

## Smoothing Projected Gradient Method for Solving Stochastic Linear Complementarity Problems

Chao Zhang      Xiaojun Chen  
Department of Mathematical Sciences  
Hirosaki University, Hirosaki 036-8561, Japan

### 1 Introduction

In this paper, we propose a smoothing projected gradient method for solving the stochastic nonlinear complementarity problem.

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

$$x \geq 0, M(\omega)x + q(\omega) \geq 0, x^T(M(\omega)x + q(\omega)) = 0, \quad \omega \in \Omega. \quad (1.1)$$

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. Throughout the paper, we always assume  $M(\omega)$  and  $q(\omega)$  are measurable functions of  $\omega$  and satisfy

$$E[\|M(\omega)\|^2 + \|q(\omega)\|^2] < \infty. \quad (1.2)$$

When  $\Omega$  is a singleton,  $\text{SLCP}(M(\omega), q(\omega))$  reduces to the well-known linear complementarity problem  $\text{LCP}(M, q)$  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 [4] is a deterministic formulation for the SLCP, which is defined as

$$\min_{x \in R_+^n} f(x) := E[\|\Phi(x, \omega)\|^2] \quad (1.3)$$

where  $E$  stands for the expectation, and

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

and  $\phi : R^2 \rightarrow R$  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 (1.3) is neither convex nor smooth. Among various NCP functions, the “min” function

$$\phi(a, b) := \min(a, b), \quad \text{for any } (a, b) \in R^2, \quad (1.4)$$

has various nice properties for (1.3). It is shown in Lemma 2.2 [6] that the ERM formulation defined by the “min” function always has a solution if  $\Omega = \{\omega^1, \omega^2, \dots, \omega^N\}$  is a finite set. However, the ERM formulation defined by the Fischer-Burmeister 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

$$\min_{x \in \mathbb{R}_+^n} f(x) := E[\|\min(x, M(\omega)x + q(\omega))\|^2]. \quad (1.5)$$

This is a nonsmooth, nonconvex constrained minimization problem.

The expected residual minimization (ERM) formulation for the SLCP discussed in [4, 6, 8]. However, it is hard to find an efficient numerical methods to solve (1.5) when  $n$  is large. In this paper, we propose a smoothing projected gradient (SPG) method, which combines the smoothing techniques and the classical PG method to solve stochastic linear complementarity problem. The SPG method is easy to implement. At each iteration, we approximate the objective function  $f$  by a smooth function  $\tilde{f}$  with a fixed smoothing parameter, and employ the classical PG method to obtain a new point. Then we update the smoothing parameter using the new point for the next iteration.

The projected gradient (PG) method was originally proposed by Goldstein [9], and Levitin and Polyak [11] in 1960s, for minimizing a continuously differentiable mapping  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  on a nonempty closed convex set  $X$ . Nonsmooth and nonconvex optimization occurs frequently in practice. The projected subgradient method [13] extends the PG method to the case that  $f$  is nonsmooth, but convex. Recently, Burke, Lewis and Overton [1] introduced a robust gradient sampling algorithm for solving nonsmooth, nonconvex unconstrained minimization problem. Kiwiel [10] slightly revised the gradient sampling algorithm in [1] and showed that any accumulation point generated by the algorithm is a Clarke stationary point with probability one.

Throughout the paper, we use  $\|\cdot\|$  to represent the Euclidean norm, and let  $\mathbb{R}_{++}^n = \{x \in \mathbb{R}^n : x > 0\}$ .  $I$  denotes the identity matrix. For a given matrix  $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ , let  $A_i$  be the  $i$ -th row of  $A$ .

## 2 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 for SLCP.

Let  $H_\omega : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\theta_\omega : \mathbb{R}^n \rightarrow \mathbb{R}$  be defined by

$$H_\omega(x) = \min(x, M(\omega)x + q(\omega)), \quad \theta_\omega(x) = \frac{1}{2}H_\omega(x)^T H_\omega(x) \quad \text{for } \omega \in \Omega.$$

Thus the ERM formulation of SLCP( $M(\omega), q(\omega)$ ) can be expressed by

$$\min_{x \in \mathbb{R}_+^n} f(x) := 2E[\theta_\omega(x)]. \quad (2.1)$$

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

$$\begin{aligned}\alpha_\omega(x) &= \{i : x_i > (M(\omega)x + q(\omega))_i\} \\ \beta_\omega(x) &= \{i : x_i = (M(\omega)x + q(\omega))_i\} \\ \gamma_\omega(x) &= \{i : x_i < (M(\omega)x + q(\omega))_i\}.\end{aligned}\tag{2.2}$$

**Proposition 2.1** *The function  $f$  is locally Lipschitz continuous, and everywhere directionally differentiable with*

$$f'(x, d) = 2E[\theta'_\omega(x, d)] \quad \text{for all } d.\tag{2.3}$$

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

$$x_i = 0; \quad \text{or} \quad (M(\omega))_{ii} = I_i, \quad \text{for any } i \in \beta_\omega(x), \quad \omega \in \Omega \text{ a.e.,}\tag{2.4}$$

then  $f$  is differentiable at  $x$  and

$$\nabla f(x) = 2E[\nabla \theta_\omega(x)].\tag{2.5}$$

Moreover,  $f$  is differentiable at  $x \in R_+^n$  if and only if (2.4) holds.

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

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

For problem (1.5), it is easy to show that  $x \in R_+^n$  is a stationary point if and only if

$$f'(x, d) \geq 0, \quad \forall d \in F(x; R_+^n),\tag{2.6}$$

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 Clarke stationary point. Denote  $e^i = I_i^T$  for  $i = 1, \dots, n$ . For an arbitrary  $x \in R_+^n$ , let us denote the index set  $\mathcal{S}_x = \{i : x_i > 0\} = \{s_1, s_2, \dots, s_{t(x)}\}$ , and  $\bar{\mathcal{S}}_x = \{1, 2, \dots, n\} \setminus \mathcal{S}_x = \{i : x_i = 0\}$ , where  $t(x)$  is the number of elements in  $\mathcal{S}_x$ . Let

$$\mathcal{D}_x = \{e^i, i = 1, \dots, n\} \cup \{-e^{s_i}, i = 1, \dots, t(x)\}.\tag{2.7}$$

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 2.1**  *$x \in R_+^n$  is a stationary point of the problem (1.5) if and only if  $f'(x, d^i) \geq 0$  for any  $d^i \in \mathcal{D}_x$ .*

**Corollary 2.1** If  $\Omega = \{\omega^1, \omega^2, \dots, \omega^N\}$ , then  $x \in R_+^n$  is a local minimizer of the problem (1.5) if and only if  $f'(x, d^l) \geq 0$  for any  $d^l \in \mathcal{D}_x$ .

**Remark 2.1** If  $x^*$  is a Clarke stationary point of (1.5) and  $f$  is differentiable at  $x^*$ , then  $(\nabla f(x^*), x^* - z) \leq 0$  for all  $z \in R_+^n$ . 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 (2.6),  $x^*$  is a stationary point of (1.5). If, in addition,  $\Omega = \{\omega^1, \omega^2, \dots, \omega^N\}$  is a finite set,  $x^*$  is a local minimizer according to Corollary 2.1.

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

**Theorem 2.2** If  $\mathcal{P}\{\omega : (M(\omega))_i \neq I_i, M_{ii}(\omega) = 1\} = 0$  for each  $i \in \{1, 2, \dots, n\}$ , then  $f$  is differentiable at any local minimizer  $z \in R_+^n$ .

### 3 Smoothing projected gradient method

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

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

**Definition 3.1** Let  $f : X \subseteq R^n \rightarrow R$  be a locally Lipschitz continuous function. We call  $\tilde{f} : X \times R_{++} \rightarrow R$  a smoothing function of  $f$ , if  $\tilde{f}$  is continuously differentiable in  $X \times R_{++}$ , and there exists a function  $\varphi : R \rightarrow R_+$  such that  $\varphi(t) \rightarrow \infty$  implies  $|t| \rightarrow \infty$ , and

$$|\tilde{f}(x, \mu) - f(x)| \leq [\varphi(f(x))]\mu \quad \text{for all } x \in X. \quad (3.1)$$

Let  $\tilde{f}$  be a smoothing function of  $f$ , and the smoothing gradient projection method is defined as follows.

**Algorithm 3.1** (Smoothing projected gradient algorithm)

Let  $\gamma_1, \gamma_2$  and  $\hat{\gamma}$  be positive constants, and  $\sigma_1, \sigma_2$  and  $\sigma$  be constants in  $(0, 1)$ , where  $\sigma_1 \leq \sigma_2$ . Choose  $x^0 \in X$  and  $\mu_0 \in R_{++}$ . For  $k \geq 0$ :

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 4. Otherwise, go to step 2.
2. Let

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

and  $x^{k+1} = x^k(\alpha_k)$  where  $\alpha_k$  is chosen so that,

$$\tilde{f}(x^{k+1}, \mu_k) \leq \tilde{f}(x^k, \mu_k) + \sigma_1(\nabla_x \tilde{f}(x^k, \mu_k), x^{k+1} - x^k) \quad (3.2)$$

and

$$\alpha_k \geq \gamma_1, \quad \text{or} \quad \alpha_k \geq \gamma_2 \bar{\alpha}_k > 0, \quad (3.3)$$

such that  $\bar{x}^{k+1} = x^k(\bar{\alpha}_k)$  satisfies

$$\bar{f}(\bar{x}^{k+1}, \mu_k) > \bar{f}(x^k, \mu_k) + \sigma_2(\nabla_x \bar{f}(x^k, \mu_k), \bar{x}^{k+1} - x^k). \quad (3.4)$$

3. If  $\frac{\|\bar{x}^{k+1} - x^k\|}{\alpha_k} > \hat{\gamma}\mu_k$ , set  $\mu_{k+1} = \mu_k$ . Otherwise, go to step 4.

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

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

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

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

$$\min\{\bar{f}(x, \mu_k) : x \in X\}, \quad (3.5)$$

that is,  $x^k$  satisfies

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

If  $x^k$  is not a stationary point of (3.5), then from the differentiability of  $\bar{f}(\cdot, \mu_k)$  and analysis in [2], we can show that there exists  $\alpha_k > 0$  such that (3.2) and (3.3) hold.

### 3.1 Smoothing function for ERM

Now we show that smoothing functions  $\bar{f}$  derived from the Chen-Mangasarian smoothing function [3] satisfy Definition 3.1. Let  $\rho : R \rightarrow [0, \infty)$  be a piecewise continuous density function satisfying

$$\rho(s) = \rho(-s) \quad \text{and} \quad \kappa := \int_{-\infty}^{\infty} |s|\rho(s)ds < \infty. \quad (3.6)$$

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

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

is built as

$$\phi(a, b, \mu) = a - \int_{-\infty}^{\infty} \max(0, a - b - \mu s)\rho(s)ds. \quad (3.7)$$

Employing (3.7) to  $f$ , we obtain the smoothing function  $\bar{f}$

$$\bar{f}(x, \mu) = 2E[\bar{\theta}_\omega(x, \mu)], \quad (3.8)$$

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

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

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

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

Let  $\partial_C H_\omega$  be the C-generalized Jacobian of  $H_\omega$  defined by

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

where  $\partial(H_\omega(x))_i$  is the Clarke generalized Jacobian [7] of  $(H_\omega(x))_i$  for  $i = 1, 2, \dots, n$ .

**Lemma 3.1** [5] For any  $\omega \in \Omega$ ,  $x \in R^n$  and  $\mu \in R_{++}$ ,

- (i)  $\|\bar{H}_\omega(x, \mu) - H_\omega(x)\| \leq \sqrt{n}\kappa\mu$ .
- (ii)  $\lim_{\mu \downarrow 0} \text{dist}((\nabla_x \bar{H}_\omega(x, \mu))^T, \partial_C H_\omega(x)) = 0$ .

Let us denote  $h : R^n \rightarrow R$  by  $h(x) = x^T x$ , then we have  $f = E[h \circ H_\omega]$ .

**Proposition 3.1** For any  $x \in R_+^n$ , and any sequence  $\{x^k\} \subset R_+^n$  and  $\{\mu_k\} \subset R_{++}$ ,

$$\lim_{x^k \rightarrow x, \mu_k \downarrow 0} \text{dist}((\nabla_x \bar{f}(x^k, \mu_k))^T, E[\nabla h \circ \partial_C H_\omega(x)]) = 0.$$

Moreover, if  $f$  is differentiable at  $x \in R_+^n$ , then

$$\{\nabla f(x)\} = E[\nabla h \circ \partial_C H_\omega(x)].$$

**Proposition 3.2** The sequence  $\{f(x^k)\}$  generated by Algorithm 2.1 is bounded.

**Theorem 3.1** Let  $\{x^{k_j}\}$  be an infinite subsequence generated by Algorithm 2.1 with  $\mu_{k_j} > \mu_{k_{j+1}}$  for any  $j$ . Then for any accumulation point  $x^*$  of  $\{x^{k_j}\}$ , there is  $V \in E[\nabla h \circ \partial_C H_\omega(x^*)]$  such that

$$(V, x^* - z) \leq 0 \quad \text{for all } z \in R_+^n.$$

If  $f$  is differentiable at  $x^*$ , then  $x^*$  is a stationary point of problem (1.5).

## References

- [1] 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.
- [2] P.H. Calamai and J.J. Moré, Projected gradient method for linearly constrained problems, *Math. Program.*, 39(1987), pp. 93-116.
- [3] 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.
- [4] X. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Math. Oper. Res.*, 30(2005), pp. 1022-1038.
- [5] 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.
- [6] X. Chen, C. Zhang and M. Fukushima, Robust Solution of Monotone Stochastic Linear Complementarity Problems, to appear in *Math. Program.*.
- [7] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, (reprinted by SIAM, Philadelphia), 1990.
- [8] H. Fang, X. Chen and M. Fukushima, Stochastic  $R_0$  matrix linear complementarity problems, *SIAM J. Optim.*, 18(2007), pp. 482-506.
- [9] A.A. Goldstein, Convex programming in Hilbert space, *Bull. Amer. Soc.*, 70(1964), pp. 709-710.
- [10] K.C. Kiwiel, Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization, *SIAM J. Optim.*, 18(2007), pp. 379-388.
- [11] E.S. Levitin and B.T. Polyak, Constrained minimization problems, *USSR. Comput. Math. Math. Phys.*, 6(1966), pp. 1-50.
- [12] 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.
- [13] N.Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer-Verlag, Berlin, 1985.