Finding a Sparse Vector in a Subspace: Linear Sparsity Using Alternating Directions

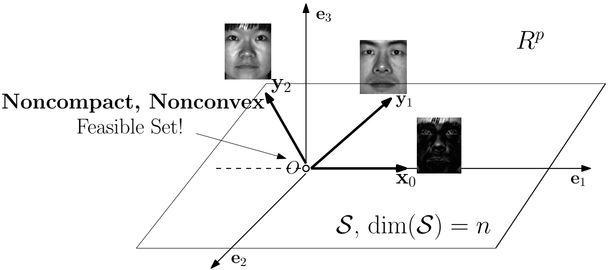
Qing Qu, Ju Sun, John Wright

Electrical Engineering, Columbia University

{*qq2105, js4038, jw2966*}*@columbia.edu*

March 12, 2015

Problem Statement



Problem (Finding a Sparse Vector in a Linear Subspace?)

Mathematically, given any basis $\mathbf{Y}$ *of* $\mathcal{S}$*, can we* ***efficiently*** *solve*

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \quad s.t. \quad \mathbf{x} \in \mathcal{S}, \quad \mathbf{x} \neq \mathbf{0} \tag{1}$$

to find the ***sparsest*** *vector* $\mathbf{x}_0$ *up to scaling and sign difference?*

Motivations and Contributions

- **Constructing sparse basis** is a **meta problem** in the context of (i) dictionary learning, sparse PCA (ii) numerical linear algebra (iii) graphical model learning.
- Previous methods either **break** beyond $O(\sqrt{n})$ **sparsity level** or are **computationally expensive**.

Our Contribution

We present the **first**, **efficient** algorithm, **provably** solving this problem in **proportional sparsity regime** $k \in O(n)$.

Method	Recovery Condition	Total Complexity
ℓ^1/ℓ^∞	~~$k \in O(\sqrt{n})$~~	$O(np^3)$
SDP	~~$k \in O(\sqrt{n})$~~	$O(p^{3.5})$
SOS	$p \geq \Omega(n^2)$, $k \in O(n)$	~~high order $poly(p)$~~
This work	$p \geq \Omega(n^4 \log n)$, $\boxed{k \in O(n)}$	$\boxed{O(n^5 p^2 \log n)}$

Model Assumption

Model Assumption for the Subspace $\mathcal{S}$

We assume a **planted sparse vector (PSV)** model, a **single sparse** vector embedded in an **otherwise random** subspace:

$$\boxed{\mathcal{S} = \operatorname{span}\left(\mathbf{x}_0, \mathbf{g}_1, \cdots, \mathbf{g}_{n-1}\right) \subseteq \mathbb{R}^p,} \tag{2}$$

where $x_0(i) \sim_{\text{i.i.d.}} \frac{1}{\sqrt{\theta p}} \text{Ber}(\theta)$ and $\mathbf{g}_k \sim_{\text{i.i.d.}} \mathcal{N}\left(\mathbf{0}, \frac{1}{p}\mathbf{I}\right)$.

Problem Relaxation

Problem (A Nonconvex Spherical Constraint Problem)

Relax the original problem as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad s.t. \quad \mathbf{x} \in \mathcal{S}, \quad \|\mathbf{x}\|_2 = 1. \tag{3}$$

Let $\mathbf{Y} \in R^{p\times n}$ *is an arbitrary* ***orthonormal basis*** *for* $\mathcal{S}$ *so that* $\mathbf{x} = \mathbf{Yq}$. *We consider an* ***equivalent*** *problem of the form*

$$\boxed{\min_{\mathbf{q}} \|\mathbf{Yq}\|_1, \quad s.t. \quad \|\mathbf{q}\|_2 = 1.} \tag{4}$$

Alternating Direction Method (ADM)

By an **auxiliary variable** $\mathbf{x} \approx \mathbf{Yq}$, further relax the problem

$$\boxed{\min_{\mathbf{q},\mathbf{x}} \frac{1}{2}\|\mathbf{Yq} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \quad \text{s.t.} \quad \|\mathbf{q}\|_2 = 1,} \tag{5}$$

where $\lambda = \frac{1}{\sqrt{p}}$. Minimize the problem by **alternating direction**:

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{Yq} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{6}$$

$$\mathbf{Q} = \arg\min_{\mathbf{q}} \frac{1}{2}\|\mathbf{Yq} - \tilde{\mathbf{x}}\|_2^2 \ \text{ s.t. } \ \|\mathbf{q}\|_2 = 1. \tag{7}$$

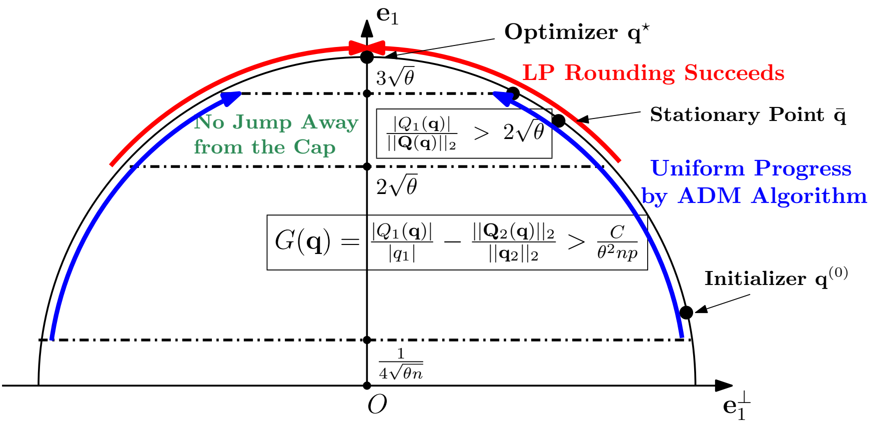
Closed form solutions of (6) and (7) leads to one ADM iterate

$$\boxed{\mathbf{Q}(\mathbf{q}) = \frac{\mathbf{Y}^\top S_\lambda[\mathbf{Yq}]}{\|\mathbf{Y}^\top S_\lambda[\mathbf{Yq}]\|_2},} \tag{8}$$

where $S[z] = \text{sign}(z)(|z| - \lambda)_+$ is the soft-thresholding

Proof Sketch for ADM

Analyze the $\mathbf{Q}(\mathbf{q})$ iterate over the sphere $\mathbb{S}^{n-1}$:



Main Theorem

Theorem (Exact Recovery for the ADM Algorithm, PSV)

Assume our subspace $\mathcal{S} \subseteq \mathbb{R}^p$ follows the PSV model:

- *Apply the ADM algorithm with $\lambda = 1/\sqrt{p}$, using all rows of $\mathbf{Y}$ as initializations for $\mathbf{q}^{(0)}$ to produce $\overline{\mathbf{q}}_1, \ldots, \overline{\mathbf{q}}_p$.*
- *Use a rounding phase to produce $\widehat{\mathbf{q}}_1, \ldots, \widehat{\mathbf{q}}_p$.*
- *Set $i^\star = \arg\min_i \left\|\mathbf{Y}\widehat{\mathbf{q}}_i\right\|_0$, w.h.p. (i.e., $1 - poly^{-1}(p)$),*

$$\boxed{\mathbf{Y}\widehat{\mathbf{q}}_{i^\star} = \gamma \mathbf{x}_0} \tag{9}$$

for some $\gamma \neq 0$, provided that

$$\boxed{p \;>\; \Omega(n^4 \log n), \qquad \text{and} \qquad \theta \;\leq\; \theta_0.}$$

Exploratory Experiment on Yale B Dataset

Our ADM algorithm works on **generic subspaces**!

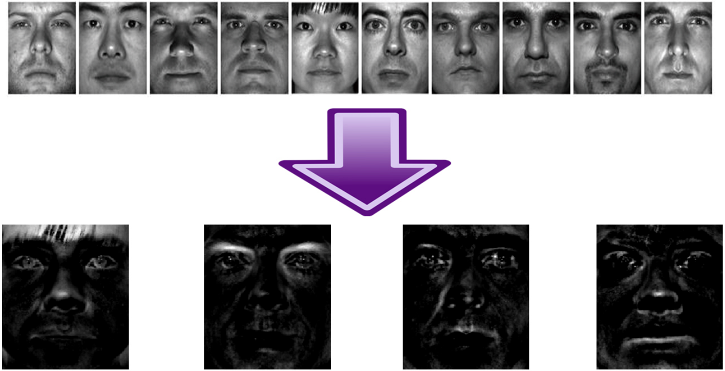


Figure: 4 sparse vectors extracted by ADM algorithm for 10 persons' face under normal illumination.

Take-home Messages

I. Applications for Finding a Sparse Vector in Subspace?

- We provide an **efficient** algorithm with **strong** guarantee.
- **Are there any other applications to use our tools?**

II. Nonconvex Optimization Possible?

- Our work joins **recent success** on **nonconvex optimization**: matrix completion [1], phase retrieval [2].
- Other **nonconvex problems** of interest?
- **Generic tools** to analyze them?

J. Sun, Q. Qu, J. Wright, **Complete Dictionary Learning Over the Sphere**.

[1] P. Jain, P. Netrapalli, and S. Sanghavi. "Low-rank matrix completion using alternating minimization." STOC'13.
[2] E. J. Candes, X. Li, and M. Soltanolkotabi. "Phase retrieval via wirtinger flow: Theory and algorithms." IEEE Trans Inform. Theory, 2014.

References and Acknowledgement

[1] **Qing Qu, Ju Sun, John Wright. "Finding a sparse vector in a subspace: Linear sparsity using alternating directions".** *NIPS'14* (extended version arXiv:1412.4659).
[2] Boaz Barak, Jonathan Kelner, and David Steurer. "Rounding sum-of-squares relaxations". arXiv preprint arXiv:1312.6652, 2013.
[3]Paul Hand and Laurent Demanet. "Scaling Law for Recovering the Sparsest Element in a Subspace". arXiv preprint arXiv:1310.1654, 2013.
[4] Daniel A Spielman, Huan Wang, and John Wright. Exact recovery of sparsely-used dictionaries. In Proceedings of the 25th Annual Conference on Learning Theory, 2012.

Thanks for Prof. **John Wright** and Mr. **Ju Sun**.

Thanks for attendance!