Augmented Lagrangian methods as dynamical systems for constrained optimization

Alberto De Marchi*

2021-03-04

This is a preprint version of a submitted paper, currently under review.

De Marchi, Augmented Lagrangian methods as dynamical systems for constrained optimization (2021).

Abstract

Dynamical systems approaches to constrained optimization often rely on a penalization term to reach feasible points. We demonstrate that the sequential homotopy method, namely taking projected backward Euler steps on a projected gradient/antigradient augmented Lagrangian flow, coincides with the classical augmented Lagrangian method without the multiplier estimate update. Then, we introduce a time-scaled flow and provide an interpretation of augmented Lagrangian methods as discrete-time dynamical systems. This approach inspires a simple yet effective method for nonlinear programming. We report on numerical results for equality-constrained problems.

*Institute for Applied Mathematics and Scientific Computing, Department of Aerospace Engineering, Universität der Bundeswehr München, Germany. e-mail: alberto.demarchi@unibw.de, orcid: 0000-0002-3545-6898.

1 Introduction

We consider nonlinear programming problems of the form

$$\min_{x \in C} \; f(x) \qquad \text{subject to } \; c(x) = 0, \tag{1}$$

with the closed convex set $C := \{x \in \mathbb{R}^n \,|\, l \leq x \leq u\}$, and vectors l, u that satisfy $l \leq u$, $l < +\infty$, and $u > -\infty$ component-wise. The objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint function $c : \mathbb{R}^n \to \mathbb{R}^m$ are assumed twice continuously differentiable and the feasible set nonempty.

Dynamical systems approaches to continuous optimization replace the minimization in (1) by an initial value problem [16]. The sequential homotopy method proposed in [12] is based on the gradient/antigradient augmented Lagrangian flow and its numerical solution via projected backward Euler steps. In such approach, the penalty term in the augmented Lagrangian may be required to drive the flow toward a feasible minimizer, although it may slow down the convergence rate. Our work stems from the question: how can we prevent, or alleviate, such side effect while maintaining the convergence guarantees of the method? In Section 3 we find that scaling the dual variable's time is a suitable remedy and improves the convergence speed. Then, we consider the projected backward Euler solution of the time-scaled flow and characterize its relationship with the augmented Lagrangian framework [13, 5, 4]. This analysis reveals interrelations and suggests different perspectives on other methods for nonlinear programming, such as nonlinearly-constrained Lagrangian (NCL) [10], primal-dual augmented Lagrangian [14] and regularized methods [1, 2]. In particular, we observe that taking a projected backward Euler step along the augmented Lagrangian flow is equivalent to performing an iteration of a classical augmented Lagrangian method without applying the first-order dual estimate update. Thus, although in continuous-time the penalty term cannot be discarded in general, the discrete-time system can be globally asymptotically stable even without penalty term. We further explore this issue in Section 5 and demonstrate that the classical augmented Lagrangian method with dual estimate update corresponds to a discretized time-scaled flow without penalty term.

Building upon these elaborations, we design an algorithm for solving (1) via a sequence of primal-dual proximally regularized subproblems. This resolves the numerical difficulties due to subproblem infeasibility and lack of constraint qualification, yielding increased robustness. Also, globalised Newton-type methods can be adopted as a local solver, achieving fast convergence. From a computational point of view, our approach can be assimilated to primal-dual augmented Lagrangian or regularized methods [14, 1].

2 Augmented Lagrangian Flow

The first-order necessary optimality (or KKT) conditions of problem (1) and the augmented Lagrangian function play a key role in this work. The former require that, at a solution $x \in C$, there exist $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ such that

$$0 = \nabla f(x) + \nabla c(x)^\top y + z, \tag{2a}$$
$$0 = c(x), \tag{2b}$$
$$0 = x - \mathrm{P}_C(x + z), \tag{2c}$$

where P_C denotes the Euclidean projection onto C. The augmented Lagrangian function is defined by

$$\mathcal{L}_\rho(x,y) := f(x) + \frac{\rho}{2}\|c(x)\|^2 + y^\top c(x), \tag{3}$$

for some fixed penalty parameter $\rho \geq 0$. In [12], a projected primal-dual gradient/antigradient flow of the augmented Lagrangian $\mathcal{L}_\rho$ is considered, according to

$$\dot{x}(t) = \mathrm{P}_C\left(-\nabla_x \mathcal{L}_\rho(x(t), y(t))\right), \tag{4a}$$
$$\dot{y}(t) = \nabla_y \mathcal{L}_\rho(x(t), y(t)), \tag{4b}$$

where the gradients with respect to x and y are

$$\nabla_x \mathcal{L}_\rho(x,y) = \nabla f(x) + \nabla c(x)^\top \left[\rho c(x) + y\right],$$
$$\nabla_y \mathcal{L}_\rho(x,y) = c(x).$$

Following the flow defined by (4) with a projected backward Euler step, from $(\hat{x}, \hat{y}) \in C \times \mathbb{R}^m$ to $(x,y) \in C \times \mathbb{R}^m$ with stepsize $\Delta t > 0$, requires solving

$$0 = x - \mathrm{P}_C\left(\hat{x} - \Delta t\, \nabla_x \mathcal{L}_\rho(x,y)\right),$$
$$0 = y - \hat{y} - \Delta t\, c(x).$$

Introducing variable $z \in \mathbb{R}^n$, an equivalent subproblem is

$$0 = \nabla_x \mathcal{L}_\rho(x,y) + \frac{x - \hat{x}}{\Delta t} + z, \tag{5a}$$
$$0 = c(x) + \frac{\hat{y} - y}{\Delta t}, \tag{5b}$$
$$0 = x - \mathrm{P}_C\left(x + z\right), \tag{5c}$$

which reduces to (2) for $\Delta t \to +\infty$.

Example 1 (pendulum)**.** *Consider the problem*

$$\min_{x \in \mathbb{R}^2} \; x_2 \qquad \text{subject to } \; x_1^2 + x_2^2 = 1. \tag{6}$$

The necessary optimality conditions imply that $x = (0, \pm 1)$, $y = \mp 1/2$ must hold at a solution, where the critical point $x = (0,1)$, $y = -1/2$, is a maximum and the critical point $x = (0,-1)$, $y = 1/2$, is the only minimum.

Fig. 1 depicts trajectories of the flow (4) for Example 1, adopting the augmented Lagrangian with increasing values of the penalty parameter $\rho \geq 0$. Starting close to the inverted pendulum position, the minimum is found in all cases. However, "for larger values of the penalty parameter ρ, the primal trajectories are driven closer to the feasible set, but the dual trajectories converge more slowly" [12]. The following interpretation accounts for this phenomenon, referring to the definition in (3): by the minimization of $\mathcal{L}_\rho$ with respect to x, increasing ρ leads to an initial faster reduction in constraint violation. This also influences the maximization of $y^\top c(x)$ with respect to y, whose flow slows down, eventually taking longer to drive the constraint violation to zero. In other words, giving emphasis to the primal feasibility, the dual flow can be badly affected, and then the primal one too.

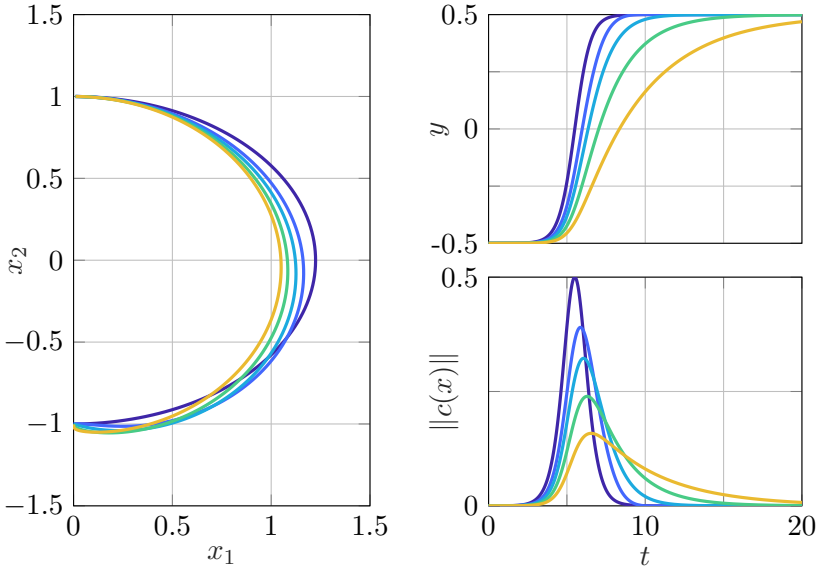

Figure 1: Primal (left), dual (top right), and constraint violation (bottom right) trajectories of the gradient/antigradient augmented Lagrangian flow for the pendulum of Example 1. The initial value is $x = (0.01, 1)$, $y = -1/2$, which is close to the inverted pendulum position. For larger values of the penalty parameter ρ, the primal trajectories are driven closer to the feasible set, but the dual trajectories converge more slowly.

3 Time-Scaled Lagrangian Flow

The gradient/antigradient flow (4) relates to the (projected) steepest descent method. On the vein of variable metric approaches, we consider a scaled flow with primal and dual scaling $\sigma_x, \sigma_y > 0$, according to

$$\dot{x}(t) = \sigma_x \, \mathrm{P}_C \left(-\nabla_x \mathcal{L}_\rho(x(t), y(t))\right), \tag{7a}$$

$$\dot{y}(t) = \sigma_y \, \nabla_y \mathcal{L}_\rho(x(t), y(t)). \tag{7b}$$

We refer to this continuous-time dynamical system as to $\mathcal{CT}(\sigma_x, \sigma_y, \rho)$. In light of (7), this can be understood as having different time scales for primal and dual variables. Intuitively, the right-hand side of (4) could be scaled by any positive definite matrix, retaining stability properties (and thus convergence guarantees) of the original flow. Moreover, the scaling could be time-varying, and state-dependent, provided it stays bounded away from zero. This resembles some dynamical systems approaches in unconstrained minimization [16]. In the following we will consider $\sigma_x = 1$ and focus on the dual scaling σ_y only.

Fig. 2 depicts trajectories of the flow (7) for Example 1, adopting increasing values of the dual scaling $\sigma_y \geq 1$. Starting close to the inverted pendulum position, the minimum is found in all cases. In contrast to the side effects of ρ shown in Fig. 1, for larger values of σ_y, the primal trajectories are driven closer to the feasible set, while the dual trajectories converge faster. Based on definition (3), this could be justified as follows: with a larger dual scaling σ_y,

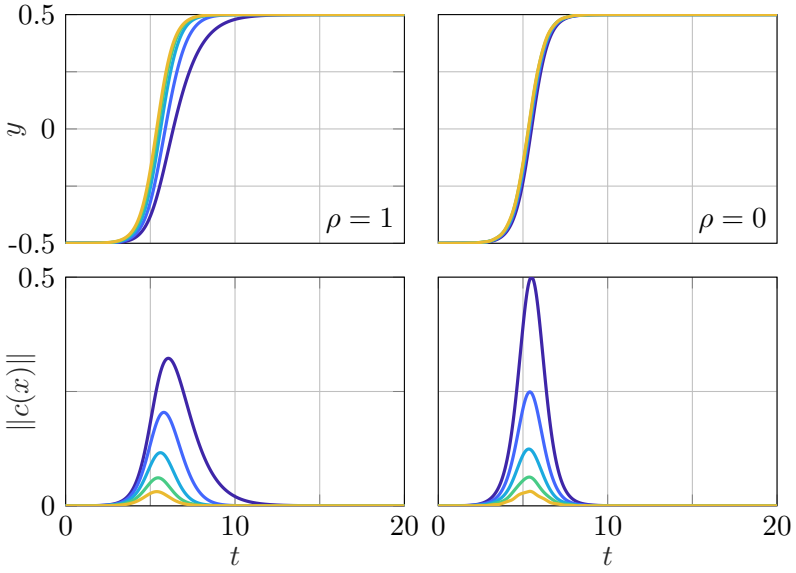

Figure 2: Dual (top) and constraint violation (bottom) trajectories of the gradient/antigradient augmented Lagrangian flow for the pendulum of Example 1, with $\rho = 1$ (left) or $\rho = 0$ (right). The initial value is $x = (0.01, 1)$, $y = -1/2$, which is close to the inverted pendulum position. For larger values of the dual scaling σ_y, the primal trajectories are driven closer to the feasible set and the dual trajectories converge faster.

the dual variable y tracks more closely the direction of $c(x)$, in order to maximize the term $y^\top c(x)$ in $\mathcal{L}_\rho$. Thus, for sufficiently large σ_y, we have $y^\top c(x) \approx \alpha \|c(x)\|^2$, for some positive scalar α that increases with σ_y. Hence, even for $\rho = 0$, the concurrent primal flow tends to minimize the objective function penalized by the constraints violation. However, although not necessary, "the augmentation term is important to guarantee convergence of the flow" [12]. In fact, the gradient/antigradient flow may require such augmentation term for convergence, as demonstrated by the following Example.

Example 2 (unbounded)**.** *Consider the problem*

$$\min_{x \in \mathbb{R}} \; -x^2/2 \qquad \text{subject to } \; x = 0. \tag{8}$$

The unique solution is its only feasible point $x = 0$, with Lagrange multiplier $y = 0$.

For Example 2, the time-scaled flow (7), namely $\mathcal{CT}(\sigma_x, \sigma_y, \rho)$, reduces to the continuous-time linear system

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \underbrace{\begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix} \begin{bmatrix} 1-\rho & -1 \\ 1 & 0 \end{bmatrix}}_{=:A_{\mathrm{CT}}} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \tag{9}$$

whose stability properties depend on the eigenvalues

$$\lambda_{1,2}(A_{\mathrm{CT}}) = \frac{\sigma_x}{2} \left((1-\rho) \pm \sqrt{(1-\rho)^2 - 4\frac{\sigma_y}{\sigma_x}} \right).$$

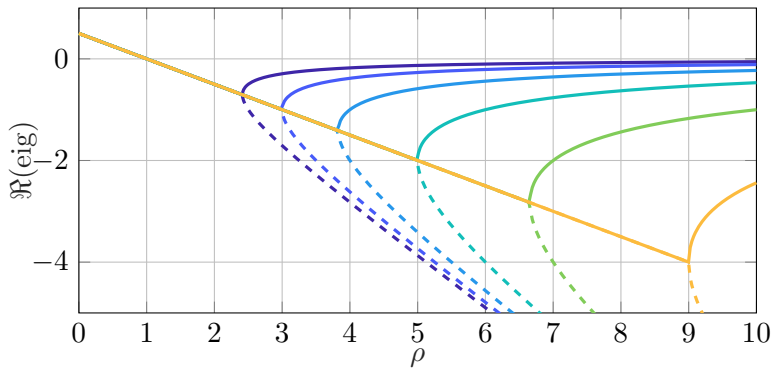

Figure 3: Real part of eigenvalues of continuous-time dynamics for Example 2 as a function of penalty parameter $\rho \geq 0$. For larger values of dual scaling σ_y, the bifurcation takes place at higher values of ρ.

Global asymptotic stability of the flow (9) is guaranteed only for $\rho > 1$, independently from σ_x and σ_y; cf. Fig. 3. In any case, according to the aforementioned interpretation, leveraging the dual scaling σ_y can alleviate, and possibly compensate for, the slowing effect of the penalty term, as demonstrated by the results in Fig. 2.

We now investigate the discrete-time system arising from the projected backward Euler solution of flow (7). In fact, the iterative nature of most optimization methods lends itself to being interpreted as a discrete-time process. We denote this discrete-time dynamical system by $\mathcal{DT}(\sigma_x, \sigma_y, \rho, \Delta t)$. Similarly to the unscaled update (5), we can obtain, from $(\hat{x}, \hat{y}) \in C \times \mathbb{R}^m$ and with stepsize $\Delta t > 0$, a projected backward Euler step $(x, y) \in C \times \mathbb{R}^m$ by solving

$$0 = \nabla_x \mathcal{L}_\rho(x, y) + \sigma(x - \hat{x}) + z, \tag{10a}$$
$$0 = c(x) + \mu(\hat{y} - y), \tag{10b}$$
$$0 = x - \mathrm{P}_C\left(x + z\right), \tag{10c}$$

where $\sigma \equiv 1/(\sigma_x \Delta t)$ and $\mu \equiv 1/(\sigma_y \Delta t)$. Subproblem (10) admits a unique solution for σ sufficiently large, that is, Δt sufficiently small. Moreover, the associated discrete-time trajectory may be globally convergent to a feasible minimizer $x^\star$ of (1), with some dual variable $y^\star$, provided ρ is sufficiently large or μ sufficiently small. Indeed, in Section 4 we argue that, under mild assumptions, there exists always a sufficiently small μ (and a sufficiently large σ) so that the discrete-time trajectory generated by (10), namely by $\mathcal{DT}(\sigma_x, \sigma_y, \rho, \Delta t)$, converges to a solution of (2), namely a KKT point for (1), even for $\rho = 0$.

Concerning Example 2, the subproblem in (10) reads

$$0 = (\rho - 1)x + y + \sigma(x - \hat{x}),$$
$$0 = x + \mu(\hat{y} - y),$$

which reduces to a discrete-time linear system that can be written as $(x, y)^\top = A_{\mathrm{DT}}(\hat{x}, \hat{y})^\top$, for some matrix A_{DT} (dependent on σ, μ, and ρ). The stability properties of such system depend on the eigenvalues $\lambda_{1,2}(A_{\mathrm{DT}})$, whose magnitude is depicted in Fig. 4. In contrast to the

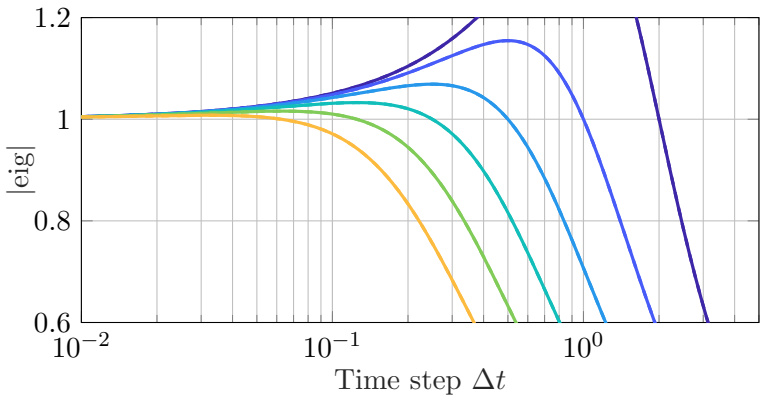

Figure 4: Magnitude of eigenvalues of discrete-time dynamics for Example 2 as a function of stepsize $\Delta t > 0$, for $\rho = 0$. For larger values of dual scaling σ_y, the curve falls earlier below the unit value.

continuous-time case, the discrete-time system can be stable even without augmentation term, that is, with $\rho = 0$. In fact, selecting a sufficiently large stepsize Δt, the discrete-time system becomes asymptotically stable, and the higher σ_y, the lower the minimum threshold on Δt. This can be related to the numerical dissipation (or diffusion) introduced by the discretization scheme adopted for solving the differential equation (7). Notice, in particular, that larger stepsize Δt and dual scaling σ_y induce stronger numerical dissipation, and thus more stabilizing effects.

4 Primal-Dual Regularization

In this Section we establish and discuss connections between $\mathcal{DT}(\sigma_x, \sigma_y, \rho, \Delta t)$, namely the discrete-time flow (10), and several Lagrangian-based approaches for nonlinear programming. In particular, we refer to the classical [5, 4] and primal-dual [14] augmented Lagrangian frameworks, and the NCL algorithm [10].

The subproblem in (10) can be interpreted as necessary optimality conditions of a regularized version of the augmented form of (1), that reads

$$\begin{aligned} \min_{x \in C, r \in \mathbb{R}^m} \quad & f(x) + \frac{\rho}{2}\|c(x)\|^2 + \frac{\sigma}{2}\|x - \hat{x}\|^2 + \frac{1}{2\mu}\|r - \mu\hat{y}\|^2 \\ \text{subject to} \quad & c(x) + r = 0. \end{aligned} \tag{11}$$

This primal-dual proximally regularized subproblem is always feasible and satisfies the linear independence constraint qualification at all its feasible points. It corresponds to the NCL subproblem [10] with additional primal proximal (or Tikhonov) regularization and penalty term. The corresponding dual variable can be recovered from a solution as $y = \hat{y} - r/\mu$. Formally solving for $r = -c(x)$ and substituting, we obtain

$$\min_{x \in C} \quad f(x) + \frac{\rho}{2}\|c(x)\|^2 + \frac{\sigma}{2}\|x - \hat{x}\|^2 + \frac{1}{2\mu}\|c(x) + \mu\hat{y}\|^2, \tag{12}$$

that goes back to the bound-constrained Lagrangian (BCL) subproblem [5]. From (12), the dual variable can be recovered as $y = \hat{y} + c(x)/\mu$. Furthermore, we notice that the necessary optimality conditions of the following bound-constrained problem are equivalent to (10):

$$\min_{x \in C, y \in \mathbb{R}^m} \quad f(x) + \frac{\rho}{2}\|c(x)\|^2 + \frac{\sigma}{2}\|x - \hat{x}\|^2 + \frac{1}{\mu}\|c(x) + \mu(\hat{y} - y/2)\|^2 + \frac{\mu}{4}\|y\|^2. \tag{13}$$

For $\rho = 0$ and $\sigma = 0$, this reduces to the primal-dual BCL (pdBCL) subproblem [14, Ch 4].

The sequential homotopy method [12] generates a sequence of proximally regularized subproblems, whose advantageous properties are based on the underlying continuous-time formulation and leverage its stability to derive convergence guarantees. However, this may require an augmentation term to insure global convergence to a feasible minimizer. In the discrete-time settings, instead, the formulations given in (11)–(13) show that the augmentation term is indeed not necessary, under mild assumptions, since they stem from the augmented Lagrangian framework [4]. Indeed, the discrete-time system based on (10) corresponds to an augmented Lagrangian method, possibly with subproblems including primal proximal regularization and a penalty term.

Neglecting the penalty term, parameters σ and μ directly control primal and dual regularization; cf. (11)–(13). The regularization can be interpreted as a proximal-point augmented Lagrangian method applied to (1), in the same vein as [15]. Indeed, the dual regularization parameter μ also controls the constraint penalization. Finally, we highlight that solving (1) via subproblems (10) asymptotically reduces to a sequence of regularized steps applied to the original, unperturbed optimality system, as in [1]. This closely relates to the exactness of the proximal primal-dual regularization, in the sense of Proposition 1. It follows from the direct comparison of (2) and (10), similarly to [2, Thm 1].

Proposition 1. *Consider problem (1), its KKT conditions in (2), subproblem (10), and let $\rho \geq 0$.*

- *Suppose $(\hat{x}, \hat{y})$ solves (10) for some $\sigma \geq 0$ and $\mu \geq 0$. Then, $(\hat{x}, \hat{y})$ is a KKT point of problem (1).*
- *Alternatively, suppose (x°, y°) solves (10) for $\sigma = 0$ and some $\mu > 0$, with x° feasible for (1). Then, $y^\circ = \hat{y}$ and (x°, y°) is a KKT point of problem (1).*
- *Conversely, suppose $(x^\star, y^\star)$ is a KKT point of problem (1). Then, $(\hat{x}, \hat{y}) := (x^\star, y^\star)$ solves (10) for any $\sigma \geq 0$ and $\mu \geq 0$.*

5 First-Order Multiplier Estimate

In the classical augmented Lagrangian framework [4], one minimizes the augmented Lagrangian function $\mathcal{L}_\rho$, given some penalty parameter and dual estimate, and then applies a first-order dual update. The latter step helps improving the successive dual estimate. As detailed in Section 4, taking a projected backward Euler step by solving (10), namely each step of $\mathcal{DT}(\sigma_x, \sigma_y, \rho, \Delta t)$, corresponds to minimizing an augmented Lagrangian function; cf. (12)–(13). Inspired by that

finding, we consider a discrete-time system consisting in the projected backward Euler step followed by an update of the dual variable. More precisely, from an estimate $(\hat{x}, \hat{y})$, we find (x, y) by solving (10) and then set $(x, \rho c(x) + y)$ as the successive estimate. We refer to the sequence of estimates as to the evolution of this discrete-time system with dual update, denoted $\mathcal{DT}^{\mathrm{upd}}(\sigma_x, \sigma_y, \rho, \Delta t)$. Clearly, such transformed system is no more associated to the numerical solution of the continuous-time flow (7), namely to $\mathcal{CT}(\sigma_x, \sigma_y, \rho)$.

Let us analyze the discrete-time system $\mathcal{DT}^{\mathrm{upd}}$ and investigate whether it derives from an underlying continuous-time system. First of all, we introduce $u \equiv \rho c(x) + y$, so that the update rule reads $(\hat{x}, \hat{y}) \leftarrow (x, u)$. Thus, taking a step for $\mathcal{DT}^{\mathrm{upd}}$ coincides with solving (10) with respect to (x, u). Since $\Delta t, \sigma_y > 0$ and $\rho \geq 0$, the subproblem in (10) can be equivalently expressed in terms of (x, u) as

$$0 = \nabla f(x) + \nabla c(x)^\top u + \sigma(x - \hat{x}) + z, \tag{14a}$$

$$0 = c(x) + \mu^{\mathrm{upd}}(\hat{y} - u), \tag{14b}$$

$$0 = x - \mathrm{P}_C(x + z), \tag{14c}$$

where $\mu^{\mathrm{upd}} = \mu/(1 + \rho\mu) \in (0, \mu]$. Comparing (10) and (14), we discover that

$$\mathcal{DT}^{\mathrm{upd}}(\sigma_x, \sigma_y, \rho, \Delta t) = \mathcal{DT}(\sigma_x, \sigma_y^{\mathrm{upd}}, 0, \Delta t),$$

with $\sigma_y^{\mathrm{upd}} = \sigma_y + \rho/\Delta t \geq \sigma_y$. Hence, applying the dual update corresponds to having $\rho = 0$, an increased dual regularization, since $\mu^{\mathrm{upd}} \leq \mu$, and a larger dual scaling, since $\sigma_y^{\mathrm{upd}} \geq \sigma_y$. This highlights once again that, in some regards, the augmentation term corresponds to, and can be compensated by, a faster dual flow. Furthermore, it also shows that $\mathcal{DT}^{\mathrm{upd}}$ stems from a projected backward Euler solution of the flow (7), since $\mathcal{DT}(\sigma_x, \sigma_y^{\mathrm{upd}}, 0, \Delta t)$ originates in $\mathcal{CT}(\sigma_x, \sigma_y^{\mathrm{upd}}, 0)$. Still, the relationship between the transformed discrete-time system and the associated continuous-time one remains ambiguous, for σ_y^{upd} depends on the stepsize Δt. Nonetheless, for a given stepsize, the evolution of $\mathcal{CT}(\sigma_x, \sigma_y^{\mathrm{upd}}, 0)$ is unequivocal. We notice, however, that it can be unstable even if the discrete-time counterpart is stable, as in Example 2; see Fig. 3 at $\rho = 0$ and Fig. 4.

Fig. 5 depicts primal-dual trajectories of continuous- and discrete-time systems for Example 2, with penalty parameter $\rho = 4$. The evolution of $\mathcal{DT}$ approaches $\mathcal{CT}$'s and converges slowly to the solution $(0, 0)$. In contrast, $\mathcal{DT}^{\mathrm{upd}}$ is not bound to approximate $\mathcal{CT}$'s trajectory, and attains faster convergence toward the solution. Although it leads to more oscillations around the solution, the dual update appears to improve the convergence rate. This is indeed shown in Fig. 6, that depicts error trajectories for Example 2, namely the Euclidean distance of a primal-dual pair $(x(t), y(t))$ to the solution $(0, 0)$. For $\rho < 1$, the trajectories diverge. With larger values of ρ, initially the error decreases faster for $\mathcal{CT}$ and $\mathcal{DT}$, but then it does so at a lower rate, suggesting there might be an optimal ρ for the augmented Lagrangian flow. Conversely, $\mathcal{DT}^{\mathrm{upd}}$ yields an essentially different behaviour: the higher the penalty parameter ρ, the faster the (linear) convergence toward the solution.

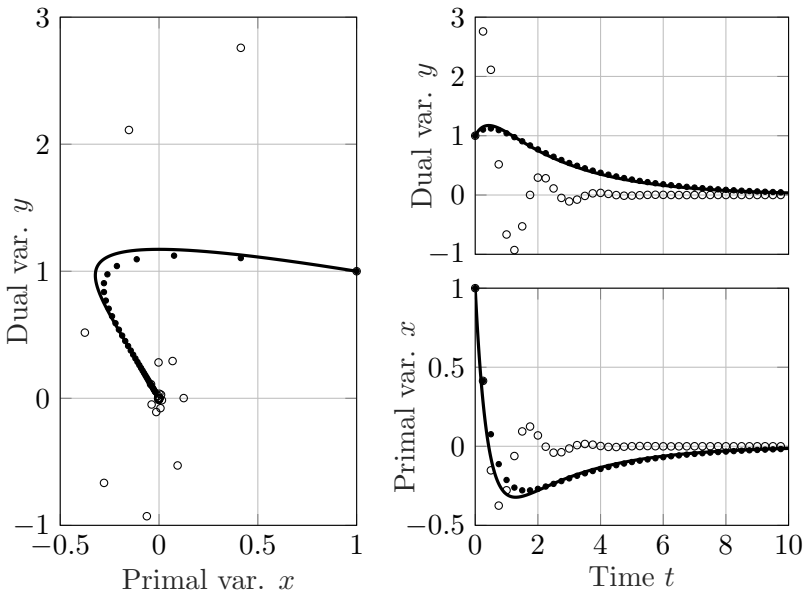

Figure 5: Primal-dual trajectories for Example 2, starting from $x = 1$ and $y = 1$, with $\rho = 4$ and stepsize $\Delta t = 0.25$. Continuous-time (line), discrete-time (dot), and discrete-time with dual update (circle).

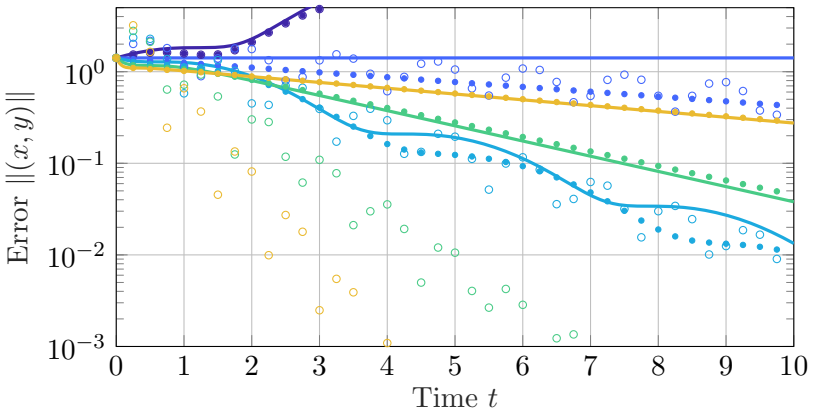

Figure 6: Error trajectories for Example 2, starting from $x = 1$ and $y = 1$, with stepsize $\Delta t = 0.25$. Continuous-time (line), discrete-time (dot), and discrete-time with dual update (circle). For $\rho = 0$, the trajectories diverge. For larger values of ρ, the discrete-time system with dual update exhibits faster convergence, whose rate increases with ρ.

6 Numerical Results

We discuss an illustrative method and present computational results on a set of equality-constrained nonlinear programs, i.e., problems of the form (1) with $C = \mathbb{R}^n$. We test and compare our implementation against IPOPT [17] and NCL [11] in Julia [3]. The code to generate the numerical results is available online [6]. All the experiments were carried out on a laptop running Ubuntu 18.04 with Intel Core i7-3537U 2.00 GHz and 8 GB RAM.

6.1 Algorithm

Algorithm 1 provides pseudocode for a prototypical homotopy augmented Lagrangian method for equality-constrained optimization, dubbed HALeqO. It builds upon the various formulations and interpretations given above, and it consists of an outer loop over proximally regularized subproblems. Seeking a feasible minimizer, each subproblem entails the unconstrained minimization of the merit function $\mathcal{M}$ corresponding to the objective function in (13), with $\rho = \sigma = 0$. Each subproblem is solved via a globalised Newton's method, corresponding to the inner loop, before updating the estimate $(\hat{x}, \hat{y})$. The dual regularization parameter μ is adapted based on the decrease in constraint violation. The primal proximal term contributes only to the Hessian regularization for the minimization of $\mathcal{M}$: this gives us a principled way to select σ, and possibly speeds up the convergence if the primal regularization is not necessary. Thus, each subproblem yields the KKT conditions

$$0 = r(x, y) := \begin{pmatrix} \nabla f(x) + \nabla c(x)^\top y, \\ c(x) + \mu(\hat{y} - y) \end{pmatrix} \tag{15}$$

and regularized linear systems of the form

$$\begin{bmatrix} H(x, y) + \sigma I & \nabla c(x)^\top \\ \nabla c(x) & -\mu I \end{bmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = -r(x, y), \tag{16}$$

where $H(x, y)$ denotes the Hessian matrix of the Lagrangian function, namely $\nabla^2_{xx}\mathcal{L}_0(x, y)$. Then, given the search direction $d = (\delta x, \delta y)$, we adopt a backtracking linesearch with Armijo's sufficient decrease condition:

$$\mathcal{M}(x + \tau\delta x, y + \tau\delta y) \leq \mathcal{M}(x, y) + \eta\tau d^\top \nabla\mathcal{M}(x, y). \tag{17}$$

For $H(x, y) + \sigma I \succ 0$ and $\mu > 0$, this procedure is well-defined since d is a descent direction for $\mathcal{M}$ at (x, y), namely $d^\top \nabla\mathcal{M}(x, y) < 0$ for every $d \neq 0$. Moreover, the conditions $d = 0$, $r(x, y) = 0$, and $\nabla\mathcal{M}(x, y) = 0$ are equivalent.

Bounds and inequality constraints can be easily accounted for by considering a projected or semismooth Newton's method, instead of the classical one in Algorithm 1.

6.2 Setup

We consider the equality-constrained CUTEst problems [9] with $1 \leq n, m \leq 100$; this yields 161 problems. We set the tolerance $\epsilon = 10^{-6}$ and deem optimal a primal-dual pair (x, y) if

Algorithm 1 HALeqO: homotopy augmented Lagrangian method for equality-constrained optimization

Initialize $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\epsilon, \sigma, \mu > 0$
Select $\theta, \beta, \kappa_\mu \in (0,1)$, $\kappa_\sigma > 1$, $\eta \in (0, 1/2)$
Set $(\hat{x}, \hat{y}) \leftarrow (x, y)$ and $\sigma_0 \leftarrow \sigma$
while (x, y) does not solve (2) to within ϵ **do**
  if (x, y) solves (15) to within ϵ **then**
    if $\|c(x)\| > \theta \|c(\hat{x})\|$ **then**
      Set $\mu \leftarrow \kappa_\mu \mu$ *dual regul. update*
    end if
    Set $(\hat{x}, \hat{y}) \leftarrow (x, y)$ *estimate update*
  end if
  Set $\sigma \leftarrow \sigma_0$
  while $H(x, y) + \sigma I \not\succeq 0$ **do** *Hessian regul.*
    Set $\sigma \leftarrow \kappa_\sigma \sigma$
  end while
  Compute $(\delta x, \delta y)$ by solving (16) *search direction*
  Set $\tau \leftarrow 1$
  while (17) does not hold **do** *linesearch*
    Set $\tau \leftarrow \beta \tau$
  end while
  Set $(x, y) \leftarrow (x, y) + \tau(\delta x, \delta y)$
end while
return

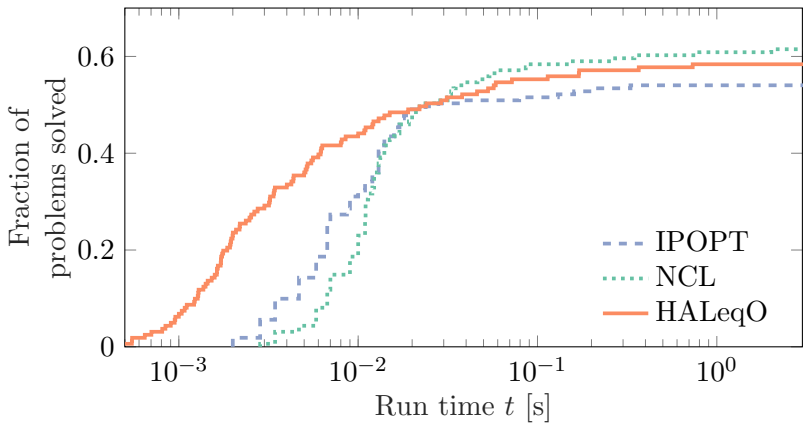

Figure 7: Comparison on equality-constrained CUTEst problems with time profiles.

it satisfies (2) within ϵ, otherwise we consider it a failure. In IPOPT and NCL, we leave all the other settings to the internal defaults. In HALeqO, we initialize $\sigma = \mu = 10^{-3}$ and set $\theta = \beta = 0.5$, $\kappa_\mu = 0.1$, $\kappa_\sigma = 10$, $\eta = 10^{-4}$; moreover, we limit the number of iterations to 3000, corresponding to IPOPT's default value. Notice that NCL performs at most 20 outer iterations, each one invoking IPOPT once.

We compare the solvers by means of time profiles. Let S, P, and $t_{s,p}$ denote the set of solvers, the set of problems, and the time required for solver $s \in S$ to return a solution for problem $p \in P$. We plot the functions $f_s : \mathbb{R} \to [0, 1]$, $s \in S$, defined by $f_s(t) := |\{p \in P : t_{s,p} \leq t\}|/|P|$. Considering $t_{s,p} = +\infty$ when solver s fails on problem p, $f_s(t)$ is the fraction of problems solved by solver s within the runtime t. Note that, in contrast to performance profiles [7], the time profile $t \mapsto f_s(t)$ is independent from other solvers and displayed with the actual timings of solver s.

6.3 Results

Computational results are summarized as time profiles in Fig. 7. NCL shows the best reliability, solving 100 problems out of 161, with HALeqO slightly behind with 95. The failure rate for IPOPT (88 solved) is higher because it does not handle problems with more constraints than variables. At the same time, HALeqO exhibits the fastest execution on many problems, as evidenced also by the performance profiles in Fig. 8, yet aware of their limitations [8]. Thus, despite its simplicity, HALeqO strikes a promising balance between speed and robustness.

References

[1] Paul Armand and Riadh Omheni. A globally and quadratically convergent primal–dual augmented Lagrangian algorithm for equality constrained optimization. *Optimization Methods and Software*, 32(1):1–21, 2017.

[2] Sylvain Arreckx and Dominique Orban. A regularized factorization-free method for equality-constrained optimization. *SIAM Journal on Optimization*, 28(2):1613–1639, 2018.

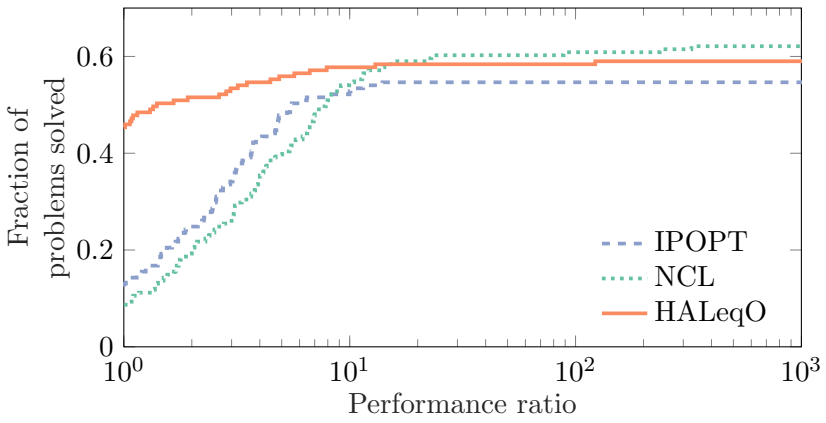

Figure 8: Performance profiles comparing runtimes on equality-constrained CUTEst problems.

[3] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.

[4] Ernesto G. Birgin and José Mario Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 4 2014.

[5] Andrew R. Conn, Nicholas I. M. Gould, and Philippe L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572, 4 1991.

[6] Alberto De Marchi. HALeqO.jl. Software on Zenodo, 3 2021.

[7] Elizabeth D. Dolan and Jorge J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

[8] Nicholas Gould and Jennifer Scott. A note on performance profiles for benchmarking software. *ACM Trans. Math. Softw.*, 43(2), 9 2016.

[9] Nicholas I. M. Gould, Dominique Orban, and Philippe L. Toint. CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60(3):545–557, 4 2015.

[10] Ding Ma, Kenneth L. Judd, Dominique Orban, and Michael A. Saunders. Stabilized optimization via an NCL algorithm. In Mehiddin Al-Baali, Lucio Grandinetti, and Anton Purnama, editors, *Numerical Analysis and Optimization*, pages 173–191, Cham, 2018. Springer International Publishing.

[11] Ding Ma, Dominique Orban, and Michael Saunders. A Julia implementation of Algorithm NCL for constrained optimization, 1 2021.

[12] Andreas Potschka and Hans Georg Bock. A sequential homotopy method for mathematical programming problems. *Mathematical Programming*, 3 2020.

[13] Michael J. D. Powell. *A method for nonlinear constraints in minimization problems*, pages 283–298. Academic Press, 1969.

[14] Daniel P. Robinson. *Primal-Dual Methods for Nonlinear Optimization*. PhD thesis, University of California, San Diego, 9 2007.

[15] Ralph Tyrrell Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of operations research*, 1(2):97–116, 5 1976.

[16] Johannes Schropp and I. Singer. A dynamical systems approach to constrained minimization. *Numerical Functional Analysis and Optimization*, 21(3-4):537–551, 2000.

[17] Andreas Wächter and Lorenz T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 3 2006.