Journal of Global Optimization

On the global convergence of the proximal gradient method for Tikhonov regularized correction of absolute value equations

--Manuscript Draft--

Manuscript Number:	JOGO-D-24-00559
Full Title:	On the global convergence of the proximal gradient method for Tikhonov regularized correction of absolute value equations
Article Type:	Manuscript
Keywords:	Absolute value equation; Optimal correction; Tikhonov regularization; Fractional programming.
Abstract:	This paper considers the Tikhonov regularized regularization for infeasible absolute value equations (AVE), whose objective function is the sum of a non-differentiable fractional term and a quadratic term. We first reformulate this problem into a well-conditioned single-ratio fractional programming problem and apply the proximal gradient method (PGM) to solve it. Notably, we demonstrate that the objective function is naturally a Kurdyka-\L ojasiewicz (K\L) function, allowing us to establish global convergence of PGM without additional assumptions, such as the function being Lipschitz continuous or the gradient being locally Lipschitz continuous, that are indispensable in the existing literature. In addition, when the proximity parameter is adaptively chosen based on the iteration, PGM with extrapolation can also be used to solve this problem. To address the computational challenges posed by the absolute value in the proximal operator, we employ the smoothing technique to derive approximate local solutions. We conclude with numerical experiments that illustrate the effectiveness of the proposed algorithms.

Journal of Global Optimization manuscript No.
(will be inserted by the editor)

On the global convergence of the proximal gradient method for Tikhonov regularized correction of absolute value equations*

Yongxin Chen · Deren Han

Received: December 29, 2024/ Accepted: date

Abstract This paper considers the Tikhonov regularized regularization for infeasible absolute value equations (AVE), whose objective function is the sum of a non-differentiable fractional term and a quadratic term. We first reformulate this problem into a well-conditioned single-ratio fractional programming problem and apply the proximal gradient method (PGM) to solve it. Notably, we demonstrate that the objective function is naturally a Kurdyka-Łojasiewicz (KŁ) function, allowing us to establish global convergence of PGM without additional assumptions, such as the function being Lipschitz continuous or the gradient being locally Lipschitz continuous, that are indispensable in the existing literature. In addition, when the proximity parameter is adaptively chosen based on the iteration, PGM with extrapolation can also be used to solve this problem. To address the computational challenges posed by the absolute value in the proximal operator, we employ the smoothing technique to derive approximate local solutions. We conclude with numerical experiments that illustrate the effectiveness of the proposed algorithms.

Keywords Absolute value equation, Optimal correction, Tikhonov regularization, Fractional programming.

Mathematics Subject Classifications(2020) 65F10, 90C30, 90C32

* This work was supported by the Ministry of Science and Technology of China (No.2021YFA1003600), the R&D project of Pazhou Lab (Huangpu) (2023K0604), and the NSFC grant 12131004 and 12126603.

Yongxin Chen
LMIB, School of Mathematical Sciences, Beihang University, Beijing 100191, P.R. China.
E-mail: chenyongxin@buaa.edu.cn

Deren Han
Corresponding author, School of Mathematical Sciences, Beihang University, LMIB and NSLSCS, Beijing 100191, China.
E-mail: handr@buaa.edu.cn

1 Introduction

The absolute value equations (AVE) take the following form:

$$Ax - |x| = b, \tag{1.1}$$

where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, and $|\cdot|$ denotes the component-wise absolute value. This problem is popular to characterize a wide range of problems arising from many areas, such as the linear complementarity problem (LCP) [16], the theory of interval computations [15], the set-partitioning problem [29], and so on.

AVE and its general forms have received considerable attention and have been studied in [29, 32, 41]. The current main focus of AVE research is centered on the solvable cases. One main topic is the existence and uniqueness of solutions. AVE (1.1) is guaranteed to have a unique solution under several conditions, including when $\|A^{-1}\| < 1$ [32, 53], $\rho(|A^{-1}|) < 1$ where $\rho(\cdot)$ denotes the spectral radius [42], or the interval matrix $[A - I, A + I]$ is regular [20, 53]. Other characterizing conditions of unique solvability were presented in [21, 49]. For discussions on cases where AVE (1.1) has many (2^n) solutions, as well as studies on specific solutions, see [19, 20, 32, 33]. Another main topic of AVE research is designing numerical methods for solvable AVE. Numerous effective methods have been developed for solving AVE (1.1), e.g., the concave minimization method [28, 52], the successive linear programming method [31], the Piard iteration method [42, 44], the Newton-like method [5, 30, 51, 53], the SOR-like method [10, 23], the operator splitting methods [11], the smoothing approaches [1, 9, 43], and so on [50]. For further studies on solvable AVE, readers can refer to [21].

Due to some factors, such as data errors, modeling inaccuracies, and noise, we encounter problems that present as systems of infeasible AVE. To the best of our knowledge, research on infeasible AVE is still limited. A natural idea is to find the least square solution for such an infeasible AVE. Motivated by this, [24] considered minimal correction in the right-hand side vector b, formulating an optimization problem as follows:

$$\min_x \|Ax - |x| - b\|^2.$$

To guarantee the feasibility of the system, [26] applied changes simultaneously to the coefficient matrix A and the right-hand side vector b, leading to a more comprehensive model for AVE correction:

$$\begin{aligned} \min_{x,E,r} \quad & \|E\|_F^2 + \|r\|^2 \\ \text{s.t.} \quad & (A + E)x - |x| = b + r. \end{aligned} \tag{1.2}$$

This problem may be ill-conditioned, potentially resulting in solutions with excessively large norms (maybe infinity; see Example 1 in [35]). To control the norm of the solution vector, [25] introduced the Tikhonov regularized correction problem:

$$\begin{aligned} \min_{x,E,r} \quad & \|E\|_F^2 + \|r\|^2 + \lambda\|x\|^2 \\ \text{s.t.} \quad & (A + E)x - |x| = b + r, \end{aligned} \tag{1.3}$$

where λ is the regularization parameter. A similar alternative Tikhonov regularization model, which incorporates a quadratic constraint to limit the size of the solution, can also be found in [35]. Moreover, the optimal correction of infeasible absolute value equations systems using ℓ_p-norm regularization was proposed in [18].

This paper considers infeasible AVE with Tikhonov regularization:

$$\min_x F(x) := \frac{\|Ax - |x| - b\|^2}{1 + \|x\|^2} + \lambda \|x\|^2. \tag{1.4}$$

which is derived by reformulating (1.3)(see Lemma 2.5). This problem belongs to the fractional optimization problems arising from various applications in many fields, including economics [27], signal processing [38], the medical field [48], and their respective extensions. In particular, problem (1.4) falls within the category of single-ratio fractional programming problems. For tackling single-ratio fractional programming problems, one classical and popular algorithm is the Dinkelbach method [13, 22, 45]. It is worth mentioning that, inspired by Dinkelbach's approach, [17] reformulated problem (1.4) to a univariate equation, and the generalized Newton method was applied to solve it. To the best of our knowledge, this is the only paper that directly deals with problem (1.4); however, it does not provide a theoretical analysis of global convergence. For more details on single-ratio fractional programming problems, readers are referred to [8,46,47,54]. In this paper, we first recast problem (1.4) into a well-conditioned single-ratio fractional programming problem. With this nice reformulation and the Dinkelbach theorem, the proximal gradient method (PGM) is used for solving the reformulated problem. Both the proximal gradient method for solving non-convex problems and general first-order methods for single-ratio fractional optimization problems typically require additional, often stringent assumptions to guarantee convergence, let alone global convergence. However, we present that $F(x)$ in (1.4) is naturally a Kurdyka-Łojasiewicz (KŁ) function, and the global convergence of PGM can be established without any additional assumption. When the proximity parameters are adaptively chosen, PGM with extrapolation can also be employed to solve this problem. Similar convergence results can be obtained via a different potential function. In both algorithms, the resulting subproblem is equivalent to computing the proximal operator, which may take an expensive cost due to the existence of the absolute value. Hence, we utilize the smoothing techniques to obtain an approximate local solution of the subproblem. Finally, we give some numerical experiments to demonstrate the advantages of the proposed algorithms.

The structure of this paper is outlined as follows. In Section 2, we list some notations and preliminaries. In Section 3, we propose a proximal gradient method and establish its global convergence without any additional conditions. An enhanced version of this method is subsequently presented and analyzed in Section 4. Finally, we present some numerical results to demonstrate the effectiveness of the proposed algorithms in Section 5 and conclude this paper in Section 6.

2 Notations and prelimiaries

In this section, we present some notations and preliminaries which will be used throughout the paper.

We use $\mathbb{R}^n$ to denote the n-dimensional real space, and the inner product is denoted by $\langle \cdot, \cdot \rangle$. The induced norm $\|\cdot\|$ is defined by $\sqrt{\langle \cdot, \cdot \rangle}$. The Frobenius norm of a matrix is denoted by $\|\cdot\|_F$. The ball $B(x,\delta)$ denotes the open ball centered at x with radius δ.

For any closed set $\Omega \subseteq \mathbb{R}^n$ and any vector $x \in \mathbb{R}^n$, the distance from x to Ω denoted by $\operatorname{dist}(x,\Omega)$ is defined as $\operatorname{dist}(x,\Omega) = \inf_{y\in\Omega}\|x-y\|$. The effective domain of $f : \Omega \to \mathbb{R}\cup\{+\infty\}$ is defined as $\operatorname{dom} f = \{x \in \Omega \mid f(x) < +\infty\}$. For any numbers a and b such that $-\infty < a < b \leq +\infty$, we set $[a < f < b] = \{x \in \mathbb{R}^n \mid a < f(x) < b\}$. Other notation will be specified in the context.

2.1 Some definitions and basic results

The following definitions and results are elementary.

Definition 2.1 ([37] Coercieveness) *A function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is called coercive at x if*

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

Definition 2.2 ([37] Lower semi-continuity) *A function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is called lower semi-continous at x if*

$$f(x) \leq \liminf_{k\to+\infty} f(x^k)$$

for any sequence $\{x^k\}$ for which $\lim_{k\to+\infty} x^k = x$. A function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is called lower semi-continuous if it is lower semi-continuous at each point in $\mathbb{R}^n$.

A differential convex function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is L-smooth if ∇f is Lipschitz continuous, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|, \quad \forall x,y \in \mathbb{R}^n,$$

where L is called the Lipschitz modulus. A differential convex function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is $\nu-$strongly convex if and only if

$$f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle + \frac{\nu}{2}\|y-x\|^2.$$

A proper lower semi-continuous function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is called weakly convex (or semi-convex) if for some $\omega > 0$, the function $f(x) + \frac{\omega}{2}\|x\|^2$ is convex.

Lemma 2.1 ([37] Descent lemma) *If $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is L-smooth, for any $x,y \in \mathbb{R}^n$, we have*

$$f(y) \leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|x-y\|^2.$$

Definition 2.3 *Given a proper closed function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, we have*

(i) *the Fréchet subdifferential of f at each $x \in \operatorname{dom} f$ is the set of the vectors $u \in \mathbb{R}^n$ which satisfy*

$$\liminf_{y\neq x, y\to x} \frac{f(y) - f(x) - \langle u, y-x\rangle}{\|x-y\|} \geq 0.$$

The Fréchet subdifferential of f at x is written as $\widehat{\partial} f(x)$ and $\widehat{\partial} f(x) = \emptyset$ if $x \notin \operatorname{dom} f$.

(ii) *the limiting-subdifferential (or simply subdifferential) of g at $x \in dom\ f$ which is written as $\partial f(x)$ is defined as*

$$\partial f(x) = \{u \in \mathbb{R}^n \mid \exists\ x^k \to x, f(x^k) \to f(x), u^k \in \widehat{\partial} f(x^k) \to u\}.$$

When f is convex, $\hat{\partial} f$ and ∂f reduce to the classical subdifferential in convex analysis, i.e.,

$$\hat{\partial} f(x) = \partial f(x) = \{d \in \mathbb{R}^n \mid f(z) \geq f(x) + \langle z - x, d\rangle, \quad \forall z \in \mathbb{R}^n\}.$$

Lemma 2.2 ([39] Robbins-Siegmund theorem) *Let $\{z_k\}, \{b_k\}, \{c_k\}$ be three nonnegative sequences. Suppose that*

$$z_{k+1} \leq (1 + \varepsilon_k)z_k - b_k + c_k, \quad \sum_{k=0}^{+\infty} \varepsilon_k < +\infty \quad \textit{and} \quad \sum_{k=0}^{+\infty} c_k < +\infty.$$

Then, we have $\sum_{k=0}^{+\infty} b_k < +\infty$ and $\lim_{k\to+\infty} z_k$ exists.

2.2 Kurdyka-Łojasiewicz property

In this subsection, we introduce the Kurdyka-Łojasiewicz (KŁ) property, which plays a crucial role in presenting global convergence in non-convex optimization.

Definition 2.4 ([2] Kurdyka-Łojasiewicz property) *The function f is said to have Kurdyka-Łojasiewicz property at $x^* \in dom\ f$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of x^* and a continuous concave function $\phi : [0, \eta) \to \mathbb{R}_+$ such that*

(i) $\phi(0) = 0$*;*
(ii) *ϕ is continuously differentiable on $(0, \eta)$;*
(iii) *for all $s \in (0, \eta), \phi'(s) > 0$;*
(iv) *for all x in $U \cap [f(x^*) < f < f(x^*) + \eta]$, the Kurdyka-Łojasiewicz inequality holds*

$$\phi'(f(x) - f(x^*))\mathit{dist}(0, \partial f(x)) \geq 1.$$

A proper closed function f, which satisfies the Kurdyka-Łojasiewicz property at each point of dom f, is called a KŁ function.

Lemma 2.3 ([7] Uniformized KŁ property) *Suppose that f is a proper closed function and Ω is a compact subset. If f is a constant on Ω and satisfies the KŁ property at each point of Ω, then there exist $\varepsilon, \eta > 0$ and ϕ with $\phi(0) = 0$ and $\phi : [0, \eta) \to \mathbb{R}_+$ is continuously differentiable on $(0, \eta)$ with $\phi' > 0$ such that*

$$\phi'(f(x) - f(x^*))\mathit{dist}(0, \partial f(x)) \geq 1,$$

for any $x^* \in \Omega$ and any x satisfying $dist(x, \Omega) < \varepsilon$ and $f(x^*) < f(x) < f(x^*) + \eta$.

Next, we discuss real semi-algebraic sets and functions, which introduce various functions that satisfy the Kurdyka-Łojasiewicz property.

Definition 2.5 (Semi-algebraic sets and functions) (i) *A subset S of $\mathbb{R}^n$ is a real semi-algebraic set if there exists a finite number of real polynomial functions $g_{ij}, h_{ij} : \mathbb{R}^n \to \mathbb{R}$ such that*

$$S = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{u \in \mathbb{R}^n \mid g_{ij}(u) = 0 \text{ and } h_{ij}(u) < 0\}.$$

(ii) *A function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is called semi-algebraic if its graph*

$$\left\{(u, t) \in \mathbb{R}^{n+1} \mid f(u) = t\right\}$$

is a semi-algebraic subset of $\mathbb{R}^{n+1}$.

Theorem 2.1 ([6]) *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper and lower semi-continuous function. If f is semi-algebraic, it satisfies the KŁ property at any point of dom f.*

Example 2.1 ([7]) *Some examples of semi-algebraic functions are provided below.*

(1) *real polynomial functions.*
(2) *indicator functions of semi-algebraic sets.*
(3) *finite sums and product of semi-algebraic functions.*
(4) *composition of semi-algebraic functions.*
(5) *generalized inverse of semi-algebraic mappings are semi-algebraic.*
(6) sup / inf *type function, e.g.,* $\sup\{g(x, y) : y \in C\}$ *is semi-algebraic when g is a semi-algebraic function and C a semi-algebraic set.*
(7) *the function $f(x) = \mathrm{dist}(x, S)^2$ is semi-algebraic whenever S is a nonempty semi-algebraic subset of $\mathbb{R}^n$.*
(8) *$\|\cdot\|_0$ is semi-algebraic. The sparsity measure of a vector x of $\mathbb{R}^n$ is defined by*

$$\|x\|_0 := \text{ number of nonzero coordinates of } x.$$

(9) *$\|\cdot\|_p$ is semi-algebraic whenever p is rational where the given number p is positive and the p norm is defined as*

$$\|x\|_p = \left(\sum_{i=1}^{n} \|x_i\|^p\right)^{1/p}, \quad x \in \mathbb{R}^n.$$

2.3 Single-ratio fractional optimization problem

In this subsection, we summarize some basic properties of the single-ratio optimization problem (1.4). For convenience, let

$$F(x) = \frac{f(x) + h(x)}{g(x)}, \tag{2.1}$$

where $g(x) := 1 + \|x\|^2$, $h(x) := \|Ax - b\|^2$, and $f(x) := -2\langle Ax - b, |x|\rangle + \|b\|^2 + \|x\|^2 + \lambda\|x\|^2(1 + \|x\|^2)$. Note that $g(x) \geq 1$, and g is 2-strongly convex and 2-smooth. Moreover, $h(x)$ is $L-$smooth and l-strongly convex with $L \geq 2\lambda_{\max}(A^T A)$

and $l = 2\lambda_{\min}(A^TA)$, where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ represent the maximum eigenvalue and minimum eigenvalue, respectively. In fact, f is ω-weakly convex (see Proposition 1 in [17]; Theorem 2.2. in [24]). Let's assume $L \geq \omega$ without loss of generality.

First, we introduce the sum and quotient rules for the limiting subdifferential of F.

Lemma 2.4 ([40] Sum and quotient rules) *Assume that the function F takes the form of (2.1). Let $x \in domF$. Then, the following statements hold.*

(i) *Suppose that h is differentiable at x. Then, we have*

$$\partial(f+g)(x) = \partial g(x) + \nabla h(x).$$

(ii) *If h and g are differentiable at x and $g(x) \neq 0$, one has*

$$\partial F(x) = \frac{g(x)[\partial f(x) + \nabla h(x)] - [f(x) + h(x)]\nabla g(x)}{[g(x)]^2}.$$

The following lemma demonstrates the relationship between the solution of (1.4) and the solution of (1.3).

Lemma 2.5 *If x^* is an optimal solution of (1.4), then*

$$x^* = x^*, \quad E^* = -\frac{Ax^* - |x^*| - b}{1 + \|x^*\|^2} x^{*T}, \quad r^* = \frac{Ax^* - |x^*| - b}{1 + \|x^*\|^2}$$

is an optimal pair to the problem (1.3).

Proof The Lagrangian function of (1.3) is

$$\mathcal{L}(x, E, r, y) = \|E\|_F^2 + \|r\|^2 + \lambda\|x\|^2 - y^T[(A+E)x - |x| - (b+r)].$$

It follows from the Karush–Kuhn–Tucker (KKT) conditions that

$$\frac{\partial \mathcal{L}}{\partial E} = 2E - yx^T = 0, \tag{2.2}$$

$$\frac{\partial \mathcal{L}}{\partial r} = 2r + y = 0, \tag{2.3}$$

$$\frac{\partial \mathcal{L}}{\partial y} = (A+E)x - |x| - (b+r) = 0. \tag{2.4}$$

From (2.2)-(2.4), we can obtain

$$E = -rx^T \quad \text{and} \quad r = Ax + Ex - |x| - b.$$

Hence, we can conclude

$$Ex = -rx^Tx = -r\|x\|^2,$$

which indicates

$$r = \frac{Ax - |x| - b}{1 + \|x\|^2} \quad \text{and} \quad E = -\frac{Ax - |x| - b}{1 + \|x\|^2} x^T. \tag{2.5}$$

Combining the above equalities yields

$$\|E\|_F^2 + \|r\|^2 + \lambda\|x\|^2 = \frac{\|Ax - |x| - b\|^2}{1 + \|x\|^2} + \lambda\|x\|^2.$$

Since x^* is an optimal solution of the problem (1.4), and the objective function of this problem corresponds to the right-hand side of the above equation, then the optimal function value of the problem (1.3) is $\|E^*\|_F^2 + \|r^*\|^2 + \lambda\|x^*\|^2$. From (2.5), we complete the proof. ∎

The following lemma shows that F is a lower semi-continuous function.

Lemma 2.6 **([54])** *If the functions $f + h$ and g do not attain 0 simultaneously, F is a lower semi-continuous function.*

We propose the following two lemmas and one theorem to induce the definition of the critical points of F [54].

Lemma 2.7 *Let $x^* \in domF$ and $\alpha^* = F(x^*)$. Then x^* is a local minimizer of (1.4) if and only if x^* is a local minimizer of*

$$\min_{x\in\mathbb{R}^n} f(x) + h(x) - \alpha^* g(x). \tag{2.6}$$

Proof Since x^* is a local minimizer of (1.4), we can obtain that there exists $\delta > 0$ such that

$$\frac{f(x) + h(x)}{g(x)} \geq \frac{f(x^*) + h(x^*)}{g(x^*)}, \quad \forall x \in B(x^*, \delta),$$

which, together with $g(x) > 0$, implies

$$0 \leq f(x) + h(x) - F(x^*)g(x) = f(x) + h(x) - \alpha^* g(x), \quad \forall x \in B(x^*, \delta).$$

It follows from $f(x^*) + h(x^*) - \alpha^* g(x^*) = 0$ that x^* is a local minimizer of (2.6). Thus, the proof is complete. ∎

Lemma 2.8 *Let $x^* \in domF$ be a local minimizer of (2.6), and let $\alpha^* = F(x^*)$. Then, there exists $\delta > 0$ such that*

$$f(x^*) \leq f(x) + \left\langle \nabla h(x^*) - \alpha^* \nabla g(x^*), x - x^* \right\rangle + \frac{L}{2}\|x - x^*\|^2, \quad \forall x \in B(x^*, \delta).$$

Proof Since x^* is a local minimizer of (2.6), there exists $\delta > 0$ such that

$$f(x^*) + h(x^*) - \alpha^* g(x^*) \leq f(x) + h(x) - \alpha^* g(x), \quad \forall x \in B(x^*, \delta).$$

It follows from the Lipschitz continuity of ∇h and Lemma 2.1 that

$$h(x) \leq h(x^*) + \left\langle \nabla h(x^*), x - x^* \right\rangle + \frac{L}{2}\|x - x^*\|^2. \tag{2.7}$$

Due to the convexity and differentiability of g, we have

$$g(x) \geq g(x^*) + \left\langle \nabla g(x^*), x - x^* \right\rangle. \tag{2.8}$$

Combining the above three inequalities and $\alpha^* \geq 0$ leads to the conclusion. ∎

Theorem 2.2 *Let $x^* \in domF$ be a local minimizer of (1.4), and let $\alpha^* = F(x^*)$. Then, we have*

$$\alpha^* \nabla g(x^*) - \nabla h(x^*) \in \partial f(x^*).$$

Proof From Lemmas 2.7 and 2.8, we have that x^* is a local minimizer of the following problem

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) + \langle \nabla h(x^*) - \alpha^* \nabla g(x^*), x - x^* \rangle + \frac{L}{2}\|x - x^*\|^2 \right\},$$

which indicates

$$0 \in \partial f(x^*) + \nabla h(x^*) - \alpha^* \nabla g(x^*).$$

■

Hence, we can define a critical point of F as follows.

Definition 2.6 (Critical point) *Let $x^* \in domF$ and $\alpha^* = F(x^*)$. We say that x^* is a critical point of F if*

$$0 \in \partial f(x^*) + \nabla h(x^*) - \alpha^* \nabla g(x^*),$$

which coincides with

$$0 \in \partial F(x^*).$$

3 PGM and convergence analysis

In this section, we first propose a proximal gradient method for solving (1.4), then prove its global convergence and convergence rate. It is noteworthy that the global convergence can be obtained without any additional assumption. The resulting algorithm is summarized in Algorithm 1.

Algorithm 1: The proximal gradient method for solving (1.4).

Input: Given $x^0 \in \text{dom}F$ and $\alpha^0 = \frac{f(x^0)+h(x^0)}{g(x^0)}$. Given sequence $\{\eta_k\}$ satisfying $L \le \eta_k \le \overline{\eta}$. Error tolerance TOL.

1: **for** $k = 0, 1, \cdots\cdots$ **do**

2: Update x^{k+1} by solving the following subproblem:

$$\arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \left\langle \nabla h(x^k) - \alpha^k \nabla g(x^k), x - x^k \right\rangle + \frac{\eta_k}{2}\|x - x^k\|^2 \right\}. \quad (3.1)$$

3: Compute

$$\alpha^{k+1} = \frac{f(x^{k+1}) + h(x^{k+1})}{g(x^{k+1})}.$$

4: **if** $\|x^{k+1} - x^k\| < \text{TOL}$ **then**

5: STOP and RETURN x^{k+1}.

6: **end if**

7: **end for**

In this algorithm, the proximity parameter sequence $\{\eta_k\}$ can be predetermined and is independent of the iteration, which differs from Algorithm 2 proposed in Section 4.

We next establish the convergence of the subsequence, the global convergence, and the convergence rate in Sections 3.1, 3.2, and 3.3, respectively.

3.1 Convergence of the subsequence

We first show the sufficient descent of $\{F(x^k)\}$ in the following lemma, which plays a key role in the subsequent analysis.

Lemma 3.1 (Sufficient descent condition) *Let $\{x^k\}$ be the sequence generated by Algorithm 1 for solving (1.4). Then, we have*

$$F(x^{k+1}) \leq F(x^k) - \frac{2\alpha^k + \eta_k - L}{2g(x^{k+1})}\|x^{k+1} - x^k\|^2.$$

Proof From the Lipschitz continuity of ∇h and Lemma 2.1, we have

$$h(x^{k+1}) \leq h(x^k) + \left\langle \nabla h(x^k), x^{k+1} - x^k \right\rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2.$$

The iterative scheme of x^{k+1} implies that

$$f(x^{k+1}) + \left\langle \nabla h(x^k) - \alpha^k \nabla g(x^k), x^{k+1} - x^k \right\rangle + \frac{\eta_k}{2}\|x^{k+1} - x^k\|^2 \leq f(x^k).$$

Due to the strong convexity and differentiability of g, we have

$$g(x^{k+1}) \geq g(x^k) + \left\langle \nabla g(x^k), x^{k+1} - x^k \right\rangle + \|x^{k+1} - x^k\|^2.$$

It follows from the above three inequalities and $\alpha^k \geq 0$ that

$$f(x^{k+1}) + h(x^{k+1}) - \alpha^k g(x^{k+1}) + \frac{\eta_k - L}{2}\|x^{k+1} - x^k\|^2 \leq f(x^k) + h(x^k) - \alpha^k g(x^k),$$

which, together with $g(x^{k+1}) \geq 1$ and the definition of α^k, implies

$$F(x^{k+1}) \leq F(x^k) - \frac{2\alpha^k + \eta_k - L}{2g(x^{k+1})}\|x^{k+1} - x^k\|^2.$$

∎

Then, the following theorem presents the convergence of the function value and the convergence of the subsequence.

Theorem 3.1 *Let $\{x^k\}$ be the sequence generated by Algorithm 1 for solving (1.4). Then, the following statements hold:*

(i) $\lim\limits_{k\to+\infty} \alpha^k = \lim\limits_{k\to+\infty} F(x^k) = \alpha$ *where α is a nonnegative constant;*

(ii) $\lim\limits_{k\to+\infty} \frac{2\alpha^k+\eta_k-L}{2g(x^{k+1})}\|x^{k+1} - x^k\|^2 = 0$*;*

(iii) *for any cluster point x^* of $\{x^k\}$, we have $\alpha = F(x^*)$. Moreover, x^* is a critical point of (1.4).*

Proof Statements (i) and (ii) can immediately obtained by utilizing Lemma 2.2 and Lemma 3.1. It is derived from statement (i) and the coerciveness of F that $\{x^k\}$ is bounded. Suppose x^* is a cluster point of $\{x^k\}$, that is to say, there exists a subsequence $\{x^{k_i}\}$ converging to x^* as $i \to +\infty$. Due to statement (ii) and $2\alpha^k + \eta_k - L \geq 2\alpha$, we have

$$\lim_{i\to+\infty} \|x^{k_i+1} - x^{k_i}\|^2 = 0 \quad \text{and} \quad \lim_{i\to+\infty} x^{k_i+1} = x^*.$$

From the optimality condition of subproblem (3.1), we have

$$0 \in \partial f(x^{k_i+1}) + \nabla h(x^{k_i}) - \alpha^{k_i}\nabla g(x^{k_i}) + \eta_{k_i}(x^{k_i+1} - x^{k_i}).$$

Thanks to the closeness of ∂f and statement (i), we can conclude

$$0 \in \partial f(x^*) + \nabla h(x^*) - \alpha \nabla g(x^*),$$

where $\alpha = \lim_{i\to+\infty} F(x^{k_i}) = F(x^*)$. From Definition 2.6, thus the proof is complete. ∎

3.2 Global convergence

Since $Ax - |x| - b$ is piecewise linear multifunction, $\|Ax - |x| - b\|^2$ is a semi-algebraic function. Note that the sum or quotient of two semi-algebraic functions is a semi-algebraic function [8], and we can conclude that $F(x)$ is semi-algebraic from Example 2.1. With the help of Theorem 2.1, $F(x)$ satisfies the KŁ property at any $x \in \mathrm{dom} f$. Now, we are ready to establish the global convergence of Algorithm 1.

Theorem 3.2 *Let $\{x^k\}$ be the sequence generated by Algorithm 1 for solving (1.4). Then, the following statements hold:*

(i) $\sum_{k=1}^{+\infty} \|x^{k+1} - x^k\| < +\infty$;

(ii) *the whole sequence $\{x^k\}$ converges to x^* which is a critical point of (1.4).*

Proof From Theorem 3.1 (i), $F(x^k)$ converges to some constant α. We first consider the case that there exists $\hat{k} > 0$ such that $F(x^{\hat{k}}) = \alpha$. By Lemma 3.1 and $2\alpha^k + \eta_k - L \geq 2\alpha$, we can obtain $F(x^k) = \alpha$ and $x^k = x^{\hat{k}}$ for any $k > \hat{k}$. In this case, statements (i) and (ii) hold.

Next, we consider the case that $F(x^k) > \alpha$ for all $k > 0$. We denote $\{x|F(x) = \alpha\}$ and $F(x^k) - \alpha$ as Ω^* and e^k, respectively. Since F is a KŁ function, there exists a concave function ϕ satisfying the conditions in Lemma 2.3. Then, there exist $\varepsilon > 0$ and $a > 0$ such that

$$\phi'\left(F(x) - \alpha\right) \mathrm{dist}\left(0, \partial F(x)\right) \geq 1,$$

for all x in U, where $U = \{x|\mathrm{dist}(x, \Omega^*) < \varepsilon\} \cap [\alpha < F < \alpha + a]$. It follows from Theorem 3.1 that there exists $\bar{k}$ such that $x^k \in U$ for all $k > \bar{k}$. The concavity of ϕ implies

$$\phi(e^{k+1}) \leq \phi(e^k) + \phi'(e^k)\left(F(x^{k+1}) - F(x^k)\right). \tag{3.2}$$

From the optimality condition of subproblem (3.1), we have

$$0 \in \partial f(x^k) + \nabla h(x^{k-1}) - \alpha^{k-1}\nabla g(x^{k-1}) + \eta_{k-1}(x^k - x^{k-1}),$$

which indicates that

$$\omega^k := \frac{\nabla h(x^k) - \nabla h(x^{k-1})}{g(x^k)} + \frac{\alpha^{k-1}\nabla g(x^{k-1}) - \alpha^k \nabla g(x^k)}{g(x^k)} - \frac{\eta_{k-1}(x^k - x^{k-1})}{g(x^k)}$$

belongs to $\partial F(x^k)$. Since $\nabla g(x) = 2x$ and $\{x^k\}$ is bounded, we can conclude that there exists a constant $M > 0$ such that $\|x^k\| \le M/2$ and $g(x^k) \le M$ for all k, which implies

$$\|\alpha^{k-1}\nabla g(x^{k-1}) - \alpha^k \nabla g(x^k)\| \le (\alpha^{k-1} - \alpha^k)M + 2\alpha^0\|x^k - x^{k+1}\|.$$

Due to the Lipschitz continuity of ∇h, $\eta_{k-1} \le \overline{\eta}$, and $g(x^k) \ge 1$, we get

$$\|w^k\| \le (L + \overline{\eta} + 2\alpha^0)\|x^k - x^{k-1}\| + (\alpha^{k-1} - \alpha^k)M. \tag{3.3}$$

Combining Lemma 2.3, Theorem 3.1, (3.2), and (3.3), we have

$$\begin{aligned}
\phi(e^k) - \phi(e^{k+1}) &\ge \phi'(e^k)\left(F(x^k) - F(x^{k+1})\right) \\
&\ge \frac{F(x^k) - F(x^{k+1})}{\mathrm{dist}(0, \partial F(x^k))} \\
&\ge \frac{\alpha\|x^{k+1} - x^k\|^2}{M(L + \overline{\eta} + 2\alpha^0)\|x^k - x^{k-1}\| + (\alpha^{k-1} - \alpha^k)M^2}.
\end{aligned}$$

Denote $c_1 := \alpha$, $c_2 := M(L + \overline{\eta} + 2\alpha^0)$, and $c_3 := \frac{M^2}{c_2}$. Hence, it follows that

$$\begin{aligned}
&\frac{c_2}{c_1}\left[\phi(e^k) - \phi(e^{k+1})\right] + \frac{1}{4}\left[\|x^k - x^{k-1}\| + c_3(\alpha^{k-1} - \alpha^k)\right] \\
\ge\ &\frac{\|x^{k+1} - x^k\|^2}{\|x^k - x^{k-1}\| + c_3(\alpha^{k-1} - \alpha^k)} + \frac{1}{4}\left[\|x^k - x^{k-1}\| + c_3(\alpha^{k-1} - \alpha^k)\right] \\
\ge\ &\|x^{k+1} - x^k\|.
\end{aligned}$$

Summing the last inequality from 1 to N yields that

$$\sum_{k=1}^{N}\left[\frac{c_2}{c_1}\left(\phi(e^k) - \phi(e^{k+1})\right) + \frac{c_3}{4}\left(\alpha^{k-1} - \alpha^k\right)\right] + \frac{1}{4}\sum_{k=1}^{N}\|x^k - x^{k-1}\| \ge \sum_{k=1}^{N}\|x^{k+1} - x^k\|,$$

which, together with the nonnegativity of ϕ and α, indicates

$$\frac{c_2}{c_1}\phi(e^1) + \frac{c_3}{4}\alpha^0 + \frac{1}{4}\|x^1 - x^0\| \ge \frac{3}{4}\sum_{k=1}^{N}\|x^{k+1} - x^k\|.$$

Since the left side of the above inequality is finite, taking the limit as $N \to +\infty$ yields

$$\sum_{k=1}^{+\infty}\|x^{k+1} - x^k\| < +\infty.$$

which is the conclusion of (i).

Finally, we prove the statement of (ii). From Theorem 3.1, any cluster point of $\{x^k\}$ is a stationary point of (1.4). Assume that $\{x^{k_i}\}$ is a subsequence converging to x^*. Then, for any $\varepsilon > 0$, there exists $\hat{i} > 0$ such that

$$\|x^{k_{\hat{i}}} - x^*\| < \frac{\varepsilon}{2}, \quad \forall\, i > \hat{i}.$$

The statement (i) demonstrates that for any $\varepsilon > 0$, there exists $\tilde{k}$ such that

$$\sum_{k=m}^{p} \|x^{k+1} - x^k\| < \frac{\varepsilon}{2}, \quad \forall\, m, p > \tilde{k}.$$

It is derived from the above two inequalities that, $\forall\ \varepsilon > 0$, for any $k_i > s > \max\{\tilde{k}, k_{\hat{i}}\}$, we have

$$\begin{aligned} \|x^s - x^*\| &\leq \|x^s - x^{k_i}\| + \|x^{k_i} - x^*\| \\ &\leq \sum_{l=s}^{k_i} \|x^{l+1} - x^l\| + \|x^{k_i} - x^*\| \\ &< \varepsilon, \end{aligned}$$

which implies that the whole sequence $\{x^k\}$ converges to x^*. Thus, the proof is complete. ∎

Remark 3.1 *In the proof of [54], the global convergence of the entire sequence was presented under some additional assumptions, such as that F is a KŁ function, F is lower semi-continuous, f is locally Lipshitz continuous, and g is continuously differentiable with a locally Lipschitz continuous gradient. However, in our proof, we can either demonstrate that the objective function (1.4) inherently satisfies these assumptions, or we can circumvent these additional assumptions through technical proof, which also constitutes one of our theoretical contributions.*

3.3 Convergence rate of Algorithm 1

In this subsection, we consider the convergence rate of the sequence $\{x^k\}$ generated by Algorithm 1 for solving (1.4). We assume that the function ϕ in Definition 2.4 takes the form $\phi(p) = cp^{1-\theta}$ for some positive number c and some $\theta \in [0, 1)$. The following theorem shows the convergence rate of Algorithm 1.

Theorem 3.3 *Let $\{x^k\}$ be the sequence generated by Algorithm 1 for solving (1.4). Suppose that the function ϕ in Definition 2.4 takes the form $\phi(s) = cs^{1-\theta}$ for some $\theta \in [0, 1)$ and $c > 0$. Then, the following statements hold.*

(i) *If $\theta = 0$, the sequences of $\{x^k\}$ and $\{F(x^k)\}$ converge in a finite number of steps.*
(ii) *If $\theta \in (0, \frac{1}{2}]$, the sequence of $\{x^k\}$ converges linearly.*
(iii) *If $\theta \in (\frac{1}{2}, 1)$, then there exists η such that $\|x^k - x^*\| < \eta k^{-\frac{1-\theta}{2\theta-1}}$ for all large k.*

Proof The proof is omitted since a similar proof can be obtained in [3]. ∎

Remark 3.2 *It is well-known that if F is a proper lower semi-continuous semi-algebraic function, then F qualifies as a KL function with an appropriate KL exponent θ [8]. From Lemma 2.6 and Example 2.1, we can conclude that (2.1) satisfies the assumption in Theorem 3.3.*

4 PGM with extrapolation and convergence analysis

It is well known that the proximal gradient method can be largely accelerated by the extrapolation technique [12,14], especially the Nesterov acceleration [36], which has been widely used. In this section, we proposed the PGM with extrapolation for solving (1.4), an improved inversion of Algorithm 1. Below, we present the specific iterative scheme of the proposed algorithm.

Algorithm 2: PGM with extrapolation for solving (1.4).

Input: Given $x^0 = x^{-1} \in \mathrm{dom}F$, $\eta_0 = L$, and $\alpha^0 = \frac{f(x^0)+h(x^0)}{g(x^0)}$. Given sequence $\{t_k\} \subseteq [0,1)$. Error tolerance TOL and a fixed constant $\epsilon > 0$.

1: **for** $k = 0, 1, \cdots\cdots$ **do**

2: Compute

$$y^k = x^k + t_k(x^k - x^{k-1}).$$

3: Update x^{k+1} by solving the following subproblem:

$$\arg\min_{x\in\mathbb{R}^n} \left\{ f(x) + \left\langle \nabla h(y^k) - \alpha^k \nabla g(x^k), x - y^k \right\rangle + \frac{\eta_k}{2}\|x - y^k\|^2 \right\}. \tag{4.1}$$

4: Compute

$$\alpha^{k+1} = \frac{f(x^{k+1}) + h(x^{k+1})}{g(x^{k+1})}.$$

5: Choose

$$L \le \eta_{k+1} \le \frac{2\alpha^k + \eta_k - \omega}{g(x^{k+1})t_{k+1}^2} + l - \epsilon. \tag{4.2}$$

6: **if** $\|x^{k+1} - x^k\| < \mathrm{TOL}$ **then**

7: STOP and RETURN x^{k+1}.

8: **end if**

9: **end for**

It is crucial to highlight that $\nabla g(x^k)$, rather than $\nabla g(y^k)$, is utilized in the subproblem (4.1). This differs from the typical approach taken by PGM with extrapolation. It is noteworthy that the most significant difference between the two algorithms lies in terms of the proximity parameters, where $\{\eta_k\}$ in Algorithm 1 is not heavily constrained and can be predetermined. In contrast, for Algorithm 2, $\{\eta_k\}$ must be determined based on the iteration. We also should make the following assumption with respect to $\{\eta_k\}$, which will be discussed in Remark 4.1.

Assumption 4.1 *The choice of η_{k+1} in (4.2) is well-defined.*

Moreover, for convenience, we define an auxiliary function

$$E_{\eta,t}(x,y) = F(x) + \frac{\eta - l}{2} t\|x - y\|^2,$$

which is used as a potential function to characterize the descent property of the algorithm.

The following lemma presents the sufficient descent property of the new potential function.

Lemma 4.1 (Sufficient descent condition) *Let $\{x^k\}$ be the sequence generated by Algorithm 2 for solving (1.4). Then, we have*

$$E_{\eta_{k+1},t_{k+1}}(x^{k+1},x^k) \le E_{\eta_k,t_k}(x^k,x^{k-1}) - c_{k+1}\|x^{k+1}-x^k\|^2,$$

where

$$c_{k+1} := \frac{2\alpha^k+\eta_k-\omega}{2g(x^{k+1})} - \frac{\eta_{k+1}-l}{2}t_{k+1}^2.$$

Proof Due to the Lipschitz continuity of ∇h and Lemma 2.1, we get

$$h(x^{k+1}) \le h(y^k) + \left\langle \nabla h(y^k), x^{k+1}-y^k \right\rangle + \frac{L}{2}\|x^{k+1}-y^k\|^2. \tag{4.3}$$

Since the subproblem of x^{k+1} is strongly convex with coefficient $\eta_k-\omega$, we can obain

$$\begin{aligned} &f(x^{k+1}) + \left\langle \nabla h(y^k) - \alpha^k \nabla g(x^k), x^{k+1}-x^k \right\rangle + \frac{\eta_k}{2}\|x^{k+1}-y^k\|^2 \\ &\le f(x^k) + \frac{\eta_k}{2}\|x^k-y^k\|^2 - \frac{\eta_k-\omega}{2}\|x^{k+1}-x^k\|^2. \end{aligned} \tag{4.4}$$

It follows from the strong convexity and differentiability of g that

$$g(x^{k+1}) \ge g(x^k) + \left\langle \nabla g(x^k), x^{k+1}-x^k \right\rangle + \|x^{k+1}-x^k\|^2. \tag{4.5}$$

Combining (4.3), (4.4), (4.5), and $\alpha^k \ge 0$ yields that

$$\begin{aligned} &f(x^{k+1}) + h(x^{k+1}) - \alpha^k g(x^{k+1}) \\ &\le f(x^k) + h(y^k) + \left\langle \nabla h(y^k), x^k - y^k \right\rangle - \alpha^k g(x^k) \\ &\quad + \frac{\eta_k}{2}\|x^k-y^k\|^2 - \left(\alpha^k + \frac{\eta_k-\omega}{2}\right)\|x^{k+1}-x^k\|^2 - \frac{\eta_k-L}{2}\|x^{k+1}-y^k\|^2 \\ &\le f(x^k) + h(x^k) - \alpha^k g(x^k) - \frac{l}{2}\|x^k-y^k\|^2 \\ &\quad + \frac{\eta_k}{2}\|x^k-y^k\|^2 - \left(\alpha^k + \frac{\eta_k-\omega}{2}\right)\|x^{k+1}-x^k\|^2 - \frac{\eta_k-L}{2}\|x^{k+1}-y^k\|^2 \\ &\le \frac{\eta_k-l}{2}\|x^k-y^k\|^2 - \left(\alpha^k + \frac{\eta_k-\omega}{2}\right)\|x^{k+1}-x^k\|^2 - \frac{\eta_k-L}{2}\|x^{k+1}-y^k\|^2 \\ &\le \frac{\eta_k-l}{2}t_k^2\|x^k-x^{k-1}\|^2 - \left(\alpha^k + \frac{\eta_k-\omega}{2}\right)\|x^{k+1}-x^k\|^2, \end{aligned} \tag{4.6}$$

where the second inequality follows from the strong convixity of h, the third inequality is derived from the definition of y^k and the definition of α^k, and the last inequality follows from $\eta_k \ge L$. It follows from (4.6) and $g(x^{k+1}) \ge 1$ that

$$F(x^{k+1}) - F(x^k) \le \frac{\eta_k-l}{2}t_k^2\|x^k-x^{k-1}\|^2 - \frac{2\alpha^k+\eta_k-\omega}{2g(x^{k+1})}\|x^{k+1}-x^k\|^2,$$

which implies

$$\begin{aligned} &F(x^{k+1}) + \frac{\eta_{k+1}-l}{2}t_{k+1}^2\|x^{k+1}-x^k\|^2 \\ &\le F(x^k) + \frac{\eta_k-l}{2}t_k^2\|x^k-x^{k-1}\|^2 - \left(\frac{2\alpha^k+\eta_k-\omega}{2g(x^{k+1})} - \frac{\eta_{k+1}-l}{2}t_{k+1}^2\right)\|x^{k+1}-x^k\|^2. \end{aligned}$$

Thus, the proof is complete. ∎

Remark 4.1 *It is worth noting that* $c_{k+1} \geq 0$ *depends on that Assumption 4.1 holds. The well-definiteness of* η_{k+1} *hinges on the condition that* $\frac{2\alpha^k+\eta_k-\omega}{g(x^{k+1})t_{k+1}^2}+l-\epsilon \geq L$ *which can be statisfied. For example, when* $L = l$*, that is to say,* $A^T A = \beta I$ *where* β *is a positive scalar.*

Remark 4.2 *In our analysis, we adaptively choose the proximity parameter* $\{\eta_k\}$ *and make some assumption that makes this choice well-defined. Another approach can be found in [8], where they assume that when* g *is bounded, extrapolation techniques can be invoked in the proximal subgradient method for solving single-ratio fractional programming problems. However, this method is not applicable to problem (1.4) since* g *is not bounded.*

Similar to Theorem 3.1, we can conclude the following theorem and omit the proof.

Theorem 4.1 *Let* $\{x^k\}$ *be the sequence generated by Algorithm 1 for solving (1.4). Then, the following statements hold:*

(i) $\lim_{k\to+\infty} E_{\eta_k,t_k}(x^k, x^{k-1}) = \alpha$ *where* α *is a nonnegative constant;*
(ii) $\lim_{k\to+\infty} c_{k+1}\|x^{k+1} - x^k\|^2 = 0$*;*
(iii) *for any cluster point* x^* *of* $\{x^k\}$*, we have* $\alpha = F(x^*)$*. Moreover,* x^* *is a critical point of (1.4).*

The global convergence and the global convergence rate can also be obtained, which are similar to that in Theorem 3.2 and Theorem 3.3.

5 Numerical experiment

In this section, we report some numerical results to demonstrate the proposed algorithms' efficiency in solving (1.4). All experiments are performed in `MATLAB R2020b` on `macOS Monterey` on an `Apple M1` CPU with 16 GB of RAM. All details regarding the code and data are available at `https://github.com/CyXinQAQ/TikAVE`.

5.1 Solving the subproblem

The subproblems (3.1) and (4.1) involve the calculation of the proximity operator of f. However, due to the presence of the absolute value, the subproblems are difficult to compute. We use a smooth function $f_\mu(x)$ to approximate $f(x)$ where $\lim_{\mu\downarrow 0} f_\mu(x) = f(x)$ and $\lim_{\mu\downarrow 0} f_\mu(x) \in \partial f(x)$. In the sense of limits, the previously discussed results on global convergence do not exert any influence. Hence, the subproblem (3.1) can be rewritten as

$$\arg\min_{x\in\mathbb{R}^n} \left\{ f_\mu(x) + \left\langle \nabla h(x^k) - \alpha^k \nabla g(x^k), x - x^k \right\rangle + \frac{\eta_k}{2}\|x - x^k\|^2 \right\},$$

which is a smoothing optimization problem. There are many approximations for $|\cdot|$. From the numerical results in [17,34], we choose the best approximation function as follows. We utilize

$$q(s,\mu)=\begin{cases}\dfrac{s^2}{2\mu}, & \text{if } \ |s|\le\mu,\\ |s|-\dfrac{\mu}{2}, & \text{if } \ |s|>\mu,\end{cases}$$

to approximate $|s|$ and obtain

$$\frac{\partial q(s,\mu)}{\partial s}=\begin{cases}\dfrac{s}{\mu}, & \text{if } \ |s|\le\mu,\\ 1, & \text{if } \ s>\mu,\\ -1, & \text{if } \ s<\mu.\end{cases} \tag{5.1}$$

In the following subsection, we use the mapping $Q(x):\mathbb{R}^n\to\mathbb{R}^n$ to approximate $|x|$ where $Q(x)=[q(x_1,\mu),q(x_2,\mu)\cdots,q(x_n,\mu)]^T$. It follows from (5.1) that

$$\nabla Q(x)=\text{diag}\left(\frac{\partial q(x_1,\mu)}{\partial x_1},\frac{\partial q(x_2,\mu)}{\partial x_2}\cdots,\frac{\partial q(x_n,\mu)}{\partial x_n}\right).$$

Now, the resulting smooth subproblem can be solved by many gradient-like methods.

5.2 Optimal correction of absolute value equations

In the following numerical comparison, we test two algorithms we proposed for solving the generated optimal correction of AVE problems. Test random infeasible absolute value equations are generated using the result from [29].

Lemma 5.1 ([29] Nonexistence of solution for AVE) *If* $\{x|(A+I)x-b\ge 0\}=\emptyset$ *or* $\{x|(A-I)x-b\ge 0\}=\emptyset$*, then AVE has no solution.*

We mainly demonstrate the performance of the two proposed algorithms (Algorithm 1 and Algorithm 2) in solving (1.4). For Algorithm 2, we consider two strategies for choosing t_k. The first one follows Nesterov's acceleration strategy [36] suggested by [4], i.e.,

$$t_k=\frac{\theta_{k-1}-1}{\theta_k} \quad \text{with} \quad \theta_{k+1}=\frac{1+\sqrt{1+4\theta_k^2}}{2} \quad \text{and} \quad \theta_{-1}=\theta_0=1, \tag{5.2}$$

while the second one simply takes $t_k\equiv 0.9$. In our numerical experiments, we set $\text{TOL}=10^{-5}$, $\lambda=10^{-2}$, and $\mu=10^{-5}$.

We consider two infeasible AVE problems, which are randomly generated by the following MATLAB codes (Listing 1 and Listing 2).

Listing 1 Generate infeasible AVE based on Lemma 6.1

```
n = input('Enter n :');
p = inline('(abs(x) + x)/2');
x = 10 * (rand(n,1) - rand(n,1));
u = p(x);
A = null(u');
A = A * A'; A = A + eye(n);
b = 5 * rand(n,1);
```

The evolution of the function value versus CPU time and the number of iterations is presented in Fig. 1. The numerical results averaged over 10 random instances are presented in Table 1. In the tables, “Iter.” denotes the number of outer iterations, “CPU” denotes the running time in seconds evaluated by the “tic-toc” command, and “Fval.” denotes the function value of the last iterative point. From Figure 1 and Table 1, we can make the following conclusions. Algorithm 2 with both extrapolation parameters outperforms Algorithm 1 in terms of the number of outer iterations and CPU time, which indicates that the extrapolation strategy can greatly improve the performance. It is worth noting that Algorithm 2 with $t_k \equiv 0.9$ exhibits the best performance in both time efficiency and the number of iterations required.

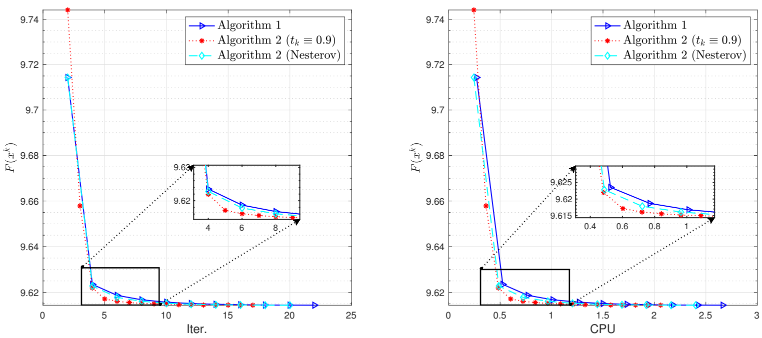

Fig. 1 Comparison results of different algorithms for solving (1.4) generated by Listing 1 on the number of iterations and running time where $n = 300$.

Table 1 Comparison results of different algorithms for solving (1.4) generated by Listing 1.

n	Algorithm 1			Algorithm 2 ($t_k \equiv 0.9$)			Algorithm 2 (Nesterov)		
	Iter.	CPU	Fval.	Iter.	CPU	Fval.	Iter.	CPU	Fval.
100	18.6	0.11	3.67e+00	**11.5**	**0.06**	3.67e+00	15.8	0.08	3.67e+00
150	10.0	0.60	4.74e+00	**13.7**	**0.40**	4.74e+00	16.5	0.48	4.48e+00
200	21.1	0.94	5.67e+00	**14.7**	**0.65**	5.67e+00	18.8	0.83	5.67e+00
250	21.3	1.18	6.33e+00	**14.7**	**0.80**	6.33e+00	18.7	1.01	6.33e+00
300	20.6	1.20	6.96e+00	**15.2**	**0.85**	6.96e+00	17.2	0.97	5.67e+00
350	20.5	1.51	7.66e+00	**15.3**	**1.14**	7.66e+00	18.3	1.35	7.66e+00
400	21.2	2.09	8.23e+00	**15.4**	**1.46**	8.23e+00	18.5	1.76	8.23e+00
450	21.1	2.23	8.86e+00	**15.6**	**1.63**	8.86e+00	18.8	1.95	8.86e+00
500	22.3	2.95	9.50e+00	**15.9**	**2.02**	9.50e+00	19.1	2.46	9.50e+00

Next, we consider another example generated by Listing 2.

Listing 2 Generate infeasible AVE based on Lemma 6.1

```
n = input('Enter n :');
u = 10 * (rand(n,1) - rand(n,1));
u = u / norm(u);
k = null(u');
k = [k,zeros(n,1)] ;
A = k + eye(n);
b = u;
```

The variation of the function value with respect to CPU time and iteration count is depicted in Figure 2. The results averaged 10 random instances are reported in Table 2. From Figure 2 and Table 2, we observe that, compared to Algorithm 1, Algorithm 2 is over 3 times faster and requires only about 1/3 the number of iterations. Another interesting observation is that, for Listing 2, Algorithm 2 with $t_k \equiv 0.9$ always performs slightly better than Algorithm 2 with (5.2). This is different from the conclusions drawn from the observations in Table 1, which demonstrates that the choice of extrapolation parameters $\{t_k\}$ is crucial for accelerating Algorithm 1.

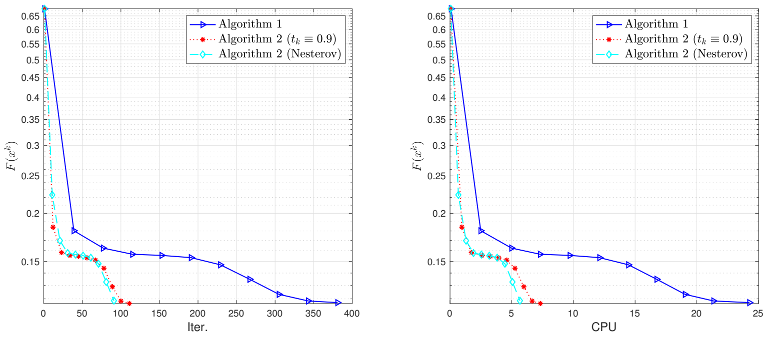

Fig. 2 Comparison results of different algorithms for solving (1.4) generated by Listing 2 on the number of iterations and running time where $n = 300$.

Table 2 Comparison results of different algorithms for solving (1.4) generated by Listing 2.

n	Algorithm 1			Algorithm 2 ($t_k \equiv 0.9$)			Algorithm 2 (Nesterov)		
	Iter.	CPU	Fval.	Iter.	CPU	Fval.	Iter.	CPU	Fval.
100	245.5	1.64	1.02e-01	87.7	0.59	9.66e-02	**77.4**	**0.51**	9.81e-02
200	263.3	11.16	1.07e-01	74.7	3.19	1.06e-01	**68.4**	**3.02**	1.09e-01
300	309.7	18.88	1.24e-01	99.5	6.01	1.10e-01	**95.9**	**5.93**	1.18e-01
400	248.2	23.14	1.21e-01	93.5	8.87	1.11e-01	**81.9**	**7.77**	1.17e-01
500	331.6	43.03	1.15e-01	104.9	13.51	1.11e-01	**93.2**	**12.02**	1.14e-01
600	316.7	64.05	1.23e-01	108.9	22.05	1.13e-01	**96.8**	**19.86**	1.18e-01
700	280.0	78.14	1.13e-01	87.7	23.85	1.16e-01	**83.0**	**23.45**	1.16e-01
800	341.6	110.92	1.11e-01	92.9	30.13	1.10e-01	**91.6**	**29.76**	1.15e-01
900	276.7	123.96	1.27e-01	101.8	44.66	1.16e-01	**97.7**	**43.70**	1.15e-01
1000	276.5	135.58	1.19e-01	91.0	44.92	1.11e-01	**90.1**	**44.04**	1.11e-01

6 Conclusion

In this paper, we mainly consider the Tikhonov regularization for infeasible absolute value equations. By reformulating the problem into a single-ratio fractional programming problem, we utilize the proximal gradient method (Algorithm 1) and the proximal gradient method with extrapolation (Algorithm 2) to solve it. We prove Algorithm 1 possesses the global convergence without any additional conditions that are indispensable in the existing work. Moreover, we also present similar results of Algorithm 2 under the condition that the proximity parameter is adaptively chosen based on the iteration. Finally, in the numerical experiment, we verify the effectiveness and practicability of the proposed algorithms for optimal correction of absolute value equations.

Declarations

Conflict of interest The authors declare that they have no conflict of interest in this work.

Data availability The datasets are generated randomly, with details and citations provided in the corresponding sections. Additionally, the MATLAB code for the numerical experiments is available from https://github.com/CyXinQAQ/TikAVE.

References

1. L. Abdallah, M. Haddou, and T. Migot, *Solving absolute value equation using complementarity and smoothing functions*, J. Comput. Appl. Math., 327 (2018), pp. 196–207.
2. H. Attouch, J. Bolte, P. Redont, et al., *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.
3. H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods*, Math. Program., 137 (2013), pp. 91–129.
4. A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imag. Sci., 2 (2009), pp. 183–202.
5. J. Y. Bello Cruz, O. P. Ferreira, and L. F. Prudente, *On the global convergence of the inexact semi-smooth Newton method for absolute value equation*, Comput. Optim. Appl., 65 (2016), pp. 93–108.

6. J. Bolte, A. Daniilidis, A. Lewis, et al., *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 18 (2007), pp. 556–572.
7. J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.
8. R. I. Boţ, M. N. Dao, and G. Y. Li, *Extrapolated proximal subgradient algorithms for nonconvex and nonsmooth fractional programs*, Math. Oper. Res., 47 (2022), pp. 2415–2443.
9. L. Caccetta, B. Qu, and G. Zhou, *A globally and quadratically convergent method for absolute value equations*, Comput. Optim. Appl., 48 (2011), pp. 45–58.
10. C. R. Chen, B. Huang, D. M. Yu, et al., *Optimal parameter of the SOR-like iteration method for solving absolute value equations*, Numerical Algorithms, 96 (2024), pp. 799–826.
11. C. R. Chen, D. M. Yu, and D. R. Han, *Exact and inexact Douglas–Rachford splitting methods for solving large-scale sparse absolute value equations*, IMA J. Numer. Anal., 43 (2023), pp. 1036–1060.
12. Y. X. Chen and D. R. Han, *Several kinds of acceleration techniques for unconstrained optimization first-order algorithms*, Math. Num. Sinica., 46 (2024), pp. 213–231.
13. W. Dinkelbach, *On nonlinear fractional programming*, Manage. Sci., 13 (1967), pp. 492–498.
14. A. d'Aspremont, D. Scieur, and A. Taylor, *Acceleration methods*, Found. Trends Optim., 5 (2021), pp. 1–245.
15. P. T. Harker and J. S. Pang, *Systems of linear interval equations*, Linear Algebra Appl., 126 (1989), pp. 39–78.
16. P. T. Harker and J. S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, Math. Program., 48 (1990), pp. 161–220.
17. F. Hashemi and S. Ketabchi, *Numerical comparisons of smoothing functions for optimal correction of an infeasible system of absolute value equations*, Numer. Algebra, Control Optim., 10 (2019), pp. 13–21.
18. F. Hashemi and S. Ketabchi, *Optimal correction of infeasible equations system as $Ax + B|x| = b$ using l_p-norm regularization*, Boletim da Sociedade Paranaense de Matemática, 40 (2019), pp. 1–16.
19. M. Hladík, *Bounds for the solutions of absolute value equations*, Comput. Optim. Appl., 69 (2018), pp. 243–266.
20. M. Hladík, *Properties of the solution set of absolute value equations and the related matrix classes*, SIAM J. Matrix Anal. Appl., 44 (2023), pp. 175–195.
21. M. Hladík, D. Moosaei, F. Hashemi, et al., *An overview of absolute value equations: From theory to solution methods and challenges*, arXiv:2404.06319v1, (2024).
22. T. Ibaraki, *Parametric approaches to fractional programs*, Math. Program., 26 (1983), pp. 345–362.
23. Y. F. Ke and C. F. Ma, *SOR-like iteration method for solving absolute value equations*, Appl. Math. Comput., 311 (2017), pp. 195–202.
24. S. Ketabchi and H. Moosaei, *An efficient method for optimal correcting of absolute value equations by minimal changes in the right hand side*, Comput. Math. Appl., 64 (2012), pp. 1882–1885.
25. S. Ketabchi and H. Moosaei, *Optimal error correction and methods of feasible directions*, J. Optim. Theory Appl., 154 (2012), pp. 209–216.
26. S. Ketabchi and H. Moosaei, *Optimal error correction of the absolute value equation using a genetic algorithm*, Math. Comput. Modell., 57 (2012), pp. 2339–2342.
27. H. Konno and M. Inori, *Bond portfolio optimization by bilinear fractional programming*, J. Oper. Res. Soc. Japan, 32 (1989), pp. 143–158.
28. O. L. Mangasarian, *Absolute value equation solution via concave minimization*, Optim. Lett., 1 (2007), pp. 3–8.
29. O. L. Mangasarian, *Absolute value programming*, Comput. Optim. Appl., 36 (2007), pp. 43–53.
30. O. L. Mangasarian, *A generalized Newton method for absolute value equations*, Optim. Lett., 3 (2009), pp. 101–108.
31. O. L. Mangasarian, *Knapsack feasibility as an absolute value equation solvable by successive linear programming*, Optim. Lett., 3 (2009), pp. 161–170.
32. O. L. Mangasarian and R. R. Meyer, *Absolute value equations*, Linear Algebra Appl., 419 (2006), pp. 359–367.

33. H. Moosaei and S. Ketabchi, *Minimum norm solution to the absolute value equation in the convex case*, J. Optim. Theory Appl., 154 (2012), pp. 1080–1087.
34. H. Moosaei and S. Ketabchi, *Optimal correcting of absolute value equations by using smoothing techniques*, J. Interdiscip. Math., 22 (2019), pp. 531–538.
35. H. Moosaei and S. Ketabchi, *Optimal correction of the absolute value equations*, J. Global Optim., 79 (2021), pp. 645–667.
36. Y. Nesterov, *A method of solving a convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 27 (1983), pp. 372–376.
37. Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, Springer Science & Business Media, 2003.
38. Y. Rahimi, C. Wang, H. Dong, et al., *A scale-invariant approach for sparse signal recovery*, SIAM J. Sci. Comput., 41 (2019), pp. 3649–3672.
39. H. Robbins and D. Siegmund, *A convergence theorem for nonnegative almost supermartingales and some applications*, in Optim. Methods Stat., Academic Press, UK, 1971, pp. 233–257.
40. R. T. Rockafellar and R. J. B. Wets, *Variational analysis*, Springer Science & Business Media, 2009.
41. J. Rohn, *A theorem of the alternatives for the equation* $Ax + B|x| = b$, Linear Multilinear Algebra, 52 (2004), pp. 421–426.
42. J. Rohn, V. Hooshyarbakhsh, and R. Farhadsefat, *An iterative method for solving absolute value equations and sufficient conditions for unique solvability*, Optim. Lett., 8 (2014), pp. 35–44.
43. B. Saheya, C. H. Yu, and J. S. Chen, *Numerical comparisons based on four smoothing functions for absolute value equation*, J. Appl. Math. Comput., 56 (2018), pp. 131–149.
44. D. K. Salkuyeh, *The Picard-HSS iteration method for absolute value equations*, Optim. Lett., 8 (2014), pp. 2191–2202.
45. S. Schaible, *Fractional programming. II, on Dinkelbach's algorithm*, Manage. Sci., 22 (1976), pp. 868–873.
46. S. Schaible, *Fractional programming*, in Handbook of Global Optimization, R. Horst and P. M. Pardalos, eds., vol. 2 of Nonconvex Optimization and Its Applications, Kluwer Academic Publishers, Dordrecht, 1995, pp. 495–608.
47. S. Schaible and T. Ibaraki, *Fractional programming*, Eur. J. Oper. Res., 12 (1983), pp. 325–338.
48. C. Wang, M. Tao, J. G. Nagy, et al., *Limited-angle CT reconstruction via the* L_1/L_2 *minimization*, SIAM J. Imag. Sci., 14 (2021), pp. 749–777.
49. S. L. Wu and S. Q. Shen, *On the unique solution of the generalized absolute value equation*, Optim. Lett., 15 (2021), pp. 2017–2024.
50. D. M. Yu, C. R. Chen, and D. R. Han, *A modified fixed point iteration method for solving the system of absolute value equations*, Optim., 71 (2022), pp. 449–461.
51. N. Zainali and T. Lotfi, *On developing a stable and quadratic convergent method for solving absolute value equation*, J. Comput. Appl. Math., 330 (2018), pp. 742–747.
52. M. Zamani and M. Hladik, *A new concave minimization algorithm for the absolute value equation solution*, Optim. Lett., 15 (2021), pp. 2241–2254.
53. C. Zhang and Q. J. Wei, *Global and finite convergence of a generalized Newton method for absolute value equations*, J. Optim. Theory Appl., 143 (2009), pp. 391–403.
54. N. Zhang and Q. Li, *First-order algorithms for a class of fractional optimization problems*, SIAM J. Optim., 32 (2022), pp. 100–129.