Robust Guarantees for Perception-Based Control

Sarah Dean[1], Nikolai Matni[2], Benjamin Recht[1], Vickie Ye[1]

[1]University of California, Berkeley EECS, [1]University of Pennsylvania ESE

Problem Setting

We consider the linear dynamical system

$$x_{k+1} = Ax_k + Bu_k + Hw_k, \quad (1)$$

with associated high-dimensional observations (e.g. images)

$$z_k = q(x_k). \quad (2)$$

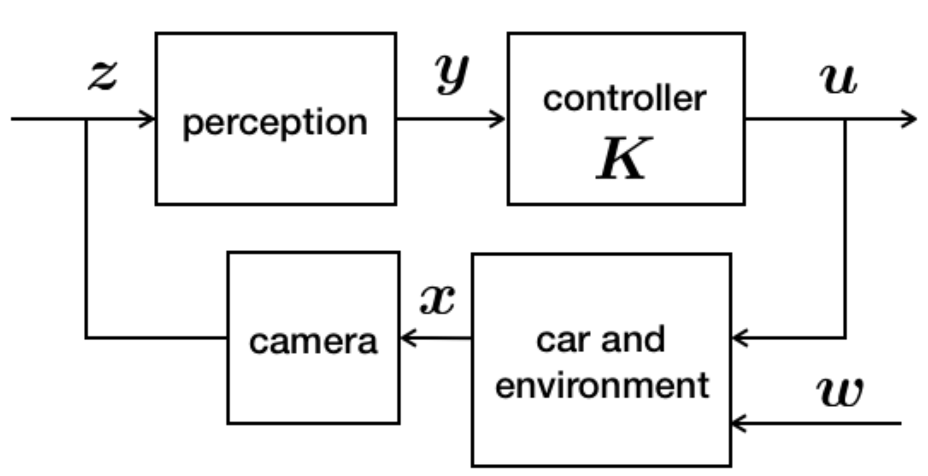
Our goal is then to solve the optimal control problem

$$\begin{aligned} \text{minimize} \quad & c(\boldsymbol{x}, \boldsymbol{u}) \\ \text{subject to} \quad & \text{dynamics (1)}, \\ & u_k = \gamma(z_{0:k}) = \pi(y_{0:k}). \end{aligned} \quad (3)$$

We suppose we have a perception map p that acts as a *virtual sensor* to yield outputs

$$y_k = p(z_k) = Cx_k + e_k. \quad (4)$$

Then the optimal control problem can be reformulated into linear output feedback control.



Perception Error Characterization

If the error function $e(x) = p(q(x)) - Cx$ is locally S-slope bounded, then within local regions of each training data point $\mathcal{X}_\gamma(x_d, z_d)$ defined as

$$\{x \mid \|p(z_d) - Cx_d\| + S\|x - x_d\| \leq \gamma\}, \quad (5)$$

it is possible to guarantee bounded errors.
We therefore define the *safe set* as the union:

$$\mathcal{X}_\gamma = \bigcup_{(x_d, z_d) \in \mathcal{S}_0} \mathcal{X}_\gamma(x_d, z_d)\,. \quad (6)$$

Within $\mathcal{X}_\gamma$ the perception error is bounded by γ.

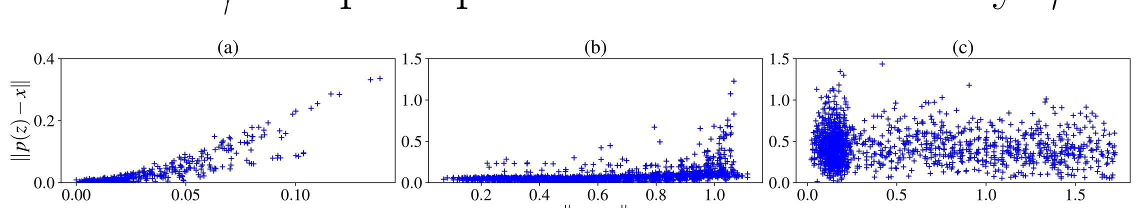


Figure: Plotting the perception errors vs. distance from nearest training point shows their slope-boundedness. Left: synthetic example, middle: visual odometry for driving example, right: CNN for driving example.

Robust Control Analysis

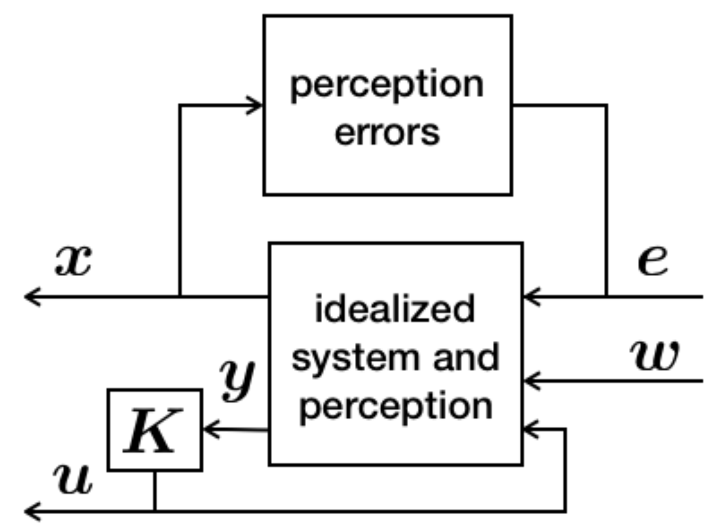


Figure: The robust control view of the closed-loop system

For any linear feedback control law $\boldsymbol{K}$, we equivalently describe the closed-loop system responses to process and measurement noise:

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi_{xw}} & \boldsymbol{\Phi_{xe}} \\ \boldsymbol{\Phi_{uw}} & \boldsymbol{\Phi_{ue}} \end{bmatrix} \begin{bmatrix} H\boldsymbol{w} \\ \boldsymbol{e} \end{bmatrix}. \quad (7)$$

Since we consider a perception map as a virtual sensor, perception errors are the measurement noise. Therefore, $\boldsymbol{\Phi_{xe}}$ describes the effect of the perception error on the state.

Main Result

Suppose the perception errors are locally S-slope bounded within a radius r and the maximum training error is R_0. Then for control law $\boldsymbol{u} = \boldsymbol{K}p(\boldsymbol{z})$, define $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_d\|$ to be the planned deviation from the training data in the absence of sensor error.
Then as long as the system is not too sensitive to measurement error,

$$\|\boldsymbol{\Phi_{xe}}\| \leq \frac{1 - \frac{1}{r}\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_d\|}{S + \frac{R_0}{r}}, \quad (8)$$

we can guarantee for all closed loop trajectories:

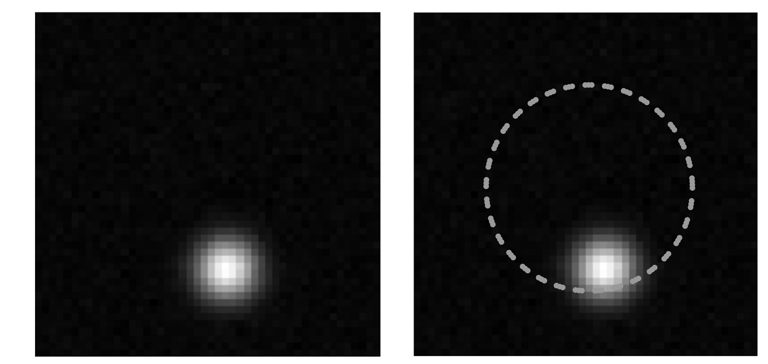
1. The perception errors remain bounded

$$\|p(\boldsymbol{z}) - C\boldsymbol{x}\| \leq \frac{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_d\| + R_0}{1 - S\|\boldsymbol{\Phi_{xe}}\|} := \gamma, \quad (9)$$

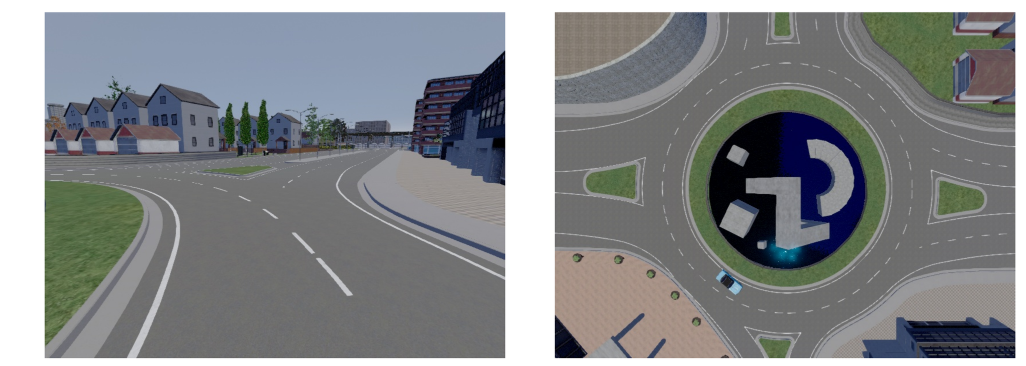
2. The trajectory lies within the safe set $\mathcal{X}_\gamma$.

This result establishes conditions for which the safe set $\mathcal{X}_\gamma$ is made *invariant* by the control law.
In designing the controller, there is a trade-off when ensuring that $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_d\|$ and $\|\boldsymbol{\Phi_{xe}}\|$ are both small, but more training data makes this easier.

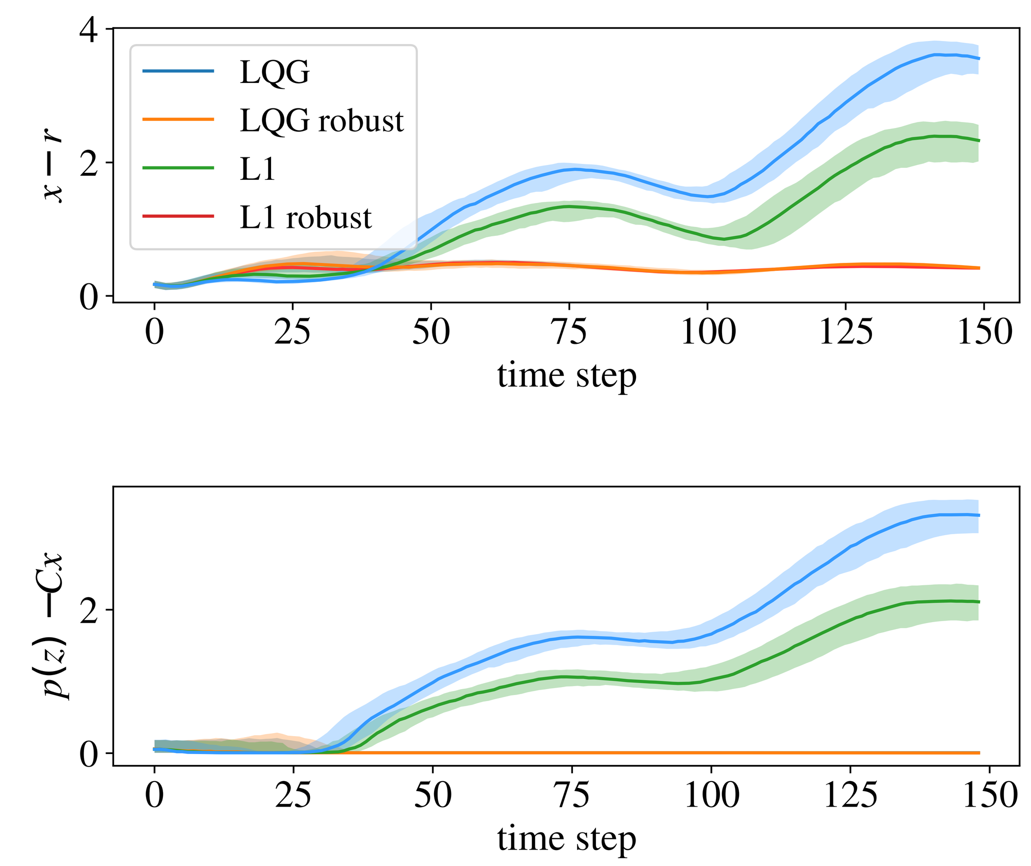
Experiments: Waypoint Following



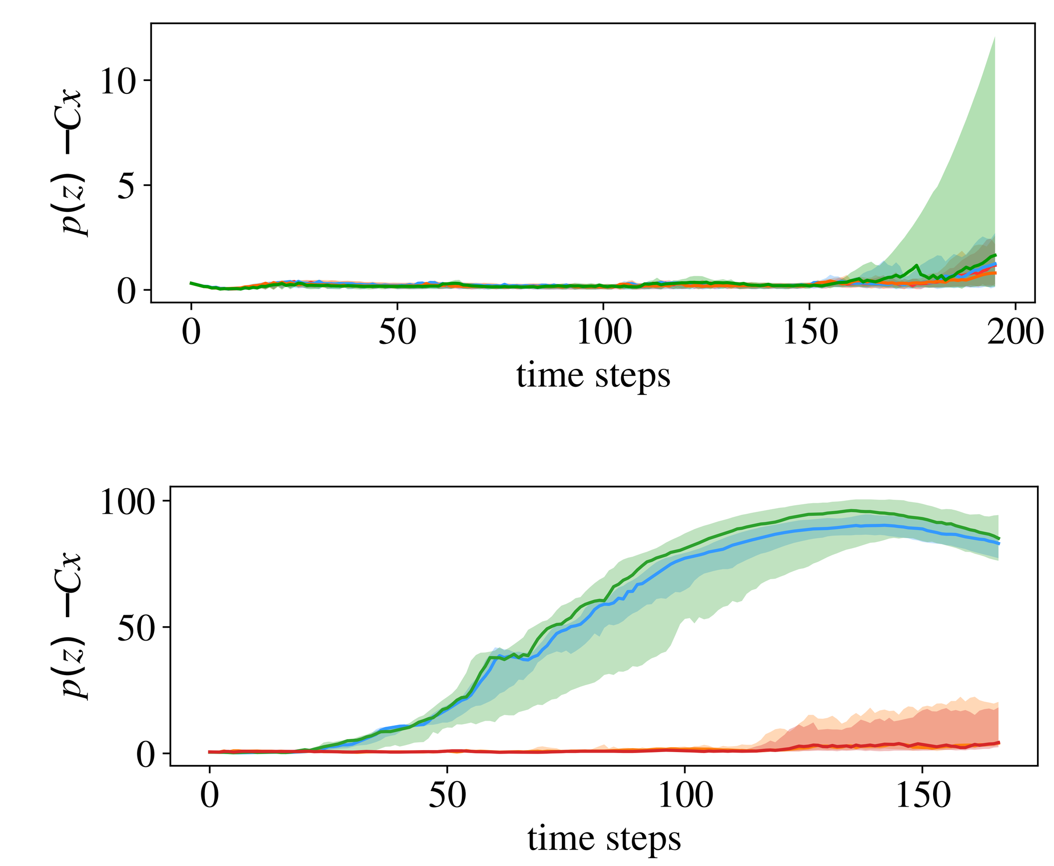
(a) Images z_t and reference to track for synthetic example.



(b) Images z_t and reference to track for vehicle task.



(c) Trajectories and perception errors for synthetic example.



(d) Perception errors for visual odometry (top) and CNN (bottom).

Figure: Experimental setting and results: tracking and perception errors for (c) 200 rollouts of the synthetic and perception errors (d) 100 rollouts of the CARLA examples with both visual odometry and CNN perception. Lines indicate median values while shaded regions indicate upper and lower quartiles.

We present numerical experiments for two experimental settings and three different perception schemes.

- **Synthetic:** White dot on black background with Gaussian noise, and a linear map from image to position.
- **Driving with visual odometry:** CARLA graphics simulator generates observations as a function of position and heading angle, and a visual odometry method maps image to position. The visual odometry method is "trained" with 200 datapoints using SLAM to construct a database of references images with known poses.
- **Driving with CNN:** Observations are generated as above, and a one-layer CNN maps image to position. The CNN is trained with 30,000 datapoints.

In all cases, we consider two dimensional double integrator dynamics. Controllers are synthesized according to the LQG and $\mathcal{L}_1$ objective functions. We compare nominal controllers to robust controllers, which are synthesized with a constraint as in (8).
We demonstrate a scenario in which the nominal controller fails to track the reference, while the robust controller is successful. Furthermore, different perception strategies matter: the failures of the visual odometry method are less frequent that those of the CNN method, which relates to the different slope characteristics of the errors maps (Figure 1).