Fixed Point Quasiconvex Subgradient Method

Kazuhiro Hishinuma[1a], Hideaki Iiduka[b]

[a]*Computer Science Course, Graduate School of Science and Technology, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571 Japan.*
[b]*Department of Computer Science, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571 Japan.*

Abstract

Constrained quasiconvex optimization problems appear in many fields, such as economics, engineering, and management science. In particular, fractional programming, which models ratio indicators such as the profit/cost ratio as fractional objective functions, is an important instance. Subgradient methods and their variants are useful ways for solving these problems efficiently. Many complicated constraint sets onto which it is hard to compute the metric projections in a realistic amount of time appear in these applications. This implies that the existing methods cannot be applied to quasiconvex optimization over a complicated set. Meanwhile, thanks to fixed point theory, we can construct a computable nonexpansive mapping whose fixed point set coincides with a complicated constraint set. This paper proposes an algorithm that uses a computable nonexpansive mapping for solving a constrained quasiconvex optimization problem. We provide convergence analyses for constant diminishing step-size rules. Numerical comparisons between the proposed algorithm and an existing algorithm show that the proposed algorithm runs stably and quickly even when the running time of the existing algorithm exceeds the time limit.

Keywords: Nonlinear programming, Fractional programming

1. Introduction

This paper considers the constrained quasiconvex optimization problem. This problem is composed of a quasiconvex objective functional and a closed convex constraint set. We call a functional of which any slice is convex a

---

[1]Research Fellow of the Japan Society for the Promotion of Science.

quasiconvex functional, and the class of this functional is a generalization of convex functionals. Quasiconvex functionals inherit some nice properties of convex functionals [13]. However, they do not have all the important properties of convex functionals, such as convexity of the sum of convex functionals, or give a guarantee of the coincidence of local optimality and global optimality. Therefore, the constrained quasiconvex optimization problem is difficult to solve in general.

Fractional programming is an important instance of constrained quasiconvex optimization problems. In economics, there are various situations in which one optimizes ratio indicators, such as the debt/equity ratio (in financial and corporate planning), inventory/sales and output/employee ratios (in production planning), and cost/patient and nurse/patient ratios (in health care and hospital planning) [35]. Under certain conditions, these ratio indicators, fractional objective functionals in other words, have quasiconvexity [22, Lemma 3]. Therefore, these problems can be dealt with as constrained quasiconvex optimizations. Here, we will examine the numerical behaviors of the existing and proposed algorithms when they are applied to the Cobb-Douglas production efficiency problem [3, Problem (3.13)], [11, Problem (6.1)], [35, Section 1.7], which is an instance of a fractional programming and constrained quasiconvex optimization problem. Furthermore, the demand for techniques to solve optimization problems is nowadays not only limited to convex objectives. In particular, optimization problems whose objective functionals are quasiconvex have appeared in economics, engineering, and management science [11, 13]. Therefore, this paper builds an algorithm that can efficiently solve constrained quasiconvex optimization problems even if they have some complexity.

Subgradient methods with the usual Fenchel subdifferential, an expansion of the gradient for nonsmooth functionals, are useful for solving problems in convex optimization [2, Section 8.2], [16, 17, 18, 19, 25]. We need to use an alternative notion of the usual Fenchel subdifferential since the usual Fenchel subdifferential is defined for a convex functional [4, Subsection 2.1], [22, Subsection 2.2], [33, Proposition 8.12]. Indeed, the usual Fenchel subdifferential may be empty even for a differential nonconvex functional [4, Subsection 2.1]. This implies that we cannot use it directly to solve quasiconvex optimization problems. Fortunately, various, extended subdifferentials for nonconvex functionals have been proposed [31, Section 4], [33, Definition 8.3]. As an instance of them, we can define subdifferentials for quasiconvex or more general functionals by the procedure described in [31, Section 4] to construct them with directional derivatives. These subdifferentials inherit some of the properties, called axioms for subdifferentials [31,

2

Axioms $(S_1–S_4)$], from the usual Fenchel subdifferential for convex functionals; however, they may not be easily computable. Furthermore, an essential issue is that a local minimizer might not coincide with the global minimizer in quasiconvex optimization. This issue reduces the subdifferential till it contains only one vector, i.e., the zero vector, meaning that the methods lose any clue as to the direction of the global minimizer. Hence, we cannot ensure the convergence of the generated sequence to the global minimizer of the quasiconvex optimization problem when the usual subgradient methods are used. For the unconstrained quasiconvex optimization problem, Konnov [23] introduced a subgradient method that uses a normalized normal vector to the slice at a current approximation as a subgradient. This idea overcomes the above issue, since there certainly exists a nonzero normal vector to the slice which indicates the direction to the minimizer even if the current approximation is a non-global local minimizer.

Kiwiel [22] proposed a subgradient method that uses a normalized normal vector to the slice as a subgradient (we will call it a subgradient throughout this paper) for solving the constrained quasiconvex optimization problem. Hu et al. [11] analyzed its convergence properties when inexact subgradients are used and/or when it includes computational errors. Furthermore, a number of subgradient-method variants exist for solving quasiconvex optimization problems, such as the conditional subgradient methods [13] and the stochastic subgradient method [12].

The existing methods assume the computability of the metric projection onto the constraint set, because they use the metric projection to guarantee that the solution is in the constraint set. The metric projection onto the constraint set is defined as a mapping which translates a given point into the nearest point inside the constraint set. Therefore, in general, we have to solve a subproblem of minimizing the distance from a given point subject to the solution being in the constraint set. Certainly, there are some sets onto which the metric projections can be computed easily, such as boxes [1, Proposition 29.15], closed balls [1, Example 3.18 and Proposition 3.19], [34, Section 4], and closed half-spaces [1, Example 29.20]. However, various complicated sets on which computing the metric projections is difficult appear in practical problems [6, 14, 19, 20, 37]. Therefore, we have to develop a new algorithm that can run lightly and quickly even when it is difficult to compute the metric projection onto the constraint set.

On the other hand, if the constraint set can be expressed as a fixed point set of or the intersection of some fixed point sets of nonexpansive mapping(s), there are algorithms that use these nonexpansive mappings instead of the metric projection for convex optimization [16, 17, 18, 19]. Fixed

point sets of nonexpansive mappings have great powers of expression. Any metric projection onto a closed convex set is also a nonexpansive mapping whose fixed point set coincides with these sets [1, Proposition 4.16]. We can build a nonexpansive mapping whose fixed point set coincides with the intersection of the fixed point sets of two or more given nonexpansive mappings [1, Proposition 4.9, 4.47]. Furthermore, there are complicated convex sets called generalized convex feasible sets that are defined by closed convex sets whose intersection may be empty. They can also be expressed using concrete nonexpansive mappings [19, Definition (8)], [37, Definition (50)]. The algorithms listed at the beginning of this paragraph use nonexpansive mappings instead of metric projections onto the constraint sets. Therefore, if these nonexpansive mappings can be more easily computed than the metric projections, it can also be expected that their algorithms will run more efficiently than algorithms which use metric projections directly.

The existing algorithms for solving convex optimization problems over fixed point sets of nonexpansive mappings are realized by combining a fixed point iterator, which generates a sequence converging to some fixed point of a given nonexpansive mapping, with the existing subgradient methods. The Krasnosel'skii-Mann iterator [24, 27] and Halpern iterator [9] are useful fixed point iterators for finding a fixed point of given nonexpansive mapping. Both generate a sequence converging to some fixed point of a given nonexpansive mapping.

In contrast to the existing literature, this paper proposes a novel algorithm which minimizes a given quasiconvex functional over the fixed point set of a given nonexpansive mapping. To realize this algorithm, we combine the Krasnosel'skii-Mann iterator [24, 27] with the existing subgradient method [22] for solving quasiconvex optimization problems. The goal of this paper is to show that our algorithm can solve constrained quasiconvex optimization problems whose constraint set is too complex for the existing algorithms to solve in a realistic amount of time.

This paper offers three contributions. The first is to provide a widely applicable algorithm for solving constrained quasiconvex optimization problems. The nonexpansive mappings are an extended notion of the metric projection, since the metric projection is also nonexpansive. Therefore, this paper allows more varied modeling for constrained quasiconvex optimization problems.

The second contribution is to present the theoretical convergence properties of our algorithm. We analyzed the convergence properties for constant and diminishing step-size rules. These results show by how much the error increases when a constant step-size rule is adopted what conditions are

4

required for the generated sequence to converge to the solution of the optimization problem.

The last contribution is to overcome the issue of the existing methods; that is, we show that the proposed algorithm can solve problems whose metric projections onto constraint sets cannot be easily computed. We conduct a numerical comparison of our algorithm and the existing algorithm. The results show that our algorithm can solve actual problems even when the constraint sets are too complex to find the metric projection onto them and when the existing algorithm cannot run in a realistic amount of time.

This paper is organized as follows. Section 2 gives the mathematical preliminaries. Section 3 defines our algorithm and presents its convergence analyses. Section 4 shows numerical comparisons between the proposed algorithm and the existing subgradient method, by solving a constrained quasiconvex optimization problem named the Cobb-Douglas production efficiency problem. Section 5 concludes this paper. The Appendices include miscellaneous propositions, lemmas, and their proofs.

2. Mathematical Preliminaries

First, we present the main problem considered in this paper, i.e., Problem 2.1, which is called a *constrained quasiconvex optimization problem*.

Problem 2.1. Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$, and let f be a continuous functional on H. In addition, suppose that the functional f has *quasiconvexity*, i.e. $f((1 - \alpha)x + \alpha y) \leq \max\{f(x), f(y)\}$ holds for any $x, y \in H$ and for any $\alpha \in [0, 1]$. Let X, D be nonempty closed convex subsets of H. Then, we would like to

$$\text{minimize } f(x) \text{ subject to } x \in X \cap D.$$

We define the *set of minima* and the *minimum value* of Problem 2.1 by $X^\star := \operatorname{argmin}_{x \in X \cap D} f(x)$ and $f_\star := \inf_{x \in X \cap D} f(x)$, respectively.

We use the following notation in this paper. $\mathbb{N}$ is the set of natural numbers without zero, and $\mathbb{R}$ is the set of real numbers. $\mathbf{B} := \{x \in H : \|x\| \leq 1\}$ is the unit ball in this Hilbert space, and $\mathbf{S} := \{x \in H : \|x\| = 1\}$ is the unit sphere in that space. Id is the identity mapping of H onto itself. The boundary of a set $C \subset H$ is denoted by $\operatorname{bd} C$, and the closure of this set is denoted by $\operatorname{cl} C$. The metric projection onto a closed, convex set $C \subset H$ is denoted by P_C and defined as $P_C(x) \in C$ and $\|x - P_C(x)\| = \inf_{y \in C} \|x - y\|$ for any $x \in H$. For any $\alpha \in \mathbb{R}$, the α-slice of a functional $f : H \to \mathbb{R}$ is

denoted by $\mathrm{lev}_{<\alpha} f := \{x \in H : f(x) < \alpha\}$. The fixed point set of a mapping $T : H \to H$ is denoted by $\mathrm{Fix}(T) := \{x \in H : T(x) = x\}$.

This paper makes full use of the nonexpansivity of some mapping for analyzing the convergence of the proposed algorithm. Hence, let us define two kinds of nonexpansive condition. A mapping $T : H \to H$ is said to be nonexpansive if $\|T(x) - T(y)\| \le \|x - y\|$ for any $x, y \in H$, and it is said to be firmly nonexpansive if $\|T(x) - T(y)\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \le \|x - y\|^2$ for any $x, y \in H$. Obviously, a firmly nonexpansive mapping is also a nonexpansive mapping [1, Subchapter 4.1]. The properties of these nonexpansivities are described in detail in [1, Chapter 4], [36, Chapter 6].

Useful algorithms for solving Problem 2.1 were proposed in [11, 13, 22, 23]. However, they assume that the metric projection onto the set X can be computed explicitly. Unfortunately, there are many complicated convex sets onto which constructing and/or computing the projection are difficult [6, 19, 20, 37]. This paper assumes a weaker and expanded condition for the constraint set X, only requiring the existence of a certain nonexpansive mapping expressing this set. Below, we list the conditions assumed throughout in this paper.

Assumption 2.1. We suppose that

(A1) the effective domain $\mathrm{dom}(f) := \{x \in H : f(x) < \infty\}$ coincides with the whole space H;

(A2) there exists some firmly nonexpansive mapping $T : H \to H$ whose fixed point set $\mathrm{Fix}(T)$ coincides with the constraint set X;

(A3) the constraint set $X = \mathrm{Fix}(T)$ and the feasible set $X \cap D$ are nonempty and there exists at least one minima, i.e. $X^\star \ne \emptyset$.

Assumptions (A2–3) mean that any closed convex sets which can be expressed as a fixed point set of some (firmly) nonexpansive mapping are accepted as constraint sets. Fixed point sets of nonexpansive mappings can express a variety of constraint sets, including not only the sets onto which the metric projections can be calculated such as is used in the existing literature [11, 13, 22], but also complicated sets onto which metric projections cannot be easily calculated [6, 19, 20, 37].

We can construct more complex sets by combining simpler nonexpansive mappings. The following proposition gives the fundamental, variously applicable transformations for building nonexpansive mappings.

Proposition 2.1. *Let $T_1, T_2, \ldots, T_N : H \to H$ be nonexpansive mappings (including metric projections onto some convex sets), and suppose that the intersection of these fixed point sets is nonempty. Let P_C be a metric projection onto a nonempty, closed, convex set $C \subset H$, and assume that $C \cap \mathrm{Fix}(T_1) \neq \emptyset$. Let $\alpha \in (0, 1/2]$. Then,*

(T1) *the mapping $\sum_{i=1}^{N} T_i / N$ is also a nonexpansive mapping, and its fixed point set coincides with $\bigcap_{i=1}^{N} \mathrm{Fix}(T_i)$ [1, Propositions 4.9 and 4.47];*

(T2) *the mapping $\alpha \, \mathrm{Id} + (1 - \alpha) T_1$ is a firmly nonexpansive mapping, and its fixed point set coincides with $\mathrm{Fix}(T_1)$ [1, Remark 4.37 and Proposition 4.47].*

The transformation (T1) ensures that we can make a nonexpansive mapping whose fixed point set coincides with the intersection of the fixed point sets of two or more nonexpansive mappings. The transformation (T2) provides us with a way to convert any nonexpansive mapping into a firmly nonexpansive mapping whose fixed point sets correspond with the given one. Our GitHub repository (URL: `https://github.com/iiduka-researches/201811-kaz`) provides these implementations of the transformations (T1 and T2) as higher-order functions `average` and `firm_up` in Python. By using our code, the reader can easily make a nonexpansive mapping expressing his or her desired constraint set.

Furthermore, let us examine an instance of convex sets that can be expressed as fixed point sets of some nonexpansive mappings, called the *generalized convex feasible sets* [19, Definition (10)], [37, Subsection 4.B]. Here, let us consider several closed convex sets $X_i \subset H$ for $i = 0, 1, \ldots, K$, and suppose that the metric projections $\{P_{X_i}\}_{i=0}^{K}$ onto these convex sets $\{X_i\}_{i=0}^{K}$ can be easily calculated. If the intersection of these sets is not empty, we can use the transformation and construction procedures described before to make a nonexpansive mapping whose fixed point set coincides with it. Hence, let us consider the opposite case; that is, there is a possibility that the intersection of the sets $\{X_i\}_{i=1}^{K}$ is empty. Then, we cannot use the straightforward way because the emptiness of the constraint set violates Assumption (A3). To design an alternative constraint set, let us define a functional [19, Definition (8)], [37, Definition (50)]

$$g(x) := \frac{1}{2} \sum_{i=1}^{K} w_i \left(\min_{y \in X_i} \|x - y\| \right)^2 \quad (x \in H), \tag{1}$$

where $\sum_{i=1}^{K} w_i = 1$. This functional g stands for the mean square value from the point x to the sets $\{X_i\}_{i=1}^{K}$ with respect to the weights $\{w_i\}_{i=1}^{K}$.

Therefore, we can consider the set of points which minimize this functional [19, Definition (10)], [37, Definition (50)],

$$X_g := \left\{ x \in X_0 : g(x) = \min_{y \in X_0} g(y) \right\},$$

as an alternative constraint set in terms of the mean square norm. This set is called the generalized convex feasible set. We can construct a nonexpansive mapping whose fixed point set coincides with this set, and thus, we can deal with the minimization problem over this constraint set by using the algorithm presented later. The way to construct this nonexpansive mapping is described in [19, Definition (9)], [37, Definition (50)].

Various subdifferentials have been proposed for quasiconvex functionals, such as the classical subdifferential [22, Definition (10)], the Greenberg-Pierskalla subdifferential [8, Section 3] and its variants such as the star subdifferential [4, Definition 7], Plastria's lower subdifferential [32, Section 2], and so on [4, Subsection 2.1], [22, Definition (6)–(9)], [31, Section 5]. The Greenberg-Pierskalla subdifferential is one of the most important concepts of subdifferentials for generalized convex functionals because it is a general notion that can be easily handled [31, Section 5]. However, it does not account for the norm of its subgradients and gives only directions. Plastria's lower subdifferential is proposed as another important concept whose properties are closer to those of the usual Fenchel subdifferential for convex functionals [31, Section 5]. In this paper, conforming to [11, 13, 23], we use the *subdifferential* defined as the normal cones to the slice of the functional f. That is, given a point $x \in H$, we call the set

$$\partial^\star f(x) := \{ g \in H : \langle g, y - x \rangle \leq 0 \ (y \in \operatorname{lev}_{<f(x)} f) \}$$

the subdifferential of the quasiconvex functional f at a point $x \in H$ [11, Definition 2.3], [13, Definition 2.1], [22, Definition (9)], [23, Section 1]. We also call its element a *subgradient.*

This subdifferential $\partial^\star f(\cdot)$ has some favorable properties, as listed in the following proposition.

Proposition 2.2 ([1, Proposition 6.2.4], [22, Lemma 3], [31, Propositions 6 and 8]). *Suppose that Assumption 2.1 holds, and assume that f is a continuous quasiconvex functional. Then, the following hold.*

(P1) $\partial^\star f(\cdot)$ *coincides with the closure of the Greenberg-Pierskalla subdifferential, i.e., the union of the Greenberg-Pierskalla subdifferential and the singleton set $\{0\}$.*

(P2) *The Greenberg-Pierskalla subdifferential, Plastria's lower subdifferential, and the usual Fenchel subdifferential are contained in the subdifferential $\partial^\star f(x)$ for any $x \in H$.*[2]

(P3) *For any $x \in H$, the subdifferential $\partial^\star f(x)$ is nonempty, and also contains some nonzero vector.*

(P4) $\partial^\star f(\cdot)$ *is a nonempty closed convex cone.*

First, we defined the subdifferential $\partial^\star f(\cdot)$ as the closure of the Greenberg-Pierskalla subdifferential, such as is shown in Proposition (P1). This implies that the subdifferential $\partial^\star f(\cdot)$ is an extension of the Greenberg-Pierskalla subdifferential, and some properties of this subdifferential can also be used. For example, $\partial^\star f(x)$ coincides with the whole space H if x is a minimizer of f. This proposition ensures that the subdifferential $\partial^\star f(\cdot)$ coincides with the closure of the Greenberg-Pierskalla subdifferential, which is not always closed [11, Subsection 2.1]. Hence, this subdifferential $\partial^\star f(\cdot)$ overcomes the problem of the non-closedness of the Greenberg-Pierskalla subdifferential; it has been used in the recent literature [11, 12, 13]. Furthermore, as shown in Proposition (P2), the subdifferential is also a superset of Plastria's lower subdifferential and the usual Fenchel subdifferential. Since the subdifferential $\partial^\star f(\cdot)$ is a cone as shown in Proposition (P4), this property ensures that every arbitrarily scaled element of the Plastria's lower subdifferential or the usual Fenchel subdifferential can be used as a subgradient in the discussion of this paper. Proposition (P3) ensures the existence of nonzero subgradients at all points. This fact guarantees that the algorithm described later can always find a subgradient, which is required for the computation. In addition, Proposition (P4) shows that the normalized vector of a subgradient is also a subgradient. Our algorithm implicitly uses this property for choosing a subgradient whose norm is 1.

A subgradient in $\partial^\star f(\cdot)$ is computable when, for example, the functional is formed as a fractional function, a typical instance of a quasiconvex function, with concrete conditions. The following proposition gives the conditions for the quasiconvexity and subgradient computability of fractional functions.

---

[2] Furthermore, the other four kinds of subdifferential presented in [31, Section 5] are also contained in the subdifferential $\partial^\star f(x)$. Please refer to [31, Proposition 6] for more details.

Proposition 2.3 ([22, Lemmas 3 (i) and 4]). *Let a be a convex functional on H, and let b be a finite, positive functional on H. Suppose that $f(x) := a(x)/b(x)$ for any $x \in H$, the interior of $\mathrm{dom}(f)$ is convex, and one of the following conditions holds:*

(i) b is affine;

(ii) a is nonnegative on the interior of $\mathrm{dom}(f)$ and b is concave;

(iii) a is nonpositive on the interior of $\mathrm{dom}(f)$ and b is convex.

Then, the functional f is a quasiconvex functional on the interior of $\mathrm{dom}(f)$. Furthermore, the functional $(a - \alpha b)(\cdot)$ is convex and $\partial(a - \alpha b)(x) \subset \partial^\star f(x)$ for any $x \in H$, where $\alpha := f(x)$ and $\partial(a - \alpha b)$ is the usual Fenchel subdifferential of the functional $(a - \alpha b)(\cdot)$.

The following defines a property named the Hölder condition of a functional. This property is used in turn to describe some of the properties of the subgradient and to establish the convergence of subgradient methods [13, Section 2].

Definition 2.1 (Hölder condition [23, Definition 1]). A functional $f : H \to \mathbb{R}$ is said to satisfy the *Hölder condition* with degree $\beta > 0$ at a point $x \in H$ on a set $M \subset H$ if there exists a number $L \in \mathbb{R}$ such that

$$|f(z) - f(x)| \leq L \, \|z - x\|^\beta \quad (z \in M).$$

The Hölder condition with degree 1 is equivalent to Lipschitz continuity. Furthermore, when f is a convex functional, it is also equivalent to the bounded subgradient assumption frequently assumed in convergence analyses of subgradient methods for solving convex optimization problems [13, Section 2]. For more details on this property, see Example 3.1 described later.

The following Proposition 2.4 is a key lemma which relates the distance to the set of minima to its functional value. While nearly the same assertion in Euclidian spaces is presented in [23, Proposition 2.1], this proposition extends it to Hilbert spaces. We should remark that the condition for a point x is slightly modified from the original one for the later discussion. Nevertheless, we can similarly prove this proposition.

Proposition 2.4 ([23, Proposition 2.1]). *Suppose that the functional f satisfies the Hölder condition with degree $\beta > 0$ at a point $x^\star \in X^\star$ on the set $\mathrm{cl}(\mathrm{lev}_{<f(x)} f)$ for some point $x \in H$ such that $f_\star < f(x)$. Then, we have*

$$f(x) - f_\star \leq L \, \langle g, x - x^\star \rangle^\beta \quad (g \in \partial^\star f(x) \cap \mathbf{S}).$$

10

The following propositions are used to prove the theorems presented later.

Proposition 2.5 ([30, Lemma 1])**.** *Let $\{x_k\}$ be a sequence in the Hilbert space H and suppose that it converges weakly to x. Then for any $y \neq x$, $\liminf_{k\to\infty}\|x_k - x\| < \liminf_{k\to\infty}\|x_k - y\|$.*

Proposition 2.6 ([1, Proposition 10.25])**.** *Every quasiconvex continuous functional on a real Hilbert space H has weakly lower semicontinuity. That is to say, we have $f(x) \leq \liminf_{n\to\infty} f(x_n)$ for any sequence $\{x_n\} \subset H$ which converges weakly to a point $x \in H$ if the functional f is a quasiconvex continuous functional on a real Hilbert space H.*

Proposition 2.7 ([1, Proposition 11.8])**.** *Let C be a convex subset of H, and let f be a strictly quasiconvex functional on the Hilbert space H, i.e., $f(\alpha x + (1 - \alpha)y) < \max\{f(x), f(y)\}$ for any $\alpha \in (0,1)$ and for any two distinct points $x, y \in H$. Then, f has at most one minimizer over C.*

3. Quasiconvex Subgradient Method over a Fixed Point Set

We propose the following Algorithm 1 for solving Problem 2.1 with Assumption 2.1. This algorithm iteratively generates the next point x_{k+1} from

---
Algorithm 1 Fixed Point Quasiconvex Subgradient Method for Solving Problem 2.1

Require:
   $f\colon H \to \mathbb{R}$, $T\colon H \to H$, $D \subset H$;
   $\{v_k\} \subset (0, \infty)$, $\{\alpha_k\} \subset (0, 1]$.
Ensure:
   $\{x_k\} \subset D$.
1: $x_1 \in D$.
2: **for** $k = 1, 2, \ldots$ **do**
3:    $g_k \in \partial^\star f(x_k) \cap \mathbf{S}$.
4:    $x_{k+1} := P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k))$.
5: **end for**

---

the current approximation x_k in order to improve it. Specifically, step 3 of this algorithm finds a regularized subgradient of the functional f at the current approximation x_k. Step 4 is composed of two improving iterators: one is the subgradient method iterator $x_k - v_k g_k$ to improve approximations with respect to the functional value, and the other is the Krasnosel'skiĭ-Mann iterator [24, 27] $\alpha_k \operatorname{Id} + (1 - \alpha_k)T$ to improve approximations with respect to

the distance to the fixed point set $\mathrm{Fix}(T)$. To ensure that the generated sequence is contained in the set D, we project each generated point onto this set (this operation is optional because the metric projection operator P_D coincides with the identity mapping Id if the set D is the whole space H). By repeating steps 3–4, this algorithm generates a sequence converging to a point in the solution set $X^\star$.

Before moving on to the convergence analyses, we will give the assumptions and lemmas describing the fundamental properties of Algorithm 1.

Assumption 3.1. (A4) For any $k \in \mathbb{N}$ such that $f_\star < f(x_k)$ and for all $x^\star \in X^\star$, the functional f satisfies the Hölder condition with degree $\beta > 0$ at the point $x^\star$ on the set $\mathrm{cl}(\mathrm{lev}_{<f(x_k)} f)$.

(A5) The generated sequence $\{x_k\}$ is bounded.

(A6) The real sequence $\{\alpha_k\} \subset (0,1]$ satisfies $0 < \liminf_{k\to\infty} \alpha_k \le \limsup_{k\to\infty} \alpha_k < 1$.

In the following, we have to ensure that Assumption (A5), i.e., the boundedness of the sequence generated by Algorithm 1, holds. If we know the estimated range of the solution candidates, the simplest way to bound the generated sequence is to let the set D be a ball with a large enough diameter. We can compute the metric projection onto a closed ball easily [1, Example 3.18]. For example, giving $10^{16}\mathbf{B}$ as the set D to Algorithm 1 satisfies Assumption (A5). Even when the boundedness of the set D cannot be guaranteed, we can ensure the boundedness of the generated sequence if the objective functional is coercive. A detailed discussion and proof of this fact is given in Appendix A.

Here let us give some examples which satisfy Assumption 3.1. The first example shows the applicability of Algorithm 1 to constrained convex optimization problems.

Example 3.1 ([1, Remark 4.34.(iii), Propositions 4.47 and 16.20]). *Suppose that f is a continuous, convex functional on H, $\tilde{T}$ is a nonexpansive mapping of H into itself, and D is a closed, bounded, convex subset of H. Set $T := (\mathrm{Id} + \tilde{T})/2$. Furthermore, assume that the feasible set $\mathrm{Fix}(T) \cap D$ is nonempty. Set $\alpha_k := 1/2$ for every $k \in \mathbb{N}$. If H is finite-dimensional, or if the usual Fenchel subdifferential of f maps every bounded subset of H to a bounded set, then T is a firmly nonexpansive mapping, $\mathrm{Fix}(T)$ coincides with the fixed point set $\mathrm{Fix}(T)$, and Assumption 3.1 holds.*

This example shows that, if the set D is bounded, our algorithm can be applied to nonsmooth convex optimization problems over fixed point sets of

nonexpansive mappings [16, 17, 18, 19]. The following discussion is predicated upon Assumption (A5), i.e., that the generated sequence is bounded, since it is required for evaluating the distance between the generated sequence and the fixed point set (Lemma 3.2). Indeed, the existing analysis of the fixed point subgradient method for convex optimization assumes the generated sequence is bounded [15, Assumption (A2)], [16, Assumption 3.1]. However, we can guarantee the boundedness. As we described above, the simplest D that satisfies the requirements is a ball with a large enough diameter.

The transformation $\tilde{T} \mapsto (\mathrm{Id} + \tilde{T})/2$ converts the given nonexpansive mapping $\tilde{T}$ into a corresponding firmly nonexpansive mapping T whose fixed point set coincides with the given one [1, Remark 4.37 and Proposition 4.47]. This implies that any nonexpansive mapping whose fixed point set coincides with the constraint set can be used as $\tilde{T}$. Of course, since any metric projection operator is a firmly nonexpansive mapping [1, Proposition 4.16], we can solve an optimization problem over the constraint set onto which the metric projection can be computed. In Section 4, we will describe a concrete example of constructing a firmly nonexpansive mapping. Furthermore, our algorithm extends the existing subgradient methods for convex optimization, since any of the usual Fenchel subgradients of a convex functional is also a subgradient as defined in this paper. Hence, the convergence analyses of our algorithm (described later) will be very useful for not only quasiconvex optimization [11, 13, 22, 23] but also convex optimization [16, 17, 18, 19].

Let us consider the simplest example of a (nonconvex) quasiconvex objective functional. The next example shows that a typical quasiconvex functional, called the capped-l_1 norm, appearing in sparse regularization of machine learning tasks [5, Equation (25)], [38, Appendix C.3.1] can be minimized using Algorithm 1.

Example 3.2. *Let $f(x) := \min\{\|x\|, \alpha\}$ for some $\alpha > 0$, and let $T := \mathrm{Id}$. Set $\{v_k\} \subset (0, \alpha]$ and $\alpha_k := 1/2$ for all $k \in \mathbb{N}$. Use $g_k := x_k / \|x_k\| \in \partial^\star f(x_k) \cap \mathbf{S}$ for each $k \in \mathbb{N}$ until x_k reaches the solution. Then, this setting satisfies Assumption 3.1.*

Here we should remark that Assumption 3.1 does not guarantee that the sequence generated by Algorithm 1 converges to some optimum. For example, let us consider the case where $f(x) := \min\{|x|, 1\}$ for any real $x \in \mathbb{R}$ and the initial point $x_1 := 3/2$. Even though it violates the assumption of Example 3.2, let us assume $v_k := 2$ for all $k \in \mathbb{N}$. Then, this setting still satisfies Assumption 3.1. However, as illustrated in Figure 1, we can see that the generated sequence does not converge to an optimum. In each
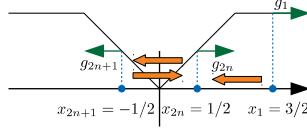
13

Figure 1: Illustration of case where the generated sequence repeats $1/2$ and $-1/2$ and does not converge to the origin.

step, the approximation is moved in the direction of the origin by 1. In this counterexample, the first step moves the initial point $3/2$ to the point $1/2$. After that, the algorithm eternally iterates so as to move approximations to their symmetric point with respect to the origin. Therefore, the generated sequence repeats $1/2$ and $-1/2$ and does not converge to the optimum 0. The convergence theorems presented later describes what is required to make the generated sequence converge to the optimum and/or how much error can occur in the solution.

Finally, we present a concrete application that can be dealt with as a quasiconvex optimization problem. The following fractional programming problem is called the Cobb-Douglas production efficiency problem and satisfies Assumption 3.1.

Example 3.3 ([3, Problem (3.13)], [11, Problem (6.1))], [35, Section 1.7]). *Let us consider the problem in an Euclidean space; i.e., suppose that $H :=$ $\mathbb{R}^n$. Set $D := [0, M]^n$ for some $M > 0$, and set $\alpha_k := 1/2$ for any $k \in \mathbb{N}$. We give two positive scalars $a_0, c_0 > 0$ and two n-dimensional positive vectors in advance, $a, c \in (0, \infty)^n$ such that $\sum_{i=1}^n a_i = 1$. Let*

$$f(x) := \begin{cases} \frac{-a_0 \prod_{j=1}^n x_j^{a_j}}{\langle c, x \rangle + c_0} & (x \in [0, \infty)^n), \\ 0 & (\text{otherwise}), \end{cases}$$

and assume that T is a nonexpansive mapping from $\mathbb{R}^n$ to itself and $\mathrm{Fix}(T) \cap D \neq \emptyset$ holds. Run Algorithm 1 with an initial point $x_1 \in [0, M]^n$. Then, Assumption 3.1 holds.

In Section 4, we will define the Cobb-Douglas production efficiency problem and show the numerical behavior of Algorithm 1 when it solves a concrete instance of this problem. Hence, we will put off explaining the problem in detail till later and limit ourselves here to a brief description.

The Cobb-Douglas production efficiency problem was introduced by Bradley and Frey [3]. Hu, Yang, and Sim proposed an algorithm for solving this problem by regarding it as a constrained quasiconvex optimization problem [11]. The fact that the objective functional f is quasiconvex allows us to treat it as a quasiconvex optimization problem. The study by Hu et al. [11] is also based on this fact; it ensures the quasiconvexity of the objective functional. However, the existing results including that of [11] assume that the projection onto the constraint set is easily computable.

Let us consider the meaning of this objective function. The numerator expresses the total profit defined by the production factors x_j for each $j = 1, 2, \ldots, n$. This numerator is modeled with the Cobb-Douglas production function. In this problem, we consider the total cost for the production activities to be an affine function with respect to the production factors $\{x_j\}_{j=1}^n$. This cost function is set as the denominator of the objective function. Hence, the objective function f represents the ratio of the total profit and the total cost. In addition, it is known that the numerator, i.e., the Cobb-Douglas production function, is convex [26, Section 2], and therefore, Proposition 2.3 guarantees that f is quasiconvex.

The sequence generated by Algorithm 1 must be contained in the box $[0, M]^n$, because its complement includes points that make the denominator of the function f zero. Therefore, we set the domain to $D = [0, M]^n \subset [0, \infty)^n$. However, from the definition of f, setting $f(x) := 0$ when x is out of the set $[0, \infty)^n$ makes it possible to expand its domain to the whole space while maintaining continuity.

The following lemmas show the fundamental properties of Algorithm 1.

Lemma 3.1. *Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1. Suppose that Assumptions 2.1 and (A4) hold. Then, for any $k \in \mathbb{N}$ that satisfies $f_\star < f(x_k)$, the following inequality holds.*

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - 2v_k(1 - \alpha_k)\left(\frac{f(x_k) - f_\star}{L}\right)^{\frac{1}{\beta}} + (1 - \alpha_k)v_k^2.$$

Similar lemmas are presented in the literature that discuss algorithms which use the metric projection onto the constraint set [12, Lemma 3.1], [13, Lemma 3.2], [22, Lemma 6]. This implies that the generated sequence is guaranteed to be contained in the constraint set. Therefore, the proofs of these lemmas use a property which ensures $x_k \in X \cap D$ and $f_\star \leq f(x_k)$. However, the algorithm presented here generates a sequence that may not be in the fixed point set of T, in other words, it may be out of the feasible set. The convergence analyses are carefully divided into cases where $f_\star < f(x_k)$

holds and cases where it does not hold. The above lemma describes that the existing results hold for the proposed algorithm only if the positive case $f_\star < f(x_k)$ holds and even if the containedness of the generated sequence in the fixed point set cannot be guaranteed.

Proof. See Appendix C for the proof of this lemma. ◻

Lemma 3.2. *Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1. Suppose that Assumptions 2.1 and (A5) hold and the real sequence $\{v_k\}$ is bounded. Then, for each $x \in \mathrm{Fix}(T) \cap D$, there exists $M_1 \geq 0$ such that*

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - (1 - \alpha_k)\|x_k - T(x_k - v_k g_k)\|^2 + v_k M_1 \quad (k \in \mathbb{N}).$$

Proof. See Appendix C for the proof of this lemma. ◻

3.1. Constant step-size rule

The following theorem shows how precise the generated solution is when the constant step-size rule is used.

Theorem 3.1. *Let $v > 0$ and $v_k := v$ for all $k \in \mathbb{N}$ and $\{x_k\} \subset H$ be a sequence generated by Algorithm 1. Suppose that Assumptions 2.1 and 3.1 hold. Then, the sequence $\{x_k\}$ satisfies*

$$\liminf_{k \to \infty} f(x_k) \leq f_\star + L\left(\frac{v}{2}\right)^\beta, \ \text{and} \ \liminf_{k \to \infty} \|x_k - T(x_k)\|^2 \leq Mv$$

for some $M \geq 0$.

Proof. See Appendix D for the proof of this theorem. ◻

3.2. Diminishing step-size rule

Finally, we prove the weak convergence theorem of Algorithm 1. To let the generated sequence weakly converge to some optimum, we can use a specific step-size called a diminishing step-size. The following theorem describes this condition and shows that the generated sequence weakly converges.

Theorem 3.2. *Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1. Suppose that*

(i) *Assumptions 2.1 and 3.1 hold,*

(ii) *and the real sequence $\{v_k\} \subset (0, \infty)$ satisfies*

$$\lim_{k \to \infty} v_k = 0, \ \text{and} \ \sum_{k=1}^{\infty} v_k = \infty.$$

16

Then, there exists a subsequence of the generated sequence $\{x_k\}$ which converges weakly to a point in $X^\star$. In addition, if

(iii) the whole space H is an N-dimensional Euclidean space $\mathbb{R}^N$,

(iv) and the solution $x^\star \in X^\star$ is unique,

then, the whole sequence $\{x_k\}$ converges to this unique solution $x^\star$.

Assumption (A3) and Proposition 2.7 show that the strict quasiconvexity of the objective function is a sufficient condition for the uniqueness of the solution $x^\star \in X^\star$.

Before proving the above theorem, we prove the following lemma which will be needed later.

Lemma 3.3. *Suppose that Assumptions 2.1 and 3.1 hold, and suppose that the real sequence $\{v_k\} \subset (0, \infty)$ satisfies*

$$\lim_{k \to \infty} v_k = 0, \ and \ \sum_{k=1}^{\infty} v_k = \infty.$$

Let $\{x_k\} \subset H$ be the sequence generated by Algorithm 1 with this real sequence $\{v_k\}$. Then,

$$\liminf_{k \to \infty} f(x_k) \leq f_\star$$

holds.

Proof. See Appendix E for the proof of this theorem. □

Now let us prove Theorem 3.2 with the above result.

Proof of Theorem 3.2. Let the limit superior of the real sequence $\{\alpha_k\}$ be denoted by $\bar{\alpha} \in (0, 1)$. Fix $x^\star \in X^\star$ arbitrarily. We will prove the assertion by separating the problem into two cases: the case where there exists a number $k_0 \in \mathbb{N}$ such that $\|x_{k+1} - x^\star\| \leq \|x_k - x^\star\|$ for all $k \geq k_0$, and its negation.

First, let us consider the positive case; i.e., there exists a number $k_0 \in \mathbb{N}$ such that $\|x_{k+1} - x^\star\| \leq \|x_k - x^\star\|$ for all $k \geq k_0$. The property of the limit superior guarantees the existence of a number $k_1 \geq k_0$ such that $\alpha_k < \bar{\alpha} + (1 - \bar{\alpha})/2$ for all $k \geq k_1$. Therefore, applying this relationship between

$\bar{\alpha}$ and α_k to Lemma 3.2, an estimate of $\|x_k - T(x_k - v_k g_k)\|$ for any $k \geq k_1$ can be obtained as follows:

$$\frac{1}{2}(1 - \bar{\alpha}) \|x_k - T(x_k - v_k g_k)\|^2 \leq (1 - \alpha_k) \|x_k - T(x_k - v_k g_k)\|^2$$
$$\leq \|x_k - x^\star\|^2 - \|x_{k+1} - x^\star\|^2 + v_k M_1. \tag{2}$$

In this case, the monotonicity and boundedness of the subsequence $\{\|x_k - x^\star\|\}_{k \geq k_1}$ are assured. Hence, this subsequence converges to some nonnegative real. Since we have assumed that the real sequence $\{v_k\}$ converges to zero, the left-hand side of inequality (2) converges to zero. On the other hand, $\|x_k - T(x_k)\|$ for each $k \in \mathbb{N}$ can be expanded with the triangle inequality and by noting the nonexpansivity of T as follows:

$$\|x_k - T(x_k)\| \leq \|x_k - T(x_k - v_k g_k)\| + \|T(x_k - v_k g_k) - T(x_k)\|$$
$$\leq \|x_k - T(x_k - v_k g_k)\| + v_k.$$

From the assumption of this theorem and the previous discussion, both terms on the right-hand side above converge to zero with respect to k. Hence, we find that the real sequence $\{\|x_k - T(x_k)\|\}$ converges to zero.

The property of the limit inferior of the real sequence $\{f(x_k)\}$ guarantees the existence of a subsequence $\{f(x_{k_i})\}$ converging to $\liminf_{k \to \infty} f(x_k)$. Note that Lemma 3.3 asserts that this limit inferior is less than or equal to the minimum value $f_\star$. There exist a point $u \in H$ and a subsequence $\{x_{k_{i_j}}\} \subset \{x_{k_i}\}$ converging weakly to the point u, since $\{x_{k_i}\}$ is a bounded sequence in H. Now, suppose that u is not a fixed point of T. Since the real sequence $\{\|x_k - T(x_k)\|\}$ converges to zero and T is a nonexpansive mapping, Proposition 2.5 produces a contradiction as follows:

$$\liminf_{j \to \infty} \left\|x_{k_{i_j}} - u\right\| < \liminf_{j \to \infty} \left\|x_{k_{i_j}} - T(u)\right\|$$
$$\leq \liminf_{j \to \infty} \left(\left\|x_{k_{i_j}} - T\left(x_{k_{i_j}}\right)\right\| + \left\|T(x_{k_{i_j}}) - T(u)\right\|\right)$$
$$\leq \liminf_{j \to \infty} \left\|x_{k_{i_j}} - u\right\|.$$

Therefore, we can see that u is a fixed point of T. Proposition 2.6 means that the objective functional f has weakly lower semicontinuity. The weak convergence of the sequence $\{x_{k_{i_j}}\}$ implies

$$f(u) \leq \liminf_{j \to \infty} f(x_{k_{i_j}}) = \lim_{i \to \infty} f(x_{k_i}) \leq f_\star;$$

that is, u is an optimum.

To deal with the positive case, let us consider the weak convergence of $\{x_k\}$ and its subsequences. Take another subsequence $\{x_{k_{i_l}}\} \subset \{x_{k_i}\}$ that converges weakly to a point $v \in H$. A similar discussion to the one for obtaining $u \in X^\star$ ensures that the point v is also an optimum. To show the uniqueness of the weak accumulation points of the sequence $\{x_{k_i}\}$, let us assume that $u \neq v$. Since the sequence $\{\|x_k - x^\star\|\}$ converges, Proposition 2.5 leads us to a contradiction:

$$
\begin{aligned}
\lim_{k \to \infty} \|x_k - u\| = \lim_{j \to \infty} \left\| x_{k_{i_j}} - u \right\| &< \lim_{j \to \infty} \left\| x_{k_{i_j}} - v \right\| \\
&= \lim_{j \to \infty} \|x_k - v\| = \lim_{l \to \infty} \left\| x_{k_{i_l}} - v \right\| < \lim_{l \to \infty} \left\| x_{k_{i_l}} - u \right\| \\
&= \lim_{k \to \infty} \|x_k - u\|.
\end{aligned}
$$

Hence, we can see that u is the same point as v, and the uniqueness of all weak accumulation points of the sequence $\{x_{k_i}\}$ is proven. This uniqueness implies that the sequence $\{x_{k_i}\}$ converges weakly to $u \in X^\star$. Now take another subsequence $\{x_{k_m}\} \subset \{x_k\}$ that converges weakly to a point $w \in H$, and suppose that u is a different point from v. From the fact that the sequence $\{\|x_k - x^\star\|\}$ converges and from Proposition 2.5, we can deduce that

$$
\begin{aligned}
\lim_{k \to \infty} \|x_k - u\| = \lim_{i \to \infty} \|x_{k_i} - u\| &< \lim_{i \to \infty} \|x_{k_i} - w\| \\
&= \lim_{j \to \infty} \|x_k - w\| = \lim_{m \to \infty} \|x_{k_m} - w\| < \lim_{m \to \infty} \|x_{k_m} - u\| \\
&= \lim_{k \to \infty} \|x_k - u\|.
\end{aligned}
$$

However, this is a contradiction. Hence, the sequence $\{x_k\}$ converges weakly to some optimum. This proves the positive case.

Next, let us consider the negative case, in other words, the case where a subsequence $\{x_{k_i}\} \subset \{x_k\}$ exists that satisfies $\|x_{k_i} - x^\star\| < \|x_{k_i+1} - x^\star\|$ for all $i \in \mathbb{N}$. A similar discussion to the one for finding the number k_1 in the positive case guarantees the existence of $i_0 \in \mathbb{N}$ satisfying $\alpha_{k_i} < \bar{\alpha} + (1 - \bar{\alpha})/2$ for all $i \geq i_0$. The distances from the point $x^\star$ to each point x_{k_i+1} where $i \geq i_0$ from Lemma 3.2 are as follows:

$$
\begin{aligned}
\|x_{k_i+1} - x\|^2 &\leq \|x_{k_i} - x\|^2 - (1 - \alpha_{k_i}) \|x_{k_i} - T(x_{k_i} - v_{k_i} g_{k_i})\|^2 + v_{k_i} M_1 \\
&\leq \|x_{k_i} - x\|^2 - \frac{1}{2}(1 - \bar{\alpha}) \|x_{k_i} - T(x_{k_i} - v_{k_i} g_{k_i})\|^2 + v_{k_i} M_1.
\end{aligned}
$$

19

Here, we have assumed that $\|x_{k_i+1} - x^\star\|$ is greater than $\|x_{k_i} - x^\star\|$ for all $i \in \mathbb{N}$ and the real sequence $\{v_k\}$ converges to zero. Thus, the above inequality implies that the real sequence $\{\|x_{k_i} - T(x_{k_i} - v_{k_i} g_{k_i})\|\}$ converges to zero with respect to i. On the other hand, the distances $\|x_{k_i} - T(x_{k_i})\|$ for each $i \in \mathbb{N}$ can be estimated from the nonexpansivity of T as follows:

$$\|x_{k_i} - T(x_{k_i})\| \leq \|x_{k_i} - T(x_{k_i} - v_{k_i} g_{k_i})\| + \|T(x_{k_i} - v_{k_i} g_{k_i}) - T(x_{k_i})\|$$
$$\leq \|x_{k_i} - T(x_{k_i} - v_{k_i} g_{k_i})\| + v_{k_i}.$$

Since both terms of the right-hand side of the above inequality converge to zero, the left-hand side also converges to zero.

We will proceed by way of contradiction; suppose that $\limsup_{i\to\infty} f(x_{k_i}) > f_\star$. This implies the existence of $\delta > 0$ and a subsequence $\{x_{k_{i_j}}\} \subset \{x_{k_i}\}$ such that $f_\star + \delta < f(x_{k_{i_j}})$ for all $j \in \mathbb{N}$. Since $\|x_{k_i} - x^\star\| < \|x_{k_i+1} - x^\star\|$ and $f_\star < f(x_{k_{i_j}})$ hold for any $j \in \mathbb{N}$, we can use Lemma 3.1 to get

$$v_{k_{i_j}} \left(1 - \alpha_{k_{i_j}}\right) \left(2\left(\frac{\delta}{L}\right)^{\frac{1}{\beta}} - v_{k_{i_j}}\right) < 0$$

for all $j \in \mathbb{N}$. The above inequality does not hold for sufficiently large $j \in \mathbb{N}$, since the real sequence $\{v_k\}$ converges to zero. Therefore, we arrive at a contradiction, and thus, $\limsup_{i\to\infty} f(x_{k_i}) \leq f_\star$.

The boundedness of the sequence $\{x_{k_i}\}$ guarantees the existence of a subsequence $\{x_{k_{i_j}}\} \subset \{x_{k_i}\}$ that weakly converges to some point $u \in H$. To show that u is a fixed point of the mapping T, let us assume that it is not. Recall that the real sequence $\{\|x_{k_i} - T(x_{k_i})\|\}$ converges to zero. Hence, the nonexpansivity of T together with Proposition 2.5 produces a contradiction,

$$\liminf_{j\to\infty} \left\|x_{k_{i_j}} - u\right\| < \liminf_{j\to\infty} \left\|x_{k_{i_j}} - T(u)\right\|$$
$$\leq \liminf_{j\to\infty} \left(\left\|x_{k_{i_j}} - T(x_{k_{i_j}})\right\| + \left\|T(x_{k_{i_j}}) - T(u)\right\|\right)$$
$$\leq \liminf_{j\to\infty} \left\|x_{k_{i_j}} - u\right\|.$$

Therefore, we have $u \in \mathrm{Fix}(T)$. In addition, Proposition 2.6 means that the objective functional f has weakly lower semicontinuity. Hence,

$$f(u) \leq \liminf_{j\to\infty} f(x_{k_{i_j}}) \leq \limsup_{i\to\infty} f(x_{k_i}) \leq f_\star$$

holds. This implies conclusively that there exists a subsequence of $\{x_k\}$ which weakly converges to the optimum $u \in X^\star$.

We omit the proof of the additional part of this theorem. The complete proof is in Appendix F. □

4. Numerical Experiments

To confirm that Algorithm 1 converges to the optimum and evaluate its performance, we ran it and an existing algorithm [22, Algorithm (14)] on a concrete constrained quasiconvex optimization problem, i.e., the Cobb-Douglas production efficiency problem [3, Problem (3.13)], [11, Problem (6.1)], [35, Section 1.7]. Let us redefine Example 3.3 with concrete constraints as follows.

Problem 4.1 ([3, Problem (3.13)], [11, Problem (6.1)], [35, Section 1.7]). Suppose that $H := \mathbb{R}^n$. Let $a_0, c_0 > 0$ and let $a, c \in (0, \infty)^n$ such that $\sum_{i=1}^n a_i = 1$. Furthermore, let $b_i \in [0, \infty)^n$, $\underline{p}_i \in [0, \infty)^n$, and $\bar{p}_i \in (0, \infty]^n$ for $i = 1, 2, \ldots, m$. Then, we would like to

$$\text{minimize } f(x) := \begin{cases} \frac{-a_0 \prod_{j=1}^n x_j^{a_j}}{\langle c, x \rangle + c_0} & (x \in [0, \infty)^n), \\ 0 & \text{(otherwise)}, \end{cases}$$

$$\text{subject to } \underline{p}_i \le \langle b_i, x \rangle \le \bar{p}_i \quad (i = 1, 2, \ldots, m),$$

$$x \in D := [0, M]^n,$$

where $M > 0$.

This problem is an instance of Example 3.3. Indeed, metric projections onto any closed half spaces, including boxes, can be computed explicitly [1, Example 29.20], and the transformations (T1–2) enable us to build a firmly nonexpansive mapping whose fixed point set coincides with their intersection. Our GitHub repository, `https://github.com/iiduka-researches/201811-kaz`, provides the means to make a metric projection onto a given half space and the transformations (T1–2). We used these implementation in the following experiments.

Before discussing our experiments, let us examine the background of this problem. The goal is to find the most efficient production factors under funding-level restrictions [11, Section 6]. As mentioned in Section 3, the objective function f represents the ratio between the total profit (what is obtained) and the total cost (how much expenditure is required) as an efficiency indicator. We also described how the total profit and the total cost are modeled in Section 3. The total profit is the numerator of the objective function and is modeled with the Cobb-Douglas production function on the production factors $x \in \mathbb{R}^n$. The total cost is the denominator of the objective function and is modeled with the affine function on the production factors $x \in \mathbb{R}^n$. There are a variety of constraints on the funding level [11,

Section 6]. These constraints represent the duties and restrictions of each production project $i = 1, 2, \ldots, m$. These indicators are modeled with affine functions and we set two parameters $\underline{p}_i, \overline{p}_i$ as lower and upper bounding constraints to the indicator of each project $i = 1, 2, \ldots, m$.

We conducted numerical experiments in three cases. First, in the unbounded constraint case, which is treated in the existing literature [11], we set $M := 100$ and did not guarantee the uniqueness of the optima, which is required for letting the generated sequence converge. Second, in the bounded constraint case, we set $M := 100$ and guaranteed the uniqueness of optima, as shown in Section 3. In practice, we cannot manufacture products infinitely because there are many restrictions on the amount of materials, capital, human resources, number and/or capacity of machines, environments, and so on. Therefore, this case has realistic experimental assumptions for optimizing production efficiency. Furthermore, we conducted an optimization over the generalized convex feasible sets.

We compared Algorithm 1 with the exact quasi-subgradient method (QSM) [22, Algorithm (14)]. In order for it to run, this algorithm requires a computation of the metric projection onto the feasible set $X \cap D$. Here, we used a trust-region algorithm for constrained optimization, `trust-constr`, implemented as the `scipy.optimize.minimize` solver provided by the SciPy fundamental library for scientific computing [21]. This algorithm also requires the step-sizes $\{v_k\}$ for it to run. We set its error tolerance a tenth of v_k for each $k = 1, 2, \ldots$. That is, we solved the subproblem to

$$\text{find } \|x_{k+1} - (x_k - v_k g_k)\|^2 \leq \min_{u \in X} \|u - (x_k - v_k g_k)\|^2 + \frac{v_k}{10}$$

subject to $x_{k+1} \in X$,

as the computation of the metric projection with the existing optimization solver.

In contrast to the existing algorithm which finds the metric projection onto the constraint set in the above way, we constructed a firmly nonexpansive mapping whose fixed point set coincides with the constraint set and gave it to Algorithm 1. Here, the constraint set is the intersection of the half-spaces $\{x : \underline{p}_i \leq \langle b_i, x \rangle\}$ and $\{x : \langle b_i, x \rangle \leq \overline{p}_i\}$ for $i = 1, 2, \ldots, m$. The metric

projection onto each half-space can be easily computed [1, Example 29.20]:

$$P_{\{x:\underline{p}_i \leq \langle b_i, x \rangle\}}(x) := \begin{cases} x & (\underline{p}_i \leq \langle b_i, x \rangle), \\ x + \frac{\underline{p}_i - \langle b_i, x \rangle}{\|b_i\|^2} b_i & \text{(otherwise)} \end{cases},$$

$$P_{\{x:\langle b_i, x \rangle \leq \overline{p}_i\}}(x) := \begin{cases} x & (\langle b_i, x \rangle \leq \overline{p}_i), \\ x + \frac{\overline{p}_i - \langle b_i, x \rangle}{\|b_i\|^2} b_i & \text{(otherwise)} \end{cases}$$

for all $i = 1, 2, \ldots, m$ and for any $x \in H$. To construct a nonexpansive mapping whose fixed point set coincides with the intersection of the above sets, we use the transformation (T1) and construct

$$\tilde{T}(x) := \frac{1}{m} \sum_{i=1}^{m} \frac{P_{\{x:\underline{p}_i \leq \langle b_i, x \rangle\}}(x) + P_{\{x:\langle b_i, x \rangle \leq \overline{p}_i\}}(x)}{2}$$

for any $x \in H$. This mapping $\tilde{T}$ is nonexpansive, but not firmly nonexpansive. Therefore, we convert the nonexpansive mapping $\tilde{T}$ into the corresponding firmly nonexpansive one T by using the transformation (T2):

$$T := \frac{\mathrm{Id} + \tilde{T}}{2}.$$

We gave T to Algorithm 1 in the experiment.

Our experimental environment was as follows: Python 3.6.6 with NumPy 1.15.0 [29] and SciPy 1.1.0 [21] libraries on macOS High Sierra version 10.13.6 on Mac Pro (Late 2013) with a 3 GHz 8 Cores Intel Xeon E5 CPU and 32GB 1800MHz DDR3 ECC memory. We used the `time.process_time` method for the evaluating computational time of each algorithm. The method was implemented with the `clock_gettime(2)` system call and had a 10^{-6} second resolution. Our GitHub repository, `https://github.com/iiduka-researches/201811-kaz`, provides the codes that were used in the experiments. It has the implementations of Algorithm 1 and QSM and miscellaneous utilities including higher-order functions to be used for composing a nonexpansive mapping.

We ran Algorithm 1 and QSM with five different randomly chosen initial points, limited their computational time to ten seconds, and evaluated the average of the computed number of iterations k and the following values:

$$V_{\text{func}} := \frac{1}{8} \sum_{i=1}^{8} f(x_{(i)}^{\star}), \quad V_{\text{dist}} := \frac{1}{8} \sum_{i=1}^{8} \left\| x_{(i)}^{\star} - T(x_{(i)}^{\star}) \right\|,$$

where $x_{(i)}^{\star}$ is the solution obtained for each sampling $i = 1, 2, \ldots, 8$.

4.1. Unbounded constraint case

Here, we ran Algorithm 1 and QSM on Problem 4.1 with the following settings: $n := 100$; $m := 100$; $a_0, c_0 \in (0, 10]$, $\tilde{a} \in (0, 1]^n$, $c \in (0, 10]^n$ were chosen randomly; $a := \tilde{a} / \sum_{i=1}^n a_i$; $b_i \in [0, 1)^n$, $\underline{p}_i \in [0, 25\|b_i\|)$ were chosen randomly for each $i = 1, 2, \ldots, m$; $\bar{p}_i := \infty$ for all $i = 1, 2, \ldots, m$; and $M := +\infty$.

The experimental results are shown in Table 1. The proposed algorithm

Table 1: Results of unbounded constraint case.

	k	V_{func}	V_{dist}
Alg. 1 ($v_k := 10^{-1}$)	11495.5	-0.00092189	$3.18571477 \times 10^{-13}$
Alg. 1 ($v_k := 10^{-2}$)	11539.5	-0.00061939	$3.17169706 \times 10^{-13}$
Alg. 1 ($v_k := 10^{-3}$)	11459.2	-0.00049238	$3.10991757 \times 10^{-13}$
Alg. 1 ($v_k := 10^{-1}/k$)	11474.5	-0.00046070	$3.05626418 \times 10^{-13}$
Alg. 1 ($v_k := 10^{-2}/k$)	11264.5	-0.00045498	$3.16790812 \times 10^{-13}$
Alg. 1 ($v_k := 10^{-3}/k$)	11318.4	-0.00045426	$3.28061131 \times 10^{-13}$
QSM ($v_k := 10^{-1}$)	160.4	-0.00050596	$4.04768150 \times 10^{-13}$
QSM ($v_k := 10^{-2}$)	39.6	-0.00045707	$4.12640954 \times 10^{-13}$
QSM ($v_k := 10^{-3}$)	25.1	-0.00045306	$4.00174170 \times 10^{-13}$
QSM ($v_k := 10^{-1}/k$)	36.1	-0.00045750	$4.23263890 \times 10^{-13}$
QSM ($v_k := 10^{-2}/k$)	26.6	-0.00045353	$3.99258739 \times 10^{-13}$
QSM ($v_k := 10^{-3}/k$)	20.5	-0.00045275	$4.10328799 \times 10^{-13}$

(Algorithm 1) can iterate the computation more times than the existing one within the same computational time. Algorithm 1 does not require any subproblem to be solved, while QSM requires one to be solved in order to find a metric projection onto the constraint set. Therefore, the required time for computing an iteration of Algorithm 1 is much less than that of QSM. According to the values of D, both Algorithm 1 and QSM for any step-size (and no matter whether a constant or diminishing step-size rule was used) can obtain the solution belonging to the constraint set. Indeed, our experimental environment (NumPy) used the `float64` data type (double precision float: sign bit, 11-bit exponent, and 52-bit mantissa) to express a real number, and its resolution is 10^{-15}. By considering the number of dimensions as well, we can regard all values of D to be almost zero. Let us examine the functional values of the obtained solutions. When we applied Algorithm 1 and QSM to the problem with the same step-size, we found that the function value of the solution obtained by Algorithm 1 is better than

that of QSM. In particular, the function value obtained by Algorithm 1 with $v_k := 10^{-1}$ is nearly twice as good as QSM with the same step-size. Since Algorithm 1 can iterate the main loop more times than QSM, it can reduce the functional value sufficiently.

4.2. Bounded constraint case

Next, we evaluated Algorithm 1 and QSM when they were run with the following settings: $n := 100$; $m := 100$; $a_0, c_0 \in (0, 10]$, $\tilde{a} \in (0, 1]^n$, $c \in (0, 10]^n$ were chosen randomly; $a := \tilde{a}/\sum_{i=1}^n a_i$; $b_i \in [0, 1)^n$, $\underline{p}_i \in [0, 25\|b_i\|)$, $\overline{p}_i \in (75\|b_i\|, 100\|b_i\|]$ were chosen randomly for each $i = 1, 2, \ldots, m$; and $M := 100$. As shown in Example 3.3, this case satisfies Assumption 3.1. Therefore, the sequence generated by Algorithm 1 is guaranteed to converge to some optimum.

The experimental results are shown in Table 2. The existing algorithm

Table 2: Results of bounded constraint case.

	k	V_{func}	V_{dist}
Alg. 1 ($v_k := 10^{-1}$)	6254.0	-0.00092536	$1.21925957 \times 10^{-13}$
Alg. 1 ($v_k := 10^{-2}$)	6208.9	-0.00050503	$5.22534774 \times 10^{-3}$
Alg. 1 ($v_k := 10^{-3}$)	6276.1	-0.00019717	$6.42589842 \times 10^{-4}$
Alg. 1 ($v_k := 10^{-1}/k$)	6293.8	-0.00014214	$1.28281187 \times 10^{-5}$
Alg. 1 ($v_k := 10^{-2}/k$)	6245.6	-0.00014163	$8.56673942 \times 10^{-6}$
Alg. 1 ($v_k := 10^{-3}/k$)	6294.0	-0.00014162	$8.33884360 \times 10^{-6}$
All Results of QSM	0.0	—	—

(QSM) with all step-size rules could not compute even one iteration within the time limit, 10 seconds. QSM required about 15 seconds to compute the first iteration. In this case, QSM must compute the metric projection onto the intersection of two hundred halfspaces and a box, but this intersection is too complex in shape to compute quickly. Therefore, it could not deal with this instance. In contrast, Algorithm 1 solved this instance. In particular, Algorithm 1 with $v_k := 10^{-1}$ found the solution having the best function value and belonging to the constraint set. Therefore, it can solve problems even if their constraint sets have complex shapes.

4.3. Optimization over generalized convex feasible sets

Finally, let us consider the case in which conflicts of constraints exist, i.e., the intersection of the constraint sets may be empty. Even in this

case, the proposed algorithm can be used if the constraints are extended to generalized convex feasible sets, such as described in Section 2.

In the two previous subsections, we computed the metric projection with the constrained smooth optimization solver provided by the SciPy library. However, it is difficult to find the metric projection onto the set of minimizers of the functional (1) due to the discontinuity of its Hessian matrix and the complexity of the problem. Therefore, in this subsection, we will examine only the performance of Algorithm 1.

The settings of this experiment were as follows: $n := 100$; $m := 100$; $a_0, c_0 \in (0, 10]$, $\tilde{a} \in (0, 1]^n$, $c \in (0, 10]^n$ were chosen randomly; $a := \tilde{a}/\sum_{i=1}^n a_i$; $b_i \in [0, 1)^n$, $\underline{p}_i, \overline{p}_i \in [0, 100\|b_i\|)$ were chosen randomly for each $i = 1, 2, \ldots, m$; and $M := +\infty$. The existence of a constraint $i \in \{1, 2, \ldots, m\}$ which satisfies $\overline{p}_i < \underline{p}_i$ was guaranteed. This implies that the intersection of at least one pair of constraints is empty. We used [19, Definition (9)] for constructing a firmly nonexpansive mapping whose fixed point set coincides with the constraint set. As in the previous subsection, this case also satisfies Assumption 3.1.

The experimental results are shown in Table 3. Algorithm 1 solved the

Table 3: Results of bounded constraint case.

	k	V_{func}	V_{dist}
Alg. 1 ($v_k := 10^{-1}$)	4988.9	-0.00015650	$2.48927983 \times 10^{-1}$
Alg. 1 ($v_k := 10^{-2}$)	4914.4	-0.00012585	$2.58656058 \times 10^{-1}$
Alg. 1 ($v_k := 10^{-3}$)	4833.6	-0.00012377	$2.59531605 \times 10^{-1}$
Alg. 1 ($v_k := 10^{-1}/k$)	4823.4	-0.00012363	$2.59665629 \times 10^{-1}$
Alg. 1 ($v_k := 10^{-2}/k$)	4818.4	-0.00012378	$2.59656373 \times 10^{-1}$
Alg. 1 ($v_k := 10^{-3}/k$)	4773.6	-0.00012380	$2.59626252 \times 10^{-1}$

problem similarly for each step-size rule; the results scarcely depended on the step-size rule. With a constant step-size $v_k = 10^{-1}$ for all $k \in \mathbb{N}$, it gave the best score in terms of V_{func} and V_{dist}. Although the time required for computing one iteration exceeded those of the two previous experiments, it approximated the solution of this complicated problem in 10 seconds.

5. Conclusion

We proposed the novel algorithm for solving the constrained quasiconvex optimization problem even if the metric projection onto its constraint set cannot be computed easily. We showed its convergence for constant and

diminishing step-size rules. When the step-size is constant, the limit inferiors of the functional value and the degree of approximation to the fixed point are guaranteed to be optimal and tolerate errors proportioned to the step-size. When the step-size is diminishing, the existence of a subsequence of the generated sequence such that it converges to the solution of the problem is ensured. Furthermore, when the problem satisfies certain conditions, the whole generated sequence converges to the solution.

The numerical experiments showed that our algorithm runs stably and lightly even if the constraint set is too complex for the existing method to run quickly. Therefore, the proposed algorithm is useful for solving complicated constrained quasiconvex optimization problems.

Acknowledgements

References

[1] Bauschke, H. H. and Combettes, P. L. (2017). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer International Publishing, second edition.

[2] Bertsekas, D. P., Nedic, A., and Ozdaglar, A. E. (2003). *Convex Analysis and Optimization.* Athena Scientific.

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

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

[5] Cheung, Y. and Lou, J. (2017). Proximal average approximated incremental gradient descent for composite penalty regularized empirical risk minimization. *Machine Learning*, 106(4):595–622.

[6] Combettes, P. L. and Bondon, P. (1999). Hard-constrained inconsistent signal feasibility problems. *IEEE Transactions on Signal Processing*, 47(9):2460–2468.

[7] Dyer, M. E. (1980). Calculating surrogate constraints. *Mathematical Programming*, 19(1):255–278.

[8] Greenberg, H. and Pierskalla, W. (1973). Quasi-conjugate functions and surrogate duality. *Cahiers Centre Études Recherche Opér*, 15:437–448.

[9] Halpern, B. (1967). Fixed points of nonexpansive maps. *Bulletin of the American Mathematical Society*, 73:957–961.

[10] Hardy, G. H., Littlewood, J. E., and Pólya, G. (1988). *Inequalities*. Cambridge University Press, second edition.

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

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

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

[14] Iiduka, H. (2013). Fixed point optimization algorithms for distributed optimization in networked systems. *SIAM Journal on Optimization*, 23(1):1–26.

[15] Iiduka, H. (2015a). Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping. *Mathematical Programming*, 149(1):131–165.

[16] Iiduka, H. (2015b). Parallel computing subgradient method for nonsmooth convex optimization over the intersection of fixed point sets of nonexpansive mappings. *Fixed Point Theory and Applications*, 2015(1):72.

[17] Iiduka, H. (2016a). Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings. *Mathematical Programming*, 159(1):509–538.

[18] Iiduka, H. (2016b). Incremental subgradient method for nonsmooth convex optimization with fixed point constraints. *Optimization Methods and Software*, 31(5):931–951.

[19] Iiduka, H. (2016c). Proximal point algorithms for nonsmooth convex optimization with fixed point constraints. *European Journal of Operational Research*, 253(2):503–513.

[20] Iiduka, H. (2017). Almost sure convergence of random projected proximal and subgradient algorithms for distributed nonsmooth convex optimization. *Optimization*, 66(1):35–59.

[21] Jones, E., Oliphant, T., Peterson, P., et al. (2001–). SciPy: Open source scientific tools for Python. [Online; accessed Aug. 16, 2018].

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

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

[24] Krasnosel'skiĭ, M. A. (1955). Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1(63)):123–127.

[25] Larsson, T., Patriksson, M., and Strömberg, A.-B. (1996). Conditional subgradient optimization—theory and applications. *European Journal of Operational Research*, 88(2):382–403.

[26] Lee, C.-Y., Johnson, A. L., Moreno-Centeno, E., and Kuosmanen, T. (2013). A more efficient algorithm for convex nonparametric least squares. *European Journal of Operational Research*, 227(2):391–400.

[27] Mann, W. R. (1953). Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510.

[28] Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer-Verlag New York, second edition.

[29] Oliphant, T. (2006). *A guide to NumPy*. Trelgol Publishing, USA.

[30] Opial, Z. (1967). Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society*, 73:591–597.

[31] Penot, J.-P. (1998). Are generalized derivatives useful for generalized convex functions? In Crouzeix, J.-P., Martinez-Legaz, J.-E., and Volle,

M., editors, *Generalized Convexity, Generalized Monotonicity: Recent Results, Nonconvex Optimization and Its Applications*, volume 27, pages 3–59. Kluwer Academic Publishers.

[32] Plastria, F. (1985). Lower subdifferentiable functions and their minimization by cutting planes. *Journal of Optimization Theory and Applications*, 46(1):37–53.

[33] Rockafellar, R. T. and Wets, R. J.-B. (2009). *Variational analysis*, volume 317. Springer Science & Business Media.

[34] Sakurai, K. and Iiduka, H. (2014). Acceleration of the Halpern algorithm to search for a fixed point of a nonexpansive mapping. *Fixed Point Theory and Applications*, 2014(1):202.

[35] Stancu-Minasian, I. M. (1997). *Fractional Programming: Theory, Methods and Applications*. Kluwer Academic Publishers.

[36] Takahashi, W. (2009). *Introduction to Nonlinear and Convex Analysis*. Yokohama Publishers.

[37] Yamada, I. (2001). The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In Butnariu, D., Censor, Y., and Reich, S., editors, *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, volume 8 of *Studies in Computational Mathematics*, pages 473–504. Elsevier.

[38] Zhang, T. (2010). Analysis of multi-stage convex relaxation for sparse regularization. *Journal of Machine Learning Research*, 11:1081–1107.

Appendix A. Sufficient condition for ensuring Assumption (A5)

Here, we present a sufficient condition for ensuring the boundedness of the sequence generated by Algorithm 1 without supposing the boundedness of the set D. Firstly, we define coerciveness, which is used as a condition for the objective functional to ensure the boundedness of the generated sequence.

Definition Appendix A.1 ([1, Definition 11.11, Proposition 11.12])**.** Let H be a real Hilbert space, and let f be a functional on H. We call f is *coercive* if

$$\lim_{\|x\| \to \infty} f(x) = \infty.$$

Furthermore, let $\text{lev}_{\leq \alpha} f$ denote the trench of f, i.e., $\text{lev}_{\leq \alpha} f := \{x \in H : f(x) \leq \alpha\}$, for any real number $\alpha \in \mathbb{R}$. Then, coerciveness is equivalent to the boundedness of all trenches of f.

Let us define the diameter of a set C as $\text{diam}(C) := \sup\{\|u - v\| : u, v \in C\}$ for the proof described later. The following proposition shows the necessary condition for ensuring the boundedness of the sequence generated by Algorithm 1.

Proposition Appendix A.1. *Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1. Suppose that Assumption 2.1 holds and there exists a number $k_0 \in \mathbb{N}$ such that $v_k < 1$ for all $k \geq k_0$. Assume that f is coercive. If one of the following holds,*

(i) Assumption (A4) holds,

(ii) the whole space H is an N-dimensional Euclidean space $\mathbb{R}^n$,

then Assumption (A5) is satisfied.

Proof. Let us prove this proposition by dividing it into case (i) and (ii). First, suppose that case (i) holds. We will proceed by way of contradiction and suppose that the sequence $\{x_k\}$ is unbounded. Then, there exists a subsequence $\{x_{k_i}\}$ of the sequence $\{x_k\}$ such that $\lim_{i \to \infty} \|x_{k_i}\| = \infty$. The coercivity of f implies that $\lim_{i \to \infty} f(x_{k_i}) = \infty$. Fix $x^\star \in X^\star$ arbitrarily. Assumption (A4) guarantees that

$$|f(z) - f_\star| \leq L \|z - x^\star\|^\beta$$

31

for all $z \in \mathrm{cl}(\mathrm{lev}_{<f(x_i)} f$ and for any $i \in \mathbb{N}$. Since $\lim_{i \to \infty} f(x_{k_i}) = \infty$, we have

$$|f(z) - f_\star| \le L \|z - x^\star\|^\beta$$

for all $z \in H$. Considering the point z appearing in the above inequality to be limited in the set $x^\star + \mathbf{B}$, we obtain

$$f(z) \le f_\star + L \|z - x^\star\|^\beta$$
$$\le f_\star + L$$

for all $z \in x^\star + \mathbf{B}$. Set δ to be the right side of the above inequality, i.e., $\delta := f_\star + L$. Then, the set $x^\star + \mathbf{B}$ is obviously a subset of the bounded trench $\mathrm{lev}_{\le \delta} f$.

From the assumption of this proposition, there exists a number $k_0 \in \mathbb{N}$ such that $v_k < 1$ for all $k \ge k_0$. For each $k \ge k_0$, let us consider the two separate cases: the case where the point x_k belongs to the trench $\mathrm{lev}_{\le \delta} f$, and its negation. First, let us consider the positive case; i.e., $x_k \in \mathrm{lev}_{\le \delta} f$ for $k \ge k_0$. The nonexpansivity of P_D and T and the fact that $x^\star \in \mathrm{Fix}(P_D) \cap \mathrm{Fix}(T)$ ensure that

$$\begin{aligned}
\|x_{k+1} - x^\star\| &= \|P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k)) - x^\star\| \\
&\le \|\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k) - x^\star\| \\
&\le \alpha_k \|x_k - x^\star\| + (1 - \alpha_k) \|T(x_k - v_k g_k) - x^\star\| \\
&\le \alpha_k \|x_k - x^\star\| + (1 - \alpha_k) \|x_k - v_k g_k - x^\star\| \\
&\le \|x_k - x^\star\| + (1 - \alpha_k)v_k
\end{aligned}$$

for any $k \ge k_0$ such that $x_k \in \mathrm{lev}_{\le \delta} f$. Here, both x_k and $x^\star$ belong to the bounded trench $\mathrm{lev}_{\le \delta} f$, and both $1 - \alpha$ and v_k are less than or equal to 1. Hence,

$$\|x_{k+1} - x^\star\| \le \mathrm{diam}(\mathrm{lev}_{\le \delta} f) + 1 < \infty$$

holds for any $k \ge k_0$ such that $x_k \in \mathrm{lev}_{\le \delta} f$. Next, let us consider the negative case; i.e., $x_k \notin \mathrm{lev}_{\le \delta} f$ for $k \ge k_0$. The nonexpansivity of P_D and T and the fact that $x^\star \in \mathrm{Fix}(P_D) \cap \mathrm{Fix}(T)$ ensure that

$$\begin{aligned}
\|x_{k+1} - x^\star\| &= \|P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k)) - x^\star\| \\
&\le \|\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k) - x^\star\| \\
&\le \alpha_k \|x_k - x^\star\| + (1 - \alpha_k) \|T(x_k - v_k g_k) - x^\star\| \\
&\le \alpha_k \|x_k - x^\star\| + (1 - \alpha_k) \|x_k - v_k g_k - x^\star\|
\end{aligned}$$

32

for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Let us consider the right term of the right side of the above inequality. Its squared value is bounded from above as follows:

$$\|x_k - v_k g_k - x^\star\|^2 = \|x_k - x^\star\|^2 - 2v_k \langle g_k, x_k - x^\star \rangle + v_k \langle g_k, v_k g_k \rangle$$
$$= \|x_k - x^\star\|^2 - v_k \langle g_k, x_k - x^\star \rangle - v_k \langle g_k, x_k - (x^\star + v_k g_k) \rangle$$

for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Here, $x_k \notin \mathrm{lev}_{\leq \delta} f$ implies that $f_\star \leq f_\star + L = \delta < f(x_k)$ for $f \geq k_0$. Therefore, we have $x^\star \in \mathrm{lev}_{\leq f(x_k)} f$ and $x^\star + v_k g_k \in x^\star + \mathbf{B} \subset \mathrm{lev}_{\leq f(x_k)} f$ for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Since the definition of $g_k \in \partial^\star f(x_k) \cap \mathbf{S}$ together with the preceding discussion implies that $\langle g_k, x_k - x^\star \rangle \geq 0$ and $\langle g_k, x_k - (x^\star + v_k g_k) \rangle \geq 0$, we have

$$\|x_k - v_k g_k - x^\star\| \leq \|x_k - x^\star\|$$

for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Hence, we have

$$\|x_{k+1} - x^\star\| \leq \|x_k - x^\star\|$$

for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. From the results for the both cases where $x_k \in \mathrm{lev}_{\leq \delta} f$ or not, we have

$$\|x_k - x^\star\| \leq \max\{\|x_1 - x^\star\|, \|x_2 - x^\star\|, \ldots, \|x_{k_0} - x^\star\|, \mathrm{diam}(\mathrm{lev}_{\leq \delta} f) + 1\}$$
$$< \infty$$

for all $k \in \mathbb{N}$. However, this contradicts the assumption that the sequence $\{x_k\}$ is unbounded. Therefore, we arrive at a contradiction and the boundedness of the sequence $\{x_k\}$ has been proved under this case.

Now let us suppose that case (ii) holds. In an N-dimensional Euclidean space $\mathbb{R}^N$, every closed, bounded set is compact [36, Problem 3.1.6]. Therefore, the nonempty, compact set $x^\star + \mathbf{B}$ contains a point $\bar{x}$ such that $f(\bar{x}) = \max_{x \in x^\star + \mathbf{B}} f(x)$ [36, Theorem 2.5.7]. Set $\delta := f(\bar{x}) \geq f_\star$. Then, the set $x^\star + \mathbf{B}$ is a subset of the bounded trench $\mathrm{lev}_{\leq \delta} f = \mathrm{lev}_{\leq f(\bar{x})} f$.

From the assumption of this proposition, there exists a number $k_0 \in \mathbb{N}$ such that $v_k < 1$ for all $k \geq k_0$. For each $k \geq k_0$, let us consider the two separate cases: the case where the point x_k belongs to the trench $\mathrm{lev}_{\leq \delta} f$, and its negation. First, let us consider the positive case; i.e., $x_k \in \mathrm{lev}_{\leq \delta} f$ for $k \geq k_0$. The nonexpansivity of P_D and T and the fact that $x^\star \in \mathrm{Fix}(P_D) \cap \mathrm{Fix}(T)$

ensure that

$$\begin{aligned}
\|x_{k+1} - x^\star\| &= \|P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k)) - x^\star\| \\
&\leq \|\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k) - x^\star\| \\
&\leq \alpha_k \|x_k - x^\star\| + (1 - \alpha_k)\|T(x_k - v_k g_k) - x^\star\| \\
&\leq \alpha_k \|x_k - x^\star\| + (1 - \alpha_k)\|x_k - v_k g_k - x^\star\| \\
&\leq \|x_k - x^\star\| + (1 - \alpha_k)v_k
\end{aligned}$$

for any $k \geq k_0$ such that $x_k \in \mathrm{lev}_{\leq \delta} f$. Here, both x_k and $x^\star$ belong to the bounded trench $\mathrm{lev}_{\leq \delta} f$, and both $1 - \alpha$ and v_k are less than or equal to 1. Hence,

$$\|x_{k+1} - x^\star\| \leq \mathrm{diam}(\mathrm{lev}_{\leq \delta} f) + 1 < \infty$$

holds for any $k \geq k_0$ such that $x_k \in \mathrm{lev}_{\leq \delta} f$. Next, let us consider the negative case; i.e., $x_k \notin \mathrm{lev}_{\leq \delta} f$ for $k \geq k_0$. The nonexpansivity of P_D and T and the fact that $x^\star \in \mathrm{Fix}(P_D) \cap \mathrm{Fix}(T)$ ensure that

$$\begin{aligned}
\|x_{k+1} - x^\star\| &= \|P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k)) - x^\star\| \\
&\leq \|\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k) - x^\star\| \\
&\leq \alpha_k \|x_k - x^\star\| + (1 - \alpha_k)\|T(x_k - v_k g_k) - x^\star\| \\
&\leq \alpha_k \|x_k - x^\star\| + (1 - \alpha_k)\|x_k - v_k g_k - x^\star\|
\end{aligned}$$

for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Let us consider the right term of the right side of the above inequality. Its squared value is bounded from above as follows:

$$\begin{aligned}
\|x_k - v_k g_k - x^\star\|^2 &= \|x_k - x^\star\|^2 - 2v_k \langle g_k, x_k - x^\star \rangle + v_k \langle g_k, v_k g_k \rangle \\
&= \|x_k - x^\star\|^2 - v_k \langle g_k, x_k - x^\star \rangle - v_k \langle g_k, x_k - (x^\star + v_k g_k) \rangle
\end{aligned}$$

for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Here, $x_k \notin \mathrm{lev}_{\leq \delta} f$ implies that $f_\star \leq f_\star + L = \delta < f(x_k)$ for $f \geq k_0$. Therefore, we have $x^\star \in \mathrm{lev}_{\leq f(x_k)} f$ and $x^\star + v_k g_k \in x^\star + \mathbf{B} \subset \mathrm{lev}_{\leq f(x_k)} f$ for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Since the definition of $g_k \in \partial^\star f(x_k) \cap \mathbf{S}$ together with the preceding discussion implies that $\langle g_k, x_k - x^\star \rangle \geq 0$ and $\langle g_k, x_k - (x^\star + v_k g_k) \rangle \geq 0$, we have

$$\|x_k - v_k g_k - x^\star\| \leq \|x_k - x^\star\|$$

for any $k \geq k_0$ such that $x_k \notin \mathrm{lev}_{\leq \delta} f$. Hence, we have

$$\|x_{k+1} - x^\star\| \leq \|x_k - x^\star\|$$

for any $k \geq k_0$ such that $x_k \notin \text{lev}_{\leq \delta} f$. From the results for the both cases where $x_k \in \text{lev}_{\leq \delta} f$ or not, we have

$$\|x_k - x^\star\| \leq \max\{\|x_1 - x^\star\|, \|x_2 - x^\star\|, \ldots, \|x_{k_0} - x^\star\|, \text{diam}(\text{lev}_{\leq \delta} f) + 1\}$$
$$< \infty$$

for all $k \in \mathbb{N}$. This implies that the sequence $\{x_k\}$ is bounded, and this completes the proof. $\quad\square$

The assumption on the step-size, i.e., the existence of a number k_0 such that $v_k < 1$ for all $k \geq k_0$, is satisfied whenever we adopt a diminishing step-size rule since the step-size sequence decreasingly converges to zero. Furthermore, it is also satisfied when we adopt a constant step-size rule with a small enough constant step-size $v < 1$. Overall, the coerciveness of the objective functional f (and the smallness of the step-sizes) is a sufficient condition for ensuring the boundedness of the generated sequence in the convergence analyses of constant or diminishing step-size rules.

Appendix B. Proofs of Proposition 2.4 and some examples

Proof of Proposition 2.4. Fix $g \in \partial^\star f(x) \cap \mathbf{S}$ arbitrarily. The continuity and quasiconvexity of f imply its level set $\text{lev}_{<f(x)} f$ is open and convex. This $\text{lev}_{<f(x)} f$ is not an empty set, since it has at least the point $x^\star$. The continuity of f also ensures $\text{bd}(\text{lev}_{<f(x)} f) \neq \emptyset$.

Set $r := \inf\{\|x^\star - u\| : u \in \text{bd}(\text{lev}_{<f(x)} f)\}$. Then, there exists a sequence $\{u_k\} \subset \text{bd}(\text{lev}_{<f(x)} f)$ such that $\|x^\star - u_k\| \leq r + 1/k$ for all $k \in \mathbb{N}$. The openness of $\text{lev}_{<f(x)} f$ implies that it is a distinct set from its boundary, i.e., $f(x) \leq f(u)$ for any $u \in \text{bd}(\text{lev}_{<f(x)} f)$. Hence,

$$\begin{aligned} f(x) - f_\star &\leq f(u_k) - f_\star \\ &\leq L \|u_k - x^\star\|^\beta \\ &< L \left(r + \frac{1}{k}\right)^\beta \quad (k \in \mathbb{N}). \end{aligned}$$

It follows that

$$f(x) - f_\star \leq L r^\beta. \tag{B.1}$$

From the definition of r, the open ball with center $x^\star$ and radius r is contained inside $\text{lev}_{<f(x)} f$. Therefore, $x^\star + (1 - 1/k)rg \in \text{lev}_{<f(x)} f$ holds

for any $k \in \mathbb{N}$, and we have

$$
\begin{aligned}
\left(1 - \frac{1}{k}\right) r - \langle g, x - x^\star \rangle &= \left(1 - \frac{1}{k}\right) r \|g\|^2 - \langle g, x - x^\star \rangle \\
&= \left\langle g, x^\star + \left(1 - \frac{1}{k}\right) rg - x \right\rangle \leq 0 \quad (k \in \mathbb{N}).
\end{aligned}
$$

This implies that $r \leq \langle g, x - x^\star \rangle$. Applying this inequality to inequality (B.1) gives $f(x) - f_\star \leq L \langle g, x - x^\star \rangle^\beta$. This completes the proof. $\qquad \square$

Proof of Example 3.2. The fixed point set $\mathrm{Fix}(T) = \mathrm{Fix}(\mathrm{Id})$ is obviously the whole space H. Therefore, the minimum value of f is 0 and its minimizer is only the origin. For any $z \in H$, we have

$$
\begin{aligned}
|f(z) - f(x^\star)| &= \min\{\|z\|, \alpha\} \\
&\leq \|z - x^\star\|.
\end{aligned}
$$

This implies that f satisfies the Hölder condition with degree 1 at its minimizer $x^\star$ on the whole space H. Expanding $\|x_{k+1}\|^2 = \|x_k - (v_k/2)g_k\|^2$ and using Proposition 2.4 with the above result, we have

$$
\begin{aligned}
\|x_{k+1}\|^2 &= \|x_k\|^2 - v_k \langle g_k, x_k - x^\star \rangle + \left(\frac{v_k}{2}\right)^2 \\
&\leq \|x_k\|^2 - v_k \min\{\|x_k\|, \alpha\} + \left(\frac{v_k}{2}\right)^2 \quad (k \in \mathbb{N}).
\end{aligned}
$$

When the index $k \in \mathbb{N}$ satisfies $\|x_k\| \leq \alpha$,

$$
\begin{aligned}
\|x_{k+1}\|^2 &\leq \|x_k\|^2 - v_k \|x_k\| + \left(\frac{v_k}{2}\right)^2 \\
&\leq \frac{5}{4}\alpha^2
\end{aligned}
$$

holds. The nonnegativeness of both sides of the above inequality ensures that $\|x_{k+1}\|$ is bounded from above by $\sqrt{5}\alpha/2$ for any $k \in \mathbb{N}$ satisfying $\|x_k\| \leq \alpha$. In the opposite case, i.e. if the index $k \in \mathbb{N}$ satisfies $\alpha < \|x_k\|$,

$$
\begin{aligned}
\|x_{k+1}\|^2 &\leq \|x_k\|^2 - v_k \alpha + \left(\frac{v_k}{2}\right)^2 \\
&= \|x_k\|^2 - v_k \left(\alpha - \frac{v_k}{4}\right)
\end{aligned}
$$

holds. Therefore, we have $\|x_{k+1}\| \leq \|x_k\|$ for any $k \in \mathbb{N}$ satisfying $\alpha < \|x_k\|$. Combining the conclusions of both cases, we can see that the sequence $\{\|x_k\|\}$ is bounded from above by $\max\{\|x_1\|, \sqrt{5}\alpha/2\}$. This completes the proof. $\qquad \square$

Proof of Example 3.3. The generated sequence $\{x_k\}$ is obviously bounded, since it is contained inside $[0, M]^n$. Fix $x^\star \in X^\star$ arbitrarily. The assumptions of this example imply that the feasible set $\mathrm{Fix}(T) \cap D$ has a point $u \in (0, \infty)^n$. Hence, the functional value of this point $f(u)$ is strictly less than $f(v) = 0$ for any $v \in \mathrm{bd}([0, \infty)^n)$. This means that $X^\star \subset (0, \infty)^n$ holds; that is, any optimum $x^\star \in X^\star$ belongs to the interior of $[0, \infty)^n$. Therefore, there exists some $\delta > 0$ such that the closed ball with center $x^\star$ and radius 2δ is contained inside $[0, \infty)^n$. Let us discuss the satisfiability of the Hölder condition on the set $[\delta, \infty)^n$ and on its complement $\overline{[\delta, \infty)^n}$ separately (see Figure B.2 for the relation between the point $x^\star$ and these two disjoint sets). First, fix $z \in [\delta, \infty)^n$ arbitrarily. Denoting the element-wise
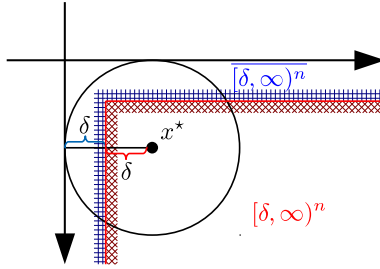

Figure B.2: Relation between the point $x^\star$, the set $\overline{[\delta, \infty)^n}$, and its complement

(Hadamard) division of x and y by $x \oslash y$, the gradient of the functional f can be written as

$$\nabla f(x) = \frac{-a_0 \prod_{j=1}^n x_j^{a_j}}{(\langle c, x \rangle + c_0)^2} c + \frac{-a_0 \prod_{j=1}^n x_j^{a_j}}{\langle c, x \rangle + c_0} (a \oslash x) \quad (x \in [\delta, \infty)^n).$$

Letting $\tilde{c}$ be a vector whose elements are each the minimum of c, an upper bound of the norms of the gradients can be evaluated by using the theorem of arithmetic and geometric means [10, Inequality (2.5.2)], as follows:

$$\begin{aligned}
\|\nabla f(x)\| &\leq \frac{a_0 \prod_{j=1}^n x_j^{a_j}}{(\langle c, x \rangle + c_0)^2} \|c\| + \frac{a_0 \prod_{j=1}^n x_j^{a_j}}{\langle c, x \rangle + c_0} \|a \oslash x\| \\
&\leq \frac{a_0}{\|\tilde{c}\|} \left(\frac{1}{\langle c, x \rangle + c_0} \|c\| + \|a \oslash x\| \right) \\
&\leq \frac{a_0}{\|\tilde{c}\|} \left(\frac{1}{\delta \|\tilde{c}\| + c_0} \|c\| + \frac{\|a\|}{\delta} \right) < \infty \quad (x \in [\delta, \infty)^n).
\end{aligned}$$

This implies that the image $(\|\nabla f(\cdot)\|)([\delta, \infty)^n)$ is bounded. The mean value theorem [28, Inequality (A.55)] ensures the existence of some $\alpha \in (0, 1)$ such

that

$$f(z) = f(x^\star) + \langle \nabla f((1 - \alpha)x^\star + \alpha z), z - x \rangle.$$

The convexity of $[\delta, \infty)$ and the boundedness of $\|\nabla f(\cdot)\|$ on $[\delta, \infty)^n$ imply that $\|\nabla f((1 - \alpha)x^\star + \alpha z)\| \leq (a_0/\|\tilde{c}\|)(\|c\|/(\delta\|\tilde{c}\| + c_0) + \|a\|/\delta)$. Therefore, the Cauchy-Schwarz inequality gives us the desired inequality as follows:

$$
\begin{aligned}
|f(z) - f(x^\star)| &= |\langle \nabla f((1 - \alpha)x^\star + \alpha z), z - x \rangle| \\
&\leq \|\nabla f((1 - \alpha)x^\star + \alpha z)\| \, \|z - x\| \\
&\leq \frac{a_0}{\|\tilde{c}\|} \left(\frac{1}{\delta\|\tilde{c}\| + c_0} \|c\| + \frac{\|a\|}{\delta} \right) \|z - x\|.
\end{aligned}
$$

This implies that the functional f satisfies the Hölder condition with degree 1 at the point $x^\star$ on the set $[\delta, \infty)^n$.

Next, fix $\bar{z} \in \overline{[\delta, \infty)^n}$ arbitrarily. The closed ball with center $x^\star$ and radius δ is contained inside $[\delta, \infty)^n$. Hence, $\|x^\star - z\| \geq \delta$ holds. From the definition of f, the maximum value of f is 0 and its range is less than or equal to 0. Therefore,

$$|f(z) - f(x^\star)| \leq \frac{f(x^\star)}{\delta} \|x^\star - z\|$$

holds. Hence, letting $L := \max\{(a_0/\|\tilde{c}\|)(\|c\|/(\delta\|\tilde{c}\| + c_0) + \|a\|/\delta), f(x^\star/\delta)\}$, it is clear that f satisfies the Hölder condition with degree 1 at the point $x^\star$ on the set $\mathbb{R}^n$. This completes the proof. $\qquad\square$

Appendix C. Proofs of Lemmas 3.1 and 3.2

Proof of Lemma 3.1. Fix $x^\star \in X^\star$ and $k \in \mathbb{N}$ arbitrarily. The convexity of $\|\cdot\|^2$ and the nonexpansivity of P_D and T ensure that

$$
\begin{aligned}
\|x_{k+1} - x^\star\|^2 &= \|P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k)) - P_D(x^\star)\|^2 \\
&\leq \|\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k) - x^\star\|^2 \\
&\leq \alpha_k \|x_k - x^\star\|^2 + (1 - \alpha_k) \|T(x_k - v_k g_k) - T(x^\star)\|^2 \\
&\leq \alpha_k \|x_k - x^\star\|^2 + (1 - \alpha_k) \|x_k - x^\star - v_k g_k\|^2 \\
&= \|x_k - x^\star\|^2 - 2v_k(1 - \alpha_k) \langle g_k, x_k - x^\star \rangle + (1 - \alpha_k)v_k^2.
\end{aligned}
$$
$$\tag{C.1}$$

On the other hand, Assumption (A4) and Proposition 2.4 ensure that

$$\left(\frac{f(x_k) - f_\star}{L} \right)^{\frac{1}{\beta}} \leq \langle g_k, x_k - x^\star \rangle.$$

Applying this inequality to inequality (C.1), we obtain

$$\|x_{k+1} - x^\star\|^2 \le \|x_k - x^\star\|^2 - 2v_k(1 - \alpha_k) \left(\frac{f(x_k) - f_\star}{L} \right)^{\frac{1}{\beta}} + (1 - \alpha_k)v_k^2.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Proof of Lemma 3.2. Fix $x \in \mathrm{Fix}(T) \cap D$ and $k \in \mathbb{N}$ arbitrarily, and set $M_1 := \sup_{k \in \mathbb{N}}(2|\langle g, x - T(x_k - v_k g_k)\rangle|)$. The boundedness of $\{x_k\}$ and $\{v_k\}$ ensures that there exist $M_2, M_3 \in \mathbb{R}$ such that

$$\|x_j\| \le M_2, \quad v_j \le M_3 \quad (j \in \mathbb{N}).$$

Therefore,

$$\begin{aligned}
\|x - T(x_j - v_j g_j)\| &\le \|x\| + \|x_j\| + v_j \\
&\le \|x\| + M_2 + M_3 < \infty \quad (j \in \mathbb{N}).
\end{aligned}$$

The Cauchy-Schwarz inequality, together with this boundedness of the real sequence $\{\|x - T(x_k - v_k g_k)\|\}$, indicates that $M_1 \le \|x\| + M_2 + M_3 < \infty$.

From the convexity of $\|\cdot\|^2$, we have

$$\begin{aligned}
\|x_{k+1} - x\|^2 &= \|P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k)) - P_D(x)\|^2 \\
&\le \|\alpha_k(x_k - x) + (1 - \alpha_k)(T(x_k - v_k g_k) - x)\|^2 \\
&\le \alpha_k \|x_k - x\|^2 + (1 - \alpha_k) \|T(x_k - v_k g_k) - x\|^2. \qquad \text{(C.2)}
\end{aligned}$$

Let us consider the term $\|T(x_k - v_k g_k) - x\|^2$. Using the firm nonexpansivity of T, we expand this term into

$$\begin{aligned}
&\|T(x_k - v_k g_k) - x\|^2 \\
&\le \|x_k - v_k g_k - x\|^2 - \|(\mathrm{Id} - T)(x_k - v_k g_k) - (\mathrm{Id} - T)(x)\|^2 \\
&= \|x_k - x\|^2 - 2v_k \langle g_k, x_k - x\rangle + v_k^2 \\
&\quad - \|x_k - T(x_k - v_k g_k)\|^2 + 2v_k \langle g_k, x_k - T(x_k - v_k g_k)\rangle - v_k^2 \\
&= \|x_k - x\|^2 - \|x_k - T(x_k - v_k g_k)\|^2 + 2v_k \langle g_k, x - T(x_k - v_k g_k)\rangle.
\end{aligned}$$

In view of the definition of M_1, the set $\{2v_k\langle g_k, x - T(x_k - v_k g_k)\rangle : k \in \mathbb{N}\}$ is bounded from above by it. Therefore, we obtain

$$\|T(x_k - v_k g_k) - x\|^2 \le \|x_k - x\|^2 - \|x_k - T(x_k - v_k g_k)\|^2 + v_k M_1.$$

Applying this inequality to inequality (C.2) yields the desired inequality:

$$\|x_{k+1} - x\|^2$$
$$\leq \alpha_k \|x_k - x\|^2 + (1 - \alpha_k)\left(\|x_k - x\|^2 - \|x_k - T(x_k - v_k g_k)\|^2 + v_k M_1\right)$$
$$\leq \|x_k - x\|^2 - (1 - \alpha_k)\|x_k - T(x_k - v_k g_k)\|^2 + v_k M_1$$

This completes the proof. □

Appendix D. Proof of Theorem 3.1

Proof. We prove each inequality in order. First, we consider whether the inequality

$$\liminf_{k \to \infty} f(x_k) \leq f_\star + L\left(\frac{v}{2}\right)^\beta \tag{D.1}$$

holds. We will proceed by way of contradiction. Suppose that the inequality does not hold; i.e.,

$$f_\star + L\left(\frac{v}{2}\right)^\beta < \liminf_{k \to \infty} f(x_k).$$

The left-hand side of this inequality is strictly less than the right-hand side. Hence, with the positivity of L, we can choose a positive δ_1 such that

$$f_\star + L\left(\frac{v}{2} + \delta_1\right)^\beta < \liminf_{k \to \infty} f(x_k)$$

holds. The property of the limit inferior guarantees that there exists $k_0 \in \mathbb{N}$ such that

$$f_\star + L\left(\frac{v}{2} + \delta_1\right)^\beta < f(x_k) \quad (k \geq k_0). \tag{D.2}$$

Obviously, this implies that $f_\star < f(x_k)$ for any $k \geq k_0$. Therefore, all assumptions of Lemma 3.1 are satisfied when $k \geq k_0$, and the following inequality holds for some $x^\star \in X^\star \neq \emptyset$.

$$\|x_{k+1} - x^\star\|^2$$
$$\leq \|x_k - x^\star\|^2 - 2v(1 - \alpha_k)\left(\frac{f(x_k) - f_\star}{L}\right)^{\frac{1}{\beta}} + (1 - \alpha_k)v^2 \quad (k \geq k_0).$$

40

Applying inequality (D.2) to the above inequality, we have

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - 2v\delta_1(1 - \alpha_k)$$

$$\leq \|x_{k_0} - x^\star\|^2 - 2v\delta_1 \sum_{j=k_0}^{k}(1 - \alpha_k) \quad (k \geq k_0). \qquad (D.3)$$

From Assumption (A6), $\limsup_{k\to\infty} \alpha_k < 1$ holds. Hence, a starting index $k_1 \in \mathbb{N}$ greater than k_0 exists such that the subsequence $\{\alpha_k\}_{k \geq k_1}$ is bounded above by some positive real that is strictly less than 1. This means that inequality (D.3) does not hold for large enough $k \geq k_1$, and we have arrived at a contradiction. Therefore, inequality (D.1) holds.

Next, let us prove the remaining part of this theorem, in other words, show that the inequality

$$\liminf_{k\to\infty} \|x_k - T(x_k)\|^2 \leq Mv$$

holds for some positive real $M > 0$. Fix $x \in \mathrm{Fix}(T)$ arbitrarily. Lemma 3.2 guarantees the existence of a nonnegative real $M_1 \geq 0$ such that

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - (1 - \alpha_k)\|x_k - T(x_k - vg_k)\|^2 + vM_1 \quad (k \in \mathbb{N}). \qquad (D.4)$$

In view of Assumption (A6), $\liminf_{k\to\infty}(1 - \alpha_k)$ is positive; i.e., it is not equal to zero. We will again proceed by way of contradiction and suppose that

$$\liminf_{k\to\infty} \|x_k - T(x_k - vg_k)\|^2 \leq \frac{2vM_1}{\liminf_{k\to\infty}(1 - \alpha_k)} \qquad (D.5)$$

does not hold and

$$\frac{2vM_1}{\liminf_{k\to\infty}(1 - \alpha_k)} < \liminf_{k\to\infty} \|x_k - T(x_k - vg_k)\|^2$$

holds. In the same ways choosing δ_1 in the first part of this proof, we can find a positive $\delta_2 > 0$ that satisfies

$$\frac{2vM_1}{\liminf_{k\to\infty}(1 - \alpha_k)} + \delta_2 < \liminf_{k\to\infty} \|x_k - T(x_k - vg_k)\|^2.$$

The property of the limit inferior guarantees that a positive number $k_2 \in \mathbb{N}$ exists such that

$$\frac{2vM_1}{\liminf_{k\to\infty}(1 - \alpha_k)} + \delta_2 < \|x_k - T(x_k - vg_k)\|^2 \quad (k \geq k_2).$$

Applying the above inequality to inequality (D.4), we obtain

$$\|x_{k+1} - x\|^2$$
$$\leq \|x_k - x\|^2 - (1 - \alpha_k) \left(\frac{2vM_1}{\liminf_{k\to\infty}(1 - \alpha_k)} + \delta_2 \right) + vM_1 \quad (k \geq k_2).$$

The fundamental property of the limit inferior also ensures the existence of a number $k_3 \in \mathbb{N}$ larger than k_2 such that $\liminf_{k\to\infty}(1 - \alpha_k)/2 < (1 - \alpha_k)$ for any $k \geq k_3$. Therefore, we have

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - \frac{\delta_2}{2} \liminf_{k\to\infty}(1 - \alpha_k)$$
$$\leq \|x_{k_3} - x\|^2 - \frac{\delta_2}{2}(k - k_3 + 1) \liminf_{k\to\infty}(1 - \alpha_k) \quad (k \geq k_3).$$

Since the above inequality does not hold for large enough $k \geq k_3$, we have arrived at a contradiction. Therefore, inequality (D.5) holds.

With this inequality, let us evaluate the squared distance between an element of the generated sequence and its transformed point by the nonexpansive mapping T. Using the triangle inequality and the nonexpansivity of T, we have

$$\|x_k - T(x_k)\|^2 \leq (\|x_k - T(x_k - vg_k)\| + \|T(x_k - vg_k) - T(x_k)\|)^2$$
$$\leq (\|x_k - T(x_k - vg_k)\| + v)^2$$
$$= \|x_k - T(x_k - vg_k)\|^2 + 2v\|x_k - T(x_k - vg_k)\| + v^2 \quad (k \in \mathbb{N}).$$

Now, since x is a fixed point of the mapping T, we expand the second term of the above expression as $\|x_k - T(x_k - vg_k)\| \leq 2\|x_k - x\| + v$ for any $k \in \mathbb{N}$. Furthermore, Assumption (A5) ensures the existence of a positive real $M_2 > 0$ bounding the set $\{\|x_k - x\| : k \in \mathbb{N}\}$ from above. Hence, we finally obtain

$$\|x_k - T(x_k)\|^2 \leq \|x_k - T(x_k - vg_k)\|^2 + 4v(M_2 + v) \quad (k \in \mathbb{N}).$$

Taking the limit inferior of both sides of the above inequality yields

$$\liminf_{k\to\infty} \|x_k - T(x_k)\|^2 \leq \liminf_{k\to\infty} \|x_k - T(x_k - vg_k)\|^2 + 4v(M_2 + v).$$

Set $M := 2(M_1/\liminf_{k\to\infty}(1 - \alpha_k) + 2(M_2 + v)) \in \mathbb{R}$. Applying inequality (D.5) to the above, we obtain the desired inequality as follows:

$$\liminf_{k\to\infty} \|x_k - T(x_k)\|^2 \leq 2 \left(\frac{M_1}{\liminf_{k\to\infty}(1 - \alpha_k)} + 2(M_2 + v) \right) v$$
$$\leq Mv.$$

This completes the proof. $\qquad\qquad\square$

Appendix E. Proof of Lemma 3.3

Proof. We will proceed by way of contradiction and suppose that the conclusion $\liminf_{k\to\infty} f(x_k) \leq f_\star$ does not hold, that is, $\liminf_{k\to\infty} f(x_k) > f_\star$. There exists a positive number $\delta > 0$ and an index $k_0 \in \mathbb{N}$ such that $f_\star + \delta < f(x_k)$ for all $k \geq k_0$. Furthermore, the assumption that the real sequence $\{v_k\}$ converges to zero guarantees the existence of an index $k_1 \geq k_0$ such that $v_k < (\delta/L)^{1/\beta}$ for all $k \geq k_1$. Applying these two inequalities to Lemma 3.1, we have

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - 2v_k(1-\alpha_k)\left(\frac{f(x_k) - f_\star}{L}\right)^{\frac{1}{\beta}} + (1-\alpha_k)v_k^2$$

$$< \|x_k - x^\star\|^2 - v_k(1-\alpha_k)\left(\frac{\delta}{L}\right)^{\frac{1}{\beta}}$$

$$< \|x_{k_1} - x^\star\|^2 - \left(\frac{\delta}{L}\right)^{\frac{1}{\beta}} \sum_{n=k_1}^{k} v_k(1-\alpha_k) \quad (k \geq k_1).$$

Assumption (A6) guarantees the existence of a positive number $\underline{\alpha}$ and an index $k_2 \geq k_1$ such that $\underline{\alpha} < 1 - \alpha_k$ for any $k \geq k_2$. This implies that the above inequality does not hold for a sufficiently large $k \geq k_2$; hence, we arrive at a contradiction. This completes the proof. $\qquad\square$

Appendix F. Proof of the Additional Statement of Theorem 3.2

Proof. Let us prove the additional statement of this theorem. The following proof is played under the assumption that the solution $x^\star \in X^\star$ is unique and the whole space is an N-dimensional Euclidean space, i.e., $H = \mathbb{R}^N$.

The existence of a subsequence $\{x_{k_i}\}$ that weakly converges to a unique solution $x^\star$ is obtained from Theorem 3.2. In Euclidean space, weak convergence coincides with strong convergence. Therefore, the sequence $\{x_{k_i}\}$ converges to a unique $x^\star$. If some number $k_0 \in \mathbb{N}$ exists such that $\|x_{k+1} - x^\star\| \leq \|x_k - x^\star\|$ for all $k \geq k_0$, the whole sequence $\{x_k\}$ converges to a point in $X^\star$. Let us consider the opposite case. Let $\{k_i\} \subset \mathbb{N}$ be the sequence of all indexes satisfying $\|x_{k_i} - x^\star\| < \|x_{k_i+1} - x^\star\|$ (and $k_i < k_{i+1}$ for any $i \in \mathbb{N}$). According to the assumption, this sequence is infinite. The sequence $\{x_{k_i}\}$ is now bounded, and this implies that it has a subsequence converging to a unique optimum $x^\star \in X^\star$. The above discussion ensures that any converging subsequence of $\{x_{k_i}\}$ converges to a unique optimum $x^\star \in X^\star$. This further implies that $\{x_{k_i}\}$ also converges to this optimum $x^\star \in X^\star$. From the assumed settings, $\|x_{j+1} - x^\star\| \leq \|x_j - x^\star\|$ holds for any index $j \in \mathbb{N}$ that does

not belong to the set $\{k_i\}$. The convergence of the sequence $\{x_{k_i}\}$ means that, for any $\epsilon > 0$, there exists an index $\hat{i} \in \mathbb{N}$ such that $\|x_{k_{\hat{i}}} - x^\star\| < \epsilon$. Furthermore, for any index $k \geq k_{\hat{i}}$ that does not belong to the set $\{k_i\}$, $\|x_{k_{\hat{i}}} - x^\star\| \leq \|x_{k_i} - x^\star\| < \epsilon$ also holds for $i := \max\{k_i : k_i \leq k\} \geq k_{\hat{i}}$. This implies that the whole sequence $\{x_k\}$ converges to a unique optimum $x^\star \in X^\star$. This completes the proof. $\qquad\square$

Appendix G. Efficiency

Here, we discuss the rate of convergence of Algorithm 1 in terms of the value of the objective functional and the distance to the fixed point set. Furthermore, we discuss a sufficient condition to obtain finite convergence to some solution.

Let us start with the convergence rate analysis in terms of the objective function. Here, we will use the following concepts originally introduced in [11, 22].

Definition Appendix G.1 ([11, Section 5], [22, Section 6]). Define the following notations:

(i) $x_k^\star :\in \operatorname{argmin}_{x \in \{x_1, x_2, \ldots, x_k\}} f(x)$,

(ii) $r_k := \sup\{r > 0 : x^\star + r\mathbf{B} \subset \operatorname{lev}_{<f(x_k^\star)} f\}$.

$x_k^\star$ expresses the best solution acquired until the k-th iteration, and r_k expresses the distance between the level set of $x_k^\star$ and the optimal solution. The difference between the above definitions and the original ones is in considering the possibility that the generated sequence may be out of the fixed point set, in other words, the constraint set.

We will use the following propositions.

Proposition Appendix G.1. *Let $C \subset H$ be a nonempty, convex set, and suppose that $x \in C$, $y \notin C$. Then, there exists $\alpha \in [0, 1]$ such that $x + \alpha(y - x) \in \operatorname{bd}(C)$.*

Proof. Define $\alpha := \sup\{\hat{\alpha} \in [0, 1] : x + \hat{\alpha}(y - x) \in C\}$ and fix $\epsilon > 0$ arbitrarily. From the properties of the supremum, there exists $\beta > \alpha - \epsilon/\|y - x\|$ such that $x + \beta(y - x) \in C$. Thus, we have

$$\|(x + \alpha(y - x)) - (x + \beta(y - x))\| = (\alpha - \beta)\|y - x\| < \epsilon.$$

Since $\epsilon > 0$ was chosen arbitrarily, the above inequality implies that the point $(x + \alpha(y - x))$ is an adherent point of C. Furthermore, there exists

$\gamma > \alpha + \epsilon/\|y - x\|$ such that $x + \gamma(y - x) \notin C$ due to the properties of the supremum. Hence, we also have

$$\|(x + \alpha(y - x)) - (x + \gamma(y - x))\| = (\gamma - \alpha)\|y - x\| < \epsilon.$$

Since $\epsilon > 0$ was chosen arbitrarily, the above inequality implies that the point $(x + \alpha(y - x))$ is an adherent point of the complement of C. Therefore, $x + \alpha(y - x)$ belongs to the boundary of C. This completes the proof. $\quad\square$

Proposition Appendix G.2 ([1, Corollary 2.15]). *Let $x, y \in H$, and let $\alpha \in \mathbb{R}$. Then,*

$$\|\alpha x + (1 - \alpha)y\|^2 = \alpha\|x\|^2 + (1 - \alpha)\|y\|^2 - \alpha(1 - \alpha)\|x - y\|^2$$

holds.

We start by proving the following lemma which leads us to the convergence rate of Algorithm 1.

Lemma Appendix G.1. *Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1, and suppose that Assumptions 2.1 and 3.1 hold. Assume that the sequence $\{v_k\}$ is bounded. Then,*

$$r_k \leq \frac{\|x_i - x^\star\|^2 + \sum_{j=i}^{k}(1 - \alpha_j)v_j^2}{2\sum_{j=i}^{k} v_j(1 - \alpha_j)}$$

for any $x^\star \in X^\star$, $k \in \mathbb{N}$, and $i \in \{1, 2, \ldots, k\}$.

Proof. Fix $x^\star \in X^\star$, $k \in \mathbb{N}$, and $i \in \{1, 2, \ldots, k\}$ arbitrarily. If r_k is nonpositive, the statement obviously holds. Therefore, let us consider the case where r_k is positive in the following. Fix $\delta \in (0, r_k)$ arbitrarily. The definition of r_k and monotonicity of the sequence $\{r_k\}$ imply that $x^\star - \delta g_j$ belongs to the level set $\text{lev}_{<f(x_j^\star)} f \subset \text{lev}_{<f(x_j)} f$ for any $j = 1, 2, \ldots, k$. Therefore, $\langle g_j, (x^\star - \delta g_j) - x_j \rangle \leq 0$ holds for all $j = 1, 2, \ldots, k$. Rearranging this inequality with the property $\|g_j\| = 1$, we have

$$\langle g_j, x^\star - x_j \rangle \leq \delta$$

for all $j = 1, 2, \ldots, k$. Here, the assumption $r_k > 0$ implies that $f_\star < f(x_j)$ for all $j = 1, 2, \ldots, k$. Hence, all assumptions of Lemma 3.1 are satisfied and

$$\begin{aligned}
\|x_{j+1} - j^\star\|^2 &\leq \|x_j - x^\star\|^2 - 2v_j(1 - \alpha_j)\langle g_j, x_j - x^\star \rangle + (1 - \alpha_j)v_j^2 \\
&\leq \|x_j - x^\star\|^2 - 2\delta v_j(1 - \alpha_j) + (1 - \alpha_j)v_j^2
\end{aligned}$$

45

is guaranteed by inequality (C.1) for all $j = 1, 2, \ldots, k$. Summing the above inequalities from $j = i$ to $j = k$ yields

$$0 \leq \|x_i - x^\star\|^2 - 2\delta \sum_{j=i}^{k} v_j(1 - \alpha_j) + \sum_{j=i}^{k}(1 - \alpha_j)v_j^2.$$

Transposing the term of δ, we get

$$\delta \leq \frac{\|x_i - x^\star\|^2 + \sum_{j=i}^{k}(1 - \alpha_j)v_j^2}{2\sum_{j=i}^{k} v_j(1 - \alpha_j)}.$$

The arbitrariness of $\delta \in (0, r_k)$ implies that

$$r_k \leq \frac{\|x_i - x^\star\|^2 + \sum_{j=i}^{k}(1 - \alpha_j)v_j^2}{2\sum_{j=i}^{k} v_j(1 - \alpha_j)}.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

In general settings, we can use this lemma to analyze the convergence rate in terms of r_k. The following two propositions present such an analysis for constant and diminishing step-size rules.

Proposition Appendix G.3 (Convergence rate analysis for constant step–size rule). *Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1, and suppose that the assumptions in Theorem 3.1 hold. Then, there exists a number $k_0 \in \mathbb{N}$ such that*

$$r_k \leq \frac{1}{1 - \limsup_{j \to \infty} \alpha_j} \left(\frac{\|x_{k_0} - x^\star\|^2}{(k - k_0 + 1)v} + \left(1 - \frac{1}{2} \liminf_{j \to \infty} \alpha_j\right) v \right)$$

holds for all $k \geq k_0$, in other words,

$$r_k = \mathcal{O}(1/k + v).$$

Furthermore,

$$r_k \leq \frac{\|x_1 - x^\star\|^2}{2(1 - \alpha)kv} + \frac{1}{2}v$$

holds for all $k \in \mathbb{N}$ when the sequence $\{\alpha_k\}$ satisfies $\alpha_k = \alpha \in (0, 1)$ for all $k \in \mathbb{N}$.

Proof. Assumption (A6) implies that there exists a number $k_0 \in \mathbb{N}$ such that

$$0 < \frac{1}{2}\left(1 - \limsup_{k\to\infty} \alpha_k\right) \leq 1 - \alpha_k \leq 1 - \frac{1}{2}\liminf_{k\to\infty} \alpha_k < 1$$

for all $k \geq k_0$. Therefore, Lemma Appendix G.1 with $i := k_0$ leads to the finding that

$$
\begin{aligned}
r_k &\leq \frac{\|x_{k_0} - x^\star\|^2 + \sum_{j=k_0}^{k}(1 - \alpha_j)v_j^2}{2\sum_{j=k_0}^{k} v_j(1 - \alpha_j)} \\
&\leq \frac{\|x_{k_0} - x^\star\|^2 + (1 - \liminf_{j\to\infty}\alpha_j/2)(k - k_0 + 1)v^2}{(1 - \limsup_{j\to\infty}\alpha_j)(k - k_0 + 1)v} \\
&= \frac{1}{1 - \limsup_{j\to\infty}\alpha_j}\left(\frac{\|x_{k_0} - x^\star\|^2}{(k - k_0 + 1)v} + \left(1 - \frac{1}{2}\liminf_{j\to\infty}\alpha_j\right)v\right)
\end{aligned}
$$

for all $k \geq k_0$.

Furthermore, if $\{\alpha_k\}$ satisfies $\alpha_k = \alpha \in (0,1)$ for all $k \in \mathbb{N}$, Lemma Appendix G.1 with $i := 1$ leads to

$$
\begin{aligned}
r_k &\leq \frac{\|x_1 - x^\star\|^2 + \sum_{j=i}^{k}(1 - \alpha_j)v_j^2}{2\sum_{j=i}^{k} v_j(1 - \alpha_j)} \\
&= \frac{\|x_1 - x^\star\|^2 + (1 - \alpha)kv^2}{2(1 - \alpha)kv} \\
&= \frac{\|x_1 - x^\star\|^2}{2(1 - \alpha)kv} + \frac{1}{2}v
\end{aligned}
$$

for all $k \in \mathbb{N}$. This completes the proof. $\qquad\square$

Proposition Appendix G.4 (Convergence rate analysis for diminishing step-size rule). *Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1, and suppose that the assumptions in Theorem 3.2 hold. Let c be a positive real number and assume that $v_k = c/k$ for all $k \in \mathbb{N}$. Then, there exists a number $k_0 \in \mathbb{N}$ such that*

$$r_k \leq \frac{\|x_{k_0} - x^\star\|^2 + 2c^2(1 - \liminf_{j\to\infty}\alpha_j/2)}{2c(1 - \limsup_{j\to\infty}\alpha_j)(\log(k+1) - \log(k_0))}$$

holds for all $k \geq k_0$, in other words,

$$r_k = \mathcal{O}(1/\log(k+1)).$$

In addition,

$$r_k \leq \frac{\|x_1 - x^\star\|^2 + 2c^2(1 - \alpha)}{2c(1 - \alpha)\log(k + 1)}$$

holds for all $k \in \mathbb{N}$ when the sequence $\{\alpha_k\}$ satisfies $\alpha_k = \alpha \in (0,1)$ for all $k \in \mathbb{N}$.

Proof. Assumption (A6) implies that there exists a number $k_0 \in \mathbb{N}$ such that

$$0 < \frac{1}{2}\left(1 - \limsup_{k \to \infty} \alpha_k\right) \leq 1 - \alpha_k \leq 1 - \frac{1}{2}\liminf_{k \to \infty} \alpha_k < 1$$

holds for all $k \geq k_0$. Therefore, Lemma Appendix G.1 with $i := k_0$ leads to the finding that

$$\begin{aligned}
r_k &\leq \frac{\|x_{k_0} - x^\star\|^2 + \sum_{j=k_0}^k (1 - \alpha_j)v_j^2}{2\sum_{j=k_0}^k v_j(1 - \alpha_j)} \\
&\leq \frac{\|x_{k_0} - x^\star\|^2 + c^2(1 - \liminf_{j \to \infty} \alpha_j/2)\sum_{j=k_0}^k 1/j^2}{2c(1 - \limsup_{j \to \infty} \alpha_j)\sum_{j=k_0}^k 1/j}
\end{aligned}$$

holds for all $k \geq k_0$. Using inequalities $\sum_{j=1}^\infty 1/j^2 \leq 2$, $\log(k+1) - \log(k_0) \leq \sum_{j=k_0}^k 1/j$, we have

$$r_k \leq \frac{\|x_{k_0} - x^\star\|^2 + 2c^2(1 - \liminf_{j \to \infty} \alpha_j/2)}{2c(1 - \limsup_{j \to \infty} \alpha_j)(\log(k + 1) - \log(k_0))}$$

for all $k \geq k_0$.

In addition, if $\{\alpha_k\}$ satisfies $\alpha_k = \alpha \in (0,1)$ for all $k \in \mathbb{N}$, Lemma Appendix G.1 with $i := 1$ leads to

$$\begin{aligned}
r_k &\leq \frac{\|x_1 - x^\star\|^2 + \sum_{j=i}^k (1 - \alpha_j)v_j^2}{2\sum_{j=i}^k v_j(1 - \alpha_j)} \\
&= \frac{\|x_1 - x^\star\|^2 + c^2(1 - \alpha)\sum_{j=i}^k 1/j^2}{2(1 - \alpha)\sum_{j=i}^k 1/j} \\
&\leq \frac{\|x_1 - x^\star\|^2 + 2c^2(1 - \alpha)}{2(1 - \alpha)\log(k + 1)}
\end{aligned}$$

for all $k \in \mathbb{N}$. This completes the proof. $\qquad\square$

48

The following theorem guarantees that the convergence rate with respect to the functional value can be bounded from above by one with respect to r_k under certain assumptions. This implies that, under the assumptions, we can deduce the convergence rate with respect to the functional value from the result of analyzing r_k. We give this result after proving the following theorem.

Theorem Appendix G.1. *Suppose that the whole space H is an N-dimensional Euclidean space $\mathbb{R}^N$. Let $\{x_k\} \subset H$ be a sequence generated by Algorithm 1 and suppose that Assumptions 2.1 and 3.1 hold. Assume that $f_\star < f(x_k^\star)$ holds. Then,*

$$f(x_k^\star) - f_\star \leq Lr_k^\beta$$

holds for all $k \in \mathbb{N}$.

Proof. Fix $k \in \mathbb{N}$ and $x^\star \in X^\star$ arbitrarily. Furthermore, fix $j \in \mathbb{N}$ arbitrarily. The complement of the slice $\text{lev}_{<f(x_k^\star)} f$, i.e., the set $\{x \in \mathbb{R}^N : f(x_k^\star) \leq f(x)\}$, is nonempty because $x_k^\star$ obviously belongs to it. Therefore, from the definition of r_k, there exists a point $v_j \in \mathbb{R}^N$ such that

$$v_j \notin \text{lev}_{<f(x_k^\star)} f \text{ and } \|x^\star - v_j\| \leq r_k + 1/j$$

hold. The assumption $f_\star < f(x_k^\star)$ implies that $x^\star \in \text{lev}_{<f(x_k^\star)} f$, and the above discussion obtained the property $v_j \notin \text{lev}_{<f(x_k^\star)} f$. Therefore, by proposition Appendix G.1, a number $\alpha_j \in [0,1]$ exists such that $w_j := x^\star + \alpha_j(v_j - x^\star) \in \text{bd}(\text{lev}_{<f(x_k^\star)} f)$. Let us consider the lower and upper bounds of the norm $\|x^\star - w_j\|$. Here, the fact $w_j \notin \text{lev}_{<f(x_k^\star)} f$ from the continuity of the objective functional f implies that $r_k \leq \|x^\star - w_j\|$ holds. The upper bound of the norm $\|x^\star - w_j\|$ is thus

$$\begin{aligned}
\|x^\star - w_j\| &= \|x^\star - (x^\star + \alpha_j(v_j - x^\star))\| \\
&= \alpha_j \|x^\star - v_j\| \\
&\leq \|x^\star - v_j\| \\
&\leq r_k + 1/j.
\end{aligned}$$

The sequence $\{w_j\}$ is bounded, since $\|w_j\| \leq \|x^\star\| + \|x^\star - w_j\| \leq \|x^\star\| + r_k + 1$ for all $j \in \mathbb{N}$. Therefore, together with the closedness of the boundary $\text{bd}(\text{lev}_{<f(x_k^\star)} f)$, there exists a subsequence $\{w_{j_t}\} \subset \{w_j\}$ and a point $u_k \in \text{bd}(\text{lev}_{<f(x_k^\star)} f)$ such that w_{j_t} converges to u_k. The continuity of $\|\cdot\|$ leads

us to the finding that

$$\|x^\star - u_k\| = \lim_{l\to\infty} \|x^\star - w_{j_l}\|$$
$$= r_k$$

due to the previous confirmation of the boundedness of the norm $\|x^\star - w_{j_t}\|$; i.e., $r_k \le \|x^\star - w_{j_t}\| \le r_k + 1/j_t$ holds for any $t \in \mathbb{N}$. Furthermore, the continuity of the objective functional f ensures the coincidence $f(u_k) = f(x_k^\star)$. Now, Assumption (A4) guarantees that the functional f satisfies the Hölder condition with degree $\beta > 0$ at the point $x^\star$ on the set $\mathrm{cl}(\mathrm{lev}_{<f(x_k)} f)$. Therefore, we have

$$f(x_k^\star) - f_\star = f(u_k) - f_\star$$
$$\le L \|x^\star - u_k\|^\beta$$
$$= L r_k^\beta.$$

This completes the proof. □

This theorem directly induces the following corollary giving the convergence rate in terms of the objective functional when the diminishing step-size rule is adopted.

Corollary Appendix G.1. *Suppose that the whole space H is an N-dimensional Euclidean space $\mathbb{R}^N$ and the assumptions in Theorem 3.2 hold. Let c be a positive real number and assume that $v_k = c/k$ for all $k \in \mathbb{N}$. Then, a number $k_0 \in \mathbb{N}$ exists such that*

$$f(x_k^\star) - f_\star \le L \left(\frac{\|x_{k_0} - x^\star\|^2 + 2c^2(1 - \liminf_{j\to\infty}\alpha_j/2)}{2c(1 - \limsup_{j\to\infty}\alpha_j)(\log(k+1) - \log(k_0))} \right)^\beta$$

holds for all $k \ge k_0$, in other words,

$$f(x_k^\star) - f_\star = \mathcal{O}\left(\frac{1}{(\log(k+1))^\beta} \right).$$

In addition,

$$f(x_k^\star) - f_\star \le L \left(\frac{\|x_1 - x^\star\|^2 + 2c^2(1 - \alpha)}{2c(1 - \alpha)\log(k+1)} \right)^\beta$$

holds for all $k \in \mathbb{N}$ when the sequence $\{\alpha_k\}$ satisfies $\alpha_k = \alpha \in (0,1)$ for all $k \in \mathbb{N}$.

Proof. This is an immediate consequence of Proposition Appendix G.4 and Theorem Appendix G.1. □

We thus far have discussed the convergence rate of Algorithm 1 in terms of the value of the objective functional. The remaining part of this section shows another convergence rate analysis of Algorithm 1, namely in terms of the distance to the fixed point set. The following theorem gives the convergence rate in terms of the distance to the fixed point set with respect to the averaged norm.

Theorem Appendix G.2. *Suppose that the assumptions in Theorem 3.2 hold. If $f_\star < f(x_k)$ for all $k \in \mathbb{N}$, then*

$$\frac{1}{k}\sum_{j=1}^{k} \|x_j - T(x_j)\| = \mathcal{O}(1/k).$$

Proof. Fix $k \in \mathbb{N}$ arbitrarily. Using Proposition Appendix G.2 and the fact that P_D is a nonexpansive mapping and $x^\star$ is its fixed point, we have

$$
\begin{aligned}
\|x_{k+1} - x^\star\|^2 &= \|P_D(\alpha_k x_k + (1 - \alpha_k)T(x_k - v_k g_k)) - P_D(x^\star)\|^2 \\
&\leq \|\alpha_k(x_k - x^\star) + (1 - \alpha_k)(T(x_k - v_k g_k) - x^\star)\|^2 \\
&= \alpha_k \|x_k - x^\star\|^2 + (1 - \alpha_k)\|T(x_k - v_k g_k) - x^\star\|^2 \\
&\quad - \alpha_k(1 - \alpha_k)\|x_k - T(x_k - v_k g_k)\|^2.
\end{aligned}
$$

Here, $x^\star$ is also a fixed point of the nonexpansive mapping T. Therefore, we can expand the second term of the above expression as follows:

$$
\begin{aligned}
\|T(x_k - v_k g_k) - x^\star\|^2 &= \|T(x_k - v_k g_k) - T(x^\star)\|^2 \\
&\leq \|x_k - x^\star - v_k g_k\|^2 \\
&= \|x_k - x^\star\|^2 - 2v_k \langle x_k - x^\star, g_k\rangle + v_k^2.
\end{aligned}
$$

Now, the assumption $f_\star < f(x_k)$ ensures that $-\langle x_k - x^\star, g_k\rangle \leq 0$ holds because g_k is a normal vector of the slice $\mathrm{lev}_{<f(x_k)} f$ at x_k. Hence, we have

$$\|T(x_k - v_k g_k) - x^\star\|^2 = \|x_k - x^\star\|^2 + v_k^2.$$

Overall, we have

$$
\begin{aligned}
\|x_{k+1} - x^\star\|^2 &\leq \alpha_k \|x_k - x^\star\|^2 + (1 - \alpha_k)(\|x_k - x^\star\|^2 + v_k^2) \\
&\quad - \alpha_k(1 - \alpha_k)\|x_k - T(x_k - v_k g_k)\|^2 \\
&= \|x_k - x^\star\|^2 - \alpha_k(1 - \alpha_k)\|x_k - T(x_k - v_k g_k)\|^2 + (1 - \alpha_k)v_k^2.
\end{aligned}
$$

51

Assumption (A6) guarantees that a number $k_0 \in \mathbb{N}$ exists such that $\liminf_{j\to\infty} \alpha_j/2 < \alpha_k < (1+\limsup_{j\to\infty} \alpha_j)/2$ holds for all $k \geq k_0$. Note that Assumption (A6) also ensures that $0 < \liminf_{j\to\infty} \alpha_k/2$ and $(1 + \limsup_{j\to\infty} \alpha_j)/2 < 1$. Therefore, together with the above inequality, we have

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 + \left(1 - \frac{1}{2}\liminf_{j\to\infty} \alpha_j\right) v_k^2$$

$$- \frac{1}{4}\left(\liminf_{j\to\infty} \alpha_j\right)\left(1 - \limsup_{j\to\infty} \alpha_j\right) \|x_k - T(x_k - v_k g_k)\|^2$$

$$\leq \|x_{k_0} - x^\star\|^2 + \left(1 - \frac{1}{2}\liminf_{j\to\infty} \alpha_j\right) \sum_{j=k_0}^{k} v_j^2$$

$$- \frac{1}{4}\left(\liminf_{j\to\infty} \alpha_j\right)\left(1 - \limsup_{j\to\infty} \alpha_j\right) \sum_{j=k_0}^{k} \|x_j - T(x_j - v_j g_j)\|^2$$

if $k \geq k_0$. This inequality implies that

$$\sum_{j=1}^{k} \|x_j - T(x_j - v_j g_j)\|^2$$

$$\leq \sum_{j=1}^{k_0-1} \|x_j - T(x_j - v_j g_j)\|^2$$

$$+ 4\left(\liminf_{j\to\infty} \alpha_j\right)^{-1}\left(1 - \limsup_{j\to\infty} \alpha_j\right)^{-1}\left(\|x_{k_0} - x^\star\|^2 + \left(1 - \frac{1}{2}\liminf_{j\to\infty} \alpha_j\right) \sum_{j=k_0}^{k} v_j^2\right)$$

holds[3]. Here, we assume that $\sum_{j=1}^{\infty} v_j^2$ converges. Therefore, the right side of the above inequality is bounded from above with respect to k. This implies that the sequence $\{\sum_{j=1}^{k} \|x_j - T(x_j - v_j g_j)\|^2\}$ is also bounded from above. Let $M \in \mathbb{R}$ denote an upper bound of this sequence.

Let us estimate the distance before and after applying the nonexpansive mapping to each approximation x_k. Using the parallelogram law and the nonexpansivity of T, we obtain

$$\|x_k - T(x_k)\|^2 = \|x_k - T(x_k - v_k g_k) + T(x_k - v_k g_k) - T(x_k)\|^2$$

$$\leq 2\|x_k - T(x_k - v_k g_k)\|^2 + 2\|T(x_k - v_k g_k) - T(x_k)\|^2$$

$$\leq 2\|x_k - T(x_k - v_k g_k)\|^2 + 2v_k^2.$$

---

[3]If k is less than k_0, we consider $\sum_{j=k_0}^{k} v_j^2 = 0$ here.

Summing the above inequalities with respect to k and dividing both sides by k, we get

$$\frac{1}{k}\sum_{j=1}^{k}\|x_j - T(x_j)\|^2 \leq \frac{1}{k}\left(2\sum_{j=1}^{k}\|x_j - T(x_j - v_j g_j)\|^2 + 2\sum_{j=1}^{k}v_j^2\right)$$

$$\leq \frac{1}{k}\left(2M + 2\sum_{j=1}^{\infty}v_j^2\right).$$

This completes the proof. □

We discussed two convergence analyses for Algorithm 1 with the diminishing step-size rule. By placing assumptions on the problem to be solved, we can also prove finite convergence. The following proposition describes the requirements to obtain an optimal solution in a finite number of iterations.

Proposition Appendix G.5. *Let* $\{x_k\} \subset H$ *be a sequence generated by Algorithm 1, and suppose that the assumptions in Theorem 3.2 hold. Furthermore, assume that* $X^\star$ *has a nonempty interior and the sequence* $\{x_k\}$ *is contained inside* $\mathrm{Fix}(T)$. *Then,* $x_k \in X^\star$ *for some* $k \in \mathbb{N}$.

Proof. We will proceed by way of contradiction and suppose that the conclusion does not hold, that is, $f_\star < f(x_k)$ for all $k \in \mathbb{N}$. Fix $x^\star \in X^\star$ arbitrarily. We deduce the following from inequality (C.1):

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - 2v_k(1 - \alpha_k)\langle g_k, x_k - x^\star\rangle + (1 - \alpha_k)v_k^2$$

$$\leq \|x_1 - x^\star\|^2 - 2\sum_{j=1}^{k}v_j(1 - \alpha_j)\langle g_j, x_j - x^\star\rangle + \sum_{j=1}^{k}(1 - \alpha_j)v_j^2$$

$$\leq \|x_1 - x^\star\|^2 - 2\left(\min_{j=1,2,\ldots,k}\langle g_j, x_j - x^\star\rangle\right)\sum_{j=1}^{k}(1 - \alpha_j)v_j + \sum_{j=1}^{k}v_j^2$$

for all $k \in \mathbb{N}$. The nonnegativity of the left side of the above inequality ensures that

$$\min_{j=1,2,\ldots,k}\langle g_j, x_j - x^\star\rangle \leq \frac{\|x_1 - x^\star\|^2}{2\sum_{j=1}^{k}(1 - \alpha_j)v_j} + \frac{\sum_{j=1}^{k}v_j^2}{2\sum_{j=1}^{k}(1 - \alpha_j)v_j} \qquad \text{(G.1)}$$

for all $k \in \mathbb{N}$. Now, there exists a positive real $\delta > 0$ which satisfies $\delta\mathbf{B} + x^\star \subset X^\star$ because $X^\star$ has a nonempty interior. Therefore, $\delta \leq \delta + \langle g_k, x_k - (x^\star +$

$\delta g_k)\rangle = \langle g_k, x_k - x^\star \rangle$ holds for any $k \in \mathbb{N}$. This property with inequality (G.1) implies

$$0 < \delta \leq \min_{j=1,2,\ldots,k} \langle g_j, x_j - x^\star \rangle$$

$$\leq \frac{\|x_1 - x^\star\|^2}{2\sum_{j=1}^{k}(1-\alpha_j)v_j} + \frac{\sum_{j=1}^{k} v_j^2}{2\sum_{j=1}^{k}(1-\alpha_j)v_j}$$

for all $k \in \mathbb{N}$. However, both terms of the right side of the above inequality converge to zero since $\limsup_{k\to\infty} \alpha_k < 1$, $\sum_{j=1}^{\infty} v_j = \infty$ and $\sum_{j=1}^{\infty} v_j^2 < \infty$. Therefore, we arrive at a contradiction. This completes the proof. $\qquad\square$

The nonemptiness of the interior of minima appears in many interesting applications, such as surrogate relaxation of discrete programming problems [7, 11]. When we construct a nonexpansive mapping that transforms a given point into a fixed point of itself (an example of such a mapping is a metric projection, but notice that the assumption of this sentence is not limited to it) and give a fixed point of the mapping as the initial point to the algorithm, the generated sequence is contained within the fixed point set of the mapping due to its convexity. Therefore, Proposition Appendix G.5 can be applied to these situations.