GENERALIZED NESTEROV'S ALGORITHM FOR CONSTRAINED MINIMIZATION PROBLEMS ON CLOSED CONVEX SETS

Pham Quy Muoi*, Vo Quang Duy, Phan Duc Tuan

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

*Corresponding author: pqmuoi@ued.udn.vn

(Received: September 14, 2023; Revised: October 13, 2023; Accepted: October 16, 2023)

Abstract - In this paper, we propose a generalized Nesterov algorithm for the constrained optimization problems on a closed convex set. We prove the convergence as well as the convergence rate of the proposed algorithm. First, we present a new algorithm based on the generalization of Nesterov's algorithm. Then, we prove the convergence as well as the convergence rate of the new algorithm. With a specific choice of parameters, the new algorithm becomes Nesterov's algorithm. Therefore, the convergence as well as the convergence rate of Nesterov's algorithm are also followed. We illustrate the effectiveness of the new algorithm as well as compare it with Nesterov's algorithm and the gradient descent algorithm through a specific example.

Key words - Constrained minimization problem; Closed convex set; Nesterov's algorithm; Convergence; Convergence rate.

1. Introduction

In this paper, we consider a constrained minimization problem

$$\min_{x\in Q} f(x), \tag{1}$$

Where, f is a strongly convex function on $Q \subset \mathbb{R}^n$ with the derivative f' being Lipschitz continuous, Q is a closed convex set. We also denote x^* and f^* as a solution and the minimum of problem (1), respectively.

Constrained minimization problems on simple closed convex sets are a fundamental class of optimization problems with widespread applications in various fields, including engineering, economics, machine learning, and many others. The goal is to find the optimal value of a function subject to constraints that limit the search space to a simple closed convex set.

Several algorithms have been developed to tackle such problems efficiently such that projected gradient descent [1], interior point methods [2, 3, 4], sequential quadratic programming [5, 6, 7], penalty and augmented Lagrangian methods [8, 9] and trust-region methods [10, 11, 12].

The choice of an algorithm depends on various factors such as the problem's characteristics (convex or non-convex), the dimensionality of the variables, the type and complexity of the constraints, and the desired trade-off between computational efficiency and accuracy.

Recently, Nesterov proposed an optimal algorithm for problem (1), which has the optimal order of convergence rate among all algorithms only use values of the objective functional and its gradient [13]. The order of convergence rate of the method is of $O(q^k)$ with $q \in (0,1)$ if problem (1) is strongly convex and is of $O(\frac{1}{k^2})$ if problem (1) is convex. However, the proof of convergence and convergence rate of the method is not available in detail. Thus, in this paper, we will generalize Nesterov's algorithm and prove its convergence and convergence rate. Note that with special choice of parameters, the generalized Nesterov's algorithm returns the original one. Thus, Nesterov's algorithm is proved as well. Note that our proposed algorithm for problem (1) with $Q = \mathbb{R}^n$ (unconstained minimization problem) return to the generalized Nestrov's algorithm that has investigated in [14].

2. Notations and preliminary results

In this part, we recall some technical terms and properties of strongly convex differentiable function. We denote $\mathcal{S}^1(\mathbb{R}^n)$ is the set of all differentiable, convex function in $\mathbb{R}^n$ and $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ is the set of all differentiable, strongly convex function and its derivative f' is Lipschitz continuous with Lipschitz constant L in $\mathbb{R}^n$.

A continuously differentiable function f is called *convex* in $\mathbb{R}^n$ if and ony if

$$f(y) \ge f(x) + \langle f'(x), y - x\rangle, \qquad \forall x, y \in \mathbb{R}^n.$$

A continuously differentiable function f is called *strongly convex* in $\mathbb{R}^n$ if and ony if there exists a constant $\mu \ge 0$ such that

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

A differentiable function f is Lipschitz continuous on $\mathbb{R}^n$ if and only if there exists $L > 0$ such that

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

Then, μ and L are respectively called strongly convexity constant and Lipschitz constant.

Note that if f is convex and Lipschitz continuously differentiable in $\mathbb{R}^n$, then for any $x, y \in \mathbb{R}^n$,

$$0 \le f(y) - f(x) - \langle f'(x), y - x\rangle \le \tfrac{L}{2}\|x - y\|^2, \tag{2}$$

and

$$0 \le \langle f'(x) - f'(y), x - y\rangle \le L\|x - y\|^2, \forall x, y \in \mathbb{R}^n. \tag{3}$$

The projection operator is the distance function from $y \in \mathbb{R}^n$ to the closed convex set Q, defined by

$$P_Q(y) = \operatorname*{argmin}_{x\in Q} \tfrac{1}{2}\|x - y\|^2. \tag{4}$$

Note that $P_Q(y) = y$ if and only if $y \in Q$.

Let us fix some $\gamma > 0$ and $y \in \mathbb{R}^n$. We define $x_Q(y;\gamma)$ and $g_Q(y;\gamma)$ by

$$x_Q(y;\gamma) = argmin_{x\in Q}\left[f(y)+\langle f'(y),x-y\rangle+\frac{\gamma}{2}\|x-y\|^2\right] = P_Q\left(y-\frac{1}{\gamma}f'(y)\right),$$

and $g_Q(y;\gamma)=\gamma\big(y-x_Q(y;\gamma)\big)$.

Theorem 2.1 *Let $f\in\mathcal{S}^1(\mathbb{R}^n)$ and Q be a closed convex set in $\mathbb{R}^n$. The point x^* is a solution of problem (1) if and only if*

$$\langle f'(x^*),x-x^*\rangle\ge 0,\forall x\in Q. \quad (5)$$

Proof. If (5) is true, then

$$f(x)\ge f(x^*)+\langle f'(x^*),x-x^*\rangle\ge f(x^*),$$

for all $x\in Q$. Therefore, x^* is the solution of problem (1).

Now let x^* be a solution of problem (1). Assume that there exists some $x\in Q$ such that

$$\langle f'(x^*),x-x^*\rangle<0.$$

Consider the function $\varphi(\alpha)=f\big(x^*+\alpha(x-x^*)\big)$, $\alpha\in[0;1]$. Since $\varphi(0)=f(x^*),\varphi'(0)=\langle f'(x^*), x-x^*\rangle<0$, $f\big(x^*+\alpha(x-x^*)\big)=\varphi(\alpha)<\varphi(0)=f(x^*)$ for α small enough. That is a contradiction. Thus, the theorem is proved.

Lemma 2.1 *Let $f\in\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n),\gamma\ge L$ and $y\in\mathbb{R}^n$. Then for any $x\in\mathbb{R}^n$ we have*

$$f(x)\ge f\left(x_Q(y;\gamma)\right)+\frac{1}{2\gamma}\left\|g_Q(y;\gamma)\right\|^2 +\langle g_Q(y;\gamma),x-y\rangle+\frac{\mu}{2}\|x-y\|^2. \quad (6)$$

Proof. Denote $x_Q=x_Q(y,\gamma), g_Q=g_Q(y;\gamma)$ and let

$$\phi(x)=f(y)+\langle f'(y),x-y\rangle+\frac{\gamma}{2}\|x-y\|^2.$$

Then $\phi'(x)=f'(y)+\gamma(x-y)$ and for any $x\in\mathbb{R}^n$ we have

$$\langle f'(y)-g_Q,x-x_Q\rangle=\langle\phi'(x_Q),x-x_Q\rangle\ge 0.$$

Hence,

$$\begin{aligned}f(x)-\frac{\mu}{2}\|x-y\|^2 &\ge f(y)+\langle f'(y),x-y\rangle\\ &= f(x)+\langle f'(y),x_Q-y\rangle+\langle f'(y),x-x_Q\rangle\\ &\ge f(x)+\langle f'(y),x_Q-y\rangle+\langle g_Q,x-x_Q\rangle\\ &=\phi(x_Q)-\frac{\gamma}{2}\left\|x_Q-y\right\|^2+\langle g_Q,x-x_Q\rangle\\ &=\phi(x_Q)-\frac{1}{2\gamma}\left\|g_Q\right\|^2+\langle g_Q,x-x_Q\rangle.\end{aligned}$$

Note that since $\gamma\ge L$, $\phi(x_Q)\ge f(x_Q)$. Thus, the lemma is proved.

3. Generalized Nesterov's algorithm

In order to find an approximate solution to problem (1), we introduce the generalized Nesterov's algorithm that is presented in Algorithm 3. Note that if we set $\beta_k=L$ for all k, then the generalized Nesterov's algorithm returns Nesterov's algorithm in [8].

Generalized Nesterov's algorithm (Algorithm 3)

[1] Initial guess Choose $x_0\in Q$ and $\gamma_0\ge\mu,\beta_0\ge L$. Set $v_0=x_0$.

[2] For $k=0,1,2,\ldots$

1. Compute $\alpha_k\in(0,1)$ from equation

$$\beta_k\alpha_k^2=(1-\alpha_k)\gamma_k+\alpha_k\mu\ .$$

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

3. Compute $y_k=\frac{\alpha_k\gamma_k}{\gamma_k+\alpha_k\mu}v_k+\frac{\gamma_{k+1}}{\gamma_k+\alpha_k\mu}x_k$.

4. Compute $x_{k+1}=x_Q(y_k;\beta_k)$ and $g_Q(y_k;\beta_k)$

5. Compute

$$v_{k+1}=\frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_kv_k+\alpha_k\mu y_k-\alpha_kg_Q(y_k;\beta_k)\right].$$

6. Compute $\beta_{k+1}\ge L$

[3] Output: $\{x_k\}$.

In order to prove the convergence and convergence rate of Algorithm 3, we introduce the pair of sequences, $\{\phi_k(x)\}_{k=0}^\infty,\{\lambda_k\}_{k=0}^\infty$, recursively defined by:

$$\lambda_0=1,\ \lambda_{k+1}=(1-\alpha_k)\lambda_k, \quad (7)$$

$$\phi_0(x)=f(x_0)+\frac{\gamma_0}{2}\|x-x_0\|^2, \quad (8)$$

$$\phi_{k+1}(x)=(1-\alpha_k)\phi_k(x) +\alpha_k\left[f\left(x_Q(y_k;\beta_k)\right)+\frac{1}{2\beta_k}\left\|g_Q(y_k;\beta_k)\right\|^2 +\langle g_Q(y_k;\beta_k),x-y_k\rangle+\frac{\mu}{2}\|x-y_k\|^2\right]. \quad (9)$$

Lemma 3.1 *A pair of sequences $\{\phi_k(x)\}_{k=0}^\infty$ and $\{\lambda_k\}_{k=0}^\infty$ satisfies that for any $x\in\mathbb{R}^n$ and all $k\ge 0$ we have*

$$\phi_k(x)\le(1-\lambda_k)f(x)+\lambda_k\phi_0(x). \quad (10)$$

Proof. We prove by induction method. For $k=0$, the statement is true since $\phi_0(x)\le(1-\lambda_0)f(x)+\lambda_0\phi_0(x)\equiv\phi_0(x)$. Further, let (10) holds for some $k\ge 0$. Then, for any $x\in\mathbb{R}^n$,

$$\begin{aligned}\phi_{k+1}(x)&\le(1-\alpha_k)\phi_k(x)+\alpha_kf(x)\\ &= (1-(1-\alpha_k)\lambda_k)f(x)+(1-\alpha_k)(\phi_k(x) -(1-\lambda_k)f(x))\\ &\le (1-(1-\alpha_k)\lambda_k)f(x)+(1-\alpha_k)\lambda_k\phi_0(x)\\ &= (1-\lambda_{k+1})f(x)+\lambda_{k+1}\phi_0(x).\end{aligned}$$

Lemma 3.2 *Assume that $f\in\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ and Q is a closed convex set in $\mathbb{R}^n$. Let $\{x_k\},\{y_k\},\{v_k\},\{\alpha_k\}$ be sequences generated by Algorithm 3. Then,*

(1) For every $k\in\mathbb{N}$, v_k is the minimizer of ϕ_k defined by (8) and (9), and the function ϕ_k has the form

$$\phi_k(x)=\phi_k^*+\frac{\gamma_k}{2}\|x-v_k\|^2, for\ any\ x\in\mathbb{R}^n, \quad (11)$$

where $\phi_0^*=f(x_0)$,

$$\phi_{k+1}^*=(1-\alpha_k)\phi_k^*+\alpha_kf\left(x_Q(y_k;\beta_k)\right) +\frac{1}{2\beta_k}(\alpha_k-1)\left\|g_Q(y_k;\beta_k)\right\|^2+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}} \left(\frac{\mu}{2}\|y_k-v_k\|^2+\langle g_Q(y_k;\beta_k),v_k-y_k\rangle\right). \quad (12)$$

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

Proof.

(1) We prove by the induction method. It is obvious if $k=0$. Assume that the statement is true for some $k \geq 0$, i.e., ϕ_k defined by (8) and (9), and the function ϕ_k has the form

$$\phi_k(x) = \phi_k^* + \frac{\gamma_k}{2}\|x - v_k\|^2, \qquad \text{for any } x \in \mathbb{R}^n.$$

Then, from (9), for any $x \in \mathbb{R}^n$,

$$\begin{aligned}\phi_{k+1}(x) &= (1-\alpha_k)\left(\phi_k^* + \frac{\gamma_k}{2}\|x - v_k\|^2\right) \\ &+\alpha_k\left[f\left(x_Q(y_k;\beta_k)\right) + \frac{1}{2\beta_k}\left\|g_Q(y_k;\beta_k)\right\|^2 \right.\\ &\left.+\langle g_Q(y_k;\beta_k), x - y_k\rangle + \frac{\mu}{2}\|x - y_k\|^2\right].\end{aligned}$$

By Step 2 and Step 3 in Algorithm 3, we have

$$\begin{aligned}&\phi_{k+1}{}'(x)\\ &= (1-\alpha_k)\gamma_k(x - v_k) + \alpha_k\left[g_Q(y_k;\beta_k) + \mu(x - y_k)\right]\\ &= [(1-\alpha_k)\gamma_k + \alpha_k\mu]x\\ &\quad -\left[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k g_Q(y_k;\beta_k)\right]\\ &= \gamma_{k+1}x - \left[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k g_Q(y_k;\beta_k)\right].\end{aligned}$$

From Step 6 in Algorithm 3, we have $\phi_{k+1}{}'(v_{k+1}) = 0$, i.e., v_{k+1} is the miniminzer of ϕ_{k+1}. Furthermore, from (9) we have

$$\phi_{k+1}''(x) = (1-\alpha_k)\phi_k''(x) + \alpha_k\mu I_n = \gamma_{k+1}I_n, \quad \forall x \in \mathbb{R}^n.$$

Thus, $\phi_{k+1}(x) = \phi_{k+1}^* + \frac{\gamma_{k+1}}{2}\|x - v_{k+1}\|^2$, for any $x \in \mathbb{R}^n$. Finally, let us compute ϕ_{k+1}^*. We have

$$\begin{aligned}\phi_{k+1}^* + \frac{\gamma_{k+1}}{2}\|y_k - v_{k+1}\|^2 &= \phi_{k+1}(y_k)\\ &= (1-\alpha_k)\left(\phi_k^* + \frac{\gamma_k}{2}\|y_k - v_k\|^2\right)\\ &+\alpha_k f\left(x_Q(y_k;\beta_k)\right) + \frac{\alpha_k}{2\beta_k}\left\|g_Q(y_k;\beta_k)\right\|^2. \quad (13)\end{aligned}$$

From Step 6 in Algorithm 3, we have

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

Therefore,

$$\begin{aligned}\frac{\gamma_{k+1}}{2}\|v_{k+1} - y_k\|^2 &= \frac{1}{2\gamma_{k+1}}\left[(1-\alpha_k)^2\gamma_k^2\|v_k - y_k\|^2\right.\\ &-2\alpha_k(1-\alpha_k)\gamma_k\left\langle v_k - y_k, g_Q(y_k;\beta_k)\right\rangle +\\ &\left.\alpha_k^2\left\|g_Q(y_k;\beta_k)\right\|^2\right].\end{aligned}$$

It remains to substitute this relation into (13).

(2) We now prove $\phi_k^* \geq f(x_k)$ for all $k \in \mathbb{N}$ by the induction method. For $k=0$, we have $\phi_0(x) = f(x_0) + \frac{\gamma_0}{2}\|x - v_0\|^2$. Thus, $f(x_0) = \phi_0^*$. Suppose that $\phi_k^* \geq f(x_k)$ is true for some $k \geq 0$. Then, from (1) and Lemma 2.1 we have

$$\begin{aligned}\phi_{k+1}^* &\geq (1-\alpha_k)f(x_k) + \alpha_k f\left(x_Q(y_k;\beta_k)\right)\\ &+\frac{1}{2\beta_k}(\alpha_k - 1)\left\|g_Q(y_k;\beta_k)\right\|^2\\ &+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle g_Q(y_k;\beta_k), v_k - y_k\rangle\\ &\geq (1-\alpha_k)[f(x_Q(y_k;\beta_k)) + \langle g_Q(y_k;\beta_k), x_k - y_k\rangle\\ &+\frac{1}{2\beta_k}\left\|g_Q(y_k;\beta_k)\right\|^2 + \frac{\mu}{2}\|x_k - y_k\|^2]\\ &+\alpha_k f\left(x_Q(y_k;\beta_k)\right) + \frac{1}{2\beta_k}(\alpha_k - 1)\left\|g_Q(y_k;\beta_k)\right\|^2\\ &+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle g_Q(y_k;\beta_k), v_k - y_k\rangle\\ &= f(x_Q(y_k;\beta_k)) + \frac{(1-\alpha_k)\mu}{2}\|x_k - y_k\|^2\\ &+(1-\alpha_k)\left\langle g_Q(y_k;\beta_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k\right\rangle.\end{aligned}$$

From Step 4 in Algorithm 3, we have $\langle g_Q(y_k;\beta_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k\rangle = 0$. Therefore,

$$\phi_{k+1}^* \geq f\left(x_Q(y_k;\beta_k)\right) = f(x_{k+1}).$$

Thus, by the induction method, we have $\phi_k^* \geq f(x_k)$ for all $k \in \mathbb{N}$.

Lemma 3.3 *Algorithm 3 generates a sequence $\{x_k\}_{k=0}^\infty$ which satisfies*

$$f(x_k) - f(x^*) \leq \lambda_k\left[f(x_0) - f(x^*) + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right], \quad \forall k$$

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

Proof. By Lemma 3.2 (2), we have $\phi_k^* \geq f(x_k)$ for all k. By Lemma 3.1, we have

$$\begin{aligned}f(x_k) \;\leq\; \phi_k^* &= \min_{x\in\mathbb{R}^n}\phi_k(x)\\ &\leq \min_{x\in\mathbb{R}^n}[(1-\lambda_k)f(x) + \lambda_k\phi_0(x)]\\ &\leq (1-\lambda_k)f(x^*) + \lambda_k\phi_0(x^*).\end{aligned}$$

Therefore,

$$f(x_k) - f(x^*) \leq \lambda_k\left[f(x_0) - f(x^*) + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right], \quad \forall k.$$

Finally, from (7) we deduce that $\lambda_k = \prod_{i=0}^{k-1}(1-\alpha_i)$ with $\lambda_0 = 1$.

To estimate the convergence rate of Algorithm 3, we need the following results.

Lemma 3.4 *Let $\{\lambda_k\}_{k=0}^\infty$ be the sequence in Lemma 3.3. Then,*

(1) If the sequence $\{\beta_k\}$ is increasing, then

$$\lambda_k \leq min\left\{\left(1 - \sqrt{\frac{\mu}{\beta_{k-1}}}\right)^k, \frac{4\beta_k}{\left(2\sqrt{\beta_k} + k\sqrt{\gamma_0}\right)^2}\right\}.$$

(2) If the sequence $\{\beta_k\}$ is bounded from above by $\overline{\beta}$, then

$$\lambda_k \leq min\left\{\left(1 - \sqrt{\frac{\mu}{\overline{\beta}}}\right)^k, \frac{4\overline{\beta}}{\left(2\sqrt{\overline{\beta}} + k\sqrt{\gamma_0}\right)^2}\right\}.$$

Proof.

We prove that $\gamma_k \geq \mu$ for all k by the induction method. It is easy to show that the inequality is true for $k=0$. Assume that $\gamma_k \geq \mu$ for some $k \geq 0$. Then $\gamma_{k+1} = \beta_k\alpha_k^2 =$

$(1-\alpha_k)\gamma_k + \alpha_k\mu \geq \mu$. Therefore, $\gamma_k \geq \mu$ for all k. Furthermore, since $\beta_k\alpha_k^2 \geq \alpha_k\mu$, $\alpha_k \geq \frac{\mu}{\beta_k} \geq \sqrt{d\frac{\mu}{\beta_k}}$.

If the sequence $\{\beta_k\}$ is increasing, then $\lambda_k = \prod_{i=0}^{k-1}(1-\alpha_i) \leq \left(1-\sqrt{\frac{\mu}{\beta_{k-1}}}\right)^k$ for all k. Similarly, we can prove that if $\{\beta_k\}$ is bounded from above by $\overline{\beta}$, then $\lambda_k = \prod_{i=0}^{k-1}(1-\alpha_i) \leq \left(1-\sqrt{\frac{\mu}{\overline{\beta}}}\right)^k$.

On the other hand, we can prove that $\gamma_k \geq \gamma_0\lambda_k$ by the induction method. With $k=0$, we have $\gamma_0 = \gamma_0\lambda_0$. Thus, the inequality is true with $k=0$. Assume that the inequality is true for some $k=m$, i.e., $\gamma_m \geq \gamma_0\lambda_m$. Then,

$$\gamma_{m+1} = (1-\alpha_m)\gamma_m + \alpha_k\mu \geq (1-\alpha_m)\gamma_0\lambda_m + \alpha_k\mu = \gamma_0\lambda_{m+1}.$$

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

Let $a_k = \frac{1}{\sqrt{\lambda_k}}$ for all k. Since $\{\lambda_k\}$ is a decreasing sequence, we have

$$a_{k+1} - a_k = \frac{1}{\sqrt{\lambda_{k+1}}} - \frac{1}{\sqrt{\lambda_k}} = \frac{\sqrt{\lambda_k}-\sqrt{\lambda_{k+1}}}{\sqrt{\lambda_k}\sqrt{\lambda_{k+1}}} = \frac{\lambda_k - \lambda_{k+1}}{\sqrt{\lambda_k}\sqrt{\lambda_{k+1}}\left(\sqrt{\lambda_k}+\sqrt{\lambda_{k+1}}\right)} \geq \frac{\lambda_k-\lambda_{k+1}}{2\lambda_k\sqrt{\lambda_{k+1}}} = \frac{\alpha_k\lambda_k}{2\lambda_k\sqrt{\lambda_{k+1}}} = \frac{\alpha_k}{2\sqrt{\lambda_{k+1}}}.$$

Using $\beta_k\alpha_k^2 = \gamma_{k+1} \geq \gamma_0\lambda_{k+1}$, we have

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

Thus, if the seqsuence $\{\beta_k\}$ is increasing, then $a_k \geq 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{\beta_k}}$ and if the sequence $\{\beta_k\}$ is bounded from above by $\overline{\beta}$, then $a_k \geq 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{\overline{\beta}}}$. The lemma is proved.

From Lemma 3.4 we observe that if $\{\beta_k\}$ is increasing and unbounded from above, then the convergence rate of $\{\lambda_k\}$ is worse than the case which $\{\beta_k\}$ is bounded from above. Thus, in the following we only consider the case the sequence $\{\beta_k\}$ bounded from above. In this case, the convegence and convergence rate of Algorithm 3 is given in the following theorem.

Theorem 3.1 *If $\gamma_0 \geq \mu$ and $\{\beta_k\} \subset [L, \overline{\beta}]$, then the sequence $\{x_k\}_{k=0}^{\infty}$ is generated by Algorithm 3 satisfies*

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

Proof. By Lemma 3.4 and the fact $\langle f'(x^*), x - x^*\rangle \geq 0$ for all $x \in Q$, we have

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

(using (2) and $\beta_k \geq L$).

$$= \lambda_k\left[\frac{\beta_k+\gamma_0}{2}\|x_0 - x^*\|^2 - 2\langle f'(x_0) - f'(x^*), x_0 - x^*\rangle\right]$$
$$\leq \frac{\lambda_k(\beta_k+\gamma_0)}{2}\|x_0 - x^*\|^2 \quad (\text{using}(2),(3)).$$

From the last inequality and Lemma 3.1, we have

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

From Lemma 3.1, the theorem is proved.

4. Simulation

In this section we illustrate the performance of the proposed algorithm (Generalized Nesterov's algorithm - GNA) and compare it with the projected gradient descent (PGD) algorithm with constant stepsizes (equal to the Lipschitz constant L). We consider an specific example with a closed-form objective functional given by

$$\min_{x\in Q} f(x) := \frac{1}{16}\left(x_1^2 + \sum_{i=1}^{N-1}(x_i - x_{i+1})^2 - 2x_1\right) + \frac{\overline{\mu}}{2}\|x\|^2, \qquad (14)$$

Where, $x = (x_1, x_2 \dots, x_N)$ and $N = 500$. It is easy to show that f is strongly convex with strong convexity parameter $\mu \leq \overline{\mu} = 0.1$ and f' is Lipschitz continuous with the Lipschitz constant $L \geq \overline{L} = 0.5$.

Firstly, we analyze the performance of PGD and GNA with $\mu = \overline{\mu}$ and $\mu = 0$ on $N-$dimensional box $Q = [-50,50]^N$ and $Q = [20,50]^N \subset \mathbb{R}^N$. For these algorithms, we use the same starting point x_0 that is generalized randomly by Matlab function $random(40,50,N)$. The Lipschitz constant is $L = \overline{L} = 0.5$ and iterator number $n = 30$.

The values of objective function at each iteration are presented in Figure 1 for two cases of Q. It shows that GNA with two different values of μ has similar convergence rate and they converge faster than PGD.

Secondly, we demonstate the performence of GNA for four case of $\{\beta_k\}$: $\beta_k = L, 0.2L, 2L$ for all k and $\beta_k = \frac{2L\left(k+\frac{1}{3}\right)}{2k+1}$. Here, we also set $L = 0.5, \mu = 0.1$ and the iterator number $n = 30$.

For $Q = [20; 50]^N$, the values of objective function at each iteration are presented in Figure 2. It shows that the convergence of GNA for $\beta_k = 0.2L$ is faster than that for $\beta_k = L$. It shows that GNA may converge for the sequence $\{\beta_k\}$ with $\beta_k < L$ in some practical cases, which is not proved by the theory.

For $Q = [-50; 50]^N$, the values of objective function at each iteration are presented in Figure 3. This shows that GNA with $\beta_k = 0.2L$ diverges rapidly. This is suitable with theoretical result. We have proved the convergence of GNA under condition $\beta_k \geq L$ for all k. Note that GNA converges very fast for $\beta_k = \frac{2L\left(k+\frac{1}{3}\right)}{2k+1}$. This observation has not obtained theoretically. For this situation, $\beta_k < L$ for all k and β_k converges to L as k tend to infinitive.

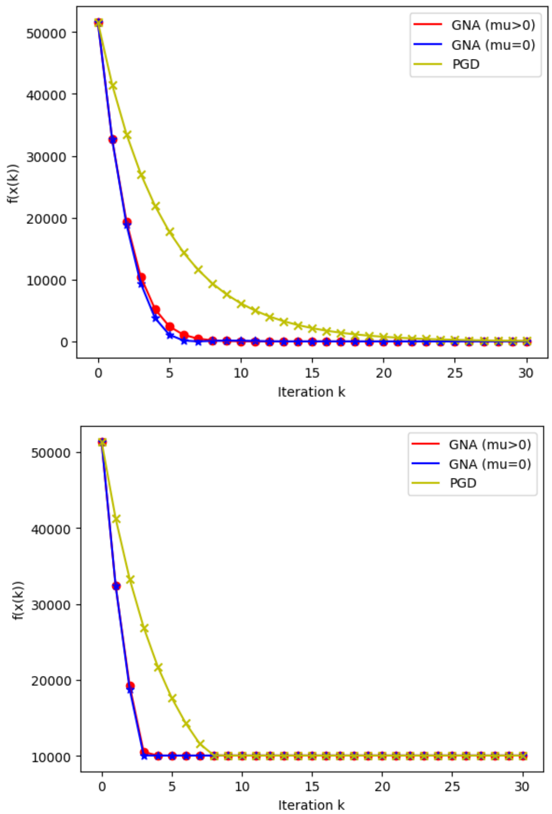

Figure 1. *Values of $f(x_k)$ in PGD and GNA with $L = 0.5$, $\mu = 0.1$. Here, $x_0 = random(40,50,N), Q = [-50; 50]^{500}$ (above) and $Q = [20; 50]^{500}$ (below)*

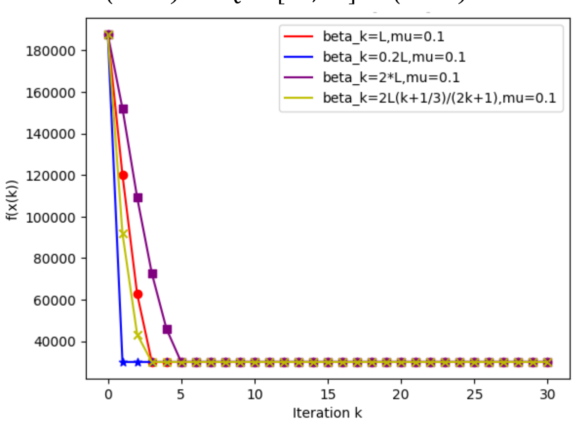

Figure 2. *Values of $f(x_k)$ in GNA with different values of $L, \mu = 0.1$. Here, $Q = [20; 50]^N, x_0 = 50 \cdot ones(N)$*

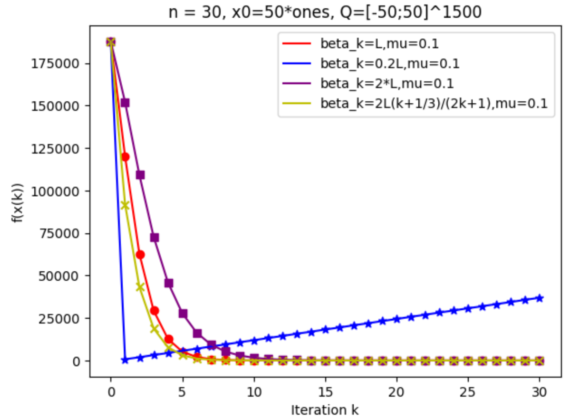

Figure 3. *Values of $f(x_k)$ in GNA with different values of $L, \mu = 0.1$. Here, At $Q = [-50; 50]^N, x_0 = 50 \cdot ones(N)$*

5. Conclusion

In this paper, we have presented the generalized Nesterov's algorithm for the constrained minimization on closed convex sets. The algorithm is presented in detail in Algirithm 3 and its convergence and convergence rate are given in Theorem 3.1. We have simulated the algrithm by one specific numerical example. We have showed that the generalized Nesterov's algorithm is faster than the projection gradient descent in both theory and specific numerical examples.

REFERENCES

[1] D. P. Bertsekas, W. Hager, and O. Mangasarian, *Nonlinear programming. Athena scientific belmont*. Massachusets, USA, 1999.

[2] N. Karmarkar, "A new polynomial-time algorithm for linear programming", in *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, New York, USA, 1984, pp. 302–311. Doi: 10.1145/800057.808695

[3] S. J. Wright, *Primal-dual interior-point methods*. SIAM, 1997.

[4] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[5] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 1999.

[6] M. JD Powell, "A view of algorithms for optimization without derivatives", *Mathematics Today-Bulletin of the Institute of Mathematics and its Applications*, vol. 43, no. 5, pp. 170–174, 2007.

[7] R. H Byrd, J. Nocedal, and R. B. Schnabel, "Representations of quasi-newton matrices and their use in limited memory methods", *Mathematical Programming*, vol. 63, no. 1-3, pp. 129–156, 1994.

[8] M. JD Powell, "A method for nonlinear constraints in minimization problems", *Optimization*, Academic Press, pp. 283–298, 1969.

[9] A. R. Conn, N. IM Gould, and P. L. Toint, *Trust region methods*. SIAM, 2000.

[10] M. JD Powell, "A fast algorithm for nonlinearly constrained optimization calculations", in *Numerical Analysis: Proceedings of the Biennial Conference Held at Dundee*, 1977, pp. 144–157.

[11] J. Nocedal and Y. Yuan, "Combining trust region and line search techniques", in *Advances in Nonlinear Programming: Proceedings of the 96 International Conference on Nonlinear Programming*, Springer, 1998, pp. 153–175.

[12] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*. SIAM, 2009.

[13] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Springer Science & Business Media, 2013.

[14] P. Q. Muoi, V. Q. Duy, and C. V. Khanh, "An optimal algorithm for convex minimization problems with nonconstant step-sizes", *The University of Danang - Journal of Science and Technology*, vol. 19, no. 12.1, pp.17–20, 2021.