A NOTE ON THE COMPLEXITY ANALYSIS OF FAST ITERATIVE SHRINKAGE-THRESHOLDING ALGORITHM

Qian Liu
College of Mathematics and Information, China West Normal University, Nanchong, Sichuan 637009, China
Email: qianliu1989@126.com

ABSTRACT
The fast iterative shrinkage-thresholding algorithm (FISTA) which was proposed by Beck and Teboulle in 2009, is a benchmark for minimization the sum of two convex functions such that one is differentiable with Lipschitz gradient. They proved the sequence generated by FISTA for which the objective is controlled, have a complexity rate \( O(1/k^2) \) which is the optimal complexity rate for first-order algorithm in the sense of Nemirovski and Yudin. However, their proof strategy is tricky. In this note, we give a new proof of the complexity analysis of FISTA. The same \( O(1/k^2) \) complexity rate of function value sequence is easily established.

1. INTRODUCTION
In this paper, we consider the following problem

\[
\min_{x \in \mathbb{R}^n} F(x) = f(x) + g(x),
\]

where \( g: \mathbb{R}^n \to \mathbb{R} \) is a convex function which is possibly nonsmooth and \( f: \mathbb{R}^n \to \mathbb{R} \) is a smooth convex function of the type \( C^{1,1} \), i.e., continuously differentiable with Lipschitz continuous gradient \( L \):

\[
\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \| \text{ for every } x, y \in \mathbb{R}^n.
\]

Specifically, when \( g(x) \equiv 0 \), then problem (1) reduces to the general unconstrained smooth convex minimization problem. Moreover, problem (1) captures a number of important applications arising in various areas. We here only mention a most commonly used application.

Example \( \ell_1 \)-norm regularization problem

\[
\min_{x \in \mathbb{R}^n} F(x) = \| Ax - b \|^2 + \lambda \| x \|_1,
\]

where \( A \in \mathbb{R}^{m \times n} \) is a matrix, \( b \in \mathbb{R}^m \) is a vector, \( \lambda > 0 \) is positive constant. \( \| x \|_1 \) represents the sum of the absolute values of the components of \( x \). Note that, the least-squares term reflects the fidelity of observed or fitting data, the \( \ell_1 \)-norm aims at inducing a sparse solution. Popular applications of (2) include the well-known Lasso model in statistics and the basis pursuit model in compressive sensing.

One of the most popular methods for solving (2) is the iterative shrinkage-thresholding algorithm (ISTA), which each iteration involves matrix-vector multiplication involving \( A \) and \( A^T \) followed by a shrinkage/soft-thresholding step, see [3, 4, 5]. The convergence analysis of ISTA has been well studied in the literature and the advantage of ISTA is in its simplicity. However, ISTA has been recognized as a slow method. In fact, it has been proved the sequence generated by ISTA shares an asymptotic rate of convergence that can be very slow and arbitrarily bad (for more details, see [2]). More recently, Beck and Teboulle in [1] construct a fast algorithm which built upon ideas of Nesterov and Güler, called FISTA, that keeps its simplicity but shares the improved rate \( O(1/k^2) \) of the optimal gradient method devised earlier in [6] for minimization smooth convex problems. Moreover, their complexity analysis are based on the more general problem (1) rather than (2). More precisely, the fast iterative Shrinkage-Thresholding algorithm (FISTA) for solving (1) reads as:

- Take \( y^1 = x^0 \), \( t_1 = 1 \).
- \((k \geq 1)\)
  \[
  \begin{align*}
  x^k &= \arg\min_{x} [g(x) + \frac{L}{2} \| x - (y^k - \frac{1}{L} \nabla f(y^k)) \|^2], \\
  t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\
  y^{k+1} &= x^k + \frac{t_{k+1}}{t_k} (x^k - x^{k-1}).
  \end{align*}
  \]

However, the complexity analysis proof in [1] is very tricky and elaborate. In this paper, we give a simple proof of the complexity analysis of FISTA. The rest of the paper is organized as follows. In section 2, we summarize some preliminary materials and useful results for further analysis. In section 3, we present the complexity analysis of FISTA. Finally, we give some concluding remarks.
2. PRELIMINARIES
Throughout this paper, we assume problem (1) is solvable, i.e., $X^*:= \text{argmin} F \neq \emptyset$, and for any $x^* \in X^*$, we set $F^* = F(x^*)$.\[4pt]

**Definition 2.1** [8] A function $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex with modulus $\beta > 0$ if for any $x, y \in \mathbb{R}^n$ and $\theta \in (0, 1)$, we have
\[
f((1-\theta)x + \theta y) \leq (1-\theta)f(x) + \theta f(y).
\]

The following lemma is useful in the derivation of the main results.

**Lemma 2.1** [7] Let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and gradient $\nabla f$ is Lipschitz continuous with the modulus $L > 0$, then for any $x, y \in \mathbb{R}^n$, we have
\[
f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \| y - x \|^2.
\]

To close this section, we give a close observation about the (3). Note that by the definition of $t_k$, it follows that
\[
t_{k+1}^2 - t_{k+1} = t_k^2, \forall k \geq 0, \text{with } t_0 = 0.
\]
Since $t_0 = 0$, then $x^{-1} = y^1 = x^0$. Moreover, by Lemma 4.3 in [1], we know $t_k \geq \frac{k+1}{2}$ for all $k \geq 1$.

3. COMPLEXITY ANALYSIS
**Lemma 3.1** Let \{x^k, y^k\} be the sequence generated by FISTA (3), then for any $x \in \mathbb{R}^n$ we have
\[
F(x) \geq F(x^k) + \frac{L}{2} \| x - x^k \|^2 - \frac{L}{2} \| x - y^k \|^2.
\]

**Proof.** By the optimality condition of (3), we have
\[
0 \in \partial g(x^k) + L(x^k - y^k) + \frac{1}{L} \nabla f(y^k) = 0,
\]
then $-\nabla f(y^k) + L(y^k - x^k) \in \partial g(x^k)$. By the convexity of $f$ and $g$, for any $x$, we have
\[
f(x) \geq f(y^k) + \langle \nabla f(y^k), x - y^k \rangle + \frac{L}{2} \| x - y^k \|^2.
\]

Adding (2) and (3), we know
\[
f(x) \geq f(x^k) + f(y^k) - f(x^k) + \langle \nabla f(y^k), y^k - x^k \rangle
+ L(y^k - x^k, x - x^k).
\]

By setting $x = x^k$ and $y = y^k$, it follows from Lemma 2.1 that
\[
f(x^k) \leq f(y^k) + \langle \nabla f(y^k), y^k - x^k \rangle + \frac{1}{2} \| y^k - x^k \|^2.
\]

Combining (??) and (5), then
\[
F(x) \geq F(x^k) + L(y^k - x^k, x - x^k) - \frac{L}{2} \| x^k - y^k \|^2.
\]

Observe that
\[
(\| y^k - x^k, x - x^k \| = \frac{1}{2} (\| y^k - x^k \|^2 + \| x - x^k \|^2 - \| x - y^k \|^2).
\]

Substituting (7) into (6), we obtain (1) immediately. This completes the proof.

Now, we give the new proof about the complexity analysis of FISTA, which is more simple than in [1].

**Theorem 3.1** Let \{x^k, y^k\} be the sequence generated by FISTA (3). Then for any $x \in \mathbb{R}^n$
\[
F(x^n) - F(x) \leq \frac{2L}{(n+1)^2} \| x - x^0 \|^2.
\]

**Proof** Set
\[
\hat{x}^k = \frac{1}{t_{k+1}} x + (1 - \frac{1}{t_{k+1}}) x^k.
\]

Since for all $k \geq 0$, we know $t_{k+1} \in (0, 1)$, by the convexity of $F$ we get
\[
F(\hat{x}^k) \leq \frac{1}{t_{k+1}} F(x) + (1 - \frac{1}{t_{k+1}}) F(x^k).
\]

Multiply $t_{k+1}^2$ on both sides of (10), we have
\[
t_{k+1}^2 F(\hat{x}^k) \leq t_{k+1}^2 F(x) + (t_{k+1}^2 - t_{k+1}) F(x^k).
\]
Since $t_{k+1}^2 - t_{k+1}^2 = t_k^2$, then (11) becomes
\[
\frac{t_{k+1}^2}{t_k^2} F(\hat{x}^k) \leq \frac{t_{k+1}^2}{t_k^2} F(x) - \frac{t_{k+1}^2}{t_k^2} F(x^k) + t_k^2 F(x^k).
\] (12)

Add $\frac{t_{k+1}^2}{t_k^2} F(x^{k+1})$ on both sides of (12) and rearrange, it follows that
\[
\frac{t_{k+1}^2}{t_k^2} F(x^{k+1}) - F(\hat{x}^k) - \frac{t_{k+1}^2}{t_k^2} F(x^k) - F(x) \leq \frac{t_{k+1}^2}{t_k^2} (F(x^{k+1}) - F(x^k)).
\] (13)

Next, we estimate $F(x^{k+1}) - F(\hat{x}^k)$. Actually, by (1) we have
\[
F(x) \geq F(x^{k+1}) + \frac{L}{2} \| x - x^{k+1} \|^2 - \frac{L}{2} \| x - y^{k+1} \|^2.
\] (14)

Specifically, set $x = \hat{x}^k$ in (14) we get
\[
F(x^{k+1}) - F(\hat{x}^k) \leq \frac{L}{2} \| \hat{x}^k - y^{k+1} \|^2 - \frac{L}{2} \| \hat{x}^k - x^{k+1} \|^2.
\] (15)

Note that,
\[
\hat{x}^k = \frac{1}{t_{k+1}} x + (1 - \frac{1}{t_{k+1}}) x^{k+1} = x^k + \frac{t_k}{t_{k+1}} (x^k - x^{k-1}).
\]

Set $u^k = x^{k-1} + t_k (x^k - x^{k-1})$, it follows from (3) and (9) that
\[
\hat{x}^k - y^{k+1} = \frac{1}{t_{k+1}} x + (1 - \frac{1}{t_{k+1}}) x^k - \frac{t_k}{t_{k+1}} (x^k - x^{k-1})
\]
\[
= \frac{1}{t_{k+1}} (x - x^{k-1} + t_k (x^{k-1} - x^k))
\]
\[
= \frac{1}{t_{k+1}} (x - u^k)
\] (16)

and
\[
\hat{x}^k - x^{k+1} = \frac{1}{t_{k+1}} x + (1 - \frac{1}{t_{k+1}}) x^k - x^{k+1}
\]
\[
= \frac{1}{t_{k+1}} (x - x^k + t_{k+1} (x^k - x^{k+1}))
\]
\[
= \frac{1}{t_{k+1}} (x - u^{k+1}).
\] (17)

Thus, substituting (15) (?) and (?) into (13), we get
\[
t_{k+1}^2 (F(x^{k+1}) - F(x)) - t_k^2 (F(x^k) - F(x)) \leq \frac{L}{2} \| x - u^k \|^2 - \frac{L}{2} \| x - u^{k+1} \|^2.
\]

Note that, $t_0 = 0$, summing the above inequality from $k = 0, 1, \ldots, n-1$, we get that
\[
t_n^2 (F(x^n) - F(x)) \leq \frac{L}{2} \| x - u^0 \|^2.
\]

Since $u^0 = x^{-1} = x^0$ and $t_n \geq \frac{n+1}{2}$, this implies (8) holds.

**Remark 3.1** Based on Theorem 2.1, we know that in order to obtain an $\varepsilon$-optimal solution $\bar{x}$, namely, $F(\bar{x}) - F^* \leq \varepsilon$, only need at most $\left[ \sqrt{\frac{2L}{\varepsilon}} \| x^0 - x^* \| \right]$ iterations.

4. **CONCLUSIONS**

In this paper, we give a simply proof of the complexity analysis of FISTA. A possible drawback of this basic schemes is that the Lipschitz constant $L$ is not always known or computable. A more practical strategy is taking advantage of backtracking stepsize rules. It is interesting to know if our analysis could be extended to this situation. We leave it as a further research topic.

5. **ACKNOWLEDGMENTS**

This research is supported by the Natural Science Foundation of China (Grant No. 11371015)

6. **REFERENCES**


