PhD Course in KTH - Sparse Signal Processing
Slides 5
Discussion Topic - Exact to Approximate

Saikat Chatterjee

Communication Theory Lab, KTH

March 13, 2014

General Motivation

1. The exact constraint $\mathbf{Ax} = \mathbf{b}$ is often relaxed, with a quadratic penalty $Q(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2$.
2. Such relaxation allows us to
   - define a quasi-solution in case no exact solution exists (even in the case of an over-determined setup),
   - exploit ideas from optimization theory, and
   - measure the quality of a candidate solution.

Therefore, we relax the (P_0) problem with the use of an error tolerance $\epsilon > 0$,

$$(P_0^\epsilon): \quad \arg\min_{\mathbf{x}\in\mathbb{R}^m} \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{b} - \mathbf{Ax}\|_2 \leq \epsilon. \tag{1}$$

Remark

A comment: When (P_0) and (P_0^ϵ) are applied on the same problem instance, the error-tolerent problem (P_0^ϵ) must always provide results at-least as sparse as those arising in the exact constrained problem (P_0), since the feasible solution set is wider.

Remark

An alternative interpretation: Interpreting the problem (P_0^ϵ) as a noise removal scheme. Consider a sufficiently sparse vector $\mathbf{x}_0$, and assume that $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$, where $\mathbf{e}$ is a nisance vector of finite energy $\|\mathbf{e}\|_2^2 = \epsilon^2$. Rougly speaking (P_0^ϵ) aims to find $\mathbf{x}_0$, i.e., to do rougly the same thing as (P_0) would be on noiseless data $\mathbf{b} = \mathbf{A}\mathbf{x}_0$.

Our rational thought process

1. We can have a rationale that the results for (P_0^ϵ) are some ways parallel to those in the noiseless case (P_0).
2. Specifically, we should discuss the uniqueness property - conditions under which a sparse solution is known to be the global minimizer of (P_0^ϵ) and hence the true solution.

Stability of the sparsest solution

Remark

A fundamental question: Suppose that a sparse vector $\mathbf{x}_0$ *is pre-multiplied by* $\mathbf{A}$*, and we obtain a noise version as* $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$ *with* $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon$*. Let*

$$\mathbf{x}_0^\epsilon = \arg\min_{\mathbf{x}} \ \|\mathbf{x}\|_0 \ \text{subject to} \ \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon. \tag{2}$$

1. *How good shall this approximation be?*
2. *How its accuracy is affected by the sparsity of* $\mathbf{x}_0$*?*

These questions are the natural extension from the uniqueness porperty of (P_0)*.*

Uniqueness versus stability - Gaining intuition

Remark

A question: Can we have the sparsest solution for (P_0^ϵ) unique?

Answer: No, we can not claim uniqueness for the solution of(P_0^ϵ).

Illustration

1. Practical illustration by following Fig. 5.1 of the book.
2. We discuss about Fig. 5.2 of the book where noise strength is high and we note
   - We can have a different support solution with same sparsity level.
   - Even null solution can be a solution.

Theoretical Study of stability of (P_0^ϵ)

- Instead of claiming uniqueness of a sparse solution, we try to be happy with a notion of stability - a claim that if a sufficiently spase solution is found, then all alternative solutions necessarily resides (very) close to it.
- Starting point: Extending the notion of 'spark' by considering a relaxed notion of linear dependency.
- Recall that *spark*$(\mathbf{A})$ is the minimum number of linearly dependent columns. Mathematically, it was defined as

$$spark(\mathbf{A}) = \min_{\mathbf{x}\in\mathbb{R}^m} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{0}, \;\; \mathbf{x} \neq \mathbf{0}. \tag{3}$$

- If $\mathbf{x}_1$ and $\mathbf{x}_2$ are two solutions for noiseless case $\mathbf{A}\mathbf{x} = \mathbf{b}$, then we can have $\mathbf{d} = \mathbf{x}_1 - \mathbf{x}_2$ and $\mathbf{A}\mathbf{d} = \mathbf{0}$. This motivates the null space characterization.

- Following the same rationale, if there exists two feasible solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ satisfying $\|\mathbf{A}\mathbf{x}_i - \mathbf{b}\|_2 \leq \epsilon$, $i = 1, 2$, then we can have $\mathbf{d} = \mathbf{x}_1 - \mathbf{x}_2$ and $\|\mathbf{A}\mathbf{d}\|_2 = \|\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)\|_2 \leq 2\epsilon$.
- Therefore, we may generalize the spark to allow for ϵ-proximity to the null-space.

Definition

Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, we consider all possible sub-sets of s columns, each such set forrms a sub-matrix $\mathbf{A}_s \in \mathbb{R}^{n \times s}$. We define spark$_\eta(\mathbf{A})$ as the smallest possible s (number of columns) that guarantees

$$\min_{s} \ \sigma_s(\mathbf{A}_s) \leq \eta. \tag{4}$$

In Words: This is the smallest (integer) number of columns that can be gathered from $\mathbf{A}$, such that the smallest singular-value of $\mathbf{A}_s$ is no larger than η.

Question: How this new stuff is connected with the spark definition?

Remark

Relation: For $\eta = 0$, $spark_0(\mathbf{A}) = spark(\mathbf{A})$. $spark_\eta(\mathbf{A})$ is monotone decreasing in η. We also have -

$$\forall 0 \leq \eta \leq 1, \;\; 1 \leq spark_\eta(\mathbf{A}) \leq spark(\mathbf{A}) \leq n+1. \tag{5}$$

Explain: The bounds in the above relation.

A fundamental property of the spark is that $\mathbf{Av} = \mathbf{0}$ implies $\|\mathbf{v}\|_0 \geq spark(\mathbf{A})$. Let us have a generalized version.

Lemma

If $\|\mathbf{Av}\|_2 \leq \eta$ *and* $\|\mathbf{v}\|_2 = 1$, *then* $\|\mathbf{v}\|_0 \geq spark_\eta(\mathbf{A})$.

Proof: We work out.

Now, we try to connect $spark_\eta(\mathbf{A})$ with $\mu(\mathbf{A})$.

Lemma

If $\mathbf{A}$ *has normalized columns and mutual coherence* $\mu(\mathbf{A})$, *then* $spark_\eta(\mathbf{A}) \geq \frac{1-\eta^2}{\mu(\mathbf{A})} + 1$.

Proof: We work out.

Lemma

If $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *satisfy* $\|\mathbf{b} - \mathbf{A}\mathbf{x}_i\|_2 \leq \epsilon$, $i = 1, 2$, *then* $\|\mathbf{x}_1\|_0 + \|\mathbf{x}_2\|_0 \geq spark_\eta(\mathbf{A})$, *where* $\eta = \frac{2\epsilon}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2}$.

Proof: We work out.

Remark

Comment: For the noiseless case (P_0), *we use the above uncertainty rule to derive a uniqueness result, but for the* (P_0^ϵ) *problem, we go for a form of localization in a single ball.*

Theorem

Given a distance $D \geq 0$ *and* ϵ, *set* $\eta = \frac{2\epsilon}{D}$. *Suppose there are two approximate representations* $\mathbf{x}_i$, $i = 1, 2$, *both obeying*

$$\|\mathbf{b} - \mathbf{A}\mathbf{x}_i\|_2 \leq \epsilon \quad \text{and,} \quad \|\mathbf{x}_i\|_0 \leq \frac{1}{2} spark_{\eta}(\mathbf{A}). \tag{6}$$

Then $\|\mathbf{x}_1 - \mathbf{x}_2\|_2 \leq D$.

Proof: We work out.

Theorem

Stability of (P_0^ϵ): Consider the instance of problem (P_0^ϵ) defined by the triplet $(\mathbf{A}, \mathbf{b}, \epsilon)$. Suppose that a sparse vector $\mathbf{x}_0 \in \mathbb{R}^m$ satisfies that sparsity constraint $\|\mathbf{x}_0\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(\mathbf{A})})$, and gives a representation of $\mathbf{b}$ within error tolerance ϵ (i.e., $\|\mathbf{b} - \mathbf{A}\mathbf{x}_0\|_2 \leq \epsilon$). Every solution $\mathbf{x}_0^\epsilon$ of (P_0^ϵ) must obey

$$\|\mathbf{x}_0^\epsilon - \mathbf{x}_0\|_2^2 \leq \frac{4\epsilon^2}{1 - \mu(\mathbf{A})(2\|\mathbf{x}_0\|_0 - 1)} \tag{7}$$

Proof: Home work

Remark

Note that this result parallels the uniqueness result for (P_0) problem, and indeed it reduces to it exactly for the case of $\epsilon = 0$.