Robust-Adaptive Control of Linear Systems: beyond Quadratic Costs

Edouard Leurent, Denis Efimov,
Odalric-Ambrym Maillard

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

- Adaptive: estimate the dynamics along the way
- Robust: avoid failures, prepare for the worst

Related work: Robust DP (finite states), quadratic costs (stabilization only)

Setting

- Linear dynamics, with unknown parameters θ

$$\dot{x}(t) = A(\theta)x(t) + Bu(t) + D\omega(t)$$

- Arbitrary bounded reward R

Algorithm

1. Model Estimation: build a confidence region $\mathcal{C}_{N,\delta}$

$$\mathbb{P}(\theta \in \mathcal{C}_{N,\delta}) \geq 1-\delta$$

2. Robust Control: solve a minimax objective

$$\sup_{u} \underbrace{\inf_{\substack{\theta\in\mathcal{C}_{N,\delta}\\ \omega\in[\underline{\omega},\overline{\omega}]}} \left(\sum_{n=N+1}^{\infty}\gamma^n R(x(t_n))\,\middle|\, u,\theta,\omega\right)}_{V^r(u)}$$

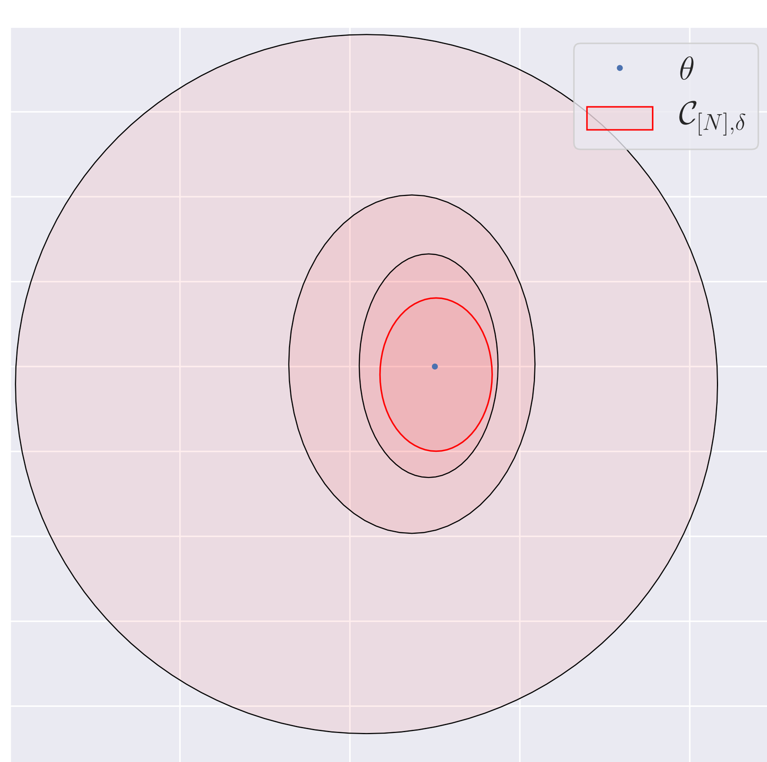
1. Model Estimation

Structure assumption: $A(\theta) = A + \sum_{i=1}^{d}\theta_i\phi_i$

Noise model: sub-Gaussian observations, bounded disturbance $\omega(t)$

Theorem (Confidence Ellipsoid, adaptation of Abbasi-Yadkori et al., 2001)
With probability at least $1-\delta$,

$$\|\theta - \theta_{N,\lambda}\|_{G_{N,\lambda}} \leq \beta_N(\delta).$$



The confidence ellipsoid shrinks as the number of samples N increases.

2.1. Interval Prediction

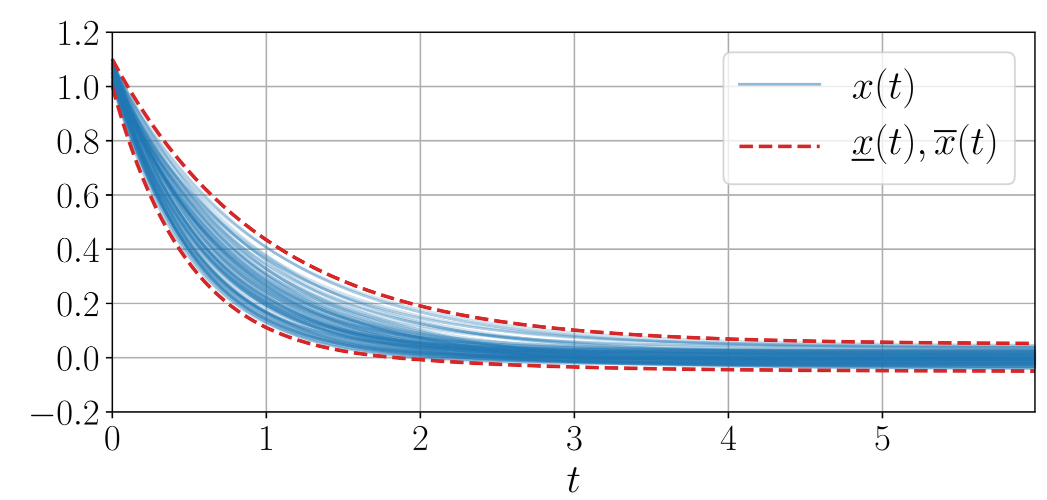
Having observed N samples, given $\theta \in \mathcal{C}_{N,\delta}$ we want to bound the trajectory

$$\underline{x}(t) \leq x(t) \leq \overline{x}(t), \qquad \forall t \geq t_N.$$

Proposition (Predictor of Leurent et al., 2019) For $A(\theta_N)$ Metzler,

$$\dot{\underline{x}}(t) = A(\theta_N)\underline{x}(t) - \Delta A_+\underline{x}^-(t) - \Delta A_-\overline{x}^+(t) + Bu(t) + D^+\underline{\omega}(t) - D^-\overline{\omega}(t),$$
$$\dot{\overline{x}}(t) = A(\theta_N)\overline{x}(t) + \Delta A_+\overline{x}^+(t) + \Delta A_-\underline{x}^-(t) + Bu(t) + D^+\overline{\omega}(t) - D^-\underline{\omega}(t),$$

ensures the inclusion property.



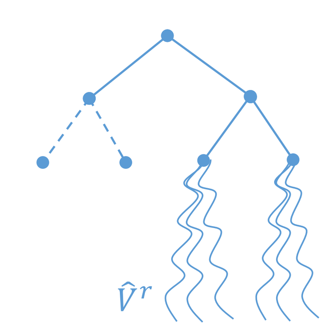
Predicted intervals for a simple example.

2.2. Robust Control

We approximate V^r by a tractable pessimistic surrogate $\hat{V}^r$ using the predicted intervals:

$$\hat{V}^r(u) = \sum_{n=N+1}^{\infty}\gamma^n \min_{\underline{x}(t_n)\leq x\leq\overline{x}(t_n)} R(x).$$

$\hat{V}^r$ is easily evaluable and can be optimized by Monte-Carlo Tree Search.



Theorem (Lower bound)

$$\underbrace{\hat{V}^r(u)}_{\text{surrogate value}} \leq \underbrace{V^r(u)}_{\text{robust value}} \leq \underbrace{V(u)}_{\text{true performance}}$$

If we find some good enough control sequence u that attains a desirable (safe) level of predicted performance $\hat{V}^r$ while planning, it is **guaranteed to perform at least as well** when executed on the true system.
But the gap can lead to suboptimal behavior.

Theorem (Bounded suboptimality) Under two conditions:

- Lipschitz reward R,
- A stability condition for $A(\theta_N)$,

we can bound the suboptimality with probability at least $1-\delta$ by

$$\underbrace{V(u^\star) - V(u_K)}_{\text{suboptimality}} \leq \underbrace{\Delta_\omega}_{\text{robustness to disturbances}} - \underbrace{\mathcal{O}\left(\frac{\beta_{N,\delta}^2}{\lambda_{\min}(G_{N,\lambda})}\right)}_{\text{estimation error}} - \underbrace{\mathcal{O}\left(K^{-\frac{\log(\kappa)}{\log 1/\gamma}}\right)}_{\text{planning error}}$$

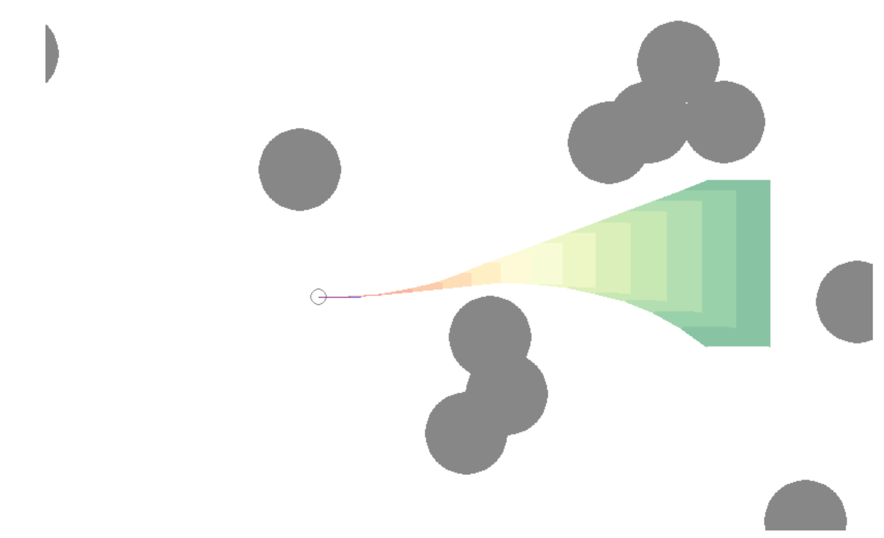
Corollary (Asymptotic near-optimality)
Under an additional persistent excitation assumption,

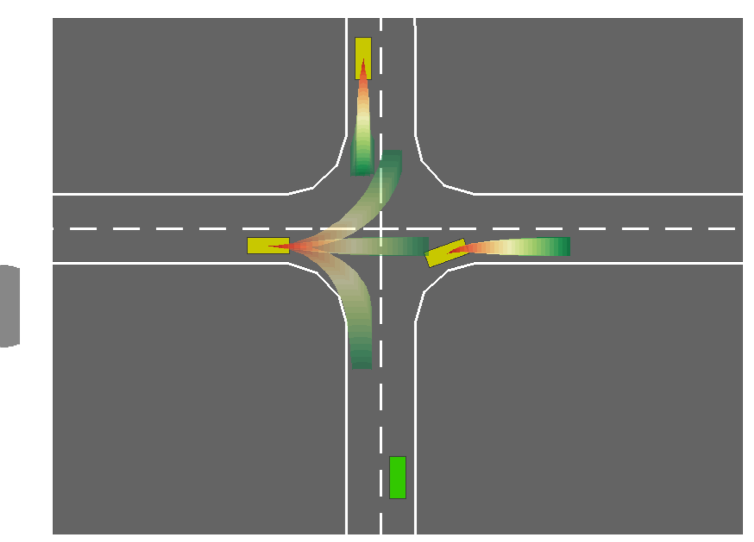
- the stability condition can be relaxed to apply to the true system $A(\theta)$
- the model estimation error takes the more explicit form

$$\frac{\beta_{N,\delta}^2}{\lambda_{\min}(G_{N,\lambda})} = \mathcal{O}\left(\frac{\log(N^{d/2}/\delta)}{N}\right)$$

which ensures asymptotic near-optimality as $N\to\infty$, $K\to\infty$.

Experiments



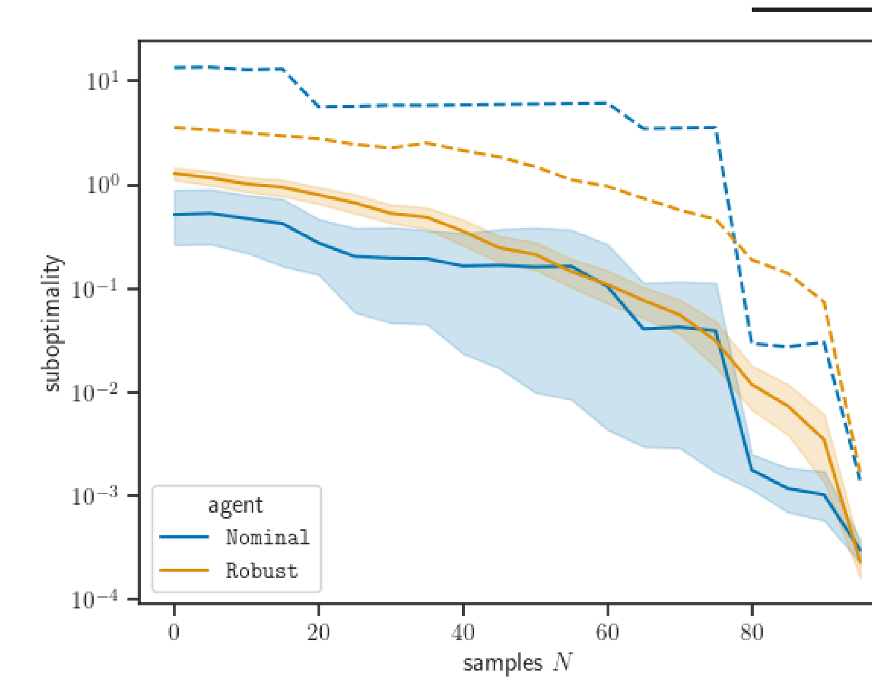


Obstacle avoidance

	failures	min	avg $\pm$ std
Oracle	0%	11.6	14.2 ± 1.3
Nominal	4%	2.8	$\mathbf{13.8} \pm 2.0$
Robust	**0%**	**10.4**	13.0 ± 1.5
DQN[a]	6%	1.7	12.3 ± 2.5

Unsignalized intersection

	failures	min	avg $\pm$ std
Oracle	0%	6.9	7.4 ± 0.5
Nominal 1	4%	5.2	$\mathbf{7.3} \pm 1.5$
Nominal 2	33%	3.5	6.4 ± 0.3
Robust	**0%**	**6.8**	7.1 ± 0.3
DQN[a]	3%	5.4	6.3 ± 0.6



Mean (solid), 95% CI for the mean (shaded), and maximum (dashed) suboptimality with respect to the number of samples N.

The robust agent gets **more efficient** as it is **more confident**, while **ensuring safety** at all times.