

Regret Analysis of Online Gradient Descent-based Iterative Learning Control with Model Mismatch

Efe C. Balta, Andrea Iannelli, Roy S. Smith, John Lygeros

Abstract—In Iterative Learning Control (ILC), a sequence of feedforward control actions is generated at each iteration on the basis of partial model knowledge and past measurements with the goal of steering the system toward a desired reference trajectory. This is framed here as an online learning task, where the decision-maker takes sequential decisions by solving a sequence of optimization problems having only partial knowledge of the cost functions. Having established this connection, the performance of an online gradient descent-based scheme using inexact gradient information is analyzed in the setting of static and dynamic regret, standard measures in online learning. Fundamental limitations of the scheme and its integration with adaptation mechanisms are further investigated, followed by numerical simulations on a benchmark ILC problem.

I. INTRODUCTION

Online learning-based optimization approaches have been increasingly studied in recent literature [1]–[4]. The online-learning setting usually assumes an unknown cost function that changes at each time-step, and an optimization algorithm that aims to minimize the unknown cost by using any prior information, e.g., a model, and observations of the cost and/or the gradient at each time-step. A natural generalization of this online learning setting is to consider an online-learning control problem, where the decision maker aims to control a dynamical system while minimizing a control cost at each time step. One of the first works studying online-learning and adaptive control was [5]. Since then, there have been many works focusing on solving the online-learning control problem under various assumptions on the type of model, uncertainty, constraints, and noise characteristics [3], [4], [6]. Regret is a common metric in many of the online-learning problems, as it provides a characterization of the cost incurred at each time step due to unknown changes to the cost function or problem structure. Additionally, since a fixed point convergence is not well-defined in many cases of online-learning problems, regret provides an alternative metric to assess the effectiveness of a given algorithm.

The class of online convex optimization (OCO) methods has been widely used for online learning problems [7]. Among the family of OCO methods, online gradient descent is of specific interest due to its simplicity and favorable guarantees on achievable regret under mild assumptions on the cost function and constraints [7]. However, many online gradient descent algorithms assume access to gradient

observations, which may not be available in many practical control applications. Recent work has considered variants of the online gradient descent using inexact gradient information for proximal-type optimization algorithms in an online setting [2] with additive errors on the gradient. Iterative approaches for control in an inexact gradient setting are studied in [1], where only additive errors to the known dynamics are considered.

Online optimization problems have a close relationship with Iterative Learning Control (ILC) methods. In ILC, the controller utilizes an input-output model of the process and learns from past iterations dealing with iteration-invariant [8]–[10] as well as iteration-varying problems [11]–[13]. While convergence properties under various assumptions on the dynamics and model uncertainty have been analyzed, regret analysis in an online learning ILC setting has not been considered yet in the literature. This work proposes an online-learning-based ILC method that utilizes a pre-conditioned online gradient descent method in the presence of model mismatch. After formulating the proposed control algorithm, its static and dynamic regret are quantified and variants are discussed and investigated. Our general analysis encompasses common ILC schemes previously proposed in the literature, and thus their regret characterization is an additional outcome of the work. The contribution of this work is therefore threefold: (i) a new online learning-based ILC methodology inspired by online gradient descent methods, (ii) a detailed regret analysis of the proposed ILC method and its variants, and (iii) regret analysis of existing ILC methods from the literature as special instances of the proposed ILC method.

Section II formulates the problem and proposes the online ILC controller. Section III provides a detailed analysis of regret in the transient and limit cases, while Section IV extends the results to the iteration-invariant ILC methods from the existing literature. Section V provides a numerical demonstration and Section VI gives concluding remarks.

Notation: Given a square matrix A , $\|A\|_2$ denotes its spectral norm and $\|A\|_P = \|P^{1/2}AP^{-1/2}\|_2$, where P is a symmetric positive definite matrix of appropriate dimension. Given a vector x , the weighted norm is $\|x\|_P = \sqrt{x^T Px}$.

II. PROBLEM FORMULATION

The considered iterative learning control problem is modelled by the following input-output dynamics in the absence of exogenous disturbances

$$y(x_k) = H_k x_k, \quad (1)$$

Research supported by the Swiss National Science Foundation under: NCCR Automation, a National Centre of Competence in Research (grant number 180545), and grant number 200021_178890.

All authors are with the Automatic Control Laboratory, ETH Zurich, 8092, Zurich, Switzerland. {ebalta,iannelli,rsmith,lygeros}@control.ee.ethz.ch

where, $y_k \in \mathbb{R}^{n_y}$ is the output, and $x_k \in \mathbb{R}^{n_x}$ is the input at iteration k . The input-output dynamics map H_k is commonly employed in the ILC literature and is referred to as the *lifted* representation of a system. Concretely, H_k may represent the temporal evolution of a linear parameter or time varying, or invariant dynamics along an iteration [8]–[10]. Importantly, H_k is assumed to be only partially known, with an uncertainty structure formally stated below, and a nominal estimate $M \approx H_k$. In each k , the goal is to minimize

$$f_k(x) = \frac{1}{2} (\|H_k x - r\|_Q^2 + \|x\|_R^2), \quad (2)$$

where r is a reference to be tracked, $Q = Q^T \succ 0$ is a weighting matrix, and the second term with $R = R^T \succ 0$ is used for regularization. This term is a flexible design choice used to penalize undesired features of the solution, such as high inputs. The weighting matrices Q, R may also be positive semi-definite in certain cases, see [10]. Note that following the same formulation, iteration varying and a priori known r_k may be used in place of r . We focus on the case with iteration invariant r in this work for simplicity. The gradient of (2) is given by

$$\nabla f_k(x) = H_k^T Q(H_k x - r) + R x. \quad (3)$$

Notice that while the term $H_k x$ can be evaluated directly by running an iteration on the true system with the input x and measuring the output $y(x)$, the adjoint dynamics of the true system H_k^T are unknown. To circumvent this problem, one can use the nominal estimate, M , to estimate the gradient, leading to

$$\tilde{\nabla} f_k(x) = M^T Q(y(x) - r) + R x. \quad (4)$$

The ILC update applied to generate new inputs at each iteration is given by the following Preconditioned Online Gradient Descent (POGD) step

$$x_{k+1} = \Pi_{\mathcal{X}}^W \left(x_k - \alpha_k W^{-1} \tilde{\nabla} f_k(x_k) \right), \quad (5)$$

where $W = W^T \succ 0$ is a preconditioner matrix, \mathcal{X} is a convex input constraint set, and α_k is the step-size at iteration k . The projection operator to the set \mathcal{X} in the weighted preconditioner norm is

$$\Pi_{\mathcal{X}}^W(x) := \operatorname{argmin}_{u \in \mathcal{X}} \|u - x\|_W. \quad (6)$$

The concrete uncertainty representation of H_k is stated next.

Assumption 1: For all k , the true dynamics H_k belongs to the set $\mathbf{H}(M, \Delta) := \{H \mid H = M + M\Delta\}$, where M is a nominal estimate with full column rank and the uncertainty Δ belongs to the unstructured norm bounded set $\mathbf{\Delta}(W, \gamma) := \{\Delta \mid \|\Delta\|_W \leq \gamma\}$, where $\gamma \geq 0$ is the known uncertainty size, and W is the preconditioner matrix.

Uncertainty representations similar to Assumption 1 have been used in the past literature [8], [14], [15]. Note that due to $y(x)$, the updates (4) and (5) denote the interconnection of a dynamical system and a controller algorithm. Additionally, (4) with Assumption 1 introduces a gradient mismatch that has not been studied in the past OGD literature.

The main technical contribution of the paper is the analysis of this POGD-ILC in terms of two notions of regret. The most general case corresponds to the dynamic regret

$$J_d(T) = \sum_{k=1}^T f_k(x_k) - \sum_{k=1}^T f_k(x_k^*), \quad (7)$$

where $x_k^* = \operatorname{argmin}_{x \in \mathcal{X}} f_k(x)$, i.e., the regret with respect to an iteration-wise optimal control policy. Additionally, we consider the traditional static regret [7]

$$J_s(T) = \sum_{k=1}^T f_k(x_k) - \min_{x \in \mathcal{X}} \sum_{k=1}^T f_k(x). \quad (8)$$

The static regret is with respect to a controller that defines a single fixed optimal input with the hindsight information about the full sequence of iteration-varying f_k . The regret analysis is based on the following assumptions:

Assumption 2: For each k , f_k is locally Lipschitz continuous in \mathcal{X} with Lipschitz constant L_k in the weighted preconditioner norm, i.e., $\|f_k(x) - f_k(y)\|_W \leq L_k \|x - y\|_W \forall x, y \in \mathcal{X}$; moreover, $\bar{L} := \sup_k \{L_k\} < \infty$.

Assumption 3: The optimal input between consecutive iterations is bounded as $\|x_k^* - x_{k+1}^*\|_W \leq e_k$.

Assumption 4: There exist a sequence σ_k such that $\|W^{-1}(M\Delta)^T Q(H_k x_k^* - r)\|_W \leq \sigma_k$ for all k with $\bar{\sigma} := \sup_k \{\sigma_k\} < \infty$.

Assumption 2 holds for example when \mathcal{X} is compact. Assumption 3 ensures that the change in the optimal inputs are bounded and an upper bound estimate is available. Assumption 4 is due to the model mismatch term Δ , and characterizes the distance between the fixed point of (5) for fixed k , and the optimizer x_k^* . We formally show how this term appears in some of the regret bounds and discuss its role under various settings in later sections. Finally, we define

$$\phi_k := \|I - \alpha_k W^{-1}(M^T Q H_k + R)\|_W, \quad \Phi_{j,k} := \prod_{i=j}^k \phi_i$$

The missing proofs in Sections III and IV are given in [16].

III. REGRET ANALYSIS

In this section we analyze the dynamic and static regrets of the sequential actions taken using the POGD algorithm (5).

A. Dynamic Regret: Transient and Asymptotic Behavior

The following theorem provides an upper bound on the dynamic regret of the POGD algorithm under the design choices and assumptions discussed so far.

Theorem 1 (Dynamic Regret of POGD-ILC): Under Assumptions 1, 2, 3, and 4, consider the choice of preconditioner $W = M^T Q M + R$ and define $w := \|W^{-1} M^T Q M\|_W$. If $w\gamma < 1$ and the step-size is chosen as $\alpha_k \in \left(0, \frac{2}{1+w\gamma}\right)$, then the dynamic regret of POGD is upper bounded by

$$J_d(T) \leq \bar{L} \delta_{x_1} \sum_{k=1}^T \Phi_{1,k} + \bar{L} \bar{\sigma} \sum_{k=1}^T \sum_{j=1}^k \alpha_j \Phi_{j+1,k} + \bar{L} \sum_{k=1}^T E_k$$

where $E_k := \sum_{j=1}^k e_j \Phi_{j+1,k}$ and $\delta_{x_1} := \|x_1 - x_1^*\|_W$.

Proof: We first bound the distance between the input updates and the corresponding optimal inputs.

$$\begin{aligned}
\|x_{k+1} - x_{k+1}^*\|_W &\leq \|x_{k+1} - x_k^*\|_W + \|x_k^* - x_{k+1}^*\|_W \\
&= \|\Pi_{\mathcal{X}}^W(x_k - \alpha_k W^{-1} \tilde{\nabla} f_k(x_k)) \\
&\quad - \Pi_{\mathcal{X}}^W(x_k^* - \alpha_k W^{-1} \nabla f_k(x_k^*))\|_W + e_k \\
&\leq \|x_k - \alpha_k W^{-1}(M^T Q(H_k x_k - r) + R x_k) \\
&\quad + x_k^* - \alpha_k W^{-1}(H_k^T Q(H_k x_k^* - r) + R x_k^*)\|_W + e_k \\
&\leq \|(I - \alpha_k W^{-1}(M^T Q H_k + R))(x_k - x_k^*)\|_W \\
&\quad + \alpha_k \|W^{-1}(M \Delta)^T Q(H_k x_k^* - r)\|_W + e_k \\
&\leq \phi_k \|x_k - x_k^*\|_W + \alpha_k \sigma_k + e_k, \tag{9}
\end{aligned}$$

where in the first inequality we use the triangle inequality and in the second equality the fact that x_k^* is a fixed point of the POGD with the true gradient ∇f_k , $x_k^* = \Pi_{\mathcal{X}}^W(x_k^* - \alpha_k W^{-1} \nabla f_k(x_k^*))$, and Assumption 3. For the other inequalities we use the fact that the weighted projection operator is nonexpansive in the weighted preconditioner norm, Cauchy-Schwartz inequality and Assumption 4. Next, we show the step-size parameters required to ensure $\phi_k < 1$. Using Assumption 1 and the specific choice of preconditioner we have

$$\begin{aligned}
\phi_k &= \|I - \alpha_k W^{-1}(M^T Q(M + M \Delta) + R)\|_W \\
&= \|(1 - \alpha_k)I - \alpha_k W^{-1} M^T Q M \Delta\|_W \\
&\leq |1 - \alpha_k| + \alpha_k \|W^{-1} M^T Q M \Delta\|_W \\
&\leq |1 - \alpha_k| + \alpha_k w \gamma, \tag{10}
\end{aligned}$$

where $\|I\|_W = 1$ was used in the first inequality. To ensure that $\phi_k < 1$, α_k must be chosen such that

$$|1 - \alpha_k| < 1 - \alpha_k w \gamma. \tag{11}$$

Then since $w \gamma < 1$, $\alpha_k \in (0, \frac{2}{1+w\gamma})$ implies $\phi_k < 1$. By iterating (9) one gets

$$\begin{aligned}
\|x_{k+1} - x_{k+1}^*\|_W &\leq \|x_1 - x_1^*\|_W \prod_{j=1}^k \phi_j \\
&\quad + \sum_{j=1}^k ((\sigma_j \alpha_j + e_j) \prod_{i=j+1}^k \phi_i),
\end{aligned}$$

where we adopt the convention $\prod_{j=1}^j a_j = 1$. Using the Lipschitz constant L_k we get

$$\begin{aligned}
f_{k+1}(x_{k+1}) - f_{k+1}(x_{k+1}^*) &\leq L_k \|x_{k+1} - x_{k+1}^*\|_W \\
&\leq L_k \left(\|x_1 - x_1^*\|_W \Phi_{1,k} + \bar{\sigma} \sum_{j=1}^k \alpha_j \Phi_{j+1,k} + \sum_{j=1}^k e_j \Phi_{1,k} \right).
\end{aligned}$$

Taking the sum from 1 to T and using the upper bound \bar{L} instead of L_k at each step gives the desired result. ■

The condition $w \gamma < 1$ can be fulfilled by choice of the regularization matrix R . To see this, define w_1 and γ_1 the values of w and γ associated with R_1 and W_1 . If $w_1 \gamma_1 \geq 1$, we can always find R_2 such that $w_2 \gamma_2 < 1$. This is because, from the choice of preconditioner W and the definition of w in Theorem 1, w scales approximately with $\|W^{-1}\|$ and thus is $\mathcal{O}(\|R^{-1}\|)$. On the other hand, a valid (possibly not the tightest) upper bound on the

uncertainty size γ is $\mathcal{O}(1)$ in R . Consider without loss of generality $R_1 = \rho_1 I$ and $R_2 = \rho_2 I$, with $\rho_2 > \rho_1 > 0$. Using the definitions, we have $\|\Delta\|_{W_1} \leq \text{cond}(W_1^{1/2}) \|\Delta\| = \gamma_1$, where $\text{cond}(\cdot)$ denotes the condition number of the matrix. Since $\rho_2 > \rho_1$, $\text{cond}(W_2) < \text{cond}(W_1)$ and thus $\|\Delta\|_{W_2} \leq \text{cond}(W_2^{1/2}) \|\Delta\| = \gamma_2 < \gamma_1$, i.e. γ_1 is still a valid uncertainty size for the new choice of R_2 .

Using the upper bound obtained in Theorem 1, we characterize the asymptotic behavior of the dynamic regret.

Corollary 2 (Average Regret of POGD-ILC): Under the same conditions as Theorem 1, if $\alpha_k = \alpha_0 k^{-c}$, with $\alpha_0 \in (0, \frac{2}{1+w\gamma})$, $0 \leq c < 1$, then

$$\lim_{T \rightarrow \infty} \frac{J_d(T)}{T} \leq \mathcal{O}(1) + \frac{\bar{L} \sum_{k=1}^T E_k}{T}. \tag{12}$$

Proof: From Theorem 1, $J_d(T)$ is bounded by

$$\underbrace{\bar{L} \delta_{x_1} \sum_{k=1}^T \Phi_{1,k}}_{\text{Term I}} + \underbrace{\bar{L} \bar{\sigma} \alpha_0 \sum_{k=1}^T \sum_{j=1}^k j^{-c} \Phi_{j+1,k}}_{\text{Term II}} + \underbrace{\bar{L} \sum_{k=1}^T E_k}_{\text{Term III}} \tag{13}$$

where in Term II the explicit expression of the step size has been used. Term I can be interpreted as the contribution to the regret due to distance of the initial decision from the optimal one. Define $\bar{\phi}_k := \sup_{i=1}^k \{\phi_i\}$ and recall that $\bar{\phi}_k < 1$ by (10), (11), and the choice of the step size. Then

$$\bar{L} \delta_{x_1} \sum_{k=1}^T \Phi_{j,k} \leq \bar{L} \delta_{x_1} \sum_{k=1}^T (\bar{\phi}_k)^k \leq \bar{L} \delta_{x_1} \sum_{k=1}^T (\bar{\phi}_T)^k \leq \frac{\bar{L} \delta_{x_1}}{1 - \bar{\phi}_T},$$

where we used the monotonicity of $\bar{\phi}_k$ in the second inequality, and the upper bound of the infinite sum in the last inequality. From (10) there exists a finite \bar{T} , which depends on α_0 and c , such that for $T > \bar{T}$

$$\bar{\phi}_T \leq 1 + (w\gamma - 1)\alpha_0 T^{-c} \tag{14}$$

and thus

$$\lim_{T \rightarrow \infty} \frac{\bar{L} \delta_{x_1}}{T(1 - \bar{\phi}_T)} \leq \lim_{T \rightarrow \infty} \frac{\bar{L} \delta_{x_1} T^c}{T(w\gamma - 1)\alpha_0} = 0 \tag{15}$$

whenever $c < 1$. Next, consider Term II and define $S_k := \sum_{j=1}^k j^{-c} \Phi_{j+1}^k$, which thus describes the growth of this term at each step k . Observe that

$$S_{k+1} = \phi_{k+1} S_k + (k+1)^{-c} \tag{16}$$

where we know from (10) that $\phi_{k+1} < 1$. For our choice of α_k , two cases should be considered. When $0 < c < 1$ (i.e., vanishing step size), $\phi_{k+1} \rightarrow 1$ for $k \rightarrow \infty$. In the limit $k \rightarrow \infty$, the sequence S_k will thus converge to a finite constant value S_∞ . When $c = 0$, $\phi_{k+1} < 1$ as $k \rightarrow \infty$, and thus S_k can be bounded between zero and the trajectory of an asymptotically stable linear system with constant input of 1. Therefore, by using the asymptotic behavior of the linear time varying system (16) we are able to characterize the asymptotic behavior of the regret for Term II. In both cases Term II achieves linear regret, leading to $\mathcal{O}(1)$. ■

It is worth noting that the presented case relaxes some assumptions in the existing literature. As an example, [2] presents a similar result for fixed step-size case and general strongly convex cost functions, which corresponds to the case with $c = 0$. Additionally, the interpretation of the regret bound in terms of the dynamical equation (16) provides additional insights in terms of algorithm design and provides a basis for developing system-level synthesis-type regret optimal design [17].

Corollary 2 shows that the POGD algorithm applied to the ILC with model mismatch does not lead to a sublinear regret. The latter is regarded as a favorable property for sequential decision-making algorithms because it suggests that on average the decisions asymptotically converge to the optimal ones at each stage. Convergence is prevented here by two terms, namely Term II and Term III. Term III with E_k is known as *complexity* [18] term in the dynamic regret literature and captures the effect of the temporal variability of the optimal sequence of actions. It is well-known that an upper bound on the dynamic regret will have an explicit dependence on it and, in this setting, little can be said about its growth without prior information or assumptions on H_k . By inspecting the derivation of the second term on the right-hand-side in the bound (9), Term II is the contribution to the regret due to the suboptimality of the direction taken to update the decision at k . More precisely, this term is related to the term upper bounded by σ_k in Assumption 4.

B. Adaptive POGD Algorithm

Modifications to the original POGD algorithm which are sufficient for achieving sublinear regret of Term II are discussed in this section.

Assumption 5: For all k , the true dynamics H_k belongs to the set $\mathbf{H}_k(M_k, \Delta_k) := \{H \mid H = M_k + M_k \Delta_k\}$. M_k is a full column rank nominal estimate at k with $M_k^T Q M_k \preceq \bar{m} I$ and the uncertainty Δ_k belongs to the unstructured norm bounded set $\mathbf{\Delta}_k(W, \gamma_k) := \{\Delta \mid \|\Delta\|_W \leq \gamma_k\}$, where $\gamma_k \leq \gamma$ for all k and $\gamma_k \rightarrow 0$ as $k \rightarrow \infty$.

This Assumption is a stronger version of Assumption 1 and requires the uncertainty size to asymptotically vanish. This could be achieved, for example, with an online identification scheme providing updated estimates of the model M_k and of the uncertainty based on input-output measurements $\{(y_i, x_i)\}_{i=1}^T$ gathered during the decision-making problem. Asymptotic convergence to zero of the estimation error $\|H_k - M_k\|_W$ would also require appropriate excitation conditions on r in the spirit of recursive parameter identification schemes used in adaptive control [19].

Assumption 6: There exist $\tilde{\sigma}_k$ such that $\|W^{-1}(M_k \Delta_k)^T Q (H_k x_k^* - r)\|_W \leq \tilde{\sigma}_k$ for all k . Moreover, $\tilde{\sigma}_k \rightarrow 0$ as $k \rightarrow \infty$.

This Assumption replaces Assumption 4 and redefines the sequence of upper bounds $\tilde{\sigma}_k$ for the case when the estimate M_k changes across iterations. The asymptotic behavior of $\tilde{\sigma}_k$ is implied by Assumptions 5. Further, define

$$\tilde{\phi}_k := \|I - \alpha_k W^{-1}(M_k^T Q H_k + R_k)\|_W, \quad \tilde{\Phi}_{j,k} := \prod_{i=j}^k \tilde{\phi}_i$$

Consider now an adaptive variation of the POGD algorithm described in Section II which, leveraging Assumption 5, uses for its decisions the updated estimate of the model M_k . The following Corollary shows that the associated dynamic regret is sublinear if the complexity term is sublinear.

Corollary 3 (Average Regret with Adaptation): Under Assumptions 2, 3, 5, and 6, consider the choice of preconditioner $W = M_1^T Q M_1 + R_1$, with R_1 chosen so that $w_k \gamma < 1$ for all k , where $w_k := \|W^{-1} M_k^T Q M_k\|_W$. Consider also the regularizer weighting matrix $R_k = W - M_k^T Q M_k \succ 0$. If the step-size is chosen as $\alpha_k = \alpha_0$ with $\alpha_0 \in \left(0, \frac{2}{1+w\gamma}\right)$, then

$$\lim_{T \rightarrow \infty} \frac{J_d(T)}{T} \leq \frac{\bar{L} \sum_{k=1}^T E_k}{T}. \quad (17)$$

The proof is sketched here. Following the proof of Theorem 1 we have a similar upper bound on the regret with the exception of the change in Term II, where now the stepsize is constant and $\tilde{\sigma}_k$ is kept inside the inner summation. As a result, the variable $\tilde{S}_k := \sum_{j=1}^k \tilde{\sigma}_k \tilde{\Phi}_{j+1,k}$ describing the growth of Term II at each step k is such that

$$\tilde{S}_{k+1} \leq \hat{\phi} \tilde{S}_k + \tilde{\sigma}_k \quad (18)$$

where, from Assumption 6, $\tilde{\sigma}_k \rightarrow 0$ as $k \rightarrow \infty$. Therefore both Term I and Term II achieve sublinear regret and we are left with Term III as given in the result. The condition $R_k \succ 0$ can be achieved by using \bar{m} .

Compared to the originally considered POGD algorithm, the adaptive version features three major changes: the model estimate is updated online; the step size is kept constant (non-diminishing); the regularization matrix is adapted as a function of the current model estimate. Note that at this stage this is not a complete algorithm, as it needs to be complemented by an online identification algorithm satisfying Assumption 5. The purpose of its presentation is primarily to establish conditions on this complementary identification procedure to make the commonly used POGD algorithm competitive from a regret perspective.

C. Static Regret

While dynamic regret provides a powerful metric for analyzing the performance of an online learning algorithm, its upper bound depends on the limiting behavior of the complexity term, Term III, which is unknown in general. This term disappears in the static regret case (8), which is studied next. The fixed input x^* is computed in hindsight to minimize the sum of observed costs, i.e. $x^* = \arg \min_{x \in \mathcal{X}} \sum_{k=1}^T f_k(x)$, see (8). The analysis is based on the following assumption.

Assumption 7: There exist η_k such that $\|W^{-1/2} \tilde{\nabla} f_k(x^*)\| \leq \eta_k$ for all k , and $\bar{\eta} = \sup_k \{\eta_k\} < \infty$.

Corollary 4: Under the conditions of Theorem 1 and Assumption 7, the static regret of POGD is bounded by

$$J_s(T) \leq \bar{L} \delta_{x_1} \sum_{k=1}^T \Phi_{j,k} + \bar{L} \bar{\eta} \sum_{k=1}^T \sum_{j=1}^k \alpha_j \Phi_{j+1,k} \quad (19)$$

Following the arguments of Corollary 2, it can be seen that the static regret grows linearly due to the new Term II

(now depending on $\bar{\eta}$). Note that the optimality of x^* is not necessary for the proof of Corollary 4. Therefore, the static regret defines a worst-case cost gap against any static policy played over the iteration horizon of T . This observation provides further insight on the meaning of regret, and its distinction with respect to other metrics such as optimality gap or convergence rates.

IV. THE ITERATION INVARIANT PROBLEM

In this section, we specialize the results of Theorem 1 to more commonly considered ILC settings featuring the assumption on constant cost function, i.e., $f_k(x) = f(x)$ for all iterations k . Specifically, we assume that $f(x) = \frac{1}{2}(\|Hx - r\|_Q^2 + \|x\|_R^2)$, where the true dynamics H has the same uncertainty description defined in Assumption 1 but is now iteration-invariant. This results in the ILC update

$$x_{k+1} = \Pi_{\mathcal{X}}^W \left(x_k - \alpha_k W^{-1} \tilde{\nabla} f(x_k) \right), \quad (20)$$

which now has a fixed point \bar{x} under suitable conditions, see [10], [15]. Following our analysis in the proof of Theorem 1, it is easy to see that the step-size rule given in the theorem with the given preconditioner choice ensures convergence to the fixed point, i.e., $\|x_{k+1} - \bar{x}\|_W \leq \phi_k \|x_k - \bar{x}\|_W$. Additionally, note that the fixed point \bar{x} is not necessarily the optimal point x^* due to the model mismatch, thus $\|\bar{x} - x^*\|$ is nonzero in the general case (see [10] for further details).

The ILC update (20) with $\mathcal{X} = \mathbb{R}^{n_x}$ and $\alpha_k = 1$ results in norm-optimal ILC under suitable preconditioner matrix design [8], [13], [14], while the case of convex $\mathcal{X} \subset \mathbb{R}^{n_x}$ with a suitably chosen fixed step-size $\alpha_k = \bar{\alpha}$ is a variant of the optimization-based ILC [10], [15]. Due to the model mismatch, and also for the cases with bounded disturbance, such ILC algorithms achieve nonzero asymptotic error. Therefore, it is desirable to design control parameters to minimize the asymptotic ratio (gain) of the fixed point mismatch $\|\bar{x} - x^*\|$ to the uncertainty size in the problem, e.g., size of the uncertainty set or the disturbance set.

Since here we have $f_k(x) = f(x)$ for all iterations k , the constant input $x^* = \arg \min_{x \in \mathcal{X}} f(x)$ is the optimal action for both the dynamic and static problems, thus the associated notions of regret coincide and will be referred to as *ILC regret* (J_{ILC}).

Proposition 5 (ILC regret for (20)): Under Assumptions 1 and 2, consider the choice of preconditioner $W = M^T Q M + R$ and define $w := \|W^{-1} M^T Q M\|_W$. If $w\gamma < 1$ and a constant step-size is chosen as $\alpha = \alpha_0 \in \left(0, \frac{2}{1+w\gamma}\right)$, the ILC regret for the controller update (20) is bounded by

$$J_{\text{ILC}}(T) \leq (1 - \phi)^{-1} (\bar{L}(\delta_{x_1} + \sigma\alpha_0 T)), \quad (21)$$

where $\phi = \|I - \alpha_0 W^{-1}(M^T Q H + R)\|_W$, $\delta_{x_1} := \|x_1 - x^*\|_W$, and $\sigma \geq 0$ is such that $\|W^{-1}(M\Delta)^T Q(Hx^* - r)\|_W \leq \sigma$.

The result follows directly from the proof of Theorem 1, with constant x^* instead of x_k^* and without the e_k term

due to the iteration invariance. Following Corollary 3, it is easy to see that the regret of the ILC update (20) becomes sublinear if model learning takes place concurrently with the controller iterations and Assumptions 5 and 6 are satisfied for the iteration invariant problem, e.g., in certain model-free ILC applications [20]. The bound for the iteration varying α_k follows similarly from Theorem 1 and is omitted here.

The linear regret is due to the mismatch term $\|W^{-1}(M\Delta)^T Q(Hx^* - r)\|_W$, which characterizes the distance $d = \|\bar{x} - x^*\|$. Therefore, we see here a clear relationship between convergence and regret, where the case of no model mismatch achieves sublinear regret. Therefore, by improving on the fixed point by reducing d , it is possible to achieve sublinear regret.

Proposition 6: Let Assumptions 1 and 2 be satisfied, and assume further that $\gamma < 1$. Consider the choice of preconditioner $W = M^T Q M$, cost $f(x) = \frac{1}{2}\|Hx - r\|_Q^2$, assume that $\min_{x \in \mathcal{X}} f(x) = 0$ with the minimizer x^* , and a constant step-size is chosen as $\alpha = \alpha_0 \in \left(0, \frac{2}{1+w\gamma}\right)$ and define $\delta_{x_1} := \|x_1 - x^*\|_W$. Then, the ILC regret for the controller update (20) is bounded by

$$J_{\text{ILC}}(T) \leq (1 - \phi)^{-1} (\bar{L}\delta_{x_1}), \quad (22)$$

where, $\phi = \|I - \alpha_0 W^{-1}(M^T Q H)\|_W$. Hence, the average regret is sublinear.

The proof of Proposition 6 follows from Proposition 5 by recognizing that we have here $\|W^{-1}(M\Delta)^T Q(Hx^* - r)\|_W = 0$ by assumption, since $\min_{x \in \mathcal{X}} f(x) = 0$ implies that $Hx^* = r$. Therefore, by having a small enough disturbance set, i.e., $\gamma < 1$, and assuming that the optimal input is feasible for the true dynamics, the ILC update (20) has sublinear regret.

V. NUMERICAL DEMONSTRATION

For the numerical demonstration, we turn to process control for a Selective Laser Melting (SLM) additive manufacturing process. In SLM, fine metal powder is deposited, melted with the help of a high-power laser, and left to solidify in layers, to build a three-dimensional object in a layer-by-layer fashion. The melt pool dynamics at the point where the laser interacts with the material is of crucial importance for the mechanical properties of the finished part. We use the high-fidelity numerical simulations of an SLM process presented in [21] to model the melt pool length output as a function of the laser power input and extract a 5 dimensional discrete-time linear time invariant single input single output model

$$\mathcal{H} : \begin{cases} \xi(t+1) = A\xi(t) + Bv(t), \\ y(t) = C\xi(t), \end{cases} \quad (23)$$

where $v(t)$ is the instantaneous power input to the system and $y(t)$ is the melt pool length. The constraint set on the input power is defined by the minimum power requirement to initiate melting, and an upper limit based on actuator constraints, given by $\mathcal{V} = [75, 165]$, in *Watt*, so that we constraint our input to $v(t) \in \mathcal{V}$. Using the model (23), we construct the lifted input-output model of the system

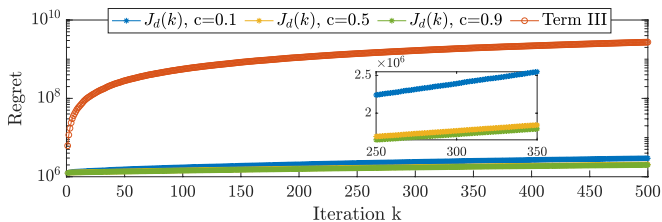


Fig. 1. Dynamic regret with different step-size rules and Term III from Theorem 1.

for an iteration duration of 100 time steps, representing a single layer of the SLM process, so that $M \in \mathbb{R}^{100 \times 100}$. The input constraint set \mathcal{X} for the POGD-ILC algorithm is then constructed using \mathcal{V} . Following our model assumption, we compute the true input-output dynamics in each iteration as $H_k = M + M\Delta$, where Δ has a diagonal structure, and is sampled from the set $\Delta(W, \gamma_0)$ for each k .

We present results (i) for the dynamic regret of non-adaptive POGD-ILC with diminishing step-size $\alpha_k = \alpha_0 k^{-c}$, chosen according to Corollary 2, and (ii) for the case with model learning Adaptive POGD-ILC, with a constant step size according to Corollary 3. For the adaptive POGD-ILC, we emulate the adaptation by a diminishing uncertainty set and $\gamma_k = \gamma_0 k^{-1/2}$, where γ_0 is the initial uncertainty, also used in the non-adaptive case.

The dynamic regret of the POGD-ILC controller under three step-size choices is shown in Fig. 1. The top plot also shows the complexity term (Term III from Theorem 1), providing part of the upper bound as predicted analytically. A close-up of the regret progression for the three step sizes is shown on the bottom plot. We see that the regret progression increases with diminishing c , which is captured analytically by the dependence of Term II in Theorem 1, suggesting that larger step sizes result in increased regret upper bounds. The melt pool lengths of the two scenarios are given in Fig. 2. The tracking performance of the adaptive POGD-ILC is much better due to the model learning and adaptation. The non-adaptive POGD-ILC still tracks the reference signal, albeit with higher error.

VI. CONCLUSION

This work analyzes the regret of online learning iterative learning controllers with model mismatch between the true process and the controller model. We propose a projected online gradient descent controller inspired by online convex optimization methods and analyze the regret performance of the proposed controller under various assumptions and conditions. Thus, we study variations of existing ILC algorithms in the context of online learning to study the closed-loop performance of given algorithms with dynamical systems. Our results motivate the need for online learning and adaptation to achieve sublinear average regret, which is desirable in many practical contexts. Accordingly, developing effective model learning methodologies and incorporating additional measurement noise, process noise, and state constraints are foreseeable extensions of this work.

Acknowledgements: The authors thank Mamzi Afrasiabi for providing the simulation data of [21].

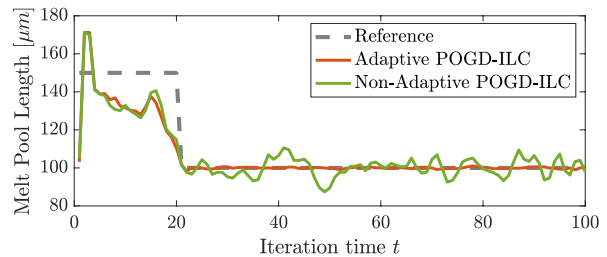


Fig. 2. Comparison of the output trajectories at the last iteration $k = 500$, for the adaptive and non-adaptive cases.

REFERENCES

- [1] N. Agarwal, E. Hazan, A. Majumdar, and K. Singh, "A regret minimization approach to iterative learning control," in *International Conference on Machine Learning*. PMLR, 2021, pp. 100–109.
- [2] R. Dixit, A. S. Bedi, R. Tripathi, and K. Rajawat, "Online learning with inexact proximal online gradient descent algorithms," *IEEE Transactions on Signal Processing*, vol. 67, no. 5, 2019.
- [3] G. Goel and B. Hassibi, "Regret-optimal control in dynamic environments," *arXiv preprint arXiv:2010.10473*, 2020.
- [4] D. Muthirayan, J. Yuan, D. Kalathil, and P. P. Khargonekar, "Online learning for receding horizon control with provable regret guarantees," *arXiv preprint arXiv:2111.15041*, 2021.
- [5] M. Raginsky, A. Rakhlin, and S. Yüksel, "Online convex programming and regularization in adaptive control," in *49th IEEE Conference on Decision and Control (CDC)*, 2010.
- [6] E. Hazan, S. Kakade, and K. Singh, "The nonstochastic control problem," in *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, vol. 117, 2020, pp. 408–421.
- [7] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends® in Optimization*, vol. 2, no. 3–4, 2016.
- [8] K. L. Barton and A. G. Alleyne, "A norm optimal approach to time-varying ILC with application to a multi-axis robotic testbed," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 1, 2010.
- [9] N. Amann, D. H. Owens, and E. Rogers, "Iterative learning control for discrete-time systems with exponential rate of convergence," *IEE Proceedings-Control Theory and Applications*, vol. 143, no. 2, 1996.
- [10] D. Liao-McPherson, E. C. Balta, A. Rupenyan, and J. Lygeros, "On robustness in optimization-based constrained iterative learning control," *IEEE Control Systems Letters*, 2022.
- [11] B. Altin, J. Willems, T. Oomen, and K. Barton, "Iterative learning control of iteration-varying systems via robust update laws with experimental implementation," *Control Engineering Practice*, 2017.
- [12] E. C. Balta, D. M. Tilbury, and K. Barton, "Switch-based iterative learning control for tracking iteration varying references," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1493–1498, 2020.
- [13] J. Van Zundert, J. Bolder, and T. Oomen, "Optimality and flexibility in iterative learning control for varying tasks," *Automatica*, 2016.
- [14] J. Van de Wijdeven, T. Donkers, and O. Bosgra, "Iterative learning control for uncertain systems: Robust monotonic convergence analysis," *Automatica*, vol. 45, no. 10, pp. 2383–2391, 2009.
- [15] T. D. Son, G. Pipeleers, and J. Swevers, "Robust monotonic convergent iterative learning control," *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 1063–1068, 2015.
- [16] E. C. Balta, A. Iannelli, R. S. Smith, and J. Lygeros, "Regret analysis of online gradient descent-based iterative learning control with model mismatch," *arXiv preprint arXiv:2204.04722*, 2022.
- [17] A. Martin, L. Furieri, F. Dörfler, J. Lygeros, and G. Ferrari-Trecate, "Safe control with minimal regret," in *Learning for Dynamics and Control Conference*. PMLR, 2022, pp. 726–738.
- [18] E. Hall and R. Willett, "Dynamical models and tracking regret in online convex programming," in *Proceedings of the 30th International Conference on Machine Learning*, vol. 28, no. 1, 2013, pp. 579–587.
- [19] K. Åström and B. Wittenmark, *Adaptive Control*, ser. Dover Books on Electrical Engineering. Dover Publications, 2008.
- [20] A. Rezaeizadeh and R. S. Smith, "Iterative learning control for the radio frequency subsystems of a free-electron laser," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 5, pp. 1567–1577, 2017.
- [21] M. Afrasiabi, C. Luthi, M. Bambach, and K. Wegener, "Smoothed particle hydrodynamics modeling of the multi-layer laser powder bed fusion process," *Procedia CIRP*, 2022.