin}\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|^2 \mid \text{supp}\,(\mathbf{z}) \subset S^{n+1}\}$
4. Output $\mathbf{x}^{\#} = \mathbf{x}^{\bar{n}}$.