SMOOTHING PROJECTED GRADIENT METHOD AND ITS APPLICATION TO STOCHASTIC LINEAR COMPLEMENTARITY PROBLEMS *

CHAO ZHANG† AND XIAOJUN CHEN‡

Abstract. A smoothing projected gradient (SPG) method is proposed for the minimization problem on a closed convex set, where the objective function is locally Lipschitz continuous but nonconvex, nondifferentiable. We show that any accumulation point generated by the SPG method is a stationary point associated with the smoothing function used in the method, which is a Clarke stationary point in many applications. We apply the SPG method to the stochastic linear complementarity problem (SLCP) and image restoration problems. We study the stationary point defined by the directional derivative and provide necessary and sufficient conditions for a local minimizer of the expected residual minimization (ERM) formulation of SLCP. Preliminary numerical experiments using the SPG method for solving randomly generated SLCP and image restoration problems of large sizes show that the SPG method is promising.

Key words. Smoothing projected gradient method; nonsmooth; nonconvex; constrained optimization; stochastic linear complementarity problem; image restoration

AMS subject classifications. 65K10, 90C26

1. Introduction. The projected gradient (PG) method was originally proposed by Goldstein [16], and Levitin and Polyak [20] in 1960s, for minimizing a continuously differentiable mapping $f : R^n \to R$ on a nonempty closed convex set X. Probably since it is quite simple to implement and attractive for large-scale problems with simple bounds constraints, ever since then, there have been various extensions which make the PG method more widely applicable and more efficient in computation, e.g., [1, 3, 28, 31].

Nonsmooth and nonconvex optimization occurs frequently in practice. The projected subgradient method [30] extends the PG method to the case that f is nonsmooth, but convex. Recently, Burke, Lewis and Overton [2] introduced a robust gradient sampling algorithm for solving nonsmooth, nonconvex unconstrained minimization problem. Kiwiel [19] slightly revised the gradient sampling algorithm in [2] and showed that any accumulation point generated by the algorithm is a Clarke stationary point with probability one.

In this paper, we propose a smoothing projected gradient (SPG) method, which combines the smoothing techniques and the classical PG method to solve the problems of the form

$$\min\{f(x) \ : \ x \in X\}, \tag{1.1}$$

where X is a nonempty closed convex set in R^n, and $f : R^n \to R$ is locally Lipschitzian, but not necessarily differentiable and convex. Many nonsmooth optimization problems are of this type, for instance, the expected residual minimization (ERM) formulation for the SLCP discussed in [6, 9, 14], and the image restoration problems studied in [15, 23]. However, it is hard to find an efficient numerical method to solve

*This work was supported in part by the National Natural Science Foundation of China (70871008), a Hong Kong Research Grant Council, and a HK PolyU grant.

†Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, China (zc.njtu@163.com).

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong (maxjchen@polyu.edu.hk).

(1.1) when n is large. The SPG method is easy to implement. At each iteration, we approximate the objective function by a smooth function with a fixed smoothing parameter, and employ the classical PG method to obtain a new point. If a certain criteria is satisfied, then we update the smoothing parameter using the new point for the next iteration. In comparison with the gradient sampling algorithm [19], we show that any accumulation point generated by the SPG method globally converges to a stationary point associated with the smoothing function used in the method, which is a Clarke stationary point in many applications.

We apply the SPG method to the stochastic linear complementarity problem (SLCP) and image restoration problems.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, where Ω is the set of random vector ω, $\mathcal{F}$ is the set of events, and $\mathcal{P}$ is the probability distribution satisfying $\mathcal{P}\{\Omega\} = 1$. The stochastic complementarity problem SLCP$(M(\omega), q(\omega))$ is defined as

$$x \geq 0,\ M(\omega)x + q(\omega) \geq 0,\ x^T(M(\omega)x + q(\omega)) = 0, \quad \omega \in \Omega. \tag{1.2}$$

Here $M(\omega) \in R^{n\times n}$ and $q(\omega) \in R^n$ are random matrix and random vector for $\omega \in \Omega$, respectively. Througout the paper, we assume $M(\omega)$ and $q(\omega)$ are measurable functions of ω and satisfy

$$E[\|M(\omega)\|^2 + \|q(\omega)\|^2] < \infty, \tag{1.3}$$

where E stands for the expectation.

When Ω is a singleton, SLCP$(M(\omega), q(\omega))$ reduces to the well-known linear complementarity problem LCP(M, q) with $M(\omega) \equiv M$ and $q(\omega) \equiv q$. In general, a deterministic formulation for the SLCP provides optimal solutions for the SLCP in some sense. The ERM formulation proposed in [6] is a deterministic formulation for the SLCP, which is defined as

$$\min_{x\in R^n_+} f(x) := E[\|\Phi(x,\omega)\|^2] \tag{1.4}$$

where

$$\Phi(x,\omega) = (\phi((M(\omega)x + q(\omega))_1, x_1), \ldots, \phi((M(\omega)x + q(\omega))_n, x_n)),$$

and $\phi : R^2 \to R$ is an NCP function, which has the property

$$\phi(a,b) = 0 \ \Leftrightarrow\ a \geq 0,\ b \geq 0,\ ab = 0.$$

The objective function in the ERM formulation (1.4) is neither convex nor smooth. Theoretical analysis including the solvability and the robustness for the ERM formulation has been studied, and preliminary numerical results have been given to show the desirable properties for the solution of the ERM formulation in [6, 9, 14]. Among various NCP functions, the "min" function

$$\phi(a,b) := \min(a,b), \quad \text{for any } (a,b) \in R^2, \tag{1.5}$$

has various nice properties for (1.4). It is shown in Lemma 2.2 [9] that the ERM formulation defined by the "min" function always has a solution if $\Omega = \{\omega^1, \omega^2, \ldots, \omega^N\}$ is a finite set. However, the ERM formulation defined by the Fischer-Burmister NCP function is not always solvable. In this paper, we concentrate on the ERM formulation defined by the "min" function, which can be expressed as

$$\min_{x\in R^n_+} f(x) := E[\|\min(x, M(\omega)x + q(\omega))\|^2]. \tag{1.6}$$

This is a nonsmooth, nonconvex constrained minimization problem.

This paper is organized as follows. In Section 2, we give definition of smoothing functions and present the SPG method for solving nonsmooth, nonconvex minimization problem on a closed convex feasible set. We show that any accumulation point generated by the SPG method is a stationary point of problem (1.1) associated with the smoothing function used in the method, which is a Clarke stationary point in many applications.

In Section 3, we consider the application of the SPG method to the problem (1.6). We establish a necessary and sufficient condition for f to be differentiable at a given point $x \in R^n_+$. We show convergence of the SPG method using the class of the Chen-Mangasarian smoothing function. Moreover, we study standard stationary point defined by the directional derivative of f. In Section 4, we illustrate the SPG method by numerical examples of the ERM formulation (1.6) and the image restoration problems. Numerical results demonstrate that the SPG method is promising.

Throughout the paper, $\|\cdot\|$ represents the Euclidean norm and $R^n_{++} = \{x \in R^n \ : \ x > 0\}$. Let $\mathbb{N}$ be the set of all natural numbers ν, and $\mathcal{N}^\sharp_\infty$ be the infinite subsets of $\mathbb{N}$. We use the notation $\underset{N}{\to}$ to denote the convergence indexed by $N \in \mathcal{N}^\sharp_\infty$. I denotes the identity matrix. For a given matrix $A = [a_{ij}] \in R^{n\times n}$, let $A_{i.}$ be the i-th row of A. For a given subset $\hat\Omega$ of Ω and a function $s : \Omega \to R$, we use $E_{\hat\Omega}[s(\omega)]$ to represent $E[s(\omega)1_{\{\omega\in\hat\Omega\}}]$, where $1_{\{\omega\in\hat\Omega\}}$ is the indicator function of the set $\hat\Omega$, which is equal to 1 if $\omega \in \hat\Omega$ and 0 if $\omega \in \Omega \setminus \hat\Omega$.

2. Smoothing projected gradient method. In this section, we present a smoothing projected gradient method for solving the minimization problem (1.1), where the objective function f is a general locally Lipschitz continuous function.

Let $P[\cdot]$ denote the orthogonal projection from R^n into X,

$$P[x] = \operatorname{argmin}\{\|z - x\| \ : \ z \in X\}.$$

Definition 2.1. *Let $f : R^n \to R$ be a locally Lipschitz continuous function. We call $\tilde f : R^n \times R_+ \to R$ a smoothing function of f, if $\tilde f(\cdot, \mu)$ is continuously differentiable in R^n for any $\mu \in R_{++}$, and for any $x \in R^n$,*

$$\lim_{z\to x,\ \mu\downarrow 0} \tilde f(z,\mu) = f(x) \tag{2.1}$$

and $\{\lim_{z\to x,\ \mu\downarrow 0} \nabla_x \tilde f(z,\mu)\}$ is nonempty and bounded.

The smoothing projected gradient method is defined as follows.

Algorithm 2.1 (Smoothing projected gradient algorithm)
Let $\hat\gamma$, γ_1 and γ_3 be positive constants, where $\gamma_1 << \gamma_3$. Let γ_2, σ, σ_1 and σ_2 be constants in $(0, 1)$, where $\sigma_1 \le \sigma_2$. Choose $x^0 \in X$ and $\mu_0 \in R_{++}$. For $k \ge 0$:

1. If $\|P[x^k - \nabla_x \tilde f(x^k, \mu_k)] - x^k\| = 0$, let $x^{k+1} = x^k$ and go to step 3. Otherwise, go to step 2.

2. (PG method)
Let $y^{0,k} = x^k$. For $j \ge 0$:

$$y^{j,k}(\alpha) = P[y^{j,k} - \alpha\nabla_x \tilde f(y^{j,k}, \mu_k)],$$

and $y^{j+1,k} = y^{j,k}(\alpha_{j,k})$ where $\alpha_{j,k}$ is chosen so that,

$$\tilde f(y^{j+1,k}, \mu_k) \le \tilde f(y^{j,k}, \mu_k) + \sigma_1(\nabla_x \tilde f(y^{j,k}, \mu_k), y^{j+1,k} - y^{j,k}) \tag{2.2}$$

and

$$\gamma_3 \geq \alpha_{j,k} \geq \gamma_1, \quad \text{or} \quad \alpha_{j,k} \geq \gamma_2 \bar{\alpha}_{j,k} > 0, \tag{2.3}$$

such that $\bar{y}^{j+1,k} = y^{j,k}(\bar{\alpha}_{j,k})$ satisfies

$$\tilde{f}(\bar{y}^{j+1,k}, \mu_k) > \tilde{f}(y^{j,k}, \mu_k) + \sigma_2(\nabla_x \tilde{f}(y^{j,k}, \mu_k), \bar{y}^{j+1,k} - y^{j,k}). \tag{2.4}$$

If $\frac{\|y^{j+1,k}-y^{j,k}\|}{\alpha_{j,k}} < \hat{\gamma}\mu_k$, set $x^{k+1} = y^{j+1,k}$ and go to step 3.

3. Choose $\mu_{k+1} \leq \sigma\mu_k$.

The smoothing projected gradient algorithm is well-defined. Note that

$$\|P[x^k - \nabla_x \tilde{f}(x^k, \mu_k)] - x^k\| = 0$$

if and only if x^k is a stationary point of

$$\min\{\tilde{f}(x, \mu_k) \; : \; x \in X\}, \tag{2.5}$$

that is, x^k satisfies

$$(\nabla_x \tilde{f}(x^k, \mu_k), x^k - z) \leq 0 \quad \text{for any} \quad z \in X.$$

If x^k is not a stationary point of (2.5), then from the continuous differentiability of $\tilde{f}(\cdot, \mu_k)$ and analysis in [12], the function $g : R_{++} \to R$ defined by

$$g(\alpha) := \frac{\tilde{f}(x^k, \mu_k) - \tilde{f}(x^k(\alpha), \mu_k)}{(\nabla_x \tilde{f}(x^k, \mu_k), x^k - x^k(\alpha))}$$

is continuous and satisfies

$$\lim_{\alpha \to 0+} g(\alpha) = 1,$$

which implies that (2.2) holds for all $\alpha_{j,k}$ sufficiently small. If there is no $\alpha_{j,k} \in [\gamma_1, \gamma_3]$ such that (2.2) holds, then there exist constants $\hat{\alpha}, \tilde{\alpha} \in (0, \gamma_1)$, $\hat{\alpha} < \tilde{\alpha}$ such that

$$g(\alpha) \geq \sigma_2 \text{ for any } \alpha \in (0, \hat{\alpha}], \quad \text{and} \quad g(\alpha) < \sigma_2 \text{ for any } \alpha \in (\hat{\alpha}, \tilde{\alpha}).$$

Thus it is easy to check that $\alpha_{j,k} = \gamma_2 \bar{\alpha}_{j,k}$ satisfies (2.2)-(2.4) for any

$$\bar{\alpha}_{j,k} \in \Big(\hat{\alpha}, \hat{\alpha} + \min(\frac{1-\gamma_2}{2\gamma_2}\hat{\alpha}, \tilde{\alpha} - \hat{\alpha})\Big).$$

Let us state some basic properties about Algorithm 2.1.

LEMMA 2.2. *[3] Let $P[\cdot]$ be the projection from R^n into X. If $z \in X$, then*

$$(P[x] - x, z - P[x]) \geq 0 \quad \textit{for all} \quad x \in R^n.$$

Set $x = y^{j,k} - \alpha \nabla_x \tilde{f}(y^{j,k}, \mu_k)$ and $z = y^{j,k}$ in Lemma 2.2. We obtain

$$(\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j,k}(\alpha) - y^{j,k}) \leq -\frac{\|y^{j,k}(\alpha) - y^{j,k}\|^2}{\alpha} \quad \text{for} \quad \alpha > 0. \tag{2.6}$$

Given constants $\mu > 0$ and $\Gamma > 0$, let us denote the level set

$$L_{\mu,\Gamma} = \{x \mid \tilde{f}(x, \mu) \leq \Gamma\}.$$

ASSUMPTION 2.1. $\tilde{f}(\cdot, \mu)$ is bounded below on X, and $\nabla_x \tilde{f}(\cdot, \mu)$ is uniformly continuous on $L_{\mu,\Gamma}$, for any $\mu > 0$ and $\Gamma > 0$.

LEMMA 2.3. *Under Assumption 2.1, we have*

$$\lim_{k\to\infty} \mu_k = 0.$$

Proof. For any fixed $\mu > 0$, $\tilde{f}(\cdot, \mu)$ is continuously differentiable, and (2.2)-(2.4) coincide with the classical projected gradient method.

From the continuous differentiability of $\tilde{f}(\cdot, \mu_k)$, we have for each fixed k,

$$\begin{aligned}&|\tilde{f}(\bar{y}^{j+1,k}, \mu_k) - \tilde{f}(y^{j,k}, \mu_k) - (\nabla_x \tilde{f}(y^{j,k}, \mu_k), \bar{y}^{j+1,k} - y^{j,k})|\\ &\quad = |(\nabla_x \tilde{f}(\theta \bar{y}^{j+1,k} + (1-\theta) y^{j,k}), \mu_k) - \nabla_x \tilde{f}(y^{j,k}, \mu_k), \bar{y}^{j+1,k} - y^{j,k})|\end{aligned}$$

for some $\theta \in [0, 1]$. Since $\{\tilde{f}(y^{j,k}, \mu_k)\}$ is a nonincreasing sequence, there must exist a constant $\Gamma > 0$ such that

$$\tilde{f}(\theta \bar{y}^{j+1,k} + (1-\theta) y^{j,k}, \mu_k) \leq \Gamma,$$

provided that $\|\bar{y}^{j+1,k} - y^{j,k}\| \to 0$. Thus the uniform continuity of $\nabla_x \tilde{f}(\cdot, \mu_k)$ on $L_{\mu_k,\Gamma}$ guaranteed by Assumption 2.1 implies that

$$|\tilde{f}(\bar{y}^{j+1,k}, \mu_k) - \tilde{f}(y^{j,k}, \mu_k) - (\nabla_x \tilde{f}(y^{j,k}, \mu_k), \bar{y}^{j+1,k} - y^{j,k})| = o(\|\bar{y}^{j+1,k} - y^{j,k}\|).$$

Following the proof of Theorem 2.3 [3], we obtain

$$\lim_{j\to\infty} \frac{\|y^{j+1,k} - y^{j,k}\|}{\alpha_{j,k}} = 0,$$

and hence $\lim_{k\to\infty} \mu_k = 0$. □

Let D_f be the subset of R^n where f is differentiable. According to Theorem 2.5.1 [10], the Clarke generalized gradient is defined by

$$\partial f(x) = \mathrm{co}\{\lim \nabla f(x^i) \ : \ x^i \to x,\ x^i \in D_f\}, \tag{2.7}$$

where "co" represents the convex hull.

DEFINITION 2.4. *We say that x^* is a Clarke stationary point of (1.1) if there is $V \in \partial f(x^*)$ such that*

$$(V, x^* - z) \leq 0 \quad \textit{for all } z \in X. \tag{2.8}$$

If f is continuously differentiable, the Clarke stationary point reduces to the stationary point for smooth optimization on a convex set. Moreover, if f is a convex function, then x^* is a Clarke stationary point if and only if x^* is a global optimal solution of problem (1.1).

For any fixed $\bar{x} \in X$, denote

$$G_{\tilde{f}}(\bar{x}) := \{V \ : \ \exists N \in \mathcal{N}_\infty^\sharp,\ x^\nu \underset{N}{\to} \bar{x},\ \mu_\nu \downarrow 0 \quad \text{with} \quad \nabla_x \tilde{f}(x^\nu, \mu_\nu) \underset{N}{\to} V\}. \tag{2.9}$$

From Definition 2.1, it is clear that $G_{\tilde{f}}(\bar{x})$ is a nonempty and bounded set. By Theorem 9.61 and (b) of Corollary 8.47 in [26], we have

$$\partial f(\bar{x}) \subseteq \text{co}G_{\tilde{f}}(\bar{x}).$$

In many cases the Clarke generalized gradient coincides with $\text{co}G_{\tilde{f}}(\bar{x})$. For instance, we consider the smoothing function $\tilde{f}$ constructed by the convolution in [26].

Let $\psi^\mu : R^n \to R_+$ satisfy $\int_{R^n} \psi^\mu(z)dz = 1$ and $B^\mu = \{z \,:\, \psi^\mu(z) > 0\}$ converging to $\{0\}$ as $\mu \downarrow 0$. In some place, ψ^μ is called a mollifier. Define the smoothing function $\tilde{f} : R^n \times R_{++} \to R$ by

$$\tilde{f}(x, \mu) := \int_{R^n} f(x - z)\psi^\mu(z)dz.$$

Thus by employing Theorem 9.67 [26], we know that

$$\text{co}G_{\tilde{f}}(x) = \partial f(x) \quad \text{for any } x \in R^n.$$

In this case,

$$G_{\tilde{f}}(x) \subseteq \partial f(x), \quad \text{for any } x \in R^n.$$

Definition 2.5. *We say that x^* is a stationary point of (1.1) associated with a smoothing function $\tilde{f}$, if there exists $V \in G_{\tilde{f}}(x^*)$ such that*

$$(V, x^* - z) \le 0 \quad \textit{for all } z \in X. \tag{2.10}$$

Theorem 2.6. *Any accumulation point x^* of $\{x^k\}$ generated by Algorithm 2.1 with a smoothing function $\tilde{f}$ is a stationary point of (1.1) associated with $\tilde{f}$.*

Proof. If there exists $K \in \mathcal{N}^\sharp_\infty$ such that for each $k \in K$, $\|P[x^k - \nabla_x \tilde{f}(x^k, \mu_k)] - x^k\| = 0$, that is, x^k is a stationary point of (2.5), and $\lim_{k\to\infty,\ k\in K} x^k = x^*$, then we have for any $k \in K$,

$$(\nabla_x \tilde{f}(x^k, \mu_k), x^k - z) \le 0 \quad \text{for any } z \in X. \tag{2.11}$$

By Definition 2.1, there exists an infinite subsequence $K_1 \subseteq K$ such that

$$\lim_{k\to\infty,\ k\in K_1} \nabla_x \tilde{f}(x^k, \mu_k) = V \in G_{\tilde{f}}(x^*),$$

which, combines with (2.11), yields that

$$(V, x^* - z) \le 0 \quad \text{for any} \quad z \in X.$$

Otherwise there exists $\hat{K} \in \mathcal{N}^\sharp_\infty$ such that for each $k \in \hat{K}$, $\|P[x^k - \nabla_x \tilde{f}(x^k, \mu_k)] - x^k\| \ne 0$, and $\lim_{k\to\infty,\ k\in\hat{K}} x^k = x^*$. Let us denote $K = \{k - 1 \mid k \in \hat{K}\}$ and thus

$$\lim_{k\to\infty,\ k\in K} x^{k+1} = x^* \quad \text{and} \quad \{x^{k+1}\}_{k\in K} = \{y^{j+1,k}\}_{k\in K}.$$

By step 2 of Algorithm 2.1 and Lemma 2.3, we have

$$\lim_{k\to\infty,\ k\in K} \frac{\|y^{j+1,k} - y^{j,k}\|}{\alpha_{j,k}} \le \lim_{k\to\infty,\ k\in K} \hat{\gamma}\mu_k = 0, \tag{2.12}$$

which, together with $\alpha_{j,k} \leq \gamma_3$, implies

$$\lim_{k\to\infty,\ k\in K} \|y^{j+1,k} - y^{j,k}\| = 0. \tag{2.13}$$

From (2.6), we have

$$(\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j+1,k} - y^{j,k}) \leq -\frac{\|y^{j+1,k} - y^{j,k}\|^2}{\alpha_{j,k}}, \quad k \in K.$$

This, together with (2.2), implies that for $k \in K$,

$$\begin{aligned} \tilde{f}(y^{j+1,k}, \mu_k) - \tilde{f}(y^{j,k}, \mu_k) &\leq \sigma_1 (\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j+1,k} - y^{j,k}) \\ &\leq -\sigma_1 \frac{\|y^{j+1,k} - y^{j,k}\|^2}{\alpha_{j,k}}. \end{aligned} \tag{2.14}$$

From $\lim\limits_{k\to\infty,\ k\in K} y^{j+1,k} = x^*$ and (2.13), we have $\lim\limits_{k\to\infty,\ k\in K} y^{j,k} = x^*$. By Definition 2.1, we know that

$$|\tilde{f}(y^{j+1,k}, \mu_k) - \tilde{f}(y^{j,k}, \mu_k)| \to 0, \quad \text{as } k \to \infty,\ k \in K.$$

This, together with (2.14), yields

$$\lim_{k\to\infty,\ k\in K} (\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j+1,k} - y^{j,k}) = 0. \tag{2.15}$$

For any $z \in X$, by using Lemma 2.2, we have for $k \in K$,

$$\begin{aligned} \alpha_{j,k}(\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j+1,k} - z) &\leq (y^{j+1,k} - y^{j,k}, z - y^{j+1,k}) \\ &\leq (y^{j+1,k} - y^{j,k}, z - y^{j,k}) \\ &\leq \|y^{j+1,k} - y^{j,k}\| \|y^{j,k} - z\|. \end{aligned}$$

Hence we have

$$(\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j,k} - z) \leq (\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j,k} - y^{j+1,k}) + \frac{\|y^{j+1,k} - y^{j,k}\|}{\alpha_{j,k}} \|y^{j,k} - z\|,$$

which, combined with (2.12) and (2.15), implies that

$$\limsup_{k\to\infty,\ k\in K} (\nabla_x \tilde{f}(y^{j,k}, \mu_k), y^{j,k} - z) \leq 0 \quad \text{for any } z \in X. \tag{2.16}$$

By Definition 2.1, there exists an infinite subsequence $K_1 \subseteq K$ such that

$$\lim_{k\to\infty,\ k\in K_1} \nabla_x \tilde{f}(y^{j,k}, \mu_k) = V \in G_{\tilde{f}}(x^*), \tag{2.17}$$

which, together with (2.16), yields that

$$(V, x^* - z) \leq 0 \quad \text{for any } z \in X.$$

Hence x^* is a stationary point of (1.1) associated with $\tilde{f}$. □

REMARK 2.1. We develop a global convergent algorithm for solving nonsmooth, nonconvex constrained optimization, by applying smoothing functions in the PG method. Algorithm 2.1 and Theorem 2.6 generalize the PG method and its convergence theorem [Theorem 2.3, 3] for continuously differentiable optimization to nonsmooth, nonconvex optimization.

3. ERM formulation for SLCP. In this section, we show that the SPG method can be applied to find a local minimizer of the ERM formulation (1.6) for SLCP. In particular, we give computable smoothing functions and show all assumptions in Section 2 hold. First we consider the functions

$$H(x) := \min(x, Mx + q) \quad \text{and} \quad \theta(x) := \frac{1}{2}H(x)^T H(x). \tag{3.1}$$

For an arbitrary vector $x \in R^n$, define the index sets

$$\begin{aligned} \alpha(x) &= \{i \ : \ x_i > (Mx+q)_i\} \\ \beta(x) &= \{i \ : \ x_i = (Mx+q)_i\} \\ \gamma(x) &= \{i \ : \ x_i < (Mx+q)_i\}. \end{aligned} \tag{3.2}$$

LEMMA 3.1. *(Proposition 5.8.4 [11])*
(i) The function θ is everywhere directionally differentiable and the directional derivative of θ at $x \in R^n$ along the direction $d \in R^n$ is given by

$$\theta'(x, d) = \sum_{i\in\alpha(x)} (Mx+q)_i(Md)_i + \sum_{i\in\gamma(x)} x_i d_i + \sum_{i\in\beta(x)} x_i \min(d_i, (Md)_i). \tag{3.3}$$

(ii) The function θ is differentiable at $x \in R^n$ if

$$x_i = 0; \quad or \quad M_{i.} = I_{i.}, \quad for\ any \quad i \in \beta(x). \tag{3.4}$$

(iii) The function H is differentiable at $x \in R^n$ if and only if $M_{i.} = I_{i.}$ for each $i \in \beta(x)$.

Now we show that (3.4) is not only a sufficient condition for θ being differentiable at $x \in R^n_+$, but also a necessary condition.

LEMMA 3.2. *If the function θ is differentiable at $x \in R^n_+$, then (3.4) holds.*

Proof. Since θ is differentiable at x, $\theta'(x, d)$ is a linear function in d. Hence for any $\lambda \in R$, $\theta'(x, \lambda d) = \lambda\theta'(x, d)$. By noting that the first two parts of the sum (3.3) in (i) of Lemma 3.1 are both linear in d, we have

$$\sum_{i\in\beta(x)} x_i[\min(\lambda d_i, \lambda(Md)_i) - \lambda(\min(d_i, (Md)_i))] = 0. \tag{3.5}$$

Moreover, it is easy to show that

$$\min(\lambda a, \lambda b) \le \lambda \min(a, b), \quad \forall \lambda, a, b \in R, \tag{3.6}$$

where the equality holds if and only if $\lambda \ge 0$, or $\lambda < 0$ and $a = b$. This together with (3.5) and $x \in R^n_+$ yields

$$x_i[\min(\lambda d_i, \lambda(Md)_i) - \lambda \min(d_i, (Md)_i)] = 0, \quad i \in \beta(x).$$

Therefore, for any $i \in \beta(x)$, either $x_i = (Mx + q)_i = 0$, or

$$\min(\lambda d_i, \lambda(Md)_i) - \lambda \min(d_i, (Md)_i) = 0, \quad \forall \lambda \in R, \ d \in R^n.$$

The latter case implies $d_i - (Md)_i = 0$ for any $d \in R^n$, which is equivalent to the fact that the i-th row of M is equal to the i-th unit vector by the arbitrariness of $d \in R^n$, i.e., $M_{i.} = I_{i.}$. Hence (3.4) holds. $\square$

LEMMA 3.3. *The function $\theta'(x,d)$ is concave in d for any fixed $x \in R^n_+$.*

Proof. Note that for any $a_1, a_2, b_1, b_2 \in R$,

$$\min(a_1 + a_2, b_1 + b_2) \geq \min(a_1, b_1) + \min(a_2, b_2),$$

and the equality in (3.6) holds for any $\lambda \in R_+$. Hence for every $i \in \beta(x)$, any $\eta_1, \eta_2 \in R_+$, and any two directions $d^1, d^2 \in R^n$,

$$\min\Big(\eta_1 d^1_i + \eta_2 d^2_i, \eta_1 (Md^1)_i + \eta_2 (Md^2)_i\Big) \geq \eta_1 \min\Big(d^1_i, (Md^1)_i\Big) + \eta_2 \min\Big(d^2_i, (Md^2)_i\Big).$$

Since $x \in R^n_+$, and the first two terms in the sum (3.3) for the directional derivative are both linear in d, we get

$$\theta'(x, \eta_1 d^1 + \eta_2 d^2) \geq \eta_1 \theta'(x, d^1) + \eta_2 \theta'(x, d^2). \tag{3.7}$$

Thus for any $\lambda \in (0, 1)$ and any two directions $d^1, d^2 \in R^n$,

$$\theta'(x, \lambda d^1 + (1-\lambda) d^2) \geq \lambda \theta'(x, d^1) + (1-\lambda) \theta'(x, d^2),$$

which implies that $\theta'(x,d)$ is concave in d for any fixed $x \in R^n_+$. ☐

3.1. Differentiability. Now we consider the objective function f of the ERM formulation of SLCP$(M(\omega), q(\omega))$. For an arbitrary $x \in R^n$, we define $H_\omega(x)$, $\theta_\omega(x)$, $\alpha_\omega(x)$, $\beta_\omega(x)$, $\gamma_\omega(x)$ by adding the subscript ω to $H(x)$ and $\theta(x)$ in (3.1), $\alpha(x)$, $\beta(x)$ and $\gamma(x)$ in (3.2) when M and q are replaced by $M(\omega)$ and $q(\omega)$, respectively. Thus the ERM formulation (1.6) of SLCP$(M(\omega), q(\omega))$ can be expressed by

$$\min_{x \in R^n_+} f(x) := 2E[\theta_\omega(x)]. \tag{3.8}$$

From Chapter 2 in [27], the expectation function $g(x) := E[G_\omega(x)]$ for any function $G_\omega : R^n \times \Omega \to R$, inherits various properties of the integrand $G_\omega(x)$ as stated in Lemma 3.4.

LEMMA 3.4. *[27] Suppose that for a fixed $x \in R^n$: (i) $G_\omega(x)$ is measurable and satisfies $E[|G_\omega(x)|] < \infty$, (ii) there exists a random variable $z_\omega(x)$ such that $E[z_\omega(x)] < \infty$, and for all x^1, x^2 in a neighborhood of x,*

$$|G_\omega(x^1) - G_\omega(x^2)| \leq z_\omega(x) \|x^1 - x^2\| \quad \text{for} \quad \omega \in \Omega \text{ a.e.},$$

(iii) G_ω is directionally differentiable at x for $\omega \in \Omega$ a.e.. Then g is locally Lipschitz continuous, everywhere directionally differentiable at x, and

$$g'(x,d) = E[G'_\omega(x,d)] \quad \text{for all } d.$$

Moreover, if G_ω is differentiable at x for $\omega \in \Omega$ a.e., then g is differentiable at x and

$$\nabla g(x) = E[\nabla G_\omega(x)].$$

In addition, if $G'_\omega(x,d)$ is convex in d for $\omega \in \Omega$ a.e., then g is differentiable at x if and only if G_ω is differentiable at x for $\omega \in \Omega$ a.e..

PROPOSITION 3.5. *The function f is locally Lipschitz continuous, and everywhere directionally differentiable with*

$$f'(x,d) = 2E[\theta'_\omega(x,d)] \quad \text{for all } d. \tag{3.9}$$

If the following condition holds at $x \in R^n$,

$$x_i = 0; \quad or \quad (M(\omega))_{i\cdot} = I_{i\cdot}, \quad for\ any\ i \in \beta_\omega(x),\ \omega \in \Omega\ a.e., \tag{3.10}$$

then f is differentiable at x and

$$\nabla f(x) = 2E[\nabla \theta_\omega(x)]. \tag{3.11}$$

Moreover, f is differentiable at $x \in R^n_+$ if and only if (3.10) holds.

Proof. For an arbitrary $x \in R^n$, it is obvious that $\theta_\omega(x)$ is measurable, and $E[|\theta_\omega(x)|] < \infty$ by (1.3). Moreover, it is known that for any fixed $\omega \in \Omega$, $H_\omega(\cdot)$ is globally Lipschitzian [11]. In particular,

$$\|H_\omega(y) - H_\omega(z)\| \le (1 + \|M(\omega)\|)\|y - z\| \quad \text{for all} \quad y, z \in R^n,$$

and for any constant $r > 0$ and $\tilde{x} \in B(x, r) := \{\tilde{x} \ : \ \|\tilde{x} - x\| \le r\}$,

$$\|H_\omega(\tilde{x})\| \le (1 + \|x\| + r)(1 + \|M(\omega)\| + \|q(\omega)\|).$$

The above two inequalities imply that for any $\omega \in \Omega$, $\theta_\omega(\cdot)$ is locally Lipschitzian with

$$\begin{aligned}|\theta_\omega(x^1) - \theta_\omega(x^2)| &\le \frac{1}{2}(\|H_\omega(x^1)\| + \|H_\omega(x^2)\|)(\|H_\omega(x^1) - H_\omega(x^2)\|)\\ &\le z_\omega(x)\|x^1 - x^2\| \quad \text{for any } x^1, x^2 \in B(x, r),\end{aligned}$$

where $z_\omega(x) = (1 + \|x\| + r)(1 + \|M(\omega)\| + \|q(\omega)\|)^2$ satisfying $E[z_\omega(x)] < \infty$ by (1.3). From Lemma 3.1, θ_ω is everywhere directionally differentiable for all $\omega \in \Omega$, and if (3.10) holds, then θ_ω is differentiable at x for $\omega \in \Omega$ a.e. Hence, we get (3.9) and (3.11) from Lemma 3.4.

Now we only need to show that if f is differentiable at $x \in R^n_+$, then (3.10) holds. According to Lemma 3.3, $\theta'_\omega(x, d)$ is a concave function of $d \in R^n$ for $x \in R^n_+$. That is, for any $\lambda \in (0, 1)$ and any two directions $d^1, d^2 \in R^n$,

$$\theta'_\omega(x, \lambda d^1 + (1 - \lambda)d^2) \ge \lambda\theta'_\omega(x, d^1) + (1 - \lambda)\theta'_\omega(x, d^2).$$

Hence the function $(-\theta_\omega)'(x, d) = -\theta'_\omega(x, d)$ is convex in d. From Lemma 3.4, we know that $-f(x) = 2E[-\theta_\omega(x)]$ is differentiable at $x \in R^n_+$ if and only if $-\theta_\omega$ is differentiable at x for $\omega \in \Omega$ a.e. Note that if f is differentiable at $x \in R^n_+$, then $-f$ is differentiable at x, and hence θ_ω is differentiable at x for $\omega \in \Omega$ a.e.. Thus by Lemma 3.2, the condition (3.10) holds. The proof is completed. □

REMARK 3.1. Proposition 3.5 provides a sufficient condition for f being differentiable at $x \in R^n$, which includes Theorem 4.3 [14] as a special case. Moreover, the sufficient condition is also a necessary condition for f being differentiable at $x \in R^n_+$.

3.2. Smoothing function for ERM. Now we show that smoothing functions $\tilde{f}$ derived from the Chen-Mangasarian smoothing function [5] satisfy Definition 2.1. Let $\rho : R \to [0, \infty)$ be a piecewise continuous density function satisfying

$$\rho(s) = \rho(-s) \quad \text{and} \quad \kappa := \int_{-\infty}^{\infty} |s|\rho(s)ds < \infty. \tag{3.12}$$

The Chen-Mangasarian family of smoothing approximation for the "min" function

$$\min(a, b) = a - \max(0, a - b)$$

is built as

$$\phi(a,b,\mu) = a - \int_{-\infty}^{\infty} \max(0, a-b-\mu s)\rho(s)ds. \tag{3.13}$$

It is worth mentioning that $\phi(a,b,\mu)$ is easy to compute, if some concrete density function is chosen. For instance, if we use the uniform density function

$$\rho(s) = \begin{cases} 1 & \text{if } -\frac{1}{2} \le s \le \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{3.14}$$

we get

$$\phi(a,b,\mu) = \begin{cases} b & \text{if } a-b \ge \frac{\mu}{2}, \\ a - \frac{1}{2\mu}(a-b+\frac{\mu}{2})^2 & \text{if } -\frac{\mu}{2} < a-b < \frac{\mu}{2}, \\ a & \text{if } a-b \le -\frac{\mu}{2}. \end{cases}$$

We refer to [7, 8] for other easily computed ϕ with concrete density functions.

Employing (3.13) to f, we obtain the smoothing function $\tilde{f}$

$$\tilde{f}(x,\mu) = 2E[\tilde{\theta}_\omega(x,\mu)], \tag{3.15}$$

where $\tilde{\theta}_\omega : R^n \times R_{++} \to R$ is defined by

$$\tilde{\theta}_\omega(x,\mu) = \frac{1}{2}\tilde{H}_\omega(x,\mu)^T \tilde{H}_\omega(x,\mu),$$

and $\tilde{H}_\omega : R^n \times R_{++} \to R^n$ is given by

$$\tilde{H}_\omega(x,\mu) = \begin{pmatrix} \phi(x_1, (M(\omega)x + q(\omega))_1, \mu) \\ \vdots \\ \phi(x_n, (M(\omega)x + q(\omega))_n, \mu) \end{pmatrix}.$$

It is easy to see that for any $x \in R^n$ and $\mu \in R_{++}$,

$$\nabla_x \tilde{f}(x,\mu) = 2E[\nabla_x \tilde{\theta}_\omega(x,\mu)] = 2E[\nabla_x \tilde{H}_\omega(x,\mu)\tilde{H}_\omega(x,\mu)], \tag{3.16}$$

where for each $i = 1, 2, \ldots, n$,

$$(\nabla_x \tilde{H}_\omega(x,\mu)^T)_{i.} = I_{i.} - [I_{i.} - (M(\omega))_{i.}] \int_{-\infty}^{\frac{x_i - (M(\omega)x + q(\omega))_i}{\mu}} \rho(s)ds. \tag{3.17}$$

Let $\partial H_\omega(x)$ be the generalized Jacobian of H_ω at x defined by Clarke [10].

LEMMA 3.6. *Denote $\eta = \sqrt{n}\kappa$. For any $\omega \in \Omega$ and $\mu \in R_{++}$,*
(i) $\|\tilde{H}_\omega(x,\mu) - H_\omega(x)\| \le \eta\mu$, $x \in R^n$.
(ii) $\lim_{\mu \downarrow 0} (\nabla_x \tilde{H}_\omega(x,\mu))^T = \tilde{H}^o_\omega(x) \in \partial H_\omega(x)$, $x \in R^n$.
(iii) $\lim_{\mu \downarrow 0} \nabla_x \tilde{f}(x,\mu) = \tilde{f}^o(x) \in \partial f(x)$, $x \in R^n_+$.

Proof. The statement (i) comes from Proposition 2.1 (i) in reference [8]. Since $\rho(s) = \rho(-s)$, we get from Proposition 2.1 (iii) that for any $x \in R^n$,

$$\lim_{\mu \downarrow 0} \nabla_x \tilde{H}_\omega(x,\mu)^T = \tilde{H}^o_\omega(x),$$

where

$$(\tilde{H}^o_\omega(x))_{i\cdot} = \begin{cases} (M(\omega))_{i\cdot} & \text{if } i \in \alpha_\omega(x), \\ I_{i\cdot} & \text{if } i \in \gamma_\omega(x), \\ \frac{1}{2}[(M(\omega))_{i\cdot} + I_{i\cdot}] & \text{if } i \in \beta_\omega(x). \end{cases} \tag{3.18}$$

Now we show the inclusion $\tilde{H}^o_\omega(x) \in \partial H_\omega(x)$ in statement (ii) holds. Consider an arbitrary $\hat{x} \in R^n$. Let D_{H_ω} be the set of points in R^n where H_ω admits differentiability. Since H_ω is locally Lipschitzian in R^n, it is differentiable almost everywhere. Hence there exists an infinite sequence $\{x^k\} \subset D_{H_\omega}$ converging to $\hat{x}$. It is known [10] that

$$\partial H_\omega(\hat{x}) = \mathrm{co}\{\lim \nabla H_\omega(z^k)^T \ : \ z^k \to \hat{x},\ z^k \in D_{H_\omega}\}. \tag{3.19}$$

Let $y^k = \hat{x} - (x^k - \hat{x})$ and we immediately find $\{y^k\}$ converging to $\hat{x}$. Note that $\beta_\omega(y^k) \subseteq \beta_\omega(\hat{x})$ for k sufficiently large. Now we claim that $y^k \in D_{H_\omega}$ for such k. In fact, for sufficiently large k, $\beta_\omega(y^k) \subseteq \beta_\omega(x^k)$ since $\beta_\omega(y^k) \subseteq \beta_\omega(\hat{x})$ implies that $(x^k - \hat{x})_i = (M(\omega)(x^k - \hat{x}))_i$ for $i \in \beta_\omega(y^k)$, and hence

$$x^k_i = (\hat{x} + (x^k - \hat{x}))_i = (M(\omega)(\hat{x} + (x^k - \hat{x})) + q(\omega))_i = (M(\omega)x^k + q(\omega))_i.$$

By Lemma 3.1 (iii), $(M(\omega))_{i\cdot} = I_{i\cdot}$ for any $i \in \beta_\omega(y^k)$, which in turn implies $y^k \in D_{H_\omega}$. Thus

$$\nabla(H_\omega(y^k))^T_i = \nabla(H_\omega(x^k))^T_i = \frac{1}{2}[(M(\omega))_{i\cdot} + I_{i\cdot}], \quad i \in \beta_\omega(y^k).$$

By direct computation, we have the index $i \in \gamma_\omega(x^k)$ if $i \in \alpha_\omega(y^k) \cap \beta_\omega(\hat{x})$; and $i \in \alpha_\omega(x^k)$ if $i \in \gamma_\omega(y^k) \cap \beta_\omega(\hat{x})$. It is then easy to see that

$$\tilde{H}^o_\omega(\hat{x}) = \frac{1}{2}\lim_{k\to\infty} \nabla H_\omega(x^k)^T + \frac{1}{2}\lim_{k\to\infty} \nabla H_\omega(y^k)^T.$$

Hence $\tilde{H}^o_\omega(\hat{x}) \in \partial H_\omega(\hat{x})$ according to (3.19).

Now we show (iii) holds. By noting (3.17) and $\int_{-\infty}^{\frac{x_i - (M(\omega)x+q(\omega))_i}{\mu}} \rho(s)ds \in [0,1]$, we have

$$\begin{aligned} \|\nabla_x \tilde{f}(x,\mu)\| &= \|E[\nabla_x \tilde{H}_\omega(x,\mu)\tilde{H}_\omega(x,\mu)]\| \\ &\le E[(2 + \|M(\omega)\|)\|\tilde{H}_\omega(x,\mu)\|] \\ &\le \sqrt{E[(2+\|M(\omega)\|)^2]}\sqrt{E[\|\tilde{H}_\omega(x,\mu)\|^2]} \\ &\le \sqrt{E[(2+\|M(\omega)\|)^2]}\sqrt{\tilde{f}(x,\mu)} \\ &< \infty, \end{aligned}$$

where the second inequality employs the Cauchy-Schwarz inequality and the last inequality uses (1.3). Thus by the Lebesgue Dominated Convergence Theorem,

$$\begin{aligned} \tilde{f}^o(x) = \lim_{\mu\downarrow 0} \nabla_x \tilde{f}(x,\mu) &= 2\lim_{\mu\downarrow 0} E[\nabla_x \tilde{H}_\omega(x,\mu)\tilde{H}_\omega(x,\mu)] \\ &= 2E[\lim_{\mu\downarrow 0} \nabla_x \tilde{H}_\omega(x,\mu)\tilde{H}_\omega(x,\mu)] \\ &= 2E[\tilde{H}^o_\omega(x)^T H_\omega(x)]. \end{aligned}$$

For an arbitrary $\hat{x} \in R^n_+$, using two sequences $\{x^k\} \subset D_f \cap R^n_{++}$ converging to $\hat{x}$, and $y^k = \hat{x} - (x^k - \hat{x})$, we can show

$$\tilde{f}^o(\hat{x}) = \frac{1}{2} \lim_{k\to\infty} \nabla f(x^k) + \frac{1}{2} \lim_{k\to\infty} \nabla f(y^k) \in \partial f(\hat{x})$$

in a similar way as that for (ii) of this lemma. The proof is completed. □

REMARK 3.2. (ii) of Lemma 3.6 improves the results of the Jacobian consistency property in [7], which states that for any fixed x,

$$\lim_{\mu\downarrow 0} (\nabla_x \tilde{H}_\omega(x,\mu))^T = \tilde{H}^o_\omega(x) \in \partial_C H_\omega(x),$$

where $\partial_C H_\omega$ is the C-generalized Jacobian of H_ω defined by

$$\partial_C H_\omega(x) = \partial(H_\omega(x))_1 \times \partial(H_\omega(x))_2 \times \cdots \times \partial(H_\omega(x))_n.$$

It is known that $\partial H_\omega(x) \subseteq \partial_C H_\omega(x)$.

PROPOSITION 3.7. *The function $\tilde{f}$ defined by (3.15) is a smoothing function of f and Assumption 2.1 holds. Moreover, if f is differentiable at $x \in R^n_+$, then*

$$\{\lim_{z\to x,\ \mu\downarrow 0} \nabla_x \tilde{f}(z,\mu)\} = \nabla f(x). \tag{3.20}$$

Proof. It is obvious from (3.16) that for any $\mu \in R_{++}$, $\tilde{f}(\cdot,\mu)$ is continuously differentiable in R^n, and for any $x \in R^n_+$,

$$\{\lim_{z\to x,\ \mu\downarrow 0} \nabla_x \tilde{f}(z,\mu)\} \subseteq 2E[\partial H_\omega(x)^T H_\omega(x)]$$

is a nonempty and bounded set. The expected value multifunctions $E[\partial H_\omega(x)^T H_\omega(x)]$ is well-defined for any $(x,\omega) \in R^n \times \Omega$ according to Theorem 2 [29], since $\partial H_\omega(\cdot)^T H_\omega(\cdot)$ is upper semicontinuous at $x \in R^n$ for P-almost every $\omega \in \Omega$, and

$$\|V_1^T H_\omega(x) - V_2^T H_\omega(x)\| \le (1 + \|M(\omega)\|)\|H_\omega(x)\|, \quad \text{for any } V_1, V_2 \in \partial H_\omega(x).$$

By the Cauchy-Schwarz inequality and (1.3),

$$\begin{aligned} E[(1+\|M(\omega)\|)\|H_\omega(x)\|] &\le \sqrt{E[(1+\|M(\omega)\|)^2]}\sqrt{E[\|H_\omega(x)\|^2]} \\ &= \sqrt{E[(1+\|M(\omega)\|)^2]}\sqrt{f(x)} \\ &< \infty. \end{aligned}$$

It is also clear that for any fixed $z \in R^n$ and $\mu \in R_{++}$,

$$\begin{aligned} |\tilde{f}(z,\mu) - f(z)| &= |E[\|\tilde{H}_\omega(z,\mu)\|^2 - \|H_\omega(z)\|^2]| \\ &\le E[(\|\tilde{H}_\omega(z,\mu)\| + \|H_\omega(z)\|)(\|\tilde{H}_\omega(z,\mu) - H_\omega(z)\|)] \\ &\le E[(2\min(\|\tilde{H}_\omega(z,\mu)\|, \|H_\omega(z)\|) + \eta\mu)\eta\mu] \\ &= E[\min(\|\tilde{H}_\omega(z,\mu)\|, \|H_\omega(z)\|)]2\eta\mu + \eta^2\mu^2 \\ &\le \sqrt{\min(\tilde{f}(z,\mu), f(z))}2\eta\mu + \eta^2\mu^2, \end{aligned} \tag{3.21}$$

where the second inequality comes from (i) of Lemma 3.6 and the third inequality follows from the Cauchy-Schwarz inequality that $E[\xi] \le \sqrt{E[\xi^2]}$ for any random variable ξ. Thus

$$\lim_{z\to x,\ \mu\downarrow 0} \tilde{f}(z,\mu) = f(x),$$

since for any $z \to x$ and $\mu \downarrow 0$,

$$\begin{aligned}|\tilde{f}(z,\mu) - f(x)| &\leq |\tilde{f}(z,\mu) - f(z)| + |f(z) - f(x)| \\ &\leq \sqrt{f(z)}2\eta\mu + \eta^2\mu^2 + |f(z) - f(x)| \\ &\to 0.\end{aligned}$$

Therefore, the function $\tilde{f}$ defined by (3.15) is a smoothing function of f.

Now we begin to show that Assumption 2.1 holds. Obviously, $\tilde{f}(\cdot,\mu)$ is bounded below, since $\tilde{f}(x,\mu) \geq 0$ for any $x \in R^n$ and $\mu \in R_{++}$. We have by simple computation that for any $x, y \in R^n$,

$$\|\nabla_x \tilde{H}_\omega(y,\mu)\| \leq 2 + \|M(\omega)\|,$$

$$\|\tilde{H}_\omega(x,\mu) - \tilde{H}_\omega(y,\mu)\| \leq (2 + \|M(\omega)\|)\|x - y\|,$$

and

$$\|\nabla_x \tilde{H}_\omega(x,\mu) - \nabla_x \tilde{H}_\omega(y,\mu)\| \leq \frac{(1 + \|M(\omega)\|)^2}{\mu}\|x - y\|.$$

Hence we have

$$\begin{aligned}&\|\nabla_x \tilde{f}(x,\mu) - \nabla_x \tilde{f}(y,\mu)\| \\ &\quad = |2E[\nabla_x \tilde{H}_\omega(x,\mu)\tilde{H}_\omega(x,\mu) - \nabla_x \tilde{H}_\omega(y,\mu)\tilde{H}_\omega(y,\mu)]| \\ &\quad \leq 2E[\|\nabla_x \tilde{H}_\omega(y,\mu)\|\|\tilde{H}_\omega(x,\mu) - \tilde{H}_\omega(y,\mu)\| + \|\tilde{H}_\omega(x,\mu)\|\|\nabla_x \tilde{H}_\omega(x,\mu) - \nabla_x \tilde{H}_\omega(y,\mu)\|] \\ &\quad \leq 2E[(2 + \|M(\omega)\|)^2 + \frac{1}{\mu}\|\tilde{H}_\omega(x,\mu)\|(1 + \|M(\omega)\|)^2]\|x - y\|. \qquad (3.22)\end{aligned}$$

The above inequality indicates that for any fixed $\mu > 0$ and $\Gamma > 0$, $\nabla_x \tilde{f}(\cdot,\mu)$ is uniformly continuous on the level set $L_{\mu,\Gamma}$, since $E[(2 + \|M(\omega))^2] < \infty$ according to (1.3), and

$$\begin{aligned}E[\|\tilde{H}_\omega(x,\mu)\|(1 + \|M(\omega)\|)^2] &\leq \sqrt{E[\|\tilde{H}_\omega(x,\mu)\|^2]}\sqrt{E[(1 + \|M(\omega)\|)^4]} \\ &\leq \sqrt{\tilde{f}(x,\mu)}E[(1 + \|M(\omega)\|)^2] < \infty,\end{aligned}$$

where the second inequality employs the Jensen's inequality that

$$\sqrt{E[(1 + \|M(\omega)\|)^4]} \leq E[\sqrt{(1 + \|M(\omega)\|)^4}] = E[(1 + \|M(\omega)\|)^2].$$

Hence Assumption 2.1 holds.

According to Proposition 3.5, f is differentiable at $x \in R^n_+$ if and only if for $\omega \in \Omega$ a.e., $\theta_\omega(x)$ is differentiable at x. Since $\partial H_\omega(x)^T H_\omega(x) = \nabla\theta_\omega(x)$ if $\theta_\omega(x)$ is differentiable at $x \in R^n$, we have

$$2E[\partial H_\omega(x)^T H_\omega(x)] \subseteq 2E[\nabla\theta_\omega(x)] = \nabla f(x)$$

provided that f admits differentiability at $x \in R^n_+$. Hence if f is differentiable at $x \in R^n_+$, then (3.20) holds. □

3.3. Stationary point. From Theorem 2.6, any accumulation point of $\{x^k\}$ generated by the SPG method is a stationary point of the problem (1.1) associated with the smoothing function $\tilde{f}$. For the ERM formulation (1.6) of SLCP, the objective function f is everywhere directionally differentiable according to Proposition 3.5. The stationary point [13, pp. 91] of the problem (1.6) is usually defined to be a feasible vector $x \in R^n_+$ such that

$$f'(x, d) \geq 0, \quad \forall\, d \in T(x; R^n_+), \tag{3.23}$$

where $T(x; R^n_+)$ is the tangent cone of R^n_+ at a vector $x \in R^n_+$, that is, the cone consists of all vectors $d \in R^n$, for which there exist a sequence of vectors $\{y^k\} \subset R^n_+$ and a sequence of positive scalars $\{\tau_k\}$ such that

$$\lim_{k\to\infty} y^k = x, \quad \lim_{k\to\infty} \tau_k = 0, \quad \text{and} \quad \lim_{k\to\infty} \frac{y^k - x}{\tau_k} = d.$$

Recall that a vector $0 \neq d \in R^n$ is called a feasible direction of the nonnegative orthant R^n_+ at a point $x \in R^n_+$, if there exists a constant $\delta > 0$ such that

$$x + td \in R^n_+ \quad \text{for any} \quad t \in [0, \delta].$$

For problem (1.6), it is easy to show that (3.23) is equivalent to

$$f'(x, d) \geq 0, \quad \forall\, d \in F(x; R^n_+), \tag{3.24}$$

where $F(x; R^n_+)$ is the set of feasible directions $d \in R^n$.

In what follows, we provide an equivalent characterization of the stationary point, and discuss its relation to the stationary point associated with the $\tilde{f}$. Denote $e^i = I^T_{i.}$ for $i = 1, \ldots, n$. For an arbitrary $x \in R^n_+$, let us denote the index set $\mathcal{S}_x = \{i \ : \ x_i > 0\} = \{s_1, s_2, \ldots, s_{t(x)}\}$, and $\bar{\mathcal{S}}_x = \{1, 2, \ldots, n\} \backslash \mathcal{S}_x = \{i \ : \ x_i = 0\}$, where $t(x)$ is the number of elements in $\mathcal{S}_x$. Let

$$\mathcal{D}_x = \{e^i,\ i = 1, \ldots, n\} \cup \{-e^{s_i},\ i = 1, \ldots, t(x)\}. \tag{3.25}$$

Note that $\mathcal{D}_x$ is determined by x, and for any $x \in R^n$, the number of vectors in $\mathcal{D}_x$ satisfies $n \leq |\mathcal{D}_x| \leq 2n$, and $\|d\| = 1$ for any $d \in \mathcal{D}_x$.

THEOREM 3.8. *$x \in R^n_+$ is a stationary point of the problem (1.6) if and only if $f'(x, d^l) \geq 0$ for any $d^l \in \mathcal{D}_x$.*

Proof. The "only if" part is obvious true since any direction $d^l \in \mathcal{D}_x$ is a feasible direction at the point $x \in R^n_+$ on the nonnegative orthant R^n_+. In what follows we prove for the "if" part.

Note that any feasible direction $d = (d_1, \ldots, d_n)^T$ at $x \in R^n_+$ can be expressed as a nonnegative linear combination of $d^l \in \mathcal{D}_x$. That is, we can write

$$d = \sum_{d^l \in \mathcal{D}_x} \eta_l d^l, \qquad \eta_l \geq 0 \ \text{ for any } \ d^l \in \mathcal{D}_x.$$

By repeating using (3.7) shown in the proof of Lemma 3.3, we have

$$\theta'_\omega(x, d) = \theta'_\omega(x, \sum_{d^l \in \mathcal{D}_x} \eta_l d^l) \geq \sum_{d^l \in \mathcal{D}_x} \eta_l \theta'_\omega(x, d^l).$$

Together with the linearity of the expectation function, we obtain

$$\begin{aligned} f'(x,d) &= 2E[\theta'_\omega(x, \sum_{d^l\in\mathcal{D}_x} \eta_l d^l)] \\ &\geq 2E[\sum_{d^l\in\mathcal{D}_x} \eta_l \theta'_\omega(x,d^l)] \\ &= 2\sum_{d^l\in\mathcal{D}_x} \eta_l E[\theta'_\omega(x,d^l)] \\ &= \sum_{d^l\in\mathcal{D}_x} \eta_l f'(x,d^l) \geq 0, \end{aligned}$$

which implies that x is a stationary point of (1.6). ∎

COROLLARY 3.9. *If $\Omega = \{\omega^1, \omega^2, \ldots, \omega^N\}$, then $x \in R^n_+$ is a local minimizer of the problem (1.6) if and only if $f'(x, d^l) \geq 0$ for any $d^l \in \mathcal{D}_x$.*

Proof. If $\Omega = \{\omega^1, \omega^2, \ldots, \omega^N\}$, then R^n_+ can be divided into finite polyhedron pieces, and f is a convex quadratic function on each piece. Hence a stationary point of the problem (1.6) coincides with a local minimizer of f. The corollary follows immediately from Theorem 3.8. ∎

REMARK 3.3. If x^* is a stationary point of (1.6) associated with $\tilde{f}$ and f is differentiable at x^*, then $(\nabla f(x^*), x^* - z) \leq 0$ for all $z \in R^n_+$ according to Proposition 3.7. Hence for any $d \in F(x; R^n_+)$, there exists a constant $\delta > 0$ such that $x + \delta d \in R^n_+$ and

$$f'(x^*, d) = \nabla f(x^*)^T d = -\frac{1}{\delta}(\nabla f(x^*), x^* - (x^* + \delta d)) \geq 0.$$

Thus by (3.24), x^* is a stationary point of (1.6). In addition, if $\Omega = \{\omega^1, \omega^2, \ldots, \omega^N\}$ is a finite set, x^* is a local minimizer according to Corollary 3.9.

Some mild conditions on initial data $M(\omega)$ for $\omega \in \Omega$ can guarantee that f is differentiable at any local minimizer.

THEOREM 3.10. *If $\mathcal{P}\{\omega \;:\; (M(\omega))_{i\cdot} \neq I_{i\cdot},\ M_{ii}(\omega) = 1\} = 0$ for each $i \in \{1, 2, \ldots, n\}$, then f is differentiable at any local minimizer $z \in R^n_+$.*

Proof. Suppose on the contrary that f is not differentiable at a local minimizer $z \in R^n_+$. According to Proposition 3.5, there exists an index $l \in \{1, 2, \ldots, n\}$ such that $z_l > 0$, and $\tilde{\Omega}_l = \{\omega \;:\; (M(\omega))_{l\cdot} \neq I_{l\cdot},\ l \in \beta_\omega(z)\}$ with $\mathcal{P}\{\tilde{\Omega}_l\} > 0$.

Since $z_l > 0$, both e^l and $-e^l$ are feasible directions of R^n_+ at z. For arbitrary $d \in R^n$, $i \in \{1, 2, \ldots, n\}$ and $\omega \in \Omega$, putting $\lambda = -1$ in (3.6), we obtain

$$\min(-d_i, -(M(\omega)d)_i) + \min(d_i, (M(\omega)d)_i) \leq 0, \tag{3.26}$$

where the equality holds if and only if $d_i = (M(\omega)d)_i$. For any $x \in R^n_+$ and $d \in R^n$, we have by direct computation that

$$\begin{aligned} f'(x,-d) &= -f'(x,d) + 2E[\sum_{i\in\beta_\omega(x)} x_i(\min(-d_i, -(M(\omega)d)_i) + \min(d_i, (M(\omega)d)_i)] \\ &\leq -f'(x,d). \end{aligned} \tag{3.27}$$

Noting that the local minimizer z is a stationary point, we get from Theorem 3.8 that $0 \leq f'(z, -e^l) \leq -f'(z, e^l) \leq 0$, which implies

$$f'(z, -e^l) = f'(z, e^l) = 0. \tag{3.28}$$

Setting $d = e^l$ in (3.26) and (3.27), and employing (3.28), we get

$$\begin{aligned}0 &= E[\sum_{i\in\beta_\omega(z)} z_i(\min(-e^l_i, -(M(\omega)e^l)_i) + \min(e^l_i, (M(\omega)e^l)_i))] \\ &\le E_{\tilde{\Omega}_l}[z_l(\min(-(e^l)_l, -(M(\omega)e^l)_l) + \min((e^l)_l, (M(\omega)(e^l)_l)))] \le 0,\end{aligned}$$

which further implies that $M_{ll}(\omega) = (M(\omega)e^l)_l = (e^l)_l = 1$ for $\omega \in \tilde{\Omega}_l$ a.e.. Hence,

$$\mathcal{P}\{\omega \ : \ (M(\omega))_{l\cdot} \neq I_{l\cdot},\ M_{ll}(\omega) = 1\} \ge \mathcal{P}\{\tilde{\Omega}_l\} > 0,$$

which contradicts to the assumption of this theorem. The proof is completed. □

4. Examples and numerical results. In this section, we illustrate the SPG algorithm and its convergence results using examples of the ERM formulation (1.6) of SLCPs, as well as image restoration problems. In the SPG algorithm for the numerical experiment, we set

$$\mu_0 = 1, \quad \gamma_1 = \frac{1}{2}, \quad \gamma_2 = \frac{1}{4}, \quad \gamma_3 = 10^3, \quad \sigma = \frac{1}{2}.$$

4.1. ERM formulation of SLCPs. The objective function in the ERM formulation of SLCPs is the expectation functions of the form

$$f(x) = 2E[\theta_\omega(x)],$$

where $\omega \in \Omega$ is a random vector with a given probability distribution $\mathcal{P}$.

When ω is a discrete random variable that takes on the distinct values $\omega^1, \dots, \omega^N$ with probabilities $p_1, \dots, p_N$, the function value is defined by

$$f(x) = 2\sum_{i=1}^N \theta_{\omega^i}(x)p_i.$$

If ω is a continuous random variable with probability density function $p(\omega)$, the function value is defined by

$$f(x) = 2\int_\Omega \theta_\omega(x)p(\omega)d\omega,$$

which, in general, is difficult to compute accurately. By assumption (1.3), the integrals satisfy the Law of Large Number and hence the integrals can be estimated from large sample averages. Note that if sampling is used then we do not need a knowledge of the distribution.

In practice, the sample average approximation (SAA) [18, 27] is usually employed, which replaces the original objective function by its approximation

$$\hat{f}_N(x) := \frac{2}{N}\sum_{i=1}^N \theta_{\omega^i}(x).$$

Here the sample $\omega^1, \dots, \omega^N$ is generated by Monte Carlo sampling method, following the same probability distribution as ω. The smoothing projected gradient method can then be applied to solve the approximation problem. The function value and gradient of the smoothing function are defined by

$$\tilde{f}_N(x) := \frac{2}{N}\sum_{i=1}^{N}\tilde{\theta}_{\omega^i}(x,\mu)$$

and

$$\nabla_x\tilde{f}_N(x) := \frac{2}{N}\sum_{i=1}^{N}\nabla_x\tilde{\theta}_{\omega^i}(x,\mu).$$

REMARK 4.1. The proper sample size may vary for different SLCPs in practice. The stochastic approximation (SA) method, which originates from [25], and is developed by [17, 24] etc., seems promising to avoid large sampling. We will investigate the SA method with the smoothing projected gradient method in our future research.

In our numerical experiment, we use the procedure described in [9] to generate test problems of monotone SLCP$(M(\omega), q(\omega))$, $\omega \in \Omega = \{\omega^j,\ j = 1,\dots,N\}$ with $p_j = \mathcal{P}\{\omega^j\} = \frac{1}{N}$ for all j. We call (1.2) a monotone SLCP if $E[M(\omega)]$ is positive semi-definite. Let us recall some notations in the procedure,

$\hat{x}$: the nominal seed point in R^n_+
n_x : the number of elements in the index set $\mathcal{J} = \{i\ :\ \hat{x}_i > 0\}$
$\mathcal{I}_j$: the index set $\{i\ :\ \hat{x}_i = 0,\ (M(\omega^j)\hat{x} + q(\omega^j))_i \geq 0\}$
$[0,\beta)$: the range of $(M(\omega^j)\hat{x} + q(\omega^j))_i$ for $i \in \mathcal{J}$
$(-\sigma,\sigma)$: the range of elements of matrix $E[M(\omega)] - M(\omega^j)$ for each j.

For each fixed (n, n_x, σ), we set $\hat{x}_i \in (0,20)$ for $i \in \mathcal{J}$, and $(M(\omega^j)\hat{x}+q(\omega^j))_i \in [0,15)$ for $i \in \mathcal{I}_j,\ j = 1,\dots,N$. Each random matrix $M(\omega^j)$ is generated by uniformly distributed random variables and the QR decomposition, which is a dense matrix. The condition number of the expectation matrix $\bar{M} = E[M(\omega)]$ is 100 for all the test problems.

We use the Chen-Mangasarian smoothing function with the uniform density function in (3.14), and set

$$\hat{\gamma} = 10^3, \quad \sigma_1 = \sigma_2 = 10^{-6}$$

in the SPG algorithm. We stop the iteration of the SPG algorithm and set the computed solution $\tilde{x} = x^k$, if $\|x^k - x^{k-1}\| \leq 10^{-12}$, or "T-iters", i.e., the number of the total iteration invoking (2.2)-(2.4) exceeds 4000. We use "O-iters" to represent the number of outer iterations, i.e., the number k such that $\tilde{x} = x^k$, and 'cpu' to represent the CPU time in seconds. To check the optimality at the terminal point, we compute

$$r(\tilde{x}) = \|\min(\tilde{x}, \nabla f(\tilde{x}))\|$$

if f is differentiable at $\tilde{x}$. Note that by Proposition 3.5, we can easily check whether f is differentiable at $\tilde{x}$. By Remark 3.3, if f is differentiable at $\tilde{x}$, then $\tilde{x}$ is a local minimizer if and only if $r(\tilde{x}) = 0$.

In Table 4.1, we compare the SPG method with the `fmincon`, a Matlab code for constrained minimization. For comparison, we fix $\beta = 0$, $N = 100$ and start from the same randomly generated initial point

$$x^0 = \text{floor}(1 + 10 * \text{rand}(n,1)),$$

TABLE 4.1
Finding a global optimal solution

	SPG				fmincon			
(n, n_x, σ)	$f(\tilde{x})$	cpu$(\tilde{x})$	err$(\tilde{x})$	$r(\tilde{x})$	$f(\check{x})$	cpu$(\check{x})$	err$(\check{x})$	$r(\check{x})$
$20, 10, 20$	4.30e-22	1.52	2.26e-14	9.29e-10	1.34e-7	9.66	3.75e-7	2.13e-2
$20, 10, 10$	6.52e-23	1.63	1.66e-14	1.67e-10	1.20e-7	7.66	7.07e-7	9.38e-3
$20, 10, \ 0$	8.63e-18	42.9	3.19e-10	4.47e-9	5.91e-7	3.06	3.05e-5	4.34e-3
$40, 20, 20$	4.03e-22	58.9	8.36e-15	1.77e-9	6.01e-7	29.3	3.64e-7	3.71e-2
$40, 20, 10$	1.42e-23	3.89	3.21e-15	1.61e-10	5.52e-7	28.6	7.05e-7	1.73e-2
$40, 20, \ 0$	1.23e-12	68.2	1.04e-7	1.48e-6	1.57e-6	13	5.05e-5	4.88e-3
$60, 30, 20$	2.07e-22	8.34	5.30e-15	1.26e-9	76.4	77.1	3.02e-3	323
$60, 30, 10$	8.37e-24	8.03	2.11e-15	1.16e-10	604	79.6	1.69e-2	863
$60, 30, \ 0$	4.71e-11	154	4.66e-7	4.60e-6	1.52e-6	39	3.23e-5	5.14e-3
$80, 40, 20$	1.36e-21	15.2	9.72e-15	2.90e-9	94	134	2.49e-3	540
$80, 40, 10$	3.89e-23	11.3	3.18e-15	2.03e-10	2.87	133	8.67e-4	55.1
$80, 40, \ 0$	2.08e-18	331	7.03e-11	2.48e-9	1.16e-6	109	3.4e-5	3.7e-3
$100, 50, 20$	3.85e-22	27.9	3.75e-15	2.26e-9	8.19e3	282	2.09e-2	3.87e3
$100, 50, 10$	9.13e-23	28.5	3.48e-15	6.05e-10	89.6	275	4.08e-3	206
$100, 50, \ 0$	1.01e-12	499	6.09e-8	2.14e-6	1.67e-4	192	4.70e-4	5.78e-2

TABLE 4.2
Finding a local minimizer

$N,\ n,\ n_x,\ \beta,\ \sigma$	$f(\bar{x})$	$f(\tilde{x})$	T-iters	O-iters	cpu	$r(\tilde{x})$
1000, 50, 25, 10, 20	1.92e6	6.16e2	57	27	3.28e1	3.18e-5
1000, 100, 50, 5, 10	4.68e5	3.54e2	25	22	1.14e2	6.16e-6
100, 500, 250, 10, 20	1.69e8	6.40e3	25	18	2.98e2	1.48e-4
100, 1000, 500, 5, 10	6.39e7	3.65e3	50	22	1.80e3	8.16e-4
50, 1500, 750, 10, 20	1.44e9	1.84e4	39	17	1.37e3	3.85e-3

and employ the SPG method and the `fmincon` to get computed solutions $\tilde{x}$ and $\check{x}$ of the ERM formulation, respectively. Note that $\beta = 0$ implies that $\hat{x}$ is the unique global solution of the test problems and $f(\hat{x}) = 0$. We record the relative error of a computed solution x, i.e.,

$$\mathrm{err}(x) = \frac{\|\hat{x} - x\|}{\|\hat{x}\|}.$$

From Table 4.1, we observe that the SPG method succeeds in finding the unique global solution and is much more efficient than the `fmincon` code, by noting that the function value, relative error and optimal condition at $\tilde{x}$ and $\check{x}$. In fact, the `fmincon` code failed in most cases when n exceeds 50 and $\sigma > 0$.

In Table 4.2, we present numerical results of the SPG method for the test problems with $\beta > 0$. In this case, the global solution is unknown. We first use the semismooth Newton method [21] to get a solution $\bar{x}$ of LCP$(E[M(\omega)], E[q(\omega)])$, that is, the expected value (EV) formulation of SLCP$(M(\omega), q(\omega))$. We then take its solution $\bar{x}$ as the initial point x^0 for the SPG algorithm to get a computed solution $\tilde{x}$ of the ERM formulation.

Table 4.2 shows that the SPG algorithm largely reduces the function value from the solution $\bar{x}$ of the EV formulation. Moreover, the value of r suggests that the SPG algorithm converges to a local minimizer. Furthermore, we find that the SPG algorithm keeps similar number of iterations when the dimension n of the problem increases from 50 to 1500.

4.2. Image restoration problems. The smoothing projected gradient method can be applied for minimizing a general nonconvex, nonsmooth function on a convex set. In this subsection, we provide its application in image restoration.

Image restoration is to reconstruct an image of an unknown scene from an observed image, which has wide applications in engineering and sciences. Large-scale nonsmooth, nonconvex constrained minimization problems often appear in the image restoration [15, 23].

The observed image is distorted from the unknown true image mainly by two factors – the blurring and the random noise. The blurring may arise from various sources such as atmospheric turbulence, motion blurs, etc.. Suppose the discretized scenes have $n = m \times m$ pixels, then the image of an object can be modeled as

$$b = Ax + \delta, \tag{4.1}$$

where the n-dimensional vectors x, b and δ are the true image, the observed image, and the additive noise, respectively. The matrix A of $n \times n$ is the corresponding blurring matrix of block Toeplitz with Toeplitz blocks (BTTB) when zero boundary conditions are applied. In this case, the fast Fourier transforms (FFTs) can be used to implement fast matrix-vector multiplications.

Solving (4.1) directly will lead to unstable solutions which are very sensitive to noise, since image restoration problems are ill-conditioned. Regularization technique is often used to get robust solution. As pointed out in [23], although convex potential functions (PFs), e.g., $\phi(t) = |t|$, can be used for the regularization term, nonconvex regularization provides better possibilities for restoring images with neat edges. In the following, we consider the image restoration model given in [23],

$$\begin{aligned} &\min \ \|Ax - b\|^2 + c\sum_{i=1}^{n} \varphi(x_i) \\ &\text{s.t. } x \geq 0, \end{aligned} \tag{4.2}$$

where c is the regularization parameter, and $\varphi : R \to R$ is a potential function (PF) defined by

$$\varphi(t) = \frac{a|t|}{1 + a|t|}, \quad a \in (0, 1). \tag{4.3}$$

It is easy to see that the objective function of (4.2) is nonsmooth nonconvex. The constraint $x \geq 0$ reflects the fact that the pixels are nonnegative.

We choose a map of island shown in Fig. 4.1 (a) as the original scene with pixels $n = 256 \times 256 = 65536$. We set the regularization parameter $c = 10^{-6}$ and the parameter $a = 0.5$ in the PF. We obtain the smoothing function of φ in (4.3) by replacing $|t|$ by its smoothing approximation based on the uniform density function ρ in (3.14).

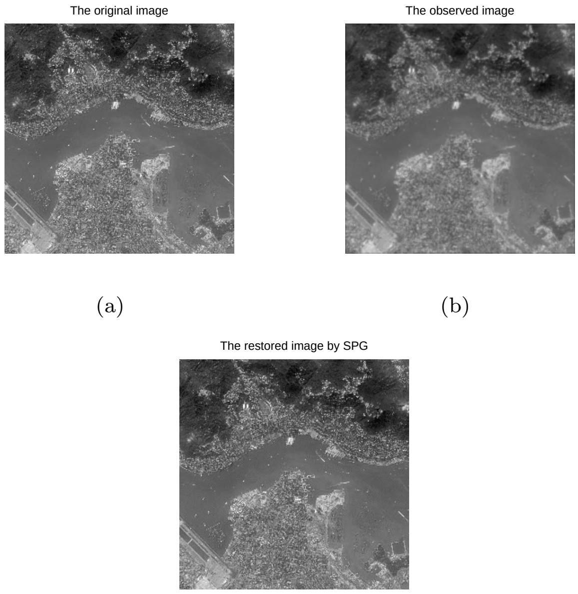
We get the observed scene in Fig. 4.1 (b) from the original image which is first blurred by a Gaussian function and then contaminated by a Gaussian white noise. We use the observed image as the initial point for the SPG algorithm, choose the parameters for this problem as

$$\hat{\gamma} = 10^5, \quad \sigma_1 = \sigma_2 = 10^{-3},$$

and stop the iteration if

$$\frac{\|x^k - x^{k-1}\|}{\|x^{k-1}\|} \leq 10^{-4}.$$

We record the restored image $\tilde{x} = x^k$ by SPG in Fig. 4.1 (c).



(a) (b)

(c)

FIG. 4.1. *(a) The original image; (b)The observed image; (c) The restored image by SPG.*

The objective value and the peak signal noise ratio (PSNR) value obtained by the SPG method are 0.0158 and 83.89, which largely improved those at the observed image $(20.6383, 72.19)$ of the observed image.

5. Final remark. We propose a globally convergent smoothing projected gradient (SPG) method for minimizing a nonconvex, nonsmooth function on a convex set. We prove that any accumulation point generated by the SPG method converges to a stationary point associated with the smoothing function used in the method, which is a Clark stationary point in many applications. The key idea of the SPG method is to use a parametric smoothing approximation function in the projected gradient method [3]. We apply the SPG method to the expected residual minimization (ERM) reformulation of the stochastic linear complementarity problems (SLCPs) and image restoration problems. Theoretical and numerical results show that the SPG method is promising. Our analysis on the SPG method and the ERM reformulation can be applied to many problems in optimization. For example, consider the following mathematical programs with equilibrium constraints [22]

$$\begin{aligned} \min\ & c(x) \\ \text{s.t.}\ & x \in X \\ & p(x) \geq 0,\ \ q(x) \geq 0,\ \ p(x)^T q(x) = 0, \end{aligned} \tag{5.1}$$

where X is a closed convex set in R^n, $c : R^n \to R$, $p, q : R^n \to R^m$ are continuously differentiable functions. (5.1) can be approximated by

$$\begin{aligned} \min\ & c(x) + \sigma \| \min(p(x), q(x)) \|^2 \\ \text{s.t.}\ & x \in X, \end{aligned} \tag{5.2}$$

where $\sigma > 0$ is a penalty parameter. The penalty function is nonconvex, nonsmooth. We can define a smoothing function for the penalty function by the Chen-Mangasarian smoothing function and use the SPG method to solve (5.2).

Acknowledgments. We would like to thank Dr. Weijun Zhou for pointing out an error in Theorem 2.6 on an earlier version of this paper. We thank Prof. Jane Ye for helpful discussion on the Clarke stationary point with respect to the smoothing function. We are very grateful to Prof. J.-S. Pang and the anonymous referees for valuable comments.

REFERENCES

[1] E. G. Birgin, J. M. Martínez, and M. Raydan, *Nonmonotone spectral projected gradient methods on convex set*, SIAM J. Optim., 10(2000), pp. 1196–1211.
[2] J. V. Burke, A. S. Lewis, and M. L. Overton, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*, SIAM J. Optim., 15(2005), pp. 751–779.
[3] P. H. Calamai and J. J. Moré, *Projected gradient method for linearly constrained problems*, Math. Program., 39(1987), pp. 93–116.
[4] B. Chen and X. Chen, *A global and local superlinear continuation-smoothing method for P_0 and R_0 NCP or monotone NCP*, SIAM J. Optim., 9(1999), pp. 624–645.
[5] C. Chen and O. L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comp. Optim. Appl., 5(1996), pp. 97–138.
[6] X. Chen and M. Fukushima, *Expected residual minimization method for stochastic linear complementarity problems*, Math. Oper. Res., 30(2005), pp. 1022–1038.
[7] X. Chen, L. Qi and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Math. Comp., 67(1998), pp. 519–540.
[8] X. Chen and Y. Ye, *On homotopy-smoothing methods for box-constrained variational inequalities*, SIAM J. Control Optim., 37(1999), pp. 589–616.
[9] X. Chen, C. Zhang and M. Fukushima, *Robust solution of monotone stochastic linear complementarity problems*, Math. Program., 117(2009), pp. 51–80.
[10] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, (reprinted by SIAM, Philadelphia), 1990.
[11] R .W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*, Academic Press, Inc. Boston, 1992.
[12] J. C. Dunn, *Global and asymptotic convergence rate estimates for a class of projected gradient processes*, SIAM J. Control. Optim., 19(1981), pp. 368–400.
[13] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.
[14] H. Fang, X. Chen and M. Fukushima, *Stochastic R_0 matrix linear complementarity problems*, SIAM J. Optim., 18(2007), pp. 482–506.
[15] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, *Efficient minimization methods of mixed $l_2 - l_1$ and $l_1 - l_1$ norms for image restoration*, SIAM J. Sci. Comput., 27(2006), pp. 1881–1902.
[16] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Amer. Soc., 70(1964), pp. 709–710.
[17] A. Juditsky, G. Lan, A. Nemirovski, and A. Shapiro, *Stochastic approach to stochastic programming*, submitted to SIAM J. Optim., E-print avaliable at http://www.optimization-online.org.
[18] A. J. Kleywegt, A. Shapiro, and T. Homen-De-Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim., 12(2001), pp. 479–502.
[19] K. C. Kiwiel, *Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization*, SIAM J. Optim., 18(2007), pp. 379–388.
[20] E. S. Levitin and B. T. Polyak, *Constrained minimization problems*, USSR. Comput. Math. Math. Phys., 6(1966), pp. 1–50.
[21] T. D. Luca, F. Facchinei and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program., 75(1996), pp. 407–439.
[22] Z.-Q. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, 1996.

[23] M. Nikolova, M. K. Ng, S. Zhang and W. Ching, *Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization*, SIAM J. Imaging Sciences, 1(2008), pp. 2–25.
[24] B. Polyak and A. Juditsky, *Acceleration of stochastic approximation by averaging*, SIAM J. Control Optim., 30(1992), pp. 838–855.
[25] H. Robbins and S. Monro, *A stochastic approximation method, Ann. Math. Statist.*, 22(1951), pp. 400–407.
[26] R. T. Rockafellar and A. J-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
[27] A. Ruszcyǹski and A. Shapiro, Eds., *Stochastic Programming*, Handbooks in OR & MS, vol. 10, North-Holland Publishing Company, Amsterdam, 2003.
[28] A. Schwartz and E. Polak, *Family of projected descent methods for optimization problems with simple bounds*, J. Optim. Theory Appl., 92(1997), pp. 1–31.
[29] A. Shapiro and H. Xu, *Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions*, J. Math. Anal. Appl., 325(2007), pp. 1390–1399.
[30] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.
[31] N. H. Xiu and J. Z. Zhang, *Some recent advances in projection-type methods for variational inequalities*, J. Comput. Appl. Math., 152(2003), pp. 559–585.