This is the Pre-Published Version.

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use (https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms), but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s10898-019-00818-6

The following publication Hu, Y., Yu, C. K. W., & Yang, X. (2019). Incremental quasi-subgradient methods for minimizing the sum of quasi-convex functions. Journal of Global Optimization, 75(4), 1003-1028 is available at https://doi.org/10.1007/s10898-019-00818-6

Incremental Subgradient Methods for Minimizing The Sum of Quasi-convex Functions

Yaohua Hu* Carisa Kwok Wai Yu† Xiaoqi Yang‡

Abstract The sum of ratios problem has a variety of important applications in economics and management science, but it is difficult to globally solve this problem. In this paper, we consider the minimization problem of a sum of a number of nondifferentiable quasi-convex component functions over a closed and convex set, which includes the sum of ratios problem as a special case. The sum of quasi-convex component functions is not necessarily to be quasi-convex, and so, this study goes beyond quasi-convex optimization. Exploiting the structure of the sum-minimization problem, we propose a new incremental subgradient method for this problem and investigate its convergence properties to a global optimal solution when using the constant, diminishing or dynamic stepsize rules and under a homogeneous assumption and the Hölder condition of order p. To economize on the computation cost of subgradients of a large number of component functions, we further propose a randomized incremental subgradient method, in which only one component function is randomly selected to construct the subgradient direction at each iteration. The convergence properties are obtained in terms of function values and distances of iterates from the optimal solution set with probability 1. The proposed incremental subgradient methods are applied to solve the sum of ratios problem, as well as the multiple Cobb-Douglas productions efficiency problem, and the numerical results show that the proposed methods are efficient for solving the large sum of ratios problem.

Keywords Quasi-convex programming, sum-minimization problem, subgradient method, incremental approach, convergence analysis

*College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China (mayhhu@szu.edu.cn). This author's work was supported in part by the National Natural Science Foundation of China (11601343) and Natural Science Foundation of Guangdong (2016A030310038).

†Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong (carisayu@hsmc.edu.hk). This author's work was supported in part by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS14/P03/14 and UGC/FDS14/P02/15).

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (mayangxq@polyu.edu.hk). This author's work was supported in part by the Research Grants Council of Hong Kong (PolyU 152342/16E).

1 Introduction

In recent years, a great amount of attention has been attracted to the research of minimizing a sum of a number of nondifferentiable component functions:

$$\begin{array}{ll} \min & f(x) := \sum_{i=1}^{m} f_i(x) \\ \text{s.t.} & x \in X, \end{array} \tag{1.1}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are real-valued functions, and $X \subseteq \mathbb{R}^n$ is a closed set.

The type of convex sum-minimization problems, i.e., problem (1.1) with each f_i being convex and X being convex, has been widely studied in various applications, such as the Lagrangian dual of the coupling constraints of large-scale separable optimization problem [9, 33], the distributed optimization problem in large-scale sensor networks [10, 39] and the empirical risk minimization problem in online machine learning [15, 32]. Motivated by vast applications of problem (1.1), the development of optimization algorithms has become an important issue of the sum-minimization problem, and many practical numerical algorithms have been proposed to solve problem (1.1); see [7, 9, 21, 32, 46] and references therein. In particular, the class of subgradient methods are popular and effective iterative algorithms for solving the large-scale convex sum-minimization problem (1.1), due to the simple formulation and low storage requirement. The subgradient method was originally introduced to solve a nondifferentiable convex optimization problem by Polyak [38] and Ermoliev [16], and until now, various variants of subgradient methods have been studied to solve structured optimization problems; see [2, 8, 27, 31, 33, 35, 43] and references therein. To meet the structure of sum-minimization problem (1.1), an idea of incremental approach has been proposed to perform the subgradient process incrementally, by sequentially taking steps along the subgradients of component functions, with intermediate adjustment of the variables after processing each component function. That is, an iteration of the incremental subgradient method can be viewed as a cycle of m subiterations, starting from $z_{k,0} := x_k$, through m steps

$$z_{k,i} := P_X\big(z_{k,i-1} - v_k g_{k,i}\big), \quad g_{i,k} \in \partial f_i(z_{k,i-1}), \quad i = 1, \ldots, m, \tag{1.2}$$

and finally arriving at $x_{k+1} := z_{k,m}$. The incremental subgradient method has gained successful applications in large-scale sensor networks and online machine learning; see, e.g., [15, 39]. So far, many articles have been devoted to the study of convergence analysis and applications of different types of incremental subgradient methods; see [7, 25, 27, 33, 36, 40, 42] and references therein. In particular, Nedić and Bertsekas [33] investigated the convergence properties of the incremental subgradient methods, including the deterministic and stochastic ones, for the constant/diminishing/dynamic stepsize rules; later the authors extended these convergence results to the inexact incremental subgradient method with the deterministic noise in [34]. Shi et al. [42] proposed a normalized incremental subgradient method, analyzed its convergence theory and demonstrated its application in wireless sensor networks. Neto and Pierro [36] explored the incremental subgradient method in a generic framework consisting of an optimality step and a feasibility step, where both approximate subgradients and approximate projections are allowed, and illustrated its application in tomographic imaging.

Most papers in the literature of subgradient methods focus on convex optimization problems. Recently, much attention have been received beyond convex optimization. Quasi-convex optimization problems can be found in many important applications in various areas, such as economics, engineering, management science and various applied sciences; see [4, 14, 20] and references therein. The subgradient methods have been well extended to solve quasi-convex optimization problems, such as the standard subgradient method [26], inexact subgradient method [22] and stochastic subgradient method [23]. The convergence results of subgradient methods, in terms of objective values and iterates, for solving quasi-convex optimization problems have been well established under the Hölder condition of order p and using the constant/diminishing/dynamic stepsize rules. However, to the best of our knowledge, there is still no study devoted to investigating subgradient methods for solving the sum-minimization problem (1.1) for the case when each f_i is quasi-convex.

The sum of ratios problem (in short, SOR) [41] has a variety of important applications in economics and management science, such as multi-stage stochastic shipping [1], government contracting [13] and bond portfolio optimization [28]. Although some numerical methods, such as the branch and bound scheme [6] and the interior point method [17], have been studied to solve the SOR, most of them are computationally expensive to implement for large-scale problems or incapable to globally solve the SOR. The absence of effective numerical algorithms for large-scale problems hinders the research and applications of the SOR. Exploiting its structure, the SOR can be formulated as a sum-maximization problem of a number of quasi-concave component functions (see section 4 for the explanation), and so it is an important application of problem (1.1). However, there is still no effective numerical algorithms for the large-scale sum-minimization problem (1.1) of quasi-convex functions, as well as the SOR.

To fill this gap, the aim of this paper is to develop the incremental subgradient methods for minimizing the sum of a number of quasi-convex component functions over a constraint set. In the remainder of this paper, we consider the sum-minimization problem (1.1) under the following hypothesis:

- $f_i : \mathbb{R}^n \to \mathbb{R}$ is quasi-convex and continuous for each $i = 1, \ldots, m$, and $X \subseteq \mathbb{R}^n$ is nonempty, closed and convex.

Note that f defined in (1.1), the sum of quasi-convex component functions, is not necessarily to be quasi-convex. The study of this paper is indeed beyond quasi-convex optimization, and so the direct application of standard subgradient method [26] to solve problem (1.1) is not necessarily convergent.

Inspired by the idea of incremental approach, we propose a new incremental subgradient method to solve the sum-minimization problem (1.1), which is different from the classical incremental subgradient method (1.2) in that the subgradient subiterations are only updated on the component functions whose minimal values are not achieved yet. Under a homogeneous assumption and the Hölder condition of order p for the component functions, we provide a proper basic inequality and establish the convergence properties of the proposed incremental subgradient method when using the constant/diminishing/dynamic stepsize rules. The

convergence properties are characterized in terms of function values and distances of iterates from the optimal solution set, and the finite convergence behavior to the optimality is further investigated when the solution set has a nonempty interior.

In the incremental subgradient method, the estimation of subgradients of all component functions at each iteration may be very expensive, especially when the number of component functions is large and no simple formulae for computing the subgradients exist. Note that the stochastic gradient descent algorithm is increasingly popular in large-scale machine learning problems; see [8, 15, 45, 46] and references therein. Employing the idea of stochastic gradient descent algorithm, we propose a randomized incremental subgradient method to save the computational cost of the incremental subgradient iteration, in which only one component function is randomly selected to construct the subgradient direction at each iteration. The convergence results show that the randomized incremental subgradient method enjoys the convergence properties with probability 1 and achieves a better tolerance than that of the deterministic incremental subgradient method.

Furthermore, we formulate the SOR as a sum-minimization problem (1.1), consider the multiple Cobb-Douglas productions efficiency problem (in short, MCDPE) [11] as an application of the SOR, and conduct numerical experiments on this problem via applying the proposed incremental subgradient methods. The numerical results show that the incremental subgradient methods are efficient for the MCDPE, especially for large-scale problems. This study may deliver a new approach for finding the global optimal solution of the large-scale SOR.

The paper is organized as follows. In section 2, we present the notations and preliminary results used in this paper. In section 3, we propose the deterministic and randomized incremental subgradient methods to solve the sum-minimization problem (1.1) and investigate their convergence properties when using the typical stepsize rules. The application to the SOR and the numerical study for the MCDPE are presented in section 4.

2 Notations and preliminary results

The notations used in this paper are standard; see, e.g., [7, 8]. We consider the n-dimensional Euclidean space $\mathbb{R}^n$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. For $x \in \mathbb{R}^n$ and $\delta \in \mathbb{R}_+$, we use $B(x, \delta)$ and $S(x, \delta)$ to denote the closed ball and the sphere of radius δ centered at x, respectively. For $x \in \mathbb{R}^n$ and $Z \subseteq \mathbb{R}^n$, we write $\operatorname{dist}(x, Z)$ and $P_Z(x)$ to denote the Euclidean distance of x from Z and the Euclidean projection of x onto Z, respectively, that is,

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

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

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

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

$$\operatorname{lev}_{<\alpha} h := \{x \in \mathbb{R}^n : h(x) < \alpha\} \quad \text{and} \quad \operatorname{lev}_{\leq\alpha} h := \{x \in \mathbb{R}^n : h(x) \leq \alpha\}$$

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

The subdifferential of quasi-convex function plays an important role in quasi-convex optimization. Several different types of subdifferentials of quasi-convex function have been introduced; see [3, 18, 22, 26] and references therein. In particular, Kiwiel [26] and Hu et al. [22] introduced a notion of quasi-subdifferential defined a normal cone to $\mathrm{lev}_{<h(x)}h$, and used the related quasi-subgradient in their proposed subgradient methods. In the following definition, we recall the notion of quasi-subdifferential taken from [22].

Definition 2.1. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex function, and let $x \in \mathbb{R}^n$. The quasi-subdifferential of h at x is defined by*

$$\partial h(x) = \left\{ g : \langle g, y - x\rangle \leq 0 \text{ for any } y \in \mathrm{lev}_{<h(x)}h \right\}. \tag{2.1}$$

The Hölder condition of order p was used in [29] to describe some properties of quasi-subgradient, and it plays an important role in the convergence study of subgradient-type methods for quasi-convex optimization problems [22, 23].

Definition 2.2. *Let $p > 0$, $L > 0$ and $x \in \mathbb{R}^n$. $h : \mathbb{R}^n \to \mathbb{R}$ is said to satisfy the Hölder condition of order p with modulus L at x if*

$$|h(y) - h(x)| \leq L\|y - x\|^p \quad \text{for any } y \in \mathbb{R}^n. \tag{2.2}$$

h is said to satisfy the Hölder condition of order p with modulus L on X if (2.2) holds for any $x \in X$.

The Hölder condition with order 1 is reduced to the Lipschitz condition, and this property holds for very broad classes of functions with various values of $p \in (0, +\infty)$. The following lemma recalls an important property of a quasi-convex function that satisfies the Hölder condition of order p. This property is a key to establish a basic inequality in convergence analysis of subgradient-type methods.

Lemma 2.1 ([30, Proposition 2.1])**.** *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex and continuous function, X be a closed and convex set, and let X^* be the set of minima of h over X. Let $p > 0$ and $L > 0$, and suppose that h satisfies the Hölder condition of order p with modulus L at some $x^* \in X^*$. Then, for any $x \in X \setminus X^*$, it holds that*

$$h(x) - h(x^*) \leq L \langle g, x - x^*\rangle^p \quad \text{for any } g \in \partial h(x) \cap S(0, 1).$$

We end this section by recalling the following two lemmas, which are useful in the convergence analysis of incremental subgradient methods.

Lemma 2.2 ([24, Lemma 4.1])**.** *Let $a_i \geq 0$ for $i = 1, 2, \ldots, n$. The following relations hold.*

(i) *If* $\gamma \in (0,1]$, *then*

$$\Big(\sum_{i=1}^{n} a_i\Big)^{\gamma} \le \sum_{i=1}^{n} a_i^{\gamma} \le n\Big(\sum_{i=1}^{n} a_i\Big)^{\gamma}.$$

(ii) *If* $\gamma \in [1,\infty)$, *then*

$$\frac{1}{n^{\gamma-1}}\Big(\sum_{i=1}^{n} a_i\Big)^{\gamma} \le \sum_{i=1}^{n} a_i^{\gamma} \le \Big(\sum_{i=1}^{n} a_i\Big)^{\gamma}.$$

Lemma 2.3 ([27, Lemma 2.1])**.** *Let* $\{a_k\}$ *be a sequence of scalars, and let* $\{v_k\}$ *be a sequence of nonnegative scalars. Suppose that* $\lim_{k\to\infty}\sum_{i=1}^{k} v_i = \infty$. *Then it holds that*

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

In particular, if $\lim_{k\to\infty} a_k = a$, *then* $\lim_{k\to\infty} \frac{\sum_{i=1}^{k} v_i a_i}{\sum_{i=1}^{k} v_i} = a$.

3 Incremental subgradient methods and convergence analysis

In this section, we propose the incremental subgradient methods, including the deterministic and randomized styles, to solve problem (1.1), and investigate their convergence properties when using typical stepsize rules. We write f^* and X^* to denote the optimal value and the (global) optimal solution set of problem (1.1) respectively, that is,

$$f^* := \min_{x\in X} \sum_{i=1}^{m} f_i(x) \quad \text{and} \quad X^* := \arg\min_{x\in X} \sum_{i=1}^{m} f_i(x),$$

and define

$$f_i^* := \min_{x\in X} f_i(x) \quad \text{and} \quad X_i^* := \arg\min_{x\in X} f_i(x) \quad \text{for } i = 1,\ldots,m.$$

To accomplish the convergence analysis, the following two assumptions are made throughout this paper.

Assumption 1. *The component functions of problem* (1.1) *have a common optimal solution.*

Assumption 2. *Let* $p > 0$ *and* $L_i > 0$ *for* $i = 1,\ldots,m$. *For each* $i = 1,\ldots,m$, f_i *satisfies the Hölder condition of order* p *with modulus* L_i *on* X.

The Hölder condition of order p was assumed in [22, 23] to develop the convergence theory of several subgradient-type methods for quasi-convex optimization. Assumption 2 consists of the Hölder condition of order p for all component functions of problem (1.1). Furthermore, we write

$$L_{\max} := \max_{i=1,\ldots,m} L_i. \tag{3.1}$$

Remark 3.1. It is easy to check that Assumption 1 is equivalent to $X^* = \cap_{i=1}^m X_i^* \neq \emptyset$. Assumption 1 is a homogeneous assumption for the component functions of problem (1.1), and it also says that $f^* = \sum_{i=1}^m f_i^*$. Shi et al. [42] used Assumption 1 to explore the convergence properties of a normalized incremental subgradient method for minimizing a sum of convex component functions. Although, under Assumption 1, we can approach the optimal value of problem (1.1) via minimizing component functions f_i over X separately, it is still difficult to find a common optimal solution, i.e., an optimal solution of problem (1.1), which is an essential issue of decision-making problems. In this paper, we propose the incremental subgradient methods to resolve this issue.

Below, we show a class of application problems, i.e., the quasi-convex feasibility problem, that can be formulated as the framework (1.1) and satisfy Assumptions 1 and 2.

Example 3.1. *The feasibility problem is at the core of the modeling of many problems in various areas of mathematics and physical sciences; see [5, 21] and references therein. In particular, the quasi-convex feasibility problem is an important class of feasibility problems (see [12, 19, 37]), which is to find a solution of the following system of inequalities:*

$$x \in X, \quad \text{and} \quad h_i(x) \leq 0 \quad \text{for each } i = 1, \ldots, m, \tag{3.2}$$

where $h_i : \mathbb{R}^n \to \mathbb{R}$ is quasi-convex and continuous for each $i = 1, \ldots, m$, and X is nonempty, closed and convex. It is always assumed that the solution set of problem (3.2) *is nonempty. The feasibility problem* (3.2) *can be cast into the framework of sum-minimization problem* (1.1) *as the following model:*

$$\min_{x \in X} f(x) := \sum_{i=1}^m f_i(x), \quad \text{where } f_i := \max\{h_i, 0\} \text{ for each } i = 1, \ldots, m. \tag{3.3}$$

It is clear that f_i is quasi-convex if h_i is quasi-convex for each $i = 1, \ldots, m$. It is easy to see that Assumption 1 is satisfied for problem (3.3) *if the solution set of problem* (3.2) *is nonempty. Assumption 2 is satisfied for problem* (3.3) *if each function h_i in* (3.2) *satisfies the Hölder condition of order p.*

The stepsize rule has a critical effect on the convergence behavior and computational capacity of subgradient methods. In this paper, we investigate convergence properties of incremental subgradient methods by using the following typical stepsize rules.

(S1) *Constant stepsize rule*:

$$v_k \equiv v \, (> 0) \quad \text{for any } k \in \mathbb{N}.$$

(S2) *Diminishing stepsize rule*:

$$v_k > 0, \quad \lim_{k \to \infty} v_k = 0, \quad \sum_{k=0}^{\infty} v_k = \infty. \tag{3.4}$$

(S3) *Dynamic stepsize rule I*:

$$v_k = \gamma_k \frac{C_{p,m}}{m^2} \left(f(x_k) - f^*\right)^{\frac{1}{p}} \quad \text{for any } k \in \mathbb{N}, \tag{3.5}$$

where $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 2$ and

$$C_{p,m} := L_{\max}^{-\frac{1}{p}} \min\left\{1, (2m)^{1-\frac{1}{p}}\right\}. \tag{3.6}$$

(S4) *Dynamic stepsize rule II*:

$$v_k = \gamma_k \frac{R_{p,m}}{m} \left(f(x_k) - f^*\right)^{\frac{1}{p}} \quad \text{for any } k \in \mathbb{N}, \tag{3.7}$$

where $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 2$ and

$$R_{p,m} := L_{\max}^{-\frac{1}{p}} \min\left\{1, m^{1-\frac{1}{p}}\right\}. \tag{3.8}$$

Type (S3) is for the deterministic incremental subgradient method, while type (S4) is for the randomized incremental subgradient method. Note that both types of dynamic stepsize rules are slightly difference from that of the classical incremental subgradient method (see [33]).

3.1 Incremental subgradient method

The aim of this subsection is to propose an incremental subgradient method to solve problem (1.1), and to study its convergence properties when using several different stepsize rules. The incremental subgradient method is formally described as follows.

Algorithm 1: Incremental subgradient method.

```
Initialize an initial point x_0 ∈ R^n, a stepsize sequence {v_k}, and let k := 0;
while f(x_k) > f* do
    Let z_{k,0} := x_k;
    for i = 1, ..., m do
        if f_i(z_{k,i-1}) = f_i* then
            Let z_{k,i} := z_{k,i-1};
        else
            Calculate g_{k,i} ∈ ∂f_i(z_{k,i-1}) ∩ S(0,1), and let z_{k,i} := P_X(z_{k,i-1} - v_k g_{k,i});
        end
    end
    Let x_{k+1} := z_{k,m} and k := k + 1.
end
```

Remark 3.2. Note that Algorithm 1 is different from the classical incremental subgradient method for convex optimization in [33] and the incremental gradient method for smooth optimization in [7]. In particular, the classical incremental gradient/subgradient method (1.2) updates subgradient subiterations in a cyclic sequence on $\{1, \ldots, m\}$; while Algorithm 1 only updates subgradient subiterations in a cyclic sequence on the index set $\{i : f_i(z_{k,i-1}) > f_i^*\}$, where the minimal value of f_i is not achieved yet.

The following example illustrates that Algorithm 1 may not converge to the optimal value of problem (1.1) if the updated sequence on $\{i : f_i(z_{k,i-1}) > f_i^*\}$ in Algorithm 1 is replaced by a cyclic sequence on $\{1, \ldots, m\}$ as in the classical incremental gradient/subgradient method (1.2), even though Assumptions 1 and 2 are satisfied.

Example 3.2. *Consider problem* (1.1)*, where* $X = \mathbb{R}$*,* $m = 2$*, and two component functions are*

$$f_1(x) := \max\{x, 0\} \quad \text{and} \quad f_2(x) := \max\{-x, 0\}.$$

Obviously, $f_1^* = f_2^* = 0$*,* $X_1^* = \mathbb{R}_-$ *and* $X_2^* = \mathbb{R}_+$*;* $f(x) := f_1(x) + f_2(x) = |x|$*,* $f^* = 0$ *and* $X^* = \{0\}$*. It is easy to see that* $X^* = X_1^* \cap X_2^*$*, and that* f_1 *and* f_2 *are quasi-convex and satisfy the Hölder condition of order 1 on* $\mathbb{R}$ *with* $L_{\max} = 1$*. Hence Assumptions 1 and 2 are satisfied.*

In this setting, for any $x_0 > 0$*, one has that* $\partial f_1(x_0) = \mathbb{R}_+$*,* $g_{1,1} = 1$ *and* $z_{1,1} = x_0 - v$*. Note that*

$$\partial f_2(x) := \begin{cases} \mathbb{R}, & x \geq 0, \\ \mathbb{R}_-, & x < 0. \end{cases}$$

Hence we can choose $g_{1,2} = -1$*, and then* $z_{1,2} = z_{1,1} + v = x_0$*. That is, a fixed sequence is generated, and so* $x_k \equiv x_0$ *and* $\lim_{k\to\infty} f(x_k) = x_0$*. Therefore, the generated sequence does not converge to the optimal value/solution of problem* (1.1)*.*

We now start the convergence analysis by providing a basic inequality of Algorithm 1, which shows a significant property of an incremental subgradient iteration. Recall that $C_{p,m}$ is defined in (3.6).

Lemma 3.1. *Suppose Assumptions 1 and 2 are satisfied. Let* $\{x_k\}$ *be a sequence generated by Algorithm 1. Then, for any* $x^* \in X^*$ *and any* $k \in \mathbb{N}$*, it holds that*

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2v_k C_{p,m}(f(x_k) - f^*)^{\frac{1}{p}} + m^2 v_k^2. \tag{3.9}$$

Proof. We first show that the following inequality holds for any $x^* \in X^*$, any $k \in \mathbb{N}$ and $i = 1, \ldots, m$.

$$\|z_{k,i} - x^*\|^2 \leq \|z_{k,i-1} - x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}} (f_i(z_{k,i-1}) - f_i^*)^{\frac{1}{p}} + v_k^2. \tag{3.10}$$

In view of Algorithm 1, if $f_i(z_{k,i-1}) = f_i^*$, then it is updated that $z_{k,i} = z_{k,i-1}$, and so (3.10) automatically holds; otherwise, $z_{k,i-1} \notin X_i^*$, and then one sees from Algorithm 1 that

$$z_{k,i} = P_X\left(z_{k,i-1} - v_k g_{k,i}\right). \tag{3.11}$$

Note by Assumption 1 that $x^* \in X^* = \cap_{i=1}^m X_i^*$, and so $x^* \in X_i^*$ for $i = 1, \ldots, m$. By the assumption that $z_{k,i-1} \notin X_i^*$, one has that $f(z_{k,i-1}) > f_i^*$. Then Lemma 2.1 is applicable (with f_i, $z_{k,i-1}$, X_i^* in place of h, x, X^*) to concluding that

$$f_i(z_{k,i-1}) - f_i^* = f_i(z_{k,i-1}) - f_i(x^*) \leq L_i \langle g_{k,i}, z_{k,i-1} - x^* \rangle^p \leq L_{\max} \langle g_{k,i}, z_{k,i-1} - x^* \rangle^p \quad (3.12)$$

(due to (3.1)). By the nonexpansive property of the projection operator, it follows from (3.11) that

$$\begin{aligned} \|z_{k,i} - x^*\|^2 &\leq \|z_{k,i-1} - v_k g_{k,i} - x^*\|^2 \\ &= \|z_{k,i-1} - x^*\|^2 - 2v_k \langle g_{k,i}, z_{k,i-1} - x^* \rangle + v_k^2 \\ &\leq \|z_{k,i-1} - x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}} \left(f_i(z_{k,i-1}) - f_i^*\right)^{\frac{1}{p}} + v_k^2, \end{aligned}$$

where the last inequality follows from (3.12). Hence (3.10) is proved.

Next, we estimate the second term in the right hand side of (3.10) in terms of $f_i(x_k) - f_i^*$. By Lemma 2.2 (with $f_i(z_{k,i-1}) - f_i^*$, $|f_i(z_{k,i-1}) - f_i(x_k)|$, $\frac{1}{p}$ in place of a_1, a_2, γ), one has that

$$\begin{aligned} (f_i(x_k) - f_i^*)^{\frac{1}{p}} &\leq ((f_i(z_{k,i-1}) - f_i^*) + |f_i(z_{k,i-1}) - f_i(x_k)|)^{\frac{1}{p}} \\ &\leq \max\left\{1, 2^{\frac{1}{p}-1}\right\} \left((f_i(z_{k,i-1}) - f_i^*)^{\frac{1}{p}} + |f_i(z_{k,i-1}) - f_i(x_k)|^{\frac{1}{p}}\right). \end{aligned}$$

Denoting

$$C_p := \left(\max\left\{1, 2^{\frac{1}{p}-1}\right\}\right)^{-1} = \min\left\{1, 2^{1-\frac{1}{p}}\right\}, \quad (3.13)$$

the above inequality is reduced to

$$(f_i(z_{k,i-1}) - f_i^*)^{\frac{1}{p}} \geq C_p (f_i(x_k) - f_i^*)^{\frac{1}{p}} - |f_i(z_{k,i-1}) - f_i(x_k)|^{\frac{1}{p}}. \quad (3.14)$$

By Assumption 2 (cf. (2.2)) and in view of Algorithm 1, it follows that

$$|f_i(z_{k,i-1}) - f_i(x_k)| \leq L_i \|z_{k,i-1} - x_k\|^p \leq L_{\max} \left(\sum_{j=1}^{i-1} \|z_{k,j} - z_{k,j-1}\|\right)^p \leq L_{\max} \left(v_k(i-1)\right)^p.$$

Hence (3.14) is reduced to

$$(f_i(z_{k,i-1}) - f_i^*)^{\frac{1}{p}} \geq C_p (f_i(x_k) - f_i^*)^{\frac{1}{p}} - L_{\max}^{\frac{1}{p}} v_k (i-1),$$

and so (3.10) yields that

$$\|z_{k,i} - x^*\|^2 \leq \|z_{k,i-1} - x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}} C_p \left(f_i(x_k) - f_i^*\right)^{\frac{1}{p}} + (2i-1)v_k^2. \quad (3.15)$$

Finally, summing (3.15) over $i = 1, \ldots, m$, we obtain that

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}} C_p \sum_{i=1}^m \left(f_i(x_k) - f_i^*\right)^{\frac{1}{p}} + m^2 v_k^2. \quad (3.16)$$

Note by Lemma 2.2 (with $f_i(x_k) - f_i^*$ and $\frac{1}{p}$ in place of a_i and γ) that

$$\sum_{i=1}^{m}(f_i(x_k) - f_i^*)^{\frac{1}{p}} \geq \min\left\{1, m^{1-\frac{1}{p}}\right\}\left(\sum_{i=1}^{m}(f_i(x_k) - f_i^*)\right)^{\frac{1}{p}} = \min\left\{1, m^{1-\frac{1}{p}}\right\}(f(x_k) - f^*)^{\frac{1}{p}}$$

(thanks to Assumption 1), and by (3.13) that

$$L_{\max}^{-\frac{1}{p}} C_p \min\left\{1, m^{1-\frac{1}{p}}\right\} = L_{\max}^{-\frac{1}{p}} \min\left\{1, 2^{1-\frac{1}{p}}\right\} \min\left\{1, m^{1-\frac{1}{p}}\right\} = C_{p,m}$$

(cf. (3.6)). Therefore, (3.9) is seen to hold by (3.16). The proof is complete. □

Remark 3.3. *In Algorithm 1, the subgradient subiteration is processed in an ordered cyclic sequence on the index set* $\{i : f_i(z_{k,i-1}) > f_i^*\}$. *It is worthy mentioning that the proof of Lemma 3.1, as well as the convergence analysis of Algorithm 1, still work if any order of component functions is assumed, as long as each component on* $\{i : f_i(z_{k,i-1}) > f_i^*\}$ *is taken into account exactly once within a cycle. Hence, in applications, we could reorder the components* f_i *by either shifting or reshuffling at the beginning of each cycle, and then proceed with the calculations until the end of this cycle.*

By virtue of Lemma 3.1, we establish the convergence results of the incremental subgradient method when using different stepsize rules in Theorems 3.1-3.3, respectively.

Theorem 3.1. *Suppose Assumptions 1 and 2 are satisfied. Let* $\{x_k\}$ *be a sequence generated by Algorithm 1 with the constant stepsize rule* (S1). *Then we have*

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \left(\frac{m^2 v}{2C_{p,m}}\right)^p. \tag{3.17}$$

Proof. We prove by contradiction, assuming that

$$\liminf_{k\to\infty} f(x_k) > f^* + \left(\frac{m^2 v}{2C_{p,m}}\right)^p.$$

Consequently, there exist $\delta > 0$ and $N \in \mathbb{N}$ such that

$$f(x_k) > f^* + \left(\frac{m^2 v}{2C_{p,m}} + \delta\right)^p \quad \text{for any } k \geq N.$$

Therefore, it follows from Lemma 3.1 that for any $k \geq N$

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2vC_{p,m}(f(x_k) - f^*)^{\frac{1}{p}} + m^2 v^2 < \|x_k - x^*\|^2 - 2v\delta C_{p,m}.$$

Summing the above inequality over $k = N, \ldots, t-1$, we have

$$\|x_t - x^*\|^2 < \|x_N - x^*\|^2 - 2(t-N)v\delta C_{p,m},$$

which yields a contradiction for a sufficiently large t. The proof is complete. □

Remark 3.4. Theorem 3.1 shows the convergence of Algorithm 1 to the optimal value of problem (1.1) within a tolerance when the constant stepsize rule is adopted. The tolerance in (3.17) is given in terms of the stepsize and circumstances of problem (1.1), including the number of component functions and parameters of Hölder conditions. In particular, when $m = 1$, problem (1.1) is reduced to a constrained quasi-convex optimization problem, and then the convergence result described in Theorem 3.1 is reduced to that of [22, Theorem 3.1] (when noise and error are vanished). When each component function in problem (1.1) is convex, the Hölder condition ($p = 1$) is equivalent to the bounded subgradient assumption, and then the convergence result described in Theorem 3.1 is reduced to that of [33, Proposition 2.1].

Theorem 3.2. *Suppose Assumptions 1 and 2 are satisfied. Let $\{x_k\}$ be a sequence generated by Algorithm 1 with the diminishing stepsize rule* (S2)*. Then the following statements hold.*

(i) $\liminf_{k\to\infty} f(x_k) = f^*$.

(ii) *If f is coercive, then $\lim_{k\to\infty} f(x_k) = f^*$ and $\lim_{k\to\infty} \operatorname{dist}(x_k, X^*) = 0$.*

(iii) *If $\sum_{k=0}^{\infty} v_k^2 < \infty$, then $\{x_k\}$ converges to an optimal solution of problem* (1.1)*.*

Proof. (i) Fix $x^* \in X^*$. Summing (3.9) over $k = 0, 1, \ldots, n-1$, we have

$$\|x_n - x^*\|^2 \le \|x_0 - x^*\|^2 - 2C_{p,m} \sum_{k=0}^{n-1} v_k (f(x_k) - f^*)^{\frac{1}{p}} + m^2 \sum_{k=0}^{n-1} v_k^2, \tag{3.18}$$

and thus

$$\frac{\sum_{k=0}^{n-1} v_k (f(x_k) - f^*)^{\frac{1}{p}}}{\sum_{k=0}^{n-1} v_k} \le \frac{\|x_0 - x^*\|^2}{2C_{p,m} \sum_{k=0}^{n-1} v_k} + \frac{m^2 \sum_{k=0}^{n-1} v_k^2}{2C_{p,m} \sum_{k=0}^{n-1} v_k}. \tag{3.19}$$

Note by (3.4) that

$$\lim_{n\to\infty} \frac{\|x_0 - x^*\|^2}{\sum_{k=0}^{n-1} v_k} = 0, \tag{3.20}$$

and by Lemma 2.3 (with v_k in place of a_k) that

$$\lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k} = \lim_{n\to\infty} v_n = 0. \tag{3.21}$$

Then, by Lemma 2.3 (with $(f(x_k) - f^*)^{\frac{1}{p}}$ in place of a_k), (3.19) implies that

$$\liminf_{n\to\infty} (f(x_k) - f^*)^{\frac{1}{p}} \le \liminf_{n\to\infty} \frac{\sum_{k=0}^{n-1} v_k (f(x_k) - f^*)^{\frac{1}{p}}}{\sum_{k=0}^{n-1} v_k} \le 0.$$

This shows the desired assertion.

(ii) Fix $\sigma > 0$. Since $\{v_k\}$ is diminishing, there exists $N \in \mathbb{N}$ be such that

$$v_k \le \frac{1}{m^2} C_{p,m} \sigma^{\frac{1}{p}} \quad \text{for any } k \ge N. \tag{3.22}$$

Define

$$X_\sigma := X \cap \mathrm{lev}_{\le f^*+\sigma} f \quad \text{and} \quad \rho(\sigma) := \max_{x \in X_\sigma} \mathrm{dist}(x, X^*). \tag{3.23}$$

By the assumption that f is coercive, it follows that its sublevel set $\mathrm{lev}_{\le f_*+\sigma} f$ is bounded, and so is X_σ. Thus, by (3.23), one has $\rho(\sigma) < \infty$. Fix $k \ge N$. We show

$$\mathrm{dist}(x_{k+1}, X^*) \le \max\{\mathrm{dist}(x_k, X^*), \rho(\sigma) + \frac{1}{m} C_{p,m} \sigma^{\frac{1}{p}}\} \tag{3.24}$$

by claiming the following two implications:

$$[f(x_k) > f_* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) \le \mathrm{dist}(x_k, X^*)]; \tag{3.25}$$

$$[f(x_k) \le f_* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) \le \rho(\sigma) + \frac{1}{m} C_{p,m} \sigma^{\frac{1}{p}}]. \tag{3.26}$$

To prove (3.25), we suppose that $f(x_k) > f_* + \sigma$. Then $x_k \notin X^*$, and so Lemma 3.1 is applicable to concluding that, for any $x^* \in X^*$,

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 2C_{p,m} v_k \sigma^{\frac{1}{p}} + m^2 v_k^2 \le \|x_k - x^*\|^2 - m^2 v_k^2$$

(due to (3.22)). Consequently, one can prove (3.25) by selecting $x^* := P_{X^*}(x_{k+1})$. To show (3.26), we suppose that $f(x_k) \le f_* + \sigma$. Then we conclude that $x_k \in X_\sigma$, and so, (3.23) says that $\mathrm{dist}(x_k, X^*) \le \rho(\sigma)$. In view of Algorithm 1, for any $x^* \in X^*$, we obtain

$$\|x_{k+1} - x^*\| \le \|x_k - x^*\| + \sum_{i=1}^{m} \|z_{k,i} - z_{k,i-1}\| \le \|x_k - x^*\| + v_k m,$$

and thus

$$\mathrm{dist}(x_{k+1}, X^*) \le \mathrm{dist}(x_k, X^*) + v_k m \le \rho(\sigma) + v_k m.$$

This, together with (3.22), shows (3.26). Therefore, (3.24) is proved as desired.

By assertion (i), we can assume, without loss of generality, that $f(x_N) \le f_* + \sigma$ (otherwise, we can choose a larger N); consequently, one has by (3.26) that $\mathrm{dist}(x_{N+1}, X^*) \le \rho(\sigma) + \frac{1}{m} C_{p,m} \sigma^{\frac{1}{p}}$. Then, we inductively obtain by (3.24) that

$$\mathrm{dist}(x_k, X^*) \le \rho(\sigma) + \frac{1}{m} C_{p,m} \sigma^{\frac{1}{p}} \quad \text{for any } k > N. \tag{3.27}$$

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 (3.27) 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) By the assumption that $\sum_{k=0}^{\infty} v_k^2 < \infty$, one sees from (3.18) that $\{\|x_k - x^*\|\}$ is bounded, and so is $\{x_k\}$. Since further it was proved in assertion (i) of this theorem that $\liminf_{k \to \infty} f(x_k) = f^*$, it follows that $\{x_k\}$ has at least a cluster point falling in X^*, assumed as $\bar{x} \in X^*$. Noting that $\lim_{n \to \infty} \sum_{k=n}^{\infty} v_k^2 = 0$, we obtain by (3.9) (with $\bar{x}$ in place of x^*) that $\{\|x_k - \bar{x}\|^2\}$ is a Cauchy sequence, and thus it converges to 0. Hence $\{x_k\}$ converges to $\bar{x}\,(\in X^*)$. The proof is complete. □

It was reported in [22, 23] that the Hölder condition of order p (i.e., Assumption 2) is essential for the convergence behavior of subgradient-type methods for quasi-convex optimization. The following example illustrates that Assumption 1 is also essential for the validity of the established convergence theorems.

Example 3.3. *Consider problem* (1.1)*, where* $X = \mathbb{R}$*,* $m = 2$*, and two component functions are*

$$f_1(x) := \max\{x + 2, 0\} \quad \text{and} \quad f_2(x) := \max\{-2x + 2, 0\}.$$

Obviously, $f_1^* = f_2^* = 0$*,* $X_1^* = (-\infty, -2]$ *and* $X_2^* = [1, +\infty)$*;* $f^* = 3$ *and* $X^* = \{1\}$*. Clearly, one sees that* $X^* \neq X_1^* \cap X_2^*$*, and so Assumption 1 is not satisfied. It is easy to verify that* f_1 *and* f_2 *are quasi-convex and satisfy the Hölder condition of order 1 on* $\mathbb{R}$ *with* $L_{\max} = 2$*, and so Assumption 2 is satisfied.*

Starting from $x_0 = 0$*, we apply Algorithm 1 to solve problem* (1.1)*. We claim that the generated sequence may not converge to the optimal value/solution of problem* (1.1) *for any stepsize. Indeed, in this setting, one has that* $\partial f_1(0) = \mathbb{R}_+$*,* $g_{1,1} = 1$ *and* $z_{1,1} = -v < 0$*, and then* $\partial f_2(z_{k,1}) = \mathbb{R}_-$*,* $g_{1,2} = -1$ *and* $z_{1,2} = z_{1,1} + v = 0$*. Consequently, a fixed sequence is generated, and so* $x_k \equiv 0$ *and* $\lim_{k\to\infty} f(x_k) = 4$*. Hence, Theorem 3.1 fails whenever* $v < \frac{1}{4}$*, and Theorem 3.2 fails for any diminishing stepsize.*

When the prior information on f^* is available, a dynamic stepsize rule is usually considered to achieve an optimal convergence property in the literature of subgradient methods; see, e.g., [7, 12, 23, 33, 34]. Below, we show the convergence of the incremental subgradient method to an optimal solution of problem (1.1) when the dynamic stepsize rule (S3) is adopted.

Theorem 3.3. *Suppose Assumptions 1 and 2 are satisfied. Let* $\{x_k\}$ *be a sequence generated by Algorithm 1 with the dynamic stepsize rule* (S3)*. Then* $\{x_k\}$ *converges to an optimal solution of problem* (1.1)*.*

Proof. By Lemma 3.1 and (3.5), we obtain that, for any $x^* \in X^*$ and any $k \in \mathbb{N}$,

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2v_k C_{p,m}(f(x_k) - f^*)^{\frac{1}{p}} + m^2 v_k^2 \\
&= \|x_k - x^*\|^2 - \gamma_k(2 - \gamma_k)\tfrac{C_{p,m}^2}{m^2}(f(x_k) - f^*)^{\frac{2}{p}} \\
&\leq \|x_k - x^*\|^2 - \underline{\gamma}(2 - \overline{\gamma})\tfrac{C_{p,m}^2}{m^2}(f(x_k) - f^*)^{\frac{2}{p}}.
\end{aligned}$$

This shows that the sequence $\{\|x_k - x^*\|\}$ is decreasing, and hence $\{x_k\}$ is bounded. It also follows from the above inequality that

$$\sum_{k=1}^{\infty}(f(x_k) - f^*)^{\frac{2}{p}} \leq \frac{m^2}{\underline{\gamma}(2 - \overline{\gamma})C_{p,m}^2}\|x_0 - x^*\|^2,$$

which is finite. Consequently, noting that $f(x_k) - f^* \geq 0$ for any $k \in \mathbb{N}$, one has that $\lim_{k\to\infty} f(x_k) = f^*$. Hence, any cluster point of $\{x_k\}$ is an optimal solution of problem (1.1), denoted by $\bar{x} \in X^*$. Since further the sequence $\{\|x_k - x^*\|\}$ is decreasing, it converges to $\|\bar{x} - x^*\|$ for any $x^* \in X^*$. Hence $\{x_k\}$ converges to an optimal solution of problem (1.1). The proof is complete. □

At the end of this subsection, we present a finite convergence property of the incremental subgradient method to the solution set X^* of problem (1.1) under the assumption that X^* has a nonempty interior.

Theorem 3.4. *Suppose Assumptions 1 and 2 are satisfied. Let $\{x_k\}$ be a sequence generated by Algorithm 1. Let $x^* \in X^*$ and $\sigma > 0$, and suppose that $\mathbf{B}(x^*, \sigma) \subseteq X^*$. Then $x_k \in X^*$ for some $k \in \mathbb{N}$, provided that one of the following conditions:*

(i) *$v_k = v \in (0, \frac{2\sigma}{m})$ for any $k \in \mathbb{N}$, and*

(ii) *the sequence $\{v_k\}$ satisfies the diminishing stepsize rule* (S2).

Proof. To proceed, we define a new process $\{\hat{x}_k\}$ via the classical incremental subgradient method starting with $\hat{x}_0 := x_0$. That is, for each iteration, we start with $\hat{z}_{k,0} := \hat{x}_k$, through $i = 1, \ldots, m$,

$$\hat{z}_{k,i} := P_X\big(\hat{z}_{k,i-1} - v_k \hat{g}_{k,i}\big), \text{ where } \hat{g}_{k,i} \in \begin{cases} \partial f_i(\hat{z}_{k,i-1}) \cap S(0,1), & \text{if } f_i(\hat{z}_{k,i-1}) > f_i^*, \\ \{0\}, & \text{otherwise,} \end{cases} \tag{3.28}$$

and finally arrive at $\hat{x}_{k+1} := \hat{z}_{k,m}$. Comparing with Algorithm 1, we observe that the process $\{\hat{x}_k\}$ is identical to $\{x_k\}$.

We prove by contradiction, assuming that $f(\hat{x}_k) > f^*$ for any $k \in \mathbb{N}$. Fixing $k \in \mathbb{N}$, we define

$$I_k := \{i \in \{1, \ldots, m\} : f_i(\hat{z}_{k,i-1}) > f_i^*\}. \tag{3.29}$$

Clearly, $I_k \neq \emptyset$; otherwise, $f(\hat{x}_k) = f^*$ and a contradiction is achieved. Fix $i \in I_k$. By the assumption that $\mathbf{B}(x^*, \sigma) \subseteq X^*$ and $\|\hat{g}_{k,i}\| = 1$, one has that $x^* + \sigma \hat{g}_{k,i} \in X^*$, and hence $f_i(x^* + \sigma \hat{g}_{k,i}) = f_i^* < f_i(\hat{z}_{k,i-1})$ (cf. (3.29)). Then it follows from Definition 2.1 that $\langle \hat{g}_{k,i}, x^* + \sigma \hat{g}_{k,i} - \hat{z}_{k,i-1} \rangle \leq 0$. Consequently,

$$\langle \hat{g}_{k,i}, \hat{z}_{k,i-1} - x^* \rangle \geq \sigma \text{ when } i \in I_k, \text{ and } \langle \hat{g}_{k,i}, \hat{z}_{k,i-1} - x^* \rangle = 0 \text{ otherwise}$$

(by (3.28) that $\hat{g}_{k,i} = 0$ when $i \notin I_k$). Therefore, we obtain that

$$\sum_{i=1}^{m} \langle \hat{g}_{k,i}, \hat{z}_{k,i-1} - x^* \rangle \geq |I_k|\sigma \geq \sigma \quad \text{for any } k \in \mathbb{N}, \tag{3.30}$$

where $|I_k| \geq 1$ since $I_k \neq \emptyset$.

On the other hand, by (3.28), it follows that

$$\begin{aligned} \|\hat{z}_{k,i} - x^*\|^2 &\leq \|\hat{z}_{k,i-1} - v_k \hat{g}_{k,i} - x^*\|^2 \\ &\leq \|\hat{z}_{k,i-1} - x^*\|^2 - 2v_k \langle \hat{g}_{k,i}, \hat{z}_{k,i-1} - x^* \rangle + v_k^2. \end{aligned}$$

Summing the above inequality over $i = 1, \ldots, m$, one has

$$v_k \sum_{i=1}^{m} \langle \hat{g}_{k,i}, \hat{z}_{k,i-1} - x^* \rangle \leq \frac{\|\hat{x}_{k+1} - x^*\|^2 - \|\hat{x}_k - x^*\|^2}{2} + \frac{m v_k^2}{2};$$

consequently,

$$\frac{\sum_{k=0}^{n-1}\left(v_k\sum_{i=1}^{m}\langle\hat{g}_{k,i},\hat{z}_{k,i-1}-x^*\rangle\right)}{\sum_{k=0}^{n-1}v_k}\leq\frac{\|x_0-x^*\|^2}{2\sum_{k=0}^{n-1}v_k}+\frac{m\sum_{k=0}^{n-1}v_k^2}{2\sum_{k=0}^{n-1}v_k}. \tag{3.31}$$

We now claim, under the assumption of (i) or (ii), that

$$\liminf_{n\to\infty}\sum_{i=1}^{m}\langle\hat{g}_{n,i},\hat{z}_{n,i-1}-x^*\rangle<\sigma. \tag{3.32}$$

(i) When a constant stepsize $v\in(0,\frac{2\sigma}{m})$ is used, (3.31) is reduced to

$$\frac{\sum_{k=0}^{n-1}\sum_{i=1}^{m}\langle\hat{g}_{k,i},\hat{z}_{k,i-1}-x^*\rangle}{n}\leq\frac{\|x_0-x^*\|^2}{2nv}+\frac{mv}{2},$$

and thus, by Lemma 2.3, we obtain that

$$\liminf_{n\to\infty}\sum_{i=1}^{m}\langle\hat{g}_{n,i},\hat{z}_{n,i-1}-x^*\rangle\leq\liminf_{n\to\infty}\frac{\sum_{k=0}^{n-1}\sum_{i=1}^{m}\langle\hat{g}_{k,i},\hat{z}_{k,i-1}-x^*\rangle}{n}\leq\frac{mv}{2}<\sigma.$$

This shows (3.32), as desired.

(ii) When a diminishing stepsize is used, by (3.20) and (3.21), it also follows from Lemma 2.3 and (3.31) that

$$\liminf_{n\to\infty}\sum_{i=1}^{m}\langle\hat{g}_{n,i},\hat{z}_{n,i-1}-x^*\rangle\leq\liminf_{n\to\infty}\frac{\sum_{k=0}^{n-1}\left(v_k\sum_{i=1}^{m}\langle\hat{g}_{k,i},\hat{z}_{k,i-1}-x^*\rangle\right)}{\sum_{k=0}^{n-1}v_k}\leq 0<\sigma.$$

Hence we proved (3.32) under the assumption of (i) or (ii), which arrives at a contradiction with (3.30). The proof is complete. □

3.2 Randomized incremental subgradient method

It could be very computationally expensive to calculate the subgradients of all component functions at each iteration of the incremental subgradient method, especially when the number of component functions is large and the calculation of subgradients is not simple. To economize on the computational cost of each iteration, we propose a randomized incremental subgradient method, in which only one component function f_{ω_i} is randomly selected to construct the subgradient direction at each iteration, rather than to take each f_i into account exactly once within an ordered cycle.

This subsection aims to explore the convergence properties of the randomized incremental subgradient method for solving problem (1.1) when using typical stepsize rules. The randomized incremental subgradient method is formally presented as follows.

Algorithm 2: Randomized incremental subgradient method.

Initialize an initial point $x_0 \in \mathbb{R}^n$, a stepsize sequence $\{v_k\}$, and let $k := 0$;
while $f(x_k) > f^*$ **do**
  Let $I_k := \{i \in \{1, \ldots, m\} : f_i(x_k) > f_i^*\}$;
  Pick up equiprobably a random variable ω_k from the set I_k, calculate $g_{k,\omega_k} \in \partial f_{\omega_k}(x_k) \cap S(0,1)$, and let $x_{k+1} := P_X\left(x_k - v_k g_{k,\omega_k}\right)$;
  Let $k := k + 1$.
end

We recall the supermartingale convergence theorem, which is useful in the convergence analysis of the randomized incremental subgradient method.

Theorem 3.5 ([10, pp. 148])**.** *Let $\{Y_k\}$, $\{Z_k\}$ and $\{W_k\}$ be three sequences of random variables, and let $\{\mathcal{F}_k\}$ be a sequence of sets of random variables such that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ for any $k \in \mathbb{N}$. Suppose for any $k \in \mathbb{N}$ that*

(a) *Y_k, Z_k and W_k are functions of nonnegative random variables in $\mathcal{F}_k$;*

(b) *$\mathbf{E}\{Y_{k+1} \mid \mathcal{F}_k\} \leq Y_k - Z_k + W_k$;*

(c) *$\sum_{k=0}^{\infty} W_k < \infty$.*

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

To begin the convergence analysis of Algorithm 2, we provide below a basic inequality of a randomized incremental subgradient iteration in terms of expectation. Recall that $R_{p,m}$ is defined in (3.8).

Lemma 3.2. *Suppose that Assumptions 1 and 2 are satisfied. Let $\{x_k\}$ be a sequence generated by Algorithm 2, and let $\mathcal{F}_k := \{x_0, x_1, \ldots, x_k\}$ for any $k \in \mathbb{N}$. Then it holds, for any $x^* \in X^*$ and any $k \in \mathbb{N}$, that*

$$\mathbf{E}\left\{\|x_{k+1} - x^*\|^2 \mid \mathcal{F}_k\right\} \leq \|x_k - x^*\|^2 - 2v_k \frac{R_{p,m}}{m}(f(x_k) - f^*)^{\frac{1}{p}} + v_k^2. \tag{3.33}$$

Proof. Fix $x^* \in X^*$ and $k \in \mathbb{N}$. In view of Algorithm 2 and by the nonexpansive property of projection operator, we have

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - v_k g_{k,\omega_k} - x^*\|^2 = \|x_k - x^*\|^2 - 2v_k \left\langle g_{k,\omega_k}, x_k - x^*\right\rangle + v_k^2. \tag{3.34}$$

Note by Assumption 1 that $x^* \in X^* = \cap_{i=1}^m X_i^*$, and so $x^* \in X_{\omega_k}^*$. By Algorithm 2, one sees that $\omega_k \in I_k$ and thus $f_{\omega_k}(x_k) > f_{\omega_k}^*$. Then, Lemma 2.1 is applicable (with f_{ω_k}, x_k, $X_{\omega_k}^*$ in place of h, x, X^*) to concluding that

$$\left\langle g_{k,\omega_k}, x_k - x^*\right\rangle \geq \left(\frac{f_{\omega_k}(x_k) - f_{\omega_k}(x^*)}{L_{\omega_k}}\right)^{\frac{1}{p}} \geq L_{\max}^{-\frac{1}{p}}\left(f_{\omega_k}(x_k) - f_{\omega_k}^*\right)^{\frac{1}{p}}.$$

Then (3.34) is reduced to

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}}\left(f_{\omega_k}(x_k)-f^*_{\omega_k}\right)^{\frac{1}{p}} + v_k^2.$$

Taking the conditional expectation with respect to $\mathcal{F}_k$, it follows that

$$\mathbf{E}\left\{\|x_{k+1}-x^*\|^2 \mid \mathcal{F}_k\right\} \le \|x_k-x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}}\mathbf{E}\left\{\left(f_{\omega_k}(x_k)-f^*_{\omega_k}\right)^{\frac{1}{p}} \mid \mathcal{F}_k\right\} + v_k^2. \tag{3.35}$$

Below, we provide an estimation of the term $\mathbf{E}\left\{\left(f_{\omega_k}(x_k)-f^*_{\omega_k}\right)^{\frac{1}{p}} \mid \mathcal{F}_k\right\}$. Noting in Algorithm 2 that ω_k is uniformly distributed on I_k, we have $P(\omega_k = i) = \frac{1}{|I_k|}$ for each $i \in I_k$, and then conclude by the elementary of probability theory that

$$\begin{aligned}\mathbf{E}\left\{\left(f_{\omega_k}(x_k)-f^*_{\omega_k}\right)^{\frac{1}{p}} \mid \mathcal{F}_k\right\} &= \tfrac{1}{|I_k|}\textstyle\sum_{i\in I_k}\left(f_i(x_k)-f_i^*\right)^{\frac{1}{p}} \\ &\ge \tfrac{1}{|I_k|}\min\left\{1, |I_k|^{1-\frac{1}{p}}\right\}\left(\textstyle\sum_{i\in I_k}(f_i(x_k)-f_i^*)\right)^{\frac{1}{p}},\end{aligned} \tag{3.36}$$

where the inequality follows from Lemma 2.2 (with $f_i(x_k)-f_i^*$ and $\frac{1}{p}$ in place of a_i and γ). By the definition of I_k (see Algorithm 2), it follows that $f_i(x_k) = f_i^*$ for each $i \notin I_k$, and so, by Assumption 1, one has

$$\sum_{i\in I_k}(f_i(x_k)-f_i^*) = \sum_{i=1}^{m}(f_i(x_k)-f_i(x^*)) = f(x_k)-f^*. \tag{3.37}$$

Note by $|I_k| \le m$ that

$$\frac{1}{|I_k|}\min\left\{1, |I_k|^{1-\frac{1}{p}}\right\} = \min\left\{|I_k|^{-1}, |I_k|^{-\frac{1}{p}}\right\} \ge \frac{1}{m}\min\left\{1, m^{1-\frac{1}{p}}\right\}. \tag{3.38}$$

Therefore, by (3.37) and (3.38), (3.36) is reduced to that

$$\mathbf{E}\left\{\left(f_{\omega_k}(x_k)-f^*_{\omega_k}\right)^{\frac{1}{p}} \mid \mathcal{F}_k\right\} \ge \frac{1}{m}\min\left\{1, m^{1-\frac{1}{p}}\right\}\left(f(x_k)-f^*\right)^{\frac{1}{p}},$$

which, together with (3.35) and (3.8), yields (3.33). The proof is complete. □

By virtue of Lemma 3.2, we explore the convergence properties (with probability 1) of the randomized incremental subgradient method when using different stepsize rules in Theorems 3.6-3.8, respectively.

Theorem 3.6. *Suppose Assumptions 1 and 2 are satisfied. Let $\{x_k\}$ be a sequence generated by Algorithm 2 with the constant stepsize rule* (S1)*. Then it holds, with probability 1, that*

$$\liminf_{k\to\infty} f(x_k) \le f^* + \left(\frac{mv}{2R_{p,m}}\right)^p. \tag{3.39}$$

Proof. Fix $\delta > 0$, and define a set X_δ by

$$X_\delta := X \cap \mathrm{lev}_{<f^* + \left(\frac{mv}{2R_{p,m}} + \delta\right)^p} f.$$

Let $y_\delta \in X$ be such that $f(y_\delta) = f^* + \delta^p$ (this y_δ is well-defined by the continuity of f). Hence $y_\delta \in X_\delta$ by construction. We define a new process $\{\hat{x}_k\}$ by letting $\hat{x}_0 := x_0$ and

$$\hat{x}_{k+1} := \begin{cases} P_X\left(\hat{x}_k - v_k \hat{g}_{k,\hat{\omega}_k}\right), & \text{if } \hat{x}_k \notin X_\delta, \\ y_\delta, & \text{otherwise,} \end{cases}$$

where $\hat{g}_{k,\hat{\omega}_k} \in \partial f_{\hat{\omega}_k}(\hat{x}_k) \cap S(0,1)$. Clearly, the process $\{\hat{x}_k\}$ is identical to $\{x_k\}$, except that $\hat{x}_k$ enters X_δ and then the process terminates with $\hat{x}_k = y_\delta \in X_\delta$.

Assume that $\hat{x}_k \notin X_\delta$ for any $k \in \mathbb{N}$, and let $\hat{\mathcal{F}}_k := \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_k\}$ for any $k \in \mathbb{N}$. It says that $f(\hat{x}_k) \geq f^* + \left(\frac{mv}{2R_{p,m}} + \delta\right)^p$, and then it follows from Lemma 3.2 that the following relation holds for any $x^* \in X^*$ and any $k \in \mathbb{N}$:

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

Then, by Theorem 3.5, we obtain that $\sum_{k=0}^{\infty} 2v\delta\frac{R_{p,m}}{m} < \infty$ with probability 1, which is impossible. Hence, $\hat{x}_k \notin X_\delta$ only occurs finitely many times, and so $\hat{x}_k \in X_\delta$ for large k. Consequently, in the original process, it holds with probability 1 that

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \left(\frac{mv}{2R_{p,m}} + \delta\right)^p.$$

Since $\delta > 0$ is arbitrary, (3.39) is obtained by letting $\delta \to 0$, and the proof is complete. □

Remark 3.5. Theorem 3.6 depicts the convergence of Algorithm 2 to the optimal value of problem (1.1) within a tolerance, expressed in terms of the stepsize, the number of component functions and parameters of Hölder conditions, when the constant stepsize rule is adopted. It is observed by (3.17) and (3.39) that the randomized incremental subgradient method (Algorithm 2) admits a much less error bound than that of the incremental subgradient method (Algorithm 1) when adopting the same stepsize. Indeed,

$$\frac{\left(\frac{mv}{2R_{p,m}}\right)^p}{\left(\frac{m^2 v}{2C_{p,m}}\right)^p} = \frac{C_{p,m}^p}{R_{p,m}^p m^p} = \frac{\min\left\{1, (2m)^{p-1}\right\}}{\min\left\{1, m^{p-1}\right\} m^p} \leq \frac{\max\left\{1, 2^{p-1}\right\}}{m^p} \ll 1.$$

The proof of the following theorem uses the property of the diminishing stepsize rule (cf. (3.4)) and a line of analysis similar to that of Theorem 3.6. Hence we omit the details.

Theorem 3.7. *Suppose Assumptions 1 and 2 are satisfied. Let $\{x_k\}$ be a sequence generated by Algorithm 2 with the diminishing stepsize rule* (S2)*. Then $\liminf_{k\to\infty} f(x_k) = f^*$ with probability 1.*

Theorem 3.8. *Suppose Assumptions 1 and 2 are satisfied. Let $\{x_k\}$ be a sequence generated by Algorithm 2 with the dynamic stepsize rule* (S4)*. Then $\{x_k\}$ converges to an optimal solution of problem* (1.1) *with probability 1.*

Proof. By Lemma 3.2 and (3.7), it follows that, for any $x^* \in X^*$ and any $k \in \mathbb{N}$,

$$\begin{aligned}\mathbf{E}\left\{\|x_{k+1}-x^*\|^2 \mid \mathcal{F}_k\right\} &\le \|x_k-x^*\|^2 - 2v_k\tfrac{R_{p,m}}{m}(f(x_k)-f^*)^{\frac{1}{p}} + v_k^2\\ &= \|x_k-x^*\|^2 - \gamma_k(2-\gamma_k)\tfrac{R_{p,m}^2}{m^2}(f(x_k)-f^*)^{\frac{2}{p}}\\ &\le \|x_k-x^*\|^2 - \underline{\gamma}(2-\overline{\gamma})\tfrac{R_{p,m}^2}{m^2}(f(x_k)-f^*)^{\frac{2}{p}}.\end{aligned}$$

Then it follows from Theorem 3.5 that $\{\|x_k-x^*\|\}$ is convergent and $\sum_{k=1}^{\infty}(f(x_k)-f^*)^{\frac{2}{p}} < \infty$ with probability 1; consequently, $\lim_{k\to\infty} f(x_k) = f^*$ with probability 1.

Let $(\Omega, \mathcal{F}, P)$ be the probability space. Let Z be a countable and dense subset of X^*, and let

$$\Theta(z) := \{\omega : \{\|x_k(\omega)-z\|\} \text{ is convergent}\} \qquad \text{for any } z \in Z,$$

and

$$\Theta := \bigcap_{z\in Z}\Theta(z).$$

Recall that $\{\|x_k - x^*\|\}$ is convergent with probability 1, that is $P(\Theta(x^*)) = 1$, for any $x^* \in X^*$. Then it follows that $P(\Theta(z)^c) = 0$ for any $z \in Z \subseteq X^*$. By the elements of probability theory, one checks that

$$P(\Theta) = 1 - P(\Theta^c) = 1 - P\left(\bigcup_{z\in Z}\Theta(z)^c\right) \ge 1 - \sum_{z\in Z} P(\Theta(z)^c) = 1. \tag{3.40}$$

For any $\omega \in \Theta$ and any $z \in Z$, it says that $\{\|x_k(\omega)-z\|\}$ is convergent; hence $\{x_k(\omega)\}$ is bounded and must have a cluster point. Define $\bar{x} : \Omega \to \mathbb{R}^n$ be such that

$$\bar{x}(\omega) \text{ is a cluster point of } \{x_k(\omega)\} \text{ for any } \omega \in \Theta.$$

Note again that $\lim_{k\to\infty} f(x_k) = f^*$ with probability 1. Without loss of generality, we can assume that $\lim_{k\to\infty} f(x_k(\omega)) = f^*$ for any $\omega \in \Theta$. Then it follows from the continuity of f that

$$\bar{x}(\omega) \in X^* \quad \text{for any } \omega \in \Theta. \tag{3.41}$$

Fix $\epsilon > 0$ and $\omega \in \Theta$. Since $\bar{x}(\omega) \in X^*$ and $Z \subseteq X^*$ is dense, there exists $z(\omega) \in Z$ such that

$$\|\bar{x}(\omega) - z(\omega)\| \le \frac{\epsilon}{3}. \tag{3.42}$$

Let $\{x_{k_i}(\omega)\}$ be a subsequence of $\{x_k(\omega)\}$ such that $\lim_{i\to\infty} x_{k_i}(\omega) = \bar{x}(\omega)$. Hence we obtain by (3.42) that $\lim_{i\to\infty}\|x_{k_i}(\omega) - z(\omega)\| \le \frac{\epsilon}{3}$. By the definition of Θ, one has that $\{\|x_k(\omega) - z(\omega)\|\}$ is convergent, and so $\lim_{k\to\infty}\|x_k(\omega) - z(\omega)\| \le \frac{\epsilon}{3}$. Then there exists $N \in \mathbb{N}$ such that

$$\|x_k(\omega) - z(\omega)\| \le \frac{2\epsilon}{3} \quad \text{for any } k \ge N.$$

This, together with (3.42), yields

$$\|x_k(\omega) - \bar{x}(\omega)\| \leq \|x_k(\omega) - z(\omega)\| + \|\bar{x}(\omega) - z(\omega)\| \leq \epsilon \quad \text{for any } k \geq N.$$

This shows that $\{x_k(\omega)\}$ converges to $\bar{x}(\omega)$ for any $\omega \in \Theta$. This, together with (3.41), says that

$$\Theta \subseteq \{\omega \in \Omega : \{x_k(\omega)\} \text{ converges to } \bar{x}(\omega)\} \cap \{\omega \in \Omega : \bar{x}(\omega) \in X^*\}.$$

Noting by (3.40) that $P(\Theta) = 1$, we conclude

$$P\left(\{\omega \in \Omega : \{x_k(\omega)\} \text{ converges to } \bar{x}(\omega),\, \bar{x}(\omega) \in X^*\}\right) \geq P(\Theta) = 1.$$

The proof is complete. □

4 Application to sum of ratios problems

This section aims to present an important application of sum-minimization problem (1.1) of a number of quasi-convex component functions. Typically, fractional programming, optimizing a certain indicator (e.g. efficiency) characterized by a ratio of technical terms, is widely applied in various areas; see [4, 14, 44] and references therein. In particular, the sum of ratios problem (SOR) [41] is a typical fractional programming and has a variety of important applications in economics and management science, which is formulated as

$$\begin{array}{ll} \max & \sum_{i=1}^{m} R_i(x) := \frac{p_i(x)}{c_i(x)} \\ \text{s.t.} & x \in X, \end{array} \tag{4.1}$$

where $p_i : \mathbb{R}^n \to \mathbb{R}$ is nonnegative and concave, $c_i : \mathbb{R}^n \to \mathbb{R}$ is positive and convex for each $i = 1, \ldots, m$. It is difficult to globally solve the SOR (4.1), especially for large-scale problems.

Exploiting the additivity structure of problem (4.1), we propose a new approach to find a global optimal solution of the SOR. By [44, Theorems 2.3.3 and 2.5.1], we have that the ratio R_i is quasi-concave for each $i = 1, \ldots, m$, and so (4.1) is a sum-maximization problem of a number of quasi-concave functions. This shows that the SOR falls in the framework (1.1). Moreover, let r_i denote the maximal ratio of R_i over X, and define $h_i(\cdot) := r_i - R_i(\cdot)$. The SOR (4.1) can also be approached by solving the resulting quasi-convex feasibility problem (3.2). In [12], Censor and Segal proposed a subgradient projection method to solve the quasi-convex feasibility problem (3.2) by using the dynamic stepsize rule. In the numerical study, we apply the incremental subgradient methods and the subgradient projection method to solve the SOR (4.1) and its reformulated quasi-convex feasibility problem (3.2), respectively, and the abbreviations of these methods are listed in Table 1.

In the numerical study, we consider the multiple Cobb-Douglas productions efficiency problem (in short, MCDPE) [11], which is an application of the SOR. Formally, consider a set of m productions with s projects and n factors. Let $x := (x_j)^T \in \mathbb{R}^n$ denote the amounts

Table 1: List of the algorithms compared in the numerical study.

Abbreviations	Algorithms
SGPM	**S**ub**G**radient **P**rojection **M**ethod in [12], which is to solve (3.2).
IncSGM	**Inc**remental **S**ub**G**radient **M**ethod (Algorithm 1) for solving (4.1).
RandSGM	**Rand**omized incremental **S**ub**G**radient **M**ethod (Algorithm 2) for solving (4.1).

of n factors. The profit function of production i can be expressed as the Cobb-Douglas production function

$$p_i(x) := a_{i,0} \prod_{j=1}^{n} x_j^{a_{i,j}},$$

where $a_{i,j} \geq 0$ for $j = 0, \ldots, n$ and $\sum_{j=1}^{n} a_{i,j} = 1$. The cost function of production i is formulated as a linear function

$$c_i(x) := \sum_{j=1}^{n} c_{i,j} x_j + c_{i,0},$$

where $c_{i,j} \geq 0$ for $j = 0, \ldots, n$. Due to the daily profit or operating cost constraints, the amounts of investment for factors should fall in the constraint set

$$X := \{x \in \mathbb{R}_+^n : \sum_{j=1}^{n} b_{tj} x_j \geq p_t, \quad t = 1, \ldots, s\}.$$

Then the MCDPE is modeled as the SOR (4.1). In the numerical experiments, the parameters of MCDPE are randomly chosen from different intervals:

$$a_{i,0} \in [0, 10], \quad a_{i,j}, b_{tj}, c_{i,0}, c_{i,j} \in [0, 1], \quad \text{and} \quad p_t \in [0, n/2].$$

The diminishing stepsize rule is chosen as

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

where v is always chosen between $[2, 5]$, while the constant stepsize is selected between $[1, 2]$. All numerical experiments are implemented in MATLAB R2014a and executed on a personal laptop (Intel Core i5, 3.20 GHz, 8.00 GB of RAM).

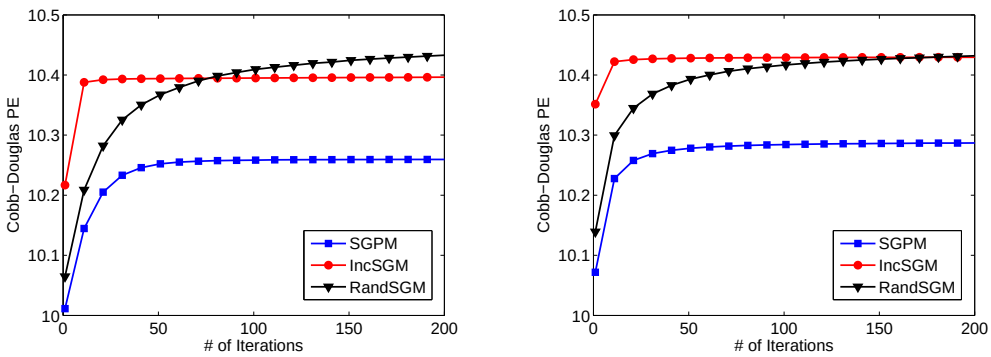
We first compare the performances (in both the obtained objective value and the CPU time) of the SGPM, IncSGM and RandSGM for different dimensions. The computation results are displayed in Table 2. In this table, the columns of Projects, Factors and Productions represent the numbers of projects (s), factors (n) and productions (m) of MCDPE, and the columns of f_{opt} and CPUtime denote the obtained optimal value and the CPU time (seconds) cost to reach f_{opt} by each algorithm, respectively. It is observed from Table 2 that the IncSGM and RandSGM outperform the SGPM in the sense that they achieve a larger production efficiency in a shorter time than the SGPM for different dimensional MCDPEs.

The second experiment is to compare the convergence behavior of the SGPM, IncSGM and RandSGM by using the constant and diminishing stepsize rules, where the problem

Table 2: Computation results for maximizing MCDPE.

Circumstance of problem			SGPM		IncSGM		RandSGM	
Projects	Factors	Productions	f_{opt}	CPUtime	f_{opt}	CPUtime	f_{opt}	CPUtime
50	50	10	23.31	0.51	23.46	0.17	23.48	0.18
50	50	100	210.22	3.38	211.86	2.41	211.84	1.74
100	100	10	11.73	0.41	11.77	0.26	11.81	0.23
100	100	100	104.20	2.62	106.52	1.40	106.49	1.03
500	500	10	2.21	1.45	2.31	0.54	2.34	0.38
500	500	100	21.01	9.61	21.25	5.93	21.24	4.28
1000	1000	10	1.15	3.23	1.19	1.69	1.21	1.47
1000	1000	100	10.56	19.64	10.62	12.48	10.60	10.41

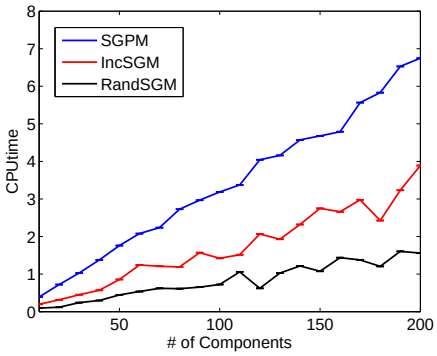
size is fixed to be $(m, n, s) = (10, 100, 100)$. We summary the averaged performance of the compared algorithms in 500 random trials. Figure 1 plots the mean of the estimated Cobb-Douglas production efficiencies along the number of the iterations in these 500 trials, from which we observe that the IncSGM converges faster (in terms of the number of iterations) to an (approximate) optimal value that the RandSGM and the SGPM. Furthermore, Figure 1(a) illustrates that the RandSGM obtains a better estimation than the IncSGM when the constant stepsize rule is adopted, which is consistent with Remark 3.5. Figure 1(b) demonstrates that both IncSGM and RandSGM converge to an optimal value when the diminishing stepsize rule is employed, which is consistent with Theorems 3.2 and 3.7. It is also shown that both IncSGM and RandSGM approach a better solution that the SGPM. Figure 2 plots the error bars of the CPU times in 500 trials when varying the number of component functions from 10 to 200. It is revealed that the RandSGM is faster (in terms of CPU time) than the IncSGM, which is faster than the SGPM. This indicates the potential applicability of the RandSGM to the large-scale SOR. Figure 3 plots the obtained maximal production efficiencies in each of these 500 trials. It is observed that the IncSGM and RandSGM outperform the SGPM consistently.



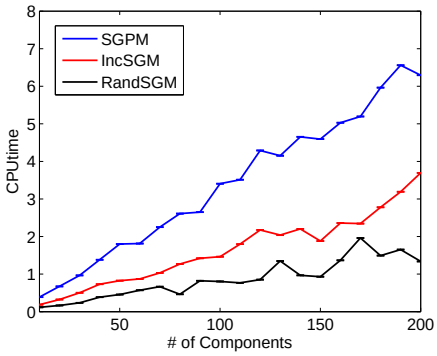
(a) The constant stepsize rule.

(b) The diminishing stepsize rule.

Figure 1: The averaged convergence behavior of SGPM, IncSGM and RandSGM in 500 random MCDPEs.

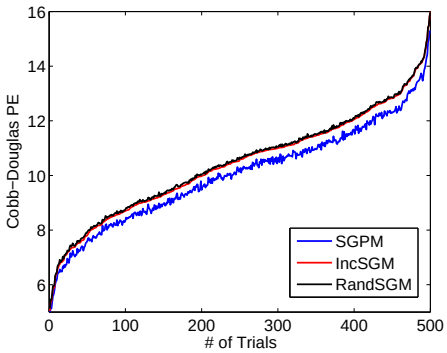


(a) The constant stepsize rule.

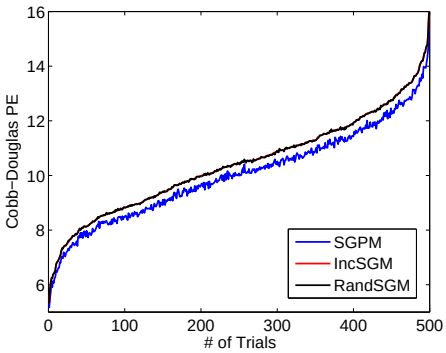


(b) The diminishing stepsize rule.

Figure 2: Variation of CPU time when varying the number of component functions.



(a) The constant stepsize rule.



(b) The diminishing stepsize rule.

Figure 3: Overall of obtained maximal values in 500 random MCDPEs.

Finally, we conduct 500 simulations to show the stability of RandSGM, which start from the same initial point, adopt the same stepsizes (constant: $v_k \equiv 1.5$ or diminishing: $v_k = 3/(1 + 0.1k)$) and solve a same MCDPE, but follow different stochastic processes. Figure 4 plots the error bars of the estimated Cobb-Douglas production efficiencies in these 500 simulations. It is shown that the RandSGM is highly stable and converges to an optimal value with probability 1.

References

[1] Y. Almogy and O. Levin. Parametric analysis of a multi-stage stochastic shipping problem. *in Operational Research 69*, Tavistok, London, 1970.

[2] 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.

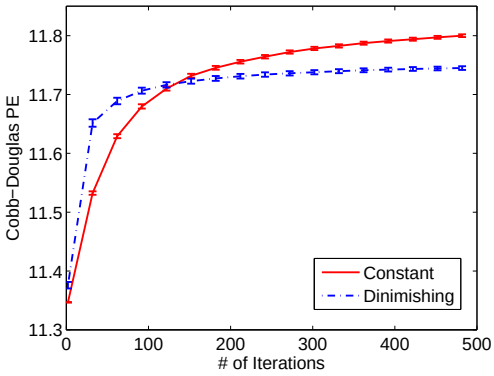


Figure 4: The error bars of RandSGM in 500 simulations.

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

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

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

[6] H. Benson. Branch-and-bound outer approximation algorithm for sum-of-ratios fractional programs. *Journal of Optimization Theory Applications*, 146:1–18, 2010.

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

[8] D. P. Bertsekas. *Convex Optimization ang Algorithms*. Athena Scientific, Massachusetts, 2015.

[9] D. P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Mathematical Programming*, 129(2):163–195, 2011.

[10] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA, 1996.

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

[12] Y. Censor and A. Segal. Algorithms for the quasiconvex feasibility problem. *Journal of Computational and Applied Mathematics*, 185(1):34–50, 2006.

[13] C. S. Colantoni, R. P. Manes, and A. Whinston. Programming, profit rates, and pricing decisions. *The Accounting Review*, 44:467–481, 1969.

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

[15] 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.

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

[17] R. W. Freund and F.Jarre. Solving the sum-of-ratios problem by an interior-point method. *Journal of Global Optimization*, 19:83–102, 2001.

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

[19] J.-L. Goffin, Z.-Q. Luo, and Y. Ye. On the complexity of a column generation algorithm for convex or quasiconvex feasibility problems. *In Large Scale Optimization: State of the Art, W. W. Hager, D. W. Hearn and P. M. Pardalos (eds.).* Kluwer Academic Publishers, Dordrecht, The Netherlands, 182–191, 1994.

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

[21] Y. Hu, C. Li, and X. Yang. On convergence rates of linearized proximal algorithms for convex composite optimization with applications. *SIAM Journal on Optimization*, 26(2):1207–1235, 2016.

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

[23] Y. 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.

[24] X. Huang and X. Yang. A unified augmented Lagrangian approach to duality and exact penalization. *Mathematics of Operations Research*, 28(3):533–552, 2003.

[25] B. Johansson, M. Rabi, and M. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Optimization*, 20(3):1157–1170, 2009.

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

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

[28] H. Konno and M. Inori. Bond portfolio optimization by bilinear fractional programming. *Journal of Operations Research Society of Japan*, 32:143–158, 1989.

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

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

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

[32] J. Mairal. Incremental majorization-minimization optimization with application to large-scale machine learning. *SIAM Journal on Optimization*, 25(2):829–855, 2015.

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

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

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

[36] 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.

[37] N. Nimana, A. P. Farajzadeh, and N. Petrot. Adaptive subgradient method for the split quasi-convex feasibility problems. *Optimization*, 65(10):1885–1898, 2016.

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

[39] 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.

[40] S. S. Ram, A. Nedić, and V. V. Veeravalli. Incremental stochastic subgradient algorithms for convex optimization. *SIAM Journal on Optimization*, 20(2):691–717, 2009.

[41] S. Schaible and J. Shi. Fractional programming: The sum-of-ratios case. *Optimization Methods and Software*, 18:219–229, 2003.

[42] Q. Shi, C. He, and L. Jiang. Normalized incremental subgradient algorithm and its application. *IEEE Transactions on Signal Processing*, 57(10):3759–3774, 2009.

[43] N. Z. Shor. *Minimization Methods for Non-differentiable Functions*. Springer-Verlag, New York, 1985.

[44] I. M. Stancu-Minasian. *Fractional Programming.* Kluwer Academic Publishers, Dordrecht, 1997.

[45] C. Tan, S. Ma, Y.-H. Dai, and Y. Qian. Barzilai-Borwein step size for stochastic gradient descent. *In NIPS*, 2016.

[46] L. Xiao and T. Zhang. A proximal stochastic gradient method with progressive variance reduction. *SIAM Journal on Optimization*, 24(4):2057–2075, 2014.