# On Equivalence of $\ell_1$ Norm Based Basic Sparse Representation Problems

Rui Jiang, Hong Qiao
Institute of Automation, CAS,
Beijing, P.R. China
Email: jiangrui627@163.com, hong.qiao@ia.ac.cn

Bo Zhang
LSEC and Institute of Applied Mathematics, AMSS, CAS,
Beijing, P.R. China
Email: b.zhang@amt.ac.cn

***Abstract*—The $\ell_1$ norm regularization problem, the $\ell_1$ norm minimization problem and the $\ell_1$ norm constraint problem are known collectively as the $\ell_1$ norm based Basic Sparse Representation Problems (BSRPs), and have been popular basic models in the field of signal processing and machine learning. The equivalence of the above three problems is one of the crucial bases for the corresponding algorithms design. However, to the best our knowledge, this equivalence issue has not been addressed appropriately in the existing literature. In this paper, we will give a rigorous proof of the equivalence of the three $\ell_1$ norm based BSRPs in the case when the dictionary is an overcomplete and row full rank matrix.**

***Index Terms*—Equivalence, $\ell_1$ norm regularization problem, $\ell_1$ norm minimization problem, $\ell_1$ norm constraint problem.**

## I. Introduction

In the field of signal processing and analysis, sparsity prior based models for signal source modeling and feature learning have been successfully applied in various signal restoration tasks and recognition tasks during the last decade [1], [2]. Among these models, the $\ell_1$ norm based Basic Sparse Representation Problems (BSRPs) are most popular and play as the basis of most generalized models. Specifically, we denote a signal and a dictionary by $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{D} \in \mathbb{R}^{n\times K}$ respectively, where $\boldsymbol{D} \in \mathbb{R}^{n\times K}$ is an overcomplete and row full rank matrix, i.e., $n < K$ and $\mathrm{rank}(\boldsymbol{D}) = n$. $\boldsymbol{x} \in \mathbb{R}^K$ is referred to as the sparse representation of the signal $\boldsymbol{y}$ over the dictionary $\boldsymbol{D}$. The $\ell_1$ norm based BSRPs have the following three mathematical formulations.

- $\ell_1$ norm regularization problem
$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \quad , \tag{1}$$
where $\lambda \geq 0$ is a parameter.
- $\ell_1$ norm minimization problem
$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2 \leq \varepsilon, \tag{2}$$
where $\varepsilon \geq 0$ is a parameter.
- $\ell_1$ norm constraint problem
$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_1 \leq \sigma, \tag{3}$$
where $\sigma \geq 0$ is a parameter.

Among these problems, problem (2) is referred to as the Basis Pursuit DeNosing (BPDN) problem [3] in the field of signal processing. And both problem (1) and problem (3) have been referred to as the Least Absolute Shrinkage and Selection Operator (LASSO) problem [4] in the statistics and machine learning literature.

To put it simply, the notion of equivalence of $\ell_1$ norm based BSRPs is that, for every two of above-mentioned problems, given certain ranges of parameters, if the solution set of one problem with some parameter value in its given range is known, then the solution set of another one with some parameter value in its corresponding range is also ready to found, and vice versa. More strictly speaking, taking the equivalence of problem (1) and problem (2) for example, given ranges of parameters, say, $\Lambda$ and $\Upsilon$, for every $\lambda_0 \in \Lambda$, there exists a unique $\varepsilon_0 \in \Upsilon$ such that, when $\lambda = \lambda_0$ and $\varepsilon = \varepsilon_0$, the two problems have the same solution set. Furthermore, in the sense that the two problems have the same solution set, there is one-to-one correspondence between $\lambda_0$ and $\varepsilon_0$.

This equivalence is one of the bases of algorithms design for $\ell_1$ norm based BSRPs. By applying cross validation to choose corresponding parameters in certain ranges, solving problem (2) or problem (3) equates to solving problem (1). Moreover, problem (1) is an unconstrained convex optimization problem, which can lead to more conveniences for algorithms design. In the literature, many fast algorithms for solving the three problems aim at solving problem (1). Typical algorithms include Feature-sign Search (Feature-sign) algorithm [5], Coordinate Descent (CD) method [6], Iterative Soft Thresholding Algorithm (ISTA) [7], Fast Iterative Shringakge-Thresholding Algorithm (FISTA) [8], Iterative Reweighted Least Square (IRLS) [9], Alternating Direction Method of Multipliers (ADMM) [10], and some stochastic optimization methods proposed in recent years, such as [11], [12].

In the literature relating to algorithms for $\ell_1$ norm based BSRPs, some papers plunged straight into the discussion on the target optimization problem, such as [5], [7], [8], [11], [12], while others, for example, [1], [2], [6], [9], [10], first mentioned the equivalence of the above three problems but have not got into the details. This results in a gap in understanding to readers, especially those unfamiliar with the sparsity topic. In [13], a blog post in Chinese, the author has also noticed this issue. In this blog post, problem (2) and problem (3) were reformulated as unconstrained convex optimization problems by using the indicator function, then the discussion on the equivalence of the three problems, more precisely, the equivalence of problem (1) and problem (3), was launched. However, in the discussion, the author has not described the properties of the dictionary, which is key conditions for the equivalence. And the uniqueness of solutions to the three problem was used in proofs as a known conclusion, but this is not always true. These issues led to a line of reasoning that is not convincing enough.

To fill this gap in the literature, in this paper, we propose a new proof for the equivalence of $\ell_1$ norm based BSRPs, in the usual case when the dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times K}$ is an overcomplete and row full rank matrix, i.e., $n < K$ and $\mathrm{rank}(\boldsymbol{D}) = n$. We first fully discuss properties of solutions of problem (1), (2) and (3) respectively and provide three corresponding lemmas. Based on these lemmas, we prove the equivalence of the three problems in Theorem 3.1 and 3.2. And we conclude the whole paper in the last section.

## II. Properties of Solutions of $\ell_1$ Norm Based BSRPs

In this section, we discuss properties of solutions of the $\ell_1$ norm regularization problem (1), the $\ell_1$ norm minimization problem (2) and the $\ell_1$ norm constraint problem (3). We first introduce some useful concepts and conclusions, then turn to prove properties of solutions of the three problems in the next subsection.

### A. Preliminaries

In this part, we first introduce the definitions of subgradient and subdifferential, and the optimality condition of unconstrained convex optimization problems, then we introduce a theorem for subdifferential calculation. Based on these concepts and conclusions, we give the optimality conditions of the $\ell_1$ norm regularization problem (1).

**Definition 2.1 (Subgradient, Subdifferential [14])** $\boldsymbol{g} \in \mathbb{R}^K$ *is said to be a subgradient of a convex function* $f(\boldsymbol{x})$ *at a point* $\boldsymbol{x}_0 \in \mathrm{dom}f$[1], *if for every* $\boldsymbol{x} \in \mathrm{dom}f$ *the following condition*

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \geq \boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_0)$$

*is satisfied. The set of all subgradients at* $\boldsymbol{x}_0$ *is called the subdifferential of* $f(\boldsymbol{x})$ *at* $\boldsymbol{x}_0$ *and is denoted by* $\partial f(\boldsymbol{x}_0)$.

Using the concept of subgradient and subdifferential, we can get the optimality condition of unconstrained convex optimization problems.

**Theorem 2.1 (Optimality condition of unconstrained convex optimization problems [14])** $\boldsymbol{x}^*$ *is a global optimal point of a convex function* $f(\boldsymbol{x})$, *if and only if* $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*)$.

Now, we introduce a rule for calculating the subdifferential of a particular convex function at any point in its domain.

**Theorem 2.2 (Moreau-Rockafellar [14])** *Let* $f_1, f_2, \cdots, f_p : \mathbb{R}^K \to [-\infty, +\infty]$ *be proper convex functions. Let*

$$f = f_1 + f_2 + \cdots + f_p.$$

*Then for every* $\boldsymbol{x} \in \bigcap_{i=1}^{p} \mathrm{dom}f_i$, *we have*

$$\partial f(\boldsymbol{x}) \supset \partial f_1(\boldsymbol{x}) + \partial f_2(\boldsymbol{x}) + \cdots + \partial f_p(\boldsymbol{x}).$$

*Furthermore, if convex sets* $\mathrm{relint}(\mathrm{dom}f_i)(i = 1, 2, \cdots, p)$ [2] *have nonempty intersection, then for every* $\boldsymbol{x} \in \bigcap_{i=1}^{p} \mathrm{dom}f_i$ *we have*

$$\partial f(\boldsymbol{x}) = \partial f_1(\boldsymbol{x}) + \partial f_2(\boldsymbol{x}) + \cdots + \partial f_p(\boldsymbol{x}).$$

By Theorem 2.1 and 2.2, we can derive the optimality conditions of problem (1).

**Theorem 2.3 (Optimality condition of the $\ell_1$ norm regularization problem** (1)**)** $\boldsymbol{x}^*$ *is a global optimal point of the* $\ell_1$ *norm regularization problem* (1) *if and only if* $\boldsymbol{x}^*$ *satisfies the following conditions*

$$\begin{cases} -[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i = \lambda, & x_i^* > 0; \\ -\lambda \leq [2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i \leq \lambda, & x_i^* = 0; \\ [2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i = \lambda, & x_i^* < 0. \end{cases} \tag{4}$$

We now briefly review the concept of the relative interior of a set and the concept of the relative interior of a convex set. Then we introduce the Slater's condition and the Karush-Kuhn-Tucker (KKT) conditions, and their relation. These concepts and conclusions are useful in the proof of properties of solution set of problem (2) and (3).

**Definition 2.2 (Relative interior of a set [15])** *The interior of a set* $\mathcal{C}$ *within its affine hull is said to be the relative interior of* $\mathcal{C}$ *and denoted by* $\mathrm{relint}(\mathcal{C})$.

The definition of the relative interior of a general set is very difficult to use in later proofs. Thus we further provide a theorem [14] on the relative interior of a convex set. And we will instead use this theorem in the following proofs.

**Theorem 2.4 (Relative interior of a convex set [14])** *Let* $\mathcal{C} \in \mathbb{R}^K$ *be a nonempty convex set. The relative interior of* $\mathcal{C}$ *can be mathematically expressed as*

$$\mathrm{relint}(\mathcal{C}) = \{\boldsymbol{x} \in \mathcal{C} | \forall \boldsymbol{x}', \exists \rho > 1 : \rho\boldsymbol{x} + (1-\rho)\boldsymbol{x}' \in \mathcal{C}\}. \tag{5}$$

Given a convex optimization problem

$$\min_{\boldsymbol{x}} \quad f_0(\boldsymbol{x}) \quad \text{s.t.} \quad f_i(\boldsymbol{x}) \leq 0, (i = 1, 2, \cdots, q). \tag{6}$$

Based on the concept of the relative interior of a set, the Slater's condition of problem (6) can be defined as follows.

**Definition 2.3 (Slater's condition for problem** (6) **[16])** *For the convex optimization problem* (6), *if its primal optimal value is equal to its dual optimal value, we say that the strong duality holds. Let* $\mathcal{D} = \bigcap_{i=0}^{q} \mathrm{dom}f_i$. *If the slater's condition holds, i.e., there exists a* $\boldsymbol{x} \in \mathrm{relint}(\mathcal{D})$ *such that* $f_i(\boldsymbol{x}) < 0 (i = 1, 2, \cdots, q)$ *holds, we say that the strong duality holds.*

Now, we introduce the KKT conditions for problem (6), and its relation to the Slate's condition.

[1]We use $\mathrm{dom}f$ to denote the domain of $f(\boldsymbol{x})$.

[2]The definition of the relative interior of a set will be given in Definition 2.2.

**Definition 2.4 (KKT conditions for problem** (6) **[16])** *Let* $\tilde{\boldsymbol{x}} \in \mathbb{R}^m$ *and* $\tilde{\boldsymbol{\eta}} \in \mathbb{R}^q$*. The conditions*

$$f_i(\tilde{\boldsymbol{x}}) \leq 0, (i = 1, 2, \cdots, q),$$

$$\tilde{\boldsymbol{\eta}} \succeq \mathbf{0},$$

$$\tilde{\eta}_i f_i(\tilde{\boldsymbol{x}}) = 0, (i = 1, 2, \cdots, q),$$

$$\mathbf{0} \in \partial \mathcal{L}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\eta}}) = \partial f_0(\tilde{\boldsymbol{x}}) + \sum_{i=1}^{q} \tilde{\eta}_i \partial f_i(\tilde{\boldsymbol{x}}),$$

*where* $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\eta})$ *is the Lagrangian, are referred to as the KKT conditions for problem* (6)*.*

**Theorem 2.5 ( [16])** *If the strong duality holds for problem* (6)*, then* $\boldsymbol{x}^* \in \mathbb{R}^K$ *and* $\boldsymbol{\eta}^* \in \mathbb{R}^q$ *are respectively the primal and dual optimal solutions of problem* (6)*, if and only if* $\boldsymbol{x}^*$ *and* $\boldsymbol{\eta}^*$ *satisfy the KKT conditions for problem* (6)*.*

According to Theorem 2.5, if the Slate's condition for problem (6) holds, the KKT conditions are necessary and sufficient conditions for the optimality of problem (6) and its dual.

### *B. Properties of Solutions of $\ell_1$ Norm Based BSRPs*

Based on the optimality condition of the $\ell_1$ norm regularization problem (1), i.e., Theorem 2.3, we can analyze properties of solutions of problem (1).

**Lemma 2.1 (Properties of solutions of the $\ell_1$ norm regularization problem** (1)**)**

1) *If* $\lambda = 0$*, the solution set of problem* (1) *is* $\{\boldsymbol{x} \in \mathbb{R}^K | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$*; If* $\lambda \geq \|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{y}\|_\infty$*, the zero vector is the only solution to problem* (1)*.*
2) *Problem* (1) *has either one solution or infinitely many solutions.*
3) *If* $0 < \lambda < \|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{y}\|_\infty$ *and problem* (1) *has infinitely many solutions, then every two different solutions, say,* $\boldsymbol{x}_1^*$ *and* $\boldsymbol{x}_2^*$*, satisfy* $\boldsymbol{D}\boldsymbol{x}_1^* = \boldsymbol{D}\boldsymbol{x}_2^*$ *and* $\|\boldsymbol{x}_1^*\|_1 = \|\boldsymbol{x}_2^*\|_1$*.*

*Proof:* First, we prove the first conclusion. If $\lambda = 0$, problem (1) equates to the least squares problem. Since that the dictionary $\boldsymbol{D}$ is an overcomplete and row full rank matrix, there is necessarily a $\boldsymbol{x} \in \mathbb{R}^K$ such that $\boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}$. Thus the solution set of problem (1) in the case that $\lambda = 0$ is $\{\boldsymbol{x} \in \mathbb{R}^K | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$. If $\lambda \geq \|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{y}\|_\infty$, then $\mathbf{0} \in \mathbb{R}^K$ satisfies the optimality conditions (4) of problem (1), which means $\mathbf{0} \in \mathbb{R}^K$ is the only solution of problem (1). As for the uniqueness of the zero vector solution, we will prove it later.

The next thing to do is proving the second conclusion. Due to the convexity of the objective of problem (1) and the fact that its feasible region is $\mathbb{R}^K$, the solution set of problem (1) is nonempty. If solutions of problem (1) are not unique and we suppose that $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ are two different solutions, that is,

$$\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}_1^*\|_2^2 + \lambda\|\boldsymbol{x}_1^*\|_1 = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}_2^*\|_2^2 + \lambda\|\boldsymbol{x}_2^*\|_1.$$

We denote this value by $\Delta$. Now, consider a point on the line segment between $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$, i.e., $\boldsymbol{x}_3 = \alpha\boldsymbol{x}_1^* + (1-\alpha)\boldsymbol{x}_2^*$, where $0 < \alpha < 1$. It follows that

$$\begin{aligned} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}_3\|_2 = & \ \|\boldsymbol{y} - \boldsymbol{D}[\alpha\boldsymbol{x}_1^* + (1-\alpha)\boldsymbol{x}_2^*]\|_2^2 \\ \leq & \ \alpha\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}_1^*\|_2 + (1-\alpha)\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}_2^*\|_2, \end{aligned} \quad (7)$$

and

$$\|\boldsymbol{x}_3\|_1 = \|\alpha\boldsymbol{x}_1^* + (1-\alpha)\boldsymbol{x}_2^*\|_1 \leq \alpha\|\boldsymbol{x}_1^*\|_1 + (1-\alpha)\|\boldsymbol{x}_2^*\|_1. \quad (8)$$

From (7) and (8), we can see that $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}_3\|_2^2 + \lambda\|\boldsymbol{x}_3\|_1 \leq \Delta$. In fact, the equality holds. Otherwise, this would contradict the assumption that $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ are two solution of problem (1), which means that each point on the line segment between $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ is a solution to problem (1).

Finally, we verify the third conclusion. Suppose that $\boldsymbol{D}\boldsymbol{x}_1^* \neq \boldsymbol{D}\boldsymbol{x}_2^*$. Since the $\ell_2$ ball is strictly convex, there is necessarily a $0 < \alpha' < 1$ making the inequality (7) strict. However, this would contradict to the assumption that both $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ are solutions of problem (1). Consequently, we infer that $\boldsymbol{D}\boldsymbol{x}_1^* = \boldsymbol{D}\boldsymbol{x}_2^*$. By the fact that $\lambda > 0$, we further get $\|\boldsymbol{x}_1^*\|_1 = \|\boldsymbol{x}_2^*\|_1$. Hence, the third conclusion is proved. Now, we turn to complete the proof of the first conclusion. If there is a non-zero vector solution, say, $\boldsymbol{x}^*$, to problem (1) when $\lambda \geq \|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{y}\|_\infty$, this would lead to the contradiction that $\|\boldsymbol{x}^*\|_1 = \|\mathbf{0}\|_1 = 0$. This completes the proof. ■

Next, we analyze the properties of solutions of the $\ell_1$ norm minimization problem (2).

**Lemma 2.2 (Properties of solutions of the $\ell_1$ norm minimization problem** (2)**)**

1) *If* $\varepsilon = 0$*, the solution set of problem* (2) *is* $\{\boldsymbol{x} \in \mathbb{R}^K | \boldsymbol{x} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \ \text{s.t.} \ \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$*; If* $\varepsilon \geq \|\boldsymbol{y}\|_2$*, the zero vector is the only solution to problem* (2)*.*
2) *Problem* (2) *has either one solution or infinitely many solutions.*
3) *If* $0 < \varepsilon < \|\boldsymbol{y}\|_2$*, all solutions of problem* (2) *are on the border of the feasible region. When problem* (2) *has infinitely many solutions, every two different solutions, say,* $\boldsymbol{x}_1^*$ *and* $\boldsymbol{x}_2^*$*, satisfy* $\boldsymbol{D}\boldsymbol{x}_1^* = \boldsymbol{D}\boldsymbol{x}_2^*$*.*

*Proof:* We first show that the first conclusion holds. When $\varepsilon = 0$, problem (2) is reduced to the following problem

$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}.$$

The proof is straightforward. If $\varepsilon \geq \|\boldsymbol{y}\|_2$, $\mathbf{0} \in \mathbb{R}^K$ is a feasible point of problem (2), it is obvious that the zero vector solution is unique.

Now, we prove the second conclusion. Since that the dictionary $\boldsymbol{D}$ is overcomplete and row full rank, then the feasible region of problem (2) is nonempty, that is, the solution set of problem (2) is nonempty. When there is not only one solution and we use $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ to denote two different solutions, i.e., $\boldsymbol{x}_1^* \neq \boldsymbol{x}_2^*$ and $\|\boldsymbol{x}_1^*\|_1 = \|\boldsymbol{x}_2^*\|_1$. We denote this value by $\sigma^*$. Consider a point, say, $\boldsymbol{x}_3$, on the line segment between $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$. We have $\boldsymbol{x}_3 = \alpha\boldsymbol{x}_1^* + (1-\alpha)\boldsymbol{x}_2^*$ and $0 < \alpha < 1$. It follows that

$$\begin{aligned} \|\boldsymbol{x}_3\|_1 &= \|\alpha\boldsymbol{x}_1^* + (1-\alpha)\boldsymbol{x}_2^*\|_1 \\ &\leq \alpha\|\boldsymbol{x}_1^*\|_1 + (1-\alpha)\|\boldsymbol{x}_2^*\|_1 = \sigma^*, \end{aligned} \quad (9)$$

and

$$
\begin{aligned}
& \|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_3\|_2 \\
= \ & \|\boldsymbol{y}-\boldsymbol{D}[\alpha\boldsymbol{x}_1^*+(1-\alpha)\boldsymbol{x}_2^*]\|_2 \\
= \ & \|\alpha(\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_1^*)+(1-\alpha)(\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_2^*)\|_2 \\
\leq \ & \alpha\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_1^*\|_2+(1-\alpha)\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_2^*\|_2 \leq \ \varepsilon.
\end{aligned} \tag{10}
$$

From the inequality (10), we can see that $\boldsymbol{x}_3$ is a feasible point of problem (2). If the inequality (9) is strict, this would be contrary to the assumption that $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ are two solution of problem (2), which means $\|\boldsymbol{x}_3\|_1=\sigma^*$, that is, each point on the line segment between $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ is a solution of problem (2).

Finally, we prove the third conclusion. Consider an equivalent problem of problem (2)

$$
\min_{\boldsymbol{x}} \quad \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}\|_2^2 \leq \varepsilon^2. \tag{11}
$$

Before proceeding further, we first show that problem (11) satisfies the Slater's condition, that is, let $\mathcal{D}=\{\boldsymbol{x}\in\mathbb{R}^K | \|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}\|_2^2 \leq \varepsilon^2\}$, then there exists a $\boldsymbol{x}\in\operatorname{relint}(\mathcal{D})$ such that $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}\|_2^2 < \varepsilon^2$. Due to the convexity of the set $\mathcal{D}$, we can see from Theorem 2.4 that $\operatorname{relint}(\mathcal{D}) := \{\boldsymbol{x}\in\mathcal{D} | \forall\boldsymbol{x}'\in\mathcal{D}, \exists\rho>1 : \rho\boldsymbol{x}+(1-\rho)\boldsymbol{x}'\in\mathcal{D}\}$. According to the fact that the dictionary $\boldsymbol{D}$ is an overcomplete and row full rank matrix, the linear equations system $\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}$ has infinitely many solutions. Denote one solution by $\boldsymbol{x}$, it is easy to see that $\boldsymbol{x}\in\mathcal{D}$. For every $\boldsymbol{x}'\in\mathcal{D}$, we have

$$
\begin{aligned}
& \|\boldsymbol{y}-\boldsymbol{D}[\rho\boldsymbol{x}+(1-\rho)\boldsymbol{x}']\|_2^2 \\
= \ & \|\rho(\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x})+(1-\rho)(\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}')\|_2^2 \\
= \ & \|(1-\rho)(\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}')\|_2^2 \\
\leq \ & (\rho-1)^2\varepsilon^2.
\end{aligned}
$$

Based on the above analysis, it follows that, for every $1<\rho\leq 2$, we have $\rho\boldsymbol{x}+(1-\rho)\boldsymbol{x}'\in\mathcal{D}$, i.e., each solution of the linear equations system $\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}$ satisfies that $\boldsymbol{x}\in\operatorname{relint}(\mathcal{D})$ and $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}\|_2^2<\varepsilon^2$. In conclusion, problem (11) satisfies the Slater's condition. Applying Theorem 2.5, we obtain that the KKT conditions of problem (11) are necessary and sufficient for its optimality.

Now we turn to prove the third conclusion. Let $\boldsymbol{x}^*$ and $\eta^*$ be the optimal points of the primal problem and the dual problem, respectively. Then $\boldsymbol{x}^*$ and $\eta^*$ satisfy the KKT conditions of problem (11), that is, $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}^*\|_2^2\leq\varepsilon^2$, $\eta^*\geq 0$, and

$$
\eta^*(\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}^*\|_2^2-\varepsilon^2)=0, \tag{12}
$$

$$
\left\{
\begin{array}{rl}
-\eta^*[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^*-\boldsymbol{y})]_i=1, & x_i^*>0; \\
-1\leq\eta^*[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^*-\boldsymbol{y})]_i\leq 1, & x_i^*=0; \\
\eta^*[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^*-\boldsymbol{y})]_i=1, & x_i^*<0.
\end{array}
\right. \tag{13}
$$

If $\varepsilon<\|\boldsymbol{y}\|_2$, we have $\boldsymbol{x}^*\neq\boldsymbol{0}$, then $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}^*\|_2^2=\varepsilon^2$. Otherwise, by $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}^*\|_2^2<\varepsilon^2$ and Eq. (12), we have $\eta^*=0$. From (13), we can obtain $\boldsymbol{x}^*=\boldsymbol{0}$ in the case that $\eta^*=0$, which would be a contradiction to $\boldsymbol{x}^*\neq\boldsymbol{0}$. When problem (2) has more than one solution, we use $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ to denote two different solutions of problem (2). Suppose that $\boldsymbol{D}\boldsymbol{x}_1^*\neq\boldsymbol{D}\boldsymbol{x}_2^*$. By the strictly convexity of the $\ell_2$ ball, there is a $0<\alpha'<1$ such that $\|\boldsymbol{y}-\boldsymbol{D}[\alpha'\boldsymbol{x}_1^*+(1-\alpha')\boldsymbol{x}_2^*]\|_2<\varepsilon$. This would be contrary to the proved second conclusion. Therefore, we have $\boldsymbol{D}\boldsymbol{x}_1^*=\boldsymbol{D}\boldsymbol{x}_2^*$. ■

Finally, the properties of solutions of the $\ell_1$ norm constraint problem (3) are given in the following Lemma 2.3.

**Lemma 2.3 (Properties of solutions of the $\ell_1$ norm constraint problem** (3)**)**

1) *If $\sigma=0$, the zero solution is the only solution to problem* (3)*; If $\sigma\geq\min\{\|\boldsymbol{x}\|_1 | \boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, the solution set of problem* (3) *is $\{\boldsymbol{x}|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}, \|\boldsymbol{x}\|_1\leq\sigma\}$.*
2) *Problem* (3) *has either one solution or infinitely many solutions.*
3) *If $0<\sigma<\min\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, the solutions of problem* (3) *are on the border of the feasible region. When problem* (3) *has infinitely many solutions, every two different solutions, say, $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$, satisfy $\boldsymbol{D}\boldsymbol{x}_1^*=\boldsymbol{D}\boldsymbol{x}_2^*$.*

*Proof:* We first prove the first conclusion. If $\sigma=0$, there is only the zero vector in the feasible region and therefore it is evident that the zero vector is the only solution to problem (3). When $\sigma\geq\min\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, it is also easy to show that the solution set of problem (3) is $\{\boldsymbol{x}|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}, \|\boldsymbol{x}\|_1\leq\sigma\}$.

Now our purpose is to prove the second conclusion. Obviously, the solution set of problem (3) is nonempty. When there are more than one solutions, we denote two of the solutions by $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$. Then we have $\|\boldsymbol{x}_1^*\|_1\leq\sigma$, $\|\boldsymbol{x}_2^*\|_1\leq\sigma$, and $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_1^*\|_2=\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_2^*\|_2$, the value of which is denoted by $\varepsilon^*$. Consider a point $\boldsymbol{x}_3$ on the line segment between $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$, that is, $\boldsymbol{x}_3=\alpha\boldsymbol{x}_1^*+(1-\alpha)\boldsymbol{x}_2^*$ and $0<\alpha<1$. Then

$$
\begin{aligned}
& \|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_3\|_2 \\
= \ & \|\boldsymbol{y}-\boldsymbol{D}[\alpha\boldsymbol{x}_1^*+(1-\alpha)\boldsymbol{x}_2^*]\|_2 \\
= \ & \|\alpha(\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_1^*)+(1-\alpha)(\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_2^*)\|_2 \\
\leq \ & \alpha\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_1^*\|_2+(1-\alpha)\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_2^*\|_2 = \ \varepsilon^*,
\end{aligned} \tag{14}
$$

and

$$
\begin{aligned}
\|\boldsymbol{x}_3\|_1 &= \|\alpha\boldsymbol{x}_1^*+(1-\alpha)\boldsymbol{x}_2^*\|_1 \\
&\leq \alpha\|\boldsymbol{x}_1^*\|_1+(1-\alpha)\|\boldsymbol{x}_2^*\|_1\leq\sigma.
\end{aligned} \tag{15}
$$

From the inequality (15), we can obtain that $\boldsymbol{x}_3$ is a feasible point of problem (3). If the inequality (14) is strict, then there would be a contradiction to the assumption that $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ are two different solutions of problem (3). Therefore, $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}_3\|_2=\varepsilon^*$, that is, each point on the line segment between $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ is a solution to problem (3).

Finally, we prove the third conclusion. We first show that problem (3) satisfies the Slater's condition in the case $\sigma>0$, that is, letting $\mathcal{D}=\{\boldsymbol{x}|\|\boldsymbol{x}\|_1\leq\sigma\}$, there is a $\boldsymbol{x}$ such that $\boldsymbol{x}\in\operatorname{relint}(\mathcal{D})$ and $\|\boldsymbol{x}\|_1<\sigma$. Since the set $\mathcal{D}$ is convex, it follows from Theorem 2.4 that $\operatorname{relint}(\mathcal{D}) := \{\boldsymbol{x}\in\mathcal{D}|\forall\boldsymbol{x}'\in\mathcal{D}, \exists\rho>1 : \rho\boldsymbol{x}+(1-\rho)\boldsymbol{x}'\in\mathcal{D}\}$. Obviously, $\boldsymbol{0}\in\mathcal{D}$. For every $\boldsymbol{x}'\in\mathcal{D}$, we have

$$
\|\rho\boldsymbol{0}+(1-\rho)\boldsymbol{x}'\|_1=\|(1-\rho)\boldsymbol{x}'\|_1\leq(\rho-1)\sigma.
$$

The above reasoning reveals that, for all $1<\rho\leq 2$, we can obtain that $\rho\boldsymbol{0}+(1-\rho)\boldsymbol{x}'\in\mathcal{D}$, i.e., $\boldsymbol{0}\in\operatorname{relint}(\mathcal{D})$ and $\|\boldsymbol{0}\|_1<\sigma$. This means that problem (3) satisfies the Slater's condition. From Theorem 2.5, we have that the KKT conditions of problem (3) are necessary and sufficient for its optimality. Let $\boldsymbol{x}^*$ and $\xi^*$ are respectively the optimal solutions of the primal problem and the dual problem. Then $\boldsymbol{x}^*$ and $\xi^*$ satisfy $\|\boldsymbol{x}^*\|_1 \leq \sigma$, $\xi^* \geq 0$, and

$$\xi^*(\|\boldsymbol{x}^*\|_1 - \sigma) = 0, \tag{16}$$

$$\begin{cases} -[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i = \xi^*, & x_i^* > 0; \\ -\xi^* \leq [2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i \leq \xi^*, & x_i^* = 0; \\ [2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i = \xi^*, & x_i^* < 0. \end{cases} \tag{17}$$

When $0 < \sigma < \min\{\|\boldsymbol{x}\|_1 | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$, we have $\|\boldsymbol{x}^*\|_1 = \sigma$. Otherwise, $\xi^* = 0$ and $\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y}) = \boldsymbol{0}$. Since that the dictionary $\boldsymbol{D}$ is overcomplete and row full rank, then $\boldsymbol{D}\boldsymbol{x}^* = \boldsymbol{y}$, which would be contrary to the assumption that $0 < \sigma < \min\{\|\boldsymbol{x}\|_1 | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$. When there are infinitely many solutions of problem (3), we use $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ to denote two different solutions. Suppose that $\boldsymbol{D}\boldsymbol{x}_1^* \neq \boldsymbol{D}\boldsymbol{x}_2^*$. Due to the strictly convexity of the $\ell_2$ ball, there exists a $0 < \alpha' < 1$ such that $\|\boldsymbol{y} - \boldsymbol{D}[\alpha'\boldsymbol{x}_1^* + (1-\alpha')\boldsymbol{x}_2^*]\|_2 < \varepsilon^*$, which would be a contraction to the proved conclusion. Therefore, we get $\boldsymbol{D}\boldsymbol{x}_1^* = \boldsymbol{D}\boldsymbol{x}_2^*$. ■

## III. Equivalence of the $\ell_1$ Norm Based BSRPs

We have established the properties of solution sets of the three $\ell_1$ norm based BSRPs. Based on the proved results, we further prove the equivalence of the three problems. We first discuss the equivalence of the $\ell_1$ norm regularization problem (1) and the $\ell_1$ norm minimization problem (2).

**Theorem 3.1 (Equivalence of the $\ell_1$ norm regularization problem** (1) **and the $\ell_1$ norm minimization problem** (2)**)** *If $\lambda > 0$ and $\boldsymbol{x}^*$ is a non-zero vector solution of problem* (1)*, then $\boldsymbol{x}^*$ is also a solution of problem* (2) *in the case that $\varepsilon = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2$. Further, if there are infinitely many solutions of problem* (1)*, then they are also solutions of problem* (2) *in the case that $\varepsilon = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2$. Conversely, if $\varepsilon > 0$ and $\boldsymbol{x}^*$ is a non-zero vector solution of problem* (2)*, then $\boldsymbol{x}^*$ is also a solution of problem* (1) *in the case that $\lambda = \|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})\|_\infty$. Further, if there are infinitely many solutions of problem* (2)*, then they are also solutions of problem* (1) *in the case that $\lambda = \|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})\|_\infty$.*

*Proof:* By $\lambda > 0$ and $\boldsymbol{x}^*$ is a solution of problem (1), we can obtain $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2 > 0$ from the optimality conditions (4) of problem (1). Together with that $\boldsymbol{x}^*$ is non-zero, we can also have $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2 < \|\boldsymbol{y}\|_2$. Otherwise, $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{0}\|_2^2 + \lambda\|\boldsymbol{0}\|_1 < \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2^2 + \lambda\|\boldsymbol{x}^*\|_1$, this would lead to a contradiction to the assertion that $\boldsymbol{x}^*$ is a solution of problem (1). According to Lemma 2.2, we can see that, when $0 < \varepsilon = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2 < \|\boldsymbol{y}\|_2$, all the solutions are on the border of the feasible region. Suppose that $\boldsymbol{x}^*$ is not a solution of problem (2) in the case that $\varepsilon = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2$. Then there would be a $\boldsymbol{x}'$ such that $\|\boldsymbol{x}'\|_1 < \|\boldsymbol{x}^*\|_1$ and $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}'\|_2 = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2$. Thus we have $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}'\|_2^2 + \lambda\|\boldsymbol{x}'\|_1 < \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2^2 + \lambda\|\boldsymbol{x}^*\|_1$, which also would lead to a contradiction to the assertion that $\boldsymbol{x}^*$ is a solution of problem (1). When there are infinitely many solutions of problem (1), according to Lemma 2.1, it follows that every two solutions $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ satisfy $\boldsymbol{D}\boldsymbol{x}_1^* = \boldsymbol{D}\boldsymbol{x}_2^*$, that is, they are both solutions of problem (2) in the case that $\varepsilon = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2$. This completes the proof of the first part of the above theorem.

By $\varepsilon > 0$ and $\boldsymbol{x}^*$ is a non-zero vector solution of problem (2), we have $0 < \varepsilon < \|\boldsymbol{y}\|_2$. Let $\eta^*$ be the dual optimal solution of problem (2). We can arrive at $\eta^* > 0$ from (13). Thus we have

$$\begin{aligned} \lambda &= \|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})\|_\infty \\ &= \max\{|[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i| | i = 1, 2, \cdots, K\} = 1/\eta^*. \end{aligned}$$

This means that (13) can be rewritten as

$$\begin{cases} -[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i = \lambda, & x_i^* > 0; \\ -\lambda \leq [2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i \leq \lambda, & x_i^* = 0; \\ [2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i = \lambda, & x_i^* < 0. \end{cases} \tag{18}$$

Therefore, we can get that $\boldsymbol{x}^*$ satisfies the optimality conditions (4) of problem (1). $\boldsymbol{x}^*$ is one solution of problem (1). When $0 < \varepsilon < \|\boldsymbol{y}\|_2$ and problem (2) has infinitely many solutions, we can obtain that each solution $\boldsymbol{x}'$ satisfies $\boldsymbol{D}\boldsymbol{x}' = \boldsymbol{D}\boldsymbol{x}^*$ from the third conclusion of Lemma 2.2. Let $\lambda' = \|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}' - \boldsymbol{y})\|_\infty$. We get $\lambda' = \lambda^*$. The second part of this theorem is thus proved. ■

Now, we discuss the equivalence of the $\ell_1$ norm regularization problem (1) and the $\ell_1$ norm constraint problem (3).

**Theorem 3.2 (Equivalence of the $\ell_1$ norm regularization problem** (1) **and the $\ell_1$ norm constraint problem** (3)**)** *If $\lambda > 0$ and $\boldsymbol{x}^*$ is a non-zero vector solution of problem* (1)*, then $\boldsymbol{x}^*$ is also a solution of problem* (3) *in the case that $\sigma = \|\boldsymbol{x}^*\|_1$. Further, if there are infinitely many solutions of problem* (1)*, then they are also solutions of problem* (3) *in the case that $\sigma = \|\boldsymbol{x}^*\|_1$. Conversely, if $\sigma < \min\{\|\boldsymbol{x}\|_1 | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$ and $\boldsymbol{x}^*$ is a non-zero vector solution of problem* (3)*, then $\boldsymbol{x}^*$ is also a solution of problem* (1) *in the case that $\lambda = \|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})\|_\infty$. Further, if there are infinitely many solutions of problem* (3)*, then they are also solutions of problem* (1) *in the case that $\lambda = \|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})\|_\infty$.*

*Proof:* By $\lambda > 0$ and $\boldsymbol{x}^*$ is a solution of problem (1), then we have $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2 > 0$. Otherwise, according to $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x}^*$ and the optimality conditions (4) of problem (1), we can obtain $\lambda = 0$. Since that $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2 > 0$ and $\boldsymbol{x}^*$ is non-zero, then $0 < \|\boldsymbol{x}^*\|_1 < \min\{\|\boldsymbol{x}\|_1 | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$. Otherwise, let $\boldsymbol{x}' = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \ \text{ s.t. } \ \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}$, we can see that $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}'\|_2^2 + \lambda\|\boldsymbol{x}'\|_1 < \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2^2 + \lambda\|\boldsymbol{x}^*\|_1$, which would be contrary to the assertion that $\boldsymbol{x}^*$ is a non-zero solution of problem (1). Applying Lemma 2.3, it follows that all the solutions of problem (3) are on the border of the feasible region. Suppose that $\boldsymbol{x}^*$ is not a solution of problem (3) in the case that $\sigma = \|\boldsymbol{x}^*\|_1$. There would be a $\boldsymbol{x}'$ such that $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}'\|_2 < \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2$ and $\|\boldsymbol{x}'\|_1 = \|\boldsymbol{x}^*\|_1$. Thus we have $\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}'\|_2^2 + \lambda\|\boldsymbol{x}'\|_1 < \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}^*\|_2^2 + \lambda\|\boldsymbol{x}^*\|_1$, which would be a contradiction to the assertion that $\boldsymbol{x}^*$ is a solution of problem (1). When problem (1) has infinitely many solutions, we can derive from Lemma 2.1 that every two solutions $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ satisfy $\|\boldsymbol{x}_1^*\|_1 = \|\boldsymbol{x}_2^*\|_1$, which means that all these solutions are also solutions of problem (3) in the case that $\sigma = \|\boldsymbol{x}^*\|_1$. The first part of the theorem is proved.

If $\sigma < \min\{\|\boldsymbol{x}\|_1 | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$, and $\boldsymbol{x}^*$ is a non-zero vector solution of problem (3), then $0 < \sigma < \min\{\|\boldsymbol{x}\|_1 | \boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}\}$. Let $\xi^*$ is the dual optimal point of problem (3), it follows that $\xi^* > 0$ from (17). Thus we have

$$\begin{aligned} \lambda &= \|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})\|_\infty \\ &= \max\{|[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^* - \boldsymbol{y})]_i| | i = 1, 2, \cdots, K\} = \xi^*, \end{aligned}$$

Therefore, (17) can be rewritten as

$$\begin{cases} -[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^{*}-\boldsymbol{y})]_i=\lambda, & x_i^{*}>0; \\ -\lambda\leq[2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^{*}-\boldsymbol{y})]_i\leq\lambda, & x_i^{*}=0; \\ [2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^{*}-\boldsymbol{y})]_i=\lambda, & x_i^{*}<0. \end{cases} \tag{19}$$

This is to say that $\boldsymbol{x}^*$ satisfies the optimality conditions (4) of problem (1), which means that $\boldsymbol{x}^*$ is a optimal point of problem (1). When $0<\sigma<\min\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, and problem (3) has infinitely many solutions, from the third conclusion of Lemma 2.3, we can get that every solution $\boldsymbol{x}'$ can satisfies $\boldsymbol{D}\boldsymbol{x}'=\boldsymbol{D}\boldsymbol{x}^*$. Let $\lambda'=\|2\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}'-\boldsymbol{y})\|_\infty$. We have $\lambda'=\lambda^*$. The proof of the second part of the theorem is completed. ∎

According to Theorem 3.1, we can see that, for every $\lambda\in\{\lambda|0<\lambda<\|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{x}\|_\infty\}$, there is necessarily a unique $\varepsilon\in\{\varepsilon|0<\varepsilon<\|\boldsymbol{y}\|_2\}$ such that the corresponding problem (1) and problem (2) have the same solution set, and vice versa. This leads to the equivalence of the two problems. We also can obtain similar conclusion on the equivalence of problem (1) and problem (3) from Theorem 3.2. That is, for every $\lambda\in\{\lambda|0<\lambda<\|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{x}\|_\infty\}$, there is necessarily a unique $\sigma\in\{\sigma|0<\sigma<\min\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}\}$ such that the corresponding problem (1) and problem (3) have the same solution set, and vice versa. Theorems 3.1 and 3.2 imply that, given certain parameter ranges, in the sense that the corresponding problems have the same solution set, there is a one-to-one correspondence between the two parameter values. The equivalence of problems (2) and (3) follows immediately from Theorems 3.1 and 3.2.

## IV. Conclusion

In this paper, we have discussed the equivalence of the three $\ell_1$ norm based BSRPs, the $\ell_1$ norm regularization problem (1), the $\ell_1$ norm minimization problem (2) and the $\ell_1$ norm constraint problem (3). In particular, we have given a rigorous proof of the equivalence of the three problems in the case when the dictionary is an overcomplete and row full rank matrix. The results obtained in the paper are summarized as follows.

- If $0<\lambda<\|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{y}\|_\infty$, $0<\varepsilon<\|\boldsymbol{y}\|_2$ and $0<\sigma<\min\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, then the three $\ell_1$ based BSRPs are equivalent. Further, in the sense of equivalence, there is one-to-one correspondence between the parameters $\lambda$, $\varepsilon$ and $\sigma$.
- If $\lambda\geq\|2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{y}\|_\infty$, $\varepsilon\geq\|\boldsymbol{y}\|_2$ and $\sigma=0$, all the three problems have only the zero solution.
- If $\lambda=0$ and $\sigma\geq\max\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, both the $\ell_1$ norm regularization problem (1) and the $\ell_1$ norm constraint problem (3) are reduced to the least squares problem.
- If $\varepsilon=0$ and $\sigma=\min\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, both the $\ell_1$ norm minimization problem (2) and the $\ell_1$ norm constraint problem (3) are reduced to the following problem
$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}.$$
- If $\min\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}<\sigma<\max\{\|\boldsymbol{x}\|_1|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$, then the solution set of the $\ell_1$ norm constraint problem (3) is a subset of $\{\boldsymbol{x}\in\mathbb{R}^K|\boldsymbol{D}\boldsymbol{x}=\boldsymbol{y}\}$.


## Acknowledgements

This work was partly supported by the NNSF of China under grants 61210009, 61033011, 61379093 and 11131006.


## References


[1] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Science & Business Media, 2010.

[2] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Foundations and Trends in Computer Graphics and Vision*, vol. 8, no. 2-3, pp. 85–283, 2014.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[5] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," in *Advances in Neural Information Processing Systems (NIPS)*, 2006, pp. 801–808.

[6] W. Fu, "Penalized regressions: the bridge versus the lasso," *Journal of Computational and Graphical Statistics*, vol. 7, no. 3, pp. 397–416, 1998.

[7] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[9] S. Canu and Y. Grandvalet, "Outcomes of the equivalence of adaptive ridge with least absolute shrinkage," in *Advances in Neural Information Processing Systems*, 1999, p. 445.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[11] S. Shalev-Shwartz and T. Zhang, "Stochastic dual coordinate ascent methods for regularized loss," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 567–599, 2013.

[12] Y. Zhang and L. Xiao, "Stochastic primal-dual coordinate method for regularized empirical risk minimization," *arXiv preprint arXiv:1409.3257*, 2014.

[13] C. Zhang, "Equivalence of several l1 sparsity problem," 2013. [Online]. Available: http://freemind.pluskid.org/machine-learning/equivalence-of-several-l1-sparsity-problem/

[14] R. Rockafellar, *Convex analysis*. Princeton University Press, 1970.

[15] C. Zalinescu, *Convex analysis in general vector spaces*. Singapore: World Scientific, 2002.

[16] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.