Nonlinear Functional Analysis and Applications
Vol. 20, No. 2 (2015), pp. 247-256

http://nfaa.kyungnam.ac.kr/jour-nfaa.htm

STRONG CONVERGENCE OF AN ITERATIVE METHOD FOR THE LASSO

Peichao Duan[1] and Miaomiao Song[2]

[1]Department of Mathematics
Civil Aviation University of China, Tianjin, China
e-mail: pcduancauc@126.com

[2]Department of Mathematics
Civil Aviation University of China, Tianjin, China
e-mail: mmsong_1306004@126.com

Abstract. The lasso of Tibshiranis(1996) is a least-squares regularized by the l_1 norm. In recent years, the lasso has been paid much attention due to the involvement of the l_1 norm, which it's property is promoted by the sparseness of the norm. Now, we have mostly article studied it's weak convergence to a solution of the lasso. It is the purpose of this paper to show that under certain conditions, the iterative sequence $\{x_n\}$ converges strongly to a solution of the lasso, is also the unique solution of the following variational inequality: $\langle (I-h)x^*, \tilde{x}-x^*\rangle \geq 0,\ \forall x^* \in Fix(V_\lambda)$, where $h : H \to H$ is a contractive mapping and $V_\lambda : H \to H$ is an averaged mapping.

1. Introduction

The lasso is short for the least absolute shrinkage and selection operator, which was introduced by Tibshiranis [11] in 1996, and is formulated as the minimization problem

$$\min_x \frac{1}{2}\|Ax-b\|_2^2 \quad \textit{subject to} \quad \|x\|_1 \leq t, \tag{1.1}$$

where A is an $n\times m$(real) matrix, $x \in R^m, b \in R^n, t \geq 0$ is a tuning parameter. An equivalent formulation of (1.1) is formulated as the following regularized minimization problem

[0]Received November 17, 2014. Revised March 26, 2015.
[0]2010 Mathematics Subject Classification: 47H09, 47H10, 49J40, 65J15, 65J22.
[0]Keywords: Lasso, proximal operators, fixed point, variational inequality, averaged mapping.

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \gamma\|x\|_1, \tag{1.2}$$

where $\gamma \geq 0$ is a regularization parameter.

From now on, let H be a Hilbert space with inner product and norm, respectively denote $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. We use $Fix(T)$ to denote the set of fixed points of a mapping T; i.e., $Fix(T) = \{x \in H : x = Tx\}$. We also use " $\to$ "and " $\rightharpoonup$ " to denote the strong convergence and the weak convergence, respectively.

A mapping $T : H \to H$ is called nonexpansive if

$$\|Tx - Ty\| \leq \|x - y\|, \quad \forall\, x, y \in H.$$

A mapping $h : H \to H$ is called ρ-contractive if there exists a contraction constant $\rho \in [0, 1)$ such that

$$\|h(x) - h(y)\| \leq \rho\|x - y\|, \quad \forall\, x, y \in H.$$

A mapping $T : H \to H$ is called L-Lipschitzian continuous if there exists a constant $L > 0$ such that

$$\|Tx - Ty\| \leq L\|x - y\|, \quad \forall\, x, y \in H.$$

A mapping $F : H \to H$ is called η-inverse strongly monotone(η-ism), if there exists a constant $\eta > 0$ satisfing the following inequality

$$\langle Fx - Fy, x - y\rangle \geq \eta\|Fx - Fy\|^2, \quad \forall\, x, y \in H.$$

A mapping $V : H \to H$ is called α-averaged(α-av for short) if

$$V = (1 - \alpha)I + \alpha T,$$

where $\alpha \in (0, 1)$, $T : H \to H$ is nonexpansive.

The lasso has been received much attention in recent years due to the involvement of the l_1 norm which promotes sparsity. If a certain appropriate sparsity condition is imposed, then we can get a well result on solving the problem of (1.1). We know some iterative methods for the lasso have been discovered by other authors. However, up to now, only weak convergence results have been discussed. But, it is well known that strongly convergent algorithms are of fundamental importance for solving infinite dimensional problems.

We get the inspiration from Xu's iterative methods for the lasso [14] and Moudafi's viscosity approximationn method [10], we transfer weak convergence of the lasso's certain iterative methods to convergence strongly. In the last part of this paper we combine the proximal operators with a viscosity procedure to get the strong convergent of the iterative sequence of iterates generated by our scheme. Meanwhile, we also get the convergent point of the sequence which is the unique solution of the variational inequality.

2. Preliminaries

For the purpose, we first give some preliminaries, which will be needed to prove our main results.

Lemma 2.1. *For all $x, y \in H$, there holds the following relation:*

$$\|x+y\|^2 \leq \|x\|^2 + 2\langle y, x+y\rangle.$$

Lemma 2.2. ([1, 6]) *Let F be μ-ism, thus we can get F is $\frac{1}{\mu}$-Lipschitzian mapping with coefficient $\mu > 0$, and λF is $\frac{\mu}{\lambda}$-ism. If $\frac{\mu}{\lambda} > \frac{1}{2}$, we know $I - \lambda F$ is $\frac{\lambda}{2\mu}$-av.*

Lemma 2.3. ([2]) *If $T_1, T_2, \cdots, T_n$ are averaged mappings, we can get that $T_n T_{n-1} \cdots T_1$ is averaged. In particular, if T_i is α_i-av, i=1,2, then T_2T_1 is $(\alpha_2 + \alpha_1 - \alpha_2\alpha_1)$-av.*

Lemma 2.4. ([12]) *Let $h : H \to H$ be a ρ-contraction with $\rho \in [0, 1)$ and $T : H \to H$ be a nonexpansive mapping. Then*

(i) *$I - h$ is $(1-\rho)$-strongly monotone:*

$$\langle (I-h)x - (I-h)y, x-y\rangle \geq (1-\rho)\|x-y\|^2, \quad \forall\, x, y \in H.$$

(ii) *$I - T$ is monotone:*

$$\langle (I-T)x - (I-T)y, x-y\rangle \geq 0, \quad \forall\, x, y \in H.$$

Lemma 2.5. ([4], Demiclosedness Principle) *Let H be a real Hilbert space, and let $T : H \to H$ be a nonexpansive mapping with $Fix(T) \neq \emptyset$. If $\{x_n\}$ is a sequence in H converges weakly to x and if $\{(I-T)x_n\}$ converges strongly to y, then $(I-T)x = y$; in particulary, if $y = 0$, then $x \in Fix(T)$.*

Lemma 2.6. ([13]) *Assume $\{a_n\}$ is a sequence of nonnegative real numbers such that*

$$a_{n+1} \leq (1-\gamma_n)a_n + \gamma_n\delta_n + \beta_n, \ n \geq 0, \tag{2.1}$$

where $\{\gamma_n\}$ and $\{\beta_n\}$ is a sequence in $(0, 1)$ and $\{\delta_n\}$ is a sequence in $\mathbb{R}$ such that

(i) *$\sum_{n=0}^{\infty} \gamma_n = \infty$;*
(ii) *$\limsup_{n\to\infty} \delta_n \leq 0$ or $\sum_{n=0}^{\infty} |\gamma_n\delta_n| < \infty$;*
(iii) *$\Sigma_{n=0}^{\infty}\beta_n < \infty$.*

Then $\lim_{n\to\infty} a_n = 0$.

3. Main results

Let H be a Hilbert space and let $\Gamma_0(H)$ be the space of convex functions in H that are proper, lower semicontinuous and convex.

Definition 3.1. ([8, 9]) The proximal operator of $\varphi \in \Gamma_0(H)$ is defined by

$$prox_{\varphi}(x) = arg\min_{\nu\in H}\left\{\varphi(\nu) + \frac{1}{2}\|\nu - x\|^2\right\}, \quad x \in H.$$

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

$$prox_{\lambda\varphi}(x) = arg\min_{\nu\in H}\left\{\varphi(\nu) + \frac{1}{2\lambda}\|\nu - x\|^2\right\}, \quad x \in H.$$

Lemma 3.2. *The proximal identity*

$$prox_{\lambda\varphi}x = prox_{\mu\varphi}(\frac{\mu}{\lambda}x + (1 - \frac{\mu}{\lambda})prox_{\lambda\varphi}x) \tag{3.1}$$

holds for $\varphi \in \Gamma_0(H),\ \lambda > 0$ *and* $\mu > 0$.

We list some of the useful properties of the proximal operator.

Lemma 3.3. ([3, 7, 13]) *Let* $\varphi \in \Gamma_0(H)$ *and* $\lambda \in (0, \infty)$.

(i) *If* C *is a nonempty closed convex subset of* H *and* $\varphi = I_C$ *is the indicator function of* C*, then the proximal operators* $prox_{\lambda\varphi} = P_C$ *for all* $\lambda > 0$*, where* P_C *is the metric projection from* C *onto* H*.*

(ii) *The operator* $prox_{\lambda\varphi}$ *is firmly nonexpansive(hence nonexpansive). Recall a mapping* $T : H \to H$ *is firmly nonexpansive if*

$$\|Tx - Ty\|^2 \le \langle Tx - Ty, x - y\rangle, \quad x, y \in H.$$

(iii) *The operator* $prox_{\lambda\varphi} = (I + \lambda\partial\varphi)^{-1} = J_{\lambda}{}^{\partial\varphi}$*, the resolvent of the subdifferential* $\partial\varphi$ *of* φ*.*

(iv) *If* $f : H \to \mathbb{R}$ *is a differentiable functional, then we denote by* ∇f *the gradient of* f*. Assume that* ∇f *is Lipschitz continuous on* H*. The operator* $V_\lambda = prox_{\lambda g}(I - \lambda\nabla f)$ *is* $\frac{2+\lambda L}{4}$*-av for each* $0 < \lambda < \frac{2}{L}$*. The proximal operator can be used to minimize the sum of two convex functions*

$$\min_{x\in H} f(x) + g(x), \tag{3.2}$$

where $f, g \in \Gamma_0(H)$*. It is often the case where one of them is differentiable. The following is an equivalent fixed point formulation of* (3.2)*.*

Lemma 3.4. ([14]) *Let $f, g \in \Gamma_0(H)$. Let $x^* \in H$ and $\lambda > 0$. Assume that f is finite-valued and differential on H. Then x^* is a solution to* (3.2) *if and only if x^* solves the fixed point equation*

$$x^* = (prox_{\lambda g}(I - \lambda \nabla f))x^*. \tag{3.3}$$

Theorem 3.5. *Let $f, g \in \Gamma_0(H)$ and assume that* (3.2) *is consist. Given $x_0 \in H$ and define the sequence $\{x_n\}$ by the following viscosity proximal algorithm*

$$x_{n+1} = \alpha_n h(x_n) + (1 - \alpha_n)V_{\lambda_n} x_n, \tag{3.4}$$

where $h : H \to H$ is ρ-contraction with constant $\rho \in [0, 1)$, $\lambda_n \in (0, \frac{2}{L})$, $\alpha_n \in (0, \frac{2+\lambda_n L}{4})$, ∇f satisfies the Lipschitz continuity, and $V_{\lambda_n} = prox_{\lambda_n g}(I - \lambda_n \nabla f)$. Suppose that

(i) $\lim_{n\to\infty} \alpha_n = 0$ *and* $\sum_{n=1}^{\infty} \alpha_n = \infty$;
(ii) $\sum_{n=1}^{\infty} |\alpha_{n+1} - \alpha_n| < \infty$;
(iii) $0 < \liminf_{n\to\infty} \lambda_n \le \limsup_{n\to\infty} \lambda_n < \frac{2}{L}$;
(iv) $\sum_{n=1}^{\infty} |\lambda_n - \lambda_{n-1}| < \infty$.

Then $\{x_n\}$ converges strongly to $\tilde{x}$, where $\tilde{x}$ is a solution of (3.2). *Meanwhile, it is also the unique solution for the following variational inequality:*

$$\langle (I - h)x^*, \tilde{x} - x^* \rangle \ge 0, \ \ \forall\, \tilde{x} \in Fix(V_\lambda). \tag{3.5}$$

Proof. Let S be the nonempty solution set of (3.2). The proof is divided into several steps.

Step 1. Show that $\{x_n\}$ is bounded.

For any $p \in S$, we get

$$\begin{aligned}
&\|x_{n+1} - p\| \\
&= \|\alpha_n h(x_n) + (1 - \alpha_n)V_{\lambda_n} x_n - p\| \\
&= \|\alpha_n (h(x_n) - p) + (1 - \alpha_n)(V_{\lambda_n} x_n - p)\| \\
&\le \alpha_n \|h(x_n) - h(p)\| + \alpha_n \|h(p) - p\| + (1 - \alpha_n)\|V_{\lambda_n} x_n - p\| \\
&\le \alpha_n \rho \|x_n - p\| + \alpha_n \|h(p) - p\| + (1 - \alpha_n)\|x_n - p\| \\
&= (1 - \alpha_n(1 - \rho))\|x_n - p\| + \alpha_n (1 - \rho)\frac{\|h(p) - p\|}{1 - \rho}.
\end{aligned} \tag{3.6}$$

So, by induction, we can conclude from (3.6) get that

$$\|x_n - p\| \le \max\left\{ \|x_0 - p\|, \frac{\|h(p) - p\|}{1 - \rho} \right\},$$

which implies that the sequence $\{x_n\}$ is bounded.

Step 2. Show that $\|x_{n+1} - x_n\| \to 0$.

By the iterative scheme (3.4), we have

$$\begin{aligned}
&\|x_{n+1}-x_n\|\\
&=\|\alpha_n h(x_n)+(1-\alpha_n)V_{\lambda_n}x_n-\alpha_{n-1}h(x_{n-1})-(1-\alpha_{n-1})V_{\lambda_{n-1}}x_{n-1}\|\\
&\le\|\alpha_n h(x_n)-\alpha_{n-1}h(x_{n-1})\|+\|(1-\alpha_n)V_{\lambda_n}x_n-(1-\alpha_{n-1})V_{\lambda_{n-1}}x_{n-1}\|\\
&\le\|\alpha_n h(x_n)-\alpha_{n-1}h(x_n)+\alpha_{n-1}h(x_n)-\alpha_{n-1}h(x_{n-1})\|\\
&\quad+\|(1-\alpha_n)V_{\lambda_n}x_n-(1-\alpha_{n-1})V_{\lambda_{n-1}}x_{n-1}\|\\
&\le|\alpha_n-\alpha_{n-1}|\|h(x_n)\|+\rho\alpha_{n-1}\|x_n-x_{n-1}\|\\
&\quad+\|(1-\alpha_n)V_{\lambda_n}x_n-(1-\alpha_{n-1})V_{\lambda_{n-1}}x_{n-1}\|. \qquad (3.7)
\end{aligned}$$

As V_{λ_n} given in Lemma 3.3 is $\frac{2+\lambda_n L}{4}$-av, we can rewrite

$$V_{\lambda_n}=prox_{\lambda_n g}(I-\lambda_n\nabla f)=(I-\gamma_n)I+\gamma_n T_n, \qquad (3.8)$$

where $\gamma_n=\frac{2+\lambda_n L}{4}$, T_n is nonexpansive. We then get

$$\begin{aligned}
&\|(1-\alpha_n)V_{\lambda_n}x_n-(1-\alpha_{n-1})V_{\lambda_{n-1}}x_{n-1}\|\\
&=\|(1-\alpha_n)V_{\lambda_n}x_n-(1-\alpha_{n-1})V_{\lambda_n}x_n+(1-\alpha_{n-1})V_{\lambda_n}x_n\\
&\quad-(1-\alpha_{n-1})V_{\lambda_{n-1}}x_{n-1}\|\\
&\le|\alpha_n-\alpha_{n-1}|\|V_{\lambda_n}x_n\|+(1-\alpha_{n-1})\|V_{\lambda_n}x_n-V_{\lambda_{n-1}}x_{n-1}\|, \qquad (3.9)
\end{aligned}$$

where, we also know

$$\begin{aligned}
&\|V_{\lambda_n}x_n-V_{\lambda_{n-1}}x_{n-1}\|\\
&=\|V_{\lambda_n}x_n-V_{\lambda_{n-1}}x_n+V_{\lambda_{n-1}}x_n-V_{\lambda_{n-1}}x_{n-1}\|\\
&\le\|V_{\lambda_n}x_n-V_{\lambda_{n-1}}x_n\|+\|x_n-x_{n-1}\| \qquad (3.10)
\end{aligned}$$

and

$$\begin{aligned}
&\|V_{\lambda_n}x_n-V_{\lambda_{n-1}}x_n\|\\
&=\|prog_{\lambda_n g}(I-\lambda_n\nabla f)x_n-prog_{\lambda_{n-1}g}(I-\lambda_{n-1}\nabla f)x_n\|\\
&=\|prox_{\lambda_{n-1}g}(\frac{\lambda_{n-1}}{\lambda_n}(I-\lambda_n\nabla f)x_n\\
&\quad+(1-\frac{\lambda_{n-1}}{\lambda_n})prox_{\lambda_n g}(I-\lambda_n\nabla f)x_n)-prox_{\lambda_{n-1}g}(I-\lambda_{n-1}\nabla f)x_n\|\\
&\le\|\frac{\lambda_{n-1}}{\lambda_n}(I-\lambda_n\nabla f)x_n+(1-\frac{\lambda_{n-1}}{\lambda_n})prox_{\lambda_n g}(I-\lambda_n\nabla f)x_n\\
&\quad-(I-\lambda_{n-1}\nabla f)x_n\|\\
&=\|(1-\frac{\lambda_{n-1}}{\lambda_n})prox_{\lambda_n g}(I-\lambda_n\nabla f)x_n-(1-\frac{\lambda_{n-1}}{\lambda_n})(I-\lambda_n\nabla f)x_n\|\\
&\quad+\|(\lambda_{n-1}-\lambda_n)\nabla f(x_n)\|
\end{aligned}$$

$$\leq |1 - \frac{\lambda_{n-1}}{\lambda_n}| \|prox_{\lambda_n g}(I - \lambda_n \nabla f)x_n - (I - \lambda_n \nabla f)x_n\| \\ + |\lambda_{n-1} - \lambda_n| \|\nabla f(x_n)\|. \tag{3.11}$$

Thus, we can get that

$$\begin{aligned} &\|x_{n+1} - x_n\| \\ &\leq (\alpha_n - \alpha_{n-1})\|h(x_n)\| + \rho\alpha_{n-1}\|x_n - x_{n-1}\| \\ &\quad + |\alpha_n - \alpha_{n-1}|\|V_{\lambda_n} x_n\| + (1 - \alpha_{n-1})\|x_n - x_{n-1}\| \\ &\quad + (1 - \alpha_{n-1})\left(\left|1 - \frac{\lambda_{n-1}}{\lambda_n}\right| \|prox_{\lambda_n} g(I - \lambda_n \nabla f)x_n - (I - \lambda_n \nabla f)x_n\| \right. \\ &\quad \left. + |\lambda_n - \lambda_{n-1}|\|\nabla f(x_n)\|\right). \end{aligned} \tag{3.12}$$

Namely,

$$\begin{aligned} &\|x_{n+1} - x_n\| \\ &\leq (1 - \alpha_{n-1}(1 - \rho))\|x_n - x_{n-1}\| + (\alpha_n - \alpha_{n-1})(\|h(x_n)\| + \|V_{\lambda_n} x_n\|) \\ &\quad + (1 - \alpha_{n-1})\left(\left|1 - \frac{\lambda_{n-1}}{\lambda_n}\right| \|prox_{\lambda_n} g(I - \lambda_n \nabla f)x_n - (I - \lambda_n \nabla f)x_n\| \right. \\ &\quad \left. + |\lambda_{n-1} - \lambda_n|\|\nabla f(x_n)\|\right), \end{aligned} \tag{3.13}$$

where

$$\begin{aligned} &\gamma_n \delta_n \\ &= (\alpha_n - \alpha_{n-1})(\|h(x_n)\| + \|V_{\lambda_n} x_n\|) \\ &\quad + (1 - \alpha_{n-1})\left(\left|1 - \frac{\lambda_{n-1}}{\lambda_n}\right| \|prox_{\lambda_n} g(I - \lambda_n \nabla f)x_n - (I - \lambda_n \nabla f)x_n\| \right. \\ &\quad \left. + |\lambda_n - \lambda_{n-1}|\|\nabla f(x_n)\|\right). \end{aligned} \tag{3.14}$$

We use the conditions to Lemma 2.6 can get that

$$\sum_{n=1}^{\infty} \gamma_n \delta_n < \infty.$$

Thus, we can get that

$$\lim_{n \to \infty} \|x_{n+1} - x_n\| = 0. \tag{3.15}$$

Step 3. Show that $\|x_n - V_{\lambda_n} x_n\| \to 0$.

It is easy to get

$$\begin{aligned}&\|x_n - V_{\lambda_n}x_n\|\\&= \|x_n - (x_{n+1} - \alpha_n h(x_n) + \alpha_n V_{\lambda_n}x_n)\|\\&\leq \|x_{n+1} - x_n\| + \alpha_n\|h(x_n) - V_{\lambda_n}x_n\| \to 0 \quad (n \to \infty).\end{aligned} \tag{3.16}$$

Step 4. Show that

$$\omega_w(x_n) \subset S. \tag{3.17}$$

Here, $\omega_w(x_n)$ is the set of all weak cluster points of $\{x_n\}$. Note that $\{x_n\}$ is bounded and (3.17) together guarantee that $\{x_n\}$ converges weakly to a point in S and then the proof is consist. To see (3.17), we prove as follows. Take $\tilde{x} \in \omega_w\{x_n\}$ and assume that $\{x_{n_j}\}$ is a subsequence of $\{x_n\}$ weakly converging to $\tilde{x}$. Hence by (3.15), $x_{n_j+1} \rightharpoonup \tilde{x}$ as well. Without loss of generality, we may assume $\lambda_{n_j} \to \lambda$, then $0 < \lambda < \frac{2}{L}$. Set $V_\lambda = prox_{\lambda g}(I - \lambda\nabla f)$, then V_λ is nonexpansive. Set

$$y_j = x_{n_j} - \lambda_{n_j}\nabla f(x_{n_j}), \;\; z_j = x_{n_j} - \lambda\nabla f(x_{n_j}).$$

Using the proximal identify of Lemma 3.2, we deduce that

$$\begin{aligned}&\|V_{\lambda_{n_j}}x_{n_j} - V_\lambda x_{n_j}\|\\&= \|prox_{\lambda_{n_j}g}y_j - prox_{\lambda g}z_j\|\\&= \left\|prox_{\lambda g}\left(\frac{\lambda}{\lambda_{n_j}}y_j + \left(1 - \frac{\lambda}{\lambda_{n_j}}\right)prox_{\lambda_{n_j}g}y_j\right) - prox_{\lambda g}z_j\right\|\\&\leq \left\|\frac{\lambda}{\lambda_{n_j}}y_j + \left(1 - \frac{\lambda}{\lambda_{n_j}}\right)prox_{\lambda_{n_j}g}y_j - z_j\right\|\\&\leq \frac{\lambda}{\lambda_{n_j}}\|y_j - z_j\| + \left(1 - \frac{\lambda}{\lambda_{n_j}}\right)\|prox_{\lambda_{n_j}g}y_j - z_j\|\\&= \frac{\lambda}{\lambda_{n_j}}|\lambda_{n_j} - \lambda|\|\nabla f(x_{n_j})\| + \left(1 - \frac{\lambda}{\lambda_{n_j}}\right)\|prox_{\lambda_{n_j}g}y_j - z_j\|.\end{aligned} \tag{3.18}$$

Since $\{x_n\}$ is bounded, ∇f is Lipschitz continuous, and $\lambda_{n_j} \to \lambda$, we immediately derive from the last relation that $\|V_{\lambda_{n_j}}x_{n_j} - V_\lambda x_{n_j}\| \to 0$. As a result, we find

$$\|x_{n_j} - V_\lambda x_{n_j}\| \leq \|x_{n_j} - V_{\lambda_{n_j}}x_{n_j}\| + \|V_{\lambda_{n_j}}x_{n_j} - V_\lambda x_{n_j}\| \to 0. \tag{3.19}$$

Now the demiclosedness of the nonexpansive mapping $I - V_\lambda$ implies that $(I - V_\lambda)\tilde{x} = 0$. Namely, $\tilde{x} \in Fix(V_\lambda) = S$. Therefore, (3.17) is proved.

Step 5. Show that $\|x_n - x^*\| \to 0$, where x^* is a unique solution of the variational inequality $\langle (I-h)\tilde{x}, x^* - \tilde{x}\rangle \geq 0, \ \forall \tilde{x} \in Fix(V_\lambda)$. Then we have

$$\begin{aligned}
&\|x_{n+1} - x^*\|^2 \\
&= \|\alpha_n h(x_n) + (1-\alpha_n)V_{\lambda_n} x_n - x^*\| \\
&= \|\alpha_n (h(x_n) - x^*) + (1-\alpha_n)(V_{\lambda_n} x_n - x^*)\|^2 \\
&\leq (1-\alpha_n)^2 \|V_{\lambda_n} x_n - x^*\|^2 + 2\alpha_n \langle h(x_n) - x^*, x_{n+1} - x^*\rangle \\
&\leq (1-\alpha_n)^2 \|x_n - x^*\|^2 + 2\alpha_n \rho \|x_n - x^*\| \|x_{n+1} - x^*\| \\
&\quad + 2\alpha_n \langle h(x^*) - x^*, x_{n+1} - x^*\rangle \\
&\leq (1-\alpha_n)^2 \|x_n - x^*\|^2 + \alpha_n \rho \|x_n - x^*\|^2 + \alpha_n \rho \|x_{n+1} - x^*\|^2 \\
&\quad + 2\alpha_n \langle h(x^*) - x^*, x_{n+1} - x^*\rangle.
\end{aligned} \tag{3.20}$$

Thus, we get

$$\begin{aligned}
&\|x_{n+1} - x^*\|^2 \\
&\leq \frac{(1-\alpha_n)^2 + \alpha_n \rho}{1 - \alpha_n \rho} \|x_n - x^*\|^2 + \frac{2\alpha_n}{1-\alpha_n \rho} \langle h(x^*) - x^*, x_{n+1} - x^*\rangle \\
&\leq \left(1 - \frac{2\alpha_n(1+\rho)}{1-\alpha_n \rho}\right) \|x_n - x^*\|^2 \\
&\quad + \frac{2\alpha_n(1+\rho)}{1-\alpha_n\rho} \frac{1}{1+\rho} \left(\langle h(x^*) - x^*, x_{n+1} - x^*\rangle + \frac{\alpha_n}{2} \|x_n - x^*\|^2 \right) \\
&= (1 - \tilde{\alpha_n}) \|x_n - x^*\|^2 + \tilde{\alpha_n} \tilde{\beta_n},
\end{aligned} \tag{3.21}$$

where

$$\begin{aligned}
\tilde{\alpha_n} &= \frac{2\alpha_n(1+\rho)}{1-\alpha_n\rho}, \\
\tilde{\beta_n} &= \frac{1}{1+\rho} \left(\langle h(x^*) - x^*, x_{n+1} - x^*\rangle + \frac{\alpha_n}{2} \|x_n - x^*\|^2 \right).
\end{aligned}$$

We next show that

$$\limsup_{n\to\infty} \langle h(x^*) - x^*, x_{n+1} - x^*\rangle \leq 0. \tag{3.22}$$

Indeed take a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that

$$\limsup_{n\to\infty} \langle h(x^*) - x^*, x_{n+1} - x^*\rangle = \lim_{k\to\infty} \langle h(x^*) - x^*, x_{n_k} - x^*\rangle. \tag{3.23}$$

We may assume that $x_{n_k} \rightharpoonup \tilde{x}$. It follows from (3.5) that we get

$$\limsup_{n\to\infty} \langle h(x^*) - x^*, x_{n+1} - x^*\rangle = \langle h(x^*) - x^*, \tilde{x} - x^*\rangle \leq 0. \tag{3.24}$$

It is easily seen that $\tilde{\alpha_n} \to 0$, $\Sigma_{n=1}^{\infty}\tilde{\alpha_n} = \infty$, and $\limsup_{n\to\infty}\tilde{\beta_n} \leq 0$ by (3.24). Hence, by Lemma 2.6, the sequence $\{x_n\}$ converges strongly to x^*. □

Acknowledgments: This work was supported by Fundamental Research Funds for the Central Universities (3122015L007).

References

[1] C. Byrne, *A unified treatment of some iterative algorithms in signal processing and image reconstruction*, Inverse Probl., **20** (2004), 103–120.
[2] P.L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Opitim., **53** (2004), 475–504.
[3] P.L. Combettes and R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., **4**(4) (2005), 1168–1200.
[4] K. Geobel and W.A. Kirk, *Topics in metric fixed point theory*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, **28** (1990).
[5] D. Kinderlehrer and G. Stampacchia, *An introduction to variational inequalities and their applications*, Academic Press, New York (1985).
[6] C. Martinez-Yanes and H.K. Xu, *Strong convergence of the CQ method for fixed-point iteration processes*, Nonlinear Anal., **64** (2006), 2400–2411.
[7] C.A. Micchellil, L. Shen and Y. Xu, *Proximity algorithms for image models: Denoising*, Inverse Probl., 045009, **27** (2011), 30.
[8] J.J. Moreau, *Proprietes des applications 'prox'*, C.R. Acad. Sci. Paris Sér.A Math., **256** (1963), 1069–1071.
[9] J.J. Moreau, *Proximite et dualite dans un espace hilbertien*, Bull. Soc. Math. France, **93** (1965), 279–299.
[10] A. Moudafi, *Viscosity approximation methods for fixed-point problems*, J. Math. Anal. Appl., **241** (2000), 46–55.
[11] R. Tibshiranis, *Regression shrinkage and selection via the lasso*, J. Royal Statist. Soc. Ser. B, **58** (1996), 267–288.
[12] H.K. Xu, *Viscosity approximation methods for nonexpansive mappings*, J. Math. Anal. Appl., **298** (2004), 279–291.
[13] H.K. Xu, *Averaged mappings and the gratident-projection algorithm*, J. Optim. Theory Appl., **150** (2011), 360–378.
[14] H.K. Xu, *Properties and iterative methods for the lasso and its variants*, Chin. Ann. Math., **35B**(3) (2014), 1–18.