2025 IEEE 64th Conference on Decision and Control (CDC)
December 10-12, 2025. Rio de Janeiro, Brazil

Remarks on the Polyak-Łojasiewicz inequality and the convergence of gradient systems

Arthur C. B. de Oliveira[1], Leilei Cui[2], and Eduardo D. Sontag[1,3]

2025 IEEE 64th Conference on Decision and Control (CDC) | 979-8-3315-2627-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CDC57313.2025.11312036

***Abstract*— This work explores generalizations of the Polyak-Łojasiewicz inequality (PŁI) and their implications for the convergence behavior of gradient flows in optimization problems. Motivated by the continuous-time linear quadratic regulator (CT-LQR) policy optimization problem – where only a weaker version of the PŁI is characterized in the literature – this work shows that while weaker conditions guarantee global convergence of gradient flows and optimality at the critical points of the cost function, the trajectory "profile" of the solutions may differ considerably depending on which "flavor" of inequality the cost satisfies. After a general theoretical analysis, we focus on fitting the CT-LQR policy optimization problem to the proposed framework, showing that, in fact, it can never satisfy a PŁI in its strongest form. We finish our analysis with a brief discussion on the difference between continuous- and discrete-time LQR policy optimization, proposing some intuition to explain why one satisfy a PŁI while the other does not.**

I. Introduction

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have rekindled interest in optimization theory, with many traditional results being revisited in light of the proposed techniques [1]–[8]. In particular, the typical model-free formulation of many successful learning techniques motivates the study of gradient-based optimization methods, which are invaluable in understanding the training of neural networks and similar architectures, typically done through back-propagation algorithms.

Gradient descent or, in continuous time, gradient flow, consists in searching for the argument x that minimizes the value of a given function $f(x)$ by "moving along" the direction of steepest descent of the cost function. Theoretical guarantees are typically desirable, and in search of balancing generality and good properties, often in the optimization literature one deals with specific classes of problems, such as convex optimization [9] or linear programming [10]. In this paper, we will focus on optimization problems that satisfy (to different degrees) a Polyak-Łojasiewicz inequality (PŁI), also known as the gradient dominance condition [11], [12].

The PŁI is a staple in nonlinear optimization analysis, as, in its strongest form, it guarantees global exponential convergence of the gradient flow to the optimal value of the cost [12]. Furthermore, satisfying a PŁI globally (gPŁI) also guarantees strong robustness properties [2]. However, characterizing such a condition might not be possible for every optimization problem. In [1] the authors noticed that more general conditions than the gPŁI can be proposed by using different classes of comparison functions so that different robustness results can be guaranteed.

In particular, the problem of policy optimization for the linear quadratic regulator (LQR) motivates the discussion around weaker versions of the PŁI [1], [2], [6], [13]–[18]. For the discrete-time version of the problem, in [14]–[16] the authors show that it satisfies a gPŁI, guaranteeing exponential convergence to the optimal feedback law for initialization in the stabilizing set of feedback matrices. However, so far in the literature for the continuous-time LQR policy optimization problem there is no characterization of a gPŁI [6], [17], with the analysis in [6] indicating that, at least for the scalar case, the continuous-time LQR does not satisfy a gPŁI.

In this work, we are interested in characterizing how generalizations of the PŁI affect the rate of convergence of the solution. We begin in Section II by revisiting common assumptions and their consequences regarding the convergence of the gradient flow. We then formally introduce the gPŁI and another weaker definition, and discuss their differences and consequences to the trajectory of the cost along a gradient flow solution. We next deepen the analysis by defining a new family of conditions closely related to, but more general than, the global PŁI. We discuss how these weaker conditions relate to each other and how they can characterize weaker forms of convergence than the gPŁI. Then, in Section III, we contextualize the theoretical analysis of this paper through the specific problem of the continuous-time LQR policy optimization problem. This problem has no guarantees of satisfying a gPŁI, and in fact we show that it can never satisfy such a condition. We characterize which sequences of points within the policy space result in an unbounded value for the gradient of the cost, and which result in a "sub-exponential" convergence profile for the solution. We follow up with a brief discussion on the difference between the continuous- and discrete-time LQR policy optimization, and finalize the paper in Section IV.

All proofs are omitted due to space constraints but are available in the extended version of this paper on arXiv [19].

II. Theoretical Setup

Along this paper, let $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the sets of real, non-negative real, and strictly positive real numbers, respectively. Let $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ be the sets of real symmetric

The work of EDS and ACO was partially supported by grants AFOSR FA9550-22-1-0316, 21-1-0289, and ONR N00014-21-1-2431

[1]Department of Electrical and Computer Engineering, Northeastern University, USA a.castello@northeastern.edu
[2]Department of Mechanical Engineering, University of New Mexico, Albuquerque, USA lcui@unm.edu
[3]Department of BioEngineering, Northeastern University, USA e.sontag@northeastern.edu

positive semi-definite and positive definite n-by-n matrices, respectively.

A given function $\alpha : \mathbb{R}_+ \to \mathbb{R}$ is said to be *positive-definite* ($\mathcal{PD}$) if $\alpha(0) = 0$ and $\alpha(x) > 0$ for all $x \neq 0$. Similarly, α is said to be of class $\mathcal{K}$ if it is continuous, positive-definite, and strictly increasing. Finally, α is of class $\mathcal{K}_\infty$ if it is of class $\mathcal{K}$ and unbounded.

For a given function $f : \mathcal{X} \to \mathbb{R}$ bounded below, let $\underline{f} = \inf_{x \in \mathcal{X}} f(x)$, and $x^* = \arg\inf_{x \in \mathcal{X}} f(x)$.

A. Optimization problems and gradient methods

Let $\mathcal{X}$ be an open subset of an Euclidean space (the analysis in this paper can be generalized to manifolds, but we refrain from it for simplicity). Then, an optimization problem consists of searching for the value of an argument/parameter $x \in \mathcal{X}$ that minimizes some cost function $f : \mathcal{X} \to \mathbb{R}$ (or minimizes the negative of a reward for maximization). Mathematically, we write such a problem as

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \mathcal{X} \end{array}, \tag{1}$$

which might have one, multiple, or no solution, requiring some assumptions about either the cost f or the search space $\mathcal{X}$ to guarantee the existence and uniqueness of the solution. A common assumption is that of compactness of $\mathcal{X}$, and continuity of $f(x)$ for $x \in \mathcal{X}$, which would guarantee the existence of a minimum (and maximum) in $\mathcal{X}$ since $f(\mathcal{X}) := \{f(x) \mid x \in \mathcal{X}\}$ would be compact. Usually, however, the search space $\mathcal{X}$ is not compact, requiring the adoption of an alternative set of assumptions, outlined next.

Assumption 1: The function f is *real analytic*, *bounded below*, and *proper* (*i.e.* coercive).

It is easy to prove that Assumption 1 guarantees the existence of a minimum $\underline{f} \in f(\mathcal{X})$ attained at a set of points $\mathcal{T} := \{x \in \mathcal{X} \mid f(x) = \underline{f}\}$. Note that, from the point of view of the optimization problem, any $x \in \mathcal{T}$ constitutes a valid solution of (1), since all such points yield the same value of the cost function. Then, the optimization problem can be thought of as finding any $x \in \mathcal{T}$.

A natural candidate for solutions to the optimization problem is the set of critical points of f, i.e. the set of points $\mathcal{Z} := \{x \in \mathcal{X} \mid \nabla f(x) = 0\}$. A common strategy for finding an $x \in \mathcal{Z}$ is "moving the parameters along the direction of steepest descent of the function". Mathematically and in continuous-time, this means imposing the following dynamics for the parameter x

$$\dot{x} = -\nabla f(x), \tag{2}$$

while in discrete time one would impose the following update law for x_k for a small enough $h > 0$

$$x_{k+1} = x_k - h\nabla f(x_k). \tag{3}$$

In this paper we focus on the continuous-time strategy, and one can easily verify that x is an equilibrium of (2) if and only if $x \in \mathcal{Z}$. Nonetheless, there is no a priori guarantee that a solution of (2) initialized in $\mathcal{X}$ will converge to a point in $\mathcal{Z}$, much less in $\mathcal{T}$. Therefore, we next examine the convergence guarantees that can be derived under Assumption 1, and explore additional assumptions that may strengthen these guarantees.

B. Convergence guarantees and the Polyak-Łojasiewicz inequality

Consider an optimization problem (1) satisfying Assumption 1, then Łojasiewicz's theorem [20] guarantees that any solution of (2) initialized in $\mathcal{X}$ will converge to a critical point of f (precompactness of trajectories follows from properness of the cost function). Despite that, $x \in \mathcal{Z}$ is only a necessary condition for the optimality of x. In fact a point $x \in \mathcal{Z}$ can be either a local minimum, a local maximum, or a saddle-point of f. Regarding that, the following result can be stated ([3], [4] appendix A):

Lemma 1 ([3], [4]): Let

$$\mathcal{S} := \{x \in \mathcal{Z} \mid \exists v \in \mathbb{R}^n \text{ s.t. } v^\top \nabla^2 f(x) v < 0\},$$

where $\nabla^2 f$ is the Hessian of f, and let $\phi : \mathbb{R}_+ \times \mathcal{X} \to \mathcal{X}$ be the solution of the initial value problem (2). Then, the set of $x_0 \in \mathcal{X}$ for which $\lim_{t\to\infty} \phi(t, x_0) \in \mathcal{S}$ has Lebesgue measure zero. In other words, the center-stable manifold of $\mathcal{S}$ has measure zero.

Notice that Lemma 1 holds even if $\mathcal{S}$ is a continuous set, or the union of continuous sets. In fact, $\mathcal{S}$ need not be even compact, as long as the condition on the value of the Hessian holds for all of its elements. However, despite this result excluding any local maxima and "strict" saddles from the result of a gradient flow solution to problem (1) (with probability one), it is still not enough to guarantee the optimality of the gradient flow solution. Typically, additional assumptions are introduced in the literature to ensure that $\lim_{t\to\infty} \phi(t, x_0) \in \mathcal{T}$, with convexity of f arguably being one of the most common. If f is convex in $\mathcal{X}$, then $\mathcal{Z} = \mathcal{T}$ and a gradient flow will eventually find the optimal solution. Furthermore, if f is strongly convex, then a solution of (2) converges to $\mathcal{T}$ exponentially.

In this paper, we first review a condition for f that is weaker than strong convexity but still ensures that a solution of (2) converges to $\mathcal{T}$ exponentially (in terms of the cost).

Definition 1 (μ-global Polyak-Łojasiewicz inequality): Given a fixed $\mu > 0$, a function f satisfies a *μ-global Polyak-Łojasiewicz inequality* (μ-gPŁI) if

$$\|\nabla f(x)\| \geq \alpha\left(f(x) - \underline{f}\right), \tag{4}$$

with $\alpha(r) = \sqrt{\mu r}$, for all $x \in \mathcal{X}$. Furthermore, f satisfies a gPŁI if the μ-gPŁI holds for some $\mu > 0$.

The property in Definition 1 (often written in the form $\|\nabla f(x)\|^2 \geq \mu\left(f(x) - \underline{f}\right)$) has been the object of much recent study, and is a natural generalization of convexity (see [12] for a thorough analysis of the relationship between convexity and the gPŁI). An immediate consequence is that all critical points of f must solve (1), *i.e.* if f is gPŁI, then $\mathcal{Z} = \mathcal{T}$, which guarantees the optimality of gradient flow solutions. Further usefulness of this property lies in the fact that it guarantees exponential convergence of the cost along any solution of the gradient flow, as we show next.

Definition 2 (μ-global exponential cost stability): Given a fixed $\mu > 0$, the gradient flow (2) of f is μ-globally exponential cost stable (μ-GECS) if

$$f(\phi(t, x_0)) - \underline{f} \leq (f(x_0) - \underline{f})e^{-\mu t} \text{ for all } t \geq 0. \quad (5)$$

Furthermore, the gradient flow of f is GECS if it is μ-GECS for some $\mu > 0$.

Lemma 2: The gradient flow (2) of f is μ-GECS if and only if f satisfies a μ-global PŁI.

Despite the favorable convergence properties guaranteed by the existence of an exponential upper-bound, proving that a cost function satisfies the μ-global PŁI is not always possible. In some relevant examples in the literature, the following weaker condition is characterized instead.

Definition 3 (Semi-global PŁI): A function f satisfies a *semi-global Polyak-Łojasiewicz inequality* (*sg*PŁI) if, for every $\epsilon > 0$, there exists a $\mu_\epsilon > 0$ such that

$$\|\nabla f(x)\| \geq \alpha_\epsilon(f(x) - \underline{f}), \quad (6)$$

with $\alpha_\epsilon(r) = \sqrt{\mu_\epsilon r}$, for all $x \in \mathcal{X}_\epsilon := \{x \in \mathcal{X} \mid f(x) - \underline{f} \leq \epsilon\}$.

Although both conditions look similar, the distinction can imply important differences in the convergence behavior of solutions. To provide a rigorous framework with good granularity for analyzing the different possible types of Polyak-Łojasiewicz inequalities and their consequences to the convergence of gradient systems, we employ comparison functions in the next section.

C. Generalizations of the PŁI

The classic PŁI condition is originally formulated as in Definition 1, using the square root comparison function. However, practical cost functions often fail to satisfy this condition globally — an example being the continuous-time LQR cost, which will be discussed later. This limitation motivated us in [1] to propose a nonlinear version of the PŁI. By simply generalizing $\alpha(r)$ in (4) from $\alpha = \sqrt{\mu r}$ to a positive definite function $\alpha \in \mathcal{PD}$, the convergence of the gradient flow can still be ensured. As an additional benefit, when the gradient flow in (2) is subject to additive noise, the error $f(\phi(t, x_0)) - \underline{f}$ is bounded by an energy-like measure of the noise. This property is formally known as integral input-to-state stability (iISS) [21]. Furthermore, if α is strengthened to a class $\mathcal{K}$ function, the error $f(\phi(t, x_0)) - \underline{f}$ not only converges to zero in the absence of noise but also remains bounded when the noise is below a certain threshold, a property referred to as small-input ISS (siISS) [1]. If α is further strengthened to be a class $\mathcal{K}_\infty$ function, then the error $f(\phi(t, x_0)) - \underline{f}$ converges to zero in the noise-free case and remains bounded under any bounded noise, which corresponds to the classical input-to-state stability (ISS) [22]. Clearly, the global and semi-global PŁIs define comparison functions of class $\mathcal{K}_\infty$ and $\mathcal{PD}$ respectively.

For our analysis, we also introduce a new class of functions, called class *sat*PŁI functions, which can be

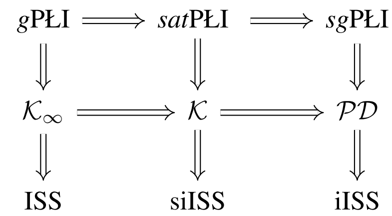


Fig. 1. Diagram of the hierarchy between types of comparison functions and their relationship with the different types of Polyak-Łojasiewicz inequality. Notice in particular that while all functions that satisfy a PŁI also have a class $\mathcal{K}_\infty$ lower-bound (as presented in definition 4), the converse is not necessarily true. Similarly, satisfying a sgPLI implies there exists a $\mathcal{PD}$ lower-bound, but the converse is not true. Also notice that the newly introduced class *sat*PŁI lower bound lies in between *g*PŁI and *sg*PŁI, and provides a better convergence guarantee. Finally, the ℓPŁI stands isolated in the graph, but one should note that it sustains the same convergence and robustness properties than the *g*PŁI, so long as the disturbance is small enough to keep the solution in a neighborhood of the optimum.

represented as

$$\alpha(r) = \sqrt{\frac{ar}{b + r}} \quad \forall r \geq 0, \quad (7)$$

where $a, b > 0$ are constants.

With this established, the following definition summarizes the "zoo" of the generalized inequalities based on different classes of comparison functions.

Definition 4: A function f satisfies a *class $\mathcal{K}_\infty$ lower bound* (resp. class *sat*PŁI , $\mathcal{K}$, or $\mathcal{PD}$) if

$$\|\nabla f(x)\| \geq \alpha\left(f(x) - f(x^*)\right), \quad (8)$$

for all $x \in \mathcal{X}$, with α being a function of class $\mathcal{K}_\infty$ (resp. class *sat*PŁI , $\mathcal{K}$, or $\mathcal{PD}$).

Notice that there is a natural order between the comparison function and a natural relation between each of them and the previously defined different types of PŁI, all illustrated in Fig. 1. In particular, notice that if the comparison function α is of class *sat*PŁI , then it lies in between a *g*PŁI and a *sg*PŁI.

An important consequence of satisfying different classes of lower bounds is the convergence behavior of the solutions. To better illustrate this effect, we first define a weaker form of convergence than the one from Definition 2.

Definition 5 (global linear-exponential cost stability): The gradient flow (2) of f is globally linear-exponential cost stable (GLECS) if there exists an $m > 0$ such that for every $x_0 \in \mathcal{X}$ and every $\underline{t} > 0$ there exists a $\mu_{x_0,\underline{t}} > 0$ for which

$$\begin{aligned}(f(\phi(t,&x_0)) - \underline{f}) \quad (9)\\ &\leq \begin{cases} (f(\phi(\underline{t}, x_0)) - \underline{f}) - m(t - \underline{t}) & \text{if } t \leq \underline{t} \\ (f(\phi(\underline{t}, x_0)) - \underline{f})e^{-\mu_{x_0,\underline{t}}(t-\underline{t})} & \text{if } t > \underline{t} \end{cases}.\end{aligned}$$

From this definition, we can derive rigorous conditions for the solution to be GLECS as follows:

Lemma 3: The gradient flow (2) of a function f is GLECS if

- The gradient ∇f is globally bounded;
- The function f satisfies a *sat*PŁI .

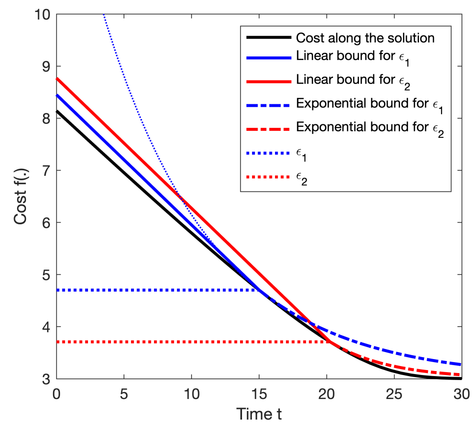


Fig. 2. Gradient Flow trajectory for a cost f that satisfies the conditions in Lemma 3.

We illustrate the results from Lemma 3 in Fig. 2. Observe that the true trajectory of the cost follows a "linear-exponential" profile, which means that for a given "margin of error" $\epsilon := f(\phi(\underline{t}, x_0)) - \underline{f}$, the solution is upper bounded by a line for values of the cost higher than ϵ and by and exponential for values smaller than ϵ. Also notice from the figure that a smaller margin of error ϵ results in a worse upper bound for the linear part of the solution, but a tighter upper bound for the exponential part.

Finally, notice a gap between GECS and GLECS: if f neither satisfies a global PŁI nor has a globally bounded gradient, then its solution is neither GECS nor GLECS. This is an important observation for the next section of the paper, as we will show that this is precisely the case for the LQR cost function.

III. Applications to LQR policy optimization

The policy optimization LQR setup is discussed thoroughly in publications such as [1], [3], [4], [6], [13], [14], [16], [17], so we provide only the necessary results, without derivation. Consider the following continuous-time linear system:

$$\dot{x} = Ax + Bu \tag{10}$$

where $A \in \mathbb{R}^{n\times n}$, and $B \in \mathbb{R}^{n\times m}$ are the system matrices, with (A, B) assumed to be controllable. Let $\mathcal{G} := \{K \in \mathbb{R}^{m\times n} \mid A - BK \text{ is Hurwitz}\}$. The objective is to determine a state feedback $u = -Kx$ with $K \in \mathcal{G}$ that minimizes

$$J(K) = \operatorname{trace}(P_K), \tag{11}$$

where P_K is the positive definite solution of

$$P_K(A - BK) + (A - BK)^\top P_K + K^\top RK + Q = 0, \tag{12}$$

with given matrices $R \in \mathbb{S}^{m\times m}_{++}$ and $Q \in \mathbb{S}^{n\times n}_{++}$. For this cost, the gradient flow $\dot{K} = -\nabla J(K)$, is given for ∇J being given by [23]:

$$\nabla J(K) = -2(B^\top P_K - RK)Y_K, \tag{13}$$

where, for any $K \in \mathcal{G}$, P_K is the solution of (12), and Y_K is the unique positive definite solution of

$$Y_K(A - BK)^\top + (A - BK)Y_K + I = 0. \tag{14}$$

In previous works in the literature [5], [6], it was established that the solution for a gradient flow dynamics for solving this problem initialized inside $\mathcal{G}_a := \{K \in \mathbb{R}^{m\times n} \mid J(K) \le a\}$, satisfies a semi-global PŁI (Lemma 1 of [5], and Theorem 3.16 of [6]), and in [1] it was shown that it actually satisfies a class *sat*PŁI lower-bound. Although a priori this does not rule out the possibility that J defined in (11) satisfies a *g*PŁI, in the following section we prove that this is not the case; that is, $J(\cdot)$ can never satisfy a *g*PŁI.

A. The LQR cost lies in the gap

Consider the following definition.

Definition 6: A matrix-valued function $\widetilde{K} : \mathbb{R}_+ \to \mathcal{G}$, is called a *high gain curve* of $\mathcal{G}$ if the eigenvalues of the closed-loop matrix are strictly in the left half-plane (LHP), *i.e.* there exists an $\epsilon > 0$ such that for all $\rho > 0$,

$$\max_i(\mathfrak{Re}(\lambda_i(A - B(\widetilde{K}(\rho) + K^*)))) < -\epsilon,$$

and the limit value of the cost is unbounded, *i.e.* $\lim_{\rho\to\infty} J(K^* + \widetilde{K}(\rho)) = \infty$.

Furthermore, we can guarantee that any (A, B) controllable has a high gain curve, as we show next:

Lemma 4: Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ be the system matrices for (10), with (A, B) being controllable. Then, there always exists a $\widetilde{K} : \mathbb{R}_+ \to \mathcal{G}$ that is a high gain curve of $\mathcal{G} := \{K \in \mathbb{R}^{m\times n} \mid (A - BK) \text{ is Hurwitz}\}$, with the additional property that $B\widetilde{K}(\rho)$ is diagonalizable for all $\rho > 0$.

With this, we can state the following lemma regarding the behavior of the gradient along high gain trajectories:

Lemma 5: Let $\widetilde{K} : \mathbb{R}_+ \to \mathcal{G}$ be a high gain curve of $\mathcal{G}$ and let there be a $\overline{\rho} > 0$ such that $B\widetilde{K}(\rho)$ is diagonalizable for all $\rho > \overline{\rho}$. Then the limit

$$\lim_{\rho\to\infty} \nabla J(\widetilde{K}(\rho) + K^*) \in \mathbb{R}^{m\times n}, \tag{15}$$

exists and is finite, *i.e.* the norm of the gradient $\|\nabla J(\widetilde{K}(\rho) + K^*)\|$ converges to a constant value as $J(\widetilde{K}(\rho) + K^*) - J(K^*)$ goes to infinity.

Informally, Lemma 5 proves the boundedness of the norm of the gradient along any "eventually diagonalizable" high gain curve. As a consequence, we state the following corollary.

Corollary 1: There is no $\mu > 0$ such that for all $K \in \mathcal{G}$ it holds that

$$\|\nabla J(K)\| \ge \sqrt{\mu(J(K) - J(K^*))},$$

i.e. , the cost function J defined in (11) can never satisfy a *g*PŁI.

From these results, one would think that J has a globally bounded gradient norm, since Lemma 5 shows that the gradient converges to a constant matrix for any high gain curve. However, that is not the case, as we show in the following lemma.

Lemma 6: For any $\overline{K} \in \partial\mathcal{G}$, let $\{K_i\}$ for $i = 1, 2, \dots$ be a sequence of matrices $K_i \in \mathcal{G}$ such that $\lim_{i\to\infty} K_i = \overline{K}$, then

$$\lim_{i\to\infty} \|\nabla J(K_i)\|_F = \infty.$$

As a consequence of this fact, gradient $\nabla J(\cdot)$ does not admit a global upper bound, and thus does not satisfy the conditions for Lemma 3.

The fact that the continuous-time LQR cost neither satisfies a gPŁI, nor has a globally bounded gradient, makes it hard to provide tight global convergence rate estimates to the policy-optimization algorithm. In fact, the solution can be either exponential or linear-exponential, depending on which region of the search-space it is initialized at. This can be circumvented by judicious restriction of the search-space as follows:

Lemma 7: For a $\delta > 0$, let $\mathcal{G}_\delta := \{K \in \mathcal{G} \mid A - BK + \delta I$ is Hurwitz$\}$, then there exists $g > 0$ such that $\|\nabla J(K)\| < g$ for all $K \in \mathcal{G}_\delta$.

The results presented so far illustrate, through the LQR policy optimization problem, the value of understanding exactly what kind of "PŁI-like" condition the cost function in question satisfies. To complement our analysis so far, and in hopes of illustrating the different possible behaviors for the solution of the policy optimization for the LQR, we next provide an analysis of the single-input single-state/output LQR case.

B. Convergence analysis of the scalar LQR policy optimization

For the scalar case, the continuous-time system dynamics is given by (10) with $A = a \in \mathbb{R}$, and $B = b = 1$, the latter being assumed without loss of generality (the magnitude of b can simply be included in the magnitude of u). The weighting matrices of the LQR cost are $Q = q \in \mathbb{R}$ and $R = r \in \mathbb{R}$.

Then, the cost and its gradient can be computed for the scalar case as in the following proposition

Proposition 1: For the scalar LQR, let k^* be the value of the feedback gain that minimizes the cost $J(k)$. Then we have that

$$J(k^* + \epsilon) - J(k^*) = \ell r \epsilon^2 =: \delta(\epsilon) \quad (16)$$

$$\partial J(k^* + \epsilon) = \ell(-\delta(\epsilon) + r\epsilon) \quad (17)$$

$$m(k^* + \epsilon) := \frac{\|\partial J(k^* + \epsilon)\|^2}{J(k^* + \epsilon) - J(k^*)} \quad (18)$$

$$= r\ell(\ell^2\epsilon^2 - 2\ell\epsilon + 1), \quad (19)$$

where p and ℓ solve (12) and (14), respectively.

These computations allow us to conclude that $m(k^*) = r\ell^* > 0$, characterizing exponential convergence near k^*. Furthermore, $\lim_{\epsilon\to a-k^*} m(k^* + \epsilon) = \infty$, indicating that the convergence rate explodes for values of k in the boundary of stability. Furthermore, the simpler form of the scalar case makes it easier to analyze numerical results.

Take $a = q = r = 1$ and notice from Fig. 3 (a) and (b), that the gradient $\partial J(k)$ behaves completely differently if $k > k^*$ or if $k < k^*$. However, notice that if we restrict the domain from $[1, \infty]$ to $[1+\epsilon, \infty]$, the value of the gradient is now globally bounded. This is the intuition behind Lemma 7.

Furthermore, the linear-exponential convergence behavior described in Lemma 3 becomes very evident for the scalar LQR if k is initialized larger than k^*, as can be seen in Fig. 3 (c). If, however, the solution is initialized near the border of instability, the convergence is much closer to a decreasing exponential, as evident in Fig. 3 (d).

C. Comments on the difference between continuous and discrete-time LQR policy optimization

We conclude the analysis of this paper with a brief overview of the behavior of the discrete-time LQR policy optimization problem. This scenario is studied in different papers in the literature [14]–[17] and for this problem, a global PŁI is characterized (see, for example, Lemma 1 in [16]). This is surprising since, by Corollary 1, the continuous-time LQR policy optimization can never admit a global PŁI.

Some intuition behind this difference can be obtained by looking at the Euler discretization of the scalar continuous case with step-size $h > 0$, and with $R_d = hr$ and $Q_d = hq$. For this problem, we can define $m_d(k_d^* + \epsilon)$ similarly to how it was done for the continuous-time case in (18).

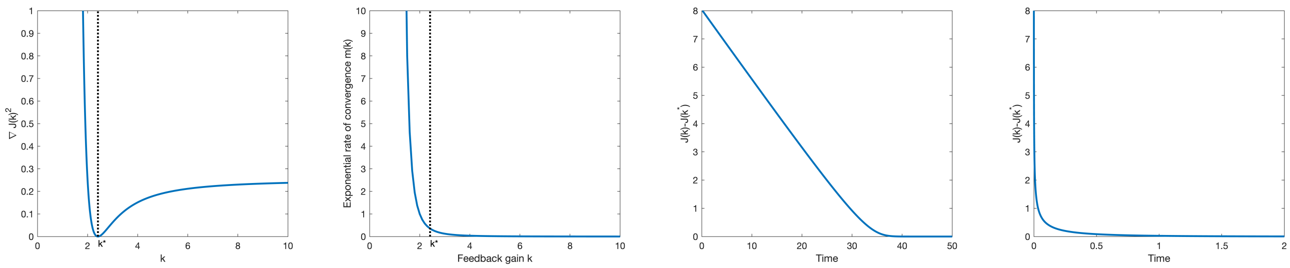
Furthermore, notice that the feedback gain k_d is bounded between $a < k_d^* + \epsilon < (2 + ha)/h$, which implies that $\mathcal{G}$ is bounded in the discrete-time. Moreover, upon explicit computation of $m_d(k_d^* + \epsilon)$, one can check that for any $h > 0$, $m_d(k_d, h) \to \infty$ if either $k_d \to a$ or $k_d \to (2+ha)/h$, which implies that $m_d(\cdot, h)$ must admit a minimum value $\underline{m_d}(h)$ attained at some point $\underline{k_d} \in \mathcal{G}$, *i.e.* $\underline{m_d}(h) = m_d(\underline{k_d}, h)$.

With these established, we pick $a = q = r = 1$ and plot $\underline{k_d}(h)$ and $\underline{m_d}(h)$ in Fig. 4. Notice that as $h \to 0$ (*i.e.* as we approach the CT LQR policy optimization problem), the value $\underline{k_d}(h)$ at which $m_d(k_d, h)$ is minimized goes to infinity, while $\underline{m_d}(h)$ goes to zero.

IV. CONCLUSIONS AND FUTURE WORKS

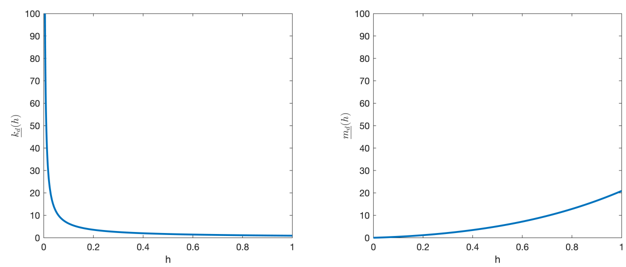
In this paper, we presented an overview of convergence guarantees for gradient methods in optimization problems. We revisited the Polyak-Łojasiewicz inequality (*i.e.* gradient dominance condition) and observed how slight changes in its characterization can imply significant changes to the convergence of the gradient flow solution. This motivated the introduction of nonlinear comparison functions as a way of characterizing the behavior of the solution, which we supported with a result that gives conditions for the solution to present a "linear-exponential" behavior in Lemma 3.

The paper follows up with a scenario where the traditional PŁI condition does not hold: the continuous-time model-free linear quadratic regulator problem. For this problem we showed that it presents neither globally exponential, nor globally linear-exponential convergence behavior, but a mixture of both depending on how close the solution is initialized to the border of instability. Despite that, we show in Lemma 5 that for any "high gain curve" in the space of stabilizing feedback matrices, the norm of the gradient



(a) (b) (c) (d)

Fig. 3. Illustration of the dual behavior of the LQR cost function (11) fort he scalar case with $a = r = q = 1$. In (a) we plot the squared norm of the gradient $\|\nabla J\|^2$ as a function of the feedback gain k, while in (b) we plot the largest exponential rate of convergence $m(k)$ as defined in (18), and in both plots the dotted vertical line indicates the optimal k^*. Plots (c) and (d) present numerical simulations for the gradient-flow initialized such that $J(k(0)) - J(k^*) \approx 8$, however (a) was initialized for $k(0) > k^*$ and (b) for $k(0) < k^*$.



(a) (b)

Fig. 4. Visualization of how the discretization step affects the global exponential rate of convergence for the scalar discrete-time LQR policy optimization problem.

is upper-bounded, which allowed us, through Lemma 7, to characterize global linear-exponential convergence behavior through a judicious restriction of the optimization search space. We then illustrate our results through numerical simulations of the scalar case, where the two regions of the parameter space are clearly defined. Finally we briefly discuss the distinct behavior of the LQR cost function when formulated for a continuous- and discrete-time linear system.

References

[1] L. Cui, Z.-P. Jiang, and E. D. Sontag, "Small-disturbance input-to-state stability of perturbed gradient flows: Applications to LQR problem," *Systems & Control Letters*, vol. 188, p. 105804, 2024.

[2] E. D. Sontag, "Remarks on input to state stability of perturbed gradient flows, motivated by model-free feedback control learning," *Systems & Control Letters*, vol. 161, p. 105138, Mar. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167691122000056

[3] A. C. B. De Oliveira, M. Siami, and E. D. Sontag, "Remarks on the gradient training of linear neural network based feedback for the LQR problem," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*, 2024, pp. 7846–7852.

[4] A. C. B. de Oliveira, M. Siami, and E. D. Sontag, "Convergence analysis of overparametrized LQR formulations," *arXiv preprint arXiv:2408.15456*, 2024.

[5] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović, "Convergence and sample complexity of gradient methods for the model-free linear-quadratic regulator problem," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2435–2450, 2021.

[6] I. Fatkhullin and B. Polyak, "Optimizing static linear feedback: Gradient method," *SIAM Journal on Control and Optimization*, vol. 59, no. 5, pp. 3887–3911, 2021.

[7] J. Bhandari and D. Russo, "Global optimality guarantees for policy gradient methods," *Operations Research*, vol. 72, no. 5, pp. 1906–1927, 2024. [Online]. Available: https://doi.org/10.1287/opre.2021.0014

[8] A. Eftekhari, "Training linear neural networks: Non-local convergence and complexity results," in *International Conference on Machine Learning*. PMLR, 2020, pp. 2836–2847.

[9] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, England: Cambridge University Press, 2004.

[10] V. Chvátal, *Linear programming*. London, UK: Macmillan, 1983.

[11] B. T. Polyak, "Gradient methods for minimizing functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 3, no. 4, pp. 643–653, 1963.

[12] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition," in *Machine Learning and Knowledge Discovery in Databases*, P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken, Eds. Cham: Springer International Publishing, 2016, pp. 795–811.

[13] Y. Watanabe and Y. Zheng, "Revisiting strong duality, hidden convexity, and gradient dominance in the linear quadratic regulator," *arXiv preprint arXiv:2503.10964*, 2025.

[14] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for the linear quadratic regulator," in *International Conference on Machine Learning*. PMLR, 2018, pp. 1467–1476.

[15] Y. Sun and M. Fazel, "Learning optimal controllers by policy gradient: Global optimality via convex parameterization," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 4576–4581.

[16] B. Hu, K. Zhang, N. Li, M. Mesbahi, M. Fazel, and T. Başar, "Toward a theoretical foundation of policy optimization for learning control policies," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 6, no. 1, pp. 123–158, 2023.

[17] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanovic, "Convergence and sample complexity of gradient methods for the model-free linear–quadratic regulator problem," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2435–2450, May 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9448427/

[18] H. Mohammadi, M. Soltanolkotabi, and M. R. Jovanović, "On the lack of gradient domination for linear quadratic Gaussian problems with incomplete state information," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 1120–1124.

[19] A. C. B. de Oliveira, L. Cui, and E. Sontag, "Remarks on the Polyak-Lojasiewicz inequality and the convergence of gradient systems," *arXiv preprint arXiv:2503.23641*, 2025.

[20] S. Łojasiewicz, "Sur les trajectoires du gradient d'une fonction analytique. (Trajectories of the gradient of an analytic function)," *Semin. Geom., Univ. Studi Bologna*, vol. 1982/1983, pp. 115–117, 1984.

[21] D. Angeli, E. D. Sontag, and Y. Wang, "A characterization of integral input-to-state stability," *IEEE Transactions on Automatic Control*, vol. 45, no. 6, pp. 1082–1097, 2000.

[22] E. D. Sontag *et al.*, "Smooth stabilization implies coprime factorization," *IEEE Transactions on Automatic Control*, vol. 34, no. 4, pp. 435–443, 1989.

[23] T. Rautert and E. W. Sachs, "Computational design of optimal output feedback controllers," *SIAM Journal on Optimization*, vol. 7, no. 3, pp. 837–852, Aug. 1997. [Online]. Available: http://epubs.siam.org/doi/10.1137/S1052623495290441