Convergence Rate of Inertial Forward-Backward Algorithms Based on the Local Error Bound Condition

Hongwei Liu¹ · Ting Wang¹ · Zexian Liu² ·

Received: date / Accepted: date

Abstract The "Inertial Forward-Backward algorithm" (IFB) is a powerful tool for convex nonsmooth minimization problems, and the "fast iterative shrinkage-thresholding algorithm" (FISTA) is one of the IFB with the property that is computational simplicity and better global convergence rate of function value, however, no convergence of iterates generated by FISTA has been proved. In this paper, we exploit some assumption conditions for the important parameter $t_k$ in IFB, and establish the strong or weak convergence of the iterates generated by the IFB algorithms with these $t_k$ satisfying the above assumption conditions in Hilbert space under the local error bound condition. Further, we discuss four options of $t_k$ to analyze the convergence of function value and to establish the sublinear convergence of the iterates generated by the corresponding IFB algorithms under the local error bound condition. It is remarkable that the sublinear convergence of the iterates generated by the original FISTA is established under the local error bound condition and the IFB algorithms with some $t_k$ mentioned above can achieve sublinear convergence rate $o\left(\frac{1}{k^p}\right)$ for any positive integer $p$. Some numerical experiments are conducted to illustrate our results.

Keywords Inertial Forward-Backward algorithm, local error bound condition, rate of convergence

Mathematics Subject Classification (2000) 94A12 · 65K10 · 94A08 · 90C25 ·

1 Introduction

Let $H$ be a real Hilbert space. $f : H \to R$ be a smooth convex function and continuously differentiable with $L_f$-Lipschitz continuous gradient, and $g : H \to R \cup \{+\infty\}$ be a proper lower semi-continuous convex

Ting Wang (✉)
E-mail: wangting76400163.com
Hongwei Liu
E-mail: hwliu@mail.xidian.edu.cn
Zexian Liu
E-mail: liuzexian2008@163.com

¹ School of Mathematics and Statistics, Xidian University, Xi’an, 710126, China
² State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering computing, AMSS, Chinese Academy of Sciences, Beijing, 100190, China.
function. We also assume that the proximal operator of \( \lambda g \), i.e.,

\[
\text{prox}_{\lambda g} (\cdot) = \arg\min_{x \in H} \left\{ g(x) + \frac{1}{2\lambda} \| x - \cdot \|^2 \right\}
\]

can be easily computed for all \( \lambda > 0 \).

In this paper, we consider the following problem:

\[
(P) \quad \min_{x \in H} F(x) := f(x) + g(x).
\]

We assume that problem (P) is solvable, i.e., \( X^* := \arg\min F \neq \emptyset \), and for \( x^* \in X^* \) we set \( F^* := F(x^*) \).

In order to solve the problem (P), several algorithms have been proposed based on the use of the proximal operator due to the non-differentiable part. One can consult [10,11,18] for a recent account on the proximal-based algorithms that play a central role in nonsmooth optimization. A typical optimization strategy for solving problem (P) is the Inertial Forward-Backward algorithm (IFB).

**Algorithm 1** Inertial Forward-Backward algorithm (IFB)

**Step 0.** Take \( y_1 = x_0 \in H^*, t_1 = 1 \). Input \( \lambda = \frac{1}{t_1^2}, \) where \( \mu \in [0,1] \).

**Step k.** Compute

\[
x_k = p_{\lambda g}(y_k) = \text{prox}_{\lambda g}(y_k - \lambda \nabla f(y_k))
\]

\[
y_{k+1} = x_k + \gamma_k (x_k - x_{k-1}) \quad \text{where} \quad \gamma_k = \frac{t_{k-1}}{t_{k+1}}.
\]

In view of the composition of IFB, we can easily found that the inertial term \( \gamma_k \) plays an important role for improving the speed of convergence of IFB. Based on Nesterov’s extrapolation techniques [13], Beck and Teboulle [4] proposed a “fast iterative shrinkage-thresholding algorithm” (FISTA) with \( t_1 = 1 \) and \( t_{k+1} = \frac{k+1+\sqrt{4k^2+1}}{2k} \) for solving (P). The remarkable property of this algorithm is the computational simplicity and the significantly better global rate of convergence of the function value, that is \( F(x_k) - F(x^*) \approx O \left( \frac{1}{k} \right) \).

Several variants of FISTA considered in works such as [3,6,7,8,12,15,16], the properties such as convergence of the iterates and rate of convergence of function value have also been studied.

Chambolle and Dossal [6] pointed out that FISTA satisfies a better worst-case estimate, however, the convergence of the iterates is not known. They proposed a new \( t_k = \frac{k+1+\alpha}{\alpha} \) (\( \alpha > 2 \)) to show that the iterates generated by the corresponding IFB, named “FISTA\text{CD}”, is converges weakly to the minimizer of \( F \).

Attouch and Peypouquet [1] further proved that the sequences generated by FISTA\text{CD} approximate the optimal value of the problem with a rate that is strictly faster than \( O \left( \frac{1}{k^2} \right) \), namely \( F(x_k) - F(x^*) = o \left( \frac{1}{k^2} \right) \).

Apidopoulos et al. [3] noticed that the basic idea of the choice of \( t_k \) in [2,4,6] is the Nesterov’s rule: \( t_k^2 - t_{k+1}^2 + t_{k+1} \geq 0 \), and they focus on the case that the Nesterov’s rule is not satisfied. They studied the \( \gamma_k = \frac{b}{n+\alpha} \) with \( 0 < b < 3 \) and found that the exact estimate bounds are the following: \( F(x_k) - F(x^*) = O \left( \frac{1}{k^{3+b}} \right) \).

Attouch and Peypouquet [2] considered various options for the sequence \( \gamma_k \) to analyze the convergence rate of the function value and weak convergence of the iterates under the given assumptions. Further, they showed that the strong convergence of iterates can be satisfied for the special options of \( f \). Wen, Chen and Pong [19] showed that under the local error bound condition [17], there exists a threshold such that if the \( \gamma_k \) are chosen below this threshold, the \( R \)-linear convergence of both the sequence \( \{x_k\} \) and the corresponding sequence of objective values \( \{F(x_k)\} \) can be satisfied for the case of \( f \) possibly unconvex.
Further, under the error bound condition, they pointed out that in the case that $f$ is convex, the sequences \( \{x_k\} \) and \( \{F(x_k)\} \) generated by FISTA with fixed restart \cite{14} are \( R \)-linearly convergent; and the local convergence rates of the iterates generated by FISTA for solving (P) is still unknown, even under the local error bound condition.

Notice that the rate of convergence of function value and the convergence of iterates are constantly improved, but the convergence of iterates for the original FISTA remains an open problem, that is the first point this work focuses.

We also pay attention to the Nesterov’s rule: 
\[
t_2 - t_2 + t_k + 1 + t_k < 0.
\]
For the \( t_k \) satisfies it, we can derive that \( t_k - t_k < 1 \) and \( \sum_{k=1}^{\infty} \frac{1}{t_k} \) is divergent, which will greatly limit the choice of \( t_k \). What our expect is whether we can find the more suitable \( t_k \) and obtain the improved theoretical results if we replace the Nesterov’s rule by some new we proposed.

**Contributions.**

In this paper, based on the local error bound condition, we exploit some assumption conditions for the important parameter \( t_k \) in IFB, and discuss the convergence results including convergence rate of function value and strong or weak convergence of iterates generated by the corresponding IFB. We discuss four choices of \( t_k \), which includes the ones in original FISTA and FISTA_CD and satisfies our assumption conditions, and separately analyze the convergence of the function value and establish the sublinear convergence of the iterates generate by the corresponding IFB. It is remarkable that the sublinear convergence rate of the iterates generated by the original FISTA is established under the local error bound condition and the IFB algorithms with some \( t_k \) mentioned above can achieve sublinear convergence rate \( o \left( \frac{1}{k^p} \right) \) for any positive integer \( p \).

## 2 Some new assumption conditions for \( t_k \) and the convergence of the corresponding IFB algorithms

In this section, we derive some assumption conditions for the abstract \( t_k \) in IFB, and analyze the convergence of iterates generated by the corresponding IFB.

We found that the theoretical analyses in \cite{1,3,4,6,19} are all based on the following inequality:
\[
F(p_{\lambda g}(y)) - F(x) \leq \frac{1}{2\lambda} \|x - y\|^2 - \frac{1}{2\lambda} \|p_{\lambda g}(y) - y\|^2 - \frac{1}{2\lambda} \|p_{\lambda g}(y) - x\|^2.
\]

Here, we derive a key result, which is stronger than the above one.

**Lemma 2.1** For any \( y \in \mathbb{R}^n \), \( \lambda = \frac{\mu}{2\gamma} \), where \( \mu \in [0, 1] \), we have,
\[
F(p_{\lambda g}(y)) \leq F(x) + \frac{1}{2\lambda} \|x - y\|^2 - \frac{1}{2\lambda} \|p_{\lambda g}(y) - y\|^2 - \frac{1}{2\lambda} \|p_{\lambda g}(y) - x\|^2.
\]

**Proof.** Following from
\[
p_{\lambda g}(y) = \arg\min_{x \in H} \left\{ f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2\lambda} \|x - y\|^2 + g(x) \right\},
\]
\[
F(p_{\lambda g}(y)) \leq F(x) + \frac{1}{2\lambda} \|x - y\|^2 - \frac{1}{2\lambda} \|p_{\lambda g}(y) - y\|^2 - \frac{1}{2\lambda} \|p_{\lambda g}(y) - x\|^2.
\]
Proof. Applying the inequality (2.2), we have
\[
g(p_\lambda(y)) + \langle \nabla f(y), p_\lambda(y) - y \rangle + \frac{1}{2\lambda} \|p_\lambda(y) - y\|^2 + \frac{1}{2\lambda} \|p_\lambda(x) - x\|^2 \leq g(x) + \langle \nabla f(y), x - y \rangle + \frac{1}{2\lambda} \|x - y\|^2. \tag{2.3}
\]
Since $\nabla f$ is Lipschitz continuous and $\lambda = \frac{1}{\gamma f}$, we have that
\[
f(p_\lambda(y)) \leq f(y) + \langle \nabla f(y), p_\lambda(y) - y \rangle + \frac{\mu}{2\lambda} \|p_\lambda(y) - y\|^2. \tag{2.4}
\]
Summing the (2.3) and (2.4), and combining the fact that $f$ is convex, we obtain that
\[
F(p_\lambda(y)) \leq f(y) + g(x) + \langle \nabla f(y), x - y \rangle - \frac{1-\mu}{2\lambda} \|p_\lambda(y) - y\|^2 + \frac{1}{2\lambda} \|x - y\|^2 - \frac{\mu}{2\lambda} \|p_\lambda(y) - x\|^2 \leq F(x) - \frac{1-\mu}{2\lambda} \|p_\lambda(y) - y\|^2 + \frac{\mu}{2\lambda} \|x - y\|^2 - \frac{\mu}{2\lambda} \|p_\lambda(y) - x\|^2. \tag{2.5}
\]

Next, we give a very weak assumption to show that the sequence $\{F(x_k)\}$, which is generated by Algorithm 1 with $0 \leq \gamma_k \leq 1$ for $k$ is large sufficiently, converges to $F(x^*)$ independent on $t_k$.

**Assumption A0:** For any $\xi_0 \geq F^*$, there exist a $\varepsilon_0 > 0$ and $\tau_0 > 0$ such that
\[
dist(x, X^*) \leq \tau_0 \tag{2.6}
\]
whenever $\|p_{\frac{1}{\gamma f}}(x) - x\| < \varepsilon_0$ and $F(x) \leq \xi_0$.

**Remark 1.** Note that Assumption A0 can be derived by the assumption that $F$ is boundedness of level sets.

**Lemma 2.2** boundedness of level sets [20] For $\lambda_1 \geq \lambda_2 > 0$, we have
\[
\|p_{\lambda_1 \gamma}(x) - x\| \geq \|p_{\lambda_2 \gamma}(x) - x\| \quad \text{and} \quad \|p_{\lambda_1 \gamma}(x) - x\| \leq \|p_{\lambda_2 \gamma}(x) - x\| \frac{\lambda_1}{\lambda_2}. \tag{2.7}
\]

**Theorem 2.1** Let $(x_k), \{y_k\}$ be generated by Algorithm 1. Suppose that there exists a positive interger $k_0$ such that for $k > k_0$, $0 \leq \gamma_k \leq 1$. Then,
1) $\sum_{k=1}^\infty \|x_{k+1} - y_{k+1}\|^2$ is convergent.
2) $\lim_{k \to \infty} F(x_k) = F(x^*)$.

**Proof.** Applying the inequality (2.1) at the point $x = x_k$, $y = y_{k+1}$, we obtain
\[
\frac{1-\mu}{2\lambda} \|x_{k+1} - y_{k+1}\|^2 \leq \left( F(x_k) + \frac{\gamma_k^2}{2\lambda} \|x_k - x_{k-1}\|^2 \right) - \left( F(x_{k+1}) + \frac{1}{2\lambda} \|x_{k+1} - x_k\|^2 \right). \tag{2.8}
\]
Then, we can easily obtain result 1).

Setting $\xi_0 = F(x_{k_0+1}) + \frac{\mu}{2\lambda} \|x_{k_0+1} - x_{k_0}\|^2$.

From Lemma 2.2, the nonexpansiveness property of the proximal operator and $\nabla f$ is Lipschitz continuous, we obtain that
\[
\|p_{\frac{1}{\gamma f}}(x_k) - x_k\| \leq \frac{1}{\lambda L_f} \|p_{\lambda \gamma}(x_k) - x_k\| = \frac{1}{\lambda L_f} \|p_{\lambda \gamma}(x_k) - p_{\lambda \gamma}(y_k)\| \leq \left( 1 + \frac{1}{\lambda L_f} \right) \|x_k - y_k\|. \tag{2.9}
\]
It follows from (2.8) that for \( k \geq k_0 \), \( \{ F(x_{k+1}) + \frac{1}{k} \| x_{k+1} - x_k \|^2 \} \) is non-increasing, then, \( F(x_k) \leq \xi_0 \). Hence, Combining the Assumption A0, (2.9) and result 1), we have for \( \xi_0 = F(x_{k_0+1}) + \frac{1}{2\xi} \| x_{k_0+1} - x_{k_0} \|^2 \), there exists a \( \tau_0 > 0 \), for \( k \) is large sufficiently, such that
\[
\text{dist} (x_k, X^*) \leq \tau_0. \tag{2.10}
\]

In addition, applying the inequality (2.1) at the point \( y = y_{k+1} \), and \( x \) be an \( x_{k+1} \in X^* \) such that \( \text{dist} (x_{k+1}, X^*) = \| x_{k+1} - x^* \| \), we obtain
\[
F(x_{k+1}) - F(x^*) \leq \frac{1}{k} \| y_{k+1} - x_{k+1} \|^2 - \frac{1}{k} \| x_{k+1} - x^* \|^2
= \frac{1}{k} \| y_{k+1} - x_{k+1} \|^2 + \frac{1}{k} \langle y_{k+1} - x_{k+1}, x_{k+1} - x^* \rangle\tag{2.11}
\]
\[
\leq \frac{1}{k} \| y_{k+1} - x_{k+1} \|^2 + \frac{1}{k} \| y_{k+1} - x_{k+1} \| \text{dist} (x_{k+1}, X^*)
\]

Then, combining the result 1) with (2.10), we have \( \lim_{k \to \infty} F(x_k) = F(x^*) \). \( \square \)

The rest of this paper is based on the following assumption.

**Assumption A1**: [17] (Local error bound condition) For any \( \xi \geq F^* \), there exist a \( \varepsilon > 0 \) and \( \bar{r} > 0 \) such that
\[
\text{dist} (x, X^*) \leq \bar{r} \| p_{\bar{r}g}(x) - x \|\tag{2.12}
\]
whenever \( \| p_{\bar{r}g}(x) - x \| < \varepsilon \) and \( F(x) \leq \xi \).

Under the local error bound condition (Assumption A1), we will analyze the convergence of iterates and convergence rate of the function value for the Algorithm 1 with a class of abstract \( t_k \), which satisfy the following assumptions.

**Assumption A2**: \( \lim_{k \to \infty} t_k = +\infty \).

**Assumption A3**: There exists a positive constant \( p \) such that \( \lim_{k \to \infty} k^p \left( \frac{t_{k+1}}{t_k} - 1 \right) = c \), where \( c > 0 \).

**Remark 2**. It follows that \( \gamma_k \in [0,1], \forall k \geq k_0 \) where \( k_0 \) is sufficiently large and \( \lim_{k \to \infty} \frac{t_{k+1}}{t_k} = 1 \) from Assumptions A2 and A3.

**Lemma 2.3** Suppose that Assumptions A1–A3 hold. Let \( \{x_k\} \) be generated by Algorithm 1 and \( x^* \in X^* \). There exists a constant \( \tau_1 > 0 \) such that
\[
\forall k \geq 1, \quad F(x_{k+1}) - F(x^*) \leq \frac{\tau_1}{4} \| y_{k+1} - x_{k+1} \|^2.
\]

**Proof**. Similar with the proof of Theorem 2.1, we can deduce that for \( k \) sufficient large,
\[
\text{dist} (x_k, X^*) \leq \bar{r} \| p_{\bar{r}g}(x) - x \| \leq \frac{\bar{r}}{2\mu} \| p_{\bar{r}g}(x) - x \| \leq \frac{2\bar{r}}{\mu} \| x - y_k \|, \tag{2.13}
\]
where the second inequality of (2.13) because Lemma 2.2, and
\[
F(x_{k+1}) - F(x^*) \leq \frac{1}{k} \| y_{k+1} - x_{k+1} \|^2 + \frac{1}{k} \| y_{k+1} - x_{k+1} \| \text{dist} (x_{k+1}, X^*) \leq \frac{1}{k} \left( \frac{2\bar{r}}{\mu} + \frac{1}{2} \right) \| y_{k+1} - x_{k+1} \|^2. \tag{2.14}
\]
Therefore, there exists a \( \tau_1 \geq \frac{2\bar{r}}{\mu} + \frac{1}{2} \) such that the conclusion holds. \( \square \)

**Lemma 2.4** Suppose that Assumptions A1–A3 hold. Let \( \{x_k\} \) be generated by Algorithm 1 and \( x^* \in X^* \). Then,
\[
\sum_{k=1}^{\infty} t_{k+1}^2 (F(x_{k+1}) - F(x^*)) \text{ is convergent and } \| x_k - x_{k-1} \| \leq O \left( \frac{1}{t_k} \right).
\]
Proof. Applying the inequality (2.1) at the point \( x := \left(1 - \frac{1}{t_{k+1}}\right)x_k + \frac{1}{t_{k+1}}x^*, \ y := y_{k+1}\), we obtain

\[
F(x_{k+1}) \leq F\left((1 - \frac{1}{t_{k+1}})x_k + \frac{1}{t_{k+1}}x^*\right) + \frac{1}{t_{k+1}}\left(1 - \frac{1}{t_{k+1}}\right)x_k + \frac{1}{t_{k+1}}x^* - y_{k+1}\right)^2
- \frac{1}{2\lambda}\left(1 - \frac{1}{t_{k+1}}\right)\|x_k + \frac{1}{t_{k+1}}x^* - x_{k+1}\|^2 - \frac{1}{2\lambda}\|x_{k+1} - y_{k+1}\|^2
\leq \left(1 - \frac{1}{t_{k+1}}\right)F(x_k) + \frac{1}{t_{k+1}}F(x^*) + \frac{1}{2\lambda t_{k+1}}\|u_k\|^2 - \frac{1}{2\lambda t_{k+1}}\|u_{k+1}\|^2 - \frac{1}{2\lambda}\|x_{k+1} - y_{k+1}\|^2
\]

where \( u_k = t_kx_k - (t_k - 1)x_{k-1} - x^*\).

Further, multiplying above inequality by \( t_{k+1}^2\), we obtain

\[
t_{k+1}^2 (F(x_k) - F(x^*)) - t_{k+1}^2 (F(x_{k+1}) - F(x^*))
\geq \frac{1}{2\lambda}\|u_{k+1}\|^2 - \frac{1}{2\lambda}\|u_k\|^2 + \frac{1}{4\lambda t_{k+1}}\|x_{k+1} - y_{k+1}\|^2 + (t_k^2 + t_k + 1 - t_{k+1}^2) (F(x_k) - F(x^*))
= \frac{1}{2\lambda}\|u_{k+1}\|^2 - \frac{1}{2\lambda}\|u_k\|^2 + \frac{1}{4\lambda t_{k+1}}\|x_{k+1} - y_{k+1}\|^2 + \frac{\rho_k^2}{4\lambda t_{k+1}}\|x_{k+1} - y_{k+1}\|^2 - \rho_k (F(x_k) - F(x^*))
\geq \frac{1}{2\lambda}\|u_{k+1}\|^2 - \frac{1}{2\lambda}\|u_k\|^2 + \frac{1}{4\lambda t_{k+1}}\|x_{k+1} - y_{k+1}\|^2 + \frac{\rho_k^2}{4\lambda t_{k+1}}\|F(x_{k+1}) - F(x^*)\| - \rho_k (F(x_k) - F(x^*))
\]

(2.15)

where \( \rho_k = t_{k+1}^2 - t_k^2 - t_{k+1}\) and the last inequality is follows from the Lemma 2.3.

Rearranging the inequality (2.15), we have

\[
(t_k^2 + \rho_k) (F(x_k) - F(x^*)) - (t_{k+1}^2 + \rho_{k+1}) (F(x_{k+1}) - F(x^*))
\geq \frac{1}{2\lambda}\|u_{k+1}\|^2 - \frac{1}{2\lambda}\|u_k\|^2 + \frac{1}{4\lambda t_{k+1}}\|x_{k+1} - y_{k+1}\|^2 + t_k^2 \left(\frac{\rho_k^2}{4\lambda t_{k+1}} - \rho_{k+1}\right) (F(x_{k+1}) - F(x^*))
\geq \frac{1}{2\lambda}\|u_{k+1}\|^2 - \frac{1}{2\lambda}\|u_k\|^2 + \frac{1}{4\lambda t_{k+1}}\|x_{k+1} - y_{k+1}\|^2,
\]

(2.16)

where \( t_{k+1}^2 \left(\frac{\rho_k^2}{4\lambda t_{k+1}} - \rho_{k+1}\right) (F(x_{k+1}) - F(x^*)) > 0\) because the fact that \( \lim_{k \to \infty} \frac{\rho_{k+1}}{t_{k+1}^2} = 0\), which can be deduced by \( \lim_{k \to \infty} \frac{t_{k+1}}{t_k} = 1\) from Assumption A3.

Denote that \( \phi_k = (t_k^2 + \rho_k) (F(x_k) - F(x^*)) + \frac{1}{2\lambda}\|u_k\|^2\).

Rearranging (2.16), we have

\[
\phi_k - \phi_{k+1} \geq \frac{1 - \mu}{4\lambda} t_{k+1}^2 \|x_{k+1} - y_{k+1}\|^2.
\]

(2.17)

Observing (2.17) that the sequence \( \{\phi_k\} \) is nonincreasing and moreover, \( \{\phi_k\} \) is convergent with the fact that \( \{\phi_k\} \) is bound below. Further, we can deduce that \( \sum_{k=1}^{\infty} t_{k+1}^2 \|x_{k+1} - y_{k+1}\|^2 \) is convergent.

Applying the Lemma 2.3, we have \( \sum_{k=1}^{\infty} t_{k+1}^2 (F(x_{k+1}) - F(x^*)) \) is obviously convergent.

Notice that \( \lim_{k \to \infty} (t_k^2 + \rho_k) (F(x_k) - F(x^*)) = 0\) by \( \lim_{k \to \infty} \frac{\rho_k}{t_k} = 0\). Combining with the convergence of \( \{\phi_k\} \), we have \( \{\|u_k\|\} \) is convergent. There exists a positive constant \( t_1 \) such that \( \|u_k\| \leq t_1 \), i.e., \( \|x_k - x^*\| - \left(1 - \frac{1}{t_k}\right)\|x_{k-1} - x^*\| \leq \frac{t_1}{t_k} \), which implies that \( \|x_k - x^*\| \leq \frac{t_1}{t_k} + \left(1 - \frac{1}{t_k}\right)\|x_{k-1} - x^*\| \). An immediate recurrence shows that \( \|x_k - x^*\| \) is bounded. Further, \( t_k \|x_k - x_{k-1}\| \) is also bounded, which means that the conclusion \( \|x_k - x_{k-1}\| \leq O\left(\frac{1}{t_k}\right) \) holds.

Remark 3. Lemma 2.4 implies that \( F(x_k) - F(x^*) = o\left(\frac{1}{t_k}\right) \).

Theorem 2.2 Suppose that Assumptions A1–A3 hold and \( \sum_{k=1}^{\infty} \frac{1}{t_k} \) is convergent. Then, the iterates \( \{x_k\} \) strongly converges to a minimizer of \( F \).
Proof. In the proof of Lemma 2.4, we have showed that there exists a constant \( t_2 > 0 \) such that \( \| x_k - x_{k-1} \| \leq \frac{t_2}{k} \). Recalling the assumption that \( \sum_{k=1}^{\infty} \frac{1}{k} \) is convergent, we can deduce that the sequence \( \{ x_k \} \) is a Cauchy series. Suppose that \( \lim_{k \to \infty} x_k = \bar{x} \), we conclude that \( \{ x_k \} \) strongly converges to \( \bar{x} \in X^* \) since \( F \) is lower semi-continuous convex.

In the following, we consider the case that \( \sum_{k=1}^{\infty} \frac{1}{\lambda_k} \) is divergent by adding the following Assumption \( A^+ \):

**Assumption \( A^+ \):** There exist \( 0 < M < 2 \) and \( m > 0 \) such that \( t_{k+1} - t_k \leq M, \forall k > m \).

**Remark 4.** We see that Assumption \( A^+ \) implies \( \sum_{k=1}^{\infty} \frac{1}{\lambda_k} \) is divergent.

**Lemma 2.5** Suppose that Assumptions \( A_1 - A_3 \) and \( A^+ \) hold. Let \( \{ x_k \} \) be generated by Algorithm 1 and \( x^* \in X^* \). Then, we have \( F(x_k) - F(x^*) = o \left( \frac{1}{\lambda_k} \right) \) and \( \| x_k - x_{k-1} \| = o \left( \frac{1}{\lambda_k} \right) \).

**Proof.** Applying (2.8) we can easily obtain that

\[
( F(x_k) - F(x^*) ) + \left( \frac{1}{2\lambda_k} \left( \frac{t_k}{t_{k+1}} - 1 \right) \right)^2 \| x_k - x_{k-1} \|^2 \geq ( F(x_{k+1}) - F(x^*) ) + \frac{1}{2\lambda_k} \| x_{k+1} - x_k \|^2. \tag{2.18}
\]

Denote \( \psi_k = ( F(x_k) - F(x^*) ) + \frac{1}{2\lambda_k} \| x_k - x_{k-1} \|^2 \) and \( \beta_k = \frac{2}{t_k} \left( (2-M) t_k^2 + 2M - 1 \right) t_k - M \).

Multiplying (2.18) by \( t_{k+1}^3 \psi_{k+1} \), it follows from the Assumption \( A^+ \) that

\[
\begin{align*}
t_{k+1}^3 \psi_{k+1} & \leq ( F(x_k) - F(x^*) ) + (t_{k+1}^2 - t_k^2) ( F(x_k) - F(x^*) ) + \frac{1}{2\lambda_k} \| x_{k+1} - 1 \| \| x_k - x_{k-1} \|^2 \\
& \leq ( F(x_k) - F(x^*) ) + (t_{k+1} + t_k - 1) ( F(x_k) - F(x^*) ) + \frac{1}{2\lambda_k} (t_k + M) (t_k - 1)^2 \| x_k - x_{k-1} \|^2 \\
& \leq t_k^2 \psi_k + M \left( t_k^2 + t_k + 1 \right) ( F(x_k) - F(x^*) ) - \beta_k \| x_k - x_{k-1} \|^2 \\
\end{align*}
\tag{2.19}
\]

Then, we have \( t_{k+1}^3 \psi_{k+1} - t_k^2 \psi_k \leq M \left( t_k^2 + t_k + 1 \right) ( F(x_k) - F(x^*) ) \). It follows that \( \lim_{k \to \infty} t_k^2 \psi_k \) exists from Lemma 2.4 and the fact that \( \{ t_k^2 \psi_k \} \) is bounded below.

Now, we prove that \( \lim_{k \to \infty} t_k^2 \psi_k = 0 \). Recalling (2.19) that

\[
\beta_k \| x_k - x_{k-1} \|^2 \leq t_k^2 \psi_k - t_{k+1}^3 \psi_{k+1} + M \left( t_k^2 + t_k + 1 \right) ( F(x_k) - F(x^*) ),
\]

which implies that \( \sum_{k=1}^{\infty} t_k^2 \psi_k \) is convergent by Lemma 2.4. Further, we obtain that \( \sum_{k=1}^{\infty} t_k^2 \psi_k \) is convergent. It follows that \( \liminf_{k \to \infty} t_k^2 \psi_k = 0 \) from the convergence of \( \sum_{k=1}^{\infty} t_k^2 \psi_k \) and the divergence of \( \sum_{k=1}^{\infty} \frac{1}{\lambda_k} \), which means that \( \lim_{k \to \infty} t_k^2 \psi_k = 0 \). Hence, we can easily obtain the result.

In order to prove the convergence of iterates, we need the following lemma.

**Lemma 2.6** [9] Let \( X^* \) be a nonempty subset of \( H \) and \( \{ x_k \} \) a sequence of elements of \( H \). Assume that

(i) every sequential weak cluster point of \( \{ x_k \} \), as \( k \to \infty \), belongs to \( X^* \);

(ii) for every \( x^* \in X^* \), \( \lim_{k \to \infty} \| x_k - x^* \| \) exists.

Then \( \{ x_k \} \) converges weakly to a point in \( X^* \) as \( k \to \infty \).

**Theorem 2.3** Suppose that Assumptions \( A_1 - A_3 \) and \( A^+ \) hold. We have that the sequence \( \{ x_k \} \) generated by Algorithm 1 converges weakly to its limit belongs to \( X^* \).
Proof. Following from the proof of Lemma 2.4 that \( \{\|u_k\|\} \) is convergent, where \( u_k = t_k x_k - (t_k - 1) x_{k-1} - x^* \), and \( \|x_k - x^*\| \) is bounded, which implies that \( \|x_k\| \) is bounded. Then, there exists a subsequence \( \{x_{k_j}\} \), which is weak convergent to \( \bar{x} \). Since the convex function \( F \) is lower semi-continuous, it is lower semi-continuous for the weak topology and hence satisfies

\[
F(\bar{x}) \leq \liminf_{j \to \infty} F(x_{k_j}) = \lim_{k \to \infty} F(x_k) = F(x^*).
\]

On the other hand, we have \( F(\bar{x}) \geq F(x^*) \), it ensures that \( \bar{x} \in X^* \), which means that the first point of the Lemma 2.6 holds.

We see that \( \|u_k\|^2 = (t_k - 1)^2 \|x_k - x_{k-1}\|^2 + \|x_k - x^*\|^2 + 2 (t_k - 1) (x_k - x_{k-1}, x_k - x^*) \). Notice Lemma 2.5 gives \( \lim_{k \to \infty} t_k \|x_k - x_{k-1}\| = 0 \). Combining with the convergence of \( \{\|u_k\|\} \) and the boundedness of \( \{\|x_k - x^*\|\} \), we deduce that \( \{\|x_k - x^*\|\} \) is convergent, that means the second point of the Lemma 2.6 holds. By Lemma 2.6, we obtain the conclusion. \( \square \)

Theorem 2.4 Suppose that Assumptions A1–A3 and \( A^+ \) hold and \( \sum_{k=1}^{\infty} \frac{1}{t_k^5} \) is convergent. Then, the iterates \( \{x_k\} \) strongly converges to a minimizer of \( F \).

Proof. In the proof of Lemma 2.5, we show that \( \lim_{k \to \infty} t_k^{1/5} \|x_k - x_{k-1}\| = 0 \), i.e., there exists a positive constant \( t_3 \) such that

\[
\|x_k - x_{k-1}\| \leq \frac{t_3}{t_k^{1/5}}.
\]

Recalling the assumption that \( \sum_{k=1}^{\infty} \frac{1}{t_k^5} \) is convergent, we can deduce that the sequence \( \{x_k\} \) is a Cauchy series. Combining with the Theorem 2.3, we conclude that \( \{x_k\} \) strongly converges to \( \bar{x} \in X^* \). \( \square \)

3 Several options for \( t_k \) and the sublinear convergence of the corresponding IFB algorithms

In Section 2, we see that both of the convergence rate of function value and the convergence of iterates generated by Algorithm 1 are greatly depend on \( t_k \). Based on this, we will discuss some options of \( t_k \) including the one in original FISTA and the famous FISTA_CD to analyze the convergence rate of function value and convergence of iterates, even the rate of convergence of iterates.

Option 1. \( t_1 = 1 \) and \( t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2t_k} \). It is noted that Algorithm 1 with the \( t_k \) is the original FISTA.

Corollary 3.1 Suppose that Assumption A1 holds. Let \( \{x_k\} \) be generated by Algorithm 1 and \( x^* \in X^* \). Then,

1) \( F(x_k) - F(x^*) = o\left(\frac{1}{k}\right) \) and \( \|x_k - x_{k-1}\| = o\left(\frac{1}{\sqrt{k}}\right) \).
2) \( \{x_k\} \) is converges sublinearly to \( \bar{x} \in X^* \) at the \( o\left(\frac{1}{k^{1/5}}\right) \) rate of convergence.

Proof. We can easily obtain that \( \lim_{k \to \infty} t_k = +\infty \), \( \lim_{k \to \infty} k \left( \frac{t_k}{t_{k+1}} - 1 \right) = 1 \), and \( t_{k+1} - t_k \leq 1 \), which means that Assumptions A2, A3 and \( A^+ \) hold. We can also obtain that \( \lim_{k \to \infty} \frac{k}{t_k} = \frac{1}{2} \), then, \( \sum_{k=1}^{\infty} \frac{1}{t_k} \) is convergent. Hence, we can deduce the result 1) from Lemma 2.5 and \( \{x_k\} \) is strongly converges to \( \bar{x} \in X^* \) form Theorem 2.4.
We have $\lim_{k \to \infty} k^{1.5} \|x_k - x_{k-1}\| = 0$, i.e., for $\forall \varepsilon > 0$, there exists a positive constant $N$ such that $\|x_k - x_{k-1}\| \leq \frac{\varepsilon}{k^{1.5}}$, $\forall k > N$. Then, we can deduce that

$$\forall p > 1, \quad \|x_{k+p} - x_k\| \leq \sum_{i=k+1}^{k+p} \|x_i - x_{i-1}\| \leq \varepsilon \sum_{i=k+1}^{k+p} \frac{1}{r_i} \leq \varepsilon \int_{k}^{k+p} \frac{1}{r}dx.$$ 

Then

$$\|x_k - \bar{x}\| \leq 2\varepsilon \frac{1}{k^{0.5}}, \text{ as } p \to \infty.$$ 

Hence, result 2) holds. \(\square\)

**Remark 5.** We see that under the Assumption A1, the strong convergence of iterates of the original FISTA with fixed stepsize $\lambda = \frac{\mu}{t^2}$, $\mu \in [0, 1]$ has been proved, which is an open question that has puzzled many scholars for a long time. We also improve the rate of convergence of function value and obtain the sublinear convergence rate of iterates, which to our knowledge has not been established previously.

**Option 2.** $t_k = \frac{k^{r-1+a}}{n}$ where $r > 0$ and $\begin{cases} a \geq 1, & \text{if } r = 1 \\ a > 0, & \text{if } r \neq 1 \end{cases}$.

It is noted that when $r = 1$ and $a > 2$, Algorithm 1 with the parameter $t_k$ reduces to FISTA [CD] [6].

**Corollary 3.2** Suppose that Assumption A1 holds. Let $\{x_k\}$ be generated by Algorithm 1 with $t_k = \frac{k^{r-1+a}}{n}$ ($0 < r \leq 1$) and $x^* \in X^*$. Then, we have

1) For any positive integer $p$, $F(x_k) - F(x^*) = o\left(\sum_{i=1}^{k} 1^\lambda \right)$ and $\|x_k - x_{k-1}\| = o\left(\sum_{i=1}^{k} 1^\lambda \right)$.

2) For any positive integer $p$, $\{x_k\}$ converges sublinearly to $\bar{x} \in X^*$ at the $o\left(\sum_{i=1}^{k} 1^\lambda \right)$ rate of convergence.

**Proof.** It is easy to verify that $\lim_{k \to \infty} t_k = +\infty$ and $\lim_{k \to \infty} k \left(\frac{t_{k+1}}{t_k} - 1\right) = r$ for the case that $0 < r \leq 1$, which means that Assumptions A2 and A3 hold. Then, Lemma 2.4 holds.

In order to prove the result 1), we first prove that for any positive integer $p$, $\sum_{k=1}^{\infty} t_k^{(2-p+\frac{p}{r})} \psi_k$ is convergent by induction, where $\psi_k = (F(x_k) - F(x^*)) + \frac{\mu}{2\lambda} \|x_k - x_{k-1}\|^2$.

For $p = 1$, and multiplying (2.18) by $t_{k+1}$, we get

$$t_{k+1}^{(2+\frac{p}{r})} \psi_k - t_k^{(2+\frac{p}{r})} \psi_k \leq Q_k \left(F(x_k) - F(x^*)\right) - P_k \frac{1}{2\lambda} \|x_k - x_{k-1}\|^2,$$ (3.1)

where $Q_k = t_{k+1}^{(2+\frac{p}{r})} - t_k^{(2+\frac{p}{r})}$ and $P_k = t_k^{(2+\frac{p}{r})} - t_{k+1}^{(2+\frac{p}{r})}$. Since $\lim_{k \to \infty} Q_k^{(2+\frac{p}{r})} = 2r + 1$ and $\lim_{k \to \infty} t_k^{(2+\frac{p}{r})} = 2$, we have $Q_k \leq 4t_k^2$ and $P_k \geq \frac{2}{t_k^{(2+\frac{p}{r})}}$ for sufficiently large $k$. Then, in view of (3.1), we obtain

$$t_{k+1}^{(2+\frac{p}{r})} \psi_k - t_k^{(2+\frac{p}{r})} \psi_k \leq 4t_k^2 \left(F(x_k) - F(x^*)\right) - \frac{1}{2\lambda} \|x_k - x_{k-1}\|^2.$$ (3.2)

Using Lemma 2.4 and the fact that $\sum_{k=1}^{\infty} t_k^{(2+\frac{p}{r})} \psi_k$ is bounded below, we obtain that $\lim_{k \to \infty} t_k^{(2+\frac{p}{r})} \psi_k$ exists and $\sum_{k=1}^{\infty} t_k^{(2+\frac{p}{r})} \|x_k - x_{k-1}\|^2$ is convergent. Since $\gamma_k < 1$, $t_{k+1} \sim t_k$ and Lemma 2.3, we have

$$\frac{\lambda}{\tau_1} t_k^{(2+\frac{p}{r})} \left(F(x_k) - F(x^*)\right) \leq t_k^{(2+\frac{p}{r})} \|x_k - y_k\|^2 \leq 2 \left(t_k^{(2+\frac{p}{r})} \|x_k - x_{k-1}\|^2 + t_{k-1}^{(2+\frac{p}{r})} \|x_{k-1} - x_k\|^2\right).$$ (3.3)
which means that \( \sum_{k=1}^{\infty} t_k^{(1+\frac{1}{r})} (F(x_k) - F(x^*)) \) is convergent. Then, \( \sum_{k=1}^{\infty} t_k^{1+\frac{1}{r}} \psi_k \) is convergent, which means that for \( p = 1 \), \( \sum_{k=1}^{\infty} t_k^{(2-p+\frac{1}{r})} \psi_k \) is convergent.

Suppose that for any \( p, \sum_{k=1}^{\infty} t_k^{(2-p+\frac{1}{r})} \psi_k \) is convergent. Now we prove that for \( p+1 \), the result still holds. Multiplying (2.18) by \( t_{k+1}^{(2-p+\frac{1}{r})} \), we get

\[
(2-p+\frac{1}{r}) \psi_{k+1} - (2-p+\frac{1}{r}) \psi_k \leq Q_k^{p+1} (F(x_k) - F(x^*)) - P_k^{p+1} \|x_k - x_{k-1}\|^2, \tag{3.4}
\]

where \( Q_k^{p+1} = t_{k+1}^{(2-p+\frac{1}{r})} - t_k^{(2-p+\frac{1}{r})} \) and \( P_k^{p+1} = t_k^{(2-p+\frac{1}{r})} - t_{k+1}^{(2-p+\frac{1}{r})} (t_k - 1)^2 \).

We see that \( \lim_{k \to \infty} \frac{Q_k^{p+1}}{t_k^{(2-p+\frac{1}{r})}} = 2r + 1 \) and \( \lim_{k \to \infty} \frac{P_k^{p+1}}{t_k^{(2-p+\frac{1}{r})}} = \begin{cases} \frac{2}{r}, & \text{if } r = 1, \\ 2, & \text{if } r < 1. \end{cases} \)

Then, similar with the proof for the case of \( p = 1 \), we can deduce that \( \lim_{k \to \infty} t_k^{(2-p+\frac{1}{r})} \psi_k \) exists and \( \sum_{k=1}^{\infty} t_k^{1+\frac{1}{r}} \|x_k - x_{k-1}\|^2 \) and \( \sum_{k=1}^{\infty} t_k^{1+\frac{1}{r}} (F(x_k) - F(x^*)) \) are convergent. Then \( \sum_{k=1}^{\infty} t_k^{(1+\frac{1}{r})} \psi_k \) is convergent, which means the result still holds for \( p+1 \). Therefore, \( \sum_{k=1}^{\infty} t_k^{(2-p+\frac{1}{r})} \psi_k \) is convergent for any positive integer \( p \).

Using the existence of \( \lim_{k \to \infty} t_k^{(2-p+\frac{1}{r})} \psi_k \), the convergence of \( \sum_{k=1}^{\infty} t_k^{(2-p+\frac{1}{r})} \psi_k \) and the divergence of \( \sum_{k=1}^{\infty} t_k^{-\frac{1}{r}} \), we have \( \lim_{k \to \infty} t_k^{(2-p+\frac{1}{r})} \psi_k = 0 \). Hence, from the fact that \( \lim_{k \to \infty} t_k = \frac{1}{a} \), result 1) holds.

We see that for any positive integer \( p \), \( \sum_{k=1}^{\infty} t_k^{(2-p+\frac{1}{r})} \psi_k \) is convergent. Similar with the proof of Corollary 3.1, we can conclude result 2).

\textbf{Remark 6.} For the \( t_k \) with \( r < 1 \), the Corollary 3.2 shows the sublinear convergence rate of Algorithm 1 is faster than the sublinear convergence rate of any order. Similar conclusion can be established for the function value. And we see that the \( t_k \) with \( r = 1 \) is the \( t_k \) proposed in FISTA,CD [6] but with a wider scope of \( a \), and the Corollary 3.2 shows better results than the existing conclusions in [1,3], in particular, \( \{x_k\} \) is sublinearly converges to \( \bar{x} \in X^* \) at the \( O\left(\frac{1}{\sqrt{n}}\right) \) rate of convergence.

\textbf{Corollary 3.3} Suppose that Assumption A1 holds. Let \( \{x_k\} \) be generated by Algorithm 1 with \( t_k = \frac{r^k - 1 + a}{a} \) (\( r > 1 \)) and \( x^* \in X^* \). Then, we have

1) \( F(x_k) - F(x^*) = o\left(\frac{1}{\sqrt{n}}\right) \) and \( \|x_k - x_{k-1}\| \leq O\left(\frac{1}{\sqrt{n}}\right) \).

2) \( \{x_k\} \) is converges sublinearly to \( \bar{x} \in X^* \) at the \( O\left(\frac{1}{\sqrt{n}}\right) \) rate of convergence.

\textbf{Proof.} It is easy to verify that \( \lim_{k \to \infty} t_k = +\infty \) and \( \lim_{k \to \infty} k\left(\frac{t_{k+1}}{t_k} - 1\right) = r \) for the case that \( r > 1 \), which means that Assumptions A2 and A3 hold. Combining with \( \lim_{k \to \infty} \frac{1}{t_k} = \frac{1}{a} \) and \( \sum_{k=1}^{\infty} t_k \) is convergent, we can deduce that the result 1) holds by Lemma 2.4 and \( \{x_k\} \) strongly converges to \( \bar{x} \in X^* \) by Theorem 2.2.

It follows from the result 1) that there exists a positive constant \( c' \) such that \( \|x_k - x_{k-1}\| \leq \frac{c'}{r} \). Then, similar with the proof of result 3) in Corollary 3.1, we can deduce that

\[
\|x_k - \bar{x}\| \leq \frac{c'}{r - 1} \frac{1}{k^{r-1}}.
\]

Hence, result 2) holds.

\textbf{Remark 7.} For the \( t_k \) in Option 2 with \( r > 1 \), we show that the convergence rate of function value and iterates can be improved to any order. The larger \( r \), the better convergence rate Algorithm 1 achieves.
Suppose that Assumption Corollary 3.5
Proof. We can easily verify that \( \lim_{k \to \infty} \) \( \frac{\ln \theta}{k} \) and \( \|x_k - x_{k-1}\| = O\left( \frac{\ln^{1.5} \theta}{k} \right) \).

2) \( \{x_k\} \) sublinearly converges to \( \bar{x} \in X^* \) at the \( O\left( \frac{\ln^{1.5} \theta}{k} \right) \) rate of convergence.

Proof. We can prove that \( \lim_{k \to \infty} t_k = +\infty \), \( \lim_{k \to \infty} k \left( \frac{\ln \theta}{k} - 1 \right) = 1 \), and \( \lim_{k \to \infty} t_{k+1} - t_k = 0 \), which means that Assumptions A2, A3 and A+ hold. Using the convergence of \( \sum_{k=1}^{\infty} \frac{1}{k} \), the result 1) is satisfied and \( \{x_k\} \) is converges strongly to \( \bar{x} \in X^* \).

It follows from the result 1) that for any \( x \in X^* \), there exists a positive integer \( N > e^{6\theta} \) such that

\[ \|x_k - x_{k-1}\| \leq \epsilon \frac{\ln^{1.5} \theta}{k}, \quad k > N. \]

Then, for \( k > e^{6\theta} \),

\[ \int_k^{+\infty} \frac{\ln^{1.5} \theta}{x^2} dx \leq \int_k^{+\infty} \left( \frac{\ln^{1.5} \theta}{x^2} + \frac{\ln^{1.5} \theta - 1}{x^2} (\ln x - 6\theta) \right) dx \]

\[ = 2 \int_k^{+\infty} \frac{\ln^{1.5} \theta - 1}{x^2} \ln x - 3\theta \right) dx \]

\[ = -4x^{-0.5} \ln^{1.5} \theta x_k^{+\infty} = 4k^{-0.5} \ln^{1.5} \theta, \]

we can deduce by (3.5) that

\[ \|x_k - \bar{x}\| \leq 4\epsilon \frac{\ln^{1.5} \theta}{k^{0.5}}, \quad k \to \infty, \]

which means that result 2) holds.

Remark 8. We notice that the \( t_k \) in Option 3 is similar with the one proposed in [2], which enjoys an improved convergence rate.

Option 4. \( t_k = e^{(k-1)^{\alpha}}, 0 < \alpha < 1 \).

Corollary 3.5 Suppose that Assumption A1 holds. Let \( \{x_k\} \) be generated by Algorithm 1 and \( x^* \in X^* \). Then, we have

1) \( F(x_k) - F(x^*) = o\left( \frac{1}{k^{(k-1)^{\alpha}}} \right) \) and \( \|x_k - x_{k-1}\| = O\left( \frac{1}{k^{(k-1)^{\alpha}}} \right) \).

2) \( \{x_k\} \) is converges sublinearly to \( \bar{x} \in X^* \) at the \( O\left( (k-1)^{\alpha} \frac{\ln^{1.5} \theta - 1}{k} e^{-(k-1)^{\alpha}} \right) \) rate of convergence.

Proof. We can easily verify that \( \lim_{k \to \infty} t_k = +\infty \), and \( \lim_{k \to \infty} k^{1-\alpha} \left( \frac{\ln \theta}{k} - 1 \right) = 0 \), which means that Assumptions A2 and A3 hold. Hence, the result 1) is satisfied.

It follows from the result 1) that there exists a positive constant \( c'' \) such that \( \|x_k - x_{k-1}\| \leq \frac{c''}{e^{(k-1)^{\alpha}}} \), we can deduce that

\[ \forall p > 1, \quad \|x_{k+p} - x_k\| \leq \sum_{i=k+1}^{k+p} \|x_i - x_{i-1}\| \leq \sum_{i=k+1}^{k+p} \frac{c''}{e^{(k-1)^{\alpha}}} \leq c'' \int_k^{k+p} \frac{1}{e^{(x-1)^{\alpha}}} dx. \]
Since the convergence of \( \int_1^{+\infty} e^{-(x-1)^\alpha} \, dx \), we see that \( \sum_{i=1}^{+\infty} \|x_i - x_{i-1}\| \) is convergent, which means that \( \{x_k\} \) is a Cauchy series and converges strongly to \( \bar{x} \in X^* \).

Then, as \( p \to \infty \), we have
\[
\|x_k - \bar{x}\| \leq \sum_{i=k+1}^{+\infty} \|x_i - x_{i-1}\| \leq c'' \sum_{i=k+1}^{+\infty} \frac{1}{e^{-\alpha r}} \leq c'' \int_k^{+\infty} e^{-(x-1)^\alpha} \, dx
\]
\[
= c'' \int_{(k-1)^\alpha}^{+\infty} \frac{1}{\alpha} y^{\frac{1}{\alpha}-1} e^{-y} \, dy \leq c'' \int_{(k-1)^\alpha}^{+\infty} y^{\frac{1}{\alpha}-1} e^{-y} \, dy.
\]
Denote \( \omega = \left[ \frac{1}{\alpha} - 1 \right] \) and \( A = (k-1)^\alpha \). We can deduce that
\[
\|x_k - \bar{x}\| \leq \frac{c''}{\alpha} \int_A^{+\infty} y^{\omega} e^{-y} \, dy = \frac{c''}{\alpha} A^\omega e^{-A} + \frac{c''}{\alpha} \sum_{j=0}^{\omega-1} \left( \prod_{i=0}^j (\omega - i) \right) (A^{\omega-j-1} e^{-A}) = O \left( A^\omega e^{-A} \right)
\]
which means that
\[
\|x_k - \bar{x}\| = O \left( (k-1)^{\alpha \left[ \frac{1}{\alpha} - 1 \right]} e^{-(k-1)^\alpha} \right).
\]
Hence, result 2) holds. \( \square \)

**Remark 9.** Notice that \( \forall p > 1, (k-1)^{\alpha \left[ \frac{1}{\alpha} - 1 \right]} e^{-(k-1)^\alpha} = o \left( \frac{1}{n^p} \right) \), which means that IFB with the \( t_k \) in Option 4 enjoys the similar sublinear convergence rate of IFB with Corollary 3.2, i.e., the sublinear convergence rate is faster than any order. Here, we give a further analysis for the convergence rate of the corresponding IFB from another aspect. We can derive that \( \gamma_k = 1 - \frac{c}{\alpha} + o \left( \frac{1}{n^p} \right) - \frac{1}{n^{1+\gamma}} \) from our Assumption 3 (denoted as Option 4). And for \( t_k \) in Option 2 with \( r < 1 \), we have corresponding \( \gamma_k = 1 - \frac{c}{\alpha} + o \left( \frac{1}{n^p} \right) \); For \( t_k \) in Option 4, we have \( \gamma_k = 1 - \frac{c}{\alpha} + o \left( \frac{1}{n^p} \right) \). Obviously, these two \( \gamma_k \) be of the similar magnitude, in particular, they should be of the same order if we choose \( r = 0.5, a = 0.5 \), and \( \alpha = 0.5 \), theoretically. Thus, it’s reasonable that the corresponding IFBs have similar numerical experiments.

4 Numerical Experiments

In this section, we conduct numerical experiments to study the numerical performance of IFB with different options of \( t_k \).

**LASSO** We first consider the LASSO
\[
\min_{x \in \mathbb{R}^n} F(x) = \frac{1}{2} \|Ax - b\|^2 + \delta \|x\|_1.
\]

We generate an \( A \in \mathbb{R}^{m \times n} \) be a Gaussian matrix and randomly generate a \( s \)-sparse vector \( \hat{x} \) and set \( b = A\hat{x} + 0.5\varepsilon \), where \( \varepsilon \) has standard i.i.d. Gaussian entries. And set \( \delta = 1 \). We observe that (4.1) is in the form of problem (P) with \( f(x) = \frac{1}{2} \|Ax - b\|^2 \) and \( g(x) = \lambda \|x\|_1 \). It is clear that \( f \) has a Lipschitz continuous gradient and \( L_f = \lambda \max \left( A^T A \right) \). Moreover, in view of (4.1) is satisfied the Assumption 1, the IFBs with \( t_k \) discussed in Option 2 with \( r < 1 \) and Option 4 of Section 3 should enjoy the rates of convergence better than any order of convergence rate. We terminate the algorithms once \( \|\partial F(x_k)\| < 10^{-8} \).

Considering Corollary 3.3 of Section 3, we know that in theory, the rate of convergence should improve constantly as \( r \) increasing. In the Fig.1, we test four choice of \( r \), which is \( r = 2, r = 4, r = 6 \) and \( r = 8 \), to show the same result in experiments as in theory. Denote that the IFB with \( t_k = \frac{k^{\gamma-1} + a}{\alpha T} \) is called as \( \text{"FISTA}_{\alpha \text{pow}(\tau)} \)’. Here we set \( a = 4 \). And the constant stepsize is \( \lambda = \frac{0.08}{nT} \).
Now, we perform numerical experiments to study the IFB with four choices of $t_k$. We consider the following five algorithms:
1) FISTA;
2) FISTA$_{\text{CD}}$ with $a = 4$;
3) FISTA$_{\text{pow}}(8)$, i.e., the IFB with $t_k = \frac{k^r - 1 + a}{a}$ ($r = 8$ and $a = 4$).
4) FISTA$_{\text{pow}}(0.5)$, i.e., the IFB with $t_k = \frac{k^r - 1 + a}{a}$ ($r = 0.5$ and $a = 0.5$).
5) FISTA$_{\text{exp}}$, i.e., the IFB with $t_k = e^{(k-1)\alpha}, 0 < \alpha < 1$. And set $\alpha = 0.5$.

We set the constant stepsize in the Algorithms 1) and 2) are $\frac{1}{L_f}$, and in the Algorithms 3), 4) and 5) are $\frac{0.98}{L_f}$.

Our computational results are presented in Fig. 2. We see that FISTA$_{\text{exp}}$ and FISTA$_{\text{pow}}(0.5)$ cost many fewer steps than FISTA$_{\text{CD}}$ and FISTA, and faster than FISTA$_{\text{pow}}(8)$. This results are same as the theoretical analyses in Section 3. And we see that the two lines of FISTA$_{\text{exp}}$ and FISTA$_{\text{pow}}(0.5)$ almost coincide, here, we give the detail number of iterations: for FISTA$_{\text{exp}}$, it’s number of iteration is 5000, and for FISTA$_{\text{pow}}(0.5)$, it’s number of iteration is 5107, which verify our theoretical analysis in Remark 9.

**Sparse Logistic Regression.** We also consider the sparse logistic regression with the $l_1$ regularized, that is

$$
\min_x \frac{1}{n} \sum_{i=1}^{n} \log (1 + \exp (-l_i \langle h_i, x \rangle)) + \delta \|x\|_1,
$$

(4.2)
where \( h_i \in \mathbb{R}^m, l_i \in \{-1, 1\}, i = 1, \ldots, n \). Define \( K_{ij} = -l_i h_{ij} \) and \( L_f = \frac{1}{m} \| K^T K \|. \) Set \( \delta = 1.e - 2 \). We take three datasets “w4a”, “a9a” and “sonar” from LIBSVM [5]. And the computational results relative to the number of iterations are reported in following Table 1.

Table 1: Comparison of the number of iterations

|       | FISTA | FISTA\_CD | FISTA\_pow(8) | FISTA\_pow(0.5) | FISTA\_exp |
|-------|-------|-----------|---------------|-----------------|------------|
| “w4a”| 1147  | 769       | 544           | 510             | 548        |
| “a9a”| 2049  | 1289      | 757           | 623             | 714        |
| “sonar”| 8405  | 3406      | 1586          | 922             | 980        |

We see from Table 1 that FISTA\_exp, FISTA\_pow(0.5) and FISTA\_pow(8) outperform FISTA and FISTA\_CD and the numerical results are consistent with the theoretical ones.

5 Conclusion

In this paper, we study the IFB with a class of abstract \( t_k \) satisfying our assumptions for solving the problem (P). Based on the local error bound condition, the sublinear convergence rate of iterates generated by the original FISTA is established. Hence, we claim that open problem about the convergence of original FISTA is solved under the local error bound condition. Further, We give other three types of \( t_k \) including the \( t_k \) in [6] to show their convergence rate of function value and strong convergence of iterates, moreover, to establish the sublinear convergence rate of iterates. Specially, we show that the sublinear convergence rates for both of function value and iterates generated by IFB with \( t_k \) in Option 2 with \( r < 1 \) and Option 4 can achieve \( o \left( \frac{1}{k^p} \right) \) for any positive integer \( p \).

Acknowledgements The work was supported by the National Natural Science Foundation of China (No.11901561), the Natural Science Foundation of Guangxi (No. 2018GXNSFBA281180) and the Postdoctoral Fund Project of China (Grant No. 2019M660833).

References

1. Attouch, H., Peyouquet, J.: The rate of convergence of Nesterov’s accelerated forwardbackward method is actually faster than \( \frac{1}{k^2} \). SIAM J. Optim. 26(3). 1824–1834 (2016)
2. Attouch, H., Cabot, A.: Convergence rates of inertial forward-backward algorithms. SIAM J. Optim. 28(1). 849–874 (2018)
3. Apidopoulos, V., Aujol, J., Dossal, C.: Convergence rate of inertial Forward-Backward algorithm beyond Nesterov’s rule. Math. Program. 180(1-2). 137–156 (2020)
4. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2(1). 183–202 (2009)
5. Chang, C. C., Lin, C. J.: LIBSVM: a library for support vector machines. ACM. Trans. Intell. Syst. Technol. 2(3). 1–27 (2011)
6. Chambolle, A., Dossal, C.: On the convergence of the iterates of the “fast iterative shrinkage-thresholding algorithm”. J. Optim. Theory Appl. 166(3). 968–982 (2015).
7. Calatroni, L., Chambolle, A.: Backtracking Strategies for Accelerated Descent Methods with Smooth Composite Objectives. SIAM J. Optim. 29(3), 1772–1798 (2019)
8. Donghwan, K., Jeffrey, A. F.: Another look at the fast iterative shrinkage/thresholding algorithm (FISTA). SIAM J. Optim. 28(1), 223–250 (2018)
9. Hai, T. N.: Error bounds and stability of the projection method for strongly pseudomonotone equilibrium problems. Int. J. Comput. Math. (2020) https://doi.org/10.1080/00207160.2019.1711374
10. Johnstone, P. R., Moulin, P.: Local and global convergence of a general inertial proximal splitting scheme for minimizing composite functions. Computational Optimization and Applications. 67, 259–292 (2017)
11. Moudafi, A., Oliny, M.: Convergence of a splitting inertial proximal method for monotone operators. J. Comput. Appl. Math. 155, 447–454 (2003)
12. Mridula, V., Shukla, K. K.: Convergence analysis of accelerated proximal extra-gradient method with applications. Neurocomputing. 388, 288–300 (2020)
13. Nesterov, Y.: A method for solving the convex programming problem with convergence rate $O\left(\frac{1}{k^2}\right)$. Dokl. Akad. Nauk SSSR. 269(3), 543–547 (1983)
14. O’Donoghue, B., Candès, E.: Adaptive restart for accelerated gradient schemes. Found Comput Math. 15(3), 715–732 (2015)
15. Su, W., Boyd, S., Candès, E. J.: A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights. J. Mach. Learn. Res. 17, 1–43 (2016)
16. Tao, S. Z., Boley, D., Zhang, S. Z.: Local Linear Convergence of ISTA and FISTA on the LASSO Problem. SIAM J. Optim. 26(1), 313–336 (2016)
17. Tseng, P., Yun, S.: A coordinate gradient descent method for nonsmooth separable minimization. Math. Program. 117, 387–423 (2009)
18. Villa, S., Salzo, S., Baldassarre, L., Verri, A.: Accelerated and Inexact Forward-Backward Algorithms. SIAM J. Optim. 23, 1607–1633 (2013)
19. Wen, B., Chen, X. J., Pong, T. K.: Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems. SIAM J. Optim. 27(1), 124–145 (2017)
20. Abdellah B.: A self-adaptive method for solving general mixed variational inequalities. J. Math. Anal. Appl. 309, 136–150 (2005)