On $\min \|\mathbf{x}\|_0$ s.t. $\mathbf{Ax} = \mathbf{b}$

Andersen Ang

Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Canada

ms$\mathbf{x}$ang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

First draft: August 18, 2017
Last update: March 21, 2021

System of linear equation

- Given $\mathbf{A}, \mathbf{b}$ find $\mathbf{x}$ such that

$$(\mathcal{P}) \;:\; \mathbf{A}\mathbf{x} = \mathbf{b}$$

- $\mathbf{A} \in \mathbb{R}^{m \times n}$ is given
  - $m = n$: square matrix
  - $m > n$: tall-thin matrix
  - $m < n$: short-fat matrix
- $\mathbf{b} \in \mathbb{R}^{m}$ is the given
- $\mathbf{x} \in \mathbb{R}^{n}$ is the unknown

When $\mathbf{A}$ is short-fat

$$(\mathcal{P}) \; : \; \mathbf{A}\mathbf{x} = \mathbf{b}.$$

- If $m < n$, $\mathbf{A}$ has more columns than rows and there are free variables in the solution, meaning there are many sols.
- In set language: if $\mathbf{x}_0$ is a sol. of $(\mathcal{P})$, then the complete set of solution for $(\mathcal{P})$ is the set

$$S = \left\{ \mathbf{x}_0 + \mathbf{v} \;\middle|\; \mathbf{A}\mathbf{x}_0 = \mathbf{b}, \mathbf{v} \in \mathsf{null}(\mathbf{A}) \right\},$$

  where $\mathsf{null}(\mathbf{A})$ denotes the null space of $\mathbf{A}$, and the size of $\mathsf{null}(\mathbf{A})$ is infinity.
- For the case $m < n$, if there is no other additional information on $(\mathcal{P})$, all the sol. in S are "equal", none of them are more important nor less important.

Ranking sol. based on their sparsity

$$(\mathcal{P}) \ : \ \mathbf{A}\mathbf{x} = \mathbf{b}, \quad S = \Big\{ \mathbf{x}_0 + \mathbf{v} \ \Big| \ \mathbf{A}\mathbf{x}_0 = \mathbf{b}, \mathbf{v} \in \mathsf{null}(\mathbf{A}) \Big\}.$$

- Among all the sol. available inside the set S, we can say that those that are "sparser" are "better".
- In this case we arrive at a new problem

  $$(\mathcal{P}') \ : \ \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}.$$

  meaning that, we want to find the "sparest" solution $\mathbf{x}$ in the set S that satisfies $\mathbf{A}\mathbf{x} = \mathbf{b}$.
- $\|\mathbf{x}\|_0$ is the L_0-norm of $\mathbf{x}$, which is the sparsity of $\mathbf{x}$.
- Problem $(\mathcal{P}')$ is called Sparse recovery problem, it is the key theme in the field of compressive sensing.

Sparsity and L_0-norm

- Given a vector $\mathbf{u}$, the sparsity of $\mathbf{u}$ is the number of non-zero elements in $\mathbf{u}$.
- Example: a 3-dimensional vector $\mathbf{w} = [1, 0.5, 0]$ has sparsity 2.
- The index set of non-zero element of a vector $\mathbf{x}$ is denoted as $\mathrm{supp}(\mathbf{x}) = \{i \mid x_i \neq 0\}$. For the example $\mathbf{w}$, $\mathrm{supp}(\mathbf{w}) = \{1, 2\}$.
- Sparsity can be represented as the size of $\mathrm{supp}(\mathbf{x})$, denoted as $|\mathrm{supp}(\mathbf{x})|$ where $|\cdot|$ denotes the cardinality.
- For shorter notation, we can just use L_0-norm to denote the sparsity of $\mathbf{u}$ as $\|\mathbf{u}\|_0$.
- A related concept of sparsity is compressibility. A vector is compressible means that vector is "nearly sparse".

The L_0 counting "norm"

- Note that L_0-norm is not really a norm.
- Definition of norm: a function $p(\mathbf{x})$ is called a norm if it fulfills the following conditions
  - $p(\mathbf{x}) \geq 0$ (nonnegativity)
  - $p(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
  - $p(\mathbf{x} + \mathbf{y}) \leq p(\mathbf{x}) + p(\mathbf{y})$ (triangle inequality)
  - $p(\alpha \mathbf{x}) = |\alpha| p(\mathbf{x})$ (Homogeneous)
- L_0-norm is not Homogeneous : e.g. $\mathbf{x} = [0, 1]$, then $\|3\mathbf{x}\|_0 = \|\mathbf{x}\|_0 = 1 \neq 3 = 3\|\mathbf{x}\|_0$.

The combinatorial nature of L_0 norm

$$(\mathcal{P}') \ : \ \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}.$$

- Problem $(\mathcal{P}')$ is NP-hard due to the combinatorial nature of L_0 norm.
- Illustration: consider the simplest case that the solution to $(\mathcal{P}')$ is 1-sparse. That is, let $\mathbf{x}$ be the sol. to $(\mathcal{P}')$, then $\mathbf{x}$ only has 1 non-zero element.
- Note that we don't know where is the location of such non-zero element, in this situation, we can only try all the possible cases to find the solution:

$$\mathbf{A}\begin{bmatrix} ? \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{b}, \quad \mathbf{A}\begin{bmatrix} 0 \\ ? \\ \vdots \\ 0 \end{bmatrix} = \mathbf{b}, \quad \ldots \mathbf{A}\begin{bmatrix} 0 \\ 0 \\ \vdots \\ ? \end{bmatrix} = \mathbf{b}.$$

We can then solve these n problems to find the unknown.

The combinatorial nature of L_0 norm

- If we assume the solution to $(\mathcal{P}')$ is 2-sparse. There are $\binom{n}{2}$ number of possibilities on the location of the non-zero entries in the solution.

- In general, we do not know the true sparsity of the solution. We only know "the solution is sparse", and we can only guess the sol. is at most k-sparse.

- If we assume the solution to $(\mathcal{P}')$ is k-sparse, the total number of possible sub-problems is $\sum_i^k \binom{n}{i}$, Such number is in factorial, which grows (much) faster than the polynomial time complexity, so L_0-norm sparse recovery problem is a NP-hard problem.

Convex relaxation

- The L_0-norm problem is NP-hard

$$(\mathcal{P}') \ : \ \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}.$$

- We can consider replacing the L_0-norm to L_1-norm

$$(\mathcal{P}'') \ : \ \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b},$$

  where L_1-norm is just the sum of absolute value of the elements

$$\|\mathbf{x}\|_1 = \sum_{i}^{n} |x_i|.$$

- $(\mathcal{P}'')$ is much easier to solve because it is a convex problem.
- It can be shown under some conditions (on $\mathbf{A}$ and $\mathbf{x}$), the solution of $(\mathcal{P}'')$ is also the solution of $(\mathcal{P}')$

Algorithms

- Matching Pursuit.
- Orthogonal Matching Pursuit.
- Proximal gradient method: soft-thresholding on L_1-norm problem.
- Proximal gradient method: hard-thresholding on L_0-norm problem.
- Iterative re-weighted least squares.
- Details of these algorithms will be discussed in other documents.

Last page - summary

- $(\mathcal{P}) \ : \ \mathbf{Ax} = \mathbf{b}$ with $m < n$ has infinitely many solutions.
- Sparsity of a vector
- The L_0-norm
- $(\mathcal{P}') \ : \ \min \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{Ax} = \mathbf{b}$ for $m < n$ is NP-hard.

End of document