

When Does MAML Objective Have Benign Landscape?

Igor Molybog and Javad Lavaei

Abstract—The paper studies the landscape of the optimization problem behind the Model-Agnostic Meta-Learning (MAML) algorithm. The goal of the study is to determine the global convergence of MAML on sequential decision-making tasks possessing a common structure. We investigate in what scenarios the benign optimization landscape of the underlying tasks results in a benign landscape of the corresponding MAML objective. For illustration, we analyze the landscape of the MAML objective on LQR tasks to determine what types of similarities in their structures enable the algorithm to converge to the globally optimal solution.

I. INTRODUCTION

Meta-learning, along with transfer learning, is a rapidly developing research area in machine learning which aims to design algorithms that gain computational advantages out of inherent similarities between learning problem instances, otherwise referred to as tasks. In this work, we study one of the most popular meta-learning algorithms, named *Model-Agnostic Meta-Learning (MAML)*, which has been developed in [1]. In the reinforcement learning domain, the algorithm is expected to rapidly adapt a pre-learned policy to a new task. However, it remains unclear how to measure the quality of adaptation and what tasks are suitable for the meta-learning. In the core of meta-learning, there is an optimization problem that is concerned with the expected return averaged among considered tasks. In this paper, we propose to measure the performance of MAML with respect to the optimality gap of the corresponding optimization problem. We consider a space of tasks to be suitable for meta-learning if the algorithm converges to a global optimizer of the meta-learning objective or to a point with a similar value. This approach enables distinguishing those meta-learning problems that are solvable by MAML from other problems. Intuitively, a particular algorithm should perform satisfactorily on a set of meta-learning problems that consist of tasks united by a particular type of similarity. For the purpose of demonstration, we consider linear quadratic control problems, although our theory applies to a broad class of RL tasks with benign optimization landscape. The term benign landscape in this paper is used to describe optimization problems which can be (approximately) solved with common tractable optimization techniques. It is formalized in Section II. We aim to theoretically study the global convergence properties of the original MAML algorithm on sequential decision-making tasks. In short, our findings can be summarised as follows:

This work was supported by grants from AFOSR, ONR, ARO and NSF. The authors are with the University of California, Berkeley (email: igormolybog@berkeley.edu, lavaei@berkeley.edu).

- Meta-Learning objective inherits a benign landscape from the objectives of the individual tasks if they are similar pointwise. As a result, the original MAML and other meta-learning algorithms that rely on local search are guaranteed to perform well on the corresponding problems.
- As a strongly negative result, those problems consisting of linear quadratic tasks that coincide up to a scaling of the reward function are not solvable by MAML. We propose an alternative scheme that addresses the issue with this type of similarity.

For the clarity of explanation, we investigate discrete stationary infinite-horizon decision problems and note that the generalization of the results to non-stationary and finite-horizon cases is straightforward. A stationary discrete dynamical system is described as

$$s_{t+1} \sim T(s_t, a_t),$$

where T is a probability distribution over the next state $s_{t+1} \in \mathbb{R}^d$ given the current state $s_t \in \mathbb{R}^d$ and action $a_t \in \mathbb{R}^r$. The initial state s_0 is assumed to follow the distribution T_0 . The objective of the infinite-horizon problem is to find a control input a_t minimizing the total discounted cost (the negation of the reward)

$$\begin{aligned} & \text{minimize} && \mathbb{E}_{s_0 \sim T_0} \left[\sum_{t=0}^{\infty} \gamma^t c(s_t, a_t) \right] \\ & \text{subject to} && s_{t+1} \sim T(s_t, a_t) \quad \forall t \in \{0, 1, \dots\} \end{aligned}$$

where $\gamma \in (0, 1]$ is the discount factor that is assumed to be 1 for the LQR problem and $\mathbb{E}[\cdot]$ denotes the expectation operator.

A. Linear-Quadratic Regulator

One of the important examples of decision-making problems is related to the control of linear dynamical systems with a quadratic objective, referred to as *linear-quadratic regulator (LQR)*. LQR and Iterative LQR (iLQR) [2] are fundamental tools for model-based reinforcement learning. Moreover, linear-quadratic problems enjoy a rich theoretical foundation with a large number of provable guarantees, which make them a compelling benchmark for the mathematical analysis of novel learning algorithms.

We consider the following infinite-horizon exact LQR problem:

$$\begin{aligned} & \text{minimize} && \mathbb{E}_{x_0} \left[\sum_{t=0}^{\infty} (s_t^\top Q s_t + a_t^\top R a_t) \right] \\ & \text{subject to} && s_{t+1} = A s_t + B a_t, \quad \forall t \in \{0, 1, \dots\} \end{aligned} \tag{1}$$

where Q and R are positive semidefinite matrices. Assuming that the matrices A and B are such that the optimal cost is finite, it is a classic result that the optimal control policy is deterministic and linear in the state, i.e.,

$$a_t = -W^* s_t$$

where $W^* \in \mathbb{R}^{r \times d}$ [3]. Moreover, the matrix W^* can be found from the model parameters by solving the Algebraic Riccati Equation (ARE)

$$P = A^\top P A + Q - A^\top P B (B^\top P B + R)^{-1} B^\top P A,$$

and substituting the positive-definite root P into

$$W^* = (B^\top P B + R)^{-1} B^\top P A.$$

This implies that in order to find the solution of LQR, it suffices to only search over deterministic policies of the form $a = -W s$ parameterized with a matrix $W \in \mathbb{R}^{r \times d}$.

In an effort to build a bridge between practical RL algorithms and the optimal control theory, [4] shows that W^* can be found by applying the policy gradient algorithm to a reformulated cost function. The LQR cost of a linear deterministic policy with respect to W can be defined as

$$C(W) := \mathbb{E}_{s_0 \sim T_0} \left[\sum_{t=0}^{\infty} (s_t^\top Q s_t + a_t^\top R a_t) \right]$$

where $a_t = -W s_t$ and $s_{t+1} = (A - BW) s_t$. This can be reformulated as

$$C(W) = \mathbb{E}_{s_0 \sim T_0} s_0^\top P_W s_0$$

where P_W is the solution of $P_W = Q + W^\top R W + (A - BW)^\top P_W (A - BW)$. We only consider the cost of stable policies, and assume the cost to be infinite for unstable ones.

B. Model-Agnostic Meta-Learning

Given a set of tasks \mathcal{T} , each represented by an objective function \mathcal{L}_τ , and a probability distribution $\mathbb{P}_\mathcal{T}$ over the tasks, [1] proposes an algorithm for finding an initialization of the policy gradient method that allows a fast adaptation to a task through just several gradient updates. In case the task consists in regression, classification, or clusterization, \mathcal{L}_τ is a risk or an empirical risk. In case of a reinforcement learning task, \mathcal{L}_τ is the cost (the negated return) of a policy.

If the space of considered policies is parameterized via $w \in \mathcal{W}$, then a single-shot MAML (which uses just one gradient update) aims to minimize the objective

$$L(w) = \mathbb{E}_{\tau \in \mathcal{T}} [f_\tau(w - \eta \nabla g_\tau(w))] \quad (2)$$

where f_τ and g_τ can be two different approximations of the objective function of the task τ and ∇ is the gradient operator. Similarly, a multi-shot version applies multiple gradient updates within the adaptation procedure. For example, if \mathcal{L}_τ is the risk of a learning problem, then f_τ can be the empirical risk conditioned on a large dataset, while g_τ is the empirical risk conditioned on a smaller dataset. The generality of this formulation will be used in Section II, but for the study of LQR we will assume that the functions \mathcal{L}_τ , f_τ and g_τ

all coincide and are equal to $C(W)$. When it is clear from the context which task is being discussed, we will omit the subscript.

Being based upon gradient descent with a constant step size, Algorithm 1 is a basic version of a few-shot MAML although other versions have been developed in the literature, e.g. FO-MAML, HF-MAML [5], iMAML [6], Reptile [7] and FTML [8]. We do not directly address the other formulations in the paper, but the conclusions of this work are applicable to them as well since they are created to minimize essentially the same objective function (2) and the convergence to at least a first-order stationary point has been proven for the majority of these variants.

An interesting modification of MAML is based on exact proximal optimization as an alternative to a gradient update. [9] proposes the proximal update MAML algorithm with the adaptation procedure that finds the best set of parameters in the neighborhood of initialization, while [10] proves that a similar idea can be shaped into a meta-RL algorithm that is proven to converge globally under some overparametrization assumption. However, the exact optimization in the adaptation procedure may be problematic in meta-learning setup, since adaptation is supposed to be fast, meaning that there are sharp constraints on sample complexity and the amount of computation allowed for it.

Algorithm 1 Model-Agnostic Meta-Learning (MAML)

Require: $p(\mathcal{T})$: Probabilistic task generator

Require: η, β : Step size hyperparameters

- 1: Randomly initialize w
 - 2: **while** not done **do**
 - 3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
 - 4: **for all** \mathcal{T}_i **do**
 - 5: Evaluate $\nabla g_{\mathcal{T}_i}(w)$
 - 6: Compute adapted parameters with gradient descent:
 $w'_i = w - \eta \nabla_w g_{\mathcal{T}_i}(w)$
 - 7: **end for**
 - 8: Update $w \leftarrow w - \beta \nabla_w \sum_{\mathcal{T}_i \sim p(\mathcal{T})} f_{\mathcal{T}_i}(w'_i)$
 - 9: **end while**
-

As meta-learning seeks to improve learning performance by exploiting similarities between tasks, it is important to understand what type of similarities a particular meta-learning algorithm can take advantage of. A highly desirable feature of a meta-learning algorithm is an acceptable meta-test performance at least on the tasks it has been meta-trained on. In other words, if the set of tasks \mathcal{T} is finite and the meta-training procedure has access to all of them, then it should succeed on the meta-testing stage. For MAML, this translates into a requirement of successful minimization of the objective (2). [8] shows that the global minimum is achieved in case the functions $f_\tau = g_\tau$ are all smooth and strongly convex. However, the objective of many practical decision-making tasks are not convex, although may possess benign landscape, like the LQR objective (1). Tasks with non-convex objectives are substantially harder to analyze,

and thus [5] only shows convergence of MAML to a first-order stationary point of (2) if f_τ are non-convex. We aim to study the global convergence properties of MAML applied to tasks with non-convex objectives.

The landscape of (1) has been studied in [4], which has observed that there exist instances of LQR that are not convex, quasi-convex, or star-convex, which means that none of the existing results on global convergence of MAML can be applied even to such a basic decision problem as LQR. However, (1) possess benign landscape, and our study shows that this property can be transferred to (2).

The purpose of this paper is not to design a new algorithm or conduct a study of its application on a specific real-world case, but rather to prove the basic properties of a popular existing algorithm for a vast variety of cases. Meta-Learning algorithms are supposed to capture similarities between tasks, although there is no clear way to determine whether or not a similarity has been captured. For MAML, we propose a criterion that is based on the properties of the landscape of the optimization associated with MAML. Specifically, we declare that a version of MAML captures the similarities between the tasks in \mathcal{T} if the objective of the algorithm on \mathcal{T} has benign landscape. Otherwise, the tasks in \mathcal{T} are recognized to be too distinct for this particular version of MAML.

The advantage of the above criterion is that it correlates with the computational complexity of the problem that MAML aims to solve. If the objective function has benign landscape, then the optimization problem has a low computational complexity and one can solve it by a local search method, which is implemented within MAML. If the objective does not have benign landscape, then MAML can become stuck in a spurious local minimum, which can potentially be arbitrarily worse than the optimal solution.

A drawback of this approach is that it does not allow to compare Meta-Learning algorithms against each other, since it does not take into account the computational complexity of the adaptation algorithm. In this paper, we do not consider this aspect because of our focus specifically on the few-shot MAML. To the best of our knowledge, this is the first paper to propose a systematic way of reasoning about the general theory of meta-adaptation between RL tasks. We consider MAML primarily with applications to linear-quadratic systems because they are realistic and yet easy to analyze although our conclusions go far beyond this application.

Notation

Given a set $\mathcal{Z} \subseteq \mathbb{R}^n$ and a point $\bar{z} \in \mathcal{Z}$, define the *open neighborhood* $\mathcal{B}_\delta(\bar{z}) = \{z \in \mathcal{Z} \mid \|z - \bar{z}\|_2 < \delta\}$. Given a function $\ell : \mathcal{Z} \rightarrow \mathbb{R}$, we call $\bar{z} \in \mathcal{Z}$ a *local minimizer* of the function ℓ if there exists $\delta > 0$ such that $\ell(z) \geq \ell(\bar{z})$ for all $z \in \mathcal{B}_\delta(\bar{z})$. The value $\ell(\bar{z})$ in this case is called a *local minimum*. A *global minimizer* of ℓ is a point \bar{z} such that $\ell(z) \geq \ell(\bar{z})$ for all $z \in \mathcal{Z}$. The *global minimum* is a value $\min_{z \in \mathcal{Z}} \ell(z)$ such that $\ell(z) \geq \min_{z \in \mathcal{Z}} \ell(z)$ for all $z \in \mathcal{Z}$. A *spurious local minimum* is a local minimum that

is not a global minimum. Given a differentiable function $\ell : \mathcal{Z} \rightarrow \mathbb{R}$, a *first-order stationary point* $\bar{z} \in \mathcal{Z}$ is such that $\nabla \ell(\bar{z}) = 0$. Given a twice differentiable function ℓ , a *first-order stationary point* \bar{z} such that $\nabla^2 \ell(\bar{z}) \succeq 0$ is called a *second-order stationary point*, where \succeq denotes the positive semidefinite matrix sign. Given a matrix $A \in \mathbb{R}^{n \times m}$, we denote its transpose as A^\top and its operator norm in the space ℓ_2 as $\|A\|$. For a vector v from an ℓ_2 -space, its norm is denoted with $\|v\|$. Cardinality of a set \mathcal{T} is denoted by $|\mathcal{T}|$.

II. MAIN RESULTS

In this section, we study the MAML algorithm under four different scenarios. We consider MAML applied to a single task and to several identical tasks. Afterwards, we introduce a metric between tasks and extend the study to a number of close tasks, and, finally, we study the convergence of MAML on a large number of distant LQR tasks.

We provide theoretical results for general multi-dimensional systems, while all of the presented examples and counter-examples are on one-dimensional LQR tasks since they are easy to visualize. Note that these examples are extendable to multi-dimensional systems as well. The details on the exact tasks used for the computations are moved to the Appendix.

A. Single task

We begin by analyzing MAML applied to a singleton task set \mathcal{T} . If MAML fails under this scenario, then its global convergence for the multiple-task scenario becomes questionable. For the single task, we rewrite the MAML objective (2) as

$$h(w) = f(w - \eta \nabla g(w))$$

where f is assumed to be a continuously differentiable function and g is assumed to be twice continuously differentiable. As noticed in Section I, the existing results on global convergence of MAML are not applicable to LQR. Figure 1 demonstrates an example of the MAML objective (2) applied to a single LQR task. It is non-convex and has three distinct strict local minimizers. Nevertheless, all of these three points are also global minimizers, which implies that Algorithm 1 would converge to its global minimizer from almost any initial point. The minimizer in the middle corresponds to W^* of the task, while the rightmost and leftmost minimizers are some points W such that $W - \eta \nabla C(W) = W^*$ and $\nabla C(W) \neq 0$. The minimizers on the both sides rely on the rapid adaptation during the meta-testing stage, which has been assumed in the design of the algorithm.

This example gives rise to a hypothesis that the MAML objective for a single LQR possesses some sort of benign landscape and, more generally, that benign landscape of the cost of the underlying task results in benign landscape of the resulting MAML objective. Following [11], we formalize the notion of benign landscape by defining global functions

Definition 1: A continuous function $\ell : \mathcal{Z} \rightarrow \mathbb{R}$ is called global if every local minimizer of ℓ is a global minimizer.

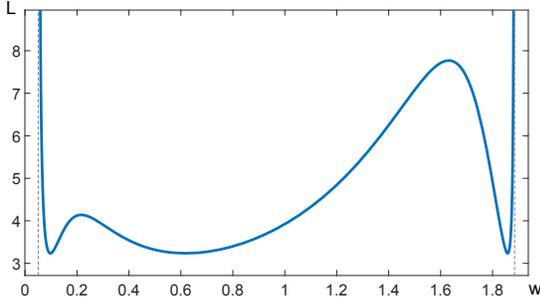


Fig. 1: MAML objective (2) on a single LQR task.

This property is also referred as having no spurious local minima and generalizes the notions of convexity, quasiconvexity, and star-convexity. As a relaxation of this property, we define ε -global function:

Definition 2: A continuous function $\ell : \mathcal{Z} \rightarrow \mathbb{R}$ is called ε -global if for every local minimizer \bar{z} of ℓ it holds that

$$\ell(\bar{z}) - \min_{z \in \mathcal{Z}} \ell(z) \leq \varepsilon$$

Being global is equivalent to being 0-global. This property is more likely to be satisfied than a perfect no-spurious property for cost functions coming from real-world applications, providing with a broader way to formalize the intuitive meaning of a benign landscape. These two notions of benign landscape characterize when a coercive function is easy to optimize using local optimization methods based on gradient descent. The following theorem shows that the benign landscape of the cost of the underlying task indeed leads to a benign landscape for the resulting MAML objective.

Theorem 1: Let $f : \mathcal{W} \rightarrow \mathbb{R}$ be global and $g : \mathcal{W} \rightarrow \mathbb{R}$ be twice continuously differentiable with $\|\nabla^2 g(w)\| \leq M < \infty$ for all $w \in \mathcal{W}$. If Algorithm 1 with the parameter η chosen to be smaller than $\frac{1}{M}$ converges to a local minimizer $w^* \in \mathcal{W}$ of h , then w^* is a global minimizer of h .

The proof of Theorem 1 is moved to [12] and relies significantly on the following technical Lemma:

Lemma 1: Let $\ell : \mathcal{Z} \rightarrow \mathbb{R}$ be an ε -global function, and consider a continuous map $\mathcal{F} : \mathcal{W} \rightarrow \mathcal{Z}$ with $\mathcal{Z} = \text{range}(\mathcal{F})$ that is locally open at a local minimizer \bar{w} of $\ell \circ \mathcal{F}$. Then, it holds that

$$\ell(\mathcal{F}(\bar{w})) - \min_{w \in \mathcal{W}} \ell(\mathcal{F}(w)) \leq \varepsilon$$

In the context of LQR, Theorem 1 results in the statement below.

Theorem 2: Let $w^* \in \mathbb{R}^{r \times d}$ be the limit point of the sequence generated by Algorithm 1 (MAML) with $\eta < \frac{1}{\|\nabla^2 C(w^*)\|_2}$ applied to a single LQR task, meaning that $h(W) = C(W - \eta \nabla C(W))$. Then, w^* is the global minimizer of h , which implies that $C(W - \eta \nabla C(W))$ is a global function.

Proof: It is shown in [5] that MAML converges to a first-order stationary point of a smooth nonconvex loss. Therefore, if the limit point w^* exists, it must be a stationary point of $C(W - \eta \nabla C(W))$. Consider the first-order station-

arity condition for w^* :

$$\begin{aligned} 0 &= \nabla C(w^* - \eta \nabla C(w^*)) = \\ &= [\mathcal{I} - \eta \nabla^2 C(w^*)]^\top \nabla C(W) \Big|_{W=w^* - \eta \nabla C(w^*)} \end{aligned}$$

Similar to the proof of Theorem 1, $\mathcal{I} - \eta \nabla^2 C(w^*)$ is a full-rank matrix, meaning that w^* is a first-order stationary point for the MAML objective if and only if $\nabla C(W) \Big|_{W=w^* - \eta \nabla C(w^*)} = 0$.

Theorem 7 of [4] states that the Gradient descent algorithm finds an ε -approximation of the global optimum of $C(W)$ in polynomial time for any initial point with a finite value. This directly implies that all first-order stationary points of $C(W)$ are the global minimizers of $C(W)$ because otherwise we could initialize the Gradient descent algorithm at a stationary point and since it converges to the point of initialization, it will lead to a contradiction. Being a first-order stationary point is a sufficient condition of local minimality, and hence guarantees that $C(W)$ is a global function. By Theorem 1, the MAML objective is global for a sufficiently small η .

Since $w^* - \eta \nabla C(w^*)$ is a first-order stationary point of C , it is a global minimizer of C and since $\min_W C(W) \leq \min_W C(W - \eta \nabla C(W))$, the point w^* is the global minimizer of the MAML objective. ■

Theorem 1 has implications far beyond the study of LQR. For example, Theorem 4.3 in [13] states that, under certain conditions, the objective of \mathcal{H}_2 linear control with \mathcal{H}_∞ robustness guarantee is a global function. Applying Theorem 1 to this objective yields that MAML on the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ state-feedback control design will also have no spurious local minima under the corresponding conditions. It is also applicable in case f is not a reinforcement learning objective but an objective of a regression or a classification task. It also has a straightforward extension to the multi-shot MAML and other variants of the MAML algorithm. Theorem 1 can also be viewed as a practical guideline for proving global convergence post-factum. If the function f is global and one runs Algorithm 1 with a parameter η that converges to a point w^* such that $\nabla h(w^*) = 0$, then one can check whether $\nabla^2 h(w^*) \succ 0$ and $\eta < \frac{1}{\|\nabla^2 g(w^*)\|_2}$ and if so, w^* is guaranteed to be a global minimizer of h . In case f is not global but its landscape has benign properties, then the following generalization takes place:

Proposition 1: If f is ε -global for some $\varepsilon > 0$ and $w^* \in \mathcal{W}$ is a local minimum and $\eta < \frac{1}{\|\nabla^2 g(w^*)\|_2}$, then

$$h(\bar{w}) - \min_{w \in \mathcal{W}} h(w) \leq \varepsilon$$

Proof: The mapping $\mathcal{F}(w) = w - \eta \nabla g(w)$ is continuously differentiable over \mathcal{W} . The Jacobian of this mapping is $\nabla \mathcal{F}(w) = \mathcal{I} - \eta \nabla^2 g(w)$. By assumption, $\eta < \frac{1}{\|\nabla^2 g(w^*)\|_2}$ and consequently there exists δ such that $\eta < \frac{1}{\|\nabla^2 g(w)\|_2}$ for all $w \in \mathcal{B}_\delta(w^*)$. Similar to the proof of Theorem 1, $\nabla \mathcal{F}(w)$ is positive definite for all $w \in \mathcal{B}_\delta(w^*)$, and therefore by Theorem 9.25 in [14], $\mathcal{F}|_{\mathcal{B}_\delta(w^*)}$ is an open mapping and hence locally open at w^* . By Lemma 1, it means that $h(\bar{w}) - \min_{w \in \mathcal{W}} h(w) \leq \varepsilon$. ■

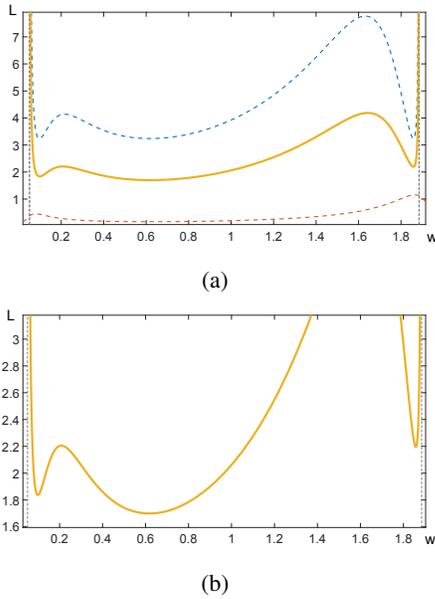


Fig. 2: Two MAML objective functions (2) for identical LQR tasks (dashed lines) and the MAML objective for the uniform distribution among them (solid line). 2b demonstrates spurious local minima of the solid line.

So far, we have shown that the benign landscape properties of MAML applied to a singleton \mathcal{T} are inherited from the benign landscape of the objective of the task to which MAML has been applied. In particular, this holds true for the LQR tasks.

B. Several identical tasks

Results obtained for the single-task scenario give rise to a hypothesis that the benign landscape of every individual task would help with the convergence of MAML in a multi-task setting as well, provided that all of the tasks have a similar structure. Thus, in this part we study multiple-task learning for which the MAML has originally been designed for. Starting with some tasks that are the most similar to each other, we consider LQR tasks that coincide up to multiplication of the cost by a positive scalar. This is the highest degree of similarity one may hope to have between sequential decision-making problems. More precisely, the landscape features of the cost function (local and global minimizers, maximizers and saddle points) are preserved under this transformation, and therefore we refer to tasks of these types as identical henceforth.

We consider a finite set of LQR tasks with a uniform distribution among them, which allows us to reduce the MAML objective (2) to the form

$$\frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} C_{\tau}(W - \eta \nabla C_{\tau}(W))$$

Figure 2 demonstrates an example of the MAML objective applied to two identical LQR tasks. Although the MAML objective of each individual task is global, only one global minimizer coincides among all of them. This is the minimizer

that corresponds to W^* . The two minimizers on the sides are shifted and after interference they produce spurious local minima for the total objective function. Hence, we can conclude that MAML fails to capture this type of similarity between the tasks. One practical lesson to learn from this figure is that keeping the cost (or the reward) function of the considered tasks normalized may improve the quality of the solution provided by MAML. Another lesson is that the design of meta-learning algorithms may benefit from considering the analysis of identical tasks as the simplest form of common structure in the tasks.

As an alternative, we propose a modification of MAML with normalized adaptation step. It turns out that this modification manages to capture the similarity between tasks with scaled rewards. The objective function for this version of MAML under single-task scenario is h' :

$$h'(w) = f\left(w - \eta \frac{\nabla g(w)}{\|\nabla g(w)\|}\right)$$

For which a statement similar to Proposition 1 holds.

Proposition 2: If f is ε -global for some $\varepsilon > 0$ and $w^* \in \mathcal{W}$ is a local minimizer of $h(w)$ with $\eta < \frac{\|\nabla g(w^*)\|}{\|\nabla^2 g(w^*)\|}$, then

$$h'(\bar{w}) - \min_{w \in \mathcal{W}} h'(w) \leq \varepsilon$$

Proof: The mapping $\mathcal{F}(w) = w - \eta \frac{\nabla g(w)}{\|\nabla g(w)\|}$ is continuously differentiable over \mathcal{W} . The Jacobian of this mapping is

$$\nabla \mathcal{F}(w) = \mathcal{I} - \eta \left[\frac{\nabla^2 g}{\|\nabla g\|^3} (\|\nabla g\|^2 \mathcal{I} - \nabla g \nabla g^T) \right]$$

Observe that

$$\left\| \frac{\nabla^2 g}{\|\nabla g\|^3} (\|\nabla g\|^2 \mathcal{I} - \nabla g \nabla g^T) \right\| \leq \frac{\|\nabla^2 g\|}{\|\nabla g\|^3} \|\|\nabla g\|^2 \mathcal{I} - \nabla g \nabla g^T\|$$

Since $\nabla g \nabla g^T$ is a rank-one matrix, the largest eigenvalue of $\|\nabla g\|^2 \mathcal{I} - \nabla g \nabla g^T$ must be equal to $\|\nabla g\|^2$ and therefore

$$\|\|\nabla g\|^2 \mathcal{I} - \nabla g \nabla g^T\| = \|\nabla g\|^2$$

Hence, $\eta < \frac{\|\nabla g(w^*)\|}{\|\nabla^2 g(w^*)\|}$ is a sufficient to conclude that there exists δ such that $\nabla \mathcal{F}(w)$ is positive definite for all $w \in \mathcal{B}_{\delta}(w^*)$, and therefore by Theorem 9.25 in [14], $\mathcal{F}|_{\mathcal{B}_{\delta}(w^*)}$ is an open mapping and thus locally open at w^* . By Lemma 1, it follows that $h'(w^*) - \min_{w \in \mathcal{W}} h'(w) \leq \varepsilon$. ■

In general, the objective of this modification of MAML can be written as

$$\mathbb{E}_{\tau} f\left(w - \eta \frac{\nabla g_{\tau}(w)}{\|\nabla g_{\tau}(w)\|}\right)$$

and its first-order stationary point can be found by utilizing the gradient descent with Armijo rule as noticed by [15] in Proposition 1.2.1. The claim that MAML with normalized gradient step inherits the benign landscape from the identical tasks can be formulated for LQR tasks as follows:

Theorem 3: Let the normalized-gradient version of MAML be applied to k LQR tasks $\{(A_i, B_i, Q_i, R_i)\}_{i=1}^k$

with scaled dynamics and rewards, meaning that there exist $\alpha_1, \dots, \alpha_k > 0$ and $\beta_1, \dots, \beta_k > 0$ such that

$$\begin{aligned} A_1 &= \dots = A_k; \\ B_1 &= \dots = B_k; \\ \alpha_1 Q_1 &= \dots = \alpha_k Q_k; \\ \alpha_1 R_1 &= \dots = \alpha_k R_k, \end{aligned}$$

and $h(W) = \sum_{i=1}^k \omega_i C_i \left(W - \eta \frac{\nabla C_i(W)}{\|\nabla C_i(W)\|} \right)$. Let $w^* \in \mathbb{R}^{r \times d}$ be a local minimal point of $h(W)$ with $\eta < \frac{\|\nabla C_i(w^*)\|}{\|\nabla^2 C_i(w^*)\|}$ for some $i \in \{1 \dots k\}$. Then, w^* is the global minimizer of h .

Proof: From the construction of the function C_i , it follows that for all $W \in \mathbb{R}^{r \times d}$ there exists $C(W)$ such that

$$C(W) = \frac{C_1(W)}{\alpha_1} = \dots = \frac{C_k(W)}{\alpha_k}$$

Consequently, $\nabla C(W) = \frac{\nabla C_1(W)}{\alpha_1} = \dots = \frac{\nabla C_k(W)}{\alpha_k}$ and $\nabla^2 C(W) = \frac{\nabla^2 C_1(W)}{\alpha_1} = \dots = \frac{\nabla^2 C_k(W)}{\alpha_k}$ for all $i \in \{1 \dots k\}$. Hence,

$$h(W) = \left[\sum_{i=1}^k \omega_i \alpha_i \right] C \left(W - \eta \frac{\nabla C(W)}{\|\nabla C(W)\|} \right)$$

Since w^* is a local minimum, by Proposition 2, we conclude that w^* is also a global minimum. ■

C. Several similar tasks

Moving forward, it is desirable to consider a different type of similarity between different tasks in the MAML setting. Therefore, we study the landscape of MAML on those tasks that are similar to each other in terms of the norm of the difference between parameters. For LQR, the parameters are the matrices A, B, Q and R . Our result states that the benign landscape of the underlying tasks carries over to the MAML objective if the tasks are sufficiently close to each other. For LQR, it will be informally stated below.

Proposition 3: For every $\varepsilon > 0$ and $k \in \mathbb{N}$, there exists $\delta > 0$ such that the MAML objective (2) is ε -global for almost any set \mathcal{T} of k LQR tasks defined through the parameters $\{(A_i, B_i, Q_i, R_i)\}_{i=1}^k$ such that for all $i, j \in \{1 \dots k\}$

$$\|A_i - A_j\| + \|B_i - B_j\| + \|Q_i - Q_j\| + \|R_i - R_j\| \leq \delta$$

The formal statement along with the proof of the result and additional discussions are provided in the Appendix. Intuitively, as long as the dependence of the cost of a single task on the parameters is continuous, for a set of tasks that are close to each other, the multi-task landscape of MAML remains close to the landscape of MAML for just one of them. However, for this reasoning to hold true, we need to determine the continuity properties of a manifold of parametric ε -global functions. In order to do that, we make an assumption below.

Assumption 1: Let $\ell : \mathcal{X} \times \mathbb{R}^m \rightarrow \mathbb{R}$ with a compact set $\mathcal{X} \subset \mathbb{R}^n$ be a twice continuously differentiable function with

respect to $x \in \mathcal{X}$ with $\ell, \nabla_x \ell$ and $\nabla_{xx}^2 \ell$ being continuous with respect to $t \in \mathbb{R}^m$. Assume that $\ell(\cdot, t)$ has a finite number of first-order stationary points for all $t \in \mathbb{R}^m$ and that the Hessian is non-singular ($\det[\nabla_{xx}^2 \ell(x, t)] \neq 0$) for all $t \in \mathbb{R}^m$ and all $x \in \mathcal{X}$ such that $\nabla_t \ell(x, t) = 0$.

The following Theorem determines the continuity property of ε -globality over the manifold of parametric functions satisfying Assumption 1.

Theorem 4: If for some \bar{t} and $\varepsilon > 0$ the function $\ell(\cdot, \bar{t})$ satisfying Assumption 1 is ε -global, then for any $\varepsilon' > 0$ such that $\varepsilon' > \varepsilon$ there exists $\delta > 0$ for which the function $\ell(\cdot, t)$ is ε' -global for all $t \in \mathcal{B}_\delta(\bar{t})$.

Proof: We prove the theorem in three steps. Step 1: $\ell(\cdot, t)$ has stationary points in the neighborhoods of the stationary points of $\ell(\cdot, \bar{t})$. Step 2: the types of the stationary points coincide. Step 3: There are no stationary points outside of the considered neighborhoods.

A finite number of stationary points means that all of them are isolated. Consider a stationary point \bar{x} of $\ell(\cdot, \bar{t})$. By Theorem 9.28 of [14] (Generalized implicit function theorem), there exist $\psi > 0$ and $\phi > 0$ such that for all $t \in \mathcal{B}_\psi(\bar{t})$ there exists a unique $x(t) \in \mathcal{B}_\phi(\bar{x})$ with the property that $\nabla_x \ell(x(t), t) = 0$. This implies that for every function $\ell(\cdot, t)$ with $t \in \mathcal{B}_\psi(\bar{t})$ there is a unique stationary point over $\mathcal{B}_\phi(\bar{x})$.

Note that for any $v \in \mathbb{R}^n$ the function $h(x, t, v) = v^\top [\nabla_{xx}^2 \ell(x, t)] v$ is continuous in both x and t . By the assumption of the theorem, $\det[\nabla_{xx}^2 \ell(\bar{x}, \bar{t})] \neq 0$, and therefore $\nabla_{xx}^2 \ell(\bar{x}, \bar{t}) \succ 0$ if \bar{x} is a local minimum, $\nabla_{xx}^2 \ell(\bar{x}, \bar{t}) \prec 0$ if it is a local maximum, and $\nabla_{xx}^2 \ell(\bar{x}, \bar{t})$ indefinite if it is a saddle point. In each of these cases, we describe how to find values δ' and ϕ' such that ℓ has a bounded value of local minima over $\mathcal{B}_{\phi'}(\bar{x}) \times \mathcal{B}_{\delta'}(\bar{t})$.

Case 1: \bar{x} is a local minimum. The value of $h(\bar{x}, \bar{t}, v)$ is positive for all $v \in \mathbb{R}^n \setminus \{0\}$. By continuity, there exist $\psi' > 0$ and $\phi' > 0$ such that $\psi' < \psi$ and $\phi' < \phi$ and $h(x, t, v) > 0$ for all $x \in \mathcal{B}_{\phi'}(\bar{x})$, $t \in \mathcal{B}_{\psi'}(\bar{t})$ and $v \in \mathbb{R}^n \setminus \{0\}$. This way, $\nabla_{xx}^2 \ell(x(t), t) \succ 0$ and therefore $x(t)$ is a local minimum of $\ell(\cdot, t)$. By continuity of $\ell(x, t)$ with respect to x and t , there exists $\delta' > 0$ such that $\delta' < \psi'$ and $|\ell(x(t), t) - \ell(\bar{x}, \bar{t})| < \frac{\varepsilon' - \varepsilon}{2}$ for all $t \in \mathcal{B}_{\delta'}(\bar{t})$.

Case 2: \bar{x} is a local maximum. The value of $h(\bar{x}, \bar{t}, v)$ is negative for all $v \in \mathbb{R}^n \setminus \{0\}$. By continuity, there exist $\psi' > 0$ and $\phi' > 0$ such that $\psi' < \psi$ and $\phi' < \phi$ and $h(x, t, v) < 0$ for all $x \in \mathcal{B}_{\phi'}(\bar{x})$; $t \in \mathcal{B}_{\psi'}(\bar{t})$ and $v \in \mathbb{R}^n \setminus \{0\}$. This way, $\nabla_{xx}^2 \ell(x(t), t) \prec 0$ and therefore $x(t)$ is a local maximum of $\ell(\cdot, t)$. In this case, we assign to the point \bar{x} the value $\delta' = \psi'$.

Case 3: \bar{x} is a saddle point. There exist $v \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$ such that $h(\bar{x}, \bar{t}, v) > 0$ and $h(\bar{x}, \bar{t}, u) < 0$. By continuity, there exist $\psi' > 0$ and $\phi' > 0$ such that $\psi' < \psi$ and $\phi' < \phi$ yet $h(x, t, v) > 0$ and $h(x, t, u) < 0$ for all $x \in \mathcal{B}_{\phi'}(\bar{x})$, $t \in \mathcal{B}_{\psi'}(\bar{t})$ and $v \in \mathbb{R}^n \setminus \{0\}$. This way, $\nabla_{xx}^2 \ell(x(t), t)$ is indefinite and therefore $x(t)$ is a saddle point of $\ell(\cdot, t)$. In this case, we assign to the point \bar{x} the value $\delta' = \psi'$.

As a result, having selected a single stationary point \bar{x} of $\ell(\cdot, \bar{t})$, we can find δ' and ϕ' such that all the stationary points

of $\ell(\cdot, t)$ for $t \in \mathcal{B}_{\delta'}(\bar{t})$ are of the same type as \bar{x} , and in case they are local minimizers, they have a value that is not too different from $\ell(\bar{x}, \bar{t})$. One can repeat this argument for all the stationary points and therefore form sets of numbers $\{\delta'_i\}_{i=1}^N$ and $\{\phi'_i\}_{i=1}^N$, where i corresponds to the index of each of the N stationary points \bar{x}_i of $\ell(\cdot, \bar{t})$.

Consider the set $\mathcal{Y} = \mathcal{X} \setminus \cup_{i \in \{1 \dots N\}} \mathcal{B}_{\phi'_i}(\bar{x}_i)$. It is a compact set as a compact set minus an open set and $\|\nabla_x \ell(x, \bar{t})\| > 0$ for all $x \in \mathcal{Y}$. Since $\|\nabla_x \ell(x, \bar{t})\|$ is continuous in x over a compact set, it is uniformly continuous and it reaches its lower bound, meaning that there exists $\xi > 0$ such that $\|\nabla_x \ell(x, \bar{t})\| \geq \xi$ for all $x \in \mathcal{Y}$. By continuity of $\|\nabla_x \ell(x, t)\|$ with respect to t , there exists δ'' such that $\|\nabla_x \ell(x, t)\| > \frac{\xi}{2}$ over $\mathcal{Y} \times \mathcal{B}_{\delta''}(\bar{t})$ and therefore there are no stationary points of ℓ over $\mathcal{Y} \times \mathcal{B}_{\delta''}(\bar{t})$.

We select $\delta = \min[\{\delta'_i | i \in \{1 \dots N\}\} \cup \{\delta''\}]$ and observe that for all $t \in \mathcal{B}_{\delta}(\bar{t})$ the only local minimizers of $\ell(\cdot, t)$ are those close to the local minimizers of $\ell(\cdot, \bar{t})$. In *Case I*, the corresponding δ' was selected such that, given a local minimizer $x(t)$ of $\ell(\cdot, t)$ that is neighboring a local minimizer \bar{x} of $\ell(\cdot, \bar{t})$ and a global minimizer $x'(t)$ of $\ell(\cdot, t)$ that is neighboring a local minimizer \bar{x}' of $\ell(\cdot, \bar{t})$, it holds that

$$\begin{aligned} \ell(x(t), t) - \min_x \ell(x, t) &= \ell(x(t), t) - \ell(x'(t), t) = \\ &= \ell(x(t), t) - \ell(\bar{x}, \bar{t}) + \ell(\bar{x}, \bar{t}) - \ell(\bar{x}', \bar{t}) + \\ &\quad + \ell(\bar{x}', \bar{t}) - \ell(x'(t), t) \leq \\ &= |\ell(x(t), t) - \ell(\bar{x}, \bar{t})| + |\ell(\bar{x}, \bar{t}) - \ell(\bar{x}', \bar{t})| + \\ &\quad + |\ell(\bar{x}', \bar{t}) - \ell(x'(t), t)| < \\ &= \frac{\varepsilon' - \varepsilon}{2} + \varepsilon + \frac{\varepsilon' - \varepsilon}{2} = \varepsilon' \end{aligned}$$

Proposition 3 goes beyond the uniform distribution and holds for any distribution on a finite number of tasks. An example of the MAML objective for five similar LQR tasks is demonstrated in Figure 3a. Similarly to the results for single-task scenario, the statements of this section are generalizable to other tasks with benign optimization landscape. In the end, we conclude that MAML is able to capture the common structure among different tasks given through similar values of the parameters.

D. A number of tasks with common dynamics

Figure 3 demonstrates the MAML objective for several LQR tasks that share the same dynamics but have different cost functions. Increasing the number of considered tasks improves the features of the landscape of the total MAML objective. In the eleven-task scenario, MAML learns the mean of the optimal policies W^* for the tasks, which implies that instead of learning to rapidly adapt, MAML learns the average policy by effectively finding the minimum of $\frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} C_{\tau}(W)$. However, for the two-task scenario, the global minimizer of the MAML objective appears to correspond to a policy W that has $\nabla C_{\tau}(W)$ far from zero for every considered τ , and thus the algorithm that converged to that point would learn to adapt to a task during the meta-testing phase.

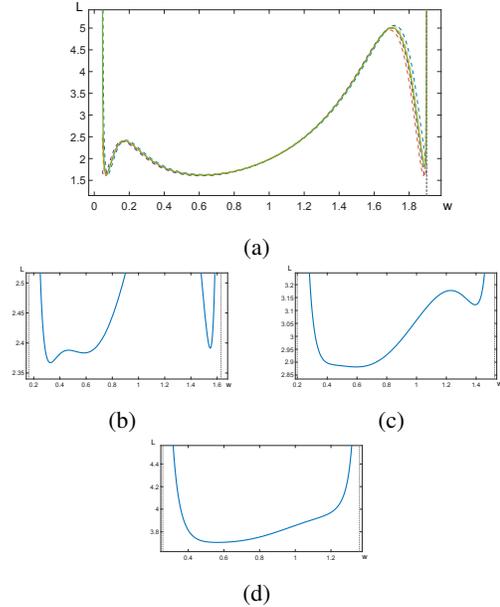


Fig. 3: Five MAML objective functions (2) for LQR tasks with similar values of A , B , Q and R (dashed lines) and the MAML objective for the uniform distribution among them (solid line) (**plot 3a**). MAML objective (2) for the uniform distribution among two (**plot 3b**), five (**plot 3c**) and eleven (**plot 3d**) different LQR tasks that share the same dynamics A and B but have different cost matrices Q and R .

We observe that the weighted sum of a large number of functions that have minimizers in a close proximity ends up being almost global with the minimizer in the same region. Therefore, in practice, increasing the number of similar tasks on meta-training stage may improve the properties of the landscape of the MAML objective and assure convergence to the global minimum.

III. CONCLUSION

In the paper, we studied the objective function (2) of the MAML algorithm constructed for different spaces of tasks with some common structure. We showed that the MAML objective inherits the benign optimization landscape from the underlying tasks if they are similar pointwise and that this desirable property fails to hold for the tasks that coincide up to a multiplication of the objective by a positive value. We proposed a modification of MAML that does not possess this drawback. During this study of the MAML we used LQR as the primary example, although a number of general results on benign landscapes were discovered, including continuity property for the newly introduced notion of ε -global function and its properties with respect to convex combinations that are described in the Appendix.

APPENDIX

Lemma 2: Given $\lambda_i > 0$ and $\sum_{i=1}^k \lambda_i = 1$, if for some \bar{t} and $\varepsilon > 0$ the function $\ell(\cdot, \bar{t})$ satisfying Assumption 1 is ε -global, then for any $\varepsilon' > 0$ such that $\varepsilon' > \varepsilon$ there exists $\delta > 0$

for which the convex combination $\lambda_1 \ell(\cdot, t_1) + \dots + \lambda_k \ell(\cdot, t_k)$ is ε' -global for all $t_1, \dots, t_k \in \mathcal{B}_\delta(\bar{t})$.

Proof: We provide the proof for $k = 2$, but the argument holds true for other finite values of k . By Theorem 4, $\ell(\cdot, t_1)$ can be assumed to be ε'' -global for $\varepsilon' > \varepsilon'' > \varepsilon$. Therefore, without loss of generality, we assume that $t_1 = \bar{t}$, and then aim to prove that for a given $\lambda \in [0, 1]$ there exists $\delta > 0$ such that $\lambda \ell(x, \bar{t}) + (1 - \lambda) \ell(x, t)$ is ε' -global. We introduce $\tau(x, t) = \lambda \ell(x, \bar{t}) + (1 - \lambda) \ell(x, t)$ and note that $\tau(x, \bar{t}) = \ell(x, \bar{t})$, which means that $\tau(\cdot, \bar{t})$ is ε -global and satisfies Assumption 1. Theorem 4 applied to τ yields that there exists $\delta > 0$ such that $\lambda \ell(x, \bar{t}) + (1 - \lambda) \ell(x, t)$ is ε' -global. ■

Proposition 4: Given $\lambda_i > 0$ and $\sum_{i=1}^k \lambda_i = 1$, if for all the values of t in a compact set $\mathcal{C} \subset \mathbb{R}^m$ and some $\varepsilon > 0$ the function $\ell(\cdot, t)$ satisfying Assumption 1 (restricted from \mathbb{R}^m to \mathcal{C}) is ε -global, then for any $\varepsilon' > 0$ such that $\varepsilon' > \varepsilon$ there exists $\delta > 0$ for which any convex combination $\lambda_1 \ell(\cdot, t_1) + \dots + \lambda_m \ell(\cdot, t_m)$ is ε' -global for all t_1, \dots, t_m such that $\|t_i - t_j\| < \delta$.

Proof: To prove by contradiction, suppose that there exists $\varepsilon' > 0$ such that $\varepsilon' > \varepsilon$ and for all $\delta > 0$ there are $t_1(\delta), \dots, t_k(\delta)$ such that $\|t_i - t_j\| < \delta$ and $\lambda_1 \ell(\cdot, t_1) + \dots + \lambda_m \ell(\cdot, t_m)$ is not ε' -global. From the sequence $t_1(\frac{1}{l})$, one can extract a converging sub-sequence $t_1(\delta_l)$ since \mathcal{C} is compact. By Lemma 2, for the point $\bar{t} = \lim_{l \rightarrow \infty} t_1(\delta_l)$ there exists $\delta' > 0$ such that for all $t_1, \dots, t_k \in \mathcal{B}_{\delta'}(\bar{t})$ the convex combination $\lambda_1 \ell(\cdot, t_1) + \dots + \lambda_k \ell(\cdot, t_k)$ is ε' -global. Therefore, for the $t_1(\delta_l), \dots, t_k(\delta_l)$ such that $\delta_l < \frac{\delta'}{2k}$ and $\|t_1(\delta_l) - \bar{t}\| < \frac{\delta'}{2k}$, this convex combination is ε' -global, which is a contradiction. ■

Proposition 5 (Formal statement of Proposition 3):

Consider the instances of LQR such that their parameters A, B, Q and R belong to a compact set. Assume that none of them produces a cost function $C(W)$ with a singular Hessian at a stationary point. For any $\varepsilon > 0$ and $k \in \mathbb{N}$, there exists $\delta > 0$ such that for any set of k considered instances of LQR defined through the parameters $\{(A_i, B_i, Q_i, R_i)\}_{i=1}^k$ with

$$\|A_i - A_j\| + \|B_i - B_j\| + \|Q_i - Q_j\| + \|R_i - R_j\| \leq \delta,$$

for all $i, j \in \{1 \dots m\}$, the MAML objective (2) is ε -global.

Proof: Take $t = (A, B, Q, R)$. By Theorem 1, every function $C(W, t)$ satisfies Assumption 1. Therefore, $C(W, t)$ satisfies all of the assumptions of Proposition 4 and the proof follows immediately. ■

Given a matrix \bar{W} , the system of equations

$$\nabla C(\bar{W}) = 0; \quad \det(\nabla^2 C(\bar{W})) = 0 \quad (3)$$

with respect to the parameters A, B, Q and R defines a low-dimensional manifold in the space of LQR tasks denoted by $HS(\bar{W})$. All the LQR systems that have at least one stationary point with a singular Hessian are contained in $\cup_W HS(W)$. Now, notice that the union occurs over an $n \times n$ dimensional space, while $HS(W)$ is defined with a system of $(n \times n) + 1$ equations. This implies that $\cup_W HS(W)$ is

a low-dimensional manifold and thus almost no LQR has a stationary point with a singular Hessian. Thus, the compact domain that is mentioned in the assumption of Proposition 5 can be a large closed set with $\cup_W HS(W)$ excluded together with its small neighborhood. Thus, Proposition 3 is an informal restatement of Proposition 5.

Details on the experiments

The one-dimensional initial state s_0 is deterministic.

Figure	LQR $(A, B, Q, R, s_0 s_0^\top)$
Figure 1	$(1, 1, 2, 2, 1), \eta = 0.1$
Figure 2	$(1, 1, 2, 2, 1), (1, 1, 0.1, 0.1, 1), \eta = 0.1$
Figure 3a	$(1.01, 1, 1, 1, 1), (1, 1.01, 1, 1, 1), (1, 1, 1.01, 1, 1), \eta = 0.1$
Figure 3b	$(1, 1, 1, 2, 1), (1, 1, 2, 1, 1), \eta = 0.1$
Figure 3c	$(1, 1, 1, 1, 1), (1, 1, 1, 2, 1), (1, 1, 2, 1, 1), (1, 1, 2, 3, 1), (1, 1, 3, 2, 1), \eta = 0.1$
Figure 3d	$(1, 1, 1, 1, 1), (1, 1, 1, 2, 1), (1, 1, 2, 1, 1), (1, 1, 2, 3, 1), (1, 1, 3, 2, 1), (1, 1, 3, 1, 1), (1, 1, 1, 3, 1), (1, 1, 4, 1, 1), (1, 1, 1, 4, 1), (1, 1, 5, 3, 1), (1, 1, 3, 5, 1), \eta = 0.1$

REFERENCES

- [1] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR.org, 2017, pp. 1126–1135.
- [2] E. Todorov and W. Li, "A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *American Control Conference*. IEEE, 2005, pp. 300–306.
- [3] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2017.
- [4] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for the linear quadratic regulator," in *Proceedings of the 35th International Conference on Machine Learning*. PMLR, 2018.
- [5] A. Fallah, A. Mokhtari, and A. Ozdaglar, "On the convergence theory of gradient-based model-agnostic meta-learning algorithms," *arXiv preprint arXiv:1908.10400*, 2019.
- [6] A. Rajeswaran, C. Finn, S. M. Kakade, and S. Levine, "Meta-learning with implicit gradients," in *Advances in Neural Information Processing Systems*, 2019, pp. 113–124.
- [7] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv preprint arXiv:1803.02999*, 2018.
- [8] C. Finn, A. Rajeswaran, S. Kakade, and S. Levine, "Online meta-learning," *arXiv preprint arXiv:1902.08438*, 2019.
- [9] P. Zhou, X. Yuan, H. Xu, S. Yan, and J. Feng, "Efficient meta learning via minibatch proximal update," in *Advances in Neural Information Processing Systems*, 2019, pp. 1534–1544.
- [10] L. Wang, Q. Cai, Z. Yang, and Z. Wang, "On the global optimality of model-agnostic meta-learning," *arXiv preprint arXiv:2006.13182*, 2020.
- [11] C. Jozs, Y. Ouyang, R. Zhang, J. Lavaei, and S. Sojoudi, "A theory on the absence of spurious solutions for nonconvex and nonsmooth optimization," in *Advances in neural information processing systems*, 2018, pp. 2441–2449.
- [12] I. Molybog and J. Lavaei, "When does maml objective have benign landscape?" *arXiv preprint arXiv:2006.00453*, 2020.
- [13] K. Zhang, B. Hu, and T. Basar, "Policy optimization for h_2 linear control with h_∞ robustness guarantee: Implicit regularization and global convergence," *arXiv preprint arXiv:1910.09496*, 2019.
- [14] W. Rudin, *Principles of mathematical analysis (third edition)*. McGraw-Hill Inc., 1976.
- [15] D. P. Bertsekas, W. Hager, and O. Mangasarian, *Nonlinear programming*. Athena Scientific Belmont, MA, 1998.