Noname manuscript No.
(will be inserted by the editor)

Acceleration Method for Convex Optimization over the Fixed Point Set of a Nonexpansive Mapping

Hideaki Iiduka

Received: date / Accepted: date

Abstract The existing algorithms for solving the convex minimization problem over the fixed point set of a nonexpansive mapping on a Hilbert space are based on algorithmic methods, such as the steepest descent method and conjugate gradient methods, for finding a minimizer of the objective function over the whole space, and attach importance to minimizing the objective function as quickly as possible. Meanwhile, it is of practical importance to devise algorithms which converge in the fixed point set quickly because the fixed point set is the set with the constraint conditions that must be satisfied in the problem. This paper proposes an algorithm which not only minimizes the objective function quickly but also converges in the fixed point set much faster than the existing algorithms and proves that the algorithm with diminishing step-size sequences strongly converges to the solution to the convex minimization problem. We also analyze the proposed algorithm with each of the Fletcher–Reeves, Polak–Ribiére–Polyak, Hestenes–Stiefel, and Dai–Yuan formulas used in the conventional conjugate gradient methods, and show that there is an inconvenient possibility that their algorithms may not converge to the solution to the convex minimization problem. We numerically compare the proposed algorithm with the existing algorithms and show its effectiveness and fast convergence.

Keywords convex optimization · fixed point set · nonexpansive mapping · conjugate gradient method · three-term conjugate gradient method · fixed point optimization algorithm

This work was supported by the Japan Society for the Promotion of Science through a Grant-in-Aid for Young Scientists (B) (23760077), and in part by the Japan Society for the Promotion of Science through a Grant-in-Aid for Scientific Research (C) (22540175).

H. Iiduka
Department of Computer Science, Meiji University
1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa, 214-8571 Japan
E-mail: iiduka@cs.meiji.ac.jp

Mathematics Subject Classification (2000) 47H07 · 47H09 · 65K05 · 65K10 · 90C25 · 90C30 · 90C52

1 Introduction

This paper discusses the following convex optimization problem over the fixed point set of a nonexpansive mapping [31]: given a convex, continuously Fréchet differentiable functional, f, on a real Hilbert space, H, and a nonexpansive mapping, N, from H into itself, which has its fixed point (i.e., $\mathrm{Fix}(N) := \{x \in H : N(x) = x\} \neq \emptyset$),

$$\text{minimize } f(x) \text{ subject to } x \in \mathrm{Fix}(N). \tag{1}$$

Problem (1) includes practical problems such as signal recovery [8], beamforming [27], and bandwidth allocation [17,19]. In particular, it plays an important role when the constraint set composed of the absolute set and the subsidiary sets is not feasible [17,19]. When we consider an optimization problem, including the problem in [17,19], it would be reasonable to deal with a constraint set in the problem as a subset [9, Section I, Framework 2], [31, Definition 4.1] of the absolute set with the elements closest to the subsidiary sets in terms of the norm. Here, we formulate a compromise solution to the problem by using the minimizer of the objective function over this subset. Since the subset can be expressed as the fixed point set of a certain nonexpansive mapping [31, Proposition 4.2], the minimization problem over the subset can be formulated as Problem (1).

We shall review the existing algorithms, called *fixed point optimization algorithms*, for solving Problem (1) when the gradient of f, denoted by $\nabla f : H \to H$, is strongly monotone and Lipschitz continuous. The first algorithm developed for solving Problem (1) is the *hybrid steepest descent method* (HSDM) [31,32]: $x_0 \in H$, $d_0^f := -\nabla f(x_0)$,

$$\begin{aligned} x_{n+1} &:= N\left(x_n + \mu \alpha_n d_n^f\right), \\ d_{n+1}^f &:= -\nabla f\left(x_{n+1}\right), \end{aligned} \tag{2}$$

for each $n \in \mathbb{N}$, where $\mu > 0$ and $(\alpha_n)_{n\in\mathbb{N}}$ is a sequence with $\lim_{n\to\infty} \alpha_n = 0$ and $\sum_{n=0}^{\infty} \alpha_n = \infty$. HSDM strongly converges to the unique solution to Problem (1) [32, Theorem 2.15, Remark 2.17 (a)]. Reference [8] proposed an effective algorithm, called the block-iterative surrogate constraint splitting method, to accelerate HSDM. The method in [8] converges strongly to the solution to Problem (1) without using diminishing sequences.

The *conjugate gradient methods* [25, Chapter 5] and *three-term conjugate gradient methods* [7,23,34–36] are the most popular methods that can accelerate the steepest descent method (i.e., $x_{n+1} := x_n - \alpha_n \nabla f(x_n)$) for large-scale unconstrained optimization problems. The search directions of the conjugate gradient method and three-term conjugate gradient method are as follows: for each $n \in \mathbb{N}$,

$$d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n^{(1)} d_n^f, \tag{3}$$

$$d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n^{(1)} d_n^f - \delta_n^{(2)} z_n, \tag{4}$$

where $(\delta_n^{(i)})_{n\in\mathbb{N}} \subset [0,\infty)$ $(i = 1,2)$ and $z_n \in H$ $(n \in \mathbb{N})$ is an arbitrary point. In general, the conjugate gradient method (i.e, $x_{n+1} := x_n + \alpha_n d_n^f$ with d_n^f defined by Equation (3)) does not generate the descent search direction,[1] which means that it does not always decrease f at each iteration. We need to set $\delta_n^{(1)}$ appropriately to ensure that $(d_n^f)_{n\in\mathbb{N}}$ defined by Equation (3) is the descent search direction. Meanwhile, the three-term conjugate gradient method (i.e., $x_{n+1} := x_n + \alpha_n d_n^f$ with d_n^f defined by Equation (4)) generates the descent search direction [23, Subsection 2.1] without depending on the choice of $\delta_n^{(1)}$ (see Footnote 2 for the well-known formulas of $\delta_n^{(1)}$). This is because the third term $\delta_n^{(2)} z_n$ in Equation (4) plays a role in generating the descent search direction (see [23, Subsection 2.1] and Subsection 3.1). On the basis of such acceleration methods for the steepest descent method, references [20] and [15] presented algorithms that respectively use Equation (2) and Equations (3) and (4) to solve Problem (1). The algorithm with Equations (2) and (3) and the algorithm with Equations (2) and (4) are referred to here as the *hybrid conjugate gradient method* (HCGM) and the *hybrid three-term conjugate gradient method* (HTCGM), respectively. HCGM and HTCGM converge strongly to the solution to Problem (1) when $\lim_{n\to\infty} \delta_n^{(i)} = 0$ $(i = 1,2)$ and $(z_n)_{n\in\mathbb{N}}$ is bounded [20, Theorem 4.1], [15, Theorem 7]. Here, we remark that the conjugate gradient methods with the well-known formulas, such as the Fletcher–Reeves (FR), Polak–Ribiére–Polyak (PRP), Hestenes–Stiefel (HS), and Dai–Yuan (DY) formulas[2], can solve unconstrained optimization problems without assuming $\lim_{n\to\infty} \delta_n = 0$. To distinguish between the conventional conjugate gradient directions with the four formulas and the directions defined by Equations (3) and (4) with $\lim_{n\to\infty} \delta_n^{(i)} = 0$, we call the latter the conjugate gradient-like directions. The numerical examples in [15,20] show that HCGM and HTCGM with slowly diminishing sequences $\delta_n^{(i)}$s (e.g., $\delta_n^{(i)} := 1/(n+1)^{0.01}$ $(i = 1,2)$) converge to the solution faster than HSDM, and that, in particular, HTCGM converges fastest.

The main advantage of fixed point optimization algorithms, such as HCGM and HTCGM, with conjugate gradient-like directions is to enable f to be decreased much faster than HSDM with the steepest descent direction. Meanwhile, the rates of convergence of the distance $d(x_n, \mathrm{Fix}(N)) := \inf_{x\in\mathrm{Fix}(N)} \|x_n - x\|$ to 0 are the same for all three algorithms because these algorithms each iterate as $x_{n+1} := N(x_n + \mu\alpha_n d_n^f)$ $(n \in \mathbb{N})$. Here, we shall discuss Problem

[1] $(d_n^f)_{n\in\mathbb{N}}$ is referred to as a descent search direction if $\langle d_n^f, \nabla f(x_n)\rangle < 0$ for all $n \in \mathbb{N}$.

[2] These are defined as follows: $\delta_n^{\mathrm{FR}} := \|\nabla f(x_{n+1})\|^2/\|\nabla f(x_n)\|^2$, $\delta_n^{\mathrm{PRP}} := v_n/\|\nabla f(x_n)\|^2$, $\delta_n^{\mathrm{HS}} := v_n/u_n$, $\delta_n^{\mathrm{DY}} := \|\nabla f(x_{n+1})\|^2/u_n$, where $u_n := \langle d_n^f, \nabla f(x_{n+1}) - \nabla f(x_n)\rangle$ and $v_n := \langle \nabla f(x_{n+1}), \nabla f(x_{n+1}) - \nabla f(x_n)\rangle$.

(1) when $\mathrm{Fix}(N)$ is the set of all minimizers of a convex, continuously Fréchet differentiable functional, g, over H and see that $x_{n+1} := N(x_n + \mu\alpha_n d_n^f)$ $(n \in \mathbb{N})$ is based on the steepest descent method to minimize g over H. Suppose that $\nabla g \colon H \to H$ is Lipschitz continuous with a constant $l > 0$ and define $N_g \colon H \to H$ by $N_g := I - \alpha \nabla g$, where $\alpha \in (0, 2/l]$ and $I \colon H \to H$ stands for the identity mapping. Accordingly, N_g satisfies the nonexpansivity condition and $\mathrm{Fix}(N_g) = \mathrm{Argmin}_{x \in H}\, g(x)$ (see, e.g., [14, Proposition 2.3]). Therefore, Equation (2) with $N_g := I - \alpha\nabla g$ is as follows:

$$
\begin{aligned}
y_n &:= x_n + \mu\alpha_n d_n^f,\\
x_{n+1} &:= N_g\left(x_n + \mu\alpha_n d_n^f\right)\\
&= [I - \alpha\nabla g]\,(y_n) = y_n + \alpha\,[-\nabla g(y_n)]\\
&= y_n + \alpha\left[\frac{N_g(y_n) - y_n}{\alpha}\right].
\end{aligned} \tag{5}
$$

Hence, Algorithm (5) has the steepest descent direction,

$$
d_{n+1}^{N_g} := -\nabla g(y_n) = \frac{N_g(y_n) - y_n}{\alpha},
$$

which implies it converges slowly in the constraint set, $\mathrm{Fix}(N_g)$. From such a viewpoint, one can expect that an algorithm with the three-term conjugate gradient direction,

$$
d_{n+1}^{N_g} := \frac{N_g(y_n) - y_n}{\alpha} + \beta_n^{(1)} d_n^{N_g} + \beta_n^{(2)} w_n, \tag{6}
$$

where $\beta_n^{(i)} \in \mathbb{R}$ $(i = 1, 2)$ and $w_n \in H$, would converge in the constraint set faster than Algorithm (5).

In this paper, we present an algorithm with both Direction (4) to accelerate the objective function minimization and Direction (6) to accelerate the search for a fixed point of a nonexpansive mapping. We also present its convergence analysis.

This paper is organized as follows. Section 2 gives mathematical preliminaries. Section 3 presents the fixed point optimization algorithm with Directions (4) and (6) to accelerate the existing algorithms and proves that the algorithm with $\lim_{n\to\infty} \delta_n^{(i)} = 0$ and $\lim_{n\to\infty} \beta_n^{(i)} = 0$ $(i = 1, 2)$ strongly converges to the solution to Problem 3.1. It also proves that HCGM with each of the FR, PRP, HS, and DY formulas (i.e., the algorithm with Equations (2) and (3) when $\delta_n^{(1)}$ is defined by one of $\delta_n^{\mathrm{FR}}, \delta_n^{\mathrm{PRP}}, \delta_n^{\mathrm{HS}}$, and δ_n^{DY}) does not satisfy $\lim_{n\to\infty} \delta_n^{(1)} = 0$ when the unique minimizer of the objective function over the whole space is not in the fixed point set of a nonexpansive mapping. This implies that there is an inconvenient possibility that HCGMs with the four formulas may not converge strongly to the unique minimizer of the objective function over the fixed point set that is not equal to the unique minimizer of the objective function over the whole space. Section 4 provides numerical comparisons of the proposed algorithm with the existing fixed point optimization algorithms and shows its effectiveness. It also describes examples such that HCGMs with the four formulas do not always converge to the solution. Section 5 concludes the paper by summarizing its key points and mentions future subjects for development of the proposed algorithm.

2 Mathematical Preliminaries

Let H be a real Hilbert space with inner product, $\langle \cdot, \cdot \rangle$, and its induced norm, $\|\cdot\|$, and let $\mathbb{N}$ be the set of zero and all positive integers; i.e., $\mathbb{N} := \{0, 1, 2, \ldots\}$. We denote the identity mapping on H by I; i.e., $I(x) := x$ for all $x \in H$.

2.1 Convexity, monotonicity, and nonexpansivity

A function, $f \colon H \to \mathbb{R}$, is said to be *convex* if, for any $x, y \in H$ and for any $\lambda \in [0,1]$, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. In particular, a convex function, $f \colon H \to \mathbb{R}$, is said to be *strongly convex* with $c > 0$ (c-strongly convex) if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - (c\lambda(1-\lambda)/2)\|x-y\|^2$ for all $x, y \in H$ and for all $\lambda \in [0,1]$.

$A \colon H \to H$ is referred to as a *monotone* operator if $\langle x-y, A(x)-A(y)\rangle \geq 0$ for all $x, y \in H$. $A \colon H \to H$ is said to be *strongly monotone* with $c > 0$ (c-strongly monotone) if $\langle x-y, A(x)-A(y)\rangle \geq c\|x-y\|^2$ for all $x, y \in H$. Let $f \colon H \to \mathbb{R}$ be a Fréchet differentiable function. If f is convex (resp. c-strongly convex), ∇f is monotone (resp. c-strongly monotone) [4, Example 22.3].

A mapping, $A \colon H \to H$, is said to be *Lipschitz continuous* with $L > 0$ (L-Lipschitz continuous) if $\|A(x) - A(y)\| \leq L\|x-y\|$ for all $x, y \in H$. When $N \colon H \to H$ is 1-Lipschitz continuous, N is referred to as a *nonexpansive* mapping [3,12,13,29]. In particular, N is said to be *firmly nonexpansive* if $\|N(x)-N(y)\|^2 \leq \langle x-y, N(x)-N(y)\rangle$ for all $x, y \in H$. The Cauchy-Schwarz inequality guarantees that any firmly nonexpansive mapping satisfies the nonexpansivity condition. We denote the *fixed point set* of $N \colon H \to H$ by

$$\mathrm{Fix}(N) := \{x \in H \colon N(x) = x\}.$$

$\mathrm{Fix}(N)$ satisfies closedness and convexity properties when N is nonexpansive [13, Proposition 5.3].

Given a nonempty, closed convex set, C $(\subset H)$, the mapping that assigns every point, $x \in H$, to its unique nearest point in C is called the *metric projection* onto C and is denoted by P_C; i.e., $P_C(x) \in C$ and $\|x - P_C(x)\| = \inf_{y \in C} \|x-y\|$. The metric projection, P_C, satisfies the firm nonexpansivity condition with $\mathrm{Fix}(P_C) = C$ [3, Facts 1.5], [28, Theorem 2.4-1 (ii)], [4, Proposition 4.8, Equation (4.8)]. Some closed convex set, C, for example, a linear variety, a closed ball, a closed cone, or a closed polytope, is simple in the sense that the explicit form of P_C is known, which implies that P_C can be computed within a finite number of arithmetic operations [4, Subchapter 28.3], [30].

The following lemmas will be used to prove the main theorem.

Lemma 2.1 (Lemma 3.1 in [31]) *Suppose that* $f\colon H \to \mathbb{R}$ *is* c*-strongly convex and Fréchet differentiable,* $\nabla f\colon H \to H$ *is* L*-Lipschitz continuous, and* $\mu \in (0, 2c/L^2)$*. Define* $T\colon H \to H$ *by* $T(x) := x - \mu\alpha\nabla f(x)$ $(x \in H)$*, where* $\alpha \in [0,1]$*. Then, for all* $x, y \in H$*,* $\|T(x) - T(y)\| \leq (1 - \tau\alpha)\|x - y\|$*, where* $\tau := 1 - \sqrt{1 - \mu(2c - \mu L^2)} \in (0,1]$*.*

Lemma 2.2 (Theorems 3.7 and 3.9 in [1]) *Suppose that* $N_1\colon H \to H$ *is firmly nonexpansive and* $N_2\colon H \to H$ *is nonexpansive with* $\mathrm{Fix}(N_1) \cap \mathrm{Fix}(N_2) \neq \emptyset$*, and* $(x_n)_{n\in\mathbb{N}}$ $(\subset H)$ *is bounded. Then,* $\lim_{n\to\infty} \|x_n - N_1(N_2(x_n))\| = 0$ *if and only if* $\lim_{n\to\infty} \|x_n - N_1(x_n)\| = 0$ *and* $\lim_{n\to\infty} \|x_n - N_2(x_n)\| = 0$*.*

2.2 Monotone variational inequality

The *variational inequality* problem [11,22] for a monotone operator, $A\colon H \to H$, over a closed convex set, C $(\subset H)$, is to find a point in

$$\mathrm{VI}(C, A) := \{x^\star \in C\colon \langle x - x^\star, A(x^\star)\rangle \geq 0 \text{ for all } x \in C\}.$$

Suppose that $f\colon H \to \mathbb{R}$ is c-strongly convex and Fréchet differentiable, and $\nabla f\colon H \to H$ is L-Lipschitz continuous. Then, $\mathrm{VI}(C, \nabla f)$ can be characterized as the set of all minimizers of f over C, which coincides with the fixed point set of $P_C(I - \alpha\nabla f)$ [6, Subsection 8.3], [11, Proposition 2.1], [33, Theorem 46.C (1) and (2)]:

$$\begin{aligned}\mathrm{VI}(C, \nabla f) &= \operatorname*{Argmin}_{x\in C} f(x) := \left\{x^\star \in C\colon f(x^\star) = \min_{x\in C} f(x)\right\} \\ &= \mathrm{Fix}\left(P_C\left(I - \alpha\nabla f\right)\right) := \left\{x^\star \in C\colon P_C\left(x^\star - \alpha\nabla f(x^\star)\right) = x^\star\right\},\end{aligned}$$

where α is an arbitrary positive real number. Since $P_C(I - \hat{\alpha}\nabla f)$ is a contraction mapping when $\hat{\alpha} \in (0, 2c/L^2)$, $P_C(I - \hat{\alpha}\nabla f)$ has a unique fixed point [12, Theorem 2.1]. Therefore, the solution to the variational inequality consists of one point.

3 Optimization over the Fixed Point Set

This section discusses the following problem:

Problem 3.1 *Under the assumptions that*

(A1) $N\colon H \to H$ *is a nonexpansive mapping with* $\mathrm{Fix}(N) \neq \emptyset$*,*

(A2) K $(\subset H)$ *is a nonempty, bounded, closed convex set onto which the metric projection is computable, and* $\mathrm{Fix}(N) \subset K$*,*[3]

[3] For example, when there is a bound on $\mathrm{Fix}(N)$, we can choose K as a closed ball with a large radius containing $\mathrm{Fix}(N)$. The metric projection onto such a K is easily computed (see also Subsection 2.1). See the final paragraph in Subsection 3.1 for a discussion of Problem 3.1 when a bound on $\mathrm{Fix}(N)$ either does not exist or is not known.

(A3) $f\colon H \to \mathbb{R}$ *is c-strongly convex and Fréchet differentiable, and* $\nabla f\colon H \to H$ *is L-Lipschitz continuous,*

$$\textit{minimize } f(x) \textit{ subject to } x \in \mathrm{Fix}(N).$$

From the closedness and convexity of $\mathrm{Fix}(N)$ and the discussion in Subsection 2.2, we get the following:

Proposition 3.1 *The existence and uniqueness of the solution to Problem 3.1 is guaranteed.*

3.1 Acceleration method for the convex optimization problem over the fixed point set

We present the following algorithm for solving Problem 3.1:

Algorithm 3.1

Step 0. Take $(\alpha_n)_{n\in\mathbb{N}}$, $(\beta_n^{(i)})_{n\in\mathbb{N}}$, $(\delta_n^{(i)})_{n\in\mathbb{N}} \subset (0,1]$ $(i=1,2)$, $\gamma \in (0,1]$, *and* $\mu > 0$, *choose* $x_0 \in H$ *arbitrarily, and let* $d_0^f := -\nabla f(x_0)$, $y_0 := x_0 + \mu\alpha_0 d_0^f$, $d_0^N := N(y_0) - y_0$, *and* $n := 0$.

Step 1. Given $x_n, d_n^f \in H$, *compute* $y_n \in H$ *as*

$$y_n := P_K\left(x_n + \mu\alpha_n d_n^f\right).$$

Compute $d_{n+1}^N \in H$ *as*

$$d_{n+1}^N := N(y_n) - y_n + \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n, \tag{7}$$

where $w_n \in H$ *is an arbitrary point.*

Step 2. Compute $x_{n+1} \in H$ *as*

$$x_{n+1} := P_K\left(y_n + \gamma d_{n+1}^N\right)$$

and update $d_{n+1}^f \in H$ *by*

$$d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n^{(1)} d_n^f - \delta_n^{(2)} z_n, \tag{8}$$

where $z_n \in H$ *is an arbitrary point. Put* $n := n+1$, *and go to Step 1.*

In unconstrained optimization problems, it is desirable to use iterative methods which generate descent search directions. This is because such methods can decrease strictly the objective function at each iteration. Generally, it is not guaranteed that the conjugate gradient method defined by $x_{n+1} := x_n + \alpha_n d_n^f$ and Equation (8) with $\delta_n^{(2)} z_n := 0$ generates the descent search direction.[4] The three-term conjugate gradient method defined by $x_{n+1} := x_n + \alpha_n d_n^f$

[4] The conjugate gradient method with the DY formula (i.e., $\delta_n^{(1)} := \delta_n^{\mathrm{DY}}$) generates the descent search direction under the Wolfe conditions [10]. Whether or not the conjugate gradient methods generate descent search directions depends on the choices of $\delta_n^{(1)}$ and α_n.

and Equation (8) with $\delta_n^{(2)} z_n \neq 0$ generates the descent search direction without depending on the choices of $\delta_n^{(1)}$ and α_n [23, Subsection 2.1]. [5] Therefore, it would be useful in Problem 3.1 to use an accelerated algorithm with Direction (8) when $\delta_n^{(2)} z_n \neq 0$. On the other hand, the discussion on Equation (5) describes that $N(y_n) - y_n$ is expressed as the steepest descent direction at y_n of a certain convex function of which a minimizer is a fixed point of N. Accordingly, Direction (7) is the three-term conjugate gradient direction for finding a fixed point of N. Hence, one can expect that an algorithm with Direction (7) when $\beta_n^{(2)} w_n \neq 0$ would converge in $\mathrm{Fix}(N)$ quickly (see also Section 1).

Let us compare Algorithm 3.1 with the existing algorithms, such as HSDM [32, Theorem 2.15, Remark 2.17 (a)], HCGM [20, Algorithm 3.4], and HTCGM [15, Algorithm 6], for solving Problem 3.1. HTCGM is as follows (see also Equations (2) and (4)): $x_0 \in H$, $d_0^f := -\nabla f(x_0)$, and

$$\begin{cases} x_{n+1} := N\left(x_n + \mu \alpha_n d_n^f\right), \\ d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n^{(1)} d_n^f - \delta_n^{(2)} z_n \ (n \in \mathbb{N}). \end{cases} \tag{9}$$

Algorithm (9) with $\delta_n^{(i)} := 0$ ($i = 1, 2$, $n \in \mathbb{N}$) coincides with HSDM, and Algorithm (9) with $\delta_n^{(2)} := 0$ ($n \in \mathbb{N}$) coincides with HCGM. Hence, the existing algorithms can be expressed as Algorithm (9). Algorithm 3.1 with $K := H$, $\gamma := 1$, and $\beta_n^{(i)} := 0$ ($i = 1, 2$) has $x_{n+1} = y_n + d_{n+1}^N = N(y_n) = N(x_n + \mu \alpha_n d_n^f)$, which means that Algorithm 3.1 in this case coincides with Algorithm (9). Algorithm 3.1 uses $d_{n+1}^N := N(y_n) - y_n + \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n$ to converge in $\mathrm{Fix}(N)$ faster than Algorithm (9), as discussed in Section 1.

The following theorem constitutes the convergence analysis of Algorithm 3.1. The proof of the theorem is given in Subsection 3.3.

Theorem 3.1 *Suppose that* (I) $\mu \in (0, 2c/L^2)$, (II) $(w_n)_{n \in \mathbb{N}}$ *and* $(z_n)_{n \in \mathbb{N}}$ *are bounded,*[6] *and* (III) $(\alpha_n)_{n \in \mathbb{N}}$, $(\beta_n^{(i)})_{n \in \mathbb{N}}$, *and* $(\delta_n^{(i)})_{n \in \mathbb{N}}$ $(i = 1, 2)$ *are sequences in* $(0, 1]$ *satisfying* (i) $\lim_{n \to \infty} \alpha_n = 0$, (ii) $\sum_{n=0}^{\infty} \alpha_n = \infty$, (iii) $\sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n| < \infty$, (iv) $\beta_n^{(i)} \leq \alpha_n^2$ $(i = 1, 2, n \in \mathbb{N})$, *and* (v) $\lim_{n \to \infty} \delta_n^{(i)} = 0$ $(i = 1, 2)$. *Then,* $(x_n)_{n \in \mathbb{N}}$ *in Algorithm 3.1 strongly converges to the unique solution to Problem 3.1.*

Let us compare Theorem 3.1 with the previously reported results in [32, Theorem 2.15, Remark 2.17 (a)], [20, Theorem 4.1], and [15, Theorem 7]. HSDM (i.e., Algorithm (9) with $\delta_n^{(i)} := 0$ $(i = 1, 2)$) with Conditions (I), (i), (ii), and (iii) in Theorem 3.1 converges strongly to the solution to Problem 3.1

[5] Reference [23, Subsection 2.1] showed that $x_{n+1} := x_n + \alpha_n d_n^f$ and $d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n^{(1)} d_n^f - \delta_n^{(2)} z_n$, where $\alpha_n, \delta_n^{(1)}$ (> 0) are arbitrary, z_n $(\in \mathbb{R}^N)$ is any vector, and $\delta_n^{(2)} := \delta_n^{(1)} (\langle \nabla f(x_{n+1}), d_n \rangle / \langle \nabla f(x_{n+1}), z_n \rangle)$, satisfy $\langle d_n^f, \nabla f(x_n) \rangle = -\|\nabla f(x_n)\|^2$ $(n \in \mathbb{N})$.

[6] We can choose, for example, $w_n := N(y_n) - y_n$ and $z_n := \nabla f(x_{n+1})$ $(n \in \mathbb{N})$ by referring to [35] and [15, Section 3]. Lemma 3.1 ensures that they are bounded.

[32, Theorem 2.15, Remark 2.17 (a)]. HCGM (i.e., Algorithm (9) with $\delta_n^{(2)} := 0$) with the conditions in Theorem 3.1 converges strongly to the solution if $(\nabla f(x_n))_{n\in\mathbb{N}}$ is bounded [20, Theorem 4.1]. Theorem 7 in [15] guarantees that, if $(\nabla f(x_n))_{n\in\mathbb{N}}$ is bounded, then HTCGM with the conditions in Theorem 3.1 converges strongly to the solution. The results in [20, Theorem 4.1] and [15, Theorem 7] and the proof of Theorem 3.1 lead us to a strong convergence of Algorithm 3.1 to the solution without assuming the boundedness of K if $(\nabla f(x_n))_{n\in\mathbb{N}}$ and $(N(y_n)-y_n)_{n\in\mathbb{N}}$ are bounded. However, it would be difficult to verify whether $(\nabla f(x_n))_{n\in\mathbb{N}}$ and $(N(y_n)-y_n)_{n\in\mathbb{N}}$ are bounded or not in advance. Hence, we assume the existence of a bounded K satisfying $\mathrm{Fix}(N) \subset K$ in place of the boundedness of $(\nabla f(x_n))_{n\in\mathbb{N}}$ and $(N(y_n)-y_n)_{n\in\mathbb{N}}$ (see Footnote 3 for the choice of K).

Let us consider the case where a bound on $\mathrm{Fix}(N)$ either does not exist or is not known. In this case, we cannot choose a bounded K satisfying $\mathrm{Fix}(N) \subset K$.[7] Even in the case, we can execute Algorithm 3.1 with $K = H$. However, we need to verify the boundedness of $(\nabla f(x_n))_{n\in\mathbb{N}}$ and $(N(y_n)-y_n)_{n\in\mathbb{N}}$ to guarantee that Algorithm 3.1 converges to the solution (see the above paragraph). When we try to apply HCGM and HTCGM to this case, we also need to verify whether or not $(\nabla f(x_n))_{n\in\mathbb{N}}$ is bounded [20, Theorem 4.1], [15, Theorem 7]. Meanwhile, we can apply HSDM to this case without any problem [32, Theorem 2.15, Remark 2.17 (a)]. Therefore, when a bound on $\mathrm{Fix}(N)$ either does not exist or is not known, we should execute HSDM. However, HSDM converges slowly. Hence, it would be desirable to execute HSDM, HCGM, HTCGM, and Algorithm 3.1 and verify whether the convergent point of HSDM that is the minimizer of f over $\mathrm{Fix}(N)$ is equal to the convergent points of HCGM, HTCGM, and Algorithm 3.1.

3.2 Analysis of Algorithm 3.1 with the conventional formulas of conjugate gradient directions

In this subsection, we analyze Algorithm 3.1 when $\delta_n^{(1)}$ is one of the well-known formulas, such as the Fletcher–Reeves (FR), Dai–Yuan (DY), Polak–Ribiére–Polyak (PRP), and Hestenes–Stiefel (HS) formulas, that are used to solve large-scale unconstrained optimization problems. Let us define the FR, DY, PRP, and HS formulas, which can be applied to constrained optimization

[7] Given a halfspace $S := \{x \in H \colon \langle a, x\rangle \leq b\}$, where $a\ (\neq 0) \in H$ and $b \in \mathbb{R}$, $N(x) := P_S(x) = x - [\max\{0, \langle a, x\rangle - b\}/\|a\|^2]a\ (x \in H)$ is nonexpansive with $\mathrm{Fix}(N) = \mathrm{Fix}(P_S) = S \neq \emptyset$ [3, p. 406], [4, Subchapter 28.3]. However, we cannot define a bounded K satisfying $\mathrm{Fix}(N) = S \subset K$.

problems, as follows: for each $n \in \mathbb{N}$,

$$\delta_n^{\mathrm{FR}} := \begin{cases} \dfrac{\|\nabla f(x_{n+1})\|^2}{\|\nabla f(x_n)\|^2} & \text{if } \|\nabla f(x_n)\| > 0, \\ 0 & \text{otherwise,} \end{cases}$$
$$\delta_n^{\mathrm{DY}} := \begin{cases} \dfrac{\|\nabla f(x_{n+1})\|^2}{\langle d_n^f, \nabla f(x_{n+1}) - (1+\eta)\nabla f(x_n)\rangle} & \text{if } u_n \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$
$$\delta_n^{\mathrm{PRP}} := \begin{cases} \dfrac{\langle \nabla f(x_{n+1}), \nabla f(x_{n+1}) - (1+\kappa)\nabla f(x_n)\rangle}{\|\nabla f(x_n)\|^2} & \text{if } \|\nabla f(x_n)\| > 0, \\ 0 & \text{otherwise,} \end{cases}$$
$$\delta_n^{\mathrm{HS}} := \begin{cases} \dfrac{\langle \nabla f(x_{n+1}), \nabla f(x_{n+1}) - (1+\kappa)\nabla f(x_n)\rangle}{\langle d_n^f, \nabla f(x_{n+1}) - (1+\eta)\nabla f(x_n)\rangle} & \text{if } u_n \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where $\eta, \kappa \geq 0$, and $u_n := \langle d_n^f, \nabla f(x_{n+1}) - (1+\eta)\nabla f(x_n)\rangle$ $(n \in \mathbb{N})$. For simplicity, we assume that $\delta_n^{(2)} := 0$ $(n \in \mathbb{N})$, i.e., d_n^f $(n \in \mathbb{N})$ in Algorithm 3.1 is defined by the conventional conjugate gradient direction:

$$d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n d_n^f \ (n \in \mathbb{N}),$$

where $\delta_n \in \mathbb{R}$ is defined as one of Formulas (10).

The following proposition is satisfied for Algorithm 3.1 with the conventional FR, DY, PRP, and HS Formulas (10):

Proposition 3.2 *Suppose that Conditions (I), (II), and (i)–(iv) in Theorem 3.1 are satisfied. Then, the following hold:*

(i) *If* $\lim_{n\to\infty} \delta_n^{\mathrm{FR}} = 0$*, then the unique minimizer of* f *over* H *belongs to* $\mathrm{Fix}(N)$ *(i.e.,* $\mathrm{Fix}(N) \cap \mathrm{Argmin}_{x\in H} f(x) \neq \emptyset$*). In this case,* $(x_n)_{n\in\mathbb{N}}$ *in Algorithm 3.1 strongly converges to the unique minimizer of* f *over* H*.*
(ii) *If* $\lim_{n\to\infty} \delta_n^{\mathrm{DY}} = 0$ *and if* $\eta \neq 0$*, then the unique minimizer of* f *over* H *belongs to* $\mathrm{Fix}(N)$*. In this case,* $(x_n)_{n\in\mathbb{N}}$ *in Algorithm 3.1 strongly converges to the unique minimizer of* f *over* H*.*
(iii) *If* $\lim_{n\to\infty} \delta_n^{\mathrm{PRP}} = 0$ *and if* $\kappa \neq 0$*, then the unique minimizer of* f *over* H *belongs to* $\mathrm{Fix}(N)$*. In this case,* $(x_n)_{n\in\mathbb{N}}$ *in Algorithm 3.1 strongly converges to the unique minimizer of* f *over* H*.*
(iv) *If* $\lim_{n\to\infty} \delta_n^{\mathrm{HS}} = 0$ *and if* $\eta, \kappa \neq 0$*, then the unique minimizer of* f *over* H *belongs to* $\mathrm{Fix}(N)$*. In this case,* $(x_n)_{n\in\mathbb{N}}$ *in Algorithm 3.1 strongly converges to the unique minimizer of* f *over* H*.*

Let us discuss Proposition 3.2 for Algorithm 3.1 with $\gamma := 1$, $\beta_n^{(i)} := 0$ $(i = 1, 2)$, and $\delta_n := \delta_n^{(1)}$ defined by one of Formulas (10), i.e.,

$$\begin{cases} x_{n+1} := P_K\left[N\left(P_K\left(x_n + \mu\alpha_n d_n^f\right)\right)\right], \\ d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n d_n^f \ (n \in \mathbb{N}). \end{cases} \tag{11}$$

Proposition 3.2 says that, in the case of $\eta, \kappa \neq 0$, if Algorithm (11) with δ_n defined by one of Formulas (10) satisfies $\lim_{n\to\infty} \delta_n = 0$, then the unique minimizer of f over H is always in $\mathrm{Fix}(N)$, and the algorithm strongly converges to the unique minimizer of f over H belonging to $\mathrm{Fix}(N)$.

Proposition 3.2 describes that Algorithm (11) satisfies $\lim_{n\to\infty} \delta_n \neq 0$ in the case where $\eta, \kappa \neq 0$ and the unique minimizer of f over H is not in $\mathrm{Fix}(N)$. According to Theorem 3.1 (or [20, Theorem 4.1]), Algorithm (11) in this case might not converge to the unique minimizer of f over $\mathrm{Fix}(N)$ that is not equal to the unique minimizer of f over H. To guarantee that Algorithm (11) converges in this case to the unique minimizer of f over $\mathrm{Fix}(N)$, for example, we need to reset $\delta_n := 0$ when n exceeds a certain number of iterations. Algorithm (11) with the steepest descent direction ($\delta_n := 0$), i.e., $x_{n+1} := P_K[N(P_K(x_n - \mu\alpha_n \nabla f(x_n)))]$ ($n \in \mathbb{N}$) (HSDM [32]), strongly converges to the unique minimizer of f over $\mathrm{Fix}(N)$ [32, Theorem 2.15, Remark 2.17 (a)], however, it converges slowly. The above observation suggests that Algorithm (11) with each of the conventional formulas would not be an efficient way to solve constrained convex optimization problems. Numerical examples in Section 4 show that the algorithms with the conventional formulas do not always converge to the unique minimizer of f over $\mathrm{Fix}(N)$.

Meanwhile, Algorithm (11) with $\lim_{n\to\infty} \delta_n = 0$ (e.g., $\delta_n := 1/(n+1)^a$ ($a > 0$)) always converges to the unique minimizer of f over $\mathrm{Fix}(N)$ [20, Theorem 4.1], which means there is no need to verify whether the unique minimizer of f over H is in $\mathrm{Fix}(N)$ or not in advance, and converges faster than HSDM (see [20] for details on the fast convergence of Algorithm (11) with $\lim_{n\to\infty} \delta_n = 0$).

In the case that the unique minimizer of f over H is not in $\mathrm{Fix}(N)$ and $\lim_{n\to\infty} \delta_n^{\mathrm{DY}} = 0$ (or $\lim_{n\to\infty} \delta_n^{\mathrm{HS}} = 0$ and $\kappa \neq 0$), we get $\eta = 0$ from Proposition 3.2 (ii) and (iv). Hence, Inequality (14) (see the proof of Proposition 3.2 (ii) and (iv) and Remark 3.1) imply that

$$\|\nabla f(x_{n+1})\| \geq \|\nabla f(x_n)\| \text{ for large enough } n.$$

Since $\|\nabla f(x_n)\|$ tends to be smaller at the unique minimizer of f over $\mathrm{Fix}(N)$, Algorithm (11) with $\delta_n := \delta_n^{\mathrm{DY}}$ (or δ_n^{HS}) will not converge to the unique minimizer of f over $\mathrm{Fix}(N)$ when the unique minimizer of f over H is not in $\mathrm{Fix}(N)$.

Proof of Proposition 3.2. Let $x^\star \in H$ be the unique minimizer of f over H.

(i) The boundedness of K ensures that $(x_n)_{n\in\mathbb{N}}$ is bounded. The Lipschitz continuity of ∇f guarantees that $\|\nabla f(x_n) - \nabla f(x^\star)\| \leq L\|x_n - x^\star\|$ for all $n \in \mathbb{N}$, which implies that $(\nabla f(x_n))_{n\in\mathbb{N}}$ is bounded. Hence, $B_1 > 0$ exists such that

$$\|\nabla f(x_n)\| \leq B_1 \text{ for all } n \in \mathbb{N}. \tag{12}$$

Assume that $x^\star \notin \mathrm{Fix}(N)$. We then can choose $\varepsilon_1, \varepsilon_2 > 0$ such that

$$d(x^\star, \mathrm{Fix}(N)) := \inf\{\|x^\star - y\| : y \in \mathrm{Fix}(N)\} \geq \varepsilon_1 + \varepsilon_2.$$

Theorem 3.1 and $\lim_{n\to\infty} \delta_n^{\mathrm{FR}} = 0$ guarantee that $(x_n)_{n\in\mathbb{N}}$ strongly converges to the unique minimizer of f over $\mathrm{Fix}(N)$, denoted by $x^* \in \mathrm{Fix}(N)$. Hence, for $\varepsilon_1 > 0$, there exists $N_0 \in \mathbb{N}$ such that, for all $n \geq N_0$,

$$d\left(x_n, \mathrm{Fix}(N)\right) := \inf\left\{\|x_n - y\| : y \in \mathrm{Fix}(N)\right\} \leq \|x_n - x^*\| \leq \varepsilon_1.$$

Fix $n \geq N_0$ arbitrarily. Then, $y(n) \in \mathrm{Fix}(N)$ exists such that $\|x_n - y(n)\| = d(x_n, \mathrm{Fix}(N))$. Hence,

$$\begin{aligned}\varepsilon_2 &\leq d\left(x^\star, \mathrm{Fix}(N)\right) - \varepsilon_1 \leq d\left(x^\star, \mathrm{Fix}(N)\right) - d\left(x_n, \mathrm{Fix}(N)\right)\\ &= \inf\left\{\|x^\star - y\| : y \in \mathrm{Fix}(N)\right\} - \|x_n - y(n)\|\\ &\leq \|x^\star - y(n)\| - \|x_n - y(n)\| \leq \|x^\star - x_n\|.\end{aligned}$$

Since the c-strong monotonicity of ∇f implies $(\nabla f)^{-1}$ is $1/c$-Lipschitz continuous, we find that, for all $x \in H$, $\|(\nabla f)^{-1}(0) - (\nabla f)^{-1}(x)\| \leq (1/c)\|x\|$, and hence,

$$\varepsilon_2 \leq \|x^\star - x_n\| = \left\|(\nabla f)^{-1}(0) - (\nabla f)^{-1}\left(\nabla f(x_n)\right)\right\| \leq \frac{1}{c}\|\nabla f(x_n)\|.$$

Therefore, we have

$$\|\nabla f(x_n)\| \geq c\varepsilon_2 =: B_2 \text{ for all } n \geq N_0. \tag{13}$$

Inequalities (12) and (13) ensure that, for all $n \geq N_0$,

$$\delta_n^{\mathrm{FR}} = \frac{\|\nabla f(x_{n+1})\|^2}{\|\nabla f(x_n)\|^2} \geq \left(\frac{B_2}{B_1}\right)^2 > 0.$$

This contradicts $\lim_{n\to\infty} \delta_n^{\mathrm{FR}} = 0$. Hence, we find that $\{x^\star\} = \mathrm{Argmin}_{x\in H} f(x) \subset \mathrm{Fix}(N)$. Moreover, since $x^\star$ is the solution to Problem 3.1, Theorem 3.1 guarantees that $(x_n)_{n\in\mathbb{N}}$ strongly converges to $x^\star$.

(ii) Assume that $x^\star \notin \mathrm{Fix}(N)$. Proposition 3.2 (i) guarantees that Inequalities (12) and (13) hold for all $n \geq N_0$. Since $\lim_{n\to\infty} \delta_n^{\mathrm{DY}} = 0$, there exists $N_1 \in \mathbb{N}$ such that $\delta_n^{\mathrm{DY}} \leq 1/2$ for all $n \geq N_1$. Put $B := \max\{B_1, \|d_{N_1}^f\|\} < \infty$. Then, $\|d_{N_1}^f\| \leq 2B$. Suppose that $\|d_m^f\| \leq 2B$ for some $m \geq N_1$. From $d_{n+1}^f := -\nabla f(x_{n+1}) + \delta_n^{\mathrm{DY}} d_n^f$ $(n \in \mathbb{N})$, we find that

$$\left\|d_{m+1}^f\right\| \leq \|\nabla f(x_{m+1})\| + \delta_m^{\mathrm{DY}} \left\|d_m^f\right\| \leq B + \frac{1}{2}(2B) = 2B.$$

Induction shows that $\|d_n^f\| \leq 2B$ for all $n \geq N_1$. Hence, the boundedness of $(d_n^f)_{n\in\mathbb{N}}$ and $(\nabla f(x_n))_{n\in\mathbb{N}}$ imply that, for all $n \geq N_1$,

$$\begin{aligned}u_n &\leq |\langle d_n^f, \nabla f(x_{n+1}) - (1+\eta)\nabla f(x_n)\rangle| \leq \left\|d_n^f\right\| \|\nabla f(x_{n+1}) - (1+\eta)\nabla f(x_n)\|\\ &\leq \left\|d_n^f\right\| \left(\|\nabla f(x_{n+1})\| + (1+\eta)\|\nabla f(x_n)\|\right) \leq 2(2+\eta)BB_1.\end{aligned}$$

If $u_n := \langle d_n^f, \nabla f(x_{n+1}) - (1+\eta)\nabla f(x_n)\rangle \neq 0$ for all $n \geq \max\{N_0, N_1\}$, then

$$\delta_n^{\mathrm{DY}} = \frac{\|\nabla f(x_{n+1})\|^2}{u_n} \geq \frac{B_2^2}{2(2+\eta)BB_1} > 0,$$

which implies $\lim_{n\to\infty} \delta_n^{\mathrm{DY}} > 0$. Therefore, we find that $u_n = 0$ for all $n \geq \max\{N_0, N_1\}$, i.e., $\delta_n^{\mathrm{DY}} = 0$ for all $n \geq \max\{N_0, N_1\}$. This implies that $d_{n+1}^f = -\nabla f(x_{n+1})$ for all $n \geq \max\{N_0, N_1\}$. From $u_n = 0$ for all $n \geq N_2 := \max\{N_0, N_1\}+1$, we have $\langle d_n^f, \nabla f(x_{n+1})\rangle = (1+\eta)\langle d_n^f, \nabla f(x_n)\rangle$, which means $\langle \nabla f(x_n), \nabla f(x_{n+1})\rangle = (1+\eta)\|\nabla f(x_n)\|^2$, and hence,

$$\|\nabla f(x_{n+1})\| \geq (1+\eta)\|\nabla f(x_n)\| \geq (1+\eta)^{n-N_2}\|\nabla f(x_{N_2})\| \geq (1+\eta)^{n-N_2} B_2. \tag{14}$$

In the case of $\eta > 0$, the right hand side of the above inequality diverges when n diverges. This contradicts the boundedness property of $(\nabla f(x_n))_{n\in\mathbb{N}}$. Accordingly, $\{x^\star\} = \mathrm{Argmin}_{x\in H} f(x) \subset \mathrm{Fix}(N)$. Theorem 3.1 guarantees that $(x_n)_{n\in\mathbb{N}}$ strongly converges to $x^\star$.

(iii) Assume that $x^\star \notin \mathrm{Fix}(N)$. Put $v_n := \langle \nabla f(x_{n+1}), \nabla f(x_{n+1}) - (1+\kappa)\nabla f(x_n)\rangle$ ($n \in \mathbb{N}$). From $\lim_{n\to\infty} \delta_n^{\mathrm{PRP}} = \lim_{n\to\infty}(v_n/\|\nabla f(x_n)\|^2) = 0$ and Inequalities (12) and (13), we have $\lim_{n\to\infty} v_n = 0$. Moreover, the strong convergence of $(x_n)_{n\in\mathbb{N}}$ to $x^* \in \mathrm{Fix}(N)$ (Theorem 3.1) and the continuity of ∇f ensure

$$\begin{aligned} 0 = \lim_{n\to\infty} v_n &= \lim_{n\to\infty}\left(\|\nabla f(x_{n+1})\|^2 - (1+\kappa)\langle \nabla f(x_{n+1}), \nabla f(x_n)\rangle\right) \\ &= \|\nabla f(x^*)\|^2 - (1+\kappa)\|\nabla f(x^*)\|^2 = -\kappa\|\nabla f(x^*)\|^2, \end{aligned}$$

which implies from $\|\nabla f(x^*)\| \neq 0$ that $\kappa = 0$. Therefore, assuming $\kappa \neq 0$ and $\lim_{n\to\infty} \delta_n^{\mathrm{PRP}} = 0$ implies that $\{x^\star\} = \mathrm{Argmin}_{x\in H} f(x) \subset \mathrm{Fix}(N)$. Theorem 3.1 guarantees that $(x_n)_{n\in\mathbb{N}}$ strongly converges to $x^\star$.

(iv) Assume that $x^\star \notin \mathrm{Fix}(N)$. A similar discussion to that of the proof of Proposition 3.2 (ii) leads us to Inequalities (12) and (13) and the boundedness of $(d_n^f)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$. The strong convergence of $(x_n)_{n\in\mathbb{N}}$ to $x^* \in \mathrm{Fix}(N)$ (Theorem 3.1), the continuity of ∇f, the boundedness of $(d_n^f)_{n\in\mathbb{N}}$, and $\lim_{n\to\infty} \delta_n^{\mathrm{HS}} = 0$ imply that

$$\begin{aligned} 0 &\leq \limsup_{n\to\infty}\left\|d_{n+1}^f + \nabla f(x^*)\right\| = \limsup_{n\to\infty}\left\|-\nabla f(x_{n+1}) + \delta_n^{\mathrm{HS}} d_n^f + \nabla f(x^*)\right\| \\ &\leq \limsup_{n\to\infty}\left(\|\nabla f(x^*) - \nabla f(x_{n+1})\| + \delta_n^{\mathrm{HS}}\left\|d_n^f\right\|\right) \leq 0, \end{aligned}$$

which implies that $\lim_{n\to\infty}\|d_{n+1}^f + \nabla f(x^*)\| = 0$. Meanwhile, we have that, for all $n \in \mathbb{N}$,

$$\begin{aligned} &u_n - \eta\|\nabla f(x^*)\|^2 \\ &= \left\langle d_n^f, \nabla f(x_{n+1}) - (1+\eta)\nabla f(x_n)\right\rangle + \langle \nabla f(x^*), \nabla f(x^*) - (1+\eta)\nabla f(x^*)\rangle \\ &= \left\langle d_n^f, \nabla f(x_{n+1})\right\rangle + \langle \nabla f(x^*), \nabla f(x^*)\rangle \\ &\quad - (1+\eta)\left(\left\langle d_n^f, \nabla f(x_n)\right\rangle + \langle \nabla f(x^*), \nabla f(x^*)\rangle\right). \end{aligned}$$

So, the triangle inequality ensures that, for all $n \in \mathbb{N}$,

$$\left| u_n - \eta \left\| \nabla f(x^*) \right\|^2 \right| \leq \left| \left\langle d_n^f, \nabla f(x_{n+1}) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) \right\rangle \right| + (1+\eta) \left| \left\langle d_n^f, \nabla f(x_n) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) \right\rangle \right|.$$

Moreover, we find that, for all $n \in \mathbb{N}$,

$$\begin{aligned}
&\left| \left\langle d_n^f, \nabla f(x_{n+1}) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) \right\rangle \right| \\
&= \left| \left\langle d_n^f + \nabla f(x^*), \nabla f(x_{n+1}) \right\rangle - \left\langle \nabla f(x^*), \nabla f(x_{n+1}) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) \right\rangle \right| \\
&= \left| \left\langle d_n^f + \nabla f(x^*), \nabla f(x_{n+1}) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) - \nabla f(x_{n+1}) \right\rangle \right| \\
&\leq \left| \left\langle d_n^f + \nabla f(x^*), \nabla f(x_{n+1}) \right\rangle \right| + \left| \left\langle \nabla f(x^*), \nabla f(x^*) - \nabla f(x_{n+1}) \right\rangle \right|,
\end{aligned}$$

which means that

$$\begin{aligned}
&\left| \left\langle d_n^f, \nabla f(x_{n+1}) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) \right\rangle \right| \\
&\leq \left\| d_n^f + \nabla f(x^*) \right\| \left\| \nabla f(x_{n+1}) \right\| + \left\| \nabla f(x^*) \right\| \left\| \nabla f(x^*) - \nabla f(x_{n+1}) \right\|.
\end{aligned}$$

We also have that, for all $n \in \mathbb{N}$,

$$\begin{aligned}
&\left| \left\langle d_n^f, \nabla f(x_n) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) \right\rangle \right| \\
&= \left| \left\langle d_n^f + \nabla f(x^*), \nabla f(x_n) \right\rangle - \left\langle \nabla f(x^*), \nabla f(x_n) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) \right\rangle \right| \\
&= \left| \left\langle d_n^f + \nabla f(x^*), \nabla f(x_n) \right\rangle + \left\langle \nabla f(x^*), \nabla f(x^*) - \nabla f(x_n) \right\rangle \right| \\
&\leq \left| \left\langle d_n^f + \nabla f(x^*), \nabla f(x_n) \right\rangle \right| + \left| \left\langle \nabla f(x^*), \nabla f(x^*) - \nabla f(x_n) \right\rangle \right| \\
&\leq \left\| d_n^f + \nabla f(x^*) \right\| \left\| \nabla f(x_n) \right\| + \left\| \nabla f(x^*) \right\| \left\| \nabla f(x^*) - \nabla f(x_n) \right\|.
\end{aligned}$$

From $\lim_{n\to\infty} \|d_n^f + \nabla f(x^*)\| = 0$, $\lim_{n\to\infty} \|\nabla f(x_n) - \nabla f(x^*)\| = 0$, and the boundedness of $(\nabla f(x_n))_{n\in\mathbb{N}}$, we have that $\lim_{n\to\infty} |\langle d_n^f, \nabla f(x_{n+1})\rangle + \langle \nabla f(x^*), \nabla f(x^*)\rangle| = 0$ and $\lim_{n\to\infty} |\langle d_n^f, \nabla f(x_n)\rangle + \langle \nabla f(x^*), \nabla f(x^*)\rangle| = 0$. Therefore, $\lim_{n\to\infty} |u_n - \eta \|\nabla f(x^*)\|^2| = 0$, i.e.,

$$\lim_{n\to\infty} u_n = \eta \left\| \nabla f(x^*) \right\|^2. \tag{15}$$

In the case of $\eta > 0$, we find from Equation (15) and $\|\nabla f(x^*)\| \neq 0$ that $\lim_{n\to\infty} u_n > 0$. Therefore, we have from $\lim_{n\to\infty} \delta_n^{\mathrm{HS}} = \lim_{n\to\infty} (v_n/u_n) = 0$ that $\lim_{n\to\infty} v_n = 0$. A discussion similar to the proof of Proposition 3.2 (iii) leads us to $\kappa = 0$, which is a contradiction. Hence, we find that $\{x^\star\} = \mathrm{Argmin}_{x\in H} f(x) \subset \mathrm{Fix}(N)$. □

Remark 3.1 Consider the case in which the minimizer of f over H is not in $\mathrm{Fix}(N)$, $\lim_{n\to\infty} \delta_n^{\mathrm{HS}} = 0$, and $\kappa \neq 0$. Then, Proposition 3.2 (iv) leads to $\eta = 0$. In the case of $\eta = 0$, Equation (15) implies that $\lim_{n\to\infty} u_n = 0$. From $\lim_{n\to\infty} \delta_n^{\mathrm{HS}} = \lim_{n\to\infty} (v_n/u_n) = 0$, we have the following cases: (A) $v_n = o(u_n)$, or (B) $u_n = 0$ for large enough n (i.e., $\delta_n^{\mathrm{HS}} = 0$ for large enough n). We have that $\lim_{n\to\infty} v_n = -\kappa \|\nabla f(x^*)\|^2 \neq 0$ because $\kappa \neq 0$ and x^* is the minimizer of f over $\mathrm{Fix}(N)$ when the minimizer of f over H is not in

$\mathrm{Fix}(N)$ (see the proof of Proposition 3.2 (iii)). This implies that Case (A) does not hold. In Case (B), from a discussion similar to Proposition 3.2 (ii) (see Inequality (14)) and $\eta = 0$, we find that $\|\nabla f(x_{n+1})\| \geq \|\nabla f(x_n)\|$ for large enough n.

3.3 Proof of Theorem 3.1

We first prove the boundedness of $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$, $(d_n^N)_{n\in\mathbb{N}}$, and $(d_n^f)_{n\in\mathbb{N}}$.

Lemma 3.1 *Suppose that Conditions (II), (i), (iv), and (v) in Theorem 3.1 are satisfied. Then,* $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$, $(d_n^N)_{n\in\mathbb{N}}$, *and* $(d_n^f)_{n\in\mathbb{N}}$ *in Algorithm 3.1 are bounded.*

Proof The boundedness of K and the definitions of $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ guarantee the boundedness of $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$. The nonexpansivity of N guarantees that $\|N(y_n) - y\| \leq \|y_n - y\|$ for all $y \in \mathrm{Fix}(N)$. Thus, the boundedness of $(y_n)_{n\in\mathbb{N}}$ implies that $(N(y_n))_{n\in\mathbb{N}}$ is bounded, i.e., $(N(y_n) - y_n)_{n\in\mathbb{N}}$ is bounded. Moreover, the Lipschitz continuity of ∇f ensures that $\|\nabla f(x_n) - \nabla f(x)\| \leq L\|x_n - x\|$ for all $x \in H$. Hence, the boundedness of $(x_n)_{n\in\mathbb{N}}$ means that $(\nabla f(x_n))_{n\in\mathbb{N}}$ is bounded.

We shall prove that $(d_n^N)_{n\in\mathbb{N}}$ is bounded. Since $\lim_{n\to\infty} \beta_n^{(i)} = 0$ $(i = 1, 2)$ from Conditions (i) and (iv), there exists $n_0 \in \mathbb{N}$ such that $\beta_n^{(1)} \leq 1/3$ and $\beta_n^{(2)} \leq 1$ for all $n \geq n_0$. Condition (II) ensures that $K_1 := \max\{\sup\{\|N(y_n) - y_n\| \colon n \in \mathbb{N}\}, \sup\{\|w_n\| \colon n \in \mathbb{N}\}\} < \infty$ and $K_2 := \max\{K_1, \|d_{n_0}^N\|\} < \infty$. Obviously, $\|d_{n_0}^N\| \leq 3K_2$. We assume that $\|d_n^N\| \leq 3K_2$ for some $n \geq n_0$. The definition of $(d_n^N)_{n\in\mathbb{N}}$ guarantees that

$$\left\|d_{n+1}^N\right\| \leq \|N(y_n) - y_n\| + \beta_n^{(1)} \left\|d_n^N\right\| + \beta_n^{(2)}\|w_n\| \leq K_2 + \frac{1}{3}(3K_2) + K_2 = 3K_2.$$

Induction shows that $\|d_n^N\| \leq 3K_2$ for all $n \geq n_0$; i.e., $(d_n^N)_{n\in\mathbb{N}}$ is bounded.

Next, we shall prove that $(d_n^f)_{n\in\mathbb{N}}$ is bounded. The boundedness of $(\nabla f(x_n))_{n\in\mathbb{N}}$ and Condition (II) imply that $K_3 := \max\{\sup\{\|\nabla f(x_n)\| \colon n \in \mathbb{N}\}, \sup\{\|z_n\| \colon n \in \mathbb{N}\}\} < \infty$. Condition (v) ensures the existence of $n_1 \in \mathbb{N}$ such that $\delta_n^{(1)} \leq 1/3$ and $\delta_n^{(2)} \leq 1$ for all $n \geq n_1$. Put $K_4 := \max\{K_3, \|d_{n_1}^f\|\} < \infty$. Then, $\|d_{n_1}^f\| \leq 3K_4$. Suppose that $\|d_n^f\| \leq 3K_4$ for some $n \geq n_1$. The definition of $(d_n^f)_{n\in\mathbb{N}}$ means that

$$\left\|d_{n+1}^f\right\| \leq \|\nabla f(x_{n+1})\| + \delta_n^{(1)} \left\|d_n^f\right\| + \delta_n^{(2)}\|z_n\| \leq K_4 + \frac{1}{3}(3K_4) + K_4 = 3K_4.$$

Induction shows that $\|d_n^f\| \leq 3K_4$ for all $n \geq n_1$; i.e., $(d_n^f)_{n\in\mathbb{N}}$ is bounded. □

Next, we prove the following:

Lemma 3.2 *Suppose that the assumptions in Theorem 3.1 are satisfied. Then,*

(i) $\lim_{n\to\infty} \|x_{n+1} - P_K(\hat{N}(x_n))\| = 0$, *where* $\hat{N} := (1-\gamma)I + \gamma N$;

(ii) $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$ *and* $\lim_{n\to\infty} \|x_n - \hat{N}(x_n)\| = 0$*;*
(iii) $\limsup_{n\to\infty}\langle x^\star - x_n, \nabla f(x^\star)\rangle \leq 0$*, where* $x^\star \in \mathrm{Fix}(N)$ *is the solution to Problem 3.1.*

Proof (i) x_{n+1} $(n \in \mathbb{N})$ in Algorithm 3.1 can be rewritten as follows:

$$\begin{aligned} x_{n+1} &= P_K\left(y_n + \gamma d_{n+1}^N\right) = P_K\left(y_n + \gamma\left(N(y_n) - y_n + \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right)\right) \\ &= P_K\left(\left((1-\gamma)I + \gamma N\right)(y_n) + \gamma\left(\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right)\right) \\ &=: P_K\left(\hat{N}(y_n) + \gamma t_n\right), \end{aligned}$$

where $\hat{N} := (1-\gamma)I + \gamma N$ and $t_n := \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n$ $(n \in \mathbb{N})$. Since N is nonexpansive, $\hat{N} := (1-\gamma)I + \gamma N$ satisfies the nonexpansivity condition with $\mathrm{Fix}(N) = \mathrm{Fix}(\hat{N})$. The nonexpansivity of P_K guarantees that, for all $n \in \mathbb{N}$,

$$\begin{aligned} &\left\|x_{n+1} - P_K\left(\hat{N}(x_n)\right)\right\| = \left\|P_K\left(\hat{N}\left(y_n\right) + \gamma t_n\right) - P_K\left(\hat{N}(x_n)\right)\right\| \\ &\leq \left\|\left(\hat{N}\left(y_n\right) + \gamma t_n\right) - \hat{N}(x_n)\right\| \leq \left\|\hat{N}\left(y_n\right) - \hat{N}(x_n)\right\| + \gamma\left\|t_n\right\|, \end{aligned}$$

which from the nonexpansivity of $\hat{N}$ means that, for all $n \in \mathbb{N}$,

$$\left\|x_{n+1} - P_K\left(\hat{N}(x_n)\right)\right\| \leq \|y_n - x_n\| + \gamma\|t_n\|.$$

Since $y_n := P_K(x_n + \mu\alpha_n d_n^f)$ and $x_n = P_K(x_n)$ (from $x_{n+1} = P_K(\hat{N}(y_n) + \gamma t_n) \in K$ $(n \in \mathbb{N})$), the nonexpansivity of P_K ensures that

$$\begin{aligned} \|y_n - x_n\| &= \left\|P_K\left(x_n + \mu\alpha_n d_n^f\right) - P_K\left(x_n\right)\right\| \leq \left\|\left(x_n + \mu\alpha_n d_n^f\right) - x_n\right\| \\ &= \mu\alpha_n\left\|d_n^f\right\|. \end{aligned}$$

Moreover, from Condition (iv) and $\alpha_n^2 \leq \alpha_n \leq 1$ $(n \in \mathbb{N})$ we have

$$\begin{aligned} \|t_n\| &= \left\|\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right\| \leq \beta_n^{(1)}\left\|d_n^N\right\| + \beta_n^{(2)}\|w_n\| \leq \alpha_n^2\left\|d_n^N\right\| + \alpha_n^2\|w_n\| \\ &\leq \alpha_n\left(\left\|d_n^N\right\| + \|w_n\|\right). \end{aligned}$$

Therefore, we find that, for all $n \in \mathbb{N}$,

$$\left\|x_{n+1} - P_K\left(\hat{N}(x_n)\right)\right\| \leq \mu\alpha_n\left\|d_n^f\right\| + \gamma\alpha_n\left(\left\|d_n^N\right\| + \|w_n\|\right) \leq 2K_5\alpha_n,$$

where $K_5 := \max\{\sup\{\mu\|d_n^f\| \colon n \in \mathbb{N}\}, \sup\{\gamma(\|d_n^N\| + \|w_n\|) \colon n \in \mathbb{N}\}\} < \infty$. Hence, Condition (i) implies that $\lim_{n\to\infty}\|x_{n+1} - P_K(\hat{N}(x_n))\| = 0$.

(ii) Let $\tau := 1 - \sqrt{1 - \mu\left(2c - \mu L^2\right)} \in (0, 1]$ (see also Lemma 2.1). Put $s_n := (x_n + \mu\alpha_n d_n^f) - (x_{n-1} + \mu\alpha_{n-1} d_{n-1}^f)$ $(n \geq 1)$, $M_1 := \sup\{2\mu|\langle s_n, \nabla f(x_{n-1})\rangle| \colon n \geq 1\}$, $M_2 := \sup\{|\langle s_n, d_{n-1}^f\rangle|/\tau \colon n \geq 1\}$, $M_3 := \sup\{|\langle s_n, z_{n-1}\rangle|/\tau \colon n \geq 1\}$, $M_4 := \sup\{|\langle s_n, d_{n-2}^f\rangle|/\tau \colon n \geq 2\}$, and $M_5 := \sup\{|\langle s_n, z_{n-2}\rangle|/\tau \colon n \geq 2\}$.

Lemma 3.1 ensures that $M_6 := \max\{M_i \colon i = 1, 2, \ldots, 5\} < \infty$. The definition of $(d_n^f)_{n\in\mathbb{N}}$ means that, for all $n \geq 2$,

$$\begin{aligned}\|s_n\|^2 &= \left\|\left(x_n + \mu\alpha_n d_n^f\right) - \left(x_{n-1} + \mu\alpha_{n-1} d_{n-1}^f\right)\right\|^2\\ &= \left\| \left(x_n + \mu\alpha_n\left(-\nabla f(x_n) + \delta_{n-1}^{(1)} d_{n-1}^f - \delta_{n-1}^{(2)} z_{n-1}\right)\right)\right.\\ &\quad \left. - \left(x_{n-1} - \mu\alpha_n \nabla f(x_{n-1})\right) - \mu\alpha_n \nabla f(x_{n-1}) - \mu\alpha_{n-1} d_{n-1}^f \right\|^2,\end{aligned}$$

which implies that

$$\begin{aligned}\|s_n\|^2 = &\left\|\left(x_n - \mu\alpha_n \nabla f(x_n)\right) - \left(x_{n-1} - \mu\alpha_n \nabla f(x_{n-1})\right)\right.\\ &\left. + \mu\left(\alpha_n\delta_{n-1}^{(1)} d_{n-1}^f - \alpha_n\delta_{n-1}^{(2)} z_{n-1} - \alpha_n\nabla f(x_{n-1}) - \alpha_{n-1} d_{n-1}^f\right)\right\|^2.\end{aligned}$$

From the inequality, $\|x+y\|^2 \leq \|x\|^2 + 2\langle x+y, y\rangle$ $(x, y \in H)$, we find that, for all $n \geq 2$,

$$\begin{aligned}\|s_n\|^2 \leq &\left\|\left(x_n - \mu\alpha_n \nabla f(x_n)\right) - \left(x_{n-1} - \mu\alpha_n \nabla f(x_{n-1})\right)\right\|^2\\ &+ 2\mu\left\langle \alpha_n\delta_{n-1}^{(1)} d_{n-1}^f - \alpha_n\delta_{n-1}^{(2)} z_{n-1} - \alpha_n\nabla f(x_{n-1}) - \alpha_{n-1} d_{n-1}^f, s_n\right\rangle.\end{aligned}$$

Moreover, Lemma 2.1 guarantees that, for all $n \geq 2$,

$$\begin{aligned}&\left\|\left(x_n - \mu\alpha_n \nabla f(x_n)\right) - \left(x_{n-1} - \mu\alpha_n \nabla f(x_{n-1})\right)\right\|^2\\ \leq &(1-\tau\alpha_n)^2\|x_n - x_{n-1}\|^2 \leq (1-\tau\alpha_n)\|x_n - x_{n-1}\|^2.\end{aligned}$$

The definition of $(d_n^f)_{n\in\mathbb{N}}$ means that, for all $n \geq 2$,

$$\begin{aligned}&2\mu\left\langle \alpha_n\delta_{n-1}^{(1)} d_{n-1}^f - \alpha_n\delta_{n-1}^{(2)} z_{n-1} - \alpha_n\nabla f(x_{n-1}) - \alpha_{n-1} d_{n-1}^f, s_n\right\rangle\\ &= 2\mu\Big\langle \alpha_n\delta_{n-1}^{(1)} d_{n-1}^f - \alpha_n\delta_{n-1}^{(2)} z_{n-1} - \alpha_n\nabla f(x_{n-1})\\ &\quad - \alpha_{n-1}\left(-\nabla f(x_{n-1}) + \delta_{n-2}^{(1)} d_{n-2}^f - \delta_{n-2}^{(2)} z_{n-2}\right), s_n\Big\rangle\\ &= 2\mu(\alpha_{n-1} - \alpha_n)\left\langle s_n, \nabla f(x_{n-1})\right\rangle\\ &\quad + 2\mu\alpha_n\delta_{n-1}^{(1)}\left\langle s_n, d_{n-1}^f\right\rangle + 2\mu\alpha_n\delta_{n-1}^{(2)}\left\langle s_n, -z_{n-1}\right\rangle\\ &\quad + 2\mu\alpha_{n-1}\delta_{n-2}^{(1)}\left\langle s_n, -d_{n-2}^f\right\rangle + 2\mu\alpha_{n-1}\delta_{n-2}^{(2)}\left\langle s_n, z_{n-2}\right\rangle\\ &\leq 2\mu\left|\alpha_{n-1} - \alpha_n\right|\left|\left\langle s_n, \nabla f(x_{n-1})\right\rangle\right|\\ &\quad + 2\mu\alpha_n\delta_{n-1}^{(1)}\left|\left\langle s_n, d_{n-1}^f\right\rangle\right| + 2\mu\alpha_n\delta_{n-1}^{(2)}\left|\left\langle s_n, z_{n-1}\right\rangle\right|\\ &\quad + 2\mu\alpha_{n-1}\delta_{n-2}^{(1)}\left|\left\langle s_n, d_{n-2}^f\right\rangle\right| + 2\mu\alpha_{n-1}\delta_{n-2}^{(2)}\left|\left\langle s_n, z_{n-2}\right\rangle\right|.\end{aligned}$$

The definition of M_6 leads one to deduce that, for all $n \geq 2$,

$$\begin{aligned} &2\mu \left\langle \alpha_n \delta_{n-1}^{(1)} d_{n-1}^f - \alpha_n \delta_{n-1}^{(2)} z_{n-1} - \alpha_n \nabla f(x_{n-1}) - \alpha_{n-1} d_{n-1}^f, s_n \right\rangle \\ &\leq M_6 \left|\alpha_{n-1} - \alpha_n\right| + 2\mu\alpha_n \delta_{n-1}^{(1)} \left(\tau M_6\right) + 2\mu\alpha_n \delta_{n-1}^{(2)} \left(\tau M_6\right) \\ &\quad + 2\mu\alpha_{n-1} \delta_{n-2}^{(1)} \left(\tau M_6\right) + 2\mu\alpha_{n-1} \delta_{n-2}^{(2)} \left(\tau M_6\right). \end{aligned}$$

From $\alpha_{n-1} \leq \alpha_n + |\alpha_n - \alpha_{n-1}|$ ($n \in \mathbb{N}$), the right hand side of the above inequality means that

$$\begin{aligned} &M_6 \left|\alpha_{n-1} - \alpha_n\right| + 2\mu\alpha_n \delta_{n-1}^{(1)} \left(\tau M_6\right) + 2\mu\alpha_n \delta_{n-1}^{(2)} \left(\tau M_6\right) \\ &\quad + 2\mu\alpha_{n-1} \delta_{n-2}^{(1)} \left(\tau M_6\right) + 2\mu\alpha_{n-1} \delta_{n-2}^{(2)} \left(\tau M_6\right) \\ &\leq M_6 \left|\alpha_{n-1} - \alpha_n\right| + 2\mu\alpha_n \delta_{n-1}^{(1)} \left(\tau M_6\right) + 2\mu\alpha_n \delta_{n-1}^{(2)} \left(\tau M_6\right) \\ &\quad + 2\mu \left\{\alpha_n + \left|\alpha_n - \alpha_{n-1}\right|\right\} \delta_{n-2}^{(1)} \left(\tau M_6\right) + 2\mu \left\{\alpha_n + \left|\alpha_n - \alpha_{n-1}\right|\right\} \delta_{n-2}^{(2)} \left(\tau M_6\right) \\ &= \left(M_6 + 2\mu \delta_{n-2}^{(1)} \left(\tau M_6\right) + 2\mu \delta_{n-2}^{(2)} \left(\tau M_6\right)\right) \left|\alpha_{n-1} - \alpha_n\right| \\ &\quad + 2\mu M_6 \tau \alpha_n \delta_{n-1}^{(1)} + 2\mu M_6 \tau \alpha_n \delta_{n-1}^{(2)} + 2\mu M_6 \tau \alpha_n \delta_{n-2}^{(1)} + 2\mu M_6 \tau \alpha_n \delta_{n-2}^{(2)}. \end{aligned}$$

Therefore, combining all the above inequalities means that, for all $n \geq 2$,

$$\begin{aligned} \|s_n\|^2 &\leq (1 - \tau\alpha_n)\|x_n - x_{n-1}\|^2 + M_7|\alpha_{n-1} - \alpha_n| \\ &\quad + 2\mu M_6 \tau \alpha_n \delta_{n-1}^{(1)} + 2\mu M_6 \tau \alpha_n \delta_{n-1}^{(2)} + 2\mu M_6 \tau \alpha_n \delta_{n-2}^{(1)} + 2\mu M_6 \tau \alpha_n \delta_{n-2}^{(2)}, \end{aligned} \tag{16}$$

where $M_7 := \sup\{M_6 + 2\mu\delta_{n-2}^{(1)}(\tau M_6) + 2\mu\delta_{n-2}^{(2)}(\tau M_6) \colon n \geq 2\} < \infty$.

Put $r_n := (\hat{N}(y_n) + \gamma t_n) - (\hat{N}(y_{n-1}) + \gamma t_{n-1})$ ($n \in \mathbb{N}$) and $M_8 := \max\{\sup\{2\gamma\|r_n\|(\|d_n^N\| + \|w_n\|) \colon n \in \mathbb{N} \colon n \geq 1\}, \sup\{2\gamma\|r_n\|(\|d_{n-1}^N\| + \|w_{n-1}\|) \colon n \in \mathbb{N} \colon n \geq 2\}\} < \infty$. From $x_{n+1} = P_K(\hat{N}(y_n) + \gamma t_n)$ ($n \in \mathbb{N}$) and the nonexpasivity of P_K, we have that, for all $n \geq 1$,

$$\begin{aligned} \|x_{n+1} - x_n\|^2 &= \left\|P_K\left(\hat{N}(y_n) + \gamma t_n\right) - P_K\left(\hat{N}(y_{n-1}) + \gamma t_{n-1}\right)\right\|^2 \\ &\leq \left\|\left(\hat{N}(y_n) + \gamma t_n\right) - \left(\hat{N}(y_{n-1}) + \gamma t_{n-1}\right)\right\|^2 = \|r_n\|^2. \end{aligned}$$

So, from the inequality, $\|x+y\|^2 \leq \|x\|^2 + 2\langle x+y, y\rangle$ ($x, y \in H$), we find that, for all $n \geq 1$,

$$\begin{aligned} \|x_{n+1} - x_n\|^2 \leq \|r_n\|^2 &= \left\|\hat{N}(y_n) - \hat{N}(y_{n-1}) + \gamma(t_n - t_{n-1})\right\|^2 \\ &\leq \left\|\hat{N}(y_n) - \hat{N}(y_{n-1})\right\|^2 + 2\gamma\langle r_n, t_n - t_{n-1}\rangle. \end{aligned}$$

The nonexpasivity of $\hat{N}$ and P_K and the Cauchy-Schwarz inequality mean that

$$\begin{aligned}
&\|x_{n+1} - x_n\|^2 \\
&\leq \|y_n - y_{n-1}\|^2 + 2\gamma \|r_n\| \|t_n - t_{n-1}\| \\
&= \left\| P_K\left(x_n + \mu\alpha_n d_n^f\right) - P_K\left(x_{n-1} + \mu\alpha_{n-1} d_{n-1}^f\right)\right\|^2 + 2\gamma \|r_n\| \|t_n - t_{n-1}\| \\
&\leq \left\| \left(x_n + \mu\alpha_n d_n^f\right) - \left(x_{n-1} + \mu\alpha_{n-1} d_{n-1}^f\right)\right\|^2 + 2\gamma \|r_n\| \|t_n - t_{n-1}\| \\
&\leq \|s_n\|^2 + 2\gamma \|r_n\| \left\{\|t_n\| + \|t_{n-1}\|\right\},
\end{aligned}$$

where $y_n := P_K(x_n + \mu\alpha_n d_n^f)$ and $s_n := (x_n + \mu\alpha_n d_n^f) - (x_{n-1} + \mu\alpha_{n-1} d_{n-1}^f)$ $(n \geq 1)$. From $t_n := \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n$ $(n \in \mathbb{N})$ and Condition (iv), we find that

$$\begin{aligned}
&\|x_{n+1} - x_n\|^2 \\
&\leq \|s_n\|^2 + 2\gamma\|r_n\| \left\{\beta_n^{(1)} \left\|d_n^N\right\| + \beta_n^{(2)} \|w_n\| + \beta_{n-1}^{(1)} \left\|d_{n-1}^N\right\| + \beta_{n-1}^{(2)} \|w_{n-1}\|\right\} \\
&\leq \|s_n\|^2 + 2\gamma\|r_n\| \left\{\alpha_n^2 \left(\left\|d_n^N\right\| + \|w_n\|\right) + \alpha_{n-1}^2 \left(\left\|d_{n-1}^N\right\| + \|w_{n-1}\|\right)\right\} \\
&\leq \|s_n\|^2 + M_8 \left(\alpha_n^2 + \alpha_{n-1}^2\right).
\end{aligned}$$

Moreover, from $\alpha_{n-1} \leq \alpha_n + |\alpha_n - \alpha_{n-1}|$ $(n \in \mathbb{N})$, we have that, for all $n \geq 1$,

$$\begin{aligned}
\|x_{n+1} - x_n\|^2 &\leq \|s_n\|^2 + \frac{M_8}{\tau}\tau\alpha_n\alpha_n + M_8\alpha_{n-1}\left(\alpha_n + |\alpha_n - \alpha_{n-1}|\right) \\
&\leq \|s_n\|^2 + \frac{M_8}{\tau}\tau\alpha_n\left(\alpha_n + \alpha_{n-1}\right) + M_8|\alpha_n - \alpha_{n-1}|.
\end{aligned}$$

Accordingly, Inequality (16) guarantees that, for all $n \geq 2$,

$$\begin{aligned}
&\|x_{n+1} - x_n\|^2 \\
&\leq (1 - \tau\alpha_n)\|x_n - x_{n-1}\|^2 + (M_7 + M_8)\left|\alpha_{n-1} - \alpha_n\right| + 2\mu M_6\tau\alpha_n\delta_{n-1}^{(1)} \\
&\quad + 2\mu M_6\tau\alpha_n\delta_{n-1}^{(2)} + 2\mu M_6\tau\alpha_n\delta_{n-2}^{(1)} + 2\mu M_6\tau\alpha_n\delta_{n-2}^{(2)} + \frac{M_8}{\tau}\tau\alpha_n\left(\alpha_n + \alpha_{n-1}\right).
\end{aligned}$$

On the other hand, Conditions (i) and (v) guarantee that, for all $\varepsilon > 0$, there exists $m_0 \in \mathbb{N}$ such that $(M_8/\tau)\alpha_n \leq \varepsilon/10$, $(M_8/\tau)\alpha_{n-1} \leq \varepsilon/10$, $\mu M_6\delta_{n-1}^{(i)} \leq \varepsilon/10$, and $\mu M_6\delta_{n-2}^{(i)} \leq \varepsilon/10$ $(i = 1, 2)$ for all $n \geq m_0$. Therefore, we find that, for all $n \geq m_0$,

$$\begin{aligned}
&\|x_{n+1} - x_n\|^2 \\
&\leq (1 - \tau\alpha_n)\|x_n - x_{n-1}\|^2 + (M_7 + M_8)\left|\alpha_{n-1} - \alpha_n\right| + \tau\alpha_n\varepsilon \\
&= (1 - \tau\alpha_n)\|x_n - x_{n-1}\|^2 + (M_7 + M_8)\left|\alpha_n - \alpha_{n-1}\right| + \left(1 - (1 - \tau\alpha_n)\right)\varepsilon.
\end{aligned}$$

Hence, for all $m, n \geq m_0$,

$$
\begin{aligned}
&\|x_{n+m+1} - x_{n+m}\|^2 \\
&\leq (1 - \tau\alpha_{n+m})\|x_{n+m} - x_{n+m-1}\|^2 + (M_7 + M_8)\,|\alpha_{n+m} - \alpha_{n+m-1}| \\
&\quad + \varepsilon\,(1 - (1 - \tau\alpha_{n+m})) \\
&\leq (1 - \tau\alpha_{n+m})\big\{(1 - \tau\alpha_{n+m-1})\|x_{n+m-1} - x_{n+m-2}\|^2 \\
&\quad + (M_7 + M_8)\,|\alpha_{n+m-1} - \alpha_{n+m-2}| + \varepsilon\,(1 - (1 - \tau\alpha_{n+m-1}))\,\big\} \\
&\quad + (M_7 + M_8)\,|\alpha_{n+m} - \alpha_{n+m-1}| + \varepsilon\,(1 - (1 - \tau\alpha_{n+m})) \\
&\leq (1 - \tau\alpha_{n+m})(1 - \tau\alpha_{n+m-1})\|x_{n+m-1} - x_{n+m-2}\|^2 \\
&\quad + (M_7 + M_8)\,(|\alpha_{n+m} - \alpha_{n+m-1}| + |\alpha_{n+m-1} - \alpha_{n+m-2}|) \\
&\quad + \varepsilon\,(1 - (1 - \tau\alpha_{n+m})(1 - \tau\alpha_{n+m-1})) \\
&\leq \prod_{k=m}^{n+m-1}(1 - \tau\alpha_{k+1})\|x_{m+1} - x_m\|^2 + (M_7 + M_8)\sum_{k=m}^{n+m-1}|\alpha_{k+1} - \alpha_k| \\
&\quad + \varepsilon\left(1 - \prod_{k=m}^{n+m-1}(1 - \tau\alpha_{k+1})\right).
\end{aligned}
$$

Since $\prod_{k=m}^{\infty}(1 - \tau\alpha_{k+1}) = 0$ from Condition (ii), we find that, for every $m \geq m_0$,

$$
\begin{aligned}
&\limsup_{n\to\infty}\|x_{n+1} - x_n\|^2 = \limsup_{n\to\infty}\|x_{n+m+1} - x_{n+m}\|^2 \\
&\leq \prod_{k=m}^{\infty}(1 - \tau\alpha_{k+1})\|x_{m+1} - x_m\|^2 + (M_7 + M_8)\sum_{k=m}^{\infty}|\alpha_{k+1} - \alpha_k| \\
&\quad + \varepsilon\left(1 - \prod_{k=m}^{\infty}(1 - \tau\alpha_{k+1})\right) \\
&\leq (M_7 + M_8)\sum_{k=m}^{\infty}|\alpha_{k+1} - \alpha_k| + \varepsilon.
\end{aligned}
$$

Moreover, since $\lim_{m\to\infty}\sum_{k=m}^{\infty}|\alpha_{k+1} - \alpha_k| = 0$ from Condition (iii), we find that $\limsup_{n\to\infty}\|x_{n+1} - x_n\|^2 \leq \varepsilon$ for all $\varepsilon > 0$. The arbitrary property of ε ensures that

$$\lim_{n\to\infty}\|x_{n+1} - x_n\| = 0.$$

From $\|x_n - P_K(\hat{N}(x_n))\| \leq \|x_n - x_{n+1}\| + \|x_{n+1} - P_K(\hat{N}(x_n))\|$, $\lim_{n\to\infty}\|x_{n+1} - P_K(\hat{N}(x_n))\| = 0$, and $\lim_{n\to\infty}\|x_{n+1} - x_n\| = 0$, we find that

$$\lim_{n\to\infty}\left\|x_n - P_K\left(\hat{N}(x_n)\right)\right\| = 0.$$

Therefore, the firm nonexpansivity of P_K, the nonexpansivity of $\hat{N}$, $\mathrm{Fix}(N) = \mathrm{Fix}(\hat{N}) \subset K = \mathrm{Fix}(P_K)$, and Lemma 2.2 ensure that

$$\lim_{n\to\infty} \left\| x_n - \hat{N}\left(x_n\right) \right\| = 0. \tag{17}$$

(iii) Suppose that $x^\star \in \mathrm{Fix}(N)$ is the unique solution to Problem 3.1. Choose a subsequence, $(x_{n_i})_{i\in\mathbb{N}}$, of $(x_n)_{n\in\mathbb{N}}$ such that

$$\limsup_{n\to\infty} \langle x^\star - x_n, \nabla f(x^\star)\rangle = \lim_{i\to\infty} \langle x^\star - x_{n_i}, \nabla f(x^\star)\rangle .$$

The boundedness of $(x_{n_i})_{i\in\mathbb{N}}$ guarantees the existence of a subsequence, $(x_{n_{i_j}})_{j\in\mathbb{N}}$, of $(x_{n_i})_{i\in\mathbb{N}}$ and a point, $x^* \in H$, such that $(x_{n_{i_j}})_{j\in\mathbb{N}}$ weakly converges to x^*. From the closedness of K and $(x_n)_{n\in\mathbb{N}} \subset K$, we find that $x^* \in K$. We may assume without loss of generality that $(x_{n_i})_{i\in\mathbb{N}}$ weakly converges to x^*. We shall prove that $x^* \in K$ is a fixed point of N. Assume that $x^* \neq \hat{N}(x^*)$. Opial's condition[8], Equation (17), and the nonexpansivity of $\hat{N}$ produce a contradiction:

$$\begin{aligned}
&\liminf_{i\to\infty} \|x_{n_i} - x^*\| < \liminf_{i\to\infty} \left\| x_{n_i} - \hat{N}(x^*) \right\| \\
&= \liminf_{i\to\infty} \left\| x_{n_i} - \hat{N}(x_{n_i}) + \hat{N}(x_{n_i}) - \hat{N}(x^*) \right\| = \liminf_{i\to\infty} \left\| \hat{N}(x_{n_i}) - \hat{N}(x^*) \right\| \\
&\leq \liminf_{i\to\infty} \|x_{n_i} - x^*\|.
\end{aligned}$$

Accordingly, we find that $x^* \in \mathrm{Fix}(\hat{N}) = \mathrm{Fix}(N)$. Since $x^\star \in \mathrm{Fix}(N)$ is the solution to Problem 3.1, $\langle x^* - x^\star, \nabla f(x^\star)\rangle \geq 0$ holds (see Subsection 2.2). Therefore,

$$\limsup_{n\to\infty} \langle x^\star - x_n, \nabla f(x^\star)\rangle = \lim_{i\to\infty} \langle x^\star - x_{n_i}, \nabla f(x^\star)\rangle = \langle x^\star - x^*, \nabla f(x^\star)\rangle \leq 0.$$

This completes the proof.

Regarding Lemma 3.2(ii), we can make the following remark.

Remark 3.2 From $\|x_n - \hat{N}(x_n)\| = \|x_n - (1-\gamma)x_n - \gamma N(x_n)\| = \gamma\|x_n - N(x_n)\|$ $(n \in \mathbb{N})$, Lemma 3.2(ii) guarantees that $\lim_{n\to\infty} \|x_n - N(x_n)\| = 0$. Let us see whether $(\|x_n - N(x_n)\|)_{n\in\mathbb{N}}$ in Algorithm 3.1 monotonically decreases or not. For simplicity, we assume that $x_n + \mu\alpha_n d_n^f$, $y_n + \gamma d_{n+1}^N \in K$ $(n \in \mathbb{N})$ and $\gamma := 1$, i.e., $y_n := x_n + \mu\alpha_n d_n^f$, $x_{n+1} := y_n + d_{n+1}^N$ $(n \in \mathbb{N})$. The definition of $(d_n^N)_{n\in\mathbb{N}}$ means that, for all $n \in \mathbb{N}$,

$$\begin{aligned}
\|x_{n+1} - N(x_{n+1})\| &= \left\| \left(y_n + d_{n+1}^N\right) - N(x_{n+1}) \right\| \\
&= \left\| N(y_n) + \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n - N(x_{n+1}) \right\| \\
&\leq \|N(y_n) - N(x_{n+1})\| + \left\| \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n \right\|,
\end{aligned}$$

[8] Suppose that $(x_n)_{n\in\mathbb{N}}$ $(\subset H)$ weakly converges to $\hat{x} \in H$ and $\bar{x} \neq \hat{x}$. Then, the following condition, called Opial's condition [26], is satisfied: $\liminf_{n\to\infty} \|x_n - \hat{x}\| < \liminf_{n\to\infty} \|x_n - \bar{x}\|$. In the above situation, Opial's condition leads to $\liminf_{i\to\infty} \|x_{n_i} - x^*\| < \liminf_{i\to\infty} \|x_{n_i} - \hat{N}(x^*)\|$.

which from the nonexpansivity of N implies that, for all $n \in \mathbb{N}$,

$$\|x_{n+1} - N(x_{n+1})\| \leq \|y_n - x_{n+1}\| + \left\|\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right\|.$$

From the definition of y_n and the triangle inequality, we also have that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|y_n - x_{n+1}\| &= \left\|x_n + \mu\alpha_n d_n^f - x_{n+1}\right\| \\ &\leq \|x_n - x_{n+1}\| + \mu\alpha_n \left\|d_n^f\right\| \\ &\leq \|x_n - N(x_n)\| + \|N(x_n) - x_{n+1}\| + \mu\alpha_n \left\|d_n^f\right\|.\end{aligned}$$

Since the the triangle inequality and the nonexpansivity of N guarantee that, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|N(x_n) - x_{n+1}\| &= \left\|N(x_n) - \left(y_n + d_{n+1}^N\right)\right\| \\ &= \left\|N(x_n) - N(y_n) - \beta_n^{(1)} d_n^N - \beta_n^{(2)} w_n\right\| \\ &\leq \|N(x_n) - N(y_n)\| + \left\|\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right\| \\ &\leq \|x_n - y_n\| + \left\|\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right\| \\ &= \mu\alpha_n \left\|d_n^f\right\| + \left\|\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right\|,\end{aligned}$$

we find that, for all $n \in \mathbb{N}$,

$$\|x_{n+1} - N(x_{n+1})\| \leq \|x_n - N(x_n)\| + 2\left(\mu\alpha_n \left\|d_n^f\right\| + \left\|\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\right\|\right).$$

This implies that $(\|x_n - N(x_n)\|)_{n\in\mathbb{N}}$ does not monotonically decrease. However, for large enough n, $\mu\alpha_n\|d_n^f\| + \|\beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n\| \approx 0$ by Conditions (i) and (iv) in Theorem 3.1. Therefore, we can see that $(\|x_n - N(x_n)\|)_{n\in\mathbb{N}}$ will monotonically decrease for large enough n. Such a trend is also observed in the numerical examples in Section 4. Figures 5 and 7 show that $(\|x_n - N(x_n)\|)_{n\leq 10}$ in Algorithm 3.1 does not monotonically decrease, $(\|x_n - N(x_n)\|)_{n>10}$ in Algorithm 3.1 monotonically decreases, and Algorithm 3.1 converges in $\mathrm{Fix}(N)$ faster than the existing algorithms. See Section 4 for the details about the numerical examples.

We can prove Theorem 3.1 by using Lemmas 3.1 and 3.2.

Proof of Theorem 3.1. Conditions (II), (i) and (v), and Lemmas 3.1 and 3.2 guarantee that, for all $\varepsilon > 0$, there exists $m_1 \in \mathbb{N}$ such that, for all $n \geq m_1$,

$$\begin{aligned}&\frac{\mu}{\tau}\langle x^\star - x_n, \nabla f(x^\star)\rangle \leq \frac{\varepsilon}{10},\ \frac{\mu\delta_{n-1}^{(1)}}{\tau}\left\langle x_n - x^\star, d_{n-1}^f\right\rangle \leq \frac{\varepsilon}{10}, \\ &\frac{\mu\delta_{n-1}^{(2)}}{\tau}\langle x^\star - x_n, z_{n-1}\rangle \leq \frac{\varepsilon}{10}, \\ &\frac{\gamma\alpha_n}{\tau}\left(\left\|d_n^N\right\| + \|w_n\|\right)\left\|\hat{N}(y_n) - \hat{N}(x^\star) + \gamma t_n\right\| \leq \frac{\varepsilon}{10}, \\ &\frac{\mu^2\alpha_n}{\tau}\left\langle d_n^f, \delta_{n-1}^{(1)} d_{n-1}^f - \delta_{n-1}^{(2)} z_{n-1} - \nabla f(x^\star)\right\rangle \leq \frac{\varepsilon}{10}.\end{aligned} \tag{18}$$

The nonexpansivity of P_K and the definition of $(d_n^f)_{n\in\mathbb{N}}$ imply that, for all $n \geq m_1$,

$$\begin{aligned}
\|y_n - x^\star\|^2 &= \left\|P_K\left(x_n + \mu\alpha_n d_n^f\right) - P_K(x^\star)\right\|^2 \\
&\leq \left\|x_n + \mu\alpha_n d_n^f - x^\star\right\|^2 \\
&= \left\|x_n + \mu\alpha_n\left(-\nabla f(x_n) + \delta_{n-1}^{(1)} d_{n-1}^f - \delta_{n-1}^{(2)} z_{n-1}\right) - x^\star\right\|^2 \\
&= \Big\|\left(x_n - \mu\alpha_n \nabla f(x_n)\right) - \left(x^\star - \mu\alpha_n \nabla f(x^\star)\right) \\
&\quad + \mu\alpha_n\left(\delta_{n-1}^{(1)} d_{n-1}^f - \delta_{n-1}^{(2)} z_{n-1} - \nabla f(x^\star)\right)\Big\|^2.
\end{aligned}$$

Accordingly, from the inequality, $\|x+y\|^2 \leq \|x\|^2 + 2\langle x+y, y\rangle$ $(x, y \in H)$ and Lemma 2.1, we have that, for all $n \geq m_1$,

$$\begin{aligned}
&\|y_n - x^\star\|^2 \\
&\leq \|(x_n - \mu\alpha_n \nabla f(x_n)) - (x^\star - \mu\alpha_n \nabla f(x^\star))\|^2 \\
&\quad + 2\mu\alpha_n\left\langle x_n + \mu\alpha_n d_n^f - x^\star, \delta_{n-1}^{(1)} d_{n-1}^f - \delta_{n-1}^{(2)} z_{n-1} - \nabla f(x^\star)\right\rangle \\
&\leq (1 - \tau\alpha_n)\|x_n - x^\star\|^2 \\
&\quad + 2\mu\alpha_n\left\langle x_n + \mu\alpha_n d_n^f - x^\star, \delta_{n-1}^{(1)} d_{n-1}^f - \delta_{n-1}^{(2)} z_{n-1} - \nabla f(x^\star)\right\rangle \\
&= (1 - \tau\alpha_n)\|x_n - x^\star\|^2 + 2\tau\alpha_n\Bigg\{\frac{\mu}{\tau}\langle x^\star - x_n, \nabla f(x^\star)\rangle + \frac{\mu\delta_{n-1}^{(1)}}{\tau}\left\langle x_n - x^\star, d_{n-1}^f\right\rangle \\
&\quad + \frac{\mu\delta_{n-1}^{(2)}}{\tau}\langle x^\star - x_n, z_{n-1}\rangle + \frac{\mu^2\alpha_n}{\tau}\left\langle d_n^f, \delta_{n-1}^{(1)} d_{n-1}^f - \delta_{n-1}^{(2)} z_{n-1} - \nabla f(x^\star)\right\rangle\Bigg\}.
\end{aligned}$$

Therefore, Inequality (18) guarantees that, for all $n \geq m_1$,

$$\|y_n - x^\star\|^2 \leq (1 - \tau\alpha_n)\|x_n - x^\star\|^2 + \frac{4}{5}\varepsilon\tau\alpha_n. \tag{19}$$

Also, from the nonexpansivity of P_K and the inequality, $\|x+y\|^2 \leq \|x\|^2 + 2\langle x+y, y\rangle$ $(x, y \in H)$, we have that, for all $n \geq m_1$,

$$\begin{aligned}
\|x_{n+1} - x^\star\|^2 &= \left\|P_K\left(\hat{N}(y_n) + \gamma t_n\right) - P_K\left(\hat{N}(x^\star)\right)\right\|^2 \\
&\leq \left\|\hat{N}(y_n) - \hat{N}(x^\star) + \gamma t_n\right\|^2 \\
&\leq \left\|\hat{N}(y_n) - \hat{N}(x^\star)\right\|^2 + 2\gamma\left\langle t_n, \hat{N}(y_n) - \hat{N}(x^\star) + \gamma t_n\right\rangle.
\end{aligned}$$

Moreover, the nonexpasivity of $\hat{N}$ and the Cauchy-Schwarz inequality mean that, for all $n \geq m_1$,

$$\begin{aligned}
&\|x_{n+1} - x^\star\|^2 \\
&\leq \|y_n - x^\star\|^2 + 2\gamma\left\langle \beta_n^{(1)} d_n^N + \beta_n^{(2)} w_n, \hat{N}(y_n) - \hat{N}(x^\star) + \gamma t_n\right\rangle \\
&\leq \|y_n - x^\star\|^2 + 2\gamma\left(\beta_n^{(1)}\left\|d_n^N\right\| + \beta_n^{(2)}\|w_n\|\right)\left\|\hat{N}(y_n) - \hat{N}(x^\star) + \gamma t_n\right\|.
\end{aligned}$$

Condition (iv) leads one to deduce that, for all $n \geq m_1$,

$$\begin{aligned}
&\|x_{n+1} - x^\star\|^2 \\
&\leq \|y_n - x^\star\|^2 + 2\gamma\alpha_n^2 \left(\left\|d_n^N\right\| + \|w_n\|\right) \left\|\hat{N}(y_n) - \hat{N}(x^\star) + \gamma t_n\right\| \\
&= \|y_n - x^\star\|^2 + 2\tau\alpha_n \frac{\gamma\alpha_n}{\tau} \left(\left\|d_n^N\right\| + \|w_n\|\right) \left\|\hat{N}(y_n) - \hat{N}(x^\star) + \gamma t_n\right\|.
\end{aligned}$$

Hence, Inequalities (19) and (18) imply that, for all $n \geq m_1$,

$$\begin{aligned}
\|x_{n+1} - x^\star\|^2 &\leq (1 - \tau\alpha_n) \|x_n - x^\star\|^2 + \frac{4}{5}\varepsilon\tau\alpha_n + 2\tau\alpha_n \frac{\varepsilon}{10} \\
&= (1 - \tau\alpha_n) \|x_n - x^\star\|^2 + \varepsilon\tau\alpha_n \\
&= (1 - \tau\alpha_n) \|x_n - x^\star\|^2 + \varepsilon(1 - (1 - \tau\alpha_n)).
\end{aligned}$$

Induction thus gives, for all $n \geq m_1$,

$$\|x_{n+1} - x^\star\|^2 \leq \prod_{k=m_1}^{n} (1 - \tau\alpha_k) \|x_{m_1} - x^\star\|^2 + \varepsilon \left(1 - \prod_{k=m_1}^{n} (1 - \tau\alpha_k)\right).$$

Since $\prod_{k=m_1}^{\infty}(1 - \tau\alpha_{k+1}) = 0$ from Condition (ii), we find that

$$\limsup_{n\to\infty} \|x_{n+1} - x^\star\|^2 \leq \varepsilon.$$

The arbitrary property of ε ensures that $\limsup_{n\to\infty} \|x_{n+1} - x^\star\|^2 \leq 0$; i.e., $\lim_{n\to\infty} \|x_{n+1} - x^\star\|^2 = 0$. This means that $(x_n)_{n\in\mathbb{N}}$ in Algorithm 3.1 strongly converges to the unique solution to Problem 3.1. □

4 Numerical Examples

This section provides numerical comparisons of the existing algorithms (HSDM, HCGM, and HTCGM) with Algorithm 3.1 for the following problem:

Problem 4.1

$$\textit{Minimize } f(x) := \frac{1}{2}\langle x, Qx\rangle + \langle b, x\rangle \textit{ subject to } x \in \mathrm{Fix}(N),$$

where $Q \in \mathbb{R}^{S\times S}$ $(S = 1000, 5000)$ *is positive definite,* $b \in \mathbb{R}^S$*, and* $N \colon \mathbb{R}^S \to \mathbb{R}^S$ *is nonexpansive with* $\mathrm{Fix}(N) \neq \emptyset$.

HSDM, HCGM, and HTCGM used in the experiment were as follows: $x_0 \in \mathbb{R}^S$, $d_0^f := -\nabla f(x_0)$,

$$x_{n+1} := N\left(x_n + \frac{10^{-4}}{\sqrt{n+1}} d_n^f\right),$$

where the directions in HSDM, HCGM, and HTCGM are, respectively,

$$d_{n+1}^f := -\nabla f(x_{n+1}),$$

$$d_{n+1}^f := -\nabla f(x_{n+1}) + \frac{1}{(n+1)^{0.01}} d_n^f, \tag{20}$$

$$d_{n+1}^f := -\nabla f(x_{n+1}) + \frac{1}{(n+1)^{0.01}} d_n^f - \frac{1}{(n+1)^{0.01}} \nabla f(x_{n+1}). \tag{21}$$

It is guaranteed that the above HSDM, HCGM, and HTCGM converge to the unique solution to Problem 4.1 [15, Theorem 7].

The directions in Algorithm 3.1 used in the experiment are given by $d_0^f := -\nabla f(x_0)$, $d_0^N := N(x_0 + 10^{-4} d_0^f) - (x_0 + 10^{-4} d_0^f)$,

$$d_{n+1}^N := N(y_n) - y_n + \frac{1}{n+1} d_n^N + \frac{1}{n+1} \left(N(y_n) - y_n\right), \tag{22}$$

$$d_{n+1}^f := -\nabla f(x_{n+1}) + \frac{1}{(n+1)^{0.01}} d_n^f - \frac{1}{(n+1)^{0.01}} \nabla f(x_{n+1}), \tag{23}$$

where $y_n := P_K(x_n + (10^{-4}/\sqrt{n+1}) d_n^f)$, $x_{n+1} := P_K(y_n + d_{n+1}^N)$, and K $(\subset \mathbb{R}^S)$ is a closed ball with a large radius. Theorem 3.1 guarantees that Algorithm 3.1 with the above directions converges to the unique solution to Problem 4.1.

We also applied HCGM with each of the FR, PRP, HS, and DY formulas (Algorithm (11) with δ_n defined by one of Formulas (10)) to Problem 4.1 and verified whether HCGMs with the FR, PRP, HS, and DY formulas converge to the solution to Problem 4.1.

We chose five random initial points and executed HSDM, HCGM, HTCGM, Algorithm 3.1, and HCGMs with the FR, PRP, HS, and DY formulas for any initial point. The following graphs plot the mean values of the fifth execution. The computer used in the experiment had an Intel Boxed Core i7 i7-870 2.93 GHz 8 M CPU and 8 GB of memory. The language was MATLAB 7.9.

4.1 Constraint set in Problem 4.1 is the intersection of two balls

Suppose that $b := 0 \in \mathbb{R}^S$, $Q \in \mathbb{R}^{S \times S}$ $(S = 1000, 5000)$ is a diagonal matrix which has eigenvalues, $1, 2, \ldots, S$, $C_1 := \{x \in \mathbb{R}^S : \|x\|^2 \leq 4\}$, and $C_2 := \{x \in \mathbb{R}^S : \|x - (2, 0, 0 \ldots, 0)^T\|^2 \leq 1\}$. Define $N : \mathbb{R}^S \to \mathbb{R}^S$ by

$$N := P_{C_1} P_{C_2}.$$

Then, N is nonexpansive because P_{C_1} and P_{C_2} are nonexpansive. Moreover, $\mathrm{Fix}(N) = C_1 \cap C_2 \neq \emptyset$. Note that the exact solution to Problem 4.1 in this case is $x^\star := (1, 0, 0, \ldots, 0)^T \in \mathbb{R}^S$. To see whether or not the algorithms used in the experiment converge to the solution, we employed the following function: for each $n \in \mathbb{N}$,

$$D_n := \|x_n - x^\star\|^2,$$

where x_n is the nth approximation to the solution. The convergence of $(D_n)_{n\in\mathbb{N}}$s to 0 implies that the algorithms converge to the solution to Problem 4.1.

Figure 1 describes the behaviors of D_n for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $S = 1000$. This figure shows that $(D_n)_{n\in\mathbb{N}}$ generated by Algorithm 3.1 converges to 0 faster than $(D_n)_{n\in\mathbb{N}}$s generated by the existing algorithms, which means that Algorithm 3.1 converges fastest to the solution. The CPU time to compute x_{2000} satisfying $D_{2000} < 10^{-6}$ in Algorithm 3.1 is about 8.1 s, while HSDM, HCGM, and HTCGM satisfy $D_n > 10^{-2}$ when the CPU time is about 8.1 s. In particular, Algorithm 3.1 converges to the solution faster than the best conventional HTCGM employing the three-term conjugate gradient-like direction. HTCGM has the direction, $d_{n+1}^N := N(y_n) - y_n$, whereas Algorithm 3.1 has the direction in Equation (22) to converge in Fix(N) quickly. It is considered that this difference between HTCGM and Algorithm 3.1 leads us to the fast convergence of Algorithm 3.1. Figure 2 plots the behaviors of D_n for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $S = 5000$ and shows that Algorithm 3.1 converges fastest, as can be seen in Figure 1.

Let us apply HCGMs employing the conventional FR, PRP, HS, and DY formulas with $\lim_{n\to\infty} \delta_n \neq 0$ to Problem 4.1 in the above cases and see whether they converge to the solution. Unfortunately, it is not guaranteed that they converge to the solution because δ_n defined by one of Formulas (10) satisfies $\lim_{n\to\infty} \delta_n \neq 0$ when $\kappa, \eta \neq 0$ and the unique minimizer of f over $\mathbb{R}^S$ satisfying $\nabla f(x^*) = Qx^* = 0$ (i.e., $x^* = Q^{-1}0 = 0$) is not

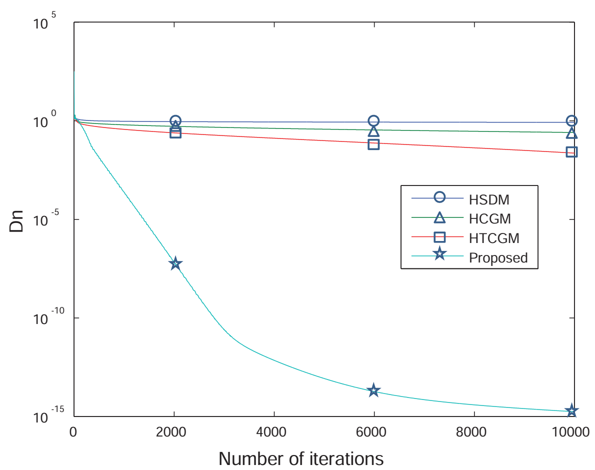

Fig. 1 Behavior of $D_n := \|x_n - x^\star\|^2$ for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $S = 1000$ and $\{x^\star\} = \mathrm{Argmin}_{x\in C_1\cap C_2} f(x)$ (CPU times to compute x_{500} in HSDM, HCGM, HTCGM, and Algorithm 3.1 are, respectively, 0.5048 s, 0.9446 s, 1.2596 s, and 2.0045 s.)

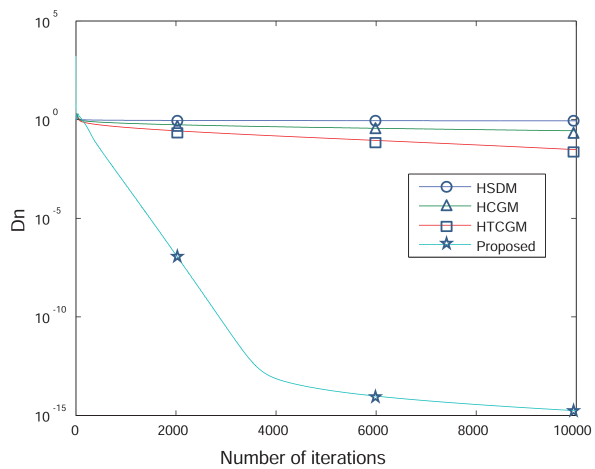

Fig. 2 Behavior of $D_n := \|x_n - x^\star\|^2$ for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $S = 5000$ and $\{x^\star\} = \mathrm{Argmin}_{x\in C_1\cap C_2} f(x)$ (CPU times to compute x_{500} in HSDM, HCGM, HTCGM, and Algorithm 3.1 are, respectively, 8.5586 s, 15.2463 s, 21.1717 s, and 33.4314 s.)

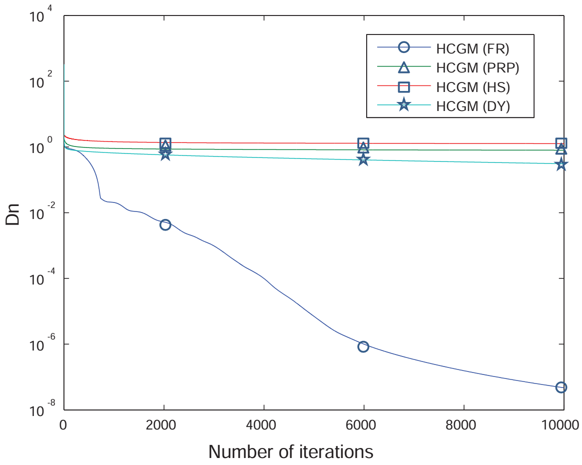

Fig. 3 Behavior of $D_n := \|x_n - x^\star\|^2$ for HCGMs with FR, PRP, HS, and DY formulas when $S = 1000$ and $\{x^\star\} = \mathrm{Argmin}_{x \in C_1 \cap C_2} f(x)$ (CPU times to compute x_{500} in FR, PRP, HS, and DY are, respectively, 1.9925 s, 2.3310 s, 2.6161 s, and 2.5345 s.)

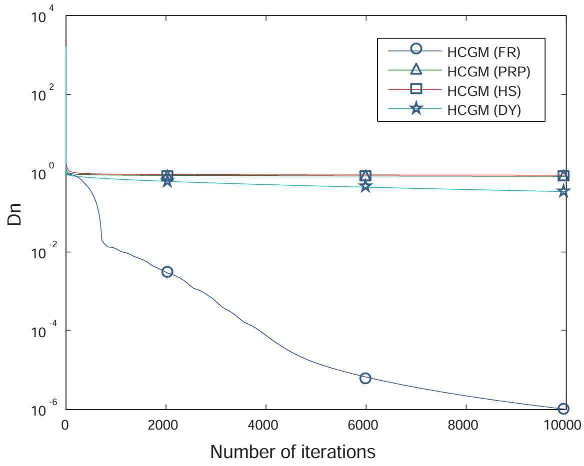

Fig. 4 Behavior of $D_n := \|x_n - x^\star\|^2$ for HCGMs with FR, PRP, HS, and DY formulas when $S = 5000$ and $\{x^\star\} = \mathrm{Argmin}_{x \in C_1 \cap C_2} f(x)$ (CPU times to compute x_{500} in FR, PRP, HS, and DY are, respectively, 25.7676 s, 29.1687 s, 32.9006 s, and 31.8607 s.)

in $C_1 \cap C_2$ (see Proposition 3.2 and Theorem 3.1). Figures 3 and 4 plot the behaviors of D_n for HCGMs with the FR, PRP, HS, and DY formulas when $\kappa = \eta := 0.01$ and $S = 1000, 5000$. We can see from these figures that $(D_n)_{n \in \mathbb{N}}$s generated by HCGMs with the PRP, HS, and DY formulas do not converge to 0, while $(D_n)_{n \in \mathbb{N}}$ generated by HCGM with the FR formula converges to 0; i.e., HCGM with the FR formula converges to the solution to Problem 4.1. To verify why HCGM with the FR formula converges to the solution, we checked the behaviors of δ_n^{FR}, δ_n^{PRR}, δ_n^{HS}, and δ_n^{DY} in the above cases. These values and behaviors were as follows: $\delta_n^{\mathrm{FR}} \approx 0.9999$, $\delta_n^{\mathrm{PRR}} \approx -0.1049, \delta_n^{\mathrm{HS}} \approx -0.9958$, and $\delta_n^{\mathrm{DY}} \approx 0.8897$ $(n \geq 2000)$. Meanwhile, HCGM, HTCGM, and Algorithm 3.1 use the slowly diminishing sequence, $\delta_n^{(i)} := 1/(n+1)^{0.01}$ $(i = 1, 2)$, in Directions (20), (21), and (23), satisfying $1/(10^4)^{0.01} \approx 0.9120$. This leads us that δ_n^{FR} is approximately $\delta_n^{(i)}$ $(i = 1, 2)$ for large enough n. Since it is guaranteed that HCGM with $\delta_n^{(1)} := 1/(n+1)^{0.01}$ converges to the solution (Theorem 4.1 in [20] or Theorem 3.1 in this paper), it is considered that HCGM with the FR formula also converges to the solution. In the case of Subsection 4.1, we can see from Figures 1–4 that HCGM with the FR formula converges faster than HCGM with $\delta_n^{(1)} := 1/(n+1)^{0.01}$. Reference [20] includes a numerical performance evaluation of HCGM associated with the choice of $\delta_n^{(1)}$. The evaluation shows that HCGM with $\delta_n^{(1)} = 1/(n+1)^{0.001}$ $(1/(10^4)^{0.001} \approx 0.9908)$ converges to the solution faster than one with $\delta_n^{(1)} = 1/(n+1)^{0.01}$ $(1/(10^4)^{0.01} \approx 0.9120)$. From this observation, it is considered that HCGM with $\delta_n^{\mathrm{FR}} \approx 0.9999 \approx 1/(10^4)^{0.001}$ converges faster than one with $\delta_n^{(1)} = 1/(n+1)^{0.01}$. In the case of $S = 1000$, the CPU time of HCGM with the FR formula satisfying $D_{6000} < 10^{-6}$ is about 23.9 s, while Algorithm 3.1 satisfies $D_{2000} < 10^{-6}$ and the CPU time of Algorithm 3.1 at 2000 iterations is about 8.1 s. This implies that Algorithm 3.1 converges to the solution fastest. We can also check from Figure 4 that, in the case of $S = 5000$, Algorithm 3.1 converges to the solution faster than HCGM with the FR formula. Therefore, we can conclude from Figures 1–4 that Algorithm 3.1 solves Problem 4.1 in the case of Subsection 4.1 faster than the existing algorithms.

4.2 Constraint set in Problem 4.1 is the generalized convex feasible set

We chose $b \in \mathbb{R}^S$ as a random number given by MATLAB, and set a box constraint set, C_3, and halfspaces, C_4 and C_5, with $C_3 \cap C_4 \cap C_5 = \emptyset$. We chose[9] a positive definite matrix, $Q \in \mathbb{R}^{S\times S}$, of which the maximum and minimum eigenvalues are, respectively, $\lambda_Q^{\max} := \lambda_Q^S = S$ and $\lambda_Q^{\min} := \lambda_Q^1 = 1$. In this case, we define a constraint set in Problem 4.1 by a subset of C_3 with the elements closest to C_4 and C_5 in terms of the mean square norm. This subset is referred to as the *generalized convex feasible set* [9, Section I, Framework 2], [31, Definition 4.1] and it is defined as follows:

$$C_\Phi := \left\{ x \in C_3 \colon \Phi(x) = \min_{y \in C_3} \Phi(y) \right\},$$

where $\Phi(x)$ stands for the mean square value of the distances from $x \in \mathbb{R}^S$ to C_4 and C_5:

$$\Phi(x) := \frac{1}{2} \sum_{k=4,5} \left(\min_{z \in C_k} \|x - z\| \right)^2 \quad \left(x \in \mathbb{R}^S\right).$$

Define $N \colon \mathbb{R}^S \to \mathbb{R}^S$ by

$$N := P_{C_3} \left[\frac{1}{2} P_{C_4} + \frac{1}{2} P_{C_5} \right].$$

Accordingly, N is nonexpansive and $\mathrm{Fix}(N) = C_\Phi$ [31, Proposition 4.2]. Since we cannot describe the solution to Problem 4.1 in this case, let us check whether the algorithms used in the experiment converge in $\mathrm{Fix}(N)$ or not. Figure 5 describes the behaviors[10] of $(\|x_n - N(x_n)\|)_{n=0}^{100}$ for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $S = 1000$. This figure shows that the behaviors of $(\|x_n - N(x_n)\|)_{n=0}^{100}$ for HSDM, HCGM, and HTCGM are

[9] We randomly chose $\lambda_Q^k \in (1, S)$ $(k = 2, 3, \ldots, S-1)$ and set $\hat{Q} \in \mathbb{R}^{S\times S}$ as a diagonal matrix with eigenvalues, $\lambda_Q^1, \lambda_Q^2, \ldots, \lambda_Q^S$. We made a positive definite matrix, $Q \in \mathbb{R}^{S\times S}$, using an orthogonal matrix and $\hat{Q}$.

[10] $x \in \mathbb{R}^S$ satisfies $\|x - N(x)\| = 0$ if and only if $x \in \mathrm{Fix}(N)$.

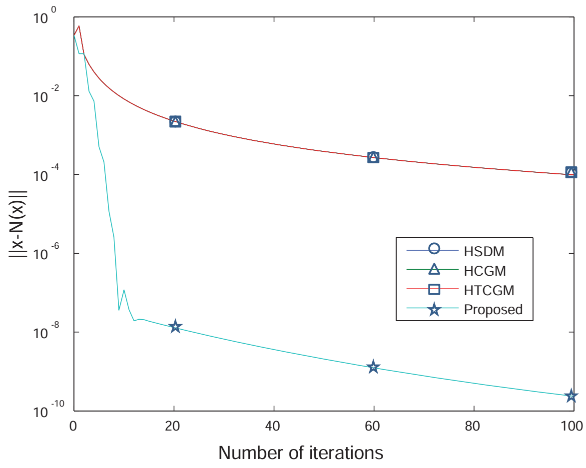

Fig. 5 Behavior of $\|x_n - N(x_n)\|$ for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $Q \in \mathbb{R}^{1000\times1000}$ is a positive definite matrix and Fix(N) is a generalized convex feasible set (CPU times to compute x_{500} in HSDM, HCGM, HTCGM, and Algorithm 3.1 are, respectively, 0.7349 s, 1.3021 s, 1.8243 s, and 2.9998 s.)

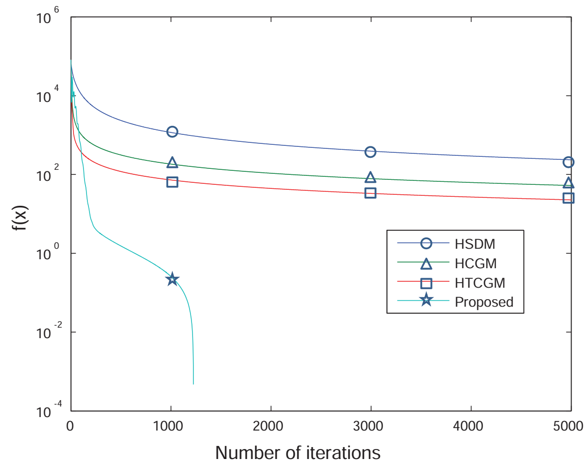

Fig. 6 Behavior of $f(x_n)$ for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $Q \in \mathbb{R}^{1000\times1000}$ is a positive definite matrix and Fix(N) is a generalized convex feasible set (Algorithm 3.1 is stable for $n \geq 2000$ and converges to $x^\star \in$ Fix(N) with $f(x^\star) \approx -1.4$.)

the same, whereas $(\|x_n - N(x_n)\|)_{n\in\mathbb{N}}$ generated by Algorithm 3.1 converges fastest to 0; i.e., Algorithm 3.1 converges fastest in Fix(N). [11] This will be because Algorithm 3.1 has Direction (22) to converge in Fix(N) quickly. The behaviors of $(f(x_n))_{n=0}^{5000}$ for the four algorithms is presented in Figure 6. We can see that Algorithm 3.1 dramatically decreases f as compared with HSDM, HCGM, and HTCGM. We checked that Algorithm 3.1 is stable for $n \geq 2000$ and converges to $x^\star$ with $f(x^\star) \approx -1.4$, and that the CPU time of Algorithm 3.1 at 2000 iterations is about 11.9 s (HSDM, HCGM, and HTCGM are not stable when the CPU time is about 11.9 s). Therefore, Figures 5 and 6 and Theorem 3.1 ensure that Algorithm 3.1 converges to the solution to Problem 4.1 faster than the existing algorithms.

Figures 7 and 8 plot the behaviors of $(\|x_n - N(x_n)\|)_{n=0}^{100}$ and $(f(x_n))_{n=0}^{5000}$ when $S = 5000$, respectively. As with the case of $S = 1000$, we can see that Algorithm 3.1 converges in Fix(N) and decreases f faster than the existing algorithms. We checked that Algorithm 3.1 is stable for $n \geq 2000$ and converges to $x^\star$ with $f(x^\star) \approx -2.5$.

Finally, we verify whether HCGMs employing the FR, PRP, HS, and DY formulas converge to the solution to Problem 4.1 under the same conditions that were discussed in Figures 5–8. Figures 9 and 10 describe the behaviors of $(f(x_n))_{n=0}^{5000}$ for HCGMs with the FR, PRP, HS, and DY formulas when

[11] See Remark 3.2 on the nonmonotonicity of $(\|x_n - N(x_n)\|)_{n\in\mathbb{N}}$ in Algorithm 3.1.

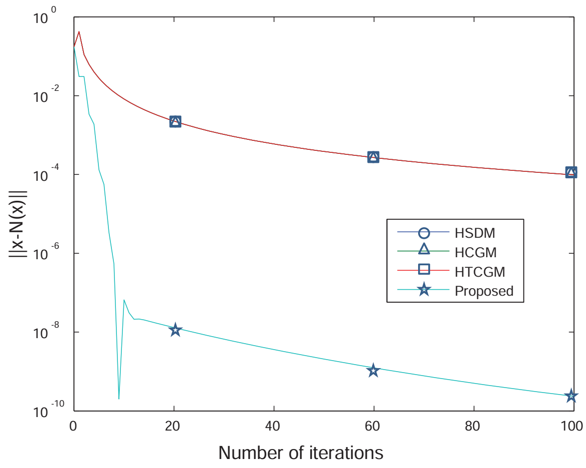

Fig. 7 Behavior of $\|x_n - N(x_n)\|$ for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $Q \in \mathbb{R}^{5000\times 5000}$ is a positive definite matrixand $\mathrm{Fix}(N)$ is a generalized convex feasible set (CPU times to compute x_{500} in HSDM, HCGM, HTCGM, and Algorithm 3.1 are, respectively, 8.7845 s, 16.9029 s, 22.1541 s, and 33.8423 s.)

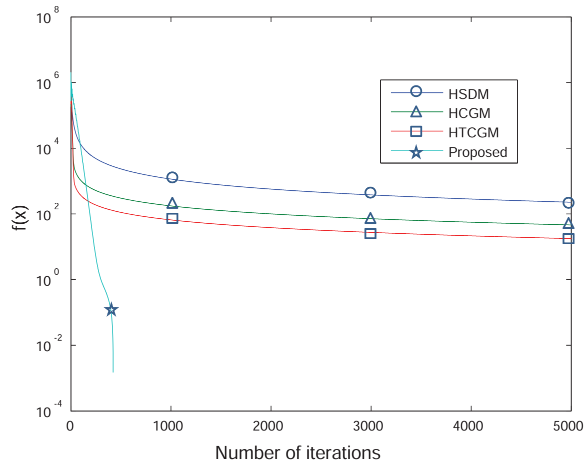

Fig. 8 Behavior of $f(x_n)$ for HSDM, HCGM, HTCGM, and Algorithm 3.1 (Proposed) when $Q \in \mathbb{R}^{5000\times 5000}$ is a positive definite matrix and $\mathrm{Fix}(N)$ is a generalized convex feasible set (Algorithm 3.1 is stable for $n \geq 2000$ and converges to $x^\star \in \mathrm{Fix}(N)$ with $f(x^\star) \approx -2.5$.)

$\kappa = \eta := 0.01$ and $S = 1000, 5000$. We can see that HCGM with the FR formula decreases f in the early stages. We checked that HCGM with the FR formula is stable for $n \geq 3000$, and converges in $\mathrm{Fix}(N)$ and to the same point to which Algorithm 3.1 converges ($f(x^\star) \approx -1.4$ when $S = 1000$, $f(x^\star) \approx -2.5$ when $S = 5000$). We also verified the behaviors and values of δ_n^{FR}, δ_n^{PRP}, δ_n^{HS}, and δ_n^{DY}: $\delta_n^{\mathrm{FR}} \approx 0.9999$, $\delta_n^{\mathrm{PRP}} \approx -0.1007$, $\delta_n^{\mathrm{HS}} \approx -0.9981$, and $\delta_n^{\mathrm{DY}} \approx 0.9076$ ($n \geq 3000$). From the same discussion in the third paragraph of Subsection 4.1, we can conclude that HCGM with the FR formula converges to the solution to Problem 4.1. Here, let us compare Algorithm 3.1 with HCGM with the FR formula. In the case of $S = 1000$, the required CPU time and number of iterations of Algorithm 3.1 are, respectively, about 11.9 s and 2000 iterations, whereas the required CPU time and number of iterations of HCGM with the FR formula are, respectively, about 15.2 s and 3000 iterations. We can also verify that, in the case of $S = 5000$, the required CPU time and number of iterations of Algorithm 3.1 are, respectively, about 135 s and 2000 iterations, whereas the required CPU time and number of iterations of HCGM with the FR formula are, respectively, about 177 s and 3000 iterations. Therefore, we can conclude form Figures 5–10 that Algorithm 3.1 solves Problem 4.1 in the case of Subsection 4.2 faster than the existing algorithms.

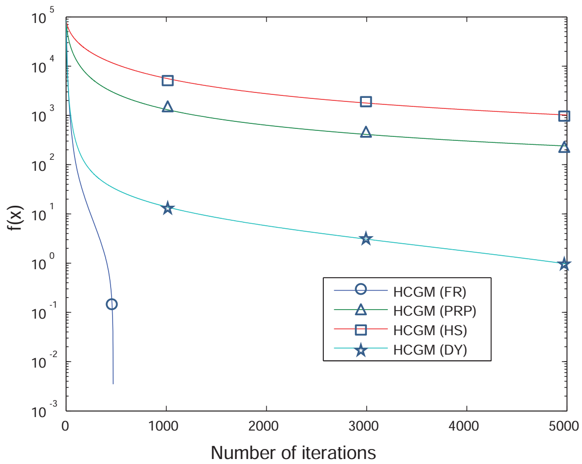

Fig. 9 Behavior of $f(x_n)$ for HCGMs with FR, PRP, HS, and DY formulas when $Q \in \mathbb{R}^{1000\times 1000}$ is a positive definite matrix and Fix(N) is a generalized convex feasible set (CPU times to compute x_{500} in FR, PRP, HS, and DY are, respectively, 2.5199 s, 2.7011 s, 3.2887 s, and 2.9388 s.)

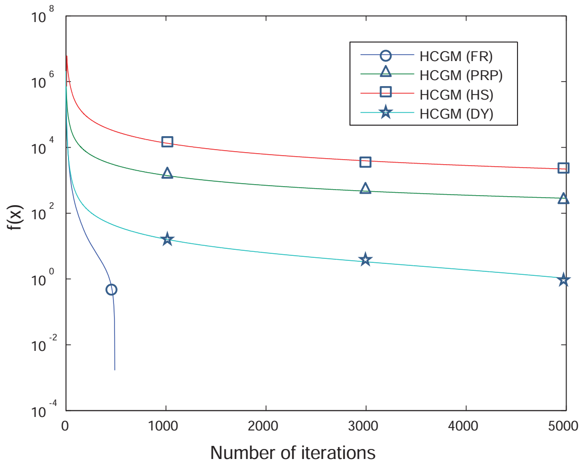

Fig. 10 Behavior of $f(x_n)$ for HCGMs with FR, PRP, HS, and DY formulas when $Q \in \mathbb{R}^{5000\times 5000}$ is a positive definite matrix and Fix(N) is a generalized convex feasible set (CPU times to compute x_{500} in FR, PRP, HS, and DY are, respectively, 29.4655 s, 33.2822 s, 34.3774 s, and 33.5166 s.)

5 Conclusion and Future Work

This paper presented a novel fixed point optimization algorithm to accelerate the existing algorithms for solving the convex minimization problem over the fixed point set of a nonexpansive mapping on a Hilbert space and showed that the algorithm with slowly diminishing step-size sequences strongly converges to the solution to the problem. It also described that conjugate gradient methods with the conventional Fletcher–Reeves, Polak–Ribiére–Polyak, Hestenes–Stiefel, and Dai–Yuan formulas do not always converge to the solution to the problem. To demonstrate the effectiveness and convergence of the algorithm, we provided numerical comparisons of the algorithm with the existing algorithms. The comparisons suggested that the algorithm is effective for solving the convex minimization problems.

The numerical examples in this paper demonstrated that, for concrete convex optimization problems, the proposed algorithm converges to the desired solutions faster than the existing algorithms. Although we need to prove that the acceleration algorithm increases the rate of convergence, it seems to be difficult to evaluate the rate of convergence for a convex optimization problem over the fixed point set of a nonexpansive mapping. This is because the constraint set is a fixed point set, which does not have a simple form; i.e., the explicit form of the metric projection onto the fixed point set is not known. However, combining the ideas of [2, Chapter 5] and [5, 24] would give a good way to evaluate the convergence rate of the proposed algorithm because reference [2, Chapter 5] discussed the convergence rate of iterative algorithms for solving nonlinear ill-posed problems with monotone operators and references [5, 24] discussed the convergence rate of algorithms for solving linear inverse problems and convex optimization problems. On the basis on these previously reported results, we first should try to evaluate the rate of convergence of the proposed algorithm for concrete convex optimization problems (e.g., a problem of minimizing a linear objective function over the generalized convex feasible set).

The problem of minimizing an objective function, which does not satisfy a strong convexity condition, over the fixed point set of a nonexpansive mapping includes important and practical engineering problems. For example, the network bandwidth allocation problem is expressed as a problem of maximizing a strictly concave or a nonconcave utility function over a certain fixed point set [17–19], the power control problem is one of maximizing a nonconcave function over a certain fixed point set [16], and the optimal control problem is one of minimizing a convex function over a certain fixed point set [21]. The algorithms presented in [16–19, 21] are based on algorithmic methods, such as the steepest descent method and conjugate gradient methods, for minimizing objective functions. Therefore, by referring to the ideas of [16–19, 21] and Algorithm 3.1 in this paper, we will be able to devise an algorithm with three-term conjugate gradient-like directions which can be applied when objective functions are convex or nonconvex. However, it was not demonstrated that the algorithm performs better in numerical experiments than the algorithms presented in [16–19, 21]. In the future, we need to apply the algorithm to real-world optimization problems in [16–19, 21] and see whether the algorithm performs better than those algorithms.

Acknowledgments

I wrote Subsection 3.2 by referring to the referee's report on the original manuscript of [20]. I am sincerely grateful to the anonymous referee that reviewed the original manuscript of [20] for helping me compile the paper. I also would like to thank the Co-Editor, Michael C. Ferris, and the two anonymous reviewers for helping me improve the original manuscript.

References

1. Aoyama, K., Kimura, Y., Takahashi, W., Toyoda, M.: On a strongly nonexpansive sequence in hilbert spaces. Journal of Nonlinear and Convex Analysis **8**, 471–489 (2007)
2. Bakushinsky, A., Goncharsky, A.: Ill-Posed Problems: Theory and Applications. Kluwer, Dordrecht (1994)
3. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Review **38**(3), 367–426 (1996)
4. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer (2011)
5. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences **2**, 183–202 (2009)

6. Borwein, J.M., Lewis, A.S.: Convex Analysis and Nonlinear Optimization: Theory and Examples. Springer (2000)
7. Cheng, W.: A two-term prp-based descent method. Numerical Functional Analysis and Optimization **28**, 1217–1230 (2007)
8. Combettes, P.L.: A block-iterative surrogate constraint splitting method for quadratic signal recovery. IEEE Transactions on Signal Processing **51**(7), 1771–1782 (2003)
9. Combettes, P.L., Bondon, P.: Hard-constrained inconsistent signal feasibility problems. IEEE Transactions on Signal Processing **47**(9), 2460–2468 (1999)
10. Dai, Y.H., Yuan, Y.: A nonlinear conjugate gradient method with a strong global convergence property. SIAM Journal on Optimization **10**, 177–182 (1999)
11. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. Classics Appl. Math. 28. SIAM (1999)
12. Goebel, K., Kirk, W.A.: Topics in Metric Fixed Point Theory. Cambridge Studies in Advanced Mathematics. Cambridge University Press (1990)
13. Goebel, K., Reich, S.: Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings. Dekker (1984)
14. Iiduka, H.: Iterative algorithm for solving triple-hierarchical constrained optimization problem. Journal of Optimization Theory and Applications **148**, 580–592 (2011)
15. Iiduka, H.: Three-term conjugate gradient method for the convex optimization problem over the fixed point set of a nonexpansive mapping. Applied Mathematics and Computation **217**, 6315–6327 (2011)
16. Iiduka, H.: Fixed point optimization algorithm and its application to power control in cdma data networks. Mathematical Programming **133**(1), 227–242 (2012)
17. Iiduka, H.: Iterative algorithm for triple-hierarchical constrained nonconvex optimization problem and its application to network bandwidth allocation. SIAM Journal on Optimization **22**(3), 862–878 (2012)
18. Iiduka, H.: Fixed point optimization algorithms for distributed optimization in networked systems. SIAM Journal on Optimization **23**(1), 1–26 (2013)
19. Iiduka, H., Uchida, M.: Fixed point optimization algorithms for network bandwidth allocation problems with compoundable constraints. IEEE Communications Letters **15**(6), 596–598 (2011)
20. Iiduka, H., Yamada, I.: A use of conjugate gradient direction for the convex optimization problem over the fixed point set of a nonexpansive mapping. SIAM Journal on Optimization **19**(4), 1881–1893 (2009)
21. Iiduka, H., Yamada, I.: Computational method for solving a stochastic linear-quadratic control problem given an unsolvable stochastic algebraic riccati equation. SIAM Journal on Control and Optimization **50**(4), 2173–2192 (2012)
22. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and Their Applications. Classics Appl. Math. 31. SIAM (2000)
23. Narushima, Y., Yabe, H., Ford, J.A.: A three-term conjugate gradient method with sufficient descent property for unconstrained optimization. SIAM Journal on Optimization **21**, 212–230 (2011)
24. Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate $o(1/k^2)$. Doklady Akademii Nauk SSSR **269**, 543–547 (1983)
25. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research and Financial Engineering. Springer (1999)
26. Opial, Z.: Weak convergence of the sequence of successive approximation for nonexpansive mappings. Bulletin of the American Mathematical Society **73**, 591–597 (1967)
27. Slavakis, K., Yamada, I.: Robust wideband beamforming by the hybrid steepest descent method. IEEE Transactions on Signal Processing **55**(9), 4511–4522 (2007)
28. Stark, H., Yang, Y.: Vector Space Projections: A Numerical Approach to Signal and Image Processing. John Wiley & Sons Inc (1998)
29. Takahashi, W.: Nonlinear Functional Analysis. Yokohama Publishers (2000)
30. Wolfe, P.: Finding the nearest point in a polytope. Mathematical Programming **11**, 128–149 (1976)
31. Yamada, I.: The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In: D. Butnariu, Y. Censor, S. Reich (eds.) Inherently Parallel Algorithms for Feasibility and Optimization and Their Applications, pp. 473–504. Elsevier (2001)

32. Yamada, I., Ogura, N., Shirakawa, N.: A numerical robust hybrid steepest descent method for the convexly constrained generalized inverse problems. Contemporary Mathematics **313**, 269–305 (2002)
33. Zeidler, E.: Nonlinear Functional Analysis ans Its Applications III. Variational Methods and Optimization. Springer (1985)
34. Zhang, L., Zhou, W., Li, D.H.: A descent modified polak-ribiére-polyak conjugate gradient method and its global convergence. IMA Journal of Numerical Analysis **26**, 629–640 (2006)
35. Zhang, L., Zhou, W., Li, D.H.: Global convergence of a modified fletcher-reeves conjugate gradient method with armijo-type line search. Numerische Mathematik **104**, 561–572 (2006)
36. Zhang, L., Zhou, W., Li, D.H.: Some descent three-term conjugate gradient methods and their global convergence. Optimization Methods and Software **22**, 697–711 (2007)