Control-Oriented Identification for the Linear Quadratic Regulator

Sean Anderson, João P. Hespanha

Abstract—Data-driven control benefits from rich datasets, but constructing such datasets becomes challenging when gathering data is limited. We consider an offline experiment design approach to gathering data where we design a control input to collect data that will most improve the performance of the control task. We show how such a control-oriented approach can be used in a setting with linear dynamics and quadratic objective and, through design of a gradient estimator, solve the problem via stochastic gradient descent. We contrast our method with a classical A-optimal experiment design approach and numerically demonstrate that our method outperforms A-optimal design in terms of improving control performance.

I. INTRODUCTION

Model-based control methods benefit from useful models of the controlled system. Consider a setting in which there is uncertainty in the model parameters and there is an opportunity to collect experimental data to learn more about the system. The data collection involves selecting a control input, and this paper focuses on the optimal selection of such an input with the goal of eventually generating a “high-performance” controller for the uncertain process. In this context, “high-performance” is defined in terms of a pre-specified criterion that we use to evaluate the performance of a given controller. This motivates the following control-oriented experiment design problem: select a control input for a data-collection experiment so that the feedback controller designed using the data acquired will lead to the best possible performance. Our approach is general in terms of the control design procedure used to generate the controller from the experimental data collected. However, in this paper we focus our attention on controllers generated through certainty equivalence, which in this context means constructing an a-posteriori estimate for the process and designing a controller for this estimate.

This paper includes two key contributions: First, we show how such a control-oriented approach to experiment design can be carried out for the control of a linear system with uncertain matrix dynamics and a quadratic objective function. While this problem does not have a closed-form solution, we show that it can be efficiently solved by stochastic gradient descent. The second contribution lies in the observation that our formulation of a control-oriented approach to experiment design generally leads to closed-loop controllers that exhibit higher performance than what would be achieved by classical forms of experiment design, which are typically aimed at minimizing a-posteriori estimation error (or other generic forms of post-experiment uncertainty), rather than focusing directly on improving closed-loop performance.

In this work, we focus on the offline experiment design setting for data-driven control. In Section III we present a general formulation for experiment design that takes into account a-priori parameter uncertainty in discrete-time dynamics. We consider an experiment that generates a dataset consisting of state and input measurements, where the dataset is used in the construction of a controller. The resulting optimal experiment design problem aims to improve the expected performance of the controller with respect to possible parameter values as well process noise during the experiment. This formulation is contrasted with classical experiment design formulations that aim to minimize a measure of the parameter uncertainty [1].

We then turn to the linear quadratic regulator (LQR) setting in Section III where we define the linear dynamics and quadratic control objective. We address the system identification step and how to handle exploding trajectories during (simulated) experiments. For the data-driven controller, we use an LQR controller with certainty equivalence in the parameters [2]. The solutions available in LQR are amenable to fast computations, which is conducive to the numerical approach we take later on.

We go on to present the solution method to the experiment design approach in Section V. We use a first-order approach by designing a pathwise gradient estimator for the purposes of stochastic gradient descent. By leveraging the structure of the problem, we find a gradient estimate that proves to be effective numerically as illustrated in Section V. In order to validate our method, we compare our approach against a classical experiment design that aims to reduce a measure of the parameter uncertainty [3].

We then turn to the linear quadratic regulator (LQR) setting in Section III where we define the linear dynamics and quadratic control objective. We address the system identification step and how to handle exploding trajectories during (simulated) experiments. For the data-driven controller, we use an LQR controller with certainty equivalence in the parameters [2]. The solutions available in LQR are amenable to fast computations, which is conducive to the numerical approach we take later on.

Related work: The issue of how to gather data through well-planned experiments has traditionally been addressed through the framework of optimal experiment design. Modern optimal experiment design is often attributed to Gustav Elfving, who designed experiments to minimize measures of parameter error covariance [3]. Later on, researchers worked on aligning experiment design with particular criteria, including control objectives [4]. See [5] for a thorough review.
of early work in this area. Recent relevant work in this area includes [6], which proposes a stochastic gradient descent approach to designing experiments that minimizes a post-experiment optimal control objective. Work in experiment design for control in the statistical learning community includes [7], [8], [9], all of which emphasize theoretical aspects of learning linear systems. As an alternative paradigm, online learning or adaptive control allow for improvement during the experiment trial. Adaptive control has been well-covered in [10], and online linear quadratic regulation has been explored from learning theory in [11].

As we focus on the linear quadratic regulator setting, the most relevant papers are from [7], in which the authors derive fundamental limits for learning the linear quadratic regulator via offline experiments, and [8], in which the authors propose an experiment design method for learning linear systems for particular tasks. The former focuses on the statistical learning guarantees associated with the LQR problem but do not broach how to construct an optimal experiment. In the latter, the authors propose a weighted-trace optimal experiment design that bears similarity to $L$-optimal experiment design for control-oriented identification as noted in [5]; the authors consider convergence of sequential experiment designs to the underlying system.

Gradient estimation has seen particular attention in the machine learning community [12], and two dominant methods are via the pathwise gradient and the score function (or log score) method [13]. Score function estimators benefit from only taking the gradient of the density function, but the structure of our problem is not amenable: as such, we focus on the pathwise gradient estimate.

II. EXPERIMENT DESIGN FOR DATA-DRIVEN CONTROL

We consider a discrete-time system with dynamics of the form

$$x_{t+1} = f(x_t, u_t, w_t; \theta),$$

with state $x_t \in \mathbb{R}^{n_x}$, control input $u_t \in \mathbb{R}^{n_u}$, and an unmeasured stochastic disturbance $w_t \in \mathbb{R}^{n_w}$ independent and identically distributed across time. The dynamics depend on stochastic parameters $\theta$ that are unknown, but for which we have an $a$-priori distribution.

An experiment will be performed to provide additional information about the parameter $\theta$. State and input measurements are collected throughout the experiment providing a sequence of $M$ triples $\mathcal{D} := \{ (x_i, x_i, u_i) : i = 1, ..., M \}$ that satisfy the basic model of the dynamical system:

$$x_{i+1}^+ = f(x_i, u_i, w_i; \theta), \quad \forall i \in \{1, \ldots, M\}$$

with the $w_i$ independent across indices $i$ and with the same distribution as the disturbance. If the experiment consists of a single run of (1) over a time horizon $t = 1$ through $t = T$, then the index $i$ is simply time and $M = T$. However, in general, “an experiment” may include multiple runs of (1) over different time horizons, in which case (2) include all the data collected. To simplify notation we collect all the columns vectors $x_i^+, x_i \in \mathbb{R}^{n_x \times r}$, $u_i \in \mathbb{R}^{n_u}$ into matrices

with $M$ columns that we denote by $X^+$, $X \in \mathbb{R}^{n_x \times M}, U \in \mathbb{R}^{n_u \times M}$, respectively.

Our goal is to design a controller $\pi$ that optimizes a given cost function $J(\pi; \theta)$ that depends both on the controller and on the actual (but unknown) value of the parameter $\theta$. We also take as given a control design procedure that maps the experiment design data $\mathcal{D}$ to a specific controller $\pi$, with the goal of minimizing the cost $J(\pi; \theta)$. A reasonable procedure would be to select a controller that minimizes the expected value of the cost $J(\pi; \theta)$, given the data $\mathcal{D}$ collected during the experiment, as in:

$$K(\mathcal{D}) = \arg \min_{\pi} \mathbb{E}[J(\pi; \theta) \mid \mathcal{D}]. \quad (3)$$

However, and because this optimization is often intractable, our presentation considers a general control design method $K : \mathcal{D} \mapsto \pi$, which may or may not be optimal.

The experiment design problem arises from the observation that the data $\mathcal{D}$ collected depends on the control inputs $U$ used during the experiment as well as on the actual realizations of the the random disturbances (also during the experiment) that we denote by $W \in \mathcal{R}^{n_w \times M}$. To emphasize this dependence, we use the notation $\mathcal{D}_{U,W}$ to express the dependence of the dataset on these variables. The optimal experiment design problem can then be formulated as

$$\min_{U \in \mathcal{U}} \mathbb{E} \left[ J(\pi; \theta) \right], \quad \pi := K(\mathcal{D}_{U,W}), \quad (4)$$

where the expectation refers to an integration over (i) the a-priori distribution of the parameter $\theta$, and (ii) the realization of the disturbance $W$ during the actual experiment. The minimization is performed over a set of admissible controls that we denote generically by $\mathcal{U}$. The experiment design criterion (4) should be contrasted with more classical formulations that set the goal of the experiment to minimize some measure of uncertainty about the unknown parameter $\theta$. For example, an $A$-optimal experiment design essentially tries to minimize

$$\min_{U \in \mathcal{U}} \mathbb{E} \left[ \| \hat{\theta}(\mathcal{D}_{U,W}) - \theta \|^2 \right], \quad \hat{\theta}(\mathcal{D}_{U,W}) := \mathbb{E} \left[ \theta \mid \mathcal{D}_{U,W} \right]. \quad (5)$$

The key distinction is that (5) ignores the impact of uncertainty on the control objective $J(\pi; \theta)$ and therefore does not take advantage of the fact that reducing uncertainty on some parameters may be much more important than on others, for our ultimate control objective of minimizing $J(\pi; \theta)$.

III. EXPERIMENT DESIGN FOR THE LINEAR QUADRATIC REGULATOR

We now specialize the general setup described above to the finite-horizon linear quadratic regulator setup. Specifically, we consider the process

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad (6)$$

such that $\theta$ contains the elements of $A$ and $B$, and $w_t$ is Gaussian noise identically distributed across time with zero mean and covariance $\Sigma_w$. 


We consider a quadratic optimization criterion of the form:

\[
J(\pi; \theta) := \mathbb{E}\left[ x_N^T Q_N x_N + \sum_{i=0}^{N-1} x_i^T Q x_i + u_i^T R u_i \mid \theta \right],
\]

where the expectation refers to an integration over the disturbances encountered by the controller \( \pi; Q, N, Q \) are positive semidefinite matrices; and \( R \) a positive definite matrix.

We consider a common option for control design generally known as certainty equivalence (CE): certainty equivalence design \( \text{K}_{CE}(D) \) computes the \( a \)-posteriori expected value of the unknown parameters \( \hat{\theta} := \mathbb{E}[\theta \mid D] \) and computes the linear optimal controller \( u_t = K_t x_t \) that minimizes \( J \), assuming that the estimate \( \hat{\theta} \) is correct.

1) System identification: In order to generate the \( a \)-posteriori estimate of the system \( \hat{\theta} \) for \( \text{K}_{CE}(D) \), we employ linear regression on a dataset \( D \), which in our case will be the dataset generated under the experiment decision variable \( U \). For identification, we express \( \Theta \) as:

\[
X^+ = \Theta Z + W
\]

with \( Z = [X; U] \in \mathbb{R}^{(n_x+n_u) \times M} \), and \( \Theta := [A, B] \in \mathbb{R}^{n_x \times (n_x+n_u)} \) to denote the vectorized version of \( \Theta \) via stacking its columns.

**Corollary 1:** [14] Consider a Gaussian prior on the parameters with mean \( \theta_0 \in \mathbb{R}^{n_x \times (n_x+n_u)} \) and covariance of the \((i,j)\)th element with the \((k,l)\)th element of \( \Theta \) given by

\[
\mathbb{E}[\Theta - \Theta_{ij} (\Theta - \Theta)_{kl}] = (\Sigma_w)_{kl} (\Lambda_n^{-1})_{ji},
\]

where \( \Sigma_w \) is the known noise covariance and \( \Lambda_n := \Lambda_0 + ZSZ^T \), and \( Z \in \mathbb{R}^{M \times M} \) is the weight matrix.

**Proof:** See [14].

The weight matrix \( S \) improves the numerics of the regression problem, particularly since simulating unstable systems can lead to exponential growth in the state that, due to large numbers, lead to deleterious performance in the inversion of \( \Lambda_n \). In particular, let

\[
S(x) := \text{diag}(s(x_0), \ldots, s(x_N))
\]

where \( s(x) \in [0,1] \) ensures that the weight matrix assigns zero weight to points on trajectories that are numerically too large. For this work, we choose \( s(x) := \arctan(|x|) \alpha_1/x_2 + 0.5 \), and \( \alpha_1, \alpha_2 \) are design parameters.

2) Certainty equivalent control: Given an estimate of the parameters \( \theta \) from \( \hat{\theta} \) with means \( \hat{\theta} \), respectively, we construct our controller \( \text{K}_{CE}(D) \) by recursively solving the Riccati difference equations given by

\[
\begin{align}
K_t &= -(R + \hat{B}^T \hat{P}_{t+1} \hat{B})^{-1} \hat{B}^T \hat{P}_{t+1} \hat{A}, \\
P_t &= Q + \hat{K}_t^T R \hat{K}_t - (\hat{A} + \hat{B} \hat{K}_t)^T \hat{P}_{t+1} (\hat{A} + \hat{B} \hat{K}_t),
\end{align}
\]

with \( P_N = Q_N; Q, Q_N \) are positive semidefinite matrices and \( R \) a positive definite matrix.

**Corollary 2:** [15] For a sequence of linear feedback gains, \( \pi := \{K_0, \ldots, K_{N-1}\} \) from \( \text{K}_{CE}(D) \), we can express the finite-horizon LQR cost \( J \) for the system in \( \Theta \) parameterized by \( \theta \) as

\[
J(\pi; \theta) = x_N^T P_0 x_0 + \sum_{i=0}^{N-1} \text{tr}(P_{t+1} \Sigma_w),
\]

with boundary condition \( P_N = Q_N \).

**Proof:** See, for example, [15].

**IV. EXPERIMENT DESIGN VIA GRADIENT DESCENT**

In order to solve the experiment design optimization \( \Theta \), we take a gradient-descent approach:

\[
U_{i+1} = \text{Proj}_d(U_i - \eta_i \nabla_U),
\]

where \( \text{Proj}_d(\cdot) \) projects \( U \) onto the set of admissible inputs \( U \), \( \eta_i \) is step size, and \( \nabla_U \) is an approximation of the true gradient \( \nabla_U \) given by differentiating the experiment criteria \( J \):

\[
\nabla_U \mathbb{E} [J(K(D); \theta)] = \nabla_U \int J(K(D); \theta) p(X \mid U, x_0, \theta)p(\theta)d\theta dX,
\]

where \( x_0 \) is the initial state for the experiment. The gradient of the integral is analytically intractable, motivating the use of a Monte Carlo gradient estimate. A gradient estimate generally requires an exchange of the gradient and integral

\[
\nabla_U \int F(X, U)p(X | U)dX = \int \nabla_U F(X, U)p(X | U)dX,
\]

for an arbitrary \( F(\cdot) \) and density \( p(\cdot) \), where the exchange is valid when (i) the magnitude of the gradient of the integrand is bounded by a function (ii) that is integrable with respect to the random variable. See [16] for further treatment of the exchange.

By introducing a change of variable \( g(\cdot) \) and using the Law of the Unconscious Statistician [17], we remove the need to differentiate the density such that

\[
\nabla_U \int F(X, U)p(X | U)dX = \int p(\epsilon) \nabla_U F(X, U)|_{X=g(\epsilon; U, x_0, \theta)}d\epsilon, \quad \epsilon \sim p(\epsilon),
\]
and only $F(\cdot)$ needs to be differentiated. This leads to the Monte
Carlo pathwise gradient estimator [12]:

$$
\frac{1}{L} \sum_{l=1}^{L} \nabla U F(X, U) \mid_{X=g(t^{(l)}; U)}, \quad \epsilon^{(l)} \sim p(\epsilon).
$$

(17)

For our problem in (14), we can find a suitable change of variable using the dynamics.

**Theorem 1:** The change of variable $g : \mathbb{R}^{n_x \times M} \times \mathbb{R}^{n_x \times M} \times \mathbb{R}^{n_x \times M} \rightarrow \mathbb{R}^{n_x \times M}$ for (14) given by the recursion of the dynamics (15):

$$
\begin{align*}
    x_1 &= f(x_0, u_0, w_0; \theta), \\
    x_2 &= f(f(u_0, x_0, w_0; \theta), u_1, w_1; \theta), \\
    &\vdots \\
    x_M &= f(f(x_{M-2}, u_{M-2}, w_{M-2}; \theta), u_{M-1}, w_{M-1}; \theta),
\end{align*}
$$

(18a)

(18b)

(18c)

where $x_M$ is the final state, recursively dependent on all preceding states, satisfies (16) such that $X = g(W; U, x_0, \theta)$ with $W$ distributed according to $p(W)$, the process noise.

**Proof:** Under the process (11), use the Markov property to write the density of $X$ as

$$
p(X \mid U, x_0, \theta) = \Pi_{t=0}^{T-1} p(x_{t+1} \mid x_t, u_t, \theta).
$$

(19a)

For any time $t$,

$$
p(x_{t+1} \mid x_t, u_t, \theta) = p(x_{t+1} \mid x_t, u_t, \theta) p(w_t),
$$

(19b)

where $p(x_{t+1} \mid x_t, u_t, \theta)$ is deterministic given the dynamics, $f(\cdot)$, in (11) such that sampling from $p(x_{t+1} \mid x_t, u_t, \theta)$ is equivalent to sampling from $f(x_t, u_t, w_t; \theta), w_t \sim p(w)$. Extend this recursively by expressing

$$
p(x_{t+1} \mid x_t, u_t, \theta) = \Pi_{t=0}^{T-1} p(x_{t+1} \mid x_t, u_t, \theta) p(w_t)\n$$

(19c)

such that $x_{t+1}$ can be expressed as $f(\ldots f(x_0, u_0, w_0; \theta), u_1, w_1; \theta)$ where the first argument is taken to mean that the state is recursively defined given an initial state $x_0$, input sequence, and $w_0, \ldots, w_t \sim p(w)$. Define this recursion as in (18) by $X = g(W; U, x_0, \theta)$ where $W \sim p(W)$ to get the desired result.

**Lemma 1:** The linear system (4) has a change of variable $G(W; U, x_0, \theta)$, such that the $t$th column of $X$ in (18) is given by

$$
x_t = A^t x_0 + \sum_{i=0}^{t-1} A^{t-i-1} (Bu_i + u_i),
$$

(20a)

$w_t \sim p(w)$, the process noise distribution.

**Proof:** See details in technical report [18].

**Lemma 2:** The change of variable (20a) and LQR experiment criteria (7), with $K_{CE}(D)$, yields an estimator of the form (17):

$$
\hat{X}_U = \frac{1}{L} \sum_{l=1}^{L} \nabla U J \left( K(D) \mid_{X=G(W^{(l)}; U, x_0, \theta^{(l)})}; \theta^{(l)} \right),
$$

(20b)

with $W^{(l)} \sim p(W)$.

**Proof:** See [18] for satisfying the differentiability of the value function.

---

### V. Numerical Experiments

#### A. Car String

![Image of a car string scenario]

Fig. 1. We focus on a scenario in which cars need to regulate to a specified gap $L$ at a desired reference velocity $v$. Deviations from the desired position are $\Delta w^{(n)}$ and deviations from the reference velocity are $\Delta u^{(n)}$.
We consider the problem of maintaining a fixed distance, \( L \), between \( n \) cars at a desired velocity \( v \) as depicted in Figure 1. We adapt the continuous-time dynamics for relative position as given in [19] to discrete-time dynamics with sampling time \( T_s \):

\[
\Delta v_{t+1}^{(n)} = \left( -\frac{\alpha^{(n)}T_s}{m^{(n)}} + 1 \right) \Delta v_{t}^{(n)} + \frac{T_s}{m^{(n)}} \Delta u_{t+1}^{(n)},
\]
\[
\Delta u_{t+1}^{(n)} = T_s (\Delta v_{t}^{(n)} - \Delta v_{t+1}^{(n+1)}) + \Delta w_{t}^{(n)},
\]

where \( \Delta v_{t}^{(n)} \) is the deviation from the reference velocity at car \( n \) and \( \Delta u_{t+1}^{(n)} \) is the deviation of the gap between cars \( n+1 \) and \( n \) from the desired gap \( L \). \( \Delta u \) is a change in force input for each car. This leads to an \( n \) car state-vector \( x_{t+1} := [\Delta u_{t+1}^{(1)}, \Delta u_{t+1}^{(1)}, \Delta u_{t+1}^{(2)}, \ldots, \Delta u_{t+1}^{(n)}]^T \). As such \( A \) and \( B \) are given by:

\[
A = \begin{bmatrix}
\frac{\alpha^{(1)}T_s}{m^{(1)}} + 1 & 0 & 0 & 0 & \cdots \\
0 & \frac{\alpha^{(2)}T_s}{m^{(2)}} + 1 & 0 & \cdots \\
& & & \ddots & \ddots \\
0 & 0 & \cdots & \frac{\alpha^{(n)}T_s}{m^{(n)}} + 1 & 0 \\
0 & 0 & \cdots & 0 & \frac{T_s}{m^{(n+1)}} \\
0 & 0 & \cdots & 0 & \frac{T_s}{m^{(n+2)}} \\
0 & 0 & \cdots & 0 & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 0 & 0 \\
\end{bmatrix}
\]
\[B = \begin{bmatrix}
\frac{T_s}{m^{(1)}} & 0 & \cdots \\
0 & \frac{T_s}{m^{(2)}} & \cdots \\
0 & 0 & \cdots \\
0 & 0 & \cdots \\
\vdots & \vdots & \ddots & \ddots \\
0 & 0 & \cdots & 0 \\
\end{bmatrix}\]

While there is a specific structure to \((A, B)\) here, we assume we do not know the structure and estimate all \((2n-1)(3n-1)\) entries. We specify the noise covariance in the dynamics as \( \Sigma_w = 1e^{-2} \times I_5 \). The prior on the parameters \( \Theta_0 = [A, B] \) with \( m^{(1)} = m^{(2)} = m^{(3)} = 1, \alpha^{(1)} = \alpha^{(3)} = 1, \alpha^{(2)} = 0, T_s = 0.1; \Lambda_0^{-1} = 1e^-2 \times I_8 \). Horizon \( N = 30 \).

**B. Experiment Design Setup**

In the results that follow, we use an experiment horizon of \( M = 20 \) time steps, and batch size \( L = 1000 \) in Algorithm 1. As in [19], for the criteria in \( Q \) includes penalties of magnitude 10 on the positions \( \Delta w \) and zero on the velocity \( \Delta v \). \( R \) has parameters \( \alpha_1 = 10^3, \alpha_2 = 10^6 \). \( U \) is initialized with \( u_t \sim U[10^{-3}, 10^{-2}] \) and is fixed across experiments.

To solve the A-optimal experiment design in (5), we use a pathwise gradient estimator of the same form as (17) to perform stochastic gradient descent, where the objective is to minimize the trace of the posterior error covariance \( \Theta \).

**C. Results and Discussion**

We start by comparing the performance of our method against A-optimal design (5) in terms of post-experiment LQR control performance (14). For the experiment design (4), we consider a feasible input set \( U \) where the Frobenius norm of the input sequence, \( U \), is bounded by a design parameter \( \beta \) such that \( U = \{ U | \|U\|_F \leq \beta \} \). We vary the allowed magnitude, \( \beta \), in Figure 2 and observe our method outperforms the A-optimal design uniformly. For any experiment design, we have a lower and upper bound on the performance. If we knew the values of \((A, B)\), we would achieve the lowest possible control cost such that this is a lower bound on achievable performance. The upper bound comes from the expected control performance associated with using a controller that uses the a-priori system estimate instead of the a-posteriori as in (3). Intuitively, as the budget \( \beta \) for an experiment increases, so should the experiment performance as the experiment can “explore more”. For small \( \beta \) the methods are nearly comparable as the additional information in minimal, but as \( \beta \) increases the performance of our method approaches that of the perfect knowledge. The A-optimal design only slowly decreases even though in the limit it should reach the lower bound.

![Figure 2: We compare the performance of our control-oriented system identification against A-optimal experiment design for a system with five states and three inputs, and known initial condition \( x_0 = [0, -4.3, 0., 2.1, 2.5]^T \) as in [19]. Any experiment design performance is lower-bounded by the optimal control given knowledge of \((A, B)\) and upper bounded by the performance of the controller given the a-priori system estimate. We include 95% confidence intervals using 10^5 samples.](image)

In Figure 3 we observe the optimal experiment inputs (\( \Delta u \)) for a three-car system. The experiment inputs for our control-oriented approach exhibit more excitation over the time horizon than the A-optimal design, which has a relatively smooth experiment input sequence, suggesting our method would perform better, which is verified by Figure 2. We also consider what the input sequence would look like if the controller given the a-priori system estimate is used during an experiment trial. Since the controller is closed-loop (11), we show the average input sequence. Using the optimal control may seem like a natural way to conduct an experiment, but the norm bound \( \beta \) on the input is not active in this case, indicating that simply performing the optimal control leaves experiment budget unused.

Figure 4 shows how the problem scales with the system dimension. In the first subplot we see the convergence of the experiment criteria in (4) as a function of iterations. The criteria is normalized by the lower bound (given by the performance if we knew \((A, B)\)) for illustrative purposes.
The number of iterations until the criteria stabilizes is roughly constant across problem dimension suggesting that the number of iterations required is independent of the system size in this case. In the second subplot, the time to compute each gradient sample is shown as a function of the state dimension. The compute time is dominated by “control solution”–the time to solve the LQR problem–thus suggesting that the time-complexity is dominated by the control problem. “Overhead” refers to the remainder of the compute tasks such as automatic differentiation and initialization of objects in the python library JAX [20].

![Fig. 3. The experiment input sequences for each car in a 3 car system are compared under the proposed method, A-optimal experiment design, and by conducting the experiment with a feedback controller given the prior.](image3)

![Fig. 4. In the upper subplot, the number of iterations to converge for the car string problem is essentially the same regardless of system size. The time to solve the classical LQR problem using automatic differentiation is shown as a function of the state dimension. The time is dominated by solving the control problem and “overhead” refers to tasks such as automatic differentiation, initial compile time, etc.](image4)

VI. CONCLUSION

We proposed a control-oriented identification approach that in expectation improves any data-driven controller by construction. Our solution method via stochastic gradient descent is shown in the LQR setting to provide solutions that outperform a typical experiment design objective in the sense that the post-experiment control performance is better with our method.

Our experiment design approach extends beyond the LQR setting, and it would be interesting to apply this to more general parametric problems. Establishing analytical results on the convergence rate and sample complexity of the stochastic gradient descent is an important direction for future research.

REFERENCES


