AN OPTIMAL ALGORITHM FOR MINIMAX PROBLEMS WITH SMOOTH COMPONENTS

Pham Quy Muoi*, Chau Vinh Khanh*

The University of Danang – University of Science and Education, Vietnam

*Corresponding author: pqmuoi@ued.udn.vn; khanhcv01@gmail.com

(Received: April 16, 2024; Revised: May 14, 2024; Accepted: May 16, 2024)

Abstract - In this paper, we propose an optimal algorithm for the convex minimax. This is an extension of the Nesterov algorithm, which allows step size parameters to be non-constant and determined automatically during algorithm execution. We present the algorithm and prove the convergence of this algorithm with the optimal order. To calculate the gradient mapping, we apply the external point penalty function method. We then propose a method of determining the parameters in the algorithm automatically. The proposed new algorithm, which is integrated with the method of calculating gradient mapping and automatic parameter determination, is detailed in Algorithm 6.1. Finally, we applied the new algorithm to solve some specific examples and compared it with Nesterov’s algorithm.

Key words - Minimax optimal problem; Optimal algorithm; Nesterov’s algorithm; Convergence; Optimal convergence rate.

1. Introduction

In this paper, we deal with the minimax problem

\[
\min_{x \in Q} f(x) = \max_{1 \le i \le m} f_i(x),
\]

(1)

where \( f_i \in S_{\mu_i L_i}(\mathbb{R}^n), i \in [1, m] \) and \( Q \) is a closed convex set.

The problem has applications in many domains such as mathematics [1], statistics [2] and optimization [3]. There are some available methods to solve the problem (1) such as Mirror-Prox [4], Nesterov’s Accelerated Gradient Descent (AGD) [5] or Efficient Algorithms combining Mirror-Prox and AGD [6]. Furthermore, in [7] Nesterov introduced an optimal scheme with a constant size step

\[
h_k = \frac{1}{L}
\]

where \( L \) is the above-mentioned parameter. Note that for \( m = 1 \), there are some generalizations of Nesterov’s algorithm, one has been published in [8] for \( Q = \mathbb{R}^n \) and the other has been published in [9] for \( Q \subset \mathbb{R}^n \). In this paper, we will generalize the scheme of Nesterov to solve the problem (1) with \( m > 1 \) by allowing size steps \( h_k \) to be nonconstant. We will prove that the proposed algorithm converges with the order of the optimal convergence rate.

2. Preliminary

We first recall some notations and preliminary results of (strongly) convex differentiable functions.

A continuously differentiable function \( h \) is called convex in \( \mathbb{R}^n \) if and only if

\[
h(y) \ge h(x) + \langle h'(x), y - x \rangle, \forall x, y \in \mathbb{R}^n.
\]

A continuously differentiable function \( h \) is called strongly convex in \( \mathbb{R}^n \) if and only if there exists a constant \( \mu \ge 0 \) such that

\[
h(y) \ge h(x) + \langle h'(x), y - x \rangle + \frac{\mu}{2} \| x - y \|^2, \forall x, y \in \mathbb{R}^n.
\]

The parameter \( \mu \) is called to be an strongly convex parameter. If \( \mu = 0 \) then \( f \) is convex. We denote \( \bar{\mu} \) the largest Lipschitz strongly convex parameter.

A function \( h \) is called Lipschitz continuous differentiable if and only if it is differentiable and there exists \( L > 0 \) such that

\[
\| \nabla h(x) - \nabla h(y) \| \le L \| x - y \|, \forall x, y \in \mathbb{R}^n.
\]

Then, \( L \) is called Lipschitz constant. We denote \( L \) the smallest Lipschitz constant. Note that if \( h \) is a Lipschitz continuous differentiable function with Lipschitz constant \( L \) and convex, then

\[
h(y) \ge h(x) + \langle h'(x), y - x \rangle + \frac{L}{2} \| x - y \|^2, \forall x, y \in \mathbb{R}^n.
\]

We denote \( h \in S_{\mu}^{1,1}(\mathbb{R}^n) \) if \( h \) is a strongly convex with the strongly convex parameter \( \mu \) and \( h \in S_{\mu L}^{1,1}(\mathbb{R}^n) \) if \( h \) is a strongly convex differentiable function with the strongly convex parameter \( \mu \) and Lipschitz continuous differentiable with Lipschitz constant \( L \). Furthermore, if \( f(x) = \max_{1 \le i \le m} f_i(x) \), where \( f_i \in S_{\mu_i L_i}^{1,1}(\mathbb{R}^n), i = 1 \ldots m \), then we also write \( f \in S_{\mu, L}^{1,1}(\mathbb{R}^n) \).

Theorem 2.1 If \( f_1 \in S_{\mu_1 L_1}^{1,1}(\mathbb{R}^n) \) and \( f_2 \in S_{\mu_2 L_2}^{1,1}(\mathbb{R}^n) \) then \( \max \{ f_1, f_2 \} = f \in S_{\mu, L}^{1,1}(\mathbb{R}^n) \), where

\[
\mu = \min \{ \mu_1, \mu_2 \}, L = \max \{ L_1, L_2 \}.
\]

Proof. Let \( \mu = \min \{ \mu_1, \mu_2 \}, L = \max \{ L_1, L_2 \} \). Since \( f_1 \in S_{\mu_1 L_1}^{1,1}(\mathbb{R}^n) \) and \( f_2 \in S_{\mu_2 L_2}^{1,1}(\mathbb{R}^n) \), we have

\[
f_1(x) + \langle f'_1(x), y - x \rangle + \frac{\mu_1}{2} \| x - y \|^2 \le f_1(y)
\]

\[
\le f_1(x) + \langle f'_1(x), y - x \rangle + \frac{L_1}{2} \| x - y \|^2, \forall x, y \in \mathbb{R}^n,
\]

\[
f_2(x) + \langle f'_2(x), y - x \rangle + \frac{\mu_2}{2} \| x - y \|^2 \le f_2(y)
\]

\[
\le f_2(x) + \langle f'_2(x), y - x \rangle + \frac{L_2}{2} \| x - y \|^2, \forall x, y \in \mathbb{R}^n.\]
\[ f_2(x) + \left( f'_2(x), y - x \right) + \frac{\mu_2}{2} \| x - y \|^2 \leq f_2(y) \]
\[ \leq f_2(x) + \left( f'_2(x), y - x \right) + \frac{L_2}{2} \| x - y \|^2, \forall x, y \in \mathbb{R}^n. \]
Therefore,
\[ f_1(x) + \left( f'_1(x), y - x \right) + \frac{\mu}{2} \| x - y \|^2 \leq f_1(y) \]
\[ \leq f_1(x) + \left( f'_1(x), y - x \right) + \frac{L}{2} \| x - y \|^2, \forall x, y \in \mathbb{R}^n, \]
\[ f_2(x) + \left( f'_2(x), y - x \right) + \frac{\mu}{2} \| x - y \|^2 \leq f_2(y) \]
\[ \leq f_2(x) + \left( f'_2(x), y - x \right) + \frac{L}{2} \| x - y \|^2, \forall x, y \in \mathbb{R}^n. \]
Hence, \( f_1, f_2 \in S_{\mu,L}(\mathbb{R}^n) \). So, we have \( f \in S_{\mu,L}(\mathbb{R}^n) \).

**Definition 2.1** Let \( f \) be a max-type function:
\[ f(x) = \max_{1 \leq i \leq m} f_i(x). \tag{2} \]
Then, function
\[ f(\bar{x}; x) = \max\{f_i(\bar{x}) + \langle f'_i(\bar{x}), x - \bar{x} \rangle\} \tag{3} \]
is called the linearization of \( f(x) \) at \( \bar{x} \).

**Lemma 2.1** Let \( f(x) = \max_{1 \leq i \leq m} f_i(x) \in S_{\mu,L}(\mathbb{R}^n) \).
Then, for any \( x \in \mathbb{R}^n \) we have
\[ f(x) \geq f(\bar{x}; x) + \frac{\mu}{2} \| x - \bar{x} \|^2, \tag{4} \]
\[ f(x) \leq f(\bar{x}; x) + \frac{L}{2} \| x - \bar{x} \|^2. \tag{5} \]
**Proof.** Since \( f \in S_{\mu,L}(\mathbb{R}^n) \), we have \( f_i \in S_{\mu,L}(\mathbb{R}^n) \), \( i = 1, n \). Then,
\[ f_i(x) \geq f_i(\bar{x}) + \langle f'_i(\bar{x}), x - \bar{x} \rangle + \frac{\mu}{2} \| x - \bar{x} \|^2, \forall i = 1, n; \]
\[ f_i(x) \leq f_i(\bar{x}) + \langle f'_i(\bar{x}), x - \bar{x} \rangle + \frac{L}{2} \| x - \bar{x} \|^2, \forall i = 1, n. \]
Therefore,
\[ \max_{1 \leq i \leq m} f_i(x) \geq \max_{1 \leq i \leq m} \left[ f_i(\bar{x}) + \langle f'_i(\bar{x}), x - \bar{x} \rangle + \frac{\mu}{2} \| x - \bar{x} \|^2 \right]; \]
\[ \max_{1 \leq i \leq m} f_i(x) \leq \max_{1 \leq i \leq m} \left[ f_i(\bar{x}) + \langle f'_i(\bar{x}), x - \bar{x} \rangle + \frac{L}{2} \| x - \bar{x} \|^2 \right]. \]
Hence, we have (4), (5).

**Theorem 2.2** Let \( f_i \) be convex and differentiable for all \( i = 1, 2, \ldots, m \). A point \( x^* \in Q \) is a solution to problem (1) if and only if for any \( x \in Q \), we have
\[ f(x^*; x) \geq f(x^*, x^*) = f(x^*). \tag{6} \]

**Proof.** If (6) is true, for any \( x \in Q \), we have:
\[ f(x) \geq f(x^*; x) \geq f(x^*; x^*) = f(x^*). \]
Now, let \( x^* \) be a solution to problem (1). Assume that there exists \( x \in Q \) such that \( f(x^*; x^*) < f(x^*) \). Consider the functions
\[ \phi_0(\alpha) = f_i(x^* + \alpha(x - x^*)), i = 1, m. \]
Note that for all \( i, 1 \leq i \leq m \), we have
\[ f_i(x^*) + \langle f'_i(x^*), x - x^* \rangle < f(x^*) = \max_{1 \leq i \leq m} f_i(x). \]
Therefore, either \( \phi_0(0) = f_i(x^*) < f(x^*) \) or
\[ \phi_0(0) = f_i(x^*); \phi_0(0) = (f'_i(x^*), x - x^*) < 0. \]
Thus, for \( \alpha \) small enough we have
\[ f_i(x^* + \alpha(x - x^*)) = \phi_0(\alpha) < f(x^*), \forall i = 1, m. \]
That is a contradiction.

**Corollary 2.1** Let \( x^* \) be a minimum of a max-type function \( f(x) \) on the set \( Q \). If \( f \in S_{\mu,L}(\mathbb{R}^n) \), then for all \( x \in Q \), we have
\[ f(x) \geq f(x^*) + \frac{\mu}{2} \| x - x^* \|^2. \]

**Proof.** Indeed, in view of Lemma 2.1 and Theorem 2.1, for any \( x \in Q \) we have
\[ f(x) \geq f(x^*; x) + \frac{\mu}{2} \| x - x^* \|^2 \geq f(x^*; x^*) + \frac{\mu}{2} \| x - x^* \|^2 = f(x^*) + \frac{\mu}{2} \| x - x^* \|^2. \]

**Theorem 2.3** Let max-type function \( f(x) \) belong to \( S_{\mu}(\mathbb{R}^n) \) with \( \mu \geq 0 \) and \( Q \) be a closed convex set. Then there exists an optimal solution \( x^* \) to the problem (1). If \( \mu > 0 \), then the solution is unique.

**Proof.** Let \( \bar{x} \in Q \). Then, for any \( x \in Q \) we have
\[ f(x) \geq f(\bar{x}; x) + \frac{\mu}{2} \| x - \bar{x} \|^2. \]
Thus, \( f \) is coercive. Since \( f \) is continuous and coercive, and \( Q \) is a closed set, \( f \) has at least one minimizer \( x^* \). Furthermore, if \( \mu > 0 \) and if \( x^*_1 \) is another solution to a problem (1), then
\[ f(x^*) = f(x^*_1; x^*_1) + \frac{\mu}{2} \| x^*_1 - x^* \|^2 \geq f(x^*) + \frac{\mu}{2} \| x^*_1 - x^* \|^2. \]
This implies that \( x^*_1 = x^* \) or the minimizer is unique.
3. Gradient mapping

Let us fix some $\gamma$ and $\bar{x} \in \mathbb{R}^n$. We assume that $f(x)$ is a max-type function and denote

$$f_{\gamma}(\bar{x}; x) = f(\bar{x}; x) + \frac{\gamma}{2} \| x - \bar{x} \|^2$$

**Definition 3.1** Let $f$ be a max-type function. We define

$$f^*(\bar{x}; \gamma) = \min_{x \in Q} f_{\gamma}(\bar{x}; x),$$

$$x_{\gamma}(\bar{x}; \gamma) = \arg \min_{x \in Q} f_{\gamma}(\bar{x}; x),$$

$$g_{\gamma}(\bar{x}; \gamma) = \gamma(\bar{x} - x_{\gamma}(\bar{x}; \gamma)).$$

We call $g_{\gamma}(\bar{x}; \gamma)$ the gradient mapping of max-type function $f$ on $Q$.

Note that $\bar{x}$ does not necessarily belong to $Q$. Furthermore, it is clear that $f_{\gamma}(\bar{x}; x)$ is a max-type function composed by the components

$$f_{\gamma}(\bar{x}) + \langle f_{\gamma}(\bar{x}), x - \bar{x} \rangle + \frac{\gamma}{2} \| x - \bar{x} \|^2 \in S_{\gamma,m}^1(\mathbb{R}^n), \quad i = 1, m.$$ 

Therefore, in view of Theorem 2.2, the gradient mapping is well defined.

**Theorem 3.1** Let $f \in S_{\mu,L}^1(\mathbb{R}^n)$. Then, for all $x \in Q$, we have

$$f(x^*; x) \geq f^*(\bar{x}; \gamma) + \langle g_{\gamma}(\bar{x}; \gamma), x - \bar{x} \rangle + \frac{1}{2\gamma} \| g_{\gamma}(\bar{x}; \gamma) \|^2.$$  

**Proof.** Denote $x_{\gamma} = x_{\gamma}(\bar{x}; \gamma), g_{\gamma} = g_{\gamma}(\bar{x}; \gamma)$. Since $f_{\gamma}(\bar{x}; x) \in S_{\gamma,m}^1(\mathbb{R}^n)$ and it is a max-type function, we can apply all results of the previous section to $f_{\gamma}$. In view of Theorem 2.1 and Corollary 2.1 we have

$$f(\bar{x}; x) = f_{\gamma}(\bar{x}; x) - \frac{\gamma}{2} \| x - \bar{x} \|^2$$

$$\geq f_{\gamma}(\bar{x}; x_{\gamma}) + \frac{\gamma}{2} \| x_{\gamma} - \bar{x} \|^2 - \| x - \bar{x} \|^2 \geq f^*(\bar{x}; \gamma) + \frac{\gamma}{2} \langle \bar{x} - x_{\gamma}, x - \bar{x} \rangle$$

$$\geq f^*(\bar{x}; \gamma) + \frac{\gamma}{2} \langle \bar{x} - x_{\gamma}, 2x - x_{\gamma} - \bar{x} \rangle$$

$$= f^*(\bar{x}; \gamma) + \frac{\gamma}{2} \langle \bar{x} - x_{\gamma}, 2(x - \bar{x}) + \bar{x} - x_{\gamma} \rangle$$

$$= f^*(\bar{x}; \gamma) + \langle g_{\gamma}, x - \bar{x} \rangle + \frac{1}{2\gamma} \| g_{\gamma} \|^2.$$ 

**Corollary 3.1** Let $f \in S_{\mu,L}^1(\mathbb{R}^n)$ via $\gamma \geq L$. Then:

1. For any $x \in Q$ and $\bar{x} \in \mathbb{R}^n$ we have

$$f(x) \geq f(x_{\gamma}(\bar{x}; \gamma)) + \langle g_{\gamma}(\bar{x}; \gamma), x - \bar{x} \rangle$$

$$+ \frac{1}{2\gamma} \| g_{\gamma}(\bar{x}; \gamma) \|^2 + \frac{\mu}{2} \| x - \bar{x} \|^2.$$  

2. If $\bar{x} \in Q$ then

$$f(x_{\gamma}(\bar{x}; \gamma)) \leq f(\bar{x}) - \frac{1}{2\gamma} \| g_{\gamma}(\bar{x}; \gamma) \|^2.$$  

3. For any $\bar{x} \in \mathbb{R}^n$ we have

$$\langle g_{\gamma}(\bar{x}; \gamma) \rangle \geq \frac{1}{2\gamma} \| g_{\gamma}(\bar{x}; \gamma) \|^2 + \frac{\mu}{2} \| x - \bar{x} \|^2.$$  

**Proof.** We assume $\gamma \geq L$. Then,

$$f^*(\bar{x}; \gamma) \geq f(x_{\gamma}(\bar{x}; \gamma)).$$

Since

$$f(x) \geq f(\bar{x}; x) + \frac{\mu}{2} \| x - \bar{x} \|^2, \quad \forall x \in \mathbb{R}^n$$

and in view of Theorem 3.1, we obtain (8). From (8), choose $x = \bar{x}$, we obtain (9). Furthermore, from (8), choose $x = \bar{x}$, we obtain (10) since

$$f(x_{\gamma}(\bar{x}; \gamma)) - f(\bar{x}) \geq 0.$$

Next, let us estimate the variation of $f^*(\bar{x}; \gamma)$ as a function of $\gamma$.

**Lemma 3.1** For any $\gamma_1, \gamma_2 > 0$ and $x \in \mathbb{R}^n$ we have

$$f^*(\bar{x}; \gamma_1) + \frac{\gamma_2}{\gamma_1} \langle g_{\gamma_1}(\bar{x}; \gamma_1), x - \bar{x} \rangle$$

$$+ \frac{1}{2\gamma_1} \| g_{\gamma_1} \|^2 + \frac{\gamma_2}{2\gamma_1} \| x - \bar{x} \|^2$$

for all $x \in Q$. In particular, for $x = x_2$ we obtain

$$f^*(\bar{x}; \gamma_2) = f(\bar{x}; x_2) + \frac{\gamma_2}{2} \| x_2 - \bar{x} \|^2$$

$$\geq f^*(\bar{x}; \gamma_1) + \langle g_{\gamma_1}(x_1, x_2 - \bar{x}), x - \bar{x} \rangle$$

$$+ \frac{1}{2\gamma_1} \| g_{\gamma_1} \|^2 + \frac{\gamma_2}{2\gamma_1} \| x_2 - \bar{x} \|^2$$

$$= f^*(\bar{x}; \gamma_1) + \frac{1}{2\gamma_1} \| g_{\gamma_1} \|^2 + \frac{\gamma_2}{2\gamma_1} \| x_2 - \bar{x} \|^2$$

$$\geq f^*(\bar{x}; \gamma_1) + \frac{1}{2\gamma_1} \| g_{\gamma_1} \|^2 + \frac{1}{2\gamma_1} \| g_{\gamma_2} \|^2.$$

4. Proposed algorithm

In this section, we generalize Nesterov’s algorithm to solve the problem (1) by introducing the sequence of $\{\beta_k\}$. The proposed algorithm is presented in Algorithm 4.1.

**Algorithm 4.1** Proposed algorithm

**Input:** Let $\{\beta_k\}_{k=0}^n$ and $\{\mu_k\}_{k=0}^n$ be two sequences such that $\beta_k \geq L$ and $0 \leq \mu_k \leq \mu$ for all $k$. Choose $x_0 \in \mathbb{R}^n, \gamma_0, \text{tol} > 0$. Set $v_0 = x_0$.
1: for $k = 0, 1, 2, \ldots, n$ do

2: Compute $\alpha_k \in (0, 1)$ from equation
   \[ \beta_k \alpha_k^2 = (1 - \alpha_k) \gamma_k + \alpha_k \mu_k. \]

3: Compute $\gamma_{k+1} = \beta_k \alpha_k^2$.

4: Compute $y_k = \frac{\alpha_k y_k V_k + \gamma_{k+1} x_k}{\gamma_k + \alpha_k \mu_k}$.

5: Compute $x_{k+1} = x_f(y_k; \beta_k)$.

6: Compute
   \[ v_{k+1} = \frac{1}{\gamma_{k+1}} \left[ (1 - \alpha_k) \gamma_k v_k + \alpha_k \mu_k - \alpha_k g_f(y_k; \beta_k) \right]. \]

7: if $\left| \frac{f(x_{k+1}) - f(x_k)}{f(x_k)} \right| \leq \text{tol}$ then

8: Stop algorithm

9: end if

10: end for

Output: $\{ x_k \}$

Theorem 4.1 Let $f \in S_{\mu_k}^1(\mathbb{R}^n)$ and $\{ \alpha_k \}, \{ y_k \}$ be sequences generated by Algorithm 4. For any $x \in \mathbb{R}^n$, the pair of sequences $\{ \phi_k(x) \}_{k=0}^\infty, \{ \lambda_k \}_{k=0}^\infty$ recursively defined by:

\[ \lambda_0 = 1, \lambda_{k+1} = (1 - \alpha_k) \lambda_k, \]

\[ \phi_0(x) = f(x_0) + \frac{\gamma_0}{2} \| x - x_0 \|^2, \]

\[ \phi_k(x) = (1 - \alpha_k) \phi_k(x) + \alpha_k \left[ f(x_f(y_k; \beta_k)) + \frac{1}{2L} \| g_f(y_k; \beta_k), x - y_k \| \right], \]

Then,

(a) the function $\phi_k$ has the form

\[ \phi_k(x) = \phi_k^* + \frac{\gamma_k}{2} \| x - v_k \|^2, \]

where

\[ \phi_0^* = f(x_0), \]

\[ \phi_k^* + \frac{\alpha_k^2}{2\gamma_k+1} \| g_f(y_k; \beta_k) \|^2, \]

\[ + \alpha_k \left[ f(x_f(y_k; \beta_k)) + \frac{1}{2L} \| g_f(y_k; \beta_k) \|^2 \right] \]

\[ + \alpha_k \left[ \frac{\mu_k}{2} \| y_k - v_k \|^2 + (g_f(y_k; \beta_k), v_k - y_k) \right], \]

(b) the sequence $\{ x_k \}$ satisfies $\phi_k^* \geq f(x_k)$ for all $k \in \mathbb{N}$.

Proof. (a) Note that $\phi_k^*(x) = \gamma_0 I_n$. Let us prove that $\phi_k^*(x) = \gamma_k I$ for all $k \geq 0$. Indeed, if that is true for some $k$, then

\[ \phi_{k+1}^*(x) = (1 - \alpha_k) \phi_k^*(x) + \alpha_k \mu_k I_n = \left( (1 - \alpha_k) \gamma_k + \alpha_k \mu_k \right) I_n = \gamma_{k+1} I_n. \]

This justifies the canonical form of the function $\phi_k(x)$.

Further,

\[ \phi_k^*(x) = (1 - \alpha_k) \left[ \frac{\gamma_k}{2} \| x - v_k \|^2 \right] \]

\[ + \alpha_k \left[ f(x_f(y_k; \beta_k)) + \frac{1}{2L} \| g_f(y_k; \beta_k) \|^2 \right] \]

\[ + (g_f(y_k; \beta_k), x - y_k) + \frac{\mu_k}{2} \| x - y_k \|^2 \]

By the first-order optimality condition for function $\phi_{k+1}(x)$, the equation $\phi_{k+1}'(x) = 0$, looks as follows:

\[ (1 - \alpha_k) \gamma_k (x - v_k) + \alpha_k \left[ g_f(y_k; \beta_k) + \mu_k (x - y_k) \right] = 0. \]

The solution of this equation is $v_{k+1}$ given in Step 6 of Algorithm 4.1, which is the minimum of the function $\phi_{k+1}(x)$.

Finally, let us compute $\phi_k^*$. In view of the recursion rule for the sequence $\{ \phi_k(x) \}$, we have

\[ \phi_k^* + \frac{\gamma_k+1}{2} \| y_k - v_{k+1} \|^2 = \phi_{k+1}(y_k) \]

\[ = (1 - \alpha_k) \left[ \phi_k^* + \frac{\gamma_k}{2} \| y_k - v_k \|^2 \right] \]

\[ + \alpha_k \left[ f(x_f(y_k; \beta_k)) + \frac{1}{2L} \| g_f(y_k; \beta_k) \|^2 \right]. \]

From Step 6 of Algorithm 4.1, we have

\[ v_{k+1} - y_k = \frac{1}{\gamma_{k+1}} \left[ (1 - \alpha_k) \gamma_k (v_k - y_k) - \alpha_k g_f(y_k; \beta_k) \right]. \]

Thus,

\[ \frac{\gamma_{k+1}}{2} \| v_{k+1} - y_k \|^2 = \frac{1}{2\gamma_{k+1}} \left[ (1 - \alpha_k)^2 \gamma_k^2 \| v_k - y_k \|^2 \right] \]

\[ - 2\alpha_k (1 - \alpha_k) \gamma_k \langle v_k - y_k, g_f(y_k; \beta_k) \rangle \]

\[ + \alpha_k^2 \| g_f(y_k; \beta_k) \|^2 \]

Therefore,

\[ \phi_{k+1}^* + \frac{1}{2\gamma_{k+1}} \left[ (1 - \alpha_k)^2 \gamma_k^2 \| v_k - y_k \|^2 \right] \]

\[ - 2\alpha_k (1 - \alpha_k) \gamma_k \langle v_k - y_k, g_f(y_k; \beta_k) \rangle + \alpha_k^2 \| g_f(y_k; \beta_k) \|^2 \]
\[
= (1 - \alpha_k)(\phi_k^* + \frac{\gamma_k}{2} \| y_k - v_k \|_2^2 )
\] + \alpha_k \left[ f(x_f(y_k; \beta_k)) + \frac{1}{2L} \left\| g_f(y_k; \beta_k) \right\|_2^2 \right]
\]

Note that \( \gamma_{k+1} = (1 - \alpha_k) \gamma_k + \alpha_k \mu_k \), we have the form of \( \phi_{k+1}^* \).

(b) We prove \( \phi_{n}^* \geq f(x_n) \) for all \( n \in \mathbb{N} \). At \( n = 0 \), \( \phi_0(x) = f(x_0) + \frac{\gamma_0}{2} \| x - v_0 \|_2^2 \), thus the statement is true for \( n = 0 \). Suppose that \( \phi_{n}^* \geq f(x_n) \) is true at \( n = k \geq 0 \), we need to prove that the inequality is still true at \( n = k + 1 \). We have
\[
\phi_{k+1}^* \geq (1 - \alpha_k) f(x_k) + \alpha_k f \left( x_f(y_k; \beta_k) \right) + \frac{1}{2L} \left\| g_f(y_k; \beta_k) \right\|_2^2
\]
\[
+ \left[ \frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}} \right] \left\| g_f(y_k; \beta_k) \right\|_2^2
\]
\[
+ \alpha_k \left( 1 - \alpha_k \gamma_k \right) \frac{\mu_k}{\gamma_{k+1}} \left\| y_k - v_k \right\|_2^2 + (g_f(y_k; \beta_k), v_k - y_k) \right]
\]

Using the inequality (8) with \( x = x_k, \gamma = y_k \), we have:
\[
f(x_k) \geq f(x_f(y_k; \beta_k)) + (g_f(y_k; \beta_k), x_k - y_k)
\]
\[
+ \frac{1}{2L} \left\| g_f(y_k; \beta_k) \right\|_2^2 + \frac{\mu}{2L} \left\| x_k - y_k \right\|_2^2.
\]

Therefore,
\[
\phi_{k+1}^* \geq (1 - \alpha_k) f(x_k) + \alpha_k f \left( x_f(y_k; \beta_k) \right)
\]
\[
+ \left[ \frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}} \right] \left\| g_f(y_k; \beta_k) \right\|_2^2
\]
\[
+ \alpha_k \left( 1 - \alpha_k \gamma_k \right) \frac{\mu_k}{\gamma_{k+1}} \left\| y_k - v_k \right\|_2^2 + (g_f(y_k; \beta_k), v_k - y_k) \right].
\]

We also have \( x_{k+1} = x_f(y_k; \beta_k), \gamma_{k+1} = \beta_k \alpha_k^2, \alpha_k \gamma_k \left( v_k - y_k \right) + x_k - y_k = 0 \). Therefore,
\[
\phi_{k+1}^* \geq f(x_{k+1}) + \left[ \frac{\alpha_k}{2L} - \frac{1}{2\beta_k} \right] \left\| g_f(y_k; \beta_k) \right\|_2^2
\]

Since \( \beta_k \geq L \), we obtain \( \phi_{k+1}^* \geq f(x_{k+1}) \).

(c) We have
\[
f(x_k) \leq \phi_k^* = \min_{x \in \mathbb{R}^n} f(x) \leq \min_{x \in \mathbb{R}^n} [ (1 - \lambda_k) f(x) + \lambda_k \phi_k(x) ]
\]
\[
\leq (1 - \lambda_k) f(x^*) + \lambda_k \phi_k(x^*).
\]

**Theorem 4.2** The Algorithm 4.1 generates a sequence \( \{ x_k \}_{k=0}^{\infty} \) such that
\[
f(x_k) - f^* \leq \lambda_k \left[ f(x_0) - f^* + \frac{\gamma_0}{2} \| x_0 - x^* \|_2^2 \right].
\]

where \( \lambda_0 = 1 \) and \( \lambda_k = \prod_{i=0}^{k-1} (1 - \alpha_i) \).

**Proof.** Since \( \phi_0(x) = f(x_0) + \frac{\gamma_0}{2} \| x - x_0 \|_2^2 \) and in view of Theorem 4.1, we obtain
\[
f(x_k) - f^* \leq \lambda_k \left[ \phi_k(x^*) - f^* \right]
\]
\[
\leq \lambda_k \left[ f(x_0) - f^* + \frac{\gamma_0}{2} \| x_0 - x^* \|_2^2 \right].
\]

**Lemma 4.1** If in Algorithm 4.1, \( 0 \leq \mu_0 \leq \gamma_0 \) and \( \mu \leq \mu_k, L \leq \beta_k \leq \beta \) for all \( k \geq 0 \), then
\[
\lambda_k \leq \min \left\{ \left( \frac{1 - \mu}{\sqrt{\beta}} \right)^k, \frac{4\beta}{(2\sqrt{\beta} + k\sqrt{\gamma}_0)^2} \right\}.
\]

**Proof.** We prove that \( \gamma_k \geq \mu_k \) for all \( k > 0 \). It is clear that the inequality is true for \( k = 0 \). Now, we suppose that \( \gamma_{k} \geq \mu_k \) for some \( k \geq 0 \). Then
\[
\gamma_{k+1} = \beta_k \alpha_k^2 = (1 - \alpha_k) \gamma_k + \alpha_k \mu_k \geq \mu_k.
\]

Hence, \( \alpha_k \geq \frac{\mu_k}{\beta_k} \geq \frac{\mu}{\beta} \).

Therefore, \( \lambda_k = \prod_{i=0}^{k-1} (1 - \alpha_i) \leq \left( 1 - \frac{\mu}{\beta} \right)^k \).

Further, let us prove that \( \gamma_k \geq \gamma_0 \lambda_k \). It is clear that the inequality is true for \( k = 0 \). Assume that the inequality is true for some \( k = m \), i.e., \( \gamma_m \geq \gamma_0 \lambda_m \). Then,
\[
\gamma_{m+1} = (1 - \alpha_m) \gamma_m + \alpha_m \mu_m \geq (1 - \alpha_m) \gamma_0 \lambda_m + \alpha_m \mu_k \geq \gamma_0 \lambda_{m+1}.
\]

Therefore, we obtain \( \beta_k \alpha_k^2 = \gamma_{k+1} \geq \gamma_0 \lambda_{k+1} \) for all \( k \in \mathbb{N} \).

Let \( a_k = \frac{1}{\sqrt{\lambda_k}} \). Since \( \{ \lambda_k \} \) is a decreasing sequence, we have
\[
a_{k+1} - a_k = \frac{1}{\sqrt{\lambda_{k+1}}} - \frac{1}{\sqrt{\lambda_k}} \geq \frac{\sqrt{\lambda_k} - \sqrt{\lambda_{k+1}}}{\sqrt{\lambda_k} \lambda_{k+1}} \]
\[
= \frac{\lambda_k - \lambda_{k+1}}{\sqrt{\lambda_k} \lambda_{k+1} (\sqrt{\lambda_k} + \sqrt{\lambda_{k+1}})} \]
\[
= \frac{\lambda_k - \lambda_{k+1}}{2\lambda_k \sqrt{\lambda_{k+1}}} \leq \frac{\alpha_k}{2\lambda_k \sqrt{\lambda_{k+1}}} = \frac{\alpha_k}{2\sqrt{\lambda_{k+1}}}.
\]
Using $\beta_k \alpha_k^2 = \gamma_{k+1} \geq \gamma_0 \lambda_{k+1}$, we have

$$a_{k+1} - a_k \geq \frac{\alpha_k}{2 \beta_{k+1}} \geq \frac{\gamma_0 \lambda_{k+1}}{2 \beta_{k+1}} = \frac{1}{2} \sqrt{\gamma_0}.$$

Thus, $a_k \geq 1 + k \frac{\sqrt{\gamma_0}}{\beta}$ and the lemma is proved.

**Theorem 4.3** Let the max-type function $f$ belong to $\mathcal{S}_{\mu,L}(\mathbb{R}^n)$. If in the algorithm we take $0 \leq \mu_0 \leq \gamma_0$ and $0 < \mu \leq \mu_0, L \leq \beta \leq \beta_0$ for all $k \geq 0$, then

$$f(x_k) - f^* \leq \frac{\beta + \gamma_0}{2} \min \left\{ \left(1 - \sqrt{\frac{\mu}{\beta}} \right)^k, \frac{4\beta}{2(\beta + k \sqrt{\gamma_0})^2} \right\} \left\| x_0 - x^* \right\|^2.$$

**Proof.** Assume that the function $f(x)$ is composed by components $f_i(x), i = 1, \ldots, m$. By Lemma 4.3, Theorem 4.2 and the fact $(f_i(x^*), x - x^*) \geq 0$ for all $x \in \mathbb{R}^n$, $i = 1, m$, we have

$$f_i(x_k) - f_i^* \leq \lambda_k \left[ f_i(x_0) - f_i^* + \frac{\gamma_0}{2} \left\| x_0 - x^* \right\|^2 \right]$$

$$\leq \lambda \left[ f_i(x_0) - f_i(x^*) + \frac{\gamma_0}{2} \left\| x_0 - x^* \right\|^2 \right]$$

$$\leq \lambda \left[ \frac{\beta_k}{2} \left\| x_0 - x^* \right\|^2 + \frac{\gamma_0}{2} \left\| x_0 - x^* \right\|^2 + 2 \left\langle f_i(x^*), x_0 - x^* \right\rangle \right]$$

$$= \lambda \left[ \frac{\beta_k + \gamma_0}{2} \left\| x_0 - x^* \right\|^2 - 2 \left\langle f_i(x_0) - f_i(x^*), x_0 - x^* \right\rangle \right]$$

$$\leq \frac{\lambda_k (\beta_k + \gamma_0)}{2} \left\| x_0 - x^* \right\|^2.$$

Note that in the third inequality, we have used the inequalities $\beta_k \geq L$ and inequality (2). Therefore,

$$f(x_k) - f^* \leq \frac{\lambda_k (\beta_k + \gamma_0)}{2} \left\| x_0 - x^* \right\|^2.$$

From the last inequality and Lemma 3.1, we have

$$f(x_k) - f^* \leq \frac{\beta + \gamma_0}{2} \min \left\{ \left(1 - \sqrt{\frac{\mu}{\beta}} \right)^k, \frac{4\beta}{2(\beta + k \sqrt{\gamma_0})^2} \right\} \left\| x_0 - x^* \right\|^2.$$

5. Computing gradient mapping

Recall, this problem of computing gradient mapping is as follows:

$$\min_{x \in \mathcal{Q}} \left\{ f(x_0, x) + \frac{\gamma}{2} \left\| x - x_0 \right\|^2 \right\}.$$

Introducing the additional variables $t \in \mathbb{R}$, we can rewrite this problem in the following way:

$$\min \left\{ t + \frac{\gamma}{2} \left\| x - x_0 \right\|^2 \right\}$$

such that

$$f_i(x_0) + \left\langle f_i(x_0), x - x_0 \right\rangle \leq t, i = 1, m,$$

$$x \in \mathcal{Q}, t \in \mathbb{R}.$$

**Lemma 5.1** Two problems (16) and (17) are equivalent. It means that if $x^*$ is a solution to (16) then $(x^*, t^*)$ where $t^* = f(x_0, x^*)$ is a solution to (17) and vice versa, if $(x^*, t^*)$ is a solution to (17) then $x^*$ is a solution to (16).

**Proof.** First, we assume that $x^*$ is a solution to (16) and $t^* = f(x_0, x^*)$. Then,

$$f_i(x_0) + \left\langle f_i(x_0), x - x_0 \right\rangle \leq t^*, i = 1, m.$$

Furthermore,

$$f(x_0, x) + \frac{\gamma}{2} \left\| x - x_0 \right\|^2 \geq f(x_0, x^*) + \frac{\gamma}{2} \left\| x^* - x_0 \right\|^2$$

$$= t^* + \frac{\gamma}{2} \left\| x^* - x_0 \right\|^2.$$

Hence, $(x^*, t^*)$ is a solution to (17).

Next, we assume that $(x^*, t^*)$ is a solution to (17). Then,

$$f_i(x_0) + \left\langle f_i(x_0), x - x_0 \right\rangle \leq t^* \quad \text{for every } i = 1, m.$$

Since $(x, t) = (x, f(x_0; x))$ is a feasible point of (17), we have

$$f(x_0; x) + \frac{\gamma}{2} \left\| x - x_0 \right\|^2 \geq t^* + \frac{\gamma}{2} \left\| x^* - x_0 \right\|^2$$

$$\geq f(x_0; x^*) + \frac{\gamma}{2} \left\| x^* - x_0 \right\|^2.$$

It points out that $x^*$ is a solution to (16).

Note that the problem (17) is a specific case of the following minimization problem

$$\begin{aligned}
\min & f(x) \\
\text{subject to} & g(x) \leq 0 \\
& x \in \mathcal{Q} \subset \mathbb{R}^n
\end{aligned}$$

where $g(x) = (g_1(x), \ldots, g_m(x))$ and $f, g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$.

To solve this problem, we apply the exterior penalty function method [2]. First, we construct the penalty function

$$p(x) = \sum_{i=1}^m \phi_i(g_i(x))$$

where $\phi$ is a continuous function on $\mathbb{R}$ satisfied $\phi(y) = 0, \forall y \leq 0; \phi(y) > 0, \forall y > 0$. Such a function used in
our paper is $\phi(y) = (\max(0, y))^2$. Then, the algorithm for the exterior penalty function method is presented in Algorithm 5.1. We denote this algorithm as a function $\text{Alg}(x_i, \mu_i, \beta, tol) = x_i$, where $(x_i, \mu_i, \beta, tol)$ is the input of the algorithm and $x_i$ is its output.

**Algorithm 5.1 Exterior penalty function algorithm**

**Input:** Let tol > 0 and $x_i \in \mathbb{Q}, \mu_i > 0, \beta > 1$

1: for $k = 0, 1, 2, ..., n$ do
2: Start with $x_i$, find the solution $x_{k+1}$ to the problem $\min_{x \in \mathbb{Q}} [f(x) + \mu_k p(x)]$
3: if $\mu_k p(x_{k+1}) < tol$ then stop. Otherwise, set $\mu_{k+1} = \beta \mu_k$
4: end for

**Output:** $\{x_k\}$

For problem (17), we have

$$x := (x,t), f := f_j(x,t) = t + \frac{\gamma}{2} \|x - x_0\|^2, \quad g_i := f_i(x_0) + \left( f_i'(x_0), x - x_0 \right) - t,$$

$$p(x,t) = \sum_{i=1}^{m} \left( \max(0, f_i(x_0) + (f_i'(x_0), x - x_0) - t) \right)^2$$

Since the function $f(x,t) + \mu_k p(x,t)$ is convex, we can deal with this problem by many different methods such as the projected gradient method, the projected Newton method and the projected Quasi-Newton method.

**6. Detailed proposed algorithm**

From Theorem 4.3, the proposed algorithm has the best convergence rate when $\beta_k = \tilde{L}$ and $\mu_k = \tilde{\mu}$ for all $k$. In many situations, the parameters $\tilde{L}$ and $\tilde{\mu}$ are not available. To overcome this situation, we propose a way to compute the sequence $\{\beta_k\}$ and $\{\mu_k\}$ such that they respectively converge (close) to these parameters automatically. The detailed algorithm is presented in Algorithm 6.

**Algorithm 6.1 Detailed proposed algorithm**

**Input:** Choose $x_0 \in \mathbb{R}^n, tol > 0, \text{eps} > 0, \text{Maxiter} > 0; \eta > 1.$

1: $\tau_0 = 0$
2: while $\tau_0 < \text{eps}$ do
3: $y_0 = x_0 + \text{rand(size}(x_0))$
4: $\tau_0 = \max_i \left( \frac{f_i'(x_0) - f_i'(y_0)}{\tau_0} \right)^2 (x_0 - y_0)$
5: end while
6: $\beta_0 = \max_i \frac{\|f_i'(x_0) - f_i'(y_0)\|^2}{\tau_0}$

7: $\mu_0 = \frac{\tau_0}{\|x_0 - y_0\|^2}$
8: Set $v_0 = x_0, \gamma'_0 = \mu_0$.
9: for $k = 0, 1, 2, ..., \text{Maxiter}$ do
10: Compute $\alpha_k \in (0, 1)$ from equation $\beta_k \alpha_k^2 = (1 - \alpha_k) \gamma'_k + \alpha_k \mu_k$.
11: Compute $\gamma_{k+1} = \beta_k \alpha_k^2$.
12: Compute $y_k = \alpha_k \gamma_k v_k + \gamma_{k+1} x_k$.
13: Compute $t_k = f(y_k : x_k)$ and
$x_{k+1} = \text{Alg}((x_i : t_i), \mu_0, \beta_k, tol)$
14: Compute $v_{k+1} = \frac{1}{\gamma_{k+1}} \left[ (1 - \alpha_k) \gamma'_k v_k + \alpha_k \mu y_k - \alpha_k g_f(y_k; \beta_k) \right]$.
15: Compute $\tau_k = \max_i \left( f_i'(x_{k+1}) - f_i'(x_k), x_{k+1} - x_k \right)$
16: if $\tau_k >= \text{eps}$ then
17: Compute $\eta_k = \max_i \frac{\|f_i'(x_{k+1}) - f_i'(x_k)\|^2}{\tau_k}$
and $\zeta_k = \frac{\tau_k}{\|x_{k+1} - x_k\|^2}$
18: else
19: Set $\eta_k = \beta_k$ and $\zeta_k = \mu_k$
20: end if
21: if $\eta_k \geq \beta_k$ then
22: $\beta_{k+1} = \eta \eta_k$
23: else
24: $\beta_{k+1} = \beta_k$
25: end if
26: if $\mu_k \geq \zeta_k$ then
27: $\mu_{k+1} = \zeta_k / \eta$
28: else
29: $\mu_{k+1} = \mu_k$
30: end if
31: if $\|f(x_{k+1}) - f(x_k)\| / \|f(x_k)\| \leq \text{tol}$ then
32: Stop
33: end if
34: end for

**Output:** $\{x_k\}$

**7. Numerical examples**

In this section, we apply our algorithm to find a numerical approximation to the solution in some specific examples.
Example 1: Let \( f_1(x) = x^2, f_2(x) = (x - 2)^2 \). Solve the problem:
\[
\min_{x \in \mathbb{R}} \left[ f(x) = \max \{ f_1(x), f_2(x) \} \right].
\]
Note that the exact solution of this problem is \( x^* = 1 \). Here, \( f_1 \in S^1_{2,2}(\mathbb{R}) \) and \( f_2 \in S^1_{2,2}(\mathbb{R}) \) so \( f \in S^1_{2,2}(\mathbb{R}) \).
In Nesterov’s algorithm, we set \( x_0 = v_0 = 4 \), \( \gamma_0 = 2, \beta_k = 2 \) for all \( k \) and \( tol = 10^{-6} \). In Algorithm 6.1, we set \( x_0 = v_0 = 4, \eta = 1.3, tol = 10^{-6} \). After some iterations, two algorithms converge. Their convergence rates are comparable.

Figure 1. The objective values, \( F(x_k) \), in Nesterov’s algorithm and the proposed algorithm with respect to the number of iterations in Example 1.

Now, we not only broaden the dimension of vector \( x \) but also the number of components of function \( f \).

Example 2: Solve the problem:
\[
\min_{x \in \mathbb{R}^4} \left[ f(x) = \max \{ f_1(x), f_2(x), f_3(x), f_4(x) \} \right],
\]
Where
\[
\begin{align*}
    f_1(x) &= x_1^2 + x_2^2 + x_3^2 + x_4^2, \\
    f_2(x) &= (x_1 - 2)^2 + (x_2 - 1)^2 + (x_3 - 1)^2 + (x_4 - 1)^2, \\
    f_3(x) &= (x_1 - 1)^2 + (x_2 - 2)^2 + (x_3 - 2)^2 + (x_4 - 1)^2, \\
    f_4(x) &= x_1^2 + (x_2 - 2)^2 + (x_3 - 1)^2 + (x_4 - 1)^2.
\end{align*}
\]
Here, \( f_1, f_2, f_3, f_4 \in S^1_{2,2}(\mathbb{R}) \) so \( f \in S^1_{2,2}(\mathbb{R}) \). In Nesterov’s algorithm, we set \( x_0 = v_0 = (4; 4; 4; 4), \gamma_0 = 2, \beta_k = 2 \) for all \( k \) and \( tol = 10^{-6} \). In Algorithm 6.1, we set \( x_0 = v_0 = (4; 4; 4; 4), \eta = 1.3, tol = 10^{-6} \). After some iterations, two algorithms converge and their convergence rates are almost the same.

Figure 2. The objective values, \( F(x_k) \), in Nesterov’s algorithm and the proposed algorithm with respect to the number of iterations in Example 2.

8. Conclusion

In this paper, we have presented the detailed proposed algorithm, Algorithm 6.1, for the minimax problem and prove its optimal convergence rate in Theorem 4.1. Our algorithm is a generalization of Nesterov’s algorithm in [7], when step size parameters are non-constants and determined automatically during algorithm execution. We also applied the new algorithm to solve some specific examples and compared it with Nesterov’s algorithm in Section 7. However, we can see in Example 2, Nesterov Algorithm’s convergence rate is still faster than Optimal Algorithm. So, it raises a new question if we can update for parameters in Algorithm 6.1 such that it converges faster than Nesterov’s algorithm. It is still an open question and motivates us to study in the future.

REFERENCES