Proximal Gradient Dynamics and Feedback Control for Equality-Constrained Composite Optimization

Veronica Centorrino* Francesca Rossi † Francesco Bullo‡ Giovanni Russo*

March 19, 2025

Abstract

This paper studies equality-constrained composite minimization problems. This class of problems, capturing regularization terms and convex inequality constraints, naturally arises in a wide range of engineering and machine learning applications. To tackle these minimization problems, we introduce the *proportional–integral proximal gradient dynamics* (PI–PGD): a closed-loop system where the Lagrange multipliers are control inputs and states are the problem decision variables. First, we establish the equivalence between the minima of the optimization problem and the equilibria of the PI–PGD. Then, leveraging tools from contraction theory, we give a comprehensive convergence analysis for the dynamics, showing linear–exponential convergence towards the equilibrium. That is, the distance between each solution and the equilibrium is upper bounded by a function that first decreases linearly and then exponentially. Our findings are illustrated numerically on a set of representative examples, which include an application to entropic-regularized optimal transport.

1 Introduction

We study equality-constrained non-smooth composite convex optimization problems (OPs), i.e., problems of the form

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x)+g(x) \\ \text{s.t.} \ & h(x)=0, \end{aligned} \tag{1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $h\colon \mathbb{R}^n \to \mathbb{R}^m$ are convex and differentiable, while $g\colon \mathbb{R}^n \to \mathbb{R}$ is convex, closed, and proper, possibly non-smooth. These problems are ubiquitous in many engineering, science, and machine learning applications as they capture regularization terms and convex inequality constraints. A common approach to solving (1) is to use projected dynamical systems or incorporate the equality constraint into the cost function and solve the resulting proximal gradient dynamics [14]. However, such approach can become challenging when the proximal operator lacks a closed form or the projection is difficult to compute.

Constrained optimization algorithms can also be interpreted as closed-loop systems, where the goal is to ensure convergence to the optimizer while enforcing feasibility. In this context, in [10] a continuous-time control-theoretic framework for equality-constrained smooth optimization has been proposed. The core idea (see Figure 1) is to consider as *plant* dynamics a system inspired by the gradient flow with respect to the primal variables of the Lagrangian. The output of this system is $y(t) = h(x(t))$ and the control variables are Lagrange multipliers. The control objective is then to design a feedback controller that drives the output to zero.

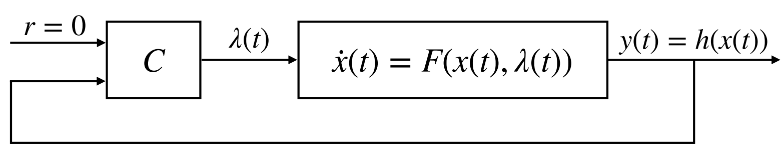

Figure 1: Closed-loop system for equality-constrained OPs: $\dot{x} = F(x, \lambda)$ has the stationary points of the Lagrangian as equilibria. The feedback controller, C, is designed to drive y toward the reference value $r = 0$, so that $h(x) = 0$.

*DIEM, University of Salerno, Italy {vcentorrino, giovarusso}@unisa.it. VC and GR supported by the European Union-Next Generation EU Mission 4 Component 1 CUP E53D23014640001.

†Scuola Superiore Meridionale, Italy. f.rossi@ssmeridionale.it.

‡Center for Control, Dynamical Systems, and Computation, UC Santa Barbara, CA, USA. bullo@ucsb.edu. FB supported in part by AFOSR grant FA9550-22-1-0059.

Inspired by the approach in [10], we propose the *proportional–integral controlled proximal gradient dynamics* (PI–PGD) to solve equality-constrained (non-smooth) composite OPs. For such dynamics, we provide a comprehensive analysis, characterizing convergence. Our findings are also illustrated via examples, which include an exploratory application to entropic-regularized optimal transport.

Literature review: Studying OPs as continuous-time dynamics is a classical problem dating back to [2], which has recently gained renewed interest thanks to developments from, e.g., online and dynamic feedback optimization [4] and the broader analysis of algorithms from a feedback control perspective [15]. The standard approach for constrained OPs is via standard and augmented primal-dual dynamics for smooth cost [20] and non-smooth composite OPs [14], respectively. The use of Lagrangian multipliers as feedback controllers has recently been proposed for smooth OPs with equality [10] and inequality [9] constraints, as well as for smooth constrained optimization via control barrier functions [1]. Global exponential convergence with PI controller is proved in [10, 9] for the standard case of full-rank linear constraints and strongly convex and strongly smooth cost. More broadly, there has been a growing interest in using strongly contracting dynamics to tackle OPs [16, 8, 12]. This is mainly due to the highly ordered transient and asymptotic behavior properties enjoyed by such dynamics [6]. The asymptotic behavior of weakly contracting dynamics has been characterized in, e.g., [11] for monotone systems, and [7] for convex OPs with a unique minimizer.

Contributions: We propose the PI–PGD for solving equality-constrained composite OPs (1). The PI–PGD is a closed-loop system, where: (i) the dynamics for the primal variables (i.e., the plant) has the stationary points of the Lagrangian of (1) as equilibria; (ii) the dual variables are control inputs and a PI controller is designed so that the closed-loop system converges to an equilibrium, which is a (feasible) minimum for (1). We prove the equivalence between the minima of the OP and the equilibria of the proposed PI–PGD. Then, we conduct a comprehensive convergence analysis for the widely considered case of linear-equality constraints and strongly convex and strongly smooth cost. As our main convergence result, we show that convergence towards the equilibrium is *linear-exponential*. That is, the distance between each solution of the dynamics and the equilibrium is upper bounded by a linear–exponential function [7]. To establish this convergence property, we leverage contraction theory and characterize global weak infinitesimal contractivity/non-expansiveness of the dynamics and local strong infinitesimal contractivity/exponential stability of the equilibrium.

Finally, we validate our results via numerical examples, including an exploratory application to entropic-regularized optimal transport. Notably, in the numerical study, our method converges even when the state-of-the-art Sinkorn algorithm fails. The code to replicate our numerical results is available at https://shorturl.at/mPpEZ.[1]

2 Mathematical Preliminaries

We denote by $\mathbb{0}_n$ and $\mathbb{1}_n \in \mathbb{R}^n$ the all-zeros and all-ones vector of size n, respectively. Vector inequalities of the form $x \leq (\geq) y$ are entrywise. We let I_n be the $n \times n$ identity matrix. The symbol $\otimes$ denotes the Kronecker product. Given $A, B \in \mathbb{R}^{n\times n}$ symmetric, we write $A \preceq B$ (resp. $A \prec B$) if $B - A$ is positive semidefinite (resp. definite). When A has only real eigenvalues, we let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be its minimun and maximum eigenvalue, respectively. We say that A is *Hurwitz* if $\alpha(A) := \max\{\mathrm{Re}(\lambda) \mid \lambda \text{ eigenvalue of } A\} < 0$, where $\mathrm{Re}(\lambda)$ denotes the real part of λ.

Norms, Logarithmic Norms and Weak Pairings We let $\|\cdot\|$ denote both a norm on $\mathbb{R}^n$ and its corresponding induced matrix norm on $\mathbb{R}^{n\times n}$. Given $x \in \mathbb{R}^n$ and $r > 0$, we let $B_p(x, r) := \{z \in \mathbb{R}^n \mid \|z - x\|_p \leq r\}$ be the *ball of radius r centered at x* computed with respect to the norm p.

Given $A \in \mathbb{R}^{n\times n}$ the *logarithmic norm* (lognorm) induced by $\|\cdot\|$ is

$$\mu(A) := \lim_{h\to 0^+} \frac{\|I_n + hA\| - 1}{h}.$$

Given a symmetric positive-definite matrix $P \in \mathbb{R}^{n\times n}$, we let $\|\cdot\|_P$ be the P-weighted ℓ_2 norm $\|x\|_P := \sqrt{x^\top P x}$, $x \in \mathbb{R}^n$ and write $\|\cdot\|_2$ if $P = I_n$. The corresponding lognorm is $\mu_P(A) = \min\{b \in \mathbb{R} \mid PA + A^\top P \preceq 2bP\}$ [6, Lemma 2.7].

Convex analysis and proximal operators Given a convex set $\mathcal{C}$, the function $\iota_{\mathcal{C}} \colon \mathbb{R}^n \to [0, +\infty]$ is the *zero-infinity indicator function on $\mathcal{C}$* and is defined by $\iota_{\mathcal{C}}(x) = 0$ if $x \in \mathcal{C}$ and $\iota_{\mathcal{C}}(x) = +\infty$ otherwise. The function $\mathrm{ReLU} \colon \mathbb{R} \to \mathbb{R}_{\geq 0}$, is defined by $\mathrm{ReLU}(x) = \max\{0, x\}$. A map $g \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is (i) *convex* if $\mathrm{epi}(g) := \{(x, y) \in \mathbb{R}^{n+1} \mid g(x) \leq y\}$ is a convex set; (ii) *proper* if its value is never $-\infty$ and there exists at least one $x \in \mathbb{R}^n$ such that $g(x) < \infty$; (iii) *closed* if it is proper and $\mathrm{epi}(g)$ is closed. A map $g \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is

(i) *strongly convex with parameter* $\rho > 0$ if the map $x \mapsto g(x) - \frac{\rho}{2}\|x\|_2^2$ is convex;

(ii) *strongly smooth with parameter* $L \geq 0$ if it is differentiable and ∇g is Lipschitz with constant L.

[1] See https://shorturl.at/mPpEZ for an extended technical report.

Next, we define the proximal operator of g, which is a map that takes a vector $x \in \mathbb{R}^n$ and maps it into a subset of $\mathbb{R}^n$, which can be either empty, contain a single element, or be a set with multiple vectors.

Definition 1 (Proximal Operator)**.** *The* proximal operator *of a function* $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ *with parameter* $\gamma > 0$*,* $\mathrm{prox}_{\gamma g}\colon \mathbb{R}^n \to \mathbb{R}^n$*, is the operator given by*

$$\mathrm{prox}_{\gamma g}(x) = \underset{z\in\mathbb{R}^n}{\arg\min}\, g(z) + \frac{1}{2\gamma}\|x - z\|_2^2, \quad \forall x \in \mathbb{R}^n. \tag{2}$$

The map $\mathrm{prox}_{\gamma g}$ is firmly nonexpansive [3, Proposition 12.28]. The subdifferential operator of g, ∂g, is linked to the proximal operator $\mathrm{prox}_{\gamma g}$ by the following relation

$$\mathrm{prox}_{\gamma g} = (I_n + \gamma \partial g)^{-1}. \tag{3}$$

The (point-to-point) map $(I_n + \gamma\partial g)^{-1}$ is called the *resolvent* of the operator ∂g with parameter $\gamma > 0$. That is, the proximal operator is the resolvent of the subdifferential operator. Moreover, when the function g is closed, convex, and proper, then the resolvent, and so the proximal map, is single-valued, even though ∂g is not.

2.1 Contraction Theory for Dynamical Systems

Consider a dynamical system

$$\dot{x}(t) = f\big(t, x(t)\big), \tag{4}$$

where $f\colon \mathbb{R}_{\geq 0} \times \mathcal{C} \to \mathbb{R}^n$, is a smooth nonlinear function with $\mathcal{C} \subseteq \mathbb{R}^n$ forward invariant set for the dynamics. We let $t \mapsto \phi_t\big(x_0\big)$ be the flow map of (4) at time t starting from initial condition $x(0) := x_0$. Then, we give the following [22, 6]:

Definition 2 (Contracting dynamics)**.** *Given a norm* $\|\cdot\|$ *with associated lognorm* μ*, a smooth function* $f\colon \mathbb{R}_{\geq 0}\times\mathcal{C} \to \mathbb{R}^n$*, with* $\mathcal{C} \subseteq \mathbb{R}^n$ *f-invariant, open and convex, and a* contraction rate $c > 0$ *($c = 0$),* f *is c-strongly (weakly) infinitesimally contracting on* $\mathcal{C}$ *if*

$$\mu\big(Df(t, x)\big) \leq -c, \tag{5}$$

for all $x \in \mathcal{C}$ and $t \in \mathbb{R}_{\geq 0}$*, where* $Df(t, x) := \partial f(t, x)/\partial x$*.*

If f is contracting, then for any two trajectories $x(\cdot)$ and $y(\cdot)$ of (4) it holds that

$$\|\phi_t\big(x_0\big) - \phi_t(y_0)\| \leq \mathrm{e}^{-ct}\|x_0 - y_0\|, \quad \text{for all } t \geq 0,$$

i.e., the distance between the two trajectories converges exponentially with rate c if f is c-strongly infinitesimally contracting, and never increases if f is weakly infinitesimally contracting.

In [13, Theorem 16] condition (5) is generalized for locally Lipschitz function, for which, by Rademacher's theorem, the Jacobian exists almost everywhere (a.e.) in $\mathcal{C}$. Specifically, if f is locally Lipschitz, then f is infinitesimally contracting on $\mathcal{C}$ if condition (5) holds for almost every $x \in \mathcal{C}$ and $t \in \mathbb{R}_{\geq 0}$.

Finally, we recall the following result on the convergence behavior of globally-weakly and locally-strongly contracting dynamics. We refer to [7] for more details.

Theorem 1 (Linear-exponential convergence of globally-weakly and locally-strongly contracting dynamics)**.** *Consider system (4). Assume that (i) f is weakly infinitesimally contracting on $\mathbb{R}^n$ w.r.t. $\|\cdot\|$ and (ii) c_{exp}-strongly infinitesimally contracting on a forward-invariant set $\mathcal{S}$ w.r.t. $\|\cdot\|$, and (iii) that there exists an equilibrium point $x^\star \in \mathcal{S}$. Also, let r be the largest radius such that $B\big(x^\star, r\big) \subseteq \mathcal{S}$. For each trajectory $x(t)$ starting from x_0, it holds that*

1. *if $x_0 \in \mathcal{S}$, then, for almost every $t \geq 0$,*

$$\|x(t) - x^\star\| \leq \mathrm{e}^{-c_{\mathrm{exp}}t}\|x_0 - x^\star\|;$$

2. *if $x_0 \notin \mathcal{S}$, then, for almost every $t \geq 0$,*

$$\|x(t) - x^\star\| \leq \text{lin-exp}(t\,;\, q, c_{\mathrm{lin}}, c_{\mathrm{exp}}, t_{\mathrm{cross}}) := \begin{cases} q - c_{\mathrm{lin}}t & \text{if } t \leq t_{\mathrm{cross}}, \\ \big(q - c_{\mathrm{lin}}t_{\mathrm{cross}}\big)\mathrm{e}^{-c_{\mathrm{exp}}(t - t_{\mathrm{cross}})} & \text{if } t > t_{\mathrm{cross}}. \end{cases}, \tag{6}$$

with exponential decay rate $c_{\mathrm{exp}} > 0$, linear decay rate $c_{\mathrm{lin}} = c_{\mathrm{exp}}\, r$, intercept $q = \|x_0 - x^\star\|$, linear-exponential crossing time $t_{\mathrm{cross}} = (q - r)/c_{\mathrm{lin}}$.

For brevity, we say that every solution of a dynamics satisfying Theorem 1 *linear-exponentially converge* towards its equilibrium. Finally, whenever it is clear from the context, we omit specifying the dependence of functions on time t.

3 Equality-constrained Composite Optimization via Feedback Control

Consider the equality-constrained composite OP (1) that we rewrite here for convenience

$$\min_{x\in\mathbb{R}^n} \; f(x)+g(x)$$
$$\text{s.t.} \;\; h(x)=0,$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $h\colon \mathbb{R}^n \to \mathbb{R}^m$ are convex and differentiable functions, while $g\colon \mathbb{R}^n \to \mathbb{R}$ is a convex, closed, and proper function, possibly non-smooth.

Remark 1. *Problem (1) includes inequality-constrained optimization problems. To see this, consider the problem*

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \;& f(x)\\ s.t. \;& g(x)<0\\ & h(x)=0,\end{aligned}$$

and note that the above minimization problem can be equivalently rewritten as

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \;& f(x)+\iota_{\{x \,\mid\, g(x)<0\}}(x)\\ s.t. \;& h(x)=0.\end{aligned}$$

To solve problem (1), we propose a continuous-time closed-loop dynamical system, for which we design a suitable feedback controller driving the dynamics towards a minimizer of (1). To this aim, consider the Lagrangian associated to problem (1), that is the map $L\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

$$L(x,\lambda)=f(x)+g(x)+\lambda^\top h(x),$$

where $\lambda \in \mathbb{R}^m$ is the vector of Lagrange multipliers.

Definition 3 (Stationary point for (1))**.** *Consider the equality-constrained OP (1). A* stationary point *for problem (1) is any pair $(x^\star,\lambda^\star)\in\mathbb{R}^{n+m}$ such that $\mathbb{0}_n \in \nabla f(x^\star)+\partial g(x^\star)+D\big(h(x^\star)\big)^\top\lambda^\star$, and $h(x^\star)=0$.*

Remark 2. *By the first-order necessary and sufficient optimality conditions for convex OPs [21], a vector $x^\star \in \mathbb{R}^n$ satisfying $h(x^\star)=0$ is a minimizer of (1) if and only if there exists $\lambda^\star\in\mathbb{R}^m$ such that $(x^\star,\lambda^\star)$ is a saddle point of the Lagrangian, that is $\mathbb{0}_n \in \nabla f(x^\star)+\partial g(x^\star)+D\big(h(x^\star)\big)^\top\lambda^\star$, and $h(x^\star)=0$. Therefore, if $(x^\star,\lambda^\star)\in\mathbb{R}^{n+m}$ is a stationary point for (1), then $x^\star$ is a minimizer of (1).*

Interpreting Lagrange multipliers $\lambda(t)\in\mathbb{R}^m$ as control input, we introduce the following dynamics:

$$\begin{cases}\dot{x}(t)=-x(t)+\operatorname{prox}_{\gamma g}\big(x(t)-\gamma(\nabla f(x(t))+D\big(h(x(t))\big)^\top\lambda(t))\big)\\ y(t)=h(x(t))\end{cases}\tag{7}$$

where $x(t)\in\mathbb{R}^n$ is the state and $y(t)\in\mathbb{R}^m$ is the output. The following result establishes the connection between the stationary point of problem (1) and the equilibrium points of the dynamics (7).

Lemma 1 (Linking the stationary point of (1) and the equilibria of (7))**.** *The point $(x^\star,\lambda^\star)\in\mathbb{R}^{n+m}$ is a stationary point for (1) if and only if is an equilibrium point of problem (7) with input $\lambda^\star$.*

Proof. Let $(x^\star,\lambda^\star)\in\mathbb{R}^{n+m}$ be a stationary point for (1). Then $\mathbb{0}_n \in \nabla f(x^\star)+\partial g(x^\star)+D\big(h(x^\star)\big)^\top\lambda^\star$. Multiplying by $\gamma>0$ and adding and subtracting $x^\star$ to the right-hand side of the above inclusion yields

$$\begin{aligned}\mathbb{0}_n \in [I_n+\gamma\partial g](x^\star)+\gamma\nabla f(x^\star)+\gamma D\big(h(x^\star)\big)^\top\lambda^\star-x^\star \iff & (I_n+\gamma\partial g)(x^\star)\in x^\star-\gamma\big(\nabla f(x^\star)+D\big(h(x^\star)\big)^\top\lambda^\star\big)\\ \iff & x^\star\in(I_n+\gamma\partial g)^{-1}\left(x^\star-\gamma\big(\nabla f(x^\star)+D\big(h(x^\star)\big)^\top\lambda^\star\big)\right).\end{aligned}$$

Recalling that $\operatorname{prox}_{\gamma g}=(I_n+\gamma\partial g)^{-1}$ and, being by assumption g convex, closed, and proper, then $\operatorname{prox}_{\gamma g}$ is single-valued [18]. Therefore, we have

$$x^\star=\operatorname{prox}_{\gamma g}\big(x^\star-\gamma(\nabla f(x^\star)+D\big(h(x^\star)\big)^\top\lambda^\star)\big).$$

That is, $x^\star$ is an equilibrium point of problem (7). Specifically, $x^\star$ is the equilibrium with input $\lambda^\star$. Since all results are equivalence, the proof is complete. □

Inspired by [10], to compute the minimizers of (1) next we design a suitable control input $\lambda(t)$ that drives the system (7) to an equilibrium point, say it $x^\star$, which is a stationary point of problem (1). Specifically, $\lambda(t)$ is the output of a PI controller, so that:

$$\lambda(t) = k_{\mathrm{p}} y(t) + k_{\mathrm{i}} \int_0^t y(\tau) d\tau,$$

where $k_{\mathrm{p}}, k_{\mathrm{i}} \in \mathbb{R}_{>0}$ are the control gains. Hence, we have

$$\dot{\lambda}(t) = k_{\mathrm{p}} \dot{y}(t) + k_{\mathrm{i}} y(t). \tag{8}$$

Then, the closed-loop dynamics composed by system (7) and the PI controller (8) – illustrated in Figure 2 – is the following continuous-time *PI proximal-gradient dynamics* (PI–PGD)

$$\begin{cases} \dot{x} = -x + \mathrm{prox}_{\gamma g}\big(x - \gamma(\nabla f(x) + D\big(h(x)\big)^\top \lambda)\big) \\ \dot{\lambda} = k_{\mathrm{p}} A \dot{x} + k_{\mathrm{i}} h(x), \end{cases}$$

or equivalently

$$\begin{cases} \dot{x} &= -x + \mathrm{prox}_{\gamma g}\big(x - \gamma(\nabla f(x) + D\big(h(x)\big)^\top \lambda)\big) \\ \dot{\lambda} &= k_{\mathrm{p}} D\big(h(x)\big)\big(-x + \mathrm{prox}_{\gamma g}\big(x - \gamma(\nabla f(x) + D\big(h(x)\big)^\top \lambda)\big)\big) + k_{\mathrm{i}} h(x). \end{cases} \tag{9}$$

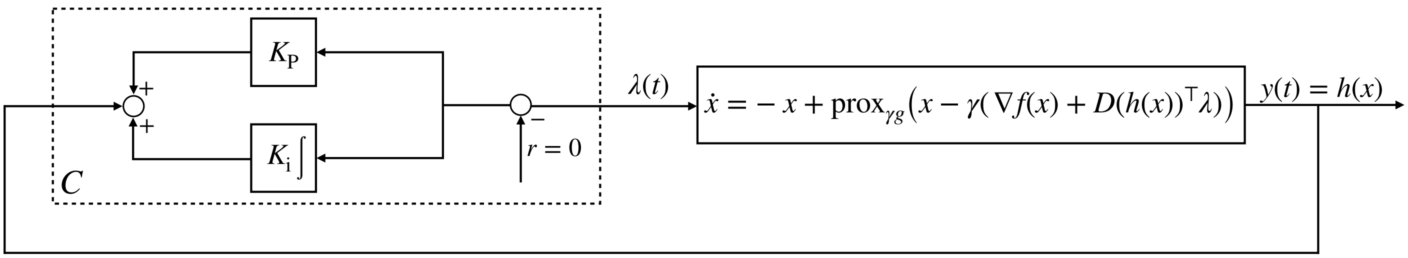

Figure 2: PI–PGD: Closed-loop dynamics composed by system (7) and the PI controller (8).

The following result is an immediate consequence of Lemma 1.

Corollary 1. *The point $(x^\star, \lambda^\star) \in \mathbb{R}^n \times \mathbb{R}^m$ is an equilibrium point of (9) if and only if $x^\star$ is a minimizer of (1).*

Proof. Follows from Lemma 1, noticing that if $(x^\star, \lambda^\star) \in \mathbb{R}^n \times \mathbb{R}^m$ is an equilibrium point of (9) then $h(x^\star) = 0$. □

4 Convergence of the PI Proximal-Gradient Dynamics

We now focus on the case of affine constraints and study the convergence properties of the resulting PI–PGD. That is, we assume that $h(x) = Ax - b$, where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. The PI–PGD then becomes

$$\begin{cases} \dot{x} = -x + \mathrm{prox}_{\gamma g}\left(x - \gamma(\nabla f(x) + A^\top \lambda)\right) \\ \dot{\lambda} = (k_{\mathrm{i}} - k_{\mathrm{p}}) A x + k_{\mathrm{p}} A \mathrm{prox}_{\gamma g}\left(x - \gamma(\nabla f(x) + A^\top \lambda)\right) - k_{\mathrm{i}} b. \end{cases} \tag{10}$$

In what follows, we let $z = (x, \lambda) \in \mathbb{R}^{n+m}$ and let $F_{\mathrm{PGD}} \colon \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ be the vector field (10) for $\dot{z} = F_{\mathrm{PGD}}(z)$. Additionally, given a point $z^\star \in \mathbb{R}^{n+m}$, we let $\Omega_{\mathrm{F}}(z^\star)$ be the set of differentiable points of F_{PGD} in a neighborhood of $z^\star$.

We make the following standard assumptions on the function f and the matrix A.

Assumption. *Given the system (10), assume*

(A1) the function $f \colon \mathbb{R}^n \to \mathbb{R}$ is strongly convex and strongly smooth with parameters ρ and L, respectively;

(A2) the matrix $A \in \mathbb{R}^{m\times n}$ satisfies $a_{\min} I_m \preceq AA^\top \preceq a_{\max} I_m$ for $a_{\min}, a_{\max} \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}^m$.

We begin our analysis by showing that, for proper parameter choice, the PI–PGD is weakly infinitesimally contracting, which in turn implies that the distance between any two trajectories of the PI–PGD never increases.

Lemma 2 (Global weak contractivity of (10))**.** *Consider the PI–PGD (10) with $k_{\mathrm{p}} = k_{\mathrm{i}} > 0$ satisfying Assumption (A1). For any $p \in \left[\max\left(\frac{k_{\mathrm{p}} L}{3}, \frac{k_{\mathrm{p}}(1-2\gamma\rho)}{\gamma}\right), \frac{k_{\mathrm{p}}}{\gamma}\right]$ and for any $\gamma \in \left]0, \frac{1}{L}\right]$, the PI–PGD (10) is weakly infinitesimally contracting on $\mathbb{R}^n$ with respect to the norm $\|\cdot\|_P$, where*

$$P = \begin{bmatrix} pI_n & 0 \\ 0 & I_m \end{bmatrix}.$$

Proof. For simplicity of notation, let $y := x - \gamma(\nabla f(x) + A^\top \lambda)$ and define $G(y) := D\mathrm{prox}_{\gamma g}(y)$. Recall that $G(y)$ is symmetric and $\mathrm{prox}_{\gamma g}(y)$ is nonexpansive [3, Proposition 12.28], therefore it holds $0 \preceq G(y) \preceq I_n$, for a.e. y. The Jacobian of F_{PGD} is

$$DF_{\mathrm{PGD}}(z) = \begin{bmatrix} -I_n + G(y)(I_n - \gamma\nabla^2 f(x)) & -\gamma G(y) A^\top \\ (k_{\mathrm{i}} - k_{\mathrm{p}})A + k_{\mathrm{p}} A G(y)\big(I_n - \gamma \nabla^2 f(x)\big) & -\gamma k_{\mathrm{p}} A G(y) A^\top \end{bmatrix},$$

which exists for almost every z. To prove our statement, we have to show that $\mu_P(DF_{\mathrm{PGD}}(z)) \leq 0$, for a.e. z. Note that

$$\sup_z \mu_P(DF_{\mathrm{PGD}}(z)) \leq \max_{\substack{0 \preceq G \preceq I_m \\ \rho I_n \preceq B \preceq L I_n}} \mu_P\Bigg(\underbrace{\begin{bmatrix} -I_n + G(I_n - \gamma B) & -\gamma G A^\top \\ (k_{\mathrm{i}} - k_{\mathrm{p}})A + k_{\mathrm{p}} A G\big(I_n - \gamma B\big) & -\gamma k_{\mathrm{p}} A G A^\top \end{bmatrix}}_{:=J}\Bigg),$$

where the sup is over all z for which $DF_{\mathrm{PGD}}(z)$ exists. Then, to prove our statement it suffices to show that the LMI $-J^\top P - PJ \succeq 0$ is satisfied. We compute

$$\begin{aligned}
-J^\top P - PJ &= \begin{bmatrix} 2pI_n - 2pG + p\gamma(BG + GB) & (k_{\mathrm{p}} - k_{\mathrm{i}})A^\top + \big(\gamma p I_n - k_{\mathrm{p}}(I_n - \gamma B)\big) G A^\top \\ (k_{\mathrm{p}} - k_{\mathrm{i}})A + AG\big(\gamma p I_n - k_{\mathrm{p}}(I_n - \gamma B)\big) & 2\gamma k_{\mathrm{p}} A G A^\top \end{bmatrix} \\
&= \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} 2pI_n - 2pG + p\gamma(BG + GB) & (k_{\mathrm{p}} - k_{\mathrm{i}})I_n + \big((\gamma p - k_{\mathrm{p}})I_n + \gamma k_{\mathrm{p}} B\big)G \\ (k_{\mathrm{p}} - k_{\mathrm{i}})I_n + G\big((\gamma p - k_{\mathrm{p}})I_n + \gamma k_{\mathrm{p}} B\big) & 2\gamma k_{\mathrm{p}} G \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & A^\top \end{bmatrix} \\
&:= \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix} \underbrace{\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & Q_{22} \end{bmatrix}}_{:=Q} \begin{bmatrix} I_n & 0 \\ 0 & A^\top \end{bmatrix} \succeq 0 \quad \iff \quad Q \succeq 0.
\end{aligned}$$

To show $Q \succeq 0$, we need to prove that (i) $Q_{11} \succ 0$ and (ii) the Schur complement of the block Q_{11} is positive semidefinite, that is, $Q_{22} - Q_{12}^\top Q_{11}^{-1} Q_{12} \succeq 0$.

(i) $\mathbf{Q_{11} \succ 0}$. We begin by noting that G is symmetric and satisfies $0 \preceq G \preceq I_n$, then there exists U satisfying $UU^\top = U^\top U = I_n$ and $g \in [0,1]^n$ such that $G = U[g]U^\top$. Substituting this equality into Q_{11} and multiplying on the left and on the right by $U^\top$ and U, respectively, we get

$$U^\top Q_{11} U = 2pI_n - 2p[g] + p\gamma(U^\top B U[g] + [g]U^\top B U) := 2p\big(I_n - [g]\big) + p\gamma(X[g] + [g]X),$$

where in the last equality we defined the matrix $X := U^\top B U$. Note that $X = X^\top$, $0 \prec \rho I_n \preceq B \preceq L I_n$ and, by assumption, $\gamma \leq \frac{1}{L}$. Then, by applying Lemma 5, for all $g \in [0,1]^n$ it holds

$$p\gamma(X[g] + [g]X) + 2p\big(I_n - [g]\big) \succ \frac{3}{2} p\gamma X. \tag{11}$$

By multiplying (11) on the left and on the right by U and $U^\top$, respectively, we get the following lower bound on Q_{11}

$$Q_{11} \succ \frac{3}{2}\gamma p B \succeq \frac{3}{2}\gamma \rho p I_n \succ 0.$$

(ii) $\mathbf{Q_{22} - Q_{12}^\top Q_{11}^{-1} Q_{12} \succeq 0}$. Let $Q_{12} = (k_{\mathrm{p}} - k_{\mathrm{i}})I_n + \gamma p G - k_{\mathrm{p}} G + \gamma k_{\mathrm{p}} B G := BR_1 + R_2$, where to simplify the notation we have introduced the matrices $R_1 := \gamma k_{\mathrm{p}} G = R_1^\top$, and $R_2 := (k_{\mathrm{p}} - k_{\mathrm{i}})I_n + (\gamma p - k_{\mathrm{p}})G = R_2^\top$. Then

$$\begin{aligned}
\frac{3\gamma p}{2} Q_{12}^\top Q_{11}^{-1} Q_{12} &\preceq (BR_1 + R_2)^\top B^{-1}(BR_1 + R_2) \preceq (R_1 B + R_2)B^{-1}(BR_1 + R_2) \\
&= R_1 B R_1 + 2R_1 R_2 + R_2 B^{-1} R_2,
\end{aligned}$$

where

$$\begin{aligned}
R_1 B R_1 &= \gamma^2 k_{\mathrm{p}}^2 G B G \preceq \gamma^2 k_{\mathrm{p}}^2 L G^2 \\
R_1 R_2 &= \gamma k_{\mathrm{p}}(k_{\mathrm{p}} - k_{\mathrm{i}})G + \gamma k_{\mathrm{p}}(\gamma p - k_{\mathrm{p}})G^2 \\
R_2 B^{-1} R_2 &= (k_{\mathrm{p}} - k_{\mathrm{i}})^2 B^{-1} + (\gamma p - k_{\mathrm{p}})^2 G B^{-1} G + (k_{\mathrm{p}} - k_{\mathrm{i}})(\gamma p - k_{\mathrm{p}})B^{-1}G + (k_{\mathrm{p}} - k_{\mathrm{i}})(\gamma p - k_{\mathrm{p}})GB^{-1} \\
&\preceq (k_{\mathrm{p}} - k_{\mathrm{i}})^2 \rho^{-1} I_n + (\gamma p - k_{\mathrm{p}})^2 \rho^{-1} G^2 + (k_{\mathrm{p}} - k_{\mathrm{i}})(\gamma p - k_{\mathrm{p}})B^{-1}G + (k_{\mathrm{p}} - k_{\mathrm{i}})(\gamma p - k_{\mathrm{p}})GB^{-1},
\end{aligned}$$

where in the last inequality we used the fact that the LMI $L^{-1}I_n \preceq B^{-1} \preceq \rho^{-1}I_n$ implies $L^{-1}G^2 I_n \preceq GB^{-1}G \preceq \rho^{-1}G^2$. Summing up, so far we have

$$\frac{3\gamma p}{2} Q_{12}^\top Q_{11}^{-1} Q_{12} \preceq \gamma^2 k_\mathrm{p}^2 L G^2 + 2\gamma k_\mathrm{p}(k_\mathrm{p} - k_\mathrm{i})G + 2\gamma k_\mathrm{p}(\gamma p - k_\mathrm{p})G^2 + (k_\mathrm{p} - k_\mathrm{i})^2 \rho^{-1} I_n \\ + (\gamma p - k_\mathrm{p})^2 \rho^{-1} G^2 + (k_\mathrm{p} - k_\mathrm{i})(\gamma p - k_\mathrm{p}) B^{-1} G + (k_\mathrm{p} - k_\mathrm{i})(\gamma p - k_\mathrm{p}) G B^{-1}.$$

Now, to ensure the LMI $Q_{22} - Q_{12}^\top Q_{11}^{-1} Q_{12} \succeq 0$ holds, the only approach is to set the term proportional to $\rho^{-1}I_n$ to zero, that is set $k_\mathrm{p} = k_\mathrm{i}$. Next, we compute

$$\begin{aligned}
\frac{3\gamma p}{2}\left(Q_{22} - Q_{12}^\top Q_{11}^{-1} Q_{12}\right) &\succeq 3\gamma^2 p k_\mathrm{p} G - \gamma^2 k_\mathrm{p}^2 L G^2 - 2\gamma k_\mathrm{p}(\gamma p - k_\mathrm{p})G^2 - (\gamma p - k_\mathrm{p})^2 \rho^{-1} G^2 \\
&\succeq \left(3\gamma^2 p k_\mathrm{p} - \gamma^2 k_\mathrm{p}^2 L - 2\gamma k_\mathrm{p}(\gamma p - k_\mathrm{p}) - (\gamma p - k_\mathrm{p})^2 \rho^{-1}\right) G^2 \succeq 0 &\quad (12)\\
&\iff 3p\gamma^2 \rho k_\mathrm{p} - \gamma^2 k_\mathrm{p}^2 L \rho - 2\gamma\rho k_\mathrm{p}(\gamma p - k_\mathrm{p}) - (\gamma p - k_\mathrm{p})^2 \geq 0 \\
&\iff \gamma^2 \rho k_\mathrm{p}(3p - k_\mathrm{p} L) - (\gamma p - k_\mathrm{p})(2\gamma\rho k_\mathrm{p} + \gamma p - k_\mathrm{p}) \geq 0 \\
&\impliedby \gamma^2 \rho k_\mathrm{p}(3p - k_\mathrm{p} L) \geq 0 \quad \text{and} \quad (\gamma p - k_\mathrm{p})(2\gamma\rho k_\mathrm{p} + \gamma p - k_\mathrm{p}) \leq 0, &\quad (13)
\end{aligned}$$

where (12) follows from the inequality $G^2 \preceq G$. Now, we have

(i) $\gamma^2 \rho k_\mathrm{p}(3p - k_\mathrm{p} L) \geq 0 \iff p \geq \frac{k_\mathrm{p} L}{3}$, and

(ii) $(\gamma p - k_\mathrm{p})(2\gamma\rho k_\mathrm{p} + \gamma p - k_\mathrm{p}) \leq 0 \iff \frac{k_\mathrm{p}(1-2\gamma\rho)}{\gamma} \leq p \leq \frac{k_\mathrm{p}}{\gamma}$, where we used the fact that $-1 \leq 1 - 2\gamma\rho \leq 1$ being $\gamma\rho \leq \frac{\rho}{L} \leq 1$.

Summing up, inequalities (13) are satisfied for any $p \in \left[\max\left(\frac{k_\mathrm{p} L}{3}, \frac{k_\mathrm{p}(1-2\gamma\rho)}{\gamma}\right), \frac{k_\mathrm{p}}{\gamma}\right]$. This concludes the proof. □

Next, we prove that the equilibrium point of (10) is not only locally exponentially stable but also locally-strongly contracting in a suitably defined norm.

Lemma 3 (Local exponential stability and local strong contractivity of (10))**.** *Suppose Assumptions (A1) and (A2) hold and let $z^\star := (x^\star, \lambda^\star)$ be an equilibrium point of (10). Then*

1. *for any γ, k_i and $k_\mathrm{p} > 0$, the point $(x^\star, \lambda^\star)$ is locally exponentially stable;*
2. *for any p, γ, k_i and $k_\mathrm{p} > 0$ satisfying*

$$4\gamma^2 p k_\mathrm{p} \rho - \max\{(\gamma p - k_\mathrm{i} + \gamma k_\mathrm{p} \rho)^2, (\gamma p - k_\mathrm{i} + \gamma k_\mathrm{p} L)^2\} > 0, \quad (14)$$

the PI–PGD (10) is strongly infinitesimally contracting with rate $c_{\exp} > 0$ with respect to the norm $\|\cdot\|_P$ in the set $B_P(z^\star, r)$, where $r := \sup\{a > 0 \mid B_P(z^\star, a) \subset \Omega_\mathrm{F}(z^\star)\}$, and

$$P = \begin{bmatrix} pI_n & 0 \\ 0 & I_m \end{bmatrix}.$$

Proof. We start noticing that $D\mathrm{prox}_{\gamma g}\left(x^\star - \gamma(\nabla f(x^\star) + A^\top \lambda^\star)\right) = I_n$. To prove statement 1, we show that $DF_\mathrm{PGD}(x^\star, \lambda^\star)$ is a Hurwitz matrix, i.e., $\alpha\left(DF_\mathrm{PGD}(x^\star, \lambda^\star)\right) < 0$, for any γ, k_i and $k_\mathrm{p} > 0$. We compute

$$DF_\mathrm{PGD}(z^\star) = \begin{bmatrix} -\gamma\nabla^2 f(x^\star) & -\gamma A^\top \\ k_\mathrm{i} A - \gamma k_\mathrm{p} A \nabla^2 f(x^\star) & -\gamma k_\mathrm{p} A A^\top \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} -\gamma\nabla^2 f(x^\star) & -\gamma I_n \\ k_\mathrm{i} I_n - \gamma k_\mathrm{p} \nabla^2 f(x^\star) & -\gamma k_\mathrm{p} I_n \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & A^\top \end{bmatrix}.$$

Now, being $-\gamma k_\mathrm{p} I_n \prec 0$, to show $DF_\mathrm{PGD}(z^\star) \prec 0$ it suffices to check that the Schur complement of the block $-\gamma k_\mathrm{p} I_n$ is negative definite. We have

$$-\gamma\nabla^2 f(x^\star) - (-\gamma I_n)(-\gamma k_\mathrm{p})^{-1}(k_\mathrm{i} I_n - \gamma k_\mathrm{p} \nabla^2 f(x^\star) \prec 0 \iff \\ -\gamma\nabla^2 f(x^\star) - k_\mathrm{i} k_\mathrm{p}^{-1} I_n + \gamma\nabla^2 f(x^\star)) \prec 0 \iff -k_\mathrm{i} k_\mathrm{p}^{-1} < 0,$$

which holds for every $k_\mathrm{i}, k_\mathrm{p} > 0$. This concludes the proof of item 1. Next, we show that the system is locally strongly contracting with respect to the norm $\|\cdot\|_P$. Being $J := DF_\mathrm{PGD}(z^\star)$ Hurwitz, there exists P invertible such that $\mu_{2,P}(J) < 0$ [6, Corollary 2.33], or equivalently if the following LMI holds $-J^\top P - PJ \succ 0$. Let

$$P = \begin{bmatrix} pI_n & 0 \\ 0 & I_m \end{bmatrix}.$$

We have

$$
\begin{aligned}
-PJ - J^\top P &= \begin{bmatrix} \gamma p \nabla^2 f(x^\star) & \gamma p A^\top \\ -k_{\mathrm{i}} A + \gamma k_{\mathrm{p}} A \nabla^2 f(x^\star) & \gamma k_{\mathrm{p}} A A^\top \end{bmatrix} + \begin{bmatrix} \gamma p \nabla^2 f(x^\star) & -k_{\mathrm{i}} A^\top + \gamma k_{\mathrm{p}} \nabla^2 f(x^\star) A^\top \\ \gamma p A & \gamma k_{\mathrm{p}} A A^\top \end{bmatrix} \\
&= \begin{bmatrix} 2\gamma p \nabla^2 f(x^\star) & \left(\gamma p I_n - k_{\mathrm{i}} I_n + \gamma k_{\mathrm{p}} \nabla^2 f(x^\star)\right) A^\top \\ A\left(\gamma p I_n - k_{\mathrm{i}} I_n + \gamma k_{\mathrm{p}} \nabla^2 f(x^\star)\right) & 2\gamma k_{\mathrm{p}} A A^\top \end{bmatrix} \\
&= \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} 2\gamma p \nabla^2 f(x^\star) I_n & \gamma p I_n - k_{\mathrm{i}} I_n + \gamma k_{\mathrm{p}} \nabla^2 f(x^\star) I_n \\ \gamma p I_n - k_{\mathrm{i}} I_n + \gamma k_{\mathrm{p}} \nabla^2 f(x^\star) I_n & 2\gamma k_{\mathrm{p}} I_m \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & A^\top \end{bmatrix} \\
&:= \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix} \underbrace{\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & Q_{22} \end{bmatrix}}_{:=Q} \begin{bmatrix} I_n & 0 \\ 0 & A^\top \end{bmatrix} \succ 0,
\end{aligned}
$$

To show $Q \succ 0$ we need to prove that (i) $Q_{22} \succ 0$ and (ii) the Schur complement of the block Q_{22} is positive semidefinite, that is, $Q_{11} - Q_{12} Q_{22}^{-1} Q_{12}^\top \succ 0$. We trivially have $Q_{22} = 2\gamma k_{\mathrm{p}} I_n \succ 0$. Next, we compute

$$
\begin{aligned}
Q_{11} - Q_{12} Q_{22}^{-1} Q_{12}^\top &= 2\gamma p \nabla^2 f(x^\star) I_n - \frac{1}{2\gamma k_{\mathrm{p}}} \left(\gamma p - k_{\mathrm{i}} + \gamma k_{\mathrm{p}} \nabla^2 f(x^\star)\right)^2 I_n \\
&\succeq 4\gamma^2 p k_{\mathrm{p}} \rho I_n - \max\{(\gamma p - k_{\mathrm{i}} + \gamma k_{\mathrm{p}} \rho)^2, (\gamma p - k_{\mathrm{i}} + \gamma k_{\mathrm{p}} L)^2\} I_n \succ 0 \qquad (15) \\
&\iff 4\gamma^2 p k_{\mathrm{p}} \rho - \max\{(\gamma p - k_{\mathrm{i}} + \gamma k_{\mathrm{p}} \rho)^2, (\gamma p - k_{\mathrm{i}} + \gamma k_{\mathrm{p}} L)^2\} > 0,
\end{aligned}
$$

where in (15) we used the fact that $\rho I_n \preceq \nabla^2 f(x) \preceq L I_n$, for all $x \in \mathbb{R}^n$. Then, by the continuity property of the lognorm, there exists $c_{\mathrm{exp}} > 0$ and $B_P(z^\star, r)$, with $r := \sup\{a > 0 \mid B_P(z^\star, a) \subset \Omega_{\mathrm{F}}(z^\star)\}$, where $DF_{\mathrm{PGD}}(z)$ exists and $\mu_{2,P}(DF_{\mathrm{PGD}}(z)) < -c_{\mathrm{exp}}$ for all $z \in B_P(z^\star, r)$. This concludes the proof. $\square$

Finally, the next result shows that the PI–PGD method converges linear-exponentially to the equilibrium point.

Theorem 2 (Linear-exponential stability of (10)). *Consider the PI–PGD (10) with $k_{\mathrm{p}} = k_{\mathrm{i}} > 0$ satisfying Assumptions (A1) and (A2) and let $z^\star := (x^\star, \lambda^\star)$ be an equilibrium point. Then, for any $\gamma \in \left]0, \frac{1}{L}\right]$ and $p \in \left[\max\left(\frac{k_{\mathrm{p}} L}{3}, \frac{k_{\mathrm{p}}(1-2\gamma\rho)}{\gamma}\right), \frac{k_{\mathrm{p}}}{\gamma}\right]$ satisfying*

$$
4\gamma^2 p k_{\mathrm{p}} \rho - (\gamma p + k_{\mathrm{p}}(\gamma L - 1))^2 > 0, \tag{16}
$$

every solution of the PI–PGD (10) linear-exponentially converges towards $z^\star$.

Proof. First note that the assumptions of Lemma 2 are satisfied, thus the system is globally weakly contracting on $\mathbb{R}^n$ with respect to the norm $\|\cdot\|_P$, where

$$
P = \begin{bmatrix} p I_n & 0 \\ 0 & I_m \end{bmatrix}. \tag{17}
$$

Next, we show that for $k_{\mathrm{p}} = k_{\mathrm{i}}$ and for any $p \in \left[\max\left(\frac{k_{\mathrm{p}} L}{3}, \frac{k_{\mathrm{p}}(1-2\gamma\rho)}{\gamma}\right), \frac{k_{\mathrm{p}}}{\gamma}\right]$ and any $\gamma \in \left]0, \frac{1}{L}\right]$, the inequality (14) is satisfied. First, note that the conditions $p \le \frac{k_{\mathrm{p}}}{\gamma}$ and $\gamma \le \frac{1}{L}$ implies $\gamma p - k_{\mathrm{p}} + \gamma k_{\mathrm{p}} \rho \le \gamma p - k_{\mathrm{p}} + \gamma k_{\mathrm{p}} L = \gamma p + k_{\mathrm{p}}(\gamma L - 1)$. The inequality (14) is then satisfied if

$$
4\gamma^2 p k_{\mathrm{p}} \rho - (\gamma p + k_{\mathrm{p}}(\gamma L - 1))^2 > 0,
$$

Then, Lemma (3) implies that there exists $c_{\mathrm{exp}} > 0$ such that the PI–PGD (10) is strongly infinitesimally contracting with rate $c_{\mathrm{exp}} > 0$ with respect to the norm $\|\cdot\|_P$ in the set $B_P(z^\star, r)$, where $r := \sup\{a > 0 \mid B_P(z^\star, a) \subset \Omega_{\mathrm{F}}(z^\star)\}$. Finally, the statement follows by applying Theorem 1. $\square$

Remark 3. *There exist parameter values satisfying the conditions in Theorem 2. For examples, by setting $k_{\mathrm{p}} = k_{\mathrm{i}} = \gamma p$, inequality (16) is satisfied for any $\gamma \in \left]0, \min\left\{4\frac{\rho}{L^2}, \frac{1}{L}\right\}\right[$. Indeed, we have*

$$
\begin{aligned}
&4\gamma^2 p K_{\mathrm{p}} \rho - \max\{(\gamma p - K_{\mathrm{i}} + \gamma K_{\mathrm{p}} \rho)^2, (\gamma p - K_{\mathrm{i}} + \gamma K_{\mathrm{p}} L)^2\} > 0 \\
&\iff 4\gamma^3 p^2 \rho - \gamma^4 p^2 L^2 > 0 \iff \gamma < 4\frac{\rho}{L^2}.
\end{aligned}
$$

5 Numerical Examples and Applications

We now demonstrate the effectiveness of the PI–PGD in solving constrained composite OPs via two applications: (i) constrained ℓ_1-regularized least squares problem, also known as the Least Absolute Shrinkage and Selection Operator (LASSO), and (ii) entropic-regularized optimal transport.

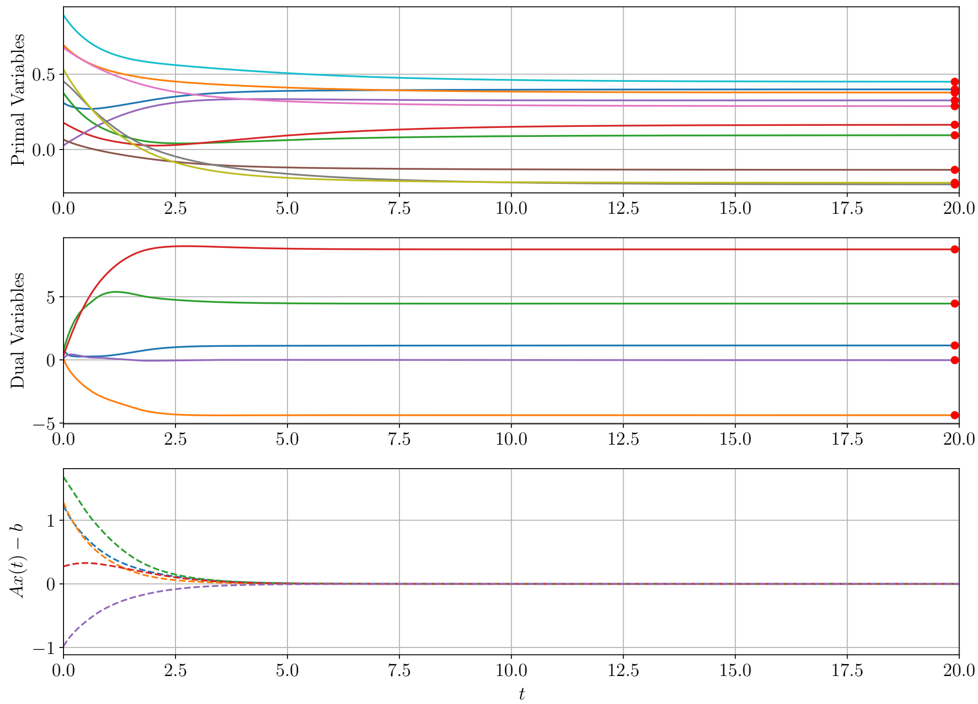

Figure 3: Trajectories of the dynamics (19) solving the constrained minimization problem (18). The figure shows the trajectories of the primal variables $x(t)$ (top) and two dual variables $\lambda(t)$ (middle) starting from z_0^1 and z_0^2 as solid curves and dashed curves, respectively. The `cvxpy` optimal values are shown as dots. The bottom panel shows the constraint $Ax(t) - b$ over time.

5.1 Equality-Constrained LASSO

Consider the following equality-constrained composite problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^\top W x + \alpha \|x\|_1 \\ & Ax = b, \end{aligned} \tag{18}$$

where $\alpha > 0$, $W \in \mathbb{R}^{n \times n}$ is positive definite, and thus the function $f(x) = \frac{1}{2} x^\top W x$ satisfies Assumption *(A1)*, and $A \in \mathbb{R}^{m \times n}$ satisfies Assumption *(A2)*. Recalling that $\text{prox}_{\gamma\alpha\|x\|_1} = \text{soft}_{\gamma\alpha}$, where $\text{soft}_{\gamma\alpha} \colon \mathbb{R}^n \to \mathbb{R}^n$ is the *soft thresholding function* defined by $(\text{soft}_{\gamma\alpha}(x))_i = \text{soft}_{\gamma\alpha}(x_i)$, and the map $\text{soft}_{\gamma\alpha} \colon \mathbb{R} \to \mathbb{R}$ is defined by

$$\text{soft}_{\gamma\alpha}(x_i) = \begin{cases} x_i - \gamma\alpha \, \text{sign}(x_i) & \text{if } |x_i| > \gamma\alpha, \\ 0 & \text{if } |x_i| \leq \gamma\alpha, \end{cases}$$

with $\text{sign} \colon \mathbb{R} \to \{-1, 0, 1\}$ being the *sign function* defined by $\text{sign}(x_i) := -1$ if $x_i < 0$, $\text{sign}(x_i) := 0$ if $x_i = 0$, and $\text{sign}(x_i) := 1$ if $x_i > 0$. Now and throughout the rest of the paper, we adopt a slight abuse of notation by using the same symbol to represent both the scalar and vector form of the activation function.

The PI–PGD (1) associated to problem (18) is then

$$\begin{cases} \dot{x} = & -x + \text{soft}_{\gamma\alpha}\big((I_n - \gamma W)x - \gamma A^\top \lambda\big) \\ \dot{\lambda} = & (k_\text{i} - k_\text{p})Ax + k_\text{p} A \, \text{soft}_{\gamma\alpha}\big((I_n - \gamma W)x - \gamma A^\top \lambda\big) - k_\text{i} b. \end{cases} \tag{19}$$

For the simulations, we set $n = 10$, $m = 5$, $\alpha = 1$, $W = 10 I_n + \tilde{W}\tilde{W}^\top \succ 0$, $\tilde{W}$, A and b with independent and normally distributed components. We solved (18) using `cvxpy` finding the optimal solution $z^\star = (x^\star, \lambda^\star)$. Next, we simulate (19) over the time interval $t \in [0, 20]$ with a forward Euler discretization with stepsize $\Delta t = 0.01$, starting from random initial conditions. In accordance with Remark 3, we set $\gamma = \min(1/L, 4\rho/L^2 - 10^{-4})$, $k_\text{i} = k_\text{p} = 20$, $p = k_\text{p}/\gamma$ and P as in (17), where ρ and L are the minimum and maximum eigenvalues of W, respectively. Simulations confirm, in accordance with our results, that the trajectories converge to $z^\star$. We plot the trajectories of the dynamics along with the optimal values found using `cvxpy`, and the constraint $Ax - b$ over time in Figure 3.

Figure 4 shows the mean and standard deviation of the lognorm of the $\|\cdot\|_P$ distance across 150 simulated trajectories of (19) w.r.t. $z^\star$. In agreement with Theorem 2, the figure shows that convergence is linearly-exponentially bounded.

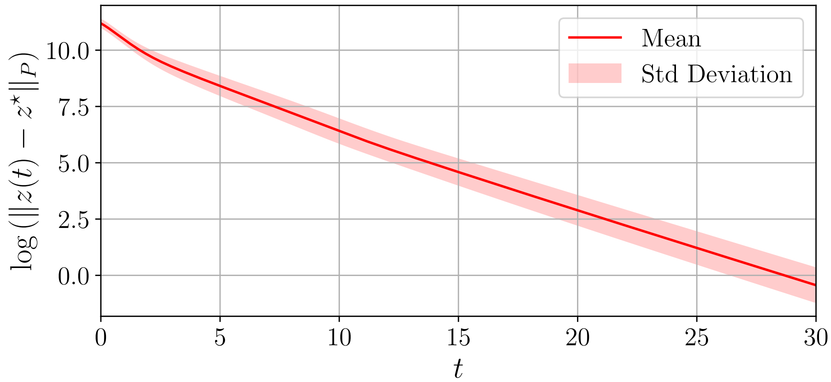

Figure 4: Mean/standard deviation of $\log(\|z(t) - z^{\star}\|_P)$ across 150 simulations. Note that, in agreement with Theorem 2, convergence is linearly-exponentially bounded.

5.2 Entropic Regularized Optimal Transport

We refer to the standard set-up from, e.g., [19]. Given a space $\mathcal{X}$, a *discrete probability measure* with weights $a \in \Sigma_n := \{q \in \mathbb{R}^n_{\geq 0} \mid \sum_{i=1}^n q_i = 1\}$ and locations $x_1, \ldots, x_n \in \mathcal{X}$ is $\alpha(a, x) = \sum_{i=1}^n a_i \delta_{x_i}$. Given $\alpha(a, x)$ and $\beta(b, y)$, the set of *admissible coupling* between them is $U(a, b) := \{P \in \mathbb{R}^{n \times m}_+ \mid P\mathbb{1}_m = a, P^\top \mathbb{1}_n = b\}$. The *entropic-regularized optimal transport* problem can be stated as

$$\min_{P \in U(a,b)} \sum_{i,j} P_{ij} C_{ij} + \epsilon \sum_{i,j} P_{ij} \log\left(P_{ij}\right), \tag{20}$$

where $C_{ij} := \mathrm{c}(x_i, y_j)$, $\mathrm{c} \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ is the cost function, $\epsilon > 0$ is a regularization parameter and with the standard convention $0 \log(0) = 0$. Problem (20) has a unique solution as its objective function is ϵ-strongly convex and the set $U(a, b)$ is convex. However, the objective function is not strongly smooth, since the Hessian is undefined if $P_{ij} = 0$, violating Assumption (*A*1). Despite this, we present a numerical example showing that the PI–PGD successfully solves problem (20) and converges even for small ϵ, in cases where the Sinkhorn algorithm fails. To write the PI–PGD we first vectorize (20). To this end, let $p = \mathrm{vec}(P) \in \mathbb{R}^{nm}$, $c = \mathrm{vec}(C) \in \mathbb{R}^{nm}$, $d = (a, b) \in \mathbb{R}^{n+m}$, and

$$A = \begin{bmatrix} \mathbb{1}_m^\top \otimes I_n \\ I_m \otimes \mathbb{1}_n^\top \end{bmatrix} \in \mathbb{R}^{(n+m) \times nm}.$$

Then, problem (20) becomes

$$\begin{aligned} \min_{p \geq 0} \; & p^\top c + \epsilon \sum_{i}^{nm} p_i \log p_i \\ \text{s.t.} \; & Ap = d, \end{aligned}$$

or equivalently in the form as problem (1):

$$\begin{aligned} \min_{p \in \mathbb{R}^{nm}} \; & p^\top c + \epsilon \sum_{i}^{nm} p_i \log p_i + \iota_{\mathbb{R}^{nm}_{\geq 0}}(p) \\ \text{s.t.} \; & Ap = d. \end{aligned} \tag{21}$$

Since A is not full row rank and $a \in \Sigma_n, b \in \Sigma_m$, any one of the $n + m$ equality constraints $Ap = d$ is redundant. Thus, without loss of generality, we remove the last constraint and let $\tilde{A}$ and $\tilde{d}$ be A without the last row and d without the last entry, respectively. This reduction ensures that $\tilde{A}$ has full row rank.

Remark 4. *The indicator function on $\mathbb{R}^{nm}_{\geq 0}$ in (21) enforces any feasible solution to be non-negative. Therefore, it is reasonable to analyze the function f_{OT} only in $\mathbb{R}^{nm}_{\geq 0}$.*

Then, the PI–PGD solving problem (21)

$$\begin{cases} \dot{p} = -p + \mathrm{ReLU}\left(p - \gamma(\epsilon \log p + \epsilon + c + \tilde{A}^\top \lambda)\right) \\ \dot{\lambda} = (k_{\mathrm{i}} - k_{\mathrm{p}})\tilde{A}p + k_{\mathrm{p}}\tilde{A}\,\mathrm{ReLU}\left(p - \gamma(\epsilon \log p + \epsilon + c + \tilde{A}^\top \lambda)\right) - k_{\mathrm{i}}\tilde{d}, \end{cases} \tag{22}$$

where $\lambda \in \mathbb{R}^{n+m-1}$. Note the positive orthant is a forward invariant set for the primal dynamics in (22). This ensures that starting with non-negative initial conditions p_0, then the primal variable $p(t)$ will remain non-negative.

We validate the PI–PGD (22) on an image morphing application from [5]. Here, optimal transport is used to transform a picture made of particles into another one by moving the particles. In our experiments, the starting number is 4 and the final desired number is 1. Source and target images are from the MNIST dataset. These pictures are stippled using the same number of particles $n = m = 100$ with the same mass, therefore a and b are uniform. We define the cost matrix as the Euclidean distance and set $\epsilon = 0.001$, $\gamma = 0.01$ and $k_p = k_i = 100$. The problem is solved using both the PI–PGD and Sinkhorn methods. Remarkably, in this set-up with small ϵ, the PI–PGD converges to the optimal transport plan – shown in Figure 5 –, while Sinkhorn does not converge. Figure 6 shows the norm of the constraint $Ap - b$ over the iterations. Finally, Figure 7 illustrates the initial, middle, and final frames of the image transformation. The gifs of the complete transformations are available at https://shorturl.at/mPpEZ.

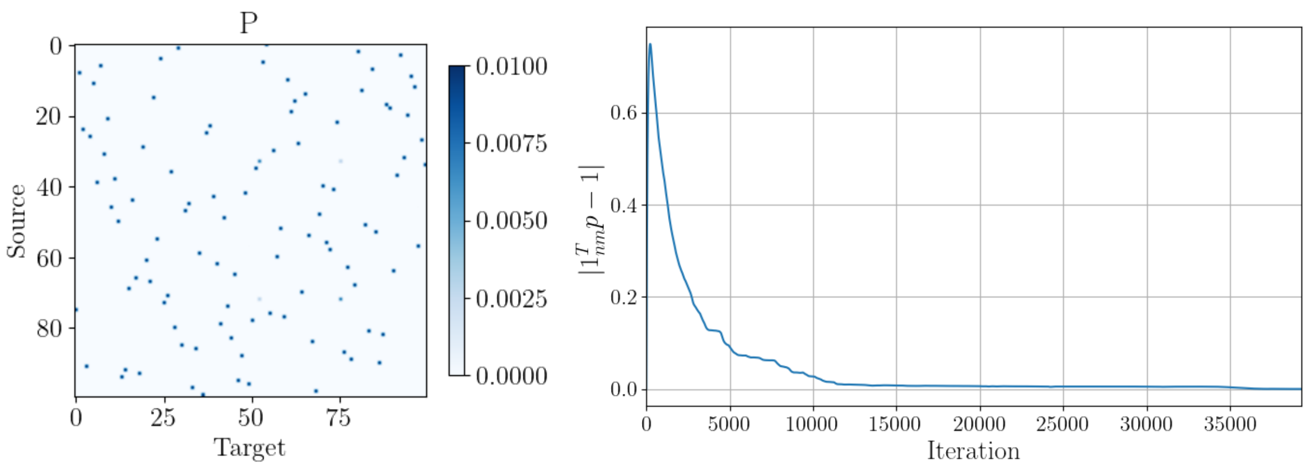

Figure 5: Optimal Transport plan obtained using the PI-PGD (22) (left). As expected the optimal vector $p \in \mathbb{R}^{nm}$ is a probability vector, that is $p \geq 0$ and $\sum_{j=1}^{nm} p_j = 1$ (right).

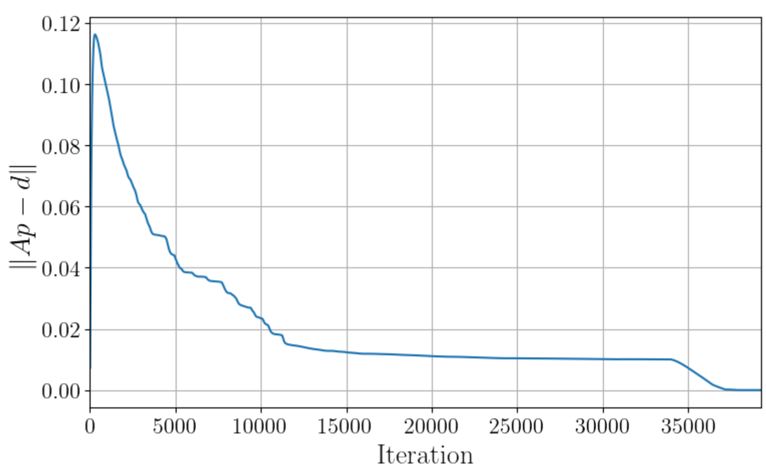

Figure 6: Norm of difference $Ap - d$ over the iterations. In agreement with our results, at convergence, the norm is zero, that is the optimal solution is feasible.

6 Discussions and Conclusions

We proposed the PI–PGD (9) for solving equality-constrained composite OPs. We established the equivalence between the minima of the OP and the equilibria of the PI–PGD. For strongly convex and strongly smooth cost functions with full row-rank constraints, we proved global linear-exponential convergence. Then, we demonstrated the effectiveness of our approach on equality-constrained LASSO and explored its application to entropic-regularized optimal transport.

(a)

(b)

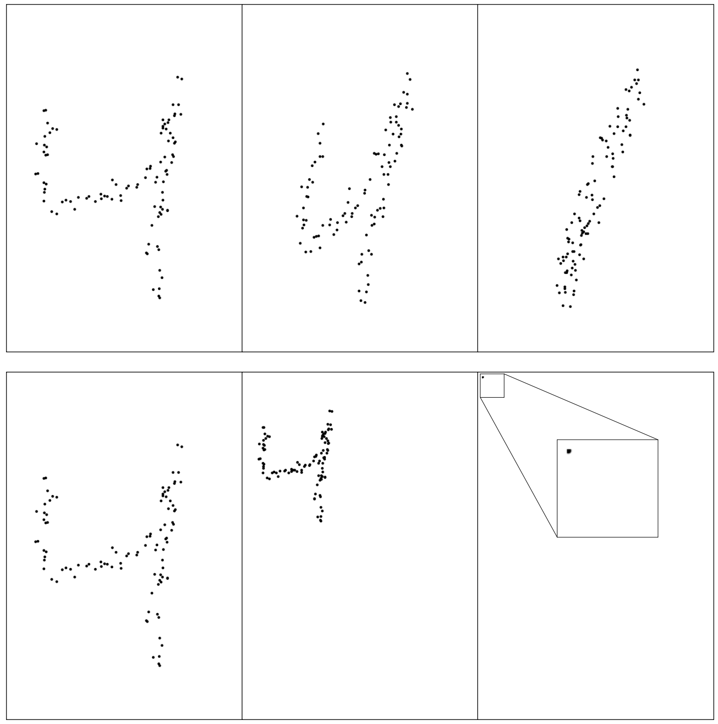

Figure 7: Image morphing via PI-PGD (a) and Sinkhorn (b).

Remarkably, the PI–PGD could solve this latter problem even when the state-of-the-art Sinkorn algorithm fails and despite Assumption $(A1)$ not being always fulfilled. Moreover, for both applications, when testing the PI–PGD for a wider range of parameters that included cases for which the conditions in Theorem 2 were not met, the PI–PGD would still converge to the optimum. Motivated by these numerical findings, future work will include: (i) obtaining sharper convergence bounds; (ii) relaxing $(A1)$–$(A2)$ to cover not-full-rank constraints [17]; relaxing $(A1)$ will extend our theoretical analysis to fully cover entropic-regularized optimal transport. As part of this research stream, we will also compare the convergence rate of PI–PGD with other optimal transport algorithms that do not require regularization.

A Instrumental Results

Lemma 4. *Consider a matrix $B = B^\top \in \mathbb{R}^{n\times n}$ satisfying $b_{\min} I_n \preceq B \preceq b_{\max} I_n$. Then, for any $a \in \mathbb{R}^n$ it holds*

$$\min\{a_{\min}^2, a_{\max}^2\} I_n \preceq (aI_n + B)^2 \preceq \max\{a_{\min}^2, a_{\max}^2\} I_n,$$

where $a_{\min} := a - b_{\min}$, and $a_{\max} := a - b_{\max}$.

Proof. We have

$$b_{\min} I_n \preceq B \preceq b_{\max} I_n \iff a_{\min} I_n \preceq aI_n + B \preceq a_{\max} I_n.$$

The proof follows noticing that if $|a_{\min}| \leq |a_{\max}|$, then $a_{\min}^2 I_n \preceq aI_n + B \preceq a_{\max}^2$, while if $|a_{\max}| \leq |a_{\min}|$, then $a_{\max}^2 I_n \preceq aI_n + B \preceq a_{\min}^2$. □

For completeness, we include a self-contained statement of the lemma proposed in [20, Lemma 6], used in the proof of Lemma 2.

Lemma 5. *Let $X = X^\top \in \mathbb{R}^{n\times n}$ satisfy $x_{\min} \preceq X \preceq x_{\max} I_n$ for some $x_{\max} \geq x_{\min} > 0$, $\gamma > 0$, and $\gamma \leq \dfrac{1}{x_{\max}}$. Then for all $g \in [0,1]^n$, the following inequality holds*

$$\gamma([g]X + X[g]) + 2(I_n - [g]) \succ \frac{3}{2}\gamma X. \tag{23}$$

Proof. See [20, Lemma 6]. □

References

[1] A. Allibhoy and J. Cortés. Control barrier function-based design of gradient flows for constrained nonlinear programming. 69(6), 2024. doi:10.1109/TAC.2023.3306492.

[2] K. J. Arrow, L. Hurwicz, and H. Uzawa, editors. *Studies in Linear and Nonlinear Programming*. Stanford University Press, 1958.

[3] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2 edition, 2017, ISBN 978-3-319-48310-8.

[4] G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese. Time-varying optimization of LTI systems via projected primal-dual gradient flows. *IEEE Transactions on Control of Network Systems*, 9(1):474–486, 2022. doi:10.1109/TCNS.2021.3112762.

[5] N. Bonneel, M. Van De Panne, S. Paris, and W. Heidrich. Displacement interpolation using Lagrangian mass transport. In *SIGGRAPH Asia conference*, number 158, pages 1–12, 2011. doi:10.1145/2024156.2024192.

[6] F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.2 edition, 2024, ISBN 979-8836646806. URL: https://fbullo.github.io/ctds.

[7] V. Centorrino, A. Davydov, A. Gokhale, G. Russo, and F. Bullo. On weakly contracting dynamics for convex optimization. 8:1745–1750, 2024. doi:10.1109/LCSYS.2024.3414348.

[8] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Euclidean contractivity of neural networks with symmetric weights. 7:1724–1729, 2023. doi:10.1109/LCSYS.2023.3278250.

[9] V. Cerone, S. M. Fosson, S. Pirrera, and D. Regruto. A feedback control approach to convex optimization with inequality constraints. *arXiv preprint arXiv:2409.07168*, 2024.

[10] V. Cerone, S. M. Fosson, S. Pirrera, and D. Regruto. A new framework for constrained optimization via feedback control of Lagrange multipliers. *arXiv preprint arXiv:2403.12738*, 2024.

[11] S. Coogan. A contractive approach to separable Lyapunov functions for monotone systems. 106:349–357, 2019. doi:10.1016/j.automatica.2019.05.001.

[12] A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Time-varying convex optimization: A contraction and equilibrium tracking approach. June 2023. Conditionally accepted. doi:10.48550/arXiv.2305.15595.

[13] A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contraction analysis of continuous-time neural networks. 70(1), 2025. doi:10.1109/TAC.2024.3422217.

[14] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović. The proximal augmented Lagrangian method for nonsmooth composite optimization. 64(7):2861–2868, 2019. doi:10.1109/TAC.2018.2867589.

[15] A. Hauswirth, Z. He, S. Bolognani, G. Hug, and F. Dörfler. Optimization algorithms as robust feedback controllers. *Annual Reviews in Control*, 57:100941, 2024. doi:10.1016/j.arcontrol.2024.100941.

[16] H. D. Nguyen, T. L. Vu, K. Turitsyn, and J.-J. E. Slotine. Contraction and robustness of continuous time primal-dual dynamics. 2(4):755–760, 2018. doi:10.1109/LCSYS.2018.2847408.

[17] I. K. Ozaslan and M. R. Jovanović. On the global exponential stability of primal-dual dynamics for convex problems with linear equality constraints. pages 210–215, San Diego, USA, 2023. doi:10.23919/ACC55779.2023.10156504.

[18] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014. doi:10.1561/2400000003.

[19] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019. doi:10.1561/2200000073.

[20] G. Qu and N. Li. On the exponential stability of primal-dual gradient dynamics. 3(1):43–48, 2019. doi:10.1109/LCSYS.2018.2851375.

[21] R. Tyrrell Rockafellar. *Convex Analysis*. 1970.

[22] G. Russo, M. Di Bernardo, and E. D. Sontag. Global entrainment of transcriptional systems to periodic inputs. 6(4):e1000739, 2010. doi:10.1371/journal.pcbi.1000739.