SUBGRADIENT METHODS FOR SADDLE POINT PROBLEMS OF QUASICONVEX OPTIMIZATION

YAOHUA HU, XIAOQI YANG, AND CARISA KWOK WAI YU

Abstract. In this paper, we consider a saddle point problem of quasiconvex optimization, which is to find a saddle point of a quasiconvex-quasiconcave function over a closed convex set. We propose a subgradient method to approach the saddle value and investigate its convergence property under a general assumption of the Hölder condition of order p and using the constant or diminishing stepsize rules. To avoid the difficulty in calculating the exact subgradient, we further propose a stochastic subgradient method, where a random noisy quasi-subgradient is employed in the subgradient approach in place of the exact quasi-subgradient. The convergence analysis shows that the stochastic subgradient method shares the same convergence behavior as that of the exact subgradient method with probability 1.

Our motivation also comes from a major application of subgradient methods: the Lagrangian duality. In general, the subgradient of the dual function is usually difficult to calculate in the quasiconvex setting, since it requires to estimate values of dual functions, that is to solve many nonconvex optimization problems. This hinders the implementation of the dual subgradient method for (quasiconvex) optimization problems. To overcome this obstacle, applying the proposed subgradient methods to solve the resulting primal-dual problem, we obtain a primal-dual subgradient method to approximate a saddle value of the Lagrangian function, which only requires to directly calculate the subgradient of the Lagrange function without solving any auxiliary problem, and thus avoiding the difficulty in computing the subgradient of the dual function.

2010 *Mathematics Subject Classification.* Primary, 65K05, 90C26; Secondary, 49M37, 74S60.

Key words and phrases. Saddle point problem; Subgradient method; Quasiconvex optimization; Primal-dual approach; Stochastic approximate.

This author's work was supported in part by the National Natural Science Foundation of China (grant 11526144), Natural Science Foundation of Guangdong (grant 2016A030310038), and Foundation for Distinguished Young Talents in Higher Education of Guangdong (grant 2015KQNCX145).

This author's work was supported in part by the Research Grants Council of Hong Kong (PolyU 152167/15E) and the National Natural Science Foundation of China (grant 11431004).

This author's work described in this paper was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS14/P03/14 and UGC/FDS14/P02/15).

1. Introduction

Saddle point problem is an important issue arising in a wide range of areas, such as duality theory of constrained optimization, zero-sum games and general equilibrium theory. The general saddle point problem is to find a saddle point (x^*, y^*) of a convex-concave function $\mathcal{F}(x, y)$ on closed convex sets $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ such that

$$\mathcal{F}(x^*, y) \leq \mathcal{F}(x^*, y^*) \leq \mathcal{F}(x, y^*) \quad \text{for any } x \in X \text{ and } y \in Y. \tag{1.1}$$

Subgradient methods provide a popular and practical decentralized computational technique for solving non-smooth saddle point and optimization problems in many disciplines. Subgradient methods originated with the works of Polyak [30] and Ermoliev [8], many extensions and generalizations have been considered and numerous applications have been proposed; see [4, 15, 16, 21, 26, 27, 29, 33] and references therein. Nowadays, because of the simple formulation and low storage requirement, subgradient methods remain important for solving nonsmooth or stochastic optimization problems, especially for large-scale problems.

Many works have been devoted to the study of subgradient methods for approaching the saddle value (or a saddle point) of a convex-concave function; see, e.g., [1, 19, 24, 28, 29, 32]. In particular, Larsson et al. [24] and Nesterov [29] studied convergence properties of primal-dual subgradient methods along with the averaging scheme and using the diminishing stepsize rule. Sen and Sherali [32] proposed a class of primal-dual subgradient methods that employed Lagrangian dual functions along with suitable penalty functions, and proved that the sequence of primal-dual iterates converges to a saddle point when using several classical types of penalty functions. Recently, Nedić and Ozdaglar [28] adopted the constant stepsize rule and estimated the convergence rate of the sequence generated by subgradient methods to the saddle value per iteration.

In recent years, a central challenge to many fields of science and engineering involves nonconvex optimization in high-dimensional spaces. One of the most important types beyond convex optimization is the quasiconvex optimization, which have many important applications in various areas, such as economics, engineering, management science and various applied sciences; see [3, 7, 14, 34] and references therein. However, the study of subgradient methods for solving quasiconvex optimization problems is limited. In particular, Kiwiel [20] studied convergence properties of the exact subgradient method for solving quasiconvex optimization problems in the use of the diminishing stepsize rule. By extending this work and further using the constant stepsize rule, Hu et al. [17] proposed a generic inexact subgradient method to solve quasiconvex optimization problems, and studied the influence of the deterministic noise by describing convergence results in both objective values and iterates and finite convergence to the approximate optimality. Furthermore, Hu et al. [18] studied convergence properties of the stochastic subgradient method for solving quasiconvex optimization problems. On the other hand, the modified dual subgradient algorithms were investigated in Gasimov [11] and Burachik et al. [6] for solving a general nonconvex optimization problem with equality constraints by virtue of a sharp augmented Lagrangian.

Extending to the quasiconvex setting, in this paper, we consider the following saddle point problem in the quasiconvex setting

$$\min_{x\in X}\max_{y\in Y}\ \mathcal{F}(x,y) \tag{1.2}$$

where $\mathcal{F}:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ is a quasiconvex-quasiconcave function, and X and Y are the nonempty, closed and convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. In particular, $\mathcal{F}(\cdot,y)$ is quasiconvex for any $y\in Y$, and $\mathcal{F}(x,\cdot)$ is quasiconvex for any $x\in X$. It is clear that a solution of problem (1.2) is a pair $(x^*,y^*)\in X\times Y$ satisfying (1.1). Such a vector pair (x^*,y^*) is also referred to as a saddle point of the function $\mathcal{F}$ on the set $X\times Y$.

In this paper, we will introduce a subgradient method to solve the constrained minimax problem (1.2) of a quasiconvex-quasiconcave function, and explore convergence properties of the subgradient method when using the constant or diminishing stepsize rules. Lacking the convexity assumed in [28], the quasiconvex optimization is more difficult to deal with, and the main technical challenge of convergence analysis of the subgradient method is to establish a proper basic inequality, which is a key tool in the literature of subgradient methods. To this end, we will adopt the quasi-subdifferential and assume the Hölder condition, as in [17]. We will show that the subgradient method converges to the optimal value of problem (1.2) within some tolerance (given in terms of the stepsize) when using the constant stepsize rule, and exactly converges to the optimal value in the use of the diminishing stepsize rule.

Due to errors in measurements or uncertainty in problem data, the direct application of the exact subgradient may not be meaningful. In such situations, we propose a stochastic subgradient method, where the noisy (unbiased) quasi-subgradient (see Definition 2.2) is adopted in each iteration, to solve problem (1.2). The convergence theory presented in this paper extends the one shown in [18] to the saddle point problem or the constrained minimax problem, and improves the one reported in [17]. In particular, our convergence results show that the stochastic subgradient method shares the same convergence behavior as that of the exact subgradient method (see Theorems 3.2 and 3.3) with probability 1, and it achieves a better tolerance than that of the inexact subgradient method reported in [17] (see Theorems 4.3 and 4.4).

The motivation of our work also stems from a major application of subgradient methods, which is to the Lagrangian function of a constrained optimization problem. For example, we usually face the following constrained quasiconvex optimization problem (the primal problem)

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \le 0, \\ & x \in X, \end{array} \tag{1.3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a quasiconvex function, $g = (g_1, \ldots, g_m)^\top$ with each $g_i : \mathbb{R}^n \to \mathbb{R}$ being quasiconvex, and $X \subseteq \mathbb{R}^n$ is a closed and convex set. Although the primal subgradient method has been investigated in [17, 18, 20] to solve the primal problem (1.3), but all these works are under an assumption that the projection onto the feasible set is easy to compute. However, the projection onto $\{x : g(x) \le 0\} \cap X$, the feasible set associated with (1.3), is not easily implemented in general. An alternative and popular technique for (1.3) is to adopt the dual approach, which is defined by the Lagrangian relaxation of the inequality constraint $g(x) \le 0$ and is given by

$$\begin{array}{ll} \max & q(\mu) \\ \text{s.t.} & \mu \in \mathbb{R}^m_+, \end{array} \tag{1.4}$$

where the dual function $q : \mathbb{R}^m \to \mathbb{R}$ is defined by

$$q(\mu) = \inf_{x \in X} \mathcal{L}(x, \mu) \quad \text{via} \quad \mathcal{L}(x, \mu) = f(x) + \mu^\top g(x).$$

The strong duality between (1.3) and (1.4) has been established in [10, Theorems 9 and 14] for quasiconvex optimization under some mild conditions. The dual subgradient method has been investigated in [6, 13, 27] for convex optimization, while the implementation of dual subgradient method deeply relies on the assumption that the subgradient of the dual function can be estimated efficiently. However, the dual function $q(\cdot)$ and its subgradient are difficult to calculate in the quasiconvex setting, since it requires to solve a nonconvex optimization problem. This hinders the implementation of the dual subgradient method for quasiconvex optimization. To overcome this obstacle, we consider the following primal-dual problem

$$\max_{\mu \in \mathbb{R}^m_+} \min_{x \in X} \mathcal{L}(x, \mu). \tag{1.5}$$

It is clear that $\mathcal{L}(x, \cdot)$ is quasiconcave (in particular, it is linear) for any $x \in X$. For some classical types of quasiconvex functions, such as fractional functions (see, e.g., [34]), $\mathcal{L}(\cdot, \mu)$ is quasiconvex for any $\mu \in \mathbb{R}^m_+$. An example is that

$$f(x) = \frac{p(x)}{r(x)} \quad \text{and} \quad g(x) = \frac{q(x)}{r(x)},$$

where $p(\cdot) > 0$ and $q(\cdot) >$ are convex, and $r(\cdot) > 0$ is concave. Directly applying the proposed subgradient methods to solve the primal-dual problem (1.5), we obtain a primal-dual subgradient method to approximate a saddle value (or a saddle point) of the Lagrangian function. In contrast to the dual subgradient method, the primal-dual subgradient method approaches a saddle value without solving any auxiliary problems, and thus avoids the difficulty in computing subgradients of the dual function.

The paper is organized as follows. In section 2, we present the notation and preliminary results used in this paper. In section 3, we introduce a subgradient method to approximate a saddle value of quasiconvex-quasiconcave function, and investigate convergence properties of the subgradient method when using the constant or diminishing rules. In section 4, we propose a stochastic subgradient method and establish its convergence properties in sense of with probability 1.

2. Notation and preliminary results

We consider the n-dimensional Euclidean space $\mathbb{R}^n$ with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. In particular, we use $\mathbf{S}$ to denote the unit sphere centered at the origin. For $x \in \mathbb{R}^n$ and $Z \subseteq \mathbb{R}^n$, we use $\mathrm{dist}(x, Z)$ and $P_Z(x)$ to denote the Euclidean distance of x from Z and the classical metric projection of x onto Z, respectively, i.e.,

$$\mathrm{dist}(x, Z) := \inf_{z\in Z} \|x - z\| \quad \text{and} \quad P_Z(x) := \arg\min_{z\in Z} \|x - z\|.$$

A function $h : \mathbb{R}^n \to \mathbb{R}$ is said to be quasiconvex if

$$h((1-\alpha)x + \alpha y) \le \max\{h(x), h(y)\} \quad \text{for any } x, y \in \mathbb{R}^n \text{ and } \alpha \in [0, 1];$$

h is said to be quasiconcave if $-h$ is quasiconvex, that is,

$$h((1-\alpha)x + \alpha y) \ge \min\{h(x), h(y)\} \quad \text{for any } x, y \in \mathbb{R}^n \text{ and } \alpha \in [0, 1].$$

For any $\alpha \in \mathbb{R}$, we denote the level sets of h by

$$\begin{aligned} \mathrm{lev}_{<\alpha}h &:= \{x \in \mathbb{R}^n : h(x) < \alpha\}, \quad \mathrm{lev}_{\le\alpha}h := \{x \in \mathbb{R}^n : h(x) \le \alpha\},\\ \mathrm{lev}_{>\alpha}h &:= \{x \in \mathbb{R}^n : h(x) > \alpha\}, \quad \mathrm{lev}_{\ge\alpha}h := \{x \in \mathbb{R}^n : h(x) \ge \alpha\}. \end{aligned}$$

It is well-known that h is quasiconvex if and only if $\mathrm{lev}_{<\alpha}h$ (and/or $\mathrm{lev}_{\le\alpha}h$) is convex for any $\alpha \in \mathbb{R}$, and that h is quasiconcave if and only if $\mathrm{lev}_{>\alpha}h$ (and/or $\mathrm{lev}_{\ge\alpha}h$) is convex for any $\alpha \in \mathbb{R}$.

The subdifferential of quasiconvex functions is an important issue of quasiconvex optimization, and several types of subdifferentials of quasiconvex functions have been introduced in the literature; see, e.g., [2, 12, 17, 20]. In particular, Kiwiel [20] and Hu et al. [17] introduced a quasi-subdifferential, which is a normal cone to a strict sublevel set of the quasiconvex function, and utilized such a subgradient in their proposed subgradient methods. We recall the definition of quasi-subdifferential as follows.

Definition 2.1. Let $h : \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathbb{R}^n$.

(a) Assume that h is quasiconvex. The quasi-subdifferential of h at x is defined by

$$\partial h(x) = \{g : \langle g, y - x\rangle \le 0, \forall y \in \mathrm{lev}_{<h(x)}h\}.$$

(b) Assume that h is quasiconcave. The quasi-subdifferential of h at x is defined by

$$\partial h(x) = \{g : \langle g, y - x\rangle \ge 0, \forall y \in \mathrm{lev}_{>h(x)}h\}.$$

Any vector $g \in \partial h(x)$ is called a quasi-subgradient of h at x.

Allowing a random noise, the following noisy quasi-subgradient was introduced and employed in the stochastic subgradient method in [18].

Definition 2.2. Let $h : \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathbb{R}^n$, and let $\tilde{g}(x) \in \mathbb{R}^n$ be a random vector.

(a) Assume that h is quasiconvex. $\tilde{g}(x)$ is called a noisy (unbiased) quasi-subgradient of h at x if $\mathbf{E}\tilde{g}(x) \in \partial h(x)$, that is,
$$\mathbf{E}\langle \tilde{g}(x), y - x\rangle \leq 0 \quad \text{for any } y \in \mathrm{lev}_{<h(x)}h,$$
where $\mathbf{E}(\cdot)$ denotes the expectation of a random variable, and $\mathbf{E}\langle \tilde{g}(x), y - x\rangle = \langle \mathbf{E}\tilde{g}(x), y - x\rangle$.

(b) Assume that h is quasiconcave. $\tilde{g}(x)$ is called a noisy quasi-subgradient of h at x if
$$\mathbf{E}\langle \tilde{g}(x), y - x\rangle \geq 0 \quad \text{for any } y \in \mathrm{lev}_{>h(x)}h.$$

(c) $\tilde{g}(x)$ is called a unit noisy quasi-subgradient of h at x if it is a noisy quasi-subgradient of f at x and satisfies $\|\mathbf{E}\tilde{g}(x)\| = 1$.

The Hölder condition is a critical assumption for the convergence study of numerical algorithms in quasiconvex optimization. The Hölder condition of order p is used to describe some properties of the quasi-subgradient in [22, 23], and assumed in [17, 18] to investigate convergence properties of the inexact quasi-subgradient method and the stochastic subgradient method.

Definition 2.3. Let $p > 0$, $L > 0$ and $\bar{x} \in \mathbb{R}^n$. $h : \mathbb{R}^n \to \mathbb{R}$ is said to satisfy the Hölder condition of order p with modulus L at $\bar{x}$ if
$$|h(x) - h(\bar{x})| \leq L\|x - \bar{x}\|^p \quad \text{for any } x \in \mathbb{R}^n. \tag{2.1}$$
h is said to satisfy the Hölder condition of order p with modulus L on X if (2.1) holds for any $\bar{x} \in X$.

A bounded subgradient assumption is usually assumed in the literature of subgradient methods for convex optimization; see, e.g., [21, 26, 27, 28]. This assumption can be guaranteed when the function h is globally Lipschitz, that is,
$$|h(x) - h(\bar{x})| \leq L\|x - \bar{x}\| \quad \text{for any } x \in \mathbb{R}^n.$$
More precisely, the Hölder condition of order 1 is equivalent to the bounded subgradient assumption whenever h is convex (see [17]). Moreover, we provide some examples of quasiconvex-quasiconcave functions that satisfy the Hölder condition.

Example 2.4. (i) $f(x, y) := \sqrt{|x|} - \sqrt{|y|}$ satisfies the Hölder condition of $\frac{1}{2}$ with modulus 1.
(ii) $f(x, y) := \|x\|^p - \|y\|^p$ with $p \in (0, 1)$ satisfies the Hölder condition of p with modulus 1.
(iii) $f(x, y) := \|x\|_p^p - \|y\|_p^p$, where $p \in (0, 1)$ and $\|x\|_p^p := \sum_{i=1}^n |x_i|^p$, satisfies the Hölder condition of p with modulus 1.

The following lemma describes an important property of a quasiconvex function that satisfies the Hölder condition. This property locally relates the quasi-subgradient with objective function values, which is a key tool to establish the basic inequality in convergence analysis. Items (i) and (ii) are taken from [23, Proposition 2.1] and [18, Lemma 2.4], respectively.

Lemma 2.5. *Let h be a quasiconvex function, X be a closed and convex set, and let X^* be set of minima of h on X. Let $p > 0$, $L > 0$ and $x \in X \setminus X^*$. Suppose that h satisfies the Hölder condition of order p with modulus L at some $x^* \in X^*$. Then it holds that*

(i) *Let $g \in \partial h(x) \cap \mathbf{S}$. Then*

$$\langle g, x - x^* \rangle \geq \left(\frac{h(x) - h(x^*)}{L} \right)^{\frac{1}{p}}.$$

(ii) *Let $\tilde{g}(x)$ be a unit noisy quasi-subgradient of h at x. Then*

$$\mathbf{E}\langle \tilde{g}(x), x - x^* \rangle \geq \left(\frac{h(x) - h(x^*)}{L} \right)^{\frac{1}{p}}.$$

3. Subgradient method for saddle point problem

The aims of this section are to introduce a subgradient method to solve problem (1.2), and to investigate its convergence properties. The subgradient method for solving (1.2) is formally presented as follows.

Algorithm 3.1. Select initial points $x_0 \in X$ and $y_0 \in Y$, and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k and y_k, we calculate the unit quasi-subgradients of $\mathcal{F}$ at (x_k, y_k) with respect to x and y, that is, compute

$$\mathcal{F}_x(x_k, y_k) \in \partial_x \mathcal{F}(x_k, y_k) \cap \mathbf{S} \quad \text{and} \quad \mathcal{F}_y(x_k, y_k) \in \partial_y \mathcal{F}(x_k, y_k) \cap \mathbf{S},$$

and update x_{k+1} and y_{k+1}, respectively, by

$$x_{k+1} = P_X(x_k - v_k \mathcal{F}_x(x_k, y_k)), \tag{3.1}$$

$$y_{k+1} = P_Y(y_k + v_k \mathcal{F}_y(x_k, y_k)). \tag{3.2}$$

The stepsize rule has a critical effect on the convergence behavior and computational performance of subgradient methods. In this paper, we consider the following two typical stepsize rules.

(a) *Constant stepsize rule.*

$$v_k \equiv v (> 0).$$

(b) *Diminishing stepsize rule.*

$$v_k > 0, \quad \lim_{k \to \infty} v_k = 0, \quad \sum_{k=0}^{\infty} v_k = +\infty. \tag{3.3}$$

To study the convergence properties of our methods, we make the following assumption[1]:

- Let (x^*, y^*) be a saddle point of (1.2). Assume that $\mathcal{F}$ satisfies the Hölder condition of order $p > 0$ with modulus $L > 0$ on $X \times Y$.

We start the convergence analysis of Algorithm 3.1 by providing the following basic inequalities, which show the behaviour of the subgradient iterations. Item (i) is for the primal subgradient approach, and item (ii) is for the dual subgradient approach.

Lemma 3.1. *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by Algorithm 3.1. Then the following two assertions hold for any $k \geq 0$:*

(i) *If $\mathcal{F}(x_k, y_k) > \mathcal{F}(x^*, y_k)$, we have*

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2v_k \left(\frac{\mathcal{F}(x_k, y_k) - \mathcal{F}(x^*, y_k)}{L} \right)^{\frac{1}{p}} + v_k^2. \tag{3.4}$$

(ii) *If $\mathcal{F}(x_k, y_k) < \mathcal{F}(x_k, y^*)$, we have*

$$\|y_{k+1} - y^*\|^2 \leq \|y_k - y^*\|^2 - 2v_k \left(\frac{\mathcal{F}(x_k, y^*) - \mathcal{F}(x_k, y_k)}{L} \right)^{\frac{1}{p}} + v_k^2. \tag{3.5}$$

Proof. (i) In view of Algorithm 3.1 (cf. (3.1)), for any $k \geq 0$, it follows from the nonexpansive property of projection operator that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - v_k \mathcal{F}_x(x_k, y_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - 2v_k \langle \mathcal{F}_x(x_k, y_k), x_k - x^* \rangle + v_k^2. \end{aligned} \tag{3.6}$$

Since $\mathcal{F}(x_k, y_k) > \mathcal{F}(x^*, y_k)$, Lemma 2.5(i) is applicable (to $\mathcal{F}(\cdot, y_k)$, x_k, $\mathcal{F}_x(x_k, y_k)$ in place of h, x, g) to concluding that

$$\langle \mathcal{F}_x(x_k, y_k), x_k - x^* \rangle \geq \left(\frac{\mathcal{F}(x_k, y_k) - \mathcal{F}(x^*, y_k)}{L} \right)^{\frac{1}{p}}.$$

Hence, (3.6) is reduced to (3.4), and this completes the proof of (i).

(ii) Similarly, by (3.2) and the nonexpansive property of projection operator, for any $k \geq 0$, we obtain that

$$\begin{aligned} \|y_{k+1} - y^*\|^2 &\leq \|y_k + v_k \mathcal{F}_y(x_k, y_k) - y^*\|^2 \\ &= \|y_k - y^*\|^2 + 2v_k \langle \mathcal{F}_y(x_k, y_k), y_k - y^* \rangle + v_k^2. \end{aligned} \tag{3.7}$$

By the assumption that $\mathcal{F}(x_k, y_k) < \mathcal{F}(x_k, y^*)$, Lemma 2.5(i) is applicable (to $-\mathcal{F}(x_k, \cdot)$, y_k, $-\mathcal{F}_y(x_k, y_k)$ in place of h, x, g); hence it follows that

$$\langle -\mathcal{F}_y(x_k, y_k), y_k - y^* \rangle \geq \left(\frac{\mathcal{F}(x_k, y^*) - \mathcal{F}(x_k, y_k)}{L} \right)^{\frac{1}{p}}.$$

Therefore, (3.7) is reduced to (3.5), and the proof of (ii) is complete.

[1]Due to the structure of saddle point problem (1.2), this assumption can be weakened to be that $\mathcal{F}(\cdot, y)$ and $\mathcal{F}(x, \cdot)$ satisfy the Hölder condition of order $p > 0$ with modulus $L > 0$ on X for any $y \in Y$ and on Y for any $x \in X$, respectively.

□

By virtue of Lemma 3.1, we will provide the convergence results of Algorithm 3.1 when using the constant and diminishing stepsize rules in Theorems 3.2 and 3.3, respectively.

Theorem 3.2. *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by Algorithm 3.1 with the constant stepsize rule. Then*

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) - L\left(\frac{v}{2}\right)^p \le \mathcal{F}(x^*, y^*) \le \limsup_{k\to\infty} \mathcal{F}(x_k, y_k) + L\left(\frac{v}{2}\right)^p.$$

Proof. It is clear that the proofs of the above two inequalities follow a similar analysis, we only show the proof of the first inequality and omit that of the second one. To do this, we prove by contradiction, assuming to the contrary that

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) > \mathcal{F}(x^*, y^*) + L\left(\frac{v}{2}\right)^p.$$

Then there exist some $\delta > 0$ and $k_0 \in \mathbb{N}$ such that, for any $k \ge k_0$,

$$\mathcal{F}(x_k, y_k) > \mathcal{F}(x^*, y^*) + L\left(\frac{v}{2} + \delta\right)^p \ge \mathcal{F}(x^*, y_k) + L\left(\frac{v}{2} + \delta\right)^p.$$

Then Lemma 3.1(i) is applicable; hence, for any $k \ge k_0$, it follows from (3.4) that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\le \|x_k - x^*\|^2 - 2v\left(\tfrac{\mathcal{F}(x_k,y_k)-\mathcal{F}(x^*,y_k)}{L}\right)^{\frac{1}{p}} + v^2 \\ &\le \|x_k - x^*\|^2 - 2v\left(\tfrac{v}{2} + \delta\right) + v^2 \\ &= \|x_k - x^*\|^2 - 2v\delta. \end{aligned}$$

Summing the above inequality over $k = k_0, \ldots, n$, we have that

$$\|x_{n+1} - x^*\|^2 \le \|x_{k_0} - x^*\|^2 - 2(n+1-k_0)v\delta,$$

which yields a contradiction for sufficiently large n. Then we obtain the first inequality, and thus, the proof is complete. □

Using the diminishing stepsize rule, the tolerance in Theorem 3.2 vanishes and the following theorem is obtained.

Theorem 3.3. *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by Algorithm 3.1 with the diminishing stepsize rule. Then*

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) \le \mathcal{F}(x^*, y^*) \le \limsup_{k\to\infty} \mathcal{F}(x_k, y_k). \tag{3.8}$$

Proof. We only show the proof of the first inequality of (3.8) and omit that of the second one. To do this, we prove by contradiction, assuming to the contrary that

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) > \mathcal{F}(x^*, y^*).$$

Then there exist some $\delta > 0$ and $k_0 \in \mathbb{N}$ such that, for any $k \ge k_0$,

$$\mathcal{F}(x_k, y_k) > \mathcal{F}(x^*, y^*) + L\delta^p \ge \mathcal{F}(x^*, y_k) + L\delta^p.$$

Then Lemma 3.1(i) is applicable; hence, for any $k \geq k_0$, it follows from (3.4) that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2v_k \left(\tfrac{\mathcal{F}(x_k,y_k)-\mathcal{F}(x^*,y_k)}{L}\right)^{\frac{1}{p}} + v_k^2 \\ &\leq \|x_k - x^*\|^2 - 2v_k\delta + v_k^2. \end{aligned} \tag{3.9}$$

Since $\{v_k\}$ diminishes (cf. (3.3)), there exists some $k_\delta \in \mathbb{N}$ such that

$$v_k \leq \delta \quad \text{for any } k \geq k_\delta.$$

Hence, for any $k \geq \tilde{k} := \max\{k_0, k_\delta\}$, (3.9) is reduced to

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - v_k\delta. \tag{3.10}$$

Summing the above inequality over $k = \tilde{k}, \ldots, n$, we have that

$$\|x_{n+1} - x^*\|^2 \leq \|x_{\tilde{k}} - x^*\|^2 - \delta \sum_{k=\tilde{k}}^{n} v_k,$$

which yields a contradiction for sufficiently large n (since $\sum_{k=0}^{\infty} v_k = +\infty$). Thus we obtain the first inequality of (3.8), and the proof is complete. □

Remark 3.4. Note in Algorithm 3.1 that we adopt the uniform stepsize in the primal and dual subgradient approaches. More general, we can also utilize the mixed stepsizes in the subgradient method for solving problem (1.2), that is, (3.1) and (3.2) in Algorithm 3.1 are replaced by

$$x_{k+1} = P_X(x_k - \alpha_k \mathcal{F}_x(x_k, y_k)), \tag{3.11}$$

$$y_{k+1} = P_Y(y_k + \beta_k \mathcal{F}_y(x_k, y_k)), \tag{3.12}$$

where $\alpha_k > 0$ and $\beta_k > 0$ are the primal and dual stepsizes, respectively. Following the convergence analysis in Theorems 3.2 and 3.3, we can establish the following convergence results for the subgradient method (3.11)-(3.12):

(i) Assume $\alpha_k \equiv \alpha > 0$ and $\{\beta_k\}$ is the diminishing stepsize rule. Then it hods that

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) - L\left(\frac{\alpha}{2}\right)^p \leq \mathcal{F}(x^*, y^*) \leq \limsup_{k\to\infty} \mathcal{F}(x_k, y_k).$$

(ii) Assume $\{\alpha_k\}$ is the diminishing stepsize rule and $\beta_k \equiv \beta > 0$. Then it hods that

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) \leq \mathcal{F}(x^*, y^*) \leq \limsup_{k\to\infty} \mathcal{F}(x_k, y_k) + L\left(\frac{\beta}{2}\right)^p.$$

4. Stochastic subgradient method

Due to errors in measurements or uncertainty in problem data, the direct application of the exact subgradient may not be meaningful. In such situations, an alternative approach is to use a noisy estimate of the subgradient. Adopting a random noisy estimate as the true subgradient, the stochastic subgradient method was pioneered by Ermoliev [8, 9] and further developed by many scholars (e.g., [5, 25, 31]). Recently, Hu et al. [18] proposed a stochastic subgradient method to solve constrained quasiconvex optimization problems. Many convergence results of the stochastic subgradient method have been established in which the generated sequence could achieve the same convergence properties as that of the exact subgradient method with probability 1, because the random behavior help "average out" the statistical noise in subgradient evaluations.

Inspired by the ideas in [18] and references therein, this section aims at the study of the stochastic subgradient method for solving problem (1.2). The only difference between the stochastic subgradient method of this section and Algorithm 3.1 is that the stochastic noisy quasi-subgradients are employed in the subgradient approach in place of the exact quasi-subgradient. The stochastic subgradient method for solving (1.2) is formally presented as follows.

Algorithm 4.1. Select initial points $x_0 \in X$ and $y_0 \in Y$, and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k and y_k, we calculate the unit noisy quasi-subgradients $\tilde{\mathcal{F}}_x(x_k, y_k)$ and $\tilde{\mathcal{F}}_y(x_k, y_k)$ of $\mathcal{F}$ at (x_k, y_k) with respect to x and y, and update x_{k+1} and y_{k+1}, respectively, by

$$x_{k+1} = P_X(x_k - v_k \tilde{\mathcal{F}}_x(x_k, y_k)), \tag{4.1}$$

$$y_{k+1} = P_Y(y_k + v_k \tilde{\mathcal{F}}_y(x_k, y_k)). \tag{4.2}$$

We recall the supermartingale convergence theorem (see [5, Proposition 4.2]), which is useful in the convergence analysis of the stochastic subgradient method.

Lemma 4.1. *Let $\{Y_k\}$, $\{Z_k\}$ and $\{W_k\}$ be three sequences of nonnegative random variables, and let $\{\mathcal{V}_k\}$ be a sequence of sets of random variables such that $\mathcal{V}_k \subseteq \mathcal{V}_{k+1}$ for any $k \geq 0$. Suppose that the following conditions are satisfied for each $k \geq 0$:*

(a) *Y_k, Z_k and W_k are functions of the random variables in $\mathcal{V}_k$;*
(b) *$\mathbf{E}\{Y_{k+1} \mid \mathcal{V}_k\} \leq Y_k - Z_k + W_k$;*
(c) *$\sum_{k=0}^{\infty} W_k < \infty$.*

Then $\sum_{k=0}^{\infty} Z_k < \infty$, and the sequence $\{Y_k\}$ converges to a nonnegative random variable Y, with probability 1.

Now we provide in the following lemma some basic inequalities, which show a significant property of a stochastic subgradient iteration.

Lemma 4.2. *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by Algorithm 4.1. Fix some $n \in \mathbb{N}$, and let $\mathcal{V}_n := \{x_0, y_0, x_1, y_1, \ldots, x_n, y_n\}$. Then the following two assertions are true:*

(i) *If $\mathcal{F}(x_n, y_n) > \mathcal{F}(x^*, y_n)$, we have*

$$\mathbf{E}\left\{\|x_{n+1} - x^*\|^2 \mid \mathcal{V}_n\right\} \leq \|x_n - x^*\|^2 - 2v_n \left(\frac{\mathcal{F}(x_n, y_n) - \mathcal{F}(x^*, y_n)}{L}\right)^{\frac{1}{p}} + v_n^2.$$

(ii) *If $\mathcal{F}(x_n, y_n) < \mathcal{F}(x_n, y^*)$, we have*

$$\mathbf{E}\left\{\|y_{n+1} - y^*\|^2 \mid \mathcal{V}_n\right\} \leq \|y_n - y^*\|^2 - 2v_n \left(\frac{\mathcal{F}(x_n, y^*) - \mathcal{F}(x_n, y_n)}{L}\right)^{\frac{1}{p}} + v_n^2.$$

Proof. In view of Algorithm 4.1 (cf. (4.1)) and by the nonexpansive property of projection operator, we have that

$$\begin{aligned}\|x_{n+1} - x^*\|^2 &\leq \|x_n - v_n\tilde{\mathcal{F}}_x(x_n, y_n) - x^*\|^2 \\ &= \|x_n - x^*\|^2 - 2v_n\langle\tilde{\mathcal{F}}_x(x_n, y_n), x_n - x^*\rangle + v_n^2.\end{aligned}$$

By taking the conditional expectation with respect to $\mathcal{V}_n$, it follows that

$$\begin{aligned}\mathbf{E}\{\|x_{n+1} - x^*\|^2 \mid \mathcal{V}_n\} &\leq \|x_n - x^*\|^2 - 2v_n\mathbf{E}\{\langle\tilde{\mathcal{F}}_x(x_n, y_n), x_n - x^*\rangle \mid \mathcal{V}_n\} + v_n^2 \\ &\leq \|x_n - x^*\|^2 - 2v_n \left(\tfrac{\mathcal{F}(x_n,y_n) - \mathcal{F}(x^*,y_n)}{L}\right)^{\frac{1}{p}} + v_n^2,\end{aligned}$$

where the last inequality follows from Lemma 2.5(ii) (to $\mathcal{F}(\cdot, y_n)$, x_n, $\tilde{\mathcal{F}}_x(x_n, y_n)$ in place of h, x, $\tilde{g}(x)$). Thus, we obtained (i) and can prove (ii) by taking the similar analysis. $\square$

By virtue of Lemma 4.2, we will establish in Theorems 4.3 and 4.4 the convergence results of Algorithm 4.1 for the constant and diminishing stepsize rules, respectively.

Theorem 4.3. *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by Algorithm 4.1 with the constant stepsize rule. Then it holds, with probability 1, that*

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) - L\left(\frac{v}{2}\right)^p \leq \mathcal{F}(x^*, y^*) \leq \limsup_{k\to\infty} \mathcal{F}(x_k, y_k) + L\left(\frac{v}{2}\right)^p. \tag{4.3}$$

Proof. We only show the proof of the first inequality of (4.3) and omit that of the second one. To do this, fix $\delta > 0$ and define a feasible level set $(X \otimes Y)_\delta$ by

$$(X \otimes Y)_\delta := \left\{(x, y) \in X \times Y : \mathcal{F}(x, y) < \mathcal{F}(x^*, y^*) + L\left(\frac{v}{2} + \delta\right)^p\right\}, \tag{4.4}$$

and let $(x_\delta, y_\delta) \in (X \otimes Y)_\delta$. We construct a new sequence $\{(\hat{x}_k, \hat{y}_k)\}$ by $(\hat{x}_0, \hat{y}_0) := (x_0, y_0)$, and

$$\begin{cases}\hat{x}_{k+1} := P_X(\hat{x}_k - v_k\tilde{\mathcal{F}}_x(\hat{x}_k, \hat{y}_k)) \\ \hat{y}_{k+1} := P_Y(\hat{y}_k + v_k\tilde{\mathcal{F}}_y(\hat{x}_k, \hat{y}_k))\end{cases} \quad \text{if } (\hat{x}_k, \hat{y}_k) \notin (X \otimes Y)_\delta;$$

otherwise, $(\hat{x}_{k+1}, \hat{y}_{k+1}) := (x_\delta, y_\delta)$. Then the sequence $\{(\hat{x}_k, \hat{y}_k)\}$ is identical to $\{(x_k, y_k)\}$, except that once $(\hat{x}_k, \hat{y}_k)$ enters $(X \otimes Y)_\delta$ and then $\{(\hat{x}_k, \hat{y}_k)\}$ terminates with $(x_\delta, y_\delta) \in (X \otimes Y)_\delta$. Assume that $(\hat{x}_k, \hat{y}_k) \notin (X \otimes Y)_\delta$ for any k and let $\hat{\mathcal{V}}_k := \{\hat{x}_0, \hat{y}_0, \hat{x}_1, \hat{y}_1, \ldots, \hat{x}_k, \hat{y}_k\}$. It follows from (4.4) that

$$\mathcal{F}(\hat{x}_k, \hat{y}_k) \geq \mathcal{F}(x^*, y^*) + L\left(\frac{v}{2} + \delta\right)^p \geq \mathcal{F}(x^*, \hat{y}_k) + L\left(\frac{v}{2} + \delta\right)^p,$$

and then Lemma 4.2(i) is applicable; hence, for any k, we obtain that

$$\begin{aligned}\mathbf{E}\left\{\|\hat{x}_{k+1}-x^*\|^2 \mid \hat{\mathcal{V}}_k\right\} &\le \|\hat{x}_k-x^*\|^2 - 2v\left(\tfrac{\mathcal{F}(\hat{x}_k,\hat{y}_k)-\mathcal{F}(x^*,\hat{y}_k)}{L}\right)^{\frac{1}{p}} + v^2 \\ &\le \|\hat{x}_k-x^*\|^2 - 2v\delta.\end{aligned}$$

Then Lemma 4.1 is applicable; hence one concludes that $\sum_{k=0}^{\infty} 2v\delta < \infty$ with probability 1, which is impossible. Therefore, $(\hat{x}_k,\hat{y}_k) \notin (X\otimes Y)_\delta$ only occurs finitely many times, and $(\hat{x}_k,\hat{y}_k) \in (X\otimes Y)_\delta$ for any k $(\ge N)$. Consequently, for the original sequence $\{x_k\}$, it holds, with probability 1, that

$$\liminf_{k\to\infty} \mathcal{F}(x_k,y_k) \le \mathcal{F}(x^*,y^*) + L\left(\frac{v}{2}+\delta\right)^p.$$

Since $\delta > 0$ is arbitrary, we arrive at the first inequality of (4.3), and the proof is complete. □

Theorem 4.3 shows the convergence of Algorithm 4.1 to the optimal value within some tolerance given in terms of the constant stepsize with probability 1. This tolerance, $L\left(\frac{v}{2}\right)^p$, is the same as the one obtained in Theorem 3.2, and is smaller than the one reported in [17, Theorem 3.1] for the inexact subgradient method that is expressed as $L\left(Rd + \frac{v}{2}(1+R)^2\right)^p + \epsilon$. This shows the advantage of adopting the randomized noise in subgradient methods.

Theorem 4.4. *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by Algorithm 4.1 with the diminishing stepsize rule. Then*

$$\liminf_{k\to\infty} \mathcal{F}(x_k,y_k) \le \mathcal{F}(x^*,y^*) \le \limsup_{k\to\infty} \mathcal{F}(x_k,y_k) \quad \textit{with probability 1.}$$

Proof. The proof of this theorem adopts the property of the diminishing stepsize rule (cf. (3.3)) and a line of analysis similar to that of Theorem 4.3. Hence we omit the details. □

Theorem 4.4 describes the exact convergence of the stochastic subgradient method for solving (1.2) when using the diminishing stepsize rule, which shares the same convergence property as that of the exact subgradient method (see Theorem 3.3) with probability 1.

Remark 4.5. Similar to Remark 3.4, we can also adopt the mixed stepsizes in the stochastic subgradient method, that is, (4.1) and (4.2) in Algorithm 4.1 are replaced by

$$x_{k+1} = P_X(x_k - \alpha_k \tilde{\mathcal{F}}_x(x_k,y_k)), \tag{4.5}$$

$$y_{k+1} = P_Y(y_k + \beta_k \tilde{\mathcal{F}}_y(x_k,y_k)). \tag{4.6}$$

The following convergence results for the subgradient method (4.5)-(4.6) follows form Theorems 4.3 and 4.4.

(i) Assume $\alpha_k \equiv \alpha > 0$ and $\{\beta_k\}$ is the diminishing stepsize rule. Then it hods, with probability 1, that

$$\liminf_{k\to\infty} \mathcal{F}(x_k,y_k) - L\left(\frac{\alpha}{2}\right)^p \le \mathcal{F}(x^*,y^*) \le \limsup_{k\to\infty} \mathcal{F}(x_k,y_k).$$

(ii) Assume $\{\alpha_k\}$ is the diminishing stepsize rule and $\beta_k \equiv \beta > 0$. Then it hods, with probability 1, that

$$\liminf_{k\to\infty} \mathcal{F}(x_k, y_k) \leq \mathcal{F}(x^*, y^*) \leq \limsup_{k\to\infty} \mathcal{F}(x_k, y_k) + L\left(\frac{\beta}{2}\right)^p.$$

5. Conclusion and future work

In this paper, we have proposed a subgradient method to solve a saddle point problem or a minimax problem of a quasiconvex-quasiconcave function over a closed convex set. The convergence theory to approach the saddle value has been established under the assumption of the Hölder condition of order p and by using the constant and diminishing stepsize rules. To adjust the uncertain noise in applications, we have proposed a stochastic subgradient method and provided its convergence analysis showing that the stochastic subgradient method shares the same convergence behavior as that of the exact subgradient method with probability 1.

Many questions maintain still open in the study of subgradient methods for solving saddle point problems of quasiconvex optimization. In our convergence study, only the convergence of objective values is provided, while the convergence of iterates is absent at this moment. Furthermore, in many applications, the computation error stems from practical considerations, and is inevitable in the computing process. The computation error usually gives rise to the calculation of an approximate subgradient, and the inexact subgradient method meets the requirement of applications. The convergence study of the inexact subgradient method has a significant influence in spreading applications of numerical optimization.

Acknowledgment. The authors are grateful to the anonymous reviewer for his/her valuable suggestions and remarks that helped to improve the quality of the paper.

References

[1] K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in Linear and Non-Linear Programming*, Stanford University Press, Stanford, 1958.
[2] D. Aussel and A. Daniilidis, *Normal characterization of the main classes of quasiconvex functions*, Set-Valued Anal. **8** (2000), 219–236.
[3] M. Avriel, W. E. Diewert, S. Schaible and I. Zang, *Generalized Concavity*, Plenum Press, New York, 1988.
[4] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Cambridge, 1999.
[5] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.
[6] R. S. Burachik, R. N. Gasimov, N. A. Ismayilova and C. Y. Kaya, *On a modified subgradient algorithm for dual problems via sharp augmented Lagrangian*, J. Global Optim. **34** (2006), 55–78.
[7] J.-P. Crouzeix, J.-E. Martinez-Legaz and M. Volle, *Generalized Convexity, Generalized Monotonicity*, Kluwer Academic Publishers, Dordrecht, 1998.

[8] Y. M. Ermoliev, *Methods of solution of nonlinear extremal problems*, Cybern. Syst. Anal. **2** (1966), 1–14.

[9] Y. M. Ermoliev, *Stochastic Programming Methods*, Nauka, Moscow, 1976.

[10] D. Fang, X. Luo and X. Wang, *Strong and total Lagrange dualities for quasiconvex programming*, Journal of Applied Mathematics **2014** (2014), Article ID 453912.

[11] R. N. Gasimov, *Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming*, J. Global Optim. **24** (2002), 187–203.

[12] H. J. Greenberg and W. P. Pierskalla, *Quasiconjugate functions and surrogate duality*, Cahiers Centre Études Recherche Opertionnelle, **15** (1973), 437–448.

[13] E. Gustavsson, M. Patriksson and A.-B. Strömberg, *Primal convergence from dual subgradient methods for convex optimization*, Math. Program. **150** (2014), 365–390.

[14] N. Hadjisavvas, S. Komlósi and S. Schaible, *Handbook of Generalized Convexity and Generalized Monotonicity*, Springer-Verlag, New York, 2005.

[15] Y. H. Hu, C. Li and X. Q. Yang, *On convergence rates of linearized proximal algorithms for convex composite optimization with applications*, SIAM J. Optim. **26** (2016), 1207–1235.

[16] Y. H. Hu, C.-K. Sim and X. Q. Yang, *A subgradient method based on gradient sampling for solving convex optimization problems*, Numer. Func. Anal. Opt. **36** (2015), 1559–1584.

[17] Y. H. Hu, X. Q. Yang and C.-K. Sim, *Inexact subgradient methods for quasi-convex optimization problems*, European J. Oper. Res., **240** (2015), 315–327.

[18] Y. H. Hu, C. K. W. Yu and C. Li, *Stochastic subgradient method for quasi-convex optimization problems*, J. Nonlinear Convex Anal. **17** (2016), 711–724.

[19] M. Kallio and C. H. Rosa, *Large-scale convex optimization via saddle point computation*, Oper. Res. **47** (1999), 93–101.

[20] K. C. Kiwiel, *Convergence and efficiency of subgradient methods for quasiconvex minimization*, Math. Program. **90** (2001), 1–25.

[21] K. C. Kiwiel, *Convergence of approximate and incremental subgradient methods for convex optimization*, SIAM J. Optim. **14** (2004), 807–840.

[22] I. V. Konnov, *On properties of supporting and quasi-supporting vectors*, J. Math. Sci. **71** (1994), 2760–2763.

[23] I. V. Konnov, *On convergence properties of a subgradient method*, Optim. Method. Softw. **18** (2003), 53–62.

[24] T. Larsson, M. Patriksson and A.-B. Strömberg, *On the convergence of conditional ϵ-subgradient methods for convex programs and convex-concave saddle-point problems*, European J. Oper. Res. **151** (2003), 461–473.

[25] S. Lee and A. Nedić, *Distributed random projection algorithm for convex optimization*, IEEE J. Sel. Top. Signa. **7** (2013), 221–229.

[26] A. Nedić and D. P. Bertsekas, *Incremental subgradient methods for nondifferentiable optimization*, SIAM J. Optim. **12** (2001), 109–138.

[27] A. Nedić and A. Ozdaglar, *Approximate primal solutions and rate analysis for dual subgradient methods*, SIAM J. Optim. **19** (2009), 1757–1780.

[28] A. Nedić and A. Ozdaglar, *Subgradient methods for saddle-point problems*, J. Optim. Theory App. **142** (2009), 205–228.

[29] Y. Nesterov, *Primal-dual subgradient methods for convex problems*, Math. Program. **120** (2009), 221–259.

[30] B. T. Polyak, *A general method for solving extremum problems*, Soviet Mathematics Doklady **8** (1967), 593–597.

[31] S. S. Ram, A. Nedić and V. V. Veeravalli, *Incremental stochastic subgradient algorithms for convex optimization*, SIAM J. Optim. **20** (2009), 691–717.

[32] S. Sen and H. D. Sherali, *A class of convergent primal-dual subgradient algorithms for decomposable convex programs*, Math. Program. **35** (1986), 279–297.

[33] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer-Verlag, New York, 1985.
[34] I. M. Stancu-Minasian, *Fractional Programming*, Kluwer Academic Publishers, Dordrecht, 1997.

(Y. H. Hu) College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China
E-mail address: mayhhu@szu.edu.cn

(X. Q. Yang) Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong
E-mail address: mayangxq@polyu.edu.hk

(Carisa K. W. Yu) Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong
E-mail address: carisayu@hsmc.edu.hk