RESEARCH   **Open Access**

CrossMark

Line search fixed point algorithms based on nonlinear conjugate gradient directions: application to constrained smooth convex optimization

Hideaki Iiduka[*]

[*]Correspondence:
iiduka@cs.meiji.ac.jp
Department of Computer Science,
Meiji University, 1-1-1 Higashimita,
Tama-ku, Kawasaki-shi, Kanagawa,
214-8571, Japan

Abstract

This paper considers the fixed point problem for a nonexpansive mapping on a real Hilbert space and proposes novel line search fixed point algorithms to accelerate the search. The termination conditions for the line search are based on the well-known Wolfe conditions that are used to ensure the convergence and stability of unconstrained optimization algorithms. The directions to search for fixed points are generated by using the ideas of the steepest descent direction and conventional nonlinear conjugate gradient directions for unconstrained optimization. We perform convergence as well as convergence rate analyses on the algorithms for solving the fixed point problem under certain assumptions. The main contribution of this paper is to make a concrete response to an issue of constrained smooth convex optimization; that is, whether or not we can devise nonlinear conjugate gradient algorithms to solve constrained smooth convex optimization problems. We show that the proposed fixed point algorithms include ones with nonlinear conjugate gradient directions which can solve constrained smooth convex optimization problems. To illustrate the practicality of the algorithms, we apply them to concrete constrained smooth convex optimization problems, such as constrained quadratic programming problems and generalized convex feasibility problems, and numerically compare them with previous algorithms based on the Krasnosel'skiĭ-Mann fixed point algorithm. The results show that the proposed algorithms dramatically reduce the running time and iterations needed to find optimal solutions to the concrete optimization problems compared with the previous algorithms.

MSC: 47H10; 65K05; 90C25

Keywords: constrained smooth convex optimization; fixed point problem; generalized convex feasibility problem; Krasnosel'skiĭ-Mann fixed point algorithm; line search method; nonexpansive mapping; nonlinear conjugate gradient methods

1 Introduction

Consider the following *fixed point problem* (see [1], Chapter 4, [2], Chapter 3, [3], Chapter 1, [4], Chapter 3):

$$\text{Find } x^\star \in \text{Fix}(T) := \{x^\star \in H : T(x^\star) = x^\star\}, \tag{1.1}$$

where H stands for a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, T is a *nonexpansive* mapping from H into itself (*i.e.*, $\|T(x) - T(y)\| \leq \|x - y\|$ ($x, y \in H$)), and one assumes $\text{Fix}(T) \neq \emptyset$. Problem (1.1) includes convex feasibility problems [5], [1], Example 5.21, constrained smooth convex optimization problems [6], Proposition 4.2, problems of finding the zeros of monotone operators [1], Proposition 23.38, and monotone variational inequalities [1], Subchapter 25.5.

There are useful algorithms for solving Problem (1.1), such as the *Krasnosel'skiĭ-Mann algorithm* [1], Subchapter 5.2, [7], Subchapter 1.2, [8–10], the *Halpern algorithm* [7], Subchapter 1.2, [11, 12], and the *hybrid method* [13] (Solodov and Svaiter [14] proposed the hybrid method to solve problems of finding the zeros of monotone operators). This paper focuses on the Krasnosel'skiĭ-Mann algorithm, which has practical applications, such as analyses of dynamic systems governed by maximal monotone operators [15] and nonsmooth convex variational signal recovery [16], defined as follows: given the current iterate $x_n \in H$ and step size $\alpha_n \in [0, 1]$, the next iterate x_{n+1} of the algorithm is

$$x_{n+1} := x_n + \alpha_n \big(T(x_n) - x_n \big). \tag{1.2}$$

Assuming that $(\alpha_n)_{n \in \mathbb{N}}$ satisfies the condition

$$\sum_{n=0}^{\infty} \alpha_n (1 - \alpha_n) = \infty, \tag{1.3}$$

the sequence $(x_n)_{n \in \mathbb{N}}$ generated by Algorithm (1.2) weakly converges to a fixed point of T (see, *e.g.*, [1], Theorem 5.14). This result indicates that Algorithm (1.2) with constant step sizes (*e.g.*, $\alpha_n := \alpha \in (0, 1)$ ($n \in \mathbb{N}$)) or diminishing step sizes (*e.g.*, $\alpha_n := 1/(n+1)$ ($n \in \mathbb{N}$)) can solve Problem (1.1). Propositions 10 and 11 in [8] indicate that Algorithm (1.2) with condition (1.3) has the following rate of convergence: for all $n \in \mathbb{N}$,

$$\big\| x_n - T(x_n) \big\| = O\left(\left\{ \sum_{k=0}^{n} \alpha_k (1 - \alpha_k) \right\}^{-\frac{1}{2}} \right) \tag{1.4}$$

(*e.g.*, $\|x_n - T(x_n)\| = O(1/\sqrt{n+1})$ when $\alpha_n := \alpha \in (0, 1)$ ($n \in \mathbb{N}$)). This fact implies that Algorithm (1.2) with (1.3) does not always have fast convergence and has motivated the development of modifications and variants for the Krasnosel'skiĭ-Mann algorithm in order to accelerate Algorithm (1.2).

One approach to accelerate Algorithm (1.2) with (1.3) is to develop line search methods that can determine a more adequate step size than a step size satisfying (1.3) at each iteration n so that the value of $\|x_{n+1} - T(x_{n+1})\|$ decreases dramatically. Magnanti and Perakis proposed an *adaptive line search framework* [17], Section 2, that can determine step sizes to satisfy weaker conditions [17], Assumptions A1 and A2, than (1.3). On the basis of this framework, they showed that Algorithm (1.2), with step sizes α_n satisfying the following *Armijo-type* condition, converges to a fixed point of T [17], Theorems 4 and 8: given $x_n \in \mathbb{R}^N$, $\beta > 0$, $D > 0$, and $b \in (0, 1)$, choose the smallest nonnegative integer l_n so that $\alpha_n = b^{l_n}$ satisfies the condition

$$g_n(\alpha_n) - g_n(0) \leq -D b^{l_n} \big\| T(x_n) - x_n \big\|^2, \tag{1.5}$$

where $g_n \colon [0,1] \to \mathbb{R}$ is a potential function [17], Scheme IV, defined for all $\alpha \in [0,1]$ by

$$g_n(\alpha) := \left\| \left(x_n + \alpha \left(T(x_n) - x_n \right) \right) - T \left(x_n + \alpha \left(T(x_n) - x_n \right) \right) \right\|^2$$
$$- \beta \alpha (1 - \alpha) \left\| T(x_n) - x_n \right\|^2. \tag{1.6}$$

Theorem 5 in [17] shows that Algorithm (1.2) with the Armijo-type condition (1.5) satisfies $\|x_{n+1} - T(x_{n+1})\|^2 \leq [1 - \beta(\alpha_n - 1/2)^2] \|x_n - T(x_n)\|^2$ ($n \in \mathbb{N}$), which implies that the algorithm has, for all $n \in \mathbb{N}$,

$$\left\| x_n - T(x_n) \right\| = O\left(\left\{ \sum_{k=0}^{n} \left(\alpha_k - \frac{1}{2} \right)^2 \right\}^{-\frac{1}{2}} \right). \tag{1.7}$$

In this paper, we introduce a line search framework using P_n defined by (1.8), (1.9), and (1.10), which is the simplest of all potential functions including g_n defined as in (1.6): given $x_n, d_n \in H$, for all $\alpha \in [0,1]$,

$$x_n(\alpha) := x_n + \alpha d_n, \tag{1.8}$$

$$Q_n(\alpha) := x_n(\alpha) - T\left(x_n(\alpha) \right), \tag{1.9}$$

$$P_n(\alpha) := \left\| Q_n(\alpha) \right\|^2. \tag{1.10}$$

When $d_n := -(x_n - T(x_n))$ and α_n is given as in (1.3), the point $x_n(\alpha_n)$ in (1.8) coincides with x_{n+1} defined by Algorithm (1.2) with (1.3).

Consider the following problem of minimizing P_n over $[0,1]$:

$$\text{Find } \alpha_n \in [0,1] \text{ such that } P_n(\alpha_n) = \min_{\alpha \in [0,1]} P_n(\alpha). \tag{1.11}$$

When the solution α_n to Problem (1.11) can be obtained in each iteration, $P_n(\alpha_n) \leq P_n(0)$ holds for all $n \in \mathbb{N}$. Accordingly, if the next iterate x_{n+1} is defined by $x_{n+1} := x_n(\alpha_n)$, $\|x_{n+1} - T(x_{n+1})\| \leq \|x_n - T(x_n)\|$ ($n \in \mathbb{N}$) holds, *i.e.*, $(\|x_n - T(x_n)\|)_{n \in \mathbb{N}}$ is monotone decreasing. Since the exact solution to Problem (1.11) cannot easily be obtained, the step size α_n can be chosen so as to yield an approximate minimum for Problem (1.11) in each iteration, specifically, to satisfy the following *Wolfe-type* conditions [18, 19]: given $x_n, d_n \in H$, and $\delta, \sigma \in (0,1)$ with $\delta \leq \sigma$,

$$P_n(\alpha_n) - P_n(0) \leq \delta \alpha_n \langle Q_n(0), d_n \rangle, \tag{1.12}$$

$$\langle Q_n(\alpha_n), d_n \rangle \geq \sigma \langle Q_n(0), d_n \rangle. \tag{1.13}$$

Condition (1.12) is the Armijo-type condition for P_n (see (1.5) for the Armijo-type condition with $d_n := -(x_n - T(x_n))$ for the potential function g_n). Under the conditions that $d_n := -(x_n - T(x_n))$ and $x_{n+1} := x_n(\alpha_n)$ ($n \in \mathbb{N}$), Algorithm (1.2) with (1.12) satisfies $\|x_{n+1} - T(x_{n+1})\|^2 \leq (1 - \delta \alpha_n) \|x_n - T(x_n)\|^2$ ($n \in \mathbb{N}$), which implies that, for all $n \in \mathbb{N}$,[a]

$$\left\| x_n - T(x_n) \right\| = O\left(\left\{ \sum_{k=0}^{n} \alpha_k \right\}^{-\frac{1}{2}} \right). \tag{1.14}$$

Here, let us see how the step size conditions (1.3), (1.5), (1.12), and (1.13) affect the efficiency of Algorithm (1.2). Algorithm (1.2) with (1.3) satisfies $\|x_{n+1} - T(x_{n+1})\|^2 \leq \|x_n - T(x_n)\|^2$ ($n \in \mathbb{N}$) [1], (5.14), while Algorithm (1.2) with each of (1.5) and (1.12) satisfies $\|x_{n+1} - T(x_{n+1})\|^2 < \|x_n - T(x_n)\|^2$ ($n \in \mathbb{N}$). Hence, it can be expected that Algorithm (1.2) with each of (1.5) and (1.12) performs better than Algorithm (1.2) with (1.3). Since the Armijo-type conditions (1.5) and (1.12) are satisfied for all sufficiently small values of α_n [20], Subchapter 3.1, there is a possibility that Algorithm (1.2) with only the Armijo-type condition (1.5) does not make reasonable progress. Meanwhile, (1.13) based on the *curvature condition* [20], Subchapter 3.1, is used to ensure that α_n is not too small and that unacceptably short steps are ruled out. Therefore, the Wolfe-type conditions (1.12) and (1.13) should be used to secure efficiency of the algorithm. Moreover, even when α_n satisfying (1.5) is not small enough, it can be expected that Algorithm (1.2) with the Wolfe-type conditions (1.12) and (1.13) will have a better convergence rate than Algorithm (1.2) with the Armijo-type condition (1.5) because of (1.7), (1.14), and $(\alpha - 1/2)^2 \leq \alpha$ ($\alpha \in [(2 - \sqrt{3})/2, 1]$). Section 3 introduces the line search algorithm [21], Algorithm 4.6, to compute step sizes satisfying (1.12) and (1.13) with appropriately chosen δ and σ and gives performance comparisons of Algorithm (1.2) with each of (1.3) and (1.5) with the one with (1.12) and (1.13).

The main concern regarding this line search is how the direction d_n should be updated to accelerate the search for a fixed point of T. To address this concern, the following problem will be discussed:

$$\text{Minimize } f(x) \text{ subject to } x \in H, \tag{1.15}$$

where $f\colon H \to \mathbb{R}$ is convex and Fréchet differentiable and $\nabla f\colon H \to H$ is Lipschitz continuous with a constant L. Let us define $T^{(f)}\colon H \to H$ by

$$T^{(f)} := \mathrm{Id} - \lambda \nabla f, \tag{1.16}$$

where Id stands for the identity mapping on H and $\lambda > 0$. The mapping $T^{(f)}$ satisfies the nonexpansivity condition for $\lambda \in (0, 2/L)$ [22], Proposition 2.3, and $\mathrm{Fix}(T^{(f)})$ coincides with the solution set of Problem (1.15). From $T^{(f)}(x) - x = (x - \lambda \nabla f(x)) - x = -\lambda \nabla f(x)$ ($\lambda > 0$, $x \in H$), Algorithm (1.2) for solving Problem (1.15) is

$$x_{n+1} = x_n + \alpha_n \bigl(T^{(f)}(x_n) - x_n \bigr) = x_n - \lambda \alpha_n \nabla f(x_n). \tag{1.17}$$

This means that the direction $d_n^{(f)} := -(x_n - T^{(f)}(x_n)) = -\lambda \nabla f(x_n)$ is the *steepest descent direction* of f at x_n and Algorithm (1.2) with $T^{(f)}$ (*i.e.*, Algorithm (1.17)) is the *steepest descent method* [20], Subchapter 3.3, for Problem (1.15).

There are many algorithms with useful search directions [20], Chapters 5-19, to accelerate the steepest descent method for unconstrained optimization. In particular, algorithms with *nonlinear conjugate gradient directions* [23], [20], Subchapter 5.2,

$$d_{n+1}^{(f)} := -\nabla f(x_{n+1}) + \beta_n d_n^{(f)}, \tag{1.18}$$

where $\beta_n \in \mathbb{R}$, have been widely used as efficient accelerated versions for most gradient methods. Well-known formulas for β_n include the Hestenes-Stiefel (HS) [24], Fletcher-

Reeves (FR) [25], Polak-Ribière-Polyak (PRP) [26, 27], and Dai-Yuan (DY) [28] formulas:

$$\beta_n^{\mathrm{HS}} := \frac{\langle \nabla f(x_{n+1}), y_n \rangle}{\langle d_n, y_n \rangle}, \qquad \beta_n^{\mathrm{FR}} := \frac{\|\nabla f(x_{n+1})\|^2}{\|\nabla f(x_n)\|^2},$$

$$\beta_n^{\mathrm{PRP}} := \frac{\langle \nabla f(x_{n+1}), y_n \rangle}{\|\nabla f(x_n)\|^2}, \qquad \beta_n^{\mathrm{DY}} := \frac{\|\nabla f(x_{n+1})\|^2}{\langle d_n, y_n \rangle}, \tag{1.19}$$

where $y_n := \nabla f(x_{n+1}) - \nabla f(x_n)$.

Motivated by these observations, we decided to use the following direction to accelerate the search for a fixed point of T, which can be obtained by replacing ∇f in (1.18) with $\mathrm{Id} - T$ (see also (1.16) for the relationship between ∇f and $T^{(f)}$): given the current direction $d_n \in H$, the current iterate $x_n \in H$, and a step size α_n satisfying (1.12) and (1.13), the next direction d_{n+1} is defined by

$$d_{n+1} := -\big(x_{n+1} - T(x_{n+1})\big) + \beta_n d_n, \tag{1.20}$$

where β_n is given by one of the formulas in (1.19) when $\nabla f = \mathrm{Id} - T$.

This paper proposes iterative algorithms (Algorithm 2.1) that use the direction (1.20) and step sizes satisfying the Wolfe-type conditions (1.12) and (1.13) for solving Problem (1.1) and describes their convergence analyses (Theorems 2.1-2.5). We also provide their convergence rate analyses (Theorem 2.6).

The main contribution of this paper is to enable us to propose nonlinear conjugate gradient algorithms for *constrained smooth convex optimization* which are examples of the proposed line search fixed point algorithms, in contrast to the previously reported results for nonlinear conjugate gradient algorithms for unconstrained smooth nonconvex optimization [20], Subchapter 5.2, [23–29]. Concretely speaking, our nonlinear conjugate gradient algorithms are obtained in the following steps. Given a nonempty, closed, and convex set $C \subset H$ and a convex function $f \colon H \to \mathbb{R}$ with the Lipschitz continuous gradient, let us define

$$T := P_C(\mathrm{Id} - \lambda \nabla f),$$

where $\lambda \in (0, 2/L]$, L is the Lipschitz constant of ∇f, and P_C stands for the metric projection onto C. Then Proposition 2.3 in [22] indicates that the mapping T is nonexpansive and satisfies

$$\mathrm{Fix}(T) = \operatorname*{argmin}_{x \in C} f(x).$$

From (1.20) with $T := P_C(\mathrm{Id} - \lambda \nabla f)$, the proposed nonlinear conjugate gradient algorithms for finding a point in $\mathrm{Fix}(T) = \operatorname{argmin}_{x \in C} f(x)$ can be expressed as follows: given $x_n, d_n \in H$ and α_n satisfying (1.12) and (1.13),

$$x_{n+1} := x_n(\alpha_n) = x_n + \alpha_n d_n,$$

$$d_{n+1} := -\big(x_{n+1} - P_C\big(x_{n+1} - \lambda \nabla f(x_{n+1})\big)\big) + \beta_n d_n,$$

where $\beta_n \in \mathbb{R}$ is each of the following formulas:[b]

$$
\begin{aligned}
\beta_n^{\text{HS+}} &:= \max\left\{ \frac{\langle x_{n+1} - P_C(x_{n+1} - \lambda \nabla f(x_{n+1})), y_n \rangle}{\langle d_n, y_n \rangle}, 0 \right\}, \\
\beta_n^{\text{FR}} &:= \frac{\|x_{n+1} - P_C(x_{n+1} - \lambda \nabla f(x_{n+1}))\|^2}{\|x_n - P_C(x_n - \lambda \nabla f(x_n))\|^2}, \\
\beta_n^{\text{PRP+}} &:= \max\left\{ \frac{\langle x_{n+1} - P_C(x_{n+1} - \lambda \nabla f(x_{n+1})), y_n \rangle}{\|x_n - P_C(x_n - \lambda \nabla f(x_n))\|^2}, 0 \right\}, \\
\beta_n^{\text{DY}} &:= \frac{\|x_{n+1} - P_C(x_{n+1} - \lambda \nabla f(x_{n+1}))\|^2}{\langle d_n, y_n \rangle},
\end{aligned}
\tag{1.21}
$$

where $y_n := (x_{n+1} - P_C(x_{n+1} - \lambda \nabla f(x_{n+1}))) - (x_n - P_C(x_n - \lambda \nabla f(x_n)))$. Our convergence analyses are performed by referring to useful results on unconstrained smooth nonconvex optimization (see [18, 19, 23, 28, 30–32] and references therein) because the proposed fixed point algorithms are based on the steepest descent and nonlinear conjugate gradient directions for unconstrained smooth nonconvex optimization (see (1.15)-(1.20)). We would like to emphasize that combining unconstrained smooth nonconvex optimization theory with fixed point theory for nonexpansive mappings enables us to develop the novel nonlinear conjugate gradient algorithms for constrained smooth convex optimization. The nonlinear conjugate gradient algorithms are a concrete response to the issue of constrained smooth convex optimization that is whether or not we can present nonlinear conjugate gradient algorithms to solve constrained smooth convex optimization problems.

To verify whether the proposed nonlinear conjugate gradient algorithms are accelerations for solving practical problems, we apply them to constrained quadratic programming problems (Section 3.2) and generalized convex feasibility problems (Section 3.3) (see [6, 33] and references therein for the relationship between the generalized convex feasibility problem and signal processing problems), which are constrained smooth convex optimization problems and particularly interesting applications of Problem (1.1). Moreover, we numerically compare their abilities to solve concrete constrained quadratic programming problems and generalized convex feasibility problems with those of previous algorithms based on the Krasnosel'skiĭ-Mann algorithm (Algorithm (1.2) with step sizes satisfying (1.3) and Algorithm (1.2) with step sizes satisfying (1.5)) and show that they can find optimal solutions to these problems faster than the previous ones.

Throughout this paper, we shall let $\mathbb{N}$ be the set of zero and all positive integers, $\mathbb{R}^d$ be a d-dimensional Euclidean space, H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, and $T \colon H \to H$ be a nonexpansive mapping with $\text{Fix}(T) := \{x \in H \colon T(x) = x\} \neq \emptyset$.

2 Line search fixed point algorithms based on nonlinear conjugate gradient directions

Let us begin by explicitly stating our algorithm for solving Problem (1.1) discussed in Section 1.

Algorithm 2.1

Step 0. Take $\delta, \sigma \in (0, 1)$ with $\delta \leq \sigma$. Choose $x_0 \in H$ arbitrarily and set $d_0 := -(x_0 - T(x_0))$ and $n := 0$.

Step 1. Compute $\alpha_n \in (0,1]$ satisfying

$$\left\| x_n(\alpha_n) - T\left(x_n(\alpha_n)\right) \right\|^2 - \left\| x_n - T(x_n) \right\|^2 \leq \delta\alpha_n \langle x_n - T(x_n), d_n \rangle, \tag{2.1}$$

$$\langle x_n(\alpha_n) - T\left(x_n(\alpha_n)\right), d_n \rangle \geq \sigma \langle x_n - T(x_n), d_n \rangle, \tag{2.2}$$

where $x_n(\alpha_n) := x_n + \alpha_n d_n$. Compute $x_{n+1} \in H$ by

$$x_{n+1} := x_n + \alpha_n d_n. \tag{2.3}$$

Step 2. If $\|x_{n+1} - T(x_{n+1})\| = 0$, stop. Otherwise, go to Step 3.

Step 3. Compute $\beta_n \in \mathbb{R}$ by using each of the following formulas:

$$\beta_n^{\mathrm{SD}} := 0,$$

$$\beta_n^{\mathrm{HS+}} := \max\left\{ \frac{\langle x_{n+1} - T(x_{n+1}), y_n \rangle}{\langle d_n, y_n \rangle}, 0 \right\}, \qquad \beta_n^{\mathrm{FR}} := \frac{\|x_{n+1} - T(x_{n+1})\|^2}{\|x_n - T(x_n)\|^2}, \tag{2.4}$$

$$\beta_n^{\mathrm{PRP+}} := \max\left\{ \frac{\langle x_{n+1} - T(x_{n+1}), y_n \rangle}{\|x_n - T(x_n)\|^2}, 0 \right\}, \qquad \beta_n^{\mathrm{DY}} := \frac{\|x_{n+1} - T(x_{n+1})\|^2}{\langle d_n, y_n \rangle},$$

where $y_n := (x_{n+1} - T(x_{n+1})) - (x_n - T(x_n))$. Generate $d_{n+1} \in H$ by

$$d_{n+1} := -\left(x_{n+1} - T(x_{n+1})\right) + \beta_n d_n.$$

Step 4. Put $n := n + 1$ and go to Step 1.

We need to use appropriate line search algorithms to compute α_n ($n \in \mathbb{N}$) satisfying (2.1) and (2.2). In Section 3, we use a useful one (Algorithm 3.1) [21], Algorithm 4.6, that can obtain the step sizes satisfying (2.1) and (2.2) whenever the line search algorithm terminates [21], Theorem 4.7. Although the efficiency of the line search algorithm depends on the parameters δ and σ, thanks to the reference [21], Section 6.1, we can set appropriate δ and σ before executing it [21], Algorithm 4.6, and Algorithm 2.1. See Section 3 for the numerical performance of the line search algorithm [21], Algorithm 4.6, and Algorithm 2.1.

It can be seen that Algorithm 2.1 is well defined when β_n is defined by β_n^{SD}, β_n^{FR}, or $\beta_n^{\mathrm{PRP+}}$. The discussion in Section 2.2 shows that Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{DY}}$ is well defined (Lemma 2.3(i)). Moreover, it is guaranteed that under certain assumptions, Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{HS+}}$ is well defined (Theorem 2.5).

2.1 Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{SD}}$

This subsection considers Algorithm 2.1 with β_n^{SD} ($n \in \mathbb{N}$), which is based on the steepest descent (SD) direction (see (1.17)), *i.e.*,

$$x_{n+1} := x_n + \alpha_n \left(T(x_n) - x_n \right) \quad (n \in \mathbb{N}). \tag{2.5}$$

Theorems 4 and 8 in [17] indicate that, if $(\alpha_n)_{n \in \mathbb{N}}$ satisfies the Armijo-type condition (1.5), Algorithm (2.5) converges to a fixed point of T. The following theorem says that Algorithm (2.5), with $(\alpha_n)_{n \in \mathbb{N}}$ satisfying the Wolfe-type conditions (2.1) and (2.2), converges to a fixed point of T.

Theorem 2.1 *Suppose that $(x_n)_{n \in \mathbb{N}}$ is the sequence generated by Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{SD}}$ ($n \in \mathbb{N}$). Then $(x_n)_{n \in \mathbb{N}}$ either terminates at a fixed point of T or*

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

In the latter situation, $(x_n)_{n \in \mathbb{N}}$ weakly converges to a fixed point of T.

2.1.1 Proof of Theorem 2.1

If $m \in \mathbb{N}$ exists such that $\|x_m - T(x_m)\| = 0$, Theorem 2.1 holds. Accordingly, it can be assumed that, for all $n \in \mathbb{N}$, $\|x_n - T(x_n)\| \neq 0$ holds.

First, the following lemma can be proven by referring to [18, 19, 32].

Lemma 2.1 *Let $(x_n)_{n \in \mathbb{N}}$ and $(d_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 2.1. Assume that $\langle x_n - T(x_n), d_n \rangle < 0$ for all $n \in \mathbb{N}$. Then*

$$\sum_{n=0}^{\infty} \left(\frac{\langle x_n - T(x_n), d_n \rangle}{\|d_n\|} \right)^2 < \infty.$$

Proof The Cauchy-Schwarz inequality and the triangle inequality ensure that, for all $n \in \mathbb{N}$, $\langle d_n, (x_{n+1} - T(x_{n+1})) - (x_n - T(x_n)) \rangle \leq \|d_n\| \|(x_{n+1} - T(x_{n+1})) - (x_n - T(x_n))\| \leq \|d_n\| (\|T(x_n) - T(x_{n+1})\| + \|x_{n+1} - x_n\|)$, which, together with the nonexpansivity of T and (2.3), implies that, for all $n \in \mathbb{N}$,

$$\left\langle d_n, \left(x_{n+1} - T(x_{n+1}) \right) - \left(x_n - T(x_n) \right) \right\rangle \leq 2\alpha_n \|d_n\|^2.$$

Moreover, (2.2) means that, for all $n \in \mathbb{N}$,

$$\left\langle d_n, \left(x_{n+1} - T(x_{n+1}) \right) - \left(x_n - T(x_n) \right) \right\rangle \geq (\sigma - 1)\langle d_n, x_n - T(x_n) \rangle.$$

Accordingly, for all $n \in \mathbb{N}$,

$$(\sigma - 1)\langle d_n, x_n - T(x_n) \rangle \leq 2\alpha_n \|d_n\|^2.$$

Since $\|d_n\| \neq 0$ ($n \in \mathbb{N}$) holds from $\langle x_n - T(x_n), d_n \rangle < 0$ ($n \in \mathbb{N}$), we find that, for all $n \in \mathbb{N}$,

$$\frac{(\sigma - 1)\langle d_n, x_n - T(x_n) \rangle}{2\|d_n\|^2} \leq \alpha_n. \tag{2.6}$$

Condition (2.1) means that, for all $n \in \mathbb{N}$, $\|x_{n+1} - T(x_{n+1})\|^2 - \|x_n - T(x_n)\|^2 \leq \delta\alpha_n \langle x_n - T(x_n), d_n \rangle$, which, together with $\langle x_n - T(x_n), d_n \rangle < 0$ ($n \in \mathbb{N}$), implies that, for all $n \in \mathbb{N}$,

$$\alpha_n \leq \frac{\|x_n - T(x_n)\|^2 - \|x_{n+1} - T(x_{n+1})\|^2}{-\delta\langle x_n - T(x_n), d_n \rangle}. \tag{2.7}$$

From (2.6) and (2.7), for all $n \in \mathbb{N}$,

$$\frac{(\sigma - 1)\langle d_n, x_n - T(x_n) \rangle}{2\|d_n\|^2} \leq \frac{\|x_n - T(x_n)\|^2 - \|x_{n+1} - T(x_{n+1})\|^2}{-\delta\langle x_n - T(x_n), d_n \rangle},$$

which implies that, for all $n \in \mathbb{N}$,

$$\frac{\delta(1-\sigma)\langle d_n, x_n - T(x_n)\rangle^2}{2\|d_n\|^2} \leq \|x_n - T(x_n)\|^2 - \|x_{n+1} - T(x_{n+1})\|^2.$$

Summing up this inequality from $n = 0$ to $n = N \in \mathbb{N}$ guarantees that, for all $N \in \mathbb{N}$,

$$\frac{\delta(1-\sigma)}{2} \sum_{n=0}^{N} \frac{\langle d_n, x_n - T(x_n)\rangle^2}{\|d_n\|^2} \leq \|x_0 - T(x_0)\|^2 - \|x_{N+1} - T(x_{N+1})\|^2$$

$$\leq \|x_0 - T(x_0)\|^2 < \infty.$$

Therefore, the conclusion in Lemma 2.1 is satisfied. □

Lemma 2.1 leads to the following.

Lemma 2.2 *Suppose that the assumptions in Theorem* 2.1 *are satisfied. Then*:
   (i) $\lim_{n \to \infty} \|x_n - T(x_n)\| = 0$.
   (ii) $(\|x_n - x\|)_{n \in \mathbb{N}}$ *is monotone decreasing for all* $x \in \text{Fix}(T)$.
   (iii) $(x_n)_{n \in \mathbb{N}}$ *weakly converges to a point in* $\text{Fix}(T)$.

Items (i) and (iii) in Lemma 2.2 indicate that Theorem 2.1 holds under the assumption that $\|x_n - T(x_n)\| \neq 0$ ($n \in \mathbb{N}$).

Proof (i) In the case where $\beta_n := \beta_n^{\text{SD}} = 0$ ($n \in \mathbb{N}$), $d_n = -(x_n - T(x_n))$ holds for all $n \in \mathbb{N}$. Hence, $\langle x_n - T(x_n), d_n \rangle = -\|x_n - T(x_n)\|^2 < 0$ ($n \in \mathbb{N}$). Lemma 2.1 thus guarantees that $\sum_{n=0}^{\infty} \|x_n - T(x_n)\|^2 < \infty$, which implies $\lim_{n \to \infty} \|x_n - T(x_n)\| = 0$.

   (ii) The triangle inequality and the nonexpansivity of T ensure that, for all $n \in \mathbb{N}$ and for all $x \in \text{Fix}(T)$, $\|x_{n+1} - x\| = \|x_n + \alpha_n(T(x_n) - x_n) - x\| \leq (1 - \alpha_n)\|x_n - x\| + \alpha_n\|T(x_n) - T(x)\| \leq \|x_n - x\|$.

   (iii) Lemma 2.2(ii) means that $\lim_{n \to \infty} \|x_n - x\|$ exists for all $x \in \text{Fix}(T)$. Accordingly, $(x_n)_{n \in \mathbb{N}}$ is bounded. Hence, there is a subsequence $(x_{n_k})_{k \in \mathbb{N}}$ of $(x_n)_{n \in \mathbb{N}}$ such that $(x_{n_k})_{k \in \mathbb{N}}$ weakly converges to a point $x^* \in H$. Here, let us assume that $x^* \notin \text{Fix}(T)$. Then Opial's condition [34], Lemma 1, Lemma 2.2(i), and the nonexpansivity of T guarantee that

$$\liminf_{k \to \infty} \|x_{n_k} - x^*\| < \liminf_{k \to \infty} \|x_{n_k} - T(x^*)\|$$

$$= \liminf_{k \to \infty} \|x_{n_k} - T(x_{n_k}) + T(x_{n_k}) - T(x^*)\|$$

$$= \liminf_{k \to \infty} \|T(x_{n_k}) - T(x^*)\|$$

$$\leq \liminf_{k \to \infty} \|x_{n_k} - x^*\|,$$

which is a contradiction. Hence, $x^* \in \text{Fix}(T)$. Let us take another subsequence $(x_{n_i})_{i \in \mathbb{N}}$ ($\subset (x_n)_{n \in \mathbb{N}}$) which weakly converges to $x_* \in H$. A similar discussion to the one for obtaining $x^* \in \text{Fix}(T)$ ensures that $x_* \in \text{Fix}(T)$. Assume that $x^* \neq x_*$. The existence of

$\lim_{n\to\infty} \|x_n - x\|$ ($x \in \text{Fix}(T)$) and Opial's condition [34], Lemma 1, imply that

$$\lim_{n\to\infty} \|x_n - x^*\| = \lim_{k\to\infty} \|x_{n_k} - x^*\| < \lim_{k\to\infty} \|x_{n_k} - x_*\|$$

$$= \lim_{n\to\infty} \|x_n - x_*\| = \lim_{i\to\infty} \|x_{n_i} - x_*\|$$

$$< \lim_{i\to\infty} \|x_{n_i} - x^*\| = \lim_{n\to\infty} \|x_n - x^*\|,$$

which is a contradiction. Therefore, $x^* = x_*$. Since any subsequence of $(x_n)_{n\in\mathbb{N}}$ weakly converges to the same fixed point of T, it is guaranteed that the whole $(x_n)_{n\in\mathbb{N}}$ weakly converges to a fixed point of T. This completes the proof.   □

2.2 Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{DY}}$

The following is a convergence analysis of Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{DY}}$.

Theorem 2.2 *Suppose that $(x_n)_{n\in\mathbb{N}}$ is the sequence generated by Algorithm* 2.1 *with $\beta_n = \beta_n^{\mathrm{DY}}$ ($n \in \mathbb{N}$). Then $(x_n)_{n\in\mathbb{N}}$ either terminates at a fixed point of T or*

$$\lim_{n\to\infty} \|x_n - T(x_n)\| = 0.$$

2.2.1 Proof of Theorem 2.2

Since the existence of $m \in \mathbb{N}$ such that $\|x_m - T(x_m)\| = 0$ implies that Theorem 2.2 holds, it can be assumed that, for all $n \in \mathbb{N}$, $\|x_n - T(x_n)\| \neq 0$ holds. Theorem 2.2 can be proven by using the ideas presented in the proof of [28], Theorem 3.3. The proof of Theorem 2.2 is divided into three steps.

Lemma 2.3 *Suppose that the assumptions in Theorem* 2.2 *are satisfied. Then*:
   (i) $\langle x_n - T(x_n), d_n \rangle < 0$ ($n \in \mathbb{N}$).
   (ii) $\liminf_{n\to\infty} \|x_n - T(x_n)\| = 0$.
   (iii) $\lim_{n\to\infty} \|x_n - T(x_n)\| = 0$.

Proof (i) From $d_0 := -(x_0 - T(x_0))$, $\langle x_0 - T(x_0), d_0 \rangle = -\|x_0 - T(x_0)\|^2 < 0$. Suppose that $\langle x_n - T(x_n), d_n \rangle < 0$ holds for some $n \in \mathbb{N}$. Accordingly, the definition of $y_n := (x_{n+1} - T(x_{n+1})) - (x_n - T(x_n))$ and (2.2) ensure that

$$\langle d_n, y_n \rangle = \langle d_n, x_{n+1} - T(x_{n+1}) \rangle - \langle d_n, x_n - T(x_n) \rangle$$

$$\geq (\sigma - 1)\langle d_n, x_n - T(x_n) \rangle > 0,$$

which implies that

$$\beta_n^{\mathrm{DY}} := \frac{\|x_{n+1} - T(x_{n+1})\|^2}{\langle d_n, y_n \rangle} > 0.$$

From the definition of $d_{n+1} := -(x_{n+1} - T(x_{n+1})) + \beta_n^{\mathrm{DY}} d_n$, we have

$$\langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle = -\|x_{n+1} - T(x_{n+1})\|^2 + \beta_n^{\mathrm{DY}}\langle d_n, x_{n+1} - T(x_{n+1}) \rangle$$

$$= \|x_{n+1} - T(x_{n+1})\|^2 \left\{ -1 + \frac{\langle d_n, x_{n+1} - T(x_{n+1}) \rangle}{\langle d_n, y_n \rangle} \right\}$$

$$= \|x_{n+1} - T(x_{n+1})\|^2 \frac{\langle d_n, (x_{n+1} - T(x_{n+1})) - y_n \rangle}{\langle d_n, y_n \rangle},$$

which, together with the definitions of y_n and β_n^{DY} (> 0), implies that

$$
\begin{aligned}
\langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle &= \| x_{n+1} - T(x_{n+1}) \|^2 \frac{\langle d_n, x_n - T(x_n) \rangle}{\langle d_n, y_n \rangle} \\
&= \beta_n^{\mathrm{DY}} \langle d_n, x_n - T(x_n) \rangle < 0.
\end{aligned}
\tag{2.8}
$$

Induction shows that $\langle x_n - T(x_n), d_n \rangle < 0$ for all $n \in \mathbb{N}$. This implies $\beta_n^{\mathrm{DY}} > 0$ ($n \in \mathbb{N}$); *i.e.*, Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{DY}}$ is well defined.

(ii) Assume that $\liminf_{n \to \infty} \| x_n - T(x_n) \| > 0$. Then there exist $n_0 \in \mathbb{N}$ and $\varepsilon > 0$ such that $\| x_n - T(x_n) \| \geq \varepsilon$ for all $n \geq n_0$. Since we have assumed that $\| x_n - T(x_n) \| \neq 0$ ($n \in \mathbb{N}$), we may further assume that $\| x_n - T(x_n) \| \geq \varepsilon$ for all $n \in \mathbb{N}$. From the definition of $d_{n+1} := -(x_{n+1} - T(x_{n+1})) + \beta_n^{\mathrm{DY}} d_n$ ($n \in \mathbb{N}$), we have, for all $n \in \mathbb{N}$,

$$
\begin{aligned}
\beta_n^{\mathrm{DY}^2} \| d_n \|^2 &= \big\| d_{n+1} + \big(x_{n+1} - T(x_{n+1}) \big) \big\|^2 \\
&= \| d_{n+1} \|^2 + 2 \langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle + \| x_{n+1} - T(x_{n+1}) \|^2.
\end{aligned}
$$

Lemma 2.3(i) and (2.8) mean that, for all $n \in \mathbb{N}$,

$$
\beta_n^{\mathrm{DY}} = \frac{\langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle}{\langle d_n, x_n - T(x_n) \rangle}.
$$

Hence, for all $n \in \mathbb{N}$,

$$
\begin{aligned}
\frac{\| d_{n+1} \|^2}{\langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle^2}
&= -\frac{\| x_{n+1} - T(x_{n+1}) \|^2}{\langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle^2} - \frac{2}{\langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle} + \frac{\| d_n \|^2}{\langle d_n, x_n - T(x_n) \rangle^2} \\
&= \frac{\| d_n \|^2}{\langle d_n, x_n - T(x_n) \rangle^2} + \frac{1}{\| x_{n+1} - T(x_{n+1}) \|^2} \\
&\quad - \left\{ \frac{1}{\| x_{n+1} - T(x_{n+1}) \|} + \frac{\| x_{n+1} - T(x_{n+1}) \|}{\langle d_{n+1}, x_{n+1} - T(x_{n+1}) \rangle} \right\}^2 \\
&\leq \frac{\| d_n \|^2}{\langle d_n, x_n - T(x_n) \rangle^2} + \frac{1}{\| x_{n+1} - T(x_{n+1}) \|^2}.
\end{aligned}
$$

Summing up this inequality from $n = 0$ to $n = N \in \mathbb{N}$ yields, for all $N \in \mathbb{N}$,

$$
\frac{\| d_{N+1} \|^2}{\langle d_{N+1}, x_{N+1} - T(x_{N+1}) \rangle^2} \leq \frac{\| d_0 \|^2}{\langle d_0, x_0 - T(x_0) \rangle^2} + \sum_{k=1}^{N+1} \frac{1}{\| x_k - T(x_k) \|^2},
$$

which, which together with $\| x_n - T(x_n) \| \geq \varepsilon$ ($n \in \mathbb{N}$) and $d_0 := -(x_0 - T(x_0))$, implies that, for all $N \in \mathbb{N}$,

$$
\frac{\| d_{N+1} \|^2}{\langle d_{N+1}, x_{N+1} - T(x_{N+1}) \rangle^2} \leq \sum_{k=0}^{N+1} \frac{1}{\| x_k - T(x_k) \|^2} \leq \frac{N+2}{\varepsilon^2}.
$$

Since Lemma 2.3(i) implies $\|d_n\| \neq 0$ ($n \in \mathbb{N}$), we have, for all $N \in \mathbb{N}$,

$$\frac{\langle d_{N+1}, x_{N+1} - T(x_{N+1})\rangle^2}{\|d_{N+1}\|^2} \geq \frac{\varepsilon^2}{N+2}.$$

Therefore, Lemma 2.1 guarantees that

$$\infty > \sum_{k=1}^{\infty}\left(\frac{\langle d_k, x_k - T(x_k)\rangle}{\|d_k\|}\right)^2 \geq \sum_{k=1}^{\infty}\frac{\varepsilon^2}{k+1} = \infty.$$

This is a contradiction. Hence, $\liminf_{n\to\infty}\|x_n - T(x_n)\| = 0$.

(iii) Condition (2.1) and Lemma 2.3(i) lead to that, for all $n \in \mathbb{N}$,

$$\left\|x_{n+1} - T(x_{n+1})\right\|^2 - \left\|x_n - T(x_n)\right\|^2 \leq \delta\alpha_n\langle x_n - T(x_n), d_n\rangle < 0.$$

Accordingly, $(\|x_n - T(x_n)\|)_{n\in\mathbb{N}}$ is monotone decreasing; *i.e.*, there exists $\lim_{n\to\infty}\|x_n - T(x_n)\|$. Lemma 2.3(ii) thus ensures that $\lim_{n\to\infty}\|x_n - T(x_n)\| = 0$. This completes the proof. $\square$

2.3 Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{FR}}$

To establish the convergence of Algorithm 2.1 when $\beta_n = \beta_n^{\mathrm{FR}}$, we assume that the step sizes α_n satisfy the *strong Wolfe-type conditions*, which are (2.1) and the following strengthened version of (2.2): for $\sigma \leq 1/2$,

$$\left|\langle x_n(\alpha_n) - T\big(x_n(\alpha_n)\big), d_n\rangle\right| \leq -\sigma\langle x_n - T(x_n), d_n\rangle. \tag{2.9}$$

See [30] on the global convergence of the FR method for unconstrained optimization under the strong Wolfe conditions.

The following is a convergence analysis of Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{FR}}$.

Theorem 2.3 *Suppose that* $(x_n)_{n\in\mathbb{N}}$ *is the sequence generated by Algorithm* 2.1 *with* $\beta_n = \beta_n^{\mathrm{FR}}$ ($n \in \mathbb{N}$), *where* $(\alpha_n)_{n\in\mathbb{N}}$ *satisfies* (2.1) *and* (2.9). *Then* $(x_n)_{n\in\mathbb{N}}$ *either terminates at a fixed point of* T *or*

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

2.3.1 Proof of Theorem 2.3

It can be assumed that, for all $n \in \mathbb{N}$, $\|x_n - T(x_n)\| \neq 0$ holds. Theorem 2.3 can be proven by using the ideas in the proof of [30], Theorem 2.

Lemma 2.4 *Suppose that the assumptions in Theorem* 2.3 *are satisfied. Then*:
  (i) $\langle x_n - T(x_n), d_n\rangle < 0$ ($n \in \mathbb{N}$).
  (ii) $\liminf_{n\to\infty}\|x_n - T(x_n)\| = 0$.
  (iii) $\lim_{n\to\infty}\|x_n - T(x_n)\| = 0$.

Proof (i) Let us show that, for all $n \in \mathbb{N}$,

$$-\sum_{j=0}^{n}\sigma^j \leq \frac{\langle x_n - T(x_n), d_n\rangle}{\|x_n - T(x_n)\|^2} \leq -2 + \sum_{j=0}^{n}\sigma^j. \tag{2.10}$$

From $d_0 := -(x_0 - T(x_0))$, (2.10) holds for $n := 0$ and $\langle x_0 - T(x_0), d_0 \rangle < 0$. Suppose that (2.10) holds for some $n \in \mathbb{N}$. Accordingly, from $\sum_{j=0}^{n} \sigma^j < \sum_{j=0}^{\infty} \sigma^j = 1/(1-\sigma)$ and $\sigma \in (0, 1/2]$, we have

$$\frac{\langle x_n - T(x_n), d_n \rangle}{\|x_n - T(x_n)\|^2} < -2 + \sum_{j=0}^{\infty} \sigma^j = \frac{-(1-2\sigma)}{1-\sigma} \le 0,$$

which implies that $\langle x_n - T(x_n), d_n \rangle < 0$. The definitions of d_{n+1} and β_n^{FR} enable us to deduce that

$$\begin{aligned}
\frac{\langle x_{n+1} - T(x_{n+1}), d_{n+1} \rangle}{\|x_{n+1} - T(x_{n+1})\|^2} &= \frac{\langle x_{n+1} - T(x_{n+1}), -(x_{n+1} - T(x_{n+1})) + \beta_n^{\mathrm{FR}} d_n \rangle}{\|x_{n+1} - T(x_{n+1})\|^2} \\
&= -1 + \frac{\|x_{n+1} - T(x_{n+1})\|^2}{\|x_n - T(x_n)\|^2} \frac{\langle x_{n+1} - T(x_{n+1}), d_n \rangle}{\|x_{n+1} - T(x_{n+1})\|^2} \\
&= -1 + \frac{\langle x_{n+1} - T(x_{n+1}), d_n \rangle}{\|x_n - T(x_n)\|^2}.
\end{aligned}$$

Since (2.9) satisfies $\sigma \langle x_n - T(x_n), d_n \rangle \le \langle x_{n+1} - T(x_{n+1}), d_n \rangle \le -\sigma \langle x_n - T(x_n), d_n \rangle$ and (2.10) holds for some n, it is found that

$$\begin{aligned}
-1 + \frac{\langle x_{n+1} - T(x_{n+1}), d_n \rangle}{\|x_n - T(x_n)\|^2} &\ge -1 + \sigma \frac{\langle x_n - T(x_n), d_n \rangle}{\|x_n - T(x_n)\|^2} \\
&\ge -1 - \sigma \sum_{j=0}^{n} \sigma^j = -\sum_{j=0}^{n+1} \sigma^j
\end{aligned}$$

and

$$\begin{aligned}
-1 + \frac{\langle x_{n+1} - T(x_{n+1}), d_n \rangle}{\|x_n - T(x_n)\|^2} &\le -1 - \sigma \frac{\langle x_n - T(x_n), d_n \rangle}{\|x_n - T(x_n)\|^2} \\
&\le -1 + \sigma \sum_{j=0}^{n} \sigma^j = -2 + \sum_{j=0}^{n+1} \sigma^j.
\end{aligned}$$

Hence,

$$-\sum_{j=0}^{n+1} \sigma^j \le \frac{\langle x_{n+1} - T(x_{n+1}), d_{n+1} \rangle}{\|x_{n+1} - T(x_{n+1})\|^2} \le -2 + \sum_{j=0}^{n+1} \sigma^j.$$

A discussion similar to the one for obtaining $\langle x_n - T(x_n), d_n \rangle < 0$ guarantees that $\langle x_{n+1} - T(x_{n+1}), d_{n+1} \rangle < 0$ holds. Induction thus shows that (2.10) and $\langle x_n - T(x_n), d_n \rangle < 0$ hold for all $n \in \mathbb{N}$.

(ii) Assume that $\liminf_{n \to \infty} \|x_n - T(x_n)\| > 0$. A discussion similar to the one in the proof of Lemma 2.3(ii) ensures the existence of $\varepsilon > 0$ such that $\|x_n - T(x_n)\| \ge \varepsilon$ for all $n \in \mathbb{N}$. From (2.9) and (2.10), we have, for all $n \in \mathbb{N}$,

$$\left| \langle x_{n+1} - T(x_{n+1}), d_n \rangle \right| < -\sigma \langle x_n - T(x_n), d_n \rangle \le \sum_{j=1}^{n+1} \sigma^j \|x_n - T(x_n)\|^2,$$

which, together with $\sum_{j=1}^{n+1} \sigma^j < \sum_{j=1}^{\infty} \sigma^j = \sigma/(1-\sigma)$ and $\beta_n^{\text{FR}} := \|x_{n+1} - T(x_{n+1})\|^2 / \|x_n - T(x_n)\|^2$ $(n \in \mathbb{N})$, implies that, for all $n \in \mathbb{N}$,

$$\beta_n^{\text{FR}} |\langle x_{n+1} - T(x_{n+1}), d_n \rangle| < \frac{\sigma}{1-\sigma} \|x_{n+1} - T(x_{n+1})\|^2.$$

Accordingly, from the definition of $d_{n+1} := -(x_{n+1} - T(x_{n+1})) + \beta_n^{\text{FR}} d_n$, we find that, for all $n \in \mathbb{N}$,

$$
\begin{aligned}
\|d_{n+1}\|^2 &= \left\| \beta_n^{\text{FR}} d_n - \left(x_{n+1} - T(x_{n+1}) \right) \right\|^2 \\
&= \beta_n^{\text{FR}^2} \|d_n\|^2 - 2\beta_n^{\text{FR}} \langle d_n, x_{n+1} - T(x_{n+1}) \rangle + \left\| x_{n+1} - T(x_{n+1}) \right\|^2 \\
&\leq \frac{\|x_{n+1} - T(x_{n+1})\|^4}{\|x_n - T(x_n)\|^4} \|d_n\|^2 + \left(\frac{2\sigma}{1-\sigma} + 1 \right) \left\| x_{n+1} - T(x_{n+1}) \right\|^2,
\end{aligned}
$$

which means that, for all $n \in \mathbb{N}$,

$$\frac{\|d_{n+1}\|^2}{\|x_{n+1} - T(x_{n+1})\|^4} \leq \frac{\|d_n\|^2}{\|x_n - T(x_n)\|^4} + \frac{1+\sigma}{1-\sigma} \frac{1}{\|x_{n+1} - T(x_{n+1})\|^2}.$$

The sum of this inequality from $n = 0$ to $n = N \in \mathbb{N}$ and $d_0 := -(x_0 - T(x_0))$ ensure that, for all $N \in \mathbb{N}$,

$$\frac{\|d_{N+1}\|^2}{\|x_{N+1} - T(x_{N+1})\|^4} \leq \frac{1}{\|x_0 - T(x_0)\|^2} + \frac{1+\sigma}{1-\sigma} \sum_{k=1}^{N+1} \frac{1}{\|x_k - T(x_k)\|^2}.$$

From $\|x_n - T(x_n)\| \geq \varepsilon$ $(n \in \mathbb{N})$, for all $N \in \mathbb{N}$,

$$\frac{\|d_{N+1}\|^2}{\|x_{N+1} - T(x_{N+1})\|^4} \leq \frac{1}{\varepsilon^2} + \frac{1+\sigma}{1-\sigma} \frac{N+1}{\varepsilon^2} = \frac{(1+\sigma)N + 2}{\varepsilon^2(1-\sigma)}.$$

Therefore, from Lemma 2.4(i) guaranteeing that $\|d_n\| \neq 0$ $(n \in \mathbb{N})$ and $\sum_{k=1}^{\infty} \varepsilon^2(1-\sigma)/((1+\sigma)(k-1) + 2) = \infty$, it is found that

$$\sum_{k=1}^{\infty} \frac{\|x_k - T(x_k)\|^4}{\|d_k\|^2} = \infty.$$

Meanwhile, since (2.10) guarantees that $\langle x_n - T(x_n), d_n \rangle \leq (-2 + \sum_{j=0}^{n} \sigma^j)\|x_n - T(x_n)\|^2 < (-(1-2\sigma)/(1-\sigma))\|x_n - T(x_n)\|^2$ $(n \in \mathbb{N})$, Lemma 2.1 and Lemma 2.4(i) lead to the deduction that

$$\infty > \sum_{k=0}^{\infty} \left(\frac{\langle x_k - T(x_k), d_k \rangle}{\|d_k\|} \right)^2 \geq \left(\frac{1-2\sigma}{1-\sigma} \right)^2 \sum_{k=0}^{\infty} \frac{\|x_k - T(x_k)\|^4}{\|d_k\|^2} = \infty,$$

which is a contradiction. Therefore, $\liminf_{n\to\infty} \|x_n - T(x_n)\| = 0$.

(iii) A discussion similar to the one in the proof of Lemma 2.3(iii) leads to Lemma 2.4(iii). This completes the proof. $\qquad \square$

2.4 Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{PRP+}}$

It is well known that the convergence of the nonlinear conjugate gradient method with β_n^{PRP} defined as in (1.19) for a general nonlinear function is uncertain [23], Section 5. To guarantee the convergence of the PRP method for unconstrained optimization, the following modification of β_n^{PRP} was presented in [35]: for β_n^{PRP} defined as in (1.19), $\beta_n^{\mathrm{PRP+}} :=$ $\max\{\beta_n^{\mathrm{PRP}}, 0\}$. On the basis of the idea behind this modification, this subsection considers Algorithm 2.1 with $\beta_n^{\mathrm{PRP+}}$ defined as in (2.4).

Theorem 2.4 *Suppose that $(x_n)_{n\in\mathbb{N}}$ and $(d_n)_{n\in\mathbb{N}}$ are the sequences generated by Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{PRP+}}$ ($n \in \mathbb{N}$) and there exists $c > 0$ such that $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ for all $n \in \mathbb{N}$. If $(x_n)_{n\in\mathbb{N}}$ is bounded, then $(x_n)_{n\in\mathbb{N}}$ either terminates at a fixed point of T or*

$$\lim_{n\to\infty} \|x_n - T(x_n)\| = 0.$$

2.4.1 Proof of Theorem 2.4

It can be assumed that $\|x_n - T(x_n)\| \neq 0$ holds for all $n \in \mathbb{N}$. Let us first show the following lemma by referring to the proof of [31], Lemma 4.1.

Lemma 2.5 *Let $(x_n)_{n\in\mathbb{N}}$ and $(d_n)_{n\in\mathbb{N}}$ be the sequences generated by Algorithm 2.1 with $\beta_n \geq 0$ ($n \in \mathbb{N}$) and assume that there exists $c > 0$ such that $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ for all $n \in \mathbb{N}$. If there exists $\varepsilon > 0$ such that $\|x_n - T(x_n)\| \geq \varepsilon$ for all $n \in \mathbb{N}$, then $\sum_{n=0}^{\infty} \|u_{n+1} - u_n\|^2 < \infty$, where $u_n := d_n/\|d_n\|$ ($n \in \mathbb{N}$).*

Proof Assuming $\|x_n - T(x_n)\| \geq \varepsilon$ and $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ ($n \in \mathbb{N}$), $\|d_n\| \neq 0$ holds for all $n \in \mathbb{N}$. Define $r_n := -(x_n - T(x_n))/\|d_n\|$ and $\delta_n := \beta_n\|d_n\|/\|d_{n+1}\|$ ($n \in \mathbb{N}$). From $\delta_n u_n = \beta_n d_n/\|d_{n+1}\|$ and $d_{n+1} = -(x_{n+1} - T(x_{n+1})) + \beta_n d_n$ ($n \in \mathbb{N}$), we have, for all $n \in \mathbb{N}$,

$$u_{n+1} = -r_{n+1} + \delta_n u_n,$$

which, together with $\|u_{n+1} - \delta_n u_n\|^2 = \|u_{n+1}\|^2 - 2\delta_n\langle u_{n+1}, u_n \rangle + \delta_n^2\|u_n\|^2 = \|u_n\|^2 - 2\delta_n\langle u_n, u_{n+1}\rangle + \delta_n^2\|u_{n+1}\|^2 = \|u_n - \delta_n u_{n+1}\|^2$ ($n \in \mathbb{N}$), implies that, for all $n \in \mathbb{N}$,

$$\|r_{n+1}\| = \|u_{n+1} - \delta_n u_n\| = \|u_n - \delta_n u_{n+1}\|.$$

Accordingly, the condition $\beta_n \geq 0$ ($n \in \mathbb{N}$) and the triangle inequality mean that, for all $n \in \mathbb{N}$,

$$\|u_{n+1} - u_n\| \leq (1 + \delta_n)\|u_{n+1} - u_n\|$$

$$\leq \|u_{n+1} - \delta_n u_n\| + \|u_n - \delta_n u_{n+1}\|$$

$$= 2\|r_{n+1}\|. \tag{2.11}$$

From Lemma 2.1, $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ ($n \in \mathbb{N}$), the definition of r_n, and $\|x_n - T(x_n)\| \geq \varepsilon$ ($n \in \mathbb{N}$), we have

$$\infty > \sum_{n=0}^{\infty} \left(\frac{\langle x_n - T(x_n), d_n \rangle}{\|d_n\|} \right)^2 \geq c^2 \sum_{n=0}^{\infty} \frac{\|x_n - T(x_n)\|^4}{\|d_n\|^2} \geq c^2\varepsilon^2 \sum_{n=0}^{\infty} \|r_n\|^2,$$

which, together with (2.11), completes the proof. □

The following property, referred to as Property ($\star$), is a result of modifying [31], Property ($*$), to conform to Problem (1.1).

Property ($\star$). Suppose that there exist positive constants γ and $\bar{\gamma}$ such that $\gamma \leq \|x_n - T(x_n)\| \leq \bar{\gamma}$ for all $n \in \mathbb{N}$. Then Property ($\star$) holds if $b > 1$ and $\lambda > 0$ exist such that, for all $n \in \mathbb{N}$,

$$|\beta_n| \leq b \quad \text{and} \quad \|x_{n+1} - x_n\| \leq \lambda \quad \text{implies} \quad |\beta_n| \leq \frac{1}{2b}.$$

The proof of the following lemma can be omitted since it is similar to the proof of [31], Lemma 4.2.

Lemma 2.6 *Let $(x_n)_{n \in \mathbb{N}}$ and $(d_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm* 2.1 *and assume that there exist $c > 0$ and $\gamma > 0$ such that $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ and $\|x_n - T(x_n)\| \geq \gamma$ for all $n \in \mathbb{N}$. Suppose also that Property ($\star$) holds. Then there exists $\lambda > 0$ such that, for all $\Delta \in \mathbb{N} \setminus \{0\}$ and for any index k_0, there is $k \geq k_0$ such that $|\mathcal{K}_{k,\Delta}^{\lambda}| > \Delta/2$, where $\mathcal{K}_{k,\Delta}^{\lambda} := \{i \in \mathbb{N} \setminus \{0\} : k \leq i \leq k + \Delta - 1, \|x_i - x_{i-1}\| > \lambda\}$ $(k \in \mathbb{N}, \Delta \in \mathbb{N} \setminus \{0\}, \lambda > 0)$ and $|\mathcal{K}_{k,\Delta}^{\lambda}|$ stands for the number of elements of $\mathcal{K}_{k,\Delta}^{\lambda}$.*

The following can be proven by referring to the proof of [31], Theorem 4.3.

Lemma 2.7 *Let $(x_n)_{n \in \mathbb{N}}$ be the sequence generated by Algorithm* 2.1 *with $\beta_n \geq 0$ $(n \in \mathbb{N})$ and assume that there exists $c > 0$ such that $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ for all $n \in \mathbb{N}$ and Property ($\star$) holds. If $(x_n)_{n \in \mathbb{N}}$ is bounded, $\liminf_{n \to \infty} \|x_n - T(x_n)\| = 0$.*

Proof Assuming that $\liminf_{n \to \infty} \|x_n - T(x_n)\| > 0$, there exists $\gamma > 0$ such that $\|x_n - T(x_n)\| \geq \gamma$ for all $n \in \mathbb{N}$. Since $c > 0$ exists such that $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ $(n \in \mathbb{N})$, $\|d_n\| \neq 0$ $(n \in \mathbb{N})$ holds. Moreover, the nonexpansivity of T ensures that, for all $x \in \text{Fix}(T)$, $\|T(x_n) - x\| \leq \|x_n - x\|$, and this, together with the boundedness of $(x_n)_{n \in \mathbb{N}}$, implies the boundedness of $(T(x_n))_{n \in \mathbb{N}}$. Accordingly, $\bar{\gamma} > 0$ exists such that $\|x_n - T(x_n)\| \leq \bar{\gamma}$ $(n \in \mathbb{N})$. The definition of x_n implies that, for all $n \geq 1$,

$$x_n - x_{n-1} = \alpha_{n-1} d_{n-1} = \alpha_{n-1} \|d_{n-1}\| u_{n-1} = \|x_n - x_{n-1}\| u_{n-1},$$

where $u_n := d_n / \|d_n\|$ $(n \in \mathbb{N})$. Hence, for all $l, k \in \mathbb{N}$ with $l \geq k > 0$,

$$x_l - x_{k-1} = \sum_{i=k}^{l} (x_i - x_{i-1}) = \sum_{i=k}^{l} \|x_i - x_{i-1}\| u_{i-1},$$

which implies that

$$\sum_{i=k}^{l} \|x_i - x_{i-1}\| u_{k-1} = x_l - x_{k-1} - \sum_{i=k}^{l} \|x_i - x_{i-1}\| (u_{i-1} - u_{k-1}).$$

From $\|u_n\| = 1$ $(n \in \mathbb{N})$ and the triangle inequality, for all $l, k \in \mathbb{N}$ with $l \geq k > 0$, $\sum_{i=k}^{l} \|x_i - x_{i-1}\| \leq \|x_l - x_{k-1}\| + \sum_{i=k}^{l} \|x_i - x_{i-1}\| \|u_{i-1} - u_{k-1}\|$. Since the boundedness of $(x_n)_{n \in \mathbb{N}}$ means

there is $M > 0$ satisfying $\|x_{n+1} - x_n\| \le M$ ($n \in \mathbb{N}$), we find that, for all $l, k \in \mathbb{N}$ with $l \ge k > 0$,

$$\sum_{i=k}^{l} \|x_i - x_{i-1}\| \le M + \sum_{i=k}^{l} \|x_i - x_{i-1}\| \|u_{i-1} - u_{k-1}\|. \tag{2.12}$$

Let $\lambda > 0$ be as given by Lemma 2.6 and define $\Delta := \lceil 4M/\lambda \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling operator. From Lemma 2.5, an index k_0 can be chosen such that $\sum_{i=k_0}^{\infty} \|u_i - u_{i-1}\|^2 \le 1/(4\Delta)$. Accordingly, Lemma 2.6 guarantees the existence of $k \ge k_0$ such that $|\mathcal{K}_{k,\Delta}^{\lambda}| > \Delta/2$. Since the Cauchy-Schwarz inequality implies that $(\sum_{i=1}^{m} a_i)^2 \le m \sum_{i=1}^{m} a_i^2$ ($m \ge 1$, $a_i \in \mathbb{R}$, $i = 1, 2, \ldots, m$), we have, for all $i \in [k, k + \Delta - 1]$,

$$\|u_{i-1} - u_{k-1}\|^2 \le \left(\sum_{j=k}^{i-1} \|u_j - u_{j-1}\| \right)^2 \le (i - k) \sum_{j=k}^{i-1} \|u_j - u_{j-1}\|^2 \le \frac{1}{4}.$$

Putting $l := k + \Delta - 1$, (2.12) ensures that

$$M \ge \frac{1}{2} \sum_{i=k}^{k+\Delta-1} \|x_i - x_{i-1}\| > \frac{\lambda}{2} |\mathcal{K}_{k,\Delta}^{\lambda}| > \frac{\lambda \Delta}{4},$$

which implies that $\Delta < 4M/\lambda$. This contradicts $\Delta := \lceil 4M/\lambda \rceil$. Therefore, $\liminf_{n\to\infty} \|x_n - T(x_n)\| = 0$. $\qquad\square$

Now we are in the position to prove Theorem 2.4.

Proof The condition $\beta_n^{\mathrm{PRP+}} \ge 0$ holds for all $n \in \mathbb{N}$. Suppose that positive constants γ and $\bar{\gamma}$ exist such that $\gamma \le \|x_n - T(x_n)\| \le \bar{\gamma}$ ($n \in \mathbb{N}$) and define $b := 2\bar{\gamma}^2/\gamma^2$ and $\lambda := \gamma^2/(4\bar{\gamma}b)$. The definition of $\beta_n^{\mathrm{PRP+}}$ and the Cauchy-Schwarz inequality mean that, for all $n \in \mathbb{N}$,

$$|\beta_n^{\mathrm{PRP+}}| \le \frac{|\langle x_{n+1} - T(x_{n+1}), y_n \rangle|}{\|x_n - T(x_n)\|^2} \le \frac{\|x_{n+1} - T(x_{n+1})\| \|y_n\|}{\|x_n - T(x_n)\|^2} \le \frac{2\bar{\gamma}^2}{\gamma^2} = b,$$

where the third inequality comes from $\|y_n\| \le \|x_{n+1} - T(x_{n+1})\| + \|x_n - T(x_n)\| \le 2\bar{\gamma}$ and $\gamma \le \|x_n - T(x_n)\| \le \bar{\gamma}$ ($n \in \mathbb{N}$). When $\|x_{n+1} - x_n\| \le \lambda$ ($n \in \mathbb{N}$), the triangle inequality and the nonexpansivity of T imply that $\|y_n\| \le \|x_{n+1} - x_n\| + \|T(x_n) - T(x_{n+1})\| \le 2\|x_{n+1} - x_n\| \le 2\lambda$ ($n \in \mathbb{N}$). Therefore, for all $n \in \mathbb{N}$,

$$|\beta_n^{\mathrm{PRP+}}| \le \frac{\bar{\gamma} \|y_n\|}{\|x_n - T(x_n)\|^2} \le \frac{2\lambda\bar{\gamma}}{\gamma^2} = \frac{1}{2b},$$

which implies that Property $(\star)$ holds. Lemma 2.7 thus guarantees that $\liminf_{n\to\infty} \|x_n - T(x_n)\| = 0$ holds. A discussion in the same manner as in the proof of Lemma 2.3(iii) leads to $\lim_{n\to\infty} \|x_n - T(x_n)\| = 0$. This completes the proof. $\qquad\square$

2.5 Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{HS+}}$

The convergence properties of the nonlinear conjugate gradient method with β_n^{HS} defined as in (1.19) are similar to those with β_n^{PRP} defined as in (1.19) [23], Section 5. On the basis of this fact and the modification of β_n^{PRP} in Section 2.4, this subsection considers Algorithm 2.1 with $\beta_n^{\mathrm{HS+}}$ defined by (2.4).

Lemma 2.7 leads to the following.

Theorem 2.5 *Suppose that $(x_n)_{n\in\mathbb{N}}$ and $(d_n)_{n\in\mathbb{N}}$ are the sequences generated by Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{HS+}}$ ($n \in \mathbb{N}$) and there exists $c > 0$ such that $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$ for all $n \in \mathbb{N}$. If $(x_n)_{n\in\mathbb{N}}$ is bounded, then $(x_n)_{n\in\mathbb{N}}$ either terminates at a fixed point of T or*

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

Proof When $m \in \mathbb{N}$ exists such that $\|x_m - T(x_m)\| = 0$, Theorem 2.5 holds. Let us consider the case where $\|x_n - T(x_n)\| \neq 0$ for all $n \in \mathbb{N}$. Suppose that $\gamma, \bar{\gamma} > 0$ exist such that $\gamma \leq \|x_n - T(x_n)\| \leq \bar{\gamma}$ ($n \in \mathbb{N}$) and define $b := 2\bar{\gamma}^2/((1-\sigma)c\gamma^2)$ and $\lambda := (1-\sigma)c\gamma^2/(4\bar{\gamma}b)$. Then (2.2) implies that, for all $n \in \mathbb{N}$,

$$\langle d_n, y_n \rangle = \langle d_n, x_{n+1} - T(x_{n+1}) \rangle - \langle d_n, x_n - T(x_n) \rangle$$

$$\geq -(1-\sigma)\langle d_n, x_n - T(x_n) \rangle,$$

which, together with the existence of $c, \gamma > 0$ such that $\langle x_n - T(x_n), d_n \rangle \leq -c\|x_n - T(x_n)\|^2$, and $\gamma \leq \|x_n - T(x_n)\|$ ($n \in \mathbb{N}$), implies that, for all $n \in \mathbb{N}$,

$$\langle d_n, y_n \rangle \geq (1-\sigma)c\left\| x_n - T(x_n) \right\|^2 \geq (1-\sigma)c\gamma^2 > 0.$$

This means Algorithm 2.1 with $\beta_n = \beta_n^{\mathrm{HS+}}$ is well defined. From $\|x_n - T(x_n)\| \leq \bar{\gamma}$ ($n \in \mathbb{N}$) and the definition of y_n, we have, for all $n \in \mathbb{N}$,

$$\left| \beta_n^{\mathrm{HS+}} \right| \leq \frac{|\langle x_{n+1} - T(x_{n+1}), y_n \rangle|}{|\langle d_n, y_n \rangle|} \leq \frac{2\bar{\gamma}^2}{(1-\sigma)c\gamma^2} = b.$$

When $\|x_{n+1} - x_n\| \leq \lambda$ ($n \in \mathbb{N}$), the triangle inequality and the nonexpansivity of T imply that $\|y_n\| \leq \|x_{n+1} - x_n\| + \|T(x_n) - T(x_{n+1})\| \leq 2\|x_{n+1} - x_n\| \leq 2\lambda$ ($n \in \mathbb{N}$). Therefore, from $\|x_n - T(x_n)\| \leq \bar{\gamma}$ ($n \in \mathbb{N}$), for all $n \in \mathbb{N}$,

$$\left| \beta_n^{\mathrm{HS+}} \right| \leq \frac{\bar{\gamma}\|y_n\|}{\langle d_n, y_n \rangle} \leq \frac{2\lambda\bar{\gamma}}{(1-\sigma)c\gamma^2} = \frac{1}{2b},$$

which in turn implies that Property ($\star$) holds. Lemma 2.7 thus ensures that $\liminf_{n\to\infty} \|x_n - T(x_n)\| = 0$ holds. A discussion similar to the one in the proof of Lemma 2.3(iii) leads to $\lim_{n\to\infty} \|x_n - T(x_n)\| = 0$. This completes the proof. □

2.6 Convergence rate analyses of Algorithm 2.1

Sections 2.1-2.5 show that Algorithm 2.1 with equations (2.4) satisfies $\lim_{n\to\infty} \|x_n - T(x_n)\| = 0$ under certain assumptions. The next theorem establishes rates of convergence for Algorithm 2.1 with equations (2.4).

Theorem 2.6

(i) *Under the Wolfe-type conditions* (2.1) *and* (2.2), *Algorithm* 2.1 *with* $\beta_n = \beta_n^{\mathrm{SD}}$ *satisfies, for all $n \in \mathbb{N}$,*

$$\left\| x_n - T(x_n) \right\| \leq \frac{\|x_0 - T(x_0)\|}{\sqrt{\delta \sum_{k=0}^{n} \alpha_k}}.$$

(ii) *Under the strong Wolfe-type conditions* (2.1) *and* (2.9), *Algorithm* 2.1 *with* $\beta_n = \beta_n^{\mathrm{DY}}$
satisfies, for all $n \in \mathbb{N}$,

$$\left\| x_n - T(x_n) \right\| \leq \frac{\|x_0 - T(x_0)\|}{\sqrt{\frac{1}{1+\sigma} \delta \sum_{k=0}^{n} \alpha_k}}.$$

(iii) *Under the strong Wolfe-type conditions* (2.1) *and* (2.9), *Algorithm* 2.1 *with* $\beta_n = \beta_n^{\mathrm{FR}}$
satisfies, for all $n \in \mathbb{N}$,

$$\left\| x_n - T(x_n) \right\| \leq \frac{\|x_0 - T(x_0)\|}{\sqrt{\frac{1}{1-\sigma} \delta \sum_{k=0}^{n} (1 - 2\sigma + \sigma^k) \alpha_k}}.$$

(iv) *Under the assumptions in Theorem* 2.4, *Algorithm* 2.1 *with* $\beta_n = \beta_n^{\mathrm{PRP+}}$ *satisfies, for
all* $n \in \mathbb{N}$,

$$\left\| x_n - T(x_n) \right\| \leq \frac{\|x_0 - T(x_0)\|}{\sqrt{c\delta \sum_{k=0}^{n} \alpha_k}}.$$

(v) *Under the assumptions in Theorem* 2.5, *Algorithm* 2.1 *with* $\beta_n = \beta_n^{\mathrm{HS+}}$ *satisfies, for
all* $n \in \mathbb{N}$,

$$\left\| x_n - T(x_n) \right\| \leq \frac{\|x_0 - T(x_0)\|}{\sqrt{c\delta \sum_{k=0}^{n} \alpha_k}}.$$

Proof (i) From $d_k = -(x_k - T(x_k))$ ($k \in \mathbb{N}$) and (2.1), we have $0 \leq \delta \alpha_k \|x_k - T(x_k)\|^2 \leq \|x_k - T(x_k)\|^2 - \|x_{k+1} - T(x_{k+1})\|^2$ ($k \in \mathbb{N}$). Summing up this inequality from $k = 0$ to $k = n$ guarantees that, for all $n \in \mathbb{N}$,

$$\delta \sum_{k=0}^{n} \alpha_k \left\| x_k - T(x_k) \right\|^2 \leq \left\| x_0 - T(x_0) \right\|^2 - \left\| x_{n+1} - T(x_{n+1}) \right\|^2 \leq \left\| x_0 - T(x_0) \right\|^2,$$

which, together with the monotone decreasing property of $(\|x_n - T(x_n)\|^2)_{n \in \mathbb{N}}$, implies that, for all $n \in \mathbb{N}$,

$$\delta \left\| x_n - T(x_n) \right\|^2 \sum_{k=0}^{n} \alpha_k \leq \left\| x_0 - T(x_0) \right\|^2.$$

This completes the proof.

(ii) Condition (2.9) and Lemma 2.3(i) ensure that $-\sigma \leq \langle x_{k+1} - T(x_{k+1}), d_k \rangle / \langle x_k - T(x_k), d_k \rangle \leq \sigma$ ($k \in \mathbb{N}$). Accordingly, (2.8) means that, for all $k \in \mathbb{N}$,

$$\begin{aligned}
\langle x_{k+1} - T(x_{k+1}), d_{k+1} \rangle &= \frac{\langle x_k - T(x_k), d_k \rangle}{\langle d_k, (x_{k+1} - T(x_{k+1})) - (x_k - T(x_k)) \rangle} \left\| x_{k+1} - T(x_{k+1}) \right\|^2 \\
&= \left(\frac{\langle x_{k+1} - T(x_{k+1}), d_k \rangle}{\langle x_k - T(x_k), d_k \rangle} - 1 \right)^{-1} \left\| x_{k+1} - T(x_{k+1}) \right\|^2 \\
&\leq -\frac{1}{1+\sigma} \left\| x_{k+1} - T(x_{k+1}) \right\|^2.
\end{aligned}$$

Hence, (2.1) implies that, for all $k \in \mathbb{N}$,

$$\left\| x_{k+1} - T(x_{k+1}) \right\|^2 - \left\| x_k - T(x_k) \right\|^2 \le -\frac{1}{1+\sigma} \delta \alpha_k \left\| x_k - T(x_k) \right\|^2.$$

Summing up this inequality from $k = 0$ to $k = n$ and the monotone decreasing property of $(\|x_n - T(x_n)\|^2)_{n \in \mathbb{N}}$ ensure that, for all $n \in \mathbb{N}$,

$$\frac{1}{1+\sigma} \delta \left\| x_n - T(x_n) \right\|^2 \sum_{k=0}^{n} \alpha_k \le \left\| x_0 - T(x_0) \right\|^2,$$

which completes the proof.

(iii) Inequality (2.10) guarantees that, for all $k \in \mathbb{N}$,

$$\left\langle x_k - T(x_k), d_k \right\rangle \le \left(-2 + \sum_{j=0}^{k} \sigma^j \right) \left\| x_k - T(x_k) \right\|^2$$

$$= -\frac{1 - 2\sigma + \sigma^k}{1 - \sigma} \left\| x_k - T(x_k) \right\|^2,$$

which, together with (2.1), implies that, for all $k \in \mathbb{N}$,

$$\left\| x_{k+1} - T(x_{k+1}) \right\|^2 - \left\| x_k - T(x_k) \right\|^2 \le -\frac{1 - 2\sigma + \sigma^k}{1 - \sigma} \delta \alpha_k \left\| x_k - T(x_k) \right\|^2.$$

Summing up this inequality from $k = 0$ to $k = n$ and the monotone decreasing property of $(\|x_n - T(x_n)\|^2)_{n \in \mathbb{N}}$ ensure that, for all $n \in \mathbb{N}$,

$$\frac{1}{1-\sigma} \delta \left\| x_n - T(x_n) \right\|^2 \sum_{k=0}^{n} (1 - 2\sigma + \sigma^k) \alpha_k \le \left\| x_0 - T(x_0) \right\|^2,$$

which completes the proof.

(iv), (v) Since there exists $c > 0$ such that $\langle x_k - T(x_k), d_k \rangle \le -c\|x_k - T(x_k)\|^2$ for all $k \in \mathbb{N}$, we have from (2.1) and the monotone decreasing property of $(\|x_n - T(x_n)\|^2)_{n \in \mathbb{N}}$, for all $n \in \mathbb{N}$,

$$c\delta \left\| x_n - T(x_n) \right\|^2 \sum_{k=0}^{n} \alpha_k \le c\delta \sum_{k=0}^{n} \alpha_k \left\| x_k - T(x_k) \right\|^2 \le \left\| x_0 - T(x_0) \right\|^2.$$

This concludes the proof. □

The conventional Krasnosel'skiĭ-Mann algorithm (1.2) with a step size sequence $(\alpha_n)_{n \in \mathbb{N}}$ obeying (1.3) satisfies the following inequality [8], Propositions 10 and 11:

$$\left\| x_n - T(x_n) \right\| \le \frac{\mathrm{d}(x_0, \mathrm{Fix}(T))}{\sqrt{\sum_{k=0}^{n} \alpha_k (1 - \alpha_k)}} \quad (n \in \mathbb{N}),$$

where $\mathrm{d}(x_0, \mathrm{Fix}(T)) := \min_{x \in \mathrm{Fix}(T)} \|x_0 - x\|$. When α_n ($n \in \mathbb{N}$) is a constant in the range of $(0, 1)$, which is the most tractable choice of step size satisfying (1.3), the Krasnosel'skiĭ-

Mann algorithm (1.2) has the rate of convergence,

$$\left\| x_n - T(x_n) \right\| = O\left(\frac{1}{\sqrt{n+1}} \right). \tag{2.13}$$

Meanwhile, according to Theorem 5 in [17], Algorithm (1.2) with $(\alpha_n)_{n \in \mathbb{N}}$ satisfying the Armijo-type condition (1.5) satisfies, for all $n \in \mathbb{N}$,

$$\left\| x_n - T(x_n) \right\| \leq \frac{\| x_0 - T(x_0) \|}{\sqrt{\beta \sum_{k=0}^{n} (\alpha_k - \frac{1}{2})^2}}. \tag{2.14}$$

In general, the step sizes satisfying (1.3) do not coincide with those satisfying the Armijo-type condition (1.5) or the Wolfe-type conditions (2.1) and (2.2). This is because the line search methods based on the Armijo-type conditions (1.5) and (2.1) determine step sizes at each iteration n so as to satisfy $\|x_{n+1} - T(x_{n+1})\| < \|x_n - T(x_n)\|$, while the constant step sizes satisfying (1.3) do not change at each iteration. Accordingly, it would be difficult to evaluate the efficiency of these algorithms by using only the theoretical convergence rates in (2.13), (2.14), and Theorem 2.6. To verify whether Algorithm 2.1 with the convergence rates in Theorem 2.6 converges faster than the previous algorithms [8], Propositions 10 and 11, [17], Theorem 5, with convergence rates (2.13) and (2.14), the next section numerically compares their abilities to solve concrete constrained smooth convex optimization problems.

3 Application of Algorithm 2.1 to constrained smooth convex optimization

This section considers the following problem:

$$\text{Minimize } f(x) \text{ subject to } x \in C, \tag{3.1}$$

where $f \colon \mathbb{R}^d \to \mathbb{R}$ is convex, $\nabla f \colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous with a constant L, and $C \subset \mathbb{R}^d$ is a nonempty, closed, and convex set onto which the metric projection P_C can be efficiently computed.

3.1 Experimental conditions and fixed point and line search algorithms used in the experiment

Problem (3.1) can be solved by using the conventional Krasnosel'skiĭ-Mann algorithm (1.2) with a nonexpansive mapping $T := P_C(\mathrm{Id} - \lambda \nabla f)$ satisfying $\mathrm{Fix}(T) = \mathrm{argmin}_{x \in C} f(x)$, where $\lambda \in (0, 2/L]$ [22], Proposition 2.3. It is represented as follows:

$$x_{n+1} = x_n + \alpha_n \big(P_C \big(x_n - \lambda \nabla f(x_n) \big) - x_n \big), \tag{3.2}$$

where $x_0 \in \mathbb{R}^d$ and $(\alpha_n)_{n \in \mathbb{N}}$ is a sequence satisfying (1.3) or the Armijo-type condition (1.5). Algorithm 2.1 with $T := P_C(\mathrm{Id} - \lambda \nabla f)$ is as follows:

$$\begin{aligned} x_{n+1} &:= x_n + \alpha_n d_n, \\ d_{n+1} &:= -\big(x_{n+1} - P_C \big(x_{n+1} - \lambda \nabla f(x_{n+1}) \big) \big) + \beta_n d_n, \end{aligned} \tag{3.3}$$

where $x_0, d_0 := -(x_0 - P_C(x_0 - \lambda\nabla f(x_0))) \in \mathbb{R}^d$, $(\alpha_n)_{n\in\mathbb{N}}$ is a sequence satisfying the Wolfe-type conditions (2.1) and (2.2), and $(\beta_n)_{n\in\mathbb{N}}$ is defined by each of equations (2.4) with $T := P_C(\mathrm{Id} - \lambda\nabla f)$ (see also (1.21)).

The best conventional nonlinear conjugate gradient method for unconstrained smooth nonconvex optimization was proposed by Hager and Zhang [29, 36], and it uses the HS formula defined as in (1.19):

$$
\beta_n^{\mathrm{HZ}} := \frac{1}{\langle d_n, y_n\rangle}\left\langle y_n - 2d_n\frac{\|y_n\|^2}{\langle d_n, y_n\rangle}, \nabla f(x_{n+1})\right\rangle
$$
$$
= \beta_n^{\mathrm{HS}} - 2\frac{\|y_n\|^2}{\langle d_n, y_n\rangle}\frac{\langle\nabla f(x_{n+1}), d_n\rangle}{\langle d_n, y_n\rangle}.
$$

Replacing ∇f in the above formula with $\mathrm{Id} - P_C(\mathrm{Id} - \lambda\nabla f)$ leads to the HZ-type formula for Problem (3.1):

$$
\beta_n^{\mathrm{HZ}} := \beta_n^{\mathrm{HS}} - 2\frac{\|y_n\|^2}{\langle d_n, y_n\rangle}\frac{\langle x_{n+1} - P_C(x_{n+1} - \lambda\nabla f(x_{n+1})), d_n\rangle}{\langle d_n, y_n\rangle}, \tag{3.4}
$$

where $y_n := (x_{n+1} - P_C(x_{n+1} - \lambda\nabla f(x_{n+1}))) - (x_n - P_C(x_n - \lambda\nabla f(x_n)))$ and β_n^{HS} is defined by $\beta_n^{\mathrm{HS}} := \langle x_{n+1} - P_C(x_{n+1} - \lambda\nabla f(x_{n+1})), y_n\rangle/\langle d_n, y_n\rangle$. We tested Algorithm (3.3) with $\beta_n := \beta_n^{\mathrm{HZ}}$ defined by (3.4) in order to see how it works on Problem (3.1).

We used the Virtual Desktop PC at the Ikuta campus of Meiji University. The PC has 8 GB of RAM memory, 1 core Intel Xeon 2.6 GHz CPU, and a Windows 8.1 operating system. The algorithms used in the experiment were written in MATLAB (R2013b), and they are summarized as follows.

SD-1:　Algorithm (3.2) with constant step sizes $\alpha_n := 0.5$ ($n \in \mathbb{N}$) [1], Theorem 5.14.

SD-2:　Algorithm (3.2) with α_n satisfying the Armijo-type condition (1.5) when $\beta = 0.5$ [17], Theorems 4 and 8.

SD-3:　Algorithm (3.3) with α_n satisfying the Wolfe-type conditions (2.1) and (2.2) and $\beta_n := \beta_n^{\mathrm{SD}}$ (Theorem 2.1).

FR:　Algorithm (3.3) with α_n satisfying the Wolfe-type conditions (2.1) and (2.2) and $\beta_n := \beta_n^{\mathrm{FR}}$ (Theorem 2.3).

PRP+:　Algorithm (3.3) with α_n satisfying the Wolfe-type conditions (2.1) and (2.2) and $\beta_n := \beta_n^{\mathrm{PRP+}}$ (Theorem 2.4).

HS+:　Algorithm (3.3) with α_n satisfying the Wolfe-type conditions (2.1) and (2.2) and $\beta_n := \beta_n^{\mathrm{HS+}}$ (Theorem 2.5).

DY:　Algorithm (3.3) with α_n satisfying the Wolfe-type conditions (2.1) and (2.2) and $\beta_n := \beta_n^{\mathrm{DY}}$ (Theorem 2.2).

HZ:　Algorithm (3.3) with α_n satisfying the Wolfe-type conditions (2.1) and (2.2) and $\beta_n := \beta_n^{\mathrm{HZ}}$ defined by (3.4) [29, 36].

The experiment used the following line search algorithm [21], Algorithm 4.6, to find step sizes satisfying the Wolfe-type conditions (2.1) and (2.2) with $\delta := 0.3$ and $\sigma := 0.5$ that were chosen by referring to [21], Section 6.1, where, for each n, $A_n(\cdot)$ and $W_n(\cdot)$ are

$$
A_n(t): \quad \|x_n(t) - T(x_n(t))\|^2 - \|x_n - T(x_n)\|^2 < \delta t\langle x_n - T(x_n), d_n\rangle,
$$
$$
W_n(t): \quad \langle x_n(t) - T(x_n(t)), d_n\rangle > \sigma\langle x_n - T(x_n), d_n\rangle.
$$

Algorithm 3.1 ([21], Algorithm 4.6)

Require: $A_n(\cdot), W_n(\cdot)$.

Ensure: $A_n(\alpha)$ and $W_n(\alpha)$.

  $\alpha \leftarrow 0, \beta \leftarrow \infty, t \leftarrow 1.$

  loop

    if $\neg A_n(t)$ **then**

      $\beta \leftarrow t.$

    else if $\neg W_n(t)$ **then**

      $\alpha \leftarrow t.$

    else

      (α: *found*).

    end if

    if $\beta < \infty$ **then**

      $t \leftarrow \frac{1}{2}(\alpha + \beta).$

    else

      $t \leftarrow 2\alpha.$

    end if

  end loop

For Algorithm SD-2, we replaced $A_n(\cdot)$ above by

$$A_n(t): \quad g_n(t) - g_n(0) < -Dt \left\| x_n - T(x_n) \right\|^2,$$

where $D := \delta = 0.3$ and g_n is defined as in (1.6), and deleted $W_n(\cdot)$ from the line search algorithm. For Algorithms FR, PRP+, HS+, DY, and HZ, if the step sizes satisfying the Wolfe-type conditions (2.1) and (2.2) were not computed by using Algorithm 3.1, the step sizes were computed by using Algorithm 3.1 when $d_n := -(x_n - T(x_n))$. This is because Algorithm 3.1 for Algorithm SD-3, which uses $d_n := -(x_n - T(x_n))$ ($n \in \mathbb{N}$), had a 100% success rate in computing the step sizes satisfying (2.1) and (2.2). Tables 1, 2, 3, and 4 indicate the satisfiability rates (defined below) of computing the step sizes for the algorithms in the experiment.

The stopping condition was

$$n = 10 \quad \text{or} \quad \left\| x_{n_0} - T(x_{n_0}) \right\| = 0 \quad \text{for some } n_0 \in [0, 10]. \tag{3.5}$$

Before describing the results, let us describe the notation used to verify the numerical performance of the algorithms.

- I: the number of initial points;
- $x_0^{(i)}$: the initial point chosen randomly ($i = 1, 2, \ldots, I$);
- ALGO: each of Algorithms SD-1, SD-2, SD-3, FR, PRP+, HS+, DY, and HZ (ALGO $\in$ {SD-1, SD-2, SD-3, FR, PRP+, HS+, DY, HZ});
- $N_1(x_0^{(i)}, \text{ALGO})$: the number of step sizes computed by Algorithm 3.1 for ALGO with $x_0^{(i)}$ before ALGO satisfies the stopping condition (3.5);
- $N_2(x_0^{(i)}, \text{ALGO})$: the number of iterations needed to satisfy the stopping condition (3.5) for ALGO with $x_0^{(i)}$.

Note that $N_1(x_0^{(i)}, \text{SD-1})$ stands for the number of iterations n satisfying $A_n(0.5)$ and $W_n(0.5)$ before Algorithm SD-1 with $x_0^{(i)}$ satisfies the stopping condition (3.5). The satisfiability rate (SR) of Algorithm 3.1 to compute the step sizes for each of the algorithms is defined by

$$\text{SR(ALGO)} := \frac{\sum_{i=1}^{I} N_1(x_0^{(i)}, \text{ALGO})}{\sum_{i=1}^{I} N_2(x_0^{(i)}, \text{ALGO})} \times 100 \ [\%]. \tag{3.6}$$

We performed 100 samplings, each starting from different random initial points (*i.e.*, $I :=$ 100) and averaged their results.

3.2 Constrained quadratic programming problem
In this subsection, let us consider the following constrained quadratic programming problem:

Problem 3.1 Suppose that C is a nonempty, closed convex subset of $\mathbb{R}^d$ onto which P_C can be efficiently computed, $Q \in \mathbb{R}^{d \times d}$ is positive semidefinite with the eigenvalues $\lambda_{\min} := \lambda_1, \lambda_2, \ldots, \lambda_d =: \lambda_{\max}$ satisfying $\lambda_i \leq \lambda_j$ $(i \leq j)$, and $b \in \mathbb{R}^d$. Our objective is to

$$\text{minimize } f(x) := \frac{1}{2} \langle x, Qx \rangle + \langle b, x \rangle \text{ subject to } x \in C.$$

Since f above is convex and $\nabla f(x) = Qx + b$ $(x \in \mathbb{R}^d)$ is Lipschitz continuous such that the Lipschitz constant of ∇f is the maximum eigenvalue $\lambda_{\max}$ of Q, Problem 3.1 is an example of Problem (3.1).

We compared the proposed algorithms SD-3, FR, PRP+, HS+, DY, and HZ with the previous algorithms SD-1 and SD-2 by applying them to Problem 3.1 (*i.e.*, the fixed point problem for $T(x) := P_C(x - (2/\lambda_{\max})(Qx + b))$ $(x \in \mathbb{R}^d)$) in the following cases:

$$d := 10^3 \text{ or } 10^4, \qquad \lambda_{\min} := 0, \qquad \lambda_{\max} := d, \qquad \lambda_i \in [0, d] \quad (i = 2, 3, \ldots, d-1),$$

$$b, c \in (-32, 32)^d, \qquad C := \left\{ x \in \mathbb{R}^d : \|x - c\| \leq 1 \right\}.$$

We randomly chose $\lambda_i \in [0, d]$ $(i = 2, 3, \ldots, d-1)$ and set Q as a diagonal matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{\max}$. The experiment used two random numbers in the range of $(-32, 32)^d$ for b and c to satisfy $C \cap \{x \in \mathbb{R}^d : \nabla f(x) = 0\} = \emptyset$. Since C is a closed ball with center c and radius 1, P_C can be computed within a finite number of arithmetic operations. More precisely, $P_C(x) := c + (x - c)/\|x - c\|$ if $\|x - c\| > 1$, or $P_C(x) := x$ if $\|x - c\| \leq 1$.

Table 1 shows the satisfiability rates as defined by (3.6) for Algorithms SD-1, SD-2, and SD-3 that are applied to Problem 3.1. It can be seen that the step sizes for SD-1 (constant step sizes $\alpha_n := 0.5$) do not always satisfy the Wolfe-type conditions (2.1) and (2.2),
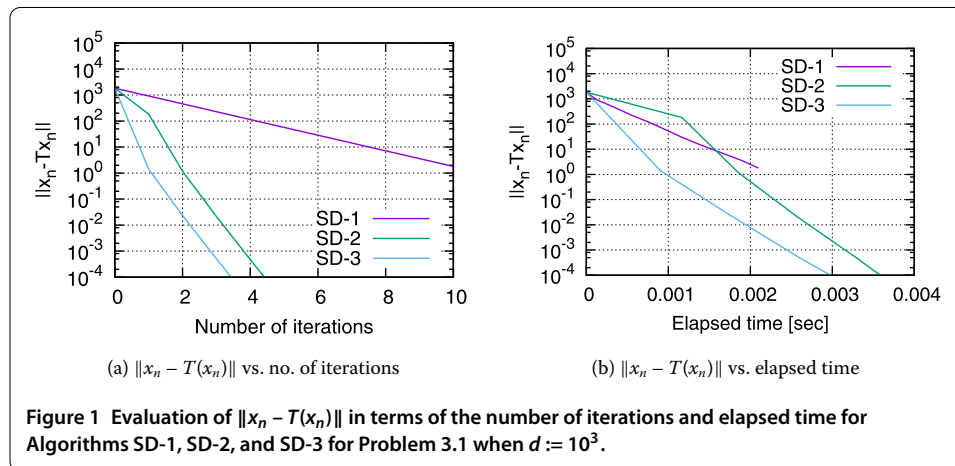
Table 1 Satisfiability rate of Algorithm 3.1 for Algorithms SD-1, SD-2, and SD-3 applied to Problem 3.1 when $d := 10^3, 10^4$

Algorithm	SR ($d := 10^3$)	SR ($d := 10^4$)
SD-1	55.9%	26.3%
SD-2	100%	100%
SD-3	100%	100%

Table 2 Satisfiability rate of Algorithm 3.1 for Algorithms FR, PRP+, HS+, DY, and HZ applied to Problem 3.1 when $d := 10^3, 10^4$

Algorithm	SR ($d := 10^3$)	SR ($d := 10^4$)
FR	19.7%	28.1%
PRP+	100%	100%
HS+	100%	98.9%
DY	21.6%	27.2%
HZ	20.0%	20.0%

(a) $\|x_n - T(x_n)\|$ vs. no. of iterations

(b) $\|x_n - T(x_n)\|$ vs. elapsed time

Figure 1 Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms SD-1, SD-2, and SD-3 for Problem 3.1 when $d := 10^3$.
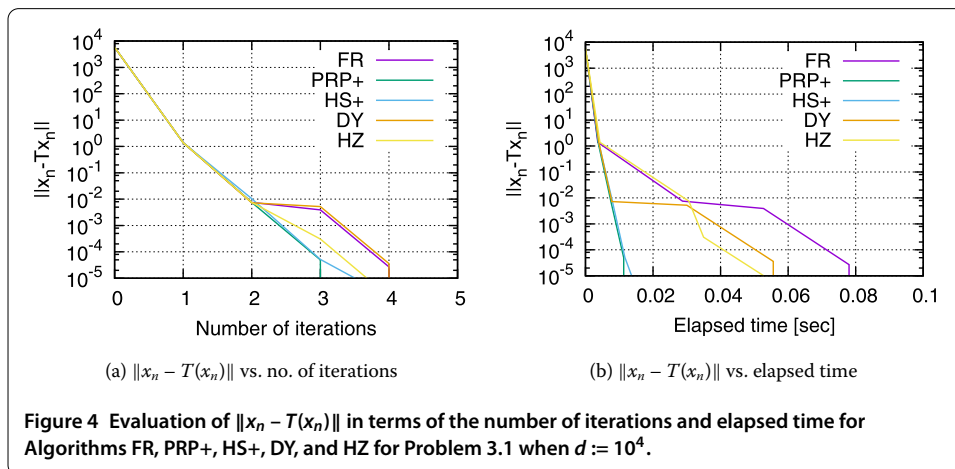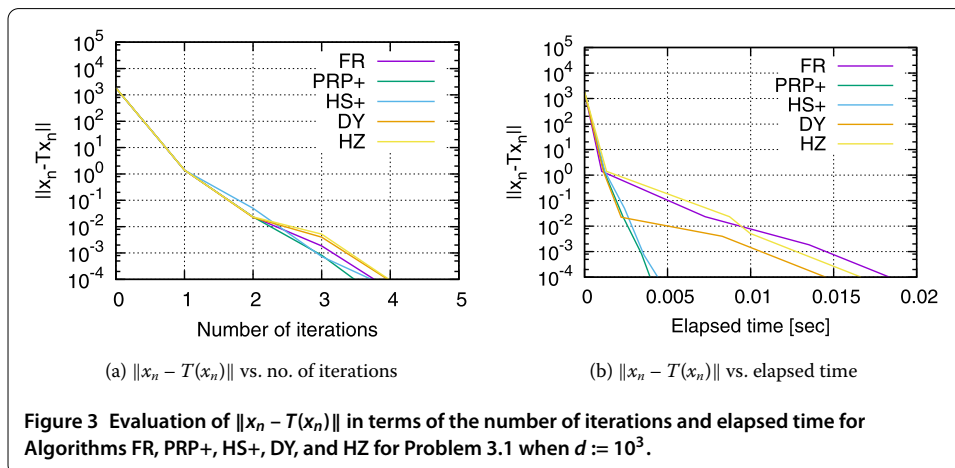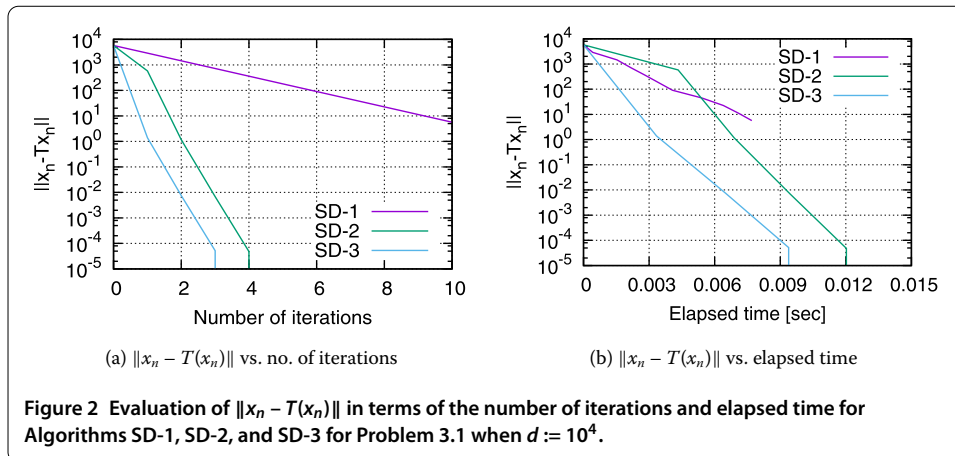
whereas the step sizes computed by Algorithm 3.1 and SD-2 (resp. Algorithm SD-3) definitely satisfy the Armijo-type condition (1.5) (resp. the Wolfe-type conditions (2.1) and (2.2)).

Table 2 showing the satisfiability rates for Algorithms FR, PRP+, HS+, DY, and HZ indicates that Algorithm 3.1 for PRP+ and HS+ has high success rates at computing the step sizes satisfying (2.1) and (2.2), while the SRs of Algorithm 3.1 for other algorithms are low. It can be seen from Tables 1 and 2 that SD-3, PRP+, and HS+ are robust in the sense that Algorithm 3.1 can compute the step sizes satisfying the Wolfe-type conditions (2.1) and (2.2).

Figure 1 indicates the behaviors of SD-1, SD-2, and SD-3 when $d := 10^3$. The y-axes in Figures 1(a) and 1(b) represent the value of $\|x_n - T(x_n)\|$. The x-axis in Figure 1(a) represents the number of iterations, and the x-axis in Figure 1(b) represents the elapsed time. If the $(\|x_n - T(x_n)\|)_{n \in \mathbb{N}}$ generated by the algorithms converges to 0, they also converge to a fixed point of T. Figure 1(a) shows that SD-2 and SD-3 terminate at fixed points of T within a finite number of iterations. It can be seen from Figure 1(a) and Figure 1(b) that SD-3 reduces the iterations and running time needed to find a fixed point compared with SD-2. These figures also show that $(\|x_n - T(x_n)\|)_{n \in \mathbb{N}}$ generated by SD-1 converges slowest and that SD-1 cannot find a fixed point of T before the tenth iteration. We can thus see that the use of the step sizes satisfying the Wolfe-type conditions is a good way to solve fixed point problems by using the Krasnosel'skiĭ-Mann algorithm. Figure 2 indicates the behaviors of SD-1, SD-2, and SD-3 when $d := 10^4$. Similarly to what is shown in Figure 1, SD-3 finds a fixed point of T faster than SD-1 and SD-2 can.

Figure 3 is the evaluation of $(\|x_n - T(x_n)\|)_{n \in \mathbb{N}}$ in terms of the number of iterations and elapsed time for Algorithms FR, PRP+, HS+, DY, and HZ when $d := 10^3$. Figure 3(a) shows

Figure 2 Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms SD-1, SD-2, and SD-3 for Problem 3.1 when $d := 10^4$.

Figure 3 Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms FR, PRP+, HS+, DY, and HZ for Problem 3.1 when $d := 10^3$.

Figure 4 Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms FR, PRP+, HS+, DY, and HZ for Problem 3.1 when $d := 10^4$.

that they can find fixed points of T within a finite number of iterations. Figure 3(b) indicates that PRP+ and HS+ find the fixed points of T faster than FR, DY, and HZ. This is because Algorithm 3.1 for each of PRP+ and HS+ has a 100% success rate at computing the step sizes satisfying (2.1) and (2.2), while the SRs of Algorithm 3.1 for FR, DY, and HZ are low (see Table 2); *i.e.*, FR, DY, and HZ require much more time to compute the step

sizes than PRP+ and HS+. In fact, we checked that the times to compute the step sizes for FR, DY, and HZ account for 92.672202%, 87.156303%, and 83.700936% of all the computational times, while the times to compute the step sizes for PRP+ and HS+ account for 60.725204% and 60.889635% of all the computational times. Figure 4 indicate the behaviors of FR, PRP+, HS+, DY, and HZ when $d := 10^4$ and PRP+ and HS+ perform better than FR, DY, and HZ, as seen in Figure 3. Such a trend can also be verified from Table 2 showing that the SRs of Algorithm 3.1 for PRP+ and HS+ are about 100%.

3.3 Generalized convex feasibility problem

This subsection considers the following generalized convex feasibility problem [33], Section I, Framework 2, [37], Section 2.2, [6], Definition 4.1:

Problem 3.2 Suppose that C_i $(i = 0, 1, \ldots, m)$ is a nonempty, closed convex subset of $\mathbb{R}^d$ onto which P_{C_i} can be efficiently computed and define the weighted mean square value of the distances from $x \in \mathbb{R}^d$ to C_i $(i = 1, 2, \ldots, m)$ as $f(x)$ below; *i.e.*, for $w_i \in (0, 1)$ $(i = 1, 2, \ldots, m)$ satisfying $\sum_{i=1}^{m} w_i = 1$,

$$f(x) := \sum_{i=1}^{m} w_i \left(\min_{y \in C_i} \|x - y\| \right)^2.$$

Our objective is to find a point in the generalized convex feasible set defined by

$$C_f := \left\{ x^\star \in C_0 : f(x^\star) = \min_{x \in C_0} f(x) \right\}.$$

C_f is a subset of C_0 having the elements closest to C_i $(i = 1, 2, \ldots, m)$ in terms of the weighted mean square norm. Even if $\bigcap_{i=0}^{m} C_i = \emptyset$, C_f is well defined because C_f is the set of all minimizers of f over C_0. The condition $C_f \neq \emptyset$ holds when C_0 is bounded [6], Remark 4.3(a). Moreover, $C_f = \bigcap_{i=0}^{m} C_i$ holds when $\bigcap_{i=0}^{m} C_i \neq \emptyset$. Accordingly, Problem 3.2 is a generalization of the convex feasibility problem [5] of finding a point in $\bigcap_{i=0}^{m} C_i \neq \emptyset$.

The convex function f in Problem 3.2 satisfies $\nabla f = \text{Id} - \sum_{i=1}^{m} w_i P_{C_i}$. Hence, ∇f is Lipschitz continuous when its Lipschitz constant is two. This means Problem 3.2 is an example of Problem (3.1). Since Problem 3.2 can be expressed as the problem of finding a fixed point of $T = P_{C_0}(\text{Id} - \lambda \nabla f) = P_{C_0}(\text{Id} - \lambda(\text{Id} - \sum_{i=1}^{m} w_i P_{C_i}))$ for $\lambda \in (0, 1]$, we used T with $\lambda = 1$; *i.e.*, $T := P_{C_0}(\sum_{i=1}^{m} w_i P_{C_i})$.

We applied SD-1, SD-2, SD-3, FR, PRP+, HS+, DY, and HZ to Problem 3.2 in the following cases:

$$d := 10^3 \text{ or } 10^4, \qquad m := 99, \qquad w_i := \frac{1}{99} \quad (i = 1, 2, \ldots, 99),$$

$$c_i \in (-32, 32)^d, \qquad C_i := \left\{ x \in \mathbb{R}^d : \|x - c_i\| \leq 1 \right\} \quad (i = 0, 1, \ldots, m).$$

The experiment used one hundred random numbers in the range of $(-32, 32)^d$ for c_i, which means $\bigcap_{i=0}^{m} C_i = \emptyset$. Since C_i $(i = 0, 1, \ldots, m)$ is a closed ball with center c_i and radius 1, P_i can be computed within a finite number of arithmetic operations.
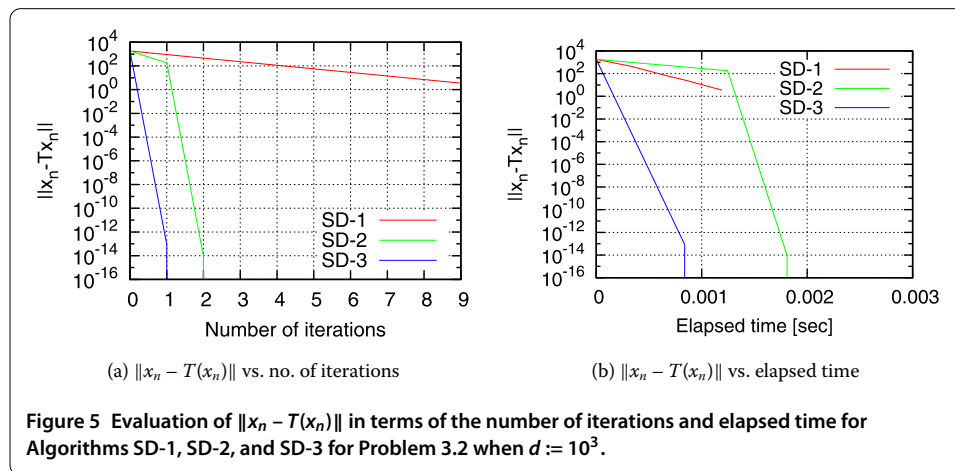
Table 3 shows the satisfiability rates as defined by (3.6) for Algorithms SD-1, SD-2, and SD-3 applied to Problem 3.2. It can be seen that the step sizes for SD-1 do not always

Table 3 Satisfiability rate of Algorithm 3.1 for Algorithms SD-1, SD-2, and SD-3 applied to Problem 3.2 when $d := 10^3, 10^4$

Algorithm	SR ($d := 10^3$)	SR ($d := 10^4$)
SD-1	80.6%	64.2%
SD-2	100%	100%
SD-3	100%	100%

Table 4 Satisfiability rate of Algorithm 3.1 for Algorithms FR, PRP+, HS+, DY, and HZ applied to Problem 3.2 when $d := 10^3, 10^4$
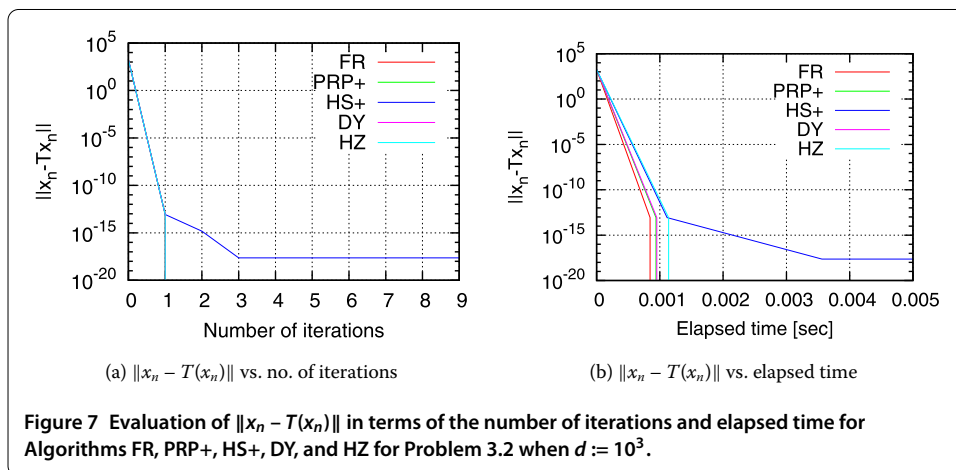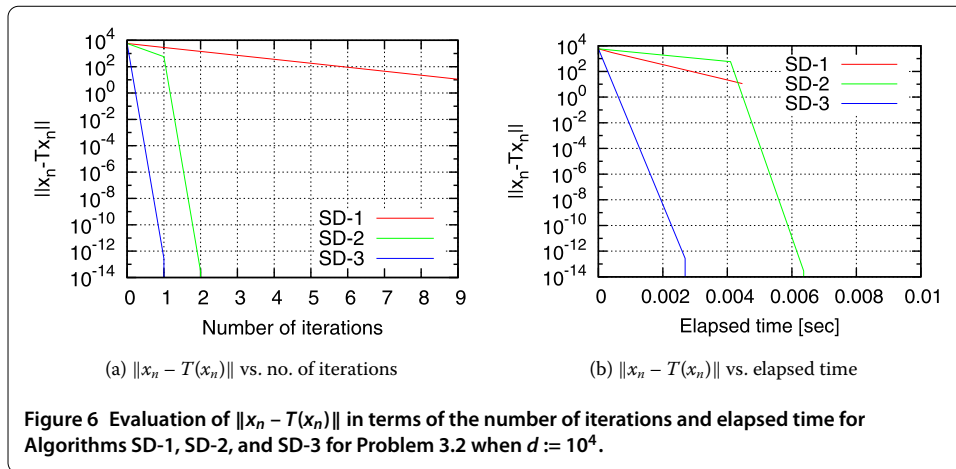
Algorithm	SR ($d := 10^3$)	SR ($d := 10^4$)
FR	50.0%	50.0%
PRP+	100%	100%
HS+	55.8%	60.4%
DY	50.0%	50.0%
HZ	50.0%	50.0%

(a) $\|x_n - T(x_n)\|$ vs. no. of iterations

(b) $\|x_n - T(x_n)\|$ vs. elapsed time

Figure 5 Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms SD-1, SD-2, and SD-3 for Problem 3.2 when $d := 10^3$.

satisfy the Wolfe-type conditions (2.1) and (2.2), whereas the step sizes computed by Algorithm 3.1 and SD-2 (resp. Algorithm SD-3) definitely satisfy the Armijo-type condition (1.5) (resp. the Wolfe-type conditions (2.1) and (2.2)). Such a trend also existed when SD-1, SD-2, and SD-3 were applied to Problem 3.1 (see Table 1).

Table 4 shows the satisfiability rates for Algorithms FR, PRP+, HS+, DY, and HZ. The table indicates that Algorithm 3.1 for PRP+ has a 100% success rate at computing the step sizes satisfying (2.1) and (2.2), while the SRs of Algorithm 3.1 for the other algorithms lie between 50% and about 60%. From Tables 3 and 4, we can see that SD-3 and PRP+ are robust in the sense that Algorithm 3.1 can compute the step sizes satisfying the Wolfe-type conditions (2.1) and (2.2).

Figure 5 indicates the behaviors of SD-1, SD-2, and SD-3 when $d := 10^3$. The y-axes represent the value of $\|x_n - T(x_n)\|$. The x-axis in Figure 5(a) represents the number of iterations, and the x-axis in Figure 5(b) represents the elapsed time. From Figure 5(a), the iterations needed to satisfy $\|x_n - T(x_n)\| = 0$ for SD-2 and SD-3 are, respectively, 3 and 2. It can be seen that SD-3 reduces the running time and iterations needed to find a fixed point compared with SD-2. These figures also show that the $(\|x_n - T(x_n)\|)_{n \in \mathbb{N}}$ generated by SD-1 converges slowest. Therefore, we can see that the use of the step sizes satisfying the Wolfe-type conditions is a good way to solve fixed point problems by using the

(a) $\|x_n - T(x_n)\|$ vs. no. of iterations

(b) $\|x_n - T(x_n)\|$ vs. elapsed time

Figure 6 **Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms SD-1, SD-2, and SD-3 for Problem 3.2 when $d := 10^4$.**

(a) $\|x_n - T(x_n)\|$ vs. no. of iterations

(b) $\|x_n - T(x_n)\|$ vs. elapsed time

Figure 7 **Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms FR, PRP+, HS+, DY, and HZ for Problem 3.2 when $d := 10^3$.**

Krasnosel'skiĭ-Mann algorithm, as seen in Figures 1 and 2 illustrating the behaviors of SD-1, SD-2, and SD-3 on Problem 3.1 when $d := 10^3, 10^4$. Figure 6 indicates the behaviors of SD-1, SD-2, and SD-3 when $d := 10^4$. Similarly to what is shown in Figure 5, SD-3 finds a fixed point of T faster than SD-1 and SD-2 can.

Figure 7(a) is the evaluation of $(\|x_n - T(x_n)\|)_{n \in \mathbb{N}}$ in terms of the number of iterations for Algorithms FR, PRP+, HS+, DY, and HZ when $d := 10^3$. Except for HS+, the algorithms approximate the fixed points of T very rapidly. It can also be seen that the algorithms other than HS+ satisfy $\|x_2 - T(x_2)\| = 0$. Figure 7(b) is the evaluation of $(\|x_n - T(x_n)\|)_{n \in \mathbb{N}}$ in terms of the elapsed time. Here, we can see that FR, PRP+, and DY can find fixed points of T faster than SD-1 and SD-2 (Figure 5). Figure 8 indicates the behaviors of FR, PRP+, HS+, DY, and HZ when $d := 10^4$. The results in these figures are almost the same as the ones in Figure 7.

From the above numerical results, we can conclude that the proposed algorithms can find optimal solutions to Problems 3.1 and 3.2 faster than the previous fixed point algorithms can. In particular, it can be seen that the algorithms for which the SRs of Algorithm 3.1 are high converge quickly to solutions of Problems 3.1 and 3.2.

(a) $\|x_n - T(x_n)\|$ vs. no. of iterations (b) $\|x_n - T(x_n)\|$ vs. elapsed time
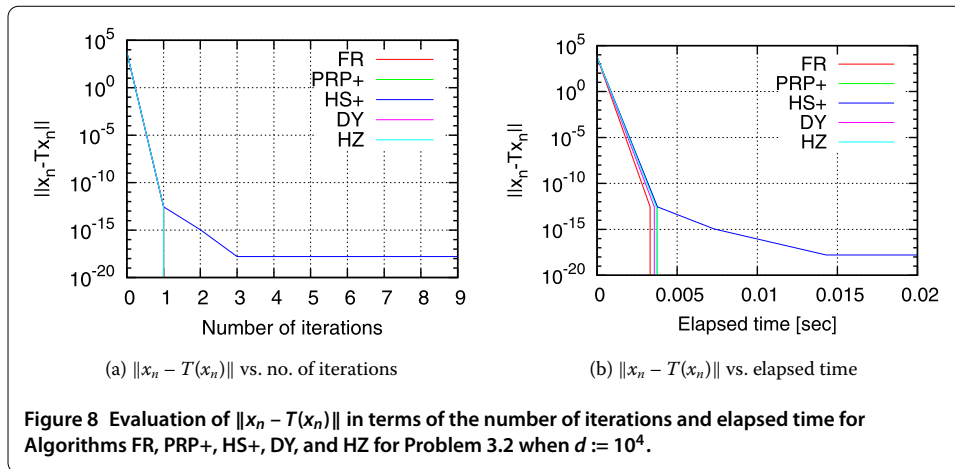
Figure 8 Evaluation of $\|x_n - T(x_n)\|$ in terms of the number of iterations and elapsed time for Algorithms FR, PRP+, HS+, DY, and HZ for Problem 3.2 when $d := 10^4$.

4 Conclusion and future work

This paper discussed the fixed point problem for a nonexpansive mapping on a real Hilbert space and presented line search fixed point algorithms for solving it on the basis of nonlinear conjugate gradient methods for unconstrained optimization and their convergence analyses and convergence rate analyses. Moreover, we used these algorithms to solve concrete constrained quadratic programming problems and generalized convex feasibility problems and numerically compared them with the previous fixed point algorithms based on the Krasnosel'skiĭ-Mann fixed point algorithm. The numerical results showed that the proposed algorithms can find optimal solutions to these problems faster than the previous algorithms.

In the experiment, the line search algorithm (Algorithm 3.1) could not compute appropriate step sizes for fixed point algorithms other than Algorithms SD-2, SD-3, and PRP+. In the future, we should consider modifying the algorithms to enable the line search to compute appropriate step sizes. Or we may need to develop new line searches that can be applied to all of the fixed point algorithms considered in this paper.

The main objective of this paper was to devise line-search fixed-point algorithms to accelerate the previous Krasnosel'skiĭ-Mann fixed point algorithm defined by (1.2), *i.e.*, $x_{n+1} := \lambda_n x_n + (1 - \lambda_n) T(x_n)$ ($n \in \mathbb{N}$), where $(\lambda_n)_{n \in \mathbb{N}} \subset [0,1]$ with $\sum_{n=0}^{\infty} \lambda_n (1 - \lambda_n) = \infty$ and $x_0 \in H$ is an initial point. Another particularly interesting problem is determining whether or not there are line search fixed point algorithms to accelerate the following Halpern fixed point algorithm [11, 12]: for all $n \in \mathbb{N}$,

$$x_{n+1} := \alpha_n x_0 + (1 - \alpha_n) T(x_n),$$

where $(\alpha_n)_{n \in \mathbb{N}} \subset (0,1)$ satisfies $\lim_{n \to \infty} \alpha_n = 0$ and $\sum_{n=0}^{\infty} \alpha_n = \infty$. The Halpern algorithm can minimize the convex function $\| \cdot -x_0 \|^2$ over Fix(T) (see, *e.g.*, [7], Theorem 6.17). A previously reported result [38], Theorem 3.1, Proposition 3.2, showed that there is an inconvenient possibility that the Halpern-type algorithm with a diminishing step size sequence (*e.g.*, $\alpha_n := 1/(n+1)^a$, where $a \in (0,1]$) and any of the FR, PRP, HS, and DY formulas used in the conventional conjugate gradient methods may not converge to the minimizer of $\| \cdot -x_0 \|^2$ over Fix(T). However, there is room for further research into devising line search fixed point algorithms to accelerate the Halpern algorithm with a diminishing step size sequence.

Endnotes
   [a] See Theorem 2.6(i) for the details of the convergence rate of the proposed algorithm when $d_n := -(x_n - T(x_n))$ $(n \in \mathbb{N})$.
   [b] To guarantee the convergence of the PRP and HS methods for unconstrained optimization, the formulas $\beta_n^{\text{PRP+}} := \max\{\beta_n^{\text{PRP}}, 0\}$ and $\beta_n^{\text{HS+}} := \max\{\beta_n^{\text{HS}}, 0\}$ were presented in [35]. We use the modifications to perform the convergence analyses on the proposed line search fixed point algorithms.

References
   1. Bauschke, HH, Combettes, PL: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
   2. Goebel, K, Kirk, WA: Topics in Metric Fixed Point Theory. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (1990)
   3. Goebel, K, Reich, S: Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings. Dekker, New York (1984)
   4. Takahashi, W: Nonlinear Functional Analysis. Yokohama Publishers, Yokohama (2000)
   5. Bauschke, HH, Borwein, JM: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367-426 (1996)
   6. Yamada, I: The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In: Butnariu, D, Censor, Y, Reich, S (eds.) Inherently Parallel Algorithms for Feasibility and Optimization and Their Applications, pp. 473-504. Elsevier, Amsterdam (2001)
   7. Berinde, V: Iterative Approximation of Fixed Points. Springer, Berlin (2007)
   8. Cominetti, R, Soto, JA, Vaisman, J: On the rate of convergence of Krasnosel'skiĭ-Mann iterations and their connection with sums of Bernoulli's. Isr. J. Math. **199**, 757-772 (2014)
   9. Krasnosel'skiĭ, MA: Two remarks on the method of successive approximations. Usp. Mat. Nauk **10**, 123-127 (1955)
   10. Mann, WR: Mean value methods in iteration. Proc. Am. Math. Soc. **4**, 506-510 (1953)
   11. Halpern, B: Fixed points of nonexpanding maps. Bull. Am. Math. Soc. **73**, 957-961 (1967)
   12. Wittmann, R: Approximation of fixed points of nonexpansive mappings. Arch. Math. **58**, 486-491 (1992)
   13. Nakajo, K, Takahashi, W: Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups. J. Math. Anal. Appl. **279**, 372-379 (2003)
   14. Solodov, MV, Svaiter, BF: Forcing strong convergence of proximal point iterations in a Hilbert space. Math. Program. **87**, 189-202 (2000)
   15. Boţ, RI, Csetnek, ER: A dynamical system associated with the fixed points set of a nonexpansive operator. J. Dyn. Differ. Equ. (2015). doi:10.1007/s10884-015-9438-x
   16. Combettes, PL, Pesquet, JC: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE J. Sel. Top. Signal Process. **1**, 564-574 (2007)
   17. Magnanti, TL, Perakis, G: Solving variational inequality and fixed point problems by line searches and potential optimization. Math. Program. **101**, 435-461 (2004)
   18. Wolfe, P: Convergence conditions for ascent methods. SIAM Rev. **11**, 226-235 (1969)
   19. Wolfe, P: Convergence conditions for ascent methods. II: some corrections. SIAM Rev. **13**, 185-188 (1971)
   20. Nocedal, J, Wright, SJ: Numerical Optimization, 2nd edn. Springer Series in Operations Research and Financial Engineering. Springer, Berlin (2006)
   21. Lewis, AS, Overton, ML: Nonsmooth optimization via quasi-Newton methods. Math. Program. **141**, 135-163 (2013)
   22. Iiduka, H: Iterative algorithm for solving triple-hierarchical constrained optimization problem. J. Optim. Theory Appl. **148**, 580-592 (2011)
   23. Hager, WW, Zhang, H: A survey of nonlinear conjugate gradient methods. Pac. J. Optim. **2**, 35-58 (2006)
   24. Hestenes, MR, Stiefel, EL: Methods of conjugate gradients for solving linear systems. J. Res. Natl. Bur. Stand. **49**, 409-436 (1952)
   25. Fletcher, R, Reeves, C: Function minimization by conjugate gradients. Comput. J. **7**, 149-154 (1964)
   26. Polak, E, Ribière, G: Note sur la convergence de directions conjugées. Rev. Fr. Autom. Inform. Rech. Opér., Anal. Numér. **3**, 35-43 (1969)
   27. Polyak, BT: The conjugate gradient method in extreme problems. USSR Comput. Math. Math. Phys. **9**, 94-112 (1969)
   28. Dai, YH, Yuan, Y: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. **10**, 177-182 (1999)
   29. Hager, WW, Zhang, H: A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim. **16**, 170-192 (2005)
   30. Al-Baali, M: Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA J. Numer. Anal. **5**, 121-124 (1985)
   31. Gilbert, JC, Nocedal, J: Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optim. **2**, 21-42 (1992)
   32. Zoutendijk, G: Nonlinear programming, computational methods. In: Abadie, J (ed.) Integer and Nonlinear Programming, pp. 37-38. North-Holland, Amsterdam (1970)
   33. Combettes, PL, Bondon, P: Hard-constrained inconsistent signal feasibility problems. IEEE Trans. Signal Process. **47**, 2460-2468 (1999)

34. Opial, Z: Weak convergence of the sequence of successive approximation for nonexpansive mappings. Bull. Am. Math. Soc. **73**, 591-597 (1967)

35. Powell, MJD: Nonconvex minimization calculations and the conjugate gradient method. In: Numerical Analysis (Dundee, 1983). Lecture Notes in Mathematics, vol. 1066, pp. 122-141. Springer, Berlin (1984)

36. Hager, WW, Zhang, H: Algorithm 851: CG_DESCENT: a conjugate gradient method with guaranteed descent. ACM Trans. Math. Softw. **32**, 113-137 (2006)

37. Iiduka, H: Iterative algorithm for triple-hierarchical constrained nonconvex optimization problem and its application to network bandwidth allocation. SIAM J. Optim. **22**, 862-878 (2012)

38. Iiduka, H: Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping. Math. Program. **149**, 131-165 (2015)