Accepted version of the article published in IEEE Transactions on Automatic Control. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2022.3146054

Restart of accelerated first order methods with linear convergence under a quadratic functional growth condition

Teodoro Alamo, Pablo Krupa, Daniel Limon

***Abstract*—Accelerated first order methods, also called fast gradient methods, are popular optimization methods in the field of convex optimization. However, they are prone to suffer from oscillatory behaviour 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 behaviour. This paper presents a restart scheme for accelerated first order methods for which we show linear convergence under the satisfaction of a quadratic functional growth condition, thus encompassing a broad class of non-necessarily strongly convex optimization problems. Moreover, the worst-case convergence rate is comparable to the one obtained using a (generally non-implementable) optimal fixed-rate restart strategy. We compare the proposed algorithm with other restart schemes by applying them to a model predictive control case study.**

***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 objective function value [5]. Some noteworthy examples are: Nesterov's fast gradient method [5], FISTA [3], accelerated ADMM [6], [7], fast Douglas-Rachford splitting [8], and fast AMA (FAMA) [9, §5].

The use of AFMOs in the field of control is a heavily researched topic, especially in the field of model predictive control (MPC), as evidenced by the following examples: [10], which employs FAMA on a condensed MPC optimization problem; [11] and [12], which consider Nesterov's fast gradient method; [13], where the infinite horizon constrained LQR problem is solved using an accelerated dual proximal method; [14], which uses the accelerated dual gradient-projection algorithm; [15] and [16], which use the fast gradient method along with the augmented Lagrangian method; and [17], where the FISTA algorithm is employed. AFOMs are also used in other areas closely related to the field of control, such as to find the solution of Lasso problems [18], which are employed, for instance, in sparse system identification [19].

A drawback of AFOMs is that they often suffer from oscillating behaviour that slows them down [20]. In order to mitigate this, restart schemes have been proposed in the literature, which have been shown to improve the convergence in a practical setting by suppressing the oscillatory behaviour. 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.

Several restart schemes have been proposed in the literature, including the following, where there are three main aspects to consider: *(i)* if they are generally implementable in practice, *(ii)* if they guarantee linear convergence, and *(iii)*, if so, under what assumptions. In [20] the authors propose two simple heuristic schemes (the functional and gradient restarts) that work well in practice but lack, for the most part, linear convergence guarantees. The results of this article were extended in [21], but the conditions for linear convergence are still restrictive. In [22, §5.2.2], the authors propose a restart scheme, that extends the scheme from [23, §5.1], for optimization problems satisfying a *quadratic functional growth* (QFG) condition [22, Definition 4], which can be viewed as a relaxation of strong convexity. This scheme, which guarantees linear convergence, restarts the AFOM after a fixed number of iterations, resulting in a very simple implementation. However, its drawback is that it requires knowledge of parameters of the optimization problem that are generally hard to obtain, such as the parameter that characterizes the QFG condition. A restart scheme for FISTA that also guarantees linear convergence for optimization problems satisfying the QFG condition is presented in [24]. The scheme requires an initial estimation of the QFG parameter, thus not requiring its exact value. However, we find that it is rather sensitive to this initial guess. Finally, the authors proposed in [25] and [26] two restart schemes with linear convergence for optimization problems satisfying the QFG condition that do not require knowledge of hard to obtain parameters of the optimization problem nor estimations of them, including the QFG parameter. However, both schemes are specific to FISTA.

This article presents a novel restart scheme for AFOMs that exhibits linear convergence for optimization problems satisfying a QFG condition, thus encompassing a broad class of non-necessarily strongly convex problems. Furthermore, it does not require hard-to-attain information of the objective function, such as the QFG parameter. We provide a theoretical upper bound on the number of iterations needed to achieve a desired accuracy and show that the obtained convergence

The authors are with the Department of Systems Engineering and Automation, Universidad de Sevilla, 41092 Seville, Spain (e-mails: talamo@us.es, pkrupa@us.es, dlm@us.es). Corresponding author: Teodoro Alamo.
This work was supported in part by Grant PID2019-106212RB-C41 funded by MCIN/AEI/10.13039/501100011033, in part by Grant P20_00546 funded by the Junta de Andalucía and by ERDF A way of making Europe, and in part by Grant PDC2021-121120-C21 funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR".

rate is comparable to the one that could be obtained using the restart scheme from [22, §5.2.2], which is optimal for the class of AFOMs and optimization problems under consideration.

This paper extends the preliminary results presented for FISTA in the conference paper [25] by providing a restart scheme that is applicable to a broad class of AFOM algorithms and with an improved worst-case convergence rate.

The remainder of this article is structured as follows. In Section II we formally present the class of optimization problem and AFOM algorithms under consideration. Section III describes the optimal fixed-rate restart strategy from [22] and provides its iteration complexity for our class of AFOMs. Section IV presents the novel implementable restart scheme with linear convergence. In Section V we compare the proposed scheme with some of the alternatives referenced above by applying them to solve an MPC problem using the FAMA algorithm from [10]. We draw conclusions in Section VI.

Notation: Given a norm $\|\cdot\|$, we denote by $\|\cdot\|_*$ its dual norm: $\|x\|_* \doteq \sup\{\ x^T z \ : \ \|z\| \leq 1\ \}$. The ℓ_1-norm is denoted by $\|\cdot\|_1$. $\mathbb{Z}_+$ denotes the set of non-negative integers. The set of integer numbers from i to j is denoted with $\mathbb{Z}_i^j$, i.e. $\mathbb{Z}_i^j \doteq \{i, i+1, \ldots, j-1, j\}$. Euler's number is denoted by e, and the natural logarithm by $\ln(\cdot)$. $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, and $\lfloor x \rfloor$ the largest integer smaller than or equal to x. The set of proper closed convex functions from $\mathbb{R}^n$ to $(-\infty, \infty]$ is denoted by Γ_n. Given $f \in \Gamma_n$, we denote by $\mathrm{dom}(f)$ its effective domain, that is, $\mathrm{dom}(f) \doteq \{\ x \in \mathbb{R}^n \ : \ f(x) < \infty\ \}$. Further notation is given in Notations 1 and 2 in the following section.

II. Problem statement

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

$$f^* = \min_{x \in \mathbb{R}^n} f(x), \tag{1}$$

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 Ω_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 Ω_f with respect to norm $\|\cdot\|$, i.e.

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

(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

$$f(x) - f^* \geq \frac{\mu}{2}\|x - \bar{x}\|^2, \ \forall x \in \mathrm{dom}(f).$$

This inequality is called the quadratic functional growth (QFG) condition and it is satisfied, at least locally, for a large class of not necessarily strongly convex functions [22], [27], [28].

Let us consider a fixed point algorithm $\mathcal{A}$ that can be applied to solve (1), i.e. given a initial point $x_0 \in \mathrm{dom}(f)$, algorithm $\mathcal{A}$ generates a sequence $\{x_k\}$ with $k \geq 0$ such that $\lim_{k\to\infty} f(x_k) = f^*$. 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 we consider.

Assumption 1. *We assume that:*

(i) For every $\rho > 0$, $f \in \Gamma_n$ satisfies the QFG condition

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

for some $\mu_\rho > 0$.

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

$$f(\mathcal{A}(x_0, 1)) \leq f(x_0) - \frac{1}{2L_f}\|g(x_0)\|_*^2, \tag{2}$$

$$f(\mathcal{A}(x_0, k)) - f^* \leq \frac{a_f}{(k+1)^2}\|x_0 - \bar{x}_0\|^2, \ \forall k \geq 1, \tag{3}$$

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 $n_\rho \doteq \max\left\{\frac{1}{2}, \sqrt{\frac{2a_f}{\mu_\rho}}\right\}$.

The conditions listed in Assumption 1 are expressed in very general terms to be able to account for a broad class of optimization problems and AFOMs. Let us start by providing some insight and examples of the conditions and terms listed in the assumption. Assumption 1.*(i)* establishes a local QFG condition on the objective function, which encompasses a broad class of non-necessarily strongly convex function (see [22, Fig. 1]). We refer the reader to [22], [27] and [28] for examples of functions satisfying this condition, including the case $f(x) = h(Ex) + c^\top x + \mathrm{I}_\mathcal{X}(x)$, where $h : \mathbb{R}^m \to \mathbb{R}$, is a smooth strictly convex function, $E \in \mathbb{R}^{m\times n}$ and $\mathrm{I}_\mathcal{X}$ is the indicator function of a polyhedral set $\mathcal{X}$, which encompasses a large family of optimization problems. Assumption 1.*(ii)* is written in terms of the constants L_f and a_f, which will depend both on the AFOM $\mathcal{A}$ being used and on the structure of f, and in terms of an operator $g : \mathbb{R}^n \to \mathbb{R}^n$, which plays the role of the *gradient* operator of $\mathcal{A}$. For instance, in the FISTA algorithm applied to $f = h + \Psi$, where h is an L-smooth differentiable convex function and Ψ is a (possibly non-smooth) convex function, we have that $L_f = L$, $a_f = 2L$ and g is the composite gradient mapping operator [25]. Condition (3) is satisfied by most AFOMs [1], [2], including the ones we list in the introduction, although under varying assumptions. Condition (2) is also satisfied by most AFOMs because the first iteration is often the result of the application of a proximal (or composite) gradient mapping operator. If this is not the case, then it can be easily enforced by taking this operator as the first step of the algorithm.

In conclusion, Assumption 1 is satisfied by a broad family of AFOMs and optimization problems of interest in the field of control, including the AFOMs listed in the introduction and optimization problems such as QPs, Lasso or those that arise from many MPC formulations.

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 [22, Subsection 5.2.2].

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

$$f(\mathcal{A}(x_0, k)) - f^* \leq \left(\frac{n_\rho}{k+1} \right)^2 (f(x_0) - f^*), \ \forall k \geq 1. \quad (4)$$

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

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

III. Optimal fixed-rate restart scheme

This section describes the optimal fixed restart scheme presented in [22, §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.

A fixed-rate restart scheme takes the recursion

$$v_{j+1} = \mathcal{A}(v_j, n), \ j \geq 0, \quad (5)$$

where $n \geq 1$ is a fixed integer, starting at a given $v_0 \in V_f(\rho)$.

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 $\lceil e n_\rho \rceil$. See also [22, §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* $M \geq \bar{M}$, *where*

$$\bar{M} \doteq 1 + \frac{1}{2(\ln n - \ln n_\rho)} \ln \left(1 + \frac{f(v_0) - f^*}{\epsilon} \right). \quad (6)$$

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

$$\bar{N}_F^* \doteq \lceil e n_\rho \rceil \left\lceil 1 + \frac{1}{2} \ln \left(1 + \frac{f(v_0) - f^*}{\epsilon} \right) \right\rceil. \quad (7)$$

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 linear optimal convergence rate provided by Nesterov's fast gradient method for strongly convex functions [22], [29, §2.2]. That is, recalling that $n_\rho = \max\{1/2, \sqrt{2a_f/\mu_\rho}\}$ we easily obtain from Property 2.*(ii)* that an ϵ accurate solution is obtained in

$$\mathcal{O}\left(n_\rho \ln \left(\frac{f(v_0) - f^*}{\epsilon} \right) \right) \quad (8)$$

iterations. However, we note that this scheme is often non-implementable because the value of n_ρ is generally not available. Nevertheless, (7) and (8) are of interest because they provide the best theoretical convergence rate that can be obtained with a fixed-rate restart strategy.

Algorithm 1: Performance-based exit condition on $\mathcal{A}$

Prototype: $[z, m] = \mathcal{A}_d(r, n)$
Require : $r \in \operatorname{dom}(f)$, $n \in \mathbb{R}$
1 $x_0 \leftarrow r$, $k \leftarrow 0$
2 Initialize $\mathcal{A}$ with x_0
3 **repeat**
4 $\quad k \leftarrow k + 1$
5 $\quad x_k \leftarrow \begin{cases} \mathcal{A}(x_0, k) & \text{if } f(\mathcal{A}(x_0, k)) \leq f(x_{k-1}) \\ x_{k-1} & \text{otherwise} \end{cases}$
6 $\quad \ell \leftarrow \left\lfloor \frac{k}{2} \right\rfloor$
7 **until** $k \geq n$ *and* $f(x_\ell) - f(x_k) \leq \frac{1}{3}(f(x_0) - f(x_\ell))$
Output: $z \leftarrow x_k$, $m \leftarrow k$

Algorithm 2: Optimal Algorithm based on $\mathcal{A}_d$

Prototype: $[z_{out}, j_{out}] = \mathcal{A}_*(z_0)$
Require : $z_0 \in \operatorname{dom}(f)$, $\epsilon > 0$
$m_0 \leftarrow 1$, $m_{-1} \leftarrow 1$, $j \leftarrow -1$
repeat
$\quad j \leftarrow j + 1$
$\quad s_j \leftarrow \begin{cases} \sqrt{\dfrac{f(z_{j-1}) - f(z_j)}{f(z_{j-2}) - f(z_j)}} & \text{if } j \geq 2 \\ 0 & \text{otherwise} \end{cases}$
$\quad n_j \leftarrow \max\{m_j, 4 s_j m_{j-1}\}$
$\quad [z_{j+1}, m_{j+1}] \leftarrow \mathcal{A}_d(z_j, n_j)$
until $f(z_j) - f(z_{j+1}) \leq \epsilon$
Output: $z_{out} \leftarrow z_{j+1}$, $j_{out} \leftarrow j$

IV. Proposed restart scheme

In this section we propose a novel restart scheme that does not require knowledge of n_ρ 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 performance-based exit condition of algorithm $\mathcal{A}$. Algorithm 1 will then be used to derive the main result of this article: Algorithm 2.

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,

$$f(x_k) = \min_{i=0,\dots,k} f(\mathcal{A}(x_0, i)). \quad (9)$$

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

$$f(x_\ell) - f(x_k) \leq \frac{1}{3}\left(f(x_0) - f(x_\ell)\right), \tag{10}$$

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, 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 constant $1/3$ was chosen because it minimized the upper bound on the maximum number of iterations that we show below in Theorem 1.*(iii)*.

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 1 holds. Then, the output $[z, m]$ from the call $[z, m] = \mathcal{A}_d(r, n)$ of Algorithm 1 satisfies, for every $r \in V_f(\rho)$:*

(i) $f(z) \leq f(r) - \frac{1}{2L_f}\|g(r)\|_*^2$,

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

(iii) $n \in (0, \lceil 4n_\rho \rceil] \implies m \in [n, \lceil 4n_\rho \rceil]$.

Proof. See Appendix B.

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 4 and 5). The main properties of the iterates of Algorithm 2 are given in the following property and theorem.

Property 4. *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) Property 3 can be applied to the iterates of Algorithm 2 (i.e., taking $r \equiv z_j$, $n \equiv n_j$ $z \equiv z_{j+1}$ and $m \equiv m_{j+1}$).

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

$$m_j \leq n_j \leq m_{j+1}, \ \forall j \in \mathbb{Z}_0^{j_{out}}. \tag{11}$$

(iii) The sequence $\{s_j\}$ satisfies $s_j \in (0, 1]$, $\forall j \in \mathbb{Z}_2^{j_{out}}$.

Proof. See Appendix C.

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 $\mathcal{A}_d$ (step 6) is bounded. That is, j_{out} is finite.

(ii) The number of iterations of $\mathcal{A}$ at each call of $\mathcal{A}_d$ (step 6) is upper bounded by $\lceil 4n_\rho \rceil$. That is,

$$m_{j+1} \leq \lceil 4n_\rho \rceil, \ \forall j \in \mathbb{Z}_0^{j_{out}}. \tag{12}$$

(iii) The total number of iterations of $\mathcal{A}$ performed by a call to Algorithm 2, which we denote by $N_\mathcal{A}$, is upper-bounded by $N_\mathcal{A} \leq \bar{N}_\mathcal{A}$, where

$$\bar{N}_\mathcal{A} \doteq \frac{e\lceil 4n_\rho \rceil}{2}\left\lceil 5 + \frac{1}{\ln 15}\ln\left(1 + \frac{f(z_0) - f^*}{\epsilon}\right)\right\rceil.$$

Proof. See Appendix C.

Remark 1. *From Property 4.(i), we have that we can rearrange Property 3.(i) 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 $\|g(z_j)\|_* \leq \tilde{\epsilon}$, where $\tilde{\epsilon} > 0$. In this case, the upper bound on the number of iterations given in Theorem 1.(iii) would be the same but replacing ϵ with $\tilde{\epsilon}/(2L_f)$.

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 ϵ accurate solution is obtained in (8) iterations. This is an important result, since we recover the optimal convergence rate for our class of optimization problems using AFOMs.

Comparing the upper bound provided in Theorem 1.*(iii)* with the upper bound $\bar{N}_F^*$ (7) of the optimal fixed-rate restart scheme presented in Section III, we have

$$\frac{\bar{N}_\mathcal{A}}{\bar{N}_F^*} = \frac{e\lceil 4n_\rho \rceil\left\lceil 5 + \frac{1}{\ln 15}\ln\left(1 + \frac{f(z_0) - f^*}{\epsilon}\right)\right\rceil}{2\lceil en_\rho \rceil\left\lceil 1 + \frac{1}{2}\ln\left(1 + \frac{f(z_0) - f^*}{\epsilon}\right)\right\rceil},$$

from where we obtain that

$$\begin{aligned}\lim_{\epsilon \to 0}\frac{\bar{N}_A}{\bar{N}_F^*} &= \frac{e\lceil 4n_\rho \rceil}{\lceil en_\rho \rceil \ln 15} \leq \frac{e(4n_\rho + 1)}{en_\rho \ln 15}\\ &= \frac{4}{\ln 15} + \frac{1}{n_\rho \ln 15} \leq \frac{3}{2}\left(1 + \frac{1}{4n_\rho}\right).\end{aligned}$$

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 $\mathcal{A}$ when ϵ tends to zero).

V. Numerical results

We compare Algorithm 2 with other restart schemes of the literature by applying them to the FAMA algorithm for MPC from [10], where we consider the MPC formulation [30, Eq. (2)] but without its terminal constraint (2f). This MPC formulation can be posed as [10, Problem 2.1], which can therefore be solved using the FAMA algorithm [10, Alg. 1]. As stated in [10, §3.2], this algorithm is equivalent to applying FISTA to the dual problem of [10, Problem 2.1]. Therefore, the dual objective function value $D(\lambda)$ (see [10, §3.2]) and its gradient $\nabla D(\lambda)$, where λ are the dual variables, satisfy Assumption 1, with $\lambda_k \equiv x_k$, $D \equiv f$, $\nabla D \equiv g$, L_f is equal to the expression $\rho(C)/\sigma_f$ described in [10, §2] and $a_f = 2L_f$.

Accepted version of the article published in IEEE Transactions on Automatic Control. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2022.3146054

We consider the system described in [30, §3], which consists of three objects connected by springs where external forces can be applied to the two outer-most objects. We use the exact same setup as in [30, §3] with the exception of the cost function matrix T, which we take as the solution of the discrete algebraic Ricatti equation, as is often the case in MPC, and the state constraints, which we do not enforce. We perform the preconditioning procedure described in [10, §4] on the resulting MPC optimization problem.

We solve the MPC's optimization problem for 1000 randomly generated system states, where the positions of each object is obtained from a uniform distribution on the interval $[0, 4]$ dm and the velocities from a uniform distribution on the interval $[-0.5, 0.5]$ m/s. We solve each problem using the FAMA algorithm [10, Alg. 1] with different restart schemes:

(*i*) ***Functional***: The heuristic restart scheme proposed in [20] that uses restart condition $D(\lambda_{k+1}) \leq D(\lambda_k)$.
(*ii*) ***Gradient***: The heuristic restart scheme proposed in [20] that uses restart condition $\langle \nabla D(\lambda_k), \lambda_k - \lambda_{k+1} \rangle \geq 0$.
(*iii*) ***Optimal***: The restart scheme proposed in [22, §5.2.2], which is typically non-implementable. To include it for comparison with the proposed scheme, we implement it using the method described at the end of the referenced subsection that requires knowing $D(\lambda^*)$, which we compute using FAMA with a very small exit tolerance.
(*iv*) ***GLCR***: The restart FISTA algorithm with linear convergence proposed in [25, Alg. 2].
(*v*) ***GBRF***: The gradient-based restart FISTA algorithm proposed in [26, Alg. 2].
(*vi*) ***Adaptive***: The adaptive scheme for FISTA from [24]. We take the initial guess of the QFG parameter as 10^{-5}.

We also use the proposed restart scheme (Algorithm 2) and the non-restarted FAMA.

Table I shows the average, median, maximum and minimum number of iterations of each scheme for the 1000 tests. The results of the table are obtained by terminating FAMA when the iterate $\mathbf{u}_k$ of the primal problem (see [10, Alg. 1]) satisfies $\|\mathbf{u}_k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2 \leq 10^{-5}$, where $\|\cdot\|_2$ stands for the standard Euclidean norm and the optimal solution $\mathbf{u}^*$ of the primal problem is obtained using the `quadprog` solver from Matlab. We use this exit condition to provide a fair comparison between the different approaches. Figure 1 shows the evolution of $\|\mathbf{u}_k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2$ for each restart scheme when taking the system state as the origin. The results indicate that the restart schemes tend to reduce the number of iterations required to obtain a solution of the MPC's optimization problem when compared with the non-restarted variant, although this is not always the case.

Additional numerical results can be found in [31, §3.3], where the above restart schemes are applied to solve random QP and Lasso problems with varying condition numbers.

TABLE I: Comparison between restart schemes.

Scheme	Avg. Iter.	Med. Iter.	Max. Iter.	Min. Iter
No restart	6262.7	5124.5	26659	48
Alg. 2	1115.9	1080.5	2902	23
GLCR	1103.0	1070.5	2963	29
GBRF	2184.0	2132.5	5991	48
Optimal	2135.3	2095.0	5893	53
Functional	845.2	801.0	4134	26
Gradient	835.5	801.5	2147	26
Adaptive	1858.1	1842.5	3827	48

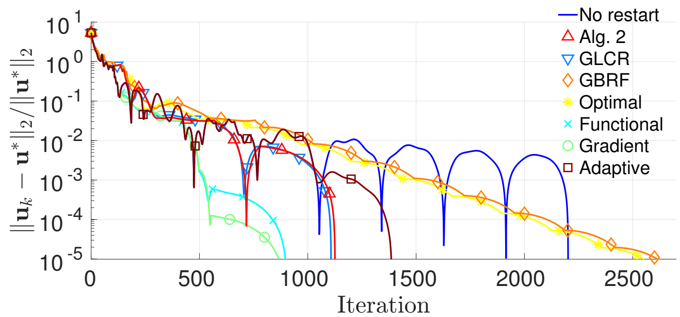

Fig. 1: Evolution of distance to optimality for each scheme.

VI. Conclusions

We propose a restart scheme applicable to a broad class of AFOMs that does not require knowledge of hard-to-obtain parameters of the optimization problem and still retains a linear convergence rate similar to the optimal one for optimization problems satisfying a QFG condition, i.e., its worst case complexity is similar to the best one that can be obtained if the parameters characterizing the convergence of the AFOM were known. The scheme is based on a performance-based exit condition that detects the degradation of the convergence of the AFOM and which we show is attained in a finite number of iterations. The numerical results indicate that the proposed algorithm is comparable, in practical terms, with other restart schemes of the literature. In our numerical results we find that the application of restart schemes may be desirable in many situations, since the non-restarted variant can require an excessive number or iterations.

Appendix

A. 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

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

Using this inequality in a recursive manner we obtain

$$\begin{aligned} \epsilon < f(v_{M-1}) - f(v_M) &\leq f(v_{M-1}) - f^* \\ &\leq \left(\frac{n_\rho}{n}\right)^{2(M-1)} (f(v_0) - f^*). \end{aligned}$$

This leads to

$$M < 1 + \frac{1}{2(\ln n - \ln n_\rho)} \ln \left(\frac{f(v_0) - f^*}{\epsilon} \right) < \bar{M}. \quad (13)$$

Thus, we conclude that if M does not satisfy (13), 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

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

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,

$$N_F(n) \leq n\left\lceil 1 + \frac{1}{2(\ln n - \ln n_\rho)} \ln\left(1 + \frac{f(v_0) - f^*}{\epsilon}\right)\right\rceil. \quad (14)$$

Simple calculus yields that the value that minimizes $\frac{n}{\ln n - \ln n_\rho}$ is $n^* = en_\rho$. 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 (14) we finally obtain

$$N_F(\lceil en_\rho \rceil) \leq \lceil en_\rho \rceil \left\lceil 1 + \frac{1}{2} \ln\left(1 + \frac{f(v_0) - f^*}{\epsilon}\right)\right\rceil. \quad \blacksquare$$

B. Proof of Property 3

From (9) and Assumption 1 we have

$$\begin{aligned} f(z) = f(x_m) &\overset{(9)}{=} \min_{i=0,\dots,m} f(\mathcal{A}(x_0, i)) \\ &\leq f(\mathcal{A}(x_0, 1)) \overset{(2)}{\leq} f(x_0) - \frac{1}{2L_f} \|g(x_0)\|_*^2. \end{aligned}$$

The first claim now follows directly from $x_0 = r$. In view of Property 1 we have, for every $k \in \mathbb{Z}_1^m$,

$$f(x_k) - f^* \overset{(9)}{\leq} f(\mathcal{A}(x_0, k)) - f^* \overset{(4)}{\leq} \left(\frac{n_\rho}{k+1}\right)^2 (f(x_0) - f^*). \quad (15)$$

The second claim now follows from $x_m = z$ and $x_0 = r$. The inequality $n \leq m$ is trivially satisfied from step 7. Thus, in order to conclude the proof we show that inequality (10) is satisfied for $\hat{k} = \lceil 4n_\rho \rceil$ and $\hat{\ell} = \left\lfloor \frac{\lceil 4n_\rho \rceil}{2} \right\rfloor \geq \lfloor 2n_\rho \rfloor \geq 1$ (where this last inequality follows from Assumption 1.*(iii)*, which states that $n_\rho \geq 1/2$).

$$\begin{aligned} f(x_{\hat{\ell}}) - f^* &\overset{(15)}{\leq} \left(\frac{n_\rho}{\hat{\ell} + 1}\right)^2 (f(x_0) - f^*) \\ &\leq \left(\frac{n_\rho}{2n_\rho}\right)^2 (f(x_0) - f^*) = \frac{1}{4}(f(x_0) - f^*). \end{aligned}$$

This implies $f(x_{\hat{\ell}}) \leq \frac{1}{4} f(x_0) + \frac{3}{4} f^* \leq \frac{1}{4} f(x_0) + \frac{3}{4} f(x_{\hat{k}})$. Thus,

$$\begin{aligned} f(x_{\hat{\ell}}) - f(x_{\hat{k}}) &\leq \frac{1}{4}(f(x_0) - f(x_{\hat{k}})) \\ &= \frac{1}{4}(f(x_0) - f(x_{\hat{\ell}})) + \frac{1}{4}(f(x_{\hat{\ell}}) - f(x_{\hat{k}})). \end{aligned}$$

We conclude that $f(x_{\hat{\ell}}) - f(x_{\hat{k}}) \leq \dfrac{1}{3}(f(x_0) - f(x_{\hat{\ell}}))$. $\blacksquare$

C. Proofs for Algorithm 2

Proof of Property 4. Since $z_0 \in \mathrm{dom}(f)$, we have that $z_0 \in V_f(\rho)$ for some $\rho > 0$. Additionally, each z_j is obtained from a call to Algorithm 1 (step 6). As such, in view of Property 3.*(i)*, we have that the iterates z_j satisfy $z_j \in V_f(\rho)$, $\forall j \in \mathbb{Z}_0^{j_{out}}$. Therefore, Property 3 can be applied to each call to $\mathcal{A}_d$, thus proving claim *(i)*. That is, for every $j \geq 0$, the iterates of Algorithm 2 satisfy:

$$f(z_{j+1}) \leq f(z_j) - \frac{1}{2L_f} \|g(z_j)\|_*^2, \quad (16a)$$

$$f(z_{j+1}) - f^* \leq \left(\frac{n_\rho}{m_{j+1} + 1}\right)^2 (f(z_j) - f^*), \quad (16b)$$

$$n_j \in (0, \lceil 4n_\rho \rceil] \Rightarrow m_{j+1} \in [n_j, \lceil 4n_\rho \rceil]. \quad (16c)$$

Next, due to step 5 we have $m_j \leq n_j$, $j \in \mathbb{Z}_0^{j_{out}}$. Moreover, from (16c), we have that $n_j \leq m_{j+1}$, $\forall j \in \mathbb{Z}_0^{j_{out}}$, which proves claim *(ii)*.

Finally, we prove claim *(iii)*. From the exit condition (step 7), we have

$$f(z_{j-1}) - f(z_j) > \epsilon, \ \forall j \in \mathbb{Z}_1^{j_{out}}. \quad (17)$$

Additionally, from (16a) we have $f(z_{j-2}) \geq f(z_{j-1})$, $\forall j \in \mathbb{Z}_2^{j_{out}}$. Thus,

$$f(z_{j-2}) - f(z_j) \geq f(z_{j-1}) - f(z_j) \overset{(17)}{>} \epsilon > 0, \ \forall j \in \mathbb{Z}_2^{j_{out}}.$$

Therefore, from step 4, taking $j \geq 2$, we have

$$0 < s_j = \sqrt{\frac{f(z_{j-1}) - f(z_j)}{f(z_{j-2}) - f(z_j)}} \leq 1, \ \forall j \in \mathbb{Z}_2^{j_{out}}. \quad \blacksquare$$

The proof of the following lemma relies upon some technical results on the iterates of Algorithm 2, namely Lemmas 2 and 3, which we include in Appendix D.

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 $j_{out} \geq D$, where*

$$D \doteq \left\lceil 5 + \frac{1}{\ln 15} \ln\left(1 + \frac{f(z_0) - f^*}{\epsilon}\right)\right\rceil.$$

Then, $m_{\ell+1} \leq \dfrac{1}{\sqrt{15}} m_{\ell+1+D}, \ \forall \ell \in \mathbb{Z}_0^{j_{out}-D}$.

Proof. The proof is obtained by reductio ad absurdum. If there is $\ell \in \mathbb{Z}_0^{j_{out}-D}$ such that $m_{\ell+1} > \frac{1}{\sqrt{15}} m_{\ell+1+D}$, then we obtain from Lemma 3.*(iv)* (see Appendix D) that

$$D < 5 + \frac{1}{\ln 15} \ln\left(1 + \frac{f(z_0) - f^*}{\epsilon}\right),$$

which contradicts the definition of D. $\blacksquare$

Proof of Theorem 1. Let $T \in \mathbb{Z}$ be such that

$$f(z_j) - f(z_{j+1}) > \epsilon, \ \forall j \in \mathbb{Z}_0^T, \quad (18)$$

is satisfied. Then, defining $d_j \doteq f(z_j) - f(z_{j+1})$, we have

$$f(z_0) - f(z_{T+1}) = \sum_{j=0}^{T} d_j \geq (T+1)\left(\min_{j=0,\dots,T} d_j\right) > (T+1)\epsilon.$$

Thus, $T + 1 < \dfrac{f(z_0) - f(z_{T+1})}{\epsilon} \leq \dfrac{f(z_0) - f^*}{\epsilon} \leq \dfrac{\rho}{\epsilon}$, from where we infer that the largest integer T satisfying (18) is bounded. Consequently, the exit condition of Algorithm 2 (see step 7) is satisfied within a finite number of iterations, thus proving claim *(i)*.

To prove claim *(ii)*, we start by noting that both m_1 and m_2 are no larger than $\lceil 4n_\rho \rceil$. 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 $\lceil 4n_\rho \rceil$ we have from (16c) that m_1 is also upper-bounded by $\lceil 4n_\rho \rceil$. Moreover, since $n_1 = m_1 \leq \lceil 4n_\rho \rceil$, we obtain by the same reasoning that $m_2 \leq \lceil 4n_\rho \rceil$. We now prove that if $j \geq 2$ and $m_j \leq \lceil 4n_\rho \rceil$, then $m_{j+1} \leq \lceil 4n_\rho \rceil$. From step 4 we have

$$\begin{aligned} s_j^2 &= \frac{f(z_{j-1}) - f(z_j)}{f(z_{j-2}) - f(z_j)} = 1 - \frac{f(z_{j-2}) - f(z_{j-1})}{f(z_{j-2}) - f(z_j)} \\ &\leq 1 - \frac{f(z_{j-2}) - f(z_{j-1})}{f(z_{j-2}) - f^*} \\ &= \frac{f(z_{j-1}) - f^*}{f(z_{j-2}) - f^*} \overset{\text{(16b)}}{\leq} \left(\frac{n_\rho}{m_{j-1} + 1} \right)^2. \end{aligned}$$

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

$$n_j = \max\{m_j, 4s_j m_{j-1}\} \leq \max\{\lceil 4n_\rho \rceil, 4n_\rho\} = \lceil 4n_\rho \rceil,$$

which, along with (16c), leads to $m_{j+1} \leq \lceil 4n_\rho \rceil$, thus proving the claim.

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

$$N_\mathcal{A} = \sum_{j=0}^{j_{out}} m_{j+1} \overset{(12)}{\leq} (1 + j_{out}) \lceil 4n_\rho \rceil. \tag{19}$$

Let us denote

$$D \doteq \left\lceil 5 + \frac{1}{\ln 15} \ln \left(1 + \frac{f(z_0) - f^*}{\epsilon} \right) \right\rceil.$$

Consider first the case $j_{out} < D$. Since both j_{out} and D are integers we infer from this inequality that $1 + j_{out} \leq D$. This, along with (19), implies that $N_\mathcal{A} \leq \lceil 4n_\rho \rceil D \leq \bar{N}_\mathcal{A}$.

Suppose now that $j_{out} \geq D$. We first recall that Property 4.*(ii)* states that the sequence $\{m_{j+1}\}_{j \geq 0}$ is non-decreasing. We now rewrite j_{out} as $j_{out} = d + tD$, where $d \in \mathbb{Z}_0^{D-1}$ and t is a non-negative integer. Thus,

$$\begin{aligned} N_\mathcal{A} &= \sum_{j=0}^{j_{out}} m_{j+1} = \sum_{j=0}^{d} m_{j+1} + \sum_{j=1}^{tD} m_{d+j+1} \\ &\leq D m_{d+1} + D \sum_{i=1}^{t} m_{d+1+iD} = D \sum_{i=0}^{t} m_{d+1+iD}. \end{aligned}$$

From Lemma 1 we have

$$m_{d+1+iD} \leq \frac{m_{d+1+(i+1)D}}{\sqrt{15}}, \ \forall i \in \mathbb{Z}_0^{t-1}.$$

Thus, $N_\mathcal{A} \leq D \sum\limits_{i=0}^{t} m_{d+1+tD} \left(\frac{1}{\sqrt{15}} \right)^{t-i}$.

Using now $m_{1+d+tD} \leq \bar{m} = \lceil 4n_\rho \rceil$ (see (12)) we obtain

$$\begin{aligned} \frac{N_\mathcal{A}}{D\bar{m}} &\leq \sum_{i=0}^{t} \left(\frac{1}{\sqrt{15}} \right)^{t-i} = \sum_{j=0}^{t} \left(\frac{1}{\sqrt{15}} \right)^{j} \\ &\leq \sum_{j=0}^{\infty} \left(\frac{1}{\sqrt{15}} \right)^{j} = \frac{\sqrt{15}}{\sqrt{15} - 1} \leq \frac{e}{2}. \end{aligned}$$

Thus, $N_\mathcal{A} \leq \frac{e}{2} \bar{m} D = \frac{e}{2} \lceil 4n_\rho \rceil D$. ∎

D. Technical results on the iterates of Algorithm 2

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

$$\varphi(s) \doteq \left(\frac{1}{s^2} - 1 \right) \cdot \max\left\{1, (4s)^4\right\},$$

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

Proof. We have that

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

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 $[\frac{1}{4}, \frac{1}{\sqrt{2}})$ 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

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

∎

Lemma 3 (Technical results on the iterates of Alg. 2)**.** *Consider Algorithm 2 with the initial condition $z_0 \in V_f(\rho)$, and $\epsilon > 0$. Suppose that Assumption 1 is satisfied and that $j_{out} \geq 2$. Suppose also that there is $T \in \mathbb{Z}_2^{j_{out}}$ and $\ell \in \mathbb{Z}_0^{j_{out}-T}$ such that $m_{\ell+1} > \frac{1}{\sqrt{15}} m_{\ell+1+T}$. Then:*

(i) $s_j \in \left(0, \frac{\sqrt{15}}{4}\right], \ \forall j \in \mathbb{Z}_{\ell+2}^{\ell+T}$.

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

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

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

Proof. Denote $f_j = f(z_j)$, $j \in \mathbb{Z}_0^{j_{out}+1}$. From $j \geq 2$ and step 4 of Algorithm 2 we have

$$s_j^2 = \frac{f_{j-1} - f_j}{f_{j-2} - f_j}, \ j \in \mathbb{Z}_2^{j_{out}}.$$

The inequality $s_j > 0, \ \forall j \in \mathbb{Z}_{\ell+2}^{\ell+T}$ follows from Property 4.*(iii)*. In order to prove the first claim it remains to prove the inequality $s_j \leq \frac{\sqrt{15}}{4}, \ \forall j \in \mathbb{Z}_{\ell+2}^{\ell+T}$. We proceed by reductio ad absurdum. Suppose that there is $j \in \mathbb{Z}_{\ell+2}^{\ell+T}$ such that $s_j > \frac{\sqrt{15}}{4}$. In this case,

$$m_{j+1} \overset{(11)}{\geq} n_j = \max\{m_j, 4s_j m_{j-1}\} \geq 4s_j m_{j-1} > \sqrt{15} m_{j-1},$$

which along the non-decreasing nature of the sequence $\{m_j\}$ (Property 4.*(ii)*) leads to

$$m_{\ell+1+T} \geq m_{j+1} > \sqrt{15} m_{j-1} \geq \sqrt{15} m_{\ell+1},$$

contradicting the assumption of the property, proving claim *(i)*.

From the non-decreasing nature of the sequence $\{m_j\}$ (Property 4.(*ii*)) we have, for every $j \in \mathbb{Z}_{\ell+2}^{\ell+T}$,

$$m_{j+1} \overset{(11)}{\geq} n_j = \max\{m_j, 4s_j m_{j-1}\} \geq m_{j-1} \cdot \max\{1, 4s_j\}.$$

Equivalently, $\ln(\max\{1, 4s_j\}) \leq \ln \frac{m_{j+1}}{m_{j-1}}$, $\forall j \in \mathbb{Z}_{\ell+2}^{\ell+T}$. This implies

$$\begin{aligned}\sum_{j=\ell+2}^{\ell+T} \ln(\max\{1, 4s_j\}) &\leq \sum_{j=\ell+2}^{\ell+T} \ln \frac{m_{j+1}}{m_{j-1}} \\ &= \ln \frac{m_{\ell+T} m_{\ell+1+T}}{m_{\ell+1} m_{\ell+2}} \leq \ln \frac{m_{\ell+1+T}^2}{m_{\ell+1}^2} = 2\ln \frac{m_{\ell+1+T}}{m_{\ell+1}} \\ &< 2\ln\sqrt{15} = \ln 15. \end{aligned} \tag{20}$$

The second claim is obtained multiplying the last inequality by 4. To prove the third claim we notice that

$$\prod_{j=\ell+2}^{\ell+T} \left(\frac{1}{s_j^2} - 1\right) = \prod_{j=\ell+2}^{\ell+T} \frac{f_{j-2} - f_{j-1}}{f_{j-1} - f_j} = \frac{f_\ell - f_{\ell+1}}{f_{\ell+T-1} - f_{\ell+T}}.$$

Since $\ell + T \leq j_{out}$ we have $f_{\ell+T-1} - f_{\ell+T} > \epsilon > 0$. Using this inequality we obtain

$$\prod_{j=\ell+2}^{\ell+T} \left(\frac{1}{s_j^2} - 1\right) < \frac{f_\ell - f_{\ell+1}}{\epsilon} \overset{(16a)}{\leq} \frac{f_0 - f_{\ell+1}}{\epsilon} \leq \frac{f_0 - f^*}{\epsilon},$$

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

$$\begin{aligned}&\sum_{j=\ell+2}^{\ell+T} \ln\left(\left(\frac{1}{s_j^2} - 1\right) \cdot \max\{1, (4s_j)^4\}\right) \\ &< \ln\left(1 + \frac{f_0 - f^*}{\epsilon}\right) + 4\ln 15. \end{aligned} \tag{21}$$

From the first claim we have $s_j \in \left(0, \frac{\sqrt{15}}{4}\right]$, $\forall j \in \mathbb{Z}_{\ell+2}^{\ell+T}$. Thus, the left term of (21) can be lower bounded by means of the following inequality (Lemma 2)

$$15 \leq \left(\frac{1}{s^2} - 1\right) \cdot \max\{1, (4s)^4\}, \ \forall s \in \left(0, \frac{\sqrt{15}}{4}\right].$$

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

Equivalently, $(T-1)\ln 15 < \ln\left(1 + \frac{f_0 - f^*}{\epsilon}\right) + 4\ln 15.$ ■

References

[1] Y. Nesterov, *Lectures on convex optimization*. Springer, 2018, vol. 137.
[2] A. Beck, *First-order methods in optimization*. SIAM, 2017, vol. 25.
[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
[5] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Sov. Math. Dokl.*, vol. 27, no. 2, pp. 372–376, 1983.
[6] I. Pejcic and C. N. Jones, "Accelerated ADMM based on accelerated Douglas-Rachford splitting," in *IEEE European Control Conference (ECC)*, 2016, pp. 1952–1957.
[7] Y. Zheng, G. Fantuzzi, A. Papachristodoulou, P. Goulart, and A. Wynn, "Fast ADMM for homogeneous self-dual embedding of sparse SDPs," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8411–8416, 2017.
[8] P. Patrinos, L. Stella, and A. Bemporad, "Douglas-Rachford splitting: Complexity estimates and accelerated variants," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 4234–4239.
[9] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal on Imaging Sciences*, vol. 7, no. 3, pp. 1588–1623, 2014.
[10] Y. Pu, M. N. Zeilinger, and C. N. Jones, "Fast alternating minimization algorithm for model predictive control," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 11 980–11 986, 2014.
[11] M. Alamir, "Monitoring control updating period in fast gradient based NMPC," in *IEEE European Control Conference*, 2013, pp. 3621–3626.
[12] S. Richter, C. N. Jones, and M. Morari, "Computational complexity certification for real-time MPC with input constraints based on the fast gradient method," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1391–1403, 2012.
[13] G. Stathopoulos, M. Korda, and C. N. Jones, "Solving the infinite-horizon constrained LQR problem using accelerated dual proximal methods," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1752–1767, 2016.
[14] P. Patrinos and A. Bemporad, "An accelerated dual gradient-projection algorithm for embedded linear model predictive control," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 18–33, 2014.
[15] M. Kögel and R. Findeisen, "Fast predictive control of linear systems combining Nesterov's gradient method and the method of multipliers," in *50th IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 501–506.
[16] V. Nedelcu, I. Necoara, and Q. Tran-Dinh, "Computational complexity of inexact gradient augmented Lagrangian methods: application to constrained MPC," *SIAM Journal on Control and Optimization*, vol. 52, no. 5, pp. 3109–3134, 2014.
[17] P. Krupa, D. Limon, and T. Alamo, "Implementation of model predictive control in programmable logic controllers," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 3, pp. 1117–1130, 2021.
[18] S. Tao, D. Boley, and S. Zhang, "Local linear convergence of ISTA and FISTA on the LASSO problem," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 313–336, 2016.
[19] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse identification of nonlinear dynamics with control (SINDYc)," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 710–715, 2016.
[20] B. O'Donoghue and E. Candes, "Adaptive restart for accelerated gradient schemes," *Foundations of computational mathematics*, vol. 15, no. 3, pp. 715–732, 2015.
[21] P. Giselsson and S. Boyd, "Monotonicity and restart in fast gradient methods," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 5058–5063.
[22] I. Necoara, Y. Nesterov, and F. Glineur, "Linear convergence of first order methods for non-strongly convex optimization," *Mathematical Programming*, vol. 175, no. 1, pp. 69–107, 2019.
[23] Y. Nesterov, "Gradient methods for minimizing composite functions," *Mathematical Programming*, vol. 140, pp. 125–161, 2013.
[24] O. Fercoq and Z. Qu, "Adaptive restart of accelerated gradient methods under local quadratic growth condition," *IMA Journal of Numerical Analysis*, vol. 39, no. 4, pp. 2069–2095, 2019.
[25] T. Alamo, P. Krupa, and D. Limon, "Restart FISTA with global linear convergence," in *18th IEEE European Control Conference*, 2019, pp. 1969–1974.
[26] ——, "Gradient based restart FISTA," in *58th IEEE Conference on Decision and Control*, 2019, pp. 3936–3941.
[27] D. Drusvyatskiy and A. S. Lewis, "Error bounds, quadratic growth, and linear convergence of proximal methods," *Mathematics of Operations Research*, vol. 43, no. 3, pp. 919–948, 2018.
[28] P.-W. Wang and C.-J. Lin, "Iteration complexity of feasible descent methods for convex optimization," *Journal of Machine Learning Research*, vol. 15, pp. 1523–1548, 2014.
[29] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*. Springer, 2004, vol. 87.
[30] P. Krupa, R. Jaouani, D. Limon, and T. Alamo, "A sparse ADMM-based solver for linear MPC subject to terminal quadratic constraint," *arXiv:2105.08419v2*, 2021.
[31] P. Krupa, "Implementation of MPC in embedded systems using first order methods," Ph.D. dissertation, University of Seville, 2021, available at arXiv:2109.02140.