



Time Rescaling of a Primal-Dual Dynamical System with Asymptotically Vanishing Damping

David Alexander Hulett¹ · Dang-Khoa Nguyen¹

Accepted: 30 March 2023 / Published online: 31 May 2023
© The Author(s) 2023

Abstract

In this work, we approach the minimization of a continuously differentiable convex function under linear equality constraints by a second-order dynamical system with an asymptotically vanishing damping term. The system under consideration is a time rescaled version of another system previously found in the literature. We show fast convergence of the primal-dual gap, the feasibility measure, and the objective function value along the generated trajectories. These convergence rates now depend on the rescaling parameter, and thus can be improved by choosing said parameter appropriately. When the objective function has a Lipschitz continuous gradient, we show that the primal-dual trajectory asymptotically converges weakly to a primal-dual optimal solution to the underlying minimization problem. We also exhibit improved rates of convergence of the gradient along the primal trajectories and of the adjoint of the corresponding linear operator along the dual trajectories. We illustrate the theoretical outcomes and also carry out a comparison with other classes of dynamical systems through numerical experiments.

Keywords Augmented Lagrangian method · Primal-dual dynamical system · Damped inertial dynamics · Nesterov's accelerated gradient method · Lyapunov analysis · Time rescaling · Convergence rate · Trajectory convergence

Mathematics Subject Classification 37N40 · 46N10 · 65K10 · 90C25

✉ Dang-Khoa Nguyen
dang-khoa.nguyen@univie.ac.at

David Alexander Hulett
david.alexander.hulett@univie.ac.at

¹ Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

1 Introduction

1.1 Problem Statement and Motivation

In this paper we will consider the optimization problem

$$\begin{aligned} \min \quad & f(x), \\ \text{subject to } & Ax = b \end{aligned} \quad (1.1)$$

where

$$\left\{ \begin{array}{l} \mathcal{X}, \mathcal{Y} \text{ are real Hilbert spaces;} \\ f: \mathcal{X} \rightarrow \mathbb{R} \text{ is a continuously differentiable convex function;} \\ A: \mathcal{X} \rightarrow \mathcal{Y} \text{ is a continuous linear operator and } b \in \mathcal{Y}; \\ \text{the set } \mathbb{S} \text{ of primal-dual optimal solutions of (1.1) is assumed to be nonempty.} \end{array} \right. \quad (1.2)$$

This model formulation underlies many important applications in various areas, such as image recovery [25], machine learning [20, 31], the energy dispatch of power grids [42, 43], distributed optimization [32, 44] and network optimization [40, 45].

In recent years, there has been a flurry of research on the relationship between continuous time dynamical systems and the numerical algorithms that arise from their discretizations. For the unconstrained optimization problem, it has been known that inertial systems with damped velocities enjoy good convergence properties. For a convex, smooth function $f: \mathcal{X} \rightarrow \mathbb{R}$, Polyak is the first to consider the *heavy ball with friction* (HBF) dynamics [37, 38]

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0. \quad (\text{HBF})$$

Alvarez and Attouch continue the line of this study, focusing on inertial dynamics with a fixed viscous damping coefficient [2–4]. Later on, Cabot et al. [21, 22] consider the system that replaces γ with a time dependent damping coefficient $\gamma(t)$. In [41], Su, Boyd, and Candès showed that it turns out one can achieve fast convergence rates by introducing a time dependent damping coefficient which vanishes in a controlled manner, neither too fast nor too slowly, as t goes to infinity

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0. \quad (\text{AVD})$$

For $\alpha \geq 3$, the authors showed that a solution $x: [t_0, +\infty) \rightarrow \mathcal{X}$ to (AVD) satisfies $f(x(t)) - f(x_*) = \mathcal{O}\left(\frac{1}{t^2}\right)$ as $t \rightarrow +\infty$. In fact, the choice $\alpha = 3$ provides a continuous limit counterpart to Nesterov's celebrated accelerated gradient algorithm [15, 34, 35]. Weak convergence of the trajectories to minimizers of f when $\alpha > 3$ has been shown by Attouch et al. in [6] and May in [33], together with the improved rates of convergence $f(x(t)) - f(x_*) = o\left(\frac{1}{t^2}\right)$ as $t \rightarrow +\infty$. In the meantime, similar

results for the discrete counterpart were also reported by Chambolle and Dossal in [23], and by Attouch and Peyrouquet in [13].

In [7], Attouch, Chbani, and Riahi proposed an inertial proximal type algorithm, which results from a discretization of the time rescaled (AVD) system

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \delta(t) \nabla f(x(t)) = 0,$$

where $\delta : [t_0, +\infty) \rightarrow \mathbb{R}_+$ is a time scaling function satisfying a certain growth condition, which enters the convergence statement by way of $f(x(t)) - f(x_*) = \mathcal{O}\left(\frac{1}{t^2\delta(t)}\right)$ as $t \rightarrow +\infty$. The resulting algorithm obtained by the authors is considerably simpler than the founding proximal point algorithm proposed by Güler in [26], while providing comparable convergence rates for the functional values.

In order to approach constrained optimization problems, Augmented Lagrangian Method (ALM) [39] (for linearly constrained problems) and Alternating Direction Method of Multipliers (ADMM) [20, 24] (for problems with separable objectives and block variables linearly coupled in the constraints) and some of their variants have been shown to enjoy substantial success. Continuous-time approaches for structured convex minimization problems formulated in the spirit of the full splitting paradigm have been recently addressed in [18] and, closely connected to our approach, in [10, 17, 27, 45], to which we will have a closer look in Subsection 2.2. The temporal discretization resulting from these dynamics gives rise to the numerical algorithm with fast convergence rates [28, 29] and with a convergence guarantee for the generated iterate [19], without additional assumptions such as strong convexity.

In this paper, we will investigate a second-order dynamical system with asymptotic vanishing damping and time rescaling term, which is associated with the optimization problem (1.1) and formulated in terms of its augmented Lagrangian. The case when the time rescaling term does not appear has been established in [17]. We show that by introducing this time rescaling function, we are able to derive faster convergence rates for the primal-dual gap, the feasibility measure, and the objective function value along the generated trajectories while still maintaining the asymptotic behaviour of the trajectories towards a primal-dual optimal solution. On the other hand, this work can also be viewed as an extension of the time rescaling technique derived in [7, 9] for the constrained case. To our knowledge, the trajectory convergence for dynamics with time scaling seems to be new in the constrained case.

1.2 Notations and a Preliminary Result

For both Hilbert spaces \mathcal{X} and \mathcal{Y} , the Euclidean inner product and the associated norm will be denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. The Cartesian product $\mathcal{X} \times \mathcal{Y}$ will be endowed with the inner product and the associated norm defined for $(x, \lambda), (z, \mu) \in \mathcal{X} \times \mathcal{Y}$ as

$$\langle (x, \lambda), (z, \mu) \rangle = \langle x, z \rangle + \langle \lambda, \mu \rangle \quad \text{and} \quad \|(x, \lambda)\| = \sqrt{\|x\|^2 + \|\lambda\|^2},$$

respectively.

Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be a continuously differentiable convex function such that ∇f is ℓ -Lipschitz continuous. For every $x, y \in \mathcal{X}$ it holds (see [35, Theorem 2.1.5])

$$0 \leq \frac{1}{2\ell} \|\nabla f(x) - \nabla f(y)\|^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{\ell}{2} \|x - y\|^2. \tag{1.3}$$

2 The Primal-Dual Dynamical Approach

2.1 Augmented Lagrangian Formulation

Consider the saddle point problem

$$\min_{x \in \mathcal{X}} \max_{\lambda \in \mathcal{Y}} \mathcal{L}(x, \lambda) \tag{2.1}$$

associated to problem (1.1), where $\mathcal{L}: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ denotes the *Lagrangian* function

$$\mathcal{L}(x, \lambda) := f(x) + \langle \lambda, Ax - b \rangle.$$

Under the assumptions (1.2), \mathcal{L} is convex with respect to $x \in \mathcal{X}$ and affine with respect to $\lambda \in \mathcal{Y}$. A pair $(x_*, \lambda_*) \in \mathcal{X} \times \mathcal{Y}$ is said to be a *saddle point* of the Lagrangian function \mathcal{L} if for every $(x, \lambda) \in \mathcal{X} \times \mathcal{Y}$

$$\mathcal{L}(x_*, \lambda) \leq \mathcal{L}(x_*, \lambda_*) \leq \mathcal{L}(x, \lambda_*). \tag{2.2}$$

If $(x_*, \lambda_*) \in \mathcal{X} \times \mathcal{Y}$ is a saddle point of \mathcal{L} then $x_* \in \mathcal{X}$ is an optimal solution of (1.1), and $\lambda_* \in \mathcal{Y}$ is an optimal solution of its Lagrange dual problem. If $x_* \in \mathcal{X}$ is an optimal solution of (1.1) and a suitable constraint qualification is fulfilled, then there exists an optimal solution $\lambda_* \in \mathcal{Y}$ of the Lagrange dual problem such that $(x_*, \lambda_*) \in \mathcal{X} \times \mathcal{Y}$ is a saddle point of \mathcal{L} . For details and insights into the topic of constraint qualifications for convex duality we refer to [14, 16].

The set of saddle points of \mathcal{L} , called also primal-dual optimal solutions of (1.1), will be denoted by \mathbb{S} and, as stated in the assumptions, it will be assumed to be nonempty. The set of feasible points of (1.1) will be denoted by $\mathbb{F} := \{x \in \mathcal{X} : Ax = b\}$ and the optimal objective value of (1.1) by f_* .

The system of primal-dual optimality conditions for (1.1) reads

$$(x_*, \lambda_*) \in \mathbb{S} \Leftrightarrow \begin{cases} \nabla_x \mathcal{L}(x_*, \lambda_*) = 0 \\ \nabla_\lambda \mathcal{L}(x_*, \lambda_*) = 0 \end{cases} \Leftrightarrow \begin{cases} \nabla f(x_*) + A^* \lambda_* = 0 \\ Ax_* - b = 0 \end{cases}, \tag{2.3}$$

where $A^*: \mathcal{Y} \rightarrow \mathcal{X}$ denotes the adjoint operator of A .

For $\beta \geq 0$, we consider also the augmented Lagrangian $\mathcal{L}_\beta : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ associated with (1.1)

$$\mathcal{L}_\beta(x, \lambda) := \mathcal{L}(x, \lambda) + \frac{\beta}{2} \|Ax - b\|^2 = f(x) + \langle \lambda, Ax - b \rangle + \frac{\beta}{2} \|Ax - b\|^2. \tag{2.4}$$

For every $(x, \lambda) \in \mathbb{F} \times \mathcal{Y}$ it holds

$$f(x) = \mathcal{L}_\beta(x, \lambda) = \mathcal{L}(x, \lambda). \tag{2.5}$$

If $(x_*, \lambda_*) \in \mathbb{S}$, then we have for every $(x, \lambda) \in \mathcal{X} \times \mathcal{Y}$

$$\mathcal{L}(x_*, \lambda) = \mathcal{L}_\beta(x_*, \lambda) \leq \mathcal{L}(x_*, \lambda_*) = \mathcal{L}_\beta(x_*, \lambda_*) \leq \mathcal{L}(x, \lambda_*) \leq \mathcal{L}_\beta(x, \lambda_*).$$

In addition, from (2.3) we have

$$\begin{aligned} (x_*, \lambda_*) \in \mathbb{S} &\Leftrightarrow \begin{cases} \nabla f(x_*) + A^*\lambda_* = 0 \\ Ax_* - b = 0 \end{cases} \Leftrightarrow \begin{cases} \nabla f(x_*) + A^*\lambda_* + \beta A^*(Ax_* - b) = 0 \\ Ax_* - b = 0 \end{cases} \\ &\Leftrightarrow \begin{cases} \nabla_x \mathcal{L}_\beta(x_*, \lambda_*) = 0 \\ \nabla_\lambda \mathcal{L}_\beta(x_*, \lambda_*) = 0. \end{cases} \end{aligned}$$

In other words, for any $\beta \geq 0$ the sets of saddle points of \mathcal{L} and \mathcal{L}_β are identical.

2.2 The Primal-Dual Asymptotic Vanishing Damping Dynamical System with Time Rescaling

In this subsection we present the system under study, and we include a brief discussion regarding the existence and uniqueness of solutions.

The dynamical system which we associate to (1.1) and investigate in this paper reads

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \delta(t) \nabla_x \mathcal{L}_\beta(x(t), \lambda(t) + \theta t \dot{\lambda}(t)) = 0 \\ \ddot{\lambda}(t) + \frac{\alpha}{t} \dot{\lambda}(t) - \delta(t) \nabla_\lambda \mathcal{L}_\beta(x(t) + \theta t \dot{x}(t), \lambda(t)) = 0, \\ (x(t_0), \lambda(t_0)) = (x_0, \lambda_0) \text{ and } (\dot{x}(t_0), \dot{\lambda}(t_0)) = (\dot{x}_0, \dot{\lambda}_0) \end{cases} \tag{2.6}$$

where $t_0 > 0$, $\alpha > 0$, $\theta > 0$, $\delta : [t_0, +\infty) \rightarrow \mathbb{R}$ is a nonnegative continuously differentiable function and $(x_0, \lambda_0), (\dot{x}_0, \dot{\lambda}_0) \in \mathcal{X} \times \mathcal{Y}$ are the initial conditions. Replacing the expressions of the partial gradients of \mathcal{L}_β into the system leads to the following formulation for (2.6):

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \delta(t)\nabla f(x(t)) + \delta(t)A^*(\lambda(t) + \theta t\dot{\lambda}(t)) + \delta(t)\beta A^*(Ax(t) - b) & = 0 \\ \ddot{\lambda}(t) + \frac{\alpha}{t}\dot{\lambda}(t) - \delta(t)(A(x(t) + \theta t\dot{x}(t)) - b) & = 0 \\ (x(t_0), \lambda(t_0)) = (x_0, \lambda_0) \text{ and } (\dot{x}(t_0), \dot{\lambda}(t_0)) = (\dot{x}_0, \dot{\lambda}_0). \end{cases}$$

The case (2.6) in which there is no time rescaling, i.e., when $\delta(t) \equiv 1$, was studied by Zeng et al. in [45], and by Boř and Nguyen in [17]. The system with more general damping, extrapolation and time rescaling coefficients was addressed by He et al. in [27, 30] and by Attouch et al. in [10]. We mention that extending the results in this paper to the multi-block case is possible. For further details, we refer the readers to [17, Sect. 2.4].

It is well known that the viscous damping term $\frac{\alpha}{t}$ has a vital role in achieving fast convergence in unconstrained minimization [6, 8, 33]. The role of the extrapolation θt is to induce more flexibility in the dynamical system and in the associated discrete schemes, as it has been recently noticed in [10, 12, 27, 45]. The time scaling function $\delta(\cdot)$ has the role to further improve the rates of convergence of the objective function value along the trajectory, as it was noticed in the context of unconstrained minimization problems in [7, 9, 11] and of linearly constrained minimization problems in [10, 30].

It is straightforward to show the existence of local solutions to (2.6), under the additional assumption that ∇f is Lipschitz continuous on every bounded subset of \mathcal{X} . First, notice that (2.6) can be rewritten as a first-order dynamical system. Indeed, $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ is a solution to (2.6) if and only if $(x, \lambda, y, v) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y}$ is a solution to

$$\begin{cases} (\dot{x}(t), \dot{\lambda}(t), \dot{y}(t), \dot{v}(t)) & = F(t, x(t), \lambda(t), y(t), v(t)) \\ (x(t_0), \lambda(t_0), y(t_0), v(t_0)) & = (x_0, \lambda_0, \dot{x}_0, \dot{\lambda}_0), \end{cases}$$

where $F : [t_0, +\infty) \times \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y}$ is given by

$$\begin{aligned} &F(t, z, \mu, w, \eta) \\ &:= \left(w, \eta, -\frac{\alpha}{t}w - \delta(t)[\nabla f(z) + A^*(\mu + \theta t\eta) + \beta A^*(Az - b)], \right. \\ &\quad \left. -\frac{\alpha}{t}\eta + \delta(t)[A(z + \theta tw) - b] \right). \end{aligned}$$

where F is evidently continuous in t , and $F(t, \cdot)$ is Lipschitz continuous on every bounded subset, provided that the same property holds for ∇f . We can then employ a theorem such as that by Cauchy-Lipschitz to obtain the existence of a unique solution to the previous system, and thus a unique solution to (2.6), defined on a maximal interval $[t_0, T_{\max})$. To go further and show the existence and uniqueness of a global solution (that is, $T_{\max} = +\infty$) we will need some energy estimates derived in the next section in a similar way as in [11, 17]. For this reason, the existence and uniqueness of a global solution is postponed to a later stage.

3 Faster Convergence Rates via Time Rescaling

In this section we will derive fast convergence rates for the primal-dual gap, the feasibility measure, and the objective function value along the trajectories generated by the dynamical system (2.6). We will make the following assumptions on the parameters α, θ, β and the function δ throughout this section.

Assumption 1 In (2.6), assume that $\delta : [t_0, +\infty) \rightarrow (0, +\infty)$ is continuously differentiable. Moreover, suppose that the parameters α, β, θ and the function δ satisfy

$$\alpha \geq 3, \quad \beta \geq 0, \quad \frac{1}{2} \geq \theta \geq \frac{1}{\alpha - 1} \quad \text{and} \quad \sup_{t \geq t_0} \frac{t\dot{\delta}(t)}{\delta(t)} \leq \frac{1 - 2\theta}{\theta}. \quad (3.1)$$

Besides the first three conditions that are known previously in [17], it is worth pointing out that we can deduce from the last one the following inequality for every $t \geq t_0$:

$$\frac{t\dot{\delta}(t)}{\delta(t)} \leq \frac{1 - 2\theta}{\theta} = \frac{1}{\theta} - 2 \leq \alpha - 3. \quad (3.2)$$

This gives a connection to the condition which appears in [7]. A few more comments regarding the function δ will come later, after the convergence rates statements.

3.1 The Energy Function

Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution of (2.6). Let $(x_*, \lambda_*) \in \mathbb{S}$ be fixed, we define the energy function $\mathcal{E} : [t_0, +\infty) \rightarrow \mathbb{R}$

$$\begin{aligned} \mathcal{E}(t) &:= \theta^2 t^2 \delta(t) (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) + \frac{1}{2} \|v(t)\|^2 \\ &\quad + \frac{\xi}{2} \|(x(t), \lambda(t)) - (x_*, \lambda_*)\|^2, \end{aligned} \quad (3.3)$$

where

$$v(t) := (x(t), \lambda(t)) - (x_*, \lambda_*) + \theta t (\dot{x}(t), \dot{\lambda}(t)), \quad (3.4)$$

$$\xi := \alpha\theta - \theta - 1 \geq 0. \quad (3.5)$$

Notice that, according to (2.4) and (2.5), we have for every $t \geq t_0$

$$\begin{aligned} \mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t)) &= \mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) + \frac{\beta}{2} \|Ax(t) - b\|^2 \\ &= \mathcal{L}(x(t), \lambda_*) - f_* + \frac{\beta}{2} \|Ax(t) - b\|^2 \\ &= f(x(t)) - f_* \end{aligned} \quad (3.6)$$

$$+ \langle \lambda_*, Ax(t) - b \rangle + \frac{\beta}{2} \|Ax(t) - b\|^2 \geq 0, \tag{3.7}$$

where f_* denotes the optimal objective value of (1.1). In addition, due to (3.7), we have

$$\mathcal{E}(t) \geq 0 \quad \forall t \geq t_0. \tag{3.8}$$

The construction of \mathcal{E} is inspired by [17]. However, one can notice that we only consider \mathcal{E} defined with respect to a fixed primal-dual solution $(x_*, \lambda_*) \in \mathbb{S}$ rather than a family of energy functions, each defined with respect to a point $(z, \mu) \in \mathbb{F} \times \mathcal{Y}$. This gives simpler proofs for some results when compared to those in [17].

Assumption 1 implies the nonnegativity of following quantity, which will appear many times in our analysis:

$$\sigma : [t_0, +\infty) \rightarrow \mathbb{R}_+, \quad \sigma(t) := \frac{1 - 2\theta}{\theta} \delta(t) - t \dot{\delta}(t). \tag{3.9}$$

The following lemma gives us the decreasing property of the energy function. As a consequence of this lemma, we obtain some integrability results which will be needed later. The proofs are postponed to the Appendix.

Lemma 3.1 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution of (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. For every $t \geq t_0$ it holds*

$$\begin{aligned} \frac{d}{dt} \mathcal{E}(t) &\leq -\theta^2 t \sigma(t) (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) - \frac{1}{2} \beta \theta t \delta(t) \|Ax(t) - b\|^2 \\ &\quad - \xi \theta t \|(\dot{x}(t), \dot{\lambda}(t))\|^2. \end{aligned}$$

Proof See ‘‘Proof of Lemma 3.1’’ in Appendix B. □

Theorem 3.2 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution of (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. The following statements are true*

(i) *it holds*

$$\int_{t_0}^{+\infty} t \sigma(t) [\mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t))] dt \leq \mathcal{E}(t_0) < +\infty, \tag{3.10}$$

$$\beta \int_{t_0}^{+\infty} t \delta(t) \|Ax(t) - b\|^2 dt \leq \frac{2\mathcal{E}(t_0)}{\theta} < +\infty, \tag{3.11}$$

$$\xi \int_{t_0}^{+\infty} t \|(\dot{x}(t), \dot{\lambda}(t))\|^2 dt \leq \frac{\mathcal{E}(t_0)}{\theta} < +\infty; \tag{3.12}$$

(ii) *if, in addition, $\alpha > 3$ and $\frac{1}{2} \geq \theta > \frac{1}{\alpha-1}$, then the trajectory $(x(t), \lambda(t))_{t \geq t_0}$ is bounded and the convergence rate of its velocity is*

$$\|(\dot{x}(t), \dot{\lambda}(t))\| = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty.$$

Proof See “Proof of Theorem 3.2” in Appendix B. □

3.2 Fast Convergence Rates for the Primal-Dual Gap, the Feasibility Measure and the Objective Function Value

The following are the main convergence rates results of the paper.

Theorem 3.3 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution of (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. The following statements are true*

(i) *for every $t \geq t_0$ it holds*

$$0 \leq \mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) \leq \frac{\mathcal{E}(t_0)}{\theta^2 t^2 \delta(t)}; \tag{3.13}$$

(ii) *for every $t \geq t_0$ it holds*

$$\|Ax(t) - b\| \leq \frac{2C_1}{t^2 \delta(t)}, \tag{3.14}$$

where

$$C_1 := \sup_{t \geq t_0} t \|\dot{\lambda}(t)\| + (\alpha - 1) \sup_{t \geq t_0} \|\lambda(t) - \lambda_*\| + t_0^2 \delta(t_0) \|Ax(t_0) - b\| + t_0 \|\dot{\lambda}(t_0)\|;$$

(iii) *for every $t \geq t_0$ it holds*

$$|f(x(t)) - f_*| \leq \left(\frac{\mathcal{E}(t_0)}{\theta^2} + 2C_1 \|\lambda_*\| \right) \frac{1}{t^2 \delta(t)}. \tag{3.15}$$

Proof (i) We have already established that \mathcal{E} is nonincreasing on $[t_0, +\infty)$. Therefore, from the expression for \mathcal{E} and relation (3.6) we deduce

$$\theta^2 t^2 \delta(t) [\mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t))] \leq \mathcal{E}(t_0) \quad \forall t \geq t_0, \tag{3.16}$$

and the first claim follows.

(ii) From the second line of (2.6), for every $t \geq t_0$ we have

$$t\ddot{\lambda}(t) + \alpha\dot{\lambda}(t) = t\delta(t) \left(A(x(t) + \theta t\dot{x}(t)) - b \right) = t\delta(t) (Ax(t) - b) + \theta t^2 \delta(t) A\dot{x}(t). \tag{3.17}$$

Fix $t \geq t_0$. On the one hand, integration by parts yields

$$\begin{aligned} \int_{t_0}^t (s\ddot{\lambda}(s) + \alpha\dot{\lambda}(s)) ds &= t\dot{\lambda}(t) - t_0\dot{\lambda}(t_0) - \int_{t_0}^t \dot{\lambda}(s) ds + \alpha \int_{t_0}^t \dot{\lambda}(s) ds \\ &= t\dot{\lambda}(t) - t_0\dot{\lambda}(t_0) + (\alpha - 1)(\lambda(t) - \lambda(t_0)). \end{aligned} \tag{3.18}$$

On the other hand, again integrating by parts leads to

$$\int_{t_0}^t s^2 \delta(s) A \dot{x}(s) ds = t^2 \delta(t) (Ax(t) - b) - t_0^2 \delta(t_0) (Ax(t_0) - b) - \int_{t_0}^t (2s \delta(s) + s^2 \dot{\delta}(s)) (Ax(s) - b) ds. \tag{3.19}$$

Now, integrating (3.17) from t_0 to t and using (3.18) and (3.19) gives us

$$\begin{aligned} & t \dot{\lambda}(t) - t_0 \dot{\lambda}(t_0) + (\alpha - 1) (\lambda(t) - \lambda(t_0)) \\ &= \int_{t_0}^t s \delta(s) (Ax(s) - b) ds + \theta \int_{t_0}^t s^2 \delta(s) A \dot{x}(s) ds \\ &= t^2 \delta(t) (Ax(t) - b) - t_0^2 \delta(t_0) (Ax(t_0) - b) \\ &\quad + \int_{t_0}^t s [(1 - 2\theta) \delta(s) - \theta s \dot{\delta}(s)] (Ax(s) - b) ds \\ &= t^2 \delta(t) (Ax(t) - b) - t_0^2 \delta(t_0) (Ax(t_0) - b) \\ &\quad + \int_{t_0}^t \frac{(1 - 2\theta) \delta(s) - \theta s \dot{\delta}(s)}{s \delta(s)} s^2 \delta(s) (Ax(s) - b) ds. \end{aligned} \tag{3.20}$$

It follows that, for every $t \geq t_0$, we have

$$\left\| t^2 \delta(t) (Ax(t) - b) + \int_{t_0}^t \frac{(1 - 2\theta) \delta(s) - \theta s \dot{\delta}(s)}{s \delta(s)} s^2 \delta(s) (Ax(s) - b) ds \right\| \leq C_1, \tag{3.21}$$

where

$$\begin{aligned} C_1 &= \sup_{t \geq t_0} t \|\dot{\lambda}(t)\| + (\alpha - 1) \sup_{t \geq t_0} \|\lambda(t) - \lambda(t_0)\| + t_0^2 \delta(t_0) \|Ax(t_0) - b\| \\ &\quad + t_0 \|\dot{\lambda}(t_0)\| < +\infty, \end{aligned}$$

and this quantity is finite in light of (B.7) and (B.5). Now, we set

$$g(t) := t^2 \delta(t) \|Ax(t) - b\|, \quad a(t) := \frac{(1 - 2\theta) \delta(t) - \theta t \dot{\delta}(t)}{t \delta(t)} \quad \forall t \geq t_0$$

and we apply Lemma A.1 to deduce that

$$t^2 \delta(t) \|Ax(t) - b\| \leq 2C_1 \quad \forall t \geq t_0. \tag{3.22}$$

(iii) For a fixed $t \geq t_0$, we have

$$\mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) = f(x(t)) - f(x_*) + \langle \lambda_*, Ax(t) - b \rangle.$$

Therefore, from using (3.22) and (3.16) we obtain, for every $t \geq t_0$,

$$|f(x(t)) - f_*| \leq \mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) + \|\lambda_*\| \|Ax(t) - b\| \leq \left(\frac{\mathcal{E}(t_0)}{\theta^2} + 2C_1 \|\lambda_*\| \right) \frac{1}{t^2 \delta(t)},$$

which leads to the last statement. □

Some comments regarding the previous proof and results are in order.

Remark 3.4 The proof we provided here is significantly shorter than the one derived in [17] thanks to Lemma A.1. This Lemma is the one used in [28] for showing the fast convergence to zero of the feasibility measure, although the authors study a different dynamical system. On the other hand, when $\delta(t) \equiv 1$, the result in [17] is more robust than the one we obtain here, as it gives the $\mathcal{O}\left(\frac{1}{t^2}\right)$ rates for the sum of primal-dual gap and feasibility measure, rather than each one individually. It also allows us to focus only on the energy function defined with respect to a primal-dual optimal solution $(x_*, \lambda_*) \in \mathbb{S}$, rather than on an arbitrary feasible point $(z, \mu) \in \mathbb{F} \times \mathcal{Y}$ as in [17].

Remark 3.5 Here are some remarks comparing our rates of convergence to those in [10, 30].

- *Primal-dual gap*: According to (3.13), the following rate of convergence for the primal-dual is exhibited:

$$\mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) = \mathcal{O}\left(\frac{1}{t^2 \delta(t)}\right) \text{ as } t \rightarrow +\infty,$$

which coincides with the findings of [10, 30].

- *Feasibility measure*: According to (3.14), we have

$$\|Ax(t) - b\| = \mathcal{O}\left(\frac{1}{t^2 \delta(t)}\right) \text{ as } t \rightarrow +\infty,$$

which improves the rate $\mathcal{O}\left(\frac{1}{t\sqrt{\delta(t)}}\right)$ reported in [10, 30].

- *Functional values*: Relation (3.15) tells us that

$$|f(x(t)) - f_*| = \mathcal{O}\left(\frac{1}{t^2 \delta(t)}\right) \text{ as } t \rightarrow +\infty.$$

In [10], only the upper bound presents this order of convergence. The lower bound obtained is of order $\mathcal{O}\left(\frac{1}{t\sqrt{\delta(t)}}\right)$ as $t \rightarrow +\infty$. In [30], there are no comments on the rate attained by the functional values in the case of a general time rescaling parameter.

- Further comparisons with [30]:* in [30, Theorem 2.16], unlike the preceding result [30, Theorem 2.15], the authors produce a rate of $\mathcal{O}\left(\frac{1}{t^{1/\theta}}\right)$ as $t \rightarrow +\infty$ for $\|Ax(t) - b\|$ and $|f(x(t)) - f(x_*)|$, provided the time rescaling parameter is chosen to be $\delta(t) = \delta_0 t^{\frac{1}{\theta}-2}$, for some $\delta_0 > 0$. This choice comes from the solution to the differential equation

$$\frac{t\dot{\delta}(t)}{\delta(t)} = \frac{1 - 2\theta}{\theta} \quad \forall t \geq t_0,$$

and thus is covered by our results when the growth condition (3.1) holds with equality. The rates are consequently

$$\mathcal{O}\left(\frac{1}{t^2 \cdot \delta_0 t^{\frac{1}{\theta}-2}}\right) = \mathcal{O}\left(\frac{1}{t^{\frac{1}{\theta}}}\right) \quad \text{as } t \rightarrow +\infty.$$

In this setting, if we wish to obtain fast convergence rates, we need to choose a small θ . In light of Assumption 1, where we have $\frac{1}{2} \geq \theta \geq \frac{1}{\alpha-1}$, this can be achieved by taking α large enough. Such behaviour can also be seen in [10] and in the unconstrained case [7, 11].

4 Weak Convergence of the Trajectory to a Primal-Dual Solution

In this section we will show that the solutions to (2.6) weakly converge to an element of \mathbb{S} . The fact that $\delta(t)$ enters the convergence rate statement suggests that one can benefit from this time rescaling function when it is at least nondecreasing on $[t_0, +\infty)$. We are, in fact, going to need this condition when showing trajectory convergence.

Assumption 2 In (2.6), assume that ∇f is ℓ -Lipschitz continuous for some $\ell > 0$ and that $\delta : [t_0, +\infty) \rightarrow (0, +\infty)$ is continuously differentiable and nondecreasing. Moreover, suppose that the parameters α, β, θ and the function δ satisfy

$$\alpha > 3, \quad \beta \geq 0, \quad \frac{1}{2} > \theta > \frac{1}{\alpha - 1}, \quad \sup_{t \geq t_0} \frac{t\dot{\delta}(t)}{\delta(t)} < \frac{1 - 2\theta}{\theta}.$$

Assumption 2 entails the existence of $C_2 > 0$ such that

$$\frac{t\dot{\delta}(t)}{\delta(t)} + C_2 \leq \frac{1 - 2\theta}{\theta} \quad \forall t \geq t_0. \tag{4.1}$$

and therefore it follows further from the nondecreasing property of δ that

$$0 < C_2\delta(t_0) \leq C_2\delta(t) \leq (1 - 2\theta)\delta(t) - \theta t\dot{\delta}(t) \quad \forall t \geq t_0. \tag{4.2}$$

Moreover, from (4.1), for every $t \geq t_0$, we have

$$0 < C_2 \leq \frac{1 - 2\theta}{\theta} - \frac{t\dot{\delta}(t)}{\delta(t)} = \frac{\sigma(t)}{\delta(t)},$$

which gives

$$\delta(t) \leq \frac{\sigma(t)}{C_2} \quad \forall t \geq t_0. \tag{4.3}$$

We can now formally state the existence and uniqueness result of the trajectory. The proof follows the same argument as in [17, Theorem 4.1], therefore we omit the details.

Theorem 4.1 *In the setting of Assumption 2, for every choice of initial conditions*

$$x(t_0) = x_0, \quad \lambda(t_0) = \lambda_0, \quad \dot{x}(t_0) = \dot{x}_0, \quad \text{and} \quad \dot{\lambda}(t_0) = \dot{\lambda}_0$$

the system (2.6) has a unique global twice continuously differentiable solution $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$.

The additional Lipschitz continuity condition of ∇f and the fact that δ is nondecreasing give rise to the following two essential integrability statements.

Proposition 4.2 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution of (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. Then it holds*

$$\int_{t_0}^{+\infty} t\delta(t) \|\nabla f(x(t)) - \nabla f(x_*)\|^2 dt < +\infty \tag{4.4}$$

and

$$\int_{t_0}^{+\infty} t\delta(t) \|Ax(t) - b\|^2 dt < +\infty. \tag{4.5}$$

Proof See ‘‘Proof of Proposition 4.2’’ in Appendix B. □

Now, for a given primal-dual solution $(x_*, \lambda_*) \in \mathbb{S}$, we define the following mappings on $[t_0, +\infty)$

$$W(t) := \delta(t) [\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))] + \frac{1}{2} \|(\dot{x}(t), \dot{\lambda}(t))\|^2 \geq 0, \tag{4.6}$$

$$\varphi(t) := \frac{1}{2} \|(x(t), \lambda(t)) - (x_*, \lambda_*)\|^2 \geq 0. \tag{4.7}$$

The following are three technical lemmas that we will need in this section. Lemma 4.4 guarantess that the first condition of Opial’s Lemma is met.

Lemma 4.3 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ a solution of (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. The following inequality holds for every $t \geq t_0$:*

$$\ddot{\varphi}(t) + \frac{\alpha}{t} \dot{\varphi}(t) + \theta t \dot{W}(t) + \frac{\delta(t)}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 + \frac{\beta\delta(t)}{2} \|Ax(t) - b\|^2 \leq 0. \tag{4.8}$$

Proof See “Proof of Lemma 4.3” in Appendix B. □

Lemma 4.4 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution to (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. Then the positive part $[\dot{\phi}]_+$ of $\dot{\phi}$ belongs to $\mathbb{L}^1 [t_0, +\infty)$ and the limit $\lim_{t \rightarrow +\infty} \varphi(t)$ exists.*

Proof For any $t \geq t_0$, we multiply (4.8) by t and drop the last two norm squared terms to obtain

$$t\ddot{\varphi}(t) + \alpha\dot{\varphi}(t) + \theta t^2 \dot{W}(t) \leq 0.$$

Recall from (4.6) that for every $t \geq t_0$ we have

$$tW(t) = t\delta(t)[\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))] + \frac{t}{2} \|(\dot{x}(t), \dot{\lambda}(t))\|^2. \tag{4.9}$$

On the one hand, according to (3.12), the second summand of the previous expression belongs to $\mathbb{L}^1 [t_0, +\infty)$. On the other hand, using (4.3) and (3.10), we assert that

$$\begin{aligned} & \int_{t_0}^{+\infty} t\delta(t)[\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))]dt \\ & \leq \frac{1}{C_2} \int_{t_0}^{+\infty} t\sigma(t)[\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))]dt < +\infty. \end{aligned}$$

Hence, the first summand of (4.9) also belongs to $\mathbb{L}^1 [t_0, +\infty)$, which implies that the mapping $t \mapsto tW(t)$ belongs to $\mathbb{L}^1 [t_0, +\infty)$ as well. For achieving the desired conclusion, we make use of Lemma A.4 with $\phi := \varphi$ and $w := \theta W$. □

Lemma 4.5 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution to (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. The following inequality holds for every $t \geq t_0$*

$$\begin{aligned} & \frac{\alpha}{t\delta(t)} \frac{d}{dt} \|(\dot{x}(t), \dot{\lambda}(t))\|^2 + 2\left\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(\lambda(t) - \lambda_*) \right\rangle \\ & + \theta \frac{d}{dt} \left(t\delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \right) + ((1 - \theta)\delta(t) - \theta t\dot{\delta}(t)) \|A^*(\lambda(t) - \lambda_*)\|^2 \\ & \leq \delta(t) \left[2\|\nabla f(x(t)) - \nabla f(x_*)\|^2 + (2\beta^2\|A\|^2 + 1) \|Ax(t) - b\|^2 \right]. \end{aligned}$$

Proof See “Proof of Lemma 4.5” in Appendix B. □

Lemma 4.6 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution to (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. Then, for every $t \geq t_0$ it holds*

$$\begin{aligned} & \theta t^{\alpha+1} \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \\ & \leq -t^\alpha \dot{\varphi}(t) + \int_{t_0}^t s^\alpha V(s) ds \end{aligned}$$

$$\begin{aligned}
 & + \int_{t_0}^t s^\alpha \left[(\theta(\alpha + 1) - 1)\delta(s) + \theta s \dot{\delta}(s) \right] \|A^*(\lambda(s) - \lambda_*)\|^2 ds \\
 & - 2t^\alpha \langle \dot{x}(t), A^*(\lambda(t) - \lambda_*) \rangle + C_5,
 \end{aligned} \tag{4.10}$$

where, for $s \geq t_0$,

$$\begin{aligned}
 V(s) := & \theta(\alpha + 1)W(s) + \left(\frac{\alpha(\alpha - 1)}{t_0^2 \delta(t_0)} + \|A\| \right) \|(\dot{x}(s), \dot{\lambda}(s))\|^2 \\
 & + C_3 \delta(s) \|\nabla f(x(s)) - \nabla f(x_*)\|^2 + C_4 \delta(s) \|Ax(s) - b\|^2,
 \end{aligned}$$

for certain nonnegative constants C_3, C_4 and C_5 .

Proof See ‘‘Proof of Lemma 4.6’’ in Appendix B. □

The following proposition provides us with the main integrability result that will be used for verifying the second condition of Opial’s Lemma.

Proposition 4.7 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution to (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. Then it holds*

$$\int_{t_0}^{+\infty} t \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 dt < +\infty. \tag{4.11}$$

Proof We divide (4.10) by t^α , thus obtaining

$$\begin{aligned}
 \theta t \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \leq & -\dot{\varphi}(t) + \frac{1}{t^\alpha} \int_{t_0}^t s^\alpha V(s) ds \\
 & + \frac{1}{t^\alpha} \int_{t_0}^t s^\alpha \left[(\theta(\alpha + 1) - 1)\delta(s) + \theta s \dot{\delta}(s) \right] \\
 & \|A^*(\lambda(s) - \lambda_*)\|^2 ds \\
 & - 2 \langle \dot{x}(t), A^*(\lambda(t) - \lambda_*) \rangle + \frac{C_5}{t^\alpha}.
 \end{aligned}$$

Now, we integrate this inequality from t_0 to r . We get

$$\begin{aligned}
 & \theta \int_{t_0}^r t \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 dt \\
 & \leq \varphi(t_0) - \varphi(r) + \int_{t_0}^r \frac{1}{t^\alpha} \left(\int_{t_0}^t s^\alpha V(s) ds \right) dt \\
 & + \int_{t_0}^r \frac{1}{t^\alpha} \left(\int_{t_0}^t s^\alpha \left[(\theta(\alpha + 1) - 1)\delta(s) + \theta s \dot{\delta}(s) \right] \|A^*(\lambda(s) - \lambda_*)\|^2 ds \right) dt \\
 & - 2 \int_{t_0}^r \langle A\dot{x}(t), \lambda(t) - \lambda_* \rangle dt + C_5 \int_{t_0}^r t^\alpha dt.
 \end{aligned} \tag{4.12}$$

We now recall some important facts. First of all, we have

$$\int_{t_0}^r \frac{1}{t^\alpha} dt \leq \frac{1}{(\alpha - 1)t_0^{\alpha-1}}. \tag{4.13}$$

In addition, according to Lemma A.2, it holds

$$\int_{t_0}^r \frac{1}{t^\alpha} \left(\int_{t_0}^t s^\alpha V(s) ds \right) dt \leq \frac{1}{\alpha - 1} \int_{t_0}^r t V(t) dt, \tag{4.14}$$

and

$$\begin{aligned} & \int_{t_0}^r \frac{1}{t^\alpha} \left(\int_{t_0}^t s^\alpha [(\theta(\alpha + 1) - 1)\delta(s) + \theta s \dot{\delta}(s)] \|A^*(\lambda(s) - \lambda_*)\|^2 ds \right) dt \\ & \leq \frac{1}{\alpha - 1} \int_{t_0}^r [(\theta(\alpha + 1) - 1)\delta(t) + \theta t \dot{\delta}(t)] \|A^*(\lambda(t) - \lambda_*)\|^2 dt, \end{aligned} \tag{4.15}$$

respectively.

Finally, integrating by parts leads to

$$\begin{aligned} & - \int_{t_0}^r \langle A\dot{x}(t), \lambda(t) - \lambda_* \rangle dt \\ & = - \langle Ax(r) - b, \lambda(r) - \lambda_* \rangle + \langle Ax(t_0) - b, \lambda(t_0) - \lambda_* \rangle + \int_{t_0}^r \langle Ax(t) - b, \dot{\lambda}(t) \rangle dt \\ & \leq \|Ax(r) - b\| \|\lambda(r) - \lambda_*\| + \|Ax(t_0) - b\| \|\lambda(t_0) - \lambda_*\| + \int_{t_0}^r \langle Ax(t) - b, \dot{\lambda}(t) \rangle dt \\ & \leq \sup_{t \geq t_0} \{ \|Ax(t) - b\| \|\lambda(t) - \lambda_*\| \} + \|Ax(t_0) - b\| \|\lambda(t_0) - \lambda_*\| \\ & \quad + \frac{1}{2} \int_{t_0}^r (\|Ax(t) - b\|^2 + \|\dot{\lambda}(t)\|^2) dt. \end{aligned} \tag{4.16}$$

The supremum term is finite due to the boundedness of the trajectory. Now, by using the nonnegativity of φ and the facts (4.13), (4.14), (4.15) and (4.16) on (4.12), we come to

$$\begin{aligned} & \frac{\theta}{\alpha - 1} \int_{t_0}^r t \sigma(t) \|A^*(\lambda(t) - \lambda_*)\|^2 dt \\ & = \int_{t_0}^r \left[\theta \delta(t) - \frac{(\theta(\alpha + 1) - 1)\delta(t) + \theta t \dot{\delta}(t)}{\alpha - 1} \right] t \|A^*(\lambda(t) - \lambda_*)\|^2 dt \\ & \leq \frac{1}{\alpha - 1} \int_{t_0}^r t V(t) dt + \int_{t_0}^r t (\|Ax(t) - b\|^2 + \|\dot{\lambda}(t)\|^2) dt + C_6, \end{aligned} \tag{4.17}$$

where

$$C_6 := \varphi(t_0) + 2 \sup_{t \geq t_0} \{ \|Ax(t) - b\| \|\lambda(t) - \lambda_*\| \} + 2 \|Ax(t_0) - b\| \|\lambda(t_0) - \lambda_*\| + \frac{C_5}{(\alpha - 1)t_0^{\alpha-1}}.$$

According to (3.11) and (3.12) in Theorem 3.2, as well as Lemma 4.4, we know that the mappings $t \mapsto tV(t)$ and $t \mapsto t(\|Ax(t) - b\|^2 + \|\dot{\lambda}(t)\|)$ belong to $\mathbb{L}^1[t_0, +\infty)$. Therefore, by taking the limit as $r \rightarrow +\infty$ in (4.17) we obtain

$$\int_{t_0}^{+\infty} t\sigma(t) \|A^*(\lambda(t) - \lambda_*)\|^2 dt < +\infty.$$

Again, from (4.3) we conclude that

$$\int_{t_0}^{+\infty} t\delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 dt \leq \frac{1}{C_2} \int_{t_0}^{+\infty} t\sigma(t) \|A^*(\lambda(t) - \lambda_*)\|^2 dt < +\infty,$$

which completes the proof. □

The following result is the final step towards the second condition of Opial’s Lemma.

Theorem 4.8 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution to (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. Then it holds*

$$\begin{aligned} \|\nabla f(x(t)) - \nabla f(x_*)\| &= o\left(\frac{1}{\sqrt{t}\sqrt[4]{\delta(t)}}\right) \text{ and} \\ \|A^*(\lambda(t) - \lambda_*)\| &= o\left(\frac{1}{\sqrt{t}\sqrt[4]{\delta(t)}}\right) \text{ as } t \rightarrow +\infty. \end{aligned} \tag{4.18}$$

Consequently,

$$\|\nabla_x \mathcal{L}(x(t), \lambda(t))\| = \|\nabla f(x(t)) + A^*\lambda(t)\| = o\left(\frac{1}{\sqrt{t}\sqrt[4]{\delta(t)}}\right) \text{ as } t \rightarrow +\infty,$$

while, as seen earlier,

$$\|\nabla_\lambda \mathcal{L}(x(t), \lambda(t))\| = \|Ax(t) - b\| = \mathcal{O}\left(\frac{1}{t^2\delta(t)}\right) \text{ as } t \rightarrow +\infty.$$

Proof We first show the gradient rate. For $t \geq t_0$, it holds

$$\begin{aligned} &\frac{d}{dt} \left(t\sqrt{\delta(t)} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 \right) \\ &= \left(\sqrt{\delta(t)} + \frac{t\dot{\delta}(t)}{2\sqrt{\delta(t)}} \right) \|\nabla f(x(t)) - \nabla f(x_*)\|^2 \end{aligned}$$

$$+ 2t\sqrt{\delta(t)} \left\langle \nabla f(x(t)) - \nabla f(x_*), \frac{d}{dt} \nabla f(x(t)) \right\rangle. \tag{4.19}$$

On the one hand, by Assumption 2, we can write

$$\begin{aligned} \left(\sqrt{\delta(t)} + \frac{t\dot{\delta}(t)}{2\sqrt{\delta(t)}} \right) &= \left(\sqrt{\delta(t)} + \frac{\sqrt{\delta(t)}}{2} \cdot \frac{t\dot{\delta}(t)}{\delta(t)} \right) \leq \left(1 + \frac{1-2\theta}{2\theta} \right) \sqrt{\delta(t)} \\ &= \frac{1}{2\theta} \sqrt{\delta(t)}. \end{aligned} \tag{4.20}$$

Since δ is nondecreasing, for $t \geq t_0$ we have $\sqrt{\delta(t)} \geq \sqrt{\delta(t_0)} > 0$. Set $t_1 := \max \left\{ t_0, \frac{1}{\sqrt{\delta_0}} \right\}$. Therefore, for $t \geq t_1$ it holds

$$\frac{1}{\sqrt{\delta(t)}} \leq \frac{1}{\sqrt{\delta(t_0)}} = t_1 \leq t$$

and thus

$$\sqrt{\delta(t)} \leq t\delta(t). \tag{4.21}$$

On the other hand, for every $t \geq t_1$ we deduce

$$\begin{aligned} &2t\sqrt{\delta(t)} \left\langle \nabla f(x(t)) - \nabla f(x_*), \frac{d}{dt} \nabla f(x(t)) \right\rangle \\ &= 2t \left\langle \sqrt{\delta(t)} [\nabla f(x(t)) - \nabla f(x_*)], \frac{d}{dt} \nabla f(x(t)) \right\rangle \\ &\leq t\delta(t) \|\nabla f(x(t)) - \nabla f(x_*)\|^2 + t \left\| \frac{d}{dt} \nabla f(x(t)) \right\|^2 \\ &\leq t\delta(t) \|\nabla f(x(t)) - \nabla f(x_*)\|^2 + \ell^2 t \|\dot{x}(t)\|^2, \end{aligned} \tag{4.22}$$

where the last inequality is a consequence of the ℓ -Lipschitz continuity of ∇f . By combining (4.20), (4.21) and (4.22), from (4.19) we assert that for every $t \geq t_1$

$$\begin{aligned} &\frac{d}{dt} \left(t\sqrt{\delta(t)} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 \right) \\ &\leq \left(1 + \frac{1}{2\theta} \right) t\delta(t) \|\nabla f(x(t)) - \nabla f(x_*)\|^2 + \ell^2 t \|\dot{x}(t)\|^2. \end{aligned}$$

The right hand side of the previous inequality belongs to $L^1[t_1, +\infty)$, according to (3.12) and (4.4). Since δ is nondecreasing, for every $t \geq t_1$ we have

$$\sqrt{\delta(t)} = \sqrt{\delta(t)} \cdot \frac{\sqrt{\delta(t)}}{\sqrt{\delta(t)}} \leq \frac{\delta(t)}{\sqrt{\delta(t_0)}}$$

and thus

$$\begin{aligned} & \int_{t_1}^{+\infty} t\sqrt{\delta(t)} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 dt \\ & \leq \frac{1}{\sqrt{\delta(t_0)}} \int_{t_1}^{+\infty} t\delta(t) \|\nabla f(x(t)) - \nabla f(x_*)\|^2 dt < +\infty. \end{aligned} \tag{4.23}$$

This means that the function being differentiated also belongs to $\mathbb{L}^1[t_1, +\infty)$. Therefore, Lemma A.3 gives us

$$t\sqrt{\delta(t)} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 \rightarrow 0 \text{ as } t \rightarrow +\infty.$$

Proceeding in the exact same way, for every $t \geq t_1$ we have

$$\begin{aligned} & \frac{d}{dt} \left(t\sqrt{\delta(t)} \|A^*(\lambda(t) - \lambda_*)\|^2 \right) \\ & = \left(\sqrt{\delta(t)} + \frac{t\dot{\delta}(t)}{2\sqrt{\delta(t)}} \right) \|A^*(\lambda(t) - \lambda_*)\|^2 + 2t\sqrt{\delta(t)} \langle A^*(\lambda(t) - \lambda_*), \dot{\lambda}(t) \rangle \\ & \leq \left(\frac{1}{2\theta} + \|A\|^2 \right) t\delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 + t \|\dot{\lambda}(t)\|^2. \end{aligned}$$

According to (3.12) and (4.11), the right hand side of the previous inequality belongs to $\mathbb{L}^1[t_1, +\infty)$. Arguing as in (4.23), we deduce that the function being differentiated also belongs to $\mathbb{L}^1[t_1, +\infty)$. Again applying Lemma A.3, we come to

$$t\sqrt{\delta(t)} \|A^*(\lambda(t) - \lambda_*)\|^2 \rightarrow 0 \text{ as } t \rightarrow +\infty.$$

Finally, recalling that $A^*\lambda_* = -\nabla f(x_*)$, we deduce from the triangle inequality that

$$\begin{aligned} \|\nabla_x \mathcal{L}(x(t), \lambda(t))\| & = \|\nabla f(x(t)) + A^*\lambda(t)\| \\ & \leq \|\nabla f(x(t)) - \nabla f(x_*)\| + \|A^*(\lambda(t) - \lambda_*)\| \\ & = o\left(\frac{1}{\sqrt{t}\sqrt[4]{\delta(t)}}\right) \text{ as } t \rightarrow +\infty, \end{aligned}$$

and the third claim follows. □

Remark 4.9 The previous theorem also has its own interest. It tells us that the time rescaling parameter also plays a role in accelerating the rates of convergence for $\|\nabla f(x(t)) - \nabla f(x_*)\|$ and $\|A^*(\lambda(t) - \lambda_*)\|$ as $t \rightarrow +\infty$. Moreover, we deduce from (4.18) that the mapping $(x, \lambda) \mapsto (\nabla f(x), A^*\lambda)$ is constant along \mathbb{S} , as reported in [17, Proposition A.4].

We now come to the final step and show weak convergence of the trajectories of (2.6) to elements of \mathbb{S} .

Theorem 4.10 *Let $(x, \lambda) : [t_0, +\infty) \rightarrow \mathcal{X} \times \mathcal{Y}$ be a solution to (2.6) and $(x_*, \lambda_*) \in \mathbb{S}$. Then $(x(t), \lambda(t))$ converges weakly to a primal-dual solution of (1.1) as $t \rightarrow +\infty$.*

Proof For proving this theorem, we make use of Opial’s Lemma (see Lemma A.5). Lemma 4.4 tells us that $\lim_{t \rightarrow +\infty} \|(x(t), \lambda(t)) - (x_*, \lambda_*)\|$ exists for every $(x_*, \lambda_*) \in \mathbb{S}$, which proves condition (i) of Opial’s Lemma. Now let $(\tilde{x}, \tilde{\lambda})$ be any weak sequential cluster point of $(x(t), \lambda(t))$ as $t \rightarrow +\infty$, which means there exists a strictly increasing sequence $(t_n)_{n \in \mathbb{N}} \subseteq [t_0, +\infty)$ such that

$$(x(t_n), \lambda(t_n)) \rightharpoonup (\tilde{x}, \tilde{\lambda}) \text{ as } n \rightarrow +\infty.$$

We want to show the remaining condition of Opial’s Lemma, which asks us to check that $(\tilde{x}, \tilde{\lambda}) \in \mathbb{S}$. In other words, we must show that

$$\mathcal{L}(\tilde{x}, \lambda) \leq \mathcal{L}(\tilde{x}, \tilde{\lambda}) \leq \mathcal{L}(x, \tilde{\lambda}) \quad \forall (x, \lambda) \in \mathcal{X} \times \mathcal{Y}. \tag{4.24}$$

Let $(x, \lambda) \in \mathcal{X} \times \mathcal{Y}$ and $(x_*, \lambda_*) \in \mathbb{S}$ be fixed. Notice that the functions

$$f(\cdot) + \langle \tilde{\lambda}, A(\cdot) - Ax \rangle : \mathcal{X} \rightarrow \mathbb{R} \quad \text{and} \quad \langle \cdot, b - Ax \rangle : \mathcal{Y} \rightarrow \mathbb{R}$$

are convex and continuous, therefore they are lower semicontinuous. According to a known result (see, for example, [14, Theorem 9.1]), they are also weakly lower semicontinuous. Therefore, we can derive that

$$\begin{aligned} \mathcal{L}(\tilde{x}, \tilde{\lambda}) - \mathcal{L}(x, \tilde{\lambda}) &= f(\tilde{x}) + \langle \tilde{\lambda}, A\tilde{x} - Ax \rangle - f(x) \\ &\leq \liminf_{n \rightarrow +\infty} \left[f(x(t_n)) + \langle \tilde{\lambda}, Ax_n - Ax \rangle \right] - f(x) \\ &= f(x_*) + \langle \tilde{\lambda}, b - Ax \rangle - f(x) \\ &\leq f(x_*) - f(x) + \liminf_{n \rightarrow +\infty} \langle A^* \lambda(t_n), x_* - x \rangle \\ &= f(x_*) - (f(x) + \langle \lambda_*, Ax - b \rangle) \\ &= \mathcal{L}(x_*, \lambda_*) - \mathcal{L}(x, \lambda_*) \leq 0, \end{aligned}$$

where in the second and third equalities we used the fact that, as $n \rightarrow +\infty$, we have $f(x(t_n)) \rightarrow f(x_*)$ and $Ax(t_n) \rightarrow b$ (Theorem 3.3), and $A^* \lambda_n \rightarrow A^* \lambda_*$ (Theorem 4.8). Similarly, the weak lower semicontinuity of the function $\langle \lambda - \tilde{\lambda}, A(\cdot) - b \rangle : \mathcal{X} \rightarrow \mathbb{R}$ yields

$$\mathcal{L}(\tilde{x}, \lambda) - \mathcal{L}(\tilde{x}, \tilde{\lambda}) = \langle \lambda - \tilde{\lambda}, A\tilde{x} - b \rangle \leq \liminf_{n \rightarrow +\infty} \langle \lambda - \tilde{\lambda}, Ax(t_n) - b \rangle = 0,$$

We have thus showed (4.24) and the proof is concluded. □

5 Numerical Experiments

We will illustrate the theoretical results by two numerical examples, with $\mathcal{X} = \mathbb{R}^4$ and $\mathcal{Y} = \mathbb{R}^2$. We will address two minimization problems with linear constraints; one with a strongly convex objective function and another with a convex objective function which is not strongly convex. In both cases, the linear constraints are dictated by

$$A = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Example 5.1 Consider the minimization problem

$$\begin{aligned} \min f(x_1, x_2, x_3, x_4) &:= (x_1 - 1)^2 + (x_2 - 1)^2 + x_3^2 + x_4^2 \\ \text{subject to } x_1 - x_2 - x_3 &= 0 \\ x_2 - x_4 &= 0. \end{aligned}$$

The optimality conditions can be calculated and lead to the following primal-dual solution pair

$$x_* = \begin{bmatrix} 0.8 \\ 0.6 \\ 0.2 \\ 0.6 \end{bmatrix} \quad \text{and} \quad \lambda_* = \begin{bmatrix} 0.4 \\ 1.2 \end{bmatrix}.$$

Example 5.2 Consider the minimization problem

$$\begin{aligned} \min f(x_1, x_2, x_3, x_4) &:= \log(1 + e^{-x_1 - x_2}) + x_3^2 + x_4^2 \\ \text{subject to } x_1 - x_2 - x_3 &= 0 \\ x_2 - x_4 &= 0. \end{aligned}$$

This problem is similar to the regularized logistic regression frequently used in machine learning. We cannot explicitly calculate the optimality conditions as in the previous case; instead, we use the last solution in the numerical experiment as the approximate solution.

To comply with Assumption 2, we choose $t_0 > 0$, $\alpha = 8$, $\beta = 10$, $\theta = \frac{1}{6}$, and we test four different choices for the rescaling parameter: $\delta(t) = 1$ (i.e., the (PD-AVD) dynamics in [17, 45]), $\delta(t) = t$, $\delta(t) = t^2$ and $\delta(t) = t^3$. In both examples, the initial conditions are

$$x(t_0) = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}, \quad \lambda(t_0) = \begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix}, \quad \dot{x}(t_0) = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} \quad \text{and} \quad \dot{\lambda}(t_0) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}.$$

For each choice of δ , we plot, using a logarithmic scale, the primal-dual gap $\mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t))$, the feasibility measure $\|Ax(t) - b\|$ and the functional values $|f(x(t)) - f_*|$, to highlight the theoretical result in Theorem 3.3.

We also illustrate the findings from Theorem 4.8, namely, we plot the quantities $\|\nabla f(x(t)) - \nabla f(x_*)\|$ and $\|A^*(\lambda(t) - \lambda_*)\|$, as well as the velocity $\|(\dot{x}(t), \dot{\lambda}(t))\|$.

Figures 1 and 2 display these plots for Examples 5.1 and 5.2, respectively. As predicted by the theory, choosing faster-growing time rescaling parameters yields better convergence rates. This is not the case for the velocities.

Next we use Example 5.2 to compare the convergence behaviour of our system (2.6) with the one where the asymptotically vanishing damping term is chosen to be $\frac{\alpha}{t^r}$, for $r \in [0, 1]$. Notice that $r = 1$ gives our system (2.6). When $r = 0$, in the setting of [30, Theorem 2.2], we know that the primal-dual gap exhibits a convergence rate of $\mathcal{O}\left(\frac{1}{t\delta(t)}\right)$ as $t \rightarrow +\infty$. This is illustrated in Fig. 3, where we plotted the combinations $(\delta(t) = t; r = 0)$, $(\delta(t) = t; r = 1)$, $(\delta(t) = t^2; r = 0)$, and $(\delta(t) = t^2; r = 1)$. In particular, observe that the rate predicted by [30, Theorem 2.2] for the primal-dual gap for the case $(\delta(t) = t^2; r = 0)$ reads $\mathcal{O}\left(\frac{1}{t \cdot t^2}\right)$, while the rate predicted by our Theorem 3.3 for the case $(\delta(t) = t; r = 1)$ reads $\mathcal{O}\left(\frac{1}{t^2-t}\right)$. It is no surprise then to see the combinations $(\delta(t) = t^2; r = 0)$ and $(\delta(t) = t; r = 1)$ exhibiting similar convergence behaviour in Fig. 3.

For better understanding, we run Example 5.2 once more to show the plots which result from fixing the time rescaling parameter $\delta(t) = t$ and varying $r \in \{0, 0.25, 0.5, 0.75, 1\}$. Notice how the convergence improves as r approaches 1. As $t \rightarrow +\infty$, [30, Theorem 2.7] predicts convergence rates of $\mathcal{O}\left(\frac{1}{t^\tau \delta(t)}\right)$ for the primal-dual gap and of $\mathcal{O}\left(\frac{1}{t^{\tau/2}}\right)$ for the velocities, which is reflected in our plots.

Acknowledgements The authors would like to thank the editor and the anonymous referees for their helpful comments and suggestions, which have led to the improvement of this paper.

Funding Open access funding provided by Austrian Science Fund (FWF). D.-K. Nguyen research supported by FWF (Austrian Science Fund), project P 34922-N.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Here we collect the auxiliary results that are required to carry out many steps in our analysis.

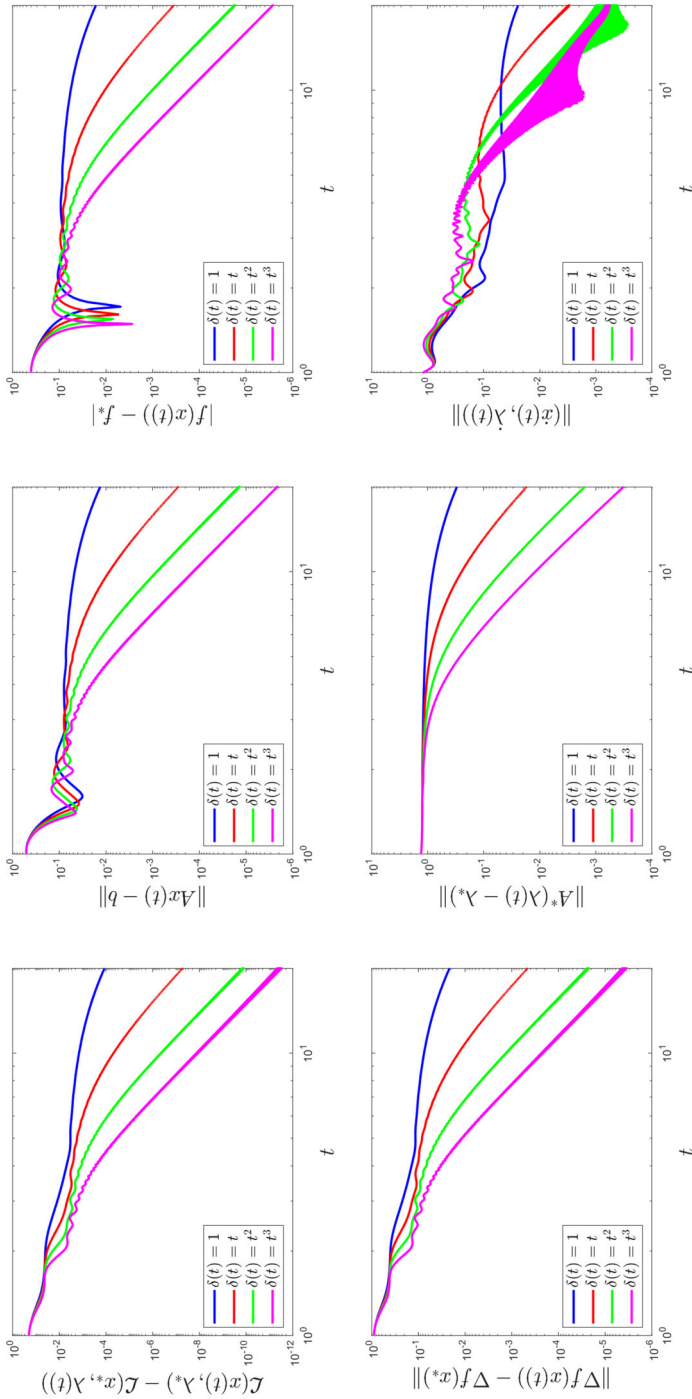


Fig. 1 The function $\delta(t)$ influences convergence behaviour in Example 5.1

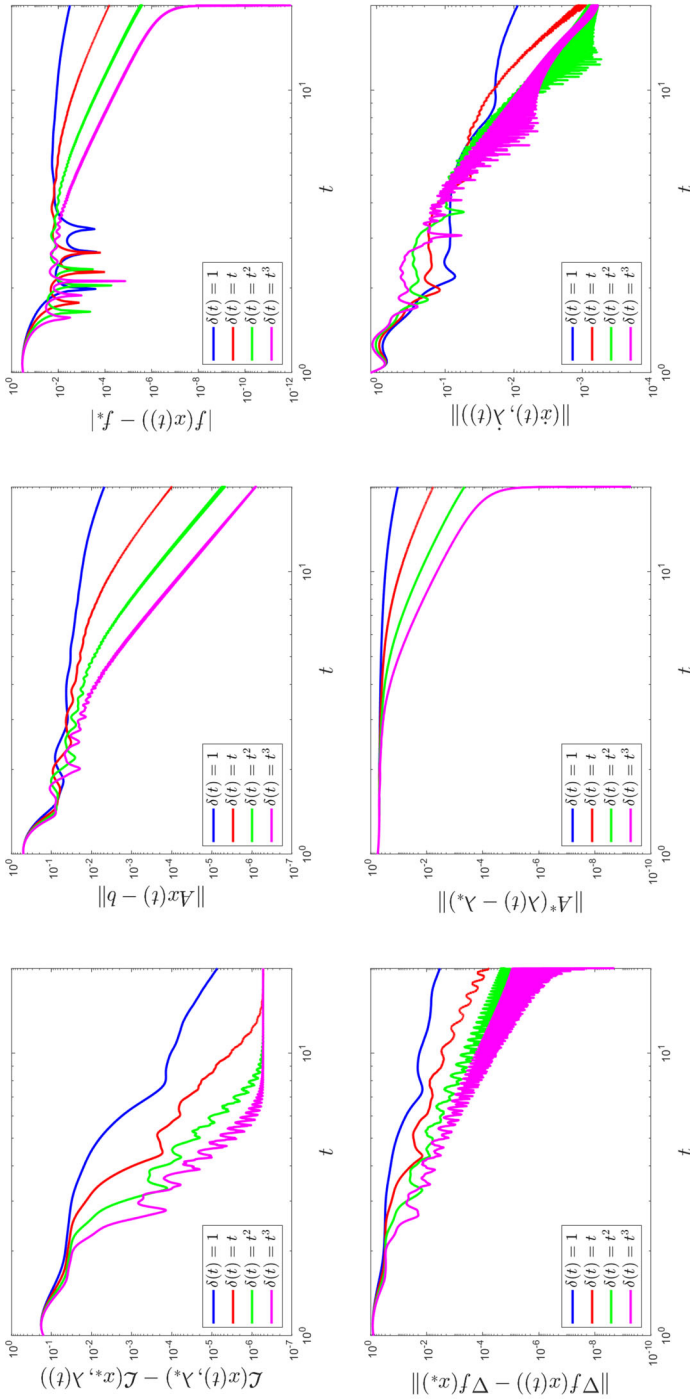


Fig. 2 The function $\delta(t)$ influences convergence behaviour in Example 5.2

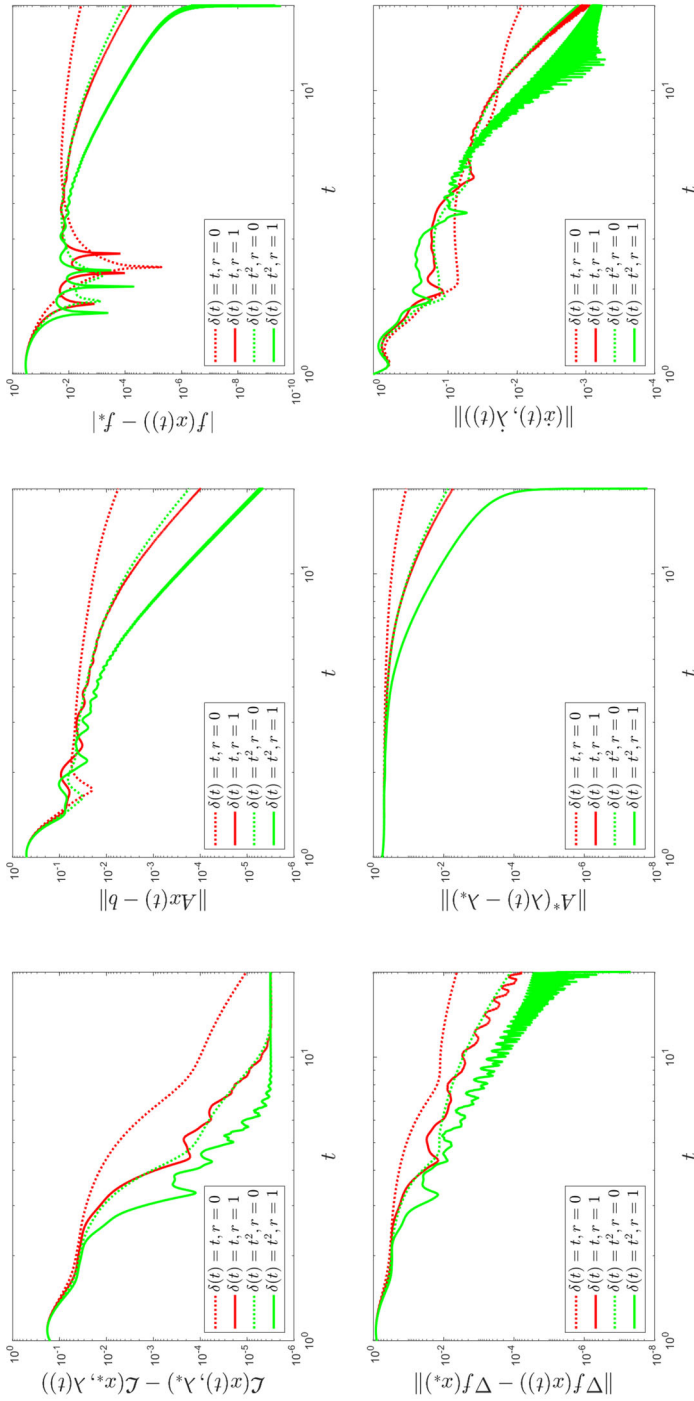


Fig. 3 The function $\delta(t)$, as well as the parameter r , influence convergence behaviour in Example 5.2

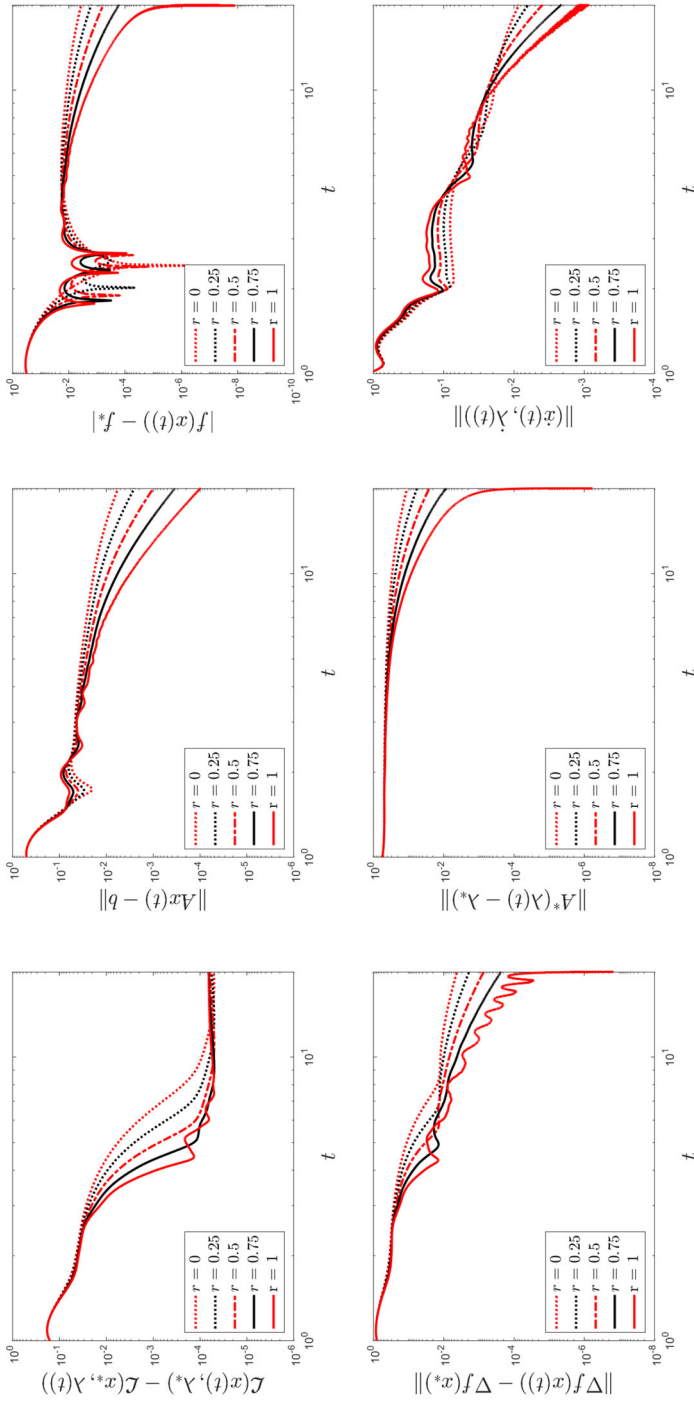


Fig. 4 The parameter r influences convergence behaviour in Example 5.2

A proof for the following lemma in the finite-dimensional case can be found in [28, Lemma 6]. The proof for the infinite-dimensional case is short and virtually identical, so we include it here for the sake of completeness.

Lemma A.1 *Assume that $t_0 > 0$, $g : [t_0, +\infty) \rightarrow \mathcal{Y}$ is a continuous differentiable function, $a : [t_0, +\infty) \rightarrow [0, +\infty)$ is a continuous function, and $C \geq 0$. If, in the sense of Bochner integrability, we have*

$$\left\| g(t) + \int_{t_0}^t a(s)g(s)ds \right\| \leq C \quad \forall t \geq t_0 \tag{A.1}$$

then

$$\sup_{t \geq t_0} \|g(t)\| \leq 2C < +\infty.$$

Proof Define, for every $t \geq t_0$,

$$G(t) := e^{\int_{t_0}^t a(s)ds} \int_{t_0}^t a(s)g(s)ds.$$

Fix $t \geq t_0$. The time derivative of G reads

$$\begin{aligned} \dot{G}(t) &= a(t)e^{\int_{t_0}^t a(s)ds} \int_{t_0}^t a(s)g(s)ds + e^{\int_{t_0}^t a(s)ds} a(t)g(t) \\ &= a(t)e^{\int_{t_0}^t a(s)ds} \left[g(t) + \int_{t_0}^t a(s)g(s)ds \right], \end{aligned}$$

so by using (A.1) and the previous equality we arrive at

$$\|\dot{G}(t)\| \leq Ca(t)e^{\int_{t_0}^t a(s)ds} = C \frac{d}{dt} \left(e^{\int_{t_0}^t a(s)ds} \right). \tag{A.2}$$

Since $G(t_0) = 0$, we have

$$G(t) = G(t) - G(t_0) = \int_{t_0}^t \dot{G}(s)ds,$$

so by employing (A.2) and the previous equality we obtain, for every $t \geq t_0$,

$$\begin{aligned} e^{\int_{t_0}^t a(s)ds} \left\| \int_{t_0}^t a(s)g(s)ds \right\| &= \|G(t)\| \leq \int_{t_0}^t \|\dot{G}(s)\|ds \leq C \int_{t_0}^t \frac{d}{ds} \left(e^{\int_{t_0}^s a(\tau)d\tau} \right) ds \\ &\leq C \left[e^{\int_{t_0}^t a(s)ds} - 1 \right] \leq Ce^{\int_{t_0}^t a(s)ds}. \end{aligned}$$

Dividing both sides of the previous inequality by $e^{\int_{t_0}^t a(s)ds}$ gives us

$$\left\| \int_{t_0}^t a(s)g(s)ds \right\| \leq C \quad \forall t \geq t_0. \tag{A.3}$$

Now, by putting (A.1) and (A.3) we finally come to

$$\|g(t)\| \leq \left\| g(t) + \int_{t_0}^t a(s)g(s)ds \right\| + \left\| \int_{t_0}^t a(s)g(s)ds \right\| \leq 2C \quad \forall t \geq t_0,$$

which leads to the announced statement. □

The proofs for the following results can be found in [17, Lemma A.1] and [1, Lemma 5.2], respectively.

Lemma A.2 *Let $0 < t_0 \leq r \leq +\infty$ and $h : [t_0, +\infty) \rightarrow [0, +\infty)$ be a continuous function. For every $\alpha > 1$ it holds*

$$\int_{t_0}^r \frac{1}{t^\alpha} \left[\int_{t_0}^t s^{\alpha-1} h(s) \right] dt \leq \frac{1}{\alpha - 1} \int_{t_0}^r h(t) dt.$$

If $r = +\infty$, then equality holds.

Lemma A.3 *Let $t_0 > 0$, $1 \leq p < +\infty$ and $1 \leq q \leq +\infty$. Suppose that $F \in \mathbb{L}^p [t_0, +\infty)$ is a locally absolutely continuous nonnegative function, $G \in \mathbb{L}^q [t_0, +\infty)$ and*

$$\dot{F}(t) \leq G(t) \quad \forall t \geq t_0.$$

Then, $\lim_{t \rightarrow +\infty} F(t) = 0$.

The following lemma is a slight variation of results already present in the literature. See, for example, [5, Lemma A.2].

Lemma A.4 *Let $t_0 > 0$, $\alpha > 1$, and let $\phi : [t_0, +\infty) \rightarrow \mathbb{R}$ be a twice continuously differentiable function bounded from below. Furthermore, assume $w : [t_0, +\infty) \rightarrow [0, +\infty)$ to be a continuously differentiable function such that $t \mapsto tw(t)$ belongs to $\mathbb{L}^1 [t_0, +\infty)$ and*

$$t\ddot{\phi}(t) + \alpha\dot{\phi}(t) + t^2\dot{w}(t) \leq 0 \quad \forall t \geq t_0.$$

Then, the positive part $[\phi]_+$ of ϕ belongs to $\mathbb{L}^1 [t_0, +\infty)$ and the limit $\lim_{t \rightarrow +\infty} \phi(t)$ is a real number.

Proof Fix $t \geq t_0$. Adding $(\alpha + 1)tw(t)$ to both sides of the previous inequality and then multiplying it by $t^{\alpha-1}$ yields

$$\frac{d}{dt}(t^\alpha \dot{\phi}(t)) + \frac{d}{dt}(t^{\alpha+1} w(t)) \leq (\alpha + 1)t^\alpha w(t).$$

Since the previous inequality holds for any $t \geq t_0$, we can integrate it from t_0 to $t \geq t_0$ to get

$$t^\alpha \dot{\phi}(t) - t_0^\alpha \dot{\phi}(t_0) + t^{\alpha+1} w(t) - t_0^{\alpha+1} w(t_0) \leq (\alpha + 1) \int_{t_0}^t s^\alpha w(s) ds.$$

After dropping the nonnegative term $t^{\alpha+1} w(t)$ and dividing by t^α we arrive at

$$\dot{\phi}(t) \leq \frac{\tilde{C}}{t^\alpha} + \frac{\alpha + 1}{t^\alpha} \int_{t_0}^t s^\alpha w(s) ds \quad \forall t \geq t_0,$$

where

$$\tilde{C} := t_0^\alpha |\dot{\phi}(t_0)| + t_0^{\alpha+1} w(t_0),$$

which further leads to

$$[\dot{\phi}(t)]_+ \leq \frac{\tilde{C}}{t^\alpha} + \frac{\alpha + 1}{t^\alpha} \int_{t_0}^t s^\alpha w(s) ds \quad \forall t \geq t_0.$$

Now, we integrate this inequality from t_0 to $r \geq t_0$ and we apply Lemma A.2 with $h : [t_0, +\infty) \rightarrow [0, +\infty)$ given by $h(s) := sw(s)$ to obtain

$$\begin{aligned} \int_{t_0}^r [\dot{\phi}(t)]_+ dt &\leq \tilde{C} \int_{t_0}^r \frac{1}{t^\alpha} dt + (\alpha + 1) \int_{t_0}^r \frac{1}{t^\alpha} \left[\int_{t_0}^t s^{\alpha-1} \cdot sw(s) ds \right] dt \\ &\leq \frac{\tilde{C}}{1 - \alpha} \left(\frac{1}{t_0^{\alpha-1}} - \frac{1}{r^{\alpha-1}} \right) + \frac{\alpha + 1}{1 - \alpha} \int_{t_0}^r tw(t) dt. \end{aligned}$$

By hypothesis, as $r \rightarrow +\infty$, the right hand side of the previous inequality is finite. In other words,

$$\int_{t_0}^{+\infty} [\dot{\phi}(t)]_+ dt < +\infty.$$

The previous statement, together with the fact that we assumed that ϕ was bounded from below, allow us to deduce that the function $\psi : [t_0, +\infty) \rightarrow \mathbb{R}$ given by

$$\psi(t) := \phi(t) - \int_{t_0}^t [\dot{\phi}(s)]_+ ds$$

is also bounded from below. An easy computation shows that ψ is nonpositive on $[t_0, +\infty)$, thus ψ is nonincreasing on $[t_0, +\infty)$. These facts imply that $\lim_{t \rightarrow +\infty} \psi(t)$ is a real number. Finally, we conclude that

$$\lim_{t \rightarrow +\infty} \phi(t) = \lim_{t \rightarrow +\infty} \psi(t) + \int_{t_0}^{+\infty} [\dot{\phi}(s)]_+ ds \in \mathbb{R}.$$

□

The proof for Opial’s Lemma can be found in [36].

Lemma A.5 (Opial’s Lemma) *Let \mathcal{H} be a real Hilbert space, $S \subseteq \mathcal{H}$ a nonempty set, $t_0 > 0$ and $z : [t_0, +\infty) \rightarrow \mathcal{H}$ a mapping that satisfies*

- (i) *for every $z_* \in S$, $\lim_{t \rightarrow +\infty} \|z(t) - z_*\|$ exists;*
- (ii) *every weak sequential cluster point of the trajectory $z(t)$ as $t \rightarrow +\infty$ belongs to S .*

Then, $z(t)$ converges weakly to an element of S as $t \rightarrow +\infty$.

Appendix B: Missing Proofs

Proof of Lemma 3.1

Let $t \geq t_0$ be fixed. Since $x_* \in \mathbb{F}$, we have

$$\begin{aligned} \nabla_x (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) &= \nabla_x \mathcal{L}_\beta(x(t), \lambda_*) \\ &= \nabla f(x(t)) + A^* \lambda_* + \beta A^*(Ax(t) - b), \\ \nabla_\lambda (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) &= -\nabla_\lambda \mathcal{L}_\beta(x_*, \lambda(t)) = 0. \end{aligned}$$

Under these expressions, the system (2.6) can be equivalently written as

$$\begin{aligned} (\ddot{x}(t), \ddot{\lambda}(t)) &= -\frac{\alpha}{t} (\dot{x}(t), \dot{\lambda}(t)) - \delta(t) (\nabla_x \mathcal{L}_\beta(x(t), \lambda_*), 0) \\ &\quad - \delta(t) (A^*(\lambda(t) - \lambda_* + \theta t \dot{\lambda}(t)), -(A(x(t) + \theta t \dot{x}(t)) - b)), \end{aligned}$$

which leads to

$$\begin{aligned} \dot{v}(t) &= (1 + \theta) (\dot{x}(t), \dot{\lambda}(t)) + \theta t (\ddot{x}(t), \ddot{\lambda}(t)) \\ &= -\xi (\dot{x}(t), \dot{\lambda}(t)) - \theta t \delta(t) (\nabla_x \mathcal{L}_\beta(x(t), \lambda_*), 0) \\ &\quad - \theta t \delta(t) (A^*(\lambda(t) - \lambda_* + \theta t \dot{\lambda}(t)), -(A(x(t) + \theta t \dot{x}(t)) - b)). \end{aligned}$$

We get from the distributive property of the inner product

$$\begin{aligned} \langle v(t), \dot{v}(t) \rangle &= -\xi \langle (x(t), \lambda(t)) - (x_*, \lambda_*), (\dot{x}(t), \dot{\lambda}(t)) \rangle - \xi \theta t \left\| (\dot{x}(t), \dot{\lambda}(t)) \right\|^2 \\ &\quad - \theta t \delta(t) \langle (\nabla_x \mathcal{L}_\beta(x(t), \lambda_*), 0), (x(t), \lambda(t)) - (x_*, \lambda_*) \rangle \\ &\quad - \theta^2 t^2 \delta(t) \langle (\nabla_x \mathcal{L}_\beta(x(t), \lambda_*), 0), (\dot{x}(t), \dot{\lambda}(t)) \rangle \\ &\quad - \theta t \delta(t) \langle \lambda(t) - \lambda_* + \theta t \dot{\lambda}(t), Ax(t) - Ax_* \rangle \\ &\quad - \theta^2 t^2 \delta(t) \langle \lambda(t) - \lambda_* + \theta t \dot{\lambda}(t), A\dot{x}(t) \rangle \end{aligned}$$

$$\begin{aligned}
 & + \theta t \delta(t) \langle A(x(t) + \theta t \dot{x}(t)) - b, \lambda(t) - \lambda_* \rangle \\
 & + \theta^2 t^2 \delta(t) \langle A(x(t) + \theta t \dot{x}(t)) - b, \dot{\lambda}(t) \rangle.
 \end{aligned}$$

Since $x_* \in \mathbb{F}$, the last four terms in the above identity vanish. Indeed,

$$\begin{aligned}
 & - \langle \lambda(t) - \lambda_* + \theta t \dot{\lambda}(t), Ax(t) - Ax_* \rangle - \theta t \langle \lambda(t) - \lambda_* + \theta t \dot{\lambda}(t), A\dot{x}(t) \rangle \\
 & + \langle A(x(t) + \theta t \dot{x}(t)) - b, \lambda(t) - \lambda_* \rangle + \theta t \langle A(x(t) + \theta t \dot{x}(t)) - b, \dot{\lambda}(t) \rangle \\
 & = - \langle \lambda(t) - \lambda_* + \theta t \dot{\lambda}(t), Ax(t) - b \rangle - \theta t \langle \lambda(t) - \lambda_* + \theta t \dot{\lambda}(t), A\dot{x}(t) \rangle \\
 & + \langle Ax(t) - b, \lambda(t) - \lambda_* \rangle + \theta t \langle A\dot{x}(t), \lambda(t) - \lambda_* \rangle \\
 & + \theta t \langle Ax(t) - b, \dot{\lambda}(t) \rangle + \theta^2 t^2 \langle A\dot{x}(t), \dot{\lambda}(t) \rangle \\
 & = 0.
 \end{aligned}$$

Therefore, differentiating \mathcal{E} with respect to t gives

$$\begin{aligned}
 \frac{d}{dt} \mathcal{E}(t) & = \theta^2 t (2\delta(t) + t\dot{\delta}(t)) (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) \\
 & + \theta^2 t^2 \delta(t) \langle (\nabla_x \mathcal{L}_\beta(x(t), \lambda_*), 0), (\dot{x}(t), \dot{\lambda}(t)) \rangle \\
 & + \langle v(t), \dot{v}(t) \rangle + \xi \langle (x(t), \lambda(t)) - (x_*, \lambda_*), (\dot{x}(t), \dot{\lambda}(t)) \rangle \\
 & = \theta^2 t (2\delta(t) + t\dot{\delta}(t)) (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) - \xi \theta t \|(\dot{x}(t), \dot{\lambda}(t))\|^2 \\
 & - \theta t \delta(t) \langle (\nabla_x \mathcal{L}_\beta(x(t), \lambda_*), 0), (x(t), \lambda(t)) - (x_*, \lambda_*) \rangle. \tag{B.1}
 \end{aligned}$$

Furthermore, the convexity of f and the fact that $x_* \in \mathbb{F}$ guarantee

$$\begin{aligned}
 & - \langle (\nabla_x \mathcal{L}_\beta(x(t), \lambda_*), 0), (x(t), \lambda(t)) - (x_*, \lambda_*) \rangle \\
 & = \langle \nabla f(x(t)), x_* - x(t) \rangle + \langle A^* \lambda_*, x_* - x(t) \rangle + \beta \langle A^*(Ax(t) - b), x_* - x(t) \rangle \\
 & \leq - (f(x(t)) - f(x_*)) - \langle \lambda_*, Ax(t) - b \rangle - \beta \|Ax(t) - b\|^2 \\
 & = - (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) - \frac{\beta}{2} \|Ax(t) - b\|^2, \tag{B.2}
 \end{aligned}$$

where we recall that the second equality comes from (2.5). By multiplying this inequality by $\theta t \delta(t)$ and combining it with (B.1), the coefficient attached to the primal-dual gap becomes

$$\theta^2 t (2\delta(t) + t\dot{\delta}(t)) - \theta t \delta(t) = -\theta^2 t \left(\frac{1 - 2\theta}{\theta} \delta(t) - t\dot{\delta}(t) \right) = -\theta^2 t \sigma(t),$$

which finally gives the desired statement. □

Proof of Theorem 3.2

(i) Recall that Assumption 1 implies $\sigma(t) \geq 0$ for all $t \geq t_0$ and $\xi \geq 0$. Moreover, $(x_*, \lambda_*) \in \mathbb{S}$ yields $x_* \in \mathbb{F}$. Therefore, we can apply (3.7) and Lemma 3.1 to obtain,

for every $t \geq t_0$,

$$\begin{aligned} \frac{d}{dt} \mathcal{E}(t) &\leq -\theta^2 t \sigma(t) (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) \\ &\quad - \frac{1}{2} \beta \theta t \delta(t) \|Ax(t) - b\|^2 - \xi \theta t \|\dot{x}(t), \dot{\lambda}(t)\|^2 \leq 0. \end{aligned} \tag{B.3}$$

This means that \mathcal{E} is nonincreasing on $[t_0, +\infty)$. For every $t \geq t_0$, by integrating (B.3) from t_0 to t , we obtain

$$\begin{aligned} &\theta^2 \int_{t_0}^t t \sigma(t) (\mathcal{L}(x(s), \lambda_*) - \mathcal{L}(x_*, \lambda(s))) ds \\ &\quad + \frac{\beta \theta}{2} \int_{t_0}^t s \delta(s) \|Ax(s) - b\|^2 ds + \xi \theta \int_{t_0}^t s \|\dot{x}(s), \dot{\lambda}(s)\|^2 ds \\ &\leq \mathcal{E}(t_0) - \mathcal{E}(t) \leq \mathcal{E}(t_0), \end{aligned}$$

where the last inequality follows from (3.8). Since all quantities inside the integrals are nonnegative, we obtain (3.10)–(3.12) by letting $t \rightarrow +\infty$.

(ii) Let $t \geq t_0$ be fixed. Inequality (B.3) tells us that \mathcal{E} is nonincreasing on $[t_0, +\infty)$.

$$\begin{aligned} &\theta^2 t^2 \delta(t) (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) \\ &\quad + \|v(t)\|^2 + \frac{\xi}{2} \|(x(t), \lambda(t)) - (x_*, \lambda_*)\|^2 \leq \mathcal{E}(t_0). \end{aligned} \tag{B.4}$$

Assuming $\alpha > 3$ and $\frac{1}{2} \geq \theta > \frac{1}{\alpha-1}$, we immediately see $\xi > 0$. From (B.4) we obtain

$$\|(x(t), \lambda(t)) - (x_*, \lambda_*)\|^2 \leq \frac{2\mathcal{E}(t_0)}{\xi}, \tag{B.5}$$

and

$$\|v(t)\| = \|(x(t), \lambda(t)) - (x_*, \lambda_*) + \theta t (\dot{x}(t), \dot{\lambda}(t))\| \leq \sqrt{2\mathcal{E}(t_0)}. \tag{B.6}$$

The estimate (B.5) leads to the boundedness of the trajectory. Moreover, applying the triangle inequality and (B.5)–(B.6), we obtain

$$\begin{aligned} t \|\dot{x}(t), \dot{\lambda}(t)\| &\leq \frac{1}{\theta} (\|(x(t), \lambda(t)) - (x_*, \lambda_*)\| + \|v(t)\|) \\ &\leq \frac{1}{\theta} \left(\sqrt{\frac{2\mathcal{E}(t_0)}{\xi}} + \sqrt{2\mathcal{E}(t_0)} \right) = \frac{1}{\theta} \left(\frac{1}{\sqrt{\xi}} + 1 \right) \sqrt{2\mathcal{E}(t_0)}, \end{aligned} \tag{B.7}$$

which gives the desired convergence rate. □

Proof of Proposition 4.2

Thanks to ∇f being ℓ -Lipschitz continuous, we can use (1.3) to refine relation (B.2) in the proof of Lemma 3.1

$$\begin{aligned}
 & - \left(\langle \nabla_x \mathcal{L}_\beta(x(t), \lambda_*) , 0 \rangle , (x(t), \lambda(t)) - (x_*, \lambda_*) \right) \\
 & = \langle \nabla f(x(t)) , x_* - x(t) \rangle + \langle A^* \lambda_* , x_* - x(t) \rangle + \beta \langle A^*(Ax(t) - b) , x_* - x(t) \rangle \\
 & \leq - (f(x(t)) - f(x_*)) - \frac{1}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 - \langle \lambda_* , Ax(t) - b \rangle \\
 & \quad - \beta \|Ax(t) - b\|^2 \\
 & = - (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) - \frac{1}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 \\
 & \quad - \frac{\beta}{2} \|Ax(t) - b\|^2 .
 \end{aligned}$$

Consequently, combining this inequality with (B.1) yields, for every $t \geq t_0$

$$\begin{aligned}
 \frac{d}{dt} \mathcal{E}(t) & \leq -\theta^2 t \sigma(t) (\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))) - \xi \theta t \|\dot{x}(t), \dot{\lambda}(t)\|^2 \\
 & \quad - \frac{\theta t \delta(t)}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 - \frac{\theta \beta t \delta(t)}{2} \|Ax(t) - b\|^2 \\
 & \leq -\frac{\theta t \delta(t)}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 .
 \end{aligned}$$

Integration of this inequality produces (4.4).

The finiteness of the second integral is only entailed by (3.11) when $\beta > 0$. For the general case $\beta \geq 0$, we use (3.14) and the fact that δ is nondecreasing on $[t_0, +\infty)$ to obtain

$$\int_{t_0}^{+\infty} t \delta(t) \|Ax(t) - b\|^2 dt \leq 4C_1^2 \int_{t_0}^{+\infty} \frac{1}{t^3 \delta(t)} dt \leq \frac{4C_1^2}{\delta(t_0)} \int_{t_0}^{+\infty} \frac{1}{t^3} dt < +\infty ,$$

and the proof is complete. □

Proof of Lemma 4.3

Let $t \geq t_0$ be fixed. Differentiating W with respect to time yields

$$\begin{aligned}
 \dot{W}(t) & = \dot{\delta}(t) \left[\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t)) \right] \\
 & \quad + \delta(t) \left[\langle \nabla_x \mathcal{L}_\beta(x(t), \lambda_*) , \dot{x}(t) \rangle - \langle \nabla_\lambda \mathcal{L}_\beta(x_*, \lambda(t)) , \dot{\lambda}(t) \rangle \right] \\
 & \quad + \langle \ddot{x}(t) , \dot{x}(t) \rangle + \langle \ddot{\lambda}(t) , \dot{\lambda}(t) \rangle .
 \end{aligned}$$

Recall the formulas for the gradients of \mathcal{L}

$$\begin{aligned}
 \nabla_x \mathcal{L}_\beta(x(t), \lambda_*) & = \nabla f(x(t)) + A^* \lambda_* + \beta A^*(Ax(t) - b) , \\
 \nabla_\lambda \mathcal{L}_\beta(x_*, \lambda(t)) & = Ax_* - b = 0 ,
 \end{aligned}$$

since $x_* \in \mathbb{F}$. Plugging this into the expression for $\dot{W}(t)$ gives us

$$\begin{aligned} \dot{W}(t) &= \dot{\delta}(t) \left[\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t)) \right] \\ &\quad + \delta(t) \left\langle \nabla_x \mathcal{L}_\beta(x(t), \lambda(t) + \theta t \dot{\lambda}(t)), \dot{x}(t) \right\rangle + \langle \ddot{x}(t), \dot{x}(t) \rangle \\ &\quad - \delta(t) \left\langle \lambda(t) - \lambda_* + \theta t \dot{\lambda}(t), A \dot{x}(t) \right\rangle \\ &\quad - \delta(t) \left\langle \nabla_\lambda \mathcal{L}_\beta(x(t) + \theta t \dot{x}(t), \lambda(t)), \dot{\lambda}(t) \right\rangle + \langle \ddot{\lambda}(t), \dot{\lambda}(t) \rangle \\ &\quad + \delta(t) \left\langle Ax(t) - b + \theta t A \dot{x}(t), \dot{\lambda}(t) \right\rangle. \end{aligned}$$

By regrouping and using (2.6), we arrive at

$$\begin{aligned} \dot{W}(t) &= \dot{\delta}(t) \left[\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t)) \right] - \frac{\alpha}{t} \|\dot{x}(t)\|^2 - \frac{\alpha}{t} \|\dot{\lambda}(t)\|^2 \\ &\quad - \delta(t) \langle \lambda(t) - \lambda_*, A \dot{x}(t) \rangle + \delta(t) \langle Ax(t) - b, \dot{\lambda}(t) \rangle. \end{aligned} \tag{B.8}$$

On the other hand, by the chain rule, we have

$$\begin{aligned} \dot{\varphi}(t) &= \langle x(t) - x_*, \dot{x}(t) \rangle + \langle \lambda(t) - \lambda_*, \dot{\lambda}(t) \rangle, \\ \ddot{\varphi}(t) &= \langle x(t) - x_*, \ddot{x}(t) \rangle + \|\dot{x}(t)\|^2 + \langle \lambda(t) - \lambda_*, \ddot{\lambda}(t) \rangle + \|\dot{\lambda}(t)\|^2. \end{aligned}$$

By combining these relations, (2.6) and the fact that $x_* \in \mathbb{F}$, we get

$$\begin{aligned} \ddot{\varphi}(t) + \frac{\alpha}{t} \dot{\varphi}(t) &= \left\langle x(t) - x_*, \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) \right\rangle + \left\langle \lambda(t) - \lambda_*, \ddot{\lambda}(t) + \frac{\alpha}{t} \dot{\lambda}(t) \right\rangle + \|\dot{x}(t)\|^2 \\ &\quad + \|\dot{\lambda}(t)\|^2 \\ &= -\langle x(t) - x_*, \delta(t) \nabla_x \mathcal{L}_\beta(x(t), \lambda(t) + \theta t \dot{\lambda}(t)) \rangle \\ &\quad + \langle \lambda(t) - \lambda_*, \delta(t) \nabla_\lambda \mathcal{L}_\beta(x(t) + \theta t \dot{x}(t), \lambda(t)) \rangle + \|\dot{x}(t)\|^2 + \|\dot{\lambda}(t)\|^2 \\ &= -\langle x(t) - x_*, \delta(t) \nabla_x \mathcal{L}_\beta(x(t), \lambda_*) \rangle \\ &\quad - \langle Ax(t) - b, \delta(t) (\lambda(t) - \lambda_* + \theta t \dot{\lambda}(t)) \rangle \\ &\quad + \langle \lambda(t) - \lambda_*, \delta(t) (Ax(t) - b + \theta t A \dot{x}(t)) \rangle + \|\dot{x}(t)\|^2 + \|\dot{\lambda}(t)\|^2 \\ &= -\delta(t) \langle x(t) - x_*, \nabla_x \mathcal{L}_\beta(x(t), \lambda_*) \rangle - \theta t \delta(t) \langle Ax(t) - b, \dot{\lambda}(t) \rangle \\ &\quad + \theta t \delta(t) \langle \lambda(t) - \lambda_*, A \dot{x}(t) \rangle + \|\dot{x}(t)\|^2 + \|\dot{\lambda}(t)\|^2 \end{aligned} \tag{B.9}$$

The Lipschitz continuity of ∇f entails

$$\begin{aligned} &-\langle x(t) - x_*, \nabla_x \mathcal{L}_\beta(x(t), \lambda_*) \rangle \\ &= -\langle x(t) - x_*, \nabla f(x(t)) \rangle - \langle x(t) - x_*, A^* \lambda_* \rangle - \beta \|Ax(t) - b\|^2 \\ &\leq -(f(x(t)) - f(x_*)) - \frac{1}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 \\ &\quad - \langle \lambda_*, Ax(t) - b \rangle - \beta \|Ax(t) - b\|^2 \end{aligned}$$

$$\begin{aligned}
 &= -\left(\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))\right) - \frac{1}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 \\
 &\quad - \frac{\beta}{2} \|Ax(t) - b\|^2.
 \end{aligned}$$

This, together with (B.9), implies

$$\begin{aligned}
 \ddot{\varphi}(t) + \frac{\alpha}{t} \dot{\varphi}(t) &\leq -\delta(t)\left(\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))\right) - \theta t \delta(t) \langle Ax(t) - b, \dot{\lambda}(t) \rangle \\
 &\quad + \theta t \delta(t) \langle \lambda(t) - \lambda_*, A\dot{x}(t) \rangle + \|\dot{x}(t)\|^2 + \|\dot{\lambda}(t)\|^2 \\
 &\quad - \frac{\delta(t)}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 - \frac{\beta\delta(t)}{2} \|Ax(t) - b\|^2. \tag{B.10}
 \end{aligned}$$

Multiplying (B.8) by $\theta t > 0$ and then adding the result to (B.10) yields

$$\begin{aligned}
 \ddot{\varphi}(t) + \frac{\alpha}{t} \dot{\varphi}(t) + \theta t \dot{W}(t) &= -(\delta(t) - \theta t \dot{\delta}(t))\left(\mathcal{L}_\beta(x(t), \lambda_*) - \mathcal{L}_\beta(x_*, \lambda(t))\right) \\
 &\quad - \frac{\delta(t)}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 - \frac{\beta\delta(t)}{2} \|Ax(t) - b\|^2 \\
 &\quad + (1 - \theta\alpha) \|\langle \dot{x}(t), \dot{\lambda}(t) \rangle\|^2 \\
 &\leq -\frac{\delta(t)}{2\ell} \|\nabla f(x(t)) - \nabla f(x_*)\|^2 - \frac{\beta\delta(t)}{2} \|Ax(t) - b\|^2,
 \end{aligned}$$

where the last inequality follows from Assumption 2

$$\begin{aligned}
 1 - \theta\alpha &\leq -\theta < 0, \\
 -\delta(t) + \theta t \dot{\delta}(t) &\leq (2\theta - 1)\delta(t) + \theta t \dot{\delta}(t) \leq 0.
 \end{aligned}$$

The desired result then follows after some rearranging. □

Proof of Lemma 4.5

Let $t \geq t_0$ be fixed. From (2.6) and the fact that $A^*\lambda_* = -\nabla f(x_*)$, we have

$$\begin{aligned}
 &\delta^2(t) \|\nabla f(x(t)) - \nabla f(x_*) + \beta A^*(Ax(t) - b)\|^2 \\
 &= \left\| \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \delta(t) A^*(\lambda(t) - \lambda_* + \theta t \dot{\lambda}(t)) \right\|^2 \\
 &= \left\| \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) \right\|^2 + \delta^2(t) \|A^*(\lambda(t) - \lambda_* + \theta t \dot{\lambda}(t))\|^2 \\
 &\quad + 2\delta(t) \left\langle \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t), A^*(\lambda(t) - \lambda_*) \right\rangle \\
 &\quad + 2\theta t \delta(t) \langle \ddot{x}(t), A^*\dot{\lambda}(t) \rangle + 2\alpha\theta t \delta(t) \langle \dot{x}(t), A^*\dot{\lambda}(t) \rangle. \tag{B.11}
 \end{aligned}$$

Again using (2.6) yields

$$\begin{aligned} \delta^2(t) \|Ax(t) - b\|^2 &= \left\| \ddot{\lambda}(t) + \frac{\alpha}{t} \dot{\lambda}(t) - \theta t \delta(t) A \dot{x}(t) \right\|^2 \\ &= \left\| \ddot{\lambda}(t) + \frac{\alpha}{t} \dot{\lambda}(t) \right\|^2 + \theta^2 t^2 \delta^2(t) \|A \dot{x}(t)\|^2 \\ &\quad - 2\theta t \delta(t) \langle \ddot{\lambda}(t), A \dot{x}(t) \rangle - 2\alpha \theta \delta(t) \langle \dot{\lambda}(t), A \dot{x}(t) \rangle \end{aligned} \tag{B.12}$$

Adding (B.11) and (B.12) together produces

$$\begin{aligned} \delta^2(t) \|\nabla f(x(t)) - \nabla f(x_*) + \beta A^*(Ax(t) - b)\|^2 + \delta^2(t) \|Ax(t) - b\|^2 \\ = \left\| (\ddot{x}(t), \ddot{\lambda}(t)) + \frac{\alpha}{t} (\dot{x}(t), \dot{\lambda}(t)) \right\|^2 + \delta^2(t) \|A^*(\lambda(t) - \lambda_* \\ + \theta t \dot{\lambda}(t))\|^2 + \theta^2 t^2 \delta^2(t) \|A \dot{x}(t)\|^2 \\ + 2\theta t \delta(t) \langle \ddot{x}(t), A^* \dot{\lambda}(t) \rangle - 2\theta t \delta(t) \langle \ddot{\lambda}(t), A \dot{x}(t) \rangle \\ + 2\delta(t) \left\langle \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t), A^*(\lambda(t) - \lambda_*) \right\rangle. \end{aligned} \tag{B.13}$$

On the one hand, we have

$$\begin{aligned} \theta^2 t^2 \delta^2(t) \|A \dot{x}(t)\|^2 + 2\theta t \delta(t) \langle \ddot{x}(t), A^* \dot{\lambda}(t) \rangle - 2\theta t \delta(t) \langle \ddot{\lambda}(t), A \dot{x}(t) \rangle \\ = \theta^2 t^2 \delta^2(t) \|(A^* \dot{\lambda}(t), -A \dot{x}(t))\|^2 - \theta^2 t^2 \delta^2(t) \|A^* \dot{\lambda}(t)\|^2 \\ + 2\theta t \delta(t) \langle (\ddot{x}(t), \ddot{\lambda}(t)), (A^* \dot{\lambda}(t), -A \dot{x}(t)) \rangle \\ = -\|(\ddot{x}(t), \ddot{\lambda}(t))\|^2 + \|(\ddot{x}(t), \ddot{\lambda}(t)) + \theta t \delta(t) A^*(A^* \dot{\lambda}(t), -A \dot{x}(t))\|^2 \\ - \theta^2 t^2 \delta^2(t) \|A^* \dot{\lambda}(t)\|^2 \\ \geq -\|(\ddot{x}(t), \ddot{\lambda}(t))\|^2 - \theta^2 t^2 \delta^2(t) \|A^* \dot{\lambda}(t)\|^2. \end{aligned} \tag{B.14}$$

On the other hand, it holds

$$\begin{aligned} \left\| (\ddot{x}(t), \ddot{\lambda}(t)) + \frac{\alpha}{t} (\dot{x}(t), \dot{\lambda}(t)) \right\|^2 - \|(\ddot{x}(t), \ddot{\lambda}(t))\|^2 \\ = \frac{\alpha^2}{t^2} \|(\dot{x}(t), \dot{\lambda}(t))\|^2 + 2\frac{\alpha}{t} \langle (\ddot{x}(t), \ddot{\lambda}(t)), (\dot{x}(t), \dot{\lambda}(t)) \rangle \geq \frac{\alpha}{t} \frac{d}{dt} \|(\dot{x}(t), \dot{\lambda}(t))\|^2. \end{aligned} \tag{B.15}$$

Moreover,

$$\begin{aligned} \delta(t) \|A^*(\lambda(t) - \lambda_* + \theta t \dot{\lambda}(t))\|^2 - \theta^2 t^2 \delta(t) \|A^* \dot{\lambda}(t)\|^2 \\ = \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 + 2\theta t \delta(t) \langle A A^*(\lambda(t) - \lambda_*), \dot{\lambda}(t) \rangle \\ = ((1 - \theta)\delta(t) - \theta t \dot{\delta}(t)) \|A^*(\lambda(t) - \lambda_*)\|^2 + \theta \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \\ + \theta t \dot{\delta}(t) \|A^*(\lambda(t) - \lambda_*)\|^2 + \theta \frac{d}{dt} \|A^*(\lambda(t) - \lambda_*)\|^2 \end{aligned}$$

$$= \left((1 - \theta)\delta(t) - \theta t\dot{\delta}(t) \right) \|A^*(\lambda(t) - \lambda_*)\|^2 + \theta \frac{d}{dt} \left(t\delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \right). \tag{B.16}$$

Now, using (B.14), (B.15) and (B.16) in (B.13) yields

$$\begin{aligned} & \delta^2(t) \| \nabla f(x(t)) - \nabla f(x_*) + \beta A^*(Ax(t) - b) \|^2 + \delta^2(t) \| Ax(t) - b \|^2 \\ & \geq \left\| \left(\ddot{x}(t), \ddot{\lambda}(t) \right) + \frac{\alpha}{t} \left(\dot{x}(t), \dot{\lambda}(t) \right) \right\|^2 + \delta^2(t) \| A^*(\lambda(t) - \lambda_* + \theta t\dot{\lambda}(t)) \|^2 \\ & \quad - \left\| \left(\ddot{x}(t), \ddot{\lambda}(t) \right) \right\|^2 - \theta^2 t^2 \delta^2(t) \| A^*\dot{\lambda}(t) \|^2 + 2\delta(t) \left\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(\lambda(t) - \lambda_*) \right\rangle \\ & \geq \frac{\alpha}{t} \frac{d}{dt} \left\| \left(\dot{x}(t), \dot{\lambda}(t) \right) \right\|^2 + 2\delta(t) \left\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(\lambda(t) - \lambda_*) \right\rangle \\ & \quad + \delta(t) \left[\left((1 - \theta)\delta(t) - \theta t\dot{\delta}(t) \right) \|A^*(\lambda(t) - \lambda_*)\|^2 \right. \\ & \quad \left. + \theta \frac{d}{dt} \left(t\delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \right) \right] \end{aligned}$$

Finally, since

$$\begin{aligned} & \| \nabla f(x(t)) - \nabla f(x_*) + \beta A^*(Ax(t) - b) \|^2 + \| Ax(t) - b \|^2 \\ & \leq 2 \| \nabla f(x(t)) - \nabla f(x_*) \|^2 + (2\beta \|A\|^2 + 1) \| Ax(t) - b \|^2, \end{aligned}$$

the conclusion follows after dividing the inequality by $\delta(t)$. □

Proof of Lemma 4.6

For every $t \geq t_0$, by summing up the two inequalities produced by Lemmas 4.3 and 4.5 we deduce that

$$\begin{aligned} & \ddot{\varphi}(t) + \frac{\alpha}{t}\dot{\varphi}(t) + \theta t\dot{W}(t) + \frac{\alpha}{t\delta(t)} \left\| \left(\dot{x}(t), \dot{\lambda}(t) \right) \right\|^2 \\ & \quad + \theta \frac{d}{dt} \left(t\delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \right) + 2 \left\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(\lambda(t) - \lambda_*) \right\rangle \\ & \leq - \left((1 - \theta)\delta(t) - \theta t\dot{\delta}(t) \right) \|A^*(\lambda(t) - \lambda_*)\|^2 \\ & \quad + \left(2 - \frac{1}{2\ell} \right) \delta(t) \| \nabla f(x(t)) - \nabla f(x_*) \|^2 \\ & \quad + \left(2\beta^2 \|A\|^2 + 1 - \frac{\beta}{2} \right) \delta(t) \| Ax(t) - b \|^2 \\ & \leq - \left((1 - \theta)\delta(t) - \theta t\dot{\delta}(t) \right) \|A^*(\lambda(t) - \lambda_*)\|^2 \\ & \quad + C_3\delta(t) \| \nabla f(x(t)) - \nabla f(x_*) \|^2 \\ & \quad + C_4\delta(t) \| Ax(t) - b \|^2, \end{aligned} \tag{B.17}$$

where

$$C_3 := \left[2 - \frac{1}{2\ell} \right]_+ \geq 0 \quad \text{and} \quad C_4 := \left[2\beta^2 \|A\|^2 + 1 - \frac{\beta}{2} \right]_+ \geq 0.$$

Multiplying (B.17) by t^α and integrating from t_0 to t , we obtain

$$\begin{aligned} & I_1(t) + \theta I_2(t) + \alpha I_3(t) + \theta I_4(t) + 2I_5(t) \\ & \leq \int_{t_0}^t s^\alpha \left((1 - \theta)\delta(s) - \theta s \dot{\delta}(s) \right) \|A^*(\lambda(s)) \\ & \quad - \lambda_*\|^2 ds + C_3 \int_{t_0}^t s^\alpha \delta(s) \|\nabla f(x(s)) - \nabla f(x_*)\|^2 ds \\ & \quad + C_4 \int_{t_0}^t s^\alpha \delta(s) \|Ax(s) - b\|^2 ds, \end{aligned} \tag{B.18}$$

where

$$\begin{aligned} I_1(t) & := \int_{t_0}^t (s^\alpha \ddot{\varphi}(s) + \alpha s^{\alpha-1} \dot{\varphi}(s)) ds, \\ I_2(t) & := \int_{t_0}^t s^{\alpha+1} \dot{W}(s) ds, \\ I_3(t) & := \int_{t_0}^t \frac{s^{\alpha-1}}{\delta(s)} \frac{d}{ds} \left\| (\dot{x}(s), \dot{\lambda}(s)) \right\|^2 ds, \\ I_4(t) & := \int_{t_0}^t s^\alpha \frac{d}{ds} \left(s \delta(s) \|A^*(\lambda(s) - \lambda_*)\|^2 \right) ds, \\ I_5(t) & := \int_{t_0}^t (s^\alpha \ddot{x}(s) + \alpha s^{\alpha-1} \dot{x}(s), A^*(\lambda(s) - \lambda_*)) ds. \end{aligned}$$

We will furnish five different inequalities from computing each of these integrals separately. Let $t \geq t_0$ be fixed.

- *The integral $I_1(t)$.* By the chain rule, for $s \geq t_0$ it holds

$$s^\alpha \ddot{\varphi}(s) + \alpha s^{\alpha-1} \dot{\varphi}(s) = \frac{d}{ds} (s^\alpha \dot{\varphi}(s)),$$

which leads to

$$0 = I_1(t) - t^\alpha \dot{\varphi}(t) + t_0^\alpha \dot{\varphi}(t_0) \leq I_1(t) - t^\alpha \dot{\varphi}(t) + |t_0^\alpha \dot{\varphi}(t_0)|. \tag{B.19}$$

- *The integrals $I_2(t)$ and $I_4(t)$.* Integration by parts gives

$$\begin{aligned} I_2(t) + I_4(t) & = t^{\alpha+1} W(t) - t_0^{\alpha+1} W(t_0) \\ & \quad - (\alpha + 1) \int_{t_0}^t s^\alpha W(s) ds + t^{\alpha+1} \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \end{aligned}$$

$$\begin{aligned}
 & - t_0^{\alpha+1} \delta(t_0) \|A^*(\lambda(t_0) - \lambda_*)\|^2 \\
 & - \alpha \int_{t_0}^t s^\alpha \delta(s) \|A^*(\lambda(s) - \lambda_*)\|^2 ds,
 \end{aligned}$$

and from here

$$\begin{aligned}
 t^{\alpha+1} \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 & \leq t^{\alpha+1} W(t) + t^{\alpha+1} \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \\
 & = I_2(t) + I_4(t) + t_0^{\alpha+1} W(t_0) + t_0^{\alpha+1} \delta(t_0) \|A^*(\lambda(t_0) - \lambda_*)\|^2 \\
 & \quad + (\alpha + 1) \int_{t_0}^t s^\alpha W(s) ds \\
 & \quad + \alpha \int_{t_0}^t s^\alpha \delta(s) \|A^*(\lambda(s) - \lambda_*)\|^2 ds. \tag{B.20}
 \end{aligned}$$

• *The integral $I_3(t)$.* Again by integrating by parts, we get

$$\begin{aligned}
 I_3(t) & = \frac{t^{\alpha-1}}{\delta(t)} \|(\dot{x}(t), \dot{\lambda}(t))\|^2 - \frac{t_0^{\alpha-1}}{\delta(t_0)} \|(\dot{x}(t_0), \dot{\lambda}(t_0))\|^2 \\
 & \quad - \int_{t_0}^t \left[\frac{(\alpha - 1)s^{\alpha-2}\delta(s) - s^{\alpha-1}\dot{\delta}(s)}{\delta^2(s)} \right] \|(\dot{x}(s), \dot{\lambda}(s))\|^2 ds.
 \end{aligned}$$

For $s \geq t_0$, according to Assumption 2 we have $\dot{\delta}(s) \geq 0$, hence δ is monotonically increasing and therefore

$$\frac{(\alpha - 1)s^{\alpha-2}\delta(s) - s^{\alpha-1}\dot{\delta}(s)}{\delta^2(s)} \leq \frac{(\alpha - 1)s^{\alpha-2}\delta(s)}{\delta^2(s)} \leq \frac{(\alpha - 1)s^\alpha}{t_0^2\delta(t_0)}.$$

It follows that

$$\begin{aligned}
 0 & \leq \frac{t^{\alpha-1}}{\delta(t)} \|(\dot{x}(t), \dot{\lambda}(t))\|^2 \\
 & = I_3(t) + \frac{t_0^{\alpha-1}}{\delta(t_0)} \|(\dot{x}(t_0), \dot{\lambda}(t_0))\|^2 \\
 & \quad + \int_{t_0}^t \left[\frac{(\alpha - 1)s^{\alpha-2}\delta(s) - s^{\alpha-1}\dot{\delta}(s)}{\delta^2(s)} \right] \|(\dot{x}(s), \dot{\lambda}(s))\|^2 ds \\
 & \leq I_3(t) + \frac{t_0^{\alpha-1}}{\delta(t_0)} \|(\dot{x}(t_0), \dot{\lambda}(t_0))\|^2 + \frac{\alpha - 1}{t_0^2\delta(t_0)} \int_{t_0}^t s^\alpha \|(\dot{x}(s), \dot{\lambda}(s))\|^2 ds. \tag{B.21}
 \end{aligned}$$

• *The integral $I_5(t)$.* Integration by parts entails

$$\begin{aligned}
 I_5(t) & = \int_{t_0}^t \left\langle \frac{d}{ds} (s^\alpha \dot{x}(s)), A^*(\lambda(s) - \lambda_*) \right\rangle ds \\
 & = t^\alpha \langle \dot{x}(t), A^*(\lambda(t) - \lambda_*) \rangle - t_0^\alpha \langle \dot{x}(t_0), A^*(\lambda(t_0) - \lambda_*) \rangle
 \end{aligned}$$

$$- \int_{t_0}^t s^\alpha \langle \dot{x}(s), A^* \dot{\lambda}(s) \rangle ds.$$

By the Cauchy–Schwarz inequality, we deduce that

$$\begin{aligned} \int_{t_0}^t s^\alpha \langle \dot{x}(s), A^* \dot{\lambda}(s) \rangle ds &\leq \|A\| \int_{t_0}^t s^\alpha \|\dot{x}(s)\| \|\dot{\lambda}(s)\| ds \\ &\leq \frac{\|A\|}{2} \int_{t_0}^t s^\alpha (\|\dot{x}(s)\|^2 + \|\dot{\lambda}(s)\|^2) ds, \end{aligned}$$

and thus

$$\begin{aligned} 0 &\leq I_5(t) - t^\alpha \langle \dot{x}(t), A^*(\lambda(t) - \lambda_*) \rangle + |t_0^\alpha \langle \dot{x}(t_0), A^*(\lambda(t_0) - \lambda_*) \rangle| \\ &\quad + \frac{\|A\|}{2} \int_{t_0}^t s^\alpha \|\langle \dot{x}(s), \dot{\lambda}(s) \rangle\|^2 ds. \end{aligned} \tag{B.22}$$

Now, summing up the inequalities (B.19), (B.20), (B.21), and (B.22), then we proceed to employ (B.18) and obtain

$$\begin{aligned} &\theta t^{\alpha+1} \delta(t) \|A^*(\lambda(t) - \lambda_*)\|^2 \\ &\leq I_1(t) + \theta I_2(t) + \alpha I_3(t) + \theta I_4(t) + 2I_5 - t^\alpha \dot{\varphi}(t) \\ &\quad + \int_{t_0}^t s^\alpha \left[\theta(\alpha + 1)W(s) + \left(\frac{\alpha(\alpha - 1)}{t_0^2 \delta(t_0)} + \|A\| \right) \|\langle \dot{x}(s), \dot{\lambda}(s) \rangle\|^2 \right] ds \\ &\quad + \theta \alpha \int_{t_0}^t s^\alpha \delta(s) \|A^*(\lambda(s) - \lambda_*)\|^2 ds - 2t^\alpha \langle \dot{x}(t), A^*(\lambda(t) - \lambda_*) \rangle + C_5 \\ &\leq -t^\alpha \dot{\varphi}(t) + \int_{t_0}^t s^\alpha V(s) ds + \int_{t_0}^t s^\alpha \left[(\theta(\alpha + 1) - 1)\delta(t) + \theta s \dot{\delta}(s) \right] \\ &\quad \|A^*(\lambda(s) - \lambda_*)\|^2 ds \\ &\quad - 2t^\alpha \langle \dot{x}(t), A^*(\lambda(t) - \lambda_*) \rangle + C_5, \end{aligned}$$

where recall that V was given by

$$\begin{aligned} V(s) &:= \theta(\alpha + 1)W(s) + \left(\frac{\alpha(\alpha - 1)}{t_0^2 \delta(t_0)} + \|A\| \right) \|\langle \dot{x}(s), \dot{\lambda}(s) \rangle\|^2 \\ &\quad + C_3 \delta(s) \|\nabla f(x(s)) - \nabla f(x_*)\|^2 + C_4 \delta(s) \|Ax(s) - b\|^2, \end{aligned}$$

and the constant C_5 is given by

$$\begin{aligned} C_5 &:= t_0^\alpha |\dot{\varphi}(t_0)| + \theta t_0^{\alpha+1} W(t_0) + \alpha \frac{t_0^{\alpha-1}}{\delta(t_0)} \|\langle \dot{x}(t_0), \dot{\lambda}(t_0) \rangle\|^2 \\ &\quad + \theta t_0^{\alpha+1} \delta(t_0) \|A^*(\lambda(t_0) - \lambda_*)\|^2 + 2t_0^\alpha |\langle \dot{x}(t_0), A^*(\lambda(t_0) - \lambda_*) \rangle| \geq 0. \end{aligned}$$

We come then to the desired result.

References

1. Abbas, B., Attouch, H., Svaiter, B.F.: Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. *J. Optim. Theory Appl.* **161**(2), 331–360 (2014). <https://doi.org/10.1007/s10957-013-0414-5>
2. Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM J. Control Optim.* **38**(4), 1102–1119 (2000). <https://doi.org/10.1137/S0363012998335802>
3. Attouch, H., Goudou, X., Redont, P.: The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Commun. Contemp. Math.* **02**(1), 1–34 (2000). <https://doi.org/10.1142/S0219199700000025>
4. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping: application to optimization and mechanics. *J. Math. Pures Appl.* **81**(8), 747–779 (2002). [https://doi.org/10.1016/S0021-7824\(01\)01253-3](https://doi.org/10.1016/S0021-7824(01)01253-3)
5. Attouch, H., Chbani, Z., Riahi, H.: Combining fast inertial dynamics for convex optimization with Tikhonov regularization. *J. Math. Anal. Appl.* **457**(2), 1065–1094 (2018). <https://doi.org/10.1016/j.jmaa.2016.12.017>
6. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.* **168**(1), 123–175 (2018). <https://doi.org/10.1007/s10107-016-0992-8>
7. Attouch, H., Chbani, Z., Riahi, H.: Fast proximal methods via time scaling of damped inertial dynamics. *SIAM J. Optim.* **29**(3), 2227–2256 (2019). <https://doi.org/10.1137/18M1230207>
8. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM Control Optim. Calc. Var.* (2019). <https://doi.org/10.1051/cocv/2017083>
9. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via time scaling of damped inertial gradient dynamics. *Pure Appl. Funct. Anal.* **6**(6), 1081–1117 (2021)
10. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics. *J. Optim. Theory Appl.* **193**(1), 704–736 (2022). <https://doi.org/10.1007/s10957-021-01859-2>
11. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with Hessian driven damping. *Math. Program.* **193**(1), 113–155 (2022). <https://doi.org/10.1007/s10107-020-01591-1>
12. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via a third-order in time evolution equation. *Optimization* **71**(5), 1275–1304 (2022). <https://doi.org/10.1080/02331934.2020.1764953>
13. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$. *SIAM J. Optim.* **26**(3), 1824–1834 (2016). <https://doi.org/10.1137/15M1046095>
14. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, New York (2011). <https://doi.org/10.1007/978-1-4419-9467-7>
15. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009). <https://doi.org/10.1137/080716542>
16. Boş, R.I.: *Conjugate Duality in Convex Optimization*, vol. 637. Lecture Notes in Economics and Mathematical Systems. Springer, Berlin (2010). <https://doi.org/10.1007/978-3-642-04900-2>
17. Boş, R.I., Nguyen, D.-K.: Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping. *J. Differ. Equ.* **303**, 369–406 (2021). <https://doi.org/10.1016/j.jde.2021.09.021>
18. Boş, R.I., Csetnek, E.R., László, S.C.: A primal-dual dynamical approach to structured convex minimization problems. *J. Differ. Equ.* **269**(12), 10717–10757 (2020). <https://doi.org/10.1016/j.jde.2020.07.039>
19. Boş, R.I., Csetnek, E.R., Nguyen, D.-K.: Fast augmented Lagrangian method in the convex regime with convergence guarantees for the iterates. *Math. Program.* (2022). <https://doi.org/10.1007/s10107-022-01879-4>
20. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2010). <https://doi.org/10.1561/22000000016>

21. Cabot, A., Engler, H., Gadat, S.: Second-order differential equations with asymptotically small dissipation and piecewise flat potentials. *Electron. J. Differ. Equ.* **2009**(17), 33–38 (2009)
22. Cabot, A., Engler, H., Gadat, S.: On the long time behavior of second order differential equations with asymptotically small dissipation. *Trans. Am. Math. Soc.* **361**(11), 5983–6017 (2009)
23. Chambolle, A., Dossal, C.: On the convergence of the iterates of the “Fast iterative shrinkage/thresholding algorithm”. *J. Optim. Theory Appl.* **166**(3), 968–982 (2015). <https://doi.org/10.1007/s10957-015-0746-4>
24. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* **2**(1), 17–40 (1976). [https://doi.org/10.1016/0898-1221\(76\)90003-1](https://doi.org/10.1016/0898-1221(76)90003-1)
25. Goldstein, T., O’Donoghue, B., Setzer, S., Baraniuk, R.: Fast alternating direction optimization methods. *SIAM J. Imaging Sci.* **7**(3), 1588–1623 (2014). <https://doi.org/10.1137/120896219>
26. Güler, O.: New proximal point algorithms for convex minimization. *SIAM J. Optim.* **2**(4), 649–664 (1992). <https://doi.org/10.1137/0802032>
27. He, X., Hu, R., Fang, Y.P.: Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems. *SIAM J. Control Optim.* **59**(5), 3278–3301 (2021). <https://doi.org/10.1137/20M1355379>
28. He, X., Hu, R., Fang, Y.-P.: Fast primal-dual algorithm via dynamical system for a linearly constrained convex optimization problem. *Automatica* **146**, 110547 (2022). <https://doi.org/10.1016/j.automatica.2022.110547>
29. He, X., Hu, R., Fang, Y.-P.: Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems. *Numer. Algorithms* **90**(4), 1669–1690 (2022). <https://doi.org/10.1007/s11075-021-01246-y>
30. He, X., Hu, R., Fang, Y.-P.: Inertial primal-dual dynamics with damping and scaling for linearly constrained convex optimization problems. *Applicable Anal.* (2022). <https://doi.org/10.1080/00036811.2022.2104260>
31. Lin, Z., Li, H., Fang, C.: *Accelerated Optimization for Machine Learning: First-Order Algorithms*. Springer, Singapore (2020). <https://doi.org/10.1007/978-981-15-2910-8>
32. Madan, R., Lall, S.: Distributed algorithms for maximum lifetime routing in wireless sensor networks. *IEEE Trans. Wirel. Commun.* **5**(8), 2185–2193 (2006). <https://doi.org/10.1109/TWC.2006.1687734>
33. May, R.: Asymptotic for a second-order evolution equation with convex potential and vanishing damping term. *Turk. J. Math.* **41**(3), 681–685 (2017). <https://doi.org/10.3906/mat-1512-28>
34. Nesterov, Y.: A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. *Proc. USSR Acad. Sci.* **269**, 543–547 (1983)
35. Nesterov, Y.: *Introductory Lectures on Convex Optimization*, vol. 87. Applied Optimization. Springer, Boston (2004). <https://doi.org/10.1007/978-1-4419-8853-9>
36. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Am. Math. Soc.* **73**(4), 591–597 (1967)
37. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **4**(5), 1–17 (1964). [https://doi.org/10.1016/0041-5553\(64\)90137-5](https://doi.org/10.1016/0041-5553(64)90137-5)
38. Polyak, B.T.: *Introduction to Optimization*. Translations Series in Mathematics and Engineering. Optimization Software, Publications Division, New York (1987)
39. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**(2), 97–116 (1976). <https://doi.org/10.1287/moor.1.2.97>
40. Shi, G., Johansson, K.H.: Randomized optimal consensus of multi-agent systems. *Automatica* **48**(12), 3018–3030 (2012). <https://doi.org/10.1016/j.automatica.2012.08.018>
41. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *J. Mach. Learn. Res.* **17**(1), 5312–5354 (2016)
42. Yi, P., Hong, Y., Liu, F.: Distributed gradient algorithm for constrained optimization with application to load sharing in power systems. *Syst. Control Lett.* **83**, 45–52 (2015). <https://doi.org/10.1016/j.sysconle.2015.06.006>
43. Yi, P., Hong, Y., Liu, F.: Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. *Automatica* **74**, 259–269 (2016). <https://doi.org/10.1016/j.automatica.2016.08.007>
44. Zeng, X., Yi, P., Hong, Y., Xie, L.: Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems. *SIAM J. Control Optim.* **56**(6), 3973–3993 (2018). <https://doi.org/10.1137/17M1118609>

45. Zeng, X., Lei, J., Chen, J.: Dynamical primal-dual accelerated method with applications to network optimization. *IEEE Trans. Autom. Control* (2022). <https://doi.org/10.1109/TAC.2022.3152720>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.