J. COMPUTATIONAL ANALYSIS AND APPLICATIONS, VOL. 29, NO.5, 2021, COPYRIGHT 2021 EUDOXUS PRESS, LLC

Rapid gradient penalty schemes and convergence for solving constrained convex optimization problem in Hilbert spaces

Natthaphon Artsawang[a,1], Kasamsuk Ungchittrakool[a,b,*]

[a] *Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand*
[b] *Research Center for Academic Excellence in Nonlinear Analysis and Optimization, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand*

Abstract

The purposes of this paper are to establish and study the convergence of a new gradient scheme with penalization terms called rapid gradient penalty algorithm (**RGPA**) for minimizing a convex differentiable function over the set of minimizers of a convex differentiable constrained function. Under the observation of some appropriate choices for the available properties of the considered functions and scalars, we can generate a suitable algorithm that weakly converges to a minimal solution of the considered constraint minimization problem. Further, we also provide a numerical example to compare the rapid gradient penalty algorithm (**RGPA**) and the algorithm introduced by Peypouquet [20].

Keywords: Rapid gradient penalty algorithm, penalization, constraint minimization, fenchel conjugate

2010 *Mathematics Subject Classification.* 65K05, 65K10, 90C25.

1. Introduction

Let $\mathcal{H}$ be a real Hilbert space with the norm and inner product given by $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$, respectively. Let $f : \mathcal{H} \to \mathbb{R}$ and $g : \mathcal{H} \to \mathbb{R}$ be convex and (Fréchet) differentiable functions on the space $\mathcal{H}$ and the gradients ∇f and ∇g are Lipschitz continuous operators with constants L_f and L_g, respectively. We consider the following constrained convex optimization problem

$$\min_{x \in \arg\min g} f(x). \tag{1.1}$$

Throughout the paper, we also assume that the solution set $\mathcal{S} := \arg\min\{f(x) : x \in \arg\min g\}$ is a nonempty set. Further, without loss of generality, we may assume that $\min g = 0$.

Due to the interesting applications of (1.1) in many branches of mathematics and sciences, many researchers have paid attention to solve the problem (1.1) which can be mentioned briefly as follows: In 2010, Attouch and Czarnecki [1] initially presented and studied a numerical algorithm called the *multiscale asymptotic gradient* (**MAG**) for solving general constrained convex optimization problem. They proved that every sequence generated by (**MAG**) converges weakly to a solution of their

*Corresponding author. Tel.:+66 55963250; fax:+66 55963201.
Email addresses: kasamsuku@nu.ac.th (Kasamsuk Ungchittrakool), na_tta_pon@hotmail.com Natthaphon Artsawang

[1] Supported by The Royal Golden Jubilee Project Grant no. PHD/0158/2557, Thailand.

considered problem. It seems that their representation is the starting point for the development of numerical algorithms in the context of solving type of constrained convex optimization problem (see, for instance [2–4, 7–10, 18]) and the references therein. Inspired by Attouch and Czarnecki [1], in 2012 Peypouquet [20] proposed and analyzed an algorithm called *diagonal gradient scheme* (**DGS**) via gradient method and exterior penalization scheme for constrained minimization of convex functions. He also provided a weak convergence to find a solution of the considered constrained minimization of convex functions. Several applications are provided such as relaxed feasibility, mathematical programming with convex inequality constraints, and Stokes equation and signal reconstruction, etc. In 2013, Shehu et al. [21] studied the problem (1.1) in the case when the constrained set is simple enough and also proposed an algorithm for solving (1.1). In the last two decades, intensive research efforts dedicated to algorithms of inertial type and their convergence behavior can be noticed (see [6, 11, 13–17, 19]). In 2017, Bot et al. [9] considered the problem of minimizing a smooth convex objective function subject to the set of minima of another differentiable convex function. They proposed a new algorithm called *gradient-type penalty with inertial effects method* (**GPIM**) for solving the problem (1.1). They also illustrated the usability of their method via a numerical experiment for image classification via support vector machines.

In the remaining part of this section, we recall some elements of convex analysis. For a function $h : \mathcal{H} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ we denote by $\operatorname{dom} h = \{x \in \mathcal{H} : h(x) < +\infty\}$ its effective domain and say that h is proper, if $\operatorname{dom} h \neq \emptyset$ and $h(x) \neq -\infty$ for all $x \in \mathcal{H}$. The *Fenchel conjugate* of h is $h^* : \mathcal{H} \to \overline{\mathbb{R}}$, which is defined by

$$h^*(z) = \sup_{x\in\mathcal{H}}\{\langle z, x\rangle - h(x)\} \ \text{ for all } z \in \mathcal{H}.$$

The subdifferential of h at $x \in \mathcal{H}$, with $h(x) \in \mathbb{R}$, is the set

$$\partial h(x) := \{v \in \mathcal{H} : h(y) - h(x) \geq \langle v, y - x\rangle \ \forall y \in \mathcal{H}\}.$$

We take by convention $\partial h(x) := \emptyset$, if $h(x) \in \{\pm\infty\}$.

The convex and differentiable function $T : \mathcal{H} \to \mathbb{R}$ has a Lipschitz continuous gradient with Lipschitz constant $L_T > 0$, if $\|\nabla T(x) - \nabla T(y)\| \leq L_T\|x - y\|$ for all $x, y \in \mathcal{H}$.

Let $\mathcal{C} \subset \mathcal{H}$ be a nonempty closed convex set. The *indicator function* is defined as:

$$\delta_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{otherwise.} \end{cases}$$

The *support function* of $\mathcal{C}$ is defined as: $\sigma_{\mathcal{C}}(x) := \sup_{c\in\mathcal{C}}\langle x, c\rangle$ for all $x \in \mathcal{H}$. The *normal cone* $\mathcal{C}$ at a point x is

$$N_{\mathcal{C}}(x) := \begin{cases} \{\overline{x} \in \mathcal{H} : \langle \overline{x}, c - x\rangle \leq 0 \ \text{ for all } c \in \mathcal{C}\}, & \text{if } x \in \mathcal{C} \\ \emptyset, \ \text{otherwise.} \end{cases}$$

We denote by $\mathbf{Ran}(N_{\mathcal{C}})$ for the range of $N_{\mathcal{C}}$. Notice that $\delta^*_{\mathcal{C}} = \sigma_{\mathcal{C}}$. Moreover, it holds that $\overline{x} \in N_{\mathcal{C}}(x)$ if and only if $\sigma_{\mathcal{C}}(\overline{x}) = \langle \overline{x}, x\rangle$.

Inspired by the research works in this direction, we are interested in the development and improvement of the method for finding solutions of the considered problem, that is, we wish to establish the algorithm called rapid gradient penalty algorithm (**RGPA**) for solving (1.1) which is generated by a controlling sequence of scalars together with the gradient of objective and feasibility gap functions as follows:

$$(\mathbf{RGPA}) \quad \begin{cases} x_1 \in \mathcal{H}; \\ y_n = x_n - \lambda_n\nabla f(x_n) - \lambda_n\beta_n\nabla g(x_n); \\ x_{n+1} = y_n + \alpha_n(y_n - x_n) \qquad \text{for all } n \geq 1, \end{cases}$$

where $\{\lambda_n\}$ and $\{\beta_n\}$ are sequences of positive parameters and $\{\alpha_n\} \subseteq (0,1)$.

For $n \geq 1$, we write $\Omega_n := f + \beta_n g$, which is also (Fréchet) differentiable function. Therefore, $\nabla \Omega_n$ is Lipschitz continuous with constant $L_n := L_f + \beta_n L_g$. In particular, if we setting $\alpha_n = 0$ for all $n \geq 1$, the algorithm (**RGPA**) can be reduced to (**DGS**) in Peypouquet [20].

In order to support our ideas, we also provide a numerical example to simulate an event for solving problem (1.1). We also compare the time and the iteration between two algorithms including (**RGPA**) and (**DGS**).

2. The Hypotheses

In this section, we will carry out the main assumptions to prove the convergence results for rapid gradient penalty algorithm (**RGPA**). In order to prove the convergence results, the following assumptions will be proposed.

Assumption A

(I) The function f is bounded from below;

(II) There exists a positive $K > 0$ such that $\beta_{n+1} - \beta_n \leq K\lambda_{n+1}\beta_{n+1}$, $\frac{L_n}{2} - \frac{1}{2\lambda_n} \leq -K$ and $\frac{\alpha_n^2 - 1}{2\lambda_n} + (1+\alpha_n)^2 K < 0$ for all $n \geq 1$;

(III) $\{\alpha_n\} \in l^2 \setminus l^1$, $\sum_{n=1}^{\infty} \lambda_n = +\infty$ and $\liminf_{n\to\infty} \lambda_n\beta_n > 0$;

(IV) For each $p \in \mathbf{Ran}(N_{\arg\min g})$, we have $\sum_{n=1}^{\infty} \lambda_n\beta_n \left[g^*\left(\frac{p}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{p}{\beta_n}\right)\right] < +\infty.$

Remark 2.1. The conditions in **Assumption A** sparsely extend the hypotheses in [20]. The differences are given by the second and third inequality in (II), which here involves a sequence $\{\alpha_n\}$ which controls the inertial terms, and by $\{\alpha_n\} \in l^2 \setminus l^1$.

In the following remark, we present some situations where **Assumption A** is verified.

Remark 2.2. Let $K > 0$, $q \in (0,1)$, $\delta > 0$ and $\gamma \in (0, \frac{1}{3L_g})$ be any given. Then we set $\alpha_n := \frac{1}{n+1}$ for all $n \geq 1$, which implies that $\lim_{n\to\infty} \alpha_n = 0$, $\sum_{n=1}^{\infty} \alpha_n^2 < +\infty$ and $\alpha_n \leq \frac{1}{2}$ for all $n \geq 1$. We also set

$$\beta_n := \frac{3\gamma[L_f + 2(K+\delta)]}{1 - 3\gamma L_g} + \gamma K n^q \quad \text{and} \quad \lambda_n := \frac{\gamma}{\beta_n} \quad \text{for all } n \geq 1.$$

Since $\beta_n \geq \frac{3\gamma[L_f+2(K+\delta)]}{1-3\gamma L_g}$, we have for each $n \geq 1$

$$\beta_n(1 - 3\gamma L_g) \geq 3\gamma[L_f + 2(K+\delta)].$$

It follows that

$$\frac{1}{3\lambda_n} - \beta_n L_g \geq L_f + 2(K_+\delta) \quad \text{for all } n \geq 1,$$

which implies that

$$-(K+\delta) \geq \frac{L_n}{2} - \frac{1}{6\lambda_n} \quad \text{for all } n \geq 1. \tag{2.1}$$

According to (2.1), we obtain that

$$-K \geq \frac{L_n}{2} - \frac{1}{2\lambda_n} \quad \text{and} \quad \frac{1}{3} > 2\lambda_n K \quad \text{for all } n \geq 1.$$

Let us consider, for each $n \geq 1$

$$\frac{\alpha_n^2 - 1}{2\lambda_n} + (1+\alpha_n)^2 K \leq \frac{-\frac{3}{4} + \frac{9}{4} 2\lambda_n K}{2\lambda_n} < \frac{-\frac{3}{4} + \frac{3}{4}}{2\lambda_n} = 0.$$

On the other hand,

$$\beta_{n+1} - \beta_n = \gamma K[(n+1)^q - n^q] \leq \gamma K = K\lambda_{n+1}\beta_{n+1}.$$

Hence, we can conclude that **Assumption A** (II) holds.

Since $q \in (0,1)$, we obtain that $\sum_{n=1}^{\infty} \frac{1}{\beta_n} = +\infty$, so $\sum_{n=1}^{\infty} \lambda_n = +\infty$. Notice that $\lambda_n \beta_n = \gamma$ for all $n \geq 1$. It follows that $\liminf_{n\to\infty} \lambda_n \beta_n = \liminf_{n\to\infty} \gamma > 0$. Thus **Assumption A** (III) holds.

Finally, since $g^* - \sigma_{\arg\min g} \geq 0$. If $g(x) \geq \frac{k}{2} \mathrm{dist}^2(x, \arg\min g)$ where $k > 0$, then $g^*(x) - \sigma_{\arg\min g}(x) \leq \frac{1}{2k}\|x\|^2$ for all $x \in \mathcal{H}$.

Therefore, for each $p \in \mathbf{Ran}(N_{\arg\min g})$, we obtain that

$$\lambda_n \beta_n \left[g^*\left(\frac{p}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{p}{\beta_n}\right)\right] \leq \frac{\lambda_n}{2k\beta_n}\|p\|^2.$$

Thus, $\sum_{n=1}^{\infty} \lambda_n \beta_n \left[g^*\left(\frac{p}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{p}{\beta_n}\right)\right]$ converges, if $\sum_{n=1}^{\infty} \frac{\lambda_n}{\beta_n}$ converges, which is equivalently to $\sum_{n=1}^{\infty} \frac{1}{\beta_n^2}$ converges. This holds for the above choices of $\{\beta_n\}$ and $\{\lambda_n\}$ when $q \in (\frac{1}{2}, 1)$.

3. Convergence analysis for convexity

In this section, we will prove the convergence of the sequence of $\{x_n\}$ generated by (**RGPA**) and of the sequence of objective values $\{f(x_n)\}$.

We start the convergence analysis of this section with three technical lemmas.

Lemma 3.1. *Let x^* be an arbitrary element in $\mathcal{S}$ and set $p^* := -\nabla f(x^*)$. Then for each $n \geq 1$*

$$\begin{aligned} \|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 + (1+\alpha_n)\lambda_n\beta_n g(x_n) \leq (1+\alpha_n)^2 \|x_n - y_n\|^2 \\ + (1+\alpha_n)\lambda_n\beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right]. \end{aligned} \tag{3.1}$$

Proof. Applying to the first-order optimality condition, we have

$$0 \in \nabla f(x^*) + N_{\arg\min g}(x^*).$$

It follows that

$$p^* = -\nabla f(x^*) \in N_{\arg\min g}(x^*).$$

Note that for each $n \geq 1$, $\frac{x_n - y_n}{\lambda_n} - \beta_n \nabla g(x_n) = \nabla f(x_n)$.

By monotonicity of ∇f, we obtain that

$$\begin{aligned} \left\langle \frac{x_n - y_n}{\lambda_n} - \beta_n \nabla g(x_n) + p^*, x_n - x^* \right\rangle &= \langle \nabla f(x_n) - \nabla f(x^*), x_n - x^* \rangle \\ &\geq 0 \quad, \forall n \geq 1, \end{aligned}$$

and hence, for each $n \geq 1$

$$2\langle x_n - y_n, x_n - x^* \rangle \geq 2\lambda_n \beta_n \langle \nabla g(x_n), x_n - x^* \rangle - 2\lambda_n \langle p^*, x_n - x^* \rangle. \tag{3.2}$$

Since g is convex and differentiable, we have for each $n \geq 1$

$$\langle \nabla g(x_n), x^* - x_n \rangle + g(x_n) \leq g(x^*) = 0,$$

whence

$$2\lambda_n \beta_n g(x_n) \leq 2\lambda_n \beta_n \langle \nabla g(x_n), x_n - x^* \rangle. \tag{3.3}$$

On the other hand,

$$2\langle x_n - y_n, x_n - x^* \rangle = \|x_n - y_n\|^2 + \|x_n - x^*\|^2 - \|y_n - x^*\|^2. \tag{3.4}$$

Combining (3.2), (3.3) and (3.4), we get that

$$\|y_n - x^*\|^2 \leq \|x_n - y_n\|^2 + \|x_n - x^*\|^2 - 2\lambda_n \beta_n g(x_n) + 2\lambda_n \langle p^*, x_n - x^* \rangle. \tag{3.5}$$

Since $x^* \in \mathcal{S}$ and $p^* \in N_{\arg\min g}(x^*)$, we have

$$\sigma_{\arg\min g}(p^*) = \langle p^*, x^* \rangle.$$

In (3.5), we observe that

$$\begin{aligned} 2\lambda_n \langle p^*, x_n - x^* \rangle - \lambda_n \beta_n g(x_n) &= 2\lambda_n \langle p^*, x_n \rangle - \lambda_n \beta_n g(x_n) - 2\lambda_n \langle p^*, x^* \rangle \\ &= \lambda_n \beta_n \left[\left\langle \frac{2p^*}{\beta_n}, x_n \right\rangle - g(x_n) - \left\langle \frac{2p^*}{\beta_n}, x^* \right\rangle \right] \\ &\leq \lambda_n \beta_n \left[g^* \left(\frac{2p^*}{\beta_n} \right) - \sigma_{\arg\min g} \left(\frac{2p^*}{\beta_n} \right) \right]. \end{aligned} \tag{3.6}$$

Combining (3.6) and (3.5), we obtain that

$$\|y_n - x^*\|^2 \leq \|x_n - y_n\|^2 + \|x_n - x^*\|^2 - \lambda_n \beta_n g(x_n) + \lambda_n \beta_n \left[g^* \left(\frac{2p^*}{\beta_n} \right) - \sigma_{\arg\min g} \left(\frac{2p^*}{\beta_n} \right) \right]. \tag{3.7}$$

On the other hand, we observe that

$$\begin{aligned} \|x_{n+1} - x^*\|^2 &= \|y_n + \alpha_n (y_n - x_n) - x^*\|^2 = \|(1 + \alpha_n)(y_n - x^*) + \alpha_n (x^* - x_n)\|^2 \\ &= (1 + \alpha_n)\|y_n - x^*\|^2 - \alpha_n \|x_n - x^*\|^2 + \alpha_n (1 + \alpha_n)\|x_n - y_n\|^2. \end{aligned} \tag{3.8}$$

By (3.7) and (3.8), we obtain the desired result. □

Lemma 3.2. *For all $n \geq 1$, we have*

$$\begin{aligned} \Omega_{n+1}(x_{n+1}) \leq \Omega_n(x_n) + (\beta_{n+1} - \beta_n) g(x_{n+1}) + \frac{\alpha_n^2 - 1}{2\lambda_n} \|y_n - x_n\|^2 \\ + \left[\frac{L_n}{2} - \frac{1}{2\lambda_n} \right] \|x_{n+1} - x_n\|^2. \end{aligned}$$

Proof. Since $\nabla \Omega$ is L_n-Lipschitz continuous and by Descent Lemma (see [5, Theorem 18.15]), we obtain that

$$\Omega_n(x_{n+1}) \leq \Omega_n(x_n) + \langle \nabla \Omega_n(x_n), x_{n+1} - x_n \rangle + \frac{L_n}{2} \|x_{n+1} - x_n\|^2.$$

Recall that $-\frac{y_n - x_n}{\lambda_n} = \nabla \Omega_n(x_n)$.

It follows that

$$\begin{aligned}
&f(x_{n+1}) + \beta_n g(x_{n+1}) \\
&\le f(x_n) + \beta_n g(x_n) - \left\langle \frac{y_n - x_n}{\lambda_n}, x_{n+1} - x_n \right\rangle + \frac{L_n}{2}\|x_{n+1} - x_n\|^2 \\
&= f(x_n) + \beta_n g(x_n) - \frac{1}{2\lambda_n}\|y_n - x_n\|^2 - \frac{1}{2\lambda_n}\|x_{n+1} - x_n\|^2 + \frac{1}{2\lambda_n}\|y_n - x_{n+1}\|^2 + \frac{L_n}{2}\|x_{n+1} - x_n\|^2 \\
&= f(x_n) + \beta_n g(x_n) + \frac{\alpha_n^2 - 1}{2\lambda_n}\|y_n - x_n\|^2 + \left[\frac{L_n}{2} - \frac{1}{2\lambda_n}\right]\|x_{n+1} - x_n\|^2.
\end{aligned}$$

Adding $\beta_{n+1} g(x_{n+1})$ to both sides, we have

$$\begin{aligned}
f(x_{n+1}) + \beta_{n+1} g(x_{n+1}) &\le f(x_n) + \beta_n g(x_n) + (\beta_{n+1} - \beta_n) g(x_{n+1}) \\
&\quad + \frac{\alpha_n^2 - 1}{2\lambda_n}\|y_n - x_n\|^2 + \left[\frac{L_n}{2} - \frac{1}{2\lambda_n}\right]\|x_{n+1} - x_n\|^2,
\end{aligned}$$

which means that

$$\Omega_{n+1}(x_{n+1}) \le \Omega_n(x_n) + (\beta_{n+1} - \beta_n) g(x_{n+1}) + \frac{\alpha_n^2 - 1}{2\lambda_n}\|y_n - x_n\|^2 + \left[\frac{L_n}{2} - \frac{1}{2\lambda_n}\right]\|x_{n+1} - x_n\|^2.$$

□

For $n \ge 1$ and $x^* \in \mathcal{S}$, we denote by

$$\begin{aligned}
\Lambda_n &:= f(x_n) + (1 - (1 + \alpha_n) K \lambda_n) \beta_n g(x_n) + K\|x_n - x^*\|^2 \\
&= \Omega_n(x_n) - (1 + \alpha_n) K \lambda_n \beta_n g(x_n) + K\|x_n - x^*\|^2.
\end{aligned}$$

Lemma 3.3. *Let $x^* \in \mathcal{S}$ and set $p^* := -\nabla f(x^*)$. Then there exists $\theta > 0$ such that for each $n \ge 1$*

$$\Lambda_{n+1} - \Lambda_n + \theta\|y_n - x_n\|^2 \le (1 + \alpha_n) K \lambda_n \beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right].$$

Proof. From Lemma 3.2 and **Assumption A** (II), we obtain that

$$\begin{aligned}
\Omega_{n+1}(x_{n+1}) - \Omega_n(x_n) &\le K \lambda_{n+1} \beta_{n+1} g(x_{n+1}) + \frac{\alpha_n^2 - 1}{2\lambda_n}\|y_n - x_n\|^2 \\
&\le (1 + \alpha_{n+1}) K \lambda_{n+1} \beta_{n+1} g(x_{n+1}) + \frac{\alpha_n^2 - 1}{2\lambda_n}\|y_n - x_n\|^2. \qquad (3.9)
\end{aligned}$$

On the other hand, multiplying (3.1) by K, we have

$$\begin{aligned}
&K\|x_{n+1} - x^*\|^2 - K\|x_n - x^*\|^2 + (1 + \alpha_n) K \lambda_n \beta_n g(x_n) \\
&\qquad \le (1 + \alpha_n)^2 K\|x_n - y_n\|^2 + (1 + \alpha_n) K \lambda_n \beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right]. \qquad (3.10)
\end{aligned}$$

Combining (3.9) and (3.10), we have

$$\Lambda_{n+1} - \Lambda_n \le \left[\frac{\alpha_n^2 - 1}{2\lambda_n} + (1 + \alpha_n)^2 K\right]\|y_n - x_n\|^2 + (1 + \alpha_n) K \lambda_n \beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right]. \qquad (3.11)$$

For each $n \ge 1$, $\frac{\alpha_n^2 - 1}{2\lambda_n} + (1 + \alpha_n)^2 K < 0$, we have there exists $\theta > 0$ such that

$$\frac{\alpha_n^2 - 1}{2\lambda_n} + (1 + \alpha_n)^2 K < -\theta.$$

From (3.11), we have

$$\Lambda_{n+1} - \Lambda_n + \theta\|y_n - x_n\|^2 \leq (1+\alpha_n)K\lambda_n\beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right].$$

This completes the proof. □

The next lemma is an important role in convergence analysis (see in [3, Lemma 2] or [12, Lemma 3.1]).

Lemma 3.4. *Let $\{\gamma_n\}$, $\{\delta_n\}$ and $\{\varepsilon_n\}$ be real sequences. Assume that $\{\gamma_n\}$ is bounded from below, $\{\delta_n\}$ is non-negative and $\sum_{n=1}^{\infty}\varepsilon_n < +\infty$ such that*

$$\gamma_{n+1} - \gamma_n + \delta_n \leq \varepsilon_n \quad \text{for all} \quad n \geq 1.$$

Then $\lim_{n\to\infty}\gamma_n$ exists and $\sum_{n=1}^{\infty}\delta_n < +\infty$.

Lemma 3.5. *Let $x^* \in \mathcal{S}$. Then the following statements hold:*

(i) *The sequence $\{\Lambda_n\}$ is bounded from below and $\lim_{n\to\infty}\Lambda_n$ exists;*

(ii) $\sum_{n=1}^{\infty}\|y_n - x_n\|^2 < +\infty$;

(iii) $\lim_{n\to\infty}\|x_n - x^*\|^2$ *exists and* $\sum_{n=1}^{\infty}\lambda_n\beta_n g(x_n) < +\infty$;

(iv) $\lim_{n\to\infty}\Omega_n(x_n)$ *exists;*

(v) $\lim_{n\to\infty} g(x_n) = 0$ *and every weak cluster point of the sequence $\{x_n\}$ lies in* $\arg\min g$.

Proof. We set $p^* := -\nabla f(x^*)$.

(i). From **Assumption A** (II) implies $1-(1+\alpha_n)K\lambda_n \geq 0$. Since f is convex and differentiable, we have for each $n \geq 1$

$$\begin{aligned}\Lambda_n &= f(x_n) + (1-(1+\alpha_n)K\lambda_n)\beta_n g(x_n) + K\|x_n - x^*\|^2 \geq f(x_n) + K\|x_n - x^*\|^2 \\ &\geq f(x^*) + \langle \nabla f(x^*), x_n - x^*\rangle + K\|x_n - x^*\|^2 = f(x^*) - \left\langle \frac{p^*}{\sqrt{2K}}, \sqrt{2K}(x_n - x^*)\right\rangle + K\|x_n - x^*\|^2 \\ &\geq f(x^*) - \frac{\|p^*\|^2}{4K} - K\|x_n - x^*\|^2 + K\|x_n - x^*\|^2 = f(x^*) - \frac{\|p^*\|^2}{4K}.\end{aligned}$$

Therefore, $\{\Lambda_n\}$ is bounded from below.

Next, we set $\gamma_n = \Lambda_n$, $\delta_n = \theta\|y_n - x_n\|^2$ and

$$\varepsilon_n = (1+\alpha_n)K\lambda_n\beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right].$$

Recall that $\min g = 0$. Thus $g \leq \delta_{\arg\min g}$. Therefore $\sigma_{\arg\min g} = (\delta_{\arg\min g})^* \leq g^*$ and hence, $g^* - \sigma_{\arg\min g} \geq 0$. It follows that

$$\varepsilon_n = (1+\alpha_n)K\lambda_n\beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right] \leq 2K\lambda_n\beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right].$$

By using **Assumption A** (IV) and $p^* \in N_{\arg\min g}(x^*)$, we have $\sum_{n=1}^{\infty}\varepsilon_n < +\infty$. Applying Lemma 3.3 and Lemma 3.4, we obtain that $\lim_{n\to\infty}\Lambda_n$ exists.

(ii). Follows immediately from Lemmas 3.3 and 3.4.
(iii). We set $\gamma_n = \|x_n - x^*\|^2$, $\delta_n = (1+\alpha_n)\lambda_n\beta_n g(x_n)$ and

$$\varepsilon_n = (1+\alpha_n)^2\|y_n - x_n\|^2 + (1+\alpha_n)\lambda_n\beta_n \left[g^*\left(\frac{2p^*}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_n}\right)\right].$$

From statement (ii), Lemma 3.4 and Lemma 3.1, we get that

$$\lim_{n\to\infty}\|x_n - x^*\| \text{ exists and } \sum_{n=1}^{\infty}\lambda_n\beta_n g(x_n) < +\infty.$$

For (iv) since for each $n \geq 1$ $\Omega_n(x_n) = \Lambda_n + (1+\alpha_n)K\lambda_n\beta_n g(x_n) - K\|x_n - x^*\|^2$, by using (i), (iii) and $\lim_{n\to\infty}\alpha_n = 0$, we have $\lim_{n\to\infty}\Omega_n(x_n)$ exists.

(v). It follows from **Assumption A** (III) that $\liminf_{n\to\infty}\lambda_n\beta_n > 0$. According to statement (iii) implies $\lim_{n\to\infty} g(x_n) = 0$. Let $\overline{x}$ be any weak cluster point of the sequence $\{x_n\}$. Therefore, there exists subsequence $\{x_{n_k}\}$ of $\{x_n\}$ converges weakly to $\overline{x}$ as $k \to \infty$. By the weak lower semicontinuity of g, we get that

$$g(\overline{x}) \leq \liminf_{k\to\infty} g(x_{n_k}) \leq \lim_{n\to\infty} g(x_n) = 0,$$

which means that $\overline{x} \in \arg\min g$. This completes the proof. □

Lemma 3.6. *Let $x^* \in \mathcal{S}$. Then*

$$\sum_{n=1}^{\infty}\lambda_n\left[\Omega_n(x_n) - f(x^*)\right] < +\infty.$$

Proof. Since f is differentiable and convex function, we obtain that for each $n \geq 1$

$$f(x^*) \geq f(x_n) + \langle \nabla f(x_n), x^* - x_n\rangle.$$

Since g is differentiable, convex function and $\min g = 0$, we obtain that for each $n \geq 1$

$$0 = g(x^*) \geq g(x_n) + \langle \nabla g(x_n), x^* - x_n\rangle,$$

which implies that

$$0 \geq \beta_n g(x_n) + \langle \beta_n \nabla g(x_n), x^* - x_n\rangle, \quad \text{for all } n \geq 1.$$

Therefore, we can conclude that

$$f(x^*) \geq \Omega_n(x_n) + \langle \nabla\Omega_n(x_n), x^* - x_n\rangle = \Omega_n(x_n) + \left\langle \frac{x_n - y_n}{\lambda_n}, x^* - x_n\right\rangle. \tag{3.12}$$

From (3.12), we obtain that

$$2\lambda_n\left[\Omega_n(x_n) - f(x^*)\right] \leq 2\langle x_n - y_n, x_n - x^*\rangle = \|x_n - y_n\|^2 + \|x_n - x^*\|^2 - \|y_n - x^*\|^2. \tag{3.13}$$

On the other hand, for each $n \geq 1$

$$\begin{aligned}
&\|y_n - x^*\|^2 \\
&= \|x_{n+1} - \alpha_n(y_n - x_n) - x^*\|^2 = \|x_{n+1} - x^*\|^2 + \alpha_n^2\|y_n - x_n\|^2 - 2\langle \alpha_n(x_{n+1} - x^*), y_n - x_n\rangle \\
&= \|x_{n+1} - x^*\|^2 + \alpha_n^2\|y_n - x_n\|^2 - \alpha_n^2\|x_{n+1} - x^*\|^2 - \|y_n - x_n\|^2 + \|\alpha_n(x_{n+1} - x^*) - (y_n - x_n)\|^2 \\
&\geq \|x_{n+1} - x^*\|^2 + \alpha_n^2\|y_n - x_n\|^2 - \alpha_n^2\|x_{n+1} - x^*\|^2 - \|y_n - x_n\|^2,
\end{aligned}$$

which implies that

$$-\|y_n - x^*\|^2 \leq -\|x_{n+1} - x^*\|^2 - \alpha_n^2\|y_n - x_n\|^2 + \alpha_n^2\|x_{n+1} - x^*\|^2 + \|y_n - x_n\|^2. \tag{3.14}$$

Combining (3.13) and (3.14), we have for all $n \geq 1$

$$\begin{aligned} 2\lambda_n \left[\Omega_n(x_n) - f(x^*)\right] &\leq (2-\alpha_n^2)\|x_n - y_n\|^2 + \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2 + \alpha_n^2\|x_{n+1} - x^*\|^2 \\ &\leq 2\|x_n - y_n\|^2 + \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2 + \alpha_n^2\|x_{n+1} - x^*\|^2. \end{aligned}$$

Therefore, according to Lemma 3.5 (iii), we get that the sequence $\{\|x_n - x^*\|\}$ is bounded, which means that there exists $M > 0$ such that $\|x_n - x^*\| \leq M$ for all $n \geq 1$.

By **Assumption A** (III) and Lemma 3.5, we obtain that

$$2\sum_{n=1}^{\infty} \lambda_n[\Omega_n(x_n) - f(x^*)] \leq 2\sum_{n=1}^{\infty} \|y_n - x_n\|^2 + \|x_1 - x^*\|^2 + M^2 \sum_{n=1}^{\infty} \alpha_n^2 < +\infty.$$

□

The following proposition will play an important role in convergence analysis, which is the main result of this paper.

Proposition 3.7 ([5, Opial Lemma]). *Let $\mathcal{H}$ be a real Hilbert space, $\mathcal{C} \subseteq \mathcal{H}$ be nonempty set and $\{x_n\}$ be any given sequence such that:*

(i) *For every $z \in \mathcal{C}$, $\lim_{n\to\infty} \|x_n - z\|$ exists;*

(ii) *Every weak cluster point of the sequence $\{x_n\}$ lies in $\mathcal{C}$.*

Then the sequence $\{x_n\}$ converges weakly to a point in $\mathcal{C}$.

Let $\{x_n\}$ be define by (**RGPA**). Then $\{x_n\}$ converges weakly to a point in $\mathcal{S}$. Moreover, the sequence $\{f(x_n)\}$ converges to the optimal objective value of the optimization problem (1.1).

Proof. From Lemma 3.5 (iii), $\lim_{n\to\infty} \|x_n - x^*\|$ exists for all $x^* \in \mathcal{S}$. Let $\overline{x}$ be any weak cluster point of $\{x_n\}$. Then there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\{x_{n_k}\}$ converges weakly to $\overline{x}$ as $k \to \infty$. According to Lemma 3.5 (v) implies $\overline{x} \in \arg\min g$. It suffices to show that $f(\overline{x}) \leq f(x)$ for all $x \in \arg\min g$. Since $\sum_{n=1}^{\infty} \lambda_n = +\infty$, and by Lemma 3.6 and Lemma 3.5 (iv), we have

$$\lim_{n\to\infty} \Omega_n(x_n) - f(x^*) \leq 0 \quad \text{for all } x^* \in \mathcal{S}.$$

Therefore, $f(\overline{x}) \leq \liminf_{k\to\infty} f(x_{n_k}) \leq \lim_{n\to\infty} \Omega_n(x_n) \leq f(x^*), \quad \forall x^* \in \mathcal{S}$, which implies that $\overline{x} \in \mathcal{S}$. Applying Opial Lemma, we obtain that $\{x_n\}$ converges weakly to a point in $\mathcal{S}$. The last statement follows immediately from the above. □

4. Numerical experiments

In this section, we present the convergence of the algorithm proposed (**RGPA**) in this paper by the minimization problem with linear equality constraints. Firstly, we are given a linear system of the form

$$\mathbf{A}x = \mathbf{b},$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. In addition, we assume that $n > m$. In this section, the system has many solutions. This leads us to the question of which solution should be considered. As a result, we may consider the following problem, say, the minimization problem with linear equality constraints.

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|x\|^2 \\ &\text{subject to } \mathbf{A}x = \mathbf{b}, \end{aligned}$$

Table 1: Comparison of the convergence of (**RGPA**) and (**DGS**) for the parameters $K = 0.001$ and $q \in (\frac{1}{2}, 1)$.

q	(RGPA)		(DGS)	
	Time (sec)	#(Iters)	Time (sec)	#(Iters)
0.6	2.38	566	10.23	2221
0.7	2.31	568	107.78	25336
0.8	2.46	581	384.00	90636
0.9	44.96	11458	447.11	103487

or , equivalently,

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|x\|^2 \\ &\text{subject to } x \in \arg\min \frac{1}{2}\|\mathbf{A}(\cdot) - \mathbf{b}\|^2. \end{aligned}$$

The above problem can be written in the form of the problem (1.1) as

$$\begin{aligned} &\text{minimize } f(x) := \frac{1}{2}\|x\|^2 \\ &\text{subject to } g(x) := \frac{1}{2}\|\mathbf{A}(x) - \mathbf{b}\|^2. \end{aligned}$$

In this setting, we have $\nabla f(x) = x$ and notice that ∇f is 1-Lipschitz continuous.

Furthermore, we get that $\nabla g(x) = \mathbf{A}^\top(\mathbf{A}x - \mathbf{b})$ and notice that ∇g is $\|\mathbf{A}\|^2$-Lipschitz continuous.

All the numerical experiments were performed under MATLAB (R2015b). We begin with the problem by random matrix $\mathbf{A}$ in $\mathbb{R}^{m\times n}$, vector $x_1 \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ with $m = 1000$ and $n = 4000$ generated by using MATLAB command **randi**, which the entries of $\mathbf{A}$, x_1 and $\mathbf{b}$ are integer in [-10,10]. Next, we are going to compare a performance of the algorithms (**RGPA**) and (**DGS**). The choice of the parameters for the computational experiment is based on Remark 2.2. We chooses $\gamma = \frac{1}{4\|\mathbf{A}\|^2}$ and $\delta = 1$. We consider different choices of the parameters $K \in (0, 1]$ and $q \in (\frac{1}{2}, 1)$. We will terminate the algorithms (**RGPA**) and (**DGS**) when the errors become small, i.e.,

$$\|x_n - x^*\| \leq 10^{-6},$$

where $x^* = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{b}$.

In Table 1 we present a comparison of the convergence between two algorithms including (**RGPA**) and (**DGS**) for the parameters $K = 0.001$ and different choices for the parameters $q \in (\frac{1}{2}, 1)$. We observe that when $q = 0.6$ leads to lowest computation time for (**RGPA**) and (**DGS**) with 2.38 second and 10.23 second, respectively. Furthermore, we also observe that (**DGS**) hit a big number of iterations than (**RGPA**) for all choices of parameter q.

In Table 2 we present a comparison of the convergence of (**RGPA**) and (**DGS**) for the parameters $q = 0.6$ and $K \in (0, 1]$. We observe that the number of iterations and computation time for (**RGPA**) smaller than the number of iterations for (**DGS**) for each choice of parameters K. Furthermore, (**RGPA**) needs tiny computation time to reach the optimality tolerance than (**DGS**) for each choice of parameter K.

We observe that our algorithm (**RGPA**) performs an advantage behavior when comparing with algorithm (**DGS**) for all different choices of parameters. Note that the number of iterations for (**RGPA**) smaller than the number of iterations for (**DGS**). Furthermore, (**RGPA**) needs tiny computation time to reach optimality tolerance than (**DGS**) for each different choice of parameters.

5. Conclusions

We have presented a new gradient penalty scheme, say, *rapid gradient penalty algorithm* **(RGPA)**. We provide sufficient conditions to guarantee the convergences of **(RGPA)** for the considered con-

Table 2: Comparison of the convergence of (**RGPA**) and (**DGS**) for the parameters $q = 0.6$ and $K \in (0, 1]$.

K	(**RGPA**)		(**DGS**)	
	Time (sec)	#(Iters)	Time (sec)	#(Iters)
0.001	2.38	566	10.23	2221
0.005	2.40	585	171.46	40888
0.01	6.63	1612	254.93	64469
0.05	83.22	20480	288.39	65722
0.1	107.41	26257	212.02	52464
0.5	79.95	18606	100.33	24419
1	51.46	13414	67.20	16616

strained convex optimization problem (1.1). We also provide a numerical example to compare the performance of the algorithms (**RGPA**) and (**DGS**). As a result, we observe that our algorithm (**RGPA**) performs an advantage behavior when comparing with (**DGS**) for all different choices of parameters.

Acknowledgements

The second author would like to thank Naresuan University and The Thailand Research Fund for financial support. Moreover, N. Artsawang is also supported by The Royal Golden Jubilee Program under Grant PHD/0158/2557, Thailand.

Disclosure statement

The authors declare that there is no conflict of interests regarding the publication of this paper.

Funding

N. Artsawang was supported by the Thailand Research Fund through the Royal Golden Jubilee PhD Program under Grant PHD/0158/2557, Thailand.

References

[1] H. Attouch, M-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, J. Differ. Equat. **248** (2010), 1315–1344.

[2] H. Attouch, M-O. Czarnecki and J. Peypouquet, *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM J. Optim. **21** (2011), 1251–1274.

[3] H. Attouch, M-O. Czarnecki and J. Peypouquet, *Prox-penalization and splitting methods for constrained variational problems*, SIAM J. Optim. **21** (2011), 149–173.

[4] S. Banert, RI. Boţ, *Backward penalty schemes for monotone inclusion problems*, J. Optim. Theory Appl. **166** (2015), 930–948.

[5] HH. Bauschke and PL. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics. New York:Springer 2011.

[6] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imag. Sci. **2** (2009), 183–202.

[7] RI. Boţ and ER. Csetnek, *Forward-backward and Tseng's type penalty schemes for monotone inclusion problems*, Set-Valued Var. Anal. **22** (2014), 313–331.

[8] RI. Boţ and ER. Csetnek, *Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems*, J. Math. Anal. Appl. **435** (2016), 1688–1700.

[9] RI. Boţ, ER. Csetnek and N. Nimana, *Gradient-type penalty method with inertial effects for solving constrained convex optimization problems with smooth data*, Optim. Lett. **12**(1) (2018), 17–33.

[10] RI. Boţ, ER. Csetnek and N. Nimana, *An Inertial Proximal-Gradient Penalization Scheme for Constrained Convex Optimization Problems*, Vietnam J. Math. **46**(1) (2018), 53–71.

[11] C. Chen, RH. Chan, S. Ma and J. Yang, *Inertial proximal ADMM for linearly constrained separable convex optimization*, SIAM J. Imaging Sci. **8**(4) (2015), 2239–2267.

[12] PL. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, In: Butnariu, D., Censor, Y., Reich, S. (eds.) Inherently Parallel Algorithms for Feasibility and Optimization. pp. 115-152. Elsevier:Amsterdam ;(2001).

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

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

[15] A. Moudafi and M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, J. Comput. Appl. Math. **155** (2003), 447–454.

[16] Y. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k2)$, Dokl. Akad. Nauk SSSR **269**(3) (1984), 543–547.

[17] Y. Nesterov, *Gradient methods for minimizing composite objective function*, Technical report, CORE DISCUSSION PAPER (2007).

[18] N. Noun and J. Peypouquet, *Forward-backward penalty scheme for constrained convex minimization without inf-compactness*, J. Optim. Theory Appl. **158** (2013), 787–795.

[19] P. Ochs, Y. Chen, T. Brox and T. Pock, *iPiano:Inertial proximal algorithm for non-convex optimization*, SIAM J. Imaging Sci. **7**(2) (2014), 1388–1419.

[20] J. Peypouquet, *Coupling the gradient method with a general exterior penalization scheme for convex minimization*, J. Optim. Theory Appl. **153** (2012), 123–138.

[21] Y. Shehu, OS. Iyiola and CD. Enyi, *Iterative approximation of solutions for constrained convex minimization problem*, Arab. J. Math. **2**(4) (2013), 393–402.