

ON THE FINITE TERMINATION OF THE GRADIENT PROJECTION METHOD

SHIN-YA MATSUSHITA

Department of Intelligent Mechatronics, Akita Prefectural University, Yuri-Honjo, Akita 015-0055, Japan

Abstract. The gradient projection method is an efficient iterative method for solving optimization problems. In this paper, we consider finite termination properties of the sequence generated by the gradient projection method in Hilbert spaces. We give estimates on the number of iterations by which the sequence terminates at a solution under mild assumptions.

Keywords. Gradient projection method; Finite termination property; Weakly sharp; Hilbert space.

2010 Mathematics Subject Classification. 90C25, 47H05, 47J25, 47H09.

1. INTRODUCTION

We consider the minimization problem

$$\text{minimize } f(x) \text{ subject to } x \in C, \quad (1.1)$$

where $f : H \rightarrow \mathbb{R}$ is a Fréchet differentiable and convex function with the Lipschitz continuous gradient and C is a nonempty, closed and convex set in H . Denote by ∇f the gradient of f , L the Lipschitz constant of ∇f , S the solution set of (1.1) and assume that S is nonempty.

A wide variety of problems can be solved by finding a solution of problem (1.1) (see, e.g., [2, 18]). Problem (1.1) is closely related to the variational inequality problem for $F : C \rightarrow H$

$$\text{find } x^* \in C \text{ such that } \langle y - x^*, F(x^*) \rangle \geq 0 \ (\forall y \in C). \quad (1.2)$$

It is known that (1.2) can be formulated equivalently to (1.1) when $F = \nabla f$ (see, e.g., [2, Proposition 26.5]). Therefore, (1.2) can be regarded as a natural extension of (1.1).

A useful method for solving (1.1) is the gradient projection method [7, 8], which has the form

$$x_k = P_C(I - \gamma_k \nabla f)(x_{k-1}), \quad k = 1, 2, \dots, \quad (1.3)$$

where $x_0 \in H$ is an initial point and $\{\gamma_k\}$ is a positive sequence. When we take gradient mapping ∇f in (1.3) to be F , the gradient projection method can be applied to solve (1.2) under appropriate assumptions. The behavior and convergence properties of the gradient projection method have been well studied (see, e.g., [2, 3, 5, 9, 13, 16, 17]). In particular, Combettes and Wajs [5, Theorem 3.4] showed that if $0 < \inf_k \gamma_k \leq \sup_k \gamma_k < 2/L$, then the sequence generated by (1.3) converges weakly to a solution of problem (1.1).

E-mail address: matsushita@akita-pu.ac.jp.

Received December 31, 2018; Accepted March 1, 2019.

We investigate the finite convergence property of the sequence generated by the gradient projection method under the solution set is weakly sharp. The notion of weakly sharp occurs in optimization problems [4, 15] and is an important tool in the analysis of the perturbation behavior of (1.1) as well as in the convergence analysis of algorithms designed to solve the problem; see [1, 6, 10, 11, 12, 15] and the references therein. This notion was extended by Patriksson [14] to the variational inequality problem and developed by Marcotte and Zhu [10]. In [11], the authors established the finite termination of the gradient projection method for (1.2) under the assumptions that S is weakly sharp and F is inverse strongly monotone. A remarkable feature of problem (1.1) is that the inverse strong monotonicity of the gradient ∇f is satisfied automatically (see, e.g., [2, Theorem 18.5]). Moreover, Al-Homidan, Ansari and Nguyen [1] gave estimates on the number of iterates by which the sequence generated by the gradient projection method converges to a solution when S is weakly sharp and F is monotone, strongly pseudomonotone [1, Definition 1 (d)] and Lipschitz continuous. More precisely, they showed that the upper bound of the iteration number needed to achieve termination depends on the modulus of the strong pseudomonotonicity and the Lipschitz constant of F , the distance from the initial point x_0 to S , and the modulus of the weak sharpness of S [1, Theorem 6]. Note that in (1.1), the gradient ∇f is not necessarily strongly pseudomonotone.

In this paper, we focus on (1.3) for solving problem (1.1). By using the descent properties of the gradient mappings, we can provide that, under mild assumptions, new estimates of the upper bound of iteration number for the generated sequence to terminate. The advantages of our results are that we do not require ∇f to be strongly pseudomonotone and our assumptions on $\{\gamma_k\}$ are mild. Our estimates complement the results of [1, Theorem 6.1] and [12, Theorem 4.1].

The rest of this paper is organized as follows. In Section 2, some preliminaries are presented. We give explicit upper bounds for the required number of iterations by which the sequence generated by the gradient projection method terminates at a solution in Section 3. Finally, the paper is concluded in Section 4.

2. PRELIMINARIES

The following notations will be used in this paper: \mathbb{R} denotes the set of real numbers; $\mathbb{N} = \{1, 2, \dots\}$ denotes the set of positive integers; H denotes a real Hilbert space; for any $x, y \in H$, $\langle x, y \rangle$ denotes the inner product of x and y ; for any $z \in H$, $\|z\|$ denotes the norm of z , i.e., $\|z\| = \sqrt{\langle z, z \rangle}$; for any $x \in H$ and $C \subset H$, $\text{dist}(x, C)$ denotes the distance between x and C , i.e. $\text{dist}(x, C) = \inf\{\|x - y\| : y \in C\}$; for any $y \in C$, $C + y$ denotes the translation of C by y , i.e., $C + y = \{u + y : u \in C\}$; for any $\varepsilon > 0$, $B(0, \varepsilon)$ denotes the closed ball of H with radius ε .

A mapping $U : H \rightarrow H$ is said to be

- (i) nonexpansive if

$$\|U(x) - U(y)\| \leq \|x - y\| \quad (\forall x, y \in H).$$

- (ii) inverse strongly monotone with constant $\beta > 0$ if

$$\beta \|U(x) - U(y)\|^2 \leq \langle x - y, U(x) - U(y) \rangle \quad (\forall x, y \in H).$$

Let $f : H \rightarrow \mathbb{R}$ be proper and convex. The function f is Fréchet differentiable at $x \in H$ if there exists $u \in H$ such that

$$\lim_{0 \neq y \rightarrow 0} \frac{f(x+y) - f(x) - \langle u, y \rangle}{\|y\|} = 0$$

(see, e.g., [2, p. 243]). In this case u is denoted by $\nabla f(x)$ and $\nabla f(x)$ is called the Fréchet gradient of f at x . Moreover, f is Fréchet differentiable on $V \subset H$ if it is Fréchet differentiable at every point of V . Let $f : H \rightarrow \mathbb{R}$ be Fréchet differentiable on H . ∇f is Lipschitz continuous with constant L if

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad (\forall x, y \in H).$$

The subdifferential of f at $x \in H$ is defined by

$$\partial f(x) = \{x^* \in H : f(y) \geq f(x) + \langle y - x, x^* \rangle \quad (\forall y \in H)\}.$$

The following is a fundamental property for Fréchet differentiable convex functions with the Lipschitz continuous gradient.

Lemma 2.1. ([2, Theorem 18.5]) *Let $f : H \rightarrow \mathbb{R}$ be a convex function. Then the following are equivalent:*

- (i) *f is Fréchet differentiable on H and ∇f is Lipschitz continuous with constant L .*
- (ii) *(descent lemma) f is Fréchet differentiable on H and*

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

- (iii) *f is Fréchet differentiable on H and ∇f is inverse strongly monotone with constant $1/L$.*

Let $C \subset H$ be a nonempty, closed and convex set. The metric projection P_C from H onto C is defined by

$$P_C(x) = \operatorname{argmin}_{y \in C} \|x - y\| \quad (\forall x \in H).$$

Note that $u = P_C(x)$ if and only if

$$\langle x - u, y - u \rangle \leq 0 \quad (\forall y \in C) \quad (2.1)$$

(see, e.g., [2, Theorem 3.14], [18, Lemma 3.1.3]). The polar C° of C is defined by

$$C^\circ = \{x^* \in H : \langle x^*, x \rangle \leq 0 \quad (\forall x \in C)\}.$$

The tangent cone to the set C at $x \in C$ is defined as

$$T_C(x) = \operatorname{cl}\left\{ \bigcup_{\lambda > 0} (C - x)/\lambda \right\}.$$

The normal cone to C at x is defined by $N_C(x) = T_C(x)^\circ$, and we can see that

$$N_C(x) = \{u \in H : \langle u, y - x \rangle \leq 0 \quad (\forall y \in C)\}.$$

The solution set S of (1.1) is said to be weakly sharp if

$$-\nabla f(x^*) \in \operatorname{int} \bigcap_{x \in S} (T_C(x) \cap N_S(x))^\circ \quad (2.2)$$

for any $x^* \in S$. From the definition, if S is weakly sharp, then there exists positive constant $\alpha > 0$ such that $B(0, \alpha) \subset \nabla f(x^*) + (T_C(x^*) \cap N_S(x^*))^\circ$ for each $x^* \in S$. We call the constant α the modulus of the weak sharpness of S . Based on this assumption, we can establish the finite convergence for (1.3).

The following results are useful.

Lemma 2.2. ([11, Remark 3.3 and Theorem 3.1]) *Let $\{x_k\}$ be the sequence generated by (1.3). Then, for any $k \in \mathbb{N} \cup \{0\}$ and for any $y \in C$, we have*

$$\begin{aligned} & \langle x_{k+1} - y, \nabla f(x_{k+1}) \rangle \\ & \leq \frac{1}{\gamma_{k+1}} (\langle x_k - x_{k+1}, x_{k+1} - y \rangle + \gamma_{k+1} \langle x_{k+1} - y, \nabla f(x_{k+1}) - \nabla f(x_k) \rangle). \end{aligned}$$

Lemma 2.3. ([1, Theorem 2]) *If the solution set S of (1.1) is weakly sharp with modulus $\alpha > 0$, then, for any $x \in C$, we have*

$$\langle x - P_S(x), \nabla f(x) \rangle \geq \alpha \text{dist}(x, S). \quad (2.3)$$

3. FINITE TERMINATION PROPERTY OF THE GRADIENT PROJECTION METHOD

We show the following lemma, which is very useful for the finite convergence analysis for (1.3).

Lemma 3.1. *Let H be a Hilbert space and let C be a nonempty, closed convex set in H . Let $f : H \rightarrow \mathbb{R}$ be a Fréchet differentiable and convex function with L -Lipschitz continuous gradient and let $T := P_C(I - \gamma \nabla f)$ with $\gamma > 0$. Then*

$$f(x) - f(T(y)) \geq \frac{1}{2} \left(\frac{1}{\gamma} - L \right) \|y - T(y)\|^2 + \frac{1}{2\gamma} (\|x - T(y)\|^2 - \|x - y\|^2)$$

for any $x \in C$ and $y \in H$.

Proof. Let $x \in C$ and $y \in H$. Applying Lemma 2.1 (ii) with $x := y$ and $y := T(y)$, we have

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

and hence

$$f(x) - f(T(y)) \geq f(x) - \{f(y) + \langle T(y) - y, \nabla f(y) \rangle + \frac{L}{2} \|T(y) - y\|^2\}. \quad (3.1)$$

On the other hand, let

$$z := \frac{1}{\gamma} (y - T(y)) - \nabla f(y).$$

Using (2.1), we have

$$\langle y - \gamma \nabla f(y) - T(y), x - T(y) \rangle \leq 0$$

and hence

$$\langle z, x - T(y) \rangle \leq 0. \quad (3.2)$$

Moreover, since $\partial f(y) = \{\nabla f(y)\}$ (see, e.g., [16, Propositions 1.26. and 3.20.]), we have

$$f(x) \geq f(y) + \langle x - y, \nabla f(y) \rangle. \quad (3.3)$$

Then, using (3.2) and (3.3), we deduce

$$f(x) \geq \langle x - T(y), z \rangle + f(y) + \langle x - y, \nabla f(y) \rangle. \quad (3.4)$$

Applying (3.4) to the right-hand side of (3.1), we have

$$\begin{aligned}
f(x) - f(T(y)) &\geq \langle x - T(y), z \rangle + f(y) + \langle x - y, \nabla f(y) \rangle \\
&\quad - \{f(y) + \langle T(y) - y, \nabla f(y) \rangle + \frac{L}{2} \|T(y) - y\|^2\} \\
&= -\frac{L}{2} \|y - T(y)\|^2 + \langle x - T(y), z + \nabla f(y) \rangle \\
&= -\frac{L}{2} \|y - T(y)\|^2 + \frac{1}{\gamma} \langle x - T(y), y - T(y) \rangle. \tag{3.5}
\end{aligned}$$

Applying the relation

$$2\langle b - a, c - a \rangle = \|c - a\|^2 + \|b - a\|^2 - \|b - c\|^2$$

to the right-hand side of (3.5) with $a := T(y)$, $b := x$ and $c := y$, we obtain

$$f(x) - f(T(y)) \geq \frac{1}{2} \left(\frac{1}{\gamma} - L \right) \|y - T(y)\|^2 + \frac{1}{2\gamma} (\|x - T(y)\|^2 - \|x - y\|^2).$$

□

Using Lemmas 2.2, 2.3 and 3.1, we show the following result. This is the main result of this paper.

Theorem 3.1. *Let H be a Hilbert space and let C be a nonempty, closed and convex set in H . Let $f : H \rightarrow \mathbb{R}$ be a Fréchet differentiable and convex function with L -Lipschitz continuous gradient and let $\{x_k\}$ be the sequence generated by (1.3). Assume that S is weakly sharp with modulus $\alpha > 0$. Then the following are hold.*

- (1) *If $\{\gamma_k\} \subset (\varepsilon, 2/L - \varepsilon)$ with $\varepsilon > 0$, then $\{x_k\}$ terminates at a solution of problem (1.1) in at most $l + 1$ iterations with*

$$l \leq \frac{18(2 - \varepsilon L)}{\alpha^2 \varepsilon^3 L^2} (f(x_0) - f^*). \tag{3.6}$$

where f^ denotes the optimal value to (1.1);*

- (2) *If $\{\gamma_k\} \subset (\varepsilon, 1/L - \varepsilon)$ with $\varepsilon > 0$, then $\{x_k\}$ terminates at a solution of problem (1.1) in at most $l + 1$ iterations with*

$$l \leq \frac{4(1 - \varepsilon L)}{\alpha^2 \varepsilon^3 L} \text{dist}(x_0, S)^2. \tag{3.7}$$

Proof. (1) Under the assumption that $\{\gamma_k\} \subset (\varepsilon, 2/L - \varepsilon)$, $\{\|x_{l+1} - x_l\|\}$ converges to 0 (e.g., see [5, Theorem 3.4 (iii)]). Let l be the smallest integer such that

$$\|x_{l+1} - x_l\| < \frac{\alpha \varepsilon}{3}. \tag{3.8}$$

If $x_{l+1} \notin S$, we obtain from Lemma 2.2, the Lipschitz continuity of ∇f and $\{\gamma_k\} \subset (\varepsilon, 2/L - \varepsilon)$ that

$$\begin{aligned}
&\langle x_{l+1} - y, \nabla f(x_{l+1}) \rangle \\
&\leq \frac{1}{\gamma_{l+1}} (\langle x_l - x_{l+1}, x_{l+1} - y \rangle + \gamma_{l+1} \langle x_{l+1} - y, \nabla f(x_{l+1}) - \nabla f(x_l) \rangle) \\
&\leq \frac{1 + \gamma_{l+1} L}{\gamma_{l+1}} \|x_l - x_{l+1}\| \|x_{l+1} - y\| \\
&\leq \frac{3}{\varepsilon} \|x_l - x_{l+1}\| \|x_{l+1} - y\| \tag{3.9}
\end{aligned}$$

for any $y \in C$. Using Lemma 2.3 with $x := x_{l+1}$, we have

$$\alpha \|x_{l+1} - P_S(x_{l+1})\| \leq \langle x_{l+1} - P_S(x_{l+1}), \nabla f(x_{l+1}) \rangle.$$

This together with (3.9) yields that

$$\alpha \|x_{l+1} - P_S(x_{l+1})\| \leq \frac{3}{\varepsilon} \|x_l - x_{l+1}\| \|x_{l+1} - P_S(x_{l+1})\|,$$

and hence

$$\frac{\alpha \varepsilon}{3} \leq \|x_l - x_{l+1}\|,$$

which contradicts (3.8). So, we must have $x_{l+1} \in S$.

Using Lemma 3.1 with $x = y := x_{k-1}$ and $T := P_C(I - \gamma_k \nabla f)$, we have

$$\begin{aligned} f(x_{k-1}) &\geq f(x_k) + \frac{1}{2} \left(\frac{1}{\gamma_k} - L \right) \|x_{k-1} - x_k\|^2 + \frac{1}{2\gamma_k} \|x_{k-1} - x_k\|^2 \\ &\geq f(x_k) + \left(\frac{1}{\gamma_k} - \frac{L}{2} \right) \|x_{k-1} - x_k\|^2 \\ &\geq f(x_k) + \left(\frac{L}{2 - \varepsilon L} - \frac{L}{2} \right) \|x_{k-1} - x_k\|^2. \end{aligned}$$

Using this inequality, we have

$$\begin{aligned} f(x_0) &\geq f(x_l) + \frac{\varepsilon L^2}{2(2 - \varepsilon L)} \sum_{j=1}^l \|x_{j-1} - x_j\|^2 \\ &\geq f^* + \frac{\varepsilon L^2}{2(2 - \varepsilon L)} \sum_{j=1}^l \|x_{j-1} - x_j\|^2 \\ &\geq f^* + \frac{\varepsilon L^2}{2(2 - \varepsilon L)} \left(\frac{\alpha \varepsilon}{3} \right)^2 l, \end{aligned}$$

where the last equality holds due to $\|x_{j-1} - x_j\| \geq \alpha \varepsilon / 3$ ($\forall j \leq l$). We can therefore conclude that

$$l \leq \frac{18(2 - \varepsilon L)}{\alpha^2 \varepsilon^3 L^2} (f(x_0) - f^*).$$

(2) By using $\{\gamma_k\} \subset (\varepsilon, 1/L - \varepsilon)$ and similar arguments used in the proof of (1), if l is the smallest integer such that

$$\|x_{l+1} - x_l\| < \frac{\alpha \varepsilon}{2}, \tag{3.10}$$

then $x_{l+1} \in S$. Let $x^* \in S$. Using Lemma 3.1 with $x := x^*$, $y := x_{k-1}$ and $T := P_C(I - \gamma_k \nabla f)$, we have

$$f^* \geq f(x_k) + \frac{1}{2} \left(\frac{1}{\gamma_k} - L \right) \|x_{k-1} - x_k\|^2 + \frac{1}{2\gamma_k} (\|x^* - x_k\|^2 - \|x^* - x_{k-1}\|^2),$$

and hence

$$\left(\frac{1}{\gamma_k} - L \right) \|x_{k-1} - x_k\|^2 \leq 2(f^* - f(x_k)) + \frac{1}{\gamma_k} (\|x^* - x_{k-1}\|^2 - \|x^* - x_k\|^2).$$

Using $\{\gamma_k\} \subset (\varepsilon, 1/L - \varepsilon)$ and $f^* - f(x_k) \leq 0$, we have

$$\frac{L}{1 - \varepsilon L} \|x_{k-1} - x_k\|^2 \leq \frac{1}{\varepsilon} (\|x^* - x_{k-1}\|^2 - \|x^* - x_k\|^2).$$

Using this inequality, we have

$$\begin{aligned} \frac{1}{\varepsilon}(\|x^* - x_0\|^2 - \|x^* - x_l\|^2) &\geq \frac{L}{1 - \varepsilon L} \sum_{j=1}^l \|x_{j-1} - x_j\|^2 \\ &\geq \frac{L}{1 - \varepsilon L} \left(\frac{\alpha\varepsilon}{2}\right)^2 l, \end{aligned}$$

where the last equality holds due to $\|x_{j-1} - x_j\| \geq \alpha\varepsilon/2$ ($\forall j \leq l$). Since x^* is an arbitrary point in S , we can therefore conclude that

$$l \leq \frac{4(1 - \varepsilon L)}{\alpha^2 \varepsilon^3 L} \text{dist}(x_0, S)^2.$$

□

Remark 3.1. When we take the gradient mapping ∇f to be a mapping $F : C \rightarrow H$, (1.3) can be written as

$$x_k = P_C(I - \gamma_k F)(x_{k-1}). \quad (3.11)$$

Then (3.11) can be applied to solve (1.2). Under the assumption that S is weakly sharp, finite convergence results have been established. In [11, Theorem 4.1], finite convergence of (3.11) was established when F is inverse strongly monotone. However, estimates on the number of iterations were not obtained. When F is monotone, strongly pseudomonotone with the modulus μ and Lipschitz continuous with the constant L , $\sigma \in (0, 1)$ and $\frac{L^2}{2\mu} \leq \gamma_k < \sigma$ ($\forall k \in \mathbb{N}$), it is shown that $\{x_k\}$ terminates at the unique solution x^* in at most l iterations with

$$l \leq \frac{(2\mu + L^3)^2}{(1 - \sigma^2)\alpha^2 L^4} \|x_0 - x^*\|^2 + 1 \quad (3.12)$$

[1, Theorem 6]. Therefore, it can be obtained an explicit upper bound for the required number of iterations. In Theorem 3.1, ∇f is not required to be strongly pseudomonotone. Moreover, our assumptions on $\{\gamma_k\}$ are very mild.

4. CONCLUSIONS

We established finite convergence results of the gradient projection method under the weak sharpness assumption. In particular, we gave two upper bounds of iteration number for the generated sequence to terminate. Our result complements the results of [11, Theorem 4.1] and [1, Theorem 6.1].

Acknowledgments

The author is grateful to Professors W. Takahashi of Tokyo Institute of Technology, D. Kuroiwa of Shimane university and L. Xu of Akita Prefectural University for their helpful support. This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology [grant number 16K05280].

REFERENCES

- [1] S. Al-Homidan, Q.H. Ansari, L.V. Nguyen, Finite convergence analysis and weak sharp solutions for variational inequalities, *Optim. Lett.* 11 (2017), 1647-1662.
- [2] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

- [3] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* 2 (2009), 183-202.
- [4] J.V. Burke, M.C. Ferris, Weak sharp minima in mathematical programming, *SIAM J. Control Optim.* 31 (1993), 1340-1359.
- [5] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.* 4 (2005), 1168-20.
- [6] M.C. Ferris, Finite termination of the proximal point algorithm, *Math. Program.* 50 (1991), 359-366.
- [7] A.A. Goldstein, Convex programming in Hilbert space, *Bull. Amer. Math. Soc.* 70 (1964), 709-710.
- [8] E.S. Levitin, B.T. Polyak, Constrained minimization methods, *Comput. Math. Math. Phys.* 6 (1966), 1-50.
- [9] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.* 16 (1979), 964-979.
- [10] P. Marcotte, D.L. Zhu, Weak sharp solutions of variational inequalities, *SIAM J. Optim.* 9 (1999), 179-189.
- [11] S. Matsushita, L. Xu, Finite convergence of the proximal point algorithm for variational inequality problems, *Set-Valued Var. Anal.* 21 (2013), 297-309.
- [12] S. Matsushita, L. Xu, On finite convergence of iterative methods for variational inequalities in Hilbert spaces, *J. Optim. Theory Appl.* 161 (2014), 701-715.
- [13] G.B. Passty, Ergodic convergence to a zero of the sum of monotone operators in Hilbert space, *J. Math. Anal. Appl.* 72 (1979), 383-390.
- [14] M. Patriksson, Unified framework of descent algorithms for nonlinear programs and variational inequalities, Ph.D. Thesis, Linköping Institute of Technology (1993).
- [15] M. Patriksson, *Nonlinear Programming and Variational Inequality Problems, a Unified Approach*, Kluwer Academic, Dordrecht 1999.
- [16] J. Peypouquet, *Convex Optimization in Normed Spaces. theory, methods and examples*, Springer, Cham, 2015.
- [17] S. Salzo, The variable metric forward-backward splitting algorithm under mild differentiability assumptions, *SIAM J. Optim.* 27 (2017), 2153-2181.
- [18] W. Takahashi, *Nonlinear Functional Analysis. Fixed Points Theory and its Applications*, Yokohama Publishers, Yokohama, 2000.