ORIGINAL ARTICLE

Abstract convergence theorem for quasi-convex optimization problems with applications

Carisa Kwok Wai Yu[a], Yaohua Hu[b], Xiaoqi Yang[c] and Siu Kai Choy[a]

[a]Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong; [b]College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China; [c]Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ARTICLE HISTORY
Compiled August 13, 2018

ABSTRACT
Quasi-convex optimization is fundamental to the modeling of many practical problems in various fields such as economics, finance and industrial organization. Subgradient methods are practical iterative algorithms for solving large-scale quasi-convex optimization problems. In the present paper, focusing on quasi-convex optimization, we develop an abstract convergence theorem for a class of sequences, which satisfy a general basic inequality, under some suitable assumptions on parameters. The convergence properties in both function values and distances of iterates from the optimal solution set are discussed. The abstract convergence theorem covers relevant results of many types of subgradient methods studied in the literature, for either convex or quasi-convex optimization. Furthermore, we propose a new subgradient method, in which a perturbation of the successive direction is employed at each iteration. As an application of the abstract convergence theorem, we obtain the convergence results of the proposed subgradient method under the assumption of the Hölder condition of order p and by using the constant, diminishing, or dynamic stepsize rules, respectively. A preliminary numerical study shows that the proposed method outperforms the standard, stochastic and primal-dual subgradient methods in solving the Cobb-Douglas production efficiency problem.

KEYWORDS
Quasi-convex programming; subgradient method; basic inequality; abstract convergence theorem; Cobb-Douglas production efficiency problem

1. Introduction

Mathematical optimization provides a unified framework for a wide variety of important problems in many disciplines and application fields. Convex optimization plays a central role in mathematical optimization; however, it is too restrictive for many real-life problems encountered in economics, finance and management science. Quasi-convex optimization usually provides a much more accurate representation of realities than convex optimization does, while it still inherits some nice properties of convex optimization. This leads to a significant increase of studies in quasi-convex optimization; see [3,9,14,34] and references therein. However, the development of numerical algorithms for quasi-convex optimization, in particular for large-scale problems, is still

CONTACT Xiaoqi Yang. Email: mayangxq@polyu.edu.hk

in its infancy. Hence, there is a great demand for developing efficient numerical algorithms for solving large-scale quasi-convex optimization problems.

Subgradient methods form a class of popular and effective iterative algorithms used to solve constrained optimization problems. The subgradient method was originally introduced by Polyak [31] and Ermoliev [12] in the 1970s to solve a nondifferentiable convex optimization problem. This technique was further developed by Shor [33]. Over the past 40 years, various features of subgradient methods have been established for convex optimization problems, and many extensions/generalizations have been devised for this case; moreover, numerous applications have been proposed; see [1,4–6,10,15,22,29,35] and references therein. For example, the conditional subgradient method [25] has been introduced to avoid the zig-zagging phenomenon of the standard subgradient method and applied to solve the uncapacitated facility location problem and the multicommodity network flow problem; incremental subgradient methods [26,30] have been proposed to minimize the summation of a large number of convex functions and widely applied to solve the distributed optimization problem in large-scale sensor networks [32] and empirical risk minimization problem in online machine learning [11]; the primal-dual subgradient methods [28,29] have been investigated to approach a saddle point of a convex-concave function and used with great success in designing decentralized network control protocols [8]. Convergence properties of subgradient methods for convex optimization, in terms of function values and distances of iterates from the optimal solution set, have been well studied by using the constant, diminishing or dynamic stepsize rules.

In the last two decades, subgradient methods have been developed to solve constrained quasi-convex optimization problems; see [16–19,21,24] and references therein. For example, Kiwiel [21] studied the convergence properties of the standard subgradient method for solving quasi-convex optimization problems by using the diminishing stepsize rule. To handle more practical problems involving computational errors and noise, Hu et al. [16] proposed a generic inexact subgradient method and investigated the influence of the deterministic noise to the inexact subgradient method by adopting the constant and diminishing stepsize rules. To avert the zig-zagging phenomenon and improve the convergence behavior, Hu et al. [19] introduced a conditional subgradient method as well as an inexact version, and presented the convergence results by further using a dynamic stepsize rule. It is worth mentioning that the basic inequality of a subgradient iteration is a key tool for convergence analysis of subgradient methods for either convex or quasi-convex optimization.

The present paper is devoted to developing the convergence theory for a class of iterative methods to solve quasi-convex optimization problems in a unified framework. In particular, we consider a constrained quasi-convex optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X, \end{array} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a quasi-convex and continuous (not necessarily differentiable) function, and $X \subseteq \mathbb{R}^n$ is a nonempty, closed and convex set. We denote the set of minima and the minimum value of problem (1) by X^* and f_*, respectively. In the following, we aim to investigate an abstract convergence theorem for a sequence satisfying a general basic inequality. That is, we fix $p > 0$, and consider a sequence $\{x_k\} \subseteq X$, as well as a sequence of nonnegative scalars $\{v_k\}$, that satisfy the following two conditions:

(H1) For each $x^* \in X^*$ and each $k \in \{i \in \mathbb{N} : x_i \notin X^*\}$,

$$\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 \leq -\alpha_k v_k (f(x_k) - f_*)^{\frac{1}{p}} + \beta_k v_k^2. \tag{2}$$

(H2) $\{\alpha_k\}$ and $\{\beta_k\}$ are two sequences of positive scalars such that

$$\lim_{k\to\infty} \alpha_k = \alpha > 0 \quad \text{and} \quad \lim_{k\to\infty} \beta_k = \beta > 0. \tag{3}$$

Condition (H1) measures the difference of the distances of iterates from any solution by the difference of function value and the minimum value, and condition (H2) characterizes some assumptions on the parameters. The nature of subgradient methods forces the generated sequence to comply with conditions (H1) and (H2) under some mild assumptions, and thus, this study provides a unified framework for various subgradient methods for either convex or quasi-convex optimization. For example, they are satisfied for the sequence generated by subgradient methods for convex optimization problems (with $p = 1$) under a bounded subgradient assumption, and for quasi-convex optimization problems under the assumption of the Hölder condition of order p. More specifically, for convex optimization problems, assuming that f has an upper bound G on its subgradients (and so $p = 1$),

- the sequence generated by either the standard subgradient method [33] or the primal-dual subgradient method [28] satisfies (H1) and (H2) with $\alpha_k = 2$ and $\beta_k = G^2$;
- the sequence generated by the incremental subgradient method (for minimizing the summation of a number of convex functions f_i) [26] satisfies (H1) and (H2) with $\alpha_k = 2$ and $\beta_k = (\sum_{i=1}^m G_i)^2$, where G_i is the upper bound of subgradients of f_i; and
- in the unified framework of subgradient methods studied in [30], (H1) and (H2) are assumed to satisfy with α_k being a positive constant and $\lim_{k\to\infty} \beta_k v_k = 0$.

For quasi-convex optimization problems, assuming the Hölder condition of order p with modulus L,

- the sequence generated by either the standard subgradient method [21] or the primal-dual subgradient method [17] satisfies (H1) and (H2) with $\alpha_k = 2L^{-\frac{1}{p}}$ and $\beta_k = 1$;
- the sequence generated by the conditional subgradient method [19] satisfies (H1) and (H2) with $\alpha_k = 2L^{-\frac{1}{p}}$ and $\beta_k = 4$.

One of the main contributions of the present paper is to establish an abstract convergence theorem for any sequence satisfying conditions (H1) and (H2) under some suitable assumptions on $\{v_k\}$, in which the convergence properties in terms of function values and distances of iterates from the optimal solution set are discussed. The abstract convergence theorem covers relevant results of many types of subgradient methods studied in the literature, for either convex or quasi-convex optimization. Another contribution of the present paper is to propose a new subgradient method to solve the constrained quasi-convex optimization problem (1). Note that the standard subgradient method usually suffers from slow convergence rate in many applications. To speed up its convergence rate, conditional subgradient methods were introduced in [25] and [19] for solving constrained convex and quasi-convex optimization problems, respectively. In these methods, the search direction consists of a subgradient and a normal vector to the constraint set. However, it might be computational expensive to calculate the normal vector to the constraint set in general, especially for large-scale problems. This hinders the application of conditional subgradient methods to large-scale (quasi-convex) optimization problems. To tackle this obstacle, we propose an implementable subgradient method, in which a perturbation of the successive direction is employed in place of the normal vector, as in the conditional subgradient method. Under the assumption of the Hölder condition of order p, we show that the generated sequence satisfies conditions (H1) and (H2); consequently, as an application of the abstract convergence theorem, we obtain the convergence results of the proposed subgradient method by using the constant, diminishing, or dynamic stepsize rules.

Furthermore, we apply the proposed algorithm to solve the Cobb-Douglas production efficiency problem [7] and compare it with the standard subgradient method [21], stochastic subgradient method [18] and primal-dual subgradient method [17]. The numerical results verify the established convergence theorem and show that the proposed algorithm outperforms these subgradient-type methods in that it obtains a larger production efficiency and converges faster by choosing a suitable algorithmic parameter.

The present paper is organized as follows. In section 2, we present the notations and preliminary results that will be used in the present paper. In section 3, we establish an abstract convergence theorem for any sequence satisfying conditions (H1) and (H2) under some suitable assumptions on $\{v_k\}$. In section 4, we propose an implementable subgradient method to solve the constrained quasi-convex optimization problem (1), establish its convergence results by applying the abstract convergence theorem, and present its numerical simulation on solving the Cobb-Douglas production efficiency problem.

2. Notations and preliminary results

The notations used in the present paper are standard; see, e.g., [5]. We consider the n-dimensional Euclidean space $\mathbb{R}^n$ with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. As usual, $\mathbf{S}$ stands for the unit sphere centered at the origin. For $x \in \mathbb{R}^n$ and $Z \subseteq \mathbb{R}^n$, we use $\operatorname{dist}(x, Z)$ and $P_Z(x)$ to denote the Euclidean distance of x from Z and the Euclidean projection of x onto Z, respectively, that is,

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

Two well-known properties of the Euclidean projection are recalled in the following proposition. Part (i) of this result shows a nonexpansive property, while part (ii) is a characterization of the projection; see [5].

Proposition 2.1. *Let $Z \subseteq \mathbb{R}^n$ be a nonempty, closed and convex set and $x, y \in \mathbb{R}^n$. Then the following assertions hold:*

(i) $\|P_Z(x) - P_Z(y)\| \le \|x - y\|$.
(ii) $\langle P_Z(x) - x, z - P_Z(x)\rangle \ge 0$ *for any* $z \in Z$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be quasi-convex if for any $x, y \in \mathbb{R}^n$ and any $\alpha \in [0, 1]$, the following inequality holds

$$f((1-\alpha)x + \alpha y) \le \max\{f(x), f(y)\}.$$

For $\alpha \in \mathbb{R}$, we denote the sublevel sets of f by

$$\text{lev}_{<\alpha} f := \{x \in \mathbb{R}^n : f(x) < \alpha\} \quad \text{and} \quad \text{lev}_{\le\alpha} f := \{x \in \mathbb{R}^n : f(x) \le \alpha\}.$$

It is well-known that f is quasi-convex if and only if $\text{lev}_{<\alpha} f$ (and/or $\text{lev}_{\le\alpha} f$) is convex for any $\alpha \in \mathbb{R}$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be coercive if $\lim_{\|x\|\to\infty} f(x) = \infty$, and so its sublevel set $\text{lev}_{\le\alpha} f$ is bounded for any $\alpha \in \mathbb{R}$.

The subdifferential of a quasi-convex function plays an important role in the study of quasi-convex optimization. Several different types of subdifferentials of quasi-convex function have been introduced in the literature, see [2,13,16,21] and references therein. In particular, Kiwiel [21] and Hu et al. [16] introduced a quasi-subdifferential defined as a normal cone to the strict sublevel set of the quasi-convex function, and applied the related quasi-subgradient in their proposed subgradient methods; see, e.g., [16,19,21]. In the following definition, we recall the notion of quasi-subdifferential taken from [16].

Definition 2.2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex function. The quasi-subdifferential of f at $x \in \mathbb{R}^n$ is defined by

$$\partial^* f(x) := \left\{g \in \mathbb{R}^n : \langle g, y - x\rangle \le 0, \forall y \in \text{lev}_{<f(x)} f\right\}.$$

Each vector $g \in \partial^* f(x)$ is called a quasi-subgradient of f at x.

Definition 2.3. Let $p > 0$ and $L > 0$. The function $f : \mathbb{R}^n \to \mathbb{R}$ is said to satisfy the Hölder condition (restricted to the set of minima X^*) of order p with modulus L on $\mathbb{R}^n$ if

$$f(x) - f_* \le L\text{dist}^p(x, X^*) \quad \text{for any } x \in \mathbb{R}^n.$$

The notion of the Hölder condition (restricted to the set of minima X^*) was introduced in [23] to describe some properties of the quasi-subgradient, and it plays an important role in the convergence analysis of subgradient methods in quasi-convex optimization [16,18,19]. In fact, the Hölder condition (restricted to the set of minima X^*) is weaker than the classical Hölder condition, which means that

$$|f(x) - f(y)| \le L\|x - y\|^p \quad \text{for any } x, y \in \mathbb{R}^n. \tag{4}$$

In fact, (4) implies the Hölder condition (restricted to the set of minima X^*) of order p. The notion of Hölder condition has been widely studied in harmonic analysis [20] and fractional analysis [34], and extensively applied in economics [34] and management science [2]. It is worth noting that the classical Hölder condition of order 1 is equivalent to the bounded subgradient assumption, always assumed in the literature of subgradient methods (see, e.g., [5,22,26]), whenever f is convex.

Some examples of quasi-convex functions that satisfy the classical Hölder condition (and so satisfy the the Hölder condition in Definition 2.3) and coercive property are provided as follows. For the sake of simplicity, we use $\mathcal{Q}(p, L)$ to denote the set of functions that are quasi-convex and coercive and satisfy the classical Hölder condition of order p with modulus L.

Example 2.4. Let $p \in (0, 1)$ and $L > 0$.

(i) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(t) := |t|^p \quad \text{for each } t \in \mathbb{R}.$$

Then $f \in \mathcal{Q}(p, 1)$. In fact, the quasi-convexity and coercive property of f are trivial, and the Hölder condition of order p is satisfied because

$$|f(u) - f(v)| = |u|^p - |v|^p \le |u - v|^p \quad \text{for any } u, v \in \mathbb{R}.$$

(ii) Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) := \|x\|^p \quad \text{for each } x \in \mathbb{R}^n,$$

Similar to (i), we have $f \in \mathcal{Q}(p, 1)$.

(iii) Let $g_i : \mathbb{R} \to \mathbb{R}$ be such that $g_i \in \mathcal{Q}(p, L)$ for $i = 1, \ldots, n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) := \max_{i=1,\ldots,n} g_i(x_i) \quad \text{for each } x \in \mathbb{R}^n.$$

Then $f \in \mathcal{Q}(p, 1)$. In fact, we can check the Hölder condition of f by $g_i \in \mathcal{Q}(p, L)$ that

$$|f(x) - f(y)| \le \max_{i=1,\ldots,n} (g_i(x_i) - g_i(y_i)) \le \max_{i=1,\ldots,n} L|x_i - y_i|^p \le L\|x - y\|^p.$$

(iv) Let $g_i : \mathbb{R}^n \to \mathbb{R}$ be such that $g_i \in \mathcal{Q}(p, L)$ for $i = 1, \ldots, m$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) := \max_{i=1,\ldots,m} g_i(x) \quad \text{for each } x \in \mathbb{R}^n.$$

Similar to (iii), we have $f \in \mathcal{Q}(p, L)$.

The following lemma describes an important property of a quasi-convex function that satisfies the Hölder condition of order p. This property locally provides a connection between the quasi-subgradient and function values, which is a key to establish the basic inequality in convergence analysis of subgradient methods.

Lemma 2.5. *Let $p > 0$ and $L > 0$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be quasi-convex, continuous and satisfy the Hölder condition of order p with modulus L on $\mathbb{R}^n$. Let $x \in X \setminus X^*$, and let $g(x)$ be a unit quasi-subgradient of f at x, that is, $g(x) \in \partial^* f(x) \cap \mathbf{S}$. Then it holds, for any $x^* \in X^*$, that*

$$\langle g(x), x - x^* \rangle \ge \left(\frac{f(x) - f_*}{L} \right)^{\frac{1}{p}}. \tag{5}$$

Proof. Fix $x \in X \setminus X^*$. The level set $\text{lev}_{<f(x)} f$ is nonempty open and convex because f is quasi-convex and continuous on $\mathbb{R}^n$. Given $x^* \in X^*$, we define

$$r := \inf \left\{ \|y - x^*\| : y \in \text{bd} \left(\text{lev}_{<f(x)} f \right) \right\}, \tag{6}$$

where $\mathrm{bd}(Z)$ denotes the boundary of the set Z. It is clear that $r > 0$. By the assumption of Hölder condition of order p, we have

$$f(y) - f_* \leq L\mathrm{dist}^p(y, X^*) \quad \text{for any } y \in \mathbb{R}^n.$$

Taking the infimun over $\mathrm{bd}\left(\mathrm{lev}_{<f(x)}f\right)$, we can show that

$$f(x) - f_* \leq L \inf\left\{\mathrm{dist}^p(y, X^*) : y \in \mathrm{bd}\left(\mathrm{lev}_{<f(x)}f\right)\right\} \leq Lr^p. \tag{7}$$

Let $\delta \in (0, 1)$. Since $g(x) \in \partial^* f(x) \cap \mathbf{S}$, we obtain by (6) that $x^* + \delta r g(x) \in \mathrm{lev}_{<f(x)}f$. Hence, it follows from Definition 2.2 that

$$\langle g(x), x - x^*\rangle = \langle g(x), x - (x^* + \delta r g(x))\rangle + \delta r \geq \delta r.$$

Since $\delta \in (0, 1)$ is arbitrary, one has $\langle g(x), x - x^*\rangle \geq r$. This, together with (7), implies (5), as desired. The proof is complete. □

We end this section by recalling the following lemma from [22, Lemma 2.1], which is useful to establish an abstract convergence theorem.

Lemma 2.6. *Let $\{a_k\}$ be a scalar sequence, and let $\{w_k\}$ be a sequence of nonnegative scalars. Suppose that $\lim_{k\to\infty} \sum_{i=1}^k w_i = \infty$. Then it holds that*

$$\liminf_{k\to\infty} a_k \leq \liminf_{k\to\infty} \frac{\sum_{i=1}^k w_i a_i}{\sum_{i=1}^k w_i} \leq \limsup_{k\to\infty} \frac{\sum_{i=1}^k w_i a_i}{\sum_{i=1}^k w_i} \leq \limsup_{k\to\infty} a_k.$$

3. Abstract convergence theorem

This section aims to investigate an abstract convergence theorem, in terms of function values and distances of iterates from the optimal solution set, for the sequence that satisfies conditions (H1) and (H2) and under some suitable assumptions on $\{v_k\}$.

Theorem 3.1. *Consider a sequence $\{x_k\} \subseteq X$ that satisfies* (H1) *and* (H2)*. The following assertions are true.*

(i) *Suppose that $v_k = v > 0$ for any $k \in \mathbb{N}$. Then*

$$\liminf_{k\to\infty} f(x_k) \leq f_* + \left(\frac{\beta v}{\alpha}\right)^p. \tag{8}$$

(ii) *Suppose that $\{v_k\}$ satisfies*

$$v_k > 0, \quad \lim_{k\to\infty} v_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} v_k = \infty. \tag{9}$$

Then

$$\liminf_{k\to\infty} f(x_k) = f_*. \tag{10}$$

Moreover, suppose that f is coercive and that $\{x_k\}$ satisfies

$$\operatorname{dist}(x_{k+1}, X^*) \le \operatorname{dist}(x_k, X^*) + \gamma_k v_k \quad \textit{for any } k \in \mathbb{N} \tag{11}$$

with $\lim_{k\to\infty} \gamma_k = \gamma > 0$. Then

$$\lim_{k\to\infty} \operatorname{dist}(x_k, X^*) = 0 \quad \text{and} \quad \lim_{k\to\infty} f(x_k) = f_*. \tag{12}$$

(iii) *Suppose that $\{v_k\}$ is given by*

$$v_k = \frac{\alpha_k \lambda_k}{2\beta_k}(f(x_k) - f_*)^{\frac{1}{p}} \quad \textit{with } 0 < \underline{\lambda} \le \lambda_k \le \overline{\lambda} < 2. \tag{13}$$

Then, either $x_k \in X^*$ for some $k \in \mathbb{N}$ or $\{x_k\}$ converges to a point in X^*.

Proof. We first claim the following assertion:

(A) Let $N \in \mathbb{N}$. If (2) is satisfied for any $k \ge N$, then it holds for any $n > N$ that

$$\frac{\sum_{k=N}^{n-1} \alpha_k v_k (f(x_k) - f_*)^{\frac{1}{p}}}{\sum_{k=N}^{n-1} \alpha_k v_k} \le \frac{\|x_N - x^*\|^2}{\sum_{k=N}^{n-1} \alpha_k v_k} + \frac{\sum_{k=N}^{n-1} \beta_k v_k^2}{\sum_{k=N}^{n-1} \alpha_k v_k}. \tag{14}$$

In fact, summing (2) over $k = N, \ldots, n-1$, we have

$$\|x_n - x^*\|^2 - \|x_N - x^*\|^2 \le -\sum_{k=N}^{n-1} \alpha_k v_k (f(x_k) - f_*)^{\frac{1}{p}} + \sum_{k=N}^{n-1} \beta_k v_k^2,$$

and then obtain (14). Next, we prove this theorem by virtue of this assertion.

(i) Without loss of generality, we assume that $x_k \in X^*$ only occurs for finitely many times; otherwise, (8) holds automatically. That is, there exists $N \in \mathbb{N}$ such that $x_k \notin X^*$ for any $k \ge N$. Then (H1) and (A) show that (14) holds for any $n > N$. By (3) and the assumption that $v_k = v$ for any $k \in \mathbb{N}$, one has

$$\lim_{n\to\infty} \sum_{k=N}^{n-1} \alpha_k v_k = \infty, \tag{15}$$

and then Lemma 2.6 (with $(f(x_k) - f_*)^{\frac{1}{p}}$ and $\alpha_k v_k$ in place of a_k and w_k) is applicable to concluding that

$$\begin{aligned} \liminf_{k\to\infty} (f(x_k) - f_*)^{\frac{1}{p}} &\le \liminf_{n\to\infty} \frac{\sum_{k=N}^{n-1} \alpha_k v_k (f(x_k) - f_*)^{\frac{1}{p}}}{\sum_{k=N}^{n-1} \alpha_k v_k} \\ &\le \liminf_{n\to\infty} \left(\frac{\|x_N - x^*\|^2}{\sum_{k=N}^{n-1} \alpha_k v_k} + \frac{\sum_{k=N}^{n-1} \beta_k v_k^2}{\sum_{k=N}^{n-1} \alpha_k v_k} \right) \end{aligned} \tag{16}$$

(due to (14)). By (15), one has

$$\lim_{n\to\infty} \frac{\|x_N - x^*\|^2}{\sum_{k=N}^{n-1} \alpha_k v_k} = 0. \tag{17}$$

On the other hand, by the assumption that $v_k = v > 0$ for any $k \in \mathbb{N}$, we obtain

$$\lim_{n\to\infty} \frac{\sum_{k=N}^{n-1} \beta_k v_k^2}{\sum_{k=N}^{n-1} \alpha_k v_k} = \lim_{n\to\infty} \frac{\sum_{k=N}^{n-1} \beta_k v}{\sum_{k=N}^{n-1} \alpha_k} = \frac{\beta v}{\alpha},$$

where the second equality follows from Lemma 2.6 and (3) (with $\frac{\beta_k}{\alpha_k}$ and α_k in place of a_k and w_k). Combining this with (17), we can reduce (16) to

$$\liminf_{k\to\infty} (f(x_k) - f_*)^{\frac{1}{p}} \leq \frac{\beta v}{\alpha},$$

and hence (8) is obtained.

(ii) Without loss of generality, we assume that there exists $N \in \mathbb{N}$ such that $x_k \notin X^*$ for any $k \geq N$ (otherwise, (10) holds automatically), and consequently, (14) holds for any $n > N$. By (3) and (9) (in particular, $\sum_{k=N}^{\infty} v_k = \infty$), one observes that (15) is satisfied, and then (16) and (17) hold. We further obtain from Lemma 2.6 that

$$\lim_{n\to\infty} \frac{\sum_{k=N}^{n-1} \beta_k v_k^2}{\sum_{k=N}^{n-1} \alpha_k v_k} = \lim_{n\to\infty} \frac{\beta_n v_n}{\alpha_n} = 0$$

by (3) and (9). This, together with (16) and (17), deduces (10).

Below, we prove (12) based on (10) and (11). By (3), there exists $N \in \mathbb{N}$ such that

$$\beta_k < 2\beta \quad \text{and} \quad \alpha_k > \frac{\alpha}{2} \quad \text{for any } k \geq N. \tag{18}$$

Fix $\sigma > 0$. Noting that $\lim_{n\to\infty} v_n = 0$ and $\lim_{k\to\infty} \gamma_k = \gamma > 0$, we can assume that

$$v_k < \frac{\alpha}{4\beta} \sigma^{\frac{1}{p}} \quad \text{and} \quad \gamma_k \leq 2\gamma \quad \text{for any } k \geq N. \tag{19}$$

Define

$$X_\sigma := X \cap \mathrm{lev}_{\leq f_* + \sigma} f \quad \text{and} \quad \rho(\sigma) := \max_{x \in X_\sigma} \mathrm{dist}(x, X^*). \tag{20}$$

By the assumption that f is coercive, it follows that its sublevel set $\mathrm{lev}_{\leq f_*+\sigma} f$ is bounded, and so is X_σ. Therefore, by (20), one has $\rho(\sigma) < \infty$. Fix $k \geq N$. We show

$$\mathrm{dist}(x_{k+1}, X^*) < \max\{\mathrm{dist}(x_k, X^*), \rho(\sigma) + \gamma \frac{\alpha}{2\beta} \sigma^{\frac{1}{p}}\} \tag{21}$$

by claiming the following two implications:

$$[f(x_k) > f_* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) < \mathrm{dist}(x_k, X^*)]; \tag{22}$$

$$[f(x_k) \leq f_* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) < \rho(\sigma) + \gamma \frac{\alpha}{2\beta} \sigma^{\frac{1}{p}}]. \tag{23}$$

To prove (22), we suppose that $f(x_k) > f_* + \sigma$. Then $x_k \notin X^*$, and so (H1) gives, for any $x^* \in X^*$, that

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \alpha_k v_k \sigma^{\frac{1}{p}} + \beta_k v_k^2 < \|x_k - x^*\|^2$$

(due to (18) and (19)). Consequently, by letting $x^* := P_{X^*}(x_k)$, it follows that

$$\mathrm{dist}^2(x_{k+1}, X^*) \leq \|x_{k+1} - P_{X^*}(x_k)\|^2 < \|x_k - P_{X^*}(x_k)\|^2 = \mathrm{dist}^2(x_k, X^*),$$

i.e., (22) is proved. To show (23), we suppose that $f(x_k) \leq f_* + \sigma$. Then, by the fact that $x_k \in X$, we conclude that $x_k \in X_\sigma$, and so, (20) says that $\mathrm{dist}(x_k, X^*) \leq \rho(\sigma)$. This, together with (11) and (19), shows (23). Therefore, (21) is proved as desired.

By (10), we can assume, without loss of generality, that $f(x_N) \leq f_* + \sigma$ (otherwise, we can choose a larger N); consequently, one has by (23) that $\mathrm{dist}(x_{N+1}, X^*) < \rho(\sigma) + \gamma \frac{\alpha}{2\beta} \sigma^{\frac{1}{p}}$. Then, we inductively obtain by (21) that

$$\mathrm{dist}(x_k, X^*) < \rho(\sigma) + \gamma \frac{\alpha}{2\beta} \sigma^{\frac{1}{p}} \quad \text{for any } k > N. \tag{24}$$

Since f is continuous and coercive, its sublevel sets are compact, and so, it is trivial to see that $\lim_{\sigma \to 0} \rho(\sigma) = 0$. Hence, we obtain by (24) that $\lim_{k \to \infty} \mathrm{dist}(x_k, X^*) = 0$, and thus $\lim_{k \to \infty} f(x_k) = f_*$ (by the continuity of f).

(iii) Without loss of generality, we assume that $x_k \notin X^*$ for any $k \in \mathbb{N}$; otherwise, assertion (iii) of this theorem follows. Then, for any $x^* \in X^*$, it follows from (2) and (13) that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 &\leq -\alpha_k v_k (f(x_k) - f_*)^{\frac{1}{p}} + \beta_k v_k^2 \\ &= -\tfrac{\alpha_k^2}{4\beta_k} \lambda_k (2 - \lambda_k)(f(x_k) - f_*)^{\frac{2}{p}} \\ &\leq -\tfrac{\alpha_k^2}{4\beta_k} \underline{\lambda}(2 - \overline{\lambda})(f(x_k) - f_*)^{\frac{2}{p}}. \end{aligned} \tag{25}$$

Note that $f(x_k) \geq f_*$ (as $x_k \in X$) for each $k \in \mathbb{N}$. Then, we can claim that

$$\lim_{k \to \infty} f(x_k) = f_*. \tag{26}$$

Fix $\sigma > 0$. Proving by contradiction, we assume that there exists a subsequence $\{x_{k_i}\}$ such that $f(x_{k_i}) > f_* + \sigma$ for any $i \in \mathbb{N}$. Then, it follows from (25) and (18) that

$$\|x_{k_{i+1}} - x^*\|^2 \leq \|x_{k_i+1} - x^*\|^2 < \|x_{k_i} - x^*\|^2 - \frac{\alpha^2}{32\beta} \underline{\lambda}(2 - \overline{\lambda}) \sigma^{\frac{2}{p}} \quad \text{for any } i \geq N.$$

Write $\Delta := \frac{\alpha^2}{32\beta} \underline{\lambda}(2 - \overline{\lambda}) \sigma^{\frac{2}{p}}$. Then one has

$$\|x_{k_{i+1}} - x^*\|^2 < \|x_{k_N} - x^*\|^2 - (i - N + 1)\Delta < 0 \quad \text{for any } i \geq N + \frac{\|x_{k_N} - x^*\|^2}{\Delta},$$

which yields a contradiction; hence (26) is proved.

It also follows from (25) that $\{\|x_k - x^*\|\}$ is decreasing, and hence, $\{x_k\}$ is bounded. Let y be a cluster point of $\{x_k\}$. Then, it follows from the decreasing property and the continuity of f and (26) that

$$\lim_{k\to\infty} \|x_k - x^*\| = \|y - x^*\| \quad \text{and} \quad y \in X^*, \tag{27}$$

respectively. Hence, for two cluster points of $\{x_k\}$, namely $\bar{x}$ and $\tilde{x}$, we obtain from (27) that $\bar{x} \in X^*$, $\tilde{x} \in X^*$, and $\|\bar{x} - x^*\| = \|\tilde{x} - x^*\|$ for any $x^* \in X^*$; letting $x^* := \tilde{x}$, we conclude that $\bar{x} = \tilde{x}$. Therefore, $\{x_k\}$ converges to a point in X^*. The proof is complete. □

Remark 1. (i) In the literature of subgradient methods, the stepsize sequence $\{v_k\}$ satisfying condition (9) or (13) is called the diminishing stepsize (see [5,16,21,26]) or the dynamic stepsize (see [18,25,27,35]), respectively.

(ii) As mentioned in section 1, conditions (H1) and (H2) are satisfied for several variants of subgradient methods for either convex or quasi-convex optimization problems. Hence, Theorem 3.1 provides a unified framework of convergence analysis for subgradient methods, and it covers the convergence results of many types of subgradient methods in the literature. For example, for convex optimization problems, Theorem 3.1 is applicable to concluding [33, Theorems 2.2 and 2.4] for the standard subgradient method, [25, Theorems 2.6 and 2.9] for the conditional subgradient method, and [26, Propositions 2.1-2.3 and 2.5] for the incremental subgradient method; for quasi-convex optimization problems, Theorem 3.1 can be directly applied to establish [21, Theorem 1] for the standard subgradient method, [17, Theorems 3.2-3.3] for the primal-dual subgradient method, and [19, Theorems 3.3-3.5] for the conditional subgradient method.

4. A new subgradient method

The standard subgradient method usually suffers from a slow convergence rate in many applications. Conditional subgradient methods were proposed in [19,25] to speed up the convergence rate of the standard subgradient method. However, the conditional subgradient method requires a normal vector to the constraint set at each iteration, which might be computational expensive to calculate for large-scale optimization problems. To avert this difficulty, we propose an implementable subgradient method to solve problem (1), in which a perturbation of the successive direction is employed in place of the normal vector, as in the conditional subgradient method.

Algorithm 4.1. Select an initial point $x_0 \in \mathbb{R}^n$, a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$, and a sequence of parameters $\{s_k\} \subseteq (0, +\infty)$. Having x_k, we calculate a unit quasi-subgradient $g_k \in \partial^* f(x_k) \cap \mathbf{S}$, and update x_{k+1} by

$$y_k := P_X(x_k - v_k g_k), \tag{28}$$

$$x_{k+1} := P_X(x_k + s_k(y_k - x_k)). \tag{29}$$

In this section, we will investigate its convergence properties by applying the abstract convergence theorem. Moreover, we will conduct some numerical experiments to demonstrate its efficiency in solving the Cobb-Douglas production efficiency problem.

4.1. *Convergence analysis*

Under the assumption of the Hölder condition of order p, we establish a basic inequality for Algorithm 4.1 to show that the generated sequence satisfies conditions (H1) and (H2).

Lemma 4.1. *Suppose that f satisfies the Hölder condition of order p with modulus L on $\mathbb{R}^n$. Let $\{x_k\}$ be a sequence generated by Algorithm 4.1. Fix $k \in \mathbb{N}$ and $x^* \in X^*$. If $x_k \notin X^*$, then it holds that*

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 2s_k v_k \left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}} + \max\{s_k, s_k^2\} v_k^2. \tag{30}$$

Proof. Fix $k \in \mathbb{N}$ and $x^* \in X^*$. In view of Algorithm 4.1 (cf. (28)), it follows from Proposition 2.1(i) that

$$\|y_k - x^*\|^2 \le \|x_k - v_k g_k - x^*\|^2 = \|x_k - x^*\|^2 - 2v_k \langle g_k, x_k - x^*\rangle + v_k^2. \tag{31}$$

Under the assumption that $x_k \notin X^*$, Lemma 2.5 is applicable (with x_k and g_k in place of x and $g(x)$) to concluding that

$$\langle g_k, x_k - x^*\rangle \ge \left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}}. \tag{32}$$

This, together with (31), implies that

$$\|y_k - x^*\|^2 \le \|x_k - x^*\|^2 - 2v_k \left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}} + v_k^2. \tag{33}$$

Below, we prove (30) in the following two cases.
Case 1: Suppose that $s_k \in (0, 1]$. By (29), it follows again from Proposition 2.1(i) that

$$\|x_{k+1} - x^*\|^2 \le \|x_k + s_k(y_k - x_k) - x^*\|^2 = \|s_k(y_k - x^*) + (1 - s_k)(x_k - x^*)\|^2. \tag{34}$$

By the convexity of $\|\cdot - x^*\|^2$, we have

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\le s_k\|y_k - x^*\|^2 + (1 - s_k)\|x_k - x^*\|^2 \\ &\le \|x_k - x^*\|^2 - 2s_k v_k \left(\tfrac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}} + s_k v_k^2 \end{aligned}$$

(due to (33)). Hence, (30) is obtained in this case.
Case 2: Suppose that $s_k \in (1, +\infty)$. It follows from the first inequality of (34) that

$$\begin{aligned} &\|x_{k+1} - x^*\|^2 \\ &\le \|x_k + s_k(y_k - x_k) - x^*\|^2 \\ &= \|(s_k - 1)(y_k - x_k) + y_k - x^*\|^2 \\ &= \|y_k - x^*\|^2 + (s_k - 1)^2\|y_k - x_k\|^2 + 2(s_k - 1)\langle y_k - x^*, y_k - x_k\rangle. \end{aligned} \tag{35}$$

Note that

$$\langle y_k - x^*, y_k - x_k\rangle = \langle y_k - x^*, y_k - x_k + v_k g_k\rangle - v_k\langle y_k - x_k, g_k\rangle - v_k\langle x_k - x^*, g_k\rangle. \quad (36)$$

By Proposition 2.1(ii) and (28) (with y_k, $x_k - v_k g_k$, x^* in place of $P_Z(x)$, x, z), we have

$$\langle y_k - x^*, y_k - x_k + v_k g_k\rangle = \langle P_X(x_k - v_k g_k) - x^*, P_X(x_k - v_k g_k) - (x_k - v_k g_k)\rangle \leq 0. \quad (37)$$

By Proposition 2.1(i), one has

$$\|y_k - x_k\| \leq \|v_k g_k\| = v_k \quad (38)$$

(noting that $\|g_k\| = 1$), and accordingly,

$$\langle y_k - x_k, g_k\rangle \geq -\|y_k - x_k\| \geq -v_k.$$

This, together with (36), (37) and (32), implies that

$$\langle y_k - x^*, y_k - x_k\rangle \leq v_k^2 - v_k \left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}}.$$

Combining this with (38), we can reduce (35) to

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &\leq \|y_k - x^*\|^2 + (s_k - 1)^2 v_k^2 + 2(s_k - 1)v_k^2 - 2(s_k - 1)v_k \left(\tfrac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}} \\ &= \|x_k - x^*\|^2 - 2s_k v_k \left(\tfrac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}} + s_k^2 v_k^2\end{aligned}$$

(by (33)). Thus, (30) is proved. The proof is complete. □

Applying the abstract convergence theorem (i.e., Theorem 3.1), we obtain the convergence results for Algorithm 4.1 when using the constant, diminishing or dynamic stepsize rules.

Theorem 4.2. *Suppose that f satisfies the Hölder condition of order p with modulus L on $\mathbb{R}^n$. Let $\{x_k\}$ be a sequence generated by Algorithm 4.1 with $\lim_{k\to\infty} s_k = s > 0$. The following assertions are true.*

(i) *If $v_k = v > 0$ for any $k \in \mathbb{N}$, then*

$$\liminf_{k\to\infty} f(x_k) \leq f_* + L\left(\frac{v}{2}\max\{1, s\}\right)^p. \quad (39)$$

(ii) *If $\{v_k\}$ satisfies (9), then $\liminf_{k\to\infty} f(x_k) = f_*$. Moreover, suppose that f is coercive. Then $\lim_{k\to\infty} \mathrm{dist}(x_k, X^*) = 0$ and $\lim_{k\to\infty} f(x_k) = f_*$.*

(iii) *If $\{v_k\}$ is given by*

$$v_k = \frac{\lambda_k}{\max\{1, s_k\}}\left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}} \quad \text{with } 0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < 2, \quad (40)$$

then $\{x_k\}$ converges to an optimal solution of problem (1).

Proof. Define three sequences of positive scalars by

$$\gamma_k := s_k, \quad \alpha_k := 2s_k L^{-\frac{1}{p}} \quad \text{and} \quad \beta_k := \max\{s_k, s_k^2\}. \tag{41}$$

By the assumption that $\lim_{k\to\infty} s_k = s > 0$, one observes that

$$\lim_{k\to\infty} \gamma_k = s > 0, \quad \lim_{k\to\infty} \alpha_k = 2sL^{-\frac{1}{p}} > 0 \quad \text{and} \quad \lim_{k\to\infty} \beta_k = \max\{s, s^2\} > 0,$$

which verifies (H2).

By Lemma 4.1 and (41), we obtain that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 &\le -2s_k v_k L^{-\frac{1}{p}} (f(x_k) - f_*)^{\frac{1}{p}} + \max\{s_k, s_k^2\} v_k^2 \\ &= -\alpha_k v_k (f(x_k) - f_*)^{\frac{1}{p}} + \beta_k v_k^2 \end{aligned}$$

for each $x^* \in X^*$ and $k \in \mathbb{N}$ with $x_k \notin X^*$. This shows that (H1) is satisfied.

Noting that $g_k \in \mathbf{S}$, it follows from (28) and Proposition 2.1(i) that

$$\|y_k - x_k\| \le \|x_k - v_k g_k - x_k\| = v_k.$$

Then, by (29), it follows from Proposition 2.1(i), for each $x^* \in \mathbb{N}$, that

$$\|x_{k+1} - x^*\| \le \|x_k + s_k(y_k - x_k) - x^*\| \le \|x_k - x^*\| + s_k\|y_k - x_k\| \le \|x_k - x^*\| + s_k v_k.$$

Since $x^* \in X^*$ is arbitrary, we obtain that

$$\operatorname{dist}(x_{k+1}, x^*) \le \operatorname{dist}(x_{k+1}, x^*) + s_k v_k = \operatorname{dist}(x_{k+1}, x^*) + \gamma_k v_k$$

(due to (41)). This verifies (11) with $\lim_{k\to\infty} \gamma_k = s > 0$. Furthermore, by (41), one checks that $\{v_k\}$ given by (40) satisfies (13). Therefore, the assumptions of Theorem 3.1 are satisfied, and the assertions of this theorem follow directly (in particular for (iii), one checks by (40) that $v_k = 0$ whenever $x_k \in X^*$, and so the generated sequence stays at this optimal solution). The proof is complete. □

4.2. *Numerical experiments*

This subsection presents an application of Algorithm 4.1 to the Cobb-Douglas production efficiency problem, which was introduced by Bradley and Frey [7] and formulated as

$$\begin{array}{ll} \max & f(x) := \frac{\text{Profit}}{\text{Cost}} = \frac{a_0 \prod_{j=1}^n x_j^{a_j}}{\sum_{j=1}^n c_j x_j + c_0} \\ \text{s.t.} & \sum_{j=1}^n b_{ij} x_j \ge p_i, \quad i = 1, \dots, m, \\ & x \ge 0, \end{array} \tag{42}$$

where $x := (x_1, \dots, x_n)^\top$ is the variable designating the vector of production factors; see [16] for details. It was shown in [16] that problem (42) is a quasi-concave maximization problem, and that the subgradient method is effective for problem (42), even for large-scale problems.

We compare Algorithm 4.1 with several existing subgradient-type methods, including the standard subgradient method (in short, SG) [21], the stochastic subgradient method (in short, StoSG) [18] and the primal-dual subgradient method (in short, PDS-G) [17], in solving the Cobb-Douglas production efficiency problem (42). All numerical experiments are implemented in MATLAB R2009a and executed on a personal laptop (Intel Core i7, 2.00 GHz, 8.00 GB of RAM). In the numerical experiments, the parameters of problem (42) are randomly chosen from different intervals:

$$a_0 \in [0, n], \quad a_j, b_{ij}, c_0, c_j \in [0, 1], \quad \text{and} \quad p_i \in [0, n/2].$$

The diminishing stepsize is chosen as

$$v_k = v/(1 + 0.1k),$$

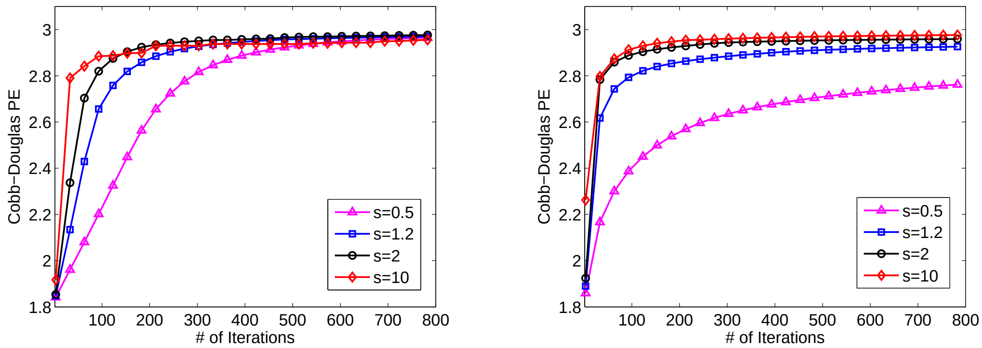
where v is always chosen between $[2, 5]$, while the constant stepsize is selected between $[0.5, 2]$. The larger the problem size, the larger the stepsize.

We first compare the performance (in terms of the obtained optimal values) of Algorithm 4.1 and the SG, StoSG, PDSG for problem (42) of different scales. Parameters s_k in Algorithm 4.1 are set to be a constant: $s = 0.5$, 1.2, 2 or 10, and the diminishing stepsize is adopted in all these subgradient-type methods. The maximal values of problem (42) are obtained and displayed in Table 1. In this table, the columns of Projects and Factors represent the numbers of projects (m) and factors (n) of problem (42). It is observed from Table 1 that Algorithm 4.1 achieves a larger production efficiency than the SG, StoSG and PDSG when $s > 1$, while a smaller value is obtained when $s < 1$. We can also observe from the results of Algorithm 4.1 that the larger the parameter s, the larger the obtained production efficiency, when using the diminishing stepsize.

Table 1. Computation results for maximizing Cobb-Douglas production efficiency.

Circumstance of problem		SG	StoSG	PDSG	Algorithm 4.1			
Projects	Factors	f_{opt}	f_{opt}	f_{opt}	$s = 0.5$	$s = 1.2$	$s = 2$	$s = 10$
50	50	2.5278	2.4429	2.4989	2.2786	2.5785	2.6084	2.6096
100	100	2.9054	2.8679	2.8887	2.8904	2.9265	2.9592	2.9766
200	200	2.8790	2.8648	2.8672	2.6756	2.8848	2.8922	2.8992
500	500	2.7535	2.6644	2.7385	2.4279	2.7712	2.7955	2.8354
1000	1000	2.5842	2.4556	2.5674	2.1383	2.6458	2.7738	2.8268
2000	2000	2.6967	2.6451	2.6801	2.4292	2.7190	2.7685	2.8235

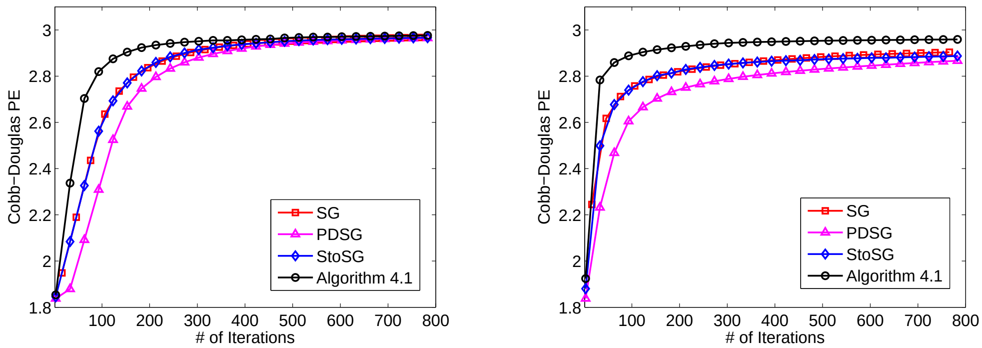
The second experiment is to show the convergence behavior of Algorithm 4.1 by choosing different parameters s and using the constant and diminishing stepsize rules, respectively, where the problem scale is fixed to be $(m, n) = (100, 100)$. Figure 1 plots the estimated Cobb-Douglas production efficiency along the number of the iterations in a random trial. Figure 1(a) illustrates that Algorithm 4.1 when $s = 0.5$, 1.2 or 2 achieve a larger production efficiency than Algorithm 4.1 with $s = 10$, when using a constant stepsize. This observation is consistent with Theorem 4.2(i) that a smaller tolerance from the optimal value is achieved for a smaller s (see (39)). It is also observed from Figure 1(a) that the larger the parameter s, the faster convergence rate Algorithm 4.1 does. Figure 1(b) demonstrates that when $s = 1.2$, 2 or 10, Algorithm 4.1 converges faster to the optimal value than that with $s = 0.5$, when using a diminishing stepsize. These observations may be because a large (constant) perturbation may cause the violation of convergence to the optimal solution, while a small one could lead to slow convergence. With regard to the trade-off, we recommend a suitable selection of parameter s_k in Algorithm 4.1, e.g., a constant in [1.5,3].



(a) The constant stepsize rule. (b) The diminishing stepsize rule.

Figure 1. Convergence behavior of Algorithm 4.1 for different parameters s.

Under the same experimental setting with the second one, we further compare the convergence behavior of Algorithm 4.1 (with $s = 2$) with SG, StoSG and PDSG. Figure 2 shows that Algorithm 4.1 outperforms the SG, StoSG and PDSG in that it converges faster to the optimal value and obtains a larger production efficiency.



(a) The constant stepsize rule. (b) The diminishing stepsize rule.

Figure 2. Convergence behavior of Algorithm 4.1 and existing subgradient-type methods.

5. Conclusion

In the present paper, we considered the quasi-convex optimization problems and established an abstract convergence theorem for a class of sequences, which satisfy conditions (H1) and (H2), under some suitable assumptions on $\{v_k\}$. The abstract convergence theorem provided a unified framework for various subgradient methods for either convex or quasi-convex optimization. Inspired by the ideas of conditional subgradient methods [19,25], we proposed an implementable subgradient method and established its convergence results by virtue of the abstract convergence theorem. The numerical results showed that the proposed method outperforms the standard, stochastic and primal-dual subgradient methods in solving the Cobb-Douglas production efficiency problem.

The standard subgradient method usually suffers from a zig-zagging phenomena and sustains slow convergence rate in many applications. It is an interesting open question how to avoid the zig-zagging phenomena of subgradient methods. Although the numerical studies showed that the conditional subgradient methods and our proposed method can speed up the convergence rate, there is still no theoretical study to guarantee their advantage in avoiding the zig-zagging phenomena. In the future work, we will contribute to investigate some subgradient methods avoiding the zig-zagging phenomena by virtue of the special structures of optimization problems.

Acknowledgements

The authors are grateful to the editor and two anonymous reviewers for their valuable suggestions and remarks which helped to improve the quality of the paper.

Funding

Carisa Kwok Wai Yu was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China [grant number UGC/FDS14/P03/14, UGC/FDS14/P02/15]; Yaohua Hu was supported in part by the National Natural Science Foundation of China [grant number 11601343, 11601344]; Natural Science Foundation of Guangdong [grant number 2016A030310038]; Foundation for Distinguished Young Talents in Higher Education of Guangdong [grant number 2015KQNCX145]; Xiaoqi Yang was supported in part by the Research Grants Council of Hong Kong [grant number PolyU 152167/15E].

References

[1] A. Auslender and M. Teboulle. Projected subgradient methods with non-Euclidean distances for non-differentiable convex minimization and variational inequalities. *Mathematical Programming*, 120:27–48, 2009.

[2] D. Aussel and A. Daniilidis. Normal characterization of the main classes of quasiconvex functions. *Set-Valued Analysis*, 8:219–236, 2000.

[3] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang. *Generalized Concavity.* Plenum Press, New York, 1988.

[4] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

[5] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Cambridge, 1999.

[6] J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples.* Springer, New York, 2006.

[7] S. P. Bradley and S. C. Frey. Fractional programming with homogeneous functions. *Operations Research*, 22(2):350–357, 1974.

[8] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, 2007.

[9] J.-P. Crouzeix, J.-E. Martinez-Legaz, and M. Volle. *Generalized Convexity, Generalized Monotonicity.* Kluwer Academic Publishers, Dordrecht, 1998.

[10] G. D'Antonio and A. Frangioni. Convergence analysis of deflected conditional approximate subgradient methods. *SIAM Journal on Optimization*, 20(1):357–386, 2009.

[11] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

[12] Y. M. Ermoliev. Methods of solution of nonlinear extremal problems. *Cybernetics and Systems Analysis*, 2:1–14, 1966.

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

[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 Journal on Optimization*, 26(2):1207–1235, 2016.

[16] Y. H. Hu, X. Q. Yang, and C.-K. Sim. Inexact subgradient methods for quasi-convex optimization problems. *European Journal of Operational Research*, 240(2):315–327, 2015.

[17] Y. H. Hu, X. Q. Yang, and C. K. W. Yu. Subgradient methods for saddle point problems of quasiconvex optimization. *Pure and Applied Functional Analysis*, 2(1):83–97, 2017.

[18] Y. H. Hu, C. K. W. Yu, and C. Li. Stochastic subgradient method for quasi-convex optimization problems. *Journal of Nonlinear and Convex Analysis*, 17(4):711–724, 2016.

[19] Y. H. Hu, C. K. W. Yu, C. Li, and X. Q. Yang. Conditional subgradient methods for constrained quasi-convex optimization problems. *Journal of Nonlinear and Convex Analysis*, 17(10):2143–2158, 2016.

[20] Y. Katznelson. *An introduction to harmonic analysis*. Cambridge University Press, Cambridge, 2004.

[21] K. C. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming*, 90:1–25, 2001.

[22] K. C. Kiwiel. Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM Journal on Optimization*, 14(3):807–840, 2004.

[23] I. V. Konnov. On properties of supporting and quasi-supporting vectors. *Journal of Mathematical Sciences*, 71:2760–2763, 1994.

[24] I. V. Konnov. On convergence properties of a subgradient method. *Optimization Methods and Software*, 18(1):53–62, 2003.

[25] T. Larsson, M. Patriksson, and A.-B. Strömberg. Conditional subgradient optimization–theory and applications. *European Journal of Operational Research*, 88:382–403, 1996.

[26] A. Nedić and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization*, 12(1):109–138, 2001.

[27] A. Nedić and D. P. Bertsekas. The effect of deterministic noise in subgradient methods. *Mathematical Programming*, 125:75–99, 2010.

[28] A. Nedić and A. Ozdaglar. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142:205–228, 2009.

[29] Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120:221–259, 2009.

[30] E. S. H. Neto and A. R. D. Pierro. Incremental subgradients for constrained convex optimization: A unified framework and new methods. *SIAM Journal on Optimization*, 20(3):1547–1572, 2009.

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

[32] M. G. Rabbat and R. D. Nowak. Quantized incremental algorithms for distributed optimization. *IEEE Journal on Selected Areas in Communications*, 23(4):798–808, 2005.

[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.

[35] J. H. Wang, Y. H. Hu, C. Li, and J.-C. Yao. Linear convergence of CQ algorithms and applications in gene regulatory network inference. *Inverse Problems*, 33(5):055017, 2017.