Supplementary Material of
Generalized Singular Value Thresholding

Canyi Lu[1], **Changbo Zhu**[1], **Chunyan Xu**[2], **Shuicheng Yan**[1], **Zhouchen Lin**[3,*]
[1] Department of Electrical and Computer Engineering, National University of Singapore
[2] School of Computer Science and Technology, Huazhong University of Science and Technology
[3] Key Laboratory of Machine Perception (MOE), School of EECS, Peking University
canyilu@nus.edu.sg, zhuchangbo@gmail.com, xuchunyan01@gmail.com, eleyans@nus.edu.sg, zlin@pku.edu.cn

1 Ananlysis of the Proximal Operator of Nonconvex Function

In the following development, we consider the following problem

$$\mathbf{Prox}_g(b) = \arg\min_{x\geq 0} f_b(x) = g(x) + \frac{1}{2}(x-b)^2, \tag{1}$$

where $g(x)$ satisfies the following assumption.

Assumption 1. *$g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, $g(0) = 0$. g is concave, nondecreasing and differentiable. The gradient ∇g is convex.*

Set $C_b(x) = b - x$ and $D(x) = \nabla g(x)$. Let $\bar{b} = \sup\{b \mid C_b(x)$ and $D(x)$ have no intersection$\}$, and $x_2^{\bar{b}} = \inf\{\, x \mid (x, y)$ is the intersection point of $C_{\bar{b}}(x)$ and $D(x)\}$.

1.1 Proof of Proposition 2

Proposition 2. *Given g satisfying **Assumption** 1 and $\nabla g(0) = +\infty$. Restricted on $[0, +\infty)$, when $b > \bar{b}$, $C_b(x)$ and $D(x)$ have two intersection points, denoted as $P_1^b = (x_1^b, y_1^b)$, $P_2^b = (x_2^b, y_2^b)$, and $x_1^b < x_2^b$. If there does not exist $b > \bar{b}$ such that $f_b(0) = f_b(x_2^b)$, then $\mathbf{Prox}_g(b) = 0$ for all $b \geq 0$. If there exists $b > \bar{b}$ such that $f_b(0) = f_b(x_2^b)$, let $b^* = \inf\{b \mid b > \bar{b}, f_b(0) = f_b(x_2^b)\ \}$. Then we have*

$$\mathbf{Prox}_g(b) = \operatorname*{argmin}_{x\geq 0} f_b(x) \begin{cases} = x_2^b, & \text{if } b > b^*, \\ \ni 0, & \text{if } b \leq b^*. \end{cases} \tag{2}$$

Remark: When b^* exists and $b > b^*$, because $D(x) = \nabla g(x)$ is convex and decreasing, we can conclude that $C_b(x)$ and $D(x)$ have exactly two intersection points. When $b \leq b^*$, $C_b(x)$ and $D(x)$ may have multiple intersection points.

Proof. When $b > \bar{b}$, since $\nabla f_b(x) = D(x) - C_b(x)$, we can easily see that f_b is increasing on $(0, x_1^b)$, decreasing on (x_1^b, x_2^b) and increasing on (x_2^b, b). So, 0 and x_2^b are two local minimum points of $f_b(x)$ on $[0, b]$.

Case 1 : If there exists $b > \bar{b}$ such that $f_b(0) = f_b(x_2^b)$, denote $b^* = \inf\{b \mid b > \bar{b}, f_b(0) = f_b(x_2^b)\ \}$.

First, we consider $b > b^*$. Let $b = b^* + \varepsilon$ for some $\varepsilon > 0$. We have

$$\begin{aligned} & f_b(x_2^{b^*}) - f_b(0) \\ =& \frac{1}{2}(x_2^{b^*} - b^* - \varepsilon)^2 + g(x^*) - \frac{1}{2}(b^* + \varepsilon)^2 \\ =& \frac{1}{2}(x_2^{b^*} - b^*)^2 - \frac{1}{2}(b^*)^2 - \varepsilon(x_2^{b^*} - b^*) - \varepsilon b^* \\ =& f_{b^*}(x_2^{b^*}) - f_{b^*}(0) - \varepsilon x_2^* \\ =& -\varepsilon x_2^* < 0. \end{aligned}$$

Since f_b is decreasing on $[x_2^{b^*}, x_2^b]$, we conclude that $f_b(0) > f_b(x_2^{b^*}) \geq f_b(x_2^b)$. So, when $b > b^*$, x_2^b is the global minimum of $f_b(x)$ on $[0, b]$.

*Corresponding author.

Second, we consider $\bar{b} < b \le b^*$. We show that $f_b(0) \le f_b(x_2^b)$ by contradiction. Suppose that there exists b such that $f_b(0) > f_b(x_2^b)$. Since $f_{\bar{b}}$ is strictly increasing on $(0, x_2^{\bar{b}})$, we have $f_{\bar{b}}(x_2^{\bar{b}}) > f_{\bar{b}}(0)$. Because we have

$$\begin{cases} f_{\bar{b}}(x_2^{\bar{b}}) > f_{\bar{b}}(0), \\ f_b(x_2^b) < f_b(0), \end{cases}$$

by a direct computation, we get

$$\begin{cases} g(x_2^{\bar{b}}) - x_2^{\bar{b}} \nabla g(x_2^{\bar{b}}) - \frac{1}{2}(x_2^{\bar{b}})^2 > 0, \\ g(x_2^b) - x_2^b \nabla g(x_2^b) - \frac{1}{2}(x_2^b)^2 < 0. \end{cases}$$

According to the intermediate value theorem, there exists $\tilde{x}$ such that $x_2^{\bar{b}} < \tilde{x} < x_2^b$ and $g(\tilde{x}) - \tilde{x}\nabla g(\tilde{x}) - \frac{1}{2}(\tilde{x})^2 = 0$. Let $\tilde{b} = \nabla g(\tilde{x}) + \tilde{x}$. Then, $(\tilde{x}, \tilde{b} - \tilde{x})$ is the intersection point of $C_{\tilde{b}}(x)$ and $D(x)$ such that $f_{\tilde{b}}(\tilde{x}) = f_{\tilde{b}}(0)$. Since $x_2^{\bar{b}} < \tilde{x} < x_2^b$ and ∇g is convex and nonincreasing, we conclude that $\bar{b} < \tilde{b} < b \le b^*$, which contradicts the minimality of b^*.

Also, when $b \le \bar{b}$, we have $\nabla f_b(x) = D(x) - C_b(x) \ge 0$, because $D(x)$ is above $C_b(x)$. So, the global minimum of $f_b(x)$ on $[0, b]$ is 0.

Case 2 : Suppose for all $b^* > \bar{b}$, $f_{b^*}(0) \ne f_{b^*}(x_2^{b^*})$. Since $f_{\bar{b}}$ is increasing on $(0, x_2^{\bar{b}})$, we have $f_{\bar{b}}(x_2^{\bar{b}}) > f_{\bar{b}}(0)$. We now show that for all $b > \bar{b}$, $f_b(x_2^b) \ge f_b(0)$. Suppose this is not true and there exists b such that $b > \bar{b}$ and $f_b(x_2^b) < f_b(0)$. Because we have

$$\begin{cases} f_{\bar{b}}(x_2^{\bar{b}}) > f_{\bar{b}}(0), \\ f_b(x_2^b) < f_b(0), \end{cases}$$

by a direct computation, we get

$$\begin{cases} g(x_2^{\bar{b}}) - x_2^{\bar{b}} \nabla g(x_2^{\bar{b}}) - \frac{1}{2}(x_2^{\bar{b}})^2 > 0, \\ g(x_2^b) - x_2^b \nabla g(x_2^b) - \frac{1}{2}(x_2^b)^2 < 0. \end{cases}$$

So, according to the intermediate value theorem, there exists $\tilde{x}$ such that $g(\tilde{x}) - \tilde{x}\nabla g(\tilde{x}) - \frac{1}{2}(\tilde{x})^2 = 0$. Let $\tilde{b} = \nabla g(\tilde{x}) + \tilde{x}$. Then, $(\tilde{x}, \tilde{b} - \tilde{x})$ is the intersection point of $C_{\tilde{b}}(x)$ and $D(x)$ such that $f_{\tilde{b}}(\tilde{x}) = f_{\tilde{b}}(0)$. Since $x_2^{\bar{b}} < \tilde{x} < x_2^b$ and ∇g is convex and nonincreasing, we conclude that $\bar{b} < \tilde{b} < b$, which contradicts $f_{b^*}(0) \ne f_{b^*}(x_2^{b^*})$ for all $b^* > \bar{b}$. So, for all $b > \bar{b}$, 0 is the minimum of $f_b(x)$ on $[0, b]$. Similarly, when $b \le \bar{b}$, we have $\nabla f_b(x) = D(x) - C_b(x) \ge 0$, because $D(x)$ is above $C_b(x)$. So, the global minimum of $f_b(x)$ on $[0, b]$ is 0. The proof is completed. □

1.2 Proof of Proposition 3

Proposition 3. *Given g satisfying **Assumption** 1 and $\nabla g(0) < +\infty$. Restricted on $[0, +\infty)$, if we have $C_{\nabla g(0)}(x) = \nabla g(0) - x \le \nabla g(x)$ for all $x \in (0, \nabla g(0))$, then $C_b(x)$ and $D(x)$ have only one intersection point (x^b, y^b) when $b > \nabla g(0)$. Furthermore,*

$$\mathbf{Prox}_g(b) = \operatorname*{argmin}_{x \ge 0} f_b(x) \begin{cases} = x^b, & \text{if } b > \nabla g(0), \\ \ni 0, & \text{if } b \le \nabla g(0). \end{cases} \tag{3}$$

Suppose there exists $0 < \hat{x} < \nabla g(0)$ such that $C_{\nabla g(0)}(\hat{x}) = \nabla g(0) - \hat{x} > \nabla g(\hat{x})$. Then, when $\nabla g(0) \ge b > \bar{b}$, $C_b(x)$ and $D(x)$ have two intersection points, which are denoted as $P_1^b = (x_1^b, y_1^b)$ and $P_2^b = (x_2^b, y_2^b)$ such that $x_1^b < x_2^b$. When $\nabla g(0) < b$, $C_b(x)$ and $D(x)$ have only one intersection point (x^b, y^b). Also, there exists $\tilde{b}$ such that $\nabla g(0) > \tilde{b} > \bar{b}$ and $f_{\tilde{b}}(0) = f_{\tilde{b}}(x_2^{\tilde{b}})$. Let $b^* = \inf\{b \mid \nabla g(0) > \tilde{b} > \bar{b}, f_b(0) = f_b(x_2^b)\ \}$. We have

$$\mathbf{Prox}_g(b) = \operatorname*{argmin}_{x \ge 0} f_b(x) \begin{cases} = x^b, & \text{if } b > \nabla g(0), \\ = x_2^b, & \text{if } \nabla g(0) \ge b > b^*, \\ \ni 0, & \text{if } b \le b^*. \end{cases} \tag{4}$$

Remark: If b^* exists, when $b \le b^*$, it is possible that $C_b(x)$ and $D(x)$ have more than two intersection points. If b^* does not exist, when $b \le \nabla g(0)$, it is also possible that $C_b(x)$ and $D(x)$ have more than two intersection points.

Proof. Case 1 : Suppose we have $C_{g'(0)}(x) = \nabla g(0) - x \le \nabla g(x)$ for all x on $(0, \nabla g(0))$. Notice for all $b \le \nabla g(0)$, we have $\nabla g(x) = D(x) - C_b(x) \ge 0$, so the minimum point of $f_b(x)$ is 0. For all $b > \nabla g(0)$, $C_b = b - x$ and $D(x)$ have only one intersection point denoted as (x^b, y^b). Then, we can easily see that f_b is decreasing on $(0, x^b)$ and increasing on (x^b, b). So, when $b > \nabla g(0)$, the minimum point of $f_b(x)$ is x^b.

Case 2 : Suppose there exists $0 < \hat{x} < \nabla g(0)$ such that $C_{\nabla g(0)}(\hat{x}) = \nabla g(0) - \hat{x} > \nabla g(\hat{x})$. Then, $D(x)$ and $C_b(x)$ have two intersection points, i.e., $(0, \nabla g(0))$ and $(x_2^{\nabla g(0)}, y_2^{\nabla g(0)})$. It is easily checked that $f_{\nabla g(0)}$ is strictly decreasing on $(0, x_2^{\nabla g(0)})$, so we have $f_{\nabla g(0)}(x_2^{\nabla g(0)}) < f_{\nabla g(0)}(0)$. Also, since $f_{\bar{b}}$ is strictly increasing on $(0, x_2^{\bar{b}})$, we have $f_{\bar{b}}(x_2^{\bar{b}}) > f_{\bar{b}}(0)$.

Because we have

$$\begin{cases} f_{\bar{b}}(x_2^{\bar{b}}) & > & f_{\bar{b}}(0), \\ f_{\nabla g(0)}(x_2^{\nabla g(0)}) & < & f_{\nabla g(0)}(0), \end{cases}$$

by a direct computation, we get

$$\begin{cases} g(x_2^{\bar{b}}) - x_2^{\bar{b}} \nabla g(x_2^{\bar{b}}) - \frac{1}{2}(x_2^{\bar{b}})^2 & > & 0, \\ g(x_2^{\nabla g(0)}) - x_2^{\nabla g(0)} \nabla g(x_2^{\nabla g(0)}) - \frac{1}{2}(x_2^{\nabla g(0)})^2 & < & 0. \end{cases}$$

So, according to the intermediate value theorem, there exists $\tilde{x}$ such that $g(\tilde{x}) - \tilde{x}\nabla g(\tilde{x}) - \frac{1}{2}(\tilde{x})^2 = 0$. Let $\tilde{b} = \nabla g(\tilde{x}) + \tilde{x}$. Then, $(\tilde{x}, \tilde{b} - \tilde{x})$ is the intersection point of $C_{\tilde{b}}(x)$ and $D(x)$ such that $f_{\tilde{b}}(\tilde{x}) = f_{\tilde{b}}(0)$. Since $x_2^{\bar{b}} < \tilde{x} < x_2^{\nabla g(0)}$ and ∇g is convex and nonincreasing, we conclude that $\bar{b} < \tilde{b} < \nabla g(0)$. Next, we set $b^* = \inf\{b \mid \bar{b} < \tilde{b} < \nabla g(0), f_b(0) = f_b(x_2^b)\ \}$.

Given $\nabla g(0) \geq b > \bar{b}$, we can easily see that f_b is increasing on $(0, x_1^b)$, decreasing on (x_1^b, x_2^b) and increasing on (x_2^b, b). So, 0 and x_2^b are two local minimum points of $f_b(x)$ on $[0, b]$.

Next, for $\nabla g(0) \geq b > b^*$, set $b = b^* + \varepsilon$ for some $\varepsilon > 0$. We have

$$\begin{aligned} & f_b(x_2^{b^*}) - f_b(0) \\ =& \frac{1}{2}(x_2^{b^*} - b^* - \varepsilon)^2 + g(x^*) - \frac{1}{2}(b^* + \varepsilon)^2 \\ =& \frac{1}{2}(x_2^{b^*} - b^*)^2 - \frac{1}{2}(b^*)^2 - \varepsilon(x_2^{b^*} - b^*) - \varepsilon b^* \\ =& f_{b^*}(x_2^{b^*}) - f_{b^*}(0) - \varepsilon x_2^* \\ =& -\varepsilon x_2^* < 0. \end{aligned}$$

Since f_b is decreasing on $(x_2^{b^*}, x_2^b)$, we conclude that $f_b(0) > f_b(x_2^{b^*}) \geq f_b(x_2^b)$. So, when $b > b^*$, x_2^b is the global minimum of $f_b(x)$ on $[0, b]$.

Then, for all $\bar{b} < b \leq b^*$, we show that $f_b(0) \leq f_b(x_2^b)$. We prove by contradiction. Suppose that there exists b such that $f_b(0) > f_b(x_2^b)$. Because we have

$$\begin{cases} f_{\bar{b}}(x_2^{\bar{b}}) > f_{\bar{b}}(0), \\ f_b(x_2^b) < f_b(0), \end{cases}$$

by a direct computation, we get

$$\begin{cases} g(x_2^{\bar{b}}) - x_2^{\bar{b}}\nabla g(x_2^{\bar{b}}) - \frac{1}{2}(x_2^{\bar{b}})^2 > 0, \\ g(x_2^b) - x_2^b \nabla g(x_2^b) - \frac{1}{2}(x_2^b)^2 < 0. \end{cases}$$

So, according to the intermediate value theorem, there exists $\tilde{x}_1$ such that $g(\tilde{x}_1) - \tilde{x}_1\nabla g(\tilde{x}_1) - \frac{1}{2}(\tilde{x}_1)^2 = 0$ and $x_2^{\bar{b}} < \tilde{x}_1 < x_2^b$. Let $\tilde{b}_1 = \nabla g(\tilde{x}_1) + \tilde{x}_1$. Then, $(\tilde{x}_1, \tilde{b}_1 - \tilde{x}_1)$ is the intersection point of $C_{\tilde{b}_1}(x)$ and $D(x)$ such that $f_{\tilde{b}_1}(\tilde{x}_1) = f_{\tilde{b}_1}(0)$. Since $x_2^{\bar{b}} < \tilde{x} < x_2^b$ and ∇g is convex and nonincreasing, we conclude that $\bar{b} < \tilde{b} < b \leq b^*$, which contradicts the minimality of b^*.

Next, when $b \leq \bar{b}$, we have $\nabla f_b(x) = D(x) - C_b(x) \geq 0$, so the global minimum of $f_b(x)$ on $[0, b]$ is 0. Also, when $b > \nabla g(0)$, $C_b = b - x$ and $D(x)$ have only one intersection point (x^b, y^b). Then, we can easily see that f_b is decreasing on $(0, x^b)$ and increasing on (x^b, b). So, when $b > \nabla g(0)$, the global minimum point of $f_b(x)$ is x^b. □

1.3 Proof of Corollary 1

Corollary 1. *Given g satisfying **Assumption** 1 in problem (1). Denote $\hat{x}^b = \max\{x|\nabla f_b(x) = 0, 0 \leq x \leq b\}$ and $x^* = \arg\min_{x\in\{0,\hat{x}^b\}} f_b(x)$. Then x^* is optimal to (1), i.e., $x^* \in \mathbf{Prox}_g(b)$.*

Proof. As shown in Proposition 2 and 3, when b is larger than a certain threshold, $\mathbf{Prox}_g(b)$ (x_2^b in (2)(4) or x^b in (3)(4)) is unique. Actually the unique solution is the largest intersection point of $C_b(x)$ and $\nabla g(x)$, i.e., $\mathbf{Prox}_g(b) = \hat{x}^b = \max\{x|\nabla f_b(x) = 0, 0 \leq x \leq b\}$. For all the other choices of b, $0 \in \mathbf{Prox}_g(b)$. Thus, 0 and $\hat{x}^b$, one of them should be optimal to (1). Thus $x^* = \arg\min_{x\in\{0,\hat{x}^b\}} f_b(x)$ is optimal to (1). □

2 Proof of Theorem 2

Theorem 2. *For any lower bounded function g, its proximal operator $\mathbf{Prox}_g(\cdot)$ is monotone, i.e., for any $p_i^* \in \mathbf{Prox}_g(x_i)$, $i = 1, 2$, $p_1^* \geq p_2^*$, when $x_1 > x_2$.*

Proof. The lower bound assumption of g guarantees a finite solution to problem (1). By the optimality of p_i^*, $i = 1, 2$, we have

$$g(p_2^*) + \frac{1}{2}(p_2^* - x_1)^2 \geq g(p_1^*) + \frac{1}{2}(p_1^* - x_1)^2, \tag{5}$$

$$g(p_1^*) + \frac{1}{2}(p_1^* - x_2)^2 \geq g(p_2^*) + \frac{1}{2}(p_2^* - x_2)^2. \tag{6}$$

Summing them together gives

$$(p_2^* - x_1)^2 + (p_1^* - x_2)^2 \geq (p_1^* - x_1)^2 + (p_2^* - x_2)^2. \tag{7}$$

It reduces to

$$(p_1^* - p_2^*)(x_1 - x_2) \geq 0. \tag{8}$$

Thus $p_1^* \geq p_2^*$ when $x_1 > x_2$. □

3 Convergence Analysis of Algorithm 1

Assume there exists

$$\hat{x}^b = \max\{x|\nabla f_b(x) = \nabla g(x) + x - b = 0, 0 \leq x \leq b\};$$

otherwise, 0 is a solution to (1).

We only need to prove that the fixed point iteration guarantees to find $\hat{x}^b$.

First, if $\nabla g(b) = 0$, then we have found $\hat{x}^b = b$.

For the case $\hat{x}^b < b$, we prove that, the fixed point iteration, starting from $x_0 = b$, converges to $\hat{x}^b$. Indeed, we have

$$b - \nabla g(x) < x, \text{ for any } x > \hat{x}^b.$$

We prove this by contradiction. Assume there exists $\tilde{x} > \hat{x}^b$ such that $b - \nabla g(\tilde{x}) > \tilde{x}$. Notice g satisfies Assumption 1. It is easy to see ∇g is continuous, decreasing and nonnegative. Then we have $b - \nabla g(b) < b$ ($\nabla g(b) > 0$ since $b > \hat{x}^b$). Thus there must exist some $\hat{x} \in (\min(b, \tilde{x}), \max(b, \tilde{x})) > \hat{x}^b$ such that $b - g(\hat{x}) = \hat{x}$. This contradicts the definition of $\hat{x}^b$.

So, we have

$$x_{k+1} = b - \nabla g(x_k) < x_k, \text{ if } x_k > \hat{x}^b.$$

On the other hand, $\{x_k\}$ is lower bounded by $\hat{x}^b$. So there must exist a limit of $\{x_k\}$, denoted as $\bar{x}$, which is no less than $\hat{x}^b$. Let $k \to +\infty$ on both sides of

$$x_{k+1} = b - \nabla g(x_k),$$

and we see that $\bar{x} = b - \nabla g(\bar{x})$. So, $\bar{x} = \hat{x}^b$, i.e., $\lim\limits_{k\to+\infty} x_k = \hat{x}^b$.

4 Convergence Analysis of Generalized Proximal Gradient Algorithm

Consider the following problem

$$\min_{\mathbf{X}} F(\mathbf{X}) = \sum_{i=1}^{m} g(\sigma_i(\mathbf{X})) + h(\mathbf{X}), \tag{9}$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, concave and nonincreasing on $[0, +\infty)$, and $h : \mathbb{R}^{m \times n} \to \mathbb{R}^+$ has Lipschitz continuous gradient with Lipschitz constant $L(h)$. The Generalized Proximal Gradient (GPG) algorithm solves the above problem by the following updating rule

$$\begin{aligned} \mathbf{X}^{k+1} &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} g(\sigma_i(\mathbf{X})) + h(\mathbf{X}^k) + \langle \nabla h(\mathbf{X}^k), \mathbf{X} - \mathbf{X}^k \rangle + \frac{\mu}{2} ||\mathbf{X} - \mathbf{X}^k||_F^2 \\ &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} g(\sigma_i(\mathbf{X})) + \frac{\mu}{2} ||\mathbf{X} - \mathbf{X}^k + \frac{1}{\mu} \nabla h(\mathbf{X}^k)||_F^2. \end{aligned} \tag{10}$$

Then we have the following results.

Theorem 3. *If $\mu > L(h)$, the sequence $\{\mathbf{X}^k\}$ generated by (10) satisfies the following properties:*

(1) $F(\mathbf{X}^k)$ is monotonically decreasing. Indeed,

$$F(\mathbf{X}^k) - F(\mathbf{X}^{k+1}) \geq \frac{\mu - L(h)}{2} ||\mathbf{X}^k - \mathbf{X}^{k+1}||_F^2 \geq 0;$$

(2) $\lim_{k \to +\infty} (\mathbf{X}^k - \mathbf{X}^{k+1}) = \mathbf{0}$;

(3) If $F(\mathbf{X}) \to +\infty$ when $||\mathbf{X}||_F \to +\infty$, then any limit point of $\{\mathbf{X}^k\}$ is a stationary point.

Proof. Since $\mathbf{X}^{k+1}$ is optimal to (10), we have

$$\begin{aligned} &\sum_{i=1}^{m} g(\sigma_i(\mathbf{X}^{k+1})) + h(\mathbf{X}^k) + \langle \nabla h(\mathbf{X}^k), \mathbf{X}^{k+1} - \mathbf{X}^k \rangle + \frac{\mu}{2} ||\mathbf{X}^{k+1} - \mathbf{X}^k||_F^2 \\ \leq &\sum_{i=1}^{m} g(\sigma_i(\mathbf{X}^k)) + h(\mathbf{X}^k) + \langle \nabla h(\mathbf{X}^k), \mathbf{X}^k - \mathbf{X}^k \rangle + \frac{\mu}{2} ||\mathbf{X}^k - \mathbf{X}^k||_F^2 \\ = &\sum_{i=1}^{m} g(\sigma_i(\mathbf{X}^k)). \end{aligned} \tag{11}$$

On the other hand, since h has Lipschitz continuous gradient, we have [1]

$$h(\mathbf{X}^{k+1}) \leq h(\mathbf{X}^k) + \langle \nabla h(\mathbf{X}^k), \mathbf{X}^{k+1} - \mathbf{X}^k \rangle + \frac{L(h)}{2} ||\mathbf{X}^{k+1} - \mathbf{X}^k||_F^2. \tag{12}$$

Combining (11) and (12) leads to

$$\begin{aligned} &F(\mathbf{X}^k) - F(\mathbf{X}^{k+1}) \\ = &\sum_{i=1}^{m} g(\sigma_i(\mathbf{X}^k)) + h(\mathbf{X}^k) - \sum_{i=1}^{m} g(\sigma_i(\mathbf{X}^{k+1})) - h(\mathbf{X}^{k+1}) \\ \geq &\frac{\mu - L(h)}{2} ||\mathbf{X}^{k+1} - \mathbf{X}^k||_F^2. \end{aligned} \tag{13}$$

Thus $\mu > L(h)$ guarantees that $F(\mathbf{X}^k) \geq F(\mathbf{X}^{k+1})$.

Summing (13) for $k = 1, 2, \cdots$, we get

$$F(\mathbf{X}^1) \geq \frac{\mu - L(h)}{2} \sum_{k=1}^{+\infty} ||\mathbf{X}^{k+1} - \mathbf{X}^k||_F^2. \tag{14}$$

This implies that

$$\lim_{k \to +\infty} (\mathbf{X}^k - \mathbf{X}^{k+1}) = \mathbf{0}. \tag{15}$$

Furthermore, since $F(\mathbf{X}) \to +\infty$ when $||\mathbf{X}||_F \to +\infty$, $\{\mathbf{X}^k\}$ is bounded. There exist $\mathbf{X}^*$ and a subsequence $\{\mathbf{X}^{k_j}\}$ such that $\lim_{j \to +\infty} \mathbf{X}^{k_j} = \mathbf{X}^*$. By using (15), we get $\lim_{j \to +\infty} \mathbf{X}^{k_j+1} = \mathbf{X}^*$. Considering that $\mathbf{X}^{k_j}$ is optimal to (10), and $-\sum_{i=1}^{m} g(\sigma_i(\mathbf{X}))$ is convex (since g is concave) [3] , there exists $\mathbf{Q}^{k_j+1} \in -\partial \left(-\sum_{i=1}^{m} g(\sigma_i(\mathbf{X}^{k_j+1}))\right)$ such that

$$\mathbf{Q}^{k_j+1} + \nabla h(\mathbf{X}^{k_j}) + \mu(\mathbf{X}^{k_j+1} - \mathbf{X}^{k_j}) = \mathbf{0}. \tag{16}$$

Let $j \to +\infty$ in (16). By the upper semi-continuous property of the subdifferential [2], there exists $\mathbf{Q}^* \in -\partial\left(-\sum_{i=1}^m g(\sigma_i(\mathbf{X}^*))\right)$, such that

$$\mathbf{0} = \mathbf{Q}^* + \nabla h(\mathbf{X}^*) \in \nabla F(\mathbf{X}^*). \tag{17}$$

Thus $\mathbf{X}^*$ is a stationary point to (9). □

References

[1] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2009.

[2] Frank Clarke. Nonsmooth analysis and optimization. In *Proceedings of the International Congress of Mathematicians*, 1983.

[3] Adrian S Lewis and Hristo S Sendov. Nonsmooth analysis of singular values. Part I: Theory. *Set-Valued Analysis*, 13(3):213–241, 2005.