PhD Course in KTH - Sparse Signal Processing
Slides 2
Discussion Topic - Uniqueness and Uncertainty

Saikat Chatterjee

Communication Theory Lab, KTH

March. 4, 2014

Questions

$$(P_0): \quad \arg\min_{\mathbf{x}\in\mathbb{R}^m} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{b} = \mathbf{A}\mathbf{x}. \tag{1}$$

1. Can uniqueness of a solution be claimed? Under what conditions?
2. If a candidate solution is available, can we say it is globally minimum?

Treating two-ortho case

1. In (P_0) problem, we seek answers for two questions.
2. Rather than answering the questions directly, we first consider special matrices $\mathbf{A}$ for which the analysis seems to be easier, and then extend our endeavor to general $\mathbf{A}$.

Special $\mathbf{A}$: Let us assume that $\mathbf{A}$ is concatenation two orthogonal bases $\mathbf{\Psi}$ and $\mathbf{\Phi}$ as $\mathbf{A} = [\mathbf{\Psi}\ \mathbf{\Phi}]$.
An example: Let $\mathbf{\Psi} = \mathbf{I}$ and $\mathbf{\Phi} = \mathbf{F}$ (Fourier matrix). That means $\mathbf{b}$ is a linear combination of few spikes and sinusoids.

Uncertainty principle

1. Classic uncertainty principle: Two conjugate variables can not be both known with arbitrary precision
2. Examples: Variables that are Fourier pair: Time and frequency, position and momentum
3. Their is a lower bound:

$$\int x^2|f(x)|^2dx \ . \ \int \omega^2|g(\omega)|^2d\omega \geq \frac{1}{2} \qquad (2)$$

Comparable claim

Remark

$\mathbf{b} \in \mathbb{R}^n$, *and* $\mathbf{\Psi}$ *and* $\mathbf{\Phi}$ *are two orthogonal complete bases. So, we can write* $\mathbf{b} = \mathbf{\Psi}\boldsymbol{\alpha}$ *and* $\mathbf{b} = \mathbf{\Phi}\boldsymbol{\beta}$*. Then, the uncertainty claim is as follows:* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *can not be both very sparse.*

To understand the above, an important point is to consider how dissimilar the two matrices are.

Definition

For an arbitrary pair of orthogonal bases $\mathbf{\Psi}$ *and* $\mathbf{\Phi}$ *that construct* $\mathbf{A} = [\mathbf{\Psi}\ \mathbf{\Phi}]$*, we define the mutual coherence* $\mu(\mathbf{A})$ *as the maximal inner product between the columns of these two bases.*

$$\mathrm{similarity}(\mathbf{\Psi}, \mathbf{\Phi}) = \mu(\mathbf{A}) = \max_{1 \le i,j \le m} |\psi_i^T \phi_j| \tag{3}$$

Remark

Bounds: $\mu(\mathbf{A})$ *satisfies* $\frac{1}{\sqrt{n}} \leq \mu(\mathbf{A}) \leq 1$

Proof: We work out

Theorem

Uncertainty principle 1: For an arbitrary pair of orthogonal bases $\mathbf{\Psi}$ *and* $\mathbf{\Phi}$ *with mutual coherence* $\mu(\mathbf{A})$, *and* $\mathbf{b} \in \mathbb{R}^n$ *and* $\mathbf{b} = \mathbf{\Psi}\boldsymbol{\alpha} = \mathbf{\Phi}\boldsymbol{\beta}$, *the following inequality holds true:*

$$\|\boldsymbol{\alpha}\|_0 + \|\boldsymbol{\beta}\|_0 \geq \frac{2}{\mu(\mathbf{A})} \tag{4}$$

Simpler proof: We work out

More involving proof: We work out. In this proof, we see an uncertainty relation with l_1 norm: $\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1 \geq \frac{2}{\sqrt{\mu(\mathbf{A})}}$

Remark

When $\mathbf{\Psi} = \mathbf{I}$ *and* $\mathbf{\Phi} = \mathbf{F}$ *(Fourier matrix: Discrete Fourier Transform), then* $\mu(\mathbf{A}) = \frac{1}{\sqrt{n}}$. *Then, in that case* $\|\boldsymbol{\alpha}\|_0 + \|\boldsymbol{\beta}\|_0 \geq 2\sqrt{n}$

Proof: Homework

Uncertainty of redundant solutions

Theorem

Uncertainty principle 2: Any two distinct solutions $\mathbf{x}_1$, $\mathbf{x}_2$ *of* $\mathbf{A}\mathbf{x} = [\boldsymbol{\Psi}\ \boldsymbol{\Phi}]\,\mathbf{x} = \mathbf{b}$ *can not be both very sparse, governed by the following uncertainty principle:*

$$\|\mathbf{x}_1\|_0 + \|\mathbf{x}_2\|_0 \geq \frac{2}{\mu\,(\mathbf{A})}. \tag{5}$$

We refer to this result as "Uncertainty of redundant solutions".

Proof: We work out

From Uncertainty to Uniqueness

A direct consequence of the previous theorem in a uniqueness result.

Theorem

Uniqueness: For $\mathbf{Ax} = [\mathbf{\Psi}\ \mathbf{\Phi}]\,\mathbf{x} = \mathbf{b}$*, if a candidate solution* $\mathbf{x}$ *satisfies* $\|\mathbf{x}\|_0 < \frac{1}{\mu(\mathbf{A})}$*, then it is necessarily the sparsest and any other solution must be denser.*

Proof: By simple arguments

Remark

- *This simple claim is wonderful and powerful. At least for special* $\mathbf{A} = [\boldsymbol{\Psi}\ \boldsymbol{\Phi}]$, *we have a complete answer for the two questions. We can claim uniqueness and global optimality.*
- *Notice that, in general for a non-convex problem a given solution can at best be verified as being locally optimal, and here we are able to verify global optimality.*

Treating general case: Uniqueness via Spark

♠ To characterize the null space of a general **A** using l_0 norm

♠ Donoho and Elad coined and defined spark in 2003

Definition

Spark (Rank): The spark (rank) of a matrix **A** *is the smallest (largest) number of linearly dependent (independent) columns*

$$spark(\mathbf{A}) = \min_{\mathbf{x}\in\mathbb{R}^m} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{0},\ \mathbf{x} \neq \mathbf{0}. \tag{6}$$

Note: By definition, the non-zero vectors in the null space of **A**, i.e., $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0},\ \mathbf{x} \neq \mathbf{0}\}$, must satisfy $\|\mathbf{x}\|_0 \geq spark(\mathbf{A})$

Uniqueness via Spark

Theorem

Uniqueness via Spark: For $\mathbf{A}\mathbf{x} = \mathbf{b}$, *if a candidate solution* $\mathbf{x}$ *satisfies* $\|\mathbf{x}\|_0 < \frac{1}{2} spark(\mathbf{A})$, *then it is necessarily the sparsest.*

Proof: We work out

Remark

Range of Spark: $2 \leq spark(\mathbf{A}) \leq n + 1$

Proof: We discuss

Home work: For two-ortho identity-Fourier pair, prove that $spark(\mathbf{A}) = 2\sqrt{n}$

Uniqueness via Mutual Coherence

1. Finding 'spark' is again NP hard. So comes mutual coherence
2. In the two-ortho case, the gram matrix is

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{\Psi}^T\mathbf{\Phi} \\ \mathbf{\Phi}^T\mathbf{\Psi} & \mathbf{I} \end{bmatrix} \quad (7)$$

   and the mutual coherence $\mu(\mathbf{A})$ is obtained as the maximum off-diagonal entry (in absolute value) in this Gram matrix.
3. Now we go for a generalization

Definition

Mutual coherence: Denoting the i-th column by $\mathbf{a}_i$

$$\mu(\mathbf{A}) = \max_{1 \le i,j \le m, i \ne j} \frac{|\mathbf{a}_i^T \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \, \|\mathbf{a}_j\|_2} \tag{8}$$

1. For an orthogonal matrix, $\mu(\mathbf{A}) = 0$
2. For two-ortho case: $\frac{1}{\sqrt{n}} \le \mu(\mathbf{A}) \le 1$
3. For a general $\mathbf{A} \in \mathbb{R}^{n \times m}$, we desire for a low $\mu(\mathbf{A})$ such that it exhibits a close behavior of an orthogonal matrix
4. Donoho and Huo's Work: For randomly constructed $\mathbf{A} \in \mathbb{R}^{n \times m}$ with full rank n, we have $\mu(\mathbf{A}) \ge \sqrt{\frac{m-n}{n(m-1)}}$.
   - When $m = n$, $\mu(\mathbf{A}) = 0$
   - As n decreases, $\mu(\mathbf{A})$ increases

Lemma

For a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, *spark*$(\mathbf{A}) \geq 1 + \frac{1}{\mu(\mathbf{A})}$

Proof: We work out (by using Gersh-gorin disk theorem). We learn the Gersh-gorin disk theorem and its relation with positive-definite property. Then we proceed with formal proof.
Challenge: Can students perform an alternate proof by using the facts that the resulting Gram matrix is symmetric, off-diagonal elements are upper bounded by mutual-coherence and the Gram matrix has to be full rank.

Theorem

Uniqueness via Mutual Coherence: For $\mathbf{Ax} = \mathbf{b}$, *if a candidate solution* $\mathbf{x}$ *satisfies* $\|\mathbf{x}\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(\mathbf{A})})$, *then it is necessarily the sparsest.*

Proof: By simple arguments

Uniqueness via Babel function

1. An intelligent approach by Joel Tropp
2. Home Work: Understand the approach

Upper bounding the Spark

For sparsest uniqueness verification

1. We can evaluate $\mu(\mathbf{A})$ and so the upper bound $\frac{1}{2}(1+\frac{1}{\mu(\mathbf{A})})$.
2. We may find a solution $\mathbf{x}$ such that $\|\mathbf{x}\|_0 < \frac{1}{2}(1+\frac{1}{\mu(\mathbf{A})})$ and conclude uniqueness. However, such a solution may not be guaranteed.
3. So comes the use of spark. We may find a solution $\mathbf{x}$ such that $\|\mathbf{x}\|_0 < \frac{1}{2}spark(\mathbf{A})$ and conclude uniqueness.
4. However, computation of spark is NP hard.
5. So, in practice, we can upper bound the spark by a computable quantity and compare.

♠ Let us evaluate such an upper bound

Constructing Grassmannian matrices

Definition

A Grassmannian (real) matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ *with* $m \geq n$ *and normalized columns satisfies that its Gram matrix* $\mathbf{G} = \mathbf{A}^T \mathbf{A}$ *has the following property*

$$\forall i \neq j, \;\; |G_{i,j}| = \sqrt{\frac{m-n}{n(m-1)}}. \tag{9}$$

For a Grassmannian matrix, $\mu(\mathbf{A}) = \sqrt{\frac{m-n}{n(m-1)}}$.

1. Grassmannian is special as the angle between each and every pairs of columns is same and smallest possible.
2. It has strong connection with packing vectors/subspaces in $\mathbb{R}^n$.
3. Important in channel coding and wireless communication.
4. Hard to construct such matrix. A numerical method was proposed by Joel Tropp.

Home Work: Read the paper and simulate the algorithm. Reproduce the experimental and/or numerical results in the paper.

Home Work Problems

1. Understand the proof for the claim related to (P_1) problem: There exist a sparse solution that has at-most n non-zeros (i.e., same as the number of constraints). According to Saikat, there is a problem in the proof. So, either identify the problem or get convinced that the proof is correct.
2. For identity-Fourier pair (two-otho case), prove that $\mu(\mathbf{A}) = \frac{1}{\sqrt{n}}$ and $\|\boldsymbol{\alpha}\|_0 + \|\boldsymbol{\beta}\|_0 \geq 2\sqrt{n}$.
3. For identity-Fourier pair (two-otho case), prove that $spark(\mathbf{A}) = 2\sqrt{n}$.
4. Understand 'Uniqueness via Babel Function' approach by Joel Tropp (From text book).
5. Construction of Grassmannian: Read the paper by Joel Tropp and simulate the algorithm. Reproduce the experimental and/or numerical results in the paper.