Applied Analysis and Optimization
Volume 6, Number 3, 2022, 431–443

Yokohama Publishers
ISSN 2189-1664 Online Journal

AN INERTIAL SPECTRAL GRADIENT PROJECTION METHOD FOR ZEROS OF MONOTONE MAPS

MA'ARUF MINJIBIR AND MUHAMMAD SALISU

Abstract. An algorithm for approximating zeros of monotone maps is developed. The algorithm incorporates an inertial term into a scheme that combines a modified spectral gradient method and projection method. The sequence generated is shown to converge globally to a zero of the said map. Numerical tests conducted exhibit great promise and significant improvement in performance of the algorithm as compared with the non-inertial methods.

1. Introduction

We consider the problem of solving the (nonlinear) equation

$$F(x) = 0, \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is *monotone*, that is, F satisfies

$$\langle F(x) - F(y), x - y \rangle \geq 0 \text{ for all } x, y \in \mathbb{R}^n,$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product in $\mathbb{R}^n$ (if we replace $\mathbb{R}^n$ with arbitrary real Hilbert space H, the same definition persists). The class of monotone functions constitutes a great part of the functions for which solutions of (1.1) are found without smoothness assumptions. In fact, several problems in applications reduce to (1.1) with F monotone. For instance, the evolution equation $\frac{du}{dt} + Au = 0$ which describes a system evolving with time, reduces to $Au = 0$ at equilibrium and the operator A is monotone (or more generally *a*ccretive). In connection with optimization, solutions of (1.1) may also correspond to the optimizers of certain functional $f : \mathbb{R}^n \to \mathbb{R}$. When the function f is convex, then the set of its minimizers coincides with the set of solutions of the inclusion $0 \in \partial f(x)$, where ∂f denotes the *subdifferential* of f. For Gâteaux differentiable, say, the mentioned inclusion becomes (1.1) with F denoting the Gâteuax derivative of f which is necessarily monotone.

Solodov and Svaiter [19] proposed a novel Newton-type method for approximation of solutions of (1.1) in the case F is monotone. In their method, utilizing the classical proximal point algorithm, in each iteration they constructed an appropriate hyperplane which strictly separates the current iterate from the set of solutions of (1.1). Then the next iterate is obtained as a projection of the current iterate onto another carefully defined hyperplane. While projections usually add to difficulties and time consumption in iterative methods, it is noted that the projection in the method of Solodov and Svaiter is onto a hyperplane which is readily and easily

---

2020 *Mathematics Subject Classification.* 47H06, 47J25, 65K15, 90C25.

Key words and phrases. Monotone map, inertial algorithm, spectral gradient method, projection method.

computed. Zhang and Zhou [20] took advantage of this projection technique and the spectral gradient method [6] to propose a new method called *spectral gradient projection method* for approximating solutions of (1.1). The method of Zhang and Zhou can also be viewed as a modification of that of Cruz and Raydan [12]. Upon implementation, the method in [20] demonstrates great efficiency and seems to have added to the robustness of the method of Solodov and Svaiter [19]. A very recent work related to [20] can be found in [1].

Implicit one-step discretization of a second order differential system describing the motion of a ball rolling under its own inertia, called "heavy ball with friction" (see, e.g., [2–5]), or "the method of a small heavy sphere" as used by Polyak [17], gives rise to the so-called inertial proximal method for maximal monotone maps, which speed up the convergence of the classical proximal algorithms. Polyak [17] is, perhaps, the first to make this connection. Recently, the idea inertial algorithms is utilized by numerous authors in accelerating convergence of many numerical algorithms for optimization and other problems, see, for example, [8–11, 13, 15, 16, 18].

In this paper, it is our purpose to incorporate inertial terms in the algorithm of Zhang and Zhou [20] and prove global convergence to a solution of (1.1). We also conduct numerical experiments to see the effects of the inertial terms with regard to speeding up convergence.

2. Algorithm and mathematical preliminaries

In this section we give some lemmas that will be used in the proof of our main theorem and state the inertial algorithm.

2.1. Mathematical Preliminaries.

Lemma 2.1 (See, e.g., [7]). *Let H be a real Hilbert space and let K_1 and K_2 be defined as $K_1 := \{x \in H : \langle u, x\rangle \leq \alpha\}$ and $K_2 := \{x \in H : \langle u, x\rangle = \alpha\}$, $u \in H$ and $\alpha \in \mathbb{R}$ fixed. If $K = K_1$ or $K = K_2$, then the projection onto K, P_K, is defined as*

$$P_K x = \begin{cases} x, & x \in K \\ x + \frac{\alpha - \langle x,\, u\rangle}{\|u\|^2} u, & x \in K^c. \end{cases}$$

Lemma 2.2 ([4]). *Let $\varphi_k \geq 0$ and $\delta_k \geq 0$ be such that $\varphi_{k+1} \leq \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k$, with $\sum_{k=0}^{\infty} \delta_k < \infty$ and $0 \leq \alpha_k \leq \alpha < 1$. Then the following hold:*

(i) $\sum_{k=0}^{\infty}[\varphi_k - \varphi_{k-1}]_+ < \infty$, *where* $[t]_+ = max\{t, 0\}$,
(ii) *there exists a real number* $\varphi^* \geq 0$ *such that* $\lim_{k\to\infty} \varphi_k = \varphi^*$.

2.2. **Algorithm.** Choose any $x_0,\ x_1 \in H$ with $x_0 \neq x_1$. Fix $r > 0$, σ, $\beta \in (0,1)$ and $\{\lambda_k\}_{k\geq 1} \subseteq [0,1)$. Define $\{x_k\}_{k\geq 0}$ in H as follows:

$$\begin{cases} u_k = x_k + \lambda_k(x_k - x_{k-1}), \\ \theta_k = \begin{cases} \frac{\|u_k-u_{k-1}\|^2}{\langle F(u_k)-F(u_{k-1}),u_k-u_{k-1}\rangle+r\|u_k-u_{k-1}\|^2}, & x_k = x_{k-1} \\ \\ \frac{\|x_k-x_{k-1}\|^2}{\langle F(x_k)-F(x_{k-1}),\ x_k-x_{k-1}\rangle+r\|x_k-x_{k-1}\|^2}, & x_k \neq x_{k-1}, \end{cases} \\ \\ d_k = \begin{cases} -F(u_k), & k = 1 \\ -\theta_k F(u_k), & k \geq 2. \end{cases} \end{cases}$$

If $d_k = 0$, then $x_{k+i} = u_k,\ i \geq 1$. Otherwise,

$$x_{k+1} = u_k - \frac{\langle F(z_k), u_k - z_k\rangle}{\|F(z_k)\|^2} F(z_k),\ k \geq 1, \tag{2.1}$$

where, $z_k = u_k + \alpha_k d_k$ and $\alpha_k = \beta^{m_k}$, with

$$m_k = \min\{m \in \mathbb{N} : -\langle F(u_k + \beta^m d_k), d_k\rangle \geq \sigma\beta^m \|d_k\|^2\}.$$

Remark 2.3.

(i) If $x_k = x_{k-1}$ for some $k > 1$ (as $x_1 \neq x_0$), then either $x_j = u_{k-1}\ \forall\, j \geq k$ (this being the case if $d_{k-1} = 0$) or $x_k \neq u_{k-1}$ (this being the case if $d_{k-1} \neq 0$). For the case $d_{k-1} = 0$, the sequence is well-defined. If $d_{k-1} \neq 0$, then $x_k = x_{k-1}$ implies $u\ _k = x_k$ and $x_k \neq u_{k-1}$ (by definition of u_k). Consequently, $u_k \neq u_{k-1}$. Thus, whenever $d_{k-1} \neq 0$, we have θ_k is well defined.

(ii) If $d_k \neq 0$, then by continuity of F and that of the inner product, there exists at least one $m \in \mathbb{N}$ such that

$$-\langle F(u_k + \beta^m d_k), d_k\rangle \geq \sigma\beta^m \|d_k\|^2. \tag{2.2}$$

Therefore, m_k is well-defined and consequently, the step-length α_k is well-defined.

(iii) In view of (i) and (ii) above, the scheme above is well-defined. Moreover, by monotonicity of F, for any $v, w \in H$, we have

$$\langle F(v) - F(w), v - w\rangle + r\|v - w\|^2 \geq r\|v - w\|^2. \tag{2.3}$$

In addition, by the Lipschitz continuity of F, there exists a constant $L > 0$ such that

$$\|F(v) - F(w)\| \leq L\|v - w\|,\ \forall\, v, w \in H.$$

Hence, we have

$$\begin{aligned} \langle F(v) - F(w), v - w\rangle + r\|v - w\|^2 &\leq L\|v - w\|^2 + r\|v - w\|^2 \\ &= (L + r)\|v - w\|^2. \end{aligned} \tag{2.4}$$

So, we have from (2.1), (2.3) and (2.4) (setting $v = x_k$ and $w = x_{k-1}$ or $v = u_k$ and $w = u_{k-1}$) we obtain

$$\frac{\|F(u_k)\|}{L+r} \leq \|d_k\| \leq \frac{\|F(u_k)\|}{r}. \tag{2.5}$$

3. Convergence property

In this section we present the main theorem of the paper. The following two lemmas will be used in the convergence analysis. We only indicate the proof of one of the lemmas as the proof of the other one is immediate.

Lemma 3.1. *Let $u, v \in X$, X a real inner product space. Then $2\langle u, v\rangle \geq -\sigma\|u\|^2 - \frac{1}{\sigma}\|v\|^2$, $\forall\, \sigma > 0$.*

Lemma 3.2. *Let H be a real Hilbert space and let $F : H \to H$ be monotone and non zero. Suppose $x', \hat{x}, y \in H$ such that $\langle F(y),\ x' + \alpha(x' - \hat{x}) - y\rangle > 0,\ \alpha > 0$. Let*

$$x^+ = x^* - \frac{\langle F(y),\ x^* - y\rangle}{\|F(y)\|^2} F(y), \ \textit{where } x^* = x' + \alpha(x' - \hat{x}).$$

Then for any $\bar{x} \in H$ such that $F(\bar{x}) = 0$, the inequality

$$\|x^+ - \bar{x}\|^2 \leq \|x^* - \bar{x}\|^2 - \|x^+ - x^*\|^2 \tag{3.1}$$

holds.

Proof. Let $\bar{x} \in H$ be any point such that $F(\bar{x}) = 0$. By monotonicity of F, $\langle F(y), \bar{x} - y\rangle \leq 0$. It follows from the hypothesis that the hyperplane

$$H_\alpha^f := \{s \in H : \langle F(y),\ s - y\rangle = 0\}$$

strictly separates x^* from $\bar{x}$, i.e.,

$$\langle F(y),\ x^* - y\rangle > 0 \geq \langle F(y),\ \bar{x} - y\rangle. \tag{3.2}$$

Also, from **Lemma** 2.1, x^+ is the projection of x^* onto the halfspace

$$T := \{s \in H : \langle F(y),\ s - y\rangle \leq 0\}.$$

Thus,

$$\langle x^* - x^+,\ x^+ - s\rangle \geq 0,\ \forall\, s \in T. \tag{3.3}$$

Therefore,

$$\begin{aligned}\|x^* - \bar{x}\|^2 &= \|x^* - x^+\|^2 + \|x^+ - \bar{x}\|^2 + 2\langle x^* - x^+,\ x^+ - \bar{x}\rangle \\ &\geq \|x^* - x^+\|^2 + \|x^+ - \bar{x}\|^2 \text{ (since } \bar{x} \in T).\end{aligned}$$

Hence,

$$\|x^+ - \bar{x}\|^2 \leq \|x^* - \bar{x}\|^2 - \|x^* - x^+\|^2.$$

□

Theorem 3.3. *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^n$ is a monotone and Lipschitz continuous non-zero map. Let $\{x_k\}$ be a sequence generated by Algorithm 1. Suppose that the set $S := \{x \in H : F(x) = 0\}$ is not empty. For $\bar{x} \in S$, it holds that*

$$\|x_{k+1} - \bar{x}\|^2 \leq \|u_k - \bar{x}\|^2 - \|x_{k+1} - u_k\|^2 \text{ for all } k.$$

Furthermore, if $\{\lambda_k\}_{k\geq 1}$ is nondecreasing and $0 \leq \lambda_k \leq \min\left\{\lambda, \frac{(1-b-a)(a-\lambda)}{\epsilon_0\lambda}\right\}$ for some $\epsilon_0 > 1$, with $0 < \lambda < a < 1$ and $0 < b < 1 - a$, then $\{x_k\}$ converges to some $x^* \in S$.

Proof. If $d_{k_0} = 0$ for some $k_0 \in \mathbb{N}$, then $x_k = u_k\ \forall\, k > k_0$. By definition of u_{k_0} in the algorithm, $d_{k_0} = 0$ implies $F(u_{k_0}) = 0$. Hence the assertions of the theorem follow trivially in this case.

We now assume that $d_k \neq 0$ for all k. We have from (2.2) that

$$\langle F(z_k), u_k - z_k\rangle = -\alpha_k\langle F(z_k), d_k\rangle \geq \sigma\alpha_k^2\|d_k\|^2 > 0. \tag{3.4}$$

Then for any $\bar{x} \in S$, by (2.1) and **Lemma** 3.2, it follows that for all $k \geq 1$,

$$\|x_{k+1} - \bar{x}\|^2 \leq \|u_k - \bar{x}\|^2 - \|x_{k+1} - u_k\|^2. \tag{3.5}$$

Using the technique of Alvarez and Attouch [4], let $\varphi_k = \|x_k - \bar{x}\|^2,\ \forall\, k \geq 1$. Then

$$\begin{aligned} \|u_k - \bar{x}\|^2 &= \|x_k + \lambda_k(x_k - x_{k-1}) - \bar{x}\|^2 \\ &= \|x_k - \bar{x}\|^2 + 2\lambda_k\langle x_k - \bar{x}, x_k - x_{k-1}\rangle + \lambda_k^2\|x_k - x_{k-1}\|^2,\ k \geq 1. \end{aligned} \tag{3.6}$$

Also,

$$\begin{aligned} \varphi_{k-1} = \|x_{k-1} - \bar{x}\|^2 &= \|(x_{k-1} - x_k) + (x_k - \bar{x})\|^2 \\ &= \|x_{k-1} - x_k\|^2 + \|x_k - \bar{x}\|^2 + 2\langle x_{k-1} - x_k, x_k - \bar{x}\rangle \\ &= \varphi_k + \|x_{k-1} - x_k\|^2 + 2\langle x_{k-1} - x_k, x_k - \bar{x}\rangle. \end{aligned} \tag{3.7}$$

This implies,

$$\varphi_k = \varphi_{k-1} - \|x_{k-1} - x_k\|^2 + 2\langle x_k - x_{k-1}, x_k - \bar{x}\rangle, \text{ for all } k \geq 1. \tag{3.8}$$

Thus, $\langle x_k - \bar{x}, x_k - x_{k-1}\rangle = \frac{1}{2}(\varphi_k - \varphi_{k-1}) + \frac{1}{2}\|x_k - x_{k-1}\|^2$ and so (3.6) yields

$$\begin{aligned} \|u_k - \bar{x}\|^2 &= \|x_k - \bar{x}\|^2 + 2\lambda_k\left(\frac{1}{2}(\varphi_k - \varphi_{k-1}) + \frac{1}{2}\|x_k - x_{k-1}\|^2\right) + \lambda_k^2\|x_k - x_{k-1}\|^2 \\ &= \|x_k - \bar{x}\|^2 + \lambda_k(\varphi_k - \varphi_{k-1}) + \lambda_k\|x_k - x_{k-1}\|^2 + \lambda_k^2\|x_k - x_{k-1}\|^2. \end{aligned} \tag{3.9}$$

Therefore,

$$\|u_k - \bar{x}\|^2 = \varphi_k + \lambda_k(\varphi_k - \varphi_{k-1}) + (\lambda_k + \lambda_k^2)\|x_k - x_{k-1}\|^2. \tag{3.10}$$

Hence, using (3.5), we have

$$\|x_{k+1} - \bar{x}\|^2 \leq \varphi_k + \lambda_k(\varphi_k - \varphi_{k-1}) + (\lambda_k + \lambda_k^2)\|x_k - x_{k-1}\|^2 - \|x_{k+1} - u_k\|^2. \tag{3.11}$$

Thus,

$$\begin{aligned}
\varphi_{k+1} &\leq \varphi_k + \lambda_k(\varphi_k - \varphi_{k-1}) + (\lambda_k + \lambda_k^2)\|x_k - x_{k-1}\|^2 \\
&\quad - \|x_{k+1} - x_k - \lambda_k(x_k - x_{k-1})\|^2 \\
&= \varphi_k + \lambda_k(\varphi_k - \varphi_{k-1}) + (\lambda_k + \lambda_k^2)\|x_k - x_{k-1}\|^2 - \|x_{k+1} - x_k\|^2 \\
&\quad - \lambda_k^2\|x_k - x_{k-1}\|^2 + 2\lambda_k\langle x_{k+1} - x_k, x_k - x_{k-1}\rangle \\
&= \varphi_k + \lambda_k(\varphi_k - \varphi_{k-1}) + \lambda_k\|x_k - x_{k-1}\|^2 \\
&\quad - \|x_{k+1} - x_k\|^2 - 2\lambda_k\langle x_{k+1} - x_k, x_{k-1} - x_k\rangle \\
&\leq \varphi_k + \lambda_k(\varphi_k - \varphi_{k-1}) + \lambda_k\|x_k - x_{k-1}\|^2 - \|x_{k+1} - x_k\|^2 \\
&\quad + \lambda_k\left(\rho\|x_{k+1} - x_k\|^2 + \frac{1}{\rho}\|x_k - x_{k-1}\|^2\right),
\end{aligned}$$

where $\rho = \frac{\epsilon_0\lambda}{a-\lambda}$. The last inequality follows from **Lemma** 3.1. It follows that

$$\varphi_{k+1} \leq \varphi_k+\lambda_k(\varphi_k-\varphi_{k-1})+\lambda_k\left(1+\frac{1}{\rho}\right)\|x_k-x_{k-1}\|^2+(\lambda_k\rho-1)\|x_{k+1}-x_k\|^2. \tag{3.12}$$

Let $\mu_k = \varphi_k - \lambda_k\varphi_{k-1} + \gamma_k\|x_k - x_{k-1}\|^2,\ k \geq 1$. Then using the assumption that $\{\lambda_k\}_k$ is monotone nondecreasing, we get

$$\begin{aligned}
\mu_{k+1} - \mu_k &= \varphi_{k+1} - \lambda_{k+1}\varphi_k + \gamma_{k+1}\|x_{k+1} - x_k\|^2 - (\varphi_k - \lambda_k\varphi_{k-1} + \gamma_k\|x_k - x_{k-1}\|^2) \\
&\leq \varphi_{k+1} - (1 + \lambda_k)\varphi_k + \lambda_k\varphi_{k-1} + \gamma_{k+1}\|x_{k+1} - x_k\|^2 - \gamma_k\|x_k - x_{k-1}\|^2.
\end{aligned} \tag{3.13}$$

From (3.12) we obtain

$$\begin{aligned}
\varphi_{k+1} - (1 + \lambda_k)\varphi_k + \lambda_k\varphi_{k-1} - \gamma_k\|x_k - x_{k-1}\|^2 + \gamma_{k+1}\|x_{k+1} - x_k\|^2 \\
\leq (\lambda_k\rho - 1 + \gamma_{k+1})\|x_{k+1} - x_k\|^2,
\end{aligned} \tag{3.14}$$

where $\gamma_k = \lambda_k\left(1 + \frac{1}{\rho}\right)$. By virtue of (3.14), we have

$$\mu_{k+1} - \mu_k \leq (\lambda_k\rho - 1 + \gamma_{k+1})\|x_{k+1} - x_k\|^2. \tag{3.15}$$

We next show that

$$\lambda_k\rho - 1 + \gamma_{k+1} \leq -b \ \forall\, k \geq 1. \tag{3.16}$$

Indeed, we first note that by the definition of ρ and λ_k,

$$\gamma_k = \lambda_k\left(1 + \frac{1}{\rho}\right) = \lambda_k + \frac{\lambda_k}{\rho} \leq \lambda + \frac{\lambda}{\rho} < a \ \forall\, k \geq 1.$$

Therefore, to justify the claim, it suffices to show that $\lambda_k\rho - 1 + a \leq -b\ \forall\, k$. By definition of $\{\lambda_k\},\ \lambda_k \leq \frac{1-b-a}{\rho}\ \forall\, k \geq 1$. So, $\rho \leq \frac{1-b-a}{\lambda_k}$. This gives $\lambda_k\rho - 1 + a \leq -b$.

From (3.15) and (3.16) we obtain,

$$\mu_{k+1} - \mu_k \leq -b\|x_{k+1} - x_k\|^2 \ \forall\, k \geq 1. \tag{3.17}$$

It follows that the sequence $(\mu_k)_{k\geq 1}$ is nonincreasing. Since $\gamma_k \geq 0$ and $\lambda_k \leq \lambda\ \forall\, k$, we have from the definition of μ_k that

$$\varphi_k - \lambda\varphi_{k-1} \leq \mu_k \leq \mu_1\ \forall\, k \geq 1. \tag{3.18}$$

We therefore have,

$$\begin{aligned}
\varphi_k &\leq \lambda\varphi_{k-1} + \mu_1 \\
&\leq \lambda(\lambda\varphi_{k-2} + \mu_1) + \mu_1 \\
&\leq \lambda(\lambda(\lambda\varphi_{k-3} + \mu_1) + \mu_1) + \mu_1 \\
&= \lambda^3\varphi_{k-3} + \lambda^2\mu_1 + \lambda\mu_1 + \mu_1 \\
&\ \ \vdots \\
&\leq \lambda^k\varphi_0 + \mu_1\sum_{i=0}^{k-1}\lambda^i
\end{aligned}$$

$$\begin{aligned}
\varphi_k &\leq \lambda^k\varphi_0 + \mu_1\cdot\frac{1(1-\lambda^k)}{1-\lambda} \\
&\leq \lambda^k\varphi_0 + \frac{\mu_1}{1-\lambda}\ \forall\, k \geq 1.
\end{aligned}$$

Thus,

$$\varphi_k \leq \lambda^k\varphi_0 + \frac{\mu_1}{1-\lambda}\ \forall\, k \geq 1.$$

Combining (3.17) and (3.18) we get for all $k \geq 1$,

$$\begin{aligned}
b\sum_{k=1}^{n}\|x_{k+1} - x_k\|^2 &\leq \mu_1 - \mu_{n+1} \\
&= \mu_1 - (\varphi_{n+1} - \lambda_{n+1}\varphi_n + \gamma_{n+1}\|x_{n+1} - x_n\|^2) \\
&\leq \mu_1 - \varphi_{n+1} + \lambda_{n+1}\varphi_n \\
&\leq \mu_1 + \lambda\varphi_n \\
&\leq \mu_1 + \lambda\left(\lambda^n\varphi_0 + \frac{\mu_1}{1-\lambda}\right) \\
&= \lambda^{n+1}\varphi_0 + \frac{\mu_1}{1-\lambda}.
\end{aligned}$$

Since $\lambda \in [0, 1)$, it follows that

$$\sum_{k=1}^{\infty}\|x_{k+1} - x_k\|^2 < +\infty. \tag{3.19}$$

Setting $\delta_k := (\lambda_k + \lambda_k^2)\|x_k - x_{k-1}\|^2,\ k \geq 1$, we obtain from (3.11) that

$$\varphi_{k+1} \leq \varphi_k + \lambda_k\left(\varphi_k - \varphi_{k-1}\right) + \delta_k,\ k \geq 1.$$

We conclude, in view of (3.19) and **Lemma** 2.2, that $\lim\varphi_k$ exists in $\mathbb{R}$ which gives $\{\|x_k - \bar{x}\|\}$ is convergent.

Convergence of $\{\|x_k - \bar{x}\|\}$ implies boundedness of $\{x_k\}$ and this in turn gives boundedness of $\{u_k\}$. Therefore, using (2.2) and the Lipschitz continuity of F, we get that $\{d_k\}$ is bounded and so is $\{z_k\}$. Using (3.11), we get

$$\|x_{k+1} - u_k\|^2 \leq \varphi_k - \varphi_{k+1} + \lambda_k (\varphi_k - \varphi_{k-1}) + \delta_k, \ \forall\, k \geq 1.$$

From the facts that $\lim \varphi_k \in \mathbb{R}$, $\sum \delta_k < +\infty$ and $\sup \lambda_k < +\infty$, we have

$$\limsup_{k\to\infty} \|x_{k+1} - u_k\|^2 \leq (\limsup \lambda_k)(0) = 0.$$

It follows that

$$\lim_{k\to\infty} \|x_{k+1} - u_k\|^2 = 0.$$

Now by Lipschitz continuity of F and boundedness of $\{z_k\}$, there exists a constant $C > 0$ such that $\|F(z_k)\| \leq C$. Thus we obtain from (2.1) and (3.4) that

$$\|x_{k+1} - u_k\| = \frac{|\langle F(z_k), u_k - z_k\rangle|}{\|F(z_k)\|} \geq \frac{\sigma}{C}\alpha_k^2\|d_k\|^2 > 0. \tag{3.20}$$

Therefore, $0 < \frac{\sigma}{C}\alpha_k^2\|d_k\|^2 \leq \|x_{k+1} - u_k\| \to 0$ as $k \to \infty$. From this inequality, we have

$$\lim_{k\to\infty} \alpha_k\|d_k\| = 0. \tag{3.21}$$

We have two cases:
Case 1. $\liminf \|d_k\| = 0$; From (2.5) we have $\liminf \|F(u_k)\| = 0$. This implies that there exists a subsequence $\{u_{k_j}\}$ of $\{u_k\}$ such that $\lim_{j\to\infty} \|F(u_{k_j})\| = 0$. Since $\{u_{k_j}\}$ is bounded, there exists a subsequence $\{u_{k_{j_l}}\}$ of $\{u_{k_j}\}$ such that $u_{k_{j_l}} \to \hat{u} \in \mathbb{R}^n$. Hence, $\lim_{l\to\infty} \|F(u_{k_{j_l}})\| = \|F(\hat{u})\|$. Since $\lim_{l\to\infty} \|F(u_{k_{j_l}})\| = \lim_{j\to\infty} \|F(u_{k_j})\| = 0$, we conclude that $\|F(\hat{u})\| = 0$. So, $\hat{u} \in S$ and therefore $\{\|x_k - \hat{u}\|\}$ converges. Since $\hat{u}$ is an accumulation point of $\{x_k\}$, with thanks to the fact that $x_{k+1} - u_k \to 0$ and $\hat{x}$ is an accumulation point of $\{u_k\}$, it holds that $\{x_k\}$ converges to $\hat{u}$ and that concludes the proof.

Case 2. $\liminf \|d_k\| > 0$; From (2.5) we have $\liminf \|F(u_k)\| > 0$. By (3.21), it holds that

$$\lim_{k\to\infty} \alpha_k = 0. \tag{3.22}$$

Definition of m_k in the **Algorithm 1** implies that

$$-\langle F(u_k + \beta^{m_k - 1} d_k), d_k\rangle < \sigma\beta^{m_k-1}\|d_k\|^2. \tag{3.23}$$

Since $\{u_k\}, \{d_k\}$ are bounded, there exist $\hat{u}$ and $\hat{d}$ such that $u_{k_j} \to \hat{u}$ and $d_{k_j} \to \hat{d}$ for some subsequences $\{u_{k_j}\}$ of $\{u_k\}$ and $\{d_{k_j}\}$ of $\{d_k\}$. From (3.23) and the fact that $\beta \in (0,1)$, we obtain

$$-\langle F(\hat{u}), \hat{d}\rangle \leq 0. \tag{3.24}$$

On the other hand,

$$-\langle F(\hat{u}), \hat{d}\rangle > 0. \tag{3.25}$$

Indeed,

$$-\langle F(u_k + \beta^m d_k), d_k\rangle \to -\langle F(u_k), d_k\rangle \text{ as } m \to \infty.$$

From $d_k = -\theta_k F(u_k)$ in the algorithm we have

$$F(u_k) = -\frac{1}{\theta_k} d_k$$
$$= -\begin{cases} \frac{\langle F(u_k)-F(u_{k-1}),\ u_k-u_{k-1}\rangle + r\|u_k-u_{k-1}\|^2}{\|u_k-u_{k-1}\|^2} d_k, & x_k = x_{k-1} \\ \\ \frac{\langle F(x_k)-F(x_{k-1}),\ x_k-x_{k-1}\rangle + r\|x_k-x_{k-1}\|^2}{\|x_k-x_{k-1}\|^2} d_k, & x_k \neq x_{k-1}. \end{cases}$$

Therefore, by virtue of the monotonicity of F,

$$-\langle F(u_k), d_k\rangle = \begin{cases} \frac{\langle F(u_k)-F(u_{k-1}),\ u_k-u_{k-1}\rangle + r\|u_k-u_{k-1}\|^2}{\|u_k-u_{k-1}\|^2} \|d_k\|^2, & x_k = x_{k-1} \\ \\ \frac{\langle F(x_k)-F(x_{k-1}),\ x_k-x_{k-1}\rangle + r\|x_k-x_{k-1}\|^2}{\|x_k-x_{k-1}\|^2} \|d_k\|^2, & x_k \neq x_{k-1} \end{cases}$$
$$\geq \begin{cases} \frac{r\|u_k-u_{k-1}\|^2}{\|u_k-u_{k-1}\|^2} \|d_k\|^2, & x_k = x_{k-1} \\ \\ \frac{r\|x_k-x_{k-1}\|^2}{\|x_k-x_{k-1}\|^2} \|d_k\|^2, & x_k \neq x_{k-1} \end{cases}$$
$$= r\|d_k\|^2 > 0.$$

The fact that $\liminf \|d_k\| > 0$, gives the desired inequality. Therefore we have (3.24) and (3.25) which is a contradiction. Hence $\liminf_{k\to\infty} \|F(u_k)\| > 0$ is not possible, that is, *Case 2* is not possible. □

4. Numerical results

In this section, we conduct numerical experiments to compare the performance of Spectral Gradient Projection Method (SGPM) of Zhang and Zhou [20] and our method (**Algorithm 1**), i.e., the inertial version of it hereafter referred to as Inertial Spectral Gradient Projection Method (ISGPM). The algorithms were coded in MATLAB R2014a and run on personal computer with 2.60GHz CPU processor.

For the purpose of implementation, **Algorithm 1** above is restated to suit implementation.

ISGPM

Step 0: Choose any $x_0,\ x_1 \in \mathbb{R}^n$ with $x_0 \neq x_1,\ r > 0,\ \sigma,\ \beta \in (0,1)$ and $\{\lambda_k\}_{k\geq 1} \subseteq [0,1)$. Let $k := 1$ and let the tolerance $\epsilon > 0$ be given.

Step 1: Compute d_k by

$$d_k = \begin{cases} -F(u_k), & k = 1 \\ -\theta_k F(u_k), & k \geq 2, \end{cases}$$

where, $u_k = x_k + \lambda_k(x_k - x_{k-1})$ and

$$\theta_k = \begin{cases} \frac{\|u_k - u_{k-1}\|^2}{\langle F(u_k)-F(u_{k-1}), u_k - u_{k-1}\rangle + r\|u_k-u_{k-1}\|^2}, & x_k = x_{k-1} \\ \\ \frac{\|x_k - x_{k-1}\|^2}{\langle F(x_k)-F(x_{k-1}),\ x_k - x_{k-1}\rangle + r\|x_k-x_{k-1}\|^2}, & x_k \neq x_{k-1}. \end{cases}$$

Stop if $d_k = 0$, else proceed to **Step 2**.

Step 2: Compute $z_k = u_k + \alpha_k d_k$, where $\alpha_k = \beta^{m_k}$, with m_k being the smallest nonnegative integer m such that

$$m_k = \min\{m \in \mathbb{N} : -\langle F(u_k + \beta^m d_k), d_k\rangle \geq \sigma\beta^m \|d_k\|^2\}.$$

Step 3: Compute

$$x_{k+1} = u_k - \frac{\langle F(z_k), u_k - z_k\rangle}{\|F(z_k)\|^2} F(z_k).$$

Step 4: Stop if $\|F(x_k)\| \leq \epsilon$, else set $k := k+1$ and go to **Step 1**.

We use the following two examples which verify the assumptions on F in **Theorem 3.3.** These example were the functions used by by Zhang and Zhou [20].

Example 4.1. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be defined as $F(x) = (F_1(x), F_2(x), \ldots, F_n(x))^T$, $x = (x_1, x_2, \ldots, x_n)^T$, where $F_i(x) = x_i - \sin|x_i|,\ i = 1, 2, \ldots, n$.

TABLE 1. Test results for SGPM and ISGPM on Example 1

S/N	Initial points		SGPM		ISGPM	
	x_0	x_1	Iter.	Time(s)	Iter.	Time(s)
1	$(10, 10, \ldots, 10)^T$	$(9, 9, \ldots, 9)^T$	143	0.0282	39	0.0090
2	$(1, 1, \ldots, 1)^T$	$(2, 2, \ldots, 2)^T$	141	0.0287	47	0.0115
3	$(1, \frac{1}{2}, \ldots, \frac{1}{500})^T$	$(-1, -1, \ldots, -1)^T$	54	0.0096	4	0.0012
4	$(-10, \ldots, -10)^T$	$(-2, -2, \ldots, -2)^T$	27	0.0081	49	0.0098
5	$(-0.1, \ldots, -0.1)^T$	$(0.009, 0.009 \ldots, 0.009)^T$	15	0.005	1	0.0002
6	$(-1, -1, \ldots, -1)^T$	$(-1.9, -1.9 \ldots, -1.9)^T$	15	0.0050	40	0.0090
7	$(5, 5, \ldots, 5)^T$	$(-2, -2, \ldots, -2)^T$	127	0.0472	61	0.0137
8	$(1, \frac{1}{2}, -1, \ldots, \frac{1}{500})^T$	$(-1, -1, \ldots, -1)^T$	120	0.0191	5	0.0002
9	$(-8, -8, \ldots, -8)^T$	$(0, 0, \ldots, 0)^T$	35	0.0102	1	0.0009
10	$(6, 6, \ldots, 6)^T$	$(2, 2, \ldots, 2)^T$	127	0.0287	56	0.0554
11	$(\frac{1}{3}, \frac{1}{3}, \ldots, \frac{1}{3})^T$	$(1.1, 1.1, \ldots, 1.1)^T$	122	0.0297	64	0.0132
12	$(0.6, 0.6, \ldots, 0.6)$	$(1, 1, \ldots, 1)^T$	126	0.0255	46	0.0088

We test **Example 4.1** with different initial points, we also set $\beta = 0.4,\ \sigma = 0.01,\ r = 0.001,\ \text{nmax} = 10,000$ and $n = 500$ (dimension) and $\lambda_n = 0.4$. We use stopping criterion $\|F(x_k)\| < 0.0001$.

Example 4.2. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be defined as $F(x) = (F_1(x), F_2(x), \ldots, F_n(x))^T$, $x = (x_1, x_2, \ldots, x_n)^T$, where

$$\begin{aligned}
F_1(x) &= 2x_1 + \sin(x_1) - 1,\\
F_i(x) &= -2x_{i-1} + 2x_i + \sin(x_i) - 1,\ i = 1, 2, \ldots, n-1,\\
F_n(x) &= 2x_n + \sin(x_n) - 1.
\end{aligned}$$

We also test **Example 4.2** with different initial points, we set $\beta = 0.4$, $\sigma = 0.01$, $r = 0.1$, nmax $= 10,000$ and $n = 500$ (dimension) and $\lambda_n = 0.4$. We use stopping criterion $\|F(x_k)\| < 0.0001$. In the two tables above, x_0 and x_1 are the

TABLE 2. Test results for SGPM and ISGPM on Example 2

S/N	Initial points		SGPM		ISGPM	
	x_0	x_1	Iter.	Time(s)	Iter.	Time(s)
1	$(\frac{1}{3}, \frac{1}{3}, \ldots, \frac{1}{3})^T$	$(1, 1, \ldots, 1)$	495	0.2000	133	0.0137
2	$(1, 1, \ldots, 1)^T$	$(0, 0, \ldots, 0)^T$	835	0.3002	311	0.1262
3	$(1, \frac{1}{2}, \ldots, \frac{1}{500})^T$	$(0.1, 0.1, \ldots, 0.1)^T$	743	0.2825	81	0.0725
4	$(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})^T$	$(-0.4, -0.4, \ldots, -0.4)^T$	631	0.2206	99	0.0870
5	$(1, \frac{1}{2}, \ldots, \frac{1}{500})^T$	$(-0.2, -0.2, \ldots, -0.2)^T$	899	0.3080	79	0.0434
6	$(0.8, 0.8, \ldots, 0.8)^T$	$(1, 1, \ldots, 1)^T$	151	0.0700	78	0.0324
7	$(0.1, 0.1, \ldots, 0.1)^T$	$(0.2, 0.2, \ldots, 0.2)^T$	736	0.2555	122	0.0671
8	$(0, 0, \ldots, 0)^T$	$(-0.3, -0.3, \ldots, -0.3)^T$	530	0.1889	88	0.0403
9	$(-1, -1, \ldots, -1)^T$	$(0.1, 0.1, \ldots, 0.1)^T$	652	0.0700	78	0.0324
10	$(\frac{1}{7}, \frac{1}{7}, \ldots, \frac{1}{7})^T$	$(1, 1, \ldots, 1)^T$	1068	0.3750	133	0.0822

two initial points for the ISGPM while x_0 (only) serves as the the initial point of SGPM. Iter. denotes number of iterations and Time denotes the time for execution of the algorithm. With regard to **Example 1**, we observe from Table 1 that except in serial number 4 and 6, the time and number of iterations for ISGPM are less than those of SGPM . The same situation obtains with regard to **Example 2** where we observe from Table 2 that in all the entries of the table, ISGPM has far less number of iterations and less time than the SGPM. These results indicate that, as expected, the inertial version of Spectral Gradient Projection Method gives better convergence performance than the method without inertial term.

5. CONCLUSION

In conclusion, we are able to incorporate inertial term in the algorithm of Zhang and Zhou [20] SGPM and prove global convergence of the resulting inertial algorithm ISGPM to a solution of problem (1.1), given existence of a solution. The proof did not put the assumption of convergence of the series $\sum_k \lambda_k \|x_k - x_{k-1}\|$, a assumption found in some inertial algorithms (see, e.g., [3, 4, 15, 16]). Furthermore, from the numerical experiments, the ISGPM has exhibited faster convergence than the SGPM thereby making the whole work of adding inertial term live up to expectation.

References

[1] A. B. Abubakar, P. Kumam and H. Mohammad, *A note on the spectral gradient projection method for nonlinear monotone equations with applications*, Comput. Appl. Math. **39** (2020), 129.

[2] F. Alvarez, *On minimizing property of a second order dissipative system in Hilbert spaces,* SIAM J. Control Optim. **38** (2000), 1102–1119.

[3] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space,* SIAM J. Optim. **14** (2004), 773–782.

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

[5] H. Attouch, X. Goudou and P. Redont, *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SIAM J. Optim. **24** (2014), 232–256.

[6] J. Barzilai and J. M. Borwein,*Two-point step size gradient methods*, IMA J. Numer. Anal. **8** (1998), 141–148.

[7] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, London, UK, 2010.

[8] A.U. Bello, H. Yusuf and N. Djitte, *Single-step algorithm for variational inequality problems in 2-uniformly convex banach spaces*, Rend. Circ. Mat. Palermo, II. Ser. doi.org/10.1007/s12215-022-00746-7, (2022).

[9] R. I Boţ, E.C. Robert and C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Appl. Math. Comput. **256** (2015), 472–487.

[10] R. I Boţ and E. C. Robert, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numer. Algor. **71** (2015), 519–540.

[11] G. Cai, Q. L. Dong and Y. Peng, *Inertial viscosity iterative method for solving pseudo-monotone variational inequality problems and fixed point problems*, Acta. Math. Sin.-English Ser. **38** (2022), 937–952.

[12] W. Cruz and M. Raydan, *Nonmonotone spectral methods for large-scale nonlinear systems,* Optim. Meth. Soft. **18** (2003), 583–599.

[13] Q. L. Dong, Y. Y. Lu and J. Yang, *The extragradient algorithm with inertial effects for solving the variational inequality*, Optimization **65** (2016), 2217–2226.

[14] Q. Dong, L. Liu, X. Qin and J. Yao, *An alternated inertial general splitting method with linearization for the split feasibility problem,* Optimization, doi.org/10.1080/02331934.2022.2069567, (2022).

[15] P. E. Maingé, *Convergence theorems for inertial KM-type algorithms*, Comput. Appl. Math. **219** (2008), 223–236.

[16] P. E. Maingé, Regularized and inertial algorithms for common fixed points of nonlinear operators, J. Math. Anal. Appl. **344** (2008), 876–887.

[17] B. T. Polyak, *Some methods of seeding up the convergence of iteration methods*, Zh. Vych. Mat. **4** (1964), 791–603.

[18] Y. Shehu, O. S. Iyiola and F. U. Ogbuisi, *Iterative method with inertial terms for nonexpansive mappings: applications to compressed sensing*, Numer. Algor. **83** (2020), 1321–1347.

[19] M. N. Solodov and B. F. Svaiter, *Hybbrid projection-proximal point algorithm*, J. Convex Anal. **6** (1997), 59–70.

[20] L. Zhang and W. Zhou, *Spectral gradient projection method for solving nonlinear monotone equations*, J. Comput. Appl. Math. **196** (2006), 478–484.

Manuscript received September 16 2022
revised October 16 2022

M. Minjiir
Department of Mathematical Sciences, Bayero University, Kano, Nigeria
E-mail address: msminjibir.mth@buk.edu.ng

M. Salisu
Department of Mathematical Sciences, Bayero University, Kano, Nigeria
E-mail address: msmuhd10@gmail.com