Perfect Recovery Conditions for Non-negative Sparse Modeling
[Y. Itoh, M. F. Duarte, and M. Parente. TSP Jan. 2017]

Lekshmi Ramesh

Signal Processing for Communications Lab
IISc, Bangalore

February 11, 2017

Problem setup

- Observations $y \in \mathbb{R}^m$ obtained from the following linear model:

$$y = Ax + e,$$

  where $A \in \mathbb{R}^{m \times N}$, $x \in \mathbb{R}^N$, and error $e \in \mathbb{R}^m$

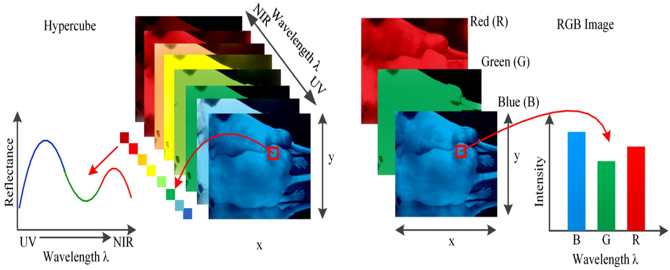
- Coefficient vector x constrained to be non-negative
- Goal: Derive Model Recovery Conditions (MRCs) for the NLASSO

$$\begin{aligned} &\underset{x \in \mathbb{R}^N}{\text{minimize}} \;\; \frac{1}{2}\|y - Ax\|_2^2 + \gamma \|x\|_1 \\ &\text{subject to} \;\; x \geq 0 \end{aligned} \tag{P}$$

MRC: conditions under which (P) correctly identifies the true support

- Contributions
  - MRCs derived for the NLASSO
  - Conditions do not make any assumptions about error e
  - Experimental validation of the proposed conditions on hyperspectral imaging dataset
- Proposed recovery guarantees are of non-uniform type (depend on a particular support)

Hyperspectral imaging



- An imaging technique
  - Device collects information about a scene form across the elctromagnetic spectrum
  - A hyperspectral data cube is created, each pixel has components in different bands
- Certain objects have unique 'fingerprints' in specific bands of the spectrum–helps in identification

Hyperspectral imaging

- In the context of hyperspectral imaging:
  y: $m \times 1$ vector denoting measured spectrum of a pixel
  A: $m \times N$ matrix containing N pure spectral signatures (called endmembers)
  x: $N \times 1$ vector containing fractional abundances of the endmembers

  Express each mixed pixel as a linear combination of endmembers

  x sparse since only few endmembers involved in the observation

Conditions for exact support recovery

- Conditions under which support of solution of NLASSO equals true support Λ
- Positive Subset Coherence (PSC) metric:

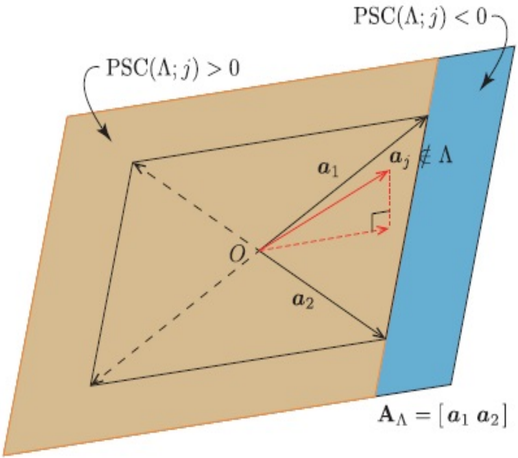
$$\mathrm{PSC}(\Lambda, j) := 1 - \mathbf{1}^\top A_\Lambda^\dagger a_j,$$

where

$$\Lambda \subset [N], |\Lambda| = k$$
$$a_j : j^{th} \text{ column of } A$$

- PSC measures alignment of a_j with the convex cone determined by the columns of A_Λ

Positive Subset Coherence



- H: hyperplane through $a_j,\ j \in \Lambda$
  $\mathrm{PSC}(\Lambda, j) > 0$ when orthogonal projection of a_j onto column space of A_Λ is on the same side of H as the origin

Positive Subset Coherence

- $\mathrm{PSC}(\Lambda, j) = 0$ when $j \in \Lambda$
- $\mathrm{PSC}(\Lambda, j) < 0$ when a_j well-aligned with convex cone determined by columns of A_Λ
- $\mathrm{PSC}(\Lambda, j) > 0$ when a_j less aligned with convex cone determined by columns of A_Λ
- A large PSC preferred

Main Result

Theorem

Let Λ denote a subset of column indices of the dictionary A such that $|\Lambda| = k$ and the columns in A_Λ are linearly independent. Let $\hat{x}$ denote the solution of NLASSO. Then $supp(\hat{x}) = \Lambda$ if

$$A_\Lambda^\dagger y > \gamma (A_\Lambda^\top A_\Lambda)^{-1} \mathbf{1}_k \tag{MCC}$$

$$\langle P_\Lambda^\perp y, a_j \rangle < \gamma PSC(\Lambda, j), \quad \forall j \in \Lambda^C. \tag{NSCC}$$

MCC: Minimum Coefficient Condition

NSCC: Non-linearity vs. Subset Coherence Condition

Interpretation of conditions

- MCC: $A_\Lambda^\dagger y > \gamma (A_\Lambda^\top A_\Lambda)^{-1} \mathbf{1}_k$

  Entries of the least squares solution should be large enough

- NSCC: $\langle P_\Lambda^\perp y, a_j \rangle < \gamma \mathrm{PSC}(\Lambda, j), \quad \forall j \in \Lambda^C$

  $P_\Lambda^\perp y$: Deviation of observation y from the span of A_Λ

  Non-linear distortion that can be tolerated for a $\mathrm{PSC}(\Lambda, j)$

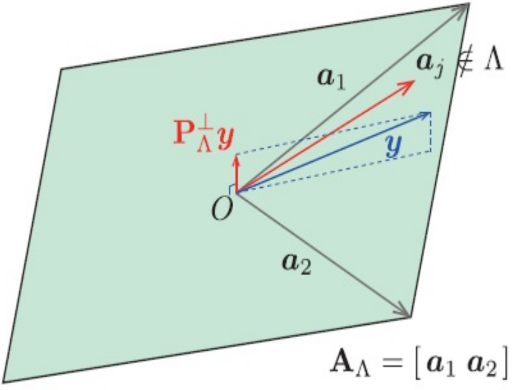
- Dependence of conditions on γ
  - γ high: sparser solution–higher number of missed detections

    To satisfy MCC, γ should be low enough

  - γ low: denser solution–higher number of false alarms

    To satisfy NSCC, γ should be high enough

Geometry of NSCC



Proof of Theorem (Outline)

- Define the restricted NLASSO problem:

$$\begin{aligned} &\underset{v_\Lambda \in \mathbb{R}^k}{\text{minimize}} \; \frac{1}{2}\|y - A_\Lambda v_\Lambda\|_2^2 + \gamma \mathbf{1}^\top v_\Lambda \\ &\text{subject to} \quad v_\Lambda \geq 0 \end{aligned}$$

- Lagrangian:

$$\mathcal{L}(v_\Lambda, \lambda) = \frac{1}{2}\|y - A_\Lambda v_\Lambda\|_2^2 + \gamma \mathbf{1}^\top v_\Lambda - \lambda^\top v_\Lambda$$

Proof of Theorem (Outline)

- KKT conditions: $\hat{v_\Lambda}$ and $\hat{\lambda}$ optimal iff
  - $\hat{v}_\Lambda \geq 0$
  - $\hat{\lambda} \geq 0$
  - $\hat{v}_\Lambda(i)\hat{\lambda}(i) = 0, \quad \forall\, i \in [k]$
  - $\frac{\partial}{\partial v_\Lambda}\mathcal{L}(v_\Lambda, \lambda) = 0$

    $\implies -A_\Lambda^\top(y - A_\Lambda \hat{v}_\Lambda) + \gamma \mathbf{1}_k - \hat{\lambda} = 0$
- Conditions hold for $\hat{\lambda} = 0$ and $\hat{v}_\Lambda = A_\Lambda^\dagger y - \gamma(A_\Lambda^\top A_\Lambda)^{-1}\mathbf{1}_k$; $\hat{v}_\Lambda > 0$ by MCC assumption

Proof of Theorem (Outline)

- By NSCC:

$$
\begin{aligned}
&(y - A_\Lambda A_\Lambda^\dagger y)^\top < \gamma(1 - \mathbf{1}_k^\top A_\Lambda^\dagger a_j) \\
&\Leftrightarrow (y - A_\Lambda A_\Lambda^\dagger y + \gamma (A_\Lambda^\dagger)^\top \mathbf{1}_k)^\top a_j < \gamma \\
&\Leftrightarrow (y - A_\Lambda (A_\Lambda^\dagger y - \gamma (A_\Lambda^\top A_\Lambda)^{-1} \mathbf{1}_k))^\top a_j < \gamma \\
&\Leftrightarrow (y - A_\Lambda \hat{v}_\Lambda)^\top a_j < \gamma
\end{aligned} \tag{A}
$$

- Can show: Condition (A) $\implies$ $\mathrm{supp}(\hat{x}) \subseteq \Lambda$, where $\hat{x}$ is the solution to NLASSO
- Finally, since $\hat{v}_\Lambda(i) > 0 \;\; \forall i \in [k]$, $\mathrm{supp}(\hat{x}) = \Lambda$

Thank you