AN OPTIMAL ALGORITHM FOR CONVEX MINIMIZATION PROBLEMS WITH NONCONSTANT STEP-SIZES

Pham Quy Muoi*, Vo Quang Duy, Chau Vinh Khanh

The University of Danang - University of Science and Education

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

(Reveived: November 16, 2021; Accepted: December 13, 2021)

Abstract - In [1], Nesterov has introduced an optimal algorithm with constant step-size, $h_k = \frac{1}{L}$ with L is the Lipschitz constant of objective function. The algorithm is proved to converge with optimal rate $O(1/k^2)$. In this paper, we propose a new algorithm, which is allowed nonconstant step-sizes h_k. We prove the convergence and convergence rate of the new algorithm. It is proved to have the convergence rate $O(1/k^2)$ as the original one. The advance of our algorithm is that it is allowed nonconstant step-sizes and give us more free choices of step-sizes, which convergence rate is still optimal. This is a generalization of Nesterov's algorithm. We have applied the new algorithm to solve the problem of finding an approximate solution to the integral equation.

Key words - Convex minimization problem; Modifed Nesterov's algorithm; Optimal convergence rate; Nonconstant step-size.

1. Introduction

In this paper, we consider an unconstrained minimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \qquad (1)$$

where $f: \mathbb{R}^n \to \mathbb{R}$ is a convex and differentiable function with the derivative f' being Lipschitz continuous. We denote L as the Lipschitz constant of f' and $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$ is the set of all such functions. We also denote x^* and f^* as a solution and the minimum of problem (1), respectively.

There are several methods to solve problem (1) such as the gradient method, conjugate gradient method, Newton and Quasi-Newton one, but these approaches are far from being optimal for class of convex minimization problems. The optimal methods for minimizing smooth convex and strongly convex functions have been proposed in [1] (see page 76, algorithm (2.2.6)). The ideas of Nesterov have been applied to nonsmooth optimization problems in [2, 3]. Although, the methods introduced by Nesterov in [1] have optimal convergent rate, he only introduce a rule for choosing constant step-size. Other possible choices of step-sizes are still missing. In this paper, we propose a new approach, which are based on the optimal method introduced in [1], but the values of step-sizes are possibly to change in each iteration. We will prove that new process converges with the convergence rate $O(1/k^2)$.

2. Notations and preliminary results

In this section, we recall some notations and properties of differentiable convex functions, differentiable functions that the gradient vectors are Lipschitz contiuous. These notations and properties are used in the proofs of main results in this paper. For more information, we refer to the references [1, 3, 4, 5, 6]. Here, the notation f' denoted for the gradient vector ∇f of function f.

A continuously differentiable function f is convex in $\mathbb{R}^n$ if and only if $f(y) \geq f(x) + \langle f'(x), y - x\rangle, \forall x, y \in \mathbb{R}^n$.

A function f is Lipschitz continuously differentiable if and only if there exists a real number $L > 0$ such that

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

If it is the case, L is called a Lipschitz constant.

Theorem 2.1 ([Theorem 2.1.5, 1]) *If $f \in F_L^{1,1}(\mathbb{R}^n)$, then $\forall x, y \in \mathbb{R}^n$,*

$$0 \leq f(y) - f(x) - \langle f'(x), y - x\rangle \leq \frac{L}{2}\|x - y\|^2 \qquad (2)$$

$$f(x) + \langle f'(x), y - x\rangle + \frac{1}{2L}\|f'(x) - f'(y)\|^2 \leq f(y) \qquad (3)$$

The schemes and efficiency bounds of optimal methods are based on the notion of *estimate sequence*.

Definition 2.1 *A pair of sequences $\{\phi_k(x)\}_{k=0}^{\infty}$ and $\{\lambda_k\}_{k=0}^{\infty}, \lambda_k \geq 0$ is called an estimate sequence of function $f(x)$ if $\lambda_k \to 0$ and for any $x \in \mathbb{R}^n$ and all $k \geq 0$ we have*

$$\phi_k(x) \leq (1 - \lambda_k) f(x) + \lambda_k \phi_0(x). \qquad (4)$$

The next statement explains why these objects could be useful.

Lemma 2.1 ([Lemma 2.2.1, 1]) *If for some sequence $\{x_k\}$, we have*

$$f(x_k) \leq \phi_k^* \equiv \min_{x\in\mathbb{R}^n} \phi_k(x), \qquad (5)$$

then $f(x_k) - f^* \leq \lambda_k[\phi_0(x^*) - f^*]$.

Thus, for any sequence $\{x_k\}$ satisfying (5) we can derive its rate of convergence directly from the rate of convergence of sequence $\{\lambda_k\}$. The next lemma gives us one choice of estimate sequences.

Lemma 2.2 ([Lemma 2.2.2, 1]) *Assume that*

1. $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$,

2. $\phi_0(x)$ is an arbitrary function on $\mathbb{R}^n$,

3. $\{y_k\}_{k=0}^{\infty}$ is an arbitrary sequence in $\mathbb{R}^n$,

4. $\{\alpha_k\}_{k=0}^{\infty}: \alpha_k \in (0,1), \sum_{k=0}^{\infty} \alpha_k = \infty$,

5. $\lambda_0 = 1$.

Then, the pair of sequences $\{\phi_k(x)\}_{k=0}^{\infty}, \{\lambda_k\}_{k=0}^{\infty}$ recursively defined by:

$$\lambda_{k+1} = (1 - \alpha_k)\lambda_k, \qquad (6)$$

$$\phi_{k+1}(x) = (1 - \alpha_k)\phi_k(x) + \alpha_k[f(y_k) + \langle f'(y_k), x - y_k\rangle] \qquad (7)$$

is an estimate sequence.

3. Optimal algorithm with nonconstant step-sizes

Lemma 2.2 provides us with some rules for updating the estimate sequence. Now we have two control sequences, which can help to ensure inequality (5). Note that we are also free in the choice of initial function $\phi_0(x)$. In [1], Nesterov has used the quadratic function for $\phi_0(x)$ and the sequence $\{\alpha_k\}$ is chosen corresponding to the constant step-size $h_k = \frac{1}{L}$. In this section, we propose a new optimal method. We still choose $\phi_0(x)$ as in [1], but the sequence $\{\alpha_k\}$ is chosen corresponding to general step-size h_k. Thus, our method is a generalization of Nesterov's algorithm, the algorithm (2.2.6) in [1] and is presented in the following theorem.

Theorem 3.1 *Let* $x_0 = v_0 \in \mathbb{R}^n, \gamma_0 > 0$ *and*

$$\phi_0(x) = f(x_0) + \frac{\gamma_0}{2} \parallel x - v_0 \parallel^2.$$

Assume that the sequence $\{\phi_k(x)\}$ *is defined by (7), where the sequences* $\{\alpha_k\}, \{y_k\}$ *are defined as follows:*

$$\alpha_k \in (0,1) \text{ and } \beta_k L \alpha_k^2 = (1-\alpha_k)\gamma_k. \quad (8)$$

$$\gamma_{k+1} = \beta_k L \alpha_k^2, \quad (9)$$

$$y_k = \alpha_k v_k + (1-\alpha_k)x_k, \quad (10)$$

$$v_{k+1} = v_k - \frac{\alpha_k}{\gamma_{k+1}} f'(y_k), \quad (11)$$

$$h_k = \frac{1}{L}\left(1 + \sqrt{1 - \frac{1}{\beta_k}}\right) \quad (12)$$

$$x_{k+1} = y_k - h_k f'(y_k), \quad (13)$$

where $\{\beta_k\}$ *with* $\beta_k \geq 1, \forall k$ *is an arbitrary sequence in* $\mathbb{R}$. *Then, the function* ϕ_k *has the form*

$$\phi_k(x) = \phi_k^* + \frac{\gamma_k}{2} \parallel x - v_k \parallel^2, \quad (14)$$

Where

$$\phi_{k+1}^* = (1-\alpha_k)\phi_k^* + \alpha_k f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}} \parallel f'(y_k) \parallel^2 + \alpha_k \langle f'(y_k), v_k - y_k \rangle$$

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

Proof. Note that $\phi''_0(x) = \gamma_0 I_n$. Let us prove that $\phi_k''(x) = \gamma_k I_n$ for all $k \geq 0$. Indeed, if that is true for some k, then

$$\phi_{k+1}''(x) = (1-\alpha_k)\phi_k''(x) = (1-\alpha_k)\gamma_k I_n \equiv \gamma_{k+1} I_n.$$

This justifies the canonical form (14) of functions $\phi_k(x)$. Further,

$$\phi_{k+1}(x) = (1-\alpha_k)\left(\phi_k^* + \frac{\gamma_k}{2} \parallel x - v_k \parallel^2\right) + \alpha_k[f(y_k) + \langle f'(y_k), x - y_k \rangle].$$

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

$$(1-\alpha_k)\gamma_k(x - v_k) + \alpha_k f'(y_k) = 0.$$

From this, we get the equation for the point v_{k+1}, which is the minimum of the function $\phi_{k+1}(x)$.

Finally, let us compute ϕ_{k+1}^*. In view of the recursion rule for the sequence $\{\phi_k(x)\}$, we have

$$\phi_{k+1}^* + \frac{\gamma_{k+1}}{2} \parallel y_k - v_{k+1} \parallel^2 = \phi_{k+1}(y_k) = (1-\alpha_k)(\phi_k^* + \frac{\gamma_k}{2} \parallel y_k - v_k \parallel^2) + \alpha_k f(y_k). \quad (15)$$

Note that in view of the relation for v_{k+1},

$$v_{k+1} - y_k = (v_k - y_k) - \frac{\alpha_k}{\gamma_{k+1}} f'(y_k).$$

Therefore

$$\frac{\gamma_{k+1}}{2} \parallel v_{k+1} - y_k \parallel^2 = \frac{\gamma_{k+1}}{2} \parallel v_k - y_k \parallel^2 - \alpha_k \langle f'(y_k), v_k - y_k \rangle + \frac{\alpha_k^2}{\gamma_{k+1}} \parallel f'(y_k) \parallel^2.$$

It remains to substitute this relation into (15). We now prove $\phi_n^* \geq f(x_n)$ for all $n \in \mathbb{N}$ by induction method. At $k = 0$, we have $\phi_0(x) = f(x_0) + \frac{\gamma_0}{2}\|x - v_0\|^2$. So, $f(x_0) = \phi_0^*$. Suppose that $\phi_n^* \geq f(x_n)$ is true at $n = k$, we need to prove that the inequality is still true at $n = k+1$.

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

By (10), we have $\frac{\alpha_k \gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k = 0$ and thus

$$(1-\alpha_k)\left\langle f'(y_k), \frac{\alpha_k \gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k \right\rangle = 0.$$

Therefore, we have $\phi_{k+1}^* \geq f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(y_k)\|^2$. To finish the proof, we need to point out that $f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(y_k)\|^2 \geq f(x_{k+1})$. Indeed, from Theorem 2.1, we have: $0 \leq f(y) - f(x) - \langle f'(x), y - x \rangle \leq \frac{L}{2}\|x - y\|^2$. Replacing x by y_k, y by x_{k+1}, we obtain $f(x_{k+1}) \leq \frac{L}{2}\|y_k - x_{k+1}\|^2 + f(y_k) + \langle f'(y_k), x_{k+1} - y_k \rangle$. Inserting $x_{k+1} - y_k = -h_k f'(y_k)$ into above inequality, we have

$$f(x_{k+1}) \leq f(y_k) + \frac{L}{2}\|h_k f'(y_k)\|^2 + \langle f'(y_k), -h_k f'(y_k) \rangle$$

$$\Leftrightarrow f(x_{k+1}) \leq f(y_k) + \frac{L}{2}\|h_k f'(y_k)\|^2 - h_k\|f'(y_k)\|^2$$

$$\Leftrightarrow f(x_{k+1}) \leq f(y_k) - \left(h_k - \frac{L}{2}h_k^2\right)\|f'(y_k)\|^2.$$

By (12), we have $\frac{\alpha_k^2}{2\gamma_{k+1}} = h_k - \frac{L}{2}h_k^2$.

Based on Theorem 3.1, we can present the optimal method with nonconstant step-sizes as the following algorithm.

Algorithm 3.1.

(3) Initial guess: Choose $x_0 \in \mathbb{R}^n$ and $\gamma_0 > 0$.

Set $v_0 = x_0$.

(2) For $k = 0,1,2, \ldots$

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

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

2. Compute $\gamma_{k+1} = \beta_k L \alpha_k^2$.
3. Compute $y_k = \alpha_k v_k + (1-\alpha_k) x_k$.
4. Compute $f(y_k)$ and $f'(y_k)$.
5. Compute $x_{k+1} = y_k - h_k f'(y_k)$ with

$$h_k = \frac{1}{L}\left(1 + \sqrt{1 - \frac{1}{\beta_k}}\right).$$

6. Compute $v_{k+1} = v_k - \frac{\alpha_k}{\gamma_{k+1}} f'(y_k)$.

(3) Output: $\{x_k\}$.

Theorem 3.2 Algorithm 3.1 generates the sequence $\{x_k\}_{k=0}^{\infty}$ that satisfies

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

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

Proof. Choose $\phi_0(x) = f(x_0) + \frac{\gamma_0}{2}\|x - v_0\|^2$ and

$$\phi_0(x) = \phi_0^* + \frac{\gamma_0}{2}\|x - v_0\|^2.$$

Therefore, $f(x_0) = \phi_0^*$. Since $f(x_k) \le \phi_k^*, \forall k > 0$ (see the proof of Lemma 2.1), we have

$$f(x_k) - f^* \le \lambda_k[\phi_0(x^*) - f^*] = \lambda_k[f(x_0) - f^*]$$
$$\le \lambda_k\left[f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right]$$

Therefore, the theorem is proved.

To estimate the convergnce rate of Algorithm 3.1, we need the following result.

Lemma 3.1 With the estimate sequence is generated by Algorithm 3.1, we have

$$\lambda_k \le \frac{4\beta_k L}{\left(2\sqrt{L} + k\sqrt{\frac{\gamma_0}{\beta_k}}\right)^2}$$

if the sequence $\{\beta_k\}$ is increasing or

$$\lambda_k \le \frac{4\bar{\beta}L}{\left(2\sqrt{L} + k\sqrt{\frac{\gamma_0}{\bar{\beta}}}\right)^2}$$

if the sequence $\{\beta_k\}$ is bounded from above by $\bar{\beta}$.

Proof. We have $\gamma_k \ge 0$ for all k. We will prove that $\gamma_k \ge \gamma_0 \lambda_k$ by induction method. At $k = 0$, we have $\gamma_0 = \gamma_0 \lambda_0$. Thus, the iniquality is true with $k = 0$. Assume that the inequality is true with $k = m$, i.e., $\gamma_m \ge \gamma_0 \lambda_m$. Then,

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

Therefore, we obtain $\beta_k L \alpha_k^2 = \gamma_{k+1} \ge \gamma_0 \lambda_{k+1}$ for all $k \in \mathbb{N}$. Let $a_k = \frac{1}{\sqrt{\lambda_k}}$. Since $\{\lambda_k\}$ is a decreasing sequence, we have

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

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

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

Thus $a_k \ge 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{\beta_k L}}$ if the sequence $\{\beta_k\}$ is increasing or $a_k \ge 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{\bar{\beta}L}}$ if the sequence $\{\beta_k\}$ is bounded from above by $\bar{\beta}$. Thus, the lemma is proved.

Theorem 3.3 *If $\gamma_0 > 0$ and the sequence $\{\beta_k\}$ with $\beta_k \ge 1$ for all k is bounded from above by $\bar{\beta}$, then Algorithm 3.1 generates the sequence $\{x_k\}_{k=0}^{\infty}$ that satisfies*

$$f(x_k) - f^* \le \frac{2(L+\gamma_0)\bar{\beta}L}{\left(2\sqrt{L} + k\sqrt{\frac{\gamma_0}{\bar{\beta}}}\right)^2}\|x_0 - x^*\|^2.$$

Proof. By Theorem 2.1, Theorem 3.1 and noting that $f'(x^*) = 0$, we have

$$f(x_k) - f^* \le \lambda_k\left[f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right]$$
$$= \lambda_k\begin{bmatrix} f(x_0) - f(x^*) - \langle f'(x^*), x_0 - x^*\rangle \\ + \frac{\gamma_0}{2}\|x_0 - x^*\|^2 \end{bmatrix}$$
$$\le \lambda_k\left[\frac{L}{2}\|x_0 - x^*\|^2 + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right]$$
$$= \frac{L+\gamma_0}{2}\lambda_k\|x_0 - x^*\|^2$$

From Lemma 3.1, the theorem is proved.

Remark 3.1 *If $\beta_k = 1$ for all k, then Algorithm 3.1 returns to the algorithm (2.2.6), page 76, with $\mu = 0$ in [1]. The advantage in Algorithm 3.1 is that we are free to choose the sequence $\{\beta_k\}$ with $\beta_k \ge 1$. As a result, the step-size h_k in Step 6 has larger value than algorithm (2.2.6) in [1] ($h_k = \frac{1}{L}$ for all k in [1]). However, by Lemma 3.1 the convergence rate of Algorithm 3.1 is reduced if the sequence $\{\beta_k\}$ has too large value. For examle, if $\beta_k = k$ for all k, then $\lambda_k = O\left(\frac{1}{k}\right)$, which losses the o ptimal convergence rate of Algorithm 3.1. Lemma 3.1 and Theorem 3.3 show that the best convergence rate for Algorithm 3.1 is obtained when the sequence $\beta_k = 1$ for all k.*

4. Numerical solution

In this section we will illustrate the algorithm in this paper and the algorithm (2.2.6) with $\mu = 0$ in [1]. Here, we apply the algorithm to find a numerical approximation to the solution of the integral equation:

$$\int_0^1 e^{ts}x(s)ds = y(t), t \in [0,1], \quad (16)$$

with $y(t) = (\exp(t+1) - 1)/(t+1)$. Note the exact solution of this equation is $x(t) = \exp(t)$.

Approximating the integral in the right hand side by trapezoidal rule, we have

$$\int_0^1 e^{ts}x(s)ds \approx h\left(\frac{1}{2}x(0) + \sum_{j=1}^{n-1} e^{jht}x(jh) + \frac{1}{2}e^t x(1)\right)$$

with $h := 1/n$. For $t = ih$, we have the following linear system

$$h\left(\frac{1}{2}x_0 + \sum_{j=1}^{n-1} e^{ijh^2}x_j + \frac{1}{2}e^{ih}x_n\right) = y(ih), \quad (17)$$

for $i = 0, \dots, n$. Here, $x_i = x(ih)$ and $y_i = y(ih)$. The last linear system can be rewrite as

$$Ax = b. \quad (18)$$

Since the problem of solving integral equation is ill-posed, the linear system is ill-conditioned [8, 9]. Using Tikhonov regularization, the regularized approximate solution to (18) is the solution of the minimization problem:

$$\min_{x\in\mathbb{R}^{n+1}} f(x) = \frac{1}{2}\parallel Ax - b \parallel^2 + \alpha \parallel x \parallel^2 \quad (19)$$

where $A \in \mathbb{R}^{(n+1)\times(n+1)}, x, b \in \mathbb{R}^{n+1}$ and $\alpha > 0$.

It is clear that problem (19) is convex and Lipschitz differentiable. Thus, all conditions for the convergence of the algorithms are satisfied. The Lipschitz constant in this example is $L = \lambda_{\max}(A^TA) + 2\alpha$.

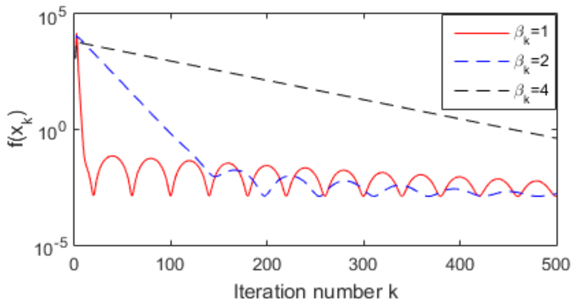

Figure 1. *The Objective function $f(x_k)$ in Algorithm 3.1 with three cases of the constant sequence $\{\beta_k\}$*

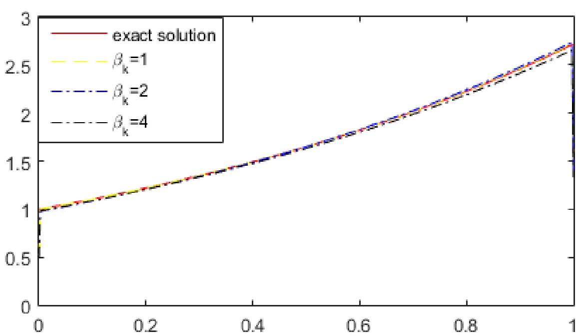

Figure 2. *The exact solution and approximate ones obtained by Algorithm 3.1 with three cases of the constant sequence $\{\beta_k\}$*

To illustrate the performance of Algorithm 3.1, we set $n = 400, \alpha = 10^{-6}$. Algorithm 3.1 is applied with three cases: $\beta_k = 1$ for all k, $\beta_k = 2$ for all k and $\beta_k = 4$ for all k. Figure 1 illustrates the behavior of objective function $f(x_k)$ in three cases of Algorithm 3.1. We see that Algorithm 3.1 works in three cases. The algorithm converges fastest when $\beta_k = 1$ for all k. However, it is hard to know when we should stop the algorithm such that the value of objective funcion is smallest since its values have violation frequently. The case of $\beta_k = 2$ for all k is a better choice in this case.

Figure 2 illustrates the approximate solutions and the exact one. In all three cases, Algorithm 3.1 gives good approximation to the exact solution, except two end points, which is normally seen by Tikhonov regularization.

5. Conclusion

In this paper, we have proposed the new algorithm, Algorithm 3.1, for the general convex minimization problem and prove the optimal convergent rate of the algorithm. Our algorithm is a generalization of Nesterov's algorithm in [1], which is allowed nonconstant step-sizes. Lemma 3.1 and Theorem 3.3 also show that the new algorithm obtain the fastest convergent rate when $\{\beta_k\}$ is the constant sequence and equal to one. Thus, it raises an new question that are there other updates for parameters in Algorithm 3.1 such that it converges faster than Nesterov's algorithm? It is still an open question and motivates us to study in the future.

Funding: This work was supported by Science and Technology Development Fund, Ministry of Training and Education under Project B2021-DNA-15.

REFERENCES

[1] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87. Springer Science & Business Media, 2013.

[2] P. Q. Muoi, D. N. Hào, S.K. Sahoo, D. Tang, N. H. Cong, and C. Dang. Inverse problems with nonnegative and sparse solutions: algorithms and application to the phase retrieval problem. *Inverse Problems*, 34(5), 055007, 2018.

[3] P.Q. Muoi, D.N. Hào, P. Maass, and M. Pidcock. Descent gradient methods for nonsmooth minimization problems in ill-posed problems. *Journal of Computational and Applied Mathematics*, *298*, 105-122, 2016.

[4] J. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.

[5] R. T. Rockafellar. *Convex analysis*, volume 36. Princeton university press, 1970.

[6] J. Stoer and C. Witzgall. *Convexity and optimization in finite dimensions I*, volume 163. Springer Science & Business Media, 2012.

[7] J. Baptiste, H. Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2004.

[8] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375. Springer Science & Business Media, 2000.

[9] A. Kirsch. An introduction to the mathematical theory of inverse problems, volume 120. Springer, 2011.