Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics

**Hedy Attouch · Zaki Chbani · Jalal Fadili ·
Hassan Riahi**

March 23, 2021

Abstract In this paper, we propose in a Hilbertian setting a second-order time-continuous dynamic system with fast convergence guarantees to solve structured convex minimization problems with an affine constraint. The system is associated with the augmented Lagrangian formulation of the minimization problem. The corresponding dynamics brings into play three general time-varying parameters, each with specific properties, and which are respectively associated with viscous damping, extrapolation and temporal scaling. By appropriately adjusting these parameters, we develop a Lyapunov analysis which provides fast convergence properties of the values and of the feasibility gap. These results will naturally pave the way for developing corresponding accelerated ADMM algorithms, obtained by temporal discretization.

Keywords augmented Lagrangian; ADMM; damped inertial dynamics; convex constrained minimization; convergence rates; Lyapunov analysis; Nesterov accelerated gradient method; temporal scaling

AMS subject classification 37N40, 46N10, 49M30, 65B99, 65K05, 65K10, 90B50, 90C25.

H. Attouch
IMAG, Univ. Montpellier, CNRS, Montpellier, France
E-mail: hedy.attouch@umontpellier.fr

Z. Chbani
Cadi Ayyad Univ., Faculty of Sciences Semlalia, Mathematics, 40000 Marrakech, Morroco
E-mail: chbaniz@uca.ac.ma

J. Fadili
GREYC CNRS UMR 6072, Ecole Nationale Supérieure d'Ingénieurs de Caen, France
E-mail: Jalal.Fadili@greyc.ensicaen.fr

H. Riahi
Cadi Ayyad Univ., Faculty of Sciences Semlalia, Mathematics, 40000 Marrakech, Morroco
E-mail: h-riahi@uca.ac.ma

1 Introduction

Our paper is part of the active research stream that studies the relationship between continuous-time dissipative dynamical systems and optimization algorithms. From this perspective, damped inertial dynamics offer a natural way to accelerate these systems. An abundant literature has been devoted to the design of the damping terms, which is the basic ingredient of the optimization properties of these dynamics. In line with the seminal work of Polyak on the heavy ball method with friction [45,46], the first studies have focused on the case of a fixed viscous damping coefficient [1,17,2]. A decisive step was taken in [52] where the authors considered inertial dynamics with an asymptotic vanishing viscous damping coefficient. In doing so, they made the link with the accelerated gradient method of Nesterov [42,41,25] for unconstrained convex minimization. This has resulted in a flurry of research activity; see *e.g.* [7,8,9,12,11,13,18,20,21,23,3,27,31,39,51,53].

In this paper, we consider the case of affinely *constrained* convex structured minimization problems. To bring back the problem to the unconstrained case, there are two main ways: either penalize the constraint (by external penalization or an internal barrier method), or use (augmented) Lagrangian multiplier methods.

Accounting for approximation/penalization terms within dynamical systems has been considered in a series of papers; see [16,29] and the references therein. It is a flexible approach which can be applied to non-convex problems and/or ill-posed problems, making it a valuable tool for inverse problems. Its major drawback is that in general, it requires a subtle tuning of the approximation/penalization parameter.

Here, we will consider the augmented Lagrangian approach and study the convergence properties of a second-order inertial dynamic with damping, which is attached to the augmented Lagrangian formulation of the affinely constrained convex minimization problem. The proposed dynamical system can be viewed as an inertial continuous-time counterpart of the ADMM method originally proposed in the mid-1970s and which has gained considerable interest in the recent years, in particular for solving large-scale composite optimization problems arising in data science. Among the novelties of our work, the dynamics we propose involves three parameters which vary in time. These are associated with viscous damping, extrapolation, and temporal scaling. By properly adjusting these parameters, we will provide fast convergence rates both for the values and the feasibility gap. The balance between the viscosity parameter (which tends towards zero) and the extrapolation parameter (which tends towards infinity) has already been developed in [54], [36] and [14], though for different problems. Temporal scaling techniques were considered in [6] for the case of convex minimization without affine constraint; see also [10,12,15]. Thus, another key contribution of this paper is to show that the temporal scaling and extrapolation can be extended to the class of ADMM-type methods with improved convergence rates. Working with general coefficients and in general Hilbert spaces allows us to encompass the results obtained in the above-mentioned papers and to broaden their scope.

It has been known for a long time that the optimality conditions of the (augmented) Lagrangian formulation of convex structured minimization problems with an affine constraint can be equivalently formulated as a monotone inclusion problem; see [50,48,49]. In turn, the problem can be converted into finding the zeros of a

maximally monotone operator, and can therefore be attacked using inertial methods for solving monotone inclusions. In this regard, let us mention the following recent works concerning the acceleration of ADMM methods via continuous-time inertial dynamics:

- In [26], the authors proposed an inertial ADMM by making use of the inertial version of the Douglas-Rachford splitting method for monotone inclusion problems recently introduced in [28], in the context of concomitantly solving a convex minimization problem and its Fenchel dual; see also [34, 43, 44, 47] in the purely discrete setting.
- Attouch [5] uses the maximally monotone operator which is associated with the augmented Lagrangian formulation of the problem, and specializes to this operator the inertial proximal point algorithm recently developed in [19] to solve general monotone inclusions. This gives rise to an inertial proximal ADMM algorithm where an appropriate adjustment of the viscosity and proximal parameters gives provably fast convergence properties, as well as the convergence of the iterates to saddle points of the Lagrangian function. This approach is in line with [22] who considered the case without inertia. But this approach fails to achieve a fully split inertial ADMM algorithm.

Contents In Section 2, we introduce the inertial second-order dynamical system with damping (coined (TRIALS)) which is attached to the augmented Lagrangian formulation. In Section 3, which is the main part of the paper, we develop a Lyapunov analysis to establish the asymptotic convergence properties of (TRIALS). This gives rise to a system of inequalities-equalities which must be satisfied by the parameters of the dynamics. From the energy estimates thus obtained, we show in Section 4 that the Cauchy problem attached to (TRIALS) is well-posed, *i.e.* existence and possibly uniqueness of a global solution. In Section 5, we examine the case of the uniformly convex objectives. In Section 6, we provide specific choices of the system parameters that satisfy our assumptions and achieve fast convergence rates. This is then supplemented by preliminary numerical illustrations. Some conclusions and perspectives are finally outlined in Section 7.

2 Problem statement

Consider the structured convex optimization problem:

$$\min_{x \in \mathcal{X}, \, y \in \mathcal{Y}} F(x, y) := f(x) + g(y) \quad \text{subject to } Ax + By = c, \tag{$\mathcal{P}$}$$

where, throughout the paper, we make the following standing assumptions:

$$\begin{cases} \mathcal{X}, \mathcal{Y}, \mathcal{Z} \text{ are real Hilbert spaces;} \\ f : \mathcal{X} \to \mathbb{R}, \; g : \mathcal{Y} \to \mathbb{R} \text{ are convex functions of class } \mathcal{C}^1; \\ A : \mathcal{X} \to \mathcal{Z}, B : \mathcal{Y} \to \mathcal{Z} \text{ are linear continuous operators, } c \in \mathcal{Z}; \\ \text{The solution set of } (\mathcal{P}) \text{ is non-empty.} \end{cases} \tag{$\mathcal{H}_{\mathcal{P}}$}$$

Throughout, we denote by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ the scalar product and corresponding norm associated to any of $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$, and the underlying space is to be understood from the context.

2.1 Augmented Lagrangian formulation

Classically, $(\mathcal{P})$ can be equivalently reformulated as the saddle point problem

$$\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \max_{\lambda\in\mathcal{Z}} \mathcal{L}(x,y,\lambda), \tag{1}$$

where $\mathcal{L} : \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\to\mathbb{R}$ is the Lagrangian associated with (1)

$$\mathcal{L}(x,y,\lambda) := F(x,y) + \langle \lambda, Ax + By - c\rangle. \tag{2}$$

Under our standing assumption $(\mathcal{H}_\mathcal{P})$, $\mathcal{L}$ is convex with respect to $(x,y)\in\mathcal{X}\times\mathcal{Y}$, and affine (and hence concave) with respect to $\lambda\in\mathcal{Z}$. A pair $(x^\star, y^\star)$ is optimal for $(\mathcal{P})$, and $\lambda^\star$ is a corresponding Lagrange multiplier if and only if $(x^\star, y^\star, \lambda^\star)$ is a saddle point of the Lagrangian function $\mathcal{L}$, *i.e.* for every $(x,y,\lambda)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$,

$$\mathcal{L}(x^\star, y^\star, \lambda) \leq \mathcal{L}(x^\star, y^\star, \lambda^\star) \leq \mathcal{L}(x, y, \lambda^\star). \tag{3}$$

We denote by $\mathscr{S}$ the set of saddle points of $\mathcal{L}$. The corresponding optimality conditions read

$$(x^\star, y^\star, \lambda^\star)\in\mathscr{S} \iff \begin{cases} \nabla_x\mathcal{L}(x^\star, y^\star, \lambda^\star) = 0 \\ \nabla_y\mathcal{L}(x^\star, y^\star, \lambda^\star) = 0 \\ \nabla_\lambda\mathcal{L}(x^\star, y^\star, \lambda^\star) = 0 \end{cases} \iff \begin{cases} \nabla f(x^\star) + A^*\lambda^\star = 0 \\ \nabla g(y^\star) + B^*\lambda^\star = 0 \\ Ax^\star + By^\star - c = 0 \end{cases}, \tag{4}$$

where we use the classical notations: ∇f and ∇g are the gradients of f and g, A^* is the adjoint operator of A, and similarly for B. The operator ∇_z is the gradient of the corresponding multivariable function with respect to variable z. Given $\mu > 0$, the augmented Lagrangian $\mathcal{L}_\mu : \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\to\mathbb{R}$ associated with the problem $(\mathcal{P})$, is defined by

$$\mathcal{L}_\mu(x,y,\lambda) := \mathcal{L}(x,y,\lambda) + \frac{\mu}{2}\|Ax + By - c\|^2. \tag{5}$$

Observe that one still has $(x^\star, y^\star, \lambda^\star)\in\mathscr{S} \iff \begin{cases} \nabla_x\mathcal{L}_\mu(x^\star, y^\star, \lambda^\star) = 0, \\ \nabla_y\mathcal{L}_\mu(x^\star, y^\star, \lambda^\star) = 0, \\ \nabla_\lambda\mathcal{L}_\mu(x^\star, y^\star, \lambda^\star) = 0. \end{cases}$

2.2 The inertial system (TRIALS)

We will study the asymptotic behaviour, as $t\to+\infty$, of the inertial system:

> (TRIALS): Temporally Rescaled Inertial Augmented Lagrangian System.
>
> $$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + b(t)\nabla_x\mathcal{L}_\mu\Big(x(t), y(t), \lambda(t) + \alpha(t)\dot{\lambda}(t)\Big) = 0 \\ \ddot{y}(t) + \gamma(t)\dot{y}(t) + b(t)\nabla_y\mathcal{L}_\mu\Big(x(t), y(t), \lambda(t) + \alpha(t)\dot{\lambda}(t)\Big) = 0 \\ \ddot{\lambda}(t) + \gamma(t)\dot{\lambda}(t) - b(t)\nabla_\lambda\mathcal{L}_\mu\Big(x(t) + \alpha(t)\dot{x}(t), y(t) + \alpha(t)\dot{y}(t), \lambda(t)\Big) = 0, \end{cases}$$

for $t\in[t_0, +\infty[$ with initial conditions $(x(t_0), y(t_0), \lambda(t_0))$ and $(\dot{x}(t_0), \dot{y}(t_0), \dot{\lambda}(t_0))$. The parameters of (TRIALS) play the following roles:

- $\gamma(t)$ is a viscous damping parameter,
- $\alpha(t)$ is an extrapolation parameter,
- $b(t)$ is attached to the temporal scaling of the dynamic.

In the sequel, we make the following standing assumption on these parameters:

$$\gamma, \alpha, b : [t_0, +\infty[\to \mathbb{R}^+ \text{ are non-negative continuously differentiable functions. } (\mathcal{H}_{\mathcal{D}})$$

Plugging the expression of the partial gradients of $\mathcal{L}_\mu$ into the above system, the Cauchy problem associated with (TRIALS) is written as follows, where we unambiguously remove the dependence of (x, y, λ) on t to lighten the formula,

$$\begin{cases} \ddot{x} + \gamma(t)\dot{x} + b(t)\Big(\nabla f(x) + A^* \left[\lambda + \alpha(t)\dot{\lambda} + \mu(Ax + By - c)\right]\Big) & = 0 \\ \ddot{y} + \gamma(t)\dot{y} + b(t)\Big(\nabla g(y) + B^* \left[\lambda + \alpha(t)\dot{\lambda} + \mu(Ax + By - c)\right]\Big) & = 0 \\ \ddot{\lambda} + \gamma(t)\dot{\lambda} - b(t)\Big(A(x + \alpha(t)\dot{x}) + B(y + \alpha(t)\dot{y}) - c\Big) & = 0 \qquad \text{(TRIALS)} \\ (x(t_0), y(t_0), \lambda(t_0)) = (x_0, y_0, \lambda_0) \text{ and} \\ (\dot{x}(t_0), \dot{y}(t_0), \dot{\lambda}(t_0)) = (u_0, v_0, \nu_0). \end{cases}$$

If in addition to $(\mathcal{H}_{\mathcal{P}})$, the gradients of f and g are Lipschitz continuous on bounded sets, we will show later in Section 4 that the Cauchy problem associated with (TRIALS) has a unique global solution on $[t_0, +\infty[$. Indeed, although the existence and uniqueness of a local solution follows from the standard non-autonomous Cauchy-Lipschitz theorem, the global existence necessitates the energy estimates derived from the Lyapunov analysis in the next section. The centrality of these estimates is the reason why the proof of well-posedness is deferred to Section 4. Thus, for the moment we take for granted the existence of classical solutions to (TRIALS).

2.3 A fast convergence result

Our Lyapunov analysis will allow us to establish convergence results and rates under very general conditions on the parameters of (TRIALS), see Section 3. In fact, there are many situations of practical interest where such conditions are easily verified, and which will be discussed in detail in Section 6. Thus for the sake of illustration and reader convenience, here we describe an important situation where convergence occurs with the fast rate $\mathcal{O}(1/t^2)$.

Theorem 1 *Suppose that the coefficients of* (TRIALS) *satisfy*

$$\alpha(t) = \alpha_0 t \text{ with } \alpha_0 > 0, \ \gamma(t) = \frac{\eta + \alpha_0}{\alpha_0 t}, \ b(t) = t^{\frac{1}{\alpha_0} - 2},$$

where $\eta > 1$. Suppose that the set of saddle points $\mathscr{S}$ is non-empty and let $(x^\star, y^\star, \lambda^\star) \in \mathscr{S}$. Then, for any solution trajectory $(x(\cdot), y(\cdot), \lambda(\cdot))$ of (TRIALS)*, the trajectory*

remains bounded, and we have the following convergence rates:

$$\mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) = \mathcal{O}\left(\frac{1}{t^{\frac{1}{\alpha_0}}}\right),$$

$$\|Ax(t) + By(t) - c\|^2 = \mathcal{O}\left(\frac{1}{t^{\frac{1}{\alpha_0}}}\right),$$

$$-\frac{C_1}{t^{\frac{1}{2\alpha_0}}} \le F(x(t), y(t)) - F(x^\star, y^\star) \le \frac{C_2}{t^{\frac{1}{\alpha_0}}},$$

$$\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\| = \mathcal{O}\left(\frac{1}{t}\right).$$

where C_1 and C_2 are positive constants.

In particular, for $\alpha_0 = \frac{1}{2}$, i.e. no time scaling $b \equiv 1$, we have

$$\mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) = \mathcal{O}\left(\frac{1}{t^2}\right),$$

$$\|Ax(t) + By(t) - c\|^2 = \mathcal{O}\left(\frac{1}{t^2}\right),$$

$$-\frac{C_1}{t} \le F(x(t), y(t)) - F(x^\star, y^\star) \le \frac{C_2}{t^2}.$$

For the ADMM algorithm (thus in discrete time $t = kh$, $k \in \mathbb{N}, h > 0$), it has been shown in [32,33] that the convergence rate of (squared) feasibility is $\mathcal{O}\left(\frac{1}{k}\right)$ and that on $|F(x_k, y_k) - F(x^\star, y^\star)|$ is $\mathcal{O}\left(\frac{1}{k^{1/2}}\right)$. These rates were shown to be essentially tight in [32]. Our results then suggest than for $\alpha_0 = \frac{1}{2}$, a proper discretization of (TRIALS) would lead to an accelerated ADMM algorithm with provably faster convergence rates (see [37,38] in this direction on specific problem instances and algorithms). These discrete algorithmic issues of (TRIALS) will be investigated in a future work.

Again, for $\alpha_0 = \frac{1}{2}$, the $\mathcal{O}\left(\frac{1}{t^2}\right)$ rate obtained on the Lagrangian is reminiscent of the fast convergence obtained with the continuous-time dynamical version of the Nesterov accelerated gradient method in which the viscous damping coefficient is of the form $\gamma(t) = \frac{\gamma_0}{t}$ and the fast rate is obtained for $\gamma_0 \ge 3$; see [11,52]. With our notations this corresponds to $\gamma_0 = \frac{\eta + \alpha_0}{\alpha_0}$, and our choice $\alpha_0 = \frac{1}{2}$ entails $\gamma_0 = 2\eta + 1 > 3$. This corresponds to the same critical value as Nesterov's but the inequality here is strict. This is not that surprising in our context since one has to handle the dual multiplier and there is an intricate interplay between γ and the extrapolation coefficient α.

2.4 The role of extrapolation

One of the key and distinctive features of (TRIALS) is that the partial gradients (with the appropriate sign) of the augmented Lagrangian function are not evaluated at $(x(t), y(t), \lambda(t))$ as it would the case in a classical continuous-time system associated to ADMM-type methods, but rather at extrapolated points. This new property will be instrumental to allow for faster convergence rates, and it can be interpreted from different standpoints: optimization, game theory, or control:

- *Optimization standpoint*: in this field, this type of extrapolation was recently studied in [14, 36, 54]. It will play a key role in the development of our Lyapunov analysis. Observe that $\alpha(t)\dot{x}(t)$ and $\alpha(t)\dot{\lambda}(t)$ point to the direction of future movement of $x(t)$ and $\lambda(t)$. Thus, (TRIALS) involves the estimated future positions $x(t) + \alpha(t)\dot{x}(t)$ and $\lambda(t) + \alpha(t)\dot{\lambda}(t)$. Explicit discretization $x_k + \alpha_k(x_k - x_{k-1})$ and $\lambda_k + \alpha_k(\lambda_k - \lambda_{k-1})$ gives an extrapolation similar to the accelerated method of Nesterov. The implicit discretization reads $x_k + \alpha_k(x_{k+1} - x_k)$ and $\lambda_k + \alpha_k(\lambda_{k+1} - \lambda_k)$. For $\alpha_k = 1$, this gives x_{k+1} and λ_{k+1}, which would yield implicit algorithms with associated stability properties.
- *Game theoretic standpoint*: let us think about (x, y) and λ as two players playing against each other, and shortly speaking, we identify the players with their actions. We can then see that in (TRIALS), each player anticipates the movement of its opponent. In the coupling term, the player (x, y) takes account of the anticipated position of the player λ, which is $\lambda(t) + \alpha(t)\dot{\lambda}(t)$, and vice versa.
- *Control theoretic standpoint*: the structure of (TRIALS) is also related to control theory and state derivative feedback. By defining $w(t) = (x(t), y(t), \lambda(t))$ the equation can be written in an equivalent way

$$\ddot{w}(t) + \gamma(t)\dot{w}(t) = K(t, w(t), \dot{w}(t)),$$

for an operator K appropriately identified from (TRIALS) in terms of the partial gradients of $\mathcal{L}_\mu$, α and b. In this system, the feedback control term K, which takes the constraint into account, is not only a function of the state $w(t)$ but also of its derivative. One can consult [40] for a comprehensive treatment of state derivative feedback. Indeed, we will use $\alpha(\cdot)$ as a control variable, which will turn to play an important role in our subsequent developments.

2.5 Associated monotone inclusion problem

The optimality system (4) can be written equivalently as

$$T_{\mathcal{L}}(x, y, z) = 0, \tag{6}$$

where $T_{\mathcal{L}} : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is the maximally monotone operator associated with the convex-concave function $\mathcal{L}$, and which is defined by

$$\begin{aligned} T_{\mathcal{L}}(x, y, \lambda) &= \left(\nabla_{x,y}\mathcal{L}, -\nabla_\lambda \mathcal{L} \right)(x, y, \lambda) \\ &= \left(\nabla f(x) + A^*\lambda, \ \nabla g(y) + B^*\lambda, \ -(Ax + By - c) \right). \end{aligned} \tag{7}$$

Indeed, it is immediate to verify that $T_{\mathcal{L}}$ is monotone using ($\mathcal{H}_{\mathcal{P}}$). Since it is continuous, it is a maximally monotone operator. Another way of seeing it is to use the standard splitting of $T_{\mathcal{L}}$ as $T_{\mathcal{L}} = T_1 + T_2$ where

$$\begin{aligned} T_1(x, y, \lambda) &= (\nabla f(x), \ \nabla g(y), 0) \\ T_2(x, y, \lambda) &= \left(A^*\lambda, \ B^*\lambda, \ -(Ax + By - c) \right). \end{aligned}$$

The operator $T_1 = \partial \Phi$ is nothing but gradient of the convex function $\Phi(x, y, \lambda) = f(x) + g(y)$, and therefore is maximally monotone owing to ($\mathcal{H}_{\mathcal{P}}$) (recall that convexity of a differentiable function implies maximal monotonicity of its gradient [50]). The operator T_2 is obtained by translating a linear continuous and

skew-symmetric operator, and therefore it is also maximally monotone. This immediately implies that $T_{\mathcal{L}}$ is maximally monotone as the sum of two maximally monotone operators, one of them being Lipschitz continuous ([30, Lemma 2.4, page 34]). In turn, $\mathscr{S}$ can be interpreted as the set of zeros of the maximally monotone operator $T_{\mathcal{L}}$. As such, it is a closed convex subset of $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.

The evolution equation associated to $T_{\mathcal{L}}$ is written

$$\begin{cases} \dot{x}(t) + \nabla f(x(t)) + A^*(\lambda(t)) & = 0 \\ \dot{y}(t) + \nabla g(y(t)) + B^*(\lambda(t)) & = 0 \\ \dot{\lambda}(t) - (A(x(t)) + B(y(t)) - c) = 0 \end{cases} \tag{8}$$

Following [30], the Cauchy problem (8) is well-posed, and the solution trajectories of (8), which define a semi-group of contractions generated by $T_{\mathcal{L}}$, converge weakly in an ergodic sense to equilibria, which are the zeros of the operator $T_{\mathcal{L}}$. Moreover, appropriate implicit discretization of (8) yields the proximal ADMM algorithm.

The situation is more complicated if we consider the corresponding inertial dynamics. Indeed, the convergence theory for the heavy ball method can be naturally extended to the case of maximally monotone cocoercive operators. Unfortunately, because of the skew-symmetric component T_2 in $T_{\mathcal{L}}$ (when $c = 0$), the operator $T_{\mathcal{L}}$ is *not* cocoercive. To overcome this difficulty, recent studies consider inertial dynamics where the operator $T_{\mathcal{L}}$ is replaced by its Yosida approximation, with an appropriate adjustment of the Yosida parameter; see [19] and [5] in the case of the Nesterov accelerated method. However, such an approach does not achieve full splitting algorithms, hence requiring an additional internal loop.

3 Lyapunov analysis

Let $(x^\star, y^\star) \in \mathcal{X} \times \mathcal{Y}$ be a solution of ($\mathcal{P}$), and denote by $F^\star := F(x^\star, y^\star)$ the optimal value of ($\mathcal{P}$). For the moment, the variable $\lambda^\star$ is chosen arbitrarily in $\mathcal{Z}$. We will then be led to specialize it. Let $t \mapsto (x(t), y(t), \lambda(t))$ be a solution trajectory of (TRIALS) defined for $t \geq t_0$. It is supposed to be a classical solution, *i.e.* of class $\mathcal{C}^2$. We are now in position to introduce the function $t \in [t_0, +\infty[\mapsto \mathcal{E}(t) \in \mathbb{R}$ that will serve as a Lyapunov function,

$$\mathcal{E}(t) := \delta^2(t) b(t) \Big(\mathcal{L}_\mu(x(t), y(t), \lambda^\star) - \mathcal{L}_\mu(x^\star, y^\star, \lambda^\star) \Big) + \frac{1}{2} \|v(t)\|^2 \tag{9}$$

$$+ \frac{1}{2} \xi(t) \|(x(t), y(t), \lambda(t)) - (x^\star, y^\star, \lambda^\star)\|^2,$$

$$v(t) := \sigma(t) \Big((x(t), y(t), \lambda(t)) - (x^\star, y^\star, \lambda^\star) \Big) + \delta(t)(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t)). \tag{10}$$

The coefficient $\sigma(t)$ is non-negative and will be adjusted later, while $\delta(t), \xi(t)$ are explicitly defined by the following formulas:

$$\begin{cases} \delta(t) := \sigma(t)\alpha(t), \\ \xi(t) := \sigma(t)^2 \Big(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1 \Big) - 2\alpha(t)\sigma(t)\dot{\sigma}(t) \end{cases} \tag{11}$$

This choice will become clear from our Lyapunov analysis. To guarantee that $\mathcal{E}$ is a Lyapunov function for the dynamical system (TRIALS), the following conditions on the coefficients $\gamma, \alpha, b, \sigma$ will naturally arise from our analysis:

Lyapunov system of inequalities/equalities on the parameters.

$(\mathcal{G}_1)$ $\sigma(t)\Big(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1\Big) - 2\alpha(t)\dot{\sigma}(t) \geq 0,$

$(\mathcal{G}_2)$ $\sigma(t)\Big(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1\Big) - \alpha(t)\dot{\sigma}(t) \geq 0,$

$(\mathcal{G}_3)$ $-\frac{d}{dt}\Big[\sigma(t)\Big(\sigma(t)\Big(\gamma(t)\alpha(t) - \dot{\alpha}(t)\Big) - 2\alpha(t)\dot{\sigma}(t)\Big)\Big] \geq 0,$

$(\mathcal{G}_4)$ $\alpha(t)\sigma(t)^2 b(t) - \frac{d}{dt}\left(\alpha^2\sigma^2 b\right)(t) = 0.$

Observe that condition $(\mathcal{G}_1)$ automatically ensures that $\xi(t)$ is a non-negative function. In most practical situations (see Section 6), we will take σ as a non-negative constant, in which case $(\mathcal{G}_1)$ and $(\mathcal{G}_2)$ coincide, and thus conditions $(\mathcal{G}_1)$–$(\mathcal{G}_4)$ reduce to a system of three differential inequalities/equalities involving only the coefficients (γ, α, b) of the dynamical system (TRIALS).

3.1 Convergence rate of the values

By relying on a Lyapunov analysis with the function $\mathcal{E}$, we are now ready to state our first main result.

Theorem 2 *Assume that $(\mathcal{H}_\mathcal{P})$ and $(\mathcal{H}_\mathcal{D})$ hold. Suppose that the growth conditions $(\mathcal{G}_1)$–$(\mathcal{G}_4)$ on the parameters $(\gamma, \alpha, \sigma, b)$ of (TRIALS) are satisfied for all $t \geq t_0$. Let $t \in [t_0, +\infty[\mapsto (x(t), y(t), \lambda(t))$ be a solution trajectory of (TRIALS). Let $\mathcal{E}$ be the function defined in (9)-(10). Then the following holds:*

(1) $\mathcal{E}$ is a non-increasing function, and for all $t \geq t_0$

$$F(x(t), y(t)) - F^\star = \mathcal{O}\left(\frac{1}{\alpha(t)^2\sigma(t)^2 b(t)}\right).$$

(2) Suppose moreover that $\mathscr{S}$, the set of saddle points of $\mathcal{L}$ in (1) is non-empty, and let $(x^\star, y^\star, \lambda^\star) \in \mathscr{S}$. Then for all $t \geq t_0$, the following rates and integrability properties are satisfied:

(i) $0 \leq \mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) = \mathcal{O}\left(\frac{1}{\alpha(t)^2\sigma(t)^2 b(t)}\right);$

(ii) $\|Ax(t) + By(t) - c\|^2 = \mathcal{O}\left(\frac{1}{\alpha(t)^2\sigma(t)^2 b(t)}\right);$

(iii) there exists positive constants C_1 and C_2 such that

$$-\frac{C_1}{\alpha(t)\sigma(t)\sqrt{b(t)}} \leq F(x(t), y(t)) - F^\star \leq \frac{C_2}{\alpha(t)^2\sigma(t)^2 b(t)};$$

(iv) $\displaystyle\int_{t_0}^{+\infty} \alpha(t)\sigma(t)^2 b(t)\|Ax(t) + By(t) - c\|^2 dt < +\infty;$

(v) $\displaystyle\int_{t_0}^{+\infty} k(t)\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\|^2 dt < +\infty$, where

$$k(t) = \alpha(t)\sigma(t)\Big(\sigma(t)\left(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1\right) - \alpha(t)\dot{\sigma}(t)\Big).$$

Proof To lighten notation, we drop the dependence on the time variable t. Recall that $(x^\star, y^\star)$ is a solution of $(\mathcal{P})$ and $\lambda^\star$ is an arbitrary vector in $\mathcal{Z}$. Let us define

$$w := (x, y, \lambda), \quad w^\star := (x^\star, y^\star, \lambda^\star), \quad \mathcal{F}_\mu(w) := \mathcal{L}_\mu(x, y, \lambda^\star) - \mathcal{L}_\mu(x^\star, y^\star, \lambda^\star).$$

With these notations we have (recall (9) and (10))

$$v = \sigma(w - w^\star) + \delta \dot{w},$$
$$\nabla \mathcal{F}_\mu(w) = (\nabla_x \mathcal{L}_\mu(x, y, \lambda^\star), \nabla_y \mathcal{L}_\mu(x, y, \lambda^\star), 0)$$
$$\mathcal{E} = \delta^2 b \mathcal{F}_\mu(w) + \frac{1}{2} \|v\|^2 + \frac{1}{2} \xi \|w - w^\star\|^2.$$

Differentiating $\mathcal{E}$ gives

$$\frac{d}{dt} \mathcal{E} = \frac{d}{dt}(\delta^2 b) \mathcal{F}_\mu(w) + \delta^2 b \langle \nabla \mathcal{F}_\mu(w), \dot{w} \rangle + \langle v, \dot{v} \rangle + \frac{1}{2} \dot{\xi} \|w - w^\star\|^2 + \xi \langle w - w^\star, \dot{w} \rangle. \quad (12)$$

Using the constitutive equation in (TRIALS), we have

$$\dot{v} = \dot{\sigma}(w - w^\star) + (\sigma + \dot{\delta}) \dot{w} + \delta \ddot{w}$$
$$= \dot{\sigma}(w - w^\star) + (\sigma + \dot{\delta}) \dot{w} - \delta \left(\gamma \dot{w} + b K_{\mu, \alpha}(w) \right)$$
$$= \dot{\sigma}(w - w^\star) + (\sigma + \dot{\delta} - \delta \gamma) \dot{w} - \delta b K_{\mu, \alpha}(w),$$

where the operator $K_{\mu, \alpha} : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is defined by

$$K_{\mu, \alpha}(w) := \begin{bmatrix} \nabla_x \mathcal{L}_\mu(x, y, \lambda + \alpha \dot{\lambda}) \\ \nabla_y \mathcal{L}_\mu(x, y, \lambda + \alpha \dot{\lambda}) \\ -\nabla_\lambda \mathcal{L}_\mu(x + \alpha \dot{x}, y + \alpha \dot{y}, \lambda) \end{bmatrix}$$

Elementary computation gives

$$K_{\mu, \alpha}(w) = \nabla \mathcal{F}_\mu(w) + \begin{bmatrix} A^*(\lambda - \lambda^\star + \alpha \dot{\lambda}) \\ B^*(\lambda - \lambda^\star + \alpha \dot{\lambda}) \\ -A(x + \alpha \dot{x}) - B(y + \alpha \dot{y}) + c \end{bmatrix}$$

According to the above formulas for v, $\dot{v}$ and $K_{\mu, \alpha}$, we get

$$\langle v, \dot{v} \rangle = \langle \dot{\sigma}(w - w^\star) + (\sigma + \dot{\delta} - \delta \gamma) \dot{w} - \delta b K_{\mu, \alpha}(w), \sigma(w - w^\star) + \delta \dot{w} \rangle$$
$$= \sigma \dot{\sigma} \|w - w^\star\|^2 + \left(\delta \dot{\sigma} + \sigma(\sigma + \dot{\delta} - \delta \gamma) \right) \langle \dot{w}, w - w^\star \rangle + \delta \left(\sigma + \dot{\delta} - \delta \gamma \right) \|\dot{w}\|^2$$
$$- \delta b \left[\sigma \langle \nabla \mathcal{F}_\mu(w), w - w^\star \rangle + \delta \langle \nabla \mathcal{F}_\mu(w), \dot{w} \rangle \right]$$
$$- \delta b \left[\sigma \langle \lambda - \lambda^\star + \alpha \dot{\lambda}, Ax - Ax^\star \rangle + \delta \langle \lambda - \lambda^\star + \alpha \dot{\lambda}, A\dot{x} \rangle \right]$$
$$- \delta b \left[\sigma \langle \lambda - \lambda^\star + \alpha \dot{\lambda}, By - By^\star \rangle + \delta \langle \lambda - \lambda^\star + \alpha \dot{\lambda}, B\dot{y} \rangle \right]$$
$$+ \delta b \sigma \langle A(x + \alpha \dot{x}) + B(y + \alpha \dot{y}) - c, \lambda - \lambda^\star \rangle$$
$$+ \delta^2 b \langle A(x + \alpha \dot{x}) + B(y + \alpha \dot{y}) - c, \dot{\lambda} \rangle.$$

Let us insert this expression in (12). We first observe that the term $\langle \nabla \mathcal{F}_\mu(w), \dot{w} \rangle$ appears twice but with opposite signs, and therefore cancels out. Moreover, the coefficient of $\langle \dot{w}, w - w^\star \rangle$ becomes $\xi + \delta \dot{\sigma} - \sigma(\gamma \delta - \dot{\delta} - \sigma)$. Thanks to the choice of δ and ξ devised in (11), the term $\langle \dot{w}, w - w^\star \rangle$ also disappears. We recall that by

virtue of $(\mathcal{G}_1)$, ξ is non-negative, and thus so is the last term in $\mathcal{E}$. Overall, the formula (12) simplifies to

$$
\frac{d}{dt}\mathcal{E} = \frac{d}{dt}(\delta^2 b)\mathcal{F}_\mu(w) + \left(\frac{1}{2}\dot{\xi} + \sigma\dot{\sigma}\right)\|w - w^\star\|^2 + \delta\left(\sigma + \dot{\delta} - \delta\gamma\right)\|\dot{w}\|^2 \tag{13}
$$
$$
- \delta b\sigma\langle\nabla\mathcal{F}_\mu(w),\, w - w^\star\rangle - \delta b\mathcal{W},
$$

where

$$
\begin{aligned}
\mathcal{W} := &\ \sigma\langle\lambda - \lambda^\star + \alpha\dot{\lambda},\, Ax - Ax^\star\rangle + \delta\langle\lambda - \lambda^\star + \alpha\dot{\lambda},\, A\dot{x}\rangle \\
&+ \sigma\langle\lambda - \lambda^\star + \alpha\dot{\lambda},\, By - By^\star\rangle + \delta\langle\lambda - \lambda^\star + \alpha\dot{\lambda},\, B\dot{y}\rangle \\
&- \sigma\langle A(x + \alpha\dot{x}) + B(y + \alpha\dot{y}) - c,\, \lambda - \lambda^\star\rangle \\
&- \delta\langle A(x + \alpha\dot{x}) + B(y + \alpha\dot{y}) - c,\, \dot{\lambda}\rangle.
\end{aligned}
$$

Since $(x^\star, y^\star) \in \mathcal{X} \times \mathcal{Y}$ is a solution of $(\mathcal{P})$, we obviously have $Ax^\star + By^\star = c$. Thus, $\mathcal{W}$ reduces to

$$
\begin{aligned}
\mathcal{W} &= \sigma\langle Ax + By - c,\, \lambda - \lambda^\star + \alpha\dot{\lambda}\rangle + \delta\langle A\dot{x} + B\dot{y},\, \lambda - \lambda^\star + \alpha\dot{\lambda}\rangle \\
&\quad - \sigma\langle Ax + By - c,\, \lambda - \lambda^\star\rangle - \sigma\alpha\langle A\dot{x} + B\dot{y},\, \lambda - \lambda^\star\rangle \\
&\quad - \delta\langle Ax + By - c,\, \dot{\lambda}\rangle - \delta\alpha\langle A\dot{x} + B\dot{y},\, \dot{\lambda}\rangle \\
&= (\sigma\alpha - \delta)\Big(\langle Ax + By - c,\, \dot{\lambda}\rangle - \langle A\dot{x} + B\dot{y},\, \lambda - \lambda^\star\rangle\Big).
\end{aligned}
$$

Since it is difficult to control the sign of the above expression, the choice of δ in (11) appears natural, which entails $\mathcal{W} = 0$.

On the other hand, by convexity of $\mathcal{L}(\cdot, \cdot, \lambda^\star)$, strong convexity of $\frac{\mu}{2}\|\cdot - c\|^2$, the fact that $Ax^\star + By^\star = c$ and $\mathcal{F}_\mu(w^\star) = 0$, it is straightforward to see that

$$
-\mathcal{F}_\mu(w) - \frac{\mu}{2}\|Ax(t) + By(t) - c\|^2 \geq \langle\nabla\mathcal{F}_\mu(w),\, w^\star - w\rangle.
$$

Collecting the above results, (13) becomes

$$
\frac{d}{dt}\mathcal{E} + \left(\delta b\sigma - \frac{d}{dt}(\delta^2 b)\right)\mathcal{F}_\mu(w) \tag{14}
$$
$$
\leq \left(\frac{1}{2}\dot{\xi} + \sigma\dot{\sigma}\right)\|w - w^\star\|^2 + \delta\left(\sigma + \dot{\delta} - \delta\gamma\right)\|\dot{w}\|^2 - \frac{\delta b\sigma\mu}{2}\|Ax(t) + By(t) - c\|^2.
$$

Since δ is non-negative (σ and α are), and in view of $(\mathcal{G}_2)$, the coefficient of the second term in the right hand side (14) is non-positive. The same conclusion holds for the coefficient of the first term since its non-positivity is equivalent to $(\mathcal{G}_3)$. Therefore, inequality (14) implies

$$
\frac{d}{dt}\mathcal{E} + \left(\delta b\sigma - \frac{d}{dt}(\delta^2 b)\right)\mathcal{F}_\mu(w) \leq 0. \tag{15}
$$

The sign of $\mathcal{F}_\mu(w)$ is unknown for arbitrary $\lambda^\star$. This is precisely where we invoke $(\mathcal{G}_4)$ which is equivalent to

$$
\delta b\sigma - \frac{d}{dt}(\delta^2 b) = 0.
$$

(1) Altogether, we have shown so far that (15) eventually reads, for any $t \geq t_0$,

$$\frac{d}{dt}\mathcal{E}(t) \leq 0, \tag{16}$$

i.e. $\mathcal{E}$ is non-increasing as claimed. Let us now turn to the rates.

$\mathcal{E}$ being non-increasing entails that for all $t \geq t_0$

$$\mathcal{E}(t) \leq \mathcal{E}(t_0). \tag{17}$$

Dropping the non-negative terms $\frac{1}{2}\|v(t)\|^2$ and $\frac{1}{2}\xi(t)\|w(t) - w^\star\|^2$ entering $\mathcal{E}$, and according to the definition of $\mathcal{L}_\mu$, we obtain that, for all $t \geq t_0$

$$\delta(t)^2 b(t)\Big(\mathcal{L}_\mu(x(t), y(t), \lambda^\star) - \mathcal{L}_\mu(x^\star, y^\star, \lambda^\star)\Big) \tag{18}$$
$$= \delta(t)^2 b(t)\Big(\mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) + \frac{\mu}{2}\|Ax(t) + By(t) - c\|^2\Big) \leq \mathcal{E}(t_0).$$

Dropping again the quadratic term in (18), we obtain

$$\delta(t)^2 b(t)\Big(F(x(t), y(t)) - F^\star + \langle \lambda^\star, \, Ax(t) + By(t) - c\rangle\Big)$$
$$\leq \delta^2(t_0) b(t_0)\Big(F(x(t_0), y(t_0)) - F^\star + \langle \lambda^\star, \, Ax(t_0) + By(t_0) - c\rangle$$
$$+ \frac{\mu}{2}\|Ax(t_0) + By(t_0) - c\|^2\Big) + \frac{1}{2}\|v(t_0)\|^2$$
$$+ \frac{1}{2}\xi(t_0)\|(x(t_0), y(t_0), \lambda(t_0)) - (x^\star, y^\star, \lambda^\star)\|^2$$
$$\leq \delta^2(t_0) b(t_0)\|\lambda^\star\|\|Ax(t_0) + By(t_0) - c\| + C_0,$$

where C_0 is the non-negative constant

$$C_0 = \delta^2(t_0) b(t_0)\Big(\big|F(x(t_0), y(t_0)) - F^\star\big| + \frac{\mu}{2}\|Ax(t_0) + By(t_0) - c\|^2\Big)$$
$$+ \frac{1}{2}\|v(t_0)\|^2 + \frac{1}{2}\xi(t_0)\|(x(t_0), y(t_0), \lambda(t_0)) - (x^\star, y^\star, \lambda^\star)\|^2. \tag{19}$$

When $Ax(t) + By(t) - c = 0$, we are done by taking, *e.g.* $\lambda^\star = 0$ and $C > C_0$. Assume now that $Ax(t) + By(t) - c \neq 0$. Since $\lambda^\star$ can be freely chosen in $\mathcal{Z}$, we take it as the unit-norm vector

$$\lambda^\star = \frac{Ax(t) + By(t) - c}{\|Ax(t) + By(t) - c\|}. \tag{20}$$

We therefore obtain

$$\delta(t)^2 b(t)\Big(F(x(t), y(t)) - F^\star + \|Ax(t) + By(t) - c\|\Big) \leq C, \tag{21}$$

where $C > \delta^2(t_0) b(t_0)\|Ax(t_0) + By(t_0) - c\| + C_0$. Since the second term in the left hand side is non-negative, the claimed rate in (1) follows immediately.

(2) Embarking from (18) and using (3) since $(x^\star, y^\star, \lambda^\star) \in \mathscr{S}$, we have the rates stated in (i) and (ii).

To show the lower bound in (iii), observe that the upper-bound of (3) entails that

$$F(x(t), y(t)) \geq F(x^\star, y^\star) - \langle Ax(t) + By(t) - c, \lambda^\star \rangle. \qquad (22)$$

Applying Cauchy-Schwarz inequality, we infer

$$F(x(t), y(t)) \geq F(x^\star, y^\star) - \|\lambda^\star\| \|Ax(t) + By(t) - c\|.$$

We now use the estimate (ii) to conclude. Finally the integral estimates of the feasibility (iv) and velocity (v) are obtained by integrating (14). $\qquad \square$

3.2 Boundedness of the trajectory and rate of the velocity

We will further exploit the Lyapunov analysis developed in the previous section to assert additional properties on the iterates and velocities.

Theorem 3 *Suppose the assumptions of Theorem 2 hold. Assume also that $\mathscr{S}$, the set of saddle points of $\mathcal{L}$ in (1) is non-empty, and let $(x^\star, y^\star, \lambda^\star) \in \mathscr{S}$. Then, each solution trajectory $t \in [t_0, +\infty[\mapsto (x(t), y(t), \lambda(t))$ of (TRIALS) satisfies the following properties:*

(1) There exists a positive constant C such that, for all $t \geq t_0$

$$\|(x(t), y(t), \lambda(t)) - (x^\star, y^\star, \lambda^\star)\|^2 \leq \frac{C}{\sigma(t)^2 \Big(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1\Big) - 2\alpha(t)\sigma(t)\dot{\sigma}(t)}$$

$$\left\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\right\| \leq \frac{C}{\alpha(t)\sigma(t)} \left(1 + \sqrt{\frac{\sigma(t)}{\sigma(t)\left(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1\right) - 2\alpha(t)\dot{\sigma}(t)}}\right).$$

(2) If $\sup_{t \geq t_0} \sigma(t) < +\infty$ and $(\mathcal{G}_1)$ is strengthened to

$$(\mathcal{G}_1^+) \ \inf_{t \geq t_0} \sigma(t) \Big(\sigma(t)\left(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1\right) - 2\alpha(t)\dot{\sigma}(t)\Big) > 0,$$

then

$$\sup_{t \geq t_0} \|(x(t), y(t), \lambda(t))\| < +\infty \ \ and \ \ \left\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\right\| = \mathcal{O}\left(\frac{1}{\alpha(t)\sigma(t)}\right).$$

If moreover,

$$(\mathcal{G}_5) \ \inf_{t \geq t_0} \alpha(t) > 0,$$

then

$$\sup_{t \geq t_0} \left\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\right\| < +\infty.$$

Proof We start from (17) in the proof of Theorem 2, which can be equivalently written

$$\delta^2(t) b(t) \left(\mathcal{L}_\mu(x(t), y(t), \lambda^\star) - \mathcal{L}_\mu(x^\star, y^\star, \lambda^\star)\right) + \frac{1}{2}\|v(t)\|^2$$

$$+ \frac{1}{2}\xi(t)\|(x(t), y(t), \lambda(t)) - (x^\star, y^\star, \lambda^\star)\|^2 \leq \mathcal{E}(t_0).$$

Since $(x^\star, y^\star, \lambda^\star) \in \mathscr{S}$, the first term is non-negative by (3), and thus

$$\frac{1}{2}\|v(t)\|^2 + \frac{1}{2}\xi(t)\|(x(t), y(t), \lambda(t)) - (x^\star, y^\star, \lambda^\star)\|^2 \leq \mathcal{E}(t_0).$$

Choosing a positive constant $C \geq \sqrt{2\mathcal{E}(t_0)}$, we immediately deduce that for all $t \geq t_0$

$$\|(x(t), y(t), \lambda(t)) - (x^\star, y^\star, \lambda^\star)\| \leq \frac{C}{\sqrt{\xi(t)}} \quad \text{and} \quad \|v(t)\| \leq C. \tag{23}$$

Set $z(t) = (x(t), y(t), \lambda(t)) - (x^\star, y^\star, \lambda^\star)$. By definition of $v(t)$, we have

$$v(t) = \sigma(t)z(t) + \delta(t)\dot{z}(t).$$

From the triangle inequality and the bound (23), we get

$$\delta(t)\|\dot{z}(t)\| \leq C\left(1 + \frac{\sigma(t)}{\sqrt{\xi(t)}}\right).$$

According to the definition (11) of $\delta(t)$ and $\xi(t)$, we get

$$\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\| \leq \frac{C}{\alpha(t)\sigma(t)}\left(1 + \sqrt{\frac{\sigma(t)}{\sigma(t)\left(\gamma(t)\alpha(t) - \dot{\alpha}(t) - 1\right) - 2\alpha(t)\dot{\sigma}(t)}}\right),$$

which ends the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.3 The role of α and time scaling

The time scaling parameter b enters the conditions on the parameters only via $(\mathcal{G}_4)$, which therefore plays a central role in our analysis. Now consider relaxing $(\mathcal{G}_4)$ to the inequality

$$(\mathcal{G}_4^+) \quad \frac{d}{dt}\left(\alpha^2\sigma^2 b\right)(t) - \alpha(t)\sigma(t)^2 b(t) \geq 0.$$

This is a weaker assumption in which case the corresponding term in (15) does not vanish. However, such an inequality can still be integrated to yield meaningful convergence rates. This is what we are about to prove.

Theorem 4 *Suppose the assumptions of Theorem 2(2) hold, where condition $(\mathcal{G}_4)$ is replaced with $(\mathcal{G}_4^+)$. Let $(x^\star, y^\star, \lambda^\star) \in \mathscr{S} \neq \emptyset$. Assume also that $\inf F(x, y) > -\infty$. Then, for all $t \geq t_0$*

$$\mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) = \mathcal{O}\left(\exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds\right)\right), \tag{24}$$

$$\|Ax(t) + By(t) - c\|^2 = \mathcal{O}\left(\exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds\right)\right), \tag{25}$$

$$-C_1\exp\left(-\int_{t_0}^t \frac{1}{2\alpha(s)}ds\right) \leq F(x(t), y(t)) - F^\star \leq C_2\exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds\right), \tag{26}$$

where C_1 and C_2 are positive constants.

Proof We embark from (15) in the proof of Theorem 2. In view of $(\mathcal{G}_4^+)$ and (11), (15) becomes

$$0 \geq \frac{d}{dt}\mathcal{E} - \left(\frac{d}{dt}\left(\alpha^2 \sigma^2 b \right) - \alpha \sigma^2 b \right) \mathcal{F}_\mu(w) \geq \frac{d}{dt}\mathcal{E} - \frac{\frac{d}{dt}\left(\alpha^2 \sigma^2 b \right) - \alpha \sigma^2 b}{\alpha^2 \sigma^2 b}\mathcal{E}. \qquad (27)$$

Since $(x^\star, y^\star, \lambda^\star) \in \mathscr{S}$, $\mathcal{F}_\mu$ is non-negative and so is the Lyapunov function $\mathcal{E}$. Integrating (27), we obtain the existence of a positive constant C such that, for all $t \geq t_0$

$$0 \leq \mathcal{E}(t) \leq C\alpha(t)^2 \sigma(t)^2 b(t) \exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds \right),$$

which entails, after dropping the positive terms in $\mathcal{E}$,

$$\mathcal{F}_\mu(w(t)) \leq C \exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds \right). \qquad (28)$$

(24) and (25) follow immediately from (28) and the definition of $\mathcal{F}_\mu$.

Let us now turn to (26). Arguing as in the proof of Theorem 2(2), we have

$$F(x(t), y(t)) - F^\star \geq -\|\lambda^\star\|\|Ax(t) + By(t) - c\|.$$

Plugging (25) in this inequality yields the lower-bound of (26).

For the upper-bound, we will argue as in the proof of Theorem 2(1) by considering $\lambda^\star$ as a free variable in $\mathcal{Z}$. By assumption, we have F is bounded from below. This together with (25) implies that $\mathcal{E}$ is also bounded from below, and we denote $\underline{\mathcal{E}}$ this lower-bound. Define $\tilde{\mathcal{E}}(t) = \mathcal{E}(t) - \underline{\mathcal{E}}$ if $\underline{\mathcal{E}}$ is negative and $\tilde{\mathcal{E}}(t) = \mathcal{E}(t)$ otherwise. Thus, from (27), it is easy to see that $\tilde{\mathcal{E}}$ verifies

$$\frac{d}{dt}\tilde{\mathcal{E}} \leq \frac{\frac{d}{dt}\left(\alpha^2 \sigma^2 b \right) - \alpha \sigma^2 b}{\alpha^2 \sigma^2 b}\tilde{\mathcal{E}}. \qquad (29)$$

Integrating (29) and arguing with the sign of $\underline{\mathcal{E}}$, we get the existence of a positive constant C such that, for all $t \geq t_0$

$$\mathcal{E}(t) \leq \tilde{\mathcal{E}}(t) \leq C\alpha(t)^2 \sigma(t)^2 b(t) \exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds \right).$$

Dropping the quadratic terms in $\mathcal{E}$, this yields

$$F(x(t), y(t)) - F^\star + \langle \lambda^\star, Ax(t) + By(t) - c \rangle \leq C \exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds \right).$$

When $Ax(t) + By(t) - c = 0$, we are done by taking, *e.g.* $\lambda^\star = 0$. Assume now that $Ax(t) + By(t) - c \neq 0$ and choose

$$\lambda^\star = \frac{Ax(t) + By(t) - c}{\|Ax(t) + By(t) - c\|}.$$

We arrive at

$$F(x(t), y(t)) - F^\star \leq F(x(t), y(t)) - F^\star + \|Ax(t) + By(t) - c\| \leq C \exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)}ds \right),$$

which completes the proof. $\qquad \square$

Remark 1 Though the rates in Theorem 2 and Theorem 4 look apparently different, it turns out that as expected, those of Theorem 2 are actually a specialisation of those in Theorem 4 when $(\mathcal{G}_4^+)$ holds as an equality, *i.e.* $(\mathcal{G}_4)$ is verified. To see this, it is sufficient to realize that, with the notation $a(t) := \alpha(t)^2 \sigma(t)^2 b(t)$, $(\mathcal{G}_4)$ is equivalent to $\dot{a}(t) = \frac{1}{\alpha(t)} a(t)$. Upon integration, we obtain $a(t) = \exp\left(\int_{t_0}^t \frac{1}{\alpha(s)} ds\right)$, or equivalently

$$\frac{1}{\alpha(t)^2 \sigma(t)^2 b(t)} = \exp\left(-\int_{t_0}^t \frac{1}{\alpha(s)} ds\right).$$

4 Well-posedness of (TRIALS)

In this section, we will show existence and uniqueness of a strong global solution to the Cauchy problem associated with (TRIALS). The main idea is to formulate (TRIALS) in the phase space as a non-autonomous first-order system. In the smooth case, we will invoke the non-autonomous Cauchy-Lipschitz theorem [35, Proposition 6.2.1]. In the non-smooth case, we will use a standard Moreau-Yosida smoothing argument.

4.1 Case of globally Lipschitz continuous gradients

We consider first the case where the gradients of f and g are globally Lipschitz continuous over $\mathcal{X}$ and $\mathcal{Y}$. Let us start by recalling the notion of strong solution.

Definition 1 Denote $\mathcal{H} := \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ equipped with the corresponding product space structure, and $w : t \in [t_0, +\infty[\mapsto (x(t), y(t), \lambda(t)) \in \mathcal{H}$. The function w is a strong global solution of the dynamical system (TRIALS) if it satisfies the following properties:

- w is in $\mathcal{C}^1([0, +\infty[; \mathcal{H})$;
- w and $\dot{w}$ are absolutely continuous on every compact subset of the interior of $[t_0, +\infty[$ (hence almost everywhere differentiable);
- for almost all $t \in [t_0, +\infty[$, (TRIALS) holds with $w(t_0) = (x_0, y_0, \lambda_0)$ and $\dot{w}(t_0) = (u_0, v_0, \nu_0)$).

Theorem 5 *Suppose that $(\mathcal{H}_{\mathcal{P}})$ holds[1] and, moreover, that ∇f and ∇g are Lipschitz continuous, respectively over $\mathcal{X}$ and $\mathcal{Y}$. Assume that γ, α, $b : [t_0, +\infty[\to \mathbb{R}^+$ are non-negative continuous functions. Then, for any given initial condition $(x(t_0), \dot{x}(t_0)) = (x_0, \dot{x}_0) \in \mathcal{X} \times \mathcal{X}$, $(y(t_0), \dot{y}(t_0)) = (y_0, \dot{y}_0) \in \mathcal{Y} \times \mathcal{Y}$, $(\lambda(t_0), \dot{\lambda}(t_0)) = (\lambda_0, \dot{\lambda}_0) \in \mathcal{Z} \times \mathcal{Z}$, the evolution system (TRIALS) has a unique strong global solution.*

Proof Recall the notations of Definition 1. Let $I = [t_0, +\infty[$ and let $Z : t \in I \mapsto (w(t), \dot{w}(t)) \in \mathcal{H}^2$. (TRIALS) can be equivalently written as the Cauchy problem on $\mathcal{H}^2$

$$\begin{cases} \dot{Z}(t) + G(t, Z(t)) = 0 & \text{for } t \in I, \\ Z(t_0) = Z_0, \end{cases} \tag{30}$$

---

[1] Actually, convexity is not needed here.

where $Z_0 = (x_0, y_0, \lambda_0, u_0, v_0, \nu_0)$, and $G : I \times \mathcal{H}^2 \to \mathcal{H}^2$ is the operator

$$G(t, (x, y, \lambda), (u, v, \nu)) = \begin{pmatrix} -u \\ -v \\ -\nu \\ \gamma(t)u + b(t)\Big(\nabla f(x) + A^*\left(\lambda + \alpha(t)\nu + \mu(Ax + By - c)\right)\Big) \\ \gamma(t)v + b(t)\Big(\nabla g(y) + B^*\left(\lambda + \alpha(t)\nu + \mu(Ax + By - c)\right)\Big) \\ \gamma(t)\nu - b(t)\Big(A(x + \alpha(t)u)\big) + B(y + \alpha(t)v) - c\Big) \end{pmatrix}.$$

(31)

To invoke [35, Proposition 6.2.1], it is sufficient to check that for a.e. $t \in I$, $G(t, \cdot)$ is $\beta(t)$-Lipschitz continuous with $\beta(\cdot) \in L^1_{loc}(I)$, and for a.e. $t \in I$, $G(t, Z) = \mathcal{O}(P(t)(1 + \|Z\|))$, $\forall Z \in \mathcal{H}^2$, with $P(\cdot) \in L^1_{loc}(I)$. Since $\nabla f, \nabla g$ are globally Lipschitz continuous, and A and B are bounded linear, elementary computation shows that there exists a constant $C > 0$ such that

$$\|G(t, Z) - G(t, \bar{Z})\| \leq C\beta(t)\|Z - \bar{Z}\|, \quad \beta(t) = 1 + \gamma(t) + b(t)(1 + \alpha(t)).$$

Owing to the continuity of the parameters $\gamma(\cdot)$, $\alpha(\cdot)$, $b(\cdot)$, $\beta(\cdot)$ is integrable on $[t_0, T]$ for all $t_0 < T < +\infty$. Similar calculation shows that

$$\|G(t, (x, y, \lambda), (u, v, \nu))\| \leq C\beta(t)\Big(\|(\nabla f(u), \nabla g(v))\| + \|(x, y, \lambda, u, v, \nu)\|\Big),$$

and we conclude similarly. It then follows from [35, Proposition 6.2.1] that there exists a unique global solution $Z(\cdot) \in W^{1,1}_{loc}(I; \mathcal{H}^2)$ of (30) satisfying the initial condition $Z(t_0) = Z_0$, and thus, by [30, Corollary A.2] that $Z(\cdot)$ is a strong global solution to (30). This in turn leads to the existence and uniqueness of a strong solution $(x(\cdot), y(\cdot), \lambda(\cdot))$ of (TRIALS). □

Remark 2 One sees from the proof that for the above result to hold, it is only sufficient to assume that the parameters γ, α, b are locally integrable instead of continuous. In addition, in the above results, we even have existence and uniqueness of a classical solution.

4.2 Case of locally Lipschitz continuous gradients

Under local Lipschitz continuity assumptions on the gradients ∇f and ∇g, the operator $Z \mapsto G(t, Z)$ defined in (31) is only Lipschitz continuous over the bounded subsets of $\mathcal{H}^2$. As a consequence, the Cauchy-Lipschitz theorem provides the existence and uniqueness of a local solution. To pass from a local solution to a global solution, we will rely on the estimates established in Theorem 3.

Theorem 6 *Suppose that ($\mathcal{H}_\mathcal{P}$) holds[2] and, moreover, that ∇f and ∇g are Lipschitz continuous over the bounded subsets of respectively $\mathcal{X}$ and $\mathcal{Y}$. Assume that γ, α, b : $[t_0, +\infty[\to \mathbb{R}^+$ are non-negative continuous functions such that the conditions ($\mathcal{G}_1^+$), ($\mathcal{G}_2$), ($\mathcal{G}_3$), ($\mathcal{G}_4$) and ($\mathcal{G}_5$) are satisfied, and that $\sup_{t \geq t_0} \sigma(t) < +\infty$. Then, for any initial condition $(x(t_0), \dot{x}(t_0)) = (x_0, \dot{x}_0) \in \mathcal{X} \times \mathcal{X}$, $(y(t_0), \dot{y}(t_0)) = (y_0, \dot{y}_0) \in \mathcal{Y} \times \mathcal{Y}$, $(\lambda(t_0), \dot{\lambda}(t_0)) = (\lambda_0, \dot{\lambda}_0) \in \mathcal{Z} \times \mathcal{Z}$, the evolution system (TRIALS) has a unique strong global solution.*

---

[2] Again, convexity is superfluous here.

Proof We use the same notation as in the proof of Theorem 5. Let us consider the maximal solution of the Cauchy problem (30), say $Z : [t_0, T[\to \mathcal{H}^2$. We have to prove that $T = +\infty$. Following a classical argument, we argue by contradiction, and suppose that $T < +\infty$. It is then sufficient to prove that the limit of $Z(t)$ exists as $t \to T$, so that it will be possible to extend Z locally to the right of T thus getting a contradiction. According to the Cauchy criterion, and the constitutive equation $\dot{Z}(t) = G(t, Z(t))$, it is sufficient to prove that $Z(t)$ is bounded over $[t_0, T[$. At this point, we use the estimates provided by Theorem 3, which gives precisely this result under the conditions imposed on the parameters. $\qquad\qquad\qquad\qquad\qquad$ □

4.3 The non-smooth case

For a large number of applications (*e.g.* data processing, machine learning, statistics), non-smooth functions are ubiquitous. To cover these practical situations, we need to consider the case where the functions f and g are non-smooth. In order to adapt the dynamic (TRIALS) to this non-smooth situation, we will consider the corresponding differential inclusion

$$
\begin{cases}
\ddot{x} + \gamma(t)\dot{x} + b(t)\Big(\partial f(x) + A^*\left[\lambda + \alpha(t)\dot{\lambda} + \mu(Ax + By - c)\right]\Big) & \ni 0 \\
\ddot{y} + \gamma(t)\dot{y} + b(t)\Big(\partial g(y) + B^*\left[\lambda + \alpha(t)\dot{\lambda} + \mu(Ax + By - c)\right]\Big) & \ni 0 \\
\ddot{\lambda} + \gamma(t)\dot{\lambda} - b(t)\Big(A(x + \alpha(t)\dot{x}) + B(y + \alpha(t)\dot{y}) - c\Big) & = 0 \\
(x(t_0), y(t_0), \lambda(t_0)) = (x_0, y_0, \lambda_0) \text{ and} \\
(\dot{x}(t_0), \dot{y}(t_0), \dot{\lambda}(t_0)) = (u_0, v_0, \nu_0),
\end{cases}
\tag{32}
$$

where ∂f and ∂g are the subdifferentials of f and g, respectively. Beyond global existence issues that we will address shortly, one may wonder whether our Lyapunov analysis in the previous sections is still valid in this case. The answer is affirmative provided one takes some care in two main steps that are central in our analysis. First, when taking the time-derivative of the Lyapunov, one has to invoke now the (generalized) chain rule for derivatives over curves (see [30]). The second ingredient is the validity of the subdifferential inequality for convex functions. In turn all our results and estimates presented in the previous sections can be transposed to this more general non-smooth context. Indeed, our approximation scheme that we will present shortly turns out to be monotonically increasing. This gives a variational convergence (epi-convergence) which allows to simply pass to the limit over the estimates established in the smooth case.

Let us now turn to the existence of a global solution to (32). We will again consider strong solutions to this problem, *i.e.* solutions that are $\mathcal{C}^1([t_0, +\infty[; \mathcal{H})$, locally absolutely continuous, and (32) holds almost everywhere on $[t_0, +\infty[$. A natural idea is to use the Moreau-Yosida regularization in order to bring the problem to the smooth case before passing to an appropriate limit. Recall that, for any $\theta > 0$, the Moreau envelopes f_θ and g_θ of f and g are defined respectively by

$$
f_\theta(x) = \min_{\xi \in \mathcal{X}} \left\{ f(\xi) + \frac{1}{2\theta}\|x - \xi\|^2 \right\}, \quad g_\theta(y) = \min_{\eta \in \mathcal{Y}} \left\{ g(\eta) + \frac{1}{2\theta}\|y - \eta\|^2 \right\}.
$$

As a classical result, f_θ and g_θ are continuously differentiable and their gradients are $\frac{1}{\theta}$-Lipschitz continuous. We are then led to consider, for each $\theta > 0$, the

dynamical system

$$
\begin{cases}
\ddot{x}_\theta + \gamma(t)\dot{x}_\theta + b(t)\Big(\nabla f_\theta(x_\theta) + A^*\big[\lambda_\theta + \alpha(t)\dot{\lambda}_\theta + \mu(Ax_\theta + By_\theta - c)\big]\Big) & = 0 \\
\ddot{y}_\theta + \gamma(t)\dot{y}_\theta + b(t)\Big(\nabla g_\theta(y_\theta) + B^*\big[\lambda_\theta + \alpha(t)\dot{\lambda}_\theta + \mu(Ax_\theta + By_\theta - c)\big]\Big) & = 0 \\
\ddot{\lambda}_\theta + \gamma(t)\dot{\lambda}_\theta - b(t)\Big(A(x_\theta + \alpha(t)\dot{x}_\theta) + B(y_\theta + \alpha(t)\dot{y}_\theta) - c\Big) & = 0.
\end{cases}
\tag{33}
$$

The system (33) comes under our previous study, and for which we have existence and uniqueness of a strong global solution. In doing so, we generate a filtered sequence $(x_\theta, y_\theta, \lambda_\theta)_\theta$ of trajectories.

The challenging question is now to pass to the limit in the system above as $\theta \to 0^+$. This is a non-trivial problem, and to answer it, we have to assume that the spaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are finite dimensional, and that f and g are convex real-valued (*i.e.* $\mathrm{dom}(f) = \mathcal{X}$ and $\mathrm{dom}(g) = \mathcal{Y}$ in which case f and g are continuous). Recall that $\partial F(x, y) = \partial f(x) \times \partial g(y)$, and denote $[\partial F(x, y)]^0$ the minimal norm selection of $\partial F(x, y)$.

Theorem 7 *Suppose that $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are finite dimensional Hilbert spaces, and that the functions $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{Y} \to \mathbb{R}$ are convex. Assume that*

(i) F is coercive on the affine feasibility set;
(ii) $\beta_F := \sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \|[\partial F(x, y)]^0\| < +\infty$;
(iii) the linear operator $L = [A\ B]$ is surjective.

Suppose also that γ, α, $b : [t_0, +\infty[\to \mathbb{R}^+$ are non-negative continuous functions such that the conditions $(\mathcal{G}_1^+)$, $(\mathcal{G}_2)$, $(\mathcal{G}_3)$ $(\mathcal{G}_4)$ and $(\mathcal{G}_5)$ are satisfied, and that $\sup_{t \geq t_0} \sigma(t) < +\infty$. Then, for any initial condition $(x(t_0), \dot{x}(t_0)) = (x_0, \dot{x}_0) \in \mathcal{X} \times \mathcal{X}$, $(y(t_0), \dot{y}(t_0)) = (y_0, \dot{y}_0) \in \mathcal{Y} \times \mathcal{Y}$, $(\lambda(t_0), \dot{\lambda}(t_0)) = (\lambda_0, \dot{\lambda}_0) \in \mathcal{Z} \times \mathcal{Z}$, the evolution system (32) admits a strong global solution .

Condition (i) is natural and ensures for instance that the solution set of $(\mathcal{P})$ is non-empty. Condition (iii) is also very mild. A simple case where (ii) holds is when f and g are Lipschitz continuous.

Proof The key property is that the estimates obtained in Theorem 2 and 3, when applied to (33), have a favorable dependence on θ. Indeed, a careful examination of the estimates shows that θ enters them through the Lyapunov function at t_0 only via $|F_\theta(x_0, y_0) - F_\theta(x_\theta^\star, y_\theta^\star)|$ and $\|(x_\theta^\star, y_\theta^\star, \lambda_\theta^\star)\|$, where $F_\theta(x, y) = f_\theta(x) + g_\theta(y)$, $(x_\theta^\star, y_\theta^\star) \in \mathrm{argmin}_{Ax+Bx=c} F_\theta(x, y)$ and $\lambda_\theta^\star$ is an associated dual multiplier; see (19). With standard properties of the Moreau envelope, see [4, Chapter 3] and [24, Chapter 12], one can show that for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$

$$
F(x, y) - \frac{\theta}{2}\|[\partial F(x, y)]^0\|^2 \leq F_\theta(x, y) \leq F(x, y).
$$

This, together with the fact that $(x_\theta^\star, y_\theta^\star) \in \mathrm{argmin}_{Ax+Bx=c} F_\theta(x, y)$ and $(x^\star, y^\star) \in \mathrm{argmin}_{Ax+Bx=c} F(x, y)$ yields

$$
F_\theta(x_\theta^\star, y_\theta^\star) \leq F_\theta(x^\star, y^\star) \leq F(x^\star, y^\star) \leq F(x_\theta^\star, y_\theta^\star).
$$

Thus

$$F(x_0, y_0) - F(x^\star, y^\star) - \frac{\theta}{2}\|[\partial F(x_0, y_0)]^0\|^2 \leq F_\theta(x_0, y_0) - F_\theta(x^\star_\theta, y^\star_\theta)$$
$$\leq F(x_0, y_0) - F(x^\star, y^\star) + \frac{\theta}{2}\|[\partial F(x^\star_\theta, y^\star_\theta)]^0\|^2.$$

This entails, owing to (ii), that

$$\left|F_\theta(x_0, y_0) - F_\theta(x^\star_\theta, y^\star_\theta)\right| \leq \left|F(x_0, y_0) - F(x^\star, y^\star)\right| + \frac{\beta_F^2 \theta}{2}$$

and thus, since we are interested in the limit as $\theta \to 0^+$,

$$\sup_{\theta \in [0, \bar\theta]} \left|F_\theta(x_0, y_0) - F_\theta(x^\star_\theta, y^\star_\theta)\right| \leq \left|F(x_0, y_0) - F(x^\star, y^\star)\right| + \frac{\beta_F^2 \bar\theta}{2} < +\infty.$$

On the other hand,

$$F(x^\star_\theta, y^\star_\theta) \leq F_\theta(x^\star_\theta, y^\star_\theta) + \frac{\beta_F^2 \theta}{2} \leq F_\theta(x^\star, y^\star) + \frac{\beta_F^2 \theta}{2} \leq F(x^\star, y^\star) + \frac{\beta_F^2 \bar\theta}{2}.$$

Thus, in view of (i), $\exists a > 0$ and $b \in \mathbb{R}$ such that

$$a\|(x^\star_\theta, y^\star_\theta)\| + b \leq F(x^\star, y^\star) + \frac{\beta_F^2 \bar\theta}{2},$$

which shows that

$$\sup_{\theta \in [0, \bar\theta]} \|(x^\star_\theta, y^\star_\theta)\| < +\infty.$$

Let us turn to $\lambda^\star_\theta$. When $\lambda^\star_\theta$ is chosen as in (20), then we are done. When $\lambda^\star_\theta$ is the optimal dual multiplier satisfying (4), then it is a solution to the Fenchel-Rockafellar dual problem

$$\min_{\lambda \in \mathcal{Z}} F_\theta^*(-L^*\lambda) + \langle c, \lambda \rangle,$$

where F_θ^* is the Legendre-Fenchel conjugate of F_θ. Without loss of generality, we assume $c = 0$. Classical conjugacy results give

$$F_\theta^*(u) = F^*(u) + \frac{\theta}{2}\|u\|^2.$$

Since f and g are convex and real-valued, the domain of F is full. This is equivalent to coercivity of F^*. This together with injectivity of L^* (see (iii)), imply that there exists $a > 0$ and $b \in \mathbb{R}$ (potentially different from those above) such that

$$a\|\lambda^\star_\theta\| + b \leq F^*(-L^*\lambda^\star_\theta) \leq F_\theta^*(-L^*\lambda^\star_\theta) \leq F_\theta^*(-L^*\lambda^\star) \leq F^*(-L^*\lambda^\star) + \frac{\bar\theta}{2}\|L^*\lambda^\star\|^2 < +\infty.$$

Altogether this shows that

$$\sup_{\theta \in [0, \bar\theta]} \|\lambda^\star_\theta\| < +\infty.$$

Combining the above with Theorem 3, we conclude that for all $T > t_0$, the trajectories $(x_\theta(.), y_\theta(.), \lambda_\theta(\cdot))$ and the velocities $(\dot{x}_\theta(.), \dot{y}_\theta(.), \dot{\lambda}_\theta(\cdot))$ are bounded in $L^2(t_0, T; \mathcal{X} \times \mathcal{Y})$ uniformly in θ. Since $\mathcal{X}$ and $\mathcal{Y}$ are finite dimensional spaces, we

deduce by the Ascoli-Arzelà theorem, that the trajectories are relatively compact for the uniform convergence over the bounded time intervals. By properties of the Moreau envelope, we also have, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$\|\nabla F_\theta(x, y)\| \nearrow \|[\partial F(x, y)]^0\| \text{ as } \theta \searrow 0,$$

and thus

$$\|\nabla F_\theta(x, y)\| \leq \beta_F.$$

Using this and the boundedness assertions of the trajectories and velocities proved above in the constitutive equations (33), the acceleration remains also bounded on the bounded time intervals. Passing to the limit as $\theta \to 0^+$ in (33) is therefore relevant by a classical maximal monotonicity argument. Indeed, we work with the canonical extension of the maximally monotone operators ∇F_θ and ∂F to $L^2(t_0, T, \mathcal{X} \times \mathcal{Y})$, and, in this functional setting, we use that ∇F_θ graph converges to ∂F in the strong-weak topology. $\square$

We conclude this section by noting that at this stage, uniqueness of the solution to (32) is a difficult open problem. In fact, even existence in infinite dimension and/or with any proper lower semicontinuous convex functions f and g is not clear. This goes far beyond the scope of the present paper and we leave it to a future work.

5 The uniformly convex case

We now turn to examine the convergence properties of the trajectories generated by (TRIALS), when the objective F in ($\mathcal{P}$) is uniformly convex on bounded sets. Recall, see *e.g.* [24], that $F : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is uniformly convex on bounded sets if, for each $r > 0$, there is an increasing function $\psi_r : [0, +\infty[\to [0, +\infty[$ vanishing only at the origin, such that

$$F(v) \geq F(w) + \langle \nabla F(w), \, v - w \rangle + \psi_r(\|v - w\|) \tag{34}$$

for all $(v, w) \in (\mathcal{X} \times \mathcal{Y})^2$ such that $\|v\| \leq r$ and $\|w\| \leq r$. The strongly convex case corresponds to $\psi_r(t) = c_F t^2/2$ for some $c_F > 0$. In finite dimension, strict convexity of F entails uniform convexity on any non-empty bounded closed convex subset of $\mathcal{X} \times \mathcal{Y}$, see [24, Corollary 10.18].

Theorem 8 *Suppose that F is uniformly convex on bounded sets, and let $(x^\star, y^\star)$ be the unique solution of the minimization problem ($\mathcal{P}$). Assume also that $\mathscr{S}$, the set of saddle points of $\mathcal{L}$ in (1) is non-empty. Suppose that the conditions $(\mathcal{G}_1^+)$–$(\mathcal{G}_4)$ on the coefficients of (TRIALS) are satisfied for all $t \geq t_0$. Then, each solution trajectory $t \in [t_0, +\infty[\mapsto (x(t), y(t), \lambda(t))$ of (TRIALS) satisfies, $\forall t \geq t_0$,*

$$\psi_r\left(\|(x(t), y(t)) - (x^\star, y^\star)\|\right) = \mathcal{O}\left(\frac{1}{\alpha(t)^2 \sigma(t)^2 b(t)}\right).$$

As a consequence, assuming that $\lim_{t \to +\infty} \alpha(t)^2 \sigma(t)^2 b(t) = +\infty$, we have that the trajectory $t \mapsto (x(t), y(t))$ converges strongly to $(x^\star, y^\star)$ as $t \to +\infty$.

Proof Uniformly convex functions are strictly convex and coercive, and thus $(x^\star, y^\star)$ is unique. From Theorem 3(2), there exists $r_1 > 0$ such that

$$\sup_{t \geq t_0} \left\| (x(t), y(t)) - (x^\star, y^\star) \right\| \leq r_1.$$

Taking $r \geq r_1 + \|(x^\star, y^\star)\|$, we have that the trajectory $(x(\cdot), y(\cdot))$ and $(x^\star, y^\star)$ are both contained in the ball of radius r centered at the origin. Let $\lambda^\star$ be a Lagrange multiplier of problem $(\mathcal{P})$, *i.e.* $(x^\star, y^\star, \lambda^\star) \in \mathscr{S}$. On the one hand, applying the uniform convexity inequality (34) at $v = (x(t), y(t))$ and $w = (x^\star, y^\star)$, we have

$$F(x(t), y(t)) \geq F(x^\star, y^\star) + \langle \nabla F(x^\star, y^\star), (x(t), y(t)) - (x^\star, y^\star) \rangle \\ + \psi_r \left(\|(x(t), y(t)) - (x^\star, y^\star)\| \right).$$

On the other hand, the optimality conditions (4) tells us that

$$\langle \nabla F(x^\star, y^\star), (x(t), y(t)) - (x^\star, y^\star) \rangle = -\langle \lambda^\star, Ax(t) + By(t)) - c \rangle$$

and obviously

$$F(x^\star, y^\star) = \mathcal{L}(x^\star, y^\star, \lambda^\star).$$

Thus,

$$\psi_r \left(\|(x(t), y(t)) - (x^\star, y^\star)\| \right) \leq \mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star).$$

Invoking the estimate in Theorem 2(2)(i) yields the claim.   □

Remark 3 The assumption $\lim_{t \to +\infty} \alpha(t)^2 \sigma(t)^2 b(t) = +\infty$ made in the above theorem is very mild. It holds in particular in all the situations discussed in Section 6. In particular, for $\alpha(t) = t^r$, $0 \leq r < 1$, σ constant, and $b(t) = \frac{1}{t^{2r}} \exp\left(\frac{1}{1-r} t^{1-r} \right)$, one has $\alpha(t)\sqrt{b(t)} = \exp\left(\frac{1}{2(1-r)} t^{1-r} \right)$. Thus, if F is strongly convex, the trajectory $t \mapsto (x(t), y(t))$ converges exponentially fast to the unique minimizer $(x^\star, y^\star)$.

6 Parameters choice for fast convergence rates

In this section, we suppose that the solution set $\mathscr{S}$ of the saddle value problem (4) is non-empty, so as to invoke Theorem 2(2), Theorem 3 and Theorem 4. The set of conditions $(\mathcal{G}_1^+)$, $(\mathcal{G}_2)$, $(\mathcal{G}_3)$, $(\mathcal{G}_4)$ and $(\mathcal{G}_5)$ imposes sufficient assumptions on the coefficients γ, α, b of the dynamical system (TRIALS), and on the coefficients σ, δ, ξ of the function $\mathcal{E}$ defined in (9), which guarantee that $\mathcal{E}$ is a Lyapunov function for the dynamical system (TRIALS).

Let us show that this system admits many solutions of practical interest which in turn will entail fast convergence rates. For this, we will organize our discussion around the coefficient α as dictated by Theorem 4. Indeed, the latter shows that the convergence rate of the Lagrangian values and feasibility is $\mathcal{O}\left(\exp\left(-\int_{t_0}^{t} \frac{1}{\alpha(s)} ds \right) \right)$. Therefore, to obtain a meaningful convergence result, we need to assume that

$$\int_{t_0}^{+\infty} \frac{1}{\alpha(s)} ds = +\infty.$$

This means that the critical growth is $\alpha(t) = at$ for $a > 0$. If $\alpha(t)$ grows faster, our analysis do not provide an instructive convergence rate. So, it is an essential ingredient of our approach to assume that $\alpha(t)$ remains positive, but not too large as $t \to +\infty$. In fact, the set of conditions $(\mathcal{G}_1^+)$, $(\mathcal{G}_2)$, $(\mathcal{G}_3)$, $(\mathcal{G}_4)$ and $(\mathcal{G}_5)$ simplifies considerably by taking σ a positive constant, and $\gamma\alpha - \dot{\alpha}$ a constant strictly greater than one. This is made precise in the following statement whose proof is immediate.

Corollary 1 *Suppose that $\sigma \equiv \sigma_0$ is a positive constant, and $\gamma\alpha - \dot{\alpha} \equiv \eta > 1$. Then the set of conditions $(\mathcal{G}_1^+)$, $(\mathcal{G}_2)$, $(\mathcal{G}_3)$, $(\mathcal{G}_4)$ and $(\mathcal{G}_5)$ reduces to*

$$b(1 + 2\eta - 2\gamma\alpha) - \alpha\dot{b} = 0 \ \ and \ \inf_{t \geq t_0} \alpha(t) > 0. \tag{35}$$

Following the above discussion, we are led to consider the following three cases.

6.1 Constant parameter α

Consider the simple situation where $\sigma \equiv \sigma_0 > 0$ and $\eta > 1$ in which case $(\mathcal{G}_1^+)$ reads $\gamma\alpha - \dot{\alpha} - 1 = \eta - 1 > 0$. Taking $\alpha \equiv \alpha_0$, a positive constant, yields $\gamma \equiv \frac{\eta}{\alpha_0}$ and (35) amounts to solving

$$b - \alpha_0\dot{b} = 0,$$

that is, $b(t) = \exp\left(\frac{t}{\alpha_0}\right)$. Capitalizing on Corollary 1, and specializing the results of Theorem 2(2) (or equivalently Theorem 4 according to Remark 1) and Theorem 3 to the current choice of parameters yields the following statement.

Proposition 1 *Suppose that $\sigma \equiv \sigma_0 > 0$, $\eta > 1$, and that the coefficients of (TRIALS) satisfy: the functions α, γ are constant with*

$$\alpha \equiv \alpha_0 > 0, \ \gamma \equiv \frac{\eta}{\alpha_0}, \ b(t) = \exp\left(\frac{t}{\alpha_0}\right).$$

Suppose that $\mathscr{S}$ is non-empty. Then, for any solution trajectory $(x(\cdot), y(\cdot), \lambda(\cdot))$ of (TRIALS), the trajectory and its velocity remain bounded, and we have

$$\mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) = \mathcal{O}\left(\exp\left(-\frac{t}{\alpha_0}\right)\right),$$

$$\|Ax(t) + By(t) - c\|^2 = \mathcal{O}\left(\exp\left(-\frac{t}{\alpha_0}\right)\right),$$

$$-C_1\exp\left(-\frac{t}{2\alpha_0}\right) \leq F(x(t), y(t)) - F^\star \leq C_2\exp\left(-\frac{t}{\alpha_0}\right),$$

$$\|(\dot{x}(\cdot), \dot{y}(\cdot), \dot{\lambda}(\cdot))\| \in L^2([t_0, +\infty[),$$

where C_1 and C_2 are positive constants.

6.2 Linearly increasing parameter α

We now take $\sigma \equiv \sigma_0 > 0$, $\eta > 1$ and $\alpha(t) = \alpha_0 t$ with $\alpha_0 > 0$. Then $(\mathcal{G}_1^+)$ is satisfied and we have $\gamma(t) = \frac{\eta + \alpha_0}{\alpha_0 t}$. Condition (35) then becomes

$$b(t)(1 - 2\alpha_0) - \alpha_0 t \dot{b}(t) = 0,$$

which admits $b(t) = t^{\frac{1}{\alpha_0} - 2}$ as a solution. We then have $b \equiv 1$ for $\alpha_0 = 1/2$, while one can distinguish two regimes for its limiting behaviour with

$$\lim_{t \to +\infty} b(t) = \begin{cases} +\infty & \alpha_0 < \frac{1}{2}, \\ 0 & \alpha_0 > \frac{1}{2}. \end{cases}$$

In view of Theorem 2(2) and Theorem 3, we obtain the following result.

Proposition 2 *Suppose that* $\sigma \equiv \sigma_0 > 0$, $\eta > 1$, *and that the coefficients of* (TRIALS) *satisfy*

$$\alpha(t) = \alpha_0 t \text{ with } \alpha_0 > 0, \ \gamma(t) = \frac{\eta + \alpha_0}{\alpha_0 t}, \ b(t) = t^{\frac{1}{\alpha_0} - 2}.$$

Suppose that $\mathcal{S}$ *is non-empty. Then, for any solution trajectory* $(x(\cdot), y(\cdot), \lambda(\cdot))$ *of* (TRIALS)*, the trajectory remains bounded, and we have the following convergence rates:*

$$\mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) = \mathcal{O}\left(\frac{1}{t^{\frac{1}{\alpha_0}}}\right),$$

$$\|Ax(t) + By(t) - c\|^2 = \mathcal{O}\left(\frac{1}{t^{\frac{1}{\alpha_0}}}\right),$$

$$-\frac{C_1}{t^{\frac{1}{2\alpha_0}}} \le F(x(t), y(t)) - F^\star \le \frac{C_2}{t^{\frac{1}{\alpha_0}}},$$

$$\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\| = \mathcal{O}\left(\frac{1}{t}\right).$$

where C_1 *and* C_2 *are positive constants.*

6.3 Power-type parameter α

Let us now take $\sigma \equiv \sigma_0 > 0$, $\eta > 1$ and consider the intermediate case between the two previous situations, where $\alpha(t) = t^r$, $0 < r < 1$. Thus $(\mathcal{G}_1^+)$ is satisfied and we have $\gamma(t) = \frac{\eta}{t^r} + \frac{r}{t}$. Condition (35) is then equivalent to

$$b(t)(1 - 2rt^{r-1}) - t^r \dot{b}(t) = 0,$$

which, after integration, shows that $b(t) = \frac{1}{t^{2r}} \exp\left(\frac{1}{1-r} t^{1-r}\right)$ is a solution. Appealing again to Theorem 2(2) and Theorem 3, we obtain the following claim.

Proposition 3 *Take $\sigma \equiv \sigma_0 > 0$ and $\eta > 1$. Suppose that the coefficients of (TRIALS) satisfy*

$$\alpha(t) = t^r \text{ with } 0 < r < 1, \ \gamma(t) = \frac{\eta}{t^r} + \frac{r}{t}, \ b(t) = \frac{1}{t^{2r}} \exp\left(\frac{1}{1-r}t^{1-r}\right).$$

Suppose that $\mathscr{S}$ is non-empty. Then, for any solution trajectory $(x(\cdot), y(\cdot), \lambda(\cdot))$ of (TRIALS), the trajectory remains bounded, and we have the convergence rates:

$$\mathcal{L}(x(t), y(t), \lambda^\star) - \mathcal{L}(x^\star, y^\star, \lambda^\star) = \mathcal{O}\left(\exp\left(-\frac{1}{1-r}t^{1-r}\right)\right),$$

$$\|Ax(t) + By(t) - c\|^2 = \mathcal{O}\left(\exp\left(-\frac{1}{1-r}t^{1-r}\right)\right),$$

$$-C_1 \exp\left(-\frac{1}{2(1-r)}t^{1-r}\right) \leq F(x(t), y(t)) - F^\star \leq C_2 \exp\left(-\frac{1}{1-r}t^{1-r}\right),$$

$$\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\| = \mathcal{O}\left(\frac{1}{t^r}\right).$$

where C_1 and C_2 are positive constants.

6.4 Numerical experiments

To support our theoretical claims, we consider in this section two numerical examples with $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}^2$, one with a strongly convex objective F and one where F is convex but not strongly so.

Example 1 We consider the quadratic programming problem

$$\min_{(x,y) \in \mathbb{R}^4} F(x, y) = \|x - (1,1)^{\mathrm{T}}\|^2 + \|y\|^2 \quad \text{subject to } y = x + (-x_2, 0)^{\mathrm{T}},$$

whose objective is strongly convex and verifies all required assumptions.

Example 2 We consider the minimization problem

$$\min_{(x,y) \in \mathbb{R}^4} F(x, y) = \log\left(1 + \exp\left(-\langle (1,1)^{\mathrm{T}}, x \rangle\right)\right) + \|y\|^2 \quad \text{subject to } y = x + (-x_2, 0)^{\mathrm{T}}.$$

The objective is convex (but not strongly so) and smooth as required. This problem is reminiscent of (regularized) logistic regression very popular in machine learning.

In all our numerical experiments, we consider the continuous time dynamical system (TRIALS), solved numerically with a Runge-Kutta adaptive method (ode45 in MATLAB) on the time interval $[1, 20]$.

For the solely convex (resp. strongly convex) objective, Figure 1 (resp. Figure 2) displays the objective error $|F(x(t), y(t)) - F^\star|$ on the left, the feasibility gap $\|Ax(t) + By(t) - c\|$ in the middle, and the velocity $\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\|$ on the right. In each figure, the first row shows the results for $\alpha(t) \equiv \alpha_0$ with $\alpha_0 \in \{1, 2, 4\}$, the second row corresponds to $\alpha(t) = \alpha_0 t$ with $\alpha_0 \in \{0.25, 0.5, 1\}$ and the third row to $\alpha(t) = t^r$ with $r \in \{0.01, 0.1, 0.5\}$. In all our experiments, we set $\mu = 10$ (recall that μ is the parameter associated with the augmented Lagrangian formulation). All these choices of the parameters comply with the requirements of Propositions 1, 2 and 3. The numerical results are in excellent agreement with our theoretical results, where the values, the velocities and the feasibility gap all converge at the predicted rates.
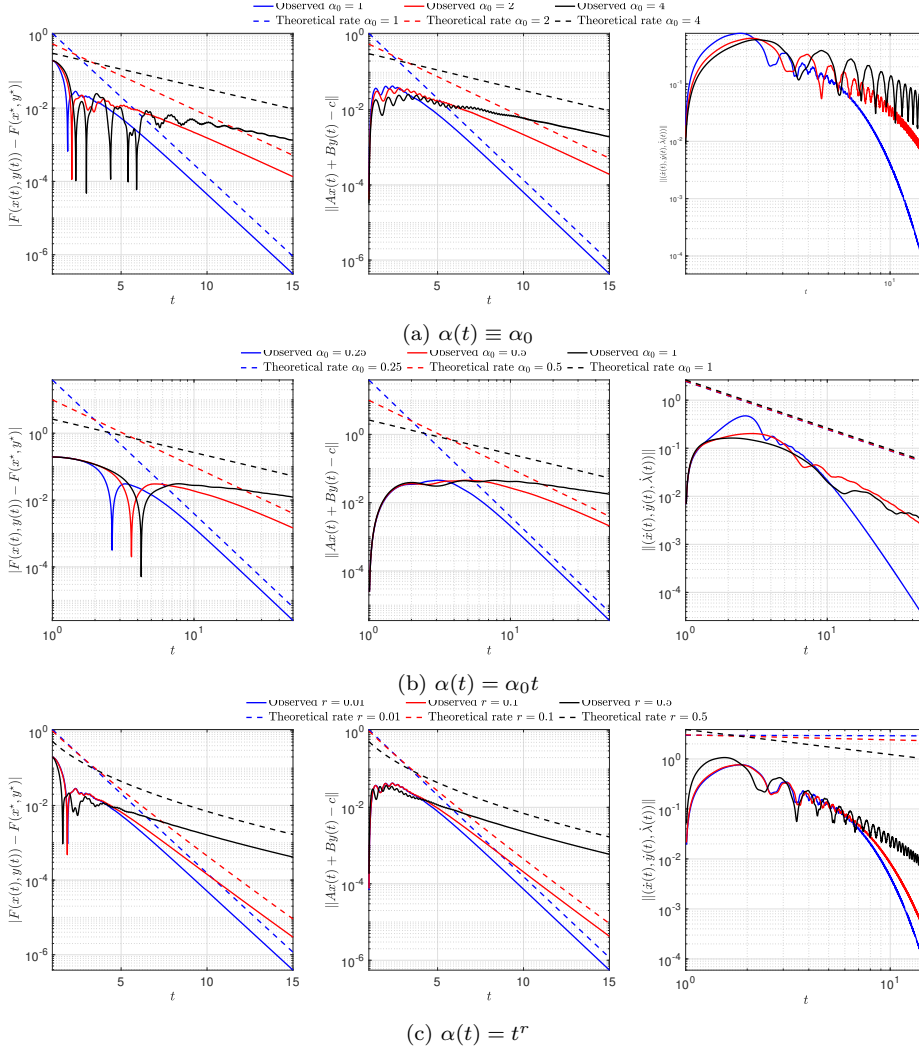
(a) $\alpha(t) \equiv \alpha_0$

(b) $\alpha(t) = \alpha_0 t$

(c) $\alpha(t) = t^r$

Fig. 1: Experiment on the solely convex objective. Observed (solid) and predicted (dashed) rates on the objective error $|F(x(t), y(t)) - F^\star|$ on the left, the feasibility gap $\|Ax(t) + By(t) - c\|$ in the middle, and the velocity $\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\|$ on the right.

7 Conclusion, perspectives

In this paper, we adopted a dynamical system perspective and we have proposed a second-order inertial system enjoying provably fast convergence rates to solve structured convex optimization problems with an affine constraint. One of the most original aspects of our study is the introduction of a damped inertial dynamic involving several time-dependent parameters with specific properties. They allow to consider a variable viscosity coefficient (possibly vanishing so making the
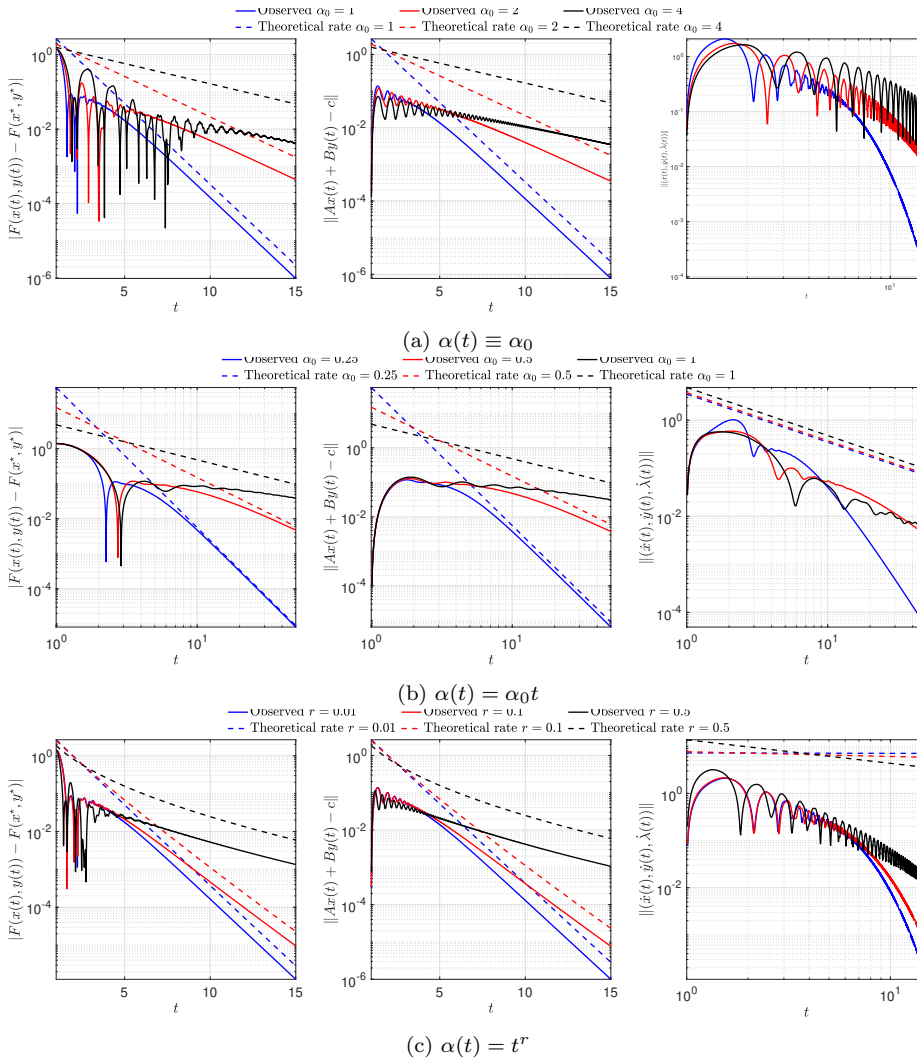
Fig. 2: Experiment on the strongly convex objective. Observed (solid) and predicted (dashed) rates on the objective error $|F(x(t), y(t)) - F^\star|$ on the left, the feasibility gap $\|Ax(t) + By(t) - c\|$ in the middle, and the velocity $\|(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))\|$ on the right.

link with the Nesterov accelerated gradient method), as well as variable extrapolation parameters (possibly large) and time scaling. The analysis of the subtle and intricate interplay between these objects together has been made possible through Lyapunov's analysis. It would have been quite difficult to undertake such an analysis directly on the algorithmic discrete form. On the other hand, as we have now gained a deeper understanding with such a powerful continuous-time framework, we believe this will serve us as a guide to design and analyze a class of inertial ADMM algorithms which can be naturally obtained by appropriate discretization

of the dynamics (TRIALS). Their full study would go beyond the scope of this paper and will be the subject of future work. Besides, several other open questions remain to be studied, among which, the introduction of geometric damping controlled by the Hessian, and the convergence of the trajectories in the general convex constrained case.

References

1. Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. SIAM J. Control Optim. **38**(4), 1102–1119 (2000). DOI 10.1137/ S0363012998335802. URL https://doi.org/10.1137/S0363012998335802

2. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. J. Math. Pures Appl. (9) **81**(8), 747–779 (2002). DOI 10.1016/S0021-7824(01)01253-3. URL https://doi.org/10.1016/S0021-7824(01)01253-3

3. Apidopoulos, V., Aujol, J.F., Dossal, C.: Convergence rate of inertial forward-backward algorithm beyond Nesterov's rule. Math. Program. **180**(1-2, Ser. A), 137–156 (2020). DOI 10.1007/s10107-018-1350-9. URL https://doi.org/10.1007/s10107-018-1350-9

4. Attouch, H.: Variational convergence for functions and operators. Applicable mathematics series. Pitman Advanced Publishing Program (1984). URL https://books.google.fr/ books?id=oxGoAAAAIAAJ

5. Attouch, H.: Fast inertial proximal ADMM algorithms for convex structured optimization with linear constraint. Minimax Theory Appl. **6**(1), 1–24 (2021)

6. Attouch, H., Balhag, A., Chbani, Z., Riahi, H.: Fast convex optimization via inertial dynamics combining viscous and hessian-driven damping with time rescaling. Evolution Equations & Control Theory (2021). URL http://aimsciences.org//article/id/ 92767e45-ae6f-4ff2-9c62-d7ead2d77844

7. Attouch, H., Cabot, A.: Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. J. Differential Equations **263**(9), 5412–5458 (2017). DOI 10.1016/j. jde.2017.06.024. URL https://doi.org/10.1016/j.jde.2017.06.024

8. Attouch, H., Cabot, A.: Convergence rates of inertial forward-backward algorithms. SIAM J. Optim. **28**(1), 849–874 (2018). DOI 10.1137/17M1114739. URL https://doi.org/10. 1137/17M1114739

9. Attouch, H., Cabot, A., Chbani, Z., Riahi, H.: Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient. Evol. Equ. Control Theory **7**(3), 353–371 (2018). DOI 10.3934/eect.2018018. URL https://doi.org/10.3934/eect. 2018018

10. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with hessian driven damping. Math. Program. (2020). URL https: //doi.org/10.1007/s10107-020-01591-1

11. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**(1-2, Ser. B), 123–175 (2018). DOI 10.1007/s10107-016-0992-8. URL https://doi.org/10.1007/ s10107-016-0992-8

12. Attouch, H., Chbani, Z., Riahi, H.: Fast proximal methods via time scaling of damped inertial dynamics. SIAM J. Optim. **29**(3), 2227–2256 (2019). DOI 10.1137/18M1230207. URL https://doi.org/10.1137/18M1230207

13. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. ESAIM Control Optim. Calc. Var. **25**, Paper No. 2, 34 (2019). DOI 10.1051/cocv/2017083. URL https://doi.org/10.1051/cocv/2017083

14. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via a third-order in time evolution equation. Optimization (2020). Preprint available at hal-02432351

15. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via time scaling of damped inertial gradient dynamics. Pure and Applied Functional Analysis (2020). To appear

16. Attouch, H., Czarnecki, M.O., Peypouquet, J.: Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities. SIAM J. Optim. **21**(4), 1251–1274 (2011). DOI 10.1137/110820300. URL https://doi.org/10.1137/ 110820300

17. Attouch, H., Goudou, X., Redont, P.: The heavy ball with friction method. I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. Commun. Contemp. Math. **2**(1), 1–34 (2000). DOI 10.1142/S0219199700000025. URL https://doi.org/10.1142/S0219199700000025

18. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**(3), 1824–1834 (2016). DOI 10.1137/15M1046095. URL https://doi.org/10.1137/15M1046095

19. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. Math. Program. **174**(1-2, Ser. B), 391–432 (2019). DOI 10.1007/s10107-018-1252-x. URL https://doi.org/10.1007/s10107-018-1252-x

20. Attouch, H., Peypouquet, J., Redont, P.: A dynamical approach to an inertial forward-backward algorithm for convex minimization. SIAM J. Optim. **24**(1), 232–256 (2014). DOI 10.1137/130910294. URL https://doi.org/10.1137/130910294

21. Attouch, H., Peypouquet, J., Redont, P.: Fast convex optimization via inertial dynamics with Hessian driven damping. J. Differential Equations **261**(10), 5734–5783 (2016). DOI 10.1016/j.jde.2016.08.020. URL https://doi.org/10.1016/j.jde.2016.08.020

22. Attouch, H., Soueycatt, M.: Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control. Pac. J. Optim. **5**(1), 17–37 (2009)

23. Aujol, J.F., Dossal, C.: Stability of over-relaxations for the forward-backward algorithm, application to FISTA. SIAM J. Optim. **25**(4), 2408–2433 (2015). DOI 10.1137/140994964. URL https://doi.org/10.1137/140994964

24. Bauschke, H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, second edn. CMS Books in Mathematics. Springer, New York (2017)

25. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009). DOI 10.1137/080716542. URL https://doi.org/10.1137/080716542

26. Boţ, R.I., Csetnek, E.R.: An inertial alternating direction method of multipliers. Minimax Theory Appl. **1**(1), 29–49 (2016)

27. Boţ, R.I., Csetnek, E.R.: An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems. J. Optim. Theory Appl. **171**(2), 600–616 (2016). DOI 10.1007/s10957-015-0730-z. URL https://doi.org/10.1007/s10957-015-0730-z

28. Boţ, R.I., Csetnek, E.R., Hendrich, C.: Inertial Douglas-Rachford splitting for monotone inclusion problems. Appl. Math. Comput. **256**, 472–487 (2015). DOI 10.1016/j.amc.2015.01.017. URL https://doi.org/10.1016/j.amc.2015.01.017

29. Boţ, R.I., Csetnek, E.R., László, S.C.: Second-order dynamical systems with penalty terms associated to monotone inclusions. Anal. Appl. (Singap.) **16**(5), 601–622 (2018). DOI 10.1142/S0219530518500021. URL https://doi.org/10.1142/S0219530518500021

30. Brézis, H.: Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York (1973). North-Holland Mathematics Studies, No. 5. Notas de Matemática (50)

31. Chambolle, A., Dossal, C.: On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". J. Optim. Theory Appl. **166**(3), 968–982 (2015). DOI 10.1007/s10957-015-0746-4. URL https://doi.org/10.1007/s10957-015-0746-4

32. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. In: Splitting methods in communication, imaging, science, and engineering, Sci. Comput., pp. 115–163. Springer, Cham (2016)

33. Davis, D., Yin, W.: Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions. Math. Oper. Res. **42**(3), 783–805 (2017). DOI 10.1287/moor.2016.0827. URL https://doi.org/10.1287/moor.2016.0827

34. Goldstein, T., O'Donoghue, B., Setzer, S., Baraniuk, R.: Fast alternating direction optimization methods. SIAM J. Imaging Sci. **7**(3), 1588–1623 (2014). DOI 10.1137/120896219. URL https://doi.org/10.1137/120896219

35. Haraux, A.: Systèmes dynamiques dissipatifs et applications, *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*, vol. 17. Masson, Paris (1991)

36. He, X., Hu, R., Fang, Y.: Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems. arXiv:2007.12428 (2020)

37. Kang, M., Kang, M., Jung, M.: Inexact accelerated augmented Lagrangian methods. Computational Optimization and Applications **62**(2), 373–404 (2015). DOI 10.1007/s10589-015-9742-8. URL https://doi.org/10.1007/s10589-015-9742-8

38. Kang, M., Yun, S., Woo, H., Kang, M.: Accelerated Bregman method for linearly constrained $\ell_1-\ell_2$ minimization. Journal of Scientific Computing **56**(3), 515–534 (2013). DOI 10.1007/s10915-013-9686-z. URL https://doi.org/10.1007/s10915-013-9686-z

39. May, R.: Asymptotic for a second-order evolution equation with convex potential and vanishing damping term. Turkish J. Math. **41**(3), 681–685 (2017). DOI 10.3906/mat-1512-28. URL https://doi.org/10.3906/mat-1512-28

40. Michiels, W., Vyhlídal, T., Huijberts, H., Nijmeijer, H.: Stabilizability and stability robustness of state derivative feedback controllers. SIAM J. Control Optim. **47**(6), 3100–3117 (2009). DOI 10.1137/070697136. URL https://doi.org/10.1137/070697136

41. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1, Ser. B), 125–161 (2013). DOI 10.1007/s10107-012-0629-5. URL https://doi.org/10.1007/s10107-012-0629-5

42. Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSR **269**(3), 543–547 (1983)

43. Patrinos, P., Stella, L., Bemporad, A.: Douglas-Rachford splitting: Complexity estimates and accelerated variants. In: 53rd IEEE Conference on Decision and Control, pp. 4234–4239 (2014). DOI 10.1109/CDC.2014.7040049

44. Pejcic, I., Jones, C.N.: Accelerated ADMM based on accelerated Douglas-Rachford splitting. In: European Control Conference (ECC), pp. 1952–1957. Aalborg, Denmark (2016)

45. Polyak, B.T.: Some methods of speeding up the convergence of iterative methods. Ž. Vyčisl. Mat i Mat. Fiz. **4**, 791–803 (1964)

46. Polyak, B.T.: Introduction to optimization. Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York (1987). Translated from the Russian, With a foreword by Dimitri P. Bertsekas

47. Poon, C., Liang, J.: Trajectory of alternating direction method of multipliers and adaptive acceleration. In: 33rd Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada (2019)

48. Rockafellar, R.T.: Monotone operators associated with saddle-functions and minimax problems. In: Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968), pp. 241–250. Amer. Math. Soc., Providence, R.I. (1970)

49. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976). DOI 10.1287/moor.1.2.97. URL https://doi.org/10.1287/moor.1.2.97

50. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–898 (1976). DOI 10.1137/0314056. URL https://doi.org/10.1137/0314056

51. Shi, B., Du, S.S., Jordan, M.I., Su, W.J.: Understanding the acceleration phenomenon via high-resolution differential equations. arXiv:2440124 (2018)

52. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. J. Mach. Learn. Res. **17**, Paper No. 153, 43 (2016)

53. Wilson, A.C., Recht, B., Jordan, M.I.: A lyapunov analysis of momentum methods in optimization. arXiv:1611.02635 (2016)

54. Zeng, X., Lei, J., Chen, J.: Dynamical primal-dual accelerated method with applications to network optimization. arXiv:1912.03690 (2019)