Augmented Lagrangian methods as dynamical systems for constrained optimization

Alberto De Marchi

***Abstract*— Dynamical systems approaches to constrained optimization often rely on a penalization term to reach feasible points, at the cost of slower convergence. However, we show one can construct a discrete-time system that maintains the same convergence guarantees without requiring such penalization term. We demonstrate that the sequential homotopy method, namely taking projected backward Euler steps on a projected gradient/anti-gradient augmented Lagrangian flow, matches 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.**

I. 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 [1]. We investigate the relationships and interactions between this approach and augmented Lagrangian (AL) methods as discrete-time counterparts. The sequential homotopy method [2] is based on the gradient/anti-gradient AL flow and its numerical solution via projected backward Euler steps. In such approach, the penalty term in the AL function 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 III 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 AL framework [3], [4]. This analysis reveals interrelations and suggests different perspectives on other methods for nonlinear programming, such as primal-dual AL [5], nonlinearly-constrained Lagrangian (NCL) [6] and regularized methods [7], [8]. In particular, we observe that taking a projected backward Euler step along the AL flow is equivalent to performing an iteration of a classical AL 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, thanks to the convergence properties inherited from the AL framework. We further explore this issue in Section V and demonstrate that the classical AL 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 AL or regularized methods [5], [7].

II. AUGMENTED LAGRANGIAN FLOW

The first-order necessary optimality (or KKT) conditions of problem (1) and the AL 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

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

where $\mathcal{P}_C$ denotes the Euclidean projection onto C. The AL 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 $\rho \geq 0$. The sequential homotopy method [2] considers the projected primal-dual gradient/anti-gradient flow of the AL function $\mathcal{L}_\rho$

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

where $T(C, x)$ denotes the tangent cone to C at x and the gradients with respect to x and y are

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

Alberto De Marchi is with the Universität der Bundeswehr München, at the Institute for Applied Mathematics and Scientific Computing (LRT-1), 85577 Neubiberg/Munich, Germany. EMAIL: alberto.demarchi@unibw.de, ORCID: 0000-0002-3545-6898.

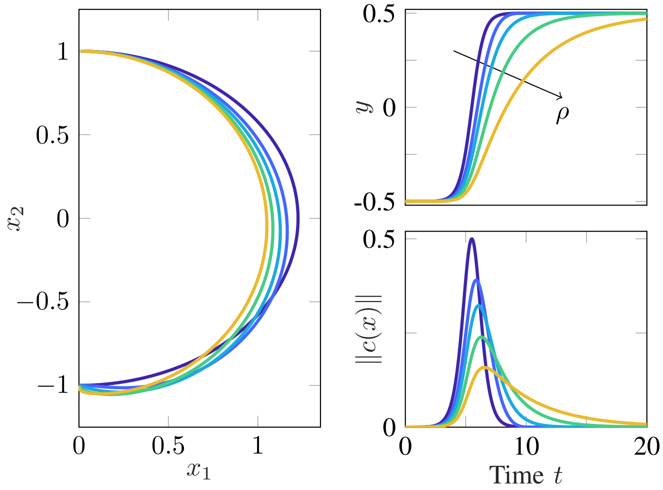

Fig. 1. Primal (left), dual (top right), and constraint violation (bottom right) trajectories of the gradient/anti-gradient AL flow for the pendulum of Example II.1. The initial value is $x = (0.01, 1)$, $y = -1/2$, which is close to the inverted pendulum position. For larger values of the penalty parameter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

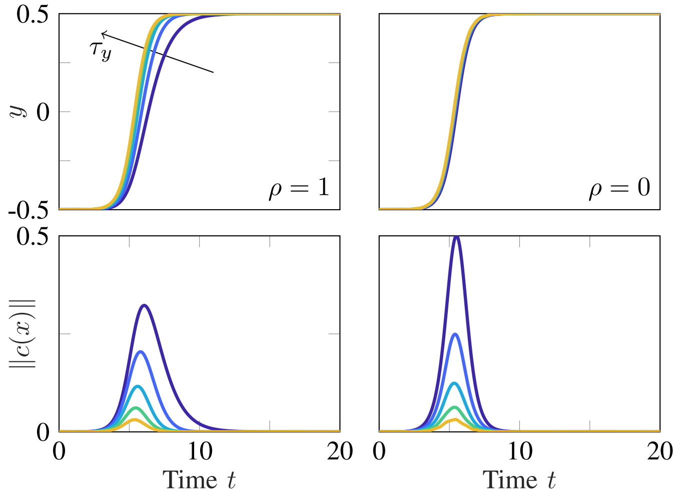

Fig. 2. Dual (top) and constraint violation (bottom) trajectories of the gradient/anti-gradient AL flow for the pendulum of Example II.1, with $\rho = 1$ (left) or $\rho = 0$ (right). The initial value is $x = (0.01, 1)$, $y = -1/2$, which is close to the inverted pendulum position. For larger values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

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 $\Delta t > 0$, by [2, Lemma 6], requires solving

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

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

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

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

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

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

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 II.1, adopting the AL function 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" [2]. Based on (3), the following interpretation accounts for this phenomenon: 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. Hence, giving emphasis to the primal feasibility can badly affect the dual flow, slowing down the primal one and the overall progress too.

III. Time-Scaled Lagrangian Flow

The gradient/anti-gradient AL 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 $\tau_x, \tau_y > 0$, according to

$$\dot{x}(t) = \tau_x\, \mathcal{P}_{T(C, x(t))}\left(-\nabla_x \mathcal{L}_\rho(x(t), y(t))\right), \tag{6a}$$
$$\dot{y}(t) = \tau_y\, \nabla_y \mathcal{L}_\rho(x(t), y(t)). \tag{6b}$$

We refer to this continuous-time dynamical system as to $\mathcal{CT}(\tau_x, \tau_y, \rho)$. In light of (6), 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 [1]. In the following we will consider $\tau_x = 1$ and focus on the dual scaling τ_y only, without loss of generality.

Fig. 2 depicts trajectories of the flow (6) for Example II.1, adopting increasing values of the dual scaling $\tau_y \geq 1$. 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 τ_y, the primal trajectories are driven closer to the feasible set, while the dual trajectories converge faster. Based on (3), this could be justified as follows: with a larger dual scaling τ_y, 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 τ_y, we have $y^\top c(x) \approx \omega \|c(x)\|^2$, for some positive scalar ω that increases with τ_y. Hence, even

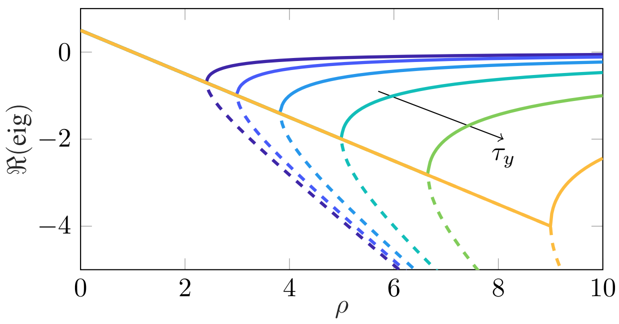

Fig. 3. Real part of eigenvalues of continuous-time dynamics for Example III.1 as a function of penalty parameter $\rho \geq 0$. For larger values of dual scaling τ_y, the bifurcation takes place at higher values of ρ.

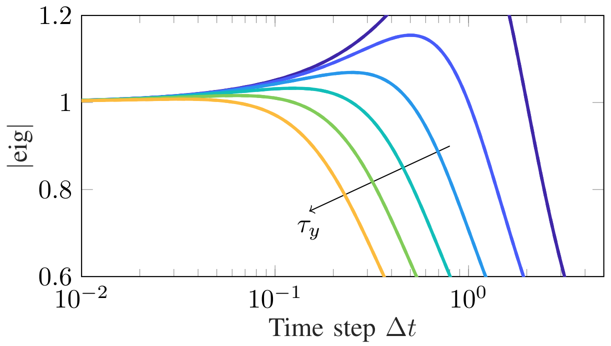

Fig. 4. Magnitude of eigenvalues of discrete-time dynamics for Example III.1 as a function of stepsize $\Delta t > 0$, for $\rho = 0$. For larger values of dual scaling τ_y, the curve falls earlier below the unit value.

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" [2]. In fact, the gradient/anti-gradient AL flow may require such augmentation, or penalty, term for convergence, as demonstrated by the following Example.

Example III.1 (unbounded)**.** Consider the problem

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

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

For Example III.1, the time-scaled flow (6), namely $\mathcal{CT}(\tau_x, \tau_y, \rho)$, reduces to the continuous-time linear system

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = A_{\text{CT}} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \quad A_{\text{CT}} := \begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix} \begin{bmatrix} 1-\rho & -1 \\ 1 & 0 \end{bmatrix},$$

whose stability properties depend on the eigenvalues

$$\lambda_{1,2}(A_{\text{CT}}) = \frac{\tau_x}{2} \left((1-\rho) \pm \sqrt{(1-\rho)^2 - 4\frac{\tau_y}{\tau_x}} \right).$$

Global asymptotic stability of this flow is guaranteed only for $\rho > 1$, independently from τ_x and τ_y; cf. Fig. 3. In any case, according to the aforementioned interpretation, leveraging the dual scaling τ_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 (6). 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 $\mathcal{DT}(\tau_x, \tau_y, \rho, \Delta t)$. 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

$$0 = \nabla_x \mathcal{L}_\rho(x, y) + \sigma(x - \hat{x}) + z, \tag{7a}$$
$$0 = c(x) + \mu(\hat{y} - y), \tag{7b}$$
$$0 = x - \mathcal{P}_C(x + z), \tag{7c}$$

where $\sigma := 1/(\tau_x \Delta t)$ and $\mu := 1/(\tau_y \Delta t)$. Subproblem (7) 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.

Concerning Example III.1, the subproblem in (7) 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 $(x, y)^\top = A_{\text{DT}}(\hat{x}, \hat{y})^\top$, for some matrix A_{DT} (dependent on σ, μ, and ρ). The stability properties of such system depend on the eigenvalues $\lambda_{1,2}(A_{\text{DT}})$, whose magnitude is depicted in Fig. 4. In general, and in stark contrast to the continuous-time case, the discrete-time system can be stable even without penalty term, namely with $\rho = 0$; cf. Section IV. In fact, selecting a sufficiently large stepsize Δt, the discrete-time system becomes asymptotically stable, and the higher τ_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 (6). Notice, in particular, that larger stepsize Δt and dual scaling τ_y induce stronger numerical dissipation, and thus more stabilizing effects.

IV. Primal-Dual Regularization

In this Section we establish and discuss connections between several Lagrangian approaches for nonlinear programming and the discrete-time system $\mathcal{DT}$ due to (7). Based on this close correspondence, we argue that, under mild assumptions, there always exists a sufficiently small μ (and a sufficiently large σ) so that the discrete-time trajectory generated by (7) converges to a solution of (2), namely $\mathcal{DT}$ approaches a KKT point for (1), even with $\rho = 0$. In particular, we refer to the classical [3], [4] and primal-dual [5] AL frameworks, and to the NCL algorithm [6].

The subproblem in (7) can be interpreted as necessary optimality conditions of a regularized version of the augmented form of (1), that reads

$$\min_{x \in C, r \in \mathbb{R}^m} \quad f(x) + \frac{\rho}{2}\|c(x)\|^2 + \frac{\sigma}{2}\|x - \hat{x}\|^2 + \frac{1}{2\mu}\|r - \mu\hat{y}\|^2$$
$$\text{subject to} \quad c(x) + r = 0. \tag{8}$$

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 [6] with additional primal proximal 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

$$\min_{x \in C} f(x) + \frac{\rho}{2}\|c(x)\|^2 + \frac{\sigma}{2}\|x - \hat{x}\|^2 + \frac{1}{2\mu}\|c(x) + \mu\hat{y}\|^2, \quad (9)$$

that goes back to the bound-constrained Lagrangian (BCL) subproblem [3]. From (9), 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 (7):

$$\begin{aligned}\min_{x \in C, y \in \mathbb{R}^m} \quad & f(x) + \frac{\rho}{2}\|c(x)\|^2 + \frac{\sigma}{2}\|x - \hat{x}\|^2 \\ & + \frac{1}{\mu}\|c(x) + \mu(\hat{y} - y/2)\|^2 + \frac{\mu}{4}\|y\|^2. \quad (10)\end{aligned}$$

For $\rho = 0$ and $\sigma = 0$, this reduces to the primal-dual BCL (pdBCL) subproblem [5, Ch 4].

The sequential homotopy method [2] 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 a penalty term to insure global convergence to a feasible minimizer. In the discrete-time settings, instead, the formulations given in (8)–(10) correspond to AL subproblems, possibly including a penalty term and a primal proximal regularization, controlled by ρ and σ respectively. Thus, analogously to the AL framework [4], we conclude that, under mild assumptions, the penalty term is not necessary to drive the discrete-time system $\mathcal{DT}$, based on (7), to a feasible minimizer of (1).

Neglecting the penalty term, parameters σ and μ directly control primal and dual regularization; cf. (8)–(10). The regularization can be interpreted as a proximal AL method applied to (1), that traces back to [9]. Indeed, the dual regularization parameter μ also controls the constraint penalization. Finally, we highlight that solving (1) via subproblems (7) asymptotically reduces to a sequence of regularized steps applied to the original, unperturbed optimality system, as in [7]. This closely relates to the exactness of the proximal primal-dual regularization [8, Thm 1], in the sense of Proposition IV.1.

Proposition IV.1. *Consider problem* (1)*, its KKT conditions in* (2)*, subproblem* (7)*, and let* $\rho \geq 0$*.*

- *Suppose* $(\hat{x}, \hat{y})$ *solves* (7) *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* (7) *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* (7) *for any* $\sigma \geq 0$ *and* $\mu \geq 0$*.*

Proof. From direct comparison of (2) and (7). □

V. FIRST-ORDER MULTIPLIER ESTIMATE

In the classical AL framework [4], one minimizes the AL 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 IV, taking a projected backward Euler step by solving (7), namely each step of $\mathcal{DT}(\tau_x, \tau_y, \rho, \Delta t)$, corresponds to minimizing an AL function; cf. (9)–(10). 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 (7) 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 $\mathcal{DT}^{\text{upd}}(\tau_x, \tau_y, \rho, \Delta t)$. Clearly, such transformed system is no more associated to the numerical solution of the continuous-time flow (6), namely to $\mathcal{CT}(\tau_x, \tau_y, \rho)$. We now analyze the discrete-time system $\mathcal{DT}^{\text{upd}}$ and investigate whether it derives from an underlying continuous-time system.

Lemma V.1. *Let* $\tau_x, \tau_y, \Delta t > 0$*,* $\rho \geq 0$ *be arbitrary but fixed. Then, with* $\tau_y^{upd} := \tau_y + \rho/\Delta t \geq \tau_y$*, the discrete-time systems* $\mathcal{DT}^{upd}(\tau_x, \tau_y, \rho, \Delta t)$ *and* $\mathcal{DT}(\tau_x, \tau_y^{upd}, 0, \Delta t)$ *coincide.*

Proof. Define $u := \rho c(x) + y$, so that the update rule reads $(\hat{x}, \hat{y}) \leftarrow (x, u)$. Hence, taking a step for $\mathcal{DT}^{\text{upd}}$ coincides with solving (7) with respect to (x, u). Since $\Delta t, \tau_y > 0$, $\rho \geq 0$, the subproblem in (7) can be equivalently expressed in terms of (x, u) as

$$\begin{aligned} 0 &= \nabla f(x) + \nabla c(x)^\top u + \sigma(x - \hat{x}) + z, \\ 0 &= c(x) + \mu^{\text{upd}}(\hat{y} - u), \\ 0 &= x - \mathcal{P}_C(x + z), \end{aligned}$$

where $\mu^{\text{upd}} := \mu/(1 + \rho\mu) \in (0, \mu]$. Comparing with (7), the result follows from recalling that $\mu := 1/(\tau_y \Delta t)$. □

The dual update leads to having $\rho = 0$, an increased dual regularization, since $\mu^{\text{upd}} \leq \mu$, and a larger dual scaling, since $\tau_y^{\text{upd}} \geq \tau_y$. This highlights once again that, in some regards, the penalty term corresponds to, and can be compensated by, a faster dual flow. Furthermore, it also shows that $\mathcal{DT}^{\text{upd}}$ stems from a projected backward Euler solution of the flow (6), since $\mathcal{DT}(\tau_x, \tau_y^{\text{upd}}, 0, \Delta t)$ originates in $\mathcal{CT}(\tau_x, \tau_y^{\text{upd}}, 0)$. Still, the relationship between the transformed discrete-time system and the associated continuous-time one remains ambiguous, for τ_y^{upd} depends on the stepsize Δt. Nonetheless, for a given stepsize, the evolution of $\mathcal{CT}(\tau_x, \tau_y^{\text{upd}}, 0)$ is unequivocal. We notice, however, that it can be unstable even if the discrete-time counterpart is stable, as in Example III.1; see Fig. 3 at $\rho = 0$ and Fig. 4.

Fig. 5 depicts primal-dual trajectories of continuous- and discrete-time systems for Example III.1, with penalty parameter $\rho = 4$. The evolution of $\mathcal{DT}$ approaches $\mathcal{CT}$'s and converges slowly to the solution $(0, 0)$. In contrast, $\mathcal{DT}^{\text{upd}}$ is not bound to approximate $\mathcal{CT}$'s trajectory, and attains faster convergence toward the solution. Although it yields more

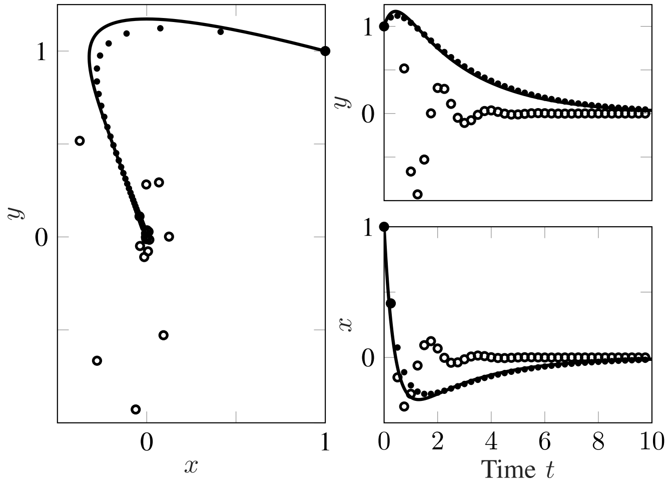

Fig. 5. Primal-dual trajectories for Example III.1, starting from $x = 1$ and $y = 1$, with $\rho = 4$ and stepsize $\Delta t = 0.25$. Continuous-time (line), discrete-time (dot), and discrete-time with dual update (circle).

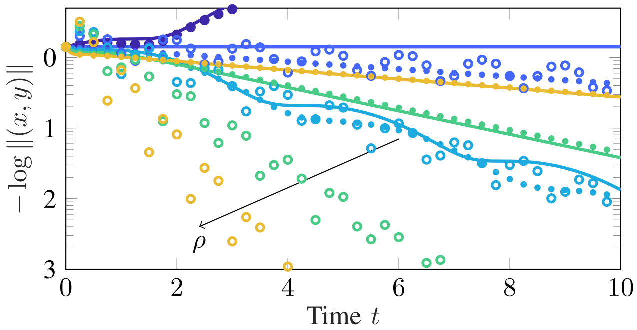

Fig. 6. Error trajectories for Example III.1, starting from $x = 1$ and $y = 1$, with stepsize $\Delta t = 0.25$. Continuous-time (line), discrete-time (dot), and discrete-time with dual update (circle). For $\rho = 0$, the trajectories diverge. For larger values of ρ, the discrete-time system with dual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

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 III.1, 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 AL flow. Conversely, $\mathcal{DT}^{\text{upd}}$ exhibits an essentially different behaviour: the higher the penalty parameter ρ, the faster the (linear) convergence toward the solution. In agreement with Fig. 2, this highlights that (suitable) multiplier estimate updates play a key role in speeding up convergence, even with stronger penalizations.

VI. 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 [10], NCL.jl [11] and Percival.jl [12] in Julia. The code to generate the numerical results is available online [13]. The tests were carried out on a desktop running Ubuntu 16.04 with Intel Core i7-8700 3.20 GHz and 16 GB RAM.

A. Algorithm

Algorithm 1 provides pseudocode for a prototypical Homotopy AL method for equality-constrained Optimization (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 a merit function $\mathcal{M}$, corresponding to the objective function in (10) with $\rho = \sigma = 0$. As such, convergence guarantees are inherited from the primal-dual AL framework [5]. The dual regularization parameter μ is adapted based on the decrease in constraint violation. Each subproblem is solved via a globalised Newton's method, corresponding to the inner loop, before updating the estimate $(\hat{x}, \hat{y})$. Bounds and inequality constraints in (1) can be easily accounted for by considering a projected or semismooth variant of Newton's method. Each subproblem yields the KKT conditions

$$0 = r(x, y) := \begin{pmatrix} \nabla f(x) + \nabla c(x)^\top y, \\ c(x) + \mu(\hat{y} - y) \end{pmatrix} \tag{11}$$

and regularized linear systems of the form

$$\begin{bmatrix} H & \nabla c(x)^\top \\ \nabla c(x) & -\mu I \end{bmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = -r(x, y), \tag{12}$$

where H is equal to, or a symmetric approximation for, the Hessian of the Lagrangian function $\mathcal{L}_0$ at (x, y). Then, given the search direction $d = (d_x, d_y)$, we perform a backtracking line search with Armijo's sufficient decrease condition:

$$\mathcal{M}(x + \alpha d_x, y + \alpha d_y) \leq \mathcal{M}(x, y) + \eta \alpha d^\top \nabla \mathcal{M}(x, y). \tag{13}$$

If $H \succ 0$ and $\mu > 0$, then d is a descent direction for $\mathcal{M}$ at (x, y), namely $d^\top \nabla \mathcal{M}(x, y) < 0$ for every $d \neq 0$, and the line search procedure is well-defined; furthermore, the conditions $d = 0$, $r(x, y) = 0$, and $\nabla \mathcal{M}(x, y) = 0$ are equivalent. The primal proximal term in (10) contributes with σI to the Hessian regularization, similarly to the strategy in [10, Eq. 13]: this gives a principled way to select a positive definite approximation H, possibly speeding up the convergence if the primal regularization is not necessary.

B. Setup

We consider the equality-constrained CUTEst problems with $1 \leq n, m \leq 100$; this yields 161 problems [14]. We set the tolerance $\varepsilon \in \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ and limit the number of iterations to 3000. In Percival, IPOPT and NCL, we leave all the other settings to the internal defaults. In HALeqO, we initialize $\mu = 10^{-3}$, set $\kappa_c = \kappa_\alpha = 0.5$, $\kappa_\mu = 0.1$, $\eta = 10^{-4}$, and use PositiveFactorizations.jl [15] to obtain a positive definite approximation H to the Hessian.

We compare the solvers by means of success rates and performance profiles [16], yet aware of their limitations [17]. 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

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

```
Initialize $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\varepsilon > 0$
Select $\mu > 0$, $\kappa_c, \kappa_\alpha, \kappa_\mu \in (0,1)$, $\eta \in (0, 1/2)$
Set $(\hat{x}, \hat{y}) \leftarrow (x, y)$
while $(x, y)$ does not solve (2) to within $\varepsilon$ do
    if $(x, y)$ solves (11) to within $\varepsilon$ then
        if $\|c(x)\| > \kappa_c \|c(\hat{x})\|$ then
            Set $\mu \leftarrow \kappa_\mu \mu$                dual regularization update
        end if
        Set $(\hat{x}, \hat{y}) \leftarrow (x, y)$              estimate update
    end if
    Select $H \succ 0$                              Hessian regularization
    Compute $(d_x, d_y)$ by solving (12)            search direction
    Set $\alpha \leftarrow 1$
    while (13) does not hold do                     line search
        Set $\alpha \leftarrow \kappa_\alpha \alpha$
    end while
    Set $(x, y) \leftarrow (x, y) + \alpha(d_x, d_y)$
end while
return $(x, y)$
```

TABLE I

COMPARISON WITH NUMBER OF PROBLEMS SOLVED, OUT OF 161.

ε	10^{-3}	10^{-4}	10^{-5}	10^{-6}
HALeqO	109	106	101	97
Percival	107	104	101	97
NCL	108	106	104	101
IPOPT	87	87	87	87

return a solution for problem $p \in P$. We plot the functions $\pi_s : \mathbb{R} \to [0,1]$, $s \in S$, defined by $\pi_s(r) := |\{p \in P : t_{s,p} \leq r t_p^{\min}\}|/|P|$, $t_p^{\min} := \min_{s \in S} t_{s,p}$. Considering $t_{s,p} = +\infty$ when solver s fails on problem p, $\pi_s(r)$ is the fraction of problems solved by solver s within r times the best timing.

C. Results

Computational results are summarized as success rates in Tab. I and performance profiles in Fig. 7. NCL shows the best reliability, with HALeqO slightly behind, followed by Percival. IPOPT lags behind as it does not handle problems with more constraints than variables. Furthermore, HALeqO exhibits the fastest execution on many problems and consistently outperforms the other solvers, as evidenced by Fig. 7. Thus, despite its simplicity, HALeqO strikes a promising balance between speed and robustness.

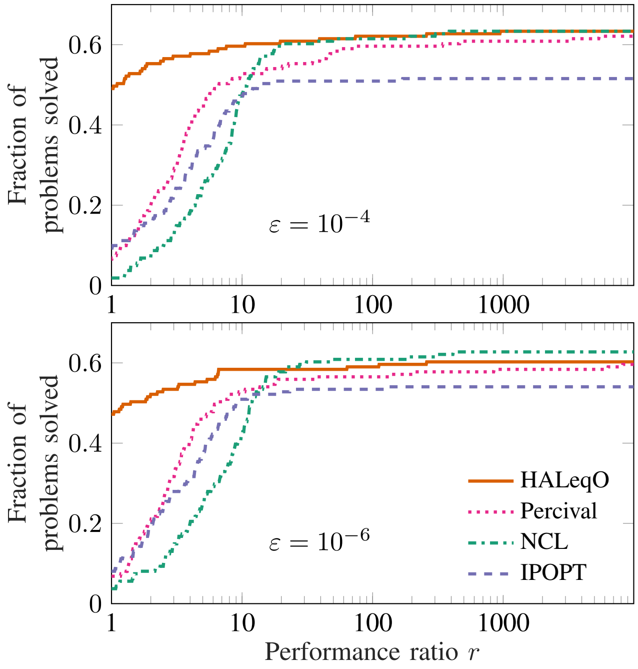

Fig. 7. Comparison with performance profiles of run times.

REFERENCES

[1] J. Schropp and I. Singer, "A dynamical systems approach to constrained minimization," *Numerical Functional Analysis and Optimization*, vol. 21, no. 3-4, pp. 537–551, 2000.

[2] A. Potschka and H. G. Bock, "A sequential homotopy method for mathematical programming problems," *Mathematical Programming*, vol. 187, no. 1, pp. 459–486, 3 2021.

[3] A. R. Conn, N. I. M. Gould, and P. L. Toint, "A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds," *SIAM Journal on Numerical Analysis*, vol. 28, no. 2, pp. 545–572, 4 1991.

[4] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 4 2014.

[5] D. P. Robinson, "Primal-dual methods for nonlinear optimization," Ph.D. dissertation, University of California, San Diego, 9 2007.

[6] D. Ma, K. L. Judd, D. Orban, and M. A. Saunders, "Stabilized optimization via an NCL algorithm," in *Numerical Analysis and Optimization*, M. Al-Baali, L. Grandinetti, and A. Purnama, Eds. Cham: Springer International Publishing, 2018, pp. 173–191.

[7] P. Armand and R. Omheni, "A globally and quadratically convergent primal–dual augmented Lagrangian algorithm for equality constrained optimization," *Optimization Methods and Software*, vol. 32, no. 1, pp. 1–21, 2017.

[8] S. Arreckx and D. Orban, "A regularized factorization-free method for equality-constrained optimization," *SIAM Journal on Optimization*, vol. 28, no. 2, pp. 1613–1639, 2018.

[9] R. T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of operations research*, vol. 1, no. 2, pp. 97–116, 5 1976.

[10] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 3 2006.

[11] D. Ma, D. Orban, and M. Saunders, "A Julia implementation of Algorithm NCL for constrained optimization," 1 2021.

[12] E. A. d. Santos and A. S. Siqueira, "Percival.jl: an augmented Lagrangian method," 7 2020. [Online]. Available: https://github.com/JuliaSmoothOptimizers/Percival.jl

[13] A. De Marchi, "HALeqO.jl," 3 2021.

[14] N. I. M. Gould, D. Orban, and P. L. Toint, "CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization," *Computational Optimization and Applications*, vol. 60, no. 3, pp. 545–557, 4 2015.

[15] T. Holy and contributors, "PositiveFactorizations.jl: Positive-definite "approximations" to matrices," 11 2020. [Online]. Available: https://github.com/timholy/PositiveFactorizations.jl

[16] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.

[17] N. Gould and J. Scott, "A note on performance profiles for benchmarking software," *ACM Trans. Math. Softw.*, vol. 43, no. 2, 9 2016.