Journal of Convex Analysis
Volume 31 (2024), No. 1, 243–254

R-Continuity with Applications to Convergence Analysis of Tikhonov Regularization and DC Programming

Ba Khiet Le
Optimization Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam
lebakhiet@tdtu.edu.vn

Received: November 15, 2021
Accepted: May 24, 2023

We study the convergence analysis of Tikhonov regularization in finding a zero of a maximal monotone operator using the notion of *R*-continuity. Applications to convex minimization and DC programming are provided.

Keywords: Convex programming, maximal monotone operators, proximal point algorithm, regularization, DC programming.

2020 Mathematics Subject Classification: 47N10, 90C26, 90C06.

1. Introduction

Our aim is to provide a new way to find $x \in H$ satisfying

$$0 \in Ax, \tag{1}$$

where H is a given real Hilbert space and $A : H \rightrightarrows H$ is a maximal monotone operator. It is one of the fundamental problems in convex optimization which has been extensively studied (see, e. g., [1, 3, 2, 7, 8, 9, 16, 18, 19, 22, 23, 24, 26, 27, 30] and the references therein). Indeed it is known that minimizing a proper convex lower semicontinuous function $f : H \to \mathbb{R} \cup \{+\infty\}$ can be reduced to the problem

$$0 \in \partial f(x). \tag{2}$$

The proximal point algorithm [18, 19, 26] is one of well-known methods which is given by

$$x_0 \in H, \;\; x_{k+1} = J_{\gamma A} x_k, \;\; k = 1, 2, 3 \dots \tag{3}$$

for some $\gamma > 0$ where $J_{\gamma A} := (Id + \gamma A)^{-1}$ denotes the resolvent of A. Weak convergence of the proximal point algorithm follows by the firm non-expansiveness of the resolvent of A and Opial's Lemma [9, 26]. Strong convergence can be obtained by combining with the Krasnoselskii–Mann iteration scheme or Halpern's procedure (see, e.g., [27, 20, 31]). Linear convergence of the algorithm is an attractive property, which was confirmed by Rockafellar [26] when A is strongly monotone or more weakly, the inverse of A is Lipschitz continuous at zero. However, the Lipschitz continuity of the inverse of A at zero in Rockafellar's sense (see [26, 29]) is quite restrictive since it requires that the inverse of A at zero is singleton, i. e., the problem must have a unique solution, which is not satisfied in many applications.

ISSN 0944-6532 / \$ 2.50 © Heldermann Verlag

Inspired by Rockafellar's idea, in this paper we establish a new way to solve the problem under only the R-continuity of the inverse of A at zero in the sense of set-valued analysis. It is known that the unique solution x_ε of the Tikhonov regularized problem

$$0 \in (A + \varepsilon Id)x_\varepsilon, \tag{4}$$

converges to the least-norm solution of (1) as $\varepsilon > 0$ tends to zero [8, 30]. However, under the truncated R-continuity of the inverse of A at zero with its explicit continuity modulus function, we can estimate how fast is this convergence. It helps us to decide whether we should use the regularization. The assumption, which is automatically satisfied when A is a maximal monotone operator, has a closed connection with the notion of metric regularity (see, e. g., [5, 11, 12]). In addition, we show that when A is not necessarily monotone, the R-continuity of the inverse of A at zero is also important in the convergence analysis of DC programming for a large class of functions.

The paper is organized as follows. First we recall some definitions and useful results concerning operators in Section 2. The convergence analysis of Tikhonov regularization is provided in Section 3. Applications to convex and DC programmings are given in Section 4. Finally, some conclusions end the paper in Section 5.

2. Notations and preliminaries

Let H be a given real Hilbert space with its norm $\|\cdot\|$ and scalar product $\langle \cdot, \cdot \rangle$. The closed unit ball is denoted by $\mathbb{B}$. Let be given a closed, convex set $K \subset H$. The distance and the projection from a point s to K are defined respectively by

$$d(s, K) := \inf_{x \in K} \|s - x\|, \;\; \mathrm{proj}_K(s) := x \in K \;\text{ such that } d(s, K) = \|s - x\|.$$

The least-norm element of K is defined by $\mathrm{proj}_K(0)$.

The domain, the range and the graph of a set-valued mapping $\mathcal{A} : H \rightrightarrows H$ are defined respectively by

$$\mathrm{dom}(\mathcal{A}) = \{x \in H : \; \mathcal{A}(x) \neq \emptyset\}, \;\; \mathrm{rge}(\mathcal{A}) = \bigcup_{x \in H} \mathcal{A}(x)$$

and
$$\mathrm{gph}(\mathcal{A}) = \{(x, y) : x \in H, y \in \mathcal{A}(x)\}.$$

It is called *monotone* provided

$$\langle x^* - y^*, x - y \rangle \geq 0 \;\; \forall\, x, y \in H, x^* \in \mathcal{A}(x) \text{ and } y^* \in \mathcal{A}(y).$$

In addition, it is called *maximal monotone* if there is no monotone operator $\mathcal{A}'$ such that the graph of $\mathcal{A}$ is strictly contained in the graph of $\mathcal{A}'$. We have the following classical result (see, e. g., [4, 6]).

Proposition 2.1. *Let $\mathcal{A} : H \rightrightarrows H$ be a maximal monotone operator. Then the graph of $\mathcal{A}$ is strongly-weakly closed, i. e., if (x_n) converges strongly to some x, (y_n) converges weakly to some y and $y_n \in \mathcal{A}(x_n)$ then $y \in \mathcal{A}(x)$.*

Suppose that $0 \in \mathrm{dom}(\mathcal{A})$. The following concept is an extension of the Lipschitz continuity introduced by Rockafellar [26].

Definition 2.2. The set-valued operator $\mathcal{A}$ is called *truncately R-continuous* at 0 if there exist $\sigma > 0$ and a non-decreasing function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $\lim_{r\to 0^+} \rho(r) = \rho(0) = 0$ such that

$$\mathcal{A}(x) \cap a\mathbb{B} \subset \mathcal{A}(0) + \rho(\|x\|)\mathbb{B}, \forall x \in \sigma\mathbb{B}, \tag{5}$$

where $a = \inf_{y\in\mathcal{A}(0)} \|y\|$. The function ρ is called a continuity modulus function of $\mathcal{A}$ at zero. It is called truncately R-Lipschitz continuous at 0 with modulus $c > 0$ if $\rho(\|x\|) = c\|x\|$.

If the left-hand side of (5) is replaced by $\mathcal{A}(x)$ then $\mathcal{A}$ is called *R-continuous at* 0. Similarly if $\rho(\|x\|) = c\|x\|$ for some $c > 0$ then $\mathcal{A}$ is called *R-Lipschitz continuous* at 0 with modulus c. □

Remark 2.3. (i) Our Lipschitz continuity is strictly weaker than the Lipschitz continuity in the classical Rockafellar's sense [26] since it allows $\mathcal{A}(0)$ to be set-valued. the two definitions coincide if $\mathcal{A}(0)$ is singleton. The *Sign* mapping, defined by

$$Sign(x) = \begin{cases} -1 & \text{if} \quad x < 0, \\ [-1, 1] & \text{if} \quad x = 0, \\ 1 & \text{if} \quad x > 0, \end{cases}$$

is Lipschitz continuous at 0 in our sense but not Lipschitz continuous at 0 in the Rockafellar's sense.

(ii) If A^{-1} is R-Lipschitz continuous at 0 with modulus c then $(kA)^{-1}$ is R-Lipschitz continuous at 0 with modulus $\frac{c}{|k|}$ for any $k \in \mathbb{R}$.

Proposition 2.4. *If $\mathcal{A} : H \rightrightarrows H$ is maximal monotone with $0 \in \operatorname{dom}(\mathcal{A})$ then $\mathcal{A}$ is truncately R-continuous at* 0.

Proof. Let $S := \mathcal{A}(0)$. Suppose that $\mathcal{A}$ is not truncately R-continuous at 0. Let us define the function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ as follows: $\rho(0) = 0$ and if $\sigma > 0$, we let

$$\rho(\sigma) = \inf\{\delta > 0 : \mathcal{A}(x) \cap a\mathbb{B} \subset \mathcal{A}(0) + \delta\mathbb{B}, \ \ \forall x \in \sigma\mathbb{B}\}$$

where $a = \|\tilde{y}\|$ and $\tilde{y}$ is the least-norm element of S.

It is easy to see that ρ is well-defined and non-decreasing. We will prove that $\lim_{\sigma\to 0^+} \rho(\sigma) = \rho(0) = 0$. Since ρ is non-decreasing and bounded below by 0, the limit $\lim_{\sigma\to 0^+} \rho(\sigma)$ exists. Suppose that $\lim_{\sigma\to 0^+} \rho(\sigma) = \delta^* > 0$ then there exists (x_n) such that $x_n \downarrow 0$

$$\mathcal{A}(x_n) \cap a\mathbb{B} \not\subset \mathcal{A}(0) + \frac{\delta^*}{2}\mathbb{B}.$$

Therefore there exists a sequence (y_n) such that $y_n \in \mathcal{A}(x_n) \ \cap \ a\mathbb{B}$ and

$$y_n \notin \mathcal{A}(0) + \frac{\delta^*}{2}\mathbb{B}. \tag{6}$$

Let y^* be a weak limit point of (y_n). The existence of y^* is assured due to the fact that $y_n \in a\mathbb{B}$. Note that the graph of $\mathcal{A}$ is strongly-weakly closed (Proposition

2.1), we must have $y^* \in S$. On the other hand, since the norm function is weakly continuous we deduce that

$$\|y^*\| \le \liminf \|y_n\| \le \limsup \|y_n\| \le a.$$

Thus $y^* = \tilde{y}$ and $\lim_{n\to+\infty} \|y_n\| = a = \|\tilde{y}\|$. Since y^* is an arbitrarily weak limit point of (y_n), this implies that y_n converges weakly to $\tilde{y}$ with $\lim_{n\to+\infty} \|y_n\| = \|\tilde{y}\|$. Consequently (y_n) convergs to $\tilde{y}$. Combining with (6), we obtain a contradiction and the conclusion follows. □

Remark 2.5. We provide not only the truncated R-continuity of a maximal monotone operator but also an associated modulus function ρ. □

Using similar arguments, one has the following result.

Proposition 2.6. *If $\mathcal{A} : H \rightrightarrows H$ has a closed graph with compact range and if $0 \in \mathrm{dom}(\mathcal{A})$ then $\mathcal{A}$ is R-continuous at* 0.

Proof. Suppose that $\mathcal{A}$ is not R-continuous at 0. We define the function ρ as follows: $\rho(0) = 0$ and if $\sigma > 0$, we let

$$\rho(\sigma) = \inf\{\delta > 0 : \mathcal{A}(x) \subset \mathcal{A}(0) + \delta\mathbb{B},\ \ \forall x \in \sigma\mathbb{B}\}.$$

Then we must have $\lim_{\sigma\to 0^+} \rho(\sigma) = \delta^* > 0$ and similarly as in the proof of Proposition 2.4, there exist two sequences (x_n), (y_n) such that $x_n \to 0$, $y_n \in \mathcal{A}(x_n)$ and

$$y_n \notin \mathcal{A}(0) + \frac{\delta^*}{2}\mathbb{B}. \tag{7}$$

Since $\mathcal{A}$ has compact range, there exists a convergent subsequence (y_{n_k}). Let $y^* = \lim_{k\to+\infty} y_{n_k}$. Then $y^* \in \mathcal{A}(0)$ because $\mathcal{A}$ has closed graph. However it is a contradiction to (7) and the conclusion follows. □

Remark 2.7. Note that $A^{-1} := \mathcal{A}$ has closed graph with compact range if and only if A has closed graph with compact domain. For example, $A = f + N_K$ where f is a continuous function and K is a convex compact set. This is the case of minimizing C^1 functions on convex compact sets, as we can see in the Corollary 4.7 when we deal with DC programming. □

A linear mapping $B : H \to H$ is called *positive semidefinite* if

$$\langle Bx, x\rangle \ge 0,\ \ \forall\, x \in H.$$

The next statement is classical (see, e. g., [28]). However we provide here an explicit constant in the estimation.

Lemma 2.8. *Let $B \in \mathbb{R}^{n\times n}$ be a symmetric positive semidefinite matrix. Then for all $x \in \mathrm{rge}(B)$, we have $\langle Bx, x\rangle \ge k\|x\|^2$, and thus*

$$\|Bx\| \ge k\|x\|, \tag{8}$$

where k is the least positive eigenvalue of B.

Proof. Since B is symmetric positive semidefinite matrix, we can write $B = O^T DO$ where D is a diagonal matrix with nonnegative eigenvalues and O is an orthogonal matrix, i. e., $OO^T = Id$. If $x \in \mathrm{rge}(B)$ then $x = By = O^T DOy$ for some $y \in \mathbb{R}^n$ and $Ox = OO^T DOy = DOy \in \mathrm{rge}(D)$. Thus

$$\langle Bx, x\rangle = \langle O^T DOx, x\rangle = \langle DOx, Ox\rangle \geq k\|Ox\|^2 = k\|x\|^2,$$

where k is the least positive eigenvalue of D and also of B. □ □

The following result is a consequence of Lemma 2.8.

Lemma 2.9. *Let $B \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then* $\ker(B) = \ker(B^2)$*,* $\mathrm{rge}(B^2) \subset \mathrm{rge}(B)$ *and for all $x \in \mathrm{rge}(B^2)$, we have*

$$\|Bx\| \geq \frac{k}{\|B\|}\|x\|, \tag{9}$$

where k is the least positive eigenvalue of B^2 and $\ker$*(B) denotes the kernel of B.*

Proof. It is easy to see that $\ker(B) = \ker(B^2)$, $\mathrm{rge}(B^2) \subset \mathrm{rge}(B)$ and B^2 is a symmetric positive semidefinite matrix. Using Lemma 2.8, for all $x \in \mathrm{rge}(B^2)$ one has

$$k\|x\|^2 \leq \langle B^2 x, x\rangle \leq \|B\|\|Bx\|\|x\|, \tag{10}$$

and the conclusion follows. □

The following propositions provides some cases such that A^{-1} is R-Lipschitz continuous at zero (see also [12] for the Hoffman constant). They can be used to estimate the error for the linear system $Bx = C$ when B is not an invertible matrix.

Proposition 2.10. *Let $A(x) := Bx - C, \forall x \in \mathbb{R}^n$ where $B \in \mathbb{R}^{n\times n}$ is a symmetric, positive semi-definite matrix, C is some vector in $\mathbb{R}^n$. Suppose that $S := A^{-1}(0) \neq \emptyset$. Then A^{-1} is R-Lipschitz continuous at zero with modulus $\frac{1}{k}$, where k is the least positive eigenvalue of B.*

Proof. Since $S \neq \emptyset$, one has $S = \ker(B) + x_r$ where x_r is the unique element of $\mathrm{rge}(B)$ such that $Bx_r = C$. Let $y \in \mathbb{R}^n$, if $A^{-1}(y)$ is empty then the conclusion holds trivially. If $A^{-1}(y)$ is nonempty, i. e., $y \in \mathrm{rge}(B)$ then $A^{-1}(y) = \ker(B) + y_r$ where y_r is the unique element of $\mathrm{rge}(B)$ such that $By_r = C + y$. Since

$$\|y_r - x_r\| \leq \tfrac{1}{k}\|B(y_r - x_r)\| = \tfrac{1}{k}\|y\|$$

(see Lemma 2.8), we conclude that A^{-1} is R-Lipschitz continuous at zero with modulus $\frac{1}{k}$. □

Proposition 2.11. *Let $A(x) := Bx - C, \forall x \in \mathbb{R}^n$ where $B \in \mathbb{R}^{n\times n}$ is a symmetric matrix and C is some vector in $\mathbb{R}^n$. Suppose that $S := A^{-1}(0) \neq \emptyset$. Then A^{-1} is R-Lipschitz continuous at zero with modulus $\frac{\|B\|}{k}$, where k is the least positive eigenvalue of B^2.*

Proof. Let $x_1, x_2 \in S$ then $x_1 - x_2 \in \ker(B) = \ker(B^2)$, i. e., x_1 and x_2 has the same projection onto $\mathrm{rge}(B^2)$, which is called x_r and one has $Bx_r = C$. Thus $S = x_r + \ker(B)$.

Next we prove that A^{-1} is R-Lipschitz continuous at zero. Let $y \in \mathbb{R}^n$, if $A^{-1}(y)$ is empty then the conclusion is trivial. If $A^{-1}(y)$ is nonempty, i. e., $y \in \mathrm{rge}(B)$ then $A^{-1}(y) = \ker(B) + y_r$ where y_r is the unique element of $\mathrm{rge}(B^2)$ such that $By_r = C + y$. Since $\|y_r - x_r\| \leq \frac{\|B\|}{k}\|B(y_r - x_r)\| = \frac{\|B\|}{k}\|y\|$ (see Lemma 2.9), the conclusion follows. □

In the sequel, we recall a general class of convex continuous differentiable functions with quadratic functional growth, that allows us to obtain linear convergence rates using first order methods [21]. We show that this class is smaller than our proposed class.

Definition 2.12. A continuously differentiable function f has a *quadratic functional growth* on H if there exists a constant $\kappa_f > 0$ such that for any $x \in H$ and $x^* = \mathrm{proj}_S(x)$, we have

$$f(x) - f^* \geq \frac{\kappa_f}{2}\|x - x^*\|^2, \tag{11}$$

where $S := \mathrm{argmin}_H f$ and $f^* = f(x^*)$.

Proposition 2.13. *If the convex continuous differentiable function $f : H \to \mathbb{R}$ has a quadratic functional growth on H, then $(\nabla f)^{-1}$ is R-Lipschitz continuous at zero.*

Proof. We prove that for given $y \in H$, we have

$$(\nabla f)^{-1}(y) \subset (\nabla f)^{-1}(0) + \frac{2}{\kappa_f}\|y\|\mathbb{B} = S + \frac{2}{\kappa_f}\|y\|\mathbb{B},$$

where k_f is defined in (11). Indeed, let $x \in (\nabla f)^{-1}(y)$ and $x^* = \mathrm{proj}_S(x)$. Using (11) and the convexity of f, one has

$$\frac{\kappa_f}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \langle \nabla f(x), x - x^* \rangle \leq \|y\|\|x - x^*\|$$

and the conclusion follows. □

3. Main result

In this section, we study the convergence analysis of Tikhonov regularization in solving (1). Traditionally, we only know that the solution x_ε of the regularized problem converges to the least-norm solution of (1). However, can we estimate how fast is this convergence?

When A is a maximal monotone operator with $0 \in \mathrm{dom}(A^{-1})$ then A^{-1} is truncately R-continuous at 0 with some continuity modulus function ρ (Proposition 2.4). If the function ρ can be estimated then we can answer the question above.

Since A^{-1} is also maximal monotone operator, $S := A^{-1}(0)$ is a closed convex set [6] and thus $\tilde{x} := \mathrm{proj}_S(0)$ is well-defined. We have the following results.

Theorem 3.1. *Suppose that $A : H \rightrightarrows H$ is a maximal monotone operator, $0 \in \mathrm{rge}(A)$ and A^{-1} is truncately R-continuous at 0 with modulus function ρ. Let $\varepsilon > 0$ small enough (e. g., $\varepsilon \leq \frac{\sigma}{\|a\|}$) and $x_\varepsilon := (A + \varepsilon Id)^{-1}(0)$, then*

$$\|x_\varepsilon\| \leq \|\tilde{x}\|, \tag{12}$$

and

$$d(x_\varepsilon, S) \leq \rho(\varepsilon\|\tilde{x}\|) = \rho(\varepsilon a) \tag{13}$$

where $\sigma > 0$, $\rho(\cdot)$ are defined in (5), $a := \|\tilde{x}\|$ where $\tilde{x}$ is the least norm element of S. In particular the sequence $(\|x_\varepsilon\|)$ is increasing as ε decreases. Furthermore

$$\lim_{\varepsilon \to 0} x_\varepsilon = \tilde{x} \tag{14}$$

with the rate

$$\|x_\varepsilon - \tilde{x}\| \le \rho(\varepsilon a) + \sqrt{\rho^2(\varepsilon a) + 2\rho(\varepsilon a)a}. \tag{15}$$

Proof. Note that x_ε is well-defined since $A + \varepsilon Id$ is strongly monotone. We have

$$0 \in A\tilde{x} \;\text{ and }\; -\varepsilon x_\varepsilon \in Ax_\epsilon.$$

The monotonicity of A implies that

$$\varepsilon \langle x_\varepsilon, x_\varepsilon - \tilde{x} \rangle \le 0, \tag{16}$$

and (12) follows. For $\delta > \varepsilon > 0$, replacing A by $A + \varepsilon I$ in the procedure above, we have $\|x_\delta\| \le \|x_\varepsilon\|$. Thus the sequence $(\|x_\varepsilon\|)$ is increasing as ε decreases and bounded above by a. In addition, we have

$$x_\varepsilon \in A^{-1}(-\varepsilon x_\varepsilon) \subset A^{-1}(0) + \rho(\varepsilon\|x_\varepsilon\|)\mathbb{B} \subset S + \rho(\varepsilon a)\mathbb{B}. \tag{17}$$

Thus

$$d(x_\varepsilon, S) \le \rho(\varepsilon a).$$

Let $y_\varepsilon := \mathrm{proj}_S(x_\varepsilon)$ then $\|x_\varepsilon - y_\varepsilon\| \le \rho(\varepsilon a)$. We have

$$\begin{aligned} \|y_\varepsilon - \tilde{x}\|^2 &\le \|y_\varepsilon\|^2 - \|\tilde{x}\|^2 \;\; (\text{since } \tilde{x} := \mathrm{proj}_S(0) \text{ and } y_\varepsilon \in S) \\ &\le (\|x_\varepsilon\| + \|x_\varepsilon - y_\varepsilon\|)^2 - a^2 \le \rho^2(\varepsilon a) + 2\rho(\varepsilon a)a. \end{aligned}$$

Thus $\|x_\varepsilon - \tilde{x}\| \le \|x_\varepsilon - y_\varepsilon\| + \|y_\varepsilon - \tilde{x}\| \le \rho(\varepsilon a) + \sqrt{\rho^2(\varepsilon a) + 2\rho(\varepsilon a)a}.$ □

Remark 3.2. The continuity modulus function ρ of A^{-1} at zero plays an essential role in the convergence analysis. The notion of R-continuity is also important in DC programming when A is not necessarily monotone as we see in the next section. □

4. Application

4.1. Convex optimization

Let us consider the minimization problem of a proper lower semicontinuous convex function $f : H \to \mathbb{R} \cup \{+\infty\}$. Suppose that the solution set $S := \partial f^{-1}(0)$ is nonempty and the minimum value $f^* := f(x^*)$ for some $x^* \in S$. Let $g = f + \frac{\varepsilon}{2}\|\cdot\|^2$ for some small $\varepsilon > 0$. Then g is strongly convex and has a unique minimizer x_ε which converges to the least-norm solution $\tilde{x} \in S$ of the original problem as $\varepsilon \to 0$ [30]. Obviously solving the strongly convex problem is easier and faster than solving the convex problem. However, it is important to know the estimations of $d(x_\varepsilon, S), \|x_\varepsilon - \tilde{x}\|$ and $f(x_\varepsilon) - f^*$ to decide for which function f, it is really better to use Tikhonov regularization and how to choose a suitable ε which should not be too small.

Theorem 4.1. *Suppose that ∂f^{-1} is truncately R-continuous at zero with modulus function ρ. For $\varepsilon > 0$ small enough, let $x_\varepsilon := (\partial f + \varepsilon I)^{-1}(0) = \partial g^{-1}(0)$. Then we have*

$$f(x_\varepsilon) - f^* \le \|\tilde{x}\|\rho(\varepsilon\|\tilde{x}\|)\varepsilon. \tag{18}$$

Proof. Using Theorem 3.1, there exists some $x^* \in S$ such that

$$\|x_\varepsilon - x^*\| \leq \rho(\varepsilon\|\tilde{x}\|).$$

Since x_ε is the unique minimizer of $g = f + \frac{\varepsilon}{2}\|\cdot\|^2$, we deduce that

$$f^* + \frac{\varepsilon}{2}\|\tilde{x}\|^2 = f(\tilde{x}) + \frac{\varepsilon}{2}\|\tilde{x}\|^2 \geq f(x_\varepsilon) + \frac{\varepsilon}{2}\|x_\varepsilon\|^2.$$

Therefore

$$\begin{aligned} f(x_\varepsilon) - f^* &\leq \frac{\varepsilon}{2}\|\tilde{x}\|^2 - \frac{\varepsilon}{2}\|x_\varepsilon\|^2 \leq \varepsilon\|\tilde{x}\|(\|\tilde{x}\| - \|x_\varepsilon\|) \\ &\leq \varepsilon\|\tilde{x}\|(\|x^*\| - \|x_\varepsilon\|) \leq \varepsilon\|\tilde{x}\|\|x^* - x_\varepsilon\| \leq \|\tilde{x}\|\rho(\varepsilon\|\tilde{x}\|)\varepsilon. \end{aligned}$$ □

The following result follows by Theorem 3.1 and 4.1.

Corollary 4.2. *Suppose that ∂f^{-1} is truncately R-Lipschitz at zero with modulus $c > 0$. For $\varepsilon > 0$ small enough, let $x_\varepsilon := (\partial f + \varepsilon I)^{-1}(0) = \partial g^{-1}(0)$. Then we have*

$$\|x_\varepsilon - x^*\| \leq c\|\tilde{x}\|\varepsilon, \quad \|x_\varepsilon - \tilde{x}\| \leq \|\tilde{x}\|(c\varepsilon + \sqrt{c^2\varepsilon^2 + 2c\varepsilon}) \tag{19}$$

for some $x^* \in S$ and

$$f(x_\varepsilon) - f^* \leq c\|\tilde{x}\|^2\varepsilon^2. \tag{20}$$

Example 4.3. Let us consider the minimization problem of a quadratic function $f(x) = \frac{1}{2}\langle Bx, x\rangle - Cx$, where $B \in \mathbb{R}^{n\times n}$ is a symmetric, positive semi-definite matrix, C is some vector in $\mathbb{R}^n$. The solution set is given by

$$S = \{x \in \mathbb{R}^n : Bx - C = 0\}.$$

Suppose that $S \neq \emptyset$. Let $A(x) := Bx - C, \forall x \in \mathbb{R}^n$ and $x_\varepsilon = (B + \varepsilon Id)^{-1}C$. Then A^{-1} is R-Lipschitz continuous at zero with modulus $\frac{1}{k}$, where k is the least positive eigenvalue of B (Proposition 2.10). Thus (x_ε) converges to the least norm element $\tilde{x}$ of S and

$$d(x_\varepsilon, S) \leq \frac{\|\tilde{x}\|\varepsilon}{k}, \quad f(x_\varepsilon) - f^* \leq \frac{\|\tilde{x}\|^2\varepsilon^2}{k}.$$

More specifically, we consider in $\mathbb{R}^3$ with

$$B = \begin{pmatrix} 22 & 46 & 68 \\ 46 & 97 & 143 \\ 68 & 143 & 211 \end{pmatrix}, \quad C = (318 \ \ 669 \ \ 987)^T.$$

Then B is singular and 330-Lipschitz continuous. Let $\varepsilon = 10^{-5}$ we can compute directly that

$$x_\varepsilon = (B + \varepsilon I)^{-1}(C) \approx (1\ 2\ 3)^T + 10^{-5}(-0.2218 \quad 0.167 \quad -0.0557)^T.$$

It is easy to check that $\tilde{x} = (1\ 2\ 3)^T$ is the least-norm solution of S and

$$f(x_\varepsilon) - f^* = f(x_\varepsilon) - f(\tilde{x}) \approx 2.7285 \times 10^{-12}.$$

If we use Nesterov's Accelerated Gradient method [22], with the initial point $y_0 = (5\ 2\ 3)^T$ after $k = 10^4$ iterations we can find $x_k \approx (2.3334 \ \ 3.3333\ 1.6667)^T$ with

$$f(x_k) - f^* \approx 9.9135 \times 10^{-10}.$$

Example 4.4. Next we want to minimize a L-smooth convex function $f : H \to \mathbb{R}$, i. e., f is convex differentiable and f' is L-Lipschitz continuous. Nesterov's method [22] can construct a sequence (x_k) such that

$$f(x_k) - f^* \leq \frac{2L\|x_0 - x^*\|^2}{(k+2)^2}, \tag{21}$$

where x^* is a minimizer of f and $f^* = f(x^*)$. Let $g = f + \frac{\varepsilon}{2}\|\cdot\|^2$ for some small $\varepsilon > 0$ then $g' = f' + \varepsilon I$ is ε-strongly monotone and $(L+\varepsilon)$-Lipschitz continuous. Another version of Nesterov's method for strongly convex and smooth functions [23, p. 94] allows us to construct a sequence (y_k) such that

$$\frac{\varepsilon}{2}\|y_k - x_\varepsilon\|^2 \leq g(y_k) - g(x_\varepsilon) \leq \frac{L+2\varepsilon}{2}\|y_0 - x_\varepsilon\|^2 e^{-k\sqrt{\varepsilon/(L+\varepsilon)}} =: S_k, \tag{22}$$

where $x_\varepsilon = (f' + \varepsilon I)^{-1}(0)$. Therefore, using Theorem 2, we have

$$\begin{aligned} f(y_k) - f^* &= g(y_k) - g(x_\varepsilon) - \frac{\varepsilon}{2}(\|y_k\|^2 - \|x_\varepsilon\|^2) + f(x_\varepsilon) - f^* \\ &\leq g(y_k) - g(x_\varepsilon) + \frac{\varepsilon}{2}(\|y_k - x_\varepsilon\|^2) + \varepsilon\|\tilde{x}\|\|y_k - x_\varepsilon\| + \varepsilon\|\tilde{x}\|\rho(\varepsilon\|\tilde{x}\|) \\ &\leq W_k + R(\varepsilon), \end{aligned}$$

where

$$W_k := g(y_k) - g(x_\varepsilon) + \frac{\varepsilon}{2}(\|y_k - x_\varepsilon\|^2) + \varepsilon\|\tilde{x}\|\|y_k - x_\varepsilon\| \text{ and } R(\varepsilon) := \varepsilon\|\tilde{x}\|\rho(\varepsilon\|\tilde{x}\|).$$

While W_k tends to zero very fast, the remain term $R(\varepsilon)$ does not depends on k. Thus it is easy to see that if $\rho(s) = cs^\alpha$ for some $\alpha > 0$ close to zero and $c > 0$, it is better to apply directly the Nesterov's method without using the regularization. However, for big problems when L is very large (e.g., $L \geq 10^{10}$) if $\rho(s) = cs^\alpha$ with $\alpha \geq 1$ (e. g., in large scale quadratic programming with $\alpha = 1$) and an acceptable ε (e.g., $\varepsilon = 10^{-5}$), using the regularization can reduce remarkably the iterations. Note that L is dominated by the exponential term in (22) and $\|x_\varepsilon\| \leq \|\tilde{x}\| \leq \|x^*\|$.

4.2. DC programming

DC (difference of convex) optimization has the following form

$$\min_{x\in H} f(x) = \min_{x\in H}(g(x) - h(x)), \tag{23}$$

where $g, h : H \to \mathbb{R}\cup\{+\infty\}$ are proper, convex, lower semicontinuous function. It is an important class of non-convex optimization and covers wide range of applications (see, e. g., [10, 13, 14, 17, 25] and the references therein). One usually wants to find a critical point x^* of f, i. e., $0 \in \partial^F f(x^*) \subset \partial g(x^*) - \partial h(x^*)$ whenever h is continuous at x^*, where ∂^F denotes the Fréchet subdifferential and ∂ denotes the subdifferential in the sense of convex analysis. Suppose that h is differentiable. Then $\partial^F f(x^*) = \partial g(x^*) - \nabla h(x^*)$ (see, e. g., [14]). Applying the *forward-backward* splitting algorithm [24], we can generate a sequence (x_k) as follows

$$x_0 \in H, x_{k+1} = J_{\gamma\partial g} L_{-\gamma\nabla h}(x_k), \tag{24}$$

where $J_{\gamma\partial g}$ is the resolvent of $\gamma\partial g$ and $L_{-\gamma\nabla h} := Id + \gamma\nabla h$.

Theorem 4.5. *Suppose that f is bounded below, ∇h is single-valued uniformly continuous, $S := (\partial g - \nabla h)^{-1}(0) \neq \emptyset$ and $(\partial g - \nabla h)^{-1}$ is R-continuous at zero with its modulus function ρ. Let (x_k) be the sequence generated by the forward-backward splitting algorithm* (24)*. Then* $\lim_{k\to+\infty} \mathrm{d}(x_k, S) = 0$.

Proof. We have

$$x_{k+1} = (I + \gamma\partial g)^{-1}(x_k + \gamma\nabla h(x_k)) \;\Leftrightarrow\; x_k + \gamma\nabla h(x_k) \in x_{k+1} + \gamma\partial g(x_{k+1})$$
$$\Leftrightarrow\; \nabla h(x_k) - \frac{x_{k+1} - x_k}{\gamma} \in \partial g(x_{k+1}). \tag{25}$$

On the other hand, since g and h are convex, we have

$$\langle \nabla h(x_k), x_{k+1} - x_k \rangle \leq h(x_{k+1}) - h(x_k) \tag{26}$$

and
$$\langle \nabla h(x_k) - \frac{x_{k+1} - x_k}{\gamma}, x_{k+1} - x_k \rangle \geq g(x_{k+1}) - g(x_k). \tag{27}$$

From (26) and (27), one has

$$\gamma(f(x_{k+1}) - f(x_k)) \leq -\|x_{k+1} - x_k\|^2. \tag{28}$$

Consequently, $\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\|^2 \leq \gamma(f(x_0) - f^*) < +\infty$, where $f^* := \inf_{x\in H} f(x)$. In particular

$$\lim_{k\to+\infty} \|x_{k+1} - x_k\| = 0. \tag{29}$$

From (25), one has

$$\nabla h(x_k) - \nabla h(x_{k+1}) - \frac{x_{k+1} - x_k}{\gamma} \in \partial g(x_{k+1}) - \nabla h(x_{k+1}).$$

Let $r_k := \nabla h(x_k) - \nabla h(x_{k+1}) - \frac{x_{k+1} - x_k}{\gamma}$. Then $r_k \to 0$ as $k \to +\infty$ and

$$x_{k+1} \in (\partial g - \nabla h)^{-1}(r_k) \subset (\partial g - \nabla h)^{-1}(0) + \rho(\|r_k\|).$$

Thus
$$\lim_{k\to+\infty} d(x_{k+1}, S) = 0. \qquad \square$$

Remark 4.6. (i) We recall the well-known DCA (see, e. g. [25]) as follows

$$y_k \in \partial h(x_k),\; x_{k+1} \in \partial g^*(y_k) = \operatorname{argmin}\{g(x) - [h(x_k) + \langle y_k, x - x_k \rangle]\}.$$

It is easy to see that DCA applied to the new couple $(\gamma g + \frac{1}{2}\|\cdot\|^2, \gamma h + \frac{1}{2}\|\cdot\|^2)$ is the *forward-backward* splitting algorithm. We can see that the form of *forward-backward* splitting algorithm and its convergence analysis are simpler since we don't need the estimations through the conjugate functions.

(ii) Theorem 4.5 provides a new condition such that $d(x_k, S) \to 0$ as $k \to +\infty$. For example, if $g = g_1 + I_K$ where g_1 is a differentiable function and K is a convex compact set (particularly in quadratic programming over convex compact sets) then $(\partial g - \nabla h)^{-1}$ is R-continuous at zero.

Corollary 4.7. *Suppose that f is bounded below, ∇h is single-valued uniformly continuous, $S := (\partial g - \nabla h)^{-1}(0) \neq \emptyset$, $g = g_1 + I_K$ where $g_1 : H \to \mathbb{R}$ is a convex differentiable function with full domain and K is a convex compact set. Let (x_k) be the sequence generated by the forward-backward splitting algorithm* (24). *Then*

$$\lim_{k \to +\infty} \mathrm{d}(x_k, S) = 0.$$

Proof. It remains to prove that $\mathcal{A} := (\partial g - \nabla h)^{-1}$ is R-continuous at zero. It is easy to see $\mathcal{A}$ has closed graph and compact range. Using Proposition 2.6, we obtain the conclusion. $\square$

5. Conclusion

We believe that Tikhonov regularization has been widely used in convex optimization. The paper provides a rigorous convergence analysis with explicit estimations using the notion of R-continuity. Consequently, it may help us to decide when we should use Tikhonov regularization instead of the direct method. The R-continuity of a set-valued mapping also play an important role in DC programming.

Acknowledgements. The author would like to show his gratitude to the handling editor and the reviewer for their careful reading and raising many interesting questions, which helped us to improve the paper significantly.

References

[1] F. Alvarez, H. Attouch: *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis 9 (2001) 3–11.

[2] H. Attouch, A. Cabot: *Convergence of a relaxed inertial proximal algorithm for maximally monotone operators*, Math. Programming A 184/1-2 (2020) 243–287.

[3] H. Attouch, J. Peypouquet: *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators*, Math. Programming 174 (2019) 391–432.

[4] J. P. Aubin, A. Cellina: *Differential Inclusions*, Grundlehren der Mathematischen Wissenschaften Vol. 264, Springer, Berlin (1984).

[5] P. Bosch, A. Jourani, R. Henrion: *Sufficient conditions for error bounds and applications*, Appl. Math. Optim. 50 (2004) 161–181.

[6] H. Brezis: *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland Mathematics Studies 5, North-Holland, Amsterdam (1973).

[7] G. H. G. Chen, R. T. Rockafellar: *Convergence rates in forward-backward splitting*, SIAM J. Optimization 7/2 (1997) 421–444.

[8] R. Cominetti, J. Peypouquet, S. Sorin: *Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization*, J. Diff. Equations 245/12 (2008) 3753–3763.

[9] J. Eckstein: *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, PhD Thesis, Mass. Inst. Technology, Boston (1989).

[10] W. Hare, C. Sagastizábal, M. Solodov: *A proximal bundle method for nonsmooth nonconvex functions with inexact information*, Comp. Opt. Appl. 63 (2016) 1–28.

[11] R. Henrion, J. V. Outrata: *Calmness of constraint systems with applications*, Math. Programming, Ser. B 104 (2005) 437–464.

[12] A. J. Hoffman: *On approximate solutions of systems of linear inequalities*, J. Res. Nat. Bureau Standards 49/4 (1952) 263–265.

[13] R. Horst, V. T. Nguyen: *DC programming: Overview*, J. Optim. Theory Appl. 103 (1999) 1–43.

[14] T. H. A. Le, V. N. Huynh, D. T. Pham: *Convergence analysis of difference-of-convex algorithm with subanalytic data*, J. Optim. Theory Appl. 179 (2018) 103–126.

[15] R. I. Leine, N. Wouw: *Stability and Convergence of Mechanical Systems with Unilateral Constraints*, Springer, Berlin (2008).

[16] P. L. Lions, B. Mercier: *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Analysis 16 (1979) 964–979.

[17] P. E. Maingé, A. Moudafi: *Convergence of new inertial proximal methods for DC programming*, SIAM J Optim. 19/1 (2008) 397–413.

[18] B. Martinet: *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Francaise Autom. Inf. Rech. Oper. 3 (1970) 154–158.

[19] B. Martinet: *Determination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox*, C. R. Acad. Sci. Paris A 274 (1972) 163–165.

[20] K. Nakajo, W. Takahashi: *Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups*, J. Math. Analysis Appl. 279 (2003) 372–379.

[21] I. Necoara, Y. Nesterov, F. Glineur: *Linear convergence of first order methods for non-strongly convex optimization*, Math. Programming 175 (2019) 69–107.

[22] Y. Nesterov: *A method of solving a convex programming problem with convergence rate O(1/k2)*, Sov. Mat. Doklady 27/2 (1983) 372–376.

[23] Y. Nesterov: *Introductory Lectures on Convex Optimization. A Basic Course*, Applied Optimization Vol. 87, Kluwer Academic Publishers, Boston (2004).

[24] G. B. Passty: *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Analysis Appl. 72/2 (1979) 383–390.

[25] D. T. Pham, L. T. H. Le: *Convex analysis approach to D.C. programming: Theory, algorithms and applications*, Acta Math. Vietnamica 22 (1997) 289–355.

[26] R. T. Rockafellar: *Monotone operators and the proximal point algorithm*, SIAM J. Control Optimization 14/5 (1976) 877–898.

[27] M. V. Solodov, B. F. Svaiter: *Forcing strong convergence of proximal point iterations in a Hilbert space*, Math. Programming A 87 (2000) 189–202.

[28] A. Tanwani, B. Brogliato, C. Prieur: *Well-posedness and output regulation for implicit time-varying evolution variational inequalities*, SIAM J. Control Optimization 56/2 (2018) 751–781.

[29] M. Tao, X. Yuan: *On the optimal linear convergence rate of a generalized proximal point algorithm*, J. Sci. Comput. 74 (2018) 826–850.

[30] A. Tikhonov, V. Arsenine: *Méthodes de Résolution de Problèmes Mal Posés*, Mir, Moscow (1974).

[31] R. Wittmann: *Approximation of fixed points of nonexpansive mappings*, Arch. Math. 58 (1992) 486–491.