Global Convergence of Control-Based Lagrangian Flows for Non-Convex Optimization

Global Convergence of Control-Based Lagrangian Flows for Non-Convex Optimization

Simone Pirrera, Francesco Ripa, Daniele Astolfi, Vito Cerone, Sophie M. Fosson, Diego Regruto

***Abstract*— This paper studies the continuous-time dynamics generated by control-theoretic Lagrangian methods for equality-constrained optimization. In particular, we consider dynamics induced by proportional-integral and feedback linearization controllers, which have recently been proposed as alternatives to primal-dual gradient methods. Unlike global convergence results for these dynamics, which rely on strong convexity of the objective function or boundedness assumptions, we exploit the geometric structure induced by the constraints. Specifically, we show global exponential convergence for non-convex problems that satisfy a suitable convexity property when restricted to the constraint manifold.**

***Index Terms*— Constrained optimization, Lagrangian methods, feedback linearization, convergence analysis.**

I. Introduction

CONSTRAINED optimization problems arise in a wide range of engineering applications, including machine learning and system identification [1], and are typically addressed via iterative algorithms. Such algorithms are either defined as discrete-time dynamical systems or obtained by discretizing continuous-time ones [2], [3]; continuous-time analysis provides valuable insights into stability, performance, and robustness, and can guide the design of novel discrete-time implementations [4]. A widely studied class of optimization dynamics is that of Lagrangian and primal-dual methods, which are derived from Lagrange multiplier theory. Among these, the primal-dual gradient dynamics (PDGD) has been extensively studied; see, e.g., [5]–[7]. An alternative perspective is to use control-theoretic tools to design optimization dynamics with rigorous convergence and performance guarantees. This approach has been applied to unconstrained optimization via integral quadratic constraints [8], [9]. In constrained optimization, the controller design regulates Lagrange multipliers to drive the system to an equilibrium satisfying first-order optimality conditions. We refer to the flows generated by the resulting dynamics as control-based Lagrangian flows.

Received XX xxxxx 2026; revised XX xxxxx 2026; accepted XX xxxxx 2026. Date of publication XX xxxx 2026

Funded by the European Union (MSCA Grant No. 101276493, Project SWELL). Research partially supported by ANR via ALLIGATOR project (ANR-22-CE48-0009-01). (Corresponding author: S. Pirrera).

S. Pirrera, F. Ripa, S. M. Fosson, V. Cerone and D. Regruto are with the Dipartimento di Automatica e Informatica, Politecnico di Torino; email: {name.surname}@polito.it.

D. Astolfi is with Université Claude Bernard Lyon 1, CNRS, LAGEPP UMR 5007, 43 boulevard du 11 novembre 1918, F-69100, Villeurbanne, France; email: daniele.astolfi@univ-lyon1.fr.

Among different control strategies, [10] and [11] propose the use of proportional-integral (PI) and feedback linearization (FL) controllers, while [12] employs control barrier functions and [13] investigates the use of anti-windup. In the context of time-varying optimization, the control approach has been adopted in [14] via robust control and in [15] using repetitive control tools. Many studies leverage convexity assumptions to simplify convergence analysis and establish global results [16].

This work focuses on the dynamics arising from PI and FL control. The PI dynamics is studied in [10] and [17], where global convergence is established under the assumption of a strongly convex objective function. Regarding the FL dynamics, [10] proves global convergence for strongly convex objective functions and local stability of the minima in the non-convex case, while the recent contribution [16] establishes global asymptotic convergence for non-convex problems under suitable boundedness assumptions. We aim to relax these assumptions by exploiting the geometric structure induced by the equality constraints. In particular, we establish global exponential convergence for a class of problems in which the objective function is not necessarily convex or bounded on the Euclidean space, but satisfies a suitable convexity property on the constraint manifold. This is the ambient-space counterpart of geodesic strong convexity, a notion central to Riemannian optimization [18], [19]. Such a property appears in the analysis of optimization methods defined on Riemannian manifolds. Methods including Riemannian gradient descent guarantee convergence when the objective is geodesically convex on the manifold. However, they operate directly on the manifold and do not naturally yield globally defined dynamical systems in Euclidean space. Similar convexity conditions are also encountered in the analysis of augmented Lagrangian methods (ALM), but are usually restricted to a neighbourhood of the optimal solution in the case of nonlinear constraints, e.g., [20].

Outline: Section II presents the problem formulation. Sections III and IV establish the global exponential convergence of the FL and PI dynamics, respectively, under the assumption of convexity on the constraint manifold. In Section V, we investigate the relationship between PI, FL, ALM and PDGD dynamics. Section VI presents a numerical example. Section VII concludes the paper.

Notation: Given $x \in \mathbb{R}^n$, we denote by $\|x\|$ its Euclidean norm. Given a function $g : \mathbb{R}^n \to \mathbb{R}^m$, we denote by $\nabla_x^\top g(x) \in \mathbb{R}^{m\times n}$ the Jacobian matrix of g and with $\nabla_x g(x)$ its transpose. Given a matrix $A \in \mathbb{R}^{n\times m}$, $\|A\|$ is the matrix norm induced by the Euclidean norm.

II. Problem statement and Background

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ be smooth functions. We consider the equality-constrained optimization problem

$$\min_{x\in\mathbb{R}^n} \; f(x) \qquad \text{s.t.} \quad h(x) = 0. \tag{1}$$

Throughout this paper, we make the following assumption.

Assumption 1. *The Jacobian matrix $\nabla_x^\top h(x) \in \mathbb{R}^{m\times n}$ is of full row rank and there exists a real constant $\underline{m} > 0$ such that $\nabla_x^\top h(x)\nabla_x h(x) \succeq \underline{m}I$ for all $x \in \mathbb{R}^n$.*

Assumption 1 is common in the constrained optimization literature; see, e.g., [2], [21], [22]. It is met in many applications, including a broad class of optimal control and system identification problems [1].

In this paper, we consider the problem of analyzing a class of continuous-time dynamical systems whose state trajectory converges to points $x^\star$ that satisfy the first-order optimality conditions of Problem (1).

Result 1 (First-order conditions [21], [23], [24])**.** *Let $x^\star \in \mathbb{R}^n$ be a local minimum of Problem* (1)*, and let Assumption 1 hold. Then, there exists a unique Lagrange multiplier vector $\lambda^\star \in \mathbb{R}^m$ such that*

$$\nabla_x f(x^\star) + \nabla_x h(x^\star)\lambda^\star = 0, \quad h(x^\star) = 0. \tag{2}$$

The continuous-time dynamics PDGD [5], described by

$$\dot{x} = -\nabla_x f(x) - \nabla_x h(x) z, \qquad \dot{z} = h(x), \tag{3}$$

is widely adopted to compute the stationary points of Problem (1). However, global exponential convergence of (3) is guaranteed only under the assumption that $f(x)$ is strongly convex on the entire ambient space $\mathbb{R}^n$ (see, e.g., [2], [7], [25]), while its trajectories may exhibit limit cycles [26] or instability [10] (see also Sec. VI) in non-convex settings.

From a control-theoretic perspective, (3) represents a feedback loop coupling the plant introduced in [10]

$$\dot{x} = -\nabla_x f(x) - \nabla_x h(x)\lambda, \qquad y = h(x), \tag{4}$$

with an integral controller driven by the constraint violation. Here, $\lambda \in \mathbb{R}^m$ is the control input and $y \in \mathbb{R}^m$ is the output; the integrator dynamics of z guarantee that any closed-loop equilibrium satisfies the feasibility condition $h(x) = 0$.

In this paper, we show that the control-based continuous-time algorithms introduced in [10] allow us to relax the requirement of global convexity on $f(x)$. Our global convergence analysis leverages a structural property of the optimization problem: there exists a suitable change of variables that removes the constraints and yields an equivalent strongly convex optimization problem.

To analyze the system dynamics within the constraint-induced coordinates, we introduce a global change of variables. Specifically, we suppose that there exists a smooth map $q : \mathbb{R}^n \to \mathbb{R}^{n-m}$ satisfying $Q(x) \doteq \nabla_x^\top q(x)\,\nabla_x q(x) \succeq \underline{q}\,I$, for some $\underline{q} > 0$ and all $x \in \mathbb{R}^n$, such that the map

$$x \mapsto \xi = \begin{bmatrix} y \\ \eta \end{bmatrix} \doteq \Phi(x) = \begin{bmatrix} h(x) \\ q(x) \end{bmatrix}, \tag{5}$$

with $y \in \mathbb{R}^m$ and $\eta \in \mathbb{R}^{n-m}$, is a global diffeomorphism of $\mathbb{R}^n$. Within this setting, we formalize our structural convexity assumption on Problem (1).

Assumption 2. *The feasible set $\Omega = \{x \in \mathbb{R}^n : h(x) = 0\}$ is connected, and the $\tilde{f} : \mathbb{R}^{n-m} \to \mathbb{R}$, defined by $\tilde{f}(\eta) \doteq f(\Phi^{-1}(0,\eta))$, is ρ_η-strongly convex on $\mathbb{R}^{n-m}$. That is, there exists a constant $\rho_\eta > 0$ such that, for all $\eta_1, \eta_2 \in \mathbb{R}^{n-m}$,*

$$\tilde{f}(\eta_2) \geq \tilde{f}(\eta_1) + \nabla_\eta^\top \tilde{f}(\eta_1)\,(\eta_2 - \eta_1) + \frac{\rho_\eta}{2}\|\eta_2 - \eta_1\|^2.$$

Remark 1. *Assumption 2 is weaker than convexity of Problem* (1) *(i.e., convexity of $f(x)$ in $\mathbb{R}^n$ and affine constraints). Under this assumption, Problem* (1) *admits a unique global optimizer $x^\star \in \mathbb{R}^n$ satisfying the conditions in Result 1.*

III. Stability analysis of FL dynamics

This section uses Assumption 2 to study the stability of the FL dynamics proposed in [10], i.e.,

$$\dot{x} = -\nabla_x f(x) - \nabla_x h(x)\lambda \tag{6a}$$
$$\lambda = [\nabla_x^\top h(x)\nabla_x h(x)]^{-1}(-\nabla_x^\top h(x)\nabla_x f(x) + \mathcal{G}(y)), \tag{6b}$$

where $\mathcal{G} : \mathbb{R}^m \to \mathbb{R}^m$ can be designed so that $\dot{y} = \mathcal{G}(y)$ is globally exponentially stable.

Under the diffeomorphism (5), system (4) takes the form:

$$\dot{y} = \nabla_x^\top h(x)\dot{x} = -\nabla_x^\top h(x)\nabla_x f(x) - \nabla_x^\top h(x)\nabla_x h(x)\lambda \tag{7a}$$
$$\dot{\eta} = \nabla_x^\top q(x)\dot{x} = -\nabla_x^\top q(x)\nabla_x f(x). \tag{7b}$$

Eq. (7b) follows from the fact that it is always possible to select $q(x)$ such that its Jacobian $\nabla_x^\top q(x)$ satisfies $\nabla_x^\top q(x)\nabla_x h(x) = 0$ for all $x \in \mathbb{R}^n$, thus making the dynamics independent of the input λ, as shown in [27].

Theorem 1 (Zero dynamics stability)**.** *Under Assumption 2, $\eta^\star = q(x^\star)$ is a global exponentially stable equilibrium for the zero dynamics associated with* (7)*, i.e., there exists a real constant $c_\eta \geq 1$ such that, for all $t \geq 0$ and all $\eta(0) \in \mathbb{R}^{n-m}$,*

$$\|\eta(t) - \eta^\star\| \leq c_\eta e^{-\underline{q}\rho_\eta t}\,\|\eta(0) - \eta^\star\|. \tag{8}$$

Proof. Define $p(\eta) = \Phi^{-1}(0,\eta)$ and $\tilde{f}(\eta) = f(p(\eta))$. We show that the zero dynamics of (7), i.e.,

$$\dot{\eta} = -\nabla_x^\top q(x)\,\nabla_x f(x), \qquad x = p(\eta), \tag{9}$$

is related to the negative gradient flow via $\dot{\eta} = -Q(p(\eta))\nabla_\eta \tilde{f}(\eta)$. Then, since $Q(x) \succeq \underline{q}I$, Assumption 2 implies global exponential stability of such a gradient flow with rate $\underline{q}\rho_\eta$ (see, e.g., [28]). We evaluate the gradient of $\tilde{f}(\eta) = f(p(\eta))$ by applying the chain rule:

$$\nabla_\eta \tilde{f}(\eta) = \nabla_\eta p(\eta)\,\nabla_x f(x)\big|_{x=p(\eta)} \tag{10}$$

Let us differentiate $p(\eta) = \Phi^{-1}(0,\eta)$ with respect to η:

$$\nabla_\eta^\top p(\eta) = \nabla_\xi^\top \Phi^{-1}(\xi)\big|_{\xi=(0,\eta)} \begin{bmatrix} 0 \\ I \end{bmatrix}. \tag{11}$$

By the inverse function theorem applied to the global diffeomorphism Φ, we obtain

$$\nabla_\xi^\top \Phi^{-1}(\xi) = \left(\nabla_x^\top \Phi(x)\right)^{-1} = \begin{bmatrix} \nabla_x^\top h(x) \\ \nabla_x^\top q(x) \end{bmatrix}^{-1} = \tag{12}$$

$$= \left[\nabla_x h(x)\,(\nabla_x^\top h(x)\,\nabla_x h(x))^{-1} \quad \nabla_x q(x)\,Q(x)^{-1}\right],$$

where the last equality uses the orthogonality condition $\nabla_x^\top q(x)\,\nabla_x h(x) = 0$, the definition of $Q(x)$, and the block-inverse formula. Combining Eqs. (11) and (12) yields

$$\nabla_\eta^\top p(\eta) = \nabla_x q(x)\,Q(x)^{-1}\big|_{x=p(\eta)} \tag{13}$$

Finally, substituting (13) into (10),

$$\nabla_\eta \tilde{f}(\eta) = Q(x)^{-1}\,\nabla_x^\top q(x)\,\nabla_x f(x)\big|_{x=p(\eta)} \tag{14}$$

so (9) reads $\dot{\eta} = -Q(p(\eta))\nabla_\eta \tilde{f}(\eta)$, concluding the proof. □

Using FL theory [27] and Theorem 1, we can prove global exponential convergence of the FL dynamics in Eq. (6) under the following additional assumptions.

Assumption 3. *The origin of $\dot{y} = \mathcal{G}(y)$ is globally exponentially stable, i.e., there exist $c_g \geq 1, \rho_g > 0$ such that*

$$\|y(t)\| \leq c_g e^{-\rho_g t}\,\|y(0)\|\,, \quad \forall y(0) \in \mathbb{R}^m, \forall\, t \geq 0. \tag{15}$$

Note that $\mathcal{G}$ can always be designed to satisfy this condition.

Assumption 4. *The global diffeomorphism defined in Eq.* (5) *and its inverse, $\Phi^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, are globally Lipschitz with constants L_Φ and L_Ψ, respectively.*

Assumption 5. *Let $\tilde{x} = \Phi^{-1}(y, \eta)$ and $p = \Phi^{-1}(0, \eta)$. The function $\nabla_x^\top q(x)\nabla f(x) : \mathbb{R}^n \to \mathbb{R}^{n-m}$ satisfies the following Lipschitz property for some real $L_1 > 0$:*

$$\left\|\nabla_x^\top q(\tilde{x})\nabla_x f(\tilde{x}) - \nabla_x^\top q(p)\nabla_x f(p)\right\| \leq L_1\,\|y\|\,. \tag{16}$$

Theorem 2. *Let Assumptions 1–5 hold. Then, the FL dynamics in Eq.* (6) *is globally exponentially stable. Specifically, for any rate $\rho_1 < \min\{\underline{q}\rho_\eta, \rho_g\}$, there exists a real $c_1 \geq 1$ such that, for all $x(0) \in \mathbb{R}^n$ and $t \geq 0$,*

$$\|x(t) - x^\star\| \leq c_1\, e^{-\rho_1 t}\,\|x(0) - x^\star\|. \tag{17}$$

Proof. By Assumption 4, since $x = \Phi^{-1}(y, \eta)$ and $x^\star = \Phi^{-1}(0, \eta^\star)$, we can bound the state error in terms of the normal form coordinates:

$$\|x - x^\star\| \leq L_\Psi \left\|\begin{bmatrix} y \\ \eta - \eta^\star \end{bmatrix}\right\| \leq L_\Psi\left(\|y\| + \|\eta - \eta^\star\|\right). \tag{18}$$

Thus, it suffices to establish exponential decay of both $\|y(t)\|$ and $\|\eta(t) - \eta^\star\|$. First, by Assumption 3, we have $\|y(t)\| \leq c_g e^{-\rho_g t}\|y(0)\|$. Next, we analyze the dynamics of η, which can be viewed as the zero dynamics of (4) perturbed by a vanishing term $\delta(t)$, i.e.,

$$\dot{\eta} = -Q(p)\nabla_\eta \tilde{f}(\eta) + \delta(t), \tag{19}$$

where $\tilde{f}(\eta)$ is defined in Theorem 1, and the perturbation term is given by $\delta(t) \doteq \nabla_x^\top q(p)\nabla_x f(p) - \nabla_x^\top q(x)\nabla_x f(x)$ with $p = \Phi^{-1}(0, \eta)$. To establish the exponential convergence of (19), we prove that it is input-to-state stable (ISS) with respect to the input $\delta(t)$, which decays exponentially to zero. Consider the ISS Lyapunov candidate $V(\eta) = \frac{1}{2}\|\eta - \eta^\star\|^2$. Using (19):

$$\dot{V} = -(\eta - \eta^\star)^\top Q(p)\nabla_\eta \tilde{f}(\eta) + (\eta - \eta^\star)^\top \delta(t). \tag{20}$$

Recalling $\tilde{f}$ is ρ_η-strongly convex and that $\nabla_\eta \tilde{f}(\eta^\star) = 0$, and since $Q(p) \succeq \underline{q}I$, we use the strong monotonicity property $-(\eta - \eta^\star)^\top Q(p)\nabla_\eta \tilde{f}(\eta) \leq -\underline{q}\rho_\eta\|\eta - \eta^\star\|^2 = -2\underline{q}\rho_\eta V$ to bound the first term. Next, applying in sequence the Cauchy–Schwarz and Young's inequalities to the second term, we obtain:

$$\dot{V} \leq -2\underline{q}\rho_\eta V + \|\eta - \eta^\star\|\|\delta(t)\| \tag{21a}$$

$$\leq (-2\underline{q}\rho_\eta + \epsilon)V + \frac{1}{2\epsilon}\|\delta(t)\|^2 \tag{21b}$$

for any $\epsilon \in (0, 2\underline{q}\rho_\eta)$, which is the required ISS dissipation inequality; see [29]. For any desired convergence rate $\rho_1 < \min\{\underline{q}\rho_\eta, \rho_g\}$, selecting ϵ sufficiently small ensures that $2\underline{q}\rho_\eta - \epsilon > 2\rho_1$ and $2\rho_g > 2\rho_1$. By Assumptions 3 and 5, $\delta(t)$ is exponentially bounded as:

$$\|\delta(t)\|^2 \leq L_1^2\|y(t)\|^2 \leq L_1^2 c_g^2 e^{-2\rho_g t}\|y(0)\|^2. \tag{22}$$

Plugging Eq. (22) in Eq. (21b) and applying the comparison lemma yields $V(t) \leq e^{-2\rho_1 t}\left(V(0) + M_\delta\|y(0)\|^2\right)$, for some positive constant M_δ. Taking the square root and using $\|\eta(t) - \eta^\star\| = \sqrt{2V(t)}$ we obtain the bound:

$$\|\eta(t) - \eta^\star\| \leq e^{-\rho_1 t}\left(\|\eta(0) - \eta^\star\| + \sqrt{2M_\delta}\|y(0)\|\right). \tag{23}$$

Finally, using the Lipschitz bounds from Assumption 4 on the initial conditions, $\|y(0)\| \leq L_h\|x(0) - x^\star\|$ and $\|\eta(0) - \eta^\star\| \leq L_q\|x(0) - x^\star\|$, we substitute (23) and the bound on $\|y(t)\|$ back into (18). This yields the desired result with $c_1 = L_\Psi(L_q + \sqrt{2M_\delta}L_h + c_g L_h)$. □

Remark 2. *Theorem 2 quantifies the convergence rate of the FL dynamics directly in terms of the problem data. In particular, we establish that ρ_1 is determined by the minimum of the two rates associated with the zero dynamics and the externally assigned y-dynamics induced by the choice of $\mathcal{G}$.*

A closely related analysis is addressed in the independent work [16]. While both works share similar technical assumptions regarding gradient boundedness (see Appendix II of the extended technical report [30]), they differ fundamentally in their core premises and resulting guarantees. Specifically, [16] relies on the assumption that $f(x)$ is lower-bounded and establishes asymptotic convergence to a first-order KKT point, which in the non-convex setting, might be a local minimum. In contrast, our work leverages the structural property in Assumption 2. This distinction is crucial, as it enables our framework to guarantee global exponential convergence to the unique global minimum. An illustrative example of a problem that fulfills Assumption 2, but $f(x)$ is not lower-bounded, is detailed in Sec. VI.

IV. Stability analysis of PI dynamics

This section studies the convergence of the PI dynamics proposed in [10] and described by the following system

$$\dot{x} = -\nabla_x f(x) - \nabla_x h(x)\,(k_p h(x) + k_i z)\,, \quad \dot{z} = h(x), \tag{24}$$

under Assumption 2.

We conduct this analysis by studying the system in the normal-form coordinates introduced in Sec. III. We start by rewriting (24) as:

$$\dot{x} = -\nabla_x f(x) - \nabla_x h(x)\lambda$$

$$\dot{\lambda} = -k_p \left(\nabla_x^\top h(x)\nabla_x h(x)\lambda + \nabla_x^\top h(x)\nabla_x f(x) + kh(x)\right),$$

where $k = k_i/k_p \in \mathbb{R}$. Next, we consider the global diffeomorphism defined by Eq. (5) and the coordinate transformation $w \doteq \lambda - k_p y - \lambda^\star$. Using these coordinates, we obtain

$$\begin{aligned} \dot{\eta} &= -\nabla_x^\top q(x)\nabla_x f(x) \\ \dot{y} &= -\nabla_x^\top h(x)\nabla_x f(x) - \nabla_x^\top h(x)\nabla_x h(x)(w + \lambda^\star) + \\ &\quad -k_p \nabla_x^\top h(x)\nabla_x h(x) y \\ \dot{w} &= k_p\, k\, y. \end{aligned} \tag{26}$$

We establish the convergence of (24) using Assumption 2 and the following additional assumption.

Assumption 6. *The function $\varphi(x) : \mathbb{R}^n \to \mathbb{R}^m$ defined by*

$$\varphi(x) \doteq -(\nabla_x^\top h(x)\nabla_x h(x))^{-1}(\nabla_x^\top h(x)\nabla_x f(x)) \tag{27}$$

is globally Lipschitz, i.e., there exists a real $L_2 > 0$ such that $\|\varphi(x_1) - \varphi(x_2)\| \leq L_2 \|x_1 - x_2\|$ for all $x_1, x_2 \in \mathbb{R}^n$.

The following lemma holds.

Lemma 1 (Lipschitz continuity of the nonlinear term r)**.** *Let $r(\eta, y) : \mathbb{R}^{n-m} \times \mathbb{R}^m \to \mathbb{R}^m$ be defined by:*

$$r(\eta, y) \doteq \nabla_x^\top h(\tilde{x})\nabla_x f(\tilde{x}) + \nabla_x^\top h(\tilde{x})\nabla_x h(\tilde{x})\lambda^\star \tag{28}$$

with $\tilde{x} = \Phi^{-1}([y^\top, \eta^\top]^\top)$. Then, there exist real constants $\ell_r^\eta, \ell_r^y > 0$ such that, for all $y \in \mathbb{R}^m$ and $\eta \in \mathbb{R}^{n-m}$, it holds

$$\|r(\eta, y)\| \leq \ell_r^\eta \|\eta - \eta^\star\| + \ell_r^y \|y\|. \tag{29}$$

Proof. Using $\lambda^\star = \varphi(x^\star)$ (see, e.g., [24]) and the definition of φ in Assumption 6, the residual can be rewritten as $r(\eta, y) = \nabla_x^\top h(\tilde{x})\nabla_x h(\tilde{x})\,[\varphi(x^\star) - \varphi(\tilde{x})]$; then Eq. (29) follows from Assumptions 4 and 6. □

Using Lemma 1, we can prove the main result of this section.

Theorem 3. *Let Assumptions 1, 2, and 4-6 hold. There exists a constant $k_p^\star$ such that, if $k_p > k_p^\star > 0$ and $0 < k_i < \underline{m}k_p^2$, then $(x^\star, \lambda^\star)$ is a globally exponentially stable equilibrium of the dynamics* (24)*, i.e., there exist real constants $c_\pi \geq 1, \rho_\pi > 0$ such that*

$$\left\| \begin{bmatrix} x(t) - x^\star \\ \lambda(t) - \lambda^\star \end{bmatrix} \right\| \leq c_\pi e^{-\rho_\pi t} \left\| \begin{bmatrix} x(0) - x^\star \\ \lambda(0) - \lambda^\star \end{bmatrix} \right\|. \tag{30}$$

Proof. Define $\tilde{\eta} = \eta - \eta^\star$ and $\zeta \doteq [y^\top, w^\top]^\top$. Consider the candidate Lyapunov function $V(\tilde{\eta}, \zeta) = \mu V_1(\tilde{\eta}) + V_2(\zeta)$, where $V_1 = \frac{1}{2}\|\tilde{\eta}\|^2$ and

$$V_2(\zeta) = \zeta^\top \Pi \zeta, \qquad \Pi = \begin{bmatrix} I & k_p^{-1} I \\ k_p^{-1} I & \theta k_p^{-1} I \end{bmatrix} \tag{31}$$

for some $\mu > 0$ and $\theta > 0$ to be chosen. Note that Π is positive definite for any $k_p > 1/\theta$ (this can be shown using the Schur complement with respect to the upper-left block). Hence, V is a valid quadratic Lyapunov function for the state $(\tilde{\eta}^\top, \zeta^\top)^\top$. Now let $\tilde{x} = \Phi^{-1}(y, \eta)$ and $p = \Phi^{-1}(0, \eta)$. Adding and subtracting $\nabla_x^\top q(p)\nabla_x f(p)$, we obtain

$$\begin{aligned} \dot{V}_1 &= -\tilde{\eta}^\top \left(\nabla_x^\top q(p)\nabla_x f(p)\right) + \\ &\quad + \tilde{\eta}^\top \left(\nabla_x^\top q(p)\nabla_x f(p) - \nabla_x^\top q(\tilde{x})\nabla_x f(\tilde{x})\right). \end{aligned} \tag{32}$$

From Theorem 1, the term $-\tilde{\eta}^\top(\nabla_x^\top q(p)\nabla_x f(p))$ is bounded by $-\underline{q}\rho_\eta \|\tilde{\eta}\|^2$, while, by Assumption 5, the term $\tilde{\eta}^\top \left(\nabla_x^\top q(p)\nabla_x f(p) - \nabla_x^\top q(\tilde{x})\nabla_x f(\tilde{x})\right)$ is bounded by $L_1\|\tilde{\eta}\|\,\|y\|$. Using Young's inequality, we obtain

$$\dot{V}_1 \leq -\frac{1}{2}\underline{q}\rho_\eta \|\tilde{\eta}\|^2 + \frac{L_1^2}{2\underline{q}\rho_\eta}\|y\|^2. \tag{33}$$

Define $H(\eta, y) \doteq \nabla_x^\top h(\tilde{x})\nabla_x h(\tilde{x})$. Then, (26) is recast to

$$\begin{aligned} \dot{y} &= -r(\eta, y) - H(\eta, y)w - k_p H(\eta, y) y, \\ \dot{w} &= k_p k y, \end{aligned} \tag{34}$$

with r defined in Lemma 1. In compact form, we obtain

$$\dot{\zeta} = \underbrace{\begin{bmatrix} -k_p H(\eta, y) & -H(\eta, y) \\ k_p k I & 0 \end{bmatrix}}_{N(\eta, y)} \zeta + \begin{bmatrix} -r(\eta, y) \\ 0 \end{bmatrix}.$$

Therefore, the time derivative of V_2 is

$$\dot{V}_2 = -\zeta^\top \Psi(\eta, y)\zeta - 2\zeta^\top \Pi \begin{bmatrix} r(\eta, y) \\ 0 \end{bmatrix},$$

where $\Psi(\eta, y) \doteq -\big(\Pi N(\eta, y) + N(\eta, y)^\top \Pi\big)$. By explicitly computing the matrix products, we have

$$\Psi(\eta, y) = \begin{bmatrix} 2k_p H(\eta, y) - 2kI & 2H(\eta, y) - \theta k I \\ \star & 2k_p^{-1} H(\eta, y) \end{bmatrix}$$

Equivalently, $\Psi(\eta, y) = 2D\overline{\Psi}D$, with

$$D = \begin{bmatrix} \sqrt{k_p}\, I & 0 \\ 0 & \frac{1}{\sqrt{k_p}} I \end{bmatrix}, \quad \overline{\Psi} = \begin{bmatrix} H - \frac{k}{k_p} I & H - \frac{1}{2}\theta k I \\ \star & H \end{bmatrix},$$

where we omitted the argument (η, y) in H for brevity. To guarantee that $\overline{\Psi} \succ 0$, we apply the Schur complement with respect to the lower-right block $H \succeq \underline{m} I \succ 0$. The Schur complement S is

$$\begin{aligned} S &= (H - \tfrac{k}{k_p} I) - \big(H - \tfrac{1}{2}\theta k I\big) H^{-1} \big(H - \tfrac{1}{2}\theta k I\big) \\ &= H - \tfrac{k}{k_p} I - \left(H - \theta k I + \tfrac{\theta^2 k^2}{4} H^{-1}\right) \\ &= k\left(\theta - \tfrac{1}{k_p}\right) I - \tfrac{\theta^2 k^2}{4} H^{-1}. \end{aligned}$$

Since $H^{-1} \preceq \underline{m}^{-1} I$, we can lower-bound S as

$$S \succeq \left[k\left(\theta - \frac{1}{k_p}\right) - \frac{\theta^2 k^2}{4\underline{m}}\right] I \succeq \left(\underline{m} - \frac{k}{k_p}\right) I,$$

where we selected $\theta = 2\underline{m}/k$. By choosing $k_p > k/\underline{m} > 1/\theta = k/(2\underline{m})$, which is equivalent to the condition $k_i < \underline{m}k_p^2$, we obtain $S \succeq c_s I \succ 0$ with $c_s = \underline{m} - k/k_p > 0$. Notice that this combined choice of θ and k_p also ensures that $\Pi \succ 0$. Consequently, there exists a constant $\overline{c}_H > 0$ such that $\overline{\Psi} \succeq \overline{c}_H I$. Using $\Psi(\eta, y) = 2D\overline{\Psi}D$, this implies

$$\Psi \succeq 2\overline{c}_H D^2 = 2\overline{c}_H \begin{bmatrix} k_p I & 0 \\ 0 & k_p^{-1} I \end{bmatrix}. \tag{35}$$

Therefore, the quadratic term is bounded as:

$$-\zeta^\top \Psi \zeta \leq -2\overline{c}_H k_p \|y\|^2 - \frac{2\overline{c}_H}{k_p}\|w\|^2. \tag{36}$$

Next, we bound the cross-term

$$-2\zeta^\top \Pi \begin{bmatrix} r(\eta, y) \\ 0 \end{bmatrix} = -2y^\top r(\eta, y) - \frac{2}{k_p} w^\top r(\eta, y). \quad (37)$$

Recalling from Lemma 1 that $r(\eta, y)$ is globally Lipschitz, i.e., $\|r(\eta, y)\| \leq \ell_r^\eta \|\tilde{\eta}\| + \ell_r^y \|y\|$, we can bound the terms using Young's inequality:

$$\begin{aligned} \left|2y^\top r(\eta, y)\right| &\leq (2\ell_r^y + \ell_r^\eta) \|y\|^2 + \ell_r^\eta \|\tilde{\eta}\|^2, \\ \frac{2}{k_p}\left|w^\top r(\eta, y)\right| &\leq \frac{2(\ell_r^y)^2}{k_p \bar{c}_H} \|y\|^2 + \frac{\bar{c}_H}{k_p} \|w\|^2 + \frac{2(\ell_r^\eta)^2}{k_p \bar{c}_H} \|\tilde{\eta}\|^2. \end{aligned}$$

Combining all the previous bounds and assuming $k_p \geq 1$, we obtain the following bound for $\dot{V}_2$:

$$\dot{V}_2 \leq -(2\bar{c}_H k_p - C_y) \|y\|^2 - \frac{\bar{c}_H}{k_p} \|w\|^2 + C_\eta \|\tilde{\eta}\|^2, \quad (38)$$

where $C_y \doteq 2\ell_r^y + \ell_r^\eta + \frac{2(\ell_r^y)^2}{\bar{c}_H}$ and $C_\eta \doteq \ell_r^\eta + \frac{2(\ell_r^\eta)^2}{\bar{c}_H}$. Consequently, the derivative of the Lyapunov function satisfies

$$\begin{aligned} \dot{V} \leq &- \left(2\bar{c}_H k_p - C_y - \mu \frac{L_1^2}{2\underline{q}\rho_\eta}\right) \|y\|^2 - \frac{\bar{c}_H}{k_p} \|w\|^2 + \\ &- \left(\frac{\mu \underline{q} \rho_\eta}{2} - C_\eta\right) \|\tilde{\eta}\|^2. \end{aligned}$$

By taking $\mu > 2C_\eta / (\underline{q}\rho_\eta)$ and

$$k_p > \max\left\{1, \frac{k}{\underline{m}}, \frac{1}{2\bar{c}_H}\left(C_y + \mu \frac{L_1^2}{2\underline{q}\rho_\eta}\right)\right\} \doteq k_p^\star, \quad (39)$$

we obtain $\dot{V} \leq -\gamma\big(\|\tilde{\eta}\|^2 + \|y\|^2 + \|w\|^2\big)$ for some constant $\gamma > 0$. This ensures the global exponential stability of the equilibrium $(\eta^\star, 0, 0)$ of the system in the normal-form coordinates. As previously stated, the global exponential convergence carries over to the original (x, λ) coordinates by virtue of the globally Lipschitz diffeomorphism Φ^{-1}. □

Assumptions 4, 5, and 6 are Lipschitz-type regularity conditions, which are required to establish the global exponential convergence results in Theorems 2 and 3 over $\mathbb{R}^n$. To verify these assumptions globally, sufficient conditions are that $\nabla_x f$ and $\nabla_x h$ are globally bounded and Lipschitz; see Appendix II of the extended version [30] for detailed derivations.

However, in a practical setting, optimization algorithms operate within bounded regions. Given any bounded set of initial conditions, one can define a sufficiently large compact sublevel set of the Lyapunov function that completely encloses them. Since $f, h \in C^2$, the Lipschitz conditions are automatically satisfied by restriction on any such compact set. Consequently, by using the local Lipschitz constants on this set, one can determine a valid $k_p^\star$ that makes the Lyapunov derivative strictly negative. This guarantees that the compact set is forward-invariant, and the exact same proof steps hold to conclude semi-global exponential stability counterparts of Theorems 1 and 3. Furthermore, since the local Lipschitz constants are generally much smaller than their global counterparts, we can significantly reduce the conservativeness of the closed-form expression of $k_p^\star$, which is provided in Appendix I of the extended version [30].

Our derivation of explicit tuning bounds distinguishes this work from singular perturbation approaches like [17], which does not provide a specific lower bound for k_p. The result in [10] differs structurally from Theorem 3 as it imposes a lower bound on k_i rather than an upper one.

V. Further comments

In this section, we investigate the relationship between the PI dynamics in (24) with continuous-time ALM [20], the FL dynamics in (6), and PDGD.

ALM can be framed as a particular case of PI dynamics (24). By applying the primal-dual flow to the quadratically augmented Lagrangian $\mathcal{L}_w(x, \lambda) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, defined as $\mathcal{L}_w(x, \lambda) \doteq f(x) + \lambda^\top h(x) + \frac{w}{2}\|h(x)\|^2$ yields the PI dynamics (24) under the choices $k_p = w$, $k_i = 1$, and $z = \lambda$. Previous works show that ALM converges locally for a sufficiently large w under Assumption 2. However, under nonlinear constraints, convergence is limited to a neighborhood of the optimum. In contrast, we establish global exponential convergence for ALM as a corollary of Theorem 3, provided that the penalty parameter is chosen as $w \geq k_p^\star$.

Second, we note that the PI dynamics (25) admits an interpretation as a gradient-flow approximation of the FL dynamics (6), yielding a more efficient implementation achieving the same optimal solution. The control law (6b) can be written equivalently as

$$\lambda(x) = \arg\min_{\sigma \in \mathbb{R}^m} \|A(x)\sigma - b(x)\|^2, \quad (40)$$

where $A(x) \doteq \nabla_x^\top h(x) \nabla_x h(x)$ and $b(x) \doteq -\nabla_x^\top h(x) \nabla_x f(x) + \mathcal{G}(h(x))$. Under Assumption 1, $A(x) \succ 0$, and (40) admits a unique minimizer. The negative gradient flow that asymptotically tracks this minimizer is

$$\dot{\sigma} = -P_0(x) \nabla_\sigma \|A(x)\sigma - b(x)\|^2, \quad (41)$$

with $P_0 = P_0^\top \succ 0$ chosen arbitrarily. Choosing $\mathcal{G}(y) = -ky$ with $k \in \mathbb{R}^+$ and $P_0 = \frac{\alpha}{2}(\nabla_x^\top h(x) \nabla_x h(x))^{-1}$ recasts (41) as $\dot{\sigma} = -\alpha(A(x)\sigma - b(x))$, which, with the identifications $k_p = \alpha$ and $\lambda = \sigma$, yields (25).

This interpretation resembles the construction in [17] for variational inequalities. We also remark that solving (6b) is the most expensive step of the FL dynamics, as it requires solving a linear system of size m; the PI dynamics bypasses this inversion, and the approximation improves as $k_p = \alpha$ increases and the gradient flow becomes faster.

Finally, PDGD coincides with the pure integral action of PI ($k_p = 0$, $k_i = 1$): no choice of $\alpha > 0$ and $k > 0$ in the gradient-flow construction recovers it, since $k_p = \alpha = 0$ forces $k_i = k\alpha = 0$ for all $k \in \mathbb{R}$. Therefore, PDGD (whose global exponential stability has been established for strongly convex problems in [7]) may fail to converge even if the optimization problem satisfies the assumptions considered in this paper, as shown in the illustrative example in Sec. VI.

VI. Illustrative Example

We consider a two-dimensional example to illustrate a scenario in which Assumption 2 is satisfied despite $f(x)$ being non-convex, $h(x)$ being nonlinear, and the two functions being nontrivially related. Specifically

$$f(x) = (x_1^2 - x_2^2 - 1)^2 + 0.7(x_1^3 + x_1 x_2^2) + 70e^{-2(x_1-1)^2 - 2x_2^2}$$

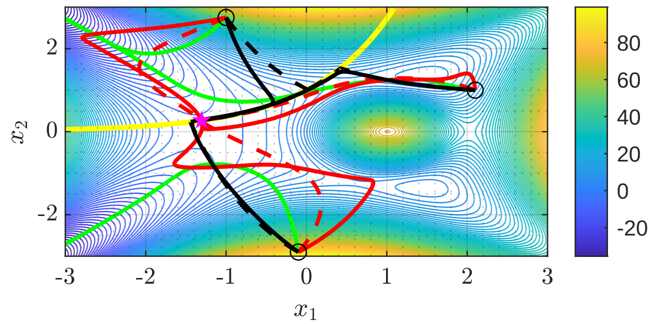

Fig. 1. Trajectories generated by PI and FL for the problem in Sec. VI. Level lines of $f(x)$ and global optimal solution ($\star$). Trajectories generated by FL with $k = 1$ (—), FL with $k = 10$ (- - -), PI with $k_p = 100$ (—), PI with $k_p = k_p^\star \approx 1.94 \times 10^{15}$ (- - -), and PDGD (—). All trajectories are initialized at points ($\circ$).

and $h(x) = x_2 - e^{x_1}$. The function $f(x)$ has three local minima, one local maximum, and one saddle point. Assumption 1 holds since $\nabla_x h(x) = [-e^{x_1}, 1]^\top \neq 0$ for all $x \in \mathbb{R}^2$. $f(x)$ is strongly convex when restricted to $\{x : h(x) = 0\}$, thereby satisfying Assumption 2. Moreover, as discussed in Sec. IV, the regularity Assumptions 4, 5, and 6 are verified on any compact forward-invariant set containing the initial conditions. Consequently, in this specific illustrative example, the developed theoretical results hold in a semiglobal sense. Fig. 1 shows the trajectories generated by FL dynamics in Eq. (6) and PI dynamics in Eq. (24) for different initial conditions and $k_i = 1$. For the PI scheme, two values of k_p are considered: the conservative threshold $k_p^\star$ provided by the explicit closed-form bound derived in Appendix I of the extended report [30] on the compact set $\mathcal{K}$, and a smaller, less conservative value. We observe that, as expected from the theoretical results established in this work, all trajectories converge to the unique global optimal solution of the optimization problem. Conversely, the trajectories generated by PDGD, initialized at the same points as for (24), diverge.

VII. Conclusion

We establish global exponential convergence of the proportional-integral and feedback linearization dynamics for a class of equality-constrained non-convex optimization problems. While previous works assume strong convexity in the ambient space, this work extends the analysis to a class of non-convex problems that satisfy a structural geometric property induced by the equality constraints. Moreover, we show that the proportional-integral dynamics is a lower-complexity approximation of the feedback-linearization dynamics that yields the same optimal solution. Future work will focus on overconstrained optimization and discrete-time models.

References

[1] V. Cerone, S. M. Fosson, S. Pirrera, and D. Regruto, "A constrained optimization approach to nonlinear system identification through simulation error minimization," *arXiv*, no. 2509.01461, 2025.

[2] M. Bin, I. Notarnicola, and T. Parisini, "Semiglobal exponential stability of the discrete-time Arrow-Hurwicz-Uzawa primal-dual algorithm for constrained optimization," *Math. Program.*, vol. 208, no. 1, pp. 629–660, 2024.

[3] S. Kelly and J. W. Simpson-Porco, "An interconnected systems approach to convergence analysis of discrete-time primal-dual algorithms," in *Proc. Amer. Control Conf. (ACC)*, 2024, pp. 662–668.

[4] W. Su, S. Boyd, and E. J. Candès, "A differential equation for modeling nesterov's accelerated gradient method: Theory and insights," *J. Mach. Learn. Res.*, vol. 17, no. 153, pp. 1–43, 2016.

[5] 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.

[6] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Syst. Control Lett.*, vol. 87, pp. 10–15, 2016.

[7] 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.

[8] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J. Optim.*, vol. 26, no. 1, pp. 57–95, 2016.

[9] C. W. Scherer and C. Ebenbauer, "A tutorial on convex design of optimization algorithms by integral quadratic constraints," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 9, 2025.

[10] V. Cerone, S. M. Fosson, S. Pirrera, and D. Regruto, "A new framework for constrained optimization via feedback control of Lagrange multipliers," *IEEE Trans. Autom. Control*, vol. 70, no. 11, pp. 7141–7156, 2025.

[11] ——, "A feedback control approach to convex optimization with inequality constraints," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2024, pp. 2538–2543.

[12] A. Allibhoy and J. Cortés, "Control-Barrier-Function-Based Design of Gradient Flows for Constrained Nonlinear Programming," *IEEE Trans. Autom. Control*, vol. 69, no. 6, pp. 3499–3514, 2024.

[13] A. Hauswirth, F. Dörfler, and A. Teel, "On the differentiability of projected trajectories and the robust convergence of non-convex anti-windup gradient flows," *IEEE Control Syst. Lett.*, vol. 4, no. 3, pp. 620–625, 2020.

[14] U. Casti, N. Bastianello, R. Carli, and S. Zampieri, "A control theoretical approach to online constrained optimization," *Automatica*, vol. 176, p. 112107, 2025.

[15] D. Astolfi and C. M. Verrelli, "Repetitive control tools for an original approach to convex optimization problems under affine periodic constraints," *Syst. Control Lett.*, vol. 201, p. 106095, 2025.

[16] R. Zhang, A. Raghunathan, J. Shamma, and N. Li, "Constrained optimization from a control perspective via feedback linearization," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 38, pp. 55 944–55 980, 2026.

[17] A. Allibhoy and J. Cortés, "Anytime solvers for variational inequalities: The (recursive) safe monotone flows," *Automatica*, vol. 177, p. 112287, 2025.

[18] H. Zhang and S. Sra, "First-order methods for geodesically convex optimization," in *Conference on Learning Theory (COLT)*. PMLR, 2016, pp. 1617–1638.

[19] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.

[20] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Belmont, MA: Athena Scientific, 1996.

[21] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 2006.

[22] J. Schropp and I. Singer, "A dynamical systems approach to constrained minimization," *Numer. Funct. Anal. Optim.*, vol. 21, pp. 537–551, 2000.

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

[24] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Massachusetts: Athena Scientific, 1999.

[25] A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo, "Time-varying convex optimization: A contraction and equilibrium tracking approach," *IEEE Trans. Autom. Control*, vol. 70, no. 11, pp. 7446–7460, 2025.

[26] T. Lin, C. Jin, and M. Jordan, "On gradient descent ascent for nonconvex-concave minimax problems," in *Int. Conf. Mach. Learn. (ICML)*. PMLR, 2020, pp. 6083–6093.

[27] A. Isidori, *Nonlinear Control Systems*. Springer London, 1995.

[28] F. Bullo, *Contraction theory for dynamical systems*. Francesco Bullo, 2022.

[29] E. D. Sontag, *Mathematical control theory: deterministic finite dimensional systems*. Springer Science & Business Media, 2013, vol. 6.

[30] S. Pirrera, F. Ripa, D. Astolfi, V. Cerone, S. M. Fosson, and D. Regruto, "Global convergence of control-based Lagrangian flows for non-convex optimization," 2026, arxiv.org/abs/2605.22486.