Certainty Equivalent Perception-Based Control

Sarah Dean DEAN_SARAH@BERKELEY.EDU
University of California, Berkeley

Benjamin Recht BRECHT@BERKELEY.EDU
University of California, Berkeley

Editors: A. Jadbabaie, J. Lygeros, G. J. Pappas, P. A. Parrilo, B. Recht, C. J. Tomlin, M. N. Zeilinger

Abstract

In order to certify performance and safety, feedback control requires precise characterization of sensor errors. In this paper, we provide guarantees on such feedback systems when sensors are characterized by solving a supervised learning problem. We show a uniform error bound on nonparametric kernel regression under a dynamically-achievable dense sampling scheme. This allows for a finite-time convergence rate on the sub-optimality of using the regressor in closed-loop for waypoint tracking. We demonstrate our results in simulation with simplified unmanned aerial vehicle and autonomous driving examples.

1. Introduction

Machine learning provides a promising avenue for incorporating rich sensing modalities into autonomous systems. However, our coarse understanding of how ML systems fail limits the adoption of data-driven techniques in real-world applications. In particular, applications involving feedback require that errors do not accumulate and lead to instability. In this work, we propose and analyze a baseline method for incorporating a learning-enabled component into closed-loop control, providing bounds on the sample complexity of a reference tracking problem.

Much recent work on developing guarantees for learning and control has focused on the case that dynamics are unknown (Dean et al., 2017; Simchowitz and Foster, 2020; Mania et al., 2020). In this work, we consider a setting in which the uncertainty comes from the sensor generating observations about system state. By considering unmodeled, nonlinear, and potentially high dimensional sensors, we capture phenomena relevant to modern autonomous systems (Codevilla et al., 2018; Lambert et al., 2018; Tang et al., 2018).

Our analysis combines contemporary techniques from statistical learning theory and robust control. We consider learning an inverse perception map with noisy measurements as supervision, and show that it is *necessary* to quantify the uncertainty pointwise to guarantee robustness. Such pointwise guarantees are more onerous than typical mean-error generalization bounds, and require a number of samples scaling exponentially with dimension. However, many interesting problems in robotics and automation are low dimensional, and we provide a high-probability pointwise error bound on nonparametric regression for such scenarios. Under a dynamically feasible dense sampling scheme, we show uniform converge for the learned map. Finally, we analyze the suboptimality of using the learned component in closed-loop, and demonstrate the utility of our method for reference tracking. We close with several numerical examples showing that our method is feasible for many problems of interest in autonomy. Full proofs of the main results, further discussion on controller

synthesis, illustrative examples, and experimental details are included in the longer version of this paper (Dean and Recht, 2020).

1.1. Problem Setting

Motivated by situations in which control is difficult due to sensing, we consider the task of waypoint tracking for a system with known, linear dynamics and complex, nonlinear observations. The setting is succinctly described by the motivating optimal control problem:

$$\min_{\pi} \quad \sup_{\substack{\{x_k^{\mathrm{ref}}\} \in \mathcal{R} \\ \|x_0\|_\infty \leq \sigma_0}} \left\| \begin{bmatrix} Q^{1/2}(x_k - x_k^{\mathrm{ref}}) \\ R^{1/2} u_k \end{bmatrix} \right\|_\infty \tag{1}$$

$$\text{s.t.} \;\; x_{k+1} = Ax_k + Bu_k, \;\; z_k = g(Cx_k), \;\; u_k = \pi(z_{0:k}, x_{0:k}^{\mathrm{ref}}), \tag{2}$$

where $x \in \mathbb{R}^n$ is the system state, $u \in \mathbb{R}^m$ is the control input, $z \in \mathbb{R}^q$ is the system observation, and we use the shorthand $x_{0:k} = (x_0, \ldots, x_k)$. The linear dynamics are specified by $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and the matrix $C \in \mathbb{R}^{p \times n}$ determines the state subspace that affects the observation, which we will refer to as the *measurement subspace*. We assume that (A, B) is controllable and (A, C) is observable. This optimal control problem seeks a causal policy π to minimize a robust waypoint tracking cost. The objective in (1) penalizes the maximum deviation from any reference trajectory in the class $\mathcal{R}$ as well as the maximum control input, with the relative importance of these terms determined by $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$. A common class of reference signal is those with bounded differences, i.e. $\|x_k^{\mathrm{ref}} - x_{k+1}^{\mathrm{ref}}\|_\infty \leq \Delta$.

As defined by the constraints of this problem, the system has linear dynamics, but nonlinear observations (2). We suppose that the parameters of the linear dynamics (A, B, C) are known, while the observation function $g : \mathbb{R}^p \to \mathbb{R}^q$ is unknown. This emulates a natural setting in which the dynamics of a system are well-specified from physical principles (e.g. vehicle dynamics), while observations from the system (e.g. camera images) are complex but encode a subset of the state (e.g. position). We assume that g is continuous and that there is a continuous inverse function $h : \mathbb{R}^q \to \mathbb{R}^p$ with $h(g(y)) = y$. For the example of a dashboard mounted camera, such an inverse exists whenever each camera pose corresponds to a unique image, which is a reasonable assumption in sufficiently feature rich environments.

If the inverse function h were known, then the optimal control problem is equivalent to a linear output-feedback control problem. We therefore pose a learning problem focused on the unknown inverse function h, which we will call a *perception map*. We suppose that during the training phase, there is an additional system output,

$$y_k^{\mathrm{train}} = Cx_k + \eta_k \tag{3}$$

where each η_k is zero mean and has independent entries bounded by σ_η. This assumption corresponds to using a simple but noisy sensor for characterizing the complex sensor. The noisy system output will both supervise the learning problem and allow for the execution of a sampling scheme where the system is driven to sample particular parts of the state space. Notice that due to its noisiness, using this sensor for waypoint tracking would be suboptimal compared using transformed observations.

We will show that the *certainty equivalent* controller, i.e. the controller which treats the approximation $\widehat{h}(z) \approx Cx$ as true, achieves a cost which converges towards optimality.

Result 1 (Informal) *Certainty equivalent control using a perception map learned with T sampled data points by our method achieves a suboptimality bounded by*

$$c(\widehat{\pi}) - c(\pi_\star) \lesssim L r_\star S_\star \left(\frac{\sigma}{T}\right)^{\frac{1}{p+4}} . \tag{4}$$

Here, $c(\cdot)$ is the cost objective in (1), L describes the continuity of the relationship between system state and the observations, $r_\star$ bounds the region of system operation under the optimal control law, $S_\star$ is the sensitivity to measurement error of the optimal closed-loop, σ is proportional to the sensor noise in (3), and p is the dimension of measurement subspace.

1.2. Related Work

The model we study is inspired by examples of control from pixels, primarily in the domain of robotics. At one end of the spectrum, calibrated camera characteristics are incorporated directly into physical models; this type of visual-inertial odometry enables agressive control maneuvers (Tang et al., 2018). On the other end are policies that map pixels directly to low-level control inputs, for example via imitation learning (Codevilla et al., 2018). Our model falls between these two approaches: relying on known system dynamics, but deferring sensor characterization to data.

The observation of a linear system through a static nonlinearity is classically studied as a *Weiner system* model (Schoukens and Tiels, 2016). While there are identification results for Weiner systems, they apply only to single-input-single-output systems, and often require assumptions that do not apply to the motivation of cameras (Hasiewicz, 1987; Wigren, 1994; Greblicki, 1997; Tiels and Schoukens, 2014). More flexible identification schemes have been proposed (Lacy and Bernstein, 2001; Salhi and Kamoun, 2016), but they lack theoretical guarantees. Furthermore, these approaches focus on identifying the full forward model, which may not be necessary for good closed-loop performance. The variant of Weiner systems that we study is closely related to our recent work, which focused on robustness in control design (Dean et al., 2019). We now extend these ideas to directly consider issues of sampling and noise.

There is much related and recent work at the intersection of learning and control. Similar in spirit is a line of work on the Linear Quadratic Regulator which focuses on issues of system identification and sub-optimality (Dean et al., 2017, 2018; Abbasi-Yadkori and Szepesvári, 2011). This style of sample complexity analysis has allowed for illuminating comparisons between model-based and model-free policy learning approaches (Tu and Recht, 2018; Abbasi-Yadkori et al., 2019). Mania et al. (2019); Simchowitz and Foster (2020) show that the simple strategy of model-based certainty equivalent control is efficient, though the argument is specialized to linear dynamics and quadratic cost. For nonlinear systems, analyses of learning often focus on ensuring safety over identification or sub-optimality (Taylor et al., 2019; Berkenkamp et al., 2017; Wabersich and Zeilinger, 2018; Cheng et al., 2019), and rely on underlying smoothness for their guarantees (Liu et al., 2019; Nakka et al., 2020). An exception is a recent result by Mania et al. (2020) which presents finite sample guarantees for parametric nonlinear system identification.

While the majority of work on learning for control focuses on settings with full state observation, output feedback is receiving growing attention for linear systems (Simchowitz et al., 2019, 2020) and for safety-critical systems (Laine et al., 2020). Recent work in closely related problem settings includes Mhammedi et al. (2020), who develop sample complexity guarantees for LQR with nonlinear observations and Misra et al. (2020), who leverage representation learning in Block MDPs; however, neither address issues of stability due to focusing on finite horizon problems.

The statistical analysis presented here focuses on nonparametric pointwise error bounds over a compact set. Distinct from mean-error generalization arguments most common in learning theory, our analysis is directly related to classical statistical results on uniform convergence (Devroye, 1978; Liero, 1989; Hansen, 2008). Our motivation is related to conformal regression (Lei and Wasserman, 2014; Barber et al., 2019), which relies on exchangeability assumptions that can be adapted to data from dynamical systems with mixing arguments (Chernozhukov et al., 2018).

2. Uniform Convergence of Perception

In this section, we introduce a sampling scheme and nonparametric regression strategy for learning the predictor $\widehat{h}$ and show that the resulting errors are uniformly bounded. While it is typical in the machine learning community to consider mean-error generalization bounds, robust control objectives require that uncertainty be quantified pointwise to guarantee. It is easy to construct examples in which errors within sets of vanishingly small measure cause systems to exit bounded regions of well-characterized perception and lead to instability (e.g. Dean and Recht (2020)).

Therefore, we begin by introducing a method for nonparametric regression and present a data-dependent pointwise error bound. Then, we propose a dense sampling scheme and show a uniform error bound over the sampled region of the measurement subspace.

2.1. Nonparametric Regression

Since uniform error bounds are necessary for robust guarantees, we now introduce a method to learn perception maps with such bounds. For simplicity of analysis and exposition, we focus on Nadarya-Watson estimators. We expect that our insights will generalize to more complex techniques, and we demonstrate similarities with additional regressors in simulation experiments presented in Section 4.

The Nadarya-Watson regression estimators with training data $\{(z_t, y_t^{\text{train}})\}_{t=0}^{T}$, bandwidth $\gamma \in \mathbb{R}_+$, and metric $\rho : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}_+$ have the form

$$\widehat{h}(z) = \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} y_t^{\text{train}}, \quad s_T(z) = \sum_{t=0}^{T} \kappa_\gamma\left(z_t, z\right), \quad \kappa_\gamma\left(z_t, z\right) = \kappa\left(\frac{\rho(z_t, z)}{\gamma}\right), \tag{5}$$

with $\widehat{h}(z) = 0$ when $s_T(z) = 0$ and $\kappa : \mathbb{R}_+ \to [0, 1]$ is a kernel function. We assume that the kernel function is Lipschitz with parameter L_κ and that $\kappa(u) = 0$ for $u > 1$, and define the quantity $V_\kappa = \int_{\mathbb{R}_+^p} \kappa\left(\|y\|_\infty\right) dy$. Thus, predictions are made by computing a weighted average over the labels y_t^{train} of training datapoints whose corresponding observations z_t are close to the current observation, as measured by the metric ρ. We assume the functions g and h are Lipschitz continuous with respect to ρ, i.e. for some L_g and L_h

$$\rho(g(y), g(y')) \leq L_g \|y - y'\|_\infty, \quad \|h(z) - h(z')\|_\infty \leq L_h \rho(z, z') \,. \tag{6}$$

While our final sub-optimality results depend on L_g and L_h, the perception map and synthesized controller do not need direct knowledge of these parameters.

For an arbitrary z with $s_T(z) \neq 0$, the prediction error can be decomposed as

$$\|\widehat{h}(z) - h(z)\|_\infty \leq \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} (Cx_t - Cx) \right\|_\infty + \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} \eta_t \right\|_\infty . \tag{7}$$

The first term is the approximation error due to finite sampling, even in the absence of noisy labels. This term can be bounded using the continuity of the true perception map h. The second term is the error due to measurement noise. We use this decomposition to state a pointwise error bound, which can be used to provide tight data-dependent estimates on error.

Lemma 1 *For a learned perception map of the form (5) with training data as in (3) collected during closed-loop operation of a system satisfying (6), we have with probability at least $1-\delta$ that for a fixed z with $s_T(z) \neq 0$,*

$$\|\widehat{h}(z) - h(z)\|_\infty \leq \gamma L_h + \frac{\sigma_\eta}{\sqrt{s_T(z)}} \sqrt{\log\left(p^2 \sqrt{s_T(z)}/\delta\right)} \,. \tag{8}$$

The expression illustrates that there is a tension between having a small bandwidth γ and ensuring that the coverage term $s_T(z)$ is large. Notice that most of the quantities in this upper bound can be readily computed from the training data; only L_h, which quantifies the continuity of the map from observation to state, is difficult to measure. We remark that while useful for building intuition, the result in Lemma 1 is only directly applicable for bounding error at a finite number of points. Since our main results handle stability over infinite horizons, they rely on a modified bound introduced in the full version of this paper (Dean and Recht, 2020) which is closely tied to continuity properties of the estimated perception map $\widehat{h}$ and the sampling scheme we propose in the next section.

2.2. Dense Sampling

We now propose a method for collecting training data and show a uniform, sample-independent bound on perception errors under the proposed scheme. This strategy relies on the structure imposed by the continuous and bijective map g, which ensures that driving the system along a dense trajectory in the measurement subspace corresponds to collecting dense samples from the space of observations. In what follows, we provide a method for driving the system along such a trajectory.

We assume that during training, the system state can be reset according to a distribution $\mathcal{D}_0$ which has has support bounded by σ_0. We do not assume that these states are observed. Between resets, an affine control law drives the system to evenly sample the measurement subspace with a combination of a stabilizing output feedback controller and reference tracking inputs:

$$u_t = \sum_{k=0}^{t} K_k y_{t-k}^{\text{train}} + u_t^{\text{ref}} \,. \tag{9}$$

The stabilizing feedback controller prevents the accumulation of errors resulting from the unobserved reset state. The closed-loop trajectories resulting from this controller are

$$x_t = \Phi_x(t) x_0 + \sum_{k=1}^{t} \Phi_{xu}(k) u_{t-k}^{\text{ref}} + \Phi_{xn}(k) \eta_{t-k} \,, \tag{10}$$

where the system response variables $\{\Phi_x(k), \Phi_{xu}(k), \Phi_{xn}(k)\}_{k\geq 0}$ arise from the interaction of the stabilizing control law with the linear dynamics; we revisit this fact in detail in Section 3.1. As long as the feedback control law $\{K_k\}_{k\geq 0}$ is chosen such that the closed-loop system is stable, the

Input: System variables C, stabilizing controller $\{K_k\}_{k=1}^n$ and system response $\{\Phi_{xu}(k)\}_{k=1}^n$, sampling radius $\bar{r}$, target dataset size T.

for $\ell \in \{1, \ldots, T\}$ **do**

reset $x_{0,\ell} \sim \mathcal{D}_0$ and sample $y_\ell^{\text{ref}} \sim \text{Unif}(\{y \mid \|y\|_\infty \leq \bar{r}\})$

design inputs $\left[(u_{0,\ell}^{\text{ref}})^\top, \ldots, (u_{n-1,\ell}^{\text{ref}})^\top\right]^\top := \begin{bmatrix} C\Phi_{xu}(1) & \ldots & C\Phi_{xu}(n) \end{bmatrix}^\dagger y_\ell^{\text{ref}}$

drive the system to states $x_{k+1,\ell}$ with $u_{k,\ell} = \sum_{j=0}^k K_j y_{j-k,t}^{\text{train}} + u_{k,\ell}^{\text{ref}}$ for $k = 0, \ldots, n-1$

end

Output: Uniformly sampled training data $\{(z_{n,\ell}, y_{n,\ell}^{\text{train}})\}_{\ell=1}^T =: \{(z_t, y_t^{\text{train}})\}_{t=1}^T$

Algorithm 1: Uniform Sampling with Resets

system response variables decay. Designing such a stabilizing controller is possible since (A, B) is controllable and (A, C) is observable. We therefore assume that for all $k \geq 0$

$$\max\left\{\|C\Phi_x(k)\|_\infty, \|C\Phi_{xn}(k)\|_\infty\right\} \leq M\rho^k, \tag{11}$$

for some $M \geq 1$ and $0 \leq \rho < 1$.

The reference inputs u_t^{ref} are chosen to ensure even sampling. Since the pair (A, B) is controllable, these reference inputs can drive the system to any state within n steps. Algorithm 1 leverages this fact to construct control sequences which drive the system to points uniformly sampled from the measurement subspace. Its use of system resets ensures independent samples; we note that since the closed-loop system is stable, such a "reset" can approximately be achieved by waiting long enough with zero control input.

As a result of the unobserved reset states and the noisy sensor, the states visited while executing Algorithm 1 do not exactly follow the desired uniform distribution. They can be decomposed into two terms:

$$Cx_{n,\ell} = \sum_{k=1}^n C\Phi_{xu}(k)u_{n-k,\ell}^{\text{ref}} + \left(C\Phi_x(n)x_{0,t} + \sum_{k=1}^n C\Phi_{xn}(k)\eta_{n-k,\ell}\right) =: y_\ell^{\text{ref}} + w_\ell$$

where y_ℓ^{ref} is uniformly sampled from $\{y \mid \|y\|_\infty \leq \bar{r}\}$, and the noise variable w_ℓ is bounded:

$$\|w_\ell\|_\infty \leq \|C\Phi_x(n)\|_\infty\|x_0\|_\infty + \sum_{\ell=1}^n \|C\Phi_{xn}(\ell)\|_\infty\|\eta_{n-\ell}\|_\infty \leq \frac{M\max\{\sigma_0, \sigma_\eta\}}{1-\rho}.$$

The following Lemma shows that uniform samples corrupted with independent and bounded noise ensure dense sampling of the measurement subspace by providing a high probability lower bound on the coverage $s_T(z)$.

Lemma 2 *Suppose that training data $\{z_t\}_{t=1}^T$ is collected with a stabilizing controller satisfying (11) according to Algorithm 1 with $\bar{r} \geq r + \frac{M\max\{\sigma_0,\sigma_\eta\}}{1-\rho} + \frac{\gamma}{L_g}$ from a system satisfying (6). Then for all z observed from a state x satisfying $\|Cx\|_\infty \leq r$,*

$$s_T(z) \geq \frac{1}{2}\sqrt{TV_\kappa}\left(\frac{\gamma}{\bar{r}L_g}\right)^{\frac{p}{2}}$$

with probability at least $1-\delta$ as long as $T \geq 8V_\kappa^{-1}\log(1/\delta)(\bar{r}L_hL_g^2)^p\gamma^{-p}$.

We use this coverage property of the training data and the error decomposition presented in Section 2.1 to show our main uniform convergence result.

Theorem 1 *If training data satisfying (3) is collected with a stabilizing controller satisfying (11) by the sampling scheme in Algorithm 1 with radius $\bar{r} = \sqrt{2}r$ from a system satisfying (6), then as long as the system remains within the set $\{x \mid \|Cx\|_\infty \le r\}$, the Nadarya-Watson regressor (5) will have bounded perception error for every observation z:*

$$\|\widehat{h}(z) - h(z)\|_\infty \le \gamma L_h + \frac{\sigma_\eta}{T^{\frac{1}{4}}} \left(\frac{L_g\sqrt{2}r}{\gamma}\right)^{\frac{p}{4}} \left(\sqrt{p\log\left(T^2/\delta\right)} + 1\right), \tag{12}$$

with probability at least $1-\delta$ as long as $\gamma \le L_g((\sqrt{2}-1)r - M\max\{\sigma_0, \sigma_\eta\}(1-\rho)^{-1})$ and

$$T \ge \max\left\{8pV_\kappa^{-1}(\sqrt{2}L_hL_g^2)^p(r/\gamma)^p\log(T^2/\delta),\ V_\kappa^{-\frac{1}{3}}(24L_\kappa L_h)^{\frac{4}{3}}L_g^{\frac{p}{3}}(r/\gamma)^{\frac{p+4}{3}}\right\}.$$

3. Closed-Loop Guarantees

The previous section shows that nonparametric regression can be successful for learning the perception map within a bounded region of the state space. But how do these bounded errors translate into closed-loop performance guarantees, and how can we ensure that states remain within the bounded region? To answer this question, we recall the waypoint tracking problem in (1).

3.1. Linear Control for Waypoint Tracking

Consider a linear control law for waypoint tracking $u_t = \sum_{k=0}^t K_k^{(y)} y_{t-k} + \sum_{k=0}^t K_k^{(r)} x_{t-k}^{\text{ref}}$ which depends on some output signal $y_k = Cx_k + \eta_k$. Similar in form to the controller used for sampling (9), this linear waypoint tracking controller can be viewed as computing u_t^{ref} based on waypoints x_k^{ref}. As first discussed in our sampling analysis (10), the closed-loop behavior of a linear system in feedback with a linear controller can be described as a linear function of noise variables.

To facilitate our discussion of the general *system response*, we will introduce boldface notation for infinite horizon signals and convolutional operators. Under this notation, we can write $u_t = \sum_{k=0}^t K_k x_{t-k}$ equivalently as $\mathbf{u} = \mathbf{K}\mathbf{x}$. We introduce the signal and linear operator norms

$$\|\mathbf{x}\|_\infty = \sup_k \|x_k\|_\infty, \quad \|\mathbf{\Phi}\|_{\mathcal{L}_1} = \sup_{\|\mathbf{w}\|_\infty \le 1} \|\mathbf{\Phi}\mathbf{w}\|_\infty .$$

These signals and operators can be concretely represented in terms of semi-infinite block-lower-triangular Toeplitz matrices acting on semi-infinite vectors or in the z-domain with $\mathbf{x} = \sum_{k=0}^\infty x_k z^{-k}$.

Using this representation, a linear controller $(\mathbf{K}_y, \mathbf{K}_r)$ induces the system responses

$$\mathbf{\Phi_x} = (zI - (A + B\mathbf{K}_y C))^{-1} \qquad \mathbf{\Phi_{xn}} = \mathbf{\Phi_x} B\mathbf{K}_y \qquad \mathbf{\Phi_{xr}} = \mathbf{\Phi_x} B\mathbf{K}_r \tag{13}$$

which are well defined as long as the interconnection is stable. Under this definition, the state signal can be written as $\mathbf{x} = \mathbf{\Phi_{xr}}\mathbf{x}^{\text{ref}} + \mathbf{\Phi_{xn}}\mathbf{n} + \mathbf{\Phi_x}x_0$. The responses $\mathbf{\Phi_{un}}$ and $\mathbf{\Phi_{ur}}$ can be similarly defined, since $\mathbf{u}$ is linear in $\mathbf{x}$ and $\mathbf{x}^{\text{ref}}$.

The robust waypoint tracking cost from (1) can be cast as an $\mathcal{L}_1$ norm on system response variables. For the purposes of our main result, it is only necessary to view system response variables as objects that arise from linear controllers. However, system response variables can also be used to synthesize linear controllers. This is a key insight of the *System Level Synthesis* framework, first introduced by Wang et al. (2016). We include further discussion in the full version of this paper (Dean and Recht, 2020).

3.2. Suboptimality of Certainty-Equivalence

Suppose that we apply a linear controller to our perception estimates,

$$\mathbf{u} = \widehat{\pi}(\mathbf{z}, \mathbf{x}^{\mathrm{ref}}) = \mathbf{K}_y \widehat{h}(\mathbf{z}) + \mathbf{K}_r \mathbf{x}^{\mathrm{ref}} =: \mathbf{K}(\widehat{h}(\mathbf{z}), \mathbf{x}^{\mathrm{ref}}). \tag{14}$$

This is the certainty equivalent controller, which treats the learned perception map as if it is true. We will compare this controller with $\pi_\star = \mathbf{K}(h(\cdot), \cdot)$, the result of perfect perception. The suboptimality depends on the following quantity, which bounds $\|C\mathbf{x}\|_\infty$ under the optimal control law

$$r_{\max}(\mathbf{\Phi}) = \sup_{\substack{\mathbf{x}^{\mathrm{ref}} \in \mathcal{R} \\ \|x_0\|_\infty \le \sigma_0}} \|C\mathbf{\Phi}_{\mathbf{xr}}\mathbf{x}^{\mathrm{ref}} + C\mathbf{\Phi}_{\mathbf{x}}x_0\|_{\mathcal{L}_1} .$$

The magnitude of this value depends on properties of the considered reference signal class. In the extended version of this paper (Dean and Recht, 2020), we present the specific form of this quantity for reference signals that are bounded and have bounded differences.

Proposition 1 *Let $\{\mathbf{\Phi}_{\mathbf{xr}}, \mathbf{\Phi}_{\mathbf{xn}}, \mathbf{\Phi}_{\mathbf{ur}}, \mathbf{\Phi}_{\mathbf{un}}\}$ denote the system responses induced by the controller $\mathbf{K}$ as in (13), and let $c(\pi_\star)$ denote the cost associated with the policy $\pi_\star = \mathbf{K}(h(\cdot), \cdot)$. Then for a perception component with error bounded by ε_h within the set $\{x \mid \|Cx\|_\infty \le r\}$, the sub-optimality of the certainty-equivalent controller (14) is bounded by*

$$c(\widehat{\pi}) - c(\pi_\star) \le \varepsilon_h \left\| \begin{bmatrix} Q^{1/2}\mathbf{\Phi}_{\mathbf{xn}} \\ R^{1/2}\mathbf{\Phi}_{\mathbf{un}} \end{bmatrix} \right\|_{\mathcal{L}_1} .$$

as long as the sampled region is large enough and the errors are small enough, $\varepsilon_h \le \frac{r - r_{\max}(\mathbf{\Phi})}{\|C\mathbf{\Phi}_{\mathbf{xn}}\|_{\mathcal{L}_1}}$.

Thus, the optimal closed-loop system's sensitivity to measurement noise is closely related to the sub-optimality. It is possible to use this insight to augment the cost of the waypoint tracking problem in (1) to make it more robust to perception errors (Dean and Recht, 2020).

We now state our main result.

Corollary 3 *Suppose that training data satisfying (3) is collected is collected with a stabilizing controller satisfying (11) according in Algorithm 1 with $\bar{r} = 2r_{\max}(\mathbf{\Phi}) \ge \max\{1, M\frac{\max\{\sigma_0, \sigma_\eta\}}{1-\rho}\}$ from a system satisfying (6), and that the Nadarya-Watson regressor (5) uses bandwidth γ chosen to minimize the upper bound in (12). Then the overall suboptimality of the certainty equivalent controller (14) is bounded by*

$$c(\widehat{\pi}) - c(\pi_\star) \le 4L_g L_h r_{\max}(\mathbf{\Phi}) \left(\frac{4p^2\sigma_\eta^4}{T}\right)^{\frac{1}{p+4}} \left\| \begin{bmatrix} Q^{1/2}\mathbf{\Phi}_{\mathbf{xn}} \\ R^{1/2}\mathbf{\Phi}_{\mathbf{un}} \end{bmatrix} \right\|_{\mathcal{L}_1} \sqrt{\log(T^2/\delta)} \tag{15}$$

with probability greater than $1-\delta$ for large enough $T \ge 4p^2\sigma_\eta^4 \left(10L_g L_h \|C\mathbf{\Phi}_{\mathbf{xn}}\|_{\mathcal{L}_1} \sqrt{\log(T^2/\delta)}\right)^{p+4}$.

4. Simulated Experiments

To illustrate our results and to probe their limits, we perform experiments in two simulated environments. Our first environment mimics a simplified unmanned aerial vehicle (UAV) scenario, in which observations are recorded from a downward pointing camera. Images are generated using the CARLA simulator (Dosovitskiy et al., 2017), and Figure 1b shows example observations. We fix the elevation and orientation, and define system dynamics using a hovercraft model, where positions along east-west and north-south axes evolve independently according to double integrator dynamics. We construct a training trajectory by applying a static reference tracking controller to follow a periodic reference tracing circles of varying radius. Figure 1c plots the positions from which training observations and measurements are collected.

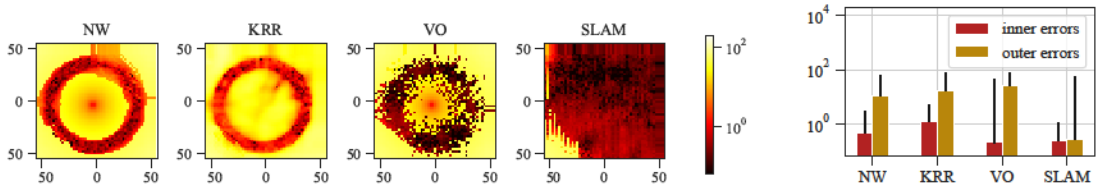
We first train a series of perception maps and evaluate their errors in the UAV setting. We consider four types of regressors: Nadarya-Watson estimators (5) with Euclidean distance on raw pixels (NW), kernel ridge regression with radial basis functions (KRR), a visual odometry method (VO), and a simultaneous localization and mapping method (SLAM). We choose these additional methods to compare the performance with a classical nonparametric approach, a classical computer vision approach, and a non-static state-of-the-art approach.

We evaluate the four learned perception maps on a grid, and the resulting errors are displayed in Figure 2a. For the three static regressors, the error heatmaps are relatively similar, with small errors within the training data coverage, and larger errors outside of it. Though VO has very small errors at many points, its heat map looks much noisier. The large errors come from failures of feature matching within a database of key frames from the training data; in contrast, NW and KRR predictions are closely related to ℓ_2 distance between pixels, which is much smoother. Because SLAM performs mapping online, it can leverage the unlabelled evaluation data to build out good perception away from training samples, and has high errors only due to a tall building obstructing the camera view (visible in Figure 1b) and violating the invertibility assumption. Figure 2b summarizes the evaluations by plotting the median and 99th percentile errors in the *inner region* of training coverage compared with the *outer region*.

Our second environment mimics an autonomous driving example with a dashboard mounted camera (Figure 1b). We use CARLA and the hovercraft dynamics model, with the elevation fixed at ground level. For this driving scenario, observations are determined as a function of vehicle pose, and thus additionally depend on the heading angle, which is determined by the ratio of velocity states.

(*a*) UAV sampling (*b*) Example observations (*c*) Car sampling

Figure 1: Coverage of training data for (a) UAV and (c) autonomous driving settings. In (b), example observations taken from positions indicated by yellow stars.



(*a*) Perception errors over region (*b*) Perception errors

Figure 2: In (a), heatmaps illustrate perception errors. In (b), median and 99th percentile errors within the inner (37-42m radius) and outer (25-55m radius, excluding inner) regions of training data.

Despite this difference, we collect training data in the same manner, resulting in a more sparsely sampled measurement subspace (Figure 1c). We use this example to illustrate potential failure modes that arise when training data is not dense. We consider NW and SLAM perception maps as virtual sensors for a static reference tracking controller. Figure 3 displays the trajectories. For NW, errors cause the system to deviate from the reference, and eventually the system exits the region covered by training data, losing stability due to the perception failure.

5. Conclusion and Future Work

We have presented a sample complexity guarantee for the task of using a complex and unmodelled sensor for waypoint tracking. Our method makes use of noisy measurements from an additional sensor to both learn an inverse perception map and collect training data. We show that evenly sampling the measurement subspace is sufficient for ensuring the success of nonparametric regression, and furthermore that using this learned component in closed-loop has bounded sub-optimality. Unlike related work that focuses on learning unknown dynamics, the task we consider incorporates both a lack of direct state observation and nonlinearity, making it relevant to modern robotic systems.

We hope that future work will continue to probe this problem setting to rigorously explore relevant trade-offs. One direction for future work is to contend with the sampling burden by collecting data in a more goal-oriented manner, perhaps with respect to a target task or the continuity of the observation map. It is also of interest to consider extensions which do not rely on the supervision of a noisy sensor or make clearer use of the structure induced by the dynamics on the observations. Making closer connections with modern computer vision methods like SLAM could lead to insights about unsupervised and semi-supervised learning, particularly when data has known structure.

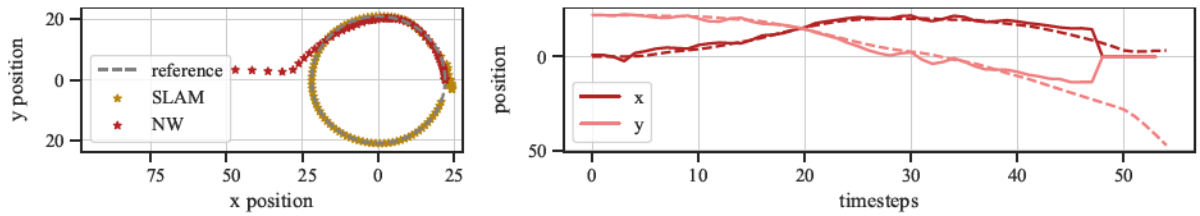


Figure 3: Success and failure of reference tracking. Left, reference and actual trajectories for NW and SLAM. Right, predicted (solid) and actual (dashed) positions for NW.

Acknowledgments

We thank Horia Mania, Pavlo Manovi, Stephen Tu, and Vickie Ye for helpful comments and assistance with experiments. This research is generously supported in part by ONR awards N00014-17-1-2191, N00014-17-1-2401, and N00014-18-1-2833, NSF CPS award 1931853, and the DARPA Assured Autonomy program (FA8750-18-C-0101). SD is supported by an NSF Graduate Research Fellowship under Grant No. DGE 1752814

References

Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26, 2011.

Yasin Abbasi-Yadkori, Nevena Lazic, and Csaba Szepesvári. Model-free linear quadratic control via reduction to expert prediction. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3108–3117, 2019.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. The limits of distribution-free conditional predictive inference. *arXiv preprint arXiv:1903.04684*, 2019.

Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in neural information processing systems*, pages 908–918, 2017.

Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. *arXiv preprint arXiv:1903.08792*, 2019.

Victor Chernozhukov, Kaspar Wuthrich, and Yinchu Zhu. Exact and robust conformal inference methods for predictive machine learning with dependent data. *arXiv preprint arXiv:1802.06300*, 2018.

Felipe Codevilla, Matthias Miiller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. End-to-end driving via conditional imitation learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–9. IEEE, 2018.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *arXiv preprint arXiv:1710.01688*, 2017.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pages 4192–4201, 2018.

Sarah Dean, Nikolai Matni, Benjamin Recht, and Vickie Ye. Robust guarantees for perception-based control. *arXiv preprint arXiv:1907.03680*, 2019.

Ssarah Dean and Benjamin Recht. Certainty-equivalent perception-based control. *arXiv preprint arXiv:2008.12332*, 2020.

Luc P Devroye. The uniform convergence of the nadaraya-watson regression function estimate. *Canadian Journal of Statistics*, 6(2):179–191, 1978.

Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. *arXiv preprint arXiv:1711.03938*, 2017.

Włodzimierz Greblicki. Nonparametric approach to wiener system identification. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 44(6):538–545, 1997.

Bruce E Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748, 2008.

Z Hasiewicz. Identification of a linear system observed through zero-memory non-linearity. *International Journal of Systems Science*, 18(9):1595–1607, 1987.

Seth L Lacy and Dennis S Bernstein. Subspace identification for nonlinear systems that are linear in unmeasured states. In *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No. 01CH37228)*, volume 4, pages 3518–3523. IEEE, 2001.

Forrest Laine, Chiu-Yuan Chiu, and Claire Tomlin. Eyes-closed safety kernels: Safety for autonomous systems under loss of observability. *arXiv preprint arXiv:2005.07144*, 2020.

Alexander Lambert, Amirreza Shaban, Amit Raj, Zhen Liu, and Byron Boots. Deep forward and inverse perceptual models for tracking and prediction. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 675–682. IEEE, 2018.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):71–96, 2014.

Hannelore Liero. Strong uniform consistency of nonparametric regression function estimates. *Probability theory and related fields*, 82(4):587–614, 1989.

Anqi Liu, Guanya Shi, Soon-Jo Chung, Anima Anandkumar, and Yisong Yue. Robust regression for safe exploration in control. *arXiv preprint arXiv:1906.05819*, 2019.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalence is efficient for linear quadratic control. In *Advances in Neural Information Processing Systems*, pages 10154–10164, 2019.

Horia Mania, Michael I. Jordan, and Benjamin Recht. Active learning for nonlinear system identification with guarantees. *arXiv preprint arXiv:2006.10277*, 2020.

Zakaria Mhammedi, Dylan J Foster, Max Simchowitz, Dipendra Misra, Wen Sun, Akshay Krishnamurthy, Alexander Rakhlin, and John Langford. Learning the linear quadratic regulator from nonlinear observations. *arXiv preprint arXiv:2010.03799*, 2020.

Dipendra Misra, Mikael Henaff, Akshay Krishnamurthy, and John Langford. Kinematic state abstraction and provably efficient rich-observation reinforcement learning. In *International conference on machine learning*, pages 6961–6971. PMLR, 2020.

Yashwanth Kumar Nakka, Anqi Liu, Guanya Shi, Anima Anandkumar, Yisong Yue, and Soon-Jo Chung. Chance-constrained trajectory optimization for safe exploration and learning of nonlinear systems. *arXiv preprint arXiv:2005.04374*, 2020.

Houda Salhi and Samira Kamoun. Combined parameter and state estimation algorithms for multivariable nonlinear systems using mimo wiener models. *Journal of Control Science and Engineering*, 2016, 2016.

Maarten Schoukens and Koen Tiels. Identification of nonlinear block-oriented systems starting from linear approximations: A survey. *CoRR, abs/1607.01217*, 2016.

Max Simchowitz and Dylan J Foster. Naive exploration is optimal for online lqr. *arXiv preprint arXiv:2001.09576*, 2020.

Max Simchowitz, Ross Boczar, and Benjamin Recht. Learning linear dynamical systems with semi-parametric least squares. *arXiv preprint arXiv:1902.00768*, 2019.

Max Simchowitz, Karan Singh, and Elad Hazan. Improper learning for non-stochastic control. *arXiv preprint arXiv:2001.09254*, 2020.

Sarah Tang, Valentin Wüest, and Vijay Kumar. Aggressive flight with suspended payloads using vision-based control. *IEEE Robotics and Automation Letters*, 3(2):1152–1159, 2018.

Andrew J Taylor, Victor D Dorobantu, Hoang M Le, Yisong Yue, and Aaron D Ames. Episodic learning with control lyapunov functions for uncertain robotic systems. *arXiv preprint arXiv:1903.01577*, 2019.

Koen Tiels and Johan Schoukens. Wiener system identification with generalized orthonormal basis functions. *Automatica*, 50(12):3147–3154, 2014.

Stephen Tu and Benjamin Recht. The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint. *arXiv preprint arXiv:1812.03565*, 2018.

Kim P Wabersich and Melanie N Zeilinger. Linear model predictive safety certification for learning-based control. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 7130–7135. IEEE, 2018.

Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. A system level approach to controller synthesis. *arXiv preprint arXiv:1610.04815*, 2016.

Torbjörn Wigren. Convergence analysis of recursive identification algorithms based on the nonlinear wiener model. *IEEE Transactions on Automatic Control*, 39(11):2191–2206, 1994.