2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Convergence of Gradient-based MAML in LQR

Negin Musavi[1], and Geir E. Dullerud[2]

I. Abstract

The main objective of this research paper is to investigate the local convergence characteristics of Model-agnostic Meta-learning (MAML) when applied to linear system quadratic optimal control (LQR). MAML and its variations have become popular techniques for quickly adapting to new tasks by leveraging previous learning knowledge in areas like regression, classification, and reinforcement learning. However, its theoretical guarantees remain unknown due to non-convexity and its structure, making it even more challenging to ensure stability in the dynamic system setting. This study focuses on exploring MAML in the LQR setting, providing its local convergence guarantees while maintaining the stability of the dynamical system. The paper also presents simple numerical results to demonstrate the convergence properties of MAML in LQR tasks.

II. Introduction

In the field of machine learning, the ability to quickly learn a new task with limited data by utilizing prior learning experiences is highly desirable. This approach is known as meta-learning or learning from learning. MAML is a well-known approach that can train machine learning models to swiftly adapt to new tasks with only a small amount of task-specific data. The concept behind MAML is to train a model on a group of related tasks so that it can acquire a useful starting parameter for a new task. MAML has gained recognition in scenarios such as few-shot image classification and reinforcement learning.

This paper focuses on applying MAML to a specific class of linear system quadratic optimal controllers, known as LQR. The goal is to examine the convergence of the algorithm for a collection of LQR tasks that vary in their system parameters or cost parameters.

There is a substantial body of research demonstrating the success of MAML through empirical studies on regression, classification, and reinforcement learning [1], [2], [3], [4]. However, there are only a few works that have analyzed the theoretical guarantees of this algorithm. To give a few examples, the work in [5], [6], [7] have established local convergence guarantees for MAML in supervised learning such as classification and regression. The works in [8], [9] have explored its global convergence in specific settings. In reinforcement learning, the local convergence properties of the algorithm have been studied in [10] under the assumption of having access to biased stochastic gradients and that stability of the dynamical system at each iteration of the algorithm. Another relevant study is the examination of the optimization landscapes of MAML in the LQR setting [11]. This work provides conditions for stability and global convergence of the algorithm in the single task LQR setting, but doesn't include conditions for the multi-task setting.

This paper presents the first examination of the convergence of the MAML algorithm for the multi-task LQR problem while also presenting conditions to guarantee the stability of the dynamical system. The MAML objective function for LQR tasks is not convex and does not inherit the gradient dominance property of the LQR cost function, which would have been helpful in analyzing convergence and ensuring the global convergence of gradient-based algorithms for the LQR problem. The study focuses on the local convergence properties of the algorithm and provides conditions to ensure the stability of the dynamical system. Whether global convergence is possible remains an outstanding research problem.

III. Preliminaries and Problem Statement

In this section, we present the basics of the standard LQR problem followed by a brief overview of the gradient-based MAML method and its modification for LQR tasks.

A. Notation

We use the following mathematical notation throughout the paper. The set of real numbers is denoted by $\mathbb{R}$. For a real matrix Z, Z^T represents its transpose, $\|Z\|$ its maximum singular value, $\|Z\|_F$ its Frobenius norm, $\text{tr}(Z)$ its trace, $\sigma_{\min}(Z)$ its minimum singular value, and $\text{vec}(Z)$ its vectorization obtained by stacking its columns. For a real square matrix Z, its spectral radius is denoted by $\text{rad}(Z)$. For a real symmetric matrix Z, $Z \succ 0$ and $Z \succeq 0$ indicate that Z is positive definite and positive semi-definite, respectively. The open ball of radius $r > 0$ centered at $Z_0 \in \mathbb{R}^{n\times m}$, is defined as $\mathcal{B}(Z_0, r) = \{Z \in \mathbb{R}^{n\times m} : \|Z - Z_0\| < r\}$. For matrices Z_1, Z_2, $\langle Z_1, Z_2\rangle = \text{tr}(Z_1^T Z_2)$ denotes their inner product and $Z_1 \otimes Z_2$ denotes their Kronecker product.

B. Standard LQR Problem

Consider the infinite horizon discrete-time LQR problem,

$$\begin{aligned} &\text{minimize}_{u(.)} \quad \mathbb{E}\left[\sum_{t=0}^{\infty}(x_t^T Q x_t + u_t^T R u_t)\right] \\ &\text{subject to} \quad x_{t+1} = Ax_t + Bu_t, \quad x_0 \sim \mathcal{D}, \end{aligned} \tag{1}$$

[1]N. Musavi is with the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign nmusavi2@illinois.edu
[2]G. E. Dullerud is with the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign dullerud@illinois.edu

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

where the initial state x_0 is randomly drawn from a distribution $\mathcal{D}$. The matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ represent the system dynamics, and $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ parameterize the cost, with Q and R being positive definite matrices. For a control policy at time $t \geq 0$ that is parameterized by a matrix $W \in \mathbb{R}^{m\times n}$ given by $u_t = -Wx_t$ the cost can be expressed as:

$$C(W) = \mathbb{E}_{x_0\sim\mathcal{D}}\left[\sum_{t=0}^{\infty} x_t^T(Q + W^T RW)x_t\right]. \tag{2}$$

This problem has an optimal solution as:

$$W^*_{lqr} = (R + B^T PB)^{-1}B^T PA,$$

where $P \succ 0$ satisfies the Algebraic Riccati Equation $P = Q + A^T PA + A^T PB(R + B^T PB)^{-1}B^T PA$.

Clearly W is an stabilizing parameter for the LQR problem in (1) with system parameters (A, B) if rad$(A-BW) < 1$. Now we present explicit formulas for LQR cost, its gradient and its Hessian at an stabilizing W which will be used later in our analysis.

- **LQR cost**: We can express the LQR cost as

  $$C(W) = \mathbb{E}_{x_0\sim\mathcal{D}}\left[x_0^T P_W x_0\right],$$

  with P_W satisfying the following Lyapunov equation

  $$P_W = Q + W^T RW + (A-BW)^T P_W(A-BW). \tag{3}$$

- **LQR cost gradient**: We can express the gradient of the LQR cost as

  $$\nabla C(W) = 2E_W\Sigma_W,$$

  with $E_W = (R + B^T P_W B)W - B^T P_W A$ and $\Sigma_W = \mathbb{E}_{x_0\sim\mathcal{D}}\left[\sum_{t=0}^{\infty} x_t x_t^T\right]$ the state correlation matrix. If we further assume that $\mathbb{E}_{x_0\sim\mathcal{D}}\left[x_0 x_0^T\right]$ is full rank, then the unique solution for $\nabla C(W) = 0$ is W^*_{lqr}(for more details see [12]).
- **LQR cost Hessian**: The action of Hessian operator on $Y \in \mathbb{R}^{n\times m}$ can be expressed as [13]:

  $$\begin{aligned}\nabla^2 C(W)[Y, Y] = {} & 2\langle(R + B^T P_W B)Y\Sigma_W, Y\rangle \\ & - 4\langle B^T P'_W[Y](A - BW)\Sigma_W, Y\rangle,\end{aligned} \tag{4}$$

  with $P'_W[Y] \succ 0$ satisfying the following equation $P'_W[Y] = (A - BW)^T P'_W[Y](A - BW) + Y^T E_W + E_W^T Y$.

C. MAML

MAML was first introduced in [14]. It is a meta-learning approach that aims to learn a good initialization for a model, such that it can quickly adapt to new tasks. More formally, suppose we have a set of tasks $\mathcal{I} = \{\mathcal{I}_1, \cdots \mathcal{I}_K\}$ drawn from distribution $p(\mathcal{I})$. Let the objective of task $\mathcal{I}_i$ be as minimizing a loss function that is parameterized by parameter W, i.e. $\min_W \mathcal{L}_i(W)$.

The idea behind MAML, under the assumption that the task can be solved by gradient descent, is to fine-tune parameter W for set of tasks $\mathcal{I}$ such that one or a few gradient steps can be taken with respect to a particular task $\mathcal{I}_i$ allowing it to serve as a good initialization for a new task drawn from the same distribution as the other tasks. For instance, with one gradient step this is achieved by solving the following optimization problem:

$$\min_W \sum_{\mathcal{I}_i\sim p(\mathcal{I})} \mathcal{L}_i(W - \eta\nabla\mathcal{L}_i(W)),$$

with $\eta > 0$ as a step-size parameter.

D. Problem Statement

We devote this section to introducing the Gradient-based MAML for set of LQR tasks. Let us first outline the LQR task set $\mathcal{I}$.

Definition 1: Each LQR task $\mathcal{I}_i$ is defined by a tuple of matrices $\big(A_i, B_i, Q_i, R_i\big)$, with $A_i \in \mathbb{R}^{n\times n}$, $B_i \in \mathbb{R}^{n\times m}$, and $Q_i \in \mathbb{R}^{n\times n}$, $R_i \in \mathbb{R}^{m\times m}$ being positive definite matrices. The cost function associated with each task, $C_i(W) : \mathbb{R}^{m\times n} \to \mathbb{R}^{\geq 0}$, is parameterized by $W \in \mathbb{R}^{m\times n}$ as in (2).

The objective is to find the best W for the tasks $\mathcal{I}$ using MAML approach such that it can be a good starting point for a new LQR task. This amounts to the following MAML objective function:

$$F(W) := \sum_{\mathcal{I}_i\sim p(\mathcal{I})} F_i(W) \tag{5}$$

with $F_i(W) = C_i(W - \eta\nabla C_i(W))$ and $\eta > 0$ as the step-size parameter. The MAML objective is to be minimized using gradient descent, with a single step update rule as follows

$$W \leftarrow W - \beta\nabla F(W), \tag{6}$$

with step-size parameter $\beta > 0$. This is called gradient-based MAML and is depicted in Algorithm 1. The step-sizes η and β are also known as the inner-loop and outer-loop step-sizes. Define $G_i(W) := \big(I - \eta\nabla^2 C_i(W)\big)$, then we have:

$$\begin{aligned}\nabla F(W) &= \sum_{\mathcal{I}_i\sim p(\mathcal{I})} \nabla F_i(W) \\ &= \sum_{\mathcal{I}_i\sim p(\mathcal{I})} G_i(W)\nabla C_i(W - \eta\nabla C_i(W)).\end{aligned} \tag{7}$$

Algorithm 1 gradient-based MAML for LQR tasks.

```
Initialize W
while not done do
    Choose step-size parameters η and β
    for all tasks ∈ I do
        Evaluate ∇C_i(W)
        Compute adapted parameter with a gradient
        step: W̃_i = W − η∇C_i(W)
        Evaluate ∇C_i(W̃_i) and G_i(W)
    Update W ← W − β Σ_{I_i∼p(I)} G_i(W)∇C_i(W̃_i)
return Estimation of MAML optimal parameter Ŵ*_maml
```

IV. Convergence Analysis of Gradient-based MAML for Multi-task LQR Problem

In this section, we will examine the local convergence of gradient-based MAML for a set of LQR tasks. We will begin by presenting definitions, assumptions, and lemmas, and then employ them to prove the desired result. For the sake of conciseness, only some of the results will be accompanied by their proofs. We commence with the following definition.

Definition 2: We say:

- parameter W is task-stabilizing for task $\mathcal{I}_i$ if $\text{rad}(A_i - B_iW) < 1$; and
- parameter W is MAML-stabilizing if it is task-stabilizing for all tasks in $\mathcal{I}$ and $\text{rad}(A - B(W - \eta\nabla C_i(W))) < 1$ holds for all tasks.

Suppose parameter W is task-stabilizing for all tasks. Define $\mu = \sigma_{min}\left(\mathbb{E}_{x_0\sim\mathcal{D}}[x_0x_0^T]\right)$, and $\delta_i(W)$ as

$$\delta_i(W) = \frac{\sigma_{\min}(Q_i)\mu}{4C_i(W)\|B_i\|(\|A_i - B_iW\| + 1)}.$$

Also let $\delta(W) = \min_{\mathcal{I}_i}\delta_i(W)$. Given these, the following lemma provides a sufficient condition on η to ensure W be also MAML-stabilizing.

Lemma 1: Given that W is task-stabilizing for all tasks, if the step-size η satisfies the condition $\eta < \min_{\mathcal{I}_i}\left\{\frac{\delta_i(W)}{\|\nabla C_i(W)\|}\right\}$, then it is also MAML-stabilizing.

This outcome is a direct consequence of the results in [12]. Before presenting the convergence theorem for Algorithm 1, we need to make certain assumptions regarding each task's cost function.

Lemma 2: Given that W and U are task-stabilizing for task $\mathcal{I}_i$, there exists $\theta_i(W)$, $\ell_i(W)$, $\rho_i(W)$, and $H_i(W)$ such that:

$$\begin{aligned}
&|C_i(U) - C_i(W)| \leq \mathbb{E}_{x_0\sim\mathcal{D}}\big[\|x_0\|^2\big]\theta_i(W)\|U - W\|,\\
&\|\nabla C_i(U) - \nabla C_i(W)\| \leq \ell_i(W)\|U - W\|,\\
&\|\nabla^2 C_i(U) - \nabla^2 C_i(W)\| \leq \rho_i(W)\|U - W\|,\\
&\text{vec}\big(\nabla C_i(U) - \nabla C_i(W)\big) = H_i(W)\text{vec}\big(U - W\big).
\end{aligned}$$

The first three inequalities indicate that the LQR cost function, its gradient, and its Hessian are locally Lipschitz continuous at any stabilizing W. Hence an upper bound for $\theta_i(W)$ and $\ell_i(W)$ are provided in references [12], [15] in terms of W, system parameters and cost parameters. We have also obtained an upper bound for $\rho_i(W)$, as presented in Lemma 7, with a brief proof that is available in [16] Appendix B for better understanding. The final statement follows directly from the mean-value theorem applied to the LQR gradient, as the LQR gradient is locally Lipschitz continuous at any stabilizing W. Lemma 9 provides the details of this and its proof is also included in [16] Appendix B. The subsequent assumption pertains to the set of LQR problems that will be considered.

Assumption 1: The set of LQR tasks $\mathcal{I}$ share the same system matrices, namely $A_i = A$ and $B_i = B$ for all $\mathcal{I}_i$. The cost parameters Q_i and R_i vary among the tasks.

Furthermore, similar to the analysis presented in [5], it is required that the variance of the LQR cost gradient be bounded over the set of tasks $\mathcal{I}$. This requirement is further elaborated in the following assumption. The set of all parameters that stabilize task $\mathcal{I}_i$ is denoted as $\mathcal{S}_i = \{W : \text{rad}(A_i - B_iW) < 1\}$. Also, let $\mathcal{S} = \cap_i\mathcal{S}_i$. Under Assumption 1, the sets $\mathcal{S}_i$ are identical for all tasks, and therefore $\mathcal{S}$ is non-empty and identical to each $\mathcal{S}_i$.

Assumption 2: Let $\mathcal{W}$ be a compact subset of $\mathcal{S}$. There exists a constant $\sigma > 0$ such that for all $W \in \mathcal{W}$

$$\mathbb{E}_i\Big[\|\nabla C_i(W) - \mathbb{E}_i\big[\nabla C_i(W)\big]\|_F^2\Big] \leq \sigma^2.$$

With these assumptions in place, we can proceed to the next two lemmas. The following lemma provides a sufficient condition on β to ensure that $W - \beta\nabla F(W)$ is task-stabilizing for all tasks given that W is MAML-stabilizing for all tasks.

Lemma 3: Let W be MAML-stabilizing parameter for all tasks in Assumption 1. Define $\bar{\delta}_i(W)$ as:

$$\bar{\delta}_i(W) = \frac{\sqrt{\|A - BW\|^2 + \frac{\mu\sigma_{min}(Q_i)}{C_i(W)}} - \|A - BW\|}{\|B\|}.$$

If β satisfies the following condition

$$\beta < \min_{\mathcal{I}_i}\left\{\frac{\bar{\delta}_i(W)}{\|\nabla F(W)\|}\right\},$$

then $W - \beta\nabla F(W)$ is task-stabilizing for all tasks in $\mathcal{I}$.

The proof of this lemma is deferred to [16] Appendix C. Now let $\ell(W) = \max_i \ell_i(W)$, $\bar{\ell}(W) = \max_i \ell_i(W - \eta\nabla C_i(W))$, $\|H(W)\| = \max_i\|H_i(W)\|$, and $\rho(W) = \max_i \rho_i(W)$ for a given W. Then we have:

Lemma 4: Consider the MAML objective in (5). Suppose W and U are task-stabilizing for all tasks in Assumption 1. If the learning rate $\eta > 0$ is such that $\eta < \min_{\mathcal{I}_i}\left\{\frac{\delta(W)}{\|\nabla C_i(W)\|}, \frac{\delta(U)}{\|\nabla C_i(U)\|}\right\}$, then we have:

$$\|\nabla F(U) - \nabla F(W)\| \leq L(W)\|U - W\|,$$

with $L(W) = \bar{\ell}(W)\big(1 + \eta(\ell(W) + \delta(W)\rho(W))\big)(1 + \eta\ell(W)) + \eta\rho(W)(1 + \eta\|H(W)\|)\|\mathbb{E}_i\big[\|\nabla C_i(W)\|_F\big]$.

The proof of this lemma is derived from the definition of the MAML objective, by applying Lemmas 1 and 2, by utilizing Assumption 2, and follows a similar approach to [5]. The proof of this lemma is deferred to [16] Appendix C. Also consider the following conditions for both step-sizes.

Condition 1: For a task-stabilizing W_j of all tasks at iteration j of Algorithm 1, consider the following conditions on η:

$$\eta_j < \min_{\mathcal{I}_i} \left\{ \frac{1}{4\|\nabla^2 C_i(W_j)\| + 2\rho_i(W_j)\delta_i(W_j)}, \frac{(1-\alpha)\delta(W_j)}{\alpha\delta(W_j)\ell_i(W_j) + \|\nabla C_i(W_j)\|} \right\},$$

for some $\alpha \in (0,1)$, and then on β:

$$\frac{a}{L(W_j)} < \beta_j < \frac{b}{L(W_j)} < \min_{\mathcal{I}_i} \left\{ \frac{\alpha\bar{\delta}_i(W_j)}{\|\nabla F(W_j)\|} \right\},$$

for some a, b such that $2a > b^2$.

The stability of the linear system during the optimization and the convergence of the algorithm, as stated in the following theorem, depend on satisfying these conditions for η and β. Specifically, the last conditions on η and β serve the former purpose, while the first condition on η and the first two conditions on β are for the latter.

Theorem 1: Let $\mathcal{W}$ be a compact subset of $\mathcal{S}$. Initialize the Algorithm 1 with a task-stabilizing W_0 for all the tasks in Assumption 1 such that $W_0 \in \mathcal{W}$. Let $W^*_{maml} := \arg\min_W F(W)$, and define the sub-optimality gap $\Delta = F(W_0) - F(W^*_{maml})$. Given that Assumption 2 holds, if the step-sizes η and β satisfy the conditions in Condition 1 at each iteration, then for any desired accuracy $\epsilon > 0$, there exist constants $\tilde{c}_1, \tilde{c}_2 > 0$ such that after running Algorithm 1 for at most $\dfrac{\Delta\big(\tilde{c}_1 + \tilde{c}_2(\sigma+\epsilon)\big)}{(2a-b^2)\epsilon^2}$ iterations, it will find a solution W_ϵ that satisfies

$$\|\nabla F(W_\epsilon)\| \le \epsilon,$$

where σ is the variance constant from Assumption 2, and a, b are constants defined in Condition 1.

The theorem states that with proper choices of step-sizes, MAML will reach a stationary point, characterized by a small gradient magnitude ($\le \epsilon$), in a number of iterations that grows inversely proportional to the square of ϵ. In other words, the computational cost of finding the stationary point scales as $\mathcal{O}(1/\epsilon^2)$. The proof of this theorem is deferred to [16] Appendix D to enhance readability.

V. Simulation Results

In this part we provide simulation results for the convergence of the gradient descent-based MAML algorithm. Specifically, we tested the algorithm on ten LQR tasks with identical A and B matrices, where $A = \begin{pmatrix} 1.5 & 0 \\ 0 & -2 \end{pmatrix}$ and $B = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$. The LQR tasks are constructed with different cost parameters Q_i and R_i, where Q_i are a linear combination of $\bar{Q}_1 = \begin{pmatrix} 0.01 & -0.5 \\ -0.5 & 200 \end{pmatrix}$ and $\bar{Q}_2 = \begin{pmatrix} 200 & 1 \\ 1 & 0.01 \end{pmatrix}$, and R_i is a linear combination of $\bar{R}_1 = 2$ and $\bar{R}_2 = 0.1$ as:

$$Q_i = \alpha_{1i}\bar{Q}_1 + \alpha_{2i}\bar{Q}_2$$
$$R_i = \alpha_{3i}\bar{R}_1 + \alpha_{4i}\bar{R}_2,$$

with coefficients α_{1i}, α_{2i}, α_{3i}, and α_{4i} that are randomly drawn from the interval $(0, 10)$.

We run Algorithm 1 with four different initialization and choices of step-sizes η, β. The results of the experiment, which ran for 50 iterations, are presented in Figure 1. In three of the runs, we kept the inner-loop step-size η constant throughout the optimization process, while in one run, we varied it. In the plot where the initialization W_0^4 is far from the converging point, we started with a small η ($= 0.00005$) and gradually increased it. The reason for this is that in this case, W_0^4 is close to the boundary of the stabilizing set $\mathcal{S}$ and requires a lower η to ensure task stability of all tasks. As W_j moves away from the boundaries of the set $\mathcal{S}$, a higher η can be chosen to speed up the convergence, so we increased the step-size accordingly. It is noted that in this case at iteration 50 the step-size $\eta = 0.0066$.

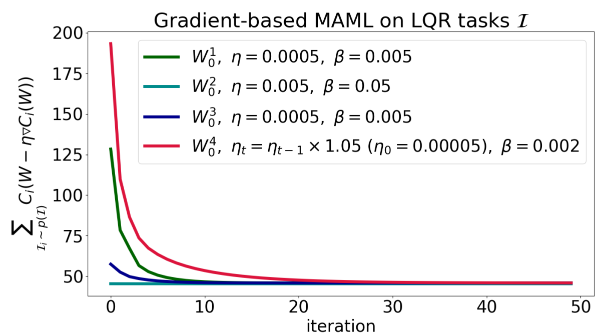

Fig. 1. Convergence of Algorithm 1 for tasks $\mathcal{I}$ with different initialization W_0^1,W_0^2, W_0^3, and W_0^4.

The optimal parameter returned by the Algorithm 1, i.e. $\tilde{W}^*_{maml}$, for the four runs is recorded and is presented in Table I. Here, w_1 and w_2 represent the first and second entries of W. To demonstrate the efficiency of Algorithm 1, it is necessary to provide a measure of how closely these estimated optimal parameters match the MAML objective in (5). Clearly, the MAML objective depends on the inner-loop step-size η. Therefore, we girded the set of stabilizing parameters $\mathcal{S}$ with a resolution of 0.05 and computed the MAML objective on the grid points using different η values. We then searched for the optimal parameter for each and recorded the results in Table II. According to these results, the optimal parameter is located within a ball of radius 0.23 centered at $[1.347, -2.102]$ which we call it MAML

TABLE I
Output of the Algorithm 1 for different runs.

W_0	W_0^1	W_0^2	W_0^3	W_0^4
$\hat{W}^*_{maml}$	$w_1 = +1.38$	$w_1 = +1.36$	$w_1 = +1.38$	$w_1 = +1.37$
	$w_2 = -2.13$	$w_2 = -2.12$	$w_2 = -2.13$	$w_2 = -2.13$

optimal neighborhood. On closer inspection of the output of Algorithm 1 recorded in Table I, we find that all the parameters lie within this optimal neighborhood. Therefore, we can conclude that Algorithm 1 converges to a MAML optimal neighborhood estimated by girding the $\mathcal{S}$.

TABLE II
THE MAML OBJECTIVE'S OPTIMAL POINT CALCULATED BY SEARCHING A GIRD CREATED OVER THE SET $\mathcal{S}$.

η	0.0005	0.0050	0.0066
W^*_{maml}	$w_1 = +1.402$	$w_1 = +1.292$	$w_1 = +1.347$
	$w_2 = -2.102$	$w_2 = -2.102$	$w_2 = -2.102$

VI. Conclusion and Future Work

We studied the convergence of MAML for a set of specific LQR tasks that share the same dynamics. We derived that with conditions on both step-sizes the linear system stability is guaranteed and MAML reach any desired accuracy level of $\epsilon \geq 0$, with a computational cost that scales as $\mathcal{O}(1/\epsilon^2)$. Our focus was on the model-based version of the LQR problem, where we have exact information about the cost, its gradient, and its Hessian for a specific policy. However, we would like to point out that these results could also be extended to model-free scenarios, where the cost gradient and Hessian must be estimated from data. We aim to explore these model-free extensions in our future work.

VII. Acknowledgement

The authors would like to thank Sanjay Shakkottai, Dawei Sun, and Sayan Mitra for the discussions that helped lead to this work.

References

[1] M. Al-Shedivat, T. Bansal, Y. Burda, I. Sutskever, I. Mordatch, and P. Abbeel, "Continuous adaptation via meta-learning in nonstationary and competitive environments," *arXiv preprint arXiv:1710.03641*, 2017.
[2] H. S. Behl, A. G. Baydin, and P. H. Torr, "Alpha maml: Adaptive model-agnostic meta-learning," *arXiv preprint arXiv:1905.07435*, 2019.
[3] E. Grant, C. Finn, S. Levine, T. Darrell, and T. Griffiths, "Recasting gradient-based meta-learning as hierarchical bayes," *arXiv preprint arXiv:1801.08930*, 2018.
[4] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv preprint arXiv:1803.02999*, 2018.
[5] A. Fallah, A. Mokhtari, and A. Ozdaglar, "On the convergence theory of gradient-based model-agnostic meta-learning algorithms," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 1082–1092.
[6] T. Johnson and S. Mitra, "Safe flocking in spite of actuator faults using directional failure detectors," *Journal of Nonlinear Systems and Applications*, vol. 2, no. 1-2, pp. 73–95, 2011.
[7] M. Abbas, Q. Xiao, L. Chen, P.-Y. Chen, and T. Chen, "Sharp-maml: Sharpness-aware model-agnostic meta learning," in *International Conference on Machine Learning*. PMLR, 2022, pp. 10–32.
[8] L. Collins, A. Mokhtari, S. Oh, and S. Shakkottai, "Maml and anil provably learn representations," in *International Conference on Machine Learning*. PMLR, 2022, pp. 4238–4310.
[9] H. Wang, Y. Wang, R. Sun, and B. Li, "Global convergence of maml and theory-inspired neural architecture search for few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9797–9808.
[10] A. Fallah, K. Georgiev, A. Mokhtari, and A. Ozdaglar, "On the convergence theory of debiased model-agnostic meta-reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3096–3107, 2021.
[11] I. Molybog and J. Lavaei, "When does maml objective have benign landscape?" in *2021 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2021, pp. 220–227.
[12] 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.
[13] J. Bu, A. Mesbahi, M. Fazel, and M. Mesbahi, "Lqr through the lens of first order methods: Discrete-time case," *arXiv preprint arXiv:1907.08921*, 2019.
[14] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.
[15] J. P. Jansch-Porto, B. Hu, and G. Dullerud, "Policy optimization for markovian jump linear quadratic control: Gradient-based methods and global convergence," *arXiv preprint arXiv:2011.11852*, 2020.
[16] N. Musavi and G. Dullerud, "Convergence of gradient-based maml in lqr," *arXiv preprint arXiv:2309.06588*, 2023.