Solution Uniqueness to Problems Involving Convex PA Functions with Applications to Constrained ℓ_1-Minimization

Ahmad Mousavi and **Jinglai Shen**

Institute for Mathematics and its Applications
University of Minnesota
Minneapolis, MN 55455, USA

Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore, MD 21250, USA

July 10, 2020

Sparsity: Most of components are zero
Sparsity Level: Number of nonzero entries
Compressiblity: Well-approximated by sparse signals

Compressed Sensing

To recover a sparse vector $x \in \mathbb{R}^N$ from a measurement vector $y \in \mathbb{R}^m$ with $y = Ax$ (possibly subject to errors) and $A \in \mathbb{R}^{m \times N}$ $(m \ll N)$ is a measurement matrix.

Problem Formulation

Let $\|x\|_0 :=$ Card(x), the original CS problem can be modeled below:

$$\min_{x \in \mathbb{R}^N} \quad \|x\|_0 \qquad \text{subject to} \qquad y = Ax \qquad (P_0).$$

Applications

Engineering, Statistics, Signal and Image Processing, and etc.

Algorithms in Compressed Sensing

Sparsity based Optimization Algorithms

Since $\ell_p \to \ell_0$ as $p \downarrow 0$, one can approximate (P_0) by the following:

$$\min_{x \in \mathbb{R}^N} \quad \|x\|_p \qquad \text{subject to} \qquad y = Ax.$$

Greedy Algorithms

They directly tackle the original problem by making a local optimal decision at each step with an attempt to find a global optimal solution.

Thresholding based Algorithms

Most of them solve the square system $A^T Ax = A^T y$ through a fixed-point method and exploit hard thresholding operator.

Main Problems Used in Sparse Optimization

Generalized Basis-Pursuit

$$\min_{x \in \mathbb{R}^N} \|x\|_p \quad \text{subject to} \quad Ax = y$$

Generalized Basis-Pursuit Denoising I

$$\min_{x \in \mathbb{R}^N} \|x\|_p \quad \text{subject to} \quad \|Ax - y\|_2 \leq \epsilon$$

Generalized Basis-Pursuit Denoising II

$$\min_{x \in \mathbb{R}^N} \|Ax - y\|_2 \quad \text{subject to} \quad \|x\|_p \leq \eta$$

Generalized Ridge Regression

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_p^p$$

Generalized Elastic Net

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Ax - y\|_2^2 + \lambda_1\|x\|_p^r + \lambda_2\|x\|_2^2$$

Geometry of BP_p and BPDN_p for Different p's

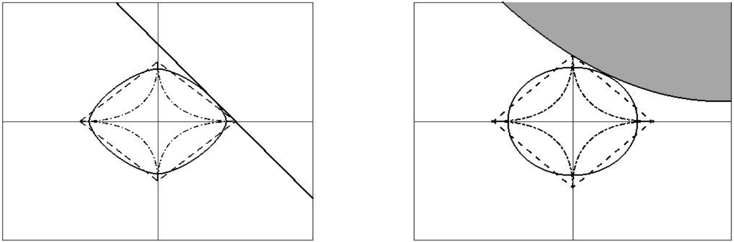

Figura: Geometries of BP and BPDN for different values of p

A Geometrical Illustration

$0 < p < 1 \longrightarrow$ A good choice but nonconvex!

$p = 1 \longrightarrow$ A good choice and results in a convex program!

$p > 1 \longrightarrow$ Not a good choice!

ℓ_1-Norm based Optimization

In sparse recovery, the desired vector is often a solution for one of the following problems:

$$\min_{x\in\mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad Ax = y \qquad \text{(BP)}$$

$$\min_{x\in\mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|Ax - y\|_2 \leq \epsilon \qquad \text{(BPD I)}$$

$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax - y\|_2^2 \quad \text{subject to} \quad \|x\|_1 \leq \eta \qquad \text{(BPD II)}$$

$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_1 \qquad \text{(LASSO)}$$

$Note\ that\ \|.\|_1\ is\ not\ strictly\ convex \implies nonunique\ solution!$

Is this important?

$$If\ not,\ recovery\ process\ is\ not\ successful!$$

A Review on Solution Uniqueness (Individual Recovery)

- Foucart established some results for the problems (BP) and (BPD I).
- Zhang et al. established necessary and sufficient conditions for the mentioned problems when $\|.\|_2^2$ is replaced with a strictly convex smooth function. Later, they replaced $\|x\|_1$ by $\|Ex\|_1$.
- Gilbert replaced $\|.\|_1$ with a polyhedral gauge function: A convex piecewise affine function that is nonnegative, positively homogeneous of degree 1, and vanishes at 0.
- Zhao established necessary and sufficient conditions for nonnegative sparse vectors that satisfy an equality linear system.

Is there a room to improve?

$$Yes!$$

Motivations and Contributions

Motivations

- To add general linear inequality constraints $\longrightarrow$ Dantzig selector:

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|A^T(Ax - y)\|_\infty \leq \epsilon.$$

- To go beyond $\|x\|_1$ and $\|Ex\|_1 \rightarrow$ fused LASSO:

$$\min_{x \in \mathbb{R}^N} \|Ax - y\|_2^2 + \lambda_1 \cdot \|x\|_1 + \lambda_2 \cdot \|D_1 x\|_1.$$

- Explicit dual-based conditions $\rightarrow$ easy and computationally favorable.

Contributions

- Added general linear inequality constraints.
- Considered convex piecewise affine functions, including ℓ_1-norm.
- Developed a unifying approach that recovers all the known results and enables us to tackle new problems.

General Framework

Let $A \in \mathbb{R}^{m\times N}, C \in \mathbb{R}^{p\times N}$ and $f : \mathbb{R}^m \to \mathbb{R}$ be a C^1 strictly convex function. Further, assume $g(x)$ is a convex piecewise affine function.

Main Question

Given a feasible point x^* for any of the below problems, under which conditions this vector is the **unique solution**?

$$\min_{x\in\mathbb{R}^N} g(x) \quad \text{subject to} \quad Ax = y \text{ and } Cx \geq d \qquad \text{(BP-like)}$$

$$\min_{x\in\mathbb{R}^N} g(x) \quad \text{subject to} \quad f(Ax - y) \leq \epsilon \text{ and } Cx \geq d \qquad \text{(BPD I-like)}$$

$$\min_{x\in\mathbb{R}^N} f(Ax - y) \quad \text{subject to} \quad g_1(x) \leq \eta_1, \ldots, g_r(x) \leq \eta_r \text{ and } Cx \geq d \qquad \text{(BPD II-like)}$$

$$\min_{x\in\mathbb{R}^N} f(Ax - y) + g(x) \quad \text{subject to} \quad Cx \geq d \qquad \text{(LASSO-like)}$$

Preliminaries (Convex Piecewise Affine Functions)

Let $g : \mathbb{R}^N \to \mathbb{R}$ be a convex piecewise affine (PA) function:

$$g(x) = \max_{i=1,2,\ldots,l} \left(p_i^T x + \gamma_i \right).$$

For $x^* \in \mathbb{R}^N$ with $Cx^* \geq d$, define $\alpha := \{ i \in \{1, \ldots, p\} \mid (Cx^* - d)_i = 0 \}$,

$$\mathcal{I} := \left\{ i \in \{1, \ldots, l\} \mid p_i^T x^* + \gamma_i = g(x^*) \right\} \text{ and } W := \begin{bmatrix} p_{i_1}^T \\ \vdots \\ p_{i_{|\mathcal{I}|}}^T \end{bmatrix} \in \mathbb{R}^{|\mathcal{I}| \times N}.$$

Finding the matrix W is equivalent to finding the convex hull generators of $\partial g(x^*)$.

A Key Lemma

Lemma

Let $A \in \mathbb{R}^{m\times N}$ *and* $H \in \mathbb{R}^{r\times N}$. *Then,*

$$\{u \in \mathbb{R}^N \mid Au = 0,\ Hu \geq 0\} = \{0\}$$

if and only if the following conditions hold:

(i) $\{u \in \mathbb{R}^N \mid Au = 0,\ Hu = 0\} = \{0\}$; *and*

(ii) There exist $z \in \mathbb{R}^m$ *and* $z' \in \mathbb{R}^r_{++}$ *such that* $A^T z = H^T z'$.

Main Idea of Proof

Define the linear program:

$$\max\ 1^T Hu \quad \text{subject to} \quad Au = 0,\ Hu \geq 0.$$

Then, $\{u \in \mathbb{R}^N \mid Au = 0,\ Hu \geq 0\} = \{0\}$ if and only if

(i) $\{u \in \mathbb{R}^N \mid Au = 0,\ Hu = 0\} = \{0\}$; and

(ii') $u^* = 0$ is the optimal value.

Basis Pursuit-like Problem

Theorem

Let x^* be a feasible point of the optimization problem (BP-like). Then x^* is its unique minimizer if and only if the following conditions hold:

(i) $\{v \in \mathbb{R}^N \mid Av = 0,\ C_{\alpha\bullet}v = 0,\ Wv = 0\} = \{0\}$*; and*

(ii) There exist $w \in \mathbb{R}^m, w' \in \mathbb{R}^{|\alpha|}_{++},$ *and* $w'' \in \mathbb{R}^{|\mathcal{I}|}$ *with* $0 < w'' < 1$ *and* $1^T w'' = 1$ *such that* $A^T w - C^T_{\alpha\bullet} w' + W^T w'' = 0$

Main Steps of Proof

1. For sufficiently small $\|v\|$, we have $g(x^* + v) = g(x^*) + \max_{i \in \mathcal{I}} p_i^T v$.
2. x^* is the unique solution if and only if $v^* = 0$ for

$$\min_{v \in \mathbb{R}^N} \left(\max_{i \in \mathcal{I}} p_i^T v \right) \quad \text{subject to} \quad Av = 0, \quad C_{\alpha\bullet} v \geq 0.$$

3. $v^* = 0$ is the unique solution of this problem if and only if

$$\{v \in \mathbb{R}^N \mid Av = 0,\ C_{\alpha\bullet} v \geq 0,\ \max_{i \in \mathcal{I}} p_i^T v \leq 0\ [\text{or } Wv \leq 0]\} = \{0\}.$$

Finding Matrix W for ℓ_1-norm Function

Let $g : \mathbb{R}^k \to \mathbb{R}$ be such that $g(z) := \|z\|_1 = \max_{1,\dots,2^k} p_i^T z$; where each

$$p_i \in \{(\pm 1, \dots, \pm 1)^T\}.$$

Given $z^* \in \mathbb{R}^k$, let $\mathcal{S} = supp(z^*)$, $\mathcal{I} := \{i \in [2^k] \mid p_i^T z^* = \|z^*\|_1\}$ and $b := \text{sign}(z^*_{\mathcal{S}}) \in \mathbb{R}^{|\mathcal{S}|}$. Then, $|\mathcal{I}| = 2^{|\mathcal{S}^c|}$.

In fact, $\widehat{W} = \left[\widehat{W}_{\bullet\mathcal{S}} \; \widehat{W}_{\bullet\mathcal{S}^c}\right] \in \mathbb{R}^{|\mathcal{I}|\times k}$ where $\widehat{W}_{\bullet\mathcal{S}} = \mathbf{1}.b^T$ and each row of $\widehat{W}_{\bullet\mathcal{S}^c}$ is of the form $(\pm 1, \dots, \pm 1) \in \mathbb{R}^{|\mathcal{S}^c|}$. For example, if $|\mathcal{S}^c| = 2$, then

$$\widehat{W}_{\bullet\mathcal{S}^c} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{bmatrix} \in \mathbb{R}^{|\mathcal{I}|\times N}.$$

Lemma

For the given $z^* \in \mathbb{R}^k$*, the matrix* $\widehat{W}_{\bullet\mathcal{S}^c} \in \mathbb{R}^{|\mathcal{I}|\times k}$ *defined above satisfies:*

1. *The columns of* $\widehat{W}_{\bullet\mathcal{S}^c}$ *are linearly independent.*
2. *For any row* $\widehat{W}_{i\mathcal{S}^c}$*, there is another row such that* $\widehat{W}_{j\mathcal{S}^c} = -\widehat{W}_{i\mathcal{S}^c}$*.*
3. $conv\left\{\widehat{W}^T_{i\mathcal{S}^c} \mid i = 1, \ldots, |\mathcal{I}|\right\} = \left\{u \in \mathbb{R}^{|\mathcal{S}^c|} \mid \|u\|_\infty \leq 1\right\}$*, and*

$$\left\{\sum_{i=1}^{|\mathcal{I}|} \lambda_i \widehat{W}^T_{i\mathcal{S}^c} \mid \sum_{i=1}^{|\mathcal{I}|} \lambda_i = 1, \lambda_i > 0, \ \forall i \in [\mathcal{I}]\right\} = \{u \mid \|u\|_\infty < 1\}.$$

What If $g(x) = \|Ex\|_1$ **with** $E \in \mathbb{R}^{k\times N}$**?**
Given x^*, let $\mathcal{S} = supp(Ex^*)$, $\mathcal{I} := \{i \in [k] \mid p_i^T Ex^* = \|Ex^*\|_1\}$ and $\tilde{b} := \text{sign}(Ex^*_{\mathcal{S}}) \in \mathbb{R}^{|\mathcal{S}|}$. Then, since $\partial g(x^*) = E^T \partial \|.\|_1(Ex^*)$, we have

$$W = \left[\widehat{W}_{\bullet\mathcal{S}} \ \widehat{W}_{\bullet\mathcal{S}^c}\right] \begin{bmatrix} E_{\mathcal{S}\bullet} \\ E_{\mathcal{S}^c\bullet} \end{bmatrix} = \mathbf{1}.b^T + \widehat{W}_{\bullet\mathcal{S}^c} E_{\mathcal{S}^c\bullet} \ \ s.t. \ \ b := E^T_{\mathcal{S}\bullet}\tilde{b}. \quad (1)$$

Lemma

Let the matrix W be defined in (1) for the function $g(x) = \|Ex\|_1$ at x^*. For a given $v \in \mathbb{R}^N, Wv = 0$ if and only if $b^T v = 0$ and $E_{\mathcal{S}^c \bullet} v = 0$

Proposition

Let $g(x) = \|Ex\|_1$, and x^* be feasible point of of the problem (BP-like). Then x^* is the unique minimizer if and only if the following conditions hold:

(a) The matrix $\begin{bmatrix} A \\ C_{\alpha\bullet} \\ E_{\mathcal{S}^c\bullet} \end{bmatrix}$ has full column rank; and

(b) There exist $u \in \mathbb{R}^m, u' \in \mathbb{R}^{|\alpha|}_{++}$, and $u'' \in \mathbb{R}^{|\mathcal{S}^c|}$ with $\|u''\|_\infty < 1$ such that $A^T u + C^T_{\alpha\bullet} u' - E_{\mathcal{S}^c\bullet} u'' = b$.

Main Idea of the Proof:

Use the above Lemma and the part (3) in Lemma on the previous page.

Nonnegative Constraints

Lemma

Let $C = I_N$ and $d = 0$. Then a feasible x^* is the unique minimizer of (BP-like) with $\mathcal{S}$ being its support if and only if the following conditions hold:

(i) The columns of $A_{\bullet\mathcal{S}}$ are linearly independent columns.

(ii) There exists $u \in \mathbb{R}^m$ such that $A^T_{\bullet\mathcal{S}} u = \mathbf{1}$ and $A^T_{\bullet\mathcal{S}^c} u < \mathbf{1}$.

Main Idea for the Proof:

In this case, we have $\hat{b} = (\text{sign}(x^*_{\mathcal{S}})) = \mathbf{1} \in \mathbb{R}^{|\mathcal{S}|}$, $\alpha = \mathcal{S}^c$, $C_{\alpha\mathcal{S}} = 0$ and $C_{\alpha\mathcal{S}^c} = I_{\mathcal{S}^c\mathcal{S}^c}$.
It suffices to show that $A^T_{\bullet\mathcal{S}^c} u < \mathbf{1}$ is equivalent to $\|A^T_{\bullet\mathcal{S}^c} u + u'\|_\infty < 1$ for some $u' > 0$.

Basis Pursuit Denoising I-like Problem

Theorem

Let x^* be a feasible point of (BPD I-like).

C1. *Suppose $f(Ax^* - y) < \epsilon$. Then x^* is the unique minimizer of (BPD I-like) if and only if $\{v \mid C_{\alpha\bullet} v = 0,\ Wv = 0\} = \{0\}$ and there exist $z \in \mathbb{R}^{|\alpha|}_{++}$ and $z' \in \mathbb{R}^{|\mathcal{I}|}$ with $0 < z' < \mathbf{1}$ and $\mathbf{1}^T z' = 1$ such that $C^T_{\alpha\bullet} z = W^T z'$.*

C2. *Suppose $f(Ax^* - y) = \epsilon$. Then x^* is the unique minimizer of (BPD I-like) if and only if the following hold:*

- 2.i $\{v \mid Av = 0,\ C_{\alpha\bullet} v = 0,\ Wv = 0\} = \{0\}$.
- 2.ii *There exist $z \in \mathbb{R}^m, z' \in \mathbb{R}^{|\alpha|}_{++}$, and $z'' \in \mathbb{R}^{|\mathcal{I}|}$ with $0 < z'' < \mathbf{1}$ and $\mathbf{1}^T z'' = 1$ such that $A^T z - C^T_{\alpha\bullet} z' + W^T z'' = 0$.*
- 3.iii *If $\mathcal{K} := \{v \mid \big(\nabla f(Ax^* - y)\big)^T Av < 0,\ C_{\alpha\bullet} v \geq 0\} \neq \emptyset$, then there exist $w \in \mathbb{R}^{|\alpha|}_{+}$, and $w' \in \mathbb{R}^{|\mathcal{I}|}_{+}$ such that $A^T \nabla f(Ax^* - y) - C^T_{\alpha\bullet} w + W^T w' = 0$.*

Main Steps of the Proof for BPD I-like

Case 1:
Since $f(Ax - y) \leq \epsilon$ is inactive, through continuity of f, consider

$$\min \ g(x) \quad \text{subject to} \quad Cx \geq d$$

Case 2:
1. $r(Av) := f(Ax^* - y + Av) - f(Ax^* - y) - \big(\nabla f(Ax^* - y)\big)^T Av$. Due to strict convexity: $r(Av) \geq 0$ and $r(Av) = 0$ if and only if $Av = 0$.
2. Let $h := A^T \nabla f(Ax^* - y)$ and $\tilde{g}(v) := \max_{i \in \mathcal{I}} p_i^T v$. Consider

$$\min \ \tilde{g}(v) \quad \text{subject to} \quad h^T v + r(Av) \leq 0, \ \ C_{\alpha\bullet} v \geq 0.$$

3. $v^* = 0$ is the unique solution of the latter if and only if
(a) Uniquely $u^* = 0 = \arg\min \ \tilde{g}(u)$ subject to $Av = 0, \ C_{\alpha\bullet} v \geq 0$.
(b) If $\mathcal{K} := \{v \mid h^T v < 0, \ C_{\alpha\bullet} v \geq 0\} \neq \emptyset$, then $\tilde{g}(u) > 0$ for all $u \in \mathcal{K}$.
4. Use the Motzkin's Transposition theorem for (b).

How to Verify These Conditions?

Solution uniqueness criteria that we found consist of:

(a) full column rank condition for a matrix $\longrightarrow$ $Linear\ Algebra$

(b) consistency of a linear system with non-strict inequalities $\longrightarrow$ LP

(c) consistency of another linear system with strict inequality $\longrightarrow$?

Lemma

Let $A \in \mathbb{R}^{m \times N},\ y \in \mathbb{R}^m$ *be given. Then, the linear inequality system*

$$Ax = y, \quad x > 0;$$

has a solution if and only if the following linear program is solvable and attains a positive optimal value:

$$\max\ \epsilon \qquad \textit{subject to} \qquad Ax = y, \quad x \geq \epsilon.1, \quad \epsilon \leq 1.$$

References

- S. Foucart, and H. Rauhut. A Mathematical Introduction to Compressed Sensing, *Basel: Birkhäuser*, 2013.
- J. C. Gilbert. On the Solution Uniqueness Characterization in the ℓ_1 Norm and Polyhedral Gauge Recovery, *Journal of Optimization Theory and Applications*, Vol. 172(1), pp. 70-101, 2017.
- S. Mousavi, and Shen. Solution Uniqueness of Convex Piecewise Affine Functions Based Optimization with Applications to Constrained ℓ_1 Minimization. *ESAIM: Control, Optimisation, and Calculus of Variations*, In Press, DOI: 10.1051/cocv/2018061. `https://arxiv.org/abs/1711.05882`.
- H. Zhang, M. Yan, and W. Yin. One Condition for Solution Uniqueness and Robustness of both ℓ1-Synthesis and ℓ1-Analysis Minimizations, *Advances in Computational Mathematics*, Vol. 42(6), pp. 1381-1399, 2016.
- H. Zhang, W. Yin, and L. Cheng. Necessary and Sufficient Conditions of Solution Uniqueness in 1-Norm Minimization, *Journal of Optimization Theory and Applications*, Vol. 164(1), pp. 109-122, 2015.