Fast Augmented Lagrangian Method in the convex regime with convergence guarantees for the iterates

Radu Ioan Boţ[*]   Ernö Robert Csetnek[†]   Dang-Khoa Nguyen[‡]

August 1, 2022

Abstract

This work aims to minimize a continuously differentiable convex function with Lipschitz continuous gradient under linear equality constraints. The proposed inertial algorithm results from the discretization of the second-order primal-dual dynamical system with asymptotically vanishing damping term addressed by Boţ and Nguyen in [8], and it is formulated in terms of the Augmented Lagrangian associated with the minimization problem. The general setting we consider for the inertial parameters covers the three classical rules by Nesterov, Chambolle-Dossal and Attouch-Cabot used in the literature to formulate fast gradient methods. For these rules, we obtain in the convex regime convergence rates of order $\mathcal{O}\left(1/k^2\right)$ for the primal-dual gap, the feasibility measure, and the objective function value. In addition, we prove that the generated sequence of primal-dual iterates converges to a primal-dual solution in a general setting that covers the two latter rules. This is the first result which provides the convergence of the sequence of iterates generated by a fast algorithm for linearly constrained convex optimization problems without additional assumptions such as strong convexity. We also emphasize that all convergence results of this paper are compatible with the ones obtained in [8] in the continuous setting.

Key Words. Augmented Lagrangian Method, primal-dual numerical algorithm, Nesterov's fast gradient method, convergence rates, iterates convergence

AMS subject classification. 49M29, 65K05, 68Q25, 90C25, 65B99

1 Introduction

1.1 Problem formulation and motivation

Consider the optimization problem

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

where $\mathcal{H}, \mathcal{G}$ are real Hilbert spaces, $f \colon \mathcal{H} \to \mathbb{R}$ is a convex and Fréchet differentiable function with $L-$Lipschitz continuous gradient, for $L > 0$, $A \colon \mathcal{H} \to \mathcal{G}$ is a continuous linear operator and $b \in \mathcal{G}$. We assume that the set $\mathcal{S}$ of primal-dual optimal solutions of (1.1) (see Section 1.2 for a precise definition) is nonempty.

---

[*]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: `radu.bot@univie.ac.at`. Research partially supported by FWF (Austrian Science Fund), project W 1260.

[†]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: `robert.csetnek@univie.ac.at`. Research partially supported by FWF (Austrian Science Fund), project P 29809-N32.

[‡]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: `dang-khoa.nguyen@univie.ac.at`. Research supported by FWF (Austrian Science Fund), project P 29809-N32.

Optimization problems of type (1.1) arise in many applications in areas like image recovery [13, 23, 26, 28], machine learning [9, 17], and network optimization [30].

Other than in the unconstrained case, for which fast continuous and discrete time approaches have been intensively investigated in the last years, the study of solution methods with fast convergence rates for linearly constrained convex optimization problems of the form (1.1) is in an incipient stage.

Zeng, Lei, and Chen (in [30]) and He, Hu, and Fang (in [16]) have investigated a dynamical system with asymptotic vanishing damping attached to (1.1), and have shown a convergence rate of order $\mathcal{O}\left(1/t^2\right)$ for the primal-dual gap, while Attouch, Chbani, Fadili and Riahi have considered in [2] a more general dynamical system with time rescaling. More recently, for a primal-dual dynamical system formulated in the spirit of [2, 16, 30], Boţ and Nguyen have obtained in [8] fast convergence rates for the primal-dual gap, the feasibility measure and the objective function value along the generated trajectory, and, additionally, have proved asymptotic convergence guarantees for the primal-dual trajectory to a primal-dual optimal solution.

Fast numerical methods for solving (1.1) have been mainly considered in the literature under additional assumptions such as strong convexity, and in several cases the convergence rate results have been formulated in terms of ergodic sequences. In the merely convex regime no convergence result for the iterates has been provided so far for fast convergence algorithms. To the works addressing fast converging methods for linearly constrained convex optimization problems belong [11, 12, 13, 15, 19, 20, 23, 24, 26, 27, 28, 29], at which we will take a closer look in Section 1.3.

The aim of this paper is to propose a numerical algorithm for solving (1.1), which results from the discretization of the dynamical system in [8], exhibits fast convergence rates for the primal-dual gap, the feasibility measure, and the objective function value as well as convergence of the sequence of iterates without additional assumptions such as strong convexity. Although there is an obvious interplay between continuous time dissipative dynamical systems and their discrete counterparts, one cannot directly and straightforwardly transfer asymptotic results from the continuous setting to numerical algorithms, thus, a separate analysis is needed for the latter. In this paper we will also comment on the similarities and the differences between the continuous and discrete time approaches.

1.2 Augmented Lagrangian formulation

Consider the saddle point problem associated to problem (1.1)

$$\min_{x\in\mathcal{H}} \max_{\lambda\in\mathcal{G}} \mathcal{L}\left(x,\lambda\right),$$

where $\mathcal{L}\colon \mathcal{H}\times\mathcal{G}\to\mathbb{R}$ denotes the Lagrangian function

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

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

$$\mathcal{L}\left(x_*,\lambda\right) \leqslant \mathcal{L}\left(x_*,\lambda_*\right) \leqslant \mathcal{L}\left(x,\lambda_*\right).$$

If $(x_*,\lambda_*)\in\mathcal{H}\times\mathcal{G}$ is a saddle point of $\mathcal{L}$, then $x_*\in\mathcal{H}$ is an optimal solution of (1.1) and $\lambda_*\in\mathcal{G}$ is an optimal solution of its Lagrange dual problem. If $x_*\in\mathcal{H}$ is an optimal solution of (1.1) and a suitable constraint qualification is fulfilled (see, for instance, [5, 7]), then there exists an optimal solution $\lambda_*\in\mathcal{G}$ of the Lagrange dual problem of (1.1) such that $(x_*,\lambda_*)\in\mathcal{H}\times\mathcal{G}$ is a saddle point of $\mathcal{L}$.

The set of saddle points of $\mathcal{L}$, called also set of primal-dual optimal solutions of (1.1), will be denoted by $\mathcal{S}$ and, as stated above, throughout this paper it will be assumed to be nonempty.

The set of feasible points of (1.1) will be denoted by $\mathcal{F} := \{x \in \mathcal{H} \colon Ax = b\}$ and the optimal objective value of (1.1) by f_*.

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

$$(x_*, \lambda_*) \in \mathcal{S} \Leftrightarrow \begin{cases} \nabla_x \mathcal{L}(x_*, \lambda_*) & = 0 \\ \nabla_\lambda \mathcal{L}(x_*, \lambda_*) & = 0 \end{cases} \Leftrightarrow \begin{cases} \nabla f(x_*) + A^* \lambda_* & = 0 \\ Ax_* - b & = 0 \end{cases}, \tag{1.2}$$

where $A^* \colon \mathcal{G} \to \mathcal{H}$ denotes the adjoint operator of A. This optimality system can be equivalently written as

$$\mathcal{T}_{\mathcal{L}}(x_*, \lambda_*) = 0,$$

where

$$\mathcal{T}_{\mathcal{L}} \colon \mathcal{H} \times \mathcal{G} \to \mathcal{H} \times \mathcal{G}, \quad \mathcal{T}_{\mathcal{L}}(x, \lambda) = \begin{pmatrix} \nabla_x \mathcal{L}(x, \lambda) \\ -\nabla_\lambda \mathcal{L}(x, \lambda) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + A^* \lambda \\ b - Ax \end{pmatrix},$$

is the maximally monotone operator associated to the convex-concave function $\mathcal{L}$. Indeed, it is immediate to verify that $\mathcal{T}_{\mathcal{L}}$ is monotone. Since it is also continuous, it is maximally monotone (see, for instance, [5, Corollary 20.28]). Therefore $\mathcal{S}$ can be interpreted as the set of zeros of the maximally monotone operator $\mathcal{T}_{\mathcal{L}}$, which means that it is a closed convex subset of $\mathcal{H} \times \mathcal{G}$ (see, for instance, [5, Proposition 23.39]).

For $\beta \geqslant 0$, we also consider the augmented Lagrangian $\mathcal{L}_\beta \colon \mathcal{H} \times \mathcal{G} \to \mathbb{R}$ associated with (1.1)

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

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

$$f(x) = \mathcal{L}_\beta(x, \lambda) = \mathcal{L}(x, \lambda). \tag{1.3}$$

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

$$\mathcal{L}(x_*, \lambda) = \mathcal{L}_\beta(x_*, \lambda) = \mathcal{L}(x_*, \lambda_*) = \mathcal{L}_\beta(x_*, \lambda_*) \leqslant \mathcal{L}(x, \lambda_*) \leqslant \mathcal{L}_\beta(x, \lambda_*). \tag{1.4}$$

In addition,

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

1.3 Related works

In this section we will recall the most significant fast primal-dual numerical approaches for linearly constrained convex optimization problems and for convex optimization problems involving compositions with continuous linear operators.

In [11], Chambolle and Pock have studied in a finite-dimensional setting the convergence rates of their celebrated primal-dual algorithm for solving the minimax problem

$$\min_{x \in \mathcal{H}} \max_{\lambda \in \mathcal{G}} \mathcal{L}(x, \lambda) := f(x) + \langle Ax, \lambda \rangle - g^*(\lambda), \tag{1.5}$$

which is naturally attached to the convex optimization problem

$$\min_{x \in \mathcal{H}} f(x) + g(Ax), \tag{1.6}$$

with $f \colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ and $g \colon \mathcal{G} \to \mathbb{R} \cup \{+\infty\}$ proper, convex and lower semicontinuous functions and $g^* \colon \mathcal{G} \to \mathbb{R} \cup \{+\infty\}$ the Fenchel conjugate of g. The problem (1.6) becomes (1.1)

3

for g the indicator function of the set $\{b\}$. For the primal-dual sequence of iterates $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ the corresponding ergodic sequence $\{(\bar{x}_k, \bar{\lambda}_k)\}_{k \geqslant 0}$ is defined for every $k \geqslant 0$ as

$$\bar{x}_k := \frac{1}{\sum_{i=0}^{k} \sigma_i} \sum_{i=0}^{k} \sigma_i x_i \qquad \text{and} \qquad \bar{\lambda}_k := \frac{1}{\sum_{i=0}^{k} \sigma_i} \sum_{i=0}^{k} \sigma_i \lambda_i,$$

where $\{\sigma_k\}_{k \geqslant 0}$ is a sequence of properly chosen positive step sizes. The Chambolle-Pock primal-dual algorithm exhibits for the restricted primal-dual gap an ergodic convergence rate of

$$\sup_{(x,\lambda) \in \mathcal{X} \times \mathcal{Y}} \left(\mathcal{L}\left(\bar{x}_k, \lambda\right) - \mathcal{L}\left(x, \bar{\lambda}_k\right) \right) = \mathcal{O}\left(\frac{1}{k}\right) \text{ as } k \to +\infty,$$

where $\mathcal{X} \subseteq \mathcal{H}$ and $\mathcal{Y} \subseteq \mathcal{G}$ are bounded sets. If f is strongly convex, then the accelerated variant of this primal-dual algorithm exhibits for the same restricted primal-dual gap an ergodic convergence rate of

$$\sup_{(x,\lambda) \in \mathcal{X} \times \mathcal{Y}} \left(\mathcal{L}\left(\bar{x}_k, \lambda\right) - \mathcal{L}\left(x, \bar{\lambda}_k\right) \right) = \mathcal{O}\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty$$

whereas, if both f and g^* are strongly convex, then even linear convergence can be achieved.

In [12], Chen, Lan and Ouyang have considered the same minimax problem (1.5), but for $f \colon \mathcal{H} \to \mathbb{R}$ a convex and Fréchet differentiable function with L-Lipschitz continuous gradient, for $L > 0$, and have proposed a primal-dual algorithm that exhibits for the restricted primal-dual gap an ergodic convergence rate of

$$\sup_{(x,\lambda) \in \mathcal{X} \times \mathcal{Y}} \left(\mathcal{L}\left(\bar{x}_k, \lambda\right) - \mathcal{L}\left(x, \bar{\lambda}_k\right) \right) = \mathcal{O}\left(\frac{L}{k^2} + \frac{\|A\|}{k}\right) \text{ as } k \to +\infty. \tag{1.7}$$

A stochastic counterpart of the primal-dual algorithm along with corresponding convergence rate results and, for both the deterministic and the stochastic setting, convergence rates when either $\mathcal{X}$ or $\mathcal{Y}$ is unbounded have been also provided.

Later on, Ouyang, Chen, Lan and Pasiliao Jr. have developed in [23] an accelerated ADMM algorithm for the optimization problem (1.6) with f assumed to be Fréchet differentiable with L-Lipschitz continuous gradient, for $L > 0$, on its effective domain. In the case when f and g^* have bounded domains this method has been proved to exhibit an ergodic convergence rate for the objective function value of type (1.7), with the coefficient of $1/k^2$ depending on L and the diameter of $\mathrm{dom} f$ and the coefficient of $1/k$ depends on $\|A\|$ and of the diameter of $\mathrm{dom} g^*$. On the other hand, without assuming boundedness for the domains of f and g^*, the accelerated ADMM algorithms has been proved to exhibit ergodic convergence rates for the feasibility measure and the objective function value of $\mathcal{O}\left(1/k\right)$ as $k \to +\infty$.

By using a smoothing approach, Tran-Dinh, Fercoq and Cevher have designed in [26] a primal-dual algorithm for solving (1.6) and its particular formulation (1.1) that exhibits last iterates convergence rates for the objective function value and the feasibility measure in the convex regime of $\mathcal{O}\left(1/k\right)$, and in the strongly convex regime of $\mathcal{O}\left(1/k^2\right)$ as $k \to +\infty$.

Goldstein, O'Donoghue, Setzer and Baraniuk have studied in [13] the two-block separable optimization problem with linear constraints

$$\begin{aligned} \min \quad & f(x) + h(y), \\ \text{subject to} \quad & Ax + By = b \end{aligned} \tag{1.8}$$

where $\mathcal{K}$ is another real Hilbert space, $f \colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ and $h \colon \mathcal{K} \to \mathbb{R} \cup \{+\infty\}$ are proper, convex and lower semicontinuous functions, $A \colon \mathcal{H} \to \mathcal{G}$ and $B \colon \mathcal{K} \to \mathcal{G}$ are continuous linear operators and $b \in \mathcal{G}$. It is obvious that (1.1) can be reformulated as (1.8) and vice versa. In

[13] a numerical algorithm for solving (1.8) has been proposed that exhibits, when f and h are strongly convex, convergence rates for the dual objective function of $\mathcal{O}\left(1/k^2\right)$ and for the feasibility measure of $\mathcal{O}\left(1/k\right)$ as $k \to +\infty$. For a fast version of the Alternating Minimization Algorithm (see [27]) a convergence rate for the dual objective function of $\mathcal{O}\left(1/k^2\right)$ as $k \to +\infty$ has been also proved.

Xu has proposed in [28] a linearized Augmented Lagrangian Method for the optimization problem (1.1) for which he has shown that it exhibits for constant step sizes ergodic convergence rates of $\mathcal{O}\left(1/k\right)$ as $k \to +\infty$ for the feasibility measure and the objective function value, whereas the sequence of primal-dual iterates has been shown to converge to a primal-dual solution. He has also proved that for appropriately chosen variable step sizes, in particular when allowing the dual step sizes to be unbounded, the convergence rates of the feasibility measure and the objective function value can be improved to $\mathcal{O}\left(1/k^2\right)$ as $k \to +\infty$, without saying anything about the convergence of the primal-dual iterates in this setting. In addition, a linearized Alternating Direction Method of Multipliers for (1.8) has been proposed in [28], for which similar statements as for the linearized Augmented Lagrangian Method have been proved, whereby the fast convergence rates have been obtained by assuming that one of the summands in the objective function is strongly convex.

In [14], He and Yuan have enhanced the Augmented Lagrangian Method for the linearly constrained convex optimization problem (1.1) with a Nesterov's momentum update rule for the sequence of dual iterates. They have proved that the expression $\mathcal{L}(x_*, \lambda_*) - \mathcal{L}(x_k, \lambda_k)$ has an upper bound of order $1/k^2$, where $(x_k, \lambda_k)_{k \geqslant 0}$ denotes the generated sequence of primal-dual iterates and (x_*, λ_*) is an arbitrary optimal solution of the Wolfe dual problem of (1.1).

In [29], Yan and He have proposed for optimization problems of type (1.1), with a proper, convex and lower semicontinuous objective function, a numerical algorithm which combines the Augmented Lagrangian Method with a Bregman proximal evaluation of the objective. When choosing the sequence of proximal parameter to fulfil $\eta_k := \eta(k+1)^p$ for every $k \geqslant 0$, where $\eta > 0$ and $p \geqslant 0$, ergodic convergence rates of

$$\sup_{(x,\lambda) \in \mathcal{X} \times \mathcal{Y}} \left(\mathcal{L}\left(\bar{x}_k, \lambda\right) - \mathcal{L}\left(x, \bar{\lambda}_k\right)\right) = \mathcal{O}\left(\frac{1}{k^{p+2}}\right) \text{ as } k \to +\infty,$$

$$\|A\bar{x}_k - b\| = \mathcal{O}\left(\frac{\log(k)}{k^{p+2}}\right) \text{ and } |f(\bar{x}_k) - f_*| = \mathcal{O}\left(\frac{\log(k)}{k^{p+2}}\right) \text{ as } k \to +\infty$$

have been obtained.

In [24], Sabach and Teboulle have considered a unified algorithmic framework for proving faster convergence rates for various Lagrangian-based methods designed to solve optimization problems of type (1.1) with a proper, convex and lower semicontinuous objective function. In the convex regime these methods exhibit a non-ergodic rate of convergence of $\mathcal{O}\left(1/k\right)$ as $k \to +\infty$ for the feasibility measure and the objective function value, namely,

$$f(x_k) - f_* \text{ has an upper bound of order } \mathcal{O}\left(\frac{1}{k}\right) \text{ and } \|Ax_k - b\| = \mathcal{O}\left(\frac{1}{k}\right) \text{ as } k \to +\infty.$$

In the strongly convex regime the convergence rates can be improved to $\mathcal{O}\left(1/k^2\right)$ as $k \to +\infty$.

For the same class of optimization problems, He, Hu, and Fang have proposed in [15] an accelerated primal-dual Lagrangian-based method, with inertial parameters following the choice of Chambolle-Dossal, that achieves a convergence rate of $\mathcal{O}\left(1/k^2\right)$ as $k \to +\infty$ for the feasibility measure and the objective function value without any strong convexity assumption.

Recently, in [19], Lou have introduced in the same context an unifying algorithmic scheme which covers both the convex and the strongly convex setting. In the convex regime a convergence rate of $\mathcal{O}\left(1/k\right)$ as $k \to +\infty$ is obtained for the primal-dual gap, the feasibility measure, and the objective function value, while in the strongly convex regime these rates are improved

to $\mathcal{O}\left(1/k^2\right)$ as $k \to +\infty$. These results have been extended to optimization problems of type (1.8) in [20], where it has been shown that, in order to achieve a convergence rate of $\mathcal{O}\left(1/k^2\right)$ as $k \to +\infty$, it is enough to assume that only one of the functions in the objective is strongly convex.

Noticeably none of theses works has addressed to convergence of the sequences of primal-dual iterates, with very few exceptions in the strongly convex regime. This phenomenon could be noticed for unconstrained convex optimization problems, too. The convergence of the sequences of iterates generated by fast numerical methods has been proved much later (by Chambolle and Dossal in [10] and by Attouch and Peypouquet in [3]) after the derivation of the convergence rates for Nesterov's accelerated gradient method ([21]) and FISTA ([6]). One explanation for this is that the analysis of the first is much more involved.

1.4   Our contributions

We consider as starting point a second-order dynamical system with asymptotic vanishing damping term associated with the optimization problem (1.1). This dynamical system is formulated in terms of the augmented Lagrangian and it has been studied in [8]. By an appropriate time discretization this system gives rise to an inertial primal-dual numerical algorithm, which allows a flexible choice of the inertial parameters. This choice covers the three classical inertial parameters rules by Nesterov ([6, 21]), Chambolle-Dossal ([10]) and Attouch-Cabot ([1]) used in the literature to formulate fast gradient methods. We show that for these rules the resulting algorithm exhibits in the convex regime convergence rates of order $\mathcal{O}\left(1/k^2\right)$ for the primal-dual gap, the feasibility measure, and the objective function value. In addition, we prove that the generated sequence of primal-dual iterates converges weakly to a primal-dual solution of the underlying problem, which is nothing else than a saddle-point of the Lagrangian. The convergence of the iterates is stated in a general setting that covers the inertial parameters rules by Chambolle-Dossal and Attouch-Cabot. This is the first result which provides the convergence of the sequence of iterates generated by a fast algorithm for linearly constrained convex optimization problems without additional assumptions such as strong convexity. All convergence and convergence rate results of this paper are compatible with the ones obtained in [8] in the continuous setting.

The proposed Fast Augmented Lagrangian Method and all convergence results can be easily extended by using the product space approach to two-block separable linearly constrained optimization problems of the form (1.8) with f and h convex and Fréchet differentiable functions with Lipschitz continuous gradients.

1.5   Notations and preliminaries

We denote by $\mathbb{B}\left(x;\varepsilon\right) := \{y \in \mathcal{H} \colon \|x - y\| \leqslant \varepsilon\}$ the closed ball centered at $x \in \mathcal{H}$ with radius $\varepsilon > 0$.

Let $x, y \in \mathcal{H}$. We have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle. \tag{1.9}$$

For every $s, t \in \mathbb{R}$ such that $s + t = 1$ it holds ([5, Corollary 2.15])

$$\|sx + ty\|^2 = s\|x\|^2 + t\|y\|^2 - st\|x - y\|^2. \tag{1.10}$$

From here one can easily deduce that for $s, t \in \mathbb{R}$ such that $s + t \neq 0$ it holds

$$\frac{1}{s + t}\|sx + ty\|^2 = s\|x\|^2 + t\|y\|^2 - \frac{st}{s + t}\|x - y\|^2. \tag{1.11}$$

We denote by $\mathbb{S}_+(\mathcal{H})$ the family of self-adjoint and positive semidefinite continuous linear operators $\mathcal{W}\colon \mathcal{H} \to \mathcal{H}$. Every $\mathcal{W} \in \mathbb{S}_+(\mathcal{H})$ induces on $\mathcal{H}$ a semi-norm defined by

$$\|x\|_{\mathcal{W}}^2 = \langle x, x\rangle_{\mathcal{W}} := \langle \mathcal{W}x, x\rangle \qquad \forall x \in \mathcal{H}.$$

The Loewner partial ordering on $\mathbb{S}_+(\mathcal{H})$ is defined for $\mathcal{W}, \mathcal{W}' \in \mathbb{S}_+(\mathcal{H})$ as

$$\mathcal{W} \succcurlyeq \mathcal{W}' \Leftrightarrow \|x\|_{\mathcal{W}}^2 \geqslant \|x\|_{\mathcal{W}'}^2 \qquad \forall x \in \mathcal{H}.$$

Thus $\mathcal{W} \in \mathbb{S}_+(\mathcal{H})$ is nothing else than $\mathcal{W} \succcurlyeq 0$. If there exists $\alpha > 0$ such that $\mathcal{W} \succcurlyeq \alpha \mathrm{Id}$ then the semi-norm $\|\cdot\|_{\mathcal{W}}$ becomes a norm.

In the spirit of (1.9) and (1.11), respectively, for every $x, y \in \mathcal{H}$ it holds

$$\|x + y\|_{\mathcal{W}}^2 = \|x\|_{\mathcal{W}}^2 + \|y\|_{\mathcal{W}}^2 + 2\langle x, y\rangle_{\mathcal{W}}, \tag{1.12}$$

and for every real numbers s, t such that $s + t \neq 0$

$$\frac{1}{s+t}\|sx + ty\|_{\mathcal{W}}^2 = s\|x\|_{\mathcal{W}}^2 + t\|y\|_{\mathcal{W}}^2 - \frac{st}{s+t}\|x - y\|_{\mathcal{W}}^2. \tag{1.13}$$

Let $f\colon \mathcal{H} \to \mathbb{R}$ be a continuously differentiable and convex function such that ∇f is L−Lipschitz continuous, for $L > 0$. For every $x, y \in \mathcal{H}$ it holds (see [22, Theorem 2.1.5] or [5, Theorem 18.15])

$$0 \leqslant \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leqslant f(x) - f(y) - \langle \nabla f(y), x - y\rangle \leqslant \frac{L}{2}\|x - y\|^2. \tag{1.14}$$

The second inequality is also known as the Descent Lemma.

The following result is a particular instance of [5, Lemma 5.31] and will be used several times in this paper.

Lemma 1.1. *Let $\{a_k\}_{k\geqslant 1}$, $\{b_k\}_{k\geqslant 1}$ and $\{d_k\}_{k\geqslant 1}$ be sequences of real numbers. Assume that $\{a_k\}_{k\geqslant 1}$ is bounded from below, and $\{b_k\}_{k\geqslant 1}$ and $\{d_k\}_{k\geqslant 1}$ are nonnegative such that $\sum_{k\geqslant 1} d_k < +\infty$. Suppose further that for every $k \geqslant 1$ it holds*

$$a_{k+1} \leqslant a_k - b_k + d_k. \tag{1.15}$$

Then the following statements are true

(i) *the sequence $\{b_k\}_{k\geqslant 1}$ is summable, namely $\sum_{k\geqslant 1} b_k < +\infty$;*

(ii) *the sequence $\{a_k\}_{k\geqslant 1}$ is convergent.*

In order to establish the weak convergence of the iterates, we will use Opial's Lemma in discrete form (see, for instance, [5, Theorem 5.5]), which we recall as follows.

Lemma 1.2. *Let $\mathcal{C}$ be a nonempty subset of $\mathcal{H}$ and $\{x_k\}_{k\geqslant 1}$ a sequence in $\mathcal{H}$. Assume that*

(i) *for every $x_* \in \mathcal{C}$, $\lim_{k\to+\infty} \|x_k - x_*\|$ exists;*

(ii) *every weak sequential cluster point of $\{x_k\}_{k\geqslant 1}$ belongs to $\mathcal{C}$.*

Then the sequence $\{x_k\}_{k\geqslant 1}$ converges weakly to an element in $\mathcal{C}$ as $k \to +\infty$.

2   Continuous time approaches and their discrete counterparts

In this section we want to derive by time discretization a primal-dual numerical algorithm from the second-order dynamical system investigated in [8]. The employed discretization technique replicates the one used when relating fast gradient algorithms with the second-order dynamical system proposed by Su, Boyd and Candès in [25] in the unconstrained case.

2.1 The primal-dual dynamical approach with vanishing damping

The second-order primal-dual dynamical system with asymptotically vanishing damping term associated in [8] with the augmented Lagrangian formulation of (1.1) reads

$$
\begin{cases}
\ddot{x}(t) + \dfrac{\alpha}{t}\dot{x}(t) + \nabla_x \mathcal{L}_\beta\Big(x(t), \lambda(t) + \theta t \dot{\lambda}(t)\Big) &= 0 \\[2mm]
\ddot{\lambda}(t) + \dfrac{\alpha}{t}\dot{\lambda}(t) - \nabla_\lambda \mathcal{L}_\beta\Big(x(t) + \theta t \dot{x}(t), \lambda(t)\Big) &= 0\,, \qquad \text{(PD-AVD)} \\[2mm]
\Big(x(t_0), \lambda(t_0)\Big) = \Big(x_0, \lambda_0\Big) \text{ and } \Big(\dot{x}(t_0), \dot{\lambda}(t_0)\Big) = \Big(\dot{x}_0, \dot{\lambda}_0\Big)
\end{cases}
$$

where $t_0 > 0, \alpha \geqslant 3, \beta \geqslant 0, \theta > 0$ and $(x_0, \lambda_0), \Big(\dot{x}_0, \dot{\lambda}_0\Big) \in \mathcal{H} \times \mathcal{G}$.

Plugging the expressions of the partial gradients of $\mathcal{L}_\beta$ into the system leads to the following formulation for (PD-AVD)

$$
\begin{cases}
\ddot{x}(t) + \dfrac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) + A^*\Big(\lambda(t) + \theta t \dot{\lambda}(t)\Big) + \beta A^*\Big(Ax(t) - b\Big) &= 0 \\[2mm]
\ddot{\lambda}(t) + \dfrac{\alpha}{t}\dot{\lambda}(t) - \Big(A\big(x(t) + \theta t \dot{x}(t)\big) - b\Big) &= 0\,. \qquad (2.1) \\[2mm]
\Big(x(t_0), \lambda(t_0)\Big) = \Big(x_0, \lambda_0\Big) \text{ and } \Big(\dot{x}(t_0), \dot{\lambda}(t_0)\Big) = \Big(\dot{x}_0, \dot{\lambda}_0\Big)
\end{cases}
$$

In [8] it has been shown that, supposing that

$$
\alpha \geqslant 3, \quad \beta \geqslant 0 \quad \text{and} \quad \frac{1}{2} \geqslant \theta \geqslant \frac{1}{\alpha - 1},
$$

for a solution $(x, \lambda) : [t_0, +\infty) \to \mathcal{H} \times \mathcal{G}$ of (PD-AVD) and $(x_*, \lambda_*) \in \mathcal{S}$ it holds for every $t \geqslant t_0$

$$
0 \leqslant \mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) + \|Ax(t) - b\| \leqslant \frac{\widehat{C}}{\theta^2 t^2}
$$

and

$$
-\frac{\|\lambda_*\| \widehat{C}}{\theta^2 t^2} \leqslant f(x(t)) - f_* \leqslant \frac{(1 + \|\lambda_*\|)\widehat{C}}{\theta^2 t^2},
$$

where $\widehat{C} > 0$.

If, in addition, ∇f is $L-$Lipschitz continuous, $\alpha > 3$ and $\dfrac{1}{2} > \theta > \dfrac{1}{\alpha - 1}$, then it holds

$$
\|A^*(\lambda(t) - \lambda_*)\| = o\left(\frac{1}{\sqrt{t}}\right) \quad \text{and} \quad \|\nabla f(x(t)) - \nabla f(x_*)\| = o\left(\frac{1}{\sqrt{t}}\right) \text{ as } t \to +\infty \quad (2.2)
$$

and, consequently,

$$
\big\|\nabla_x \mathcal{L}\big(x(t), \lambda(t)\big)\big\| = \|\nabla f(x(t)) + A^*\lambda(t)\| = o\left(\frac{1}{\sqrt{t}}\right) \qquad \text{as } t \to +\infty,
$$

whereas

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

By additionally requiring that $\beta > 0$, it has been also proved in [8] that the trajectory $\big(x(t), \lambda(t)\big)$ converges weakly to a primal-dual optimal solution of (1.1) as $t \to +\infty$.

2.2 Fast gradient scheme: from continuous to discrete time

We recall in this section for reader's convenience the connection between the second-order dynamical system by Su, Boyd and Candès ([25]) and the fast gradient numerical methods formulated in [10, 1] in the spirit of Nesterov's accelerated gradient algorithm ([21]). To this end we consider the unconstrained optimization problem

$$\min_{x \in \mathcal{H}} f(x), \tag{2.3}$$

where $f : \mathcal{H} \to \mathbb{R}$ is a convex and Fréchet differentiable function with L-Lipschitz continuous gradient, for $L > 0$.

The continuous time approach proposed in [25] in connection with this optimization problem reads

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

where $t_0 > 0$ and $\alpha \geqslant 3$. One can easily notice that for $A = 0$ and $b = 0$ the optimization problem (1.1) becomes (2.3), while (PD-AVD) reduces to (AVD).

For every $t \geqslant t_0$, we define

$$z(t) := x(t) + \frac{t}{\alpha - 1}\dot{x}(t).$$

This leads to

$$\dot{z}(t) = \dot{x}(t) + \frac{1}{\alpha - 1}\dot{x}(t) + \frac{t}{\alpha - 1}\ddot{x}(t) = \frac{t}{\alpha - 1}\ddot{x}(t) + \frac{\alpha}{\alpha - 1}\dot{x}(t) = -\frac{t}{\alpha - 1}\nabla f(x(t))$$

and (AVD) can be written as a first-order ordinary differential equation

$$\begin{cases} \dot{z}(t) & = -\dfrac{t}{\alpha - 1}\nabla f(x(t)) \\ z(t) & = x(t) + \dfrac{t}{\alpha - 1}\dot{x}(t). \end{cases} \tag{2.4}$$

Let $\sigma > 0$. For every $k \geqslant 1$ we take as time step

$$\sigma_k := \sigma\left(1 + \frac{\alpha - 1}{k}\right) > 0,$$

and set $\tau_k := \sqrt{\sigma_k}k = \sqrt{\sigma k(k + \alpha - 1)} \approx \sqrt{\sigma}(k + 1)$, $x(\tau_k) \approx x_{k+1}$ and $z(\tau_k) \approx z_{k+1}$. We "approximate" τ_k with $\sqrt{\sigma}(k + 1)$ since it is closer to this value than to $\sqrt{\sigma}k$. This also explains why we consider $x(\tau_k) \approx x_{k+1}$ and $z(\tau_k) \approx z_{k+1}$ instead of the seemingly more natural choices $x(\tau_k) \approx x_k$ and $z(\tau_k) \approx z_k$, respectively.

The implicit finite-difference scheme for (2.4) at time $t := \tau_k$ gives

$$\begin{cases} \dfrac{z_{k+1} - z_k}{\sqrt{\sigma_k}} & = -\dfrac{\sqrt{\sigma_k}k}{\alpha - 1}\nabla f(y_k) \\ z_{k+1} & = x_{k+1} + \dfrac{\sqrt{\sigma_k}k}{\alpha - 1}\dfrac{x_{k+1} - x_k}{\sqrt{\sigma_k}} \end{cases}$$

or, equivalently,

$$\begin{cases} z_{k+1} - z_k & = -\sigma\left(1 + \dfrac{\alpha - 1}{k}\right)\dfrac{k}{\alpha - 1}\nabla f(y_k) \\ z_{k+1} & = x_{k+1} + \dfrac{k}{\alpha - 1}(x_{k+1} - x_k), \end{cases} \tag{2.5}$$

9

where the gradient ∇f is evaluated at the point y_k, which is to be determined as a suitable convex combination of x_k and z_k such that $x_{k+1} - y_k \to 0$ as $k \to +\infty$. Notice that, since ∇f is $L-$Lipschitz continuous, this implies that $\nabla f(x_{k+1}) - \nabla f(y_k) \to 0$ as $k \to +\infty$.

The second equation in (2.5) is equivalent with

$$x_{k+1} = \frac{\alpha - 1}{k + \alpha - 1} z_{k+1} + \frac{k}{k + \alpha - 1} x_k$$

and consequently suggests the following choice for y_k

$$y_k = \frac{\alpha - 1}{k + \alpha - 1} z_k + \frac{k}{k + \alpha - 1} x_k. \tag{2.6}$$

From the second equation in (2.5) we further obtain

$$y_k = \frac{\alpha - 1}{k + \alpha - 1} z_k + \frac{k}{k + \alpha - 1} x_k = \frac{\alpha - 1}{k + \alpha - 1} \left(x_k + \frac{k - 1}{\alpha - 1} (x_k - x_{k-1}) \right) + \frac{k}{k + \alpha - 1} x_k$$

$$= x_k + \frac{k - 1}{k + \alpha - 1} (x_k - x_{k-1}).$$

In addition,

$$z_{k+1} - z_k = \frac{k + \alpha - 1}{\alpha - 1} (x_{k+1} - y_k) = \left(1 + \frac{k}{\alpha - 1} \right) (x_{k+1} - y_k).$$

Consequently, (2.5) can be equivalently written as

$$\begin{cases} y_k & := x_k + \dfrac{k - 1}{k + \alpha - 1} (x_k - x_{k-1}) \\ x_{k+1} & := y_k - \sigma \nabla f(y_k). \end{cases} \tag{2.7}$$

This is nothing else than the algorithm considered by Chambolle and Dossal in [10] (see also [3]).

Furthermore, if we write for every $k \geqslant 1$

$$t_k := 1 + \frac{k - 1}{\alpha - 1} = \frac{k + \alpha - 2}{\alpha - 1}, \tag{2.8}$$

so that

$$t_{k+1} - 1 = \frac{k}{\alpha - 1} \qquad \text{and} \qquad \frac{t_k - 1}{t_{k+1}} = \frac{k - 1}{k + \alpha - 1},$$

then (2.7) becomes

$$(\forall k \geqslant 1) \quad \begin{cases} y_k & := x_k + \dfrac{t_k - 1}{t_{k+1}} (x_k - x_{k-1}) \\ x_{k+1} & := y_k - \sigma \nabla f(y_k). \end{cases} \tag{2.9}$$

Modifications of the sequence $\{t_k\}_{k \geqslant 1}$ which preserve its asymptotic behaviour lead to various acceleration schemes from the literature.

For instance, the classical Nesterov's accelerated gradient method ([21]) is precisely (2.9), where the sequence $\{t_k\}_{k \geqslant 1}$ satisfies the recurrence rule

$$t_1 := 1 \qquad \text{and} \qquad t_{k+1} := \frac{1 + \sqrt{1 + 4t_k^2}}{2} \qquad \forall k \geqslant 1. \tag{2.10}$$

Another example is the algorithm proposed by Attouch and Cabot in [1] that corresponds to (2.9) with the choice

$$t_k := \frac{k - 1}{\alpha - 1} \qquad \forall k \geqslant 1. \tag{2.11}$$

It can also be interpreted as a discretization of (2.4) with time step

$$\sigma_k := \frac{\sigma k}{k - \alpha + 1} \quad \forall k \geqslant 1,$$

and by setting $\tau_k := \sqrt{\sigma_k}\,(k - \alpha + 1) = \sqrt{\sigma k\,(k - \alpha + 1)} \approx \sqrt{\sigma}\,(k + 1)$, $x\,(\tau_k) \approx x_{k+1}$ and $z\,(\tau_k) \approx z_{k+1}$.

2.3 The time discretization of (PD-AVD)

In order to provide a useful time discretization of the dynamical system (PD-AVD) we follow the approach of the previous section and define for every $t \geqslant t_0$

$$z\,(t) := x\,(t) + \frac{t}{\alpha - 1}\dot{x}\,(t) \qquad \text{and} \qquad \nu\,(t) := \lambda\,(t) + \frac{t}{\alpha - 1}\dot{\lambda}\,(t). \tag{2.12}$$

Further, we set

$$\gamma := \frac{1}{\theta\,(\alpha - 1)} \in \left[\frac{2}{\alpha - 1}, 1\right]. \tag{2.13}$$

The parameter γ will play an essential role in our analysis. For every $t \geqslant t_0$ we define

$$z^\gamma\,(t) := \gamma\,(x\,(t) + \theta t\dot{x}\,(t)) = \gamma x\,(t) + \frac{t}{\alpha - 1}\dot{x}\,(t) = z\,(t) + (\gamma - 1)\,x\,(t), \tag{2.14a}$$

$$\nu^\gamma\,(t) := \gamma\left(\lambda\,(t) + \theta t\dot{\lambda}\,(t)\right) = \gamma\lambda\,(t) + \frac{t}{\alpha - 1}\dot{\lambda}\,(t) = \nu\,(t) + (\gamma - 1)\,\lambda\,(t). \tag{2.14b}$$

Using these notations, the system (PD-AVD) (see also its equivalent formulation (2.1)) can be written as

$$\begin{cases} \ddot{x}\,(t) + \dfrac{\alpha}{t}\dot{x}\,(t) + \nabla f\,(x\,(t)) + \dfrac{1}{\gamma}A^*\nu^\gamma\,(t) + \beta A^*\left(Ax\,(t) - b\right) & = 0 \\[2mm] \ddot{\lambda}\,(t) + \dfrac{\alpha}{t}\dot{\lambda}\,(t) - \dfrac{1}{\gamma}\left(Az^\gamma\,(t) - \gamma b\right) & = 0 \,. \\[2mm] \left(x\,(t_0), \lambda\,(t_0)\right) = \left(x_0, \lambda_0\right) \text{ and } \left(\dot{x}\,(t_0), \dot{\lambda}\,(t_0)\right) = \left(\dot{x}_0, \dot{\lambda}_0\right) \end{cases} \tag{2.15}$$

Using that for every $t \geqslant t_0$

$$\dot{z}\,(t) = \frac{t}{\alpha - 1}\left(\ddot{x}\,(x) + \frac{\alpha}{t}\dot{x}\,(t)\right) \text{ and } \dot{\nu}\,(t) = \frac{t}{\alpha - 1}\left(\ddot{\lambda}\,(x) + \frac{\alpha}{t}\dot{\lambda}\,(t)\right),$$

the first two lines in (2.15) can be equivalently written as

$$\begin{cases} \dot{z}\,(t) = -\dfrac{t}{\alpha - 1}\nabla f\,(x\,(t)) - \dfrac{t}{\alpha - 1}\dfrac{1}{\gamma}A^*\nu^\gamma\,(t) - \dfrac{t}{\alpha - 1}\beta A^*\left(Ax\,(t) - b\right) \\[2mm] \dot{\nu}\,(t) = \dfrac{1}{\gamma}\dfrac{t}{\alpha - 1}\left(Az^\gamma\,(t) - \gamma b\right) \\[2mm] z\,(t) = x\,(t) + \dfrac{t}{\alpha - 1}\dot{x}\,(t) \\[2mm] z^\gamma\,(t) = \gamma x\,(t) + \dfrac{t}{\alpha - 1}\dot{x}\,(t) \\[2mm] \nu\,(t) = \lambda\,(t) + \dfrac{t}{\alpha - 1}\dot{\lambda}\,(t) \\[2mm] \nu^\gamma\,(t) = \gamma\lambda\,(t) + \dfrac{t}{\alpha - 1}\dot{\lambda}\,(t) \,. \end{cases} \tag{2.16}$$

Let $\sigma, \rho > 0$. For every $k \geqslant 1$ we take for x and λ two different time steps

$$\sigma_k := \sigma\left(1 + \frac{\alpha - 1}{k}\right) > 0 \qquad \text{and} \qquad \rho_k := \rho\left(1 + \frac{\alpha - 1}{k}\right) > 0,$$

respectively, and set $\tau_k := \sqrt{\sigma_k}k \approx \sqrt{\sigma}\left(k+1\right), x\left(\tau_k\right) \approx x_{k+1}, z\left(\tau_k\right) \approx z_{k+1}, z^\gamma\left(\tau_k\right) \approx z^\gamma_{k+1}$, and $\tau'_k := \sqrt{\rho_k}k \approx \sqrt{\rho}\left(k+1\right), \lambda\left(\tau'_k\right) \approx \lambda_{k+1}, \nu\left(\tau'_k\right) \approx \nu_{k+1}$ and $\nu^\gamma\left(\tau'_k\right) \approx \nu^\gamma_{k+1}$.

The implicit finite-difference scheme for (2.16) at time $t := \tau_k$ for x and time $t := \tau'_k$ for λ gives

$$
\begin{cases}
\dfrac{z_{k+1} - z_k}{\sqrt{\sigma_k}} &= -\dfrac{\sqrt{\sigma_k}k}{\alpha-1}\nabla f\left(y_k\right) - \dfrac{\sqrt{\sigma_k}k}{\alpha-1}A^*\dfrac{1}{\gamma}\widetilde{\nu}_{k+1} - \dfrac{\sqrt{\sigma_k}k}{\alpha-1}\beta A^*\left(Ay_k - b\right) \\[3mm]
\dfrac{\nu_{k+1} - \nu_k}{\sqrt{\rho_k}} &= \dfrac{1}{\gamma}\dfrac{\sqrt{\rho_k}k}{\alpha-1}\left(Az^\gamma_{k+1} - \gamma b\right) \\[3mm]
z_{k+1} &= x_{k+1} + \dfrac{\sqrt{\sigma_k}k}{\alpha-1}\dfrac{x_{k+1} - x_k}{\sqrt{\sigma_k}} \\[3mm]
z^\gamma_{k+1} &= \gamma x_{k+1} + \dfrac{\sqrt{\sigma_k}k}{\alpha-1}\dfrac{x_{k+1} - x_k}{\sqrt{\sigma_k}} \\[3mm]
\nu_{k+1} &= \lambda_{k+1} + \dfrac{\sqrt{\sigma_k}k}{\alpha-1}\dfrac{\lambda_{k+1} - \lambda_k}{\sqrt{\sigma_k}} \\[3mm]
\nu^\gamma_{k+1} &= \gamma\lambda_{k+1} + \dfrac{\sqrt{\sigma_k}k}{\alpha-1}\dfrac{\lambda_{k+1} - \lambda_k}{\sqrt{\sigma_k}},
\end{cases}
\tag{2.17}
$$

where y_k and $\widetilde{\nu}_{k+1}$ will be chosen appropriately to obtain an easily implementable iterative scheme. Notice that $\widetilde{\nu}_{k+1}$ must be an approximation of ν^γ_{k+1}.

Once again we take as in the previous section (see (2.6))

$$
y_k = \frac{\alpha-1}{k+\alpha-1}z_k + \frac{k}{k+\alpha-1}x_k = x_k + \frac{k-1}{k+\alpha-1}\left(x_k - x_{k-1}\right),
$$

which, by using the third equation in (2.17), gives

$$
z_{k+1} - z_k = \frac{k+\alpha-1}{\alpha-1}\left(x_{k+1} - y_k\right) = \left(1 + \frac{k}{\alpha-1}\right)\left(x_{k+1} - y_k\right).
$$

Following (2.6) we set also for the sequence of dual variables

$$
\mu_k = \frac{\alpha-1}{k+\alpha-1}\nu_k + \frac{k}{k+\alpha-1}\lambda_k = \lambda_k + \frac{k-1}{k+\alpha-1}\left(\lambda_k - \lambda_{k-1}\right),
$$

which, by using the fifth equation in (2.17), gives

$$
\nu_{k+1} - \nu_k = \frac{k+\alpha-1}{\alpha-1}\left(\lambda_{k+1} - \mu_k\right) = t_{k+1}\left(\lambda_{k+1} - \mu_k\right).
\tag{2.18}
$$

For these choices, and by taking into consideration the definition of $\{t_k\}_{k\geqslant 1}$ in (2.8), (2.17) becomes

$$
\begin{cases}
y_k &= x_k + \dfrac{t_k - 1}{t_{k+1}}\left(x_k - x_{k-1}\right) \\[3mm]
x_{k+1} &= y_k - \sigma\nabla f\left(y_k\right) - \dfrac{\sigma}{\gamma}A^*\widetilde{\nu}_{k+1} - \sigma\beta A^*\left(Ay_k - b\right) \\[3mm]
\mu_k &= \lambda_k + \dfrac{t_k - 1}{t_{k+1}}\left(\lambda_k - \lambda_{k-1}\right) \\[3mm]
z^\gamma_{k+1} &= \gamma x_{k+1} + \left(t_{k+1} - 1\right)\left(x_{k+1} - x_k\right) \\[3mm]
\lambda_{k+1} &= \mu_k + \dfrac{\rho}{\gamma}\left(Az^\gamma_{k+1} - \gamma b\right) \\[3mm]
\nu^\gamma_{k+1} &= \gamma\lambda_{k+1} + \left(t_{k+1} - 1\right)\left(\lambda_{k+1} - \lambda_k\right),
\end{cases}
\tag{2.19}
$$

where $\widetilde{\nu}_{k+1}$ is still to be chosen such that $\widetilde{\nu}_{k+1} - \nu_{k+1}^{\gamma} \to 0$ as $k \to +\infty$. We will not opt for $\widetilde{\nu}_{k+1} = \nu_{k+1}^{\gamma}$ in order to avoid an implicit iterative scheme, but choose instead (see also (2.18))

$$\widetilde{\nu}_{k+1} := \nu_{k+1}^{\gamma} + (1-\gamma)(\lambda_{k+1} - \lambda_k) = \gamma\lambda_{k+1} + (t_{k+1} - \gamma)(\lambda_{k+1} - \lambda_k) = \gamma\lambda_k + t_{k+1}(\lambda_{k+1} - \lambda_k)$$

$$= \gamma\lambda_k + (t_k - 1)(\lambda_k - \lambda_{k-1}) + t_{k+1}\left(\lambda_{k+1} - \lambda_k - \frac{t_k - 1}{t_{k+1}}(\lambda_k - \lambda_{k-1})\right)$$

$$= \nu_k^{\gamma} + t_{k+1}(\lambda_{k+1} - \mu_k) = \nu_k^{\gamma} + \nu_{k+1} - \nu_k = \nu_k^{\gamma} + \frac{\rho}{\gamma}t_{k+1}\left(Az_{k+1}^{\gamma} - \gamma b\right)$$

$$= \nu_k^{\gamma} + \frac{\rho}{\gamma}t_{k+1}\left((t_{k+1} - 1 + \gamma)Ax_{k+1} - (t_{k+1} - 1)Ax_k - \gamma b\right)$$

$$= \nu_k^{\gamma} + \frac{\rho}{\gamma}t_{k+1}(t_{k+1} - 1 + \gamma)\left(Ax_{k+1} - \frac{1}{t_{k+1} - 1 + \gamma}\left((t_{k+1} - 1)Ax_k + \gamma b\right)\right),$$

Such a choice is reasonable as long as $\lambda_{k+1} - \lambda_k \to 0$ as $k \to +\infty$, which will then imply that $\widetilde{\nu}_{k+1} - \nu_{k+1}^{\gamma} \to 0$ as $k \to +\infty$. By setting

$$s_{k+1} := \frac{\rho}{\gamma}t_{k+1}(t_{k+1} - 1 + \gamma) \text{ and } \eta_k := \frac{1}{t_{k+1} - 1 + \gamma}\left((t_{k+1} - 1)Ax_k + \gamma b\right),$$

the second line in (2.19) becomes

$$x_{k+1} = y_k - \sigma\nabla f(y_k) - \frac{\sigma}{\gamma}A^*\left(\nu_k^{\gamma} + s_{k+1}(Ax_{k+1} - \eta_k)\right) - \sigma\beta A^*(Ay_k - b)$$

or, equivalently,

$$x_{k+1} + \frac{\sigma}{\gamma}s_{k+1}A^*Ax_{k+1} = \left(\mathrm{Id} + \frac{\sigma}{\gamma}s_{k+1}A^*A\right)x_{k+1}$$

$$= y_k - \sigma\nabla f(y_k) - \frac{\sigma}{\gamma}A^*\nu_k^{\gamma} + \frac{\sigma}{\gamma}s_{k+1}A^*\eta_k - \sigma\beta A^*(Ay_k - b).$$

After rearranging the order in which the sequences are updated, (2.19) leads to the fast Augmented Lagrangian Method which we propose in this paper, and also investigate from the point of view of its convergence properties.

3   Fast Augmented Lagrangian Method

In this section we will give a precise formulation of the Augmented Lagrangian Method for solving (1.1) and prove that it exhibits convergence rates of order $\mathcal{O}\left(1/k^2\right)$ for the primal-dual gap, the feasibility measure, and the objective function value.

3.1   The algorithm

Algorithm 1. *Let* $\beta \geqslant 0$ *and* $m, \gamma, \rho, \sigma > 0$ *be such that*

$$0 < m \leqslant \gamma \leqslant 1 \qquad and \qquad 0 < \sigma \leqslant \frac{\gamma}{L + \gamma\beta\|A\|^2}. \tag{3.1}$$

Let $\{t_k\}_{k \geqslant 1}$ *be a nondecreasing sequence such that*

$$t_1 := 1 \qquad and \qquad t_{k+1}^2 - mt_{k+1} \leqslant t_k^2 \quad \forall k \geqslant 1. \tag{3.2}$$

Given $x_0 = x_1 \in \mathcal{H}$ and $\lambda_0 = \lambda_1 \in \mathcal{G}$, for every $k \geqslant 1$ we set

$$y_k := x_k + \frac{t_k - 1}{t_{k+1}} \left(x_k - x_{k-1} \right), \tag{3.3a}$$

$$\mu_k := \lambda_k + \frac{t_k - 1}{t_{k+1}} \left(\lambda_k - \lambda_{k-1} \right), \tag{3.3b}$$

$$\eta_k := A x_k + \frac{\gamma}{t_{k+1} - 1 + \gamma} (b - A x_k), \tag{3.3c}$$

$$\nu_k^\gamma := \gamma \lambda_k + \left(t_k - 1 \right) \left(\lambda_k - \lambda_{k-1} \right), \tag{3.3d}$$

$$s_{k+1} := \frac{\rho}{\gamma} t_{k+1} \left(t_{k+1} - 1 + \gamma \right), \tag{3.3e}$$

$$x_{k+1} := \arg\min_{x \in \mathcal{H}} \left\{ \langle \nabla f \left(y_k \right) + \beta A^* \left(A y_k - b \right), x - y_k \rangle + \frac{1}{\gamma} \langle \nu_k^\gamma, A x - b \rangle \right.$$
$$\left. + \frac{1}{2\gamma} s_{k+1} \left\| A x - \eta_k \right\|^2 + \frac{1}{2\sigma} \left\| x - y_k \right\|^2 \right\}, \tag{3.3f}$$

$$z_{k+1}^\gamma := \gamma x_{k+1} + \left(t_{k+1} - 1 \right) \left(x_{k+1} - x_k \right), \tag{3.3g}$$

$$\lambda_{k+1} := \mu_k + \frac{\rho}{\gamma} \left(A z_{k+1}^\gamma - \gamma b \right). \tag{3.3h}$$

One can notice that Algorithm 1 can be written in a concise way only in terms of the sequences of primal-dual iterates $\{(x_k, \lambda_k)\}_{k \geqslant 0}$, however, this elaborated formulation using auxiliary sequences is more convenient for its analysis.

Even though the choice $\gamma = 1$ would give a simplified version of Algorithm 1, without affecting its fast convergence properties, we will see that in order to guarantee the convergence of $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ to a primal-dual optimal solution it will be crucial to choose $\gamma \in (0, 1)$. A similar phenomenon is known from the continuous and discrete schemes in the unconstrained case, where fast convergence rates have been obtained for $\alpha \geqslant 3$, while the convergence of the sequence of iterates/trajectory could be shown only for $\alpha > 3$. In view of (2.13), in order to be allowed to choose $\gamma \in (0, 1)$, one must have $\alpha > 3$.

Example 3.1. (The case $A = 0$ and $b = 0$) If $A = 0$ and $b = 0$, then (1.1) becomes the unconstrained optimization problems (2.3) and Algorithm 1 reduces to the well-known accelerated gradient scheme which, given $0 < \sigma \leqslant \frac{1}{L}$, $\{t_k\}_{k \geqslant 1}$ a nondecreasing sequence fulfilling (3.2) and $x_0 = x_1 \in \mathcal{H}$, reads for every $k \geqslant 1$

$$y_k := x_k + \frac{t_k - 1}{t_{k+1}} \left(x_k - x_{k-1} \right)$$
$$x_{k+1} := y_k - \sigma \nabla f \left(y_k \right),$$

as the dual sequence $\{\lambda_k\}_{k \geqslant 0}$ can be neglected.

Remark 3.2. By denoting for every $k \geqslant 1$

$$z_k := x_k + \left(t_k - 1 \right) \left(x_k - x_{k-1} \right) \tag{3.4a}$$

$$= x_k + t_{k+1} \left(y_k - x_k \right) \tag{3.4b}$$

$$= t_k x_k - \left(t_k - 1 \right) x_{k-1}, \tag{3.4c}$$

it yields

$$y_k = \left(1 - \frac{1}{t_{k+1}} \right) x_k + \frac{1}{t_{k+1}} z_k = \left(1 - \frac{1}{t_{k+1}} \right) x_k + \frac{1}{t_{k+1}} \left(x_k + \left(t_k - 1 \right) \left(x_k - x_{k-1} \right) \right). \tag{3.5}$$

On the other hand, (3.4c) with index $k + 1$ reads

$$z_{k+1} = t_{k+1}x_{k+1} - (t_{k+1} - 1)\, x_k, \tag{3.6}$$

which is equivalent to

$$x_{k+1} = \left(1 - \frac{1}{t_{k+1}}\right) x_k + \frac{1}{t_{k+1}} z_{k+1}. \tag{3.7}$$

Subtracting (3.5) from (3.7), we obtain

$$x_{k+1} - y_k = \frac{1}{t_{k+1}} \left(z_{k+1} - z_k\right). \tag{3.8}$$

Furthermore, by the definition of z_k^γ and z_k in (3.3g) and (3.4a), it holds

$$z_k^\gamma = z_k + (\gamma - 1)\, x_k,$$

which leads to

$$z_{k+1}^\gamma - z_k^\gamma = z_{k+1} - z_k + (\gamma - 1)\,(x_{k+1} - x_k) \tag{3.9a}$$
$$= t_{k+1}\,(x_{k+1} - y_k) + (\gamma - 1)\,(x_{k+1} - x_k). \tag{3.9b}$$

By a similar argument, denoting for every $k \geqslant 1$

$$\nu_k := \lambda_k + (t_k - 1)\,(\lambda_k - \lambda_{k-1}), \tag{3.10}$$

we can derive that

$$\nu_{k+1}^\gamma - \nu_k^\gamma = t_{k+1}\,(\lambda_{k+1} - \mu_k) + (\gamma - 1)\,(\lambda_{k+1} - \lambda_k) \text{ and } \lambda_{k+1} - \mu_k = \frac{1}{t_{k+1}}\,(\nu_{k+1} - \nu_k). \tag{3.11}$$

Example 3.3. (The choice $\gamma = 1$) In case $\gamma = 1$ we denote $z_k := z_k^1$ and $\nu_k := \nu_k^1$ for every $k \geqslant 1$, which is consistent with the notations in the remark above. Given $0 < \sigma \leqslant \dfrac{1}{L + \beta\|A\|^2}$, $\{t_k\}_{k \geqslant 1}$ a nondecreasing sequence fulfilling (3.2) $x_0 = x_1 \in \mathcal{H}$ and $\lambda_0 = \lambda_1 \in \mathcal{G}$, Algorithm 1 simplifies for every $k \geqslant 1$ to

$$y_k := x_k + \frac{t_k - 1}{t_{k+1}}\,(x_k - x_{k-1}),$$

$$\mu_k := \lambda_k + \frac{t_k - 1}{t_{k+1}}\,(\lambda_k - \lambda_{k-1}),$$

$$\eta_k := \left(1 - \frac{1}{t_{k+1}}\right) A x_k + \frac{1}{t_{k+1}} b,$$

$$\nu_k := \lambda_k + (t_k - 1)\,(\lambda_k - \lambda_{k-1}),$$

$$x_{k+1} := \arg\min_{x \in \mathcal{H}} \Big\{ \langle \nabla f(y_k) + \beta A^*(Ay_k - b), x - y_k \rangle + \langle \nu_k, Ax - b \rangle$$

$$+ \frac{1}{2} \rho t_{k+1}^2 \|Ax - \eta_k\|^2 + \frac{1}{2\sigma} \|x - y_k\|^2 \Big\},$$

$$\lambda_{k+1} := \mu_k + \rho\,(Ax_{k+1} - b + (t_{k+1} - 1)\, A\,(x_{k+1} - x_k)).$$

The fact that this iterative scheme exhibits fast convergence rates for the primal-dual gap, the feasibility measure, and the objective function value will follow from the analysis we will do for Algorithm 1. However, nothing can be said about the convergence of the primal-dual iterates. To this end we will have to assume that $\gamma \in (0, 1)$, which will be a crucial assumption.

Remark 3.4. He, Hu and Fang have considered in [15] for $\alpha > 3$, $\sigma, \sigma' > 0$ and

$$t_k := 1 + \frac{k-2}{\alpha-1} \quad \forall k \geqslant 1$$

the following iterative scheme which reads for every $k \geqslant 1$

$$y_k := x_k + \frac{t_k - 1}{t_{k+1}} \left(x_k - x_{k-1} \right),$$

$$\mu_k := \lambda_k + \frac{t_k - 1}{t_{k+1}} \left(\lambda_k - \lambda_{k-1} \right),$$

$$\eta_k := \left(1 - \frac{1}{t_{k+1}} \right) A x_k + \frac{1}{t_{k+1}} b,$$

$$\nu_k := \lambda_k + (t_k - 1) \left(\lambda_k - \lambda_{k-1} \right),$$

$$x_{k+1} := \arg\min_{x \in \mathcal{H}} \left\{ \langle \nabla f(y_k), x - y_k \rangle + \langle \nu_k, Ax - b \rangle + \frac{\sigma(t_{k+2} - 1) t_{k+1}}{2} \| Ax - \eta_k \|^2 \right.$$
$$\left. + \frac{\sigma' t_{k+1}}{\sigma(t_{k+2} - 1)} \| x - y_k \|^2 \right\},$$

$$\lambda_{k+1} := \mu_k + \frac{\sigma(t_{k+2} - 1)}{t_{k+1}} \left(A x_{k+1} - b + (t_{k+1} - 1) A \left(x_{k+1} - x_k \right) \right).$$

This algorithm differs from Algorithm 1 for the choice $\gamma = 1$ (as formulated in the above example) in the way the primal-dual iterates $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ are defined. The formulation of the first allows a more direct derivation of the fast convergence rates for the feasibility measure and the objective function value. The convergence of $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ has been not addressed in [15], and it is by far not clear whether this sequence converges.

The following lemma collects some properties of the sequence $\{t_k\}_{k \geqslant 1}$ fulfilling (3.2).

Lemma 3.5. *Let* $0 < m \leqslant 1$ *and* $\{t_k\}_{k \geqslant 1}$ *a nondecreasing sequence fulfilling*

$$t_1 := 1 \qquad and \qquad t_{k+1}^2 - m t_{k+1} \leqslant t_k^2 \quad \forall k \geqslant 1.$$

Then for every $k \geqslant 1$ *it holds*

$$t_{k+1} - t_k \leqslant \varphi_m := \frac{m - 2 + \sqrt{m^2 + 4}}{2} \leqslant \frac{\sqrt{5} - 1}{2} < 1, \tag{3.13}$$

$$t_{k+1} \leqslant (1 + \varphi_m) t_k \qquad and \qquad t_{k+1} \leqslant 1 + k\varphi_m \leqslant (1 + \varphi_m) k. \tag{3.14}$$

Proof. Let $k \geqslant 1$. From the assumption we have that

$$1 \leqslant t_{k+1} \leqslant \frac{m + \sqrt{m^2 + 4t_k^2}}{2},$$

which further gives

$$t_{k+1} - t_k \leqslant \frac{m + \sqrt{m^2 + 4t_k^2}}{2} - t_k.$$

We define the function $\psi : [1, +\infty) \to \mathbb{R}$ as $s \mapsto \dfrac{m + \sqrt{m^2 + 4s^2}}{2} - s$. Since

$$\psi'(s) = \frac{2s}{\sqrt{m^2 + 4s^2}} - 1 < 0,$$

ψ is nonincreasing, consequently

$$t_{k+1} - t_k \leqslant \psi(1) = \frac{m + \sqrt{m^2 + 4}}{2} - 1 = \varphi_m \leqslant \frac{\sqrt{5} - 1}{2}.$$

The statements in (3.14) follow from the fact that $t_k \geqslant 1$ for every $k \geqslant 1$ and $\varphi_m \geqslant 0$ and by telescoping arguments, respectively. $\qquad \square$

16

3.2 Some important estimates and an energy function

In this section we will provide some important estimates which will be useful when proving that the sequence of values of a discrete energy function, which we will associate with Algorithm 1, takes at a saddle point is nonincreasing.

Lemma 3.6. *Let $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ be the sequence generated by Algorithm 1. Then for every $x \in \mathcal{H}$ and every $k \geqslant 1$ the following inequality holds*

$$f(x_{k+1}) \leqslant f(x) - \frac{1}{\gamma} \langle \nu_{k+1}^{\gamma}, Ax_{k+1} - Ax \rangle + \frac{1}{\gamma}(1 - \gamma) \langle \lambda_k - \lambda_{k+1}, Ax_{k+1} - Ax \rangle$$

$$+ \frac{1}{\sigma} \langle y_k - x_{k+1}, x_{k+1} - x \rangle - \beta \langle Ay_k - b, Ax_{k+1} - Ax \rangle$$

$$+ \frac{L}{2} \|x_{k+1} - y_k\|^2 - \frac{1}{2L} \|\nabla f(y_k) - \nabla f(x)\|^2. \tag{3.15}$$

Proof. Let $x \in \mathcal{H}$ and $k \geqslant 1$ be fixed. According to (3.3f) we have that

$$\nabla f(y_k) + \frac{1}{\gamma} A^* \nu_k^{\gamma} + \frac{1}{\gamma} s_{k+1} A^* (Ax_{k+1} - \eta_k) + \frac{1}{\sigma}(x_{k+1} - y_k) + \beta A^* (Ay_k - b) = 0. \tag{3.16}$$

On the other hand, from (3.3c), (3.3e) and (3.3h) we have

$$\frac{1}{\gamma} s_{k+1}(Ax_{k+1} - \eta_k) = \frac{\rho}{\gamma^2} t_{k+1}((t_{k+1} - 1 + \gamma) Ax_{k+1} - (t_{k+1} - 1) Ax_k - \gamma b)$$

$$= \frac{\rho}{\gamma^2} t_{k+1}(Az_{k+1}^{\gamma} - \gamma b) = \frac{1}{\gamma} t_{k+1}(\lambda_{k+1} - \mu_k)$$

$$= \frac{1}{\gamma}(\nu_{k+1}^{\gamma} - \nu_k^{\gamma} + (1 - \gamma)(\lambda_{k+1} - \lambda_k)), \tag{3.17}$$

where the last equation follows from (3.11). Hence, replacing (3.17) in (3.16) we have

$$\nabla f(y_k) = -\frac{1}{\gamma} A^* \nu_{k+1}^{\gamma} + \frac{1}{\gamma}(1 - \gamma) A^* (\lambda_k - \lambda_{k+1}) + \frac{1}{\sigma}(y_k - x_{k+1}) - \beta A^* (Ay_k - b). \tag{3.18}$$

The Descent Lemma inequality (1.14) provides

$$f(x_{k+1}) \leqslant f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{L}{2} \|x_{k+1} - y_k\|^2$$

and

$$f(y_k) \leqslant f(x) + \langle \nabla f(y_k), y_k - x \rangle - \frac{1}{2L} \|\nabla f(y_k) - \nabla f(x)\|^2.$$

By summing up these relations it yields

$$f(x_{k+1}) \leqslant f(x) + \langle \nabla f(y_k), x_{k+1} - x \rangle + \frac{L}{2} \|x_{k+1} - y_k\|^2 - \frac{1}{2L} \|\nabla f(y_k) - \nabla f(x)\|^2$$

$$= f(x) - \frac{1}{\gamma} \langle \nu_{k+1}^{\gamma}, Ax_{k+1} - Ax \rangle + \frac{1}{\gamma}(1 - \gamma) \langle \lambda_k - \lambda_{k+1}, Ax_{k+1} - Ax \rangle$$

$$+ \frac{1}{\sigma} \langle y_k - x_{k+1}, x_{k+1} - x \rangle - \beta \langle Ay_k - b, Ax_{k+1} - Ax \rangle$$

$$+ \frac{L}{2} \|x_{k+1} - y_k\|^2 - \frac{1}{2L} \|\nabla f(y_k) - \nabla f(x)\|^2,$$

which is nothing else than (3.15). $\qquad\square$

In the following we denote

$$\mathcal{Q} := \frac{1}{\sigma}\mathrm{Id} - \beta A^* A. \tag{3.19}$$

Assumption (3.1) guarantees that $\gamma\mathcal{Q} - L\mathrm{Id} \in \mathbb{S}_+(\mathcal{H})$, as

$$\gamma\mathcal{Q} - L\mathrm{Id} = \left(\frac{\gamma}{\sigma} - L\right)\mathrm{Id} - \gamma\beta A^* A \geqslant \left(\frac{\gamma}{\sigma} - L - \gamma\beta\|A\|^2\right)\mathrm{Id}. \tag{3.20}$$

Lemma 3.7. *Let $\{(x_k, \lambda_k)\}_{k\geqslant 0}$ be the sequence generated by Algorithm 1. Then for every $(x, \lambda) \in \mathcal{F} \times \mathcal{G}$ and every $k \geqslant 1$ the following two inequalities hold*

$$\mathcal{L}_\beta(x_{k+1}, \lambda) - \mathcal{L}_\beta(x, \lambda_{k+1})$$
$$\leqslant \frac{1}{\gamma}(1-\gamma)\langle\lambda_k - \lambda_{k+1}, Ax_{k+1} - b\rangle + \langle y_k - x_{k+1}, x_{k+1} - x\rangle_\mathcal{Q} + \frac{1}{\rho}\langle\mu_k - \lambda_{k+1}, \lambda_{k+1} - \lambda\rangle$$
$$- \frac{\beta}{2}\|Ax_{k+1} - b\|^2 + \frac{L}{2}\|x_{k+1} - y_k\|^2 - \frac{1}{2L}\|\nabla f(y_k) - \nabla f(x)\|^2$$
$$- \frac{1}{\gamma}(t_{k+1} - 1)\langle\lambda_{k+1} - \lambda_k, Ax_{k+1} - b\rangle + \frac{1}{\gamma}(t_{k+1} - 1)\langle\lambda_{k+1} - \lambda, Ax_{k+1} - Ax_k\rangle, \tag{3.21}$$

and

$$\mathcal{L}_\beta(x_{k+1}, \lambda) - \mathcal{L}_\beta(x, \lambda_{k+1})$$
$$\leqslant \mathcal{L}_\beta(x_k, \lambda) - \mathcal{L}_\beta(x, \lambda_k) + \frac{1}{\gamma}(1-\gamma)\langle\lambda_k - \lambda_{k+1}, Ax_{k+1} - Ax_k\rangle + \langle y_k - x_{k+1}, x_{k+1} - x_k\rangle_\mathcal{Q}$$
$$+ \frac{1}{\rho}\langle\mu_k - \lambda_{k+1}, \lambda_{k+1} - \lambda_k\rangle - \frac{\beta}{2}\|Ax_{k+1} - Ax_k\|^2 + \frac{L}{2}\|x_{k+1} - y_k\|^2$$
$$- \frac{1}{2L}\|\nabla f(y_k) - \nabla f(x_k)\|^2 + \langle\lambda - \lambda_{k+1}, Ax_{k+1} - Ax_k\rangle + \langle\lambda_{k+1} - \lambda_k, Ax_{k+1} - b\rangle. \tag{3.22}$$

Proof. Let $x \in \mathcal{F}$, which means that $Ax = b$, $\lambda \in \mathcal{G}$, and $k \geqslant 1$ be fixed. We deduce from Lemma 3.6 that

$$f(x_{k+1}) + \langle\lambda, Ax_{k+1} - b\rangle$$
$$\leqslant f(x) + \langle\lambda_{k+1}, Ax - b\rangle + \frac{\beta}{2}\|Ax - b\|^2 + \left\langle\lambda - \frac{1}{\gamma}\nu_{k+1}^\gamma, Ax_{k+1} - b\right\rangle$$
$$+ \frac{1}{\gamma}(1-\gamma)\langle\lambda_k - \lambda_{k+1}, Ax_{k+1} - b\rangle + \frac{1}{\sigma}\langle y_k - x_{k+1}, x_{k+1} - x\rangle$$
$$- \beta\langle Ay_k - b, Ax_{k+1} - Ax\rangle + \frac{L}{2}\|x_{k+1} - y_k\|^2 - \frac{1}{2L}\|\nabla f(y_k) - \nabla f(x)\|^2$$
$$= f(x) + \langle\lambda_{k+1}, Ax - b\rangle + \frac{\beta}{2}\|Ax - b\|^2 + \left\langle\lambda - \frac{1}{\gamma}\nu_{k+1}^\gamma, Ax_{k+1} - b\right\rangle$$
$$+ \frac{1}{\gamma}(1-\gamma)\langle\lambda_k - \lambda_{k+1}, Ax_{k+1} - b\rangle + \langle y_k - x_{k+1}, x_{k+1} - x\rangle_\mathcal{Q}$$
$$- \beta\|Ax_{k+1} - b\|^2 + \frac{L}{2}\|x_{k+1} - y_k\|^2 - \frac{1}{2L}\|\nabla f(y_k) - \nabla f(x)\|^2, \tag{3.23}$$

where, by using the definition of $\mathcal{Q}$, the last identity follows from

$$\frac{1}{\sigma}\langle y_k - x_{k+1}, x_{k+1} - x\rangle - \beta\langle Ay_k - b, Ax_{k+1} - Ax\rangle$$
$$= \frac{1}{\sigma}\langle y_k - x_{k+1}, x_{k+1} - x\rangle - \beta\langle Ay_k - Ax_{k+1}, Ax_{k+1} - Ax\rangle - \beta\|Ax_{k+1} - b\|^2$$
$$= \frac{1}{\sigma}\langle y_k - x_{k+1}, x_{k+1} - x\rangle - \beta\langle y_k - x_{k+1}, A^*A(x_{k+1} - x)\rangle - \beta\|Ax_{k+1} - b\|^2$$
$$= \langle y_k - x_{k+1}, x_{k+1} - x\rangle_\mathcal{Q} - \beta\|Ax_{k+1} - b\|^2.$$

18

Recall that from (3.3h) we have

$$0 = \mu_k - \lambda_{k+1} + \frac{\rho}{\gamma}\left(Az_{k+1}^\gamma - \gamma b\right),$$

which yields further

$$0 = \frac{1}{\rho}\left\langle \mu_k - \lambda_{k+1}, \lambda_{k+1} - \lambda \right\rangle + \frac{1}{\gamma}\left\langle \lambda_{k+1} - \lambda, Az_{k+1}^\gamma - \gamma b \right\rangle. \tag{3.24}$$

Moreover, from (3.3d) and (3.3g) we have

$$\left\langle \lambda - \frac{1}{\gamma}\nu_{k+1}^\gamma, Ax_{k+1} - b \right\rangle + \frac{1}{\gamma}\left\langle \lambda_{k+1} - \lambda, Az_{k+1}^\gamma - \gamma b \right\rangle$$

$$= \left\langle \lambda - \lambda_{k+1} - \frac{1}{\gamma}(t_{k+1} - 1)(\lambda_{k+1} - \lambda_k), Ax_{k+1} - b \right\rangle$$

$$+ \left\langle \lambda_{k+1} - \lambda, Ax_{k+1} + \frac{1}{\gamma}(t_{k+1} - 1)A(x_{k+1} - x_k) - b \right\rangle$$

$$= \left\langle \lambda - \lambda_{k+1}, Ax_{k+1} - b \right\rangle - \frac{1}{\gamma}(t_{k+1} - 1)\left\langle \lambda_{k+1} - \lambda_k, Ax_{k+1} - b \right\rangle$$

$$+ \left\langle \lambda_{k+1} - \lambda, Ax_{k+1} - b \right\rangle + \frac{1}{\gamma}(t_{k+1} - 1)\left\langle \lambda_{k+1} - \lambda, Ax_{k+1} - Ax_k \right\rangle$$

$$= -\frac{1}{\gamma}(t_{k+1} - 1)\left\langle \lambda_{k+1} - \lambda_k, Ax_{k+1} - b \right\rangle + \frac{1}{\gamma}(t_{k+1} - 1)\left\langle \lambda_{k+1} - \lambda, Ax_{k+1} - Ax_k \right\rangle,$$

therefore, by summing up (3.23) and (3.24) and after rearranging the terms, the estimate (3.21) follows.

Next we will prove the second estimate. By take $x := x_k$ in inequality (3.15) we get

$$f(x_{k+1}) + \left\langle \lambda, Ax_{k+1} - b \right\rangle$$

$$\leqslant f(x_k) + \left\langle \lambda, Ax_k - b \right\rangle + \left\langle \lambda - \frac{1}{\gamma}\nu_{k+1}^\gamma, Ax_{k+1} - Ax_k \right\rangle + \frac{1}{\gamma}(1 - \gamma)\left\langle \lambda_k - \lambda_{k+1}, Ax_{k+1} - Ax_k \right\rangle$$

$$+ \frac{1}{\sigma}\left\langle y_k - x_{k+1}, x_{k+1} - x_k \right\rangle - \beta\left\langle Ay_k - b, Ax_{k+1} - Ax_k \right\rangle$$

$$+ \frac{L}{2}\|x_{k+1} - y_k\|^2 - \frac{1}{2L}\|\nabla f(y_k) - \nabla f(x_k)\|^2$$

$$= f(x_k) + \left\langle \lambda, Ax_k - b \right\rangle + \left\langle \lambda - \frac{1}{\gamma}\nu_{k+1}^\gamma, Ax_{k+1} - Ax_k \right\rangle + \frac{1}{\gamma}(1 - \gamma)\left\langle \lambda_k - \lambda_{k+1}, Ax_{k+1} - Ax_k \right\rangle$$

$$+ \left\langle y_k - x_{k+1}, x_{k+1} - x_k \right\rangle_{\mathcal{Q}} - \beta\left\langle Ax_{k+1} - b, Ax_{k+1} - Ax_k \right\rangle$$

$$+ \frac{L}{2}\|x_{k+1} - y_k\|^2 - \frac{1}{2L}\|\nabla f(y_k) - \nabla f(x_k)\|^2, \tag{3.25}$$

where, by using again the definition of $\mathcal{Q}$, the last identity follows from

$$\frac{1}{\sigma}\left\langle y_k - x_{k+1}, x_{k+1} - x_k \right\rangle - \beta\left\langle Ay_k - b, Ax_{k+1} - Ax_k \right\rangle$$

$$= \frac{1}{\sigma}\left\langle y_k - x_{k+1}, x_{k+1} - x_k \right\rangle - \beta\left\langle Ay_k - Ax_{k+1}, Ax_{k+1} - Ax_k \right\rangle - \beta\left\langle Ax_{k+1} - b, Ax_{k+1} - Ax_k \right\rangle$$

$$= \frac{1}{\sigma}\left\langle y_k - x_{k+1}, x_{k+1} - x_k \right\rangle - \beta\left\langle y_k - x_{k+1}, A^*A(x_{k+1} - x_k) \right\rangle - \beta\left\langle Ax_{k+1} - b, Ax_{k+1} - Ax_k \right\rangle$$

$$= \left\langle y_k - x_{k+1}, x_{k+1} - x_k \right\rangle_{\mathcal{Q}} - \beta\left\langle Ax_{k+1} - b, Ax_{k+1} - Ax_k \right\rangle.$$

The identity (1.9) gives us

$$-\beta\left\langle Ax_{k+1} - b, Ax_{k+1} - Ax_k \right\rangle = -\frac{\beta}{2}\|Ax_{k+1} - b\|^2 - \frac{\beta}{2}\|Ax_{k+1} - Ax_k\|^2 + \frac{\beta}{2}\|Ax_k - b\|^2,$$

hence, by recalling relation (1.3), (3.25) can be equivalently written as

$$
\begin{aligned}
&\mathcal{L}_\beta\left(x_{k+1}, \lambda\right) - \mathcal{L}_\beta\left(x, \lambda_{k+1}\right) \\
&\leqslant \mathcal{L}_\beta\left(x_k, \lambda\right) - \mathcal{L}_\beta\left(x, \lambda_k\right) + \left\langle \lambda - \frac{1}{\gamma}\nu_{k+1}^\gamma, Ax_{k+1} - Ax_k \right\rangle + \frac{1}{\gamma}(\gamma - 1)\langle \lambda_{k+1} - \lambda_k, Ax_{k+1} - Ax_k \rangle \\
&\quad + \langle y_k - x_{k+1}, x_{k+1} - x_k \rangle_\mathcal{Q} - \frac{\beta}{2}\left\| Ax_{k+1} - Ax_k \right\|^2 + \frac{L}{2}\left\| x_{k+1} - y_k \right\|^2 \\
&\quad - \frac{1}{2L}\left\| \nabla f\left(y_k\right) - \nabla f\left(x_k\right) \right\|^2.
\end{aligned} \tag{3.26}
$$

In addition, by taking $\lambda := \lambda_k$ in (3.24) gives

$$
0 = \frac{1}{\rho}\left\langle \mu_k - \lambda_{k+1}, \lambda_{k+1} - \lambda_k \right\rangle + \frac{1}{\gamma}\left\langle \lambda_{k+1} - \lambda_k, Az_{k+1}^\gamma - \gamma b \right\rangle. \tag{3.27}
$$

Moreover, we have from (3.3d) and (3.3g)

$$
\begin{aligned}
&\left\langle \lambda - \frac{1}{\gamma}\nu_{k+1}^\gamma, Ax_{k+1} - Ax_k \right\rangle + \frac{1}{\gamma}\left\langle \lambda_{k+1} - \lambda_k, Az_{k+1}^\gamma - \gamma b \right\rangle \\
&= \left\langle \lambda - \lambda_{k+1}, Ax_{k+1} - Ax_k \right\rangle - \frac{1}{\gamma}(t_{k+1} - 1)\langle \lambda_{k+1} - \lambda_k, Ax_{k+1} - Ax_k \rangle \\
&\quad + \langle \lambda_{k+1} - \lambda_k, Ax_{k+1} - b \rangle + \frac{1}{\gamma}(t_{k+1} - 1)\langle \lambda_{k+1} - \lambda_k, Ax_{k+1} - Ax_k \rangle \\
&= \left\langle \lambda - \lambda_{k+1}, Ax_{k+1} - Ax_k \right\rangle + \left\langle \lambda_{k+1} - \lambda_k, Ax_{k+1} - b \right\rangle,
\end{aligned}
$$

therefore, by summing up (3.26) and (3.27) and after rearranging terms, the estimate (3.22) follows. $\qquad\square$

For $(x, \lambda) \in \mathcal{F} \times \mathcal{G}$ and $k \geqslant 1$ we introduce the following energy function associated with Algorithm 1

$$
\begin{aligned}
\mathcal{E}_k\left(x, \lambda\right) := {}& t_k\left(t_k - 1 + \gamma\right)\left(\mathcal{L}_\beta\left(x_k, \lambda\right) - \mathcal{L}_\beta\left(x, \lambda_k\right)\right) + \frac{1}{2}\left\| z_k^\gamma - \gamma x \right\|_\mathcal{Q}^2 + \frac{1}{2\rho}\left\| \nu_k^\gamma - \gamma \lambda \right\|^2 \\
&+ \frac{1}{2}\gamma\left(1 - \gamma\right)\left\| x_k - x \right\|_\mathcal{Q}^2 + \frac{1}{2\rho}\gamma\left(1 - \gamma\right)\left\| \lambda_k - \lambda \right\|^2 + \frac{1 - \gamma}{2\rho}\left(t_k - 1\right)\left\| \lambda_k - \lambda_{k-1} \right\|^2.
\end{aligned}
$$

According to (1.4), for every $(x_*, \lambda_*) \in \mathcal{S}$ and every $k \geqslant 1$ it holds

$$
\mathcal{E}_k\left(x_*, \lambda_*\right) \geqslant 0.
$$

The following estimate for the energy function will play a fundamental role in our analysis.

Proposition 3.8. *Let $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ be the sequence generated by Algorithm 1. Then for every $(x, \lambda) \in \mathcal{F} \times \mathcal{G}$ and every $k \geqslant 1$ it holds*

$$
\begin{aligned}
\mathcal{E}_{k+1}\left(x, \lambda\right) \leqslant {}& \mathcal{E}_k\left(x, \lambda\right) + \left(t_{k+1}^2 - t_{k+1} - t_k^2 + (1 - \gamma)t_k\right)\left(\mathcal{L}_\beta\left(x_k, \lambda\right) - \mathcal{L}_\beta\left(x, \lambda_k\right)\right) \\
&- \frac{\beta\gamma}{2}t_{k+1}\left\| Ax_{k+1} - b \right\|^2 - \frac{\beta}{2}t_{k+1}\left(t_{k+1} - 1\right)\left\| Ax_{k+1} - Ax_k \right\|^2 \\
&- \frac{\gamma}{2L}t_{k+1}\left\| \nabla f\left(y_k\right) - \nabla f\left(x\right) \right\|^2 - \frac{1}{2L}t_{k+1}\left(t_{k+1} - 1\right)\left\| \nabla f\left(y_k\right) - \nabla f\left(x_k\right) \right\|^2 \\
&- (1 - \gamma)\left(t_{k+1} - 1\right)\left\| x_{k+1} - x_k \right\|_\mathcal{Q}^2 - \frac{1 - \gamma}{2\rho}\left(2t_{k+1} - 1\right)\left\| \lambda_{k+1} - \lambda_k \right\|^2 \\
&- \frac{1}{2}t_{k+1}^2\left\| x_{k+1} - y_k \right\|_{\gamma\mathcal{Q} - L\mathrm{Id}}^2 - \frac{\gamma}{2\rho}t_{k+1}^2\left\| \lambda_{k+1} - \mu_k \right\|^2.
\end{aligned} \tag{3.28}
$$

Proof. Let $(x, \lambda) \in \mathcal{F} \times \mathcal{G}$ and $k \geqslant 1$ be fixed. Multiplying (3.22) by $t_{k+1}(t_{k+1} - 1) \geqslant 0$ and (3.21) by γt_{k+1}, and adding the resulting inequalities, yields

$$
\begin{aligned}
&t_{k+1}\left(t_{k+1} - 1 + \gamma\right)\left(\mathcal{L}_\beta\left(x_{k+1}, \lambda\right) - \mathcal{L}_\beta\left(x, \lambda_{k+1}\right)\right) \\
&\leqslant t_{k+1}\left(t_{k+1} - 1\right)\left(\mathcal{L}_\beta\left(x_k, \lambda\right) - \mathcal{L}_\beta\left(x, \lambda_k\right)\right) \\
&\quad + \frac{1}{\gamma}\left(1 - \gamma\right) t_{k+1}\left\langle\lambda_k - \lambda_{k+1}, \gamma\left(Ax_{k+1} - b\right) + \left(t_{k+1} - 1\right)\left(Ax_{k+1} - Ax_k\right)\right\rangle \\
&\quad + t_{k+1}\left\langle y_k - x_{k+1}, \gamma\left(x_{k+1} - x\right) + \left(t_{k+1} - 1\right)\left(x_{k+1} - x_k\right)\right\rangle_Q \\
&\quad + \frac{1}{\rho} t_{k+1}\left\langle\mu_k - \lambda_{k+1}, \gamma\left(\lambda_{k+1} - \lambda\right) + \left(t_{k+1} - 1\right)\left(\lambda_{k+1} - \lambda_k\right)\right\rangle \\
&\quad - \frac{\beta\gamma}{2} t_{k+1}\left\|Ax_{k+1} - b\right\|^2 - \frac{\beta}{2} t_{k+1}\left(t_{k+1} - 1\right)\left\|Ax_{k+1} - Ax_k\right\|^2 \\
&\quad - \frac{\gamma}{2L} t_{k+1}\left\|\nabla f\left(y_k\right) - \nabla f\left(x\right)\right\|^2 - \frac{1}{2L} t_{k+1}\left(t_{k+1} - 1\right)\left\|\nabla f\left(y_k\right) - \nabla f\left(x_k\right)\right\|^2 \\
&\quad + \frac{L}{2} t_{k+1}\left(t_{k+1} - 1 + \gamma\right)\left\|x_{k+1} - y_k\right\|^2. \tag{3.29}
\end{aligned}
$$

According to (3.3b), (3.3g) and (3.3h) we have

$$
\begin{aligned}
&\frac{1}{\gamma}\left(1 - \gamma\right) t_{k+1}\left\langle\lambda_k - \lambda_{k+1}, \gamma\left(Ax_{k+1} - b\right) + \left(t_{k+1} - 1\right)\left(Ax_{k+1} - Ax_k\right)\right\rangle \\
&= \frac{1}{\gamma}\left(1 - \gamma\right) t_{k+1}\left\langle\lambda_k - \lambda_{k+1}, Az_{k+1}^\gamma - \gamma b\right\rangle = \frac{1}{\rho}\left(1 - \gamma\right) t_{k+1}\left\langle\lambda_k - \lambda_{k+1}, \lambda_{k+1} - \mu_k\right\rangle \\
&= -\frac{1 - \gamma}{2\rho} t_{k+1}\left\|\lambda_{k+1} - \lambda_k\right\|^2 - \frac{1 - \gamma}{2\rho} t_{k+1}\left\|\lambda_{k+1} - \mu_k\right\|^2 + \frac{1 - \gamma}{2\rho} t_{k+1}\left\|\mu_k - \lambda_k\right\|^2 \\
&\leqslant -\frac{1 - \gamma}{2\rho} t_{k+1}\left\|\lambda_{k+1} - \lambda_k\right\|^2 + \frac{1 - \gamma}{2\rho}\frac{\left(t_k - 1\right)^2}{t_{k+1}}\left\|\lambda_k - \lambda_{k-1}\right\|^2 \\
&\leqslant -\frac{1 - \gamma}{2\rho} t_{k+1}\left\|\lambda_{k+1} - \lambda_k\right\|^2 + \frac{1 - \gamma}{2\rho}\left(t_k - 1\right)\left\|\lambda_k - \lambda_{k-1}\right\|^2, \tag{3.30}
\end{aligned}
$$

where in the last inequality we use that $\{t_k\}_{k \geqslant 1}$ is nondecreasing and that $t_k \geqslant 1$ for every $k \geqslant 1$.

On the other hand, (3.3g), (3.9a) and (1.12) give

$$
\begin{aligned}
&t_{k+1}\left\langle y_k - x_{k+1}, \gamma\left(x_{k+1} - x\right) + \left(t_{k+1} - 1\right)\left(x_{k+1} - x_k\right)\right\rangle_Q \\
&= \left\langle z_k^\gamma - z_{k+1}^\gamma, z_{k+1}^\gamma - \gamma x\right\rangle_Q + \left(\gamma - 1\right)\left\langle x_{k+1} - x_k, \gamma\left(x_{k+1} - x\right) + \left(t_{k+1} - 1\right)\left(x_{k+1} - x_k\right)\right\rangle_Q \\
&= -\frac{1}{2}\left\|z_{k+1}^\gamma - z_k^\gamma\right\|_Q^2 - \frac{1}{2}\left\|z_{k+1}^\gamma - \gamma x\right\|_Q^2 + \frac{1}{2}\left\|z_k^\gamma - \gamma x\right\|_Q^2 + \frac{1}{2}\gamma\left(\gamma - 1\right)\left\|x_{k+1} - x_k\right\|_Q^2 \\
&\quad + \frac{1}{2}\gamma\left(\gamma - 1\right)\left\|x_{k+1} - x\right\|_Q^2 - \frac{1}{2}\gamma\left(\gamma - 1\right)\left\|x_k - x\right\|_Q^2 + \left(\gamma - 1\right)\left(t_{k+1} - 1\right)\left\|x_{k+1} - x_k\right\|_Q^2. \tag{3.31}
\end{aligned}
$$

From (1.13), (3.9b) and (3.8) we have

$$
\begin{aligned}
-\frac{1}{2}\left\|z_{k+1}^\gamma - z_k^\gamma\right\|_Q^2 &= -\frac{1}{2}\left\|z_{k+1} - z_k + \left(\gamma - 1\right)\left(x_{k+1} - x_k\right)\right\|_Q^2 \\
&= -\frac{1}{2}\gamma\left\|z_{k+1} - z_k\right\|_Q^2 - \frac{1}{2}\gamma\left(\gamma - 1\right)\left\|x_{k+1} - x_k\right\|_Q^2 \\
&\quad + \frac{1}{2}\left(\gamma - 1\right)\left\|z_{k+1} - z_k - \left(x_{k+1} - x_k\right)\right\|_Q^2 \\
&\leqslant -\frac{1}{2}\gamma t_{k+1}^2\left\|x_{k+1} - y_k\right\|_Q^2 - \frac{1}{2}\gamma\left(\gamma - 1\right)\left\|x_{k+1} - x_k\right\|_Q^2,
\end{aligned}
$$

21

which we combine with (3.31) to obtain

$$t_{k+1} \langle y_k - x_{k+1}, \gamma (x_{k+1} - x) + (t_{k+1} - 1)(x_{k+1} - x_k) \rangle_{\mathcal{Q}}$$
$$\leqslant -\frac{1}{2} \gamma t_{k+1}^2 \|x_{k+1} - y_k\|_{\mathcal{Q}}^2 - \frac{1}{2} \|z_{k+1}^\gamma - \gamma x\|_{\mathcal{Q}}^2 + \frac{1}{2} \|z_k^\gamma - \gamma x\|_{\mathcal{Q}}^2 - \frac{1}{2} \gamma (1-\gamma) \|x_{k+1} - x\|_{\mathcal{Q}}^2$$
$$+ \frac{1}{2} \gamma (1-\gamma) \|x_k - x\|_{\mathcal{Q}}^2 + (\gamma - 1)(t_{k+1} - 1) \|x_{k+1} - x_k\|_{\mathcal{Q}}^2. \tag{3.32}$$

By using the same technique, we can derive that

$$\frac{1}{\rho} t_{k+1} \langle \mu_k - \lambda_{k+1}, \gamma (\lambda_{k+1} - \lambda) + (t_{k+1} - 1)(\lambda_{k+1} - \lambda_k) \rangle$$
$$\leqslant -\frac{\gamma}{2\rho} t_{k+1}^2 \|\lambda_{k+1} - \mu_k\|^2 - \frac{1}{2\rho} \|\nu_{k+1}^\gamma - \gamma \lambda\|^2 + \frac{1}{2} \|\nu_k^\gamma - \gamma \lambda\|^2 - \frac{\gamma}{2\rho} (1-\gamma) \|\lambda_{k+1} - \lambda\|^2$$
$$+ \frac{\gamma}{2\rho} (1-\gamma) \|\lambda_k - \lambda\|^2 + \frac{1}{\rho} (\gamma - 1)(t_{k+1} - 1) \|\lambda_{k+1} - \lambda_k\|^2. \tag{3.33}$$

Plugging (3.30), (3.32) and (3.33) into (3.29), and taking into consideration the fact that $\gamma \in (0, 1]$, gives the desired statement. $\qquad \square$

Next we record some direct consequences of the above estimate.

Proposition 3.9. *Let* $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ *be the sequence generated by Algorithm 1 and* $(x_*, \lambda_*) \in \mathcal{S}$. *Then the sequence* $\{\mathcal{E}_k(x_*, \lambda_*)\}_{k \geqslant 1}$ *is nonincreasing and the following statements are true*

$$\left(1 - \frac{m}{\gamma}\right) \sum_{k \geqslant 1} t_k \left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k)\right) < +\infty$$

$$\sum_{k \geqslant 1} t_{k+1} \left(\beta \|Ax_{k+1} - b\|^2 + \frac{1}{L} \|\nabla f(y_k) - \nabla f(x_*)\|^2\right) < +\infty$$

$$\sum_{k \geqslant 1} t_{k+1}(t_{k+1} - 1) \left(\beta \|Ax_{k+1} - Ax_k\|^2 + \frac{1}{L} \|\nabla f(y_k) - \nabla f(x_k)\|^2\right) < +\infty$$

$$(1-\gamma) \sum_{k \geqslant 1} (t_{k+1} - 1) \|x_{k+1} - x_k\|_{\mathcal{Q}}^2 < +\infty$$

$$(1-\gamma) \sum_{k \geqslant 1} (2t_{k+1} - 1) \|\lambda_{k+1} - \lambda_k\|^2 < +\infty$$

$$\sum_{k \geqslant 1} t_{k+1}^2 \left(\|x_{k+1} - y_k\|_{\gamma\mathcal{Q} - L\mathrm{Id}}^2 + \frac{\gamma}{\rho} \|\lambda_{k+1} - \mu_k\|^2\right) < +\infty.$$

Proof. Since $\{t_k\}_{k \geqslant 1}$ is an nondecreasing sequence that satisfies (3.2) and $0 < m \leqslant \gamma \leqslant 1$, we have for every $k \geqslant 1$

$$t_{k+1}^2 - t_{k+1} - t_k^2 + (1-\gamma) t_k \leqslant (m-1) t_{k+1} + (1-\gamma) t_k \leqslant (m - \gamma) t_k \leqslant 0.$$

Moreover, as $(x_*, \lambda_*) \in \mathcal{S}$, we must have $x_* \in \mathcal{F}$ and $\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k) \geqslant 0$ for every $k \geqslant 1$ due to (1.4). Combining these observations, we deduce from the inequality (3.28) applied to $(x, \lambda) = (x_*, \lambda_*)$ that for every $k \geqslant 1$

$$\mathcal{E}_{k+1}(x_*, \lambda_*) \leqslant \mathcal{E}_k(x_*, \lambda_*) - (\gamma - m) t_k \left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k)\right)$$
$$- \frac{\beta\gamma}{2} t_{k+1} \|Ax_{k+1} - b\|^2 - \frac{\beta}{2} t_{k+1}(t_{k+1} - 1) \|Ax_{k+1} - Ax_k\|^2$$
$$- \frac{\gamma}{2L} t_{k+1} \|\nabla f(y_k) - \nabla f(x_*)\|^2 - \frac{1}{2L} t_{k+1}(t_{k+1} - 1) \|\nabla f(y_k) - \nabla f(x_k)\|^2$$
$$- (1-\gamma)(t_{k+1} - 1) \|x_{k+1} - x_k\|_{\mathcal{Q}}^2 - \frac{1-\gamma}{2\rho} (2t_{k+1} - 1) \|\lambda_{k+1} - \lambda_k\|^2$$
$$- \frac{1}{2} t_{k+1}^2 \|x_{k+1} - y_k\|_{\gamma\mathcal{Q} - L\mathrm{Id}}^2 - \frac{\gamma}{2\rho} t_{k+1}^2 \|\lambda_{k+1} - \mu_k\|^2. \tag{3.35}$$

22

By applying Lemma 1.1 we obtain all conclusions. □

Remark 3.10. Since the sequence $\{\mathcal{E}_k(x_*, \lambda_*)\}_{k \geqslant 1}$ is nonincreasing and for every $k \geqslant 1$

$$\gamma t_k^2 \leqslant t_k (t_k - 1 + \gamma) \Leftrightarrow t_k \geqslant 1, \tag{3.36}$$

we deduce that

$$\begin{aligned}
\frac{\beta\gamma}{2} t_k^2 \|Ax_k - b\|^2 &\leqslant \gamma t_k^2 \left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k) \right) \\
&\leqslant t_k (t_k - 1 + \gamma) \left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k) \right) \\
&\leqslant \mathcal{E}_k(x_*, \lambda_*) \leqslant \cdots \leqslant \mathcal{E}_1(x_*, \lambda_*).
\end{aligned}$$

Consequently, for every $k \geqslant 1$ we have

$$t_k^2 \left(\mathcal{L}(x_k, \lambda_*) - \mathcal{L}(x_*, \lambda_k) \right) \leqslant t_k^2 \left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k) \right) \leqslant \frac{\mathcal{E}_1(x_*, \lambda_*)}{\gamma}$$

and, when $\beta > 0$, $\quad t_k \|Ax_k - b\| \leqslant \sqrt{\dfrac{2\mathcal{E}_1(x_*, \lambda_*)}{\beta\gamma}}.$ \hfill (3.37)

Remark 3.11. Recall that from Proposition 3.9 we have

$$\sum_{k \geqslant 1} (t_{k+1} - 1) \left(\|x_{k+1} - x_k\|_{\mathcal{Q}}^2 + \frac{1}{2\rho} \|\lambda_{k+1} - \lambda_k\|^2 \right) < +\infty, \tag{3.38}$$

whenever $\gamma < 1$. Taking into account the way γ has arisen in the context of the dynamical system (PD-AVD) (see (2.13)), this corresponds to

$$\gamma = \frac{1}{\theta(\alpha - 1)} < 1 \Leftrightarrow \frac{1}{\alpha - 1} < \theta.$$

In the continuous case it has been proved (see [8, Theorem 3.2]) that, if $\dfrac{1}{\alpha - 1} < \theta$, then

$$\int_{t_0}^{+\infty} t \left\| \left(\dot{x}(t), \dot{\lambda}(t) \right) \right\|^2 < +\infty,$$

which can be seen as the continuous counterpart of (3.38). Both statements play a crucial role in the proof of the convergence of the sequence of iterates generated by Algorithm 1 and of the trajectory generated by (PD-AVD), respectively.

The following result, which complements the statements of Proposition 3.9, will also play a crucial role in the proof of the convergence of the sequence of iterates.

Proposition 3.12. *Let* $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ *be the sequence generated by Algorithm 1 with*

$$0 < \sigma < \frac{\gamma}{L + \gamma\beta\|A\|^2},$$

and $(x_*, \lambda_*) \in \mathcal{S}$. *Then the following statements are true*

$$\left(1 - \frac{m}{\gamma} \right) \sum_{k \geqslant 1} t_k \|A^*(\lambda_k - \lambda_*)\|^2 < +\infty, \tag{3.39a}$$

$$t_{k+1} (t_{k+1} - 1)^2 \sum_{k \geqslant 1} \|A^*(\lambda_{k+1} - \lambda_k)\|^2 < +\infty. \tag{3.39b}$$

In addition, there exists $C_0 > 0$ *such that for every* $k \geqslant 1$

$$\|A^*(\lambda_k - \lambda_*)\| \leqslant \frac{C_0}{t_k}.$$

Proof. From (3.18), after rearranging some terms, we have for every $k \geqslant 1$

$$A^* \left(\frac{1}{\gamma} \nu_{k+1}^\gamma - \lambda_* \right) = \frac{1}{\gamma} (1 - \gamma) A^* (\lambda_k - \lambda_{k+1}) + \nabla f (x_*) - \nabla f (y_k) + \frac{1}{\sigma} (y_k - x_{k+1})$$
$$+ \beta A^* A (x_{k+1} - y_k) - \beta A^* (A x_{k+1} - b).$$

It follows from Proposition 3.9, by using (3.20) and the fact that $t_k \geqslant 1$, that for every $k \geqslant 1$

$$\sum_{k \geqslant 1} t_{k+1} \left\| A^* \left(\frac{1}{\gamma} \nu_{k+1}^\gamma - \lambda_* \right) \right\|^2$$

$$\leqslant \frac{5}{\gamma^2} (1 - \gamma)^2 \|A\|^2 \sum_{k \geqslant 1} t_{k+1} \|\lambda_{k+1} - \lambda_k\|^2 + 5 \sum_{k \geqslant 1} t_{k+1} \|\nabla f (y_k) - \nabla f (x_*)\|^2$$

$$+ \frac{5}{\sigma^2} \sum_{k \geqslant 1} t_{k+1} \|x_{k+1} - y_k\|^2 + 5 \beta^2 \|A^* A\|^2 \sum_{k \geqslant 1} t_{k+1} \|x_{k+1} - y_k\|^2$$

$$+ 5 \beta^2 \|A\|^2 \sum_{k \geqslant 1} t_{k+1} \|A x_{k+1} - b\|^2 < +\infty.$$

According to (3.3d) we have for every $k \geqslant 1$

$$A^* \left(\frac{1}{\gamma} \nu_{k+1}^\gamma - \lambda_* \right) = \left(1 + \frac{1}{\gamma} (t_{k+1} - 1) \right) A^* (\lambda_{k+1} - \lambda_*) - \frac{1}{\gamma} (t_{k+1} - 1) A^* (\lambda_k - \lambda_*),$$

hence, by applying the identity (1.10), we get

$$\left\| A^* \left(\frac{1}{\gamma} \nu_{k+1}^\gamma - \lambda_* \right) \right\|^2 = \left(1 + \frac{1}{\gamma} (t_{k+1} - 1) \right) \|A^* (\lambda_{k+1} - \lambda_*)\|^2 - \frac{1}{\gamma} (t_{k+1} - 1) \|A^* (\lambda_k - \lambda_*)\|^2$$

$$+ \frac{1}{\gamma} (t_{k+1} - 1) \left(1 + \frac{1}{\gamma} (t_{k+1} - 1) \right) \|A^* (\lambda_{k+1} - \lambda_k)\|^2. \qquad (3.40)$$

On the other hand, it follows from condition (3.2) and the fact that $\{t_k\}_{k \geqslant 1}$ is nondecreasing that for every $k \geqslant 1$

$$\frac{1}{\gamma} t_{k+1} (t_{k+1} - 1) - t_k \left(1 + \frac{1}{\gamma} (t_k - 1) \right) = \frac{1}{\gamma} \left(t_{k+1}^2 - t_{k+1} - t_k^2 + t_k \right) - t_k$$

$$\leqslant \frac{1}{\gamma} \left((m - 1) t_{k+1} + t_k \right) - t_k$$

$$= \frac{m - 1}{\gamma} (t_{k+1} - t_k) + \left(\frac{m}{\gamma} - 1 \right) t_k$$

$$\leqslant \left(\frac{m}{\gamma} - 1 \right) t_k, \qquad (3.41)$$

Combining (3.40) and (3.41), it yields for every $k \geqslant 1$

$$t_{k+1} \left(1 + \frac{1}{\gamma} (t_{k+1} - 1) \right) \|A^* (\lambda_{k+1} - \lambda_*)\|^2$$

$$= t_k \left(1 + \frac{1}{\gamma} (t_k - 1) \right) \|A^* (\lambda_k - \lambda_*)\|^2 + \left(\frac{1}{\gamma} t_{k+1} (t_{k+1} - 1) - t_k \left(1 + \frac{1}{\gamma} (t_k - 1) \right) \right) \|A^* (\lambda_k - \lambda_*)\|^2$$

$$- \frac{1}{\gamma} t_{k+1} (t_{k+1} - 1) \left(1 + \frac{1}{\gamma} (t_{k+1} - 1) \right) \|A^* (\lambda_{k+1} - \lambda_k)\|^2 + t_{k+1} \left\| A^* \left(\frac{1}{\gamma} \nu_{k+1}^\gamma - \lambda_* \right) \right\|^2$$

$$\leqslant t_k \left(1 + \frac{1}{\gamma} (t_k - 1) \right) \|A^* (\lambda_k - \lambda_*)\|^2 - \left(1 - \frac{m}{\gamma} \right) t_k \|A^* (\lambda_k - \lambda_*)\|^2$$

$$- \frac{1}{\gamma^2} t_{k+1} (t_{k+1} - 1)^2 \|A^* (\lambda_{k+1} - \lambda_k)\|^2 + t_{k+1} \left\| A^* \left(\frac{1}{\gamma} \nu_{k+1}^\gamma - \lambda_* \right) \right\|^2.$$

We are in the setting of inequality (1.15) with

$$a_k := t_k \left(1 + \frac{1}{\gamma}\left(t_k - 1\right)\right) \|A^* \left(\lambda_k - \lambda_*\right)\|^2 \geqslant 0,$$

$$b_k := \left(1 - \frac{m}{\gamma}\right) t_k \|A^* \left(\lambda_k - \lambda_*\right)\|^2 + \frac{1}{\gamma^2} t_{k+1} \left(t_{k+1} - 1\right)^2 \|A^* \left(\lambda_{k+1} - \lambda_k\right)\|^2 \geqslant 0,$$

$$d_k := t_{k+1} \left\|A^* \left(\frac{1}{\gamma}\nu_{k+1}^\gamma - \lambda_*\right)\right\|^2 \geqslant 0,$$

for every $k \geqslant 1$. According to Lemma 1.1, (3.39a) and (3.39b) are fulfilled and the sequence $\left\{t_k \left(1 + \frac{1}{\gamma}\left(t_k - 1\right)\right) \|A^* \left(\lambda_k - \lambda_*\right)\|^2\right\}_{k \geqslant 1}$ is convergent, therefore it is bounded. Consequently, there exists $C_0 > 0$ such that for every $k \geqslant 1$

$$t_k^2 \|A^* \left(\lambda_k - \lambda_*\right)\|^2 \leqslant t_k \left(1 + \frac{1}{\gamma}\left(t_k - 1\right)\right) \|A^* \left(\lambda_k - \lambda_*\right)\|^2 \leqslant C_0,$$

which provides the conclusion. $\qquad\qquad\square$

3.3 On the boundedness of the sequences

In this section we will discuss the boundedness of the sequence of primal-dual iterates $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ and also of other related sequences which play a role in the convergence analysis.

To this end we define on $\mathcal{H} \times \mathcal{G}$ the inner product

$$\langle u, u' \rangle_{\mathcal{W}} = \langle (x, \lambda), (x', \lambda') \rangle_{\mathcal{W}} = \langle x, x' \rangle_{\mathcal{Q}} + \frac{1}{\rho}\langle \lambda, \lambda' \rangle \quad \forall u := (x, \lambda), u' := \left(x', \lambda'\right) \in \mathcal{H} \times \mathcal{G},$$

where $\mathcal{Q}$ is the operator defined in (3.19) which we proved to be positive definite under assumption (3.1). The norm induced by this scalar product is

$$\|u\|_{\mathcal{W}} = \|(x, \lambda)\|_{\mathcal{W}} = \sqrt{\|x\|_{\mathcal{Q}}^2 + \frac{1}{\rho}\|\lambda\|^2} \quad \forall u := (x, \lambda).$$

The condition on the sequence $\{t_k\}_{k \geqslant 1}$ which we will assume in the next proposition in order to guarantee boundedness for the sequences generated by Algorithm 1 has been proposed in [4]. Later we will see that it is satisfied by the three classical inertial parameters rules by Nesterov, Chambolle-Dossal and Attouch-Cabot.

Proposition 3.13. *Let $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ be the sequence generated by Algorithm 1. Suppose that*

$$\kappa := \inf_{k \geqslant 1} \frac{t_k}{k} > 0. \tag{3.42}$$

Then the sequences $\{(x_k, \lambda_k)\}_{k \geqslant 0}$, $\left\{\left(z_k^\gamma, \nu_k^\gamma\right)\right\}_{k \geqslant 1}$ and $\{(t_{k+1}(x_{k+1} - x_k), t_{k+1}(\lambda_{k+1} - \lambda_k))\}_{k \geqslant 0}$ are bounded. If, in addition $\beta > 0$, then the sequence $\{t_{k+1}(t_{k+1} - 1)(Ax_{k+1} - Ax_k)\}_{k \geqslant 0}$ is also bounded.

Proof. Let $(x_*, \lambda_*) \in \mathcal{S}$ be fixed. For brevity we will write

$$u_* := (x_*, \lambda_*) \in \mathcal{S} \text{ and } u_k := (x_k, \lambda_k) \in \mathcal{H} \times \mathcal{G} \quad \forall k \geqslant 0.$$

By applying (1.13), we have from (3.3g) that for every $k \geqslant 1$

$$\begin{aligned}
\left\|z_k^\gamma - \gamma x_*\right\|_{\mathcal{Q}}^2 &= \|\left(t_k - 1 + \gamma\right)\left(x_k - x_*\right) - \left(t_k - 1\right)\left(x_{k-1} - x_*\right)\|_{\mathcal{Q}}^2 \\
&= \gamma\left(t_k - 1 + \gamma\right)\|x_k - x_*\|_{\mathcal{Q}}^2 - \gamma\left(t_k - 1\right)\|x_{k-1} - x_*\|_{\mathcal{Q}}^2 \\
&\quad + \left(t_k - 1 + \gamma\right)\left(t_k - 1\right)\|x_k - x_{k-1}\|_{\mathcal{Q}}^2.
\end{aligned}$$

By applying (1.11), we have from (3.3d) that for every $k \geqslant 1$

$$\begin{aligned}
\left\| \nu_k^\gamma - \gamma \lambda_* \right\|^2 &= \left\| (t_k - 1 + \gamma)(\lambda_k - \lambda_*) - (t_k - 1)(\lambda_{k-1} - \lambda_*) \right\|^2 \\
&= \gamma (t_k - 1 + \gamma) \left\| \lambda_k - \lambda_* \right\|^2 - \gamma (t_k - 1) \left\| \lambda_{k-1} - \lambda_* \right\|^2 \\
&\quad + (t_k - 1 + \gamma)(t_k - 1) \left\| \lambda_k - \lambda_{k-1} \right\|^2.
\end{aligned}$$

This means the energy function at (x_*, λ_*) can be written for every $k \geqslant 1$ as

$$\begin{aligned}
\mathcal{E}_k(x_*, \lambda_*) &= t_k(t_k - 1 + \gamma)\left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k) \right) \\
&\quad + \frac{\gamma}{2} t_k \left\| u_k - u_* \right\|_{\mathcal{W}}^2 - \frac{\gamma}{2}(t_k - 1) \left\| u_{k-1} - u_* \right\|_{\mathcal{W}}^2 \\
&\quad + \frac{1}{2}(t_k - 1 + \gamma)(t_k - 1) \left\| u_k - u_{k-1} \right\|_{\mathcal{W}}^2 + \frac{1 - \gamma}{2\rho}(t_k - 1) \left\| \lambda_k - \lambda_{k-1} \right\|^2. \quad (3.43)
\end{aligned}$$

According to Proposition 3.9, the sequence $\{ \mathcal{E}_k(x_*, \lambda_*) \}_{k \geqslant 1}$ is nonincreasing, therefore for every $k \geqslant 1$

$$\frac{\gamma}{2} t_k \left\| u_k - u_* \right\|_{\mathcal{W}}^2 - \frac{\gamma}{2}(t_k - 1) \left\| u_{k-1} - u_* \right\|_{\mathcal{W}}^2 + \frac{1}{2}(t_k - 1 + \gamma)(t_k - 1) \left\| u_k - u_{k-1} \right\|_{\mathcal{W}}^2$$

$$\leqslant \frac{1}{2} \left\| z_k^\gamma - \gamma x_* \right\|_{\mathcal{Q}}^2 + \frac{1}{2\rho} \left\| \nu_k^\gamma - \gamma \lambda_* \right\|^2 \leqslant \mathcal{E}_k(x_*, \lambda_*) \leqslant \cdots \leqslant \mathcal{E}_1(x_*, \lambda_*) < +\infty.$$

From here we conclude that the sequence $\left\{ (z_k^\gamma, \nu_k^\gamma) \right\}_{k \geqslant 1}$ is bounded. In addition, for every $k \geqslant 1$ it holds

$$\frac{\gamma}{2} t_k \left\| u_k - u_* \right\|_{\mathcal{W}}^2 \leqslant \frac{\gamma}{2}(t_k - 1) \left\| u_{k-1} - u_* \right\|_{\mathcal{W}}^2 + \mathcal{E}_1(x_*, \lambda_*) \leqslant \frac{\gamma}{2} t_{k-1} \left\| u_{k-1} - u_* \right\|_{\mathcal{W}}^2 + \mathcal{E}_1(x_*, \lambda_*),$$

where the last inequality is due to (3.13), with the convention $t_0 := 0$. After telescoping, we get

$$\frac{\gamma}{2} t_k \left\| u_k - u_* \right\|_{\mathcal{W}}^2 \leqslant k \mathcal{E}_1(x_*, \lambda_*) \quad \forall k \geqslant 1.$$

Then thanks to (3.42) we obtain

$$\left\| u_k - u_* \right\|_{\mathcal{W}}^2 \leqslant \frac{2k}{\gamma t_k} \mathcal{E}_1(x_*, \lambda_*) \leqslant \frac{2}{\gamma \kappa} \mathcal{E}_1(x_*, \lambda_*) < +\infty,$$

which means that $\{ u_k := (x_k, \lambda_k) \}_{k \geqslant 0}$ is bounded. That $\{ (t_{k+1}(x_{k+1} - x_k), t_{k+1}(\lambda_{k+1} - \lambda_k)) \}_{k \geqslant 0}$ is bounded follows from the fact that for all $k \geqslant 1$

$$\begin{aligned}
t_k(x_k - x_{k-1}) &= z_k^\gamma - (\gamma - 1)x_k - x_{k-1}, \\
t_k(\lambda_k - \lambda_{k-1}) &= \nu_k^\gamma - (\gamma - 1)\lambda_k - \lambda_{k-1}
\end{aligned}$$

Finally, recall that from (3.11), (3.3h) and (3.3g), we have for every $k \geqslant 1$

$$\begin{aligned}
\nu_{k+1}^\gamma - \nu_k^\gamma + (1 - \gamma)(\lambda_{k+1} - \lambda_k) &= t_{k+1}(\lambda_{k+1} - \mu_k) = \frac{\rho}{\gamma} t_{k+1}\left(A z_{k+1}^\gamma - \gamma b \right) \\
&= \frac{\rho}{\gamma} \left(\gamma t_{k+1}(A x_{k+1} - b) + t_{k+1}(t_{k+1} - 1)(A x_{k+1} - A x_k) \right).
\end{aligned}$$

The last statement of the proposition follows from here and (3.37). $\qquad \square$

In the following, we will see that the two most prominent choices for the sequence $\{ t_k \}_{k \geqslant 1}$ from the literature, namely, the ones following the rules by Nesterov and by Chambolle-Dossal satisfy not only (3.2), but also (3.42).

Example 3.14. (Nesterov rule) The classical construction proposed Nesterov in [21] for $\{t_k\}_{k \geqslant 1}$ satisfies the following rule

$$t_1 := 1 \qquad \text{and} \qquad t_{k+1} := \frac{1 + \sqrt{1 + 4t_k^2}}{2} \qquad \forall k \geqslant 1. \tag{3.44}$$

The sequence $\{t_k\}_{k \geqslant 1}$ is strictly increasing and verifies relation (3.2) for $m := 1$ with equality. In addition (see, for instance, [6, Lemma 4.3]), it holds $t_k \geqslant \frac{k+1}{2}$ for every $k \geqslant 1$, which means that (3.42) is satisfied for $\kappa \geqslant \frac{1}{2}$.

Example 3.15. (Chambolle-Dossal rule) The construction proposed by Chambolle and Dossal in [10] (see also [3]) for $\{t_k\}_{k \geqslant 1}$ satisfies for $\alpha \geqslant 3$ the following rule

$$t_k := 1 + \frac{k-1}{\alpha - 1} = \frac{k + \alpha - 2}{\alpha - 1} \qquad \forall k \geqslant 1. \tag{3.45}$$

First we show that this sequence fulfills (3.2) with $m := \frac{2}{\alpha - 1} \leqslant 1$. Indeed, for every $k \geqslant 1$ we have

$$t_{k+1}^2 - mt_{k+1} - t_k^2 = (t_{k+1} - t_k)(t_{k+1} + t_k) - mt_{k+1} = \frac{1}{\alpha - 1}\left(2 + \frac{2k-1}{\alpha - 1}\right) - \frac{2}{\alpha - 1}\frac{k + \alpha - 1}{\alpha - 1}$$

$$= -\frac{1}{(\alpha - 1)^2} < 0. \tag{3.46}$$

Furthermore, one can see that for every $k \geqslant 1$ it holds

$$\frac{t_k}{k} = \frac{1}{\alpha - 1} + \frac{\alpha - 2}{k(\alpha - 1)},$$

which proves that (3.42) is verified for $\kappa = \frac{1}{\alpha - 1}$.

Finally, we observe that, by taking into consideration the choice of γ in (2.13) in the context of the dynamical system (PD-AVD) and assumption (3.1) in Algorithm 1, it holds

$$m = \frac{2}{\alpha - 1} \leqslant \gamma = \frac{1}{\theta(\alpha - 1)} \Leftrightarrow \theta \leqslant \frac{1}{2}. \tag{3.47}$$

This connects the choice of the parameter m in Algorithm 1 with the one of the parameter θ in (PD-AVD).

3.4 Fast convergence rates for the primal-dual gap, the feasibility measure and the objective function value

We have seen in Remark 3.10 that, for the general choice of the sequence $\{t_k\}_{k \geqslant 1}$ in (3.2), the convergence rate of the primal-dual gap is of order $\mathcal{O}\left(1/t_k^2\right)$ as $k \to +\infty$. In addition, if $\beta > 0$, then the convergence rate of the feasibility measure is of order $\mathcal{O}\left(1/t_k\right)$ as $k \to +\infty$. In this section we will prove that convergence rates of the feasibility measure and of the objective function value are $\mathcal{O}\left(1/t_k^2\right)$ as $k \to +\infty$ when the sequence $\{t_k\}_{k \geqslant 1}$ is chosen by following the rules by Nesterov, Chambolle-Dossal and also Attouch-Cabot.

In view of (3.42), this will lead for the primal-dual sequence $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ generated by Algorithm 1 and a given primal-dual solution (x_*, λ_*) to the following fast convergence rates

$$\mathcal{L}(x_k, \lambda_*) - \mathcal{L}(x_*, \lambda_k) = \mathcal{O}\left(\frac{1}{k^2}\right) \quad \text{as } k \to +\infty,$$

$$\|Ax_k - b\| = \mathcal{O}\left(\frac{1}{k^2}\right) \text{ and } |f(x_k) - f_*| = \mathcal{O}\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty.$$

We start with the following lemma which holds in the very general setting of Algorithm 1.

Lemma 3.16. *Let* $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ *be the sequence generated by Algorithm 1 and* $(x_*, \lambda_*) \in \mathcal{S}$. *Then the quantity*

$$C_1 := \sup_{\mu \in \mathbb{B}(\lambda_*; 1)} \mathcal{E}_1(x_*, \mu) < +\infty.$$

Proof. Let $(x_*, \lambda_*) \in \mathcal{S}$ and $\mu \in \mathbb{B}(\lambda_*; 1)$. The Cauchy-Schwarz inequality gives

$$\mathcal{L}_\beta(x_1, \mu) - \mathcal{L}_\beta(x_*, \lambda_1) = f(x_1) - f(x_*) + \langle \mu, Ax_1 - b \rangle + \frac{\beta}{2}\|Ax_1 - b\|^2$$

$$\leqslant f(x_1) - f(x_*) + \|\mu\| \|Ax_1 - b\| + \frac{\beta}{2}\|Ax_1 - b\|^2$$

$$\leqslant C_2 := f(x_1) - f(x_*) + (1 + \|\lambda_*\|)\|Ax_1 - b\| + \frac{\beta}{2}\|Ax_1 - b\|^2.$$

On the other hand, as $\nu_1^\gamma = \gamma\lambda_1$ and $\mu \in \mathbb{B}(\lambda_*; 1)$, it holds

$$\frac{1}{2\rho}\|\nu_1^\gamma - \gamma\mu\|^2 + \frac{\gamma}{2\rho}(1 - \gamma)\|\lambda_1 - \mu\|^2$$

$$\leqslant \frac{1}{\rho}\left(\|\nu_1^\gamma - \gamma\lambda_*\|^2 + \gamma^2\|\mu - \lambda_*\|^2\right) + \frac{\gamma}{\rho}(1 - \gamma)\left(\|\lambda_1 - \lambda_*\|^2 + \|\mu - \lambda_*\|^2\right)$$

$$\leqslant C_3 := \frac{\gamma^2}{\rho}\|\lambda_1 - \lambda_*\|^2 + \frac{\gamma}{\rho}(1 - \gamma)\|\lambda_1 - \lambda_*\|^2 + \frac{\gamma}{\rho} = \frac{\gamma}{\rho}\left(\|\lambda_1 - \lambda_*\|^2 + 1\right).$$

Combining these estimates, as $z_1^\gamma = \gamma x_1$, we have

$$\mathcal{E}_1(x_*, \mu) = t_1(t_1 - 1 + \gamma)(\mathcal{L}_\beta(x_1, \mu) - \mathcal{L}_\beta(x_*, \lambda_1)) + \frac{1}{2}\|z_1^\gamma - \gamma x_*\|_{\mathcal{Q}}^2 + \frac{1}{2\rho}\|\nu_1^\gamma - \gamma\mu\|^2$$

$$+ \frac{1}{2}\gamma(1 - \gamma)\|x_1 - x_*\|_{\mathcal{Q}}^2 + \frac{\gamma}{2\rho}(1 - \gamma)\|\lambda_1 - \mu\|^2 + \frac{1 - \gamma}{2\rho}(t_1 - 1)\|\lambda_1 - \lambda_0\|^2$$

$$\leqslant \gamma C_2 + C_3 + \frac{\gamma}{2}\|x_1 - x_*\|_{\mathcal{Q}}^2 < +\infty,$$

which proves the statement. $\qquad\square$

3.4.1 The Nesterov ([21]) rule

We have seen that by choosing $\{t_k\}_{k \geqslant 1}$ as in (3.44), (3.2) is fulfilled as equality for $m = 1$, which also yields $\gamma = 1$ due to (3.1). Consequently, from Proposition 3.8 it follows that for every $(x, \lambda) \in \mathcal{F} \times \mathcal{G}$ and every $k \geqslant 1$ it holds

$$\mathcal{E}_{k+1}(x, \lambda) \leqslant \mathcal{E}_k(x, \lambda), \tag{3.48}$$

which means that the sequence $\{\mathcal{E}_k(x, \lambda)\}_{k \geqslant 1}$ is nonincreasing. This statement is stronger than the one in Proposition 3.9, where we have proved that the sequence of function values of the energy function taken at a primal-dual optimal solution is nonincreasing, and will play an important role in the following.

Theorem 3.17. *Let* $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ *be the sequence generated by Algorithm 1, with the sequence* $\{t_k\}_{k \geqslant 1}$ *chosen to satisfy Nesterov rule (3.44), and* $(x_*, \lambda_*) \in \mathcal{S}$. *Then for every* $k \geqslant 1$ *it holds*

$$0 \leqslant \mathcal{L}(x_k, \lambda_*) - \mathcal{L}(x_*, \lambda_k) + \|Ax_k - b\| \leqslant \frac{C_1}{t_k^2} \tag{3.49}$$

and

$$-\frac{\|\lambda_*\| C_1}{t_k^2} \leqslant f(x_k) - f(x_*) \leqslant \frac{(1 + \|\lambda_*\|) C_1}{t_k^2}. \tag{3.50}$$

Proof. As mentioned earlier in (3.48), for every $(x, \lambda) \in \mathcal{F} \times \mathcal{G}$ and every $k \geq 1$ we have (take into account that $\gamma = 1$)

$$t_k^2 \left(f(x_k) - f(x) + \langle \lambda, Ax_k - b \rangle \right) \leq \mathcal{E}_k(x, \lambda) \leq \cdots \leq \mathcal{E}_1(x, \lambda). \tag{3.51}$$

We fix $n \geq 1$ and define

$$r_n := \begin{cases} \lambda_*, & \text{if } Ax_n - b = 0 \\ \lambda_* + \dfrac{Ax_n - b}{\|Ax_n - b\|}, & \text{if } Ax_n - b \neq 0 \end{cases}.$$

Then $x_* \in \mathcal{F}$ and $r_n \in \mathbb{B}(\lambda_*; 1)$. Hence, $(x_*, r_n) \in \mathcal{F} \times \mathbb{B}(\lambda_*; 1)$, therefore, according to (3.51) and Lemma 3.16,

$$t_n^2 \left(f(x_n) - f(x_*) + \langle r_n, Ax_n - b \rangle \right) \leq \mathcal{E}_1(x_*, r_n) \leq \sup_{\mu \in \mathbb{B}(\lambda_*; 1)} \mathcal{E}_1(x_*, \mu) = C_1. \tag{3.52}$$

If $Ax_n - b \neq 0$, then

$$\begin{aligned} f(x_n) - f(x_*) + \langle r_n, Ax_n - b \rangle &= f(x_n) - f(x_*) + \langle \lambda_*, Ax_n - b \rangle + \|Ax_n - b\| \\ &= \mathcal{L}(x_n, \lambda_*) - \mathcal{L}(x_*, \lambda_n) + \|Ax_n - b\|. \end{aligned}$$

On the other hand, if $Ax_n - b = 0$, we have

$$\begin{aligned} f(x_n) - f(x_*) + \langle r_n, Ax_n - b \rangle &= f(x_n) - f(x_*) + \langle \lambda_*, Ax_n - b \rangle = \mathcal{L}(x_n, \lambda_*) - \mathcal{L}(x_*, \lambda_n) \\ &= \mathcal{L}(x_n, \lambda_*) - \mathcal{L}(x_*, \lambda_n) + \|Ax_n - b\|, \end{aligned}$$

thus, in both scenarios, (3.52) becomes

$$0 \leq t_n^2 \left(\mathcal{L}(x_n, \lambda_*) - \mathcal{L}(x_*, \lambda_n) + \|Ax_n - b\| \right) \leq C_1.$$

Since $n \geq 1$ has been arbitrarily chosen, we obtain (3.49).

As $\mathcal{L}(x_k, \lambda_*) - \mathcal{L}(x_*, \lambda_k) \geq 0$, a direct consequent of (3.49) is that for every $k \geq 1$

$$0 \leq \|Ax_k - b\| \leq \frac{C_1}{t_k^2}.$$

From (3.49) and the Cauchy-Schwarz inequality, we deduce from here that for every $k \geq 1$

$$\begin{aligned} f(x_k) - f(x_*) \leq \frac{C_1}{t_k^2} - \langle \lambda_*, Ax_k - b \rangle &\leq \frac{C_1}{t_k^2} + \|\lambda_*\| \|Ax_k - b\| \\ &\leq \frac{(1 + \|\lambda_*\|) C_1}{t_k^2}. \end{aligned} \tag{3.53}$$

On the other hand, the convexity of f together with (1.2) guarantee that for every $k \geq 1$

$$\begin{aligned} f(x_k) - f(x_*) &\geq \langle \nabla f(x_*), x_k - x_* \rangle = -\langle A^* \lambda_*, x_k - x_* \rangle \\ &= -\langle \lambda_*, Ax_k - b \rangle \geq -\|\lambda_*\| \|Ax_k - b\| \geq -\frac{\|\lambda_*\| C_1}{t_k^2}. \end{aligned} \tag{3.54}$$

By combining (3.53) and (3.54), we obtain (3.50). $\qquad \square$

3.4.2 The Chambolle-Dossal ([10]) rule

In this section we prove fast convergence rates for the primal-dual gap, the feasibility measure and the objective function value for the sequence of inertial parameters $\{t_k\}_{k \geqslant 1}$ following for $\alpha \geqslant 3$ the Chambole-Dossal rule (3.45). We have seen in Example 3.15 that in this case $\{t_k\}_{k \geqslant 1}$ fulfills (3.2) for $m := \frac{2}{\alpha-1}$ and (3.42) for $\kappa := \frac{1}{\alpha-1}$.

For the beginning we observe that for $\frac{2}{\alpha-1} = m \leqslant \gamma \leqslant 1$ and every $k \geqslant 1$ it holds (see (3.46))

$$
\begin{aligned}
t_k \left(t_k - 1 + \gamma\right) - t_{k+1} \left(t_{k+1} - 1\right) &= t_k^2 - t_{k+1}^2 + (1 - \gamma)\left(t_{k+1} - t_k\right) + \gamma t_{k+1} \\
&= -\frac{2}{\alpha - 1} t_{k+1} + \frac{1}{(\alpha - 1)^2} + \frac{1 - \gamma}{\alpha - 1} + \gamma t_{k+1} \\
&= \frac{1}{\alpha - 1}(\gamma(\alpha - 1) - 2)t_{k+1} + \frac{1}{(\alpha - 1)^2}(1 - \gamma(\alpha - 1)) + \frac{1}{\alpha - 1} \\
&= \frac{1}{(\alpha - 1)^2}\Big((\gamma(\alpha - 1) - 2)\,k + (\gamma(\alpha - 1) - 1)\,(\alpha - 2)\Big) \\
&= \frac{1}{(\alpha - 1)^2}\Big((\gamma(\alpha - 1) - 2)\,(k + \alpha - 2) + \alpha - 2\Big).
\end{aligned}
\tag{3.55}
$$

Next we are going to consider two separate cases depending on the relation between $m := \frac{2}{\alpha-1}$ and γ. First we will assume that they are equal, which will then also cover the case $\alpha = 3$.

Theorem 3.18. *Let $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ be the sequence generated by Algorithm 1 with the sequence $\{t_k\}_{k \geqslant 1}$ chosen to satisfy Chambolle-Dossal rule (3.45), $m := \frac{2}{\alpha-1} = \gamma \leqslant 1$, $\beta > 0$, and $(x_*, \lambda_*) \in \mathcal{S}$. Then for every $k \geqslant 2$ it holds*

$$
0 \leqslant \mathcal{L}\left(x_k, \lambda_*\right) - \mathcal{L}\left(x_*, \lambda_k\right) + \|Ax_k - b\| \leqslant \frac{C_4}{t_k^2}
\tag{3.56}
$$

and

$$
-\frac{\|\lambda_*\| \, C_5}{t_k^2} \leqslant f\left(x_k\right) - f\left(x_*\right) \leqslant \frac{\left(1 + \|\lambda_*\|\right) C_5}{t_k^2},
\tag{3.57}
$$

where

$$
C_5 := \frac{C_1}{\gamma} + \frac{2(\alpha - 2)}{\gamma^2 \kappa^2 (\alpha - 1)^2}\left(C_1 + \frac{\alpha - 2}{\kappa (\alpha - 1)^2}\sqrt{\frac{2\mathcal{E}_1\left(x_*, \lambda_*\right)}{\beta \gamma}}\right)\sum_{i \geqslant 1}\frac{1}{i^{3/2}} \in \mathbb{R}_+.
$$

Proof. We fix $n \geqslant 2$ and define

$$
r_n :=
\begin{cases}
\lambda_*, & \text{if } Ax_n - b = 0 \\
\lambda_* + \dfrac{Ax_n - b}{\|Ax_n - b\|}, & \text{if } Ax_n - b \neq 0
\end{cases}.
$$

Then $x_* \in \mathcal{F}$ and $r_n \in \mathbb{B}\left(\lambda_*; 1\right)$. Since $\gamma(\alpha - 1) = 2$, according to (3.55), we have for every $k \geqslant 1$

$$
\begin{aligned}
&\left(t_{k+1}^2 - t_{k+1} - t_k^2 + (1 - \gamma)t_k\right)\left(\mathcal{L}_\beta\left(x_k, r_n\right) - \mathcal{L}_\beta\left(x_*, \lambda_k\right)\right) \\
&= -\frac{\alpha - 2}{(\alpha - 1)^2}\left(\mathcal{L}_\beta\left(x_k, \lambda_*\right) - \mathcal{L}_\beta\left(x_*, \lambda_k\right) + \langle r_n - \lambda_*, Ax_k - b\rangle\right) \\
&\leqslant -\frac{\alpha - 2}{(\alpha - 1)^2}\langle r_n - \lambda_*, Ax_k - b\rangle.
\end{aligned}
$$

30

By taking $(x, \lambda) := (x_*, r_n) \in \mathcal{F} \times \mathbb{B}(\lambda_*; 1)$ in (3.28), we obtain for every $k \geqslant 1$

$$\mathcal{E}_{k+1}(x_*, r_n) \leqslant \mathcal{E}_k(x_*, r_n) + \left(t_{k+1}^2 - t_{k+1} - t_k^2 + (1 - \gamma) t_k \right) \left(\mathcal{L}_\beta(x_k, r_n) - \mathcal{L}_\beta(x_*, \lambda_k) \right)$$

$$\leqslant \mathcal{E}_k(x_*, r_n) - \frac{\alpha - 2}{(\alpha - 1)^2} \langle r_n - \lambda_*, Ax_k - b \rangle$$

$$\leqslant \mathcal{E}_k(x_*, r_n) + \frac{\alpha - 2}{(\alpha - 1)^2} \|r_n - \lambda_*\| \|Ax_k - b\| \tag{3.58a}$$

$$\leqslant \mathcal{E}_k(x_*, r_n) + \frac{\alpha - 2}{(\alpha - 1)^2} \sqrt{\frac{2\mathcal{E}_1(x_*, \lambda_*)}{\beta\gamma}} \frac{1}{t_k} \tag{3.58b}$$

$$\leqslant \mathcal{E}_k(x_*, r_n) + \frac{\alpha - 2}{\kappa(\alpha - 1)^2} \sqrt{\frac{2\mathcal{E}_1(x_*, \lambda_*)}{\beta\gamma}} \frac{1}{k}, \tag{3.58c}$$

where (3.58b) follows from (3.37) and (3.58c) is due to (3.42). By a telescoping sum argument and Lemma 3.16 we conclude that for every $k \geqslant 1$

$$\mathcal{E}_{k+1}(x_*, r_n) \leqslant \mathcal{E}_1(x_*, r_n) + \frac{\alpha - 2}{\kappa(\alpha - 1)^2} \sqrt{\frac{2\mathcal{E}_1(x_*, \lambda_*)}{\beta\gamma}} \sum_{i=1}^k \frac{1}{i}$$

$$\leqslant C_1 + \frac{\alpha - 2}{\kappa(\alpha - 1)^2} \sqrt{\frac{2\mathcal{E}_1(x_*, \lambda_*)}{\beta\gamma}} \Big(\log(k) + 1 \Big) \leqslant C_4 \Big(\log(k) + 1 \Big),$$

where

$$C_4 := C_1 + \frac{\alpha - 2}{\kappa(\alpha - 1)^2} \sqrt{\frac{2\mathcal{E}_1(x_*, \lambda_*)}{\beta\gamma}} > 0.$$

By choosing $k := n - 1$, it yields

$$t_n (t_n - 1 + \gamma)(f(x_n) - f(x_*) + \langle r_n, Ax_n - b \rangle) \leqslant \mathcal{E}_n(x_*, r_n) \leqslant C_4 \Big(\log(n - 1) + 1 \Big).$$

We have seen in the proof of Theorem 3.17 that

$$f(x_n) - f(x_*) + \langle r_n, Ax_n - b \rangle = \mathcal{L}(x_n, \lambda_*) - \mathcal{L}(x_*, \lambda_n) + \|Ax_n - b\|, \tag{3.59}$$

thus, by taking into account (3.42), we obtain

$$\gamma\kappa^2 n^2 \|Ax_n - b\| \leqslant \gamma t_n^2 \|Ax_n - b\| \leqslant t_n (t_n - 1 + \gamma)(\mathcal{L}(x_n, \lambda_*) - \mathcal{L}(x_*, \lambda_n) + \|Ax_n - b\|)$$

$$\leqslant \mathcal{E}_n(x_*, r_n) \leqslant C_4 \Big(\log(n - 1) + 1 \Big),$$

therefore, since $2 + \log(n - 1) \leqslant 2(n - 1)^{1/2}$,

$$\|Ax_n - b\| \leqslant \frac{C_4 \Big(\log(n - 1) + 1 \Big)}{\gamma\kappa^2 n^2} \leqslant \frac{2C_4}{\gamma\kappa^2 n^{3/2}}.$$

Taking into account also Lemma 3.16 and the definition of C_4, we have that for every $k \geqslant 1$

$$\|Ax_k - b\| \leqslant \frac{2C_4}{\gamma\kappa^2 k^{3/2}}.$$

Using this estimate in (3.58a), we obtain for every $k \geqslant 1$

$$\mathcal{E}_{k+1}(x_*, r_n) \leqslant \mathcal{E}_k(x_*, r_n) + \frac{\alpha - 2}{(\alpha - 1)^2} \frac{2C_4}{\gamma\kappa^2 k^{3/2}}.$$

By using once again a the telescoping sum argument, we conclude that for every $k \geqslant 1$

$$\mathcal{E}_{k+1}\left(x_*, r_n\right) \leqslant \mathcal{E}_1\left(x_*, r_n\right) + \frac{2C_4(\alpha-2)}{\gamma\kappa^2\left(\alpha-1\right)^2} \sum_{i=1}^{k} \frac{1}{i^{3/2}}$$

$$\leqslant C_1 + \frac{2C_4(\alpha-2)}{\gamma\kappa^2\left(\alpha-1\right)^2} \sum_{i\geqslant1} \frac{1}{i^{3/2}} < +\infty.$$

From here, (3.56) follows by choosing $k := n - 1$, and by using that $\gamma t_n^2 \leqslant t_n(t_n - 1 + \gamma)$ and (3.59). Statement (3.57) follows from (3.56) by repeating the arguments at the end of the proof of Theorem 3.17. $\qquad\square$

Now we come to the second case, namely, when $m := \frac{2}{\alpha-1} < \gamma \leqslant 1$, which implicitly requires that $\alpha > 3$. For the proof of the fast convergence rates we will make use of the following result which can be found in [18, Lemma 2] (see, also, [17, Lemma 3.18]).

Lemma 3.19. *Let $\{\zeta_k\}_{k\geqslant1} \subseteq \mathcal{G}$ be a sequence such that there exist $\delta > 1$ and $M > 0$ with the property that for every $K \geqslant 1$*

$$\left\| ((\delta-1)K + \delta)\zeta_{K+1} + \sum_{k=1}^{K} \zeta_k \right\| \leqslant M.$$

Then for every $K \geqslant 1$ it holds

$$\left\| \sum_{k=1}^{K} \zeta_k \right\| \leqslant M.$$

Theorem 3.20. *Let $\{(x_k, \lambda_k)\}_{k\geqslant0}$ be the sequence generated by Algorithm 1 with the sequence $\{t_k\}_{k\geqslant1}$ chosen to satisfy Chambolle-Dossal rule (3.45), $m := \frac{2}{\alpha-1} < \gamma \leqslant 1$, $\beta > 0$, and $(x_*, \lambda_*) \in \mathcal{S}$. Then for every $k \geqslant 1$ it holds*

$$0 \leqslant \mathcal{L}\left(x_k, \lambda_*\right) - \mathcal{L}\left(x_*, \lambda_k\right) \leqslant \frac{\mathcal{E}_1\left(x_*, \lambda_*\right)}{\gamma t_k^2}, \tag{3.60}$$

$$0 \leqslant \|Ax_k - b\| \leqslant \frac{C_6}{t_k^2}, \tag{3.61}$$

and

$$-\frac{\|\lambda_*\| C_6}{t_k^2} \leqslant f\left(x_k\right) - f\left(x_*\right) \leqslant \frac{1}{t_k^2}\left(\frac{\mathcal{E}_1\left(x_*, \lambda_*\right)}{\gamma} + \|\lambda_*\| C_6\right), \tag{3.62}$$

where

$$C_6 := 2\left(1 + \varphi_m\right)^2\left(2\left(\alpha-1\right)^2 \frac{\gamma}{\rho} \sup_{k\geqslant1} \|\nu_k\| + \left(\alpha-1\right)^2\gamma\|Ax_1 - b\| + \frac{1}{\kappa}\left(|\omega_0| + |\omega_1|\right)\sqrt{\frac{2\mathcal{E}_1\left(x_*, \lambda_*\right)}{\beta\gamma}}\right),$$

with

$$\delta := 1 + \frac{1}{\gamma(\alpha-1)-2} > 1,$$

$$\omega_0 := \delta\left(\alpha-2\right) - 2\left(\alpha-1\right) \quad and \quad \omega_1 := \left(\delta-1\right)\left(\alpha-2\right) - 1.$$

Proof. Relation (3.60) follows from (3.37). We fix $K \geqslant 1$. For every $1 \leqslant k \leqslant K$, according to (3.3g), we have

$$t_{k+1}\left(Az_{k+1}^\gamma - \gamma b\right) = t_{k+1}\left(t_{k+1} - 1 + \gamma\right)\left(Ax_{k+1} - b\right) - t_{k+1}\left(t_{k+1} - 1\right)\left(Ax_k - b\right)$$

$$= t_{k+1}\left(t_{k+1} - 1 + \gamma\right)\left(Ax_{k+1} - b\right) - t_k\left(t_k - 1 + \gamma\right)\left(Ax_k - b\right)$$

$$+ \left(t_k\left(t_k - 1 + \gamma\right) - t_{k+1}\left(t_{k+1} - 1\right)\right)\left(Ax_k - b\right).$$

Taking into consideration (3.11), (3.3h) and (3.55), by a telescoping argument it yields

$$
\begin{aligned}
(\alpha - 1)^2 \frac{\gamma}{\rho} \left(\nu_{K+1} - \nu_1\right) &= (\alpha - 1)^2 \frac{\gamma}{\rho} \sum_{k=1}^{K} \left(\nu_{k+1} - \nu_k\right) = (\alpha - 1)^2 \sum_{k=1}^{K} t_{k+1} \left(A z_{k+1}^{\gamma} - \gamma b\right) \\
&= (\alpha - 1)^2 t_{K+1} \left(t_{K+1} - 1 + \gamma\right) \left(A x_{K+1} - b\right) - (\alpha - 1)^2 \gamma \left(A x_1 - b\right) \\
&\quad + (\alpha - 1)^2 \sum_{k=1}^{K} \left(t_k \left(t_k - 1 + \gamma\right) - t_{k+1} \left(t_{k+1} - 1\right)\right) \left(A x_k - b\right) \\
&= (K + \alpha - 1) \left(K + \gamma(\alpha - 1)\right) \left(A x_{K+1} - b\right) - (\alpha - 1)^2 \gamma \left(A x_1 - b\right) \\
&\quad + \sum_{k=1}^{K} \left(\left(\gamma(\alpha - 1) - 2\right) \left(k + \alpha - 2\right) + \alpha - 2 \right) \left(A x_k - b\right). \quad (3.63)
\end{aligned}
$$

We define

$$
\delta := 1 + \frac{1}{\gamma(\alpha - 1) - 2} > 1,
$$

$$
\omega_0 := \delta \left(\alpha - 2\right) - 2 \left(\alpha - 1\right) \quad \text{and} \quad \omega_1 := (\delta - 1) \left(\alpha - 2\right) - 1,
$$

and

$$
\zeta_k := \left(\left(\gamma(\alpha - 1) - 2\right) \left(k + \alpha - 2\right) + \alpha - 2 \right) \left(A x_k - b\right) \text{ for } k = 1, ..., K.
$$

It holds

$$
\begin{aligned}
& (K + \alpha - 1) \left(K + \gamma(\alpha - 1)\right) \left(A x_{K+1} - b\right) \\
&= (\delta - 1) K \left(\left(\gamma(\alpha - 1) - 2\right) \left(K + \alpha - 1\right) + \alpha - 2 \right) \left(A x_{K+1} - b\right) \\
&\quad + \delta \left(\left(\gamma(\alpha - 1) - 2\right) \left(K + \alpha - 1\right) + \alpha - 2 \right) \left(A x_{K+1} - b\right) - \left(\omega_1 K + \omega_0\right) \left(A x_{K+1} - b\right) \\
&= \left((\delta - 1) K + \delta\right) \zeta_{K+1} - \left(\omega_1 K + \omega_0\right) \left(A x_{K+1} - b\right). \quad (3.64)
\end{aligned}
$$

Furthermore, it follows from (3.42) and (3.37) that

$$
\begin{aligned}
\left\| \left(\omega_1 K + \omega_0\right) \left(A x_{K+1} - b\right) \right\| &\leqslant \left(|\omega_0| + |\omega_1|\right) \left(K + 1\right) \left\| A x_{K+1} - b \right\| \\
&\leqslant \frac{1}{\kappa} \left(|\omega_0| + |\omega_1|\right) t_{K+1} \left\| A x_{K+1} - b \right\| \\
&\leqslant \frac{1}{\kappa} \left(|\omega_0| + |\omega_1|\right) \sqrt{\frac{2 \mathcal{E}_1 \left(x_*, \lambda_*\right)}{\beta \gamma}}. \quad (3.65)
\end{aligned}
$$

Combining the relations (3.63), (3.64) and (3.65), we get via the triangle inequality

$$
\begin{aligned}
& \left\| \left((\delta - 1) K + \delta\right) \zeta_{K+1} + \sum_{k=1}^{K} \zeta_k \right\| \\
&= \left\| (\alpha - 1)^2 \frac{\gamma}{\rho} \left(\nu_{K+1} - \nu_1\right) + (\alpha - 1)^2 \gamma \left(A x_1 - b\right) + \left(\omega_1 K + \omega_0\right) \left(A x_{K+1} - b\right) \right\| \\
&\leqslant (\alpha - 1)^2 \frac{\gamma}{\rho} \left\| \nu_{K+1} - \nu_1 \right\| + (\alpha - 1)^2 \gamma \left\| A x_1 - b \right\| + \left\| \left(\omega_1 K + \omega_0\right) \left(A x_{K+1} - b\right) \right\| \\
&\leqslant C_7 := 2 \left(\alpha - 1\right)^2 \frac{\gamma}{\rho} \sup_{k \geqslant 1} \left\| \nu_k \right\| + (\alpha - 1)^2 \gamma \left\| A x_1 - b \right\| + \frac{1}{\kappa} \left(|\omega_0| + |\omega_1|\right) \sqrt{\frac{2 \mathcal{E}_1 \left(x_*, \lambda_*\right)}{\beta \gamma}} < +\infty, \\
& \hspace{13cm} (3.66)
\end{aligned}
$$

where we also recall that, due to Proposition 3.13, it holds $\sup_{k \geqslant 1} \left\| \nu_k \right\| < +\infty$.

Inequality (3.66) holds for every $K \geqslant 1$ (notice that C_7 is independent of K), consequently, we can apply Lemma 3.19 to conclude that $\left\| \sum_{k=1}^{K} \zeta_k \right\| \leqslant C_7$ for every $K \geqslant 1$. By using again the triangle inequality and (3.66), we obtain for every $K \geqslant 1$ that

$$(\delta - 1) K \left\| \zeta_{K+1} \right\| \leqslant \left\| ((\delta - 1) K + \delta) \zeta_{K+1} \right\| \leqslant 2C_7. \tag{3.67}$$

Using the inequality (3.14) in Lemma 3.5, we see that for every $K \geqslant 1$ it holds

$$\frac{t_{K+1}^2}{(1 + \varphi_m)^2} \left\| Ax_{K+1} - b \right\| \leqslant K^2 \left\| Ax_{K+1} - b \right\| \leqslant (\delta - 1) K \left\| \zeta_{K+1} \right\|. \tag{3.68}$$

By combining (3.67) and (3.68), we obtain (3.61).

Statement (3.62) follows from (3.60) and (3.61) by repeating the arguments at the end of the proof of Theorem 3.17. $\qquad\square$

3.4.3 The Attouch-Cabot ([1]) rule

Another inertial parameter rule used in the literature in the context of fast numerical algorithms is the one proposed by Attouch and Cabot in [1], which reads for $\alpha \geqslant 3$

$$t_k := \frac{k-1}{\alpha - 1} \quad \forall k \geqslant 1.$$

This sequence is monotonically increasing and it fulfills (3.2) with $m := \dfrac{2}{\alpha - 1} \leqslant 1$ as, for every $k \geqslant 1$, it holds

$$t_{k+1}^2 - mt_{k+1} - t_k^2 = (t_{k+1} - t_k)(t_{k+1} + t_k) - mt_{k+1} = \frac{1}{\alpha-1}\frac{2k-1}{\alpha-1} - \frac{2}{\alpha-1}\frac{k}{\alpha-1}$$

$$= -\frac{1}{(\alpha-1)^2} < 0.$$

This shows that the sequence $\{t_k\}_{k \geqslant 1}$ has very much in common with the Chambolle-Dossal parameter rule. The only significant difference is that is starts at 0 and that $t_k \geqslant 1$ holds only for $k \geqslant k_1 := \lfloor \alpha \rfloor + 1$. Consequently, the fast convergence rate results for the primal-dual gap, the feasibility measure and the objective function value are valid also for the Attouch-Cabot rule. This can be easily seen by slightly adapting the proofs made in the setting of the Chambolle-Dossal rule by taking into consideration that some of the estimates hold only for $k \geqslant k_1$. This exercise is left to the reader.

4 Convergence of the iterates

In this section we will turn our attention to the convergence of the sequence of primal-dual iterates generated by Algorithm 1 to a primal-dual solution of (1.1). First, we will prove that the first assumption in the Opial Lemma is verified and to this end we will need the following technical lemma.

Lemma 4.1. *Let $\{\theta_k\}_{k \geqslant 1}, \{a_k\}_{k \geqslant 1}, \{t_k\}_{k \geqslant 1}$ be real sequences such that $\{a_k\}_{k \geqslant 1}$ is bounded from below and $\{t_k\}_{k \geqslant 1}$ is nondecreasing and bounded from below by 1, and $\{d_k\}_{k \geqslant 1}$ be a nonnegative sequence such that for every $k \geqslant 1$*

$$a_{k+1} \leqslant a_k + \theta_{k+1}, \tag{4.1a}$$

$$t_{k+1}\theta_{k+1} \leqslant (t_k - 1)\theta_k + d_k. \tag{4.1b}$$

If $\sum_{k \geqslant 1} d_k < +\infty$, then the sequence $\{a_k\}_{k \geqslant 1}$ is convergent.

Proof. It follows from (4.1b) that for every $k \geqslant 1$

$$t_{k+1}\theta_{k+1} \leqslant (t_k - 1)\,\theta_k + d_k \leqslant (t_k - 1)\,[\theta_k]_+ + d_k, \tag{4.2}$$

where $[\cdot]_+$ denotes the positive part. Since the right-hand side of this inequality is nonnegative, it yields that for every $k \geqslant 1$

$$[\theta_k]_+ \leqslant t_k\,[\theta_k]_+ - t_{k+1}\,[\theta_{k+1}]_+ + d_k.$$

which, by telescoping cancellation, gives $\sum_{k \geqslant 1}[\theta_k]_+ < +\infty$.

According to (4.1a), we have that for every $k \geqslant 1$ it holds

$$a_{k+1} \leqslant a_k + \theta_{k+1} \leqslant a_k + [\theta_{k+1}]_+.$$

By using Lemma 1.1 we obtain from here that the sequence $\{a_k\}_{k \geqslant 1}$ is convergent. $\qquad\square$

Proposition 4.2. *Let $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ be the sequence generated by Algorithm 1 with $0 < m < \gamma < 1$. Then for every $(x_*, \lambda_*) \in \mathcal{S}$ the limit $\lim\limits_{k \to +\infty} \|(x_k, \lambda_k) - (x_*, \lambda_*)\|_{\mathcal{W}}$ exists.*

Proof. Let $(x_*, \lambda_*) \in \mathcal{S}$ be fixed. For brevity we will write

$$u_* := (x_*, \lambda_*) \in \mathcal{S} \text{ and } u_k := (x_k, \lambda_k) \in \mathcal{H} \times \mathcal{G} \quad \forall k \geqslant 0.$$

It follows from (3.35) that $\mathcal{E}_{k+1}(x_*, \lambda_*) \leqslant \mathcal{E}_k(x_*, \lambda_*)$ for every $k \geqslant 1$. In view of (3.43), after rearranging some terms, we get for every $k \geqslant 1$

$$t_{k+1}(t_{k+1} - 1 + \gamma)\left(\mathcal{L}_\beta(x_{k+1}, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_{k+1}) + \frac{1}{2}\|u_{k+1} - u_k\|_{\mathcal{W}}^2\right)$$

$$+ \frac{\gamma}{2}t_{k+1}\left(\|u_{k+1} - u_*\|_{\mathcal{W}}^2 - \|u_k - u_*\|_{\mathcal{W}}^2\right) + \frac{1-\gamma}{2\rho}t_{k+1}\|\lambda_{k+1} - \lambda_k\|^2$$

$$\leqslant (t_k - 1)(t_k - 1 + \gamma)\left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k) + \frac{1}{2}\|u_k - u_{k-1}\|_{\mathcal{W}}^2\right)$$

$$+ \frac{\gamma}{2}(t_k - 1)\left(\|u_k - u_*\|_{\mathcal{W}}^2 - \|u_{k-1} - u_*\|_{\mathcal{W}}^2\right) + \frac{1-\gamma}{2\rho}(t_k - 1)\|\lambda_k - \lambda_{k-1}\|^2$$

$$+ (t_k - 1 + \gamma)\left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k)\right) + \frac{1}{2}(t_{k+1} - 1 + \gamma)\|u_{k+1} - u_k\|_{\mathcal{W}}^2$$

$$+ \frac{1-\gamma}{2\rho}\|\lambda_{k+1} - \lambda_k\|^2. \tag{4.3}$$

Set $a_0 := \dfrac{\gamma}{2}\|u_0 - u_*\|_{\mathcal{W}}^2 \geqslant 0$ and for every $k \geqslant 1$

$$a_k := \frac{\gamma}{2}\|u_k - u_*\|_{\mathcal{W}}^2 \geqslant 0,$$

$$\theta_k := (t_k - 1 + \gamma)\left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k) + \frac{1}{2}\|u_k - u_{k-1}\|_{\mathcal{W}}^2\right)$$

$$+ (a_k - a_{k-1}) + \frac{1-\gamma}{2\rho}\|\lambda_k - \lambda_{k-1}\|^2,$$

$$d_k := (t_k - 1 + \gamma)\left(\mathcal{L}_\beta(x_k, \lambda_*) - \mathcal{L}_\beta(x_*, \lambda_k)\right) + \frac{1}{2}(t_{k+1} - 1 + \gamma)\|u_{k+1} - u_k\|_{\mathcal{W}}^2$$

$$+ \frac{1-\gamma}{2\rho}\|\lambda_{k+1} - \lambda_k\|^2 \geqslant 0.$$

We notice that for every $k \geqslant 1$ the estimate (4.3) becomes (4.1b), while (4.1a) obviously holds. As $0 < m < \gamma < 1$, it follows from Proposition 3.9 that $\sum_{k \geqslant 1} d_k < +\infty$.

Hence, we can apply Lemma 4.1 to conclude that $\{\|(x_k, \lambda_k) - (x_*, \lambda_*)\|_{\mathcal{W}}\}_{k \geqslant 1}$ is convergent. $\qquad\square$

The following result is the discrete counterpart of [8, Theorem 4.7] (see (2.2)). Its proof is a direct consequence of Proposition 3.9 and Proposition 3.12.

Theorem 4.3. *Let $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ be the sequence generated by Algorithm 1 with the sequence $\{t_k\}_{k \geqslant 1}$ chosen to satisfy (3.42), $0 < m < \gamma \leqslant 1$, $0 < \sigma < \dfrac{\gamma}{L + \gamma\beta\|A\|^2}$, $\beta > 0$, and $(x_*, \lambda_*) \in \mathcal{S}$. Then it holds*

$$\|\nabla f(x_k) - \nabla f(x_*)\| = o\left(\frac{1}{\sqrt{k}}\right) \quad and \quad \|A^*\lambda_k - A^*\lambda_*\| = o\left(\frac{1}{\sqrt{k}}\right) \ as \ k \to +\infty.$$

consequently,

$$\|\nabla_x \mathcal{L}(x_k, \lambda_k)\| = \|\nabla f(x_k) + A^*\lambda_k\| = o\left(\frac{1}{\sqrt{k}}\right) \ as \ k \to +\infty,$$

and

$$\|\nabla_\lambda \mathcal{L}(x_k, \lambda_k)\| = \|Ax_k - b\| = o\left(\frac{1}{\sqrt{k}}\right) \ as \ k \to +\infty.$$

As seen in Section 3.4, if, in addition, $\{t_k\}_{k \geqslant 1}$ is chosen to satisfy Chambolle-Dossal or Attouch-Cabot rule and $m := \frac{2}{\alpha - 1}$, then

$$\|\nabla_\lambda \mathcal{L}(x_k, \lambda_k)\| = \|Ax_k - b\| = \mathcal{O}\left(\frac{1}{k^2}\right) \ as \ k \to +\infty.$$

Now we can prove the main theorem of this section establishing the convergence of the sequence of iterates generated by Algorithm 1.

Theorem 4.4. *Let $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ be the sequence generated by Algorithm 1 with the sequence $\{t_k\}_{k \geqslant 1}$ chosen to satisfy (3.42), $0 < m < \gamma < 1$, $0 < \sigma < \dfrac{\gamma}{L + \gamma\beta\|A\|^2}$ and $\beta > 0$. Then the sequence $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ converges weakly to a primal-dual optimal solution of (1.1).*

Proof. From Proposition 4.2 it follows that the limit $\lim_{k \to +\infty} \|(x_k, \lambda_k) - (x_*, \lambda_*)\|$ exists for every $(x_*, \lambda_*) \in \mathcal{S}$. This proves the first condition of Lemma 1.2.

In order to prove condition (ii), let $\left(\widetilde{x}, \widetilde{\lambda}\right) \in \mathcal{H} \times \mathcal{G}$ be an arbitrary weak sequential cluster point of $\{(x_k, \lambda_k)\}_{k \geqslant 0}$. This means that there exists a subsequence $\{(x_{k_n}, \lambda_{k_n})\}_{n \geqslant 0}$ which converges weakly to $\left(\widetilde{x}, \widetilde{\lambda}\right)$ as $n \to +\infty$. According to Theorem 4.3 we have $\nabla f(x_k) + A^*\lambda_k \to 0$ and $Ax_k - b \to 0$ as $k \to +\infty$, hence,

$$\nabla f(x_{k_n}) + A^*\lambda_{k_n} \to 0 \quad and \quad Ax_{k_n} - b \to 0 \qquad as \qquad n \to +\infty.$$

Since the graph of the operator $\mathcal{T}_\mathcal{L}$ is sequentially closed in $(\mathcal{H} \times \mathcal{G})^{\text{weak}} \times (\mathcal{H} \times \mathcal{G})^{\text{strong}}$ (cf. [5, Proposition 20.38]), it follows from here that

$$\begin{cases} \nabla f(\widetilde{x}) + A^*\widetilde{\lambda} & = 0 \\ A\widetilde{x} - b & = 0 \end{cases}.$$

In other words, $\left(\widetilde{x}, \widetilde{\lambda}\right) \in \mathcal{S}$ and the proof is complete. $\qquad \square$

Remark 4.5. If the sequence $\{t_k\}_{k \geqslant 1}$ is chosen to satisfy the Chambolle-Dossal or the Attouch-Cabot rule with

$$\alpha > 3, \quad m := \frac{1}{\alpha - 2} < \gamma < 1, \quad 0 < \sigma < \frac{\gamma}{L + \gamma\beta\|A\|^2} \quad and \quad \beta > 0,$$

then Theorem 4.4 guarantees that the sequence $\{(x_k, \lambda_k)\}_{k \geqslant 0}$ converges weakly to a primal-dual optimal solution of (1.1). This statement is in addition to the fast convergence rates of order $\mathcal{O}\left(1/k^2\right)$ for the primal-dual gap, the feasibility measure, and the objective function value.

If the sequence $\{t_k\}_{k \geqslant 1}$ is chosen to satisfy the Nesterov rule, then, as we have seen, the fast convergence rate results also hold, however, since in this setting $m = \gamma = 1$, one cannot apply Theorem 4.4 to obtain the convergence of the iterates. This is consistent with the unconstrained case for which it is also not known if the sequence of iterates generated by the fast gradient method with inertial parameters following the Nesterov rule converges.

References

[1] **H. Attouch, A. Cabot**. *Convergence rates of inertial forward-backward algorithms.* SIAM Journal on Optimization 28(1), 849–874 (2018)

[2] **H. Attouch, Z. Chbani, J. Fadili, H. Riahi**. *Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics.* Journal of Optimization Theory and Applications 193, 704–736 (2022)

[3] **H. Attouch, J. Peypouquet** *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$.* SIAM Journal on Optimization 26(3), 1824–1834 (2016)

[4] **H.H. Bauschke, M.N. Bui, X. Wang** *Applying FISTA to optimization problems (with or) without minimizers.* Mathematical Programming 184, 349–381 (2020)

[5] **H.H. Bauschke, P.L. Combettes**. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics. Springer, New York (2017)

[6] **A. Beck, M. Teboulle**. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems.* SIAM Journal on Imaging Sciences 2(1), 183–202 (2009)

[7] **R.I. Boţ**. *Conjugate Duality in Convex Optimization.* Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer, Berlin Heidelberg (2010)

[8] **R.I. Boţ, D.-K. Nguyen**. *Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping.* Journal of Differential Equations 303, 369–406 (2021)

[9] **S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein**. *Distributed optimization and statistical learning via the alternating direction method of multipliers.* Foundations and Trends in Machine Learning 3(1), 1–122 (2010)

[10] **A. Chambolle, C. Dossal**. *On the convergence of the iterates of the "Fast Iterative Shrinkage/Thresholding Algorithm".* Journal of Optimization theory and Applications 166(3), 968–982 (2016)

[11] **A. Chambolle, T. Pock**. *On the ergodic convergence rates of a first-order primal-dual algorithm.* Mathematical Programming 159, 253–287 (2016)

[12] **Y. Chen, G. Lan, Y. Ouyang**. *Optimal primal-dual methods for a class of saddle point problems.* SIAM Journal on Optimization 24(4), 1779–1814 (2014)

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

[14] **B.S. He, X.M. Yuan**. *On the acceleration of augmented Lagrangian method for linearly constrained optimization.* Optimization Online `http://www.optimization-online.org/DB_HTML/2010/10/2760.html`

[15] **X. He, R. Hu, Y.-P. Fang**. *Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems.* Numerical Algorithms `doi.org/10.1007/s11075-021-01246-y`

[16] **X. He, R. Hu, Y.-P. Fang**. *Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems.* SIAM Journal on Control and Optimization 59(5), 3278–3301 (2021)

[17] **H. Li, Z. Lin, C. Fang**. *Accelerated Optimization for Machine Learning.* Springer, Singapore (2019)

[18] **H. Li, Z. Lin**. *Accelerated Alternating Direction Method of Multipliers: an optimal $O(1/K)$ nonergodic analysis.* Journal of Scientific Computing 79, 671–699 (2019)

[19] **H. Lou**. *Accelerated primal-dual methods for linearly constrained convex optimization problems.* `arXiv:2109.12604`

[20] **H. Lou**. *A unified differential equation solver approach for separable convex optimization: splitting, acceleration and nonergodic rate.* `arXiv:2109.13467`

[21] **Y. Nesterov**. *A method of solving a convex programming problem with convergence rate $\mathcal{O}\left(1/k^2\right)$.* Soviet Mathematics Doklady 27, 372–376 (1983)

[22] **Y. Nesterov**. *Introductory Lectures on Convex Optimization.* Springer, New York (2004)

[23] **Y. Ouyang, Y. Chen, G. Lan, E. Pasiliao Jr.**. *An accelerated linearized Alternating Direction Method of Multipliers.* SIAM Journal on Imaging Sciences 8(1), 644–681 (2015)

[24] **S. Sabach, M. Teboulle**. *Faster Lagrangian-based methods in convex optimization.* SIAM Journal on Optimization 32(1), 204–227 (2022)

[25] **W. Su, S. Boyd, E. Candès**. *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights.* Journal of Machine Learning Research 17(153), 1–43 (2016)

[26] **Q. Tran-Dinh, O. Fercoq, V. Cevher**. *A smooth primal-dual optimization framework for nonsmooth convex optimization.* SIAM Journal on Optimization 28(1), 96–134 (2018)

[27] **P. Tseng**. *Applications of splitting algorithm to decomposition in convex programming and variational inequalities.* SIAM Journal on Control and Optimization 29(1), 119–138 (1991)

[28] **Y. Xu**. *Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming.* SIAM Journal on Optimization 27(3), 1459–1484 (2017)

[29] **S. Yan, N. He**. *Bregman Augmented Lagrangian and its acceleration.* `arXiv:2002.06315`

[30] **X. Zeng, J. Lei, J. Chen**. *Dynamical primal-dual accelerated method with applications to network optimization.* IEEE Transactions on Automatic Control, `doi.org/10.1109/TAC.2022.3152720`