A Sparse Solution of $\left\{A\underline{x}=\underline{b},\ \underline{x}\geq 0\right\}$ is Necessarily Unique !!

Alfred M. Bruckstein, Michael Elad & Michael Zibulevsky

The Computer Science Department
The Technion – Israel Institute of technology
Haifa 32000, Israel

ICASSP LAS VEGAS NEVADA 2008
2008 IEEE International Conference on Acoustics, Speech, and Signal Processing
March 30 - April 4, 2008 • Las Vegas, Nevada, U.S.A.
IEEE Signal Processing Society

Overview

- We are given an underdetermined linear system of equations $\mathbf{A}\underline{x}=\underline{b}$ (k>n) with a full-rank $\mathbf{A}$.
- There are infinitely many possible solutions in the set $S=\{\underline{x} \mid \mathbf{A}\underline{x}=\underline{b}\}$.
- What happens when we demand positivity $\underline{x}\geq 0$? Surely we should have $S_+=\{\underline{x} \mid \mathbf{A}\underline{x}=\underline{b},\ \underline{x}\geq 0\}\subseteq S$.

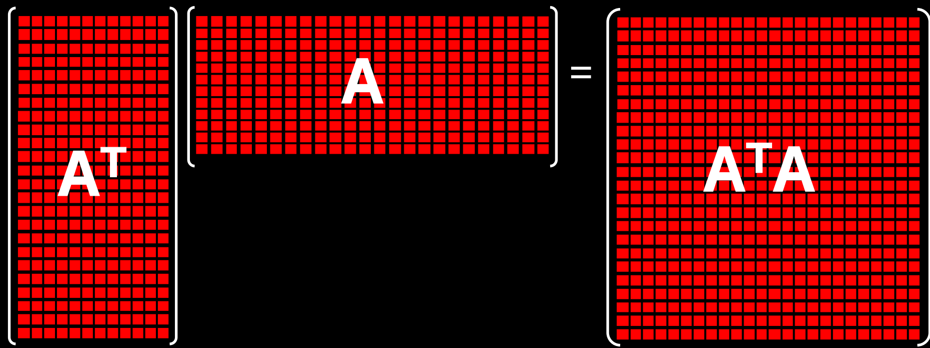
In this talk we shall briefly explain how this result is obtained, and discuss some of its implications

- Our result: For a specific type of matrices $\mathbf{A}$, if a sparse enough solution is found, we get that S_+ is a singleton (i.e. there is only one solution).
- In such a case, the regularized problem $\min f(\underline{x}) \text{ s.t. } \mathbf{A}\underline{x}=\underline{b},\ \underline{x}\geq 0$ gets to the same solution, regardless of the choice of the regularization $f(\underline{x})$ (e.g., L_0, L_1, L_2, L_∞, entropy, etc.).

Preliminaries

Stage 1: Coherence Measures

- Consider the Gram matrix $\mathbf{G}=\mathbf{A}^T\mathbf{A}$:



- Prior work on L_0-L_1 equivalence relies on a mutual-coherence measure defined by

$$\mu(\mathbf{A}) = \max_{i \neq j} \frac{\left|\underline{a}_i^T \underline{a}_j\right|}{\left\|\underline{a}_i\right\|_2 \left\|\underline{a}_j\right\|_2} = \max_{i \neq j} \frac{\left|\mathbf{G}_{ij}\right|}{\sqrt{\mathbf{G}_{ii}\mathbf{G}_{jj}}}$$

($\underline{a}_i$ – the i-th column of $\mathbf{A}$)

- In our work we need a slightly different measure: $\rho(\mathbf{A}) = \max_{i \neq j} \frac{\left|\underline{a}_i^T \underline{a}_j\right|}{\left\|\underline{a}_i\right\|_2^2} = \frac{\left|\mathbf{G}_{ij}\right|}{\mathbf{G}_{ii}}$.

- Note that this **one-sided measure** is weaker (i.e. $\mu(\mathbf{A}) \leq \rho(\mathbf{A})$), but necessary for our analysis.

- Both behave like $1/\sqrt{n}$ for random $\mathbf{A}$ with (0,1)-normal and i.i.d. entries.

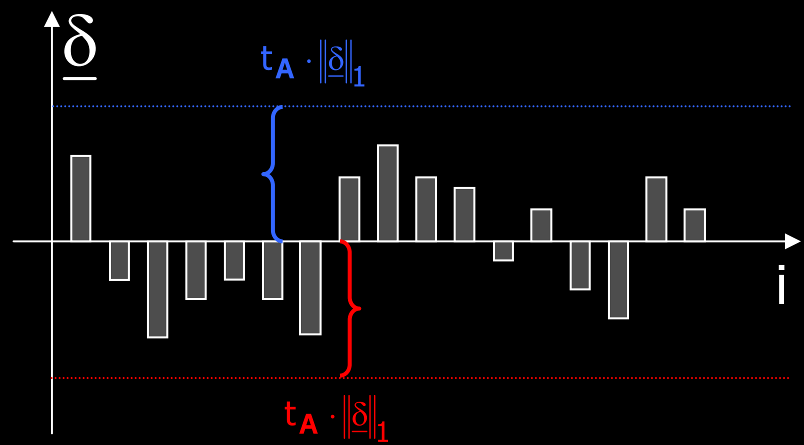
Stage 2: The Null-Space of A

$$N(\mathbf{A}) = \left\{ \underline{\delta} \,\middle|\, \mathbf{A}\,\underline{\delta} = 0 \right\}$$

It is relatively easy to show that

$$\left\|\underline{\delta}\right\|_\infty \le \frac{\rho(\mathbf{A})}{\rho(\mathbf{A})+1} \cdot \left\|\underline{\delta}\right\|_1 = t_{\mathbf{A}} \cdot \left\|\underline{\delta}\right\|_1$$

In words: a vector in the null-space of **A** cannot have arbitrarily large entries relative to its (L_1) length. The smaller the coherence, the stronger this limit becomes.



Stage 3: An Equivalence Theorem

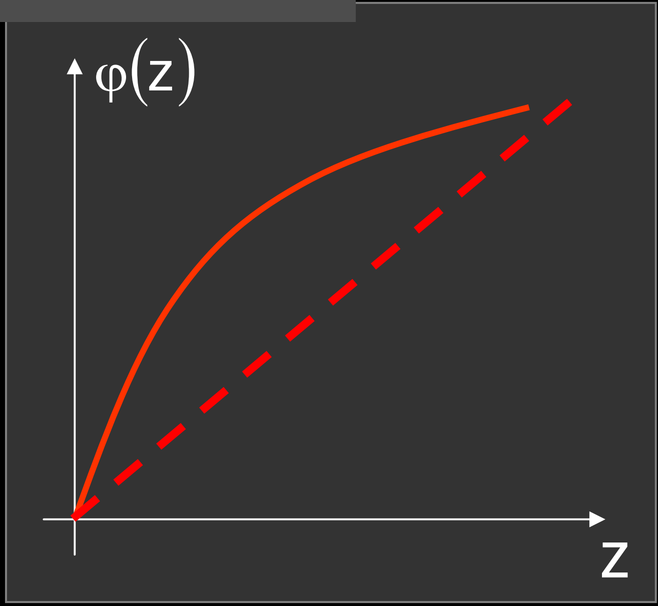
Consider the following problem with a concave semi-monotonic increasing function $\varphi(z)$:

$$\min_{\underline{x}} \sum_{i=1}^{k} \varphi(|x_i|) \quad \text{s.t.} \quad \mathbf{A}\underline{x} = \underline{b}$$

Then

A feasible solution $\hat{\underline{x}}$ (i.e. $\mathbf{A}\hat{\underline{x}}=\underline{b}$) to this problem is the **unique** global optimum if it is sparse enough:

$$\|\hat{\underline{x}}\|_0 < \frac{1}{2t_{\mathbf{A}}}$$



Note: Similar (but somewhat different!) results appear elsewhere [Donoho & Elad `03] [Gribonval & Nielsen `03, `04] [Escoda, Granai and Vandergheynst `04].

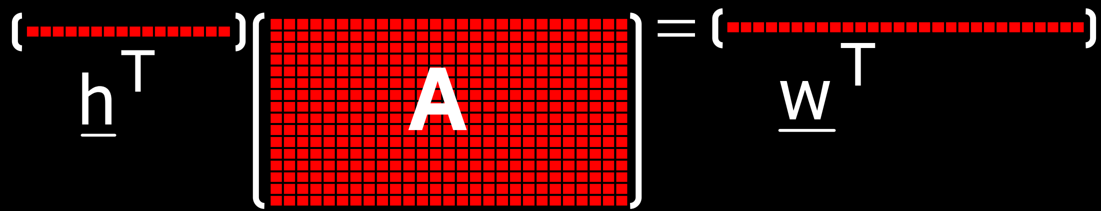
The Main Result

Stage 1: Limitations on A

- So far we considered a general matrix **A.**
- From now on, we shall assume that **A** satisfies the condition:

$$\mathbf{A} \in O^{+} = \left\{ \mathbf{A} \,\middle|\, \exists \underline{h} \in \Re^{n},\ \underline{h}^{T}\mathbf{A} = \underline{w}^{T} > 0 \right\}$$

The span of the rows in **A** intersects the **positive orthant**.



- O^{+} includes
  - All the positive matrices **A**,
  - All matrices having at least one strictly positive (or negative) row.
- Note: If $\mathbf{A} \in O^{+}$ then so does the product **P·A·Q** for any invertible matrix **P** and any diagonal and strictly positive matrix **Q.**

Stage 2: Canonization of $\mathbf{A}\underline{x}=\underline{b}$

- Suppose that we found $\underline{h}$ such that $\underline{h}^T\mathbf{A} = \underline{w}^T > 0$.
- Thus, $\mathbf{A}\underline{x}=\underline{b} \Rightarrow \underline{h}^T\mathbf{A}\underline{x}=\underline{h}^T\underline{b} \Rightarrow \underline{w}^T\underline{x}=\text{Const}$.
- Using the element-wise positive scale mapping $\underline{z} = \text{diag}(\underline{w})\underline{x} = \mathbf{W}\underline{x}$ we get a system of the form:

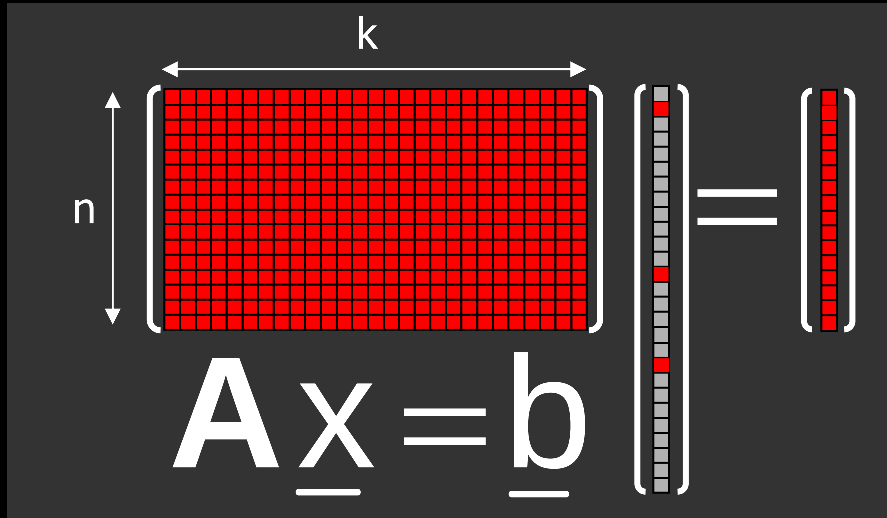
$$\mathbf{A}\mathbf{W}^{-1}\underline{z} = \mathbf{D}\underline{z} = \underline{b}$$

- Implication: we got a linear system of equations for which we also know that every solution for it must sum to Const.
- If x≥0, the additional requirement is equivalent to $\|\underline{z}\|_1 = \text{Const}$. This brings us to the next result …

Stage 3: The Main Result

Given a system of linear equations $\mathbf{A}\underline{x}=\underline{b}$ with $\mathbf{A} \in O^{+}$, we consider the set of non-negative solutions,

$$S_{+} = \{\underline{x} \mid \mathbf{A}\underline{x} = \underline{b} \;\&\; \underline{x} \geq 0\}.$$



Assume that a canonization $\mathbf{D}\underline{z}=\underline{b}$ is performed by finding a suitable $\underline{h}$ and computing the diagonal positive matrix $\mathbf{W}$:

1. $\mathbf{A}^{T}\underline{h} = \underline{w} > 0$
2. $\mathbf{W} = \text{diag}(\underline{w})$
3. $\mathbf{D} = \mathbf{A}\mathbf{W}^{-1}, \underline{z} = \mathbf{W}\underline{x}$

Then

If a sparse solution $\hat{\underline{x}} \in S_{+}$ is found such that

$$\|\hat{\underline{x}}\|_0 < \frac{1}{2t_{\mathbf{D}}}$$

then S_{+} is a **singleton**.

Sketch of the Proof

$\mathbf{A} \in \mathrm{O}^{+}$ →

Find $\underline{h}$ such that $\mathbf{A}^T \underline{h} = \underline{w} > 0$ →

Canonize the system to become $\mathbf{D}\underline{z}=\underline{b}$ using $\mathbf{W} = \mathrm{diag}(\underline{w}), \mathbf{D} = \mathbf{A}\mathbf{W}^{-1}, \underline{z} = \mathbf{W}\underline{x}$ ↓

There is a one-to-one mapping between solutions of $\mathbf{D}\underline{z}=\underline{b}$ and $\mathbf{A}\underline{x}=\underline{b}$ ←

and … any solution of $\mathbf{D}\underline{z}=\underline{b}$ satisfies $\underline{1}^T \underline{z} = \underline{h}^T \underline{b} = \mathrm{Const}$ ←

Non-negativity of the solutions implies that $\|\underline{z}\|_1 = \mathrm{Const}$ ↓

We are given a sparse solution $\hat{\underline{z}} \geq 0$, satisfying $\|\hat{\underline{z}}\|_0 < 1/2t_{\mathbf{D}}$ →

Consider Q_1, the optimization problem $Q_1 : \min \|\underline{z}\|_1 \text{ s.t. } \mathbf{D}\underline{z} = \underline{b}$ →

By Theorem 1, $\hat{\underline{z}}$ is the **unique** global minimizer of Q_1 ↓

No other solution can give the same L_1 length →

The set S_+ contains only $\hat{\underline{z}}$

Some Thoughts

How About Pre-Coherencing?

- The above result is true to **A** or any variant of it that is obtained by **A**′=**PAQ** (**P** invertible, **Q** diagonal and positive).
- Thus, we better evaluate the coherence for a "better-conditioned" matrix **A** (after canonization), with the smallest possible coherence:

$$\rho\left\{ C_{\underline{h}}\left[[\mathbf{P}][\mathbf{A}][\mathbf{Q}] \right] \right\}$$

- One trivial option is **P** that nulls the mean of the columns.
- Better choices of **P** and **Q** can be found numerically [Elad `07].

Solve $\{\mathbf{A}\underline{x} = \underline{b},\ \underline{x} \geq 0\}$

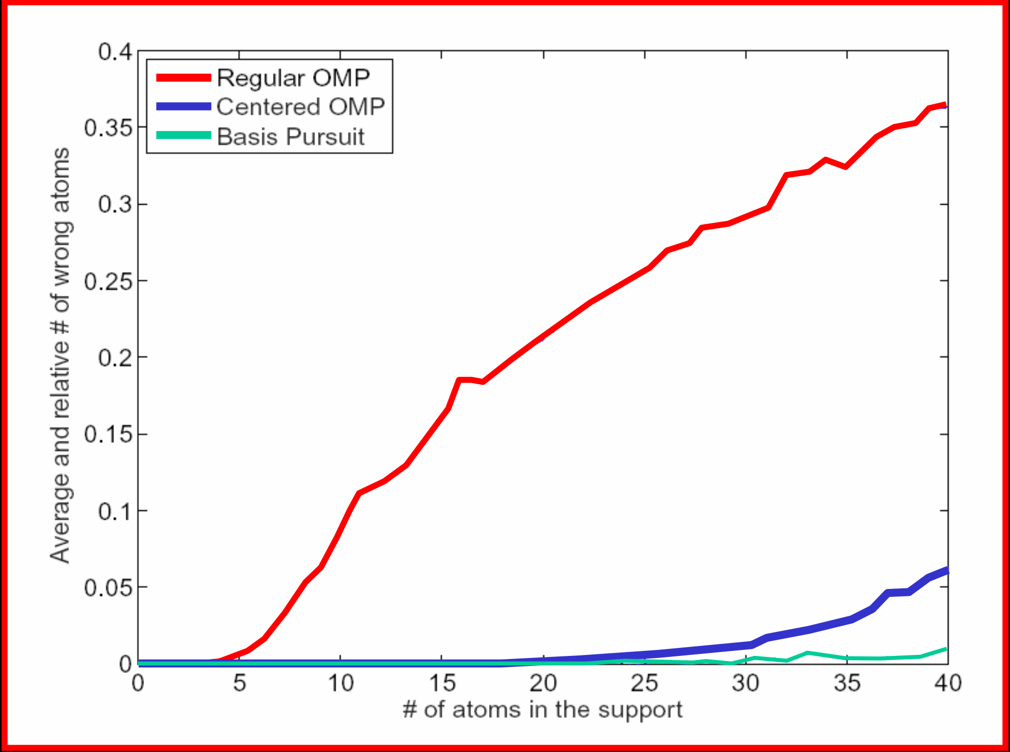
SKIP?

If we are interested in a **sparse result** (which apparently may be unique), we could:

- Trust the uniqueness and regularize in whatever method we want.
- Solve an L_1-regularized problem: min $\|\underline{x}\|_1$ s.t. $\mathbf{A}\underline{x} = \underline{b}$ & $\underline{x} \geq 0$
- Use a greedy algorithm (e.g. OMP):
  - Find one atom at a time by minimizing the residual $\|\mathbf{A}\underline{x}_j - \underline{b}\|_2$,
  - Positivity is enforced both in:
    - Checking which atom to choose,
    - The LS step after choosing an atom.
- OMP is guaranteed to find the sparse result of $\{\mathbf{A}\underline{x} = \underline{b},\ \underline{x} \geq 0\}$, if it sparse enough [Tropp `06], [Donoho, Elad, Temlyakov, `06].

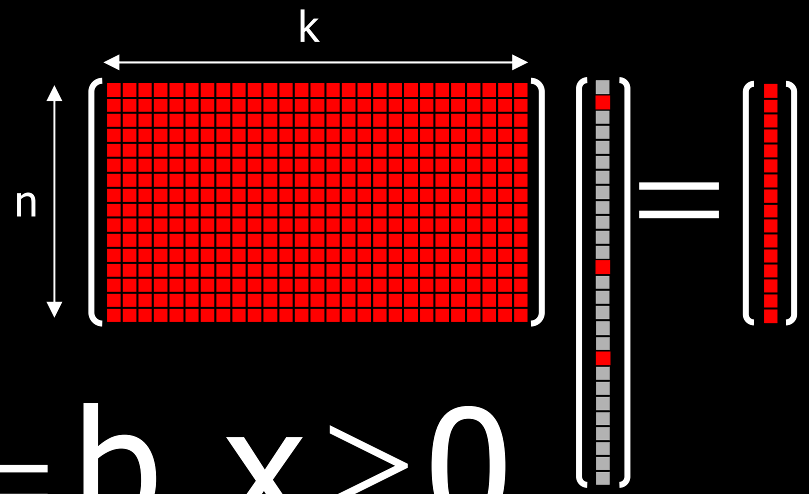
OMP and Pre-Coherencing?

- As opposed to the L_1 approach, pre-cohenercing helps the OMP and improves its performance.
- Experiment:
  - Random non-negative matrix **A** of size 100×200,
  - Generate 40 random positive solutions $\underline{x}$ with varying cardinalities,
  - Check average performance.
  - L_1 performs better BUT takes much longer ($\sim 500/\|\underline{x}\|_0$).
  - OMP may be further improved by better pre-coherencing.



Relation to Compressed Sensing?

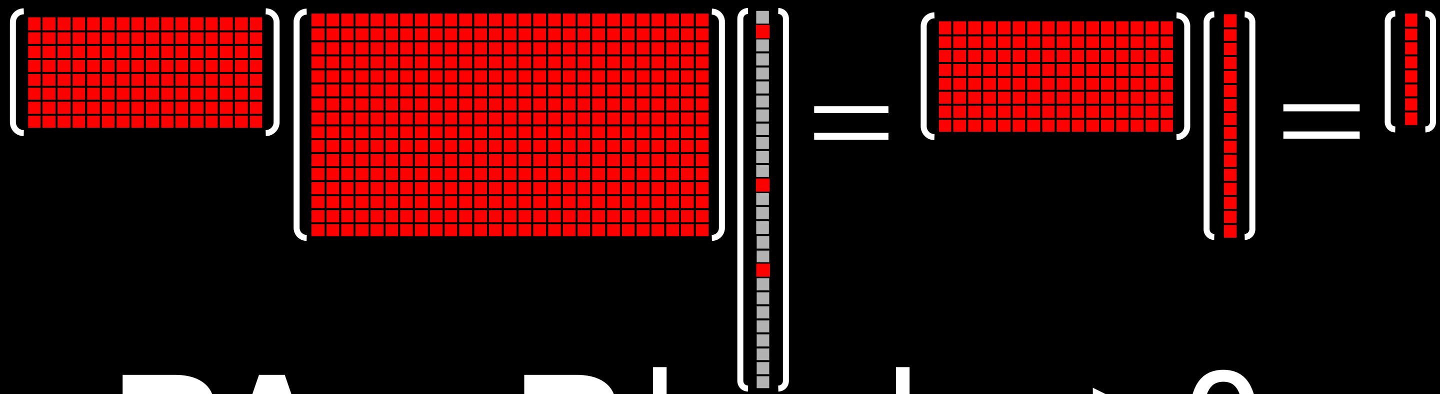
- **A Signal Model:** $\underline{b}$ belongs to a signal family that have a (very) sparse and non-negative representation over the dictionary **A.**



$$\mathbf{A}\underline{x} = \underline{b},\ \underline{x} \geq 0$$

Relation to Compressed Sensing?

- **A Signal Model:** b belongs to a signal family that have a (very) sparse and non-negative representation over the dictionary **A.**
- **CS Measurement:** Instead of measuring b we measure a projected version of it **R**b=d.



$$\mathbf{RA}\underline{x}=\mathbf{R}\underline{b}=\underline{d},\ \underline{x}\geq 0$$

Relation to Compressed Sensing?

- **A Signal Model:** b belongs to a signal family that have a (very) sparse and non-negative representation over the dictionary **A.**
- **CS Measurement:** Instead of measuring b we measure a projected version of it **R**b=d.
- **CS Reconstruction:** We seek the sparsest & non-negative solution of the system **RA**x=d – the scenario describe in this work!!
- **Our Result:** We know that if the (non-negative) representation x was sparse enough to begin with, ANY method that solves this system necessarily finds it exactly.
- **Little Bit of Bad News:** We require too strong sparsity for this claim to be true. Thus, further work is required to strengthen this result.

Conclusions

- Non-negative sparse and redundant representation models are useful in analysis of multi-spectral imaging, astronomical imaging, …
- In our work we show that when a sparse representation exists, it may be the only one possible.
- This explains various regularization methods (entropy, L_2 and even L_∞) that were found to lead to a sparse outcome.
- Future work topics:
  - Average performance (replacing the presented worst-case)?
  - Influence of noise (approximation instead of representation)?
  - Better pre-coherencing?
  - Show applications?