**Noname manuscript No.**
(will be inserted by the editor)

# Lipschitz-Free Mirror Descent Methods for Non-Smooth Optimization Problems

**Bowen Yuan, Mohammad S. Alkousa**

Received: date / Accepted: date

**Abstract** The part of the analysis of the convergence rate of the mirror descent method that is connected with the adaptive time-varying step size rules due to Alkousa et al. [1] is corrected. Moreover, a Lipschitz-free mirror descent method that achieves weak ergodic convergence is presented, generalizing the convergence results of the mirror descent method in the absence of the Lipschitz assumption.

**Keywords** Mirror descent · Mirror-C descent · Time-varying step size · Non-smooth convex optimization · Optimal convergence rate

**Mathematics Subject Classification (2000)** 90C25, 90C30

## 1 Introduction

Mirror descent method originated in [2,3] and was later analyzed in [4]. It is considered as the non-Euclidean extension of the standard subgradient method by employing a nonlinear distance function with an optimal step size in the nonlinear projection step. The mirror descent method is also applicable to optimization problems in Banach spaces where gradient descent is not [5]. An extension of the mirror descent method for constrained problems was proposed in [2,6]. This method is used in many applications, see [7,8] and references therein.

One of the focuses of the research on the projected subgradient method and the mirror descent method is the convergence rate of the algorithms under different step size rules. When the step size is fixed as $\gamma_k = O(1)/\sqrt{N}$,

Bowen Yuan
School of Mathematical Sciences, Beihang University, Beijing 100191, People's Republic of China. E-mail: yuanbowen@buaa.edu.cn
Mohammad S. Alkousa
Innopolis University, Innopolis, Universitetskaya Str., 1, 420500, Russia. E-mail: m.alkousa@innopolis.ru

where $N$ denotes the maximum number of iterations, the optimal convergence rate of the projected subgradient method has been established at $O\left(N^{-1/2}\right)$. Moreover, the projected subgradient method has been considered to have only sub-optimal convergence rate $O(N^{-1/2}\log N)$ for time-varying step sizes $\gamma_k = O(1)/\sqrt{k}$, which is well-documented in [9,10,11].

By proposing a new weighting scheme for the iteration points, which is called weak ergodic convergence, Z. Zhu et al. [12] first proved that the projected subgradient method can also achieve the optimal convergence rate of $O(N^{-1/2})$ under the time-varying step size rules. Based on the result of [12], Alkousa et al. proved that the optimal convergence rate under the time-varying step sizes still holds for the mirror descent method [1].

By devising a novel class of time-varying step size rules, Y. Xia et al. [13] successfully extended the convergence results of the projected subgradient method to the non-Lipschitz case. In this paper, we aim to further generalize the findings of [13] to the mirror descent method. Specifically, we will establish the optimal convergence results of the mirror descent method under more general conditions, without relying on the listed Assumption 2 below.

## 2 Fundamentals for the mirror descent method

Let $(\mathbf{E}, \|\cdot\|)$ be a normed finite-dimensional vector space, with an arbitrary norm $\|\cdot\|$, and $\mathbf{E}^*$ be the conjugate space of $\mathbf{E}$ with the following norm

$$\|y\|_* = \max_{x\in\mathbf{E}}\{\langle y, x\rangle : \|x\| \leq 1\},$$

where $\langle y, x\rangle$ is the value of the continuous linear functional $y \in \mathbf{E}^*$ at $x \in \mathbf{E}$.

Let $Q \subset \mathbf{E}$ be a compact convex set, and $\psi : Q \longrightarrow \mathbb{R}$ be a proper closed differentiable and $\sigma$-strongly convex (called prox-function or distance generating function) with $\sigma > 0$. The corresponding Bregman divergence is defined as

$$V_\psi(x, y) = \psi(x) - \psi(y) - \langle \nabla\psi(y), x - y\rangle \quad \forall x, y \in Q.$$

For the Bregman divergence, it holds the following inequality

$$V_\psi(x, y) \geq \frac{\sigma}{2}\|y - x\|^2 \quad \forall x, y \in Q. \tag{1}$$

In what follows, we denote the subdifferential of $f$ at $x$ by $\partial f(x)$, and the subgradient of $f$ at any point $x$ by $\nabla f(x) \in \partial f(x)$. Let $\operatorname{dom} f$ denote the domain of the function $f$, and $\operatorname{dom}(\partial f)$ denote the set of points of subdifferentiability of $f$, i.e.,

$$\operatorname{dom}(\partial f) = \{x \in \mathbf{E} : \partial f(x) \neq \emptyset\}.$$

The following identity, known as the three-point identity, is essential in the analysis of the mirror descent method.

**Lemma 1 (Three-point lemma)[14]** *Suppose that* $\psi : \mathbf{E} \longrightarrow (-\infty, \infty]$ *is proper closed, convex, and differentiable over* $\mathrm{dom}(\partial\psi)$*. Let* $a, b \in \mathrm{dom}(\partial(\psi))$ *and* $c \in \mathrm{dom}(\psi)$*. Then it holds*

$$\langle \nabla\psi(b) - \nabla\psi(a), c - a\rangle = V_\psi(c, a) + V_\psi(a, b) - V_\psi(c, b). \tag{2}$$

The following lemma is an extension of the second prox theorem to the case of non-Euclidean distances.

**Lemma 2 (Non-Euclidean second prox theorem)[14]** *Let*

1. $\psi : \mathbf{E} \longrightarrow (-\infty, \infty]$ *be a proper closed and convex function differentiable over* $\mathrm{dom}(\partial\psi)$*,*
2. $\varphi : \mathbf{E} \longrightarrow (-\infty, \infty]$ *be a proper closed and convex function satisfying* $\mathrm{dom}(\partial\varphi) \subseteq \mathrm{dom}(\partial\psi)$*,*
3. $\psi + \mathbb{I}_{\mathrm{dom}(\varphi)}$ *be a* $\sigma$*-strongly convex, where* $\mathbb{I}_A$ *is the indicator function of the set* $A$*.*

*Assume that* $b \in \mathrm{dom}(\partial\psi)$*, and let* $a$ *be defined by*

$$a = \arg\min_{x\in\mathbf{E}} \left\{\varphi(x) + V_\psi(x, b)\right\}.$$

*Then we have*

$$\langle \nabla\psi(b) - \nabla\psi(a), u - a\rangle \leq \varphi(u) - \varphi(a) \quad \forall u \in \mathrm{dom}(\varphi). \tag{3}$$

**Fenchel-Young inequality.** For any $a \in \mathbf{E}, b \in \mathbf{E}^*$, it holds the following inequality

$$|\langle a, b\rangle| \leq \frac{\|a\|^2}{2\lambda} + \frac{\lambda\|b\|_*^2}{2} \quad \forall\lambda > 0. \tag{4}$$

## 3 Lipschitz-free mirror descent method for constrained problem

In this section, we consider the convex optimization problem

$$\min_{x\in Q} f(x) \tag{5}$$

where $f : Q \longrightarrow \mathbb{R}$ is a convex continuous function.

The mirror descent method can be represented by the steps in Algorithm 1 below, which is excerpted from [1].

**Algorithm 1** Mirror Descent Method.

**Require:** Step sizes $\{\gamma_k\}_{k\geq 1}$, initial point $x^1 \in Q$, number of iterations $N$.
**for** $k = 1, 2, \dots, N$ **do**
  Calculate $\nabla f(x^k) \in \partial f(x^k)$,
  $x^{k+1} = \arg\min_{x\in Q} \left\{ \langle x, \nabla f(x^k)\rangle + \frac{1}{\gamma_k} V_\psi(x, x^k) \right\}$.
**end for**

### 3.1 Incorrectness in the analysis for adaptive step size in [1]

In the main convergence result in [1], it is necessary to require that $\{\gamma_k\}_{k\geq 1}$ is a positive non-increasing sequence of step sizes. In [1, Corollary 3.3] (also in [1, Corollary 4.3] for the composite problems), Alkousa et al. analyzed the optimal convergence rate of the mirror descent method with the adaptive time-varying step size rule (which we will call the Nesterov step size)

$$\gamma_k = \frac{\sqrt{2\sigma}}{\|\nabla f(x^k)\|_* \sqrt{k}}, \quad k = 1, 2, \dots, N. \tag{6}$$

But the Nesterov step sizes (6) are not necessarily non-increasing, see Example 1 as a counterexample in one-dimensional Euclidean space. Fortunately, this conflict can be avoided by making a modification to the step size setting. One of the goals of this note is to provide a corrective proof of the optimal convergence rate of the mirror descent method under the adaptive time-varying step size rules.

**Example 1** *Consider the convex optimization problem* (5)*in one-dimensional Euclidean space. Let the function* $f(x) = \frac{1}{2}x^2$ *and the distance generating function* $\psi(x) = \frac{1}{2}x^2$. *In this case, the mirror descent method is equivalent to the projected subgradient method. Set the initial point as* $x^1 = 10$, *the constraint set* $Q$ *as* $|x| \leq 10$. *Select the Nesterov step sizes defined in* (6).*The experimental results in Table 1 demonstrate that the Nesterov step sizes* $\gamma_k$ *are not necessarily non-increasing.*

Let $x^*$ be an optimal solution of (5). Since $x^*$ is unknown, it is hard to precisely determine how $x^1$ should be chosen to satisfy the assumption [1]

$$\max_{x\in Q} V_\psi(x^*, x) \leq V_\psi(x^*, x^1) < \infty.$$

Therefore, we restate the assumption as follows.

**Assumption 1** *There exists* $R > 0$ *such that*

$$V_\psi(x^*, x) \leq R < \infty \quad \forall x \in Q. \tag{7}$$

Table 1: Nesterov step sizes in Example 1.

| $k$ | $x^k$ | $\gamma_k$ |
|---|---|---|
| 1 | 10 | 0.141421356237310 |
| 2 | 8.58578643762690 | 0.116471566962991 |
| 3 | 7.58578643762690 | 0.107635060338339 |
| 4 | 6.76928985669918 | 0.104458044515078 |
| 5 | 6.06218307551263 | 0.104328015857587 |
| ... | ... | ... |
| 13 | 2.06458695099841 | 0.189980988733214 |
| 14 | 1.67235468072204 | 0.226007363967817 |
| ... | ... | ... |
| 24 | 0.209552285731976 | 1.37758046201432 |
| 25 | -0.0791228488628367 | 3.57472862187939 |
| ... | ... | ... |
| 48 | 0.166305589462573 | 1.22740399701280 |
| 49 | -0.0378185557693590 | 5.34210005645243 |
| ... | ... | ... |
| 60 | 0.155379438403268 | 1.17502153252226 |
| 61 | -0.0271947474317873 | 6.65832593368331 |
| ... | ... | ... |
| 80 | 0.143015997988010 | 1.10556780523025 |
| 81 | -0.0150978850204088 | 10.4077385707513 |

### 3.2 Lipschitz-free mirror descent method

Another important assumption in [1, Theorem 3.1] is that the convex function $f$ should be $M_f$-Lipschitz, i.e., satisfy the following assumption.

**Assumption 2** *There exists an $M_f > 0$ such that for any subgradient $\nabla f(x) \in \partial f(x) \neq \emptyset$ and $x \in Q$, it holds that $\|\nabla f(x)\|_* \leq M_f$.*

Although one of the advantages of the adaptive step size rules is that there is no need to know the Lipschitz coefficient $M_f$ of the objective function in advance, Assumption 1 is still necessary for the proof of the convergence result (see in [9,10,11]). However, many convex functions fail to satisfy Assumption 2 on a compact convex set.

In this section, we aim to offer a corrective proof for the optimal convergence result of the mirror descent method under the adaptive step size rules presented in [1, Corollary 3.3]. Moreover, we extend the optimal convergence rate of the mirror gradient descent method to the case where the Lipschitz condition is not assumed. Building upon the concept of weak ergodic convergence, we now introduce the following theorem.

**Theorem 1** *Let $f$ be a convex continuous function, then for problem* (5)*, by Algorithm 1, for any fixed $a \in [0, 1]$ and $m \geq -1$, with positive non-increasing* *step sizes*

$$\gamma_k = \frac{\sqrt{2\sigma R}}{G_k k^{\frac{a}{2}}}, \quad k = 1, \ldots, N, \tag{8}$$

*where*

$$G_k = \max\left\{G_{k-1}, \|\nabla f(x^k)\|_* k^{\frac{1-a}{2}}\right\} \quad (G_0 := -\infty), \tag{9}$$

*and the weak ergodic convergence weight is defined as*

$$\omega_k^{(m)} = \begin{cases} 1/\gamma_k^m, & \text{if } -1 \leq m \leq 0, \\ k^{m/2}, & \text{if } m > 0, \end{cases} \tag{10}$$

*it satisfies the following inequality*

$$f(\widehat{x}) - f(x^*) \leq \sqrt{\frac{R}{2\sigma}} \cdot \frac{N^{\frac{m+1}{2}} + \sum_{k=1}^N k^{\frac{m-1}{2}}}{\sum_{k=1}^N k^{\frac{m}{2}}} \cdot \max_{k=1,\ldots,N} \|\nabla f(x^k)\|_*, \tag{11}$$

*where*

$$\widehat{x} = \frac{1}{\sum_{k=1}^N \omega_k^{(m)}} \sum_{k=1}^N \omega_k^{(m)} x^k.$$

*Proof* Because $f$ is a convex function, we can obtain

$$\begin{aligned} \gamma_k(f(x^k) - f(x^*)) &\leq \gamma_k\langle \nabla f(x^k), x^k - x^*\rangle \\ &= \gamma_k\langle \nabla f(x^k), x^{k+1} - x^*\rangle + \gamma_k\langle \nabla f(x^k), x^k - x^{k+1}\rangle. \end{aligned} \tag{12}$$

By Lemma 2, and choosing $\varphi(x) = \gamma_k\langle \nabla f(x^k), x\rangle$, $a = x^k$, $b = x^{k+1}$, $u = x^*$, we can rewrite (3) as

$$\left\langle \nabla\psi(x^k) - \nabla\psi(x^{k+1}), x^* - x^k\right\rangle \leq \gamma_k\langle \nabla f(x^k), x^* - x^{k+1}\rangle. \tag{13}$$

From Lemma 1, we get

$$\left\langle \nabla\psi(x^k) - \nabla\psi(x^{k+1}), x^* - x^k\right\rangle = V_\psi(x^*, x^{k+1}) + V_\psi(x^{k+1}, x^k) - V_\psi(x^*, x^k). \tag{14}$$

Combining (13) and (14), and since $\psi$ is $\sigma$-strongly convex, we have

$$\begin{aligned} \gamma_k\langle \nabla f(x^k), x^{k+1} - x^*\rangle &\leq V_\psi(x^*, x^k) - V_\psi(x^*, x^{k+1}) - V_\psi(x^{k+1}, x^k) \\ &\leq V_\psi(x^*, x^k) - V_\psi(x^*, x^{k+1}) - \frac{\sigma}{2}\|x^{k+1} - x^k\|^2. \end{aligned} \tag{15}$$

Based on the Fenchel-Young inequality, we get

$$\gamma_k\langle \nabla f(x^k), x^k - x^{k+1}\rangle \leq \frac{\gamma_k^2}{2\sigma}\|\nabla f(x^k)\|_*^2 + \frac{\sigma}{2}\|x^k - x^{k+1}\|^2. \tag{16}$$

Combining (12), (15) and (16), we obtain

$$f(x^k) - f(x^*) \leq \frac{1}{\gamma_k}\left(V_\psi(x^*, x^k) - V_\psi(x^*, x^{k+1})\right) + \frac{\gamma_k}{2\sigma}\|\nabla f(x^k)\|_*^2. \tag{17}$$

By the definition of weak ergodic convergence weight in (10), because $\{\gamma_k\}_{k=1}^N$ is a positive non-increasing step sizes, we have

$$\frac{\omega_k^{(m)}}{\gamma_k} - \frac{\omega_{k-1}{}^{(m)}}{\gamma_{k-1}} = \begin{cases} \frac{1}{\gamma_k^{m+1}} - \frac{1}{\gamma_{k-1}^{m+1}} \geq 0, & \text{if } -1 \leq m \leq 0, k \geq 2, \\ \frac{k^{m/2}}{\gamma_k} - \frac{(k-1)^{m/2}}{\gamma_{k-1}} \geq 0, & \text{if } m > 0, k \geq 2. \end{cases} \tag{18}$$

We can rewrite $G_k$ in the following equivalent form

$$G_k = \max_{j=1,\ldots,k} \left\{ \|\nabla f(x^j)\|_* \cdot j^{\frac{1-a}{2}} \right\}. \tag{19}$$

Then by Jensen's inequality, (17), the boundedness of $V_\psi(x^*, x^k)$, and (18), we obtain

$$\begin{aligned}
&\left(\sum_{k=1}^N \omega_k^{(m)}\right) \left[ f\left( \frac{\sum_{k=1}^N \omega_k^{(m)} x^k}{\sum_{k=1}^N \omega_k^{(m)}} \right) - f(x^*) \right] \\
&\leq \sum_{k=1}^N \omega_k^{(m)} \left( f(x^k) - f(x^*) \right) \\
&\leq \sum_{k=1}^N \frac{\omega_k^{(m)}}{\gamma_k} \left( V_\psi(x^*, x^k) - V_\psi(x^*, x^{k+1}) \right) + \sum_{k=1}^N \frac{\omega_k^{(m)} \gamma_k}{2\sigma} \|\nabla f(x^k)\|_*^2 \\
&= \frac{\omega_1^{(m)}}{\gamma_1} V_\psi(x^*, x^1) + \sum_{k=2}^N \left( \frac{\omega_k^{(m)}}{\gamma_k} - \frac{\omega_k^{(m)}}{\gamma_{k-1}} \right) V_\psi(x^*, x^k) \\
&\quad - \frac{\omega_N^{(m)}}{\gamma_N} V_\psi(x^*, x^N) + \sum_{k=1}^N \frac{\omega_k^{(m)} \gamma_k}{2\sigma} \|\nabla f(x^k)\|_*^2 \\
&\leq R \left( \frac{\omega_1^{(m)}}{\gamma_1} + \sum_{k=2}^N \left( \frac{\omega_k^{(m)}}{\gamma_k} - \frac{\omega_k^{(m)}}{\gamma_{k-1}} \right) \right) + \sum_{k=1}^N \frac{\omega_k^{(m)} \gamma_k}{2\sigma} \|\nabla f(x^k)\|_*^2 \\
&\leq R \frac{\omega_N^{(m)}}{\gamma_N} + \sum_{k=1}^N \frac{\omega_k^{(m)} \gamma_k}{2\sigma} \|\nabla f(x^k)\|_*^2.
\end{aligned} \tag{20}$$

Let $\widehat{x} := \frac{\sum_{k=1}^N \omega_k^{(m)} x^k}{\sum_{k=1}^N \omega_k^{(m)}}$. Because the weak ergodic convergence weight is non-negative, for $-1 \leq m \leq 0$, by the definition of $\gamma_k$ in (8) and (9), we have,

$$\begin{aligned}
&f(\widehat{x}) - f(x^*) \\
&\leq \frac{R \frac{\omega_N^{(m)}}{\gamma_N} + \sum_{k=1}^N \frac{\omega_k^{(m)} \gamma_k}{2\sigma} \|\nabla f(x^k)\|_*^2}{\sum_{k=1}^N \omega_k^{(m)}} \\
&\leq \frac{\frac{R}{\gamma_N^{m+1}} + \sum_{k=1}^N \frac{\gamma_k^{1-m}}{2\sigma} \|\nabla f(x^k)\|_*^2}{\sum_{k=1}^N \gamma_k^{-m}}
\end{aligned}$$

$$\leq \sqrt{\frac{R}{2\sigma}} \frac{\left(G_N N^{\frac{a}{2}}\right)^{m+1} + \sum_{k=1}^{N} \left(G_k k^{\frac{a}{2}}\right)^{m-1} \|\nabla f(x^k)\|_*^2}{\sum_{k=1}^{N} \left(G_k k^{\frac{a}{2}}\right)^m}$$

$$\leq \sqrt{\frac{R}{2\sigma}} \frac{\left(\max\limits_{k=1,\dots,N} \|\nabla f(x^k)\|_*\right)^{1+m} \cdot N^{\frac{m+1}{2}} + \sum_{k=1}^{N} \|\nabla f(x^k)\|_*^{1+m} \cdot k^{\frac{m-1}{2}}}{\left(\max\limits_{k=1,\dots,N} \|\nabla f(x^k)\|_*\right)^m \sum_{k=1}^{N} k^{\frac{m}{2}}}$$

$$\leq \sqrt{\frac{R}{2\sigma}} \cdot \frac{N^{\frac{m+1}{2}} + \sum_{k=1}^{N} k^{\frac{m-1}{2}}}{\sum_{k=1}^{N} k^{\frac{m}{2}}} \cdot \max_{k=1,\dots,N} \|\nabla f(x^k)\|_*. \tag{21}$$

When $m > 0$, by the definition of $\omega_k^{(m)}$ and $\gamma_k$, we can get

$$f(\widehat{x}) - f(x^*)$$

$$\leq \frac{\frac{R \cdot N^{\frac{m}{2}}}{\gamma_N} + \sum_{k=1}^{N} \frac{k^{\frac{m}{2}} \gamma_k}{2\sigma} \|\nabla f(x^k)\|_*^2}{\sum_{k=1}^{N} k^{\frac{m}{2}}}$$

$$\leq \sqrt{\frac{R}{2\sigma}} \frac{G_N N^{\frac{m+a}{2}} + \sum_{k=1}^{N} \|\nabla f(x^k)\|_*^2 G_k^{-1} k^{\frac{m-a}{2}}}{\sum_{k=1}^{N} k^{\frac{m}{2}}}$$

$$\leq \sqrt{\frac{R}{2\sigma}} \frac{\max\limits_{k=1,\dots,N} \|\nabla f(x^k)\|_* \cdot N^{\frac{m+1}{2}} + \sum_{k=1}^{N} \|\nabla f(x^k)\|_* \cdot k^{\frac{m-1}{2}}}{\sum_{k=1}^{N} k^{\frac{m}{2}}}$$

$$\leq \sqrt{\frac{R}{2\sigma}} \cdot \frac{N^{\frac{m+1}{2}} + \sum_{k=1}^{N} k^{\frac{m-1}{2}}}{\sum_{k=1}^{N} k^{\frac{m}{2}}} \cdot \max_{k=1,\dots,N} \|\nabla f(x^k)\|_*.$$

This completes the proof. □

By setting $m = 0$, we can obtain the optimal convergence rate result of the Lipschitz-free mirror descent method under the adaptive time-varying step size rules:

**Corollary 1** *Let $f$ be a convex continuous function, then for problem* (5)*, by Algorithm 1, for any fixed $a \in [0, 1]$ and $m = 0$, with the step sizes* (8)*, it satisfies the following inequality*

$$f\left(\frac{1}{N}\sum_{k=1}^{N} x^k\right) - f(x^*) \leq 3\sqrt{\frac{R}{2\sigma}} \cdot \frac{1}{\sqrt{N}} \cdot \max_{k=1,\dots,N} \|\nabla f(x^k)\|_*. \tag{22}$$

**Remark 1** *If we choose the distance generating function $\psi = \frac{1}{2}\|x\|_2^2$, then the results in Theorem 1 and Corollary 1 reduce to the result in [13]. If further given Assumption 2, Corollary 1 would be the same as [12, Corollary 3.2].*

## 4 Lipschitz-free mirror descent method for constrained composite problem

In this section, we consider the composite convex problem

$$\min_{x\in Q} F(x) = f(x) + h(x), \tag{23}$$

where $f$ is convex function, and $h$ is a non-negative, continuous, and convex function. Compared with Algorithm 1, we just need to change the iterative strategy to the following

**Algorithm 2** Composite Mirror Descent Method.

**Require:** Step sizes $\{\gamma_k\}_{k\geq 1}$, initial point $x^1 \in Q$, number of iterations $N$.
  **for** $k = 1, 2, \ldots, N$ **do**
2:  Calculate $\nabla f(x^k) \in \partial f(x^k)$,
  $x^{k+1} = \arg\min_{x\in Q} \left\{ \langle \nabla f(x^k), x\rangle + \frac{1}{\gamma_k} h(x) + \frac{1}{\gamma_k} V_\psi(x, x^k) \right\}$.
4: **end for**

Now we introduce the following theorem, while Assumption 2 is still not necessary.

**Theorem 2** *Let $f$ be a convex continuous function and $h$ be a non-negative convex function, then for problem* (23)*, by Algorithm 2, for any fixed $a \in [0,1]$ and $-1 \leq m \leq 0$, with positive non-increasing step sizes $\{\gamma_k\}_{k=1}^N$ defined in* (8) *and weak ergodic convergence weight $\left\{\omega_k^{(m)}\right\}_{k=1}^N$ defined in* (10)*, it satisfies the following inequality*

$$\begin{aligned} F(\widehat{x}) - F(x^*) \leq \sqrt{\frac{R}{2\sigma}} \cdot \frac{N^{\frac{m+1}{2}} + \sum_{k=1}^N k^{\frac{m-1}{2}}}{\sum_{k=1}^N k^{\frac{m}{2}}} \cdot \max_{k=1,\ldots,N} \|\nabla f(x^k)\|_* \\ + \left( \frac{\|\nabla f(x^1)\|_*}{\max\limits_{k=1,\ldots,N} \|\nabla f(x^k)\|_*} \right)^m \cdot \frac{h(x^1)}{\sum_{k=1}^N k^{\frac{m}{2}}}, \end{aligned}$$

*where* $\widehat{x} = \frac{1}{\sum_{k=1}^N \omega_k^{(m)}} \sum_{k=1}^N \omega_k^{(m)} x^k$.

*Proof* By Lemma 2, by choosing $\varphi(x) = \gamma_k\langle\nabla f(x^k), x\rangle + \gamma_k h(x)$, $a = x^k$, $b = x^{k+1}$, $u = x^*$, then (3) can be rewritten as

$$\left\langle \nabla\psi(x^k) - \nabla\psi(x^{k+1}), x^* - x^k \right\rangle \leq \gamma_k\langle\nabla f(x^k), x^* - x^{k+1}\rangle + \gamma_k h(x^*) - \gamma_k h(x^{k+1}). \tag{24}$$

Combining (24), (15) and Lemma 1, we can get

$$\begin{aligned} \gamma_k\langle\nabla f(x^k), x^{k+1} - x^*\rangle \leq V_\psi(x^*, x^k) - V_\psi(x^*, x^{k+1}) - \frac{\sigma}{2}\|x^{k+1} - x^k\|^2 \\ + \gamma_k h(x^*) - \gamma_k h(x^{k+1}). \end{aligned} \tag{25}$$

Note that (12) and (16) still hold, so we have

$$\gamma_k(f(x^k)+h(x^{k+1})-F(x^*)) = \gamma_k(f(x^k)-f(x^*)+h(x^{k+1})-h(x^*)) \leq V_\psi(x^*,x^k)-V_\psi(x^*,x^{k+1})+\frac{\gamma_k^2}{2\sigma}\|\nabla f(x^k)\|_*^2. \quad (26)$$

Then we can obtain the following important inequality

$$f(x^k)+h(x^{k+1})-F(x^*) \leq \frac{1}{\gamma_k}\left(V_\psi(x^*,x^k)-V_\psi(x^*,x^{k+1})\right)+\frac{\gamma_k}{2\sigma}\|\nabla f(x^k)\|_*^2. \quad (27)$$

Similar to the analysis in (20), since the $\gamma_k$ we selected is exactly the same, we get

$$\sum_{k=1}^N \omega_k^{(m)}\left(f(x^k)+h(x^{k+1})-F(x^*)\right) \leq R\frac{\omega_N^{(m)}}{\gamma_N}+\sum_{k=1}^N \frac{\omega_k^{(m)}\gamma_k}{2\sigma}\|\nabla f(x^k)\|_*^2. \quad (28)$$

Now, we shall handle the left-hand side of (28). Because $\omega_k^{(m)}$ is non-increasing when $-1 \leq m \leq 0$ and $h$ is non-negative, we have

$$\begin{aligned}
&\sum_{k=1}^N \omega_k^{(m)}\left(f(x^k)+h(x^{k+1})-F(x^*)\right)\\
&=\sum_{k=1}^N \omega_k^{(m)}\left(f(x^k)+h(x^k)-h(x^k)+h(x^{k+1})-F(x^*)\right)\\
&=\sum_{k=1}^N \omega_k^{(m)}\left(F(x^k)-F(x^*)\right)+\sum_{k=1}^N \omega_k^{(m)}\left(h(x^{k+1})-h(x^k)\right)\\
&=\sum_{k=1}^N \omega_k^{(m)}\left(F(x^k)-F(x^*)\right)-\omega_1^{(m)}h(x^1)+\sum_{k=2}^N\left(\omega_{k-1}^{(m)}-\omega_k^{(m)}\right)h(x^k)+\omega_N^{(m)}h(x^{N+1})\\
&\geq \sum_{k=1}^N \omega_k^{(m)}\left(F(x^k)-F(x^*)\right)-\omega_1^{(m)}h(x^1). \quad (29)
\end{aligned}$$

Then by Jensen's Inequality, we have:

$$\begin{aligned}
&\left(\sum_{k=1}^N \omega_k^{(m)}\right)\left[F\left(\frac{\sum_{k=1}^N \omega_k^{(m)}x^k}{\sum_{k=1}^N \omega_k^{(m)}}\right)-F(x^*)\right]\\
&\leq \sum_{k=1}^N \omega_k^{(m)}(F(x^k)-F(x^*))\\
&\leq \sum_{k=1}^N \omega_k^{(m)}(f(x^k)+h(x^{k+1})-F(x^*))+\omega_1^{(m)}h(x^1)\\
&\leq R\frac{\omega_N^{(m)}}{\gamma_N}+\sum_{k=1}^N \frac{\omega_k^{(m)}\gamma_k}{2\sigma}\|\nabla f(x^k)\|_*^2+\omega_1^{(m)}h(x^1).
\end{aligned}$$

According to the analysis in (21), we know that the result below has been established when $-1 \le m \le 0$

$$\begin{aligned}&\frac{R\frac{\omega_N^{(m)}}{\gamma_N}+\sum_{k=1}^N\frac{\omega_k^{(m)}\gamma_k}{2\sigma}\|\nabla f(x^k)\|_*^2}{\sum_{k=1}^N\omega_k^{(m)}}\\ &\le\sqrt{\frac{R}{2\sigma}}\cdot\frac{N^{\frac{m+1}{2}}+\sum_{k=1}^N k^{\frac{m-1}{2}}}{\sum_{k=1}^N k^{\frac{m}{2}}}\cdot\max_{k=1,\dots,N}\|\nabla f(x^k)\|_*.\end{aligned}$$

Recall that $G_k$ can be rewritten as (19), so we get

$$\begin{aligned}\frac{\omega_1^{(m)}h(x^1)}{\sum_{k=1}^N\omega_k^{(m)}}&=\frac{\frac{1}{\gamma_k^m}h(x^k)}{\sum_{k=1}^N\frac{1}{\gamma_k^m}}=\frac{G_1^m h(x^1)}{\sum_{k=1}^N(G_k k^{\frac{a}{2}})^m}\\ &=\frac{G_1^m h(x^1)}{\sum_{k=1}^N\left(\max\limits_{j=1,\dots,k}(\|\nabla f(x^j)\|_*\cdot j^{\frac{1-a}{2}})k^{\frac{a}{2}}\right)^m}\\ &\le\frac{\|\nabla f(x^1)\|_*^m h(x^1)}{\sum_{k=1}^N\left(\max\limits_{k=1,\dots,N}\|\nabla f(x^k)\|_*\cdot k^{\frac{1-a}{2}}\cdot k^{\frac{a}{2}}\right)^m}\\ &=\left(\frac{\|\nabla f(x^1)\|_*}{\max\limits_{k=1,\dots,N}\|\nabla f(x^k)\|_*}\right)^m\cdot\frac{h(x^1)}{\sum_{k=1}^N k^{\frac{m}{2}}}.\end{aligned}$$

Thus, by setting $\widehat{x}=\frac{1}{\sum_{k=1}^N\omega_k^{(m)}}\sum_{k=1}^N\omega_k^{(m)}x^k$, we derive the following inequality

$$\begin{aligned}F(\widehat{x})-F(x^*)\le\sqrt{\frac{R}{2\sigma}}&\cdot\frac{N^{\frac{m+1}{2}}+\sum_{k=1}^N k^{\frac{m-1}{2}}}{\sum_{k=1}^N k^{\frac{m}{2}}}\cdot\max_{k=1,\dots,N}\|\nabla f(x^k)\|_*\\ &+\left(\frac{\|\nabla f(x^1)\|_*}{\max\limits_{k=1,\dots,N}\|\nabla f(x^k)\|_*}\right)^m\cdot\frac{h(x^1)}{\sum_{k=1}^N k^{\frac{m}{2}}}.\end{aligned}$$

This completes the proof. □

By setting $m = 0$, we can obtain the following corollary, which reveals the optimal convergence rate of the composite mirror descent method under the adaptive time-varying step size rules.

**Corollary 2** *Let $f$ be a convex continuous function and $h$ be a non-negative convex function, then for problem (23), by Algorithm 2, for any fixed $a \in [0,1]$ and $m = 0$, with the step sizes (8), it satisfies the following inequality*

$$F\left(\frac{1}{N}\sum_{k=1}^N x^k\right)-F(x^*)\le 3\sqrt{\frac{R}{2\sigma}}\cdot\frac{1}{\sqrt{N}}\cdot\max_{k=1,\dots,N}\|\nabla f(x^k)\|_*+\frac{h(x^1)}{N}.\quad(30)$$

**Remark 2** *In [1, Corollary 4.3], it is required to select the initial point* $x^1$ *satisfying* $h(x^1) = 0$*, which is not necessary for our theory.*

**Remark 3** *As revealed by the formula of the corollary, if we can separate a non-negative convex function in a convex optimization problem, then the part corresponding to the non-negative convex function* $h(x)$ *will converge at a faster rate* $O(1/N)$*, and this perspective is not mentioned in [1].*

## Conclusion

In this work, we correct the proofs in the part of the recent study [1] connected with the analysis of the convergence rate of the mirror descent method and composite mirror descent method under adaptive time-varying step size rules. Furthermore, we introduce a Lipschitz-free mirror descent algorithm that achieves weak ergodic convergence, extending the results of mirror descent beyond the traditional Lipschitz continuity setting. This advancement broadens the understanding of mirror descent methods and opens new possibilities for their application in non-smooth and non-Lipschitz optimization problems.

## Funding

This research was supported by the National Natural Science Foundation of China under grant 12171021, and the Fundamental Research Funds for the Central Universities.

## Data Availability

The manuscript has no associated data.

## References

1. Alkousa, M., Stonyakin, F., Abdo, A., Alcheikh, M. (2024). Optimal Convergence Rate for Mirror Descent Methods with Special Time-Varying Step Sizes Rules. In: Eremeev, A., Khachay, M., Kochetov, Y., Mazalov, V., Pardalos, P. (eds) Mathematical Optimization Theory and Operations Research: Recent Trends. MOTOR 2024. Communications in Computer and Information Science, vol 2239. Springer, Cham. https://doi.org/10.1007/978-3-031-73365-9_1.
2. Nemirovsky, A., Yudin D.: Problem Complexity and Method Efficiency in Optimization. J. Wiley & Sons, New York 1983.
3. Nemirovskii, A.: Efficient methods for large-scale convex optimization problems. Ekonomika i Matematicheskie Metody, 1979. (in Russian)
4. Beck, A., Teboulle, M.: Mirror descent and nonlinear projected subgradient methods for convex optimization. Oper. Res. Lett., 31(3), pp. 167–175, 2003.
5. Doan, T.T., Bose, S., Nguyen, D.H., Beck, C.L.: Convergence of the Iterates in Mirror Descent Methods. IEEE Control Systems Letters, **3**(1), pp. 114–119, 2019.

6. Beck, A., Ben-Tal, A., Guttmann-Beck, N., Tetruashvili, L.: The comirror algorithm for solving nonsmooth constrained convex problems. Operations Research Letters, **38**(6), pp. 493–498, 2010.
7. Ben-Tal, A., Margalit, T., Nemirovski, A.: The Ordered Subsets Mirror Descent Optimization Method with Applications to Tomography. SIAM Journal on Optimization, **12**(1), pp. 79–108, 2001.
8. Nazin, A., Anulova, S., Tremba, A.: Application of the Mirror Descent Method to Minimize Average Losses Coming by a Poisson Flow. European Control Conference (ECC) June, pp. 24–27, 2014.
9. Y. Nesterov, Introductory lectures on convex optimization: A basic course, Springer Science & Business Media, 2004.
10. Y. Nesterov, Lectures on convex optimization, volume 137, Springer, 2018.
11. S. Bubeck, Convex optimization: Algorithms and complexity. Foundations and Trends Trends® in Machine Learning, 8(3-4): 231-357, 2015.
12. Z. Zhu, Y. Zhang, Y. Xia, Convergence rate of projected subgradient method with time-varying step-sizes, Optim Lett 19, 1027–1031 (2025). https://doi.org/10.1007/s11590-024-02142-9.
13. Y. Xia, Y. Zhang, Z. Zhu, Lipschitz-free Projected Subgradient Method with Time-varying Step-size, 2024. https://optimization-online.org/?p=27932. arXiv:2410.22336.
14. A.Beck, First-order methods in optimization. Society for Industrial and Applied Mathematics, 2017.
15. G. Lan, First-order and Stochastic Optimization Methods for Machine Learning, Springer-Nature, 2020