Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2013, Article ID 250943, 8 pages
http://dx.doi.org/10.1155/2013/250943

Research Article

Properties and Iterative Methods for the *Q*-Lasso

Maryam A. Alghamdi,[1] Mohammad Ali Alghamdi,[2] Naseer Shahzad,[2] and Hong-Kun Xu[2,3]

[1] *Department of Mathematics, Sciences Faculty for Girls, King Abdulaziz University, P.O. Box 4087, Jeddah 21491, Saudi Arabia*
[2] *Department of Mathematics, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia*
[3] *Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung 80424, Taiwan*

Correspondence should be addressed to Hong-Kun Xu; xuhk@math.nsysu.edu.tw

Received 24 September 2013; Accepted 27 November 2013

Academic Editor: Chi-Keung Ng

We introduce the Q-lasso which generalizes the well-known lasso of Tibshirani (1996) with Q a closed convex subset of a Euclidean m-space for some integer $m \geq 1$. This set Q can be interpreted as the set of errors within given tolerance level when linear measurements are taken to recover a signal/image via the lasso. Solutions of the Q-lasso depend on a tuning parameter γ. In this paper, we obtain basic properties of the solutions as a function of γ. Because of ill posedness, we also apply l_1-l_2 regularization to the Q-lasso. In addition, we discuss iterative methods for solving the Q-lasso which include the proximal-gradient algorithm and the projection-gradient algorithm.

1. Introduction

The lasso of Tibshirani [1] is the minimization problem:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2 + \gamma\|x\|_1, \tag{1}$$

where A is an $m \times n$ (real) matrix, $b \in \mathbb{R}^m$, and $\gamma > 0$ is a tuning parameter. It is equivalent to the basis pursuit (BP) of Chen et al. [2]:

$$\min_{x\in\mathbb{R}^n} \|x\|_1 \quad \text{subject to } Ax = b. \tag{2}$$

It is well known that both lasso and BP model a number of applied problems arising from machine learning, signal/image processing, and statistics, due to the fact that they promote the sparsity of a signal $x \in \mathbb{R}^n$. Sparsity is popular phenomenon that occurs in practical problems since a solution may have a sparse representation in terms of an appropriate basis and therefore has been paid much attention.

Observe that both the lasso (1) and BP (2) can be viewed as the ℓ_1 regularization applied to the inverse linear system in $\mathbb{R}^n$:

$$Ax = b. \tag{3}$$

In sparse recovery, the system (3) is underdetermined (i.e., $m < n$ and often $m \ll n$ indeed). The theory of compressed sensing of Donoho [3] and Candès et al. [4, 5] makes a breakthrough that under certain conditions the underdetermined system (3) can determine a unique k-sparse solution. (Recall that a signal $x \in \mathbb{R}^n$ is said to be k-sparse if the number of nonzero entries of x is no bigger than k.)

However, due to errors of measurements, the system (3) is actually inexact: $Ax \approx b$. It turns out that the BP (2) is reformulated as

$$\min_{x\in\mathbb{R}^n} \|x\|_1 \quad \text{subject to } \|Ax-b\| \leq \varepsilon, \tag{4}$$

where $\varepsilon > 0$ is the tolerance level of errors and $\|\cdot\|$ is a norm on $\mathbb{R}^n$ (often it is the ℓ_p norm $\|\cdot\|_p$ for $p = 1, 2, \infty$; a solution to (4) when the tolerance is measured by the ℓ_∞ norm $\|\cdot\|_\infty$ is known as the Dantzig selector by Candès and Tao [6]; see also [7]).

Note that if we let $Q := B_\varepsilon(b)$ be the closed ball in $\mathbb{R}^m$ around b and with radius of ε, then (4) is rewritten as

$$\min_{x\in\mathbb{R}^n} \|x\|_1 \quad \text{subject to } Ax \in Q. \tag{5}$$

Let now Q be a nonempty closed convex subset of $\mathbb{R}^m$ and let P_Q be the projection from $\mathbb{R}^m$ onto Q. Then noticing the condition $Ax \in Q$ being equivalent to the condition $Ax - P_Q(Ax) = 0$, we see that the problem (5) is solved via

$$\min_{x\in\mathbb{R}^n}\|x\|_1 \quad \text{subject to } (I - P_Q)\, Ax = 0. \tag{6}$$

Applying the Lagrangian method, we arrive at the following equivalent minimization:

$$\min_{x\in\mathbb{R}^n}\frac{1}{2}\left\|(I - P_Q)\, Ax\right\|_2^2 + \gamma\|x\|_1, \tag{7}$$

where $\gamma > 0$ is a Lagrangian multiplier (also interpreted as a regularization parameter).

Alternatively, we may view (7) as the ℓ_1 regularization of the inclusion

$$Ax \in Q \quad (\text{equivalently, the equation } (I - P_Q)\, Ax = 0) \tag{8}$$

which extends the linear system (3) in an obvious way. We refer to the problem (7) as the Q-lasso since it is the ℓ_1 regularization of inclusion (8) as lasso (1) is the ℓ_1 regularization of the linear system (3). Throughout the rest of this paper, we always assume that (8) is consistent (i.e., solvable).

Q-lasso (7) is also connected with the so-called split feasibility problem (SFP) of Censor and Elfving [8] (see also [9]) which is stated as finding a point x with the property

$$x \in C, \qquad Ax \in Q, \tag{9}$$

where C and Q are closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. An equivalent minimization formulation of the SFP (9) is given as

$$\min_{x\in C}\frac{1}{2}\left\|Ax - P_Q Ax\right\|_2^2. \tag{10}$$

Its ℓ_1 regularization is given as the minimization

$$\min_{x\in C}\frac{1}{2}\left\|Ax - P_Q Ax\right\|_2^2 + \gamma\|x\|_1, \tag{11}$$

where $\gamma > 0$ is a regularization parameter. Problem (7) is a special case of (11) when the set of constraints, C, is taken to be the entire space $\mathbb{R}^n$.

The purpose of this paper is to study the behavior, in terms of $\gamma > 0$, of solutions to the regularized problem (7). (We leave the more general problem (11) to further work, due to the fact that the involvement of another closed convex set C brings some technical difficulties which are not easy to overcome.) We discuss iterative methods for solving the Q-lasso, including the proximal-gradient method and the projection-gradient method, the latter being derived via a duality technique. Due to ill posedness, we also apply the ℓ_1-ℓ_2 regularization to the Q-lasso.

2. Preliminaries

Let $n \geq 1$ be an integer and let $\mathbb{R}^n$ be the Euclidean n-space. If $p \geq 1$, we use $\|\cdot\|_p$ to denote the ℓ_p norm on $\mathbb{R}^n$. Namely, for $x = (x_j)^t \in \mathbb{R}^n$,

$$\begin{gathered}\|x\|_p = \left(\sum_{j=1}^{n}\left|x_j\right|^p\right)^{1/p} \quad (1 \leq p < \infty),\\ \|x\|_\infty = \max_{1\leq j\leq n}\left|x_j\right|.\end{gathered} \tag{12}$$

Let K be a closed convex subset of $\mathbb{R}^n$. Recall that the projection from $\mathbb{R}^n$ to K is defined as the operator

$$P_K(x) = \arg\min_{u\in K}\|x - u\|_2^2, \quad x \in \mathbb{R}^n. \tag{13}$$

The projection P_K is characterized as follows:

$$\text{given } x \in \mathbb{R}^n \text{ and } z \in K : z = P_K x \Longleftrightarrow \langle x - z, y - z\rangle \leq 0, \quad y \in K. \tag{14}$$

Projections are nonexpansive. Namely, we have the following.

Proposition 1. *One has that P_K is firmly nonexpansive in the sense that*

$$\langle x - y, P_K x - P_K y\rangle \geq \left\|P_K x - P_K y\right\|^2, \quad x, y \in \mathbb{R}^n. \tag{15}$$

In particular, P_K is nonexpansive; that is, $\|P_K x - P_K y\| \leq \|x - y\|$ for all $x, y \in \mathbb{R}^n$.

Recall that function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda) f(x) + \lambda f(y) \tag{16}$$

for all $\lambda \in (0,1)$ and $x, y \in \mathbb{R}^n$. (Note that we only consider finite-valued functions.)

The subdifferential of a convex function f is defined as the operator ∂f given by

$$\partial f(x) = \{\xi \in \mathbb{R}^n : f(y) \geq f(x) + \langle \xi, y - x\rangle, y \in \mathbb{R}^n\}. \tag{17}$$

The inequality in (17) is referred to as the subdifferential inequality of f at x. We say that f is subdifferentiable at x if $\partial f(x)$ is nonempty. It is well known that, for an everywhere finite-valued convex function f on $\mathbb{R}^n$, f is everywhere subdifferentiable.

Examples. (i) If $f(x) = |x|$ for $x \in \mathbb{R}$, then $\partial f(0) = [-1,1]$; (ii) of $f(x) = \|x\|_1$ for $x \in \mathbb{R}^n$, then $\partial f(x)$ is given componentwise by

$$(\partial f(x))_j = \begin{cases} \operatorname{sgn}(x_j), & \text{if } x_j \neq 0, \\ \xi_j \in [-1,1], & \text{if } x_j = 0, \end{cases} \quad 1 \leq j \leq n. \tag{18}$$

Here sgn is the sign function; that is, for $a \in \mathbb{R}$,

$$\operatorname{sgn}(a) = \begin{cases} 1, & \text{if } a > 0, \\ 0, & \text{if } a = 0, \\ -1, & \text{if } a < 0. \end{cases} \tag{19}$$

Consider the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} f(x). \tag{20}$$

The following are well known.

Proposition 2. *Let f be everywhere finite-valued on $\mathbb{R}^n$.*

(i) *If f is strictly convex, then* (20) *admits at most one solution.*

(ii) *If f is convex and satisfies the coercivity condition*

$$\|x\| \longrightarrow \infty \Longrightarrow f(x) \longrightarrow \infty, \tag{21}$$

then there exists at least one solution to (20)*. Therefore, if f is both strictly convex and coercive, there exists one and only one solution to* (20).

Proposition 3. *Let f be everywhere finite-valued convex on $\mathbb{R}^n$ and $z \in \mathbb{R}^n$. Suppose f is bounded below (i.e., $\inf\{f(x) : x \in \mathbb{R}^n\} > -\infty$). Then z is a solution to minimization* (20) *if and only if it satisfies the first-order optimality condition:*

$$0 \in \partial f(z). \tag{22}$$

3. Properties of the Q-Lasso

We study some basic properties of the Q-lasso which is repeated below

$$\min_{x \in \mathbb{R}^n} \varphi_\gamma(x) := \frac{1}{2}\left\|Ax - P_Q Ax\right\|_2^2 + \gamma\|x\|_1, \tag{23}$$

where $\gamma > 0$ is a regularization parameter. We also consider the following minimization (we call it Q-least squares problem):

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\left\|Ax - P_Q Ax\right\|_2^2. \tag{24}$$

Denote by S and S_γ the solution sets of (24) and (23), respectively. Since φ_γ is continuous, convex, and coercive (i.e., $\varphi_\gamma(x) \to \infty$ as $\|x\|_2 \to \infty$), S_γ is closed, convex, and nonempty. Notice also that since we assume the consistency of (8), we have $S \neq \emptyset$; moreover, the solution sets of (8) and (24) coincide.

Observe that the assumption that $S \neq \emptyset$ actually implies that S_γ is uniformly bounded in $\gamma > 0$, as shown by the lemma below.

Lemma 4. *Assume that* (24) *is consistent (i.e., $S \neq \emptyset$). Then, for $\gamma > 0$ and $x_\gamma \in S_\gamma$, one has $\|x_\gamma\|_1 \le \inf_{x \in S}\|x\|_1$.*

Proof. Let $x_\gamma \in S_\gamma$. In the relation

$$\begin{aligned} &\frac{1}{2}\left\|\left(I - P_Q\right) Ax_\gamma\right\|_2^2 + \gamma\left\|x_\gamma\right\|_1 \\ &\quad \le \frac{1}{2}\left\|\left(I - P_Q\right) Ax\right\|_2^2 + \gamma\|x\|_1, \quad x \in \mathbb{R}^n, \end{aligned} \tag{25}$$

taking $x \in S$ yields (for $P_Q x \in Q$)

$$\frac{1}{2}\left\|\left(I - P_Q\right) Ax_\gamma\right\|_2^2 + \gamma\left\|x_\gamma\right\|_1 \le \gamma\|x\|_1, \quad x \in S. \tag{26}$$

It follows that

$$\left\|x_\gamma\right\|_1 \le \|x\|_1, \quad x \in S. \tag{27}$$

This proves the conclusion of the lemma. □

Proposition 5. *One has the following.*

(i) *The functions*

$$\rho(\gamma) := \left\|x_\gamma\right\|_1, \qquad \eta(\gamma) := \frac{1}{2}\left\|(I - P_Q)Ax_\gamma\right\|_2^2 \tag{28}$$

are well defined for $\gamma > 0$. That is, they do not depend upon particular choice of $x_\gamma \in S_\gamma$.

(ii) *The function $\rho(\gamma)$ is decreasing in $\gamma > 0$.*

(iii) *The function $\eta(\gamma)$ is increasing in $\gamma > 0$.*

(iv) *$(I - P_Q)Ax_\gamma$ is continuous in $\gamma > 0$.*

Proof. For $x_\gamma \in S_\gamma$, we have the optimality condition:

$$0 \in \partial\varphi_\gamma\left(x_\gamma\right) = A^t\left(I - P_Q\right) Ax_\gamma + \gamma\partial\left\|x_\gamma\right\|_1. \tag{29}$$

Here A^t is the transpose of A and ∂ stands for the subdifferential in the sense of convex analysis. Equivalently,

$$-\frac{1}{\gamma}A^t\left(I - P_Q\right) Ax_\gamma \in \partial\left\|x_\gamma\right\|_1. \tag{30}$$

It follows by the subdifferential inequality that

$$\begin{aligned} \gamma\|x\|_1 \ge \gamma\left\|x_\gamma\right\|_1 - \left\langle A^t\left(I - P_Q\right) Ax_\gamma, x - x_\gamma\right\rangle, \\ \forall x \in \mathbb{R}^n. \end{aligned} \tag{31}$$

In particular, for $\hat{x}_\gamma \in S_\gamma$,

$$\gamma\left\|\hat{x}_\gamma\right\|_1 \ge \gamma\left\|x_\gamma\right\|_1 - \left\langle A^t\left(I - P_Q\right) Ax_\gamma, \hat{x}_\gamma - x_\gamma\right\rangle. \tag{32}$$

Interchange x_γ and $\hat{x}_\gamma$ to get

$$\gamma\left\|x_\gamma\right\|_1 \ge \gamma\left\|\hat{x}_\gamma\right\|_1 - \left\langle A^t\left(I - P_Q\right) A\hat{x}_\gamma, x_\gamma - \hat{x}_\gamma\right\rangle. \tag{33}$$

Adding up (32) and (33) yields

$$0 \ge \left\langle A\hat{x}_\gamma - Ax_\gamma, A\hat{x}_\gamma - Ax_\gamma\right\rangle = \left\|A\hat{x}_\gamma - Ax_\gamma\right\|_2^2. \tag{34}$$

Consequently, $A\widehat{x}_\gamma = Ax_\gamma$. Moreover, (32) and (33) imply that $\|\widehat{x}_\gamma\|_1 \geq \|x_\gamma\|_1$ and $\|x_\gamma\|_1 \geq \|\widehat{x}_\gamma\|_1$, respectively. Hence $\|\widehat{x}_\gamma\|_1 = \|x_\gamma\|_1$, and it follows that the functions

$$\rho(\gamma) := \|x_\gamma\|_1, \qquad \eta(\gamma) := \frac{1}{2}\left\|(I - P_Q) Ax_\gamma\right\|_2^2 \qquad (x_\gamma \in S_\gamma) \tag{35}$$

are well defined for $\gamma > 0$.

Now substituting $x_\beta \in S_\beta$ for x in (31), we get

$$\gamma\|x_\beta\|_1 \geq \gamma\|x_\gamma\|_1 - \left\langle A^t (I - P_Q) Ax_\gamma, x_\beta - x_\gamma\right\rangle. \tag{36}$$

Interchange γ and β and x_γ and x_β to find

$$\beta\|x_\gamma\|_1 \geq \beta\|x_\beta\|_1 - \left\langle A^t (I - P_Q) Ax_\beta, x_\gamma - x_\beta\right\rangle. \tag{37}$$

Adding up (36) and (37) and using the fact that $(I - P_Q)$ is firmly nonexpansive, we deduce that

$$\begin{aligned}(\gamma - \beta)&\left(\|x_\beta\|_1 - \|x_\gamma\|_1\right)\\ &\geq \left\langle (I - P_Q) Ax_\gamma - (I - P_Q) Ax_\beta, Ax_\gamma - Ax_\gamma\right\rangle\\ &\geq \left\|(I - P_Q) Ax_\gamma - (I - P_Q) Ax_\beta\right\|_2^2.\end{aligned} \tag{38}$$

We therefore find that if $\gamma > \beta$, then $\|x_\beta\|_1 \geq \|x_\gamma\|_1$. This proves that $\rho(\gamma)$ is nonincreasing in $\gamma > 0$. From (38) it also follows that $(I-P_Q)Ax_\gamma$ is continuous for $\gamma > 0$, which implies the continuity of $\eta(\gamma)$ for $\gamma > 0$.

To see that $\eta(\gamma)$ is increasing, we use the inequality (as $x_\gamma \in S_\gamma$)

$$\begin{aligned}&\frac{1}{2}\left\|(I - P_Q)Ax_\gamma\right\|_2^2 + \gamma\|x_\gamma\|_1\\ &\quad\leq \frac{1}{2}\left\|(I - P_Q)Ax_\beta\right\|_2^2 + \gamma\|x_\beta\|_1\end{aligned} \tag{39}$$

which implies that

$$\eta(\gamma) \leq \eta(\beta) + \gamma\left(\|x_\beta\|_1 - \|x_\gamma\|_1\right). \tag{40}$$

Now if $\beta > \gamma > 0$, then, as $\|x_\beta\|_1 \leq \|x_\gamma\|_1$, we immediately get that $\eta(\gamma) \leq \eta(\beta)$ and the increase of η is proven. □

Proposition 6. *One has the following.*

(i) $\lim_{\gamma\to 0}\eta(\gamma) = \inf_{x\in\mathbb{R}^n}(1/2)\|(I - P_Q)Ax\|_2^2$.

(ii) $\lim_{\gamma\to 0}\rho(\gamma) = \min_{x\in S}\|x\|_1$.

Proof. (i) Taking the limit as $\gamma \to 0$ in the inequality (and using the boundedness of (x_γ))

$$\begin{aligned}&\frac{1}{2}\left\|(I - P_Q) Ax_\gamma\right\|_2^2 + \gamma\|x_\gamma\|_1\\ &\quad\leq \frac{1}{2}\|(I - P_Q) Ax\|_2^2 + \gamma\|x\|_1, \quad \forall x \in \mathbb{R}^n,\end{aligned} \tag{41}$$

yields

$$\lim_{\gamma\to 0}\eta(\gamma) \leq \frac{1}{2}\|(I - P_Q)Ax\|_2^2, \quad \forall x \in \mathbb{R}^n. \tag{42}$$

The result in (i) then follows.

As for (ii), we have, by (27), $\|x_\gamma\|_1 \leq \|\widetilde{x}\|_1$ for any $\widetilde{x} \in S$. In particular, $\|x_\gamma\|_1 \leq \|x^\dagger\|_1$, where $x^\dagger$ is an ℓ_1 minimum-norm element of S; that is, $\|x^\dagger\|_1 = \min_{x\in S}\|x\|_1$.

Assume $\gamma_k \to 0$ is such that $x_{\gamma_k} \to \widehat{x}$. Then for any x,

$$\begin{aligned}\frac{1}{2}\|(I - P_Q) A\widehat{x}\|_2^2 &= \lim_{k\to\infty}\frac{1}{2}\left\|(I - P_Q) Ax_{\gamma_k}\right\|_2^2\\ &= \lim_{k\to\infty}\frac{1}{2}\left\|(I - P_Q) Ax_{\gamma_k}\right\|_2^2 + \gamma_k\left\|x_{\gamma_k}\right\|_1\\ &\leq \lim_{k\to\infty}\frac{1}{2}\|(I - P_Q) Ax\|_2^2 + \gamma_k\|x\|_1\\ &= \frac{1}{2}\|(I - P_Q)Ax\|_2^2.\end{aligned} \tag{43}$$

It follows that $\widehat{x}$ solves the minimization problem: $\min_x(1/2)\|(I - P_Q)Ax\|_2^2$; that is, $\widehat{x} \in S$. Consequently,

$$\begin{aligned}\lim_{\gamma\to 0}\rho(\gamma) &= \lim_{k\to\infty}\rho(\gamma_k) = \lim_{k\to\infty}\left\|x_{\gamma_k}\right\|_1\\ &= \|\widehat{x}\|_1 \leq \left\|x^\dagger\right\|_1 = \min_{x\in S}\|x\|_1.\end{aligned} \tag{44}$$

This suffices to ensure that the conclusion of (ii) holds. □

It is a challenging problem how to select the tuning (i.e., regularizing) parameter γ in lasso (1) and Q-lasso (7). There is no general rule to universally select γ which should instead be selected in a case-to-case manner. The following result however points out that γ cannot be large.

Proposition 7. *Let Q be a nonempty closed convex subset of $\mathbb{R}^m$ and assume that Q-lasso (7) is consistent (i.e., solvable). If $\gamma > \max\{\|A^tP_QAx\|_\infty : \|x\|_1 \leq \min_{v\in S}\|v\|_1\}$ (note that this condition is reduced to $\gamma > \|A^tb\|_\infty$ for lasso (1) for which $Q = \{b\}$), then $x_\gamma = 0$. (Here S is, as before, the solution set of the Q-least squares problem* (24)*.)*

Proof. Let $x_\gamma \in S_\gamma$. The optimality condition

$$-A^t (I - P_Q) Ax_\gamma \in \gamma\partial\|x_\gamma\|_1 \tag{45}$$

implies that

$$\begin{aligned}-\left(A^t (I - P_Q) Ax_\gamma\right)_j &= \gamma\cdot\operatorname{sgn}\left[(x_\gamma)_j\right], \quad \text{if } (x_\gamma)_j \neq 0,\\ \left|\left(A^t (I - P_Q) Ax_\gamma\right)_j\right| &\leq \gamma, \quad \text{if } (x_\gamma)_j = 0.\end{aligned} \tag{46}$$

Taking $x = 2x_\gamma$ in subdifferential inequality (31) yields

$$\begin{aligned}\gamma\left\|x_\gamma\right\|_1 &\geq -\left\langle A^t(I-P_Q)Ax_\gamma, x_\gamma\right\rangle \\ &= -\sum_{(x_\gamma)_j \neq 0}\left(A^t(I-P_Q)Ax_\gamma\right)_j\left(x_\gamma\right)_j \\ &= \sum_{(x_\gamma)_j \neq 0}\gamma\cdot\left[\operatorname{sgn}\left(x_\gamma\right)\right]_j\left(x_\gamma\right)_j \\ &= \gamma\sum_{(x_\gamma)_j \neq 0}\left|\left(x_\gamma\right)_j\right| = \gamma\|x\|_1.\end{aligned} \tag{47}$$

It follows that

$$\begin{aligned}\gamma\left\|x_\gamma\right\|_1 &= -\left\langle A^t(I-P_Q)Ax_\gamma, x_\gamma\right\rangle \\ &= -\left\langle (I-P_Q)Ax_\gamma, Ax_\gamma\right\rangle,\end{aligned} \tag{48}$$

$$\begin{aligned} &= -\left\|Ax_\gamma\right\|_2^2 + \left\langle P_QAx_\gamma, Ax_\gamma\right\rangle \\ &\leq \left\langle P_QAx_\gamma, Ax_\gamma\right\rangle = \left\langle A^tP_QAx_\gamma, x_\gamma\right\rangle \\ &\leq \left\|A^tP_QAx_\gamma\right\|_\infty\left\|x_\gamma\right\|_1.\end{aligned} \tag{49}$$

Now by Lemma 4, we have $\|x_\gamma\|_1 \leq \min_{v\in S}\|v\|_1$. Hence, from (49) it follows that if $x_\gamma \neq 0$, we must have $\gamma \leq \max\{\|A^tP_QAx\|_\infty : \|x\|_1 \leq \min_{v\in S}\|v\|_1\}$. This completes the proof. □

Notice that (48) shows that $\rho(\lambda) = \|x_\gamma\|_1$ can be determined by Ax_γ. Hence, we arrive at the following characterization of solutions of Q-lasso (23).

Proposition 8. *Let Q be a nonempty closed convex subset of $\mathbb{R}^m$ and let $\gamma > 0$ and $x_\gamma \in S_\gamma$. Then $\hat{x} \in \mathbb{R}^n$ is a solution of the Q-lasso* (23) *if and only if $A\hat{x} = Ax_\gamma$ and $\|\hat{x}\| \leq \|x_\gamma\|$. It turns out that*

$$S_\gamma = x_\gamma + N(A) \cap B_{\rho(\gamma)}, \tag{50}$$

where $N(A) = \{x \in \mathbb{R}^n : Ax = 0\}$ is the null space of A and where B_r denotes the closed ball centered at the origin and with radius of $r > 0$. This shows that if one can find one solution to Q-lasso (23)*, then all solutions are found by* (50).

Proof. If $A\hat{x} = Ax_\gamma$, then from the relations

$$\begin{aligned}\varphi_\gamma\left(x_\gamma\right) &= \frac{1}{2}\left\|(I-P_Q)Ax_\gamma\right\|_2^2 + \gamma\left\|x_\gamma\right\|_1 \\ &\leq \frac{1}{2}\left\|(I-P_Q)A\hat{x}\right\|_2^2 + \gamma\|\hat{x}\|_1 \\ &= \frac{1}{2}\left\|(I-P_Q)Ax_\gamma\right\|_2^2 + \gamma\|\hat{x}\|_1,\end{aligned} \tag{51}$$

we obtain $\|x_\gamma\|_1 \leq \|\hat{x}\|_1$. This together with the assumption of $\|\hat{x}\|_1 \leq \|x_\gamma\|_1$ yields that $\|\hat{x}\|_1 = \|x_\gamma\|_1$ which in turn implies that $\varphi_\gamma(\hat{x}) = \varphi_\gamma(x_\gamma)$ and hence $\hat{x} \in S_\gamma$. □

4. Iterative Methods

In this section we discuss the proximal iterative methods for solving Q-lasso (7). The basics are Moreau's concept of proximal operators and their fundamental properties which are briefly mentioned below. (For the sake of our purpose, we however confine ourselves to the finite-dimensional setting.)

4.1. Proximal Operators. Let $\Gamma_0(\mathbb{R}^n)$ be the space of convex functions in $\mathbb{R}^n$ that are proper, lower semicontinuous and convex.

Definition 9 (see [10, 11]). The proximal operator of $\varphi \in \Gamma_0(\mathbb{R}^n)$ is defined by

$$\operatorname{prox}_\varphi(x) := \arg\min_{v\in\mathbb{R}^n}\left\{\varphi(v) + \frac{1}{2}\|v-x\|^2\right\}, \quad x \in \mathbb{R}^n. \tag{52}$$

The proximal operator of φ of order $\lambda > 0$ is defined as the proximal operator of $\lambda\varphi$; that is,

$$\operatorname{prox}_{\lambda\varphi}(x) := \arg\min_{v\in\mathbb{R}^n}\left\{\varphi(v) + \frac{1}{2\lambda}\|v-x\|^2\right\}, \quad x \in \mathbb{R}^n. \tag{53}$$

For fundamental properties of proximal operators, the reader is referred to [12, 13] for details. Here we only mention the fact that the proximal operator $\operatorname{prox}_{\lambda\varphi}$ can have a closed-form expression in some important cases as shown in the examples below [12].

(a) If we take φ to be any norm $\|\cdot\|$ of $\mathbb{R}^n$, then

$$\operatorname{prox}_{\lambda\|\cdot\|}(x) = \begin{cases}\left(1 - \dfrac{\lambda}{\|x\|}\right)x, & \text{if } \|x\| > \lambda. \\ 0, & \text{if } \|x\| \leq \lambda.\end{cases} \tag{54}$$

In particular, if we take φ to be the absolute value function of the real line $\mathbb{R}$, we get

$$\operatorname{prox}_{\lambda|\cdot|}(x) = \operatorname{sgn}(x)\max\{|x| - \lambda, 0\} \tag{55}$$

which is also known as the scalar soft-thresholding operator.

(b) Let $\{e_k\}_{k=1}^n$ be an orthonormal basis of $\mathbb{R}^n$ and let $\{\omega_k\}_{k=1}^n$ be real positive numbers. Define φ by

$$\varphi(x) = \sum_{k=1}^n \omega_k\left|\left\langle x, e_k\right\rangle\right|. \tag{56}$$

Then $\operatorname{prox}_\varphi(x) = \sum_{k=1}^n \alpha_k e_k$, where

$$\alpha_k = \operatorname{sgn}\left(\left\langle x, e_k\right\rangle\right)\max\left\{\left|\left\langle x, e_k\right\rangle\right| - \omega_k, 0\right\}. \tag{57}$$

In particular, if $\varphi(x) = \|x\|_1$ for $x \in \mathbb{R}^n$, then

$$\begin{aligned}\operatorname{prox}_{\lambda\|\cdot\|_1}(x) &= \left(\operatorname{prox}_{\lambda|\cdot|}(x_1), \ldots, \operatorname{prox}_{\lambda|\cdot|}(x_n)\right) \\ &= \left(\alpha_1, \ldots, \alpha_n\right),\end{aligned} \tag{58}$$

where $\alpha_k = \operatorname{sgn}(x_k)\max\{|x_k| - \lambda, 0\}$ for $k = 1, \ldots, n$.

4.2. Proximal-Gradient Algorithm. The proximal operators can be used to minimize the sum of two convex functions:

$$\min_{x\in\mathbb{R}^n} f(x) + g(x), \tag{59}$$

where $f, g \in \Gamma_0(\mathbb{R}^n)$. It is often the case where one of them is differentiable. The following is an equivalent fixed point formulation of (59).

Proposition 10 (see [12, 14]). *Let $f, g \in \Gamma_0(\mathbb{R}^n)$. Let $x^* \in \mathbb{R}^n$ and $\lambda > 0$. Assume f is finite valued and differentiable on $\mathbb{R}^n$. Then x^* is a solution to* (59) *if and only if x^* solves the fixed point equation:*

$$x^* = \left(\operatorname{prox}_{\lambda g} \circ (I - \lambda \nabla f)\right) x^*. \tag{60}$$

Fixed point equation (60) immediately yields the following fixed point algorithm which is also known as the proximal-gradient algorithm for solving (59).

Initialize $x_0 \in \mathbb{R}^n$ and iterate

$$x_{k+1} = \left(\operatorname{prox}_{\lambda_k g} \circ (I - \lambda_k \nabla f)\right) x_k, \tag{61}$$

where $\{\lambda_k\}$ is a sequence of positive real numbers.

Theorem 11 (see [12, 14]). *Let $f, g \in \Gamma_0(\mathbb{R}^n)$ and assume* (59) *is consistent. Assume in addition the following.*

(i) *∇f is Lipschitz continuous on $\mathbb{R}^n$:*

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad x, y \in \mathbb{R}^n. \tag{62}$$

(ii) $0 < \liminf_{n\to\infty} \lambda_n \le \limsup_{n\to\infty} \lambda_n < 2/L$.

Then the sequence (x_k) generated by the proximal-gradient algorithm (61) *converges to a solution of* (59).

4.3. The Relaxed Proximal-Gradient Algorithm. The relaxed proximal-gradient algorithm generates a sequence (x_k) by the following iteration process.

Initialize $x_0 \in \mathbb{R}^n$ and iterate

$$x_{k+1} = (1 - \alpha_k) x_k + \alpha_k \left(\operatorname{prox}_{\lambda_k g} \circ (I - \lambda_k \nabla f)\right) x_k, \tag{63}$$

where $\{\alpha_k\}$ is the sequence of relaxation parameters and $\{\lambda_k\}$ is a sequence of positive real numbers.

Theorem 12 (see [14]). *Let $f, g \in \Gamma_0(\mathbb{R}^n)$ and assume* (59) *is consistent. Assume in addition the following.*

(i) *∇f is Lipschitz continuous on $\mathbb{R}^n$:*

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad x, y \in \mathbb{R}^n. \tag{64}$$

(ii) $0 < \liminf_{n\to\infty} \lambda_n \le \limsup_{n\to\infty} \lambda_n < 2/L$.

(iii) $0 < \liminf_{n\to\infty} \alpha_n \le \limsup_{n\to\infty} \alpha_n < 4/(2 + L \cdot \limsup_{n\to\infty} \lambda_n)$.

Then the sequence (x_k) generated by proximal-gradient algorithm (61) *converges to a solution of* (59).

If we take $\lambda_n \equiv \lambda \in (0, 2/L)$, then the relaxation parameters α_k can be chosen from a larger pool; they are allowed to be close to zero. More precisely, we have the following theorem.

Theorem 13 (see [14]). *Let $f, g \in \Gamma_0(\mathbb{R}^n)$ and assume* (59) *is consistent. Define the sequence (x_k) by the following relaxed proximal algorithm:*

$$x_{k+1} = (1 - \alpha_n) x_k + \alpha_k \operatorname{prox}_{\lambda g}\left(x_k - \lambda \nabla f(x_k)\right). \tag{65}$$

Suppose that

(a) *∇f satisfies the Lipschitz continuity condition (i) in Theorem 12;*

(b) $0 < \lambda < 2/L$ *and* $0 \le \alpha_k \le (2 + \lambda L)/4$ *for all k;*

(c) $\sum_{n=1}^{\infty} \alpha_n((4/(2 + \lambda L)) - \alpha_k) = \infty$.

Then (x_n) converges to a solution of (59).

4.4. Proximal-Gradient Algorithms Applied to Lasso. For Q-lasso (7), we take $f(x) = (1/2)\|(I - P_Q)Ax\|_2^2$ and $g(x) = \gamma\|x\|_1$. Noticing that $\nabla f(x) = A^t(I - P_Q)Ax$ which is Lipschitz continuous with constant $L = \|A\|_2^2$ for $I - P_Q$ is nonexpansive, we find that proximal-gradient algorithm (61) is reduced to the following algorithm for solving Q-lasso (7):

$$x_{k+1} = \operatorname{prox}_{\lambda_k \gamma \|\cdot\|_1}\left(I - \lambda_k A^t (I - P_Q) A\right) x_k. \tag{66}$$

The convergence theorem of general proximal-gradient algorithm (61) reads the following for Q-lasso (7).

Theorem 14. *Assume* $0 < \liminf_{k\to\infty} \lambda_k \le \limsup_{k\to\infty} \lambda_k < 2/\|A\|_2^2$. *Then the sequence (x_k) generated by proximal-gradient algorithm* (66) *converges to a solution of lasso* (7).

Remark 15. Relaxed proximal-gradient algorithms (63) and (65) also apply to Q-lasso (7). We however do not elaborate on them in detail.

Remark 16. Proximal-gradient algorithm (61) can be reduced to a projection-gradient algorithm in the case where the convex function g is homogeneous (i.e., $g(tx) = tg(x)$ for $t \ge 0$ and $x \in \mathbb{R}^n$) because the homogeneity of g implies that the proximal operator of g is actually a projection; more precisely, we have

$$\operatorname{prox}_{\lambda g} = P_{\lambda K}, \quad \lambda > 0, \tag{67}$$

where $K = \partial g(0)$. As a result, proximal-gradient algorithm (61) is reduced to the following projection-gradient algorithm:

$$x_{k+1} = \left(I - P_{\lambda_k K}\right)\left(I - \lambda_k \nabla f\right) x_k. \tag{68}$$

Now we apply projection-gradient algorithm (68) to Q-lasso (7). In this case, we have $f(x) = (1/2)\|(I - P_Q)Ax\|_2^2$ and $g(x) = \gamma\|x\|_1$ (homogeneous). Thus, $\nabla f(x) = A^t(I - P_Q)Ax$

and the convex set $K = \partial g(0)$ is given as $K = \gamma\partial(\|z\|_1)|_{z=0} = \gamma[-1,1]^n$. We find that, for each positive number $\lambda > 0$, $P_{\lambda K}$ is the projection of the Euclidean space $\mathbb{R}^n$ to the ℓ_∞ ball with radius of $\lambda\gamma$; that is, $\{x \in \mathbb{R}^n : \|x\|_\infty \le \lambda\gamma\}$. It turns out that proximal-projection algorithm (66) is rewritten as a projection algorithm below:

$$x_{k+1} = \left(I - P_{\lambda_k\gamma[-1,1]^n}\right)\left(I - \lambda_k A^t\left(I - P_Q\right)A\right)x_k. \quad (69)$$

5. An ℓ_1-ℓ_2 Regularization for the Q-Lasso

Q-lasso (7) may be ill posed and therefore needs to be regularized. Inspired by the elastic net [15] which regularizes lasso (1), we introduce an ℓ_1-ℓ_2 regularization for the Q-Lasso as the minimization

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|(I-P_Q)Ax\|_2^2 + \gamma\|x\|_1 + \delta\frac{1}{2}\|x\|_2^2 =: \varphi_{\gamma,\delta}(x), \quad (70)$$

where $\gamma > 0$ and $\delta > 0$ are regularization parameters. This is indeed the traditional Tikhonov regularization applied to Q-lasso (7).

Let $x_{\gamma,\delta}$ be the unique solution of (70) and set

$$\begin{aligned}\varphi_\gamma(x) &:= \frac{1}{2}\left\|\left(I-P_Q\right)Ax\right\|_2^2 + \gamma\|x\|_1,\\ \psi_\delta(x) &:= \frac{1}{2}\|(I-P_Q)Ax\|_2^2 + \delta\frac{1}{2}\|x\|_2^2\end{aligned} \quad (71)$$

which are the limits of $\varphi_{\gamma,\delta}(x)$ as $\delta \to 0$ and $\gamma \to 0$, respectively. Let

$$S_\gamma = \arg\min_{x\in\mathbb{R}^n}\varphi_\gamma(x), \qquad \widehat{x}_\delta = \arg\min_{x\in\mathbb{R}^n}\psi_\delta(x). \quad (72)$$

Proposition 17. *Assume the Q-least-squares problem*

$$\min_{x\in\mathbb{R}^n}\frac{1}{2}\left\|\left(I-P_Q\right)Ax\right\|_2^2 \quad (73)$$

is consistent (i.e., solvable) and let S be its nonempty set of solutions.

(i) *As $\delta \to 0$ (for each fixed $\gamma > 0$), $x_{\gamma,\delta} \to x_\gamma^\dagger$ which is the (ℓ_2) minimum-norm solution to Q-lasso (7). Moreover, as $\gamma \to 0$, every cluster point of $x_\gamma^\dagger$ is a (ℓ_1) minimum-norm solution of Q-least squares problem (73), that is, a point in the set $\arg\min_{x\in S}\|x\|_1$.*

(ii) *As $\gamma \to 0$ (for each fixed $\delta > 0$), $x_{\gamma,\delta} \to \widehat{x}_\delta$ which is the unique solution to the ℓ_2 regularized problem:*

$$\min_{x\in\mathbb{R}^n}\frac{1}{2}\|(I-P_Q)Ax\|_2^2 + \delta\frac{1}{2}\|x\|_2^2. \quad (74)$$

Moreover, as $\delta \to 0$, $\widehat{x}_\delta \to \widehat{x}$ which is the ℓ_2 minimal norm solution of (73); that is, $\widehat{x} = \arg\min_{x\in S}\|x\|_2$.

Proof. We have that $x_{\gamma,\delta}$ satisfies the optimality condition:

$$0 \in \partial\varphi_{\gamma,\delta}\left(x_{\gamma,\delta}\right), \quad (75)$$

where the subdifferential of $\varphi_{\gamma,\delta}$ is given by

$$\partial\varphi_{\gamma,\delta}(x) = A^t\left(I-P_Q\right)Ax + \delta x + \gamma\partial\|x\|_1. \quad (76)$$

It turns out that the above optimality condition is reduced to

$$-\frac{1}{\gamma}\left(A^t\left(I-P_Q\right)Ax_{\gamma,\delta} + \delta x_{\gamma,\delta}\right) \in \partial\left\|x_{\gamma,\delta}\right\|_1. \quad (77)$$

Using the subdifferential inequality, we obtain

$$\gamma\|x\|_1 \ge \gamma\left\|x_{\gamma,\delta}\right\|_1 - \left\langle A^t\left(I-P_Q\right)Ax_{\gamma,\delta} + \delta x_{\gamma,\delta}, x - x_{\gamma,\delta}\right\rangle \quad (78)$$

for $x \in \mathbb{R}^n$. Replacing x with $x_{\gamma',\delta'}$ for $\gamma' > 0$ and $\delta' > 0$ yields

$$\begin{aligned}&\gamma\left\|x_{\gamma',\delta'}\right\|_1\\ &\quad\ge \gamma\left\|x_{\gamma,\delta}\right\|_1 - \left\langle A^t\left(I-P_Q\right)Ax_{\gamma,\delta} + \delta x_{\gamma,\delta}, x_{\gamma',\delta'} - x_{\gamma,\delta}\right\rangle.\end{aligned} \quad (79)$$

Interchange γ and γ' and δ and δ' to get

$$\begin{aligned}&\gamma'\left\|x_{\gamma,\delta}\right\|_1\\ &\quad\ge \gamma'\left\|x_{\gamma',\delta'}\right\|_1 - \left\langle A^t\left(I-P_Q\right)Ax_{\gamma',\delta'} + \delta' x_{\gamma',\delta'}, x_{\gamma,\delta} - x_{\gamma',\delta'}\right\rangle.\end{aligned} \quad (80)$$

Adding up (79) and (80) results in

$$\begin{aligned}&\left(\gamma'-\gamma\right)\left(\left\|x_{\gamma,\delta}\right\|_1 - \left\|x_{\gamma',\delta'}\right\|_1\right)\\ &\quad\ge \left\langle\left(I-P_Q\right)Ax_{\gamma',\delta'} - \left(I-P_Q\right)Ax_{\gamma,\delta}, A^t x_{\gamma',\delta'} - A^t x_{\gamma,\delta}\right\rangle\\ &\qquad+ \left\langle\delta' x_{\gamma',\delta'} - \delta x_{\gamma,\delta}, x_{\gamma',\delta'} - x_{\gamma,\delta}\right\rangle\\ &\quad\ge \left\|\left(I-P_Q\right)Ax_{\gamma,\delta} - \left(I-P_Q\right)Ax_{\gamma',\delta'}\right\|_2^2\\ &\qquad+ \left(\delta'-\delta\right)\left\langle x_{\gamma',\delta'}, x_{\gamma',\delta'} - x_{\gamma,\delta}\right\rangle + \delta\left\|x_{\gamma',\delta'} - x_{\gamma,\delta}\right\|_2^2.\end{aligned} \quad (81)$$

Since ℓ_1-ℓ_2 regularization (70) is the Tikhonov regularization of Q-lasso (7), we get

$$\begin{aligned}\left\|x_{\gamma,\delta}\right\|_2 \le \left\|x_\gamma\right\|_2 \le c\left\|x_\gamma\right\|_1 &\le c\|x\|_1,\\ x_\gamma \in S_\gamma,\ x &\in S.\end{aligned} \quad (82)$$

Here $c > 0$ is a constant. It follows that $\{x_{\gamma,\delta}\}$ is bounded.

(i) For fixed $\gamma > 0$, we can use the theory of Tikhonov regularization to conclude that $x_{\gamma,\delta}$ is continuous in $\delta > 0$ and converges, as $\delta \to 0$, to $x_\gamma^\dagger$ which is the (ℓ_2) minimum-norm solution to Q-lasso (7), that is, the unique element $x_\gamma^\dagger := \arg\min_{x\in S_\gamma}\|x\|_2$. By Proposition 6, we also find that every cluster point of $x_\gamma^\dagger$, as $\gamma \to 0$, lies in the set $\arg\min_{x\in S}\|x\|_1$.

(ii) Fix $\delta > 0$ and use Proposition 6 to see that $x_{\gamma,\delta} \rightarrow \widehat{x}_\delta$ as $\gamma \rightarrow 0$. Now the standard property of Tikhonov's regularization ensures that $\widehat{x}_\delta \rightarrow \arg\min_{x\in S}\|x\|_2$ as $\delta \rightarrow 0$.

□

ℓ_1-ℓ_2 regularization (70) can be solved by proximal-gradient algorithm (61). Take $f(x) = (1/2)\|(I-P_Q)Ax\|_2^2 + (1/2)\delta\|x\|_2^2$ and $g(x) = \gamma\|x\|_1$; then algorithm (61) is reduced to

$$x_{k+1} = \operatorname{prox}_{\lambda_k\gamma\|\cdot\|_1}\left(x_k - \lambda_k\left[A^t\left(I-P_Q\right)Ax_k + \delta x_k\right]\right). \quad (83)$$

The convergence of this algorithm is given as follows.

Theorem 18. *Assume*

$$0 < \liminf_{k\to\infty}\lambda_k \le \limsup_{k\to\infty}\lambda_k < \frac{2}{\|A\|_2^2 + \delta}. \quad (84)$$

Then the sequence (x_k) *generated by algorithm* (83) *converges to the solution of* ℓ_1-ℓ_2 *regularization* (70).

We can also take $f(x) = (1/2)\|(I-P_Q)Ax\|_2^2$ and $g(x) = \gamma\|x\|_1 + (1/2)\delta\|x\|_2^2$. Then $\operatorname{prox}_{\mu g}(x) = \operatorname{prox}_{\nu\|\cdot\|_1}((1/(1+\mu\delta))x)$ with $\nu = \mu\gamma/(1+\mu\delta)$, and the proximal algorithm (61) is reduced to

$$x_{k+1} = \operatorname{prox}_{\nu_k\|\cdot\|_1}\left(\frac{1}{1+\delta\gamma_k}\left(x_k - \lambda_k A^t\left(I-P_Q\right)Ax_k\right)\right). \quad (85)$$

Here $\nu_k = \gamma\lambda_k/(1+\delta\gamma_k)$. Convergence of this algorithm is given below.

Theorem 19. *Assume*

$$0 < \liminf_{k\to\infty}\lambda_k \le \limsup_{k\to\infty}\lambda_k < \frac{2}{\|A\|_2^2}. \quad (86)$$

Then the sequence (x_k) *generated by the algorithm* (85) *converges to the solution of* ℓ_1-ℓ_2 *regularization* (70).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors were grateful to the anonymous referees for their helpful comments and suggestions which improved the presentation of this paper. This work was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, under Grant no. 2-363-1433-HiCi. The authors, therefore, acknowledge the technical and financial support of KAU.

References

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[2] 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.

[3] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[5] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[6] E. Candès and T. Tao, "The Dantzig selector: statistical estimation when *p* is much larger than *n*," *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[7] T. T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via l_1 minimization," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3388–3397, 2009.

[8] Y. Censor and T. Elfving, "A multiprojection algorithm using Bregman projections in a product space," *Numerical Algorithms*, vol. 8, no. 2–4, pp. 221–239, 1994.

[9] H.-K. Xu, "Iterative methods for the split feasibility problem in infinite-dimensional Hilbert spaces," *Inverse Problems*, vol. 26, no. 10, Article ID 105018, 17 pages, 2010.

[10] J.-J. Moreau, "Propriétés des applications 'prox'," *Comptes Rendus de l'Académie des Sciences*, vol. 256, pp. 1069–1071, 1963.

[11] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[12] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[13] C. A. Micchelli, L. Shen, and Y. Xu, "Proximity algorithms for image models: denoising," *Inverse Problems*, vol. 27, no. 4, Article ID 045009, 30 pages, 2011.

[14] H. K. Xu, "Properties and iterative methods for the Lasso and its variants," *Chinese Annals of Mathematics B*, vol. 35, no. 3, 2014.

[15] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society B*, vol. 67, no. 2, pp. 301–320, 2005.