Sparse Optimization
Lecture: Dual Certificate in ℓ_1 Minimization

Instructor: Wotao Yin
Department of Mathematics, UCLA

July 2013

Note scriber: Zheng Sun

Those who complete this lecture will know

- what is a dual certificate for ℓ_1 minimization
- a strictly complementary dual certificate guarantees exact recovery
- it also guarantees stable recovery

Sparse Optimization
Lecture: Dual Certificate in ℓ_1 Minimization
Instructor: Wotao Yin
Department of Mathematics, UCLA
July 2013
Note scriber: Zheng Sun
Those who complete this lecture will know
- what is a dual certificate for ℓ_1 minimization
- a strictly complementary dual certificate guarantees exact recovery
- it also guarantees stable recovery

What is covered

- A review of dual certificate in the context of conic programming
- A condition that guarantees recovering a set of sparse vectors (whose entries have the same signs), *not* for all k-sparse vectors ☹
- The condition depends on $\mathrm{sign}(\mathbf{x}^o)$, but not $\mathbf{x}^o$ itself or $\mathbf{b}$
- The condition is sufficient and necessary ☺
- It also guarantees robust recovery against measurement errors ☺
- The condition can be numerically verified (in polynomial time) ☺

The underlying techniques are Lagrange duality, strict complementarity, and LP strong duality.

Results in this lecture are drawn from various papers. For references, see:

H. Zhang, M. Yan, and W. Yin, One condition for all: solution uniqueness and robustness of ℓ_1-synthesis and ℓ_1-analysis minimizations

What is covered

- A review of dual certificate in the context of conic programming
- A condition that guarantees recovering a set of sparse vectors (whose entries have the same signs), *not* for all k-sparse vectors ☹
- The condition depends on $\mathrm{sign}(\mathbf{x}^o)$, but not $\mathbf{x}^o$ itself or $\mathbf{b}$
- The condition is sufficient and necessary ☺
- It also guarantees robust recovery against measurement errors ☺
- The condition can be numerically verified (in polynomial time) ☺

The underlying techniques are Lagrange duality, strict complementarity, and LP strong duality.

Results in this lecture are drawn from various papers. For references, see:
H. Zhang, M. Yan, and W. Yin, One condition for all: solution uniqueness and robustness of ℓ_1-synthesis and ℓ_1-analysis minimizations

Lagrange dual for conic programs

Let $\mathcal{K}_i$ be a first-orthant, second-order, or semi-definite cone. It is self-dual.

(Suppose $\mathbf{a}, \mathbf{b} \in \mathcal{K}_i$. Then, $\mathbf{a}^T\mathbf{b} \geq 0$. If $\mathbf{a}^T\mathbf{b} = 0$, either $\mathbf{a} = 0$ or $\mathbf{b} = 0$.)

- Primal:

$$\min \mathbf{c}^T\mathbf{x} \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x}_i \in \mathcal{K}_i\ \forall i.$$

- Lagrangian relaxation:

$$\mathcal{L}(\mathbf{x}; \mathbf{s}) = \mathbf{c}^T\mathbf{x} + \mathbf{s}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Dual function:

$$d(\mathbf{s}) = \min_{\mathbf{x}}\{\mathcal{L}(\mathbf{x}; \mathbf{s}) : \mathbf{x}_i \in \mathcal{K}_i\ \forall i\} = -\mathbf{b}^T\mathbf{s} - \iota_{\{(\mathbf{A}^T\mathbf{s}+\mathbf{c})_i \in \mathcal{K}_i\ \forall i\}}$$

- Dual problem:

$$\min_{\mathbf{s}} -d(\mathbf{s}) \iff \min_{\mathbf{s}} \mathbf{b}^T\mathbf{s} \quad \text{s.t. } (\mathbf{A}^T\mathbf{s} + \mathbf{c})_i \in \mathcal{K}_i\ \forall i$$

One problem might be simpler to solve than the other; solving one might help solve the other.

Lagrange dual for conic programs

Let $\mathcal{K}_i$ be a first-orthant, second-order, or semi-definite cone. It is self-dual.
(Suppose $\mathbf{a}, \mathbf{b} \in \mathcal{K}_i$. Then, $\mathbf{a}^T\mathbf{b} \geq 0$. If $\mathbf{a}^T\mathbf{b} = 0$, either $\mathbf{a} = 0$ or $\mathbf{b} = 0$.)

- Primal: $\min \mathbf{c}^T\mathbf{x}$ s.t. $\mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x}_i \in \mathcal{K}_i\ \forall i.$
- Lagrangian relaxation: $\mathcal{L}(\mathbf{x}; \mathbf{s}) = \mathbf{c}^T\mathbf{x} + \mathbf{s}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$
- Dual function: $d(\mathbf{s}) = \min_{\mathbf{x}}\{\mathcal{L}(\mathbf{x}; \mathbf{s}) : \mathbf{x}_i \in \mathcal{K}_i\ \forall i\} = -\mathbf{b}^T\mathbf{s} - \iota_{\{(\mathbf{A}^T\mathbf{s}+\mathbf{c})_i \in \mathcal{K}_i\ \forall i\}}$
- Dual problem: $\min_{\mathbf{s}} -d(\mathbf{s}) \iff \min_{\mathbf{s}} \mathbf{b}^T\mathbf{s}$ s.t. $(\mathbf{A}^T\mathbf{s} + \mathbf{c})_i \in \mathcal{K}_i\ \forall i$

One problem might be simpler to solve than the other; solving one might help solve the other.

- $d(s) = \inf_{\mathbf{x}}\{(\mathbf{c} + \mathbf{A}^T\mathbf{s})^T\mathbf{x} - \mathbf{s}^T\mathbf{b} : \mathbf{x}_i \in \mathcal{K}_i\}$
  If $(\mathbf{c}^T + \mathbf{A}^T\mathbf{s})_i \in \mathcal{K}_i$ the inner product is no less than zero; otherwise, it will crash down to $-\infty$. Thus $d(\mathbf{s}) = -\mathbf{b}^T\mathbf{y} - \iota_{\{(\mathbf{A}^T\mathbf{s}+\mathbf{c})_i \in \mathcal{K}_i\ \forall i\}}$ Where

$$\iota_{\{(\mathbf{A}^T\mathbf{s}+\mathbf{c})_i \in \mathcal{K}_i\ \forall i\}} = \begin{cases} 0, & \{(\mathbf{A}^T\mathbf{s} + \mathbf{c})_i \in \mathcal{K}_i\ \forall i\}. \\ \infty, & \text{otherwise} \end{cases}$$

- $\min -d(\mathbf{s}) \Leftrightarrow \max d(\mathbf{s})$

Dual certificate

Given that $\mathbf{x}^*$ is *primal feasible*, i.e., obeying $\mathbf{A}\mathbf{x}^* = \mathbf{b},\ \mathbf{x}_i^* \in \mathcal{K}_i\ \forall i$.

Question: is $\mathbf{x}^*$ optimal?

Answer: One does *not* need to compare $\mathbf{x}^*$ to all other feasible $\mathbf{x}$.

A dual vector $\mathbf{y}^*$ will certify the optimality of $\mathbf{x}^*$.

Dual certificate

Given that $\mathbf{x}^*$ is *primal feasible*, i.e., obeying $\mathbf{A}\mathbf{x}^* = \mathbf{b},\ \mathbf{x}_i^* \in \mathcal{K}_i\ \forall i$.

Question: is $\mathbf{x}^*$ optimal?

Answer: One does *not* need to compare $\mathbf{x}^*$ to all other feasible $\mathbf{x}$.

A dual vector $\mathbf{y}^*$ will certify the optimality of $\mathbf{x}^*$.

Dual certificate

Theorem

Suppose $\mathbf{x}^*$ is feasible (i.e., $\mathbf{A}\mathbf{x}^* = \mathbf{b},\ \mathbf{x}_i^* \in \mathcal{K}_i\ \forall i$). If $\mathbf{s}^*$ obeys

1. *vanished duality gap: $-\mathbf{b}^T\mathbf{s}^* = \mathbf{c}^T\mathbf{x}^*$, and*
2. *dual feasibility: $(\mathbf{A}^T\mathbf{s}^* + \mathbf{c})_i \in \mathcal{K}_i$,*

then $\mathbf{x}^*$ is primal optimal.

Pick any *primal feasible* $\mathbf{x}$ (i.e., $\mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x}_i \in \mathcal{K}_i\ \forall i$), we have

$$(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)^T\mathbf{x} = \sum_i \underbrace{(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T}_{\in\mathcal{K}_i} \underbrace{\mathbf{x}_i}_{\in\mathcal{K}_i} \geq 0$$

and thus due to $\mathbf{A}\mathbf{x} = \mathbf{b}$,

$$\mathbf{c}^T\mathbf{x} = (\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)^T\mathbf{x} - (\mathbf{A}^T\mathbf{s}^*)^T\mathbf{x} \geq -(\mathbf{A}^T\mathbf{s}^*)^T\mathbf{x} = -\mathbf{b}^T\mathbf{s}^* = \mathbf{c}^T\mathbf{x}^*.$$

Therefore, $\mathbf{x}^*$ is optimal.

Corollary: $(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)^T\mathbf{x}^* = 0$ and $(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T\mathbf{x}_i^* = 0,\ \forall i$.

Bottom line: dual vector $\mathbf{y}^* = \mathbf{A}^T\mathbf{s}^*$ certifies the optimality of $\mathbf{x}^*$.

- The illustration of "vanished gap"

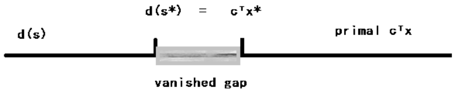


Figure: vanished gap

- To verify $\sum_i (\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T\mathbf{x}_i \geq 0$, we use the property of $\mathcal{K}_i$: Suppose $\mathbf{a}, \mathbf{b} \in \mathcal{K}_i$. Then, $\mathbf{a}^T\mathbf{b} \geq 0$. If $\mathbf{a}^T\mathbf{b} = 0$, either $\mathbf{a} = 0$ or $\mathbf{b} = 0$
- We could prove the corollary by substituting $\mathbf{x} = \mathbf{x}^*$ on the left part of the former inequality, forcing $-(\mathbf{A}^T\mathbf{s}^*)^T\mathbf{x} = \mathbf{c}^T\mathbf{x}^*$

Dual certificate

A related claim:

Theorem

If any primal feasible $\mathbf{x}^*$ and dual feasible $\mathbf{s}^*$ have no duality gap, then $\mathbf{x}$ is primal optimal and $\mathbf{s}$ is dual optimal.

Reason: the primal objective value of any primal feasible $\mathbf{x} \geq$ the dual objective value of any dual feasible $\mathbf{s}$. Therefore, assuming both primal and dual feasibilities, a pair of primal/dual objectives must be optimal.

Dual certificate

A related claim:

Theorem

If any primal feasible $\mathbf{x}^*$ and dual feasible $\mathbf{s}^*$ have no duality gap, then $\mathbf{x}$ is primal optimal and $\mathbf{s}$ is dual optimal.

Reason: the primal objective value of any primal feasible $\mathbf{x} \geq$ the dual objective value of any dual feasible $\mathbf{s}$. Therefore, assuming both primal and dual feasibilities, a pair of primal/dual objectives must be optimal.

Complementarity and strict complementarity

From

$$\sum_i (\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T \mathbf{x}_i^* = (\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)^T\mathbf{x}^* = \mathbf{c}^T\mathbf{x}^* + \mathbf{b}^T\mathbf{s}^* = 0$$

and

$$\underbrace{(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T}_{\in\mathcal{K}_i} \underbrace{\mathbf{x}_i^*}_{\in\mathcal{K}_i} \geq 0,\ \forall i.$$

we get

$$(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T \mathbf{x}_i^* = 0,\ \forall i.$$

Hence, at least one of $(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T$ and $\mathbf{x}_i^*$ is 0 (but they can be both zero.)

- If *exactly* one of $(\mathbf{c} + \mathbf{A}^T\mathbf{s}^*)_i^T$ and $\mathbf{x}_i^*$ is zero (the other is nonzero), then they are *strictly complementary.*

Certifying the uniqueness of $\mathbf{x}^*$ requires a strictly complementary $\mathbf{s}^*$.

ℓ_1 duality and dual certificate

Primal:

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \tag{1}$$

Dual:

$$\max \mathbf{b}^T\mathbf{s} \quad \text{s.t. } \|\mathbf{A}^T\mathbf{s}\|_\infty \le 1$$

- Given a feasible $\mathbf{x}^*$, if $\mathbf{s}^*$ obeys

1. $\|\mathbf{A}^T\mathbf{s}^*\|_\infty \le 1$, and
2. $\|\mathbf{x}^*\|_1 - \mathbf{b}^T\mathbf{s}^* = 0$,

then $\mathbf{y}^* = \mathbf{A}^T\mathbf{s}^*$ certifies the optimality of $\mathbf{x}^*$.

- Any primal optimal $\mathbf{x}^*$ must satisfy $\|\mathbf{x}^*\|_1 - \mathbf{b}^T\mathbf{s}^* = 0$.

- The dual problem for basis pursuit. (a special case in the 3rd slide)
  The LP formulation of $\min_{\mathbf{x}}\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ is:

$$\min_{\mathbf{x}}\{\mathbf{1}^T\mathbf{x}^1 + \mathbf{1}^T\mathbf{x}^2\ \mathbf{A}\mathbf{x}^1 - \mathbf{A}\mathbf{x}^2 = \mathbf{b}, \mathbf{x}^1, \mathbf{x}^2 \succeq \mathbf{0}\}$$

Since $L(\mathbf{x},\mathbf{s}) = \|\mathbf{x}\|_1 + \mathbf{s}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$,

$$\min_{\mathbf{x}} L(\mathbf{x},\mathbf{s}) \Leftrightarrow \min_{\mathbf{x}}\{\|\mathbf{x}\|_1 + \mathbf{s}^T\mathbf{A}\mathbf{x}\} - \mathbf{s}^T\mathbf{b}$$

$\|\mathbf{x}\|_1 + \mathbf{s}^T\mathbf{A}\mathbf{x} = \sum_i |x_i| + (\mathbf{A}^T\mathbf{s})_i x_i$. It has a lower bound if and only if $(\mathbf{A}^T\mathbf{s})_i$ is no more than 1 for each i.
Thus the dual function is :

$$d(\mathbf{s}) = -\mathbf{s}^T\mathbf{b} + \iota_{\{\|\mathbf{A}^T\mathbf{s}\|_\infty \le 1\}}$$

And the dual problem is:

$$\max_{\mathbf{s}} \mathbf{b}^T\mathbf{s} \quad \text{s.t. } \|\mathbf{A}^T\mathbf{s}\|_\infty \le 1$$

ℓ_1 duality and complementarity

- $|a| \le 1 \implies ab \le |b|$. ▶ If $ab = |b|$, then
  1. $|a| < 1 \;\Rightarrow\; b = 0$
  2. $a = 1 \;\Rightarrow\; b \ge 0$
  3. $a = -1 \;\Rightarrow\; b \le 0$
- From $\|\mathbf{A}^T\mathbf{s}^*\|_\infty \le 1$, we get $\|\mathbf{x}^*\|_1 = \mathbf{b}^T\mathbf{s}^* = (\mathbf{A}^T\mathbf{s}^*)^T\mathbf{x}^* \le \|\mathbf{x}^*\|_1$ and

$$(\mathbf{A}^T\mathbf{s}^*)_i \cdot x_i = |x_i|, \quad \forall i.$$

Therefore,

1. if $|(\mathbf{A}^T\mathbf{s}^*)_i| < 1$, then $\mathbf{x}_i^* = 0$
2. if $(\mathbf{A}^T\mathbf{s}^*)_i = 1$, then $\mathbf{x}_i^* \ge 0$
3. if $(\mathbf{A}^T\mathbf{s}^*)_i = -1$, then $\mathbf{x}_i^* \le 0$

Strict complementarity holds if for each i, $1 - |(\mathbf{A}^T\mathbf{s}^*)_i|$ or $\mathbf{x}_i$ is zero but not both.

ℓ_1 duality and complementarity

▶ $|a| \le 1 \implies ab \le |b|$. ▶ If $ab = |b|$, then
1. $|a| < 1 \Rightarrow b = 0$
2. $a = 1 \Rightarrow b \ge 0$
3. $a = -1 \Rightarrow b \le 0$
▶ From $\|\mathbf{A}^T\mathbf{s}^*\|_\infty \le 1$, we get $\|\mathbf{x}^*\|_1 = \mathbf{b}^T\mathbf{s}^* = (\mathbf{A}^T\mathbf{s}^*)^T\mathbf{x}^* \le \|\mathbf{x}^*\|_1$ and
$(\mathbf{A}^T\mathbf{s}^*)_i \cdot x_i = |x_i|, \quad \forall i.$
Therefore,
1. if $|(\mathbf{A}^T\mathbf{s}^*)_i| < 1$, then $\mathbf{x}_i^* = 0$
2. if $(\mathbf{A}^T\mathbf{s}^*)_i = 1$, then $\mathbf{x}_i^* \ge 0$
3. if $(\mathbf{A}^T\mathbf{s}^*)_i = -1$, then $\mathbf{x}_i^* \le 0$
Strict complementarity holds if for each i, $1 - |(\mathbf{A}^T\mathbf{s}^*)_i|$ or $\mathbf{x}_i$ is zero but not both.

- $(\mathbf{A}^T\mathbf{s}^*)_i \cdot x_i = |x_i|$ implies $(\mathbf{A}^T\mathbf{s}^*)_i$ and x_i has the same sign. Thus $|x_i| + (\mathbf{A}^T\mathbf{s}^*)_i \cdot x_i = (1 - |(\mathbf{A}^T\mathbf{s}^*)_i|)|x_i|$. The definition of strictly complementarity requires $1 - |(\mathbf{A}^T\mathbf{s}^*)_i|$ or x_i is zero but not both.

Uniqueness of $\mathbf{x}^*$

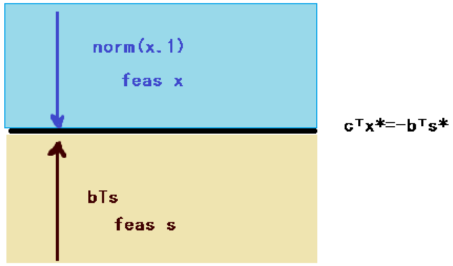
Suppose $\mathbf{x}^*$ is a solution to the basis pursuit model.

Question: Is it the unique solution?

Define $I := \mathrm{supp}(\mathbf{x}^*) = \{i : \mathbf{x}_i^* \neq 0\}$ and $J = I^c$.

- If $\mathbf{s}^*$ is a dual certificate and $\|(\mathbf{A}^T\mathbf{s}^*)_J\|_\infty < 1$, $\mathbf{x}_J = 0$ for all optimal $\mathbf{x}$.
- For $i \in I$, $(\mathbf{A}^T\mathbf{s}^*)_i = \pm 1$ cannot determine $x_i \stackrel{?}{=} 0$ for optimal $\mathbf{x}$. It is possible that $(\mathbf{A}^T\mathbf{s}^*)_i = \pm 1$ yet $x_i = 0$ (this is called *degenerate*.)
- On the other hand, if $\mathbf{A}_I\mathbf{x}_I = \mathbf{b}$ has a *unique* solution, denoted by $\mathbf{x}_I^*$, then since $\mathbf{x}_J^* = 0$ is unique, $\mathbf{x}^* = [\mathbf{x}_I^*; \mathbf{x}_J^*] = [\mathbf{x}_I^*; \mathbf{0}]$ is the unique solution to the basis pursuit model.
- $\mathbf{A}_I\mathbf{x}_I = \mathbf{b}$ has a *unique* solution provided that $\mathbf{A}_I$ has independent columns, or equivalently, $\ker(\mathbf{A}_I) = \{0\}$.

- Another illustration of the problem



Optimality and uniqueness

Condition

For a given $\bar{\mathbf{x}}$, *the index sets* $I = \mathrm{supp}(\bar{\mathbf{x}})$ *and* $J = I^c$ *satisfy*

1. $\ker(\mathbf{A}_I) = \{0\}$
2. *there exists* $\mathbf{y}$ *such that* $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T)$, $\mathbf{y}_I = \mathrm{sign}(\bar{\mathbf{x}}_I)$, *and* $\|\mathbf{y}_J\|_\infty < 1$.

Comments:

- part 1 guarantees unique $\mathbf{x}_I^*$ as the solution to $\mathbf{A}_I\mathbf{x}_I = \mathbf{b}$
- part 2 guarantees $\mathbf{x}_J^* = 0$
- $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T)$ means $\mathbf{y} = \mathbf{A}^T\mathbf{s}$ for some $\mathbf{s}$
- the condition involves I and $\mathrm{sign}(\bar{\mathbf{x}}_I)$, *not* the values of $\bar{\mathbf{x}}_I$ or $\mathbf{b}$; but different I and $\mathrm{sign}(\bar{\mathbf{x}}_I)$ require a different condition
- RIP guarantees the condition hold for all small I and arbitrary signs
- the condition is easy to verify

Optimality and uniqueness

Condition

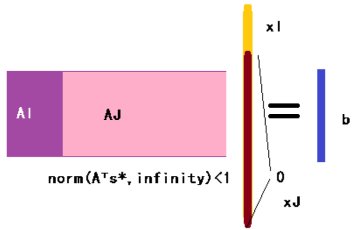
For a given $\bar{\mathbf{x}}$, the index sets $I = \mathrm{supp}(\bar{\mathbf{x}})$ and $J = I^c$ satisfy

1. $\ker(\mathbf{A}_I) = \{0\}$
2. there exists $\mathbf{y}$ such that $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T)$, $\mathbf{y}_I = \mathrm{sign}(\bar{\mathbf{x}}_I)$, and $\|\mathbf{y}_J\|_\infty < 1$.

Comments:

- part 1 guarantees unique $\mathbf{x}_I^*$ as the solution to $\mathbf{A}_I\mathbf{x}_I = \mathbf{b}$
- part 2 guarantees $\mathbf{x}_J^* = 0$
- $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T)$ means $\mathbf{y} = \mathbf{A}^T\mathbf{s}$ for some $\mathbf{s}$
- the condition involves I and $\mathrm{sign}(\bar{\mathbf{x}}_I)$, not the values of $\bar{\mathbf{x}}_I$ or $\mathbf{b}$; but different I and $\mathrm{sign}(\bar{\mathbf{x}}_I)$ require a different condition
- RIP guarantees the condition hold for all small I and arbitrary signs
- the condition is easy to verify

- Now we care about whether the optimal solution $\mathbf{x}^*$ is unique. Suppose $\|(\mathbf{A}^T\mathbf{s}^*)_J\|_\infty < 1$, all the elements on J are forced to be zero. Thus, the uniqueness is determined by the property of $\mathbf{A}_I$. If $\ker(\mathbf{A}_I) = \{0\}$, the nonhomogeneous equation $\mathbf{A}_I\mathbf{x}_I = \mathbf{b}_I$ has a unique solution. Thus, all the elements on I could be uniquely determined.
- An illustration:



Optimality and uniqueness

Theorem

Suppose $\bar{\mathbf{x}}$ *obeys* $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$ *and the above Condition, then* $\bar{\mathbf{x}}$ *is the unique solution to* $\min\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$.

In fact, the converse is also true, namely, the Condition is also necessary.

Optimality and uniqueness
Theorem
Suppose $\bar{\mathbf{x}}$ obeys $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$ and the above Condition, then $\bar{\mathbf{x}}$ is the unique solution to $\min\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$.
In fact, the converse is also true, namely, the Condition is also necessary.

- To ensure the uniqueness, $\mathbf{A}_I$ should be thin enough and at most a square. In fact, for a good recovery, we usually require the columns of A_I (corresponding to the sparsity of the signal) to be fewer than $(m+1)/2$which could be interpreted from the concept of "spark".

Uniqueness of $\mathbf{x}^*$

Part 1 $\ker(\mathbf{A}_I) = \{0\}$ is necessary.

Lemma

If $0 \neq \mathbf{h} \in \ker(\mathbf{A}_I)$, *then all* $\mathbf{x}_\alpha = \mathbf{x}^* + \alpha[\mathbf{h}; \mathbf{0}]$ *for small* α *is optimal.*

Proof.

- $\mathbf{x}_\alpha$ is feasible since $\mathbf{A}\mathbf{x}_\alpha = \mathbf{A}\mathbf{x}^* = \mathbf{b}$.
- We know $\|\mathbf{x}_\alpha\|_1 \geq \|\mathbf{x}^*\|_1$, but for small α around 0, we also have $\|\mathbf{x}_\alpha\|_1 = \|\mathbf{x}_I^* + \alpha\mathbf{h}\|_1 = (\mathbf{A}^T\mathbf{s}^*)_I^T(\mathbf{x}_I^* + \alpha\mathbf{h}) = \|\mathbf{x}^*\|_1 + \alpha(\mathbf{A}^T\mathbf{s}^*)_I^T\mathbf{h}$.
- Hence, $(\mathbf{A}^T\mathbf{s}^*)_I^T\mathbf{h} = 0$ and thus $\|\mathbf{x}_\alpha\|_1 = \|\mathbf{x}^*\|_1$. So, $\mathbf{x}_\alpha$ is also optimal.

□

Uniqueness of $\mathbf{x}^*$

Necessity

- **Is part 2 necessary?**

Introduce

$$\min_{\mathbf{y}} \|\mathbf{y}_J\|_\infty \quad \text{s.t.} \quad \mathbf{y} \in \mathcal{R}(\mathbf{A}^T), \; \mathbf{y}_I = \operatorname{sign}(\bar{\mathbf{x}}_I). \tag{2}$$

If the optimal objective value < 1, then there exists $\mathbf{y}$ obeying part 2, so part 2 is also necessary.

We shall translate (2) and rewrite $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T)$.

- $\|\mathbf{y}_J\|_\infty$ determines the stability of the signal recovery. As shown below, the $\|\mathbf{y}_J\|_\infty$ is related with the angle θ between the hyperplane and the diamond. The smaller $\|\mathbf{y}_J\|_\infty$ is , the larger angle θ is, and a more stable recovery we may have.

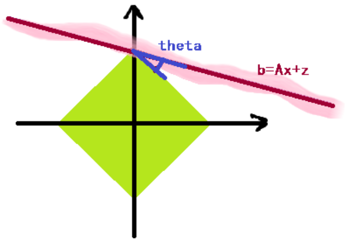


Figure: The intersection of the hyperplane and the diamond

Necessity

Define $\mathbf{a} = [\mathrm{sign}(\bar{\mathbf{x}}_I); \mathbf{0}]$ and basis $\mathbf{Q}$ of $\mathrm{Null}(\mathbf{A})$.

- If $\mathbf{a} \in \mathcal{R}(\mathbf{A}^T)$, set $\mathbf{y} = \mathbf{a}$. done.
- Otherwise, let $\mathbf{y} = \mathbf{a} + \mathbf{z}$. Then
  - $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T) \;\Leftrightarrow\; \mathbf{Q}^T\mathbf{y} = 0 \;\Leftrightarrow\; \mathbf{Q}^T\mathbf{z} = -\mathbf{Q}^T\mathbf{a}$
  - $\mathbf{y}_I = \mathrm{sign}(\bar{\mathbf{x}}_I) = \mathbf{a}_I \;\Leftrightarrow\; \mathbf{z}_I = 0$
  - $\mathbf{a}_J = 0 \;\Rightarrow\; \|\mathbf{y}_J\|_\infty = \|\mathbf{z}_J\|_\infty$

Equivalent problem:

$$\min_{\mathbf{z}} \|\mathbf{z}_J\|_\infty \quad \text{s.t. } \mathbf{Q}^T\mathbf{z} = -\mathbf{Q}^T\mathbf{a},\ \mathbf{z}_I = 0. \tag{3}$$

If the optimal objective value < 1, then part 2 is necessary.

Necessity

Define $\mathbf{a} = [\mathrm{sign}(\bar{\mathbf{x}}_I); \mathbf{0}]$ and basis $\mathbf{Q}$ of $\mathrm{Null}(\mathbf{A})$.

- If $\mathbf{a} \in \mathcal{R}(\mathbf{A}^T)$, set $\mathbf{y} = \mathbf{a}$. done.
- Otherwise, let $\mathbf{y} = \mathbf{a} + \mathbf{z}$. Then
  - $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T) \;\Leftrightarrow\; \mathbf{Q}^T\mathbf{y} = 0 \;\Leftrightarrow\; \mathbf{Q}^T\mathbf{z} = -\mathbf{Q}^T\mathbf{a}$
  - $\mathbf{y}_I = \mathrm{sign}(\bar{\mathbf{x}}_I) = \mathbf{a}_I \;\Leftrightarrow\; \mathbf{z}_I = 0$
  - $\mathbf{a}_J = 0 \;\Rightarrow\; \|\mathbf{y}_J\|_\infty = \|\mathbf{z}_J\|_\infty$

Equivalent problem:

$$\min_{\mathbf{z}} \|\mathbf{z}_J\|_\infty \quad \text{s.t. } \mathbf{Q}^T\mathbf{z} = -\mathbf{Q}^T\mathbf{a},\ \mathbf{z}_I = 0. \tag{3}$$

If the optimal objective value < 1, then part 2 is necessary.

Necessity

Theorem (LP strong duality)

If a linear program has a finite solution, its Lagrange dual has a finite solution. The two solutions achieve the same primal and dual optimal objective.

Problem (3) is feasible and has a finite objective value. The dual of (3) is

$$\max_{\mathbf{p}} (\mathbf{Q}^T \mathbf{a})^T \mathbf{p} \quad \text{s.t. } \|(\mathbf{Q}\mathbf{p})_J\|_1 \leq 1.$$

If its optimal objective value < 1, then part 2 is necessary.

2013-07-12

└─Necessity

Necessity

Theorem (LP strong duality)

If a linear program has a finite solution, its Lagrange dual has a finite solution. The two solutions achieve the same primal and dual optimal objective.

Problem (3) is feasible and has a finite objective value. The dual of (3) is

$$\max_{\mathbf{p}} (\mathbf{Q}^T \mathbf{a})^T \mathbf{p} \quad \text{s.t. } \|(\mathbf{Q}\mathbf{p})_J\|_1 \leq 1.$$

If its optimal objective value < 1, then part 2 is necessary.

Necessity

Lemma

If $\mathbf{x}^*$ is unique, then the optimal objective of the following primal-dual problems is strictly less than 1.

$$\min_{\mathbf{z}} \|\mathbf{z}_J\|_\infty \quad \text{s.t. } \mathbf{Q}^T\mathbf{z} = -\mathbf{Q}^T\mathbf{a},\ \mathbf{z}_I = 0.$$

$$\max_{\mathbf{p}} (\mathbf{Q}^T\mathbf{a})^T\mathbf{p} \quad \text{s.t. } \|(\mathbf{Q}\mathbf{p})_J\|_1 \le 1.$$

Proof.

Uniqueness of $\mathbf{x}^* \implies \forall\, \mathbf{h} \in \ker(\mathbf{A}) \setminus \{0\},\ \|\mathbf{x}^*\|_1 < \|\mathbf{x}^* + \mathbf{h}\|_1$
$\implies \mathbf{a}_I^T\mathbf{h}_I < \|\mathbf{h}_J\|_1$

Therefore,

- if $\mathbf{p}^* = 0$, then $\|\mathbf{z}_J^*\|_\infty = (\mathbf{Q}^T\mathbf{a})^T\mathbf{p}^* = 0$.
- if $\mathbf{p}^* \neq 0$, then $\mathbf{h} := \mathbf{Q}\mathbf{p}^* \in \ker(\mathbf{A}) \setminus \{0\}$ obeys $\|\mathbf{z}_J^*\|_\infty = (\mathbf{Q}^T\mathbf{a})^T\mathbf{p}^* = \mathbf{a}_I^T\mathbf{h}_I < \|\mathbf{h}_J\|_1 \le \|(\mathbf{Q}\mathbf{p})_J\|_1 \le 1$.

In both cases, the optimal objective value < 1. □

Necessity

Lemma

If $\mathbf{x}^*$ is unique, then the optimal objective of the following primal-dual problems is strictly less than 1.

$$\min_{\mathbf{z}} \|\mathbf{z}_J\|_\infty \quad \text{s.t. } \mathbf{Q}^T\mathbf{z} = -\mathbf{Q}^T\mathbf{a},\ \mathbf{z}_I = 0.$$

$$\max_{\mathbf{p}} (\mathbf{Q}^T\mathbf{a})^T\mathbf{p} \quad \text{s.t. } \|(\mathbf{Q}\mathbf{p})_J\|_1 \le 1.$$

Proof.

Uniqueness of $\mathbf{x}^* \implies \forall\, \mathbf{h} \in \ker(\mathbf{A}) \setminus \{0\},\ \|\mathbf{x}^*\|_1 < \|\mathbf{x}^* + \mathbf{h}\|_1$
$\implies \mathbf{a}_I^T\mathbf{h}_I < \|\mathbf{h}_J\|_1$

Therefore,

- if $\mathbf{p}^* = 0$, then $\|\mathbf{z}_J^*\|_\infty = (\mathbf{Q}^T\mathbf{a})^T\mathbf{p}^* = 0$.
- if $\mathbf{p}^* \neq 0$, then $\mathbf{h} := \mathbf{Q}\mathbf{p}^* \in \ker(\mathbf{A}) \setminus \{0\}$ obeys $\|\mathbf{z}_J^*\|_\infty = (\mathbf{Q}^T\mathbf{a})^T\mathbf{p}^* = \mathbf{a}_I^T\mathbf{h}_I < \|\mathbf{h}_J\|_1 \le \|(\mathbf{Q}\mathbf{p})_J\|_1 \le 1$.

In both cases, the optimal objective value < 1.

Theorem

Suppose $\bar{\mathbf{x}}$ *obeys* $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$. *Then,* $\bar{\mathbf{x}}$ *is the unique solution to* $\min\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ *if and only if the Condition holds.*

Comments:

- the uniqueness requires strong duality result for problems involving $\|\mathbf{z}_J\|_\infty$
- strong duality does not hold for all convex programs
- strong duality does hold for convex polyhedral functions $f(\mathbf{z}_J)$, as well as those with constraint qualifications (e.g., the Slater condition)
- indeed, the theorem generalizes to analysis ℓ_1 minimization: $\|\Psi^T\mathbf{x}\|_1$
- does it generalize to $\sum \|\mathbf{x}_{\mathcal{G}_i}\|_2$ or $\|\mathbf{X}\|_*$? the key is strong duality for $\|\cdot\|_2$ and $\|\cdot\|_*$
- also, the theorem generalizes to the noisy ℓ_1 models (next part...)

Theorem

Suppose $\bar{\mathbf{x}}$ *obeys* $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$. *Then,* $\bar{\mathbf{x}}$ *is the unique solution to* $\min\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ *if and only if the Condition holds.*

Comments:

- the uniqueness requires strong duality result for problems involving $\|\mathbf{z}_J\|_\infty$
- strong duality does not hold for all convex programs
- strong duality does hold for convex polyhedral functions $f(\mathbf{z}_J)$, as well as those with constraint qualifications (e.g., the Slater condition)
- indeed, the theorem generalizes to analysis ℓ_1 minimization: $\|\Psi^T\mathbf{x}\|_1$
- does it generalize to $\sum \|\mathbf{x}_{\mathcal{G}_i}\|_2$ or $\|\mathbf{X}\|_*$? the key is strong duality for $\|\cdot\|_2$ and $\|\cdot\|_*$
- also, the theorem generalizes to the noisy ℓ_1 models (next part...)

Noisy measurements

Suppose $\mathbf{b}$ is contaminated by noise: $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{w}$

Appropriate models to recover a sparse $\mathbf{x}$ include

$$\min \lambda \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \tag{4}$$

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta \tag{5}$$

Theorem

Suppose $\bar{\mathbf{x}}$ is a solution to either (4) or (5). Then, $\bar{\mathbf{x}}$ is the unique solution if and only if the Condition holds for $\bar{\mathbf{x}}$.

Key intuition: reduce (4) to (1) with a specific $\mathbf{b}$. Let $\hat{\mathbf{x}}$ be any solution to (4) and $\mathbf{b}^* := \mathbf{A}\hat{\mathbf{x}}$. All solutions to (4) are solutions to

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}^*.$$

The same applies to (5). Recall that the Condition does not involve $\mathbf{b}$.

Noisy measurements

Suppose $\mathbf{b}$ is contaminated by noise: $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{w}$

Appropriate models to recover a sparse $\mathbf{x}$ include

$$\min \lambda \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \tag{4}$$

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta \tag{5}$$

Theorem

Suppose $\bar{\mathbf{x}}$ is a solution to either (4) or (5). Then, $\bar{\mathbf{x}}$ is the unique solution if and only if the Condition holds for $\bar{\mathbf{x}}$.

Key intuition: reduce (4) to (1) with a specific $\mathbf{b}$. Let $\hat{\mathbf{x}}$ be any solution to (4) and $\mathbf{b}^* := \mathbf{A}\hat{\mathbf{x}}$. All solutions to (4) are solutions to

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}^*.$$

The same applies to (5). Recall that the Condition does not involve $\mathbf{b}$.

Stable recovery

Assumptions:

- $\bar{\mathbf{x}}$ and $\mathbf{y}$ satisfy the Condition. $\bar{\mathbf{x}}$ is the *original signal*.
- $\mathbf{b} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{w}$, where $\|\mathbf{w}\|_2 \leq \delta$
- $\mathbf{x}^*$ is the solution to

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta.$$

Goal: obtain a bound $\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2 \leq C\delta$.

Constant C shall be independent of δ.

Stable recovery
Assumptions:
- $\bar{\mathbf{x}}$ and $\mathbf{y}$ satisfy the Condition. $\bar{\mathbf{x}}$ is the *original signal*.
- $\mathbf{b} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{w}$, where $\|\mathbf{w}\|_2 \leq \delta$
- $\mathbf{x}^*$ is the solution to
$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta.$$
Goal: obtain a bound $\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2 \leq C\delta$.
Constant C shall be independent of δ.

Stable recovery

Lemma

Define $I = \mathrm{supp}(\bar{\mathbf{x}})$ *and* $J = I^c$.

$$\|\mathbf{x}^* - \bar{\mathbf{x}}\|_1 \le C_3\delta + C_4\|\mathbf{x}^*_J\|_1,$$

where $C_3 = 2\sqrt{|I|}\cdot r(I)$ *and* $C_4 = \|\mathbf{A}\|\sqrt{|I|}\cdot r(I) + 1$.

Proof.

- $\|\mathbf{x}^* - \bar{\mathbf{x}}\|_1 = \|\mathbf{x}^*_I - \bar{\mathbf{x}}_I\|_1 + \|\mathbf{x}^*_J\|_1$
- $\|\mathbf{x}^*_I - \bar{\mathbf{x}}_I\|_1 \le \sqrt{|I|}\cdot\|\mathbf{x}^*_I - \bar{\mathbf{x}}_I\|_2 \le \sqrt{|I|}\cdot r(I)\cdot\|\mathbf{A}_I(\mathbf{x}^*_I - \bar{\mathbf{x}}_I)\|_2$, where

$$r(I) := \sup_{\mathrm{supp}(\mathbf{u})=I,\mathbf{u}\neq 0} \frac{\|\mathbf{u}\|}{\|\mathbf{A}\mathbf{u}\|}$$

($r(I)$ is related to one side of the RIP bound)

- introduce $\hat{\mathbf{x}} = [\mathbf{x}^*_I; \mathbf{0}]$.
- $\|\mathbf{A}_I(\mathbf{x}^*_I - \bar{\mathbf{x}}_I)\|_2 = \|\mathbf{A}(\hat{\mathbf{x}} - \bar{\mathbf{x}})\|_2 \le \|\mathbf{A}(\hat{\mathbf{x}} - \mathbf{x}^*)\|_2 + \underbrace{\|\mathbf{A}(\mathbf{x}^* - \bar{\mathbf{x}})\|_2}_{\le 2\delta}$
- $\|\mathbf{A}(\hat{\mathbf{x}} - \mathbf{x}^*)\|_2 \le \|\mathbf{A}\|\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 \le \|\mathbf{A}\|\|\hat{\mathbf{x}} - \mathbf{x}^*\|_1 = \|\mathbf{A}\|\|\mathbf{x}^*_J\|_1$ $\square$

Stable recovery

Recall in the Condition, $\mathbf{y}_I = \text{sign}(\bar{\mathbf{x}})$ and $\|\mathbf{y}_J\|_\infty < 1$

- $\|\mathbf{x}_I^*\|_1 \geq \langle \mathbf{y}_I, \mathbf{x}_I^* \rangle$
- $\|\mathbf{x}_J^*\|_1 \leq (1 - \|\mathbf{y}_J\|_\infty)^{-1}(\|\mathbf{x}_J^*\|_1 - \langle \mathbf{y}_J, \mathbf{x}^* \rangle)$

Therefore,

- $\|\mathbf{x}_J^*\|_1 \leq (1 - \|\mathbf{y}_J\|_\infty)^{-1}(\|\mathbf{x}^*\|_1 - \langle \mathbf{y}, \mathbf{x}^* \rangle) = (1 - \|\mathbf{y}_J\|_\infty)^{-1} d_y(\mathbf{x}^*, \bar{\mathbf{x}})$,

where

$$d_{\mathbf{y}}(\mathbf{x}^*, \bar{\mathbf{x}}) = \|\mathbf{x}^*\|_1 - \|\bar{\mathbf{x}}\|_1 - \langle \mathbf{y}, \mathbf{x}^* - \bar{\mathbf{x}} \rangle$$

is the *Bregman distance* induced by $\|\cdot\|_1$.

Recall in the Condition, $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T)$ so $\mathbf{y} = \mathbf{A}^T\beta$ for some vector β.

- $d_{\mathbf{y}}(\mathbf{x}^*, \bar{\mathbf{x}}) \leq 2\|\beta\|_2\delta$.

Lemma

Under the above assumptions,

$$\|\mathbf{x}_J^*\|_1 \leq 2(1 - \|\mathbf{y}_J\|_\infty)^{-1}\|\beta\|_2\delta.$$

Stable recovery

Theorem

Assumptions:

- $\bar{\mathbf{x}}$ *and* $\mathbf{y}$ *satisfy the Condition.* $\bar{\mathbf{x}}$ *is the original signal.* $\mathbf{y} = \mathbf{A}^T\beta$.
- $\mathbf{b} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{w}$, *where* $\|\mathbf{w}\|_2 \le \delta$
- $\mathbf{x}^*$ *is the solution to*

$$\min \|\mathbf{x}\|_1 \quad \textit{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \le \delta.$$

Conclusion:

$$\|\mathbf{x}^* - \bar{\mathbf{x}}\|_1 \le C\delta,$$

where

$$C = 2\sqrt{|I|}\cdot r(I) + \frac{2\|\beta\|_2(\|\mathbf{A}\|\sqrt{|I|}\cdot r(I) + 1)}{1 - \|\mathbf{y}_J\|_\infty}$$

Comment: a similar bound can be obtained for $\min \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ with a condition on λ.

- If $\|\mathbf{y}_J\|_\infty$ approaches 1, the constant C will blow up.
- $\sqrt{I}$ should be corrected by $\sqrt{|I|}$.

Generalization

All the previous results (exact and stable recovery) generalize to the following models:

$$\min \|\Psi^T \mathbf{x}\|_1 \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\min \lambda \|\Psi^T \mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

$$\min \|\Psi^T \mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \le \delta$$

Assume that $\mathbf{A}$ and Ψ each has independent rows, the update conditions are

Condition

For a given $\bar{\mathbf{x}}$, the index sets $I = \operatorname{supp}(\Psi^T \bar{\mathbf{x}})$ and $J = I^c$ satisfy

1. $\ker(\Psi_J^T) \cap \ker(\mathbf{A}_I) = \{0\}$
2. *there exists $\mathbf{y}$ such that $\Psi \mathbf{y} \in \mathcal{R}(\mathbf{A}^T)$, $\mathbf{y}_I = \operatorname{sign}(\Psi_I^T \bar{\mathbf{x}})$, and $\|\mathbf{y}_J\|_\infty < 1$.*