Abstract—Accelerated first order methods, also called fast gradient methods, are popular optimization methods in the field of convex optimization. However, they appear to be prone to suffer from oscillatory behavior that slows their convergence when medium to high accuracy is desired. In order to address this, restart schemes have been proposed in the literature, which seek to improve the practical convergence by suppressing the oscillatory behavior. This paper presents a restart scheme applicable to a broad class of accelerated first order methods. Under a quadratic functional growth condition, linear convergence rate is proved for a large class of non-strongly convex functions. Moreover, the worst-case convergence rate is comparable to the one obtained using a (generally non-implementable) optimal fixed-rate restart strategy. We show numerical results comparing the proposed algorithm with other restart schemes.

Index Terms—Convex Optimization, Accelerated First Order Methods, Restart Schemes, Linear Convergence.

I. INTRODUCTION

In the field of convex optimization, first order methods (FOM) are a widespread class of optimization algorithms which only require evaluations of the objective function and its gradient [1], [2]. Some examples of these methods include: gradient descent [1], ISTA [3] and ADMM [4]. A subclass of FOM are the accelerated first order methods (AFOM), which are characterised by providing a convergence rate $O(1/k^2)$ in terms of the value of the objective function [5]. Some noteworthy examples are: Nesterov’s fast gradient method [5], FISTA [3], MFISTA [6, §V.A] and accelerated ADMM [7], [8], [9], [10]. The use of AFOMs in the field of control is a heavily researched topic, especially in the field of model predictive control [11], [12], [13], [14], [15].

A drawback of AFOMs is that they often suffer from oscillating behaviour that slows them down [16]. In order to mitigate this, restart schemes have been proposed in the literature, which have shown to improve the convergence in a practical setting by suppressing the oscillatory behaviour [16], [17]. In a restart scheme, the AFOM is stopped when a certain criterion is met and then restarted using the last value provided by the algorithm as the new initial condition. However, most of the restart schemes proposed in the literature do not guarantee linear convergence for non-strongly convex optimization problems.

A notable exception is the restart scheme presented in [18], which we label the optimal fixed-rate restart scheme, as it exhibits global linear convergence for convex optimization problems with non-strongly convex objective functions that satisfy a quadratic functional growth condition. This scheme restarts the AFOM after a fixed number of iterations. However, its drawback is that it requires prior knowledge of either the optimal value of the objective function or of the parameter that characterizes the quadratic functional growth, both of which are not easily available in most practical cases.

In this paper we present a novel restart scheme for AFOMs which exhibits linear convergence for non-strongly convex objective functions that satisfy a quadratic functional growth condition. Furthermore, it does not require hard-to-attain information about the objective function, such as its optimal value or the quadratic functional growth parameter. We provide a theoretical upper bound on the number of iterations needed to achieve a desired accuracy and show that the obtained convergence rate is comparable to the one that could be obtained using the optimal fixed-rate restart strategy [18], which is optimal for the class of AFOMs and optimization problems under consideration. We also show numerical results comparing the proposed scheme with other restart strategies of the literature. This paper extends on the preliminary results presented for FISTA in the conference papers [19] and [20] by providing a restart algorithm with improved worst-case convergence rate that can be applied to a broad class of AFOM algorithms.

In Section II we formally present the class of optimization problems and AFOM algorithms under consideration. Section III describes the optimal fixed-rate restart strategy and provides its iteration complexity for our class of AFOMs. Section IV presents the novel implementable restart scheme with linear convergence. We show numerical results comparing the proposed scheme with other restart strategies in Section V. Finally, conclusions are drawn in Section VI.

Notation: Given a norm $\| \cdot \|$, we denote by $\| \cdot \|_*$ its dual norm: $\|x\|_* = \sup\{ x^T z : \|z\| \leq 1 \}$. The $l_1$-norm is denoted by $\| \cdot \|_1$. $\mathbb{Z}_+$ denotes the set of non-negative integers.

The authors are with the Department of Systems Engineering and Automation, University of Seville, 41092 Seville, Spain (e-mails: talamo@us.es, pkrupa@us.es, dlimon@us.es). Corresponding author: Teodoro Alamo.

This work was supported in part by the Agencia Estatal de Investigación (AEI) under Grant PID2019-106212RB-C41/AEI/10.13039/501100011033, by MINERCO-Spain and FEDER funds under Grant DPI2016-76493-C3-1-R and by the MCIU-Spain and FSE under Grant FPI-2017.
II. PROBLEM STATEMENT

In this paper we are concerned with finding the solution of optimization problems given by

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

(1)

which we assume are solvable and where \( f \in \Gamma_n \).

We use the following notation for the optimal set of (1), the projection operation onto it, and the level sets of \( f \).

Notation 1. Given the solvable problem (1):

(i) The optimal set is denoted by \( \Omega_f \). That is,

\[ \Omega_f = \{ x \in \mathbb{R}^n : f(x) = f^* \}. \]

(ii) For every \( x \in \mathbb{R}^n \) we denote \( \bar{x} \) the closest element to \( x \) in the optimal set \( \Omega_f \) with respect to norm \( \| \cdot \| \), i.e.

\[ \bar{x} = \arg \min_{x \in \Omega_f} \| x - z \|. \]

(iii) Given \( \rho \in [0, \infty) \) we denote the level set

\[ V_f(\rho) = \{ x \in \mathbb{R}^n : f(x) - f^* \leq \rho \}. \]

It is well known that if \( f \in \Gamma_n \) is strongly convex, then there exists \( \mu > 0 \) such that

\[ f(x) - f^* \geq \frac{\mu}{2} \| x - \bar{x} \|^2, \quad \forall x \in \text{dom}(f). \]

This inequality is called the quadratic functional growth condition and it is satisfied, at least locally, for a large class of not necessarily strictly convex functions \cite{18, 21}. For example, when \( f(x) = h(E\bar{x}) + c^\top x + 1_1(x) \), where \( h : \mathbb{R}^m \rightarrow \mathbb{R} \), is a smooth strictly convex function, \( E \in \mathbb{R}^{m \times n} \) and \( 1_1 \) is the indicator function of a polyhedral set \( \mathcal{X} \) \cite{18, 22}.

Let us consider a fixed point algorithm \( \mathcal{A} \) that can be applied to solve (1), i.e. given a initial point \( x_0 \in \text{dom}(f) \), algorithm \( \mathcal{A} \) generates a sequence \( \{ x_k \} \) with \( k \geq 0 \) such that \( \lim_{k \rightarrow \infty} f(x_k) = f^* \). We use the following notation to refer to the iterates provided by algorithm \( \mathcal{A} \).

Notation 2. Suppose that the fixed point algorithm \( \mathcal{A} \) is applied to solve problem (1) using as initial condition \( x_0 \). Given the integer \( k \geq 1 \), we denote with \( \mathcal{A}(x_0, k) \) the vector in \( \mathbb{R}^n \) corresponding to iteration \( k \) of the algorithm.

The following assumption characterizes the class of optimization problems and AFOM algorithms considered in this article.

Assumption 1. We assume that:

(i) For every \( \rho > 0 \), \( f \in \Gamma_n \) satisfies a quadratic growth condition of the form

\[ f(x_0) - f^* \geq \frac{\mu_\rho}{2} \| x_0 - \bar{x}_0 \|^2, \quad \forall x_0 \in V_f(\rho), \]

for some \( \mu_\rho > 0 \).

(ii) For every \( x_0 \in \text{dom}(f) \), algorithm \( \mathcal{A} \) satisfies,

\[ f(\mathcal{A}(x_0, 1)) \leq f(x_0) - \frac{1}{2L_f} \| g(x_0) \|^2, \]  

(2)

\[ f(\mathcal{A}(x_0, k)) - f^* \leq \frac{a_f}{(k + 1)^2} \| x_0 - \bar{x}_0 \|^2, \quad \forall k \geq 1, \]  

(3)

where \( a_f > 0 \), \( L_f > 0 \), and \( g(\cdot) \) is a gradient operator satisfying \( g(x) = 0 \Leftrightarrow x \in \Omega_f \).

(iii) We denote \( \eta_\rho = \max \left\{ \frac{1}{2} \left( \frac{2a_f}{\mu_\rho (k + 1)} \right) \right\} \).

The improvement with respect to the initial condition \( x_0 \) stated in (2) is satisfied for most AFOMs because the first iteration often results from the application of a proximal gradient operator \( T_{L_f} (\cdot) \) on \( x_0 \), thus resulting in the satisfaction of (2). In this case \( L_f > 0 \) is a Lipschitz constant and \( g(\cdot) \) is the gradient mapping \( g(x) = L_f (x - T_{L_f}(x)) \) (see \cite{18, 21} and \cite{22} Chapter 10). In any case, condition (2) can be easily enforced using as initial condition for \( \mathcal{A} \) the result of the application of a proximal gradient operator on \( x_0 \). The convergence rate given in (3) is satisfied by most AFOM algorithms \cite{11, 22}. Constant \( a_f \) is equal to \( 2L_f \) in FISTA and MFISTA and a multiple of \( L_f \) in other cases, e.g. when the Lipschitz constant \( L_f \) is not known and a backtracking strategy is implemented (see \cite{22} Chapter 10).

We now present a property on the iterates of \( \mathcal{A} \) which serves as the basis for the development and convergence analysis of the optimization schemes of the following sections. An equivalent result can be found in \cite{18} Subsection 5.2.2).

Property 1. Suppose that Assumption 1 holds. Then, for every \( x_0 \in V_f(\rho) \),

\[ f(\mathcal{A}(x_0, k)) - f^* \leq \left( \frac{\eta_\rho}{k + 1} \right)^2 (f(x_0) - f^*), \quad \forall k \geq 1. \]

(4)

Proof. Denote \( f_0 = f(x_0) \), \( f_k = f(\mathcal{A}(x_0, k)) \), \( \forall k \geq 1 \). Then,

\[ f_k - f^* \leq \frac{a_f}{(k + 1)^2} \| x_0 - \bar{x}_0 \|^2 \leq \frac{2a_f}{\mu_\rho (k + 1)^2} (f_0 - f^*) \leq \frac{\eta_\rho^2}{(k + 1)^2} (f_0 - f^*). \]

\[ \square \]

III. OPTIMAL FIXED-RATE RESTART SCHEME

This section describes the optimal fixed restart scheme presented in \cite{18} §5.2.2, in which \( \mathcal{A} \) is restarted each time the iteration counter attains an optimal fixed number of iterations. We analyze, under Assumption 1 its iteration complexity.

Given \( v_0 \in V_f(\rho) \), a fixed-rate restart scheme takes the recursion

\[ v_{j+1} = \mathcal{A}(v_j, n), \quad j \geq 0, \]

(5)

where \( n \geq 1 \) is a fixed integer.

Under assumption 1 the sequence \( \{ f(v_j) \}_{j \geq 0} \) is non increasing and converges monotonically to \( f^* \) if \( n \geq n_\rho \) (see Property 1). Given an accuracy parameter \( \epsilon > 0 \), the following property states the number \( M \) of restarts required to satisfy \( f(v_{M-1}) - f(v_M) \leq \epsilon \), and shows that the bound on the total number of iterations of \( \mathcal{A} \) is minimized if \( n \) is chosen equal to \( [en_\rho] \). See also \cite{18} §5.2.2 for a similar result.

Property 2 (Optimal fixed-rate restart scheme). Let Assumption 1 hold. Given \( v_0 \in V_f(\rho) \) and an integer \( n \) satisfying \( n > n_\rho \), consider the recursion (5). Then, given \( \epsilon > 0 \):

(i) The inequality \( f(v_{M-1}) - f(v_M) \leq \epsilon \) is satisfied for every \( M \geq \hat{M} \), where

\[ \hat{M} = 1 + \frac{1}{2(ln n - ln n_\rho)} \ln \left( \frac{f(v_0) - f^*}{\epsilon} \right). \]

(6)
implementable because the value of \( n \) is generally not available. However, the obtained bound is important because it provides the best theoretical convergence rate that could be obtained with a fixed-rate restart strategy. We also remark that this scheme can be implemented without requiring knowledge of the value of \( n_p \) if the value of \( f^* \) is known in advance. In this case, a similar convergence result can be obtained if a restart is implemented each time the iteration counter \( k \) satisfies \( \mathcal{A}(x_0, k) - f^* \leq \frac{f(x_0) - f^*}{\epsilon} \). See [25, §5.2.2] for further details.

\[ \mathcal{O} \left( n_p \ln \left( \frac{f(x_0) - f^*}{\epsilon} \right) \right) \] (8)

\[ N_p = \left[ \frac{\ln \left( \frac{f(x_0) - f^*}{\epsilon} \right)}{2} \right] + 1 \] (7)

In this case, we call recursion (5) the optimal fixed-rate restart scheme.

**Proof.** See Appendix A

One of the key properties of the optimal fixed-rate restart scheme is that it recovers the optimal linear convergence rate provided by Nesterov’s fast gradient method for strongly convex functions [18, 24, §2.2]. That is, recalling that \( n_p = \max \{1/2, \sqrt{2a_f/\mu_p} \} \) we easily obtain from Property 2(ii) that an \( n \) accurate solution is obtained in

\[ f(\mathcal{A}(x_0, i)) \]

**(ii)** If \( n = \lceil en_p \rceil \), the total number of iterations of \( \mathcal{A} \) required to attain \( f(v_{j-1}) - f(v_j) \leq \epsilon \) is upper bounded by

**IV. PROPOSED RESTART SCHEME**

In this section we propose a novel restart scheme (see Algorithm 1) that does not require knowledge of \( n_p \) and that attains a convergence rate similar to the one of the optimal fixed-rate restart strategy described in Section III. We start by presenting Algorithm 1 which implements a delayed exit condition of algorithm \( \mathcal{A} \). Algorithm 1 will then be used to derive the main result of this article: Algorithm 2.

![Algorithm 1: Delayed exit condition on \( \mathcal{A} \)](image)

\[ \text{Algorithm 1: Delayed exit condition on } \mathcal{A} \]

| Prototype: \([z, m] = \mathcal{A}_d(r, n)\) |
| Require: \( r \in \text{dom}(f), n \in \mathbb{R} \) |
| \( x_0 \leftarrow r, k \leftarrow 0 \) |
| Initialize \( \mathcal{A} \) with \( x_0 \) |
| repeat |
| \( k \leftarrow k + 1 \) |
| \( x_k \leftarrow \left\{ \begin{array}{ll} \mathcal{A}(x_0, k) & \text{if } f(\mathcal{A}(x_0, k)) \leq f(x_{k-1}) \\ x_{k-1} & \text{otherwise} \end{array} \right. \) |
| \( \ell \leftarrow \left\lfloor \frac{k}{2} \right\rfloor \) |
| until \( k \geq n \) and \( f(x_{\ell}) - f(x_k) \leq \frac{1}{3} (f(x_0) - f(x_{\ell})) \) |
| Output: \( z \leftarrow x_{\ell}, m \leftarrow k \) |

![Algorithm 2: Optimal Algorithm based on \( \mathcal{A}_d \)](image)

\[ \text{Algorithm 2: Optimal Algorithm based on } \mathcal{A}_d \]

| Prototype: \([z_{out}, j_{out}] = \mathcal{A}_d(z_0)\) |
| Require: \( z_0 \in \text{dom}(f), \epsilon > 0 \) |
| \( m_0 \leftarrow 1, m_{-1} \leftarrow 1, j \leftarrow -1 \) |
| repeat |
| \( j \leftarrow j + 1 \) |
| \( s_j \leftarrow \left\{ \begin{array}{ll} \frac{f(z_{j-1}) - f(z_j)}{\sqrt{f(z_{j-2}) - f(z_j)}} & \text{if } j \geq 2 \\ 0 & \text{otherwise} \end{array} \right. \) |
| \( n_j \leftarrow \max \{m_j, 4s_jm_{j-1} \} \) |
| \( [z_{j+1}, m_{j+1}] \leftarrow \mathcal{A}_d(z_j, n_j) \) |
| until \( f(z_j) - f(z_{j+1}) \leq \epsilon \) |
| Output: \( z_{out} \leftarrow z_{j+1}, j_{out} \leftarrow j \) |

![Fig. 1: Satisfaction of the delayed exit condition](image)

Given an initial condition \( x_0 \) and a scalar \( n \), which serves as a lower bound on the number of iterations, Algorithm 1 generates a sequence \( \{x_k\}_{k \geq 0} \) that satisfies (see step 5)

\[ f(x_k) = \min \{f(x_{k-1}), f(\mathcal{A}(x_0, k))\}, \forall k \geq 1. \]

Therefore,

\[ f(x_k) = \min_{i=0, \ldots, k} f(\mathcal{A}(x_0, i)). \] (9)

The algorithm exits after \( k \geq n \) iterations if the following inequality is satisfied (see step 7):

\[ f(x_k) - f(x_\ell) \leq \frac{1}{3} (f(x_0) - f(x_\ell)), \] (10)

where \( \ell = \left\lfloor \frac{k}{2} \right\rfloor \). The outputs of the algorithm are \( z \in \mathbb{R}^n \) and \( m \in \mathbb{Z} \), where \( z = x_m \) and \( m \geq n \) is the number of iterations required to satisfy the exit condition (10).

Intuitively, as illustrated in Figure 1 the exit condition (10) detects a degradation in the performance of the iterations of \( \mathcal{A} \). Notice that at iteration \( m \), the reduction corresponding to the last half of the iterations (from \( \left\lfloor \frac{m}{2} \right\rfloor \) to \( m \)) is no larger than one third of the reduction achieved in the first half of the iterations (from 0 to \( \left\lfloor \frac{m}{2} \right\rfloor \)).

The following property characterizes the number of iterations required to attain the exit condition (10) of Algorithm 1. This result is instrumental to prove the convergence results of Algorithm 2.
Property 3. Suppose that Assumption [7] holds. Then, the output \([z, m]\) from the call \([z, m] = A_\epsilon(r, n)\) of Algorithm [1] satisfies, for every \(r \in V_f(\rho)\):

(i) \(f(z) \leq f(r) - \frac{\epsilon}{2L_f} \|g(r)\|^2\).

(ii) \(f(z) - f^* \leq \left(\frac{n_{\rho}}{m + 1}\right)^2 (f(r) - f^*)\).

(iii) \(n \in (0, [4n_{\rho}]) \Rightarrow m \in [n, [4n_{\rho}]]\).

Proof. See Appendix [A]

We now introduce the main contribution of the article: Algorithm [2]. This algorithm makes successive calls to Algorithm [1] (see step 6) using a minimum number of iterations \(n_j\) that is determined by the past evolution of the iterates \(z_j\) (see steps 3 and 5). The main properties of the iterates of Algorithm [2] are given in the following property and theorem.

Property 4. Suppose that Assumption [7] holds and consider Algorithm [2] for a given initial condition \(z_0 \in V_f(\rho)\) and accuracy parameter \(\epsilon > 0\) then:

(i) Property [2] can be applied to the iterates of Algorithm [2] (i.e., taking \(r = z_j\), \(n = n_j\) \(s = z_{j+1}\) and \(m = m_{j+1}\)).

(ii) The sequence \(\{m_j\}\) produced is non-decreasing. In particular,

\[m_j \leq n_j \leq m_{j+1}, \quad \forall j \in Z_0^{j_{out}}.\]  \hspace{1cm} (11)

(iii) The sequence \(\{s_j\}\) satisfies \(s_j \in (0, 1), \quad \forall j \in Z_2^{j_{out}}\).

Proof. See Appendix [B]

Theorem 1. Suppose that Assumption [1] holds and consider Algorithm [2] for a given initial condition \(z_0 \in V_f(\rho)\) and accuracy parameter \(\epsilon > 0\) then:

(i) The number of calls to \(A_\epsilon\) (step 6) is bounded. That is, \(j_{out}\) is finite.

(ii) The number of iterations of \(A\) at each call of \(A_\epsilon\) (step 9) is upper bounded by \([4n_{\rho}]\). That is,

\[m_{j+1} \leq [4n_{\rho}], \quad \forall j \in Z_0^{j_{out}}.\]  \hspace{1cm} (12)

(iii) The total number of iterations of \(A\) performed by a call to Algorithm [2] which we denote by \(N_A\), is upper bounded by \(N_A \leq N_{\epsilon}\), where

\[N_A \leq \frac{e [4n_{\rho}]}{2} \left(5 + \frac{1}{\ln 15} \ln \left(1 + \frac{f(z_0) - f^*}{\epsilon}\right)\right).\]

Proof. See Appendix [B]

Remark 2. From Property [4(i)] we have that we can rearrange Property [3(ii)] to read as

\[\|g(z_j)\|^2 \leq 2L_f(f(z_j) - f(z_{j+1})).\]

Therefore, the exit condition \(f(z_j) - f(z_{j+1}) \leq \epsilon\) implies \(\|g(z_j)\|^2 \leq 2L_f \epsilon\). Since, as per Assumption [1(ii)] \(g(z_j)\) serves to characterize the optimality of \(z_j\), we conclude that the exit condition of Algorithm [2] also serves to characterize the optimality of \(z_{j+1}\). This means that the exit condition could be replaced by \(\|g(z_j)\|_2 \leq \epsilon\), where \(\epsilon > 0\). In this case, the upper bound on the number of iterations given in Theorem [1(iii)] would be the same but replacing \(\epsilon\) with \(\epsilon/(2L_f)\).

Algorithm 3: MFISTA

Prototype: \([z, m] = A_{MFISTA}(x, n, E_c)\)

Require : \(x \in \text{dom}(f)\), \(n \in Z \geq 1\), exit condition \(E_c\)

1 \(y_0 = x_0\), \(t_0 = 1, k = 0\)

2 repeat

3 \(k = k + 1\)

4 \(v_k = T_{L_f}(y_{k-1})\)

5 \(t_k = \frac{1}{2} \left(1 + \sqrt{1 + 4t_{k-1}^2}\right)\)

6 \(x_k = \{\)

7 \(y_k = x_k + \frac{t_{k-1}}{t_k} (v_k - x_k) + \frac{t_{k-1} - 1}{t_k} (x_k - x_{k-1})\)

8 Compute exit condition \(E_c\)

9 until \(k \geq n\) and \(E_c\) is true

Output: \(z = x_k, m = k\)

Note that Theorem [1(iii)] shows that the proposed algorithm attains the optimal linear convergence rate of the optimal fixed-rate restart scheme, in the sense that an \(\epsilon\) accurate solution is obtained in \(\tilde{N}_A\) iterations.

Comparing the upper bound provided in Theorem [1(iii)] with the upper bound \(N_A^F\) of the optimal fixed-rate restart scheme presented in Section III, we have

\[N_A = \frac{e [4n_{\rho}]}{2 [en_{\rho}] \ln 15} \frac{5 + \frac{1}{\ln 15} \ln \left(1 + \frac{f(z_0) - f^*}{\epsilon}\right)}{1 + \frac{1}{\ln 15} \ln \left(1 + \frac{f(z_0) - f^*}{\epsilon}\right)},\]

from where we obtain that

\[\lim_{\epsilon \to 0} \frac{N_A}{N_A^F} = \frac{e [4n_{\rho}]}{[en_{\rho}] \ln 15} \leq \frac{e(4n_{\rho} + 1)}{en_{\rho} \ln 15} = \frac{4}{\ln 15} + \frac{1}{n_{\rho} \ln 15} \leq 3 \left(1 + \frac{1}{4n_{\rho}}\right).\]

We conclude that the worst case complexity of (the implementable) Algorithm [2] is comparable to the (generally non-implementable) optimal fixed-rate restart scheme (approximately 50% more iterations of \(A\) when \(\epsilon\) tends to zero).

V. NUMERICAL RESULTS

We compare the proposed Algorithm [2] with other restart schemes of the literature by applying them to weighted Lasso problem

\[\min_x \frac{1}{2N}\|Ax - b\|_2^2 + ||Wx||_1,\]  \hspace{1cm} (13)

where \(x \in \mathbb{R}^n, A \in \mathbb{R}^{N \times n}\) is sparse with an average of 90% of its entries being zero (sparsity was generated by setting a 0.9 probability for each element of the matrix to be 0), \(n > N\), and \(b \in \mathbb{R}^N\). Each nonzero element in \(A\) and \(b\) is obtained from a Gaussian distribution with zero mean and variance 1. \(W \in \mathbb{R}^{N \times n}\) is a diagonal matrix with elements obtained from a uniform distribution on the interval \([0, \alpha]\).

We note that problems [13] can be reformulated in such a way that they satisfy the quadratic growth condition [18, §6.3]. The restart schemes used for comparison are:
TABLE I: Comparison between restart schemes.

| Exit Cond. | Alg. 2 | Functional | Gradient | Opt. | GLCR |
|------------|--------|------------|----------|------|------|
| Avg. Iter. | 1108.1 | 1158       | 1114.1   | 1704.2 | 1788.2 |
| Med. Iter. | 1094   | 1153       | 1107.5   | 1679.5 | 1756  |
| Max. Iter. | 1420   | 1655       | 1460     | 2154  | 2265  |
| Min. Iter. | 907    | 898        | 860      | 1403  | 1446  |

Fig. 2: Evolution of composite gradient mapping for a problem of Test 1.

(i) **Functional**: The restart scheme proposed in [16] that uses restart condition $f(x_{k+1}) \geq f(x_k)$.

(ii) **Gradient**: The restart scheme proposed in [16] that uses restart condition $\langle g(y_k), x_k - x_{k+1} \rangle \leq 0$, where $g : \mathbb{R}^n \to \mathbb{R}^n$ denotes the gradient mapping operator [3].

(iii) **Optimal**: The restart scheme proposed in [18, §5.2.2] which requires knowing $f^*$.

(iv) **GLCR**: The restart FISTA algorithm with linear convergence proposed in [20, Alg. 2].

We use the MFISTA algorithm [9, 2] as the algorithm $\mathcal{A}$ for these tests. Given an initial point $x \in \text{dom}(f)$, a minimum number of iterations $n \geq 0$ and an exit condition $E_n$, MFISTA algorithm is given by Algorithm [3] where $T_{ik}$ is the proximal gradient operator. This algorithm is a monotone AFOM, i.e., the sequence $\{f(x_k)\}_{k \geq 0}$ it produces is non-increasing, that satisfies Assumption [4]. Since it is monotone, it suffices to set the exit condition $E_n$ as [10] and then directly use Algorithm [3] as $\mathcal{A}_d$ in step 6 of Algorithm [2].

In order to provide a fair comparison between the different restart schemes, we run each one of them until the iterate $x_k$ satisfies $\|g(x_k)\|_* \leq 10^{-8}$, where $g(\cdot)$ is the gradient mapping operator, which, as stated in Remark 2, is a valid characterization of the optimality of $x_k$.

Table I shows the results of solving 100 randomly generated problems (13) that share the values of $N = 600$, $n = 800$, and $\alpha = 0.003$.

Figure 2 shows the evolution of $\|g(x_k)\|_*$ of each one of the restart schemes for one of the Lasso problems used to obtain the results of Table I. Additionally, it also shows the result of applying MFISTA without a restart scheme. As can be seen, the use of restart schemes can greatly improve the convergence of AFOMs, especially when small exit tolerances are desired.

VI. CONCLUSIONS

The main contribution of the paper is two-fold. We propose a delayed exit condition to detect degradation of the convergence of an accelerated algorithm $\mathcal{A}$. We show that, under a quadratic growth condition, this delayed exit condition is attained in a finite number of iterations. Based on this exit condition we propose a restart scheme for accelerated first order methods that retains their optimal linear convergence rate in the sense discussed above. Moreover, its worst case complexity is similar to the best one that can be obtained if the parameters characterizing the convergence of the base algorithm $\mathcal{A}$ were known. That is, we show that the upper bound of the number of iterations of $\mathcal{A}$ of the proposed algorithm is similar to the one obtained for the optimal fixed-rate restart scheme, but without requiring the knowledge of the aforementioned parameters. Finally, the numerical results indicate that the proposed algorithm is comparable, in practical terms, with other restart schemes of the literature.

APPENDIX

A. Proof of Properties 2 and 3

**Proof of Property 2** Suppose that the integer $M$ is such that the inequality $f(v_{M-1}) - f(v_M) > \epsilon$ is satisfied. From Property 1 we have

$$f(v_{j+1}) - f^* \leq \left(\frac{\rho_n}{n}\right)^2 (f(v_j) - f^*), \forall j \geq 0.$$  

Using this inequality in a recursive manner we obtain

$$\epsilon < f(v_{M-1}) - f(v_M) \leq f(v_{M-1}) - f^* \leq \left(\frac{\rho_n}{n}\right)^2 (f(v_0) - f^*).$$

This leads to

$$M - 1 < \frac{1}{2(\ln n - \ln n^\rho)} \ln \left(\frac{f(v_0) - f^*}{\epsilon}\right).$$

Thus, we conclude that if $M$ does not satisfy (14), then $f(v_{M-1}) - f(v_M) \leq \epsilon$. This proves the first claim.

Given $\epsilon > 0$, $v_0 \in V_f(\rho)$ and $n > n^\rho$, denote $S \geq 0$ the smallest number of restarts required to satisfy the condition $f(v_{S-1}) - f(v_S) \leq \epsilon$. We infer from the first claim of the property that $S \leq \max\{1, M\}$, which, making use of the expression of $M$ (6) allows us to write:

$$S \leq \left[1 + \frac{1}{2(\ln n - \ln n^\rho)} \ln \left(1 + \frac{f(v_0) - f^*}{\epsilon}\right)\right].$$

where a 1 has been added to the argument of the logarithm in the expression of $M$ to guarantee that the above bound is no smaller than $\max\{1, M\}$. Since each restart requires $n$ iterations of $\mathcal{A}$, we conclude that $N_F(n)$, the total number of iterations of $\mathcal{A}$, is equal to $nS$. Thus,

$$N_F(n) \leq n \left[1 + \frac{1}{2(\ln n - \ln n^\rho)} \ln \left(1 + \frac{f(v_0) - f^*}{\epsilon}\right)\right].$$ (15)

Simple calculus yields that the value that minimizes the coefficient

$$\frac{n}{\ln n - \ln n^\rho}$$

is $n^\rho = en^\rho$. Since $n$ has to be a positive integer, we choose the fixed restart rate given by $n = \lceil en^\rho \rceil$. Introducing this value in the bound (15) we finally obtain

$$N_F(\lceil en^\rho \rceil) \leq \lceil en^\rho \rceil \left[1 + \frac{1}{2} \ln \left(1 + \frac{f(v_0) - f^*}{\epsilon}\right)\right].$$
Proof of Property 3. From (9) and Assumption 1 we have
\[ f(z) = f(x_m) - \min_{i=0,...,m} f(A(x_0, i)) \leq f(A(x_0, 1)) - \frac{1}{2L_f} \| g(x_0) \|^2. \]
The first claim now follows directly from \( x_0 = r \). In view of (9) and Property 1 we have, for every \( k \in Z^n_m \),
\[ f(x_k) - f^* \leq f(A(x_0, k)) - f^* \leq \left( \frac{n_p}{k+1} \right)^2 (f(x_0) - f^*). \]
The second claim now follows from \( x_m = z \) and \( x_0 = r \).

Finally, we prove claim (iii). From the exit condition (step 7), we have
\[ f(z_{j-1}) - f(z_j) > \epsilon, \forall j \in Z_{1}^{\text{out}}. \]

Additionally, from (17a) we have \( f(z_{j-2}) \geq f(z_{j-1}) \), \( \forall j \in Z_2^{\text{out}} \). Thus,
\[ f(z_{j-2}) - f(z_j) \geq f(z_{j-1}) - f(z_j) \geq \epsilon > 0, \forall j \in Z_2^{\text{out}}. \]
Therefore, from step 4 taking \( j \geq 2 \), we have
\[ 0 < s_j = \sqrt{\frac{f(z_{j-1}) - f(z_j)}{f(z_{j-2}) - f(z_j)}} \leq 1, \forall j \in Z_2^{\text{out}}. \]

The proof of the following lemma relies upon some technical results on the iterates of Algorithm 2 namely Lemmas 4 and 5 which we include in Appendix C.

Lemma 1. Consider Algorithm 2 with the initial condition \( z_0 \in V_f(\rho) \), and \( \epsilon > 0 \). Suppose that Assumption 1 is satisfied and that \( j_{\text{out}} \geq D \), where
\[ D = \left[ 5 + \frac{1}{\ln 15} \ln \left( 1 + \frac{|f(z_0) - f^*|}{\epsilon} \right) \right]. \]
Then,
\[ m_{\ell+1} \leq \frac{1}{\sqrt{15}} m_{\ell+1+D}, \forall \ell \in Z_0^{\text{out}} - D. \]

Proof. The proof is obtained by reductio ad absurdum. If there is \( \ell \in Z_0^{\text{out}} - D \) such that \( m_{\ell+1} > \frac{1}{\sqrt{15}} m_{\ell+1+D} \), then we obtain from Lemma 3(iv) (see Appendix C) that
\[ D < 5 + \frac{1}{\ln 15} \ln \left( 1 + \frac{|f(z_0) - f^*|}{\epsilon} \right), \]
which contradicts the definition of \( D \).

Proof of Theorem 1. Let \( T \in Z \) be such that
\[ f(z_j) - f(z_{j+1}) > \epsilon, \forall j \in Z_0^T, \]
is satisfied. Then, defining \( d_j \doteq f(z_j) - f(z_{j+1}) \), we have
\[ f(z_0) - f(z_{T+1}) = \sum_{j=0}^{T} d_j \geq (T+1) \left( \min_{j=0,...,T} d_j \right) > (T+1) \epsilon. \]
Thus,
\[ T + 1 < \frac{f(z_0) - f(z_{T+1})}{\epsilon} \leq \frac{f(z_0) - f^*}{\epsilon} \leq \frac{\rho}{\epsilon}, \]
from where we infer that the largest integer \( T \) satisfying (19) is bounded. Consequently, the exit condition of Algorithm 2 (step 7) is satisfied within a finite number of iterations, thus proving claim (ii).

To prove claim (iii) we start by noting that both \( m_1 \) and \( m_2 \) are no larger than \( \lceil 4n_p \rceil \). Indeed, from step 4 we have that \( s_0 = s_1 = 0 \), which, in virtue of step 5, implies that \( n_0 = m_0 = 1 \) and \( n_1 = m_1 \). Since \( n_0 = 1 \) is no larger than \( \lceil 4n_p \rceil \) we have from (17c) that \( m_1 \) is also upper-bounded by \( \lceil 4n_p \rceil \). Moreover, since \( n_1 = m_1 \leq \lceil 4n_p \rceil \), we obtain by the
same reasoning that \( m_2 \leq [4n_\rho] \). We now prove that if \( j \geq 2 \) and \( m_j \leq [4n_\rho] \), then \( m_{j+1} \leq [4n_\rho] \). From step \( \ref{eq:17c} \) we have

\[
\begin{align*}
 s_j^2 & = \frac{f(z_j) - f(z_j)}{f(z_j) - f(z_j)} = 1 - \frac{f(z_j) - f(z_j)}{f(z_j) - f(z_j)} \\
 & \leq 1 - \frac{f(z_j) - f(z_j)}{f(z_j) - f(z_j)} \\
 & = \frac{f(z_j) - f(z_j)}{f(z_j) - f(z_j)} = \left( \frac{n_\rho}{m_{j-1}} + 1 \right)^2.
\end{align*}
\]

Thus, we have \( s_j m_{j-1} \leq n_\rho \). Therefore,

\[
n_j = \max\{m_j, 4s_j m_{j-1}\} \leq \max\{[4n_\rho], 4n_\rho\} = [4n_\rho],
\]

which, along with \( \ref{eq:17c} \), leads to \( m_{j+1} \leq [4n_\rho] \), thus proving the claim.

Finally, to prove claim \( \ref{eq:iii} \) we start by noting that the computation of each \( z_{j+1} \) is obtained from \( m_{j+1} \) iterations of \( A \). Thus,

\[
N_A = \sum_{j=0}^{j_{\text{out}}} m_{j+1} \leq (1 + j_{\text{out}}) [4n_\rho].
\]

Let us denote

\[
D = \left[ 5 + \frac{1}{\ln 15} \left( 1 + \frac{f(z_0) - f^\star}{\epsilon} \right) \right].
\]

Consider first the case \( j_{\text{out}} < D \). Since both \( j_{\text{out}} \) and \( D \) are integers we infer from this inequality that \( 1 + j_{\text{out}} < D \). This, along with \( \ref{eq:20} \), implies that \( N_A \leq [4n_\rho] \).

Suppose now that \( j_{\text{out}} \geq D \). We first recall that Property \( \ref{eq:iii} \) states that the sequence \( \{m_{j+1}\}_{j \geq 0} \) is non-decreasing. We now rewrite \( j_{\text{out}} \) as \( j_{\text{out}} = d + T \), where \( d \in \mathbb{N}_{0, D-1} \) and \( t \) is a non-negative integer. Thus,

\[
N_A = \sum_{j=0}^{j_{\text{out}}} m_{j+1} = \sum_{j=0}^{d} m_{j+1} + \sum_{i=1}^{t \epsilon_{i+1}} m_{d+i+1} + 1 \\
\leq D m_{d+1} + D \left( \sum_{i=1}^{t \epsilon_{i+1}} m_{d+i+1} + 1 \right) = D \left( \sum_{i=1}^{t \epsilon_{i+1}} m_{d+i+1} + 1 \right).
\]

From Lemma \( \ref{eq:1} \) we have

\[
m_{d+i+1} \leq \frac{\sqrt{15}}{15} \epsilon_{i+1}^{-1}, \quad \forall i \in \mathbb{Z}_{t \epsilon_{i+1} + 1}.
\]

Thus,

\[
N_A \leq D \left( \sum_{i=1}^{t \epsilon_{i+1}} \frac{\sqrt{15}}{15} \right)^{t-i}.
\]

Using now \( m_{1+d+i+T} \leq m = [4n_\rho] \) (see \( \ref{eq:12} \)) we obtain

\[
N_A \leq D \left( \sum_{i=0}^{t} \left( \frac{\sqrt{15}}{15} \right)^{t-i} \right) = D \left( \frac{\sqrt{15}}{15} \right)^{t} \\
\leq \sum_{i=0}^{t} \left( \frac{\sqrt{15}}{15} \right)^{t-i} \leq \frac{\epsilon}{2}.
\]

Thus, \( N_A \leq \frac{\epsilon}{2} m_D \leq \frac{\epsilon}{2} [4n_\rho] D \).

C. Technical results on the iterates of Algorithm \( \ref{alg:2} \)

Lemma 2. The function \( \varphi(s) : \mathbb{R} \to \mathbb{R} \), defined as

\[
\varphi(s) = \left( \frac{1}{s^2} - 1 \right) \max \{1, (4s)^4\},
\]

satisfies \( \varphi(s) \geq 15 \), \( \forall s \in (0, \sqrt{15}/4) \).

Proof. We have that

\[
\varphi(s) = \begin{cases} 
4^4(s^2 - s^4) & \text{if } s > \frac{1}{4}, \\
\frac{1}{s^2} - 1 & \text{if } s \leq \frac{1}{4}.
\end{cases}
\]

It is clear that \( \varphi(\cdot) \) is monotonically decreasing in \( (0, \frac{1}{4}) \). Thus,

\[
\min_{s \in (0, \frac{\sqrt{15}}{4})} \varphi(s) = \min_{s \in [\frac{1}{4}, \frac{\sqrt{15}}{4})} \varphi(s) = \min_{s \in [\frac{1}{4}, \frac{\sqrt{15}}{4})} 4^4(s^2 - s^4).
\]

We notice that the derivative of \( s^2 - s^4 \) is \( 2s(1 - 2s^2) \), which vanishes only once in the interval of interest (at \( s = \frac{1}{\sqrt{2}} \)). From here we infer that \( s^2 - s^4 \) is increasing in \( (\frac{1}{\sqrt{2}}, \frac{\sqrt{15}}{4}) \) and decreasing in \( (\frac{1}{\sqrt{2}}, \frac{\sqrt{15}}{4}) \). Thus, the minimum is attained at the extremes of the interval \( (\frac{1}{4}, \frac{\sqrt{15}}{4}) \). That is, we conclude that

\[
\min_{s \in (0, \frac{\sqrt{15}}{4})} \varphi(s) = \min_{s \in (\frac{1}{4}, \frac{\sqrt{15}}{4})} \varphi(s) = \min_{s \in (0, \frac{\sqrt{15}}{4})} \varphi(s) = 15, 15 = 15.
\]

Lemma 3 (Technical results on the iterates of Alg. \( \ref{alg:2} \))

Consider Algorithm \( \ref{alg:2} \) with the initial condition \( z_0 \in V_\ell(\rho) \), and \( \epsilon > 0 \). Suppose that Assumption \( \ref{eq:1} \) is satisfied and that \( j_{\text{out}} \geq 2 \). Suppose also that there is \( T \in \mathbb{Z}_{j_{\text{out}}+1}^+ \) such that

\[
m_{\ell+1} > \frac{1}{\sqrt{15}} m_{\ell+1+1}.
\]

Then:

(i) \( s_j \in \big(0, \frac{\sqrt{15}}{4}\big) \), \( \forall j \in \mathbb{Z}_{\ell+1}^+ \).

(ii) \( \sum_{j=\ell+2}^{\ell+T} \ln \left( \max \{1, (4s_j)^4\} \right) < 4 \ln 15. \)

(iii) \( \sum_{j=\ell+2}^{\ell+T} \ln \left( \frac{1}{s_j^2} - 1 \right) \leq \ln \left( 1 + \frac{f(z_0) - f^\star}{\epsilon} \right). \)

(iv) \( T < 5 + \frac{1}{\ln 15} \ln \left( 1 + \frac{f(z_0) - f^\star}{\epsilon} \right). \)

Proof. Denote \( f_j = f(z_j) \), \( j \in \mathbb{Z}_{j_{\text{out}}+1}^+ \). From \( j \geq 2 \) and step \( \ref{eq:iii} \) of Algorithm \( \ref{alg:2} \) we have

\[
s_j^2 = \frac{f_j - f_j}{f_j - f_j} = 1 - \frac{f_j - f_j}{f_j - f_j} = \left( \frac{n_\rho}{m_{j-1}+1} \right)^2.
\]

The inequality \( s_j > 0 \), \( \forall j \in \mathbb{Z}_{\ell+1}^+ \) follows from Property \( \ref{eq:iii} \). In order to prove the first claim it remains to prove the inequality \( s_j \leq \frac{\sqrt{15}}{4} \), \( \forall j \in \mathbb{Z}_{\ell+1}^+ \). We proceed by reductio ad absurdum. Suppose that there is \( j \in \mathbb{Z}_{\ell+1}^+ \) such that \( s_j > \frac{\sqrt{15}}{4} \). In this case,

\[
m_{j+1} \geq n_j = \max\{m_j, 4s_j m_{j-1}\} \geq 4s_j m_{j-1} > \sqrt{15} m_{j-1}.
\]
Thus, the left term of (22) can be lower bounded by means of
(Property 4(ii)) we obtain

\[ m_{\ell + 1 + T} \geq m_{\ell + 1} > \sqrt{15} m_{\ell + 1 - 1} \geq \sqrt{15} m_{\ell + 1}. \]

This contradicts the assumptions of the property, thus proving the first claim.

From the non-decreasing nature of the sequence \( \{ m_j \} \)
(Property 4(ii)) we have, for every \( j \in \mathbb{Z}_{\ell + T}^+ \),
\[ m_{j + 1} \geq n_j = \max \{ m_j, 4 s_j m_{j - 1} \} \geq m_{j + 1} \cdot \max \{ 1, 4 s_j \}. \]

Equivalently,
\[ \ln \left( \max \{ 1, 4 s_j \} \right) \leq \ln \frac{m_{j + 1}}{m_{j - 1}}, \forall j \in Z_{\ell + T}. \]

This implies
\[ \sum_{j=\ell + 2}^{\ell + T} \ln \left( \max \{ 1, 4 s_j \} \right) \leq \sum_{j=\ell + 2}^{\ell + T} \ln \frac{m_{j + 1}}{m_{j - 1}} = \ln \frac{m_{\ell + 1} T m_{\ell + 1 + T}}{m_{\ell + 1} m_{\ell + 2}} \leq \ln \frac{m_{\ell + 1 + T}}{m_{\ell + 1}} = 2 \ln \frac{m_{\ell + 1 + T}}{m_{\ell + 1}} < 2 \ln \sqrt{15} = \ln 15. \]

The second claim is obtained multiplying the last inequality by 4. To prove the third claim we notice that
\[ \prod_{j=\ell + 2}^{\ell + T} \left( \frac{1}{s_j^4} - 1 \right) = \prod_{j=\ell + 2}^{\ell + T} \frac{f_{j + 2} - f_{j - 1}}{f_{j - 1} - f_j} = \frac{f_\ell - f_{\ell + 1}}{f_{\ell + 1 + T} - f_{\ell + 1}}. \]

Since \( \ell + T \leq j_{\text{out}} \) we have \( f_{\ell + 1 + T} - f_\ell > \epsilon > 0 \). Using this inequality we obtain
\[ \prod_{j=\ell + 2}^{\ell + T} \left( \frac{1}{s_j^4} - 1 \right) < \frac{f_\ell - f_{\ell + 1}}{\epsilon} \leq \frac{f_0 - f_{\ell + 1}}{\epsilon} \leq \frac{f_0 - f^*}{\epsilon}, \]
from where the third claim directly follows. In order to prove the last claim of the property we sum the inequalities given by the second and third claims to obtain
\[ \sum_{j=\ell + 2}^{\ell + T} \ln \left( \frac{1}{s_j^4} - 1 \right) \cdot \max \{ 1, \{ 4 s_j \}^4 \} \leq \ln \left( 1 + \frac{f_0 - f^*}{\epsilon} \right) + 4 \ln 15. \]

From the first claim we have \( s_j \in \left( 0, \frac{\sqrt{15}}{4} \right), \forall j \in \mathbb{Z}_{\ell + T}^+ \). Thus, the left term of (22) can be lower bounded by means of the following inequality (Lemma 2)
\[ 15 \leq \left( \frac{1}{s_j^4} - 1 \right) \cdot \max \{ 1, \{ 4 s_j \}^4 \}, \forall s \in \left( 0, \frac{\sqrt{15}}{4} \right). \]

That is,
\[ \sum_{j=\ell + 2}^{\ell + T} \ln 15 \leq \ln \left( 1 + \frac{f_0 - f^*}{\epsilon} \right) + 4 \ln 15. \]

Equivalently,
\[ (T - 1) \ln 15 \leq \ln \left( 1 + \frac{f_0 - f^*}{\epsilon} \right) + 4 \ln 15. \]

Therefore, \( T < 5 + \frac{1}{\ln 15} \ln \left( 1 + \frac{f_0 - f^*}{\epsilon} \right) \).