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 a feedback controller. 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 II 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 IV. 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 uncertainty in the parameters. We first illustrate our method's advantages when compared against the classical experiment design and then go on to show how the problem can be scaled to systems of a moderate order.

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

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2139319 and by the U.S. Office of Naval Research MURI grant No. N00014-23-1-2708. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation or U.S. Office of Naval Research.

The authors are with Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA, 93106 USA seananderson@ucsb.edu, hespanha@ece.ucsb.edu

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

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

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 θ that are unknown, but for which we have an *a-priori* distribution.

An experiment will be performed to provide additional information about the parameter θ. 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:

$$x_i^+ = f(x_i, u_i, w_i; \theta), \quad \forall i \in \{1, \dots, M\} \tag{2}$$

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 $x_i^+, x_i \in \mathbb{R}^{n_x}, u_i \in \mathbb{R}^{n_u}$ 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 π 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 θ. We also take as given a control design procedure that maps the experiment design data $\mathcal{D}$ to a specific controller π, 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:

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

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

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

where the expectation refers to an integration over (i) the *a-priori* distribution of the parameter θ, 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 θ. For example, an A-optimal experiment design essentially tries to minimize

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

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

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{6}$$

such that θ contains the elements of A and B, and w_t is Gaussian noise identically distributed across time with zero mean and covariance Σ_w.

We consider a quadratic optimization criterion of the form:

$$J(\pi;\theta) := \mathbb{E}\Big[x_N^T Q_N x_N + \sum_{t=0}^{N-1} x_t^T Q x_t + u_t^T R u_t \,|\, \theta\Big], \quad (7)$$

where the expectation refers to an integration over the disturbances encountered by the controller π; 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 $K_{CE}(\mathcal{D})$ computes the *a-posteriori* expected value of the unknown parameters $\hat{\theta} := \mathbb{E}[\theta \,|\, \mathcal{D}]$ and computes the linear optimal controller $u_t = K_t x_t$ that minimizes (7), *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 $K_{CE}(\mathcal{D})$, we employ linear regression on a dataset $\mathcal{D}$, which in our case will be the dataset generated under the experiment decision variable U. For identification, we express (6) as:

$$X^+ = \Theta Z + W \quad (8)$$

with $Z = [X; U] \in \mathbb{R}^{(n_x+n_u)\times M}$, and $\Theta := [A, B] \in \mathbb{R}^{n_x \times (n_x+n_u)}$. For ease of notation, we use $\theta \in \mathbb{R}^{n_x(n_x+n_u)}$ to denote the vectorized version of Θ 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 Θ given by $\mathbb{E}[(\hat{\Theta}-\Theta)_{ij}(\hat{\Theta}-\Theta)_{kl}] = (\Sigma_w)_{ki}(\Lambda_0^{-1})_{jl}$, where Σ_w is the known noise covariance and $\Lambda_0^{-1} \in \mathbb{R}^{(n_x+n_u)\times(n_x+n_u)}$ is a prior on the parameter covariance. The weighted Bayesian estimator for Θ is

$$\hat{\Theta} = (\Theta_0 \Lambda_0 + X^+ S Z^T)\Lambda_n^{-1}, \quad (9a)$$

and the error covariance of the estimate $\hat{\Theta}$ is

$$\mathbb{E}[(\hat{\Theta}-\Theta)_{ij}(\hat{\Theta}-\Theta)_{kl}] = (\Sigma_w)_{ki}(\Lambda_n^{-1})_{jl}, \quad (9b)$$

where $\Lambda_n := \Lambda_0 + ZSZ^T$, and $S \in \mathbb{R}^{M\times M}$ 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 Λ_n. In particular, let

$$S(X) := \text{diag}([s(x_0), ..., s(x_N)]) \quad (10)$$

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 $s(x) := \arctan(\|x_t - \alpha_1\|\alpha_2)/\pi + 0.5$, and α_1, α_2 are design parameters.

2) Certainty equivalent control: Given an estimate of the parameters θ from (9) with means $\hat{A}$ and $\hat{B}$, respectively, we construct our controller $K_{CE}(\mathcal{D})$ by recursively solving the Riccati difference equations given by

$$K_t = -(R + \hat{B}^T P_{t+1} \hat{B})^{-1} \hat{B}^T P_{t+1} \hat{A}, \quad (11a)$$

$$P_t = Q + K_t^T R K_t - (\hat{A} + \hat{B}K_t)^T P_{t+1} (\hat{A} + \hat{B}K_t), \quad (11b)$$

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, $\pi := \{K_0, ..., K_{N-1}\}$ from $K_{CE}(\mathcal{D})$, we can express the finite-horizon LQR cost (7) for the system in (6) parameterized by θ as

$$J(\pi;\theta) = x_0^T P_0 x_0 + \sum_{t=0}^{N-1} \text{tr}(P_{t+1}\Sigma_w), \quad (12a)$$

where

$$P_t = Q + K_t^T R K_t - (A + BK_t)^T P_{t+1}(A + BK_t), \quad (12b)$$

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 (4), we take a gradient-descent approach:

$$U_{i+1} = \text{Proj}_{\mathcal{U}}(U_i - \eta_i \hat{\nabla}_U), \quad (13)$$

where $\text{Proj}_{\mathcal{U}}(\cdot)$ projects U onto the set of admissible inputs $\mathcal{U}$, η_i is step size, and $\hat{\nabla}_U$ is an approximation of the true gradient ∇_U given by differentiating the experiment criteria (4):

$$\nabla_U \,\mathbb{E}\Big[J\Big(K\big(\mathcal{D}\big);\theta\Big)\Big] = \nabla_U \int J\big(K\big(\mathcal{D}\big);\theta\big) p(X \,|\, U, x_0, \theta) p(\theta) d\theta dX, \quad (14)$$

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

$$\nabla_U \int F(X,U) p(X|U) dX = \int \nabla_U \big(F(X,U) p(X|U)\big) dX, \quad (15)$$

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

$$\nabla_U \int F(X,U) p(X \,|\, U) dX = \int p(\epsilon) \nabla_U F(X,U)|_{X=g(\epsilon;U,x_0,\theta)} d\epsilon, \quad \epsilon \sim p(\epsilon), \quad (16)$$

and only $F(\cdot)$ needs to differentiated. This leads to the Monte Carlo pathwise gradient estimator [12]:

$$\frac{1}{L}\sum_{l=1}^{L}\nabla_U F(X,U)|_{X=g(\epsilon^{(l)};U)}, \quad \epsilon^{(l)}\sim p(\epsilon). \tag{17}$$

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

Theorem 1: The change of variable $g:\mathbb{R}^{n_x\times M}\times\mathbb{R}^{n_u\times M}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_\theta}\to\mathbb{R}^{n_x\times M}$ for (14) given by the recursion of the dynamics (1):

$$x_1 = f(x_0,u_0,w_0;\theta), \tag{18a}$$
$$x_2 = f(f(u_0,x_0,w_0;\theta),u_1,w_1;\theta), \tag{18b}$$
$$\vdots$$
$$x_M = f(f(x_{M-2},u_{M-2},w_{M-2};\theta),u_{M-1},w_{M-1};\theta), \tag{18c}$$

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 (1), use the Markov property to write the density of X as

$$p(X\,|\,U,x_0,\theta)=\Pi_{t=0}^{T-1}p(x_{t+1}\,|\,x_t,u_t,\theta). \tag{19a}$$

For any time t,

$$p(x_{t+1}\,|\,x_t,u_t,\theta)=p(x_{t+1}\,|\,x_t,u_t,w_t,\theta)p(w_t), \tag{19b}$$

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

$$p(x_{t+1}\,|\,x_t,u_t,w_t,\theta)=\Pi_{i=0}^{t}p(x_{i+1}\,|\,x_i,u_i,\theta)p(w_i) \tag{19c}$$

such that x_{t+1} can be expressed as $f\big(...f(x_0,u_0,w_0;\theta),u_t,w_t;\theta\big)$ where the first argument is taken to mean that the state is recursively defined given an initial state x_0, input sequence, and $w_0,...,w_t\sim p(w)$. 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 (6) has a change of variable $G(W;U,x_0,\theta)$, such that the tth column of X in (18) is given by

$$x_t = A^t x_0+\sum_{l=0}^{t-1}A^{t-1-l}(Bu_l+w_l), \tag{20a}$$

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

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

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

$$\hat{\nabla}_U=\frac{1}{L}\sum_{l=1}^{L}\nabla_U J\Big(K\big(\mathcal{D}|_{X=G(W^{(l)};U,x_0,\theta^{(l)})}\big);\theta^{(l)}\Big), \tag{20b}$$

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

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

A. Algorithm for Experiment Design Problem

In the LQR setting we derived a pathwise estimator (20). For more general problems as in (4), if we assume the exchange of integral and gradient is valid, we can express the pathwise gradient estimate as

$$\hat{\nabla}_U=\frac{1}{L}\sum_{l=1}^{L}\nabla_U J\Big(K\big(\mathcal{D}|_{X=G(W^{(l)};U,x_0,\theta^{(l)})}\big);\theta^{(l)}\Big). \tag{21}$$

For each sample, l, we obtain a single experiment realization under sampled noise W for a sampled system, θ. For this realization, we compute an *a-posteriori* system estimate, $\hat{\theta}$, and compute the control, π, as in (3). Computing the gradient $\nabla_U J$ can be done analytically in some cases. Given the structure of J, it may be easier to use automatic differentiation as we go on to do in Section V.

Algorithm 1 Control-Oriented Experiment Design

Input p_θ (prior on θ), U_0 (initialization), L (batch size), $\mathcal{U}$ (feasible set), x_0 (initial condition), p_W (noise dist.)
Output U^*

function ConstructControl($\mathcal{D}, p_\theta$)
  $\hat{\theta}\leftarrow$ estimate system given $\mathcal{D}, p_\theta$
  $K\leftarrow$ Solve control problem given $\hat{\theta}$
end function
function SampleGrad(U,θ,W,p_θ,x_0)
  $X\leftarrow G(W;U,x_0,\theta)$
  $\mathcal{D}\leftarrow X,U$
  $K\leftarrow$ ConstructControl($\mathcal{D},p_\theta$)
  $\nabla_U J\leftarrow$ Compute gradient of $J(K;\theta)$
end function
while not converged **do**
  for i=1 to L **do**
    $\theta\sim p_\theta, W\sim p_W$
    $\nabla_U J_i\leftarrow$ SampleGrad($U_j,\theta,W,p_\theta,x_0$)
  end for
  $U_{j+1}\leftarrow \text{Proj}_{\mathcal{U}}\big(U_j-\eta_j\frac{1}{L}\sum_{l=1}^{L}\nabla_U J_l\big)$
end while

V. Numerical Experiments

A. Car String

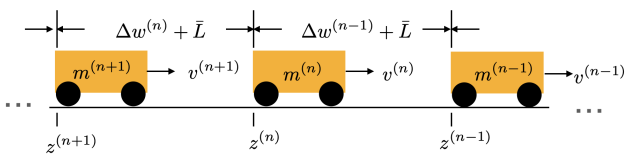

Fig. 1. We focus on a scenario in which cars need to regulate to a specified gap $\bar{L}$ at a desired reference velocity v. Deviations from the desired position are $\Delta w^{(n)}$ and deviations from the reference velocity are $\Delta v^{(n)}$.

We consider the problem of maintaining a fixed distance, $\bar{L}$, 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:

$$\Delta v_{t+1}^{(n)} = \left(-\frac{\alpha^{(n)} T_s}{m^{(n)}} + 1 \right) \Delta v_t^{(n)} + \frac{T_s}{m^{(n)}} \Delta u_t^{(n)}, \quad (22)$$

$$\Delta w_{t+1}^{(n)} = T_s (\Delta v_t^{(n)} - \Delta v_t^{(n+1)}) + \Delta w_t^{(n)}, \quad (23)$$

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

$$A = \begin{bmatrix} -\frac{\alpha^{(1)} T_s}{m^{(1)}} + 1 & 0 & 0 & 0 & \dots \\ T_s & 1 & -T_s & 0 & \dots \\ 0 & 0 & -\frac{\alpha^{(2)} T_s}{m^{(2)}} + 1 & 0 & \dots \\ 0 & 0 & T_s & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (24)$$

$$B = \begin{bmatrix} \frac{T_s}{m^{(1)}} & 0 & \dots \\ 0 & 0 & \dots \\ 0 & \frac{T_s}{m^{(2)}} & \cdots \\ 0 & 0 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}. \quad (25)$$

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 (6) as $\Sigma_w = 1\text{e}{-2} \times I_5$. The prior on the parameters (9) is $\Theta_0 = [A, B]$ with $m^{(1)} = m^{(2)} = m^{(3)} = 1$, $\alpha^{(1)} = \alpha^{(2)} = \alpha^{(3)} = 1$, $T_s = 0.1$; $\Lambda_0^{-1} = 1\text{e}{-2} \times I_8$. 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 (7) Q includes penalties of magnitude 10 on the positions Δw and zero on the velocity Δv. R is the identity matrix. The weight matrix S 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 (9).

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 (7). For the experiment design (4), we consider a feasible input set $\mathcal{U}$ where the Frobenius norm of the input sequence, U, is bounded by a design parameter β such that $\mathcal{U} = \{U \,|\, \|U\|_F \leq \beta\}$. We vary the allowed magnitude, β, 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 β for an experiment increases, so should the experiment performance as the experiment can "explore more". For small β the methods are nearly comparable as the additional information in minimal, but as β increases the performance of our method approaches that of the perfect knowledge case. The A-optimal design only slowly decreases even though in the limit it should reach the lower bound.

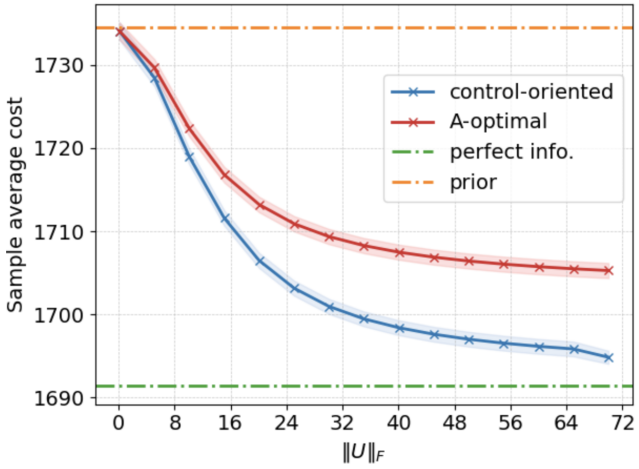

Fig. 2. We compare the performance of our control-oriented system identification against A-optimal experiment design for a system with five states and three inputs, and known initial condition $x_0 = [0., -4.3, 0., 2.1, 2.5]^T$ as in [19]. Any experiment design performance is lower-bounded by the optimal 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.

In Figure 3, we observe the optimal experiment inputs (Δu) 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 β 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].

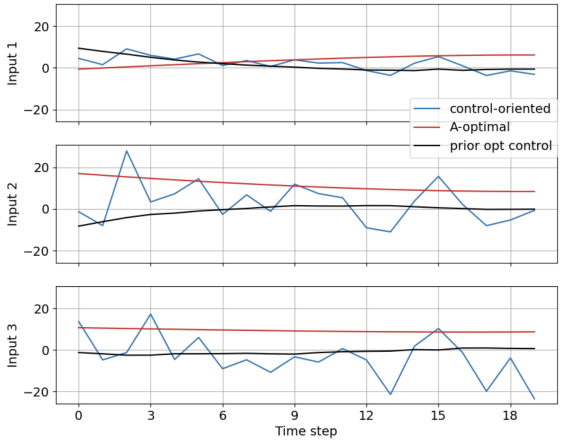

Fig. 3. The experiment input sequences for each car in a 3 car system are compared under the proposed method, A-optimal experiment design, and by conducting the experiment with a feedback controller given the prior.

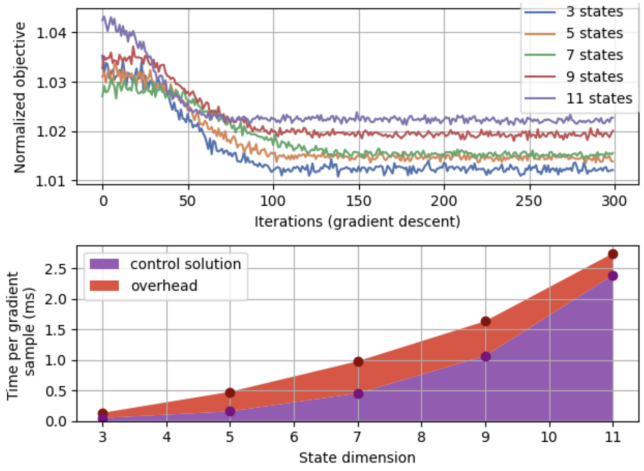

Fig. 4. In the upper subplot, the number of iterations to converge for the car string problem is essentially the same regardless of system size suggesting good scaling properties of our method. In the lower subplot, we observe the average time to compute a single gradient sample. The time is dominated by solving the control problem and "overhead" refers to tasks such as automatic differentiation, initial compile time, etc.

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

[1] H. Chernoff, "Locally Optimal Designs for Estimating Parameters," *The Annals of Mathematical Statistics*, vol. 24, no. 4, pp. 586–602, Dec. 1953, publisher: Institute of Mathematical Statistics.

[2] H. A. Simon, "Dynamic Programming Under Uncertainty with a Quadratic Criterion Function," *Econometrica*, vol. 24, no. 1, pp. 74–81, 1956, publisher: [Wiley, Econometric Society].

[3] G. Elfving, "Optimum Allocation in Linear Regression Theory," *The Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 255–262, Jun. 1952, publisher: Institute of Mathematical Statistics.

[4] K. Lindqvist and H. Hjalmarsson, "Identification for control: adaptive input design using convex optimization," in *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No.01CH37228)*, vol. 5. Orlando, FL, USA: IEEE, 2001, pp. 4326–4331.

[5] M. Gevers, "Identification for Control: From the Early Achievements to the Revival of Experiment Design," in *Proceedings of the 44th IEEE Conference on Decision and Control*, Dec. 2005, pp. 12–12, iSSN: 0191-2216.

[6] S. Anderson, K. Byl, and J. P. Hespanha, "Experiment design with gaussian process regression with applications to chance-constrained control," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 3931–3938.

[7] B. D. Lee, I. Ziemann, A. Tsiamis, H. Sandberg, and N. Matni, "The Fundamental Limitations of Learning Linear-Quadratic Regulators," Mar. 2023, arXiv:2303.15637 [cs, eess].

[8] A. J. Wagenmaker, M. Simchowitz, and K. Jamieson, "Task-Optimal Exploration in Linear Dynamical Systems," in *Proceedings of the 38th International Conference on Machine Learning*. PMLR, Jul. 2021, pp. 10 641–10 652, iSSN: 2640-3498.

[9] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the Sample Complexity of the Linear Quadratic Regulator," *Foundations of Computational Mathematics*, vol. 20, no. 4, pp. 633–679, Aug. 2020.

[10] K. J. Åström and B. Wittenmark, *Adaptive control*. Courier Corporation, 2008.

[11] M. Simchowitz and D. Foster, "Naive Exploration is Optimal for Online LQR," in *Proceedings of the 37th International Conference on Machine Learning*. PMLR, Nov. 2020, pp. 8937–8948, iSSN: 2640-3498.

[12] S. Mohamed, M. Rosca, M. Figurnov, and A. Mnih, "Monte Carlo Gradient Estimation in Machine Learning," Sep. 2020.

[13] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural Networks*, vol. 21, no. 4, pp. 682–697, May 2008.

[14] P. Rossi, G. Allenby, and R. McCulloch, *Bayesian Statistics and Marketing*. Wiley, Oct. 2006.

[15] D. Bertsekas, *Dynamic programming and optimal control: Volume I*. Athena scientific, 2012, vol. 4.

[16] P. Glasserman, *Monte Carlo methods in financial engineering*. Springer, 2004, vol. 53.

[17] G. Grimmett and D. Stirzaker, *Probability and random processes*. Oxford university press, 2020.

[18] S. Anderson and J. P. Hespanha, "Control-oriented identification for the linear quadratic regulator: Technical report," Santa Barbara, Mar. 2024. [Online]. Available: https://arxiv.org/abs/2403.05455

[19] W. Levine and M. Athans, "On the optimal error regulation of a string of moving vehicles," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 355–361, Jul. 1966, conference Name: IEEE Transactions on Automatic Control.

[20] "JAX: Autograd and XLA," Feb. 2023, original-date: 2018-10-25T21:25:02Z.