A feedback control approach to convex optimization with inequality constraints

# A feedback control approach to convex optimization with inequality constraints

V. Cerone, S. M. Fosson, S. Pirrera, D. Regruto*

***Abstract*— We propose a novel continuous-time algorithm for inequality-constrained convex optimization inspired by proportional-integral control. Unlike the popular primal-dual gradient dynamics, our method includes a proportional term to control the primal variable through the Lagrange multipliers. This approach has both theoretical and practical advantages. On the one hand, it simplifies the proof of the exponential convergence in the case of smooth, strongly convex problems, with a more straightforward assessment of the convergence rate concerning prior literature. On the other hand, through several examples, we show that the proposed algorithm converges faster than primal-dual gradient dynamics. This paper aims to illustrate these points by thoroughly analyzing the algorithm convergence and discussing some numerical simulations.**

## I. Introduction

Primal-dual gradient dynamics (PDGD) is a well-established continuous-time algorithm that solves constrained optimization problems. Introduced in [1], [2], it consists of a primal-descent, dual-ascent gradient method achieving the saddle point of the Lagrangian of the problem.

In the last years, we have witnessed a renewed interest in PDGD thanks to its effectiveness in several engineering applications and control problems, e.g., game theory [3], power systems [4], [5] and model predictive control [6]. A particular focus is on its use in distributed optimization; see, e.g., [7], [3], [8], [9]. Gradient-based algorithms are notably well-suited for implementation over decentralized networks.

In the recent literature, several works have addressed the study of the stability and convergence of PDGD. This algorithm is globally exponentially convergent for smooth, strongly convex problems (see, e.g., [10], [11], [12]) and for problems that combine strongly and non-strongly convex terms in [13], [14]. In [15] and [16], the authors study the asymptotic convergence for general saddle functions not directly related to constrained optimization. In [17] and [18], the analysis envisages also nonsmooth composite optimization problems. Among the mentioned works, [11] and [14] consider equality-constrained problems, while [12] and [10] also consider inequality constraints.

This paper proposes a novel continuous-time algorithm for smooth, strongly convex problems with inequality constraints. By starting from the definition of a suitable augmented Lagrangian, the key idea is to control the dynamics of the primal variable through the Lagrange multipliers of the problem by implementing a feedback control method inspired by proportional-integral (PI) control. The contribution of the paper is twofold. On the one hand, we prove the exponential convergence of the proposed method for strongly convex functions. On the other hand, we show its practical effectiveness through numerical simulations. In particular, we analyze its behavior when compared to PDGD.

This work partially extends the framework proposed in [19], where we develop a feedback control approach for equality-constrained problems, specializing in PI control and feedback linearization. In this paper, we retrieve the key ideas of the PI control algorithm proposed in [19], and we develop a novel PI approach for the case of inequality constraints. In particular, this extension requires a novel convergence analysis starting from a peculiar augmented Lagrangian.

We organize the paper as follows. Section II states the problem and reviews its solution through PDGD. In Section III, we develop the proposed algorithm while we study its convergence in IV. Section V shows the effectiveness of the proposed method through numerical experiments, with particular attention to the convergence speed. Finally, Section VI concludes the paper.

## II. Problem statement and related work

Let $f:\mathbb{R}^n\to\mathbb{R}$ be a smooth, strongly convex function. We consider the constrained optimization problem

$$\begin{aligned}&\min_{x\in\mathbb{R}^n} f(x)\\ &\quad\text{s.t.}\\ &\quad h(x)=Cx-d\leq 0\end{aligned}\tag{1}$$

where $C\in\mathbb{R}^{m,n}$, $d\in\mathbb{R}^m$ and "$\leq$" denotes the componentwise inequality. As in [10], we consider the following augmented Lagrangian

$$\mathcal{L}(x,\lambda)=f(x)+g(x,\lambda)\tag{2}$$

where we define $g:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ as

$$\begin{aligned}&g(x,\lambda)=\sum_{j=1}^m g_j(x,\lambda_j)\quad\text{with}\\ &g_j(x,\lambda_j)=\\ &=\begin{cases}\lambda_j h_j(x)+\dfrac{\rho}{2}h_j^2(x) & \text{if } h_j(x)\geq-\dfrac{\lambda_j}{\rho};\\ -\dfrac{1}{2\rho}\lambda_j^2 & \text{otherwise}\end{cases}\end{aligned}\tag{3}$$

where $\rho>0$ is a design hyperparameter.

The function $g_j(x,\lambda_j):\mathbb{R}^{n+1}\to\mathbb{R}$ penalizes the constraint violation. We notice that it is continuous and has

* Corresponding author. The authors are with the Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; e-mail: diego.regruto@polito.it.

a continuous gradient for each $(x, \lambda_j) \in \mathbb{R}^{n+1}$. As shown in [10, Eq. (9)], PDGD for problem (2)-(3) corresponds to the dynamic system

$$\begin{aligned} \dot{x} &= -\nabla_x \mathcal{L}(x, \lambda) = -\nabla f(x) - \nabla_x g(x, \lambda) \\ \dot{\lambda} &= \eta \nabla_\lambda g(x, \lambda) \end{aligned} \tag{4}$$

for some $\eta > 0$. Let $J_h(x) \in \mathbb{R}^{m,n}$ be the Jacobian matrix of $h$, i.e., $J_h(x) = C$ in the linear case. Then,

$$\begin{aligned} \nabla_x g_j(x, \lambda_j) &= \max\{\rho h_j(x) + \lambda_j, 0\}(J_h(x))_j^\top \\ \frac{\partial g_j}{\partial \lambda_j}(x, \lambda_j) &= \frac{\max\{\rho h_j(x) + \lambda_j, 0\} - \lambda_j}{\rho}. \end{aligned} \tag{5}$$

PDGD in (4) is a system switching between two modes for each $j$. In the first mode, i.e., when $h_j(x) \geq -\frac{\lambda_j}{\rho}$, the Lagrange multipliers $\lambda(t)$ *control* the dynamics of the state $x(t)$, in order to push the system towards the feasible set $h(x) \leq 0$. In the second mode, i.e., when the constraint is satisfied, $x(t)$ is not controlled and evolves based on the gradient of $f$; contextually, $\lambda(t)$ converges to zero.

We remark that (2) is not the standard Lagrangian for inequality-constrained problems used, e.g., in [7], [12]. As noticed in [10], the standard Lagrangian gives rise to a PDGD with a discontinuous projection step, which creates numerical issues when implementing the algorithm. Moreover, using the standard Lagrangian makes the convergence analysis more challenging, and only stability is proven; the authors of [10] conjecture that this discontinuous PDGD is not globally exponentially stable.

As noticed in [10], the saddle point of (2)-(3) corresponds to the saddle point of the standard Lagrangian; see, e.g., [20, Chapter 3] for details.

The first-order Karush-Kuhn-Tucker (KKT) conditions for problem (1) are as follows, see, e.g., [21, Section 11.8]:

*Theorem 1 (KKT first-order conditions):* Let $x^\star \in \mathbb{R}^n$ be a minimum of $f$ subject to the constraints $h(x) \leq 0$. Then, there exists a unique $\lambda^\star \in \mathbb{R}^m$ such that

$$\begin{aligned} &\nabla f(x^\star) + J_h(x^\star)^\top \lambda^\star = 0 \\ &\lambda^{\star\top} h(x^\star) = 0 \\ &\lambda^\star \geq 0. \end{aligned} \tag{6}$$

Proposition 1 in [10] states that PDGD has a unique equilibrium point $(x^\star, \lambda^\star)$, and it satisfies the first-order KKT conditions for problem (1). Furthermore, Theorem 2 in [10] proves that PDGD is globally exponentially convergent. The computed convergence rate $\tau_{ineq}$ depends on several constants arising from repeated upper/lower bounding of the eigenvalues of the matrices involved in the proof; therefore, it is difficult to explicitly assess $\tau_{ineq}$ from the formula given in Theorem 2 in [10]. The proof is rather technical because it needs a non-diagonal quadratic Lyapunov function; see [10, Eq. (10)].

As to strongly convex problems with $h(x) = 0$, in [19], we show that PDGD corresponds to an integral control system that regulates $h(x)$ to zero, based on standard Lagrangian for equality constraints problem, i.e., $f(x) + g_{eq}(x, \lambda)$ with $g_{eq}(x, \lambda) = \lambda^\top h(x)$. In this system, the Lagrange multipliers $\lambda$ play the role of the control input. In [19], we design a PI control, which usually has a faster convergence rate than PDGD. This PI control system reads as follows:

$$\begin{aligned} \dot{x} &= -\nabla_x \mathcal{L}(x, \lambda) = -\nabla f(x) - \nabla_x g_{eq}(x, \lambda) \\ \dot{\lambda} &= K_i \nabla_\lambda g_{eq}(x, \lambda) + K_p \frac{d}{dt} \nabla_\lambda g_{eq}(x, \lambda) \end{aligned} \tag{7}$$

where $K_i$ and $K_p$ are design hyperparameters. Since $\nabla_\lambda g_{eq}(x, \lambda) = h(x)$, this a PI control regulating $h(x)$ to zero.

## III. Proposed approach

This section proposes a novel PI control approach to solve (1). A natural extension of (7) to the case of inequality constraints would be the application of (7) to the augmented Lagrangian (2) with (3), i.e.,

$$\begin{aligned} \dot{x} &= -\nabla_x \mathcal{L}(x, \lambda) = -\nabla f(x) - \nabla_x g(x, \lambda) \\ \dot{\lambda} &= K_i \nabla_\lambda g(x, \lambda) + K_p \frac{d}{dt} \nabla_\lambda g(x, \lambda) \end{aligned} \tag{8}$$

We can interpret the dynamic system (8) as a PI controller that pushes $h_j(x)$ towards zero whenever $h_j(x) \geq 0$ and provides no control action when $h_j(x) \leq 0$, letting $\lambda_j$ converge to zero.

Even if (8) is a natural extension of (7), proving its exponential convergence is challenging. In particular, we notice that $\frac{d}{dt} \nabla_\lambda g_j(x, \lambda_j)$ is discontinuous, which also means the right-hand side of the differential equation describing the closed-loop system is discontinuous and, in turn, potentially, the solution may be non-unique. For this motivation, we modify (8) as follows:

$$\begin{aligned} \dot{x} &= -\nabla_x \mathcal{L}(x, \lambda) = -\nabla f(x) - \nabla_x g(x, \lambda) \\ \dot{\lambda} &= K_i \nabla_\lambda g(x, \lambda) + K_p J_h(x) \dot{x} \end{aligned} \tag{9}$$

where the involved gradients are explicitly computed in (5). We replace $\frac{d}{dt} \nabla_\lambda g(x, \lambda)$ by $J_h(x)\dot{x}$, which we can interpret as taking a continuous approximation of (8). In (8), $\dot{\lambda}$ does not depend on $\dot{x}$ when the constraints are satisfied. In contrast, in (9), the dependence on $\dot{x}$ is always present.

As observed for (4), PI in (9) is a system that switches between two modes for each $j$. In the first mode, i.e., when $h_j(x) \geq -\frac{\lambda_j}{\rho}$, the Lagrange multipliers $\lambda(t)$ *control* the dynamics of the state $x(t)$ to achieve $h(x) \leq 0$. In the second mode, $\lambda(t)$ does not control $x(t)$. On the other hand, $\lambda(t)$ still depends on $\dot{x}(t)$.

The difference between the proposed approach and PDGD in (4) and (9) is the presence of the additional term $K_p J_h(x)\dot{x}$ in the dynamics of $\lambda$. To understand the rationale of this term, we go through a feedback control interpretation, as introduced in [19] for the equality-constrained case. By extending this feedback control framework to the case $h(x) \leq 0$, we can interpret PDGD as an algorithm with integral control on $\lambda_j$ representing a non-satisfied constraint. On the other hand, in the presence of a satisfied constraint, we do not control the system through $\lambda$.

In (9), we modify the dynamics of $\lambda$ by adding $K_p J_h(x)\dot{x}$. In the following, we show the benefits of this adjustment in terms of convergence rate.

We remark that (9) is not the direct extension of the PI method for $h(x) = 0$ reported in (7). Although feasible, the use of (7) produces a non-causal system with switched dynamics, creating issues in the proof of convergence. In other words, through the term $K_p J_h(x)\dot{x}$, we control $\lambda$ via state feedback even when the constraints are satisfied. This modification enhances convergence and simplifies analysis, as shown in the following sections.

## IV. Convergence analysis

In this section, we analyse the convergence of the dynamic system (9). We define

$$z(t) := (x(t)^\top, \lambda(t)^\top)^\top \tag{10}$$

and

$$z^\star := (x^{\star\top}, \lambda^{\star\top})^\top \tag{11}$$

is the equilibrium point of (9), which corresponds to a saddle point of $\mathcal{L}(x, \lambda)$. The following result holds.

*Proposition 1:* The equilibrium point of (9) satisfies the KKT conditions (6) for problem (1).

*Proof:* Since at the equilibrium point the time derivatives are null, i.e., $\dot{x^\star} = \dot{\lambda^\star} = 0$, we have $\nabla_\lambda g(x^\star, \lambda^\star) = 0$, i.e., for each $j = 1, \ldots, m$,

$$\max\{\rho h_j(x^\star) + \lambda_j^\star, 0\} = \lambda_j^\star,$$

which implies $\lambda_j^\star \geq 0$, $h_j(x^\star) \leq 0$ and $(h_j(x^\star))\lambda_j^\star = 0$. Finally,

$$\begin{aligned} \nabla_x g(x^\star, \lambda^\star) &= \sum_{j=1}^{m} \max\{\rho h_j(x^\star) + \lambda_j^\star, 0\} C_j^\top \\ &= \sum_{j=1}^{m} \lambda_j^\star C_j^\top = C^\top \lambda^\star = J_h(x)^\top \lambda^\star \end{aligned} \tag{12}$$

Therefore, $\dot{x^\star} = -\nabla f(x^\star) - \nabla_x g(x^\star, \lambda^\star) = -\nabla f(x^\star) + J_h(x)^\top \lambda^\star = 0$. ■

As a consequence of the strong convexity of $f(x)$, from [10, Lemma 1], there exists a symmetric, positive definite $B = B(x) \in \mathbb{R}^{n,n}$ such that

$$\nabla f(x) - \nabla f(x^\star) = B(x - x^\star). \tag{13}$$

*Theorem 2 (Global exponential convergence):* Let us assume that $C$ is full rank and there exists $0 < \underline{c} \leq \overline{c}$ such that

$$\underline{c} I \preceq CC^\top \preceq \overline{c} I. \tag{14}$$

Let $\rho < \overline{c}^{-1}$. Then, there exist real positive constants $\alpha_1$ and $\alpha_2$ such that

$$\|x(t) - x^\star\|_2 \leq \alpha_1 e^{-\frac{1}{2}\mu t}, \quad \|\lambda(t) - \lambda^\star\|_2 \leq \alpha_2 e^{-\frac{1}{2}\mu t} \tag{15}$$

where

$$\mu \leq \min\left\{\frac{1}{2} K_p \underline{c}, \frac{2K_i \underline{g} - K_p \overline{g}^2}{K_i}\right\} \tag{16}$$

and $0 < \underline{g} \leq \overline{g}$ are assessed in the proof.

*Proof:* We define the candidate Lyapunov function

$$V\big(z(t)\big) = \big(z(t) - z^\star\big)^\top P\big(z(t) - z^\star\big) \tag{17}$$

where

$$P := \begin{pmatrix} \sigma I_n & 0 \\ 0 & I_m \end{pmatrix} \in \mathbb{R}^{m+n, m+n} \tag{18}$$

for some $\sigma > 0$. If

$$\dot{V}\big(z(t)\big) \leq -\mu V\big(z(t)\big) \tag{19}$$

then the theorem statement holds. Therefore, let us study the conditions that guarantee (19).

Let us consider the diagonal matrix $\Gamma = \Gamma(z) \in [0, 1]^{m,m}$ as defined in Lemma 3 in [10]. Since $\nabla_x \mathcal{L}(x^\star, \lambda^\star) = \nabla_\lambda \mathcal{L}(x^\star, \lambda^\star) = 0$, $J_h(x) = C$ and by using (13),

$$\begin{aligned} \dot{x} &= -\nabla_x \mathcal{L}(x, \lambda) = -\nabla_x \mathcal{L}(x, \lambda) + \nabla_x \mathcal{L}(x^\star, \lambda^\star) \\ &= -B(x - x^\star) - \rho C^\top \Gamma C(x - x^\star) - C^\top \Gamma(\lambda - \lambda^\star) \end{aligned} \tag{20}$$

as obtained for PDGD, see Sec. III-B in [10] for details. Furthermore,

$$\begin{aligned} \dot{\lambda} =& \nabla_\lambda \mathcal{L}(x, \lambda) = \nabla_\lambda \mathcal{L}(x, \lambda) - \nabla_\lambda \mathcal{L}(x^\star, \lambda^\star) \\ &= K_i \Gamma C(x - x^\star) + \frac{K_i}{\rho}(\Gamma - I)(\lambda - \lambda^\star) + K_p C\dot{x}. \end{aligned} \tag{21}$$

Equations (20)-(21) represent (9) in a “linear” form.

Let $G_1 := B + \rho C^\top \Gamma C$ and $G_2 := C^\top \Gamma$. Since $B$ is positive definite, $G_1$ is positive definite; let $\underline{g} I \preceq G_1^\top \preceq \overline{g} I$. Then, we can rewrite (20)-(21) in a matrix form

$$\dot{z}(t) = G\big(z(t) - z^\star\big) \tag{22}$$

where

$$G := \begin{pmatrix} -G_1 & -G_2 \\ K_i G_2^\top - K_p C G_1 & \frac{K_i}{\rho}(\Gamma - I) - K_p C G_2 \end{pmatrix} \tag{23}$$

Since

$$\begin{aligned} \dot{V}\big(z(t)\big) &= \dot{z}(t)^\top P\big(z(t) - z^\star\big) + \big(z(t) - z^\star\big)^\top P \dot{z}(t) \\ &= \big(z(t) - z^\star\big)^\top \big(G^\top P + PG\big)\big(z(t) - z^\star\big) \end{aligned} \tag{24}$$

a sufficient condition for $\dot{V}\big(z(t)\big) \leq -\mu V\big(z(t)\big)$, see (19), is

$$-G^\top P - PG - \mu P \succeq 0. \tag{25}$$

As a consequence, our next goal is to provide sufficient conditions for (25). We have

$$-G^\top P - PG - \mu P = \begin{pmatrix} Q_1 & Q_2 \\ Q_2^\top & Q_3 \end{pmatrix}. \tag{26}$$

where

$$\begin{aligned} Q_1 &= 2\sigma G_1 - \sigma\mu I_n \\ Q_2 &= (\sigma - K_i) G_2 + K_p G_1^\top C^\top \\ Q_3 &= K_p C G_2 + K_p G_2^\top C^\top + 2\frac{K_i}{\rho}(I - \Gamma) - \mu I_m \end{aligned} \tag{27}$$

If $K_i \geq K_p$, by applying [10, Lemma 6] for $\frac{1}{\rho} > \overline{c}$,

$$K_p C G_2 + K_p G_2^\top C^\top + 2\frac{K_i}{\rho}(I - \Gamma) \succeq \frac{3}{2} K_p C C^\top. \quad (28)$$

Thus,

$$Q_3 \succeq \frac{3}{2} K_p C C^\top - \mu I_m \succeq K_p C C^\top \quad (29)$$

where the last step is a consequence of the assumption $\mu \leq \frac{1}{2} K_p \underline{c}$.

To simplify the computations, we set $K_i = \sigma$. Then,

$$Q_2 = K_p G_1^\top C^\top. \quad (30)$$

In conclusion,

$$-G^\top P - PG - \mu P \succeq \begin{pmatrix} 2\sigma G_1 - \sigma\mu I_n & K_p G_1^\top C^\top \\ \left[K_p G_1^\top C^\top\right]^\top & K_p C C^\top \end{pmatrix}. \quad (31)$$

Since $CC^\top \succ 0$ is invertible from (14), we can apply the Schur complement argument: the matrix in (31) is positive semidefinite if and only if

$$2\sigma G_1 - \sigma\mu I_n - K_p G_1^\top C^\top \left(K_p C C^\top\right)^{-1} K_p C G_1 \succeq 0. \quad (32)$$

Since $CC^\top$ is invertible, then $C^\top (CC^\top)^{-1} C \preceq I$. Therefore, a sufficient condition for (32) is

$$2\sigma G_1 - \sigma\mu I_n - K_p G_1^\top G_1 \succeq 0. \quad (33)$$

Finally, (33) holds if

$$2\sigma \underline{g} - \sigma\mu - K_p \overline{g}^2 \geq 0 \quad (34)$$

which is equivalent to

$$\mu \leq \frac{2K_i \underline{g} - K_p \overline{g}^2}{K_i}, \quad (35)$$

which completes the proof. ■

*Remark 1:* In the proof, we set $K_i = \sigma$ to simplify the computations. Other choices may enhance the convergence rate.

*Remark 2:* A theoretical comparison of the convergence rates $\mu$ and $\tau_{ineq}$ in [10] is challenging because $\tau_{ineq}$ depends on many constants that are not easy to assess. Conversely, we can estimate $\mu$ straightforwardly, given some knowledge of the given optimization problem.

*Remark 3:* The proof of Theorem 2 of this manuscript is more straightforward than the proof of Theorem 2 in [10] because the diagonal form of the Lyapunov function (19) reduces the computations when compared to the Lyapunov function with cross terms in [10].

Theorem 2 also suggests some insights on the selection of $K_i, K_p$. We notice that the ratio $K_p$ must be kept small to avoid reducing the convergence rate, while $K_i$ plays the same role of $\eta$ in (4).

### A. Example on the convergence rate of PI and PDGD

To conclude this section, we report an example to compare the convergence rates of PDGD in (4) and PI in (9).

We consider the quadratic optimization problem

$$\begin{aligned} &\min_{x \in \mathbb{R}} \frac{1}{2} w x^2 \\ &\text{s.t.} \\ &x \leq 0 \end{aligned} \quad (36)$$

where $w > 0$.

Since the cost function is quadratic and the constraints are linear, the closed-loop dynamics with the PI control corresponds to a switched linear time-invariant system with the following two modes:

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \mathrm{A}_1 \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad \begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \mathrm{A}_2 \begin{pmatrix} x \\ \lambda \end{pmatrix} \quad (37)$$

where

$$\begin{aligned} \mathrm{A}_1 &= \begin{pmatrix} -w - \rho & -1 \\ K_i - K_p(w+\rho) & -K_p \end{pmatrix} \\ \mathrm{A}_2 &= \begin{pmatrix} -w & 0 \\ -wK_p & -\frac{K_i}{\rho} \end{pmatrix}. \end{aligned} \quad (38)$$

In the first mode, the control through $\lambda$ is active; in the second mode, $x(t)$ satisfies the constraints and evolves according to the derivative of the cost function.

Similarly, for PDGD, the dynamics correspond to a switched linear time-invariant system of kind (37) with two modes described by

$$\tilde{\mathrm{A}}_1 = \begin{pmatrix} -w - \rho & -1 \\ \eta & 0 \end{pmatrix}, \quad \tilde{\mathrm{A}}_2 = \begin{pmatrix} -w & 0 \\ 0 & -\frac{\eta}{\rho} \end{pmatrix} \quad (39)$$

which correspond to $\mathrm{A}_1$ and $\mathrm{A}_2$ with $K_p = 0$ and $K_i = \eta$. Now, we notice that $\mathrm{A}_2$ and $\tilde{\mathrm{A}}_2$ have the same eigenvalues for $K_i = \eta$, i.e., $-w$ and $-\frac{K_i}{\rho}$. Therefore, PI and PDGD enjoy the same convergence rate in the second mode. Concerning the first mode, the eigenvalues of $\mathrm{A}_1$ are

$$\frac{-K_p - w - \rho \pm \sqrt{(K_p + w + \rho)^2 - 4K_i}}{2}. \quad (40)$$

Therefore, if $K_i$ is sufficiently large, the eigenvalues are complex conjugate, and the convergence rate is $K_p + w + \rho$.

On the other hand, the eigenvalues of $\tilde{\mathrm{A}}_1$ are

$$\frac{-w - \rho \pm \sqrt{(w+\rho)^2 - 4\eta}}{2} \quad (41)$$

with the best convergence rate equal to $w + \rho$. In conclusion, PI has a better convergence rate than PDGD, provided a suitable $K_p$ is selected. In particular, for PDGD, increasing the convergence rate beyond $w + \rho$ is impossible.

We remark that although it is always possible to increase $\rho$ both in (9) and (4), this may cause numerical issues during the integration of the differential equations.

## V. Numerical results

In this section, we illustrate two numerical simulations to support the effectiveness of the proposed algorithm (9) compared to PDGD.

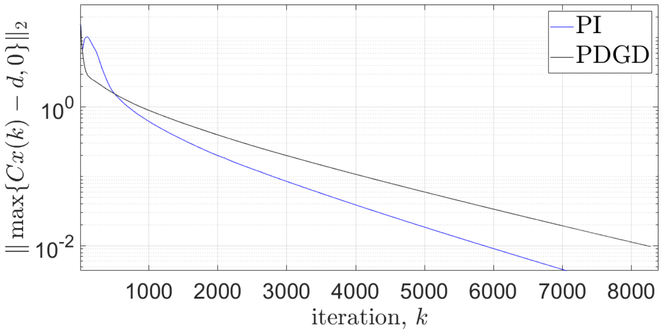


Fig. 1: Example 1: constraints violation.

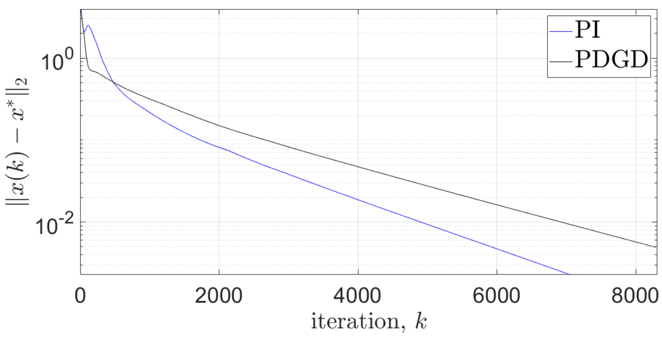


Fig. 2: Example 1: distance from the optimum.

### A. Example 1: Quadratic programming

In this simulation, we consider a strongly convex quadratic programming (QP) in the following form:

$$\begin{aligned} x^\star = \arg\min_{x\in\mathbb{R}^n} & \frac{1}{2}x^\top\left(I+W^\top W\right)x + b^\top x \\ \text{s.t.} & \\ & Cx - d \leq 0 \end{aligned} \tag{42}$$

where $W \in \mathbb{R}^{n,n}, b \in \mathbb{R}^n, C \in \mathbb{R}^{m,n}, d \in \mathbb{R}^m$ are randomly generated vectors or matrices, with independent, normally distributed components. We set $n = 50$ and $m = 45$.

We solve the optimization problem using the proposed algorithm (9) and PDGD. We integrate the differential equations (4) and (9) in the time interval $[0, 30]$ s through the *ode45* MATLAB command to select the discretization step size optimally. We set $K_i = \eta = 1$ and $K_p = -0.7$.

Fig. 1 shows the time evolution of $\|\max Cx(k) - d, 0\|_2$, where $k$ is the current iteration. This metric represents the violation of the constraints and is equal to zero when the state $x$ satisfies the constraints.

Fig. 2 shows the $\ell_2$ distance from the global optimum $x^\star$, computed through the MATLAB package CVX [22].

We perform 100 random runs with different realizations of $W, b, C, d$. PI requires fewer iterations than PDGD in all the runs and, consequently, less computational time. In Table I, we report some statistics that show the enhanced convergence speed of PI with respect to PDGD.

Fig. 1 and Fig. 2, obtained from one randomly selected run, show us that the proposed approach converges more quickly than PDGD, either in terms of fulfilment of the constraints and achievement of the minimum.

| | mean | standard deviation | worst case |
|---|---|---|---|
| $N$ PDGD | 8128.3 | 491.9 | 9361 |
| $N$ PI | 6903.2 | 312.1 | 7613 |
| $T$ PDGD | $1.23 \times 10^{-1}$ | $8.04 \times 10^{-3}$ | $1.53 \times 10^{-1}$ |
| $T$ PI | $1.02 \times 10^{-1}$ | $4.74 \times 10^{-3}$ | $1.15 \times 10^{-1}$ |

TABLE I: Example 1: results over 100 random runs. $N$ is the required numbers of iteration; $T$ the computational time (in seconds).

### B. Example 2: Linear system identification

We apply our approach to a problem of system identification. We consider the problem of identifying an unknown stable linear dynamic system $H(z)$ using uncertain input-output measurements $\{u_k, \tilde{y}_k\}$, for $k \in \{1, \ldots, N\}$, where $\tilde{y}_k = y_k + \eta_k$, $y_k$ is the $k$-th noise-free output sample and $\eta_k$ is the $k$-th sample of the noise sequence.

To perform the identification, we select a linear-in-the-parameters model structure $\tilde{H}(z, \theta)$ of the form

$$\tilde{H}(z) = \sum_{i=1}^{P} \theta_i B_i(z), \tag{43}$$

which is a standard choice in the context of system identification, and commonly considered choices of $B_i(z)$ are Laguerre or Kautz filters; see, e.g., [23],[24]. For the sake of simplicity, in this example, we select $B_i(z)$ to be first-order transfer functions with poles uniformly distributed in $[-0.9, 0.9]$. More precisely, we select

$$B_i(z) = \frac{z}{z - p_i}, \quad p_i \in \{-0.9, -0.85, \ldots, 0.9\}. \tag{44}$$

To generate time-domain data, we simulate the randomly selected system

$$H(z) = \frac{-0.4z^2 + 0.32z + 0.26}{z^3 - 1.9z^2 + 1.21z - 0.259} \tag{45}$$

excited by a random input uniformly distributed in $[0, 1]$. We corrupt the output data with normally distributed noise with variance $\sigma_\eta^2 = 0.1$.

We look for the value of the parameter $\theta$ that minimizes the $\ell_\infty$-norm of the simulation error

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^P} \|y_k(\theta) - \tilde{y}_k\|_\infty . \tag{46}$$

By adding a slack variable $\Delta \in \mathbb{R}$, we recast problem (46) to the following linear programming problem:

$$\begin{aligned} \theta^*, \Delta^* = \arg & \min_{\theta\in\mathbb{R}^P, \Delta\in\mathbb{R}} \Delta \\ & \text{s.t.} \\ & -\Delta \leq Z_k\theta - \tilde{y}_k \leq \Delta, \quad k \in \{1, \ldots, N\} \end{aligned} \tag{47}$$

where $Z_k = [z_1(k), \ldots, z_P(k)] \in \mathbb{R}^P$, $z_i(k) = B_i(q^{-1})u(k)$.

We solve the optimization problem using the proposed PI algorithm in (9) and PDGD in (4). We integrate the differential equations (4) and (9) in the time interval $[0, 1000]$ s

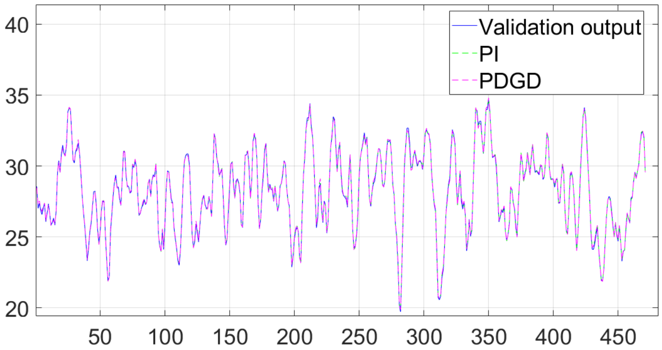


Fig. 3: Example 2: validation of identified models

| | $N$ | $T$ |
|---|---|---|
| PDGD | $10^6$ | 679.46 |
| PI | $5 \times 10^5$ | 233.09 |

TABLE II: Example 2: $N$ is the required numbers of iteration; $T$ the computational time (in seconds).

through the *ode23* MATLAB command to select the discretization step size optimally. We set $K_i = \nu = 1$ and $K_p = -0.5$.

In Fig. 3, we compare the outputs of the true model and the ones estimated using the PDGD and PI algorithms on data not used for identification. The outputs of the three models are almost exactly overlapped. We also evaluate the validation performances of the two algorithms in terms of the FIT index, defined as

$$FIT = 100\left(1 - \sqrt{\frac{\left\|y^{val} - \hat{y}^{val}\right\|_2^2}{\left\|y^{val} - m_y^{val}\right\|_2^2}}\right) \quad (48)$$

where $y^{val}$ is the true output, $m_y^{val} = \frac{1}{N}\sum_{k=1}^{N} y_k^{val}$ and $\hat{y}^{val}$ is the response of the identified model. We obtain the same value of FIT $= 98.5\%$ for both the identified models. Fig. 3 and the computed FIT values show that both algorithms converge to the optimal solution as expected from the theory. We report the comparison of the two algorithms in terms of computational effort in Table II, where we show the required number of iterations and computational time. Such results show that the PI algorithm is about two times faster than PDGD.

## VI. Conclusion

This paper proposes a novel continuous-time algorithm to solve smooth, strongly convex optimization problems with inequality constraints. As for the primal-dual gradient dynamics, the proposed algorithm consists of a dynamic system in the optimization variable and the Lagrange multipliers of the problem. By elaborating on the feedback PI control approach proposed for the equality-constrained case in [19], we develop a variant of primal-dual gradient dynamics in which an additional term adjusts the dynamics of the Lagrange multipliers and enhances the convergence speed. We prove that the proposed method is globally exponentially convergent. Finally, examples and numerical simulations show its effectiveness and velocity concerning the PDGD algorithm. Current work envisages the relaxation of affinity and smoothness requirements in the considered model.

## References


[1] T. Kose, "Solutions of saddle value problems by differential equations," *Econometrica*, vol. 24, no. 1, pp. 59–70, 1956.

[2] K. J. Arrow, L. Hurwicz, and H. B. Chenery, *Studies in linear and non-linear programming*. Stanford University Press, 1958.

[3] B. Gharesifard and J. Cortés, "Distributed convergence to nash equilibria in two-network zero-sum games," *Automatica*, vol. 49, no. 6, pp. 1683–1692, 2013.

[4] C. Zhao, U. Topcu, N. Li, and S. Low, "Design and stability of load-side primary frequency control in power systems," *IEEE Trans. Autom. Control*, vol. 59, no. 5, pp. 1177–1189, 2014.

[5] N. Li, C. Zhao, and L. Chen, "Connecting automatic generation control and economic dispatch from an optimization view," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 3, pp. 254–264, 2016.

[6] T. Skibik and M. M. Nicotra, "Analysis of time-distributed model predictive control when using a regularized primal-dual gradient optimizer," *IEEE Control Syst. Lett.*, vol. 7, pp. 235–240, 2023.

[7] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.

[8] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *J. Nonlin. Sci.*, vol. 29, pp. 1247–1272, 2019.

[9] X. Shi, X. Xu, J. Cao, and X. Yu, "Finite-time convergent primal-dual gradient dynamics with applications to distributed optimization," *IEEE Trans. Cybernetics*, vol. 53, no. 5, pp. 3240–3252, 2023.

[10] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 43–48, 2019.

[11] S. A. Alghunaim and A. H. Sayed, "Linear convergence of primal-dual gradient methods and their performance in distributed optimization," *Automatica*, vol. 117, p. 109003, 2020.

[12] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal-dual dynamics," *Systems & Control Letters*, vol. 87, pp. 10–15, 2016.

[13] S. S. Du and W. Hu, "Linear convergence of the primal-dual gradient method for convex-concave saddle point problems without strong convexity," in *Proc. Int. Conf. Artificial Intell. Stat. (AISTATS)*, vol. 89, 2019, pp. 196–205.

[14] X. Chen and N. Li, "Exponential stability of primal-dual gradient dynamics with non-strong convexity," in *Amer. Control Conf. (ACC)*, 2020, pp. 1612–1618.

[15] A. Cherukuri, B. Gharesifard, and J. Cortés, "Saddle-point dynamics: Conditions for asymptotic stability of saddle points," *SIAM J. Control Optim.*, vol. 55, no. 1, pp. 486–511, 2017.

[16] A. Cherukuri, E. Mallada, S. Low, and J. CortÃ©s, "The role of convexity in saddle-point dynamics: Lyapunov function and robustness," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2449–2464, 2018.

[17] D. Ding and M. R. Jovanović, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented lagrangian," in *Proc. Amer. Control Conf. (ACC)*, 2019, pp. 3414–3419.

[18] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented lagrangian method for nonsmooth composite optimization," *IEEE Trans. Autom. Control*, vol. 64, no. 7, pp. 2861–2868, 2019.

[19] V. Cerone, S. M. Fosson, S. Pirrera, and D. Regruto, "A new framework for constrained optimization via feedback control of lagrange multipliers," *submitted to IEEE Trans. Autom. Control*, 2024. [Online]. Available: https://arxiv.org/abs/2403.12738

[20] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, NY, USA., 1982.

[21] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 4th ed. Springer Publishing Company, Inc., 2015.

[22] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," https://cvxr.com/cvx, 2014.

[23] P. M. Van Den Hof, P. S. Heuberger, and J. Bokor, "System identification with generalized orthonormal basis functions," *Automatica*, vol. 31, no. 12, pp. 1821–1834, 1995.

[24] B. Wahlberg and P. Mäkilä, "On approximation of stable linear dynamical systems using laguerre and kautz functions," *Automatica*, vol. 32, no. 5, pp. 693–708, 1996.