Incremental Quasi-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 the sum of a number of nondifferentiable quasi-convex component functions over a closed and convex set. 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 quasi-subgradient method for this problem and investigate its convergence properties to a global optimal value/solution when using the constant, diminishing or dynamic stepsize rules and under a homogeneous assumption and the Hölder condition. To economize on the computation cost of subgradients of a large number of component functions, we further propose a randomized incremental quasi-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 iterates with probability 1. The proposed incremental quasi-subgradient methods are applied to solve the quasi-convex feasibility problem and 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-scale sum of ratios problem.
Keywords Quasi-convex programming, sum-minimization problem, sum of ratios problem, subgradient method, incremental approach

*Shenzhen Key Laboratory of Advanced Machine Learning and Applications, 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, 11871347), Natural Science Foundation of Guangdong (2016A030310038), Natural Science Foundation of Shenzhen (J-CYJ20170817100950436, JCYJ20170818091621856) and Interdisciplinary Innovation Team of Shenzhen University.

†Department of Mathematics and Statistics, The Hang Seng University of Hong Kong, Hong Kong (carisayu@hsu.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/P02/15 and UGC/FDS14/P02/17).

‡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 the 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 [8, 33], the distributed optimization problem in large-scale sensor networks [9, 39] and the empirical risk minimization problem in online machine learning [14, 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, 8, 20, 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 [15], and until now, various variants of subgradient methods have been studied to solve structured optimization problems; see [2, 7, 27, 31, 33, 35, 43] and references therein. To meet the structure of the sum-minimization problem (1.1), the 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., [14, 39]. So far, many articles have been devoted to the convergence study 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.

Recently, much attention has 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, 13, 19] 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 [21], the conditional subgradient method [23] and stochastic subgradient method [22]. More recently, a unified framework of convergence analysis of subgradient methods for quasi-convex optimization problems was provided in [47], in which the convergence theorem was established for a certain class of sequences satisfying a general basic inequality. 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 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 [12] and bond portfolio optimization [28]. Although some numerical methods, such as the branch and bound scheme [6] and the interior point method [16], 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 $i = 1, \dots, m$, and $X \subseteq \mathbb{R}^n$ is nonempty, closed and convex.

Note that the objective function 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 the standard subgradient method [26] to solve problem (1.1) is not necessarily convergent, and the convergence results of the literature [21, 22, 47] cannot be directly applied to this paper.

Inspired by the idea of incremental approach, we propose a new incremental quasi-subgradient method to solve the sum-minimization problem (1.1), which is different from the classical incremental subgradient method (1.2) in that the Greenberg-Pierskalla quasi-subgradient is employed (in place of the convex subgradient) at each subiteration and 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 for the component functions, we provide a proper basic inequality and establish the convergence properties of the proposed incremental quasi-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 calculations 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 [7, 14, 45, 46] and references therein. Employing the idea of the stochastic gradient descent algorithm, we propose a randomized incremental quasi-subgradient method to save the computational cost of the incremental quasi-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 quasi-subgradient method enjoys the convergence properties with probability 1 and achieves a much less tolerance than that of the deterministic incremental quasi-subgradient method. To the best of our knowledge, this paper seems the first study of stochastic incremental subgradient method for the sum-minimization problem of quasi-convex functions.

Furthermore, we introduce two important classes of applications of sum-minimization problem (1.1) of a number of quasi-convex component functions: the quasi-convex feasibility problem (in short, QCFP) and the SOR. We cast the QCFP into a sum-minimization problem (1.1), and then extend the (deterministic or randomized) incremental quasi-subgradient methods to solve the QCFP, as well as the convergence theorems. For the SOR, we formulate it as a sum-minimization problem (1.1), consider the multiple Cobb-Douglas productions efficiency problem (in short, MCDPE) [10] as an application of the SOR, and conduct numerical experiments on this problem via applying the proposed incremental quasi-subgradient methods. The numerical results show that the incremental quasi-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 quasi-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 QCFP and the SOR and the numerical study for the MCDPE is presented in section 4.

2 Notations and preliminary results

The notations used in this paper are standard; see, e.g., [7]. 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

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

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

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

It is well-known that h is quasi-convex if and only if $\operatorname{lev}_{<\alpha} h$ (and/or $\operatorname{lev}_{\le\alpha} h$) 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 $\operatorname{lev}_{\le\alpha} h$ is bounded for any $\alpha \in \mathbb{R}$.

The subdifferential of a 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, 17, 21, 26] and references therein. The earliest one is the Greenberg-Pierskalla quasi-subdifferential proposed in [17], and recently, Kiwiel [26] and Hu et al. [21] introduced a quasi-subdifferential defined as a normal cone to its level set. In the following definition, we recall the notions of subdifferentials for quasi-convex function taken from [17, 21].

Definition 2.1. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex function and $x \in \mathbb{R}^n$.*

(i) *The Greenberg-Pierskalla quasi-subdifferential of h at x is defined by*

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

(ii) *The quasi-subdifferential of h at x is defined by*

$$\partial^{Q} h(x) = \left\{ g : \langle g, y - x \rangle \le 0 \text{ for any } y \in \operatorname{lev}_{<h(x)} h \right\}. \tag{2.2}$$

It is clear from definition that $\partial^{GP} h(x) \subseteq \partial^Q h(x)$ for any $x \in \mathbb{R}^n$. More specifically, the existence and relationship between the Greenberg-Pierskalla quasi-subdifferential and the quasi-subdifferential are recalled as follows.

Lemma 2.1 ([21, Lemma 2.1])**.** *Let $h : \mathbb{R}^n \to \mathbb{R}$ be quasi-convex on $\mathbb{R}^n$ and $x \in \mathbb{R}^n$. Then the following statements are true.*

(i) $\partial^Q h(x) \setminus \{0\} \neq \varnothing$.

(ii) *If* h *is upper semicontinuous on* $\mathbb{R}^n$, *then* $\partial^{GP} h(x) \neq \varnothing$ *and* $\partial^Q h(x) = \partial^{GP} h(x) \cup \{0\}$.

The following lemma will be useful in the application of the quasi-convex feasibility problem, which shows that the positive part operator (partially) preserves the quasi-convexity and the quasi-subdifferentials.

Lemma 2.2. *Let* $h : \mathbb{R}^n \to \mathbb{R}$ *be a quasi-convex function and* $z \notin \mathrm{lev}_{\leq 0} h$, *and let* $f : \mathbb{R}^n \to \mathbb{R}$ *be defined by*

$$f(x) := \max\{h(x), 0\} \quad \text{for any } x \in \mathbb{R}^n. \tag{2.3}$$

Then f *is quasi-convex and* $\partial^{GP} f(z) = \partial^{GP} h(z)$, $\partial^Q f(z) = \partial^Q h(z)$.

Proof. It is clear by (2.3) that

$$\mathrm{lev}_{<\alpha} f = \begin{cases} \mathrm{lev}_{<\alpha} h, & \alpha > 0, \\ \emptyset, & \text{otherwise}. \end{cases}$$

Note by the quasi-convexity of h that $\mathrm{lev}_{<\alpha} h$ is convex, as is $\mathrm{lev}_{<\alpha} f$, for any $\alpha \in \mathbb{R}$. Hence f is also quasi-convex. By the assumption that $h(z) > 0$, one has by (2.3) that $f(z) = h(z)$ and $\mathrm{lev}_{<f(z)} f = \mathrm{lev}_{<h(z)} h$, and thus, by (2.1) and (2.2) that $\partial^{GP} f(z) = \partial^{GP} h(z)$ and $\partial^Q f(z) = \partial^Q h(z)$, respectively. The proof is complete. $\square$

The Hölder condition was used in [29] to describe some properties of quasi-subgradients, and it plays an important role in the convergence study of subgradient-type methods for quasi-convex optimization problems [21, 22, 23].

Definition 2.2. *Let* $p \in (0, 1]$, $L > 0$ *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*

$$|h(y) - h(x)| \leq L\|y - x\|^p \quad \text{for any } y \in \mathbb{R}^n. \tag{2.4}$$

h *is said to satisfy the Hölder condition of order* p *with modulus* L *on* X *if* (2.4) *holds for any* $x \in X$.

The Hölder condition of order 1 can be guaranteed by the (global) Lipschitz continuity, and this property holds for very broad classes of functions with various values of $p \in (0, 1]$. The following lemma recalls an important property (in convergence analysis) of a quasi-convex function that satisfies the Hölder condition (as $\partial^{GP} h \subseteq \partial^Q h$).

Lemma 2.3 ([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* $p \in (0, 1]$ *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*

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

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

Lemma 2.4 ([24, Lemma 4.1]). *Let $\gamma \geq 1$ and $a_i \geq 0$ for $i = 1, \ldots, n$. Then it holds that*

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

Lemma 2.5 ([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*

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

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 quasi-subgradient methods and convergence analysis

In this section, we propose the incremental quasi-subgradient methods, including the deterministic and stochastic 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. The applications satisfying these assumptions will be presented in section 4.

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

Assumption 2. *Let $p \in (0, 1]$ 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.*

Remark 3.1. (i) It is easy to see 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 the 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 quasi-subgradient methods to resolve this issue.

(ii) The Hölder condition was assumed in [21, 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

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

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 quasi-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*:

$$v_k > 0, \quad \lim_{k\to\infty} v_k = 0, \quad \sum_{k=0}^{\infty} v_k = \infty. \tag{3.2}$$

(S3) *Dynamic stepsize rule I*:

$$v_k := \frac{2}{m}\gamma_k C_{p,m}\left(f(x_k) - f^*\right)^{\frac{1}{p}} \quad \text{for any } k \in \mathbb{N}, \tag{3.3}$$

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

$$C_{p,m} := (2mL_{\max})^{-\frac{1}{p}}. \tag{3.4}$$

(S4) *Dynamic stepsize rule II*:

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

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

$$R_{p,m} := (mL_{\max})^{-\frac{1}{p}}. \tag{3.6}$$

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

3.1 Incremental quasi-subgradient method

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

Algorithm 1: Incremental quasi-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 $z_{k,0} := x_k$;
  for $i = 1, \dots, m$ **do**
    if $f_i(z_{k,i-1}) = f_i^*$ **then**
      Let $z_{k,i} := z_{k,i-1}$;
    else
      Calculate $g_{k,i} \in \partial^{GP} f_i(z_{k,i-1}) \cap S(0,1)$, and let $z_{k,i} := P_X\left(z_{k,i-1} - v_k g_{k,i}\right)$;
    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 [33] and the incremental gradient method for smooth optimization [7]. In particular, the classical incremental gradient/subgradient method (1.2) updates subgradient subiterations in a cyclic sequence on $\{1, \dots, m\}$; 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 $\{1, \dots, m\}$ as in the classical incremental gradient/subgradient method (1.2), even though Assumptions 1 and 2 are satisfied. This is the reason why we change the updating control in the proposed incremental quasi-subgradient methods.

Example 3.1. *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 *(i.e., Lipschitz continuous) 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 $\partial^{GP} f_1(x_0) = \mathbb{R}_+$, $g_{1,1} = 1$ and $z_{1,1} = x_0 - v$. Note that

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

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 quasi-subgradient iteration. Recall that $C_{p,m}$ is defined in (3.4).

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 $k \in \mathbb{N}$, it holds that*

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 4mv_k C_{p,m}(f(x_k) - f^*)^{\frac{1}{p}} + m^2 v_k^2. \tag{3.7}$$

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

$$\|z_{k,i} - x^*\|^2 \leq \|z_{k,i-1} - x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}} (f_i(z_{k,i-1}) - f_i^*)^{\frac{1}{p}} + v_k^2. \tag{3.8}$$

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 (3.8) holds automatically; otherwise, $z_{k,i-1} \notin X_i^*$, and then one sees from Algorithm 1 that

$$z_{k,i} = P_X (z_{k,i-1} - v_k g_{k,i}). \tag{3.9}$$

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 2.3 is applicable (with f_i, $z_{k,i-1}$, X_i^* in place of h, x, X^*) to concluding that

$$f_i(z_{k,i-1}) - f_i^* = f_i(z_{k,i-1}) - f_i(x^*) \leq L_i \langle g_{k,i}, z_{k,i-1} - x^*\rangle^p \leq L_{\max} \langle g_{k,i}, z_{k,i-1} - x^*\rangle^p \tag{3.10}$$

(due to (3.1)). By the nonexpansive property of the projection operator, it follows from (3.9) that

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

where the last inequality follows from (3.10). Hence (3.8) is proved.

Next, we estimate the second term in the right hand side of (3.8) in terms of $f_i(x_k) - f_i^*$. Since $p \in (0, 1]$, by Lemma 2.4 (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

$$\begin{aligned} (f_i(x_k) - f_i^*)^{\frac{1}{p}} &\leq ((f_i(z_{k,i-1}) - f_i^*) + |f_i(z_{k,i-1}) - f_i(x_k)|)^{\frac{1}{p}} \\ &\leq 2^{\frac{1}{p}-1} \left((f_i(z_{k,i-1}) - f_i^*)^{\frac{1}{p}} + |f_i(z_{k,i-1}) - f_i(x_k)|^{\frac{1}{p}} \right). \end{aligned} \tag{3.11}$$

By Assumption 2 (cf. (2.4)) and in view of Algorithm 1, it follows that

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

Hence (3.11) is reduced to

$$(f_i(z_{k,i-1}) - f_i^*)^{\frac{1}{p}} \ge 2^{1-\frac{1}{p}} (f_i(x_k) - f_i^*)^{\frac{1}{p}} - L_{\max}^{\frac{1}{p}} v_k(i-1),$$

and so (3.8) yields that

$$\|z_{k,i} - x^*\|^2 \le \|z_{k,i-1} - x^*\|^2 - 4v_k(2L_{\max})^{-\frac{1}{p}} (f_i(x_k) - f_i^*)^{\frac{1}{p}} + (2i-1)v_k^2. \tag{3.12}$$

Finally, summing (3.12) over $i = 1, \ldots, m$, we obtain that

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 4v_k(2L_{\max})^{-\frac{1}{p}} \sum_{i=1}^{m} (f_i(x_k) - f_i^*)^{\frac{1}{p}} + m^2 v_k^2. \tag{3.13}$$

Note by Lemma 2.4 (with $f_i(x_k) - f_i^*$ and $\frac{1}{p}$ in place of a_i and γ) that

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

(thanks to Assumption 1), and thus, (3.7) is seen to hold by (3.4) and (3.13). 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 quasi-subgradient method when using different stepsize rules in Theorems 3.1-3.3, respectively.

Theorem 3.1. *Suppose that 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*

$$\liminf_{k\to\infty} f(x_k) \le f^* + \left(\frac{mv}{4C_{p,m}} \right)^p. \tag{3.14}$$

Proof. We prove by contradiction, assuming that

$$\liminf_{k\to\infty} f(x_k) > f^* + \left(\frac{mv}{4C_{p,m}}\right)^p.$$

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

$$f(x_k) > f^* + \left(\frac{mv}{4C_{p,m}} + \delta\right)^p \quad \text{for any } k \geq N.$$

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

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 4mvC_{p,m}(f(x_k) - f^*)^{\frac{1}{p}} + m^2v^2 < \|x_k - x^*\|^2 - 4mv\delta C_{p,m}.$$

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

$$\|x_t - x^*\|^2 < \|x_N - x^*\|^2 - 4m(t-N)v\delta C_{p,m},$$

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 (3.14) 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 [21, 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 [33, Proposition 2.1].

Theorem 3.2. *Suppose that 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 (3.7) over $k = 0, 1, \ldots, n-1$, we have

$$\|x_n - x^*\|^2 \leq \|x_0 - x^*\|^2 - 4mC_{p,m}\sum_{k=0}^{n-1} v_k(f(x_k) - f^*)^{\frac{1}{p}} + m^2\sum_{k=0}^{n-1} v_k^2, \tag{3.15}$$

and thus,

$$\frac{\sum_{k=0}^{n-1} v_k(f(x_k) - f^*)^{\frac{1}{p}}}{\sum_{k=0}^{n-1} v_k} \leq \frac{\|x_0 - x^*\|^2}{4mC_{p,m}\sum_{k=0}^{n-1} v_k} + \frac{m\sum_{k=0}^{n-1} v_k^2}{4C_{p,m}\sum_{k=0}^{n-1} v_k}. \tag{3.16}$$

Note by (3.2) that

$$\lim_{n\to\infty} \frac{\|x_0 - x^*\|^2}{\sum_{k=0}^{n-1} v_k} = 0, \tag{3.17}$$

and by Lemma 2.5 (with v_k in place of a_k) that

$$\lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k} = \lim_{n\to\infty} v_n = 0. \tag{3.18}$$

Consequently, by Lemma 2.5 (with $(f(x_k) - f^*)^{\frac{1}{p}}$ in place of a_k), (3.16) 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 $N \in \mathbb{N}$ such that

$$v_k \le \frac{2}{m} C_{p,m} \sigma^{\frac{1}{p}} \quad \text{for any } k \ge N. \tag{3.19}$$

Define

$$X_\sigma := X \cap \mathrm{lev}_{\le f^* + \sigma} f \quad \text{and} \quad \rho(\sigma) := \max_{x \in X_\sigma} \mathrm{dist}(x, X^*). \tag{3.20}$$

By the assumption that f is coercive, it follows that its level set $\mathrm{lev}_{\le f_* + \sigma} f$ is bounded, and so is X_σ. Thus, by (3.20), one has $\rho(\sigma) < \infty$. Fix $k \ge N$. We show

$$\mathrm{dist}(x_{k+1}, X^*) \le \max\{\mathrm{dist}(x_k, X^*), \rho(\sigma) + 2C_{p,m} \sigma^{\frac{1}{p}}\} \tag{3.21}$$

by claiming the following two implications:

$$[f(x_k) > f_* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) \le \mathrm{dist}(x_k, X^*)]; \tag{3.22}$$

$$[f(x_k) \le f_* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) \le \rho(\sigma) + 2C_{p,m} \sigma^{\frac{1}{p}}]. \tag{3.23}$$

To prove (3.22), we suppose that $f(x_k) > f_* + \sigma$. Then Lemma 3.1 is applicable to concluding that, for any $x^* \in X^*$,

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 4mC_{p,m} v_k \sigma^{\frac{1}{p}} + m^2 v_k^2 \le \|x_k - x^*\|^2 - m^2 v_k^2$$

(due to (3.19)). Consequently, one can prove (3.22) by selecting $x^* := P_{X^*}(x_k)$. To show (3.23), we suppose that $f(x_k) \le f_* + \sigma$. Then we conclude that $x_k \in X_\sigma$, and so, (3.20) 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 (3.19), shows (3.23). Therefore, (3.21) is proved as desired.

By assertion (i), 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 (3.23) that $\mathrm{dist}(x_{N+1}, X^*) \leq \rho(\sigma) + 2C_{p,m}\sigma^{\frac{1}{p}}$. Then, we inductively obtain by (3.21) that

$$\mathrm{dist}(x_k, X^*) \leq \rho(\sigma) + 2C_{p,m}\sigma^{\frac{1}{p}} \quad \text{for any } k > N. \tag{3.24}$$

Since f is continuous and coercive, its level sets are compact, and so, it is trivial to see that $\lim_{\sigma\to 0}\rho(\sigma) = 0$. Hence we obtain by (3.24) 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) By the assumption that $\sum_{k=0}^{\infty} v_k^2 < \infty$, one sees from (3.15) that $\{\|x_k - x^*\|\}$ is bounded, and so is $\{x_k\}$. Since further it was proved in assertion (i) of this theorem that $\liminf_{k\to\infty} f(x_k) = f^*$, $\{x_k\}$ 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 (3.7) (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 [21, Examples 3.1 and 3.3] that the Hölder condition (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 3.2. *Consider problem* (1.1)*, where* $X = \mathbb{R}$*,* $m = 2$*, and the 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* $\partial^{GP} f_1(0) = \mathbb{R}_+$*,* $g_{1,1} = 1$ *and* $z_{1,1} = -v < 0$*, and then* $\partial^{GP} f_2(z_{k,1}) = \mathbb{R}_-$*,* $g_{1,2} = -1$ *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, 11, 22, 33, 34]. Below, we show the convergence of the incremental quasi-subgradient method to an optimal solution of problem (1.1) when the dynamic stepsize rule (S3) is adopted.

Theorem 3.3. *Suppose that 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 (3.3), we obtain that, for any $x^* \in X^*$ and $k \in \mathbb{N}$,

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 4\gamma_k(2-\gamma_k)C_{p,m}^2(f(x_k) - f^*)^{\frac{2}{p}} \\ &\leq \|x_k - x^*\|^2 - 4\underline{\gamma}(2-\overline{\gamma})C_{p,m}^2(f(x_k) - f^*)^{\frac{2}{p}}. \end{aligned}$$

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

$$\sum_{k=1}^{\infty}(f(x_k) - f^*)^{\frac{2}{p}} \leq \frac{1}{4\underline{\gamma}(2-\overline{\gamma})C_{p,m}^2}\|x_0 - x^*\|^2,$$

which is finite. Consequently, noting that $f(x_k) - f^* \geq 0$ for any $k \in \mathbb{N}$, one has $\lim_{k\to\infty} f(x_k) = f^*$. Hence, any cluster point of $\{x_k\}$ is an optimal solution of problem (1.1), denoted by $\bar{x} \in X^*$. Since further $\{\|x_k - x^*\|\}$ 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 quasi-subgradient method to the solution set X^* of problem (1.1) under the assumption that X^* has a nonempty interior.

Theorem 3.4. *Suppose that Assumptions 1 and 2 are satisfied, and let $\{x_k\}$ be a sequence generated by Algorithm 1. Suppose $x^* \in X^*$ and $\sigma > 0$ are such 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 hold:*

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

(ii) $\{v_k\}$ *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 $i = 1, \ldots, m$,

$$\hat{z}_{k,i} := P_X\big(\hat{z}_{k,i-1} - v_k\hat{g}_{k,i}\big), \text{ where } \hat{g}_{k,i} \in \begin{cases} \partial^{GP} f_i(\hat{z}_{k,i-1}) \cap S(0,1), & \text{if } f_i(\hat{z}_{k,i-1}) > f_i^*, \\ \{0\}, & \text{otherwise}, \end{cases} \tag{3.25}$$

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

$$I_k := \{i \in \{1, \ldots, m\} : f_i(\hat{z}_{k,i-1}) > f_i^*\}. \tag{3.26}$$

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. (3.26)). Then it follows from (2.1) that $\langle \hat{g}_{k,i}, x^* + \sigma \hat{g}_{k,i} - \hat{z}_{k,i-1} \rangle < 0$. Consequently,

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

(by (3.25), $\hat{g}_{k,i} = 0$ when $i \notin I_k$). Therefore, we obtain that

$$\sum_{i=1}^{m} \langle \hat{g}_{k,i}, \hat{z}_{k,i-1} - x^* \rangle > |I_k| \sigma \geq \sigma \quad \text{for any } k \in \mathbb{N}, \tag{3.27}$$

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

On the other hand, by (3.25), 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

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

consequently,

$$\frac{\sum_{k=0}^{n-1} \left(v_k \sum_{i=1}^{m} \langle \hat{g}_{k,i}, \hat{z}_{k,i-1} - x^* \rangle \right)}{\sum_{k=0}^{n-1} v_k} \leq \frac{\|x_0 - x^*\|^2}{2 \sum_{k=0}^{n-1} v_k} + \frac{m \sum_{k=0}^{n-1} v_k^2}{2 \sum_{k=0}^{n-1} v_k}. \tag{3.28}$$

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

$$\liminf_{n \to \infty} \sum_{i=1}^{m} \langle \hat{g}_{n,i}, \hat{z}_{n,i-1} - x^* \rangle < \sigma. \tag{3.29}$$

(i) When a constant stepsize $v \in (0, \frac{2\sigma}{m})$ is used, (3.28) 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 2.5, we obtain that

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

(ii) When a diminishing stepsize is used, by (3.17) and (3.18), it also follows from Lemma 2.5 and (3.28) 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 (3.29) under the assumption of (i) or (ii), which arrives at a contradiction with (3.27). The proof is complete. □

3.2 Randomized incremental quasi-subgradient method

It could be very computationally expensive to calculate the subgradients of all component functions at each iteration of the incremental quasi-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 quasi-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 quasi-subgradient method for solving problem (1.1) when using typical stepsize rules. The randomized incremental quasi-subgradient method is formally presented as follows.

Algorithm 2: Randomized incremental quasi-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 $g_{k,\omega_k} \in \partial^{GP} f_{\omega_k}(x_k) \cap S(0, 1)$, 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 quasi-subgradient method.

Theorem 3.5 ([9, p. 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) *$\mathbf{E}\{Y_{k+1} \mid \mathcal{F}_k\} \le Y_k - Z_k + W_k$;*

(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 with probability 1.

To begin the convergence analysis of Algorithm 2, we provide below a basic inequality of a randomized incremental quasi-subgradient iteration in terms of expectation. Recall that $R_{p,m}$ is defined in (3.6).

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 $k \in \mathbb{N}$, that*

$$\mathbf{E}\left\{\|x_{k+1} - x^*\|^2 \mid \mathcal{F}_k\right\} \le \|x_k - x^*\|^2 - 2v_k R_{p,m}(f(x_k) - f^*)^{\frac{1}{p}} + v_k^2. \tag{3.30}$$

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

$$\|x_{k+1} - x^*\|^2 \le \|x_k - v_k g_{k,\omega_k} - x^*\|^2 = \|x_k - x^*\|^2 - 2v_k \langle g_{k,\omega_k}, x_k - x^* \rangle + v_k^2. \tag{3.31}$$

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 2.3 is applicable (with f_{ω_k}, x_k, $X_{\omega_k}^*$ in place of h, x, X^*) to concluding that

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

Then (3.31) 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

$$\mathbf{E}\left\{ \|x_{k+1} - x^*\|^2 \mid \mathcal{F}_k \right\} \le \|x_k - x^*\|^2 - 2v_k L_{\max}^{-\frac{1}{p}} \mathbf{E}\left\{ \left(f_{\omega_k}(x_k) - f_{\omega_k}^* \right)^{\frac{1}{p}} \mid \mathcal{F}_k \right\} + v_k^2. \tag{3.32}$$

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

$$\begin{aligned} \mathbf{E}\left\{ \left(f_{\omega_k}(x_k) - f_{\omega_k}^* \right)^{\frac{1}{p}} \mid \mathcal{F}_k \right\} &= \tfrac{1}{|I_k|} \textstyle\sum_{i \in I_k} \left(f_i(x_k) - f_i^* \right)^{\frac{1}{p}} \\ &\ge m^{-\frac{1}{p}} \left(\textstyle\sum_{i \in I_k} (f_i(x_k) - f_i^*) \right)^{\frac{1}{p}}, \end{aligned} \tag{3.33}$$

where the inequality follows from Lemma 2.4 (with $f_i(x_k) - f_i^*$ and $\frac{1}{p}$ in place of a_i and γ) and $|I_k| \le m$. 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 $\sum_{i \in I_k} (f_i(x_k) - f_i^*) = \sum_{i=1}^m (f_i(x_k) - f_i(x^*)) = f(x_k) - f^*$. Therefore, (3.33) reduces to

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

which, together with (3.32) and (3.6), yields (3.30). The proof is complete. $\square$

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

Theorem 3.6. *Suppose that 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*

$$\liminf_{k \to \infty} f(x_k) \le f^* + \left(\frac{v}{2R_{p,m}} \right)^p. \tag{3.34}$$

Proof. Fix $\delta > 0$, and define a set $X_\delta \subseteq \mathbb{R}^n$ by

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

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 $\hat{g}_{k,\hat{\omega}_k} \in \partial^{GP} f_{\hat{\omega}_k}(\hat{x}_k) \cap S(0,1)$. 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 $f(\hat{x}_k) \geq f^* + \left(\frac{v}{2R_{p,m}} + \delta\right)^p$ and follows from Lemma 3.2 that the following relation holds for any $x^* \in X^*$ and $k \in \mathbb{N}$:

$$\mathbf{E}\left\{\|\hat{x}_{k+1} - x^*\|^2 \mid \hat{\mathcal{F}}_k\right\} \leq \|\hat{x}_k - x^*\|^2 - 2v\delta R_{p,m}.$$

Then, by Theorem 3.5, we obtain that $\sum_{k=0}^{\infty} 2v\delta R_{p,m} < \infty$ with probability 1, which is impossible. Hence, $\hat{x}_k \in X_\delta$ must occur for infinitely many times; consequently, in the original process, it holds with probability 1 that

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

Since $\delta > 0$ is arbitrary, (3.34) is obtained by letting δ tend 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 (3.14) and (3.34) that the randomized incremental quasi-subgradient method (Algorithm 2) admits a much less tolerance than that of the incremental quasi-subgradient method (Algorithm 1) when adopting the same stepsize. Indeed, by (3.4) and (3.6),

$$\frac{\left(\frac{v}{2R_{p,m}}\right)^p}{\left(\frac{mv}{4C_{p,m}}\right)^p} = \frac{2^{p-1}}{m^p} \ll 1.$$

The proof of the following theorem uses the property of the diminishing stepsize rule (cf. (3.2)) and a line of analysis similar to that of Theorem 3.6. Hence we omit the details.

Theorem 3.7. *Suppose that 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 that 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 (3.5), it follows that, for any $x^* \in X^*$ and any $k \in \mathbb{N}$,

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

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

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

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 $P\left(\Theta(z)^c\right) = 0$ for any $z \in Z \subseteq X^*$. By the elements of probability theory, one checks that

$$P(\Theta) = 1 - P\left(\Theta^c\right) = 1 - P\left(\bigcup_{z\in Z}\Theta(z)^c\right) \ge 1 - \sum_{z\in Z} P\left(\Theta(z)^c\right) = 1. \tag{3.35}$$

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 $\bar{x} : \Omega \to \mathbb{R}^n$ 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 $\lim_{k\to\infty} f\left(x_k(\omega)\right) = f^*$ for any $\omega \in \Theta$. Then it follows from the continuity of f that

$$\bar{x}(\omega) \in X^* \quad \text{for any } \omega \in \Theta. \tag{3.36}$$

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

$$\|\bar{x}(\omega) - z(\omega)\| \le \frac{\epsilon}{3}. \tag{3.37}$$

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 (3.37) 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 $\|x_k(\omega) - z(\omega)\| \le \frac{2\epsilon}{3}$ for any $k \ge N$. This, together with (3.37), yields

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

This shows that $\{x_k(\omega)\}$ converges to $\bar{x}(\omega)$ for any $\omega \in \Theta$. This, together with (3.36), 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 (3.35) 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 Applications

This section aims to present two important classes of applications of the sum-minimization problem (1.1) of a number of quasi-convex component functions: the quasi-convex feasibility problem and the sum of ratios problem.

4.1 Quasi-convex feasibility problem

The feasibility problem is at the core of the modeling of many problems in various areas of mathematics and physical sciences; see [5, 20] and references therein. In particular, the quasi-convex feasibility problem (QCFP) is an important class of feasibility problems (see [11, 18, 37]), which is to find a solution of the following system of inequalities:

$$x \in X, \quad \text{and} \quad h_i(x) \leq 0 \quad \text{for each } i = 1, \ldots, m, \tag{4.1}$$

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 (4.1) is nonempty, that is,

$$S := \{x \in X : h_i(x) \leq 0 \text{ for each } i = 1, \ldots, m\} \neq \emptyset.$$

The feasibility problem (4.1) can be cast into the framework of the sum-minimization problem (1.1) as the following model:

$$\min_{x \in X} f(x) := \sum_{i=1}^{m} f_i(x), \quad \text{where } f_i := \max\{h_i, 0\}. \tag{4.2}$$

For each $i = 1, \ldots, m$, we obtain by Lemma 2.2 that f_i is quasi-convex and $\partial^{GP} f_i(x) = \partial^{GP} h_i(x)$ if h_i is quasi-convex and $x \notin \mathrm{lev}_{\leq 0} h_i$. It is also clear that S is the optimal solution set of problem (4.2) if $S \neq \emptyset$. As a direct application of Algorithm 1 to problem (4.2), the incremental quasi-subgradient method for QCFP (4.1) is presented as follows.

Algorithm 3: Incremental quasi-subgradient method - QCFP.

Initialize an initial point $x_0 \in \mathbb{R}^n$, a stepsize sequence $\{v_k\}$, and let $k := 0$;
while $\max_{i=1,\ldots,m} h_i(x_k) > 0$ **do**
  Let $z_{k,0} := x_k$;
  for $i = 1, \ldots, m$ **do**
    if $h_i(z_{k,i-1}) \leq 0$ **then**
      Let $z_{k,i} := z_{k,i-1}$;
    else
      Calculate $g_{k,i} \in \partial^{GP} h_i(z_{k,i-1}) \cap S(0,1)$ and $z_{k,i} := P_X\left(z_{k,i-1} - v_k g_{k,i}\right)$;
    end
  end
  Let $x_{k+1} := z_{k,m}$ and $k := k+1$.
end

It is easy to see that Assumption 1 is satisfied for problem (4.2) if $S \neq \emptyset$; Assumption 2 is satisfied for problem (4.2) if each function h_i in (4.1) satisfies the Hölder condition of order p. Then, as direction applications of Theorems 3.1-3.3, the convergence results of Algorithm 3 when using different stepsize rules are obtained in the following theorem. Recall that $C_{p,m}$ and $R_{p,m}$ are defined by (3.4) and (3.6), respectively.

Theorem 4.1. *Suppose that $S \neq \emptyset$ and that h_i satisfies the Hölder condition of order p with modulus L_i on X for each $i = 1, \ldots, m$. Let $\{x_k\}$ be a sequence generated by Algorithm 3. Then the following assertions are true.*

(i) *If the constant stepsize rule* (S1) *is selected, then* $\liminf_{k\to\infty} \max_{i=1,\ldots,m} h_i(x_k) \leq \left(\frac{mv}{4C_{p,m}}\right)^p$.

(ii) *If the diminishing stepsize rule* (S2) *is selected, then*

- (ii-a) $\liminf_{k\to\infty} \max_{i=1,\ldots,m} h_i(x_k) = 0$*;*
- (ii-b) *If each h_i is coercive, then* $\lim_{k\to\infty} \max_{i=1,\ldots,m} h_i(x_k) = 0$ *and* $\lim_{k\to\infty} \operatorname{dist}(x_k, S) = 0$*;*
- (ii-c) *If $\sum_{k=0}^{\infty} v_k^2 < \infty$, then $\{x_k\}$ converges to a solution in S.*

(iii) *If the dynamic stepsize rule* (S3) *(with 0 in place of f^*) is selected, then $\{x_k\}$ converges to a solution in S.*

As a direct application of Algorithm 2 to problem (4.2), the randomized incremental quasi-subgradient method for QCFP (4.1) is presented as follows. Then, as direct applications of Theorems 3.6-3.8, the convergence results of Algorithm 4 when using different stepsize rules are obtained in the following theorem.

Algorithm 4: Randomized incremental quasi-subgradient method - QCFP.

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

Theorem 4.2. *Suppose that $S \neq \emptyset$ and that h_i satisfies the Hölder condition of order p with modulus L_i on X for each $i = 1, \ldots, m$. Let $\{x_k\}$ be a sequence generated by Algorithm 4. Then the following assertions are true.*

(i) *If the constant stepsize rule* (S1) *is selected, then* $\liminf_{k\to\infty} \max_{i=1,\ldots,m} h_i(x_k) \leq \left(\frac{v}{2R_{p,m}}\right)^p$ *with probability 1.*

(ii) *If the diminishing stepsize rule* (S2) *is selected, then* $\liminf_{k\to\infty} \max_{i=1,\ldots,m} h_i(x_k) = 0$ *with probability 1.*

(iii) *If the dynamic stepsize rule* (S4) *(with 0 in place of f^*) is selected, then $\{x_k\}$ converges to a solution in S with probability 1.*

4.2 Sum of ratios problem

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, 13, 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

$$\begin{array}{ll} \max & \sum_{i=1}^m R_i(x) := \frac{p_i(x)}{c_i(x)} \\ \text{s.t.} & x \in X, \end{array} \tag{4.3}$$

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 (4.3), especially for large-scale problems.

Exploiting the additivity structure of problem (4.3), we propose a new approach to find a global optimal solution of the SOR by virtue of sum-minimization formula. Indeed, 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 problem (4.3) 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 (4.3) can also be approached by solving the resulting QCFP (4.1). In [11], Censor and Segal proposed a subgradient projection method to solve the QCFP (4.1) by using the most violated control. In the numerical study, we apply the incremental quasi-subgradient methods and the subgradient projection method to solve the SOR (4.3) and its reformulated QCFP (4.1), respectively, and the abbreviations of these methods are listed in Table 1.

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

Abbreviations	Algorithms
SGPM	**Sub****G**radient **P**rojection **M**ethod in [11], which is to solve (4.1).
IncQSGM	**Inc**remental **Q**uasi-**S**ub**G**radient **M**ethod (Algorithm 1) for solving (4.3).
RandQSGM	**Rand**omized incremental **Q**uasi-**S**ub**G**radient **M**ethod (Algorithm 2) for solving (4.3).

In the numerical study, we consider the multiple Cobb-Douglas production efficiency problem (in short, MCDPE) [10], which is an application of the SOR. Formally, consider a set of m productions with s projects and n factors. Let $x := (x_j)^\top \in \mathbb{R}^n$ denote the amounts 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 (4.3). 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 in $[2, 5]$, while the constant stepsize is selected in $[1, 2]$. 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).

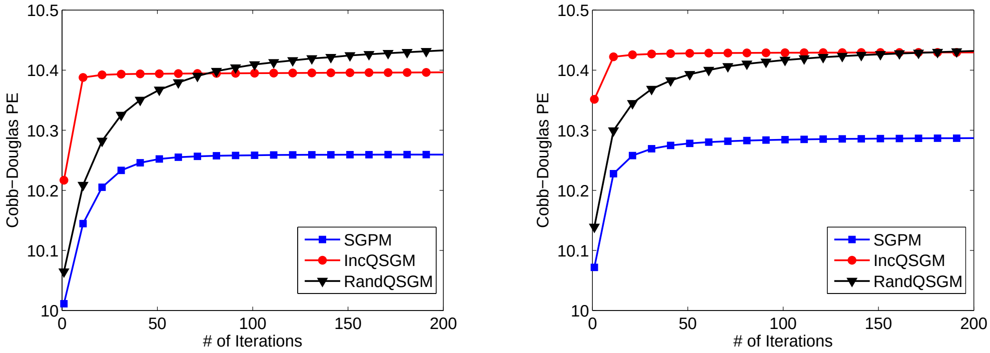
We first compare the performances (in terms of the obtained objective value and the CPU time) of the SGPM, IncQSGM and RandQSGM for different dimensions. The computation results are displayed in Table 2. In this table, the columns of Projects, Factors and Productions represent the numbers of projects (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 IncQSGM and RandQSGM outperform the SGPM in the sense that they achieve a larger production efficiency in a shorter time than the SGPM for different dimensional MCDPEs.

Table 2: Computation results for maximizing MCDPE.

Circumstance of problem			SGPM		IncQSGM		RandQSGM	
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

The second experiment is to compare the convergence behavior of the SGPM, IncQSGM and RandQSGM by using the constant and diminishing stepsize rules, where the problem 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 IncQSGM converges faster (in terms of the number of iterations) to an (approximate) optimal value that the RandQSGM and the SGPM. Furthermore, Figure 1(a) illustrates that the RandQSGM obtains a better estimation than the IncQSGM when the constant stepsize rule is adopted, which is consistent with Remark 3.5. Figure 1(b) demonstrates that both IncQSGM and RandQSGM 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 IncQSGM and RandQSGM 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 RandQSGM is faster (in terms of CPU time) than the IncQSGM, which is faster than the SGPM. This indicates the potential applicability of the RandQSGM to the large-scale SOR. Figure 3 plots the obtained maximal production efficiencies in each of these 500 trials. It is observed that the IncQSGM and RandQSGM outperform the SGPM consistently.

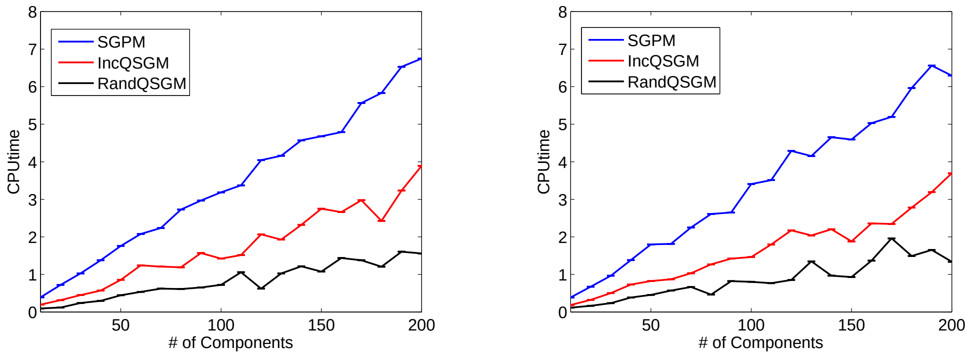
Finally, we conduct 500 simulations to show the stability of RandQSGM, which start from the same initial point, adopt the same stepsizes (constant: $v_k \equiv 1.5$ or diminishing: $v_k = 3/(1 + 0.1k)$) and solve a same MCDPE, but follow different stochastic processes. Figure 4 plots the error bars of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



(a) The constant stepsize rule.

(b) The diminishing stepsize rule.

Figure 1: The (averaged) convergence behavior of SGPM, IncQSGM and RandQSGM 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.

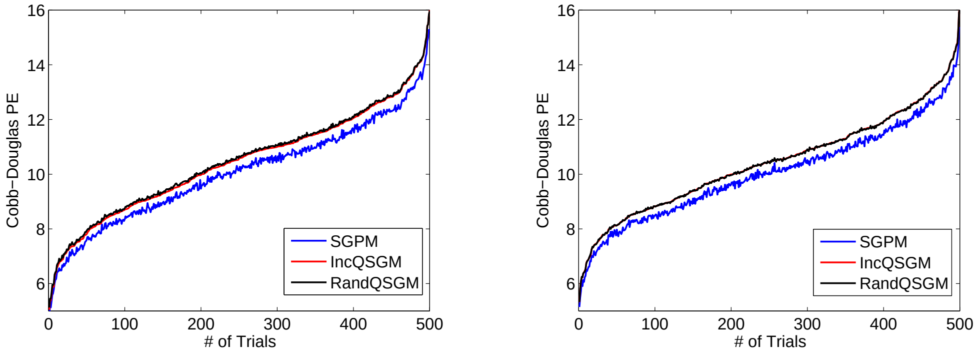
References

[1] Y. Almogy and O. Levin. Parametric analysis of a multi-stage stochastic shipping problem. *in Operational Research 69*, 359–370, 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.

[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.



(a) The constant stepsize rule.

(b) The diminishing stepsize rule.

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

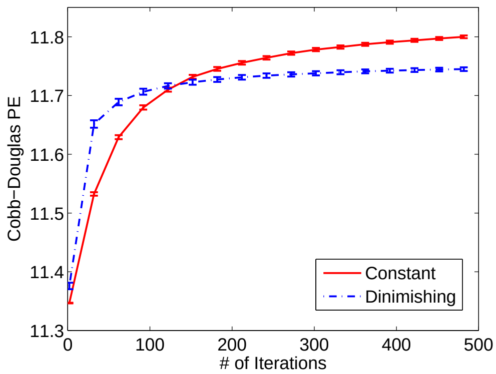


Figure 4: The error bars of RandQSGM in 500 simulations.

[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. *Convex Optimization Algorithms*. Athena Scientific, Massachusetts, 2015.

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

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

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

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

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

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

[14] 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.

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

[16] 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.

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

[18] 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.

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

[20] 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.

[21] 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.

[22] 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.

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

[47] C. K. W. Yu, Y. Hu, X. Yang and S. K. Choy. Abstract convergence theorem for quasi-convex optimization problems with applications. *Optimization*, DOI: 10.1080/02331934.2018.1455831, 2019.