Accelerated ADMM: Automated Parameter Tuning and Improved Linear Convergence

M. Tavakoli [*,†] **F. Jakob** [**,†] **G. Carnevale** [*] **G. Notarstefano** [*]
A. Iannelli [**]

[*] *Università di Bologna, Department of Electrical, Electronic and Information Engineering, Bologna, Italy (e-mail: meisam.tavakoli@studio.unibo.it, guido.carnevale, giuseppe.notarstefano@unibo.it).*
[**] *University of Stuttgart, Institute for Systems Theory and Automatic Control, Stuttgart, Germany (e-mail: fabian.jakob, andrea.iannelli@ist.uni-stuttgart.de).*
[†] *The first two authors contributed equally to this work.*

Abstract: This work studies the linear convergence of an accelerated scheme of the Alternating Direction Method of Multipliers (ADMM) for strongly convex and Lipschitz-smooth problems. We use the methodology of expressing the accelerated ADMM as a Lur'e system, i.e., an interconnection of a linear dynamical system in feedback with a slope-restricted operator, and we use Integral Quadratic Constraints to establish linear convergence. We leverage this machinery to systematically explore parameter tuning heuristics, including Nesterov-inspired choices and configurations identified via grid search, and analyze their impact on the convergence rate. Our new bounds show improved linear convergence rates compared to the vanilla algorithm and previously proposed accelerated variants, which is also empirically validated on a LASSO regression benchmark.

Keywords: ADMM, Composite optimization, Lur'e systems, Integral Quadratic Constraints

1. INTRODUCTION

The Alternating Direction Method of Multipliers (ADMM) (Glowinski and Marroco, 1975; Boyd et al., 2011) has emerged as a benchmark algorithm for solving convex composite optimization problems. Its ability to decompose the original problem into subproblems that can be solved in parallel made it especially popular in computation-heavy domains such as signal processing, machine learning, or distributed control.

It has long been established that ADMM enjoys provable convergence guarantees in the convex setting. Deng and Yin (2016) showed that for general convex problems, ADMM achieves an $\mathcal{O}(1/k)$ convergence rate, while for the strongly convex and smooth case it attains a linear convergence rate, with optimally tuned parameters given in Giselsson and Boyd (2017). An alternative proof of linear convergence was presented by Nishihara et al. (2015), who applied the Integral Quadratic Constraints (IQC) framework (Megretski and Rantzer, 1997; Lessard et al., 2016; Michalowsky et al., 2021) to numerically derive convergence rates and tune parameters. The proof frames the optimization algorithm as a Lur'e system (a feedback interconnection of an LTI system and monotone operators), and casts the convergence analysis as a stability problem. This framework offers the main advantage that the analysis of altered schemes, such as over-relaxed variants, can also be seamlessly carried out in an automated fashion as long as they can be framed in the Lur'e setting.

Accelerated variants of ADMM (A-ADMM) have also been proposed and analyzed in recent years. Goldstein et al. (2014) showed that Nesterov acceleration leads to a convergence rate of $\mathcal{O}(1/k^2)$ in the strongly convex setting. Wang et al. (2021) provided linear convergence guarantees when Anderson acceleration is applied. Patrinos et al. (2014); Pejcic and Jones (2016) proved linear convergence of accelerated Douglas–Rachford splitting and ADMM for strongly convex quadratic problems, showing that the optimal rate matches that of Nesterov's fast gradient method. However, an analysis of the convergence rates of A-ADMM for general strongly convex and smooth problems is still unaddressed, as well as the question of whether other known rates from classical accelerated gradient methods (AGMs) can be matched.

In this work, we extend the IQC-based analysis of vanilla ADMM (Nishihara et al., 2015) to the accelerated case, when one of the objectives is strongly convex and smooth. By casting the A-ADMM algorithm as a Lur'e system, we propose a semi-definite program (SDP) that provides a numerical tool to theoretically verify worst-case convergence rates of accelerated schemes using different parameter selections. Crucially, we show that dynamic O'Shea-Zames-Falb (OZF) IQCs (Scherer, 2023) are essential to

[1] F. Jakob acknowledges the support of the International Max Planck Research School for Intelligent Systems (IMPRS-IS).

Table 1. Contextualization of the results. Here, $\rho_1 = 1 - \frac{1}{\sqrt{\kappa}}$, $\rho_2 = \sqrt{1 - \frac{\sqrt{2\kappa-1}}{\kappa}}$, with $\kappa = \frac{L}{m}$. NM stands for Nesterov parameter selection (Table 2).

Setting	Objective f	Algorithm	Rate ρ	Reference
$\min f(x)$	Quadr.	Nesterov Method	ρ_1	Nesterov (2004)
$\min f(x)$	$S(m, L)$	Nesterov Method	ρ_2	Safavi et al. (2018)
(1)	Quadr.	A-ADMM (NM)	ρ_1	Patrinos et al. (2014); Pejcic and Jones (2016)
(1)	$S(m, L)$	A-ADMM (NM)	Fig. 1	This work

certify convergence of the accelerated scheme. Based on this, we propose a systematic parameter tuning procedure and demonstrate that Nesterov-inspired parameter selections match the rates of the analogous AGM in the strongly convex and smooth setting (cf. Table 1). Finally, we identify through grid search a new A-ADMM configuration with the fastest certified rate. Our theoretical results are validated on a LASSO regression case study, demonstrating consistently improved convergence speed of our schemes compared to existing benchmark algorithms.

The remainder of the paper is organized as follows. Section 2 presents the problem setup, the ADMM formulation, and the required IQC preliminaries. Section 3 provides a Lur'e representation of the A-ADMM scheme. In Section 4, we study the convergence rates of this A-ADMM scheme under different parameter configurations. In Section 5, we provide a case study on LASSO regression. Finally, Section 6 concludes the paper.

Notation. The identity matrix of dimension p is denoted as I_p. The gradient and subdifferential of a function f are denoted by ∇f and ∂f, respectively. For $0 < m \leq L < \infty$, we let $S_p(m, L)$ denote the class of functions $f : \mathbb{R}^p \to \mathbb{R}$ that are m-strongly convex and have L-Lipschitz continuous gradients. The special case $S_p(0, \infty)$ corresponds to the set of proper, closed and convex functions. For a matrix M, its condition number is defined as $\kappa_M = \bar{\sigma}(M)/\underline{\sigma}(M)$, where $\bar{\sigma}(M)$ and $\underline{\sigma}(M)$ denote the largest and smallest singular values of M, respectively. A discrete-time LTI system maps the signals $u \mapsto y$ via the recursion $\xi_{k+1} = A\xi_k + Bu_k,\ y_k = C\xi_k + Du_k$ as a ξ_0 dependent mapping, and will be compactly expressed as $y = Gu$ with $G = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. We will write $G \otimes I_d$ to define a lifted system $\left[\begin{array}{c|c} A\otimes I_d & B\otimes I_d \\ \hline C\otimes I_d & D\otimes I_d \end{array}\right]$. The series interconnection between two LTI systems $G_1 : u \mapsto y$ and $G_2 : y \mapsto w$ will be denoted as $G_1 \cdot G_2 : u \mapsto w$. The forward shift operator is denoted as $\mathbf{z}$.

2. PRELIMINARIES

2.1 Accelerated ADMM

We consider constrained convex optimization problems of the form

$$\min_{x\in\mathbb{R}^p,\ z\in\mathbb{R}^q} \ f(x) + g(z) \quad \text{s.t.} \quad Ax + Bz = c, \tag{1}$$

where $f : \mathbb{R}^p \to \mathbb{R}$, $f \in S_p(m, L)$ and $g : \mathbb{R}^q \to \mathbb{R}$, $g \in S_q(0, \infty)$. The matrices $A \in \mathbb{R}^{p\times p}$, $B \in \mathbb{R}^{p\times q}$, and $c \in \mathbb{R}^p$ define linear equality constraints coupling the variables x and z, where we particularly assume that A is invertible and that B has full column rank. Having a strongly convex and smooth component in (1) is common in many practical applications (Deng and Yin, 2016), while restricting A to be square and invertible is slightly restrictive, but also done in Nishihara et al. (2015). We note that many practical problem instances still satisfy this structure, including consensus problems and various distributed optimization tasks (Notarstefano et al., 2019).

To address problem (1), we use the over-relaxed ADMM algorithm (OR-ADMM) (Boyd et al., 2011)

$$x_{k+1} = \arg\min_{x\in\mathbb{R}^p} \ f(x) + \tfrac{1}{2\nu_1}\left\|Ax + B\hat{z}_k - c + \hat{\lambda}_k\right\|^2 \tag{2a}$$

$$z_{k+1} = \arg\min_{z\in\mathbb{R}^q} \ g(z) + \tfrac{1}{2\nu_1}\left\|\alpha Ax_{k+1} - (1-\alpha)B\hat{z}_k + Bz - \alpha c + \hat{\lambda}_k\right\|^2 \tag{2b}$$

$$\lambda_{k+1} = \hat{\lambda}_k + \alpha Ax_{k+1} - (1-\alpha)B\hat{z}_k + Bz_{k+1} - \alpha c \tag{2c}$$

as a starting point, and augment it with a *momentum term*

$$\hat{z}_k = z_k + \nu_2(z_k - z_{k-1}) \tag{3a}$$

$$\hat{\lambda}_k = \lambda_k + \nu_2(\lambda_k - \lambda_{k-1}), \tag{3b}$$

as in Goldstein et al. (2014); Pejcic and Jones (2016). Here, $\nu_1 > 0$ is a penalty parameter, $\nu_2 \geq 0$ a momentum parameter, and $\alpha \geq 0$ the so-called over-relaxation parameter. We recover the nominal A-ADMM by setting $\alpha = 1$, and the vanilla, non-accelerated, ADMM with $\nu_2 = 0$. Both α and ν_2 have been shown to improve the convergence speed for suitable choices (Eckstein and Bertsekas, 1992; Patrinos et al., 2014). The over-relaxation parameter α is chosen in the interval $(0, 2]$, while ν_2 is often selected according to Nesterov's schemes (Nesterov, 2004), e.g., as in Patrinos et al. (2014). To explore alternative parameter selection strategies, we employ the IQC framework.

2.2 Integral Quadratic Constraints for Convex Functions

Algorithm (2)-(3) will be analyzed as a dynamical system converging to points that satisfy the first-order optimality conditions of (1). We will make use of the well-known fact that (sub)gradients of (strongly) convex functions are monotone and, thus, their input/output relation satisfies IQCs for slope-restricted operators (Lessard et al., 2016). To be more precise, they have been shown to fulfill so-called ρ-hard O'Shea-Zames–Falb IQCs.

Proposition 1. (Scherer (2023)). Let $f \in S_p(m, L)$, $g \in S_p(0, \infty)$ and let $\rho \in (0, 1)$ be an exponential discount factor. For $\bar{n} \in \mathbb{N}_0$, let $\{h^f_\tau\}_{\tau=-\bar{n}}^{\bar{n}}$ and $\{h^g_\tau\}_{\tau=-\bar{n}}^{\bar{n}}$ be sequences of filter coefficients satisfying $h^f_0, h^g_0 \geq 0$ and

$$h^*_\tau \leq 0 \ \ \forall \tau \neq 0, \ \ \sum_{\tau=-\bar{n}}^{\bar{n}} \rho^{-2\tau} h^*_\tau \geq 0, \ \ \sum_{\tau=-\bar{n}}^{\bar{n}} \rho^{2\tau} h^*_\tau \geq 0 \tag{4}$$

for $* \in \{f, g\}$, and define the parametrized LTI filters

$$\Psi_g(h^g) = \begin{bmatrix} \sum_{\tau=0}^{\bar{n}} h_\tau^g \mathbf{z}^{-\tau} & 0 \\ 0 & \sum_{\tau=-\bar{n}}^{0} h_\tau^g \mathbf{z}^{\tau} \end{bmatrix} \tag{5}$$
$$\Psi_f(h^f) = \Psi_g(h^f) \begin{bmatrix} L & -1 \\ -m & 1 \end{bmatrix}$$

Let $\{a_k\}, \{b_k\}$ be some p-dim. square summable sequences and let $a^\star, b^\star \in \mathbb{R}^p$ be constant references. For $\gamma_k \in \partial g(b_k)$ and $\gamma^\star \in \partial g(b^\star)$, define

$$\tilde{a}_k = a_k - a^\star, \qquad \nabla \tilde{f}_k = \nabla f(a_k) - \nabla f(a^\star),$$
$$\tilde{b}_k = b_k - b^\star, \qquad \tilde{\gamma}_k = \gamma_k - \gamma^\star.$$

Finally, define the filtered sequences

$$\psi_{1,k} = (\Psi_f \otimes I_p) \begin{bmatrix} \tilde{a}_k \\ \nabla \tilde{f}_k \end{bmatrix}, \quad \psi_{2,k} = (\Psi_g \otimes I_p) \begin{bmatrix} \tilde{b}_k \\ \tilde{\gamma}_k \end{bmatrix}$$

and $M = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. Then, for all $T \geq 0$, it holds

$$\sum_{k=0}^{T} \rho^{-2k} \psi_{i,k}^\top (M \otimes I_p) \psi_{i,k} \geq 0, \quad i = 1, 2. \tag{6}$$

The filter coefficients h^f, h^g will serve as degrees-of-freedom in the resulting SDP, subject to the convex constraint (4). In theory, the filter dimension $\bar{n}$ can be infinite, but is chosen finite in practice to yield finite-dimensional state-space realizations. A choice of $\bar{n} = 0$ leads to static filters Ψ_f, Ψ_g and pointwise satisfaction of the inequalities (6), which is often referred to in the literature as a sector IQC (Lessard et al., 2016). While sector IQCs are sufficient to certify the convergence of vanilla ADMM (Nishihara et al., 2015), we will show in Section 4 that the more general dynamic OZF IQCs ($\bar{n} > 0$) are essential for the accelerated case.

2.3 Stability Analysis of First-Order Algorithms

Consider the following Lur'e representation of a first-order algorithm

$$\begin{aligned} \xi_{k+1} &= (\hat{A} \otimes I_p)\xi_k + (\hat{B} \otimes I_p) w_k \\ \hat{v}_k &= (\hat{C} \otimes I_p)\xi_k + (\hat{D} \otimes I_p) w_k \end{aligned} \tag{7a}$$

with state $\xi_k \in \mathbb{R}^{n_\xi}$, output $\hat{v}_k = v_k + \bar{v}$ for some constant offset $\bar{v}$, and

$$v_k = \begin{bmatrix} v_{1,k} \\ v_{2,k} \end{bmatrix}, \quad w_k = \begin{bmatrix} \nabla \hat{f}(v_{1,k}) \\ \gamma_k \end{bmatrix}, \gamma_k \in \partial \hat{g}(v_{2,k}), \tag{7b}$$

where $\hat{f} \in S_p(\hat{m}, \hat{L})$, $\hat{g} \in S_p(0, \infty)$ for some $\hat{m}, \hat{L}$. It has been shown that ADMM, as well as many other composite optimization algorithms, can be represented as (7) (Nishihara et al., 2015) when $\hat{f}, \hat{g}$ are suitably chosen. In general, $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ fulfills structural assumptions such that its fixed-point $(\xi^\star, \hat{v}^\star, w^\star)$ is unique and satisfies first-order optimality (Upadhyaya et al., 2024). Showing exponential stability of (7) is therefore the same as showing linear convergence of the underlying algorithm. With a coordinate shift $(\tilde{\xi}_k, \tilde{v}_k, \tilde{w}_k) \triangleq (\xi_k - \xi^\star, \hat{v}_k - \hat{v}^\star, w_k - w^\star)$ it is straightforward to show that the error coordinates evolve with the same state-space description (7a), i.e.,

$$\tilde{v} = \Bigg(\underbrace{\left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array}\right]}_{=:G} \otimes I_p \Bigg) \tilde{w}. \tag{8}$$

Note that $\tilde{v}_k = \hat{v}_k - \hat{v}^\star = v_k - v^\star$ (eliminating the constant offset), so that by Proposition 1 the gradient and subgradient components of $(\tilde{v}, \tilde{w})$ satisfy a ρ-OZF IQC. With a suitable permutation and stacking of the filters Ψ_1, Ψ_2, we can form a compact filter Ψ such that

$$\psi_k \triangleq \begin{bmatrix} \psi_{1,k} \\ \psi_{2,k} \end{bmatrix} = (\Psi \otimes I_p) \begin{bmatrix} \tilde{v}_k \\ \tilde{w}_k \end{bmatrix}, \tag{9}$$

with $\psi_{1,k}, \psi_{2,k}$ as in Proposition 1 (with $a \triangleq v_1$, $b \triangleq v_2$). Define the augmented plant as the mapping $\tilde{w} \mapsto \psi$, realized as

$$\Psi \cdot \begin{bmatrix} G \\ I_2 \end{bmatrix} \triangleq \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right], \tag{10}$$

where $\left[\begin{smallmatrix} G \\ I_2 \end{smallmatrix}\right]$ denotes the dynamical system arising by vertically concatenating the transfer matrix of G and a static feedthrough, mapping $\tilde{w} \to \left[\begin{smallmatrix} \tilde{v} \\ \tilde{w} \end{smallmatrix}\right]$. Then the following convergence theorem holds.

Theorem 2. Take $\rho \in (0, 1)$, and filters Ψ_1, Ψ_2 that satisfy Proposition 1 with $m = \hat{m}, L = \hat{L}$. Form the augmented plant (10). If there exist $P \succ 0$ and filter coefficients of Ψ_1, Ψ_2 such that the matrix inequality

$$\begin{bmatrix} \mathbf{A}^\top P \mathbf{A} - \rho^2 P & \mathbf{A}^\top P \mathbf{B} \\ \mathbf{B}^\top P \mathbf{A} & \mathbf{B}^\top P \mathbf{B} \end{bmatrix} + \begin{bmatrix} \mathbf{C}^\top \\ \mathbf{D}^\top \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} \mathbf{C} & \mathbf{D} \end{bmatrix} \preceq 0 \tag{11}$$

holds, then (7) is exponentially stable with rate ρ, i.e.,

$$\|\xi_k - \xi^\star\| \leq \sqrt{\kappa_P}\, \rho^k \|\xi_0 - \xi^\star\|. \tag{12}$$

The proof comes as a straightforward extension of (Lessard et al., 2016, Theorem 4) to the two-operator case, which particularly exploits the positivity constraint (6). Eq. (11) is independent of the problem dimension p, and can be used to determine the minimum worst-case convergence rate ρ via bisection. Note that Theorem 2 recovers (Nishihara et al., 2015, Theorem 6) for $\bar{n} = 0$.

3. A-ADMM AS A DYNAMICAL SYSTEM

Building on the previous derivations, we now formulate the A-ADMM (2)-(3) as a Lur'e system (7). In line with Nishihara et al. (2015), we introduce the coordinate change $r_k = Ax_k$, $s_k = Bz_k$. Moreover, define

$$\hat{f} = f \circ A^{-1}, \quad \hat{g} = g \circ B^\dagger + \mathcal{I}_{\operatorname{im} B}, \tag{13}$$

where $B^\dagger$ is a left-inverse of B and $\mathcal{I}_{\operatorname{im} B}$ is the indicator of the image of B. It is straightforward to verify that $\hat{g} \in S_p(0, \infty)$ and $\hat{f} \in S_p(\hat{m}, \hat{L})$ with

$$\hat{m} = \tfrac{m}{\bar{\sigma}^2(A)}, \quad \hat{L} = \tfrac{L}{\underline{\sigma}^2(A)}. \tag{14}$$

Accordingly, we define the condition number of the problem as $\kappa = \frac{\hat{L}}{\hat{m}} = \kappa_f \kappa_A^2$.

Consequently, the updates (2a), (2b), (3a) can be rewritten as

$$x_{k+1} = A^{-1} \arg\min_{r \in \mathbb{R}^p} \ \hat{f}(r) + \tfrac{1}{2\nu_1} \|r + \hat{s}_k - c + \hat{\lambda}_k\|^2 \tag{15a}$$

$$z_{k+1} = B^\dagger \arg\min_{s \in \mathbb{R}^p} \ \hat{g}(s) + \tfrac{1}{2\nu_1} \|\alpha r_{k+1} - (1-\alpha)\hat{s}_k + s - \alpha c + \hat{\lambda}_k\|^2 \tag{15b}$$

$$\hat{s}_k = s_k + \nu_2 (s_k - s_{k-1}). \tag{15c}$$

Using the definition of the proximal operator

$$\operatorname{prox}_f(z) := \arg\min_x f(x) + \frac{1}{2}\|x - z\|^2,$$

we can summarize the A-ADMM in transformed coordinates compactly as

$$r_{k+1} = \mathrm{prox}_{\nu_1 \hat{f}}\big(c - \hat{s}_k - \hat{\lambda}_k\big) \tag{16a}$$
$$s_{k+1} = \mathrm{prox}_{\nu_1 \hat{g}}\big(\alpha c - \alpha r_{k+1} - (\alpha-1)\hat{s}_k - \hat{\lambda}_k\big) \tag{16b}$$
$$\lambda_{k+1} = \alpha r_{k+1} + (\alpha-1)\hat{s}_k + s_{k+1} - \alpha c + \hat{\lambda}_k, \tag{16c}$$

with $\hat{s}_k, \hat{\lambda}_k$ as in (15c), (3b). Next, we bring the recursion (16) into the form (7).

Proposition 3. Define the state $\xi_k \triangleq \begin{bmatrix} \lambda_{k-1}^\top & s_{k-1}^\top & \lambda_k^\top & s_k^\top \end{bmatrix}^\top$, output $v_k \triangleq \begin{bmatrix} r_{k+1}^\top & s_{k+1}^\top \end{bmatrix}^\top$, offset $\bar{v} \triangleq \begin{bmatrix} -c^\top & 0 \end{bmatrix}^\top$, and input $w_k \triangleq \begin{bmatrix} \nabla \hat{f}(r_{k+1})^\top & \gamma_k^\top \end{bmatrix}^\top$ for any $\gamma_k \in \partial \hat{g}(s_{k+1})$. Then the sequences ξ_k, w_k, and $\hat{v}_k = v_k + \bar{v}$ generated by the A-ADMM recursion (16) satisfy (7) with the matrices

$$\hat{A} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -\nu_2(\alpha-1) & -\nu_2 & (\alpha-1)(1+\nu_2) & 1+\nu_2 \end{pmatrix},$$
$$\hat{C} = \begin{pmatrix} \nu_2 & \nu_2 & -(1+\nu_2) & -(1+\nu_2) \\ -\nu_2(\alpha-1) & -\nu_2 & (\alpha-1)(1+\nu_2) & 1+\nu_2 \end{pmatrix},$$
$$\hat{B} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & -\nu_1 \\ \alpha\nu_1 & -\nu_1 \end{pmatrix}, \quad \hat{D} = \begin{pmatrix} -\nu_1 & 0 \\ \alpha\nu_1 & -\nu_1 \end{pmatrix}. \tag{17}$$

Proof. Apply the first-order optimality condition to the r_{k+1} and s_{k+1} updates (16a)-(16b)

$$0 = \nu_1 \nabla \hat{f}(r_{k+1}) + r_{k+1} - c + \hat{s}_k + \hat{\lambda}_k, \tag{18a}$$
$$0 \in \nu_1 \partial \hat{g}(s_{k+1}) + s_{k+1} - \alpha c + \alpha r_{k+1} + (\alpha-1)\hat{s}_k + \hat{\lambda}_k. \tag{18b}$$

Take some $\gamma_k \in \partial \hat{g}(s_{k+1})$, solve for r_{k+1} and s_{k+1}, and plug them into the dual update λ_{k+1} to get

$$r_{k+1} = -\hat{s}_k - \hat{\lambda}_k + c - \nu_1 \nabla \hat{f}(r_{k+1}), \tag{19a}$$
$$s_{k+1} = \hat{s}_k + (\alpha-1)\hat{\lambda}_k + \alpha\nu_1 \nabla \hat{f}(r_{k+1}) - \nu_1 \gamma_k, \tag{19b}$$
$$\lambda_{k+1} = -\nu_1 \gamma_k. \tag{19c}$$

Now plug in the definitions of $\hat{s}_k, \hat{\lambda}_k$ and observe that s_{k+1}, λ_{k+1}, and $r_{k+1} - c$ can all be written as linear combinations of the state ξ_k and input w_k. It is then straightforward to bring ξ_{k+1} and $\hat{v}_k$ into the matrix form (7). □

Theorem 2 together with Proposition 3 provides a numerical tool to certify upper bounds on the worst-case convergence rate of A-ADMM. Next, we explore the influence of acceleration and the algorithm parameters on these rates.

4. PARAMETER SELECTION

We investigate different parameter selection heuristics for ν_1 and ν_2. As initial heuristics, we adopt the same parameter tuning schemes of AGMs. Next, we perform a grid search to identify potentially superior configurations. As a baseline for comparison, we use a vanilla ADMM instance with the parameter ν_1 tuned according to the optimal choice in the strongly convex and smooth setting characterized by Giselsson and Boyd (2017).

Table 2. AGM-inspired tuning of A-ADMM, where $\rho := 1 - 1/\sqrt{\kappa}$.

Method	ν_1	ν_2
NM	$\frac{1}{\hat{L}}$	$\frac{\sqrt{\hat{L}}-\sqrt{\hat{m}}}{\sqrt{\hat{L}}+\sqrt{\hat{m}}}$
TM	$\frac{1+\rho}{\hat{L}}$	$\frac{\rho^2}{2-\rho}$

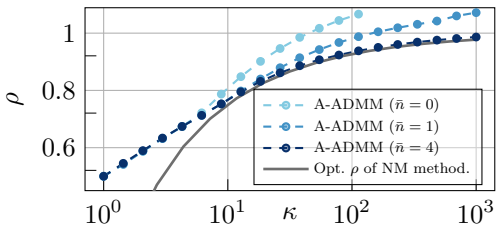


Fig. 1. Comparison of ρ values and influence of O'Shea-Zames-Falb order for A-ADMM with NM tuning.

4.1 AGM-inspired Parameters

Since Nesterov-inspired (NM) tuning (cf. Table 2) is a standard benchmark among AGMs and has proven effective for quadratic ADMM problems (Patrinos et al., 2014), it is also a natural candidate for the strongly-convex-smooth ADMM. We use the machinery developed in the previous sections to numerically verify the convergence rates ρ for this choice of (ν_1, ν_2). By bisection, we find the smallest ρ rendering (11) feasible and plot its value as a function of the condition number $\kappa = \hat{L}/\hat{m}$ [2].

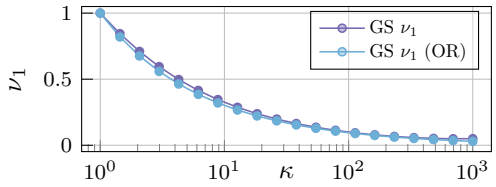
The resulting convergence rate curve is reported in Fig. 1. To highlight the role of dynamic IQC multipliers, we plot the curve resulting from different choices of the OZF order $\bar{n}$. We emphasize that the choice of a sector IQC ($\bar{n} = 0$) is not sufficient to certify convergence for all condition numbers κ. This observation is consistent with AGMs (Lessard et al., 2016). Increasing the OZF order to $\bar{n} = 4$ improves the rate substantially, which then closely tracks the classical NM rate $\rho = \sqrt{1 - \frac{\sqrt{2\kappa-1}}{\kappa}}$ (Safavi et al., 2018). The rate is matched for condition ratios larger than $\kappa \approx 10$; with deviations for low κ that are also typical of vanilla ADMM (Nishihara et al., 2015).

Our framework also allows us to seamlessly evaluate alternative parameters. For instance, since the Triple Momentum (TM) algorithm (Van Scoy et al., 2018) is known to outperform NM in the standard strongly-convex-smooth setting, we also investigate its application to A-ADMM. The comparison curves for vanilla ADMM, and A-ADMM with NM- and TM-tuning (cf. Table 2) are reported in Fig. 3. We observe that while TM parameters offer advantages in certain κ regimes, the certified rate exceeds one for high condition numbers, even with high-order OZF multipliers. Thus, this configuration does not yet provide a globally certified improvement over the NM scheme in the ADMM setting.

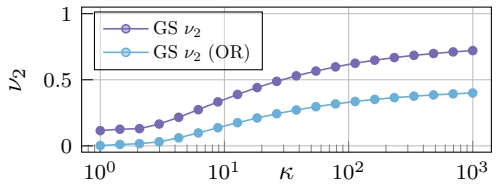
4.2 Grid Search Results

Next, we perform a grid search to identify optimal (ν_1, ν_2) configurations. For each κ, we set up a fixed grid from

[2] Code for all experiments can be found on https://github.com/col-tasas/2025-accelerated-admm



(a) Results for ν_1.



(b) Results for ν_2.

Fig. 2. Results of the grid search for (ν_1, ν_2).

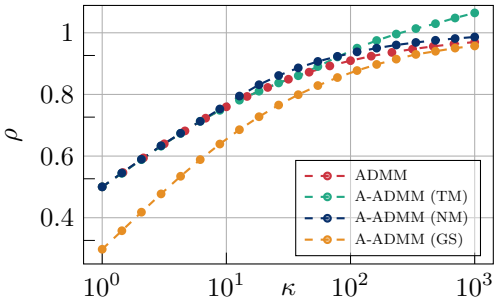


Fig. 3. Comparison of optimal ρ values for ADMM and A-ADMM with different parameter configurations.

which (ν_1, ν_2) is selected as the pair minimizing ρ via bisection on Theorem 2. The procedure is repeated for the over-relaxed version (OR-A-ADMM) with an additional grid over $\alpha \in (0, 4]$.

Fig. 2 shows the results for ν_1 and ν_2. Applying symbolic regression (Cranmer, 2023) reveals that ν_1 and ν_2 can be approximated by

$$\nu_1 \approx \frac{2.26}{\log(\kappa + 5.49)}, \qquad \nu_2 \approx 0.27\sqrt{\log(\kappa)},$$
$$\nu_1^{\mathrm{OR}} = \nu_1, \qquad \nu_2^{\mathrm{OR}} \approx 0.079\ \sqrt[4]{\kappa},$$

for the nominal and over-relaxed settings, respectively. The over-relaxation parameter α was found to be optimal for $\alpha \approx 1.45$ across all condition numbers.

Fig. 3 displays the resulting convergence rates obtained from the grid search, in the context of the vanilla and accelerated schemes. By design, A-ADMM with GS parameters gives the best linear convergence rate among all settings. Fig. 4 complements the results with the over-relaxed variants. We observe that the over-relaxed versions consistently improve upon their non-over-relaxed counterparts across the full range of condition numbers. Combining over-relaxation with the GS-tuned acceleration yields the fastest certified linear convergence rate among all configurations considered in this work.

5. CASE STUDY

We test the A-ADMM schemes designed based on the theoretical analyses in the previous section on the ℓ_1-regularized least-squares problem, also known as LASSO regression, and compare them with available benchmark algorithms (Boyd et al., 2011). We consider the problem

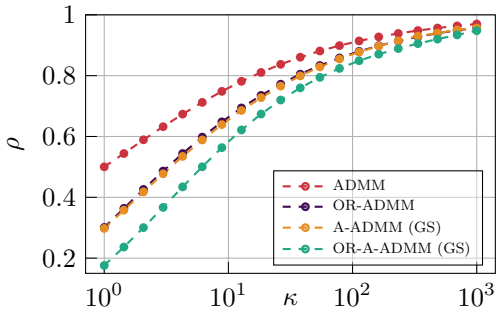


Fig. 4. Comparison of optimal ρ values for nominal and over-relaxed ADMM versions.

Table 3. Average iterations to reach $\Delta_k < 10^{-4}$ over 200 random LASSO instances.

Method	Average iterations
ADMM	197.66
OR-ADMM	163.63
A-ADMM (NM)	137.30
FISTA	136.71
A-ADMM (GS)	108.03
OR-A-ADMM (GS)	98.77

$$\min_{x,z\in\mathbb{R}^{500}} \quad \tfrac{1}{2}\|Fx - b\|_2^2 + \tau\|z\|_1$$
$$\text{s.t.} \quad x = z, \tag{20}$$

where $F \in \mathbb{R}^{600\times 500}$ is chosen as a full rank matrix with entries that are first sampled from an isotropic Gaussian distribution $\mathcal{N}(0, 1)$, whose columns are then normalized. The vector b is generated as $b = Fw^0 + \varepsilon$, where $w^0 \in \mathbb{R}^{500}$ is chosen sparse, containing 250 non-zeros drawn from $\mathcal{N}(0, 1)$ and $\varepsilon \sim \mathcal{N}(0, 10^{-3}I_{600})$. The regularization parameter $\tau > 0$ is set to $\tau = 1$. Problem (20) fits (1), where the first term is m-strongly convex and L-smooth with $m = \underline{\sigma}^2(F)$ and $L = \bar{\sigma}^2(F)$.

Applying the accelerated over-relaxed ADMM results in

$$x_{k+1} = (\nu_1 F^\top F + I)^{-1}\left(-\hat{z}_k - \hat{\lambda}_k + \nu_1 F^\top b\right), \tag{21a}$$
$$z_{k+1} = \mathcal{S}_{\tau\nu_1}\left(-\alpha\, x_{k+1} - (\alpha - 1)\hat{z}_k - \hat{\lambda}_k\right), \tag{21b}$$
$$\lambda_{k+1} = \alpha\, x_{k+1} + (1 - \alpha)\,\hat{z}_k - z_{k+1} + \hat{\lambda}_k, \tag{21c}$$

with $\hat{z}_k$ and $\hat{\lambda}_k$ as in (3), and $\mathcal{S}_{\tau\nu_1}$ being the soft-thresholding operator, which is (element-wise) defined as

$$\left[\mathcal{S}_{\tau\nu_1}(\mathbf{y})\right]_i = \mathrm{sign}(\mathbf{y}_i)\,\max\big(|\mathbf{y}_i| - \tau\nu_1, 0\big). \tag{22}$$

The convergence is assessed through the normalized iterate error

$$\Delta_k := \frac{\|x_k - x^\star\|_2}{\|x_0 - x^\star\|_2}, \tag{23}$$

where $x^\star$ is the solution of (20).

Specifically, we compare A-ADMM using NM and GS parameter selections against the widely used LASSO benchmark solver FISTA (Beck and Teboulle, 2009) and the non-accelerated baselines ADMM and OR-ADMM. The iterate error computed on one random instance of (20) is shown in Fig. 5, while Table 3 summarizes the average number of iterations required to reach $\Delta_k < 10^{-4}$ over 200 random instances.

Both Fig. 5 and Table 3 show that the performance of the A-ADMM methods differs profoundly based on their

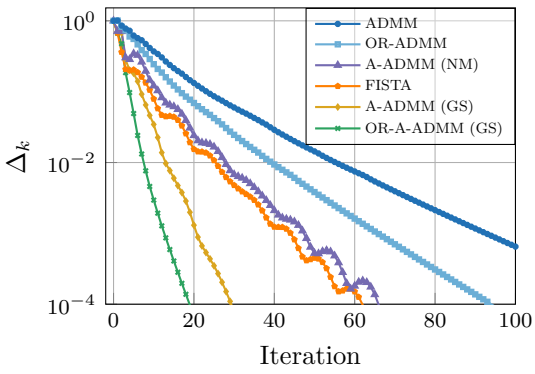


Fig. 5. Empirical performance of different A-ADMM schemes and other benchmark algorithms applied to a random instance of LASSO regression.

parameter selection. We also observe that the accelerated versions improve the convergence speed substantially over the vanilla baselines. Finally, OR-A-ADMM with GS tuning consistently outperforms all other methods. Notably, the ordering of algorithms evaluated on the empirical performance and evaluated on the worst-case convergence rates (cf. Fig. 4) stays the same, suggesting that the worst-case convergence rate is a suitable metric for algorithm comparison.

6. CONCLUSIONS

This paper extended the IQC-based algorithm analysis framework to certify worst-case linear convergence rates of A-ADMM for any choice of the algorithm parameters, thereby enabling a systematic investigation of A-ADMM performance under different parameter selections. We showed that dynamic O'Shea-Zames-Falb multipliers are essential to certify convergence in the accelerated setting. We further proposed different parameter tunings and showed that the additional momentum parameters allow A-ADMM to outperform vanilla ADMM. The fastest configuration, in terms of both certified convergence rate and empirical performance, was found using a grid search over the parameter space, giving the fastest version of A-ADMM in the strongly-convex-smooth setting. Future work could focus on closed-form characterizations of the optimal A-ADMM parameters and on a systematic A-ADMM synthesis procedure.

REFERENCES

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*.

Cranmer, M. (2023). Interpretable machine learning for science with PySR and SymbolicRegression.jl. arXiv:2305.01582.

Deng, W. and Yin, W. (2016). On the global and linear convergence of the generalized alternating direction method of multipliers. *Journal of Scientific Computing*, 66(3), 889–916.

Eckstein, J. and Bertsekas, D.P. (1992). On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1), 293–318.

Giselsson, P. and Boyd, S. (2017). Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *IEEE Transactions on Automatic Control*, 62(2), 532–544.

Glowinski, R. and Marroco, A. (1975). Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 9(2), 41–76.

Goldstein, T., O'Donoghue, B., Setzer, S., and Baraniuk, R. (2014). Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, 7(3), 1588–1623.

Lessard, L., Recht, B., and Packard, A. (2016). Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1), 57–95.

Megretski, A. and Rantzer, A. (1997). System analysis via integral quadratic constraints. *IEEE Transactions on Automatic Control*, 42(6), 819–830.

Michalowsky, S., Scherer, C., and Ebenbauer, C. (2021). Robust and structure exploiting optimisation algorithms: an integral quadratic constraint approach. *International Journal of Control*, 94(11), 2956–2979.

Nesterov, Y. (2004). *Introductory lectures on convex optimization*. Springer Science & Business Media.

Nishihara, R., Lessard, L., Recht, B., Packard, A., and Jordan, M. (2015). A general analysis of the convergence of ADMM. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, 343–352.

Notarstefano, G., Notarnicola, I., and Camisa, A. (2019). Distributed optimization for smart cyber-physical networks. *Foundations and Trends in Systems and Control*.

Patrinos, P., Stella, L., and Bemporad, A. (2014). Douglas-Rachford splitting: Complexity estimates and accelerated variants. In *53rd IEEE Conference on Decision and Control*.

Pejcic, I. and Jones, C.N. (2016). Accelerated ADMM based on accelerated Douglas-Rachford splitting. *Proc. European Control Conference (ECC)*.

Safavi, S., Joshi, B., França, G., and Bento, J. (2018). An explicit convergence rate for Nesterov's method from SDP. In *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE.

Scherer, C.W. (2023). Robust exponential stability and invariance guarantees with general dynamic O'Shea-Zames-Falb multipliers. *IFAC-PapersOnLine*, 56(2), 5799–5804.

Upadhyaya, M., Banert, S., Taylor, A.B., and Giselsson, P. (2024). Automated tight Lyapunov analysis for first-order methods. *Mathematical Programming*, 209(1), 133–170.

Van Scoy, B., Freeman, R.A., and Lynch, K.M. (2018). The fastest known globally convergent first-order method for minimizing strongly convex functions. *IEEE Control Systems Letters*, 2(1), 49–54.

Wang, D., He, Y., and Sterck, H.D. (2021). On the asymptotic linear convergence speed of Anderson acceleration applied to ADMM. *Journal of Scientific Computing*, 88(2), 38.