

A First Order Algorithm on an Optimization Problem with Improved Convergence when Problem is Convex

Chee-Khian Sim*

School of Mathematics and Physics
University of Portsmouth
Lion Gate Building, Lion Terrace
Portsmouth PO1 3HF

14 August 2025

Abstract

We propose a first order algorithm, a modified version of FISTA, to solve an optimization problem with an objective function that is the sum of a possibly nonconvex function, with Lipschitz continuous gradient, and a convex function which can be nonsmooth. The algorithm is shown to have an iteration complexity of $\mathcal{O}(\epsilon^{-2})$ to find an ϵ -approximate solution to the problem, and this complexity improves to $\mathcal{O}(\epsilon^{-2/3})$ when the objective function turns out to be convex. We further provide asymptotic convergence rate for the algorithm of worst case $\mathcal{O}(\epsilon^{-2})$ iterations to find an ϵ -approximate solution to the problem, with worst case $\mathcal{O}(\epsilon^{-2/3})$ iterations when its objective function is convex.

Keywords. Fast iterative shrinkage thresholding algorithm (FISTA); nonconvex optimization problem; convex optimization problem; iteration complexity; asymptotic convergence rate.

1 Introduction

First order methods are frequently used to solve optimization problems, especially when these problems are large scale [1]. Fast iterative shrinkage thresholding algorithm (FISTA), a first order method, is proposed in [2] to solve a convex minimization problem whose objective function is the sum of two convex functions, where one function can be nonsmooth; see also [19, 21, 22]. Algorithms, including first order algorithms, to solve minimization problems whose objective functions are nonconvex are also studied in the literature (see for example [4]). The objective function can be a sum of two functions, where one function is nonconvex, has Lipschitz continuous gradient, and the other function is convex, but nonsmooth. Solving this minimization problem has application in areas such as training a neural network, as discussed for example in [8]. Existing works that study first order algorithms on nonconvex optimization problems include [3, 6, 7, 11, 13, 14, 15, 16, 17, 27].

In the works mentioned in the previous paragraph, such as [6, 17], the iteration complexity to find an ϵ -approximate solution to the nonconvex optimization problem, whose objective function is the sum of two functions, where one function can be nonconvex, has Lipschitz

*Email address: chee-khian.sim@port.ac.uk

continuous gradient, and the other function is convex, but nonsmooth, is shown to be $\mathcal{O}(\epsilon^{-2})$. This is also the iteration complexity found for the first order algorithm, which is a modified version of FISTA, proposed in this paper on the optimization problem. Our algorithm also has the added feature that when the objective function to the problem turns out to be convex, it is able to “detect” this and its iteration complexity improves to $\mathcal{O}(\epsilon^{-2/3})$. In [6], the iteration complexity of the algorithm there also improves to $\mathcal{O}(\epsilon^{-2/3})$ when the objective function is convex, but the parameter inputs to the algorithm need to be adjusted accordingly. In [7], the authors improve on [6] by introducing a parameter free accelerated gradient method to solve nonconvex optimization problems, although the authors only show optimal complexity on function values when the objective function is convex. In [17], the accelerated composite gradient (ACG) variant introduced in the paper reduces to FISTA when the objective function is convex. The algorithm in this paper however is different from that in [17] in the way nonconvexity is handled, with different analysis as well. As with the algorithm in [17], our algorithm performs exactly one resolvent evaluation of h in each iteration, unlike papers found in the literature such as [6], where its ACG variant requires two resolvent evaluations of h in each iteration. We note that in [24], Nesterov’s accelerated gradient method is shown to have the iteration complexity of $\mathcal{O}(\epsilon^{-2/3})$ when the objective function consists only of a single convex function with Lipschitz continuous gradient. The approach for analysis in the paper is via high-resolution differential equations.

In this paper, we also provide asymptotic convergence rate¹ results for our algorithm, both when the objective function is nonconvex and convex. Our asymptotic convergence rate for the convex case is also found in [5] through high-resolution differential equations. In our case, we analyze our algorithm directly to obtain the result. In the recent published paper [9], the authors show that it takes $o(\epsilon^{-2})$ iterations to find an ϵ -approximate solution to a nonconvex optimization problem using steepest descent algorithm and other linesearch methods. In this paper, we add further to this by showing asymptotic convergence rate of $o(n^{-1/2})$ for our algorithm, a first order algorithm of the accelerated type², when we solve the nonconvex optimization problem we considered in the paper, under the assumption of convergence of iterates. To the best of our knowledge, this is the first time asymptotic convergence rate result is presented on first order algorithms of the accelerated type when solving nonconvex optimization problems.

In the next section, Section 2, we introduce the optimization problem we are solving in this paper, and also present the first order algorithm we are proposing to solve the problem. In Section 3, we state and prove results that are needed for the iteration complexities and asymptotic convergence rate results in its next section and associated subsection. We conclude the paper with Section 5.

1.1 Notations and facts

Given Ω a closed convex set of \mathbb{R}^n , we denote the standard projection map from \mathbb{R}^n onto Ω by $P_\Omega(\cdot)$. We have the following result on P_Ω :

$$\|P_\Omega(x) - y\| \leq \|x - y\|, \quad (1)$$

¹The study of asymptotic convergence is not new. For example, local superlinear convergence of interior point methods has been analyzed by researchers [10, 18, 20, 23, 25, 26].

²Examples of first order algorithms of the accelerated type on nonconvex optimization problems include accelerated composite gradient variants [17] and accelerated proximal gradient methods as found in [13].

where $x \in \mathbb{R}^n$ and $y \in \Omega$.

The following is a simple fact used in this paper:

$$\|a - b\|^2 - 2\langle a - b, a - c \rangle + \|a - c\|^2 = \|b - c\|^2, \quad (2)$$

where $a, b, c \in \mathbb{R}^n$.

Given functions $f : \mathcal{S} \rightarrow \mathbb{Z}_{++}$ and $g : \mathcal{S} \rightarrow \mathbb{R}_{++}$, where \mathcal{S} is an arbitrary set. For a subset $\hat{\mathcal{S}} \subseteq \mathcal{S}$, we write $f(w) = \mathcal{O}(g(w))$ for all $w \in \hat{\mathcal{S}}$ to mean that $f(w) \leq Mg(w)$ for all $w \in \hat{\mathcal{S}}$, where $M > 0$ is a positive constant. For $\hat{\mathcal{S}} = (0, w_0]$, where $w_0 > 0$, we write $f(w) = o(g(w))$ to mean that $f(w)/g(w) \rightarrow 0$, as $w \rightarrow 0$.

2 An Optimization Problem and Modified FISTA

We consider the following optimization problem:

$$\min_{y \in \mathbb{R}^n} \phi(y) := f(y) + h(y), \quad (3)$$

where f can be nonconvex on $\Omega \subset \mathbb{R}^n$, has Lipschitz continuous gradient with Lipschitz constant L , and Ω is a closed convex set in \mathbb{R}^n . Hence, we have

$$\|\nabla f(u_1) - \nabla f(u_2)\| \leq L\|u_1 - u_2\|, \quad \forall u_1, u_2 \in \Omega.$$

As a consequence, the following holds:

$$|f(u_1) - l_f(u_1; u_2)| \leq \frac{L}{2}\|u_1 - u_2\|^2, \quad \forall u_1, u_2 \in \Omega, \quad (4)$$

where

$$l_f(u_1; u_2) := f(u_2) + \langle \nabla f(u_2), u_1 - u_2 \rangle.$$

In particular,

$$f(u_1) - l_f(u_1; u_2) \leq \frac{L}{2}\|u_1 - u_2\|^2, \quad \forall u_1, u_2 \in \Omega. \quad (5)$$

Furthermore, h is a proper lower semi-continuous convex function, which can be non-smooth, with $\text{dom } h \subset \mathbb{R}^n$ closed and bounded. We assume that $\text{dom } h \subseteq \Omega$, and $\|y\| \leq C$ for $y \in \text{dom } h$, where C is some positive constant. An example of h is the indicator function of a closed and bounded convex set.

Observe that from (4), we have

$$-\frac{L}{2}\|u_1 - u_2\|^2 \leq f(u_1) - l_f(u_1; u_2), \quad \forall u_1, u_2 \in \Omega. \quad (6)$$

Therefore, there exists $m \geq 0$ such that

$$-\frac{m}{2}\|u_1 - u_2\|^2 \leq f(u_1) - l_f(u_1; u_2), \quad \forall u_1, u_2 \in \Omega. \quad (7)$$

An example of such m is L as seen from (6). Let $\underline{m} \geq 0$ be the smallest m such that (7) holds. It is clear that $\underline{m} \leq L$. Furthermore, $\underline{m} = 0$ if and only if f is convex on Ω .

Note that there exists an optimal solution to Problem (3), which we denote by y^* , since $\text{dom } h$ is closed and bounded. A necessary condition for $y \in \text{dom } h$ to be an optimal solution of Problem (3) is $0 \in \nabla f(y) + \partial h(y)$. Motivated by this condition, we have the notion of an ϵ -approximate solution³ of Problem (3), which is an (y, v) which satisfies

$$v \in \nabla f(y) + \partial h(y), \quad \|v\| \leq \epsilon, \quad (8)$$

where $\epsilon > 0$ is a given tolerance.

We now propose a first order algorithm, which we called Algorithm mFISTA, to find an ϵ -approximate solution to Problem (3), for a given tolerance $\epsilon > 0$. This algorithm is a modification of FISTA for convex optimization as found in [2] (see also [19]) to the setting when f in Problem (3) can be nonconvex.

Algorithm mFISTA

Initialization: Let tolerance $\epsilon > 0$, initial point $x_1 = y_0 \in \text{dom } h$ and $a_0 = 1$ be given. Set $k = 1$ with $L_k = 0$.

Step 1. Set

$$f_k(y) = f(y) + \frac{L_k}{2} \|y - y_{k-1}\|^2. \quad (9)$$

Step 2. Compute

$$y_k = \operatorname{argmin}_{y \in \mathbb{R}^n} \{l_{f_k}(y; x_k) + h(y) + 2L\|y - x_k\|^2\}. \quad (10)$$

Step 3. Find $a_k > 0$ such that

$$a_k(a_k - 1) = a_{k-1}^2. \quad (11)$$

Step 4. Compute

$$v_k = \nabla f(y_k) - \nabla f(x_k) + L_k[y_{k-1} - x_k] + 4L[x_k - y_k], \quad (12)$$

$$x_{k+1} = P_\Omega \left(y_k + \frac{a_{k-1} - 1}{a_k} (y_k - y_{k-1}) \right). \quad (13)$$

Step 5. Set

$$A_k = \begin{cases} \frac{2(l_f(y_k; x_{k+1}) - f(y_k))}{\|y_k - x_{k+1}\|^2} & , \text{ if } \|y_k - x_{k+1}\| \neq 0 \\ 0 & , \text{ otherwise} \end{cases}.$$

Compute $L_{k+1} = \max\{0, A_k\}$. Increase the value of k by 1, and return to Step 1.

Termination: If $\|v_k\| \leq \epsilon$, then output (y_k, v_k) and **exit**.

³This notion of closeness to optimality is employed in [11, 12, 17] and other papers that can be found in the literature.

Note that in Step 2 of the algorithm, y_k obtained is in $\text{dom } h$. It is easy to see that v_k obtained in Step 4 of the algorithm is in $\nabla f(y_k) + \partial h(y_k)$. Therefore, if $\|v_k\| \leq \epsilon$ when the algorithm terminates, then (y_k, v_k) is an ϵ -approximate solution to Problem (3). Furthermore, observe that if $\Omega = \mathbb{R}^n$, then there is no need to perform a projection in (13) to obtain x_{k+1} , which is not necessarily in $\text{dom } h$.

Remark 2.1 We observe that in Algorithm mFISTA, $0 \leq L_k \leq \underline{m} (\leq L)$ by (7). When f is convex, then $L_k = \underline{m} = 0$ for all k . It is worth noting that \underline{m} is not an input parameter to the algorithm, and the algorithm is able to “detect” when $\underline{m} = 0$ (that is, f is convex), resulting in an improved iteration complexity, as shown in Theorem 4.2, in this case.

We end this section with the following result on the sequence $\{a_k\}$:

Proposition 2.2 We have for $k \geq 0$,

$$a_k \geq \frac{k+1}{4} \quad (14)$$

and

$$a_k \leq 1 + k. \quad (15)$$

Hence, $\left\{ \frac{a_{k+1}}{a_k} \right\}$ is bounded. In fact, $\frac{a_{k+1}}{a_k} \rightarrow 1$ as $k \rightarrow \infty$.

Proof: Note that from (11), we have

$$a_i = \frac{1 + \sqrt{1 + 4a_{i-1}^2}}{2}. \quad (16)$$

Hence,

$$a_i = \frac{1 + \sqrt{1 + 4a_{i-1}^2}}{2} \geq \frac{1}{2} + a_{i-1}.$$

Summing the above from $i = 1$ to k , we have

$$a_k \geq \frac{k}{2} + a_0 \geq \frac{k}{2} \geq \frac{k+1}{4}.$$

On the other hand,

$$a_i = \frac{1 + \sqrt{1 + 4a_{i-1}^2}}{2} \leq \frac{1 + 1 + \sqrt{4a_{i-1}^2}}{2} = 1 + a_{i-1}. \quad (17)$$

Hence, summing the above from $i = 1$ to k , we have

$$a_k \leq a_0 + k = 1 + k.$$

Finally, boundedness of $\left\{ \frac{a_{k+1}}{a_k} \right\}$ follows from (14) and (15). It further follows from (16) and (14) that $\frac{a_{k+1}}{a_k} \rightarrow 1$ as $k \rightarrow \infty$. \square

3 Technical Preliminaries

Proposition 3.1 *We have for $k \geq 1$ and $y \in \Omega$,*

$$\hat{L}_k \|y - x_k\|^2 \leq f_k(y) - l_{f_k}(y; x_k) \leq L \|y - x_k\|^2,$$

where

$$\hat{L}_k = \begin{cases} 0 & , \text{ if } y = y_{k-1} \\ \frac{L_k - m}{2} & , \text{ otherwise} \end{cases}.$$

Proof: For $y \in \Omega$,

$$\begin{aligned} f_k(y) - l_{f_k}(y; x_k) &= f(y) + \frac{L_k}{2} \|y - y_{k-1}\|^2 - f(x_k) - \frac{L_k}{2} \|x_k - y_{k-1}\|^2 - \\ &\quad \langle \nabla f(x_k) + L_k(x_k - y_{k-1}), y - x_k \rangle \\ &= f(y) - f(x_k) - \langle \nabla f(x_k), y - x_k \rangle + \\ &\quad \frac{L_k}{2} [\|y - y_{k-1}\|^2 + 2\langle x_k - y_{k-1}, x_k - y \rangle - \|x_k - y_{k-1}\|^2] \\ &= f(y) - l_f(y; x_k) + \frac{L_k}{2} \|y - x_k\|^2, \end{aligned}$$

where the last equality follows from (2) and the definition of l_f . We have by (5) and Remark 2.1 that

$$f(y) - l_f(y; x_k) + \frac{L_k}{2} \|y - x_k\|^2 \leq L \|y - x_k\|^2.$$

On the other hand, when $y = y_{k-1}$, by definition of L_k ,

$$\begin{aligned} f(y) - l_f(y; x_k) + \frac{L_k}{2} \|y - x_k\|^2 &= f(y_{k-1}) - l_f(y_{k-1}; x_k) + \frac{L_k}{2} \|y_{k-1} - x_k\|^2 \\ &\geq f(y_{k-1}) - l_f(y_{k-1}; x_k) + \frac{A_{k-1}}{2} \|y_{k-1} - x_k\|^2 \\ &= 0. \end{aligned}$$

Otherwise, by (7),

$$\begin{aligned} f(y) - l_f(y; x_k) + \frac{L_k}{2} \|y - x_k\|^2 &\geq f(y) - l_f(y; x_k) + \frac{m}{2} \|y - x_k\|^2 + \frac{L_k - m}{2} \|y - x_k\|^2 \\ &\geq \frac{L_k - m}{2} \|y - x_k\|^2. \end{aligned}$$

□

The above proposition leads to the following consequence:

Proposition 3.2 *We have for $k \geq 1$ and $y \in \text{dom } h$,*

$$\begin{aligned} f_k(y_k) + h(y_k) - f_k(y) - h(y) + L \|y_k - x_k\|^2 &\leq 2L [\|y - x_k\|^2 - \|y - y_k\|^2] - \\ &\quad \hat{L}_k \|y - x_k\|^2. \end{aligned} \tag{18}$$

Proof: By Proposition 3.1, for $y \in \text{dom } h$, we have

$$f_k(y) + h(y) + (2L - \hat{L}_k)\|y - x_k\|^2 \geq l_{f_k}(y; x_k) + h(y) + 2L\|y - x_k\|^2.$$

Now, since y_k is the optimal solution to $l_{f_k}(y; x_k) + h(y) + 2L\|y - x_k\|^2$ over $y \in \mathbb{R}^n$, we have

$$l_{f_k}(y; x_k) + h(y) + 2L\|y - x_k\|^2 \geq l_{f_k}(y_k; x_k) + h(y_k) + 2L\|y_k - x_k\|^2 + 2L\|y - y_k\|^2.$$

Hence,

$$\begin{aligned} & f_k(y) + h(y) + (2L - \hat{L}_k)\|y - x_k\|^2 \\ & \geq l_{f_k}(y_k; x_k) + h(y_k) + 2L\|y_k - x_k\|^2 + 2L\|y - y_k\|^2 \\ & = l_{f_k}(y_k; x_k) + L\|y_k - x_k\|^2 + h(y_k) + L\|y_k - x_k\|^2 + 2L\|y - y_k\|^2 \\ & \geq f_k(y_k) + h(y_k) + L\|y_k - x_k\|^2 + 2L\|y - y_k\|^2, \end{aligned}$$

where the second inequality follows from Proposition 3.1, and we prove (18). \square

Proposition 3.3 *We have for $k \geq 2$,*

$$\begin{aligned} & a_{k-1}(a_{k-1} - 1)[\|y_{k-1} - x_k\|^2 - \|y_{k-1} - y_k\|^2] + a_{k-1}[\|x_k - y^*\|^2 - \|y_k - y^*\|^2] \\ & \leq \|a_{k-2}[y_{k-1} - y_{k-2}] + y_{k-2} - y^*\|^2 - \|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^*\|^2. \end{aligned} \quad (19)$$

Proof: Note that

$$\begin{aligned} & a_{k-1}(a_{k-1} - 1)[\|y_{k-1} - x_k\|^2 - \|y_{k-1} - y_k\|^2] + a_{k-1}[\|x_k - y^*\|^2 - \|y_k - y^*\|^2] \\ & = a_{k-1}(a_{k-1} - 1)[\|y_{k-1} - x_k\|^2 - \|(y_{k-1} - x_k) + (x_k - y_k)\|^2] + \\ & \quad a_{k-1}[\|x_k - y^*\|^2 - \|(y_k - x_k) + (x_k - y^*)\|^2] \\ & = a_{k-1}(a_{k-1} - 1)[- \|x_k - y_k\|^2 - 2\langle y_{k-1} - x_k, x_k - y_k \rangle] + \\ & \quad a_{k-1}[- \|x_k - y_k\|^2 - 2\langle y_k - x_k, x_k - y^* \rangle] \\ & = -a_{k-1}^2 \|x_k - y_k\|^2 - 2\langle a_{k-1}[x_k - y_k], y^* - x_k + (a_{k-1} - 1)(y_{k-1} - x_k) \rangle \\ & = -a_{k-1}^2 \|y_k - x_k\|^2 - 2\langle a_{k-1}[y_k - x_k], a_{k-1}x_k - ((a_{k-1} - 1)y_{k-1} + y^*) \rangle \\ & = \|a_{k-1}x_k - ((a_{k-1} - 1)y_{k-1} + y^*)\|^2 - \|a_{k-1}y_k - ((a_{k-1} - 1)y_{k-1} + y^*)\|^2, \end{aligned}$$

where the last equality follows from (2). Now,

$$\begin{aligned} & \left\| x_k - \left(\left(1 - \frac{1}{a_{k-1}} \right) y_{k-1} + \frac{1}{a_{k-1}} y^* \right) \right\| \\ & = \left\| P_\Omega \left(y_{k-1} + \frac{a_{k-2} - 1}{a_{k-1}} (y_{k-1} - y_{k-2}) \right) - \left(\left(1 - \frac{1}{a_{k-1}} \right) y_{k-1} + \frac{1}{a_{k-1}} y^* \right) \right\| \\ & \leq \left\| \left(y_{k-1} + \frac{a_{k-2} - 1}{a_{k-1}} (y_{k-1} - y_{k-2}) \right) - \left(\left(1 - \frac{1}{a_{k-1}} \right) y_{k-1} + \frac{1}{a_{k-1}} y^* \right) \right\| \\ & = \left\| \frac{a_{k-2}}{a_{k-1}} y_{k-1} - \frac{1}{a_{k-1}} [(a_{k-2} - 1)y_{k-2} + y^*] \right\|, \end{aligned}$$

where the inequality follows from (1). The result in the proposition then follows from the above derivations. \square

4 Iteration Convergence Rates of Algorithm mFISTA - Non-asymptotic and Asymptotic

We first state and prove the following key inequality, which prepares us for the iteration complexity results for Algorithm mFISTA in Theorem 4.2:

Lemma 4.1 *We have for $k \geq 2$,*

$$\begin{aligned} & a_{k-1}^2[\phi(y_k) - \phi(y^*)] - a_{k-2}^2[\phi(y_{k-1}) - \phi(y^*)] + \frac{L_k}{2}a_{k-1}^2\|y_k - y_{k-1}\|^2 + La_{k-1}^2\|y_k - x_k\|^2 \\ & \leq 2L[\|a_{k-2}[y_{k-1} - y_{k-2}] + y_{k-2} - y^*\|^2 - \|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^*\|^2] + \\ & \quad \frac{m - L_k}{2}a_{k-1}\|y^* - y_k\|^2 + \frac{L_k}{2}a_{k-1}\|y_{k-1} - y^*\|^2. \end{aligned} \quad (20)$$

Proof: Letting $y = y_{k-1}$ and $y = y^*$ in (18) lead to

$$\begin{aligned} & f_k(y_k) + h(y_k) - f_k(y_{k-1}) - h(y_{k-1}) + L\|y_k - x_k\|^2 \\ & \leq 2L[\|y_{k-1} - x_k\|^2 - \|y_{k-1} - y_k\|^2], \end{aligned} \quad (21)$$

$$\begin{aligned} & f_k(y_k) + h(y_k) - f_k(y^*) - h(y^*) + L\|y_k - x_k\|^2 \\ & \leq 2L[\|y^* - x_k\|^2 - \|y^* - y_k\|^2] + \frac{m - L_k}{2}\|y^* - x_k\|^2, \end{aligned} \quad (22)$$

respectively. Now, multiply (21) by $a_{k-1} - 1$ and add the resulting inequality to (22) gives rise to

$$\begin{aligned} & a_{k-1}[f_k(y_k) + h(y_k) - f_k(y^*) - h(y^*)] - (a_{k-1} - 1)[f_k(y_{k-1}) + h(y_{k-1}) - f_k(y^*) - h(y^*)] + \\ & La_{k-1}\|y_k - x_k\|^2 \\ & \leq 2L[(a_{k-1} - 1)(\|y_{k-1} - x_k\|^2 - \|y_{k-1} - y_k\|^2) + 2L[\|y^* - x_k\|^2 - \|y^* - y_k\|^2]] + \\ & \quad \frac{m - L_k}{2}\|y^* - x_k\|^2. \end{aligned} \quad (23)$$

Next, multiply the inequality (23) by a_{k-1} and noting (11) then leads to

$$\begin{aligned} & a_{k-1}^2[f_k(y_k) + h(y_k) - f_k(y^*) - h(y^*)] - a_{k-2}^2[f_k(y_{k-1}) + h(y_{k-1}) - f_k(y^*) - h(y^*)] + \\ & La_{k-1}^2\|y_k - x_k\|^2 \\ & \leq 2La_{k-1}[(a_{k-1} - 1)(\|y_{k-1} - x_k\|^2 - \|y_{k-1} - y_k\|^2) + [\|y^* - x_k\|^2 - \|y^* - y_k\|^2]] + \\ & \quad \frac{m - L_k}{2}a_{k-1}\|y^* - x_k\|^2. \end{aligned} \quad (24)$$

Applying (19) in Proposition 3.3 on the right hand side of the inequality in (24), we have

$$\begin{aligned} & a_{k-1}^2[f_k(y_k) + h(y_k) - f_k(y^*) - h(y^*)] - a_{k-2}^2[f_k(y_{k-1}) + h(y_{k-1}) - f_k(y^*) - h(y^*)] + \\ & La_{k-1}^2\|y_k - x_k\|^2 \\ & \leq 2L[\|a_{k-2}[y_{k-1} - y_{k-2}] + y_{k-2} - y^*\|^2 - \|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^*\|^2] + \\ & \quad \frac{m - L_k}{2}a_{k-1}\|y^* - y_k\|^2. \end{aligned} \quad (25)$$

Now,

$$a_{k-1}^2[f_k(y_k) + h(y_k) - f_k(y^*) - h(y^*)] - a_{k-2}^2[f_k(y_{k-1}) + h(y_{k-1}) - f_k(y^*) - h(y^*)]$$

$$\begin{aligned}
&= a_{k-1}^2[\phi(y_k) - \phi(y^*)] + \frac{L_k}{2}a_{k-1}^2[\|y_k - y_{k-1}\|^2 - \|y^* - y_{k-1}\|^2] - a_{k-2}^2[\phi(y_{k-1}) - \phi(y^*)] + \\
&\quad \frac{L_k}{2}a_{k-2}^2\|y^* - y_{k-1}\|^2 \\
&= a_{k-1}^2[\phi(y_k) - \phi(y^*)] - a_{k-2}^2[\phi(y_{k-1}) - \phi(y^*)] + \frac{L_k}{2}a_{k-1}^2\|y_k - y_{k-1}\|^2 - \\
&\quad \frac{L_k}{2}(a_{k-2}^2 - a_{k-1}^2)\|y^* - y_{k-1}\|^2 \\
&= a_{k-1}^2[\phi(y_k) - \phi(y^*)] - a_{k-2}^2[\phi(y_{k-1}) - \phi(y^*)] + \frac{L_k}{2}a_{k-1}^2\|y_k - y_{k-1}\|^2 - \frac{L_k}{2}a_{k-1}\|y_{k-1} - y^*\|^2,
\end{aligned}$$

where the last equality follows from (11). From the above and (25), upon algebraic manipulations, we have

$$\begin{aligned}
&a_{k-1}^2[\phi(y_k) - \phi(y^*)] - a_{k-2}^2[\phi(y_{k-1}) - \phi(y^*)] + \frac{L_k}{2}a_{k-1}^2\|y_k - y_{k-1}\|^2 + La_{k-1}^2\|y_k - x_k\|^2 \\
&\leq 2L[\|a_{k-2}[y_{k-1} - y_{k-2}] + y_{k-2} - y^*\|^2 - \|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^*\|^2] + \\
&\quad \frac{\underline{m} - L_k}{2}a_{k-1}\|y^* - y_k\|^2 + \frac{L_k}{2}a_{k-1}\|y_{k-1} - y^*\|^2,
\end{aligned}$$

and the lemma is proved. \square

Summing the inequality in Lemma 4.1 from $k = 2$ to n leads to

$$\begin{aligned}
&a_{n-1}^2[\phi(y_n) - \phi(y^*)] - a_0^2[\phi(y_1) - \phi(y^*)] + \frac{1}{2} \sum_{k=2}^n L_k a_{k-1}^2 \|y_k - y_{k-1}\|^2 + L \sum_{k=2}^n a_{k-1}^2 \|y_k - x_k\|^2 \\
&\leq 2L[\|a_0[y_1 - y_0] + y_0 - y^*\|^2 - \|a_{n-1}[y_n - y_{n-1}] + y_{n-1} - y^*\|^2] + \\
&\quad \frac{1}{2} \sum_{k=2}^n (\underline{m} - L_k) a_{k-1} \|y^* - y_k\|^2 + \frac{1}{2} \sum_{k=2}^n L_k a_{k-1} \|y_{k-1} - y^*\|^2. \tag{26}
\end{aligned}$$

We are now ready for the main theorem in this paper:

Theorem 4.2 *Given tolerance $\epsilon > 0$. Algorithm mFISTA outputs an ϵ -approximate solution to Problem (3) in at most $\mathcal{O}(\epsilon^{-2})$ iterations. In the case when f in Problem (3) is convex, the iteration complexity becomes $\mathcal{O}(\epsilon^{-2/3})$.*

Proof: From (12), we have

$$\begin{aligned}
\|v_k\| &= \|\nabla f(y_k) - \nabla f(x_k) + L_k[y_{k-1} - x_k] + 4L[x_k - y_k]\| \\
&\leq (5L + L_k)\|x_k - y_k\| + L_k\|y_k - y_{k-1}\| \\
&\leq 6L\|x_k - y_k\| + L_k\|y_k - y_{k-1}\|.
\end{aligned}$$

Hence,

$$\min_{1 \leq k \leq n} \|v_k\| \leq 2\sqrt{L} \left(\min_{2 \leq k \leq n} (36L\|x_k - y_k\|^2 + L_k\|y_k - y_{k-1}\|^2) \right)^{1/2}. \tag{27}$$

On the other hand, from (26), the following holds:

$$\frac{1}{36} \left(\min_{2 \leq k \leq n} (36L\|x_k - y_k\|^2 + L_k\|y_k - y_{k-1}\|^2) \right) \sum_{k=2}^n a_{k-1}^2$$

$$\begin{aligned}
&\leq 2L\|a_0[y_1 - y_0] + y_0 - y^*\|^2 + \frac{m}{2} \sum_{k=2}^n a_{k-1} \|y^* - y_k\|^2 + \frac{1}{2} \sum_{k=2}^n L_k a_{k-1} \|y_{k-1} - y^*\|^2 \\
&\leq 2L\|a_0[y_1 - y_0] + y_0 - y^*\|^2 + 2C^2 \left(\underline{m} + \max_{2 \leq k \leq n} L_k \right) \sum_{k=2}^n a_{k-1}, \tag{28}
\end{aligned}$$

where the last inequality follows from $\|y\| \leq C$ for $y \in \text{dom } h$. Note that by (14) and (15),

$$\frac{1}{16} \sum_{k=2}^n k^2 \leq \sum_{k=2}^n a_{k-1}^2, \quad \sum_{k=2}^n a_{k-1} \leq \sum_{k=2}^n k.$$

Using the above two inequalities, (28) implies that

$$\begin{aligned}
&\frac{1}{576} \left(\min_{2 \leq k \leq n} (36L\|x_k - y_k\|^2 + L_k\|y_k - y_{k-1}\|^2) \right) \sum_{k=2}^n k^2 \\
&\leq 2L\|a_0[y_1 - y_0] + y_0 - y^*\|^2 + 2C^2 \left(\underline{m} + \max_{2 \leq k \leq n} L_k \right) \sum_{k=2}^n k. \tag{29}
\end{aligned}$$

When f is convex, $\underline{m} = 0$ and $\max_{2 \leq k \leq n} L_k = 0$, hence (29) becomes

$$\frac{1}{576} \left(\min_{2 \leq k \leq n} (36L\|x_k - y_k\|^2 + L_k\|y_k - y_{k-1}\|^2) \right) \sum_{k=2}^n k^2 \leq 2L\|a_0[y_1 - y_0] + y_0 - y^*\|^2. \tag{30}$$

From (27) and (29), we deduce that

$$\min_{1 \leq k \leq n} \|v_k\| = \mathcal{O}(n^{-1/2}),$$

and when f is convex, from (27) and (30), we deduce that

$$\min_{1 \leq k \leq n} \|v_k\| = \mathcal{O}(n^{-3/2}).$$

The theorem is hence proved. \square

Remark 4.3 It is easy to see from (26), where we have $\underline{m} = L_k = 0$ when f is convex, that we can obtain the optimal complexity on function values of $\mathcal{O}(\epsilon^{-1/2})$ using Algorithm mFISTA to solve Problem (3). That is, it takes a worst case $\mathcal{O}(\epsilon^{-1/2})$ iterations using the algorithm to have $\phi(y_n) - \phi(y^*) \leq \epsilon$, when f is convex.

4.1 Further Discussions

In this subsection, we analyze the convergence behavior of Algorithm mFISTA further.

We first consider the case when f is convex. It turns out that we can get from the iteration complexity of Algorithm mFISTA of $\mathcal{O}(\epsilon^{-2/3})$ to an asymptotic $\mathcal{O}(\epsilon^{-2/3})$ iterations to find an ϵ -approximate solution to Problem (3) in this case.

When f is convex, it holds that $\underline{m} = L_k = 0$ for all k . Hence, (20) becomes

$$a_{k-1}^2 [\phi(y_k) - \phi(y^*)] - a_{k-2}^2 [\phi(y_{k-1}) - \phi(y^*)] + La_{k-1}^2 \|y_k - x_k\|^2$$

$$\leq 2L[\|a_{k-2}[y_{k-1} - y_{k-2}] + y_{k-2} - y^*\|^2 - \|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^*\|^2]. \quad (31)$$

We see from (31) that $\{a_{k-1}^2[\phi(y_k) - \phi(y^*)] + 2L\|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^*\|^2\}$ is a nonincreasing sequence, and furthermore each term in the sequence nonnegative. Hence, the sequence is convergent.

By summing (31) from $k = n + 1$ to $2n$, we have

$$\begin{aligned} & L \left(\min_{n+1 \leq k \leq 2n} \|y_k - x_k\|^2 \right) \sum_{k=n+1}^{2n} a_{k-1}^2 \leq L \sum_{k=n+1}^{2n} a_{k-1}^2 \|y_k - x_k\|^2 \\ & \leq (a_{n-1}^2[\phi(y_n) - \phi(y^*)] + 2L\|a_{n-1}[y_n - y_{n-1}] + y_{n-1} - y^*\|^2) - \\ & \quad (a_{2n-1}^2[\phi(y_{2n}) - \phi(y^*)] + 2L\|a_{2n-1}[y_{2n} - y_{2n-1}] + y_{2n-1} - y^*\|^2). \end{aligned} \quad (32)$$

We are then led to the following theorem:

Theorem 4.4 *When f in Problem (3) is convex, Algorithm mFISTA finds an ϵ -approximate solution to the problem in at most $o(\epsilon^{-2/3})$ iterations.*

Proof: When f is convex, v_k in (12) is given by

$$v_k = \nabla f(y_k) - \nabla f(x_k) + 4L[x_k - y_k].$$

Hence,

$$\|v_k\| \leq 5L\|y_k - x_k\|. \quad (33)$$

On the other hand, from (32) and (14), we have

$$\begin{aligned} & \frac{1}{16}L \left(\min_{n+1 \leq k \leq 2n} \|y_k - x_k\|^2 \right) \sum_{k=n+1}^{2n} k^2 \leq L \left(\min_{n+1 \leq k \leq 2n} \|y_k - x_k\|^2 \right) \sum_{k=n+1}^{2n} a_{k-1}^2 \\ & \leq (a_{n-1}^2[\phi(y_n) - \phi(y^*)] + 2L\|a_{n-1}[y_n - y_{n-1}] + y_{n-1} - y^*\|^2) - \\ & \quad (a_{2n-1}^2[\phi(y_{2n}) - \phi(y^*)] + 2L\|a_{2n-1}[y_{2n} - y_{2n-1}] + y_{2n-1} - y^*\|^2), \end{aligned}$$

where the right hand side in the above inequality tends to zero as n tends to infinity since $\{a_{k-1}^2[\phi(y_k) - \phi(y^*)] + 2L\|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^*\|^2\}$ is a convergent sequence. Hence, we have

$$\min_{n+1 \leq k \leq 2n} \|y_k - x_k\|^2 = o(n^{-3}).$$

The above equality together with (33) imply that

$$\min_{n+1 \leq k \leq 2n} \|v_k\| = o(n^{-3/2}).$$

The theorem is hence proved. \square

The above theorem takes existing non-asymptotic results in the literature, as appeared in [6, 17] (see also [24]), and in Theorem 4.2, where we have an iteration complexity to find an ϵ -approximate solution to Problem (3) of $\mathcal{O}(\epsilon^{-2/3})$, a step forward by stating an asymptotic convergence rate result - at most $o(\epsilon^{-2/3})$ iterations to find an ϵ -approximate solution to Problem (3).

In the general case when f can be nonconvex, we have the following result:

Theorem 4.5 Suppose $\{y_k\}$ is convergent. Then Algorithm mFISTA finds an ϵ -approximate solution to Problem (3) in at most $o(\epsilon^{-2})$ iterations.

Proof: Suppose $y_k \rightarrow y^{**}$ as $k \rightarrow \infty$. It is easy to check that Lemma 4.1, in particular (20), still holds with y^* replaced by y^{**} . Hence, we have for $k \geq 2$,

$$\begin{aligned} & a_{k-1}^2[\phi(y_k) - \phi(y^{**})] - a_{k-2}^2[\phi(y_{k-1}) - \phi(y^{**})] + \frac{L_k}{2}a_{k-1}^2\|y_k - y_{k-1}\|^2 + La_{k-1}^2\|y_k - x_k\|^2 \\ \leq & 2L[a_{k-2}[y_{k-1} - y_{k-2}] + y_{k-2} - y^{**}]^2 - \|a_{k-1}[y_k - y_{k-1}] + y_{k-1} - y^{**}\|^2 + \\ & \frac{\underline{m} - L_k}{2}a_{k-1}\|y^{**} - y_k\|^2 + \frac{L_k}{2}a_{k-1}\|y_{k-1} - y^{**}\|^2. \end{aligned}$$

Summing the above inequality from $k = n+1$ to $2n$ leads to

$$\begin{aligned} & \frac{1}{2} \sum_{k=n+1}^{2n} L_k a_{k-1}^2 \|y_k - y_{k-1}\|^2 + L \sum_{k=n+1}^{2n} a_{k-1}^2 \|y_k - x_k\|^2 \\ \leq & a_{n-1}^2[\phi(y_n) - \phi(y^{**})] - a_{2n-1}^2[\phi(y_{2n}) - \phi(y^{**})] \\ & 2L[a_{n-1}^2[y_n - y_{n-1}] + y_{n-1} - y^{**}]^2 - \|a_{2n-1}^2[y_{2n} - y_{2n-1}] + y_{2n-1} - y^{**}\|^2 + \\ & \frac{1}{2} \sum_{k=n+1}^{2n} (\underline{m} - L_k)a_{k-1}\|y^{**} - y_k\|^2 + \frac{1}{2} \sum_{k=n+1}^{2n} L_k a_{k-1}\|y_{k-1} - y^{**}\|^2. \end{aligned} \quad (34)$$

From (34), using (14) and (15), and following similar arguments which appear earlier in the paper, it is easy to check that we have

$$\begin{aligned} & \left(\min_{n+1 \leq k \leq 2n} (36L\|x_k - y_k\|^2 + L_k\|y_k - y_{k-1}\|^2) \right) \sum_{k=n+1}^{2n} k^2 \\ \leq & Kn^2[|\phi(y_n) - \phi(y^{**})| + |\phi(y_{2n}) - \phi(y^{**})| + \|y_n - y_{n-1}\|^2 + \|y_{2n} - y_{2n-1}\|^2 + \\ & \max_{n \leq k \leq 2n} \|y_k - y^{**}\|^2], \end{aligned}$$

where K is some positive constant independent of n . Note that

$$|\phi(y_n) - \phi(y^{**})| + |\phi(y_{2n}) - \phi(y^{**})| + \|y_n - y_{n-1}\|^2 + \|y_{2n} - y_{2n-1}\|^2 + \max_{n \leq k \leq 2n} \|y_k - y^{**}\|^2$$

tends to zero as n tends to infinity, since $y_k \rightarrow y^{**}$ as $k \rightarrow \infty$. On the other hand, we have

$$\min_{1 \leq k \leq 2n} \|v_k\| \leq 2\sqrt{L} \left(\min_{n+1 \leq k \leq 2n} (36L\|x_k - y_k\|^2 + L_k\|y_k - y_{k-1}\|^2) \right)^{1/2}$$

which can be derived in a similar way as (27). Putting everything together, we obtain

$$\min_{1 \leq k \leq 2n} \|v_k\| = o(n^{-1/2}),$$

and the theorem is proved. \square

The result in the above theorem contributes further to the literature complementing the paper [9] on first order algorithms to solve nonconvex optimization problems.

5 Conclusion

In this paper, we propose a first order algorithm, Algorithm mFISTA, to solve an optimization problem whose objective function is the sum of a possibly nonconvex function, with Lipschitz continuous gradient, and a convex function which can be nonsmooth. A feature of the algorithm is its ability to “detect” when the objective function is convex to obtain an improved iteration complexity. We analyze the algorithm to provide iteration complexity results for the algorithm in the nonconvex and convex case. We further provide asymptotic convergence rate results for the algorithm also in the nonconvex and convex case that take the iteration complexity results for these cases a step forward.

Data Availability Statement

No datasets were generated or analyzed during the current study.

References

- [1] A. Beck. *First-Order Methods in Optimization*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.
- [2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2:183–202, 2009.
- [3] Y. Carmon, J. C. Duchi, O. Hinder, and A. Sidford. Accelerated methods for nonconvex optimization. *SIAM Journal on Optimization*, 28:1751–1772, 2018.
- [4] C. Cartis, N. I. M. Gould, and P. L. Toint. *Evaluation Complexity of Algorithms for Nonconvex Optimization: Theory, Computation and Perspectives*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, 2022.
- [5] S. Chen, B. Shi, and Y.-X. Yuan. Gradient norm minimization of Nesterov acceleration: $o(1/k^3)$, 2022. arXiv:2209.08862v1.
- [6] S. Ghadimi and G. Lan. Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Mathematical Programming*, 156:59–99, 2016.
- [7] S. Ghadimi, G. Lan, and H. Zhang. Generalized uniformly optimal methods for nonlinear programming. *Journal of Scientific Computing*, 79:1854–1881, 2019.
- [8] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive Computation and Machine Learning Series. MIT Press, Cambridge, MA, 2017.
- [9] S. Gratton, C.-K. Sim, and P. L. Toint. Refining asymptotic complexity bounds for nonconvex optimization methods, including why steepest descent is $o(\epsilon^{-2})$ rather than $\mathcal{O}(\epsilon^{-2})$. *Computational Optimization and Applications*, July 2025. Published online.
- [10] M. Kojima, M. Shida, and S. Shindoh. Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs. *Mathematical Programming, Series A*, 80:129–160, 1998.
- [11] W. Kong, J. G. Melo, and R. D. C. Monteiro. Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs. *SIAM Journal on Optimization*, 29:2566–2593, 2019.

- [12] W. Kong, J. G. Melo, and R. D. C. Monteiro. Iteration complexity of a proximal augmented lagrangian method for solving nonconvex composite optimization problems with nonlinear convex constraints. *Mathematics of Operations Research*, 48:1066–1094, 2023.
- [13] H. Li and Z. Lin. Accelerated proximal gradient methods for nonconvex programming. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.
- [14] Q. Li, Y. Zhou, Y. Liang, and P. K. Varshney. Convergence analysis of proximal gradient with momentum for nonconvex optimization. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2111–2119, 2017.
- [15] J. Liang and R. D. C. Monteiro. A doubly accelerated inexact proximal point method for nonconvex composite optimization problems, 2018. arXiv:811.11378v2.
- [16] J. Liang and R. D. C. Monteiro. An average curvature accelerated composite gradient method for nonconvex smooth composite optimization problems. *SIAM Journal on Optimization*, 31:217–243, 2021.
- [17] J. Liang, R. D. C. Monteiro, and C.-K. Sim. A FISTA-type accelerated gradient algorithm for solving smooth nonconvex composite optimization problems. *Computational Optimization and Applications*, 79:649–679, 2021.
- [18] Z.-Q. Luo, J. F. Sturm, and S. Zhang. Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8:59–81, 1998.
- [19] Y. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Doklady AN SSSR*, 269:543–547, 1983.
- [20] Y. Nesterov and L. Tunçel. Local superlinear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems. *SIAM Journal on Optimization*, 26:139–170, 2016.
- [21] Yu. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming, Series A*, 103:127–152, 2005.
- [22] Yu. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming, Series B*, 140:125–161, 2013.
- [23] F. A. Potra and R. Sheng. A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8:1007–1028, 1998.
- [24] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming, Series A*, 195:79–148, 2022.
- [25] C.-K. Sim. Superlinear convergence of an infeasible predictor-corrector path-following interior point algorithm for a semidefinite linear complementarity problem using the Helmberg-Kojima-Monteiro direction. *SIAM Journal on Optimization*, 21:102–126, 2011.

- [26] S. J. Wright. *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.
- [27] Q. Yao, J. T. Kwok, F. Gao, W. Chen, and T.-Y. Liu. Efficient inexact proximal gradient algorithm for nonconvex problems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 3308–3314. IJCAI, 2017.