Proceedings of Machine Learning Research vol 120:1–9, 2020 2nd Annual Conference on Learning for Dynamics and Control

Robust Learning-Based Control via Bootstrapped Multiplicative Noise

Benjamin Gravell BENJAMIN.GRAVELL@UTDALLAS.EDU
University of Texas at Dallas

Tyler Summers TYLER.SUMMERS@UTDALLAS.EDU
University of Texas at Dallas

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

Abstract

Despite decades of research and recent progress in adaptive control and reinforcement learning, there remains a fundamental lack of understanding in designing controllers that provide robustness to inherent *non-asymptotic* uncertainties arising from models estimated with finite, noisy data. We propose a robust adaptive control algorithm that explicitly incorporates such non-asymptotic uncertainties into the control design. The algorithm has three components: (1) a least-squares nominal model estimator; (2) a bootstrap resampling method that quantifies non-asymptotic variance of the nominal model estimate; and (3) a non-conventional robust control design method using an optimal linear quadratic regulator (LQR) with multiplicative noise. A key advantage of the proposed approach is that the system identification and robust control design procedures both use stochastic uncertainty representations, so that the actual inherent statistical estimation uncertainty directly aligns with the uncertainty the robust controller is being designed against. We show through numerical experiments that the proposed robust adaptive controller can significantly outperform the certainty equivalent controller on both expected regret and measures of regret risk.

Keywords: Robust adaptive control, bootstrap, multiplicative noise, regret

1. Introduction

Recent high profile successes and the resulting hype in machine learning and reinforcement learning are generating renewed interest in adaptive control and system identification, which have their own decades-long histories (Åström and Wittenmark (2013); Ljung (1998)). Classical work on adaptive control and system identification largely focused on asymptotics, including stability, consistency, asymptotic variance, etc. Emerging research at the intersection of learning and control shifts focus to non-asymptotic statistical analyses, including regret and sample efficiency in various adaptive control and learning algorithms (Abbasi-Yadkori and Szepesvári (2011); Dean et al. (2018, 2019)).

In both classical and emerging work in learning and control, robustness has been a key issue. The classical focus on asymptotics led to a strong emphasis on certainty-equivalent adaptive control, where inevitable uncertainty in model estimates is ignored for control design and unsurprisingly can lead to serious lack of robustness. It remains poorly understood how to best interface both asymptotic and non-asymptotic uncertainty descriptions of model estimates with robust control design methods. One of the main difficulties is a mismatch in uncertainty descriptions: system identification almost universally uses stochastic data models[1], whereas robust control traditionally uses

1. A notable exception is set membership identification (Milanese and Vicino (1991)), which maintains a *set* of models that could have produced the data within some error bound and interfaces somewhat more naturally with certain traditional robust control design methods.

© 2020 B. Gravell & T. Summers.

set-based descriptions and worst-case design. This can lead to unnecessary conservatism when the assumed model uncertainty sets are poorly aligned with inherent statistical model uncertainty.

The contributions of the present work are as follows:

1. We propose a robust adaptive control algorithm where the model uncertainty description and robust control design method both use stochastic uncertainty representations.

2. We show via numerical experiments that the proposed robust adaptive controller can significantly outperform the certainty equivalent controller on both expected regret and measures of regret risk.

The algorithm has three components: (1) a least-squares nominal model estimator; (2) a bootstrap resampling method that quantifies non-asymptotic variance of the nominal model estimate; and (3) a non-conventional robust control design method using an optimal linear quadratic regulator (LQR) with multiplicative noise. This approach provides a natural interface between two widely used and highly effective methods from statistics and optimal control theory (namely, bootstrap sample variance and LQR). It is known that certainty equivalent adaptive control can achieve asymptotic optimality and statistical efficiency with an *order* optimal rate (Chen and Guo (2012); Kumar and Varaiya (2015); Mania et al. (2019)). However, neither of these imply anything about non-asymptotic optimality or robustness.

2. Problem Formulation

We consider adaptive control of the discrete-time linear dynamical system

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

where $x_t \in \mathbf{R}^n$ is the system state, $u_t \in \mathbf{R}^m$ is the control input, and w_t is i.i.d. process noise with zero mean and covariance matrix W. The system matrices (A, B) are assumed unknown, so an adaptive controller is to be designed based only on state-input trajectory data $x_{0:t} := [x_0, ..., x_t]$, $u_{0:t-1} = [u_0, ..., u_{t-1}]$. We consider the linear quadratic adaptive optimal control problem

$$\min_{\pi} \mathbb{E}\left[\sum_{t=0}^{T}(x_t^\intercal Q x_t + u_t^\intercal R u_t - J^*)\right] \tag{2}$$

where $Q \succeq 0$ and $R \succeq 0$ are cost matrices, and the optimization is over (measurable) history dependent feedback policies $\pi = \{\pi_t\}_{t=0}^{T-1}$ with $u_t = \pi_t(x_{0:t}, u_{0:t-1})$. The constant J^* in the stage costs represents the optimal infinite-horizon average steady state cost when the system matrices (A, B) are known, which results from a static linear state feedback, $u_t = Kx_t$, whose gain matrix K can be computed via several known methods, including value iteration, policy iteration, and semidefinite programming. This constant gives the stage cost an interpretation as *regret*.

The finite horizon objective in (2) emphasizes the *non-asymptotic* performance of the adaptive controller. This stands in contrast to a majority of classical work on adaptive control, which tends to focus on asymptotic performance and stability. Note that regret is a random variable that ideally should be small, and there are various ways to measure its size, including expected regret, regret variance, and measures of regret risk, such as value at risk or conditional value at risk.

It has been long known that this problem can be solved *in principle* by redefining the state as the (infinite-dimensional) joint conditional distribution over the original state and unknown model

parameters and applying dynamic programming (Bellman (1961)). However, this approach is intractable even for the most trivial instances. Since computing the optimal policy exactly appears to be intractable, we instead aim to design a computationally implementable controller with good performance and robustness properties, i.e., one that achieves both small expected regret and small regret risk. In particular, our algorithm accounts for uncertainty in various *directions* by modeling them as multiplicative noises, in contrast to the isotropic robustness afforded by the system-level synthesis in (Dean et al. (2018, 2019)). We compare with a certainty equivalent adaptive controller, where uncertainty is ignored and a controller is designed as if the point model estimates were exact.

3. Preliminaries: Multiplicative Noise LQR

Here we will represent parameter uncertainty in model estimates stochastically using covariance matrices estimated from bootstrap resampling of finite data records. This representation interfaces quite naturally with a variant of the linear quadratic regulator that incorporates multiplicative noise, which has a long history in control theory but is far less widely known than its additive noise counterpart (Wonham (1967); Gravell et al. (2019)). Consider the linear quadratic regulator problem with dynamics perturbed by multiplicative noise

$$\begin{aligned} \underset{\pi\in\Pi}{\text{minimize}} \quad & \mathbb{E}_{x_0,\{\bar{A}_t\},\{\bar{B}_t\}} \sum_{t=0}^{\infty} (x_t^\intercal Q x_t + u_t^\intercal R u_t), \\ \text{subject to} \quad & x_{t+1} = (A+\bar{A}_t)x_t + (B+\bar{B}_t)u_t, \end{aligned} \tag{3}$$

where $\bar{A}_t$ and $\bar{B}_t$ are i.i.d. zero-mean random matrices with a joint covariance structure over their entries governed by the covariance matrices $\Sigma_A := \mathbb{E}[\mathbf{vec}(\bar{A})\mathbf{vec}(\bar{A})^\intercal] \in \mathbf{R}^{n^2\times n^2}$ and $\Sigma_B := \mathbb{E}[\mathbf{vec}(\bar{B})\mathbf{vec}(\bar{B})^\intercal] \in \mathbf{R}^{nm\times nm}$, which quantify uncertainty in the nominal system matrices (A, B) and can be estimated from data using bootstrap methods.

Just as in additive noise LQR problems, dynamic programming can be used to show that the optimal policy is linear state feedback $u_t = Kx_t$. Given the problem data $(A, B, Q, R, \Sigma_A, \Sigma_B)$, the optimal quadratic cost matrix is given by the solution of the *generalized* Riccati equation

$$P = Q + A^\intercal PA + \sum_{i=1}^{n^2} \alpha_i A_i^\intercal PA_i - A^\intercal PB(R + B^\intercal PB + \sum_{j=1}^{nm} \beta_j B_j^\intercal PB_j)^{-1} B^\intercal PA, \tag{4}$$

and the associated optimal gain matrix is $K = -\big(R + B^\intercal PB + \sum_{j=1}^{nm} \beta_j B_j^\intercal PB_j\big)^{-1} B^\intercal PA$, where $\{\alpha_i, A_i\}_{i=1}^{n^2}$ and $\{\beta_j, B_j\}_{j=1}^{nm}$ are the eigenvalues and reshaped eigenvectors of Σ_A and Σ_B, respectively. The solutions are denoted $(P, K) = \text{GDARE}(A, B, Q, R, \Sigma_A, \Sigma_B)$. The optimal cost and policy can be computed via value iteration, policy iteration, or semidefinite programming (El Ghaoui (1995); Gravell et al. (2019)). Note that like traditional robust control but unlike additive noise LQR, the optimal cost matrix and control gain depend explicitly on the model uncertainty.

4. Robust Adaptive Control via Bootstrapping and Multiplicative Noise

Our robust adaptive control algorithm is summarized in Algorithm 1. The algorithm has three main components: (1) a least-squares nominal model estimator; (2) a bootstrap resampling method that quantifies non-asymptotic variance of the nominal model estimate; and (3) a non-conventional robust control design method using an optimal LQR with multiplicative noise.

Algorithm 1 Robust Adaptive Control

Input: exploration time t_{explore}, input excitation covariance U, number of bootstrap resamples N_b, model uncertainty scaling parameter γ, cost matrices $Q \succeq 0, R \succeq 0$

$x_0 \sim \mathcal{N}(0, X_0)$
for $t = 0, \ldots, T-1$ **do**
  if $t \leq t_{\text{explore}}$ **then**
    $u_t \sim \mathcal{N}(0, U)$
    $x_{t+1} = Ax_t + Bu_t + w_t$
  else
    $(\hat{A}_t, \hat{B}_t) = \texttt{OrdinaryLeastSquares}(x_{0:t}, u_{0:t-1})$
    $(\hat{\Sigma}_{A_t}, \hat{\Sigma}_{B_t}) = \texttt{BootstrapModelVariance}(x_{0:t}, u_{0:t-1})$
    $\hat{K}_t = \texttt{MultiplicativeNoiseLQR}(\hat{A}_t, \hat{B}_t, Q, R, \gamma\hat{\Sigma}_{A_t}, \gamma\hat{\Sigma}_{B_t})$
    $e_t \sim \mathcal{N}(0, \|(\hat{\Sigma}_{A_t}, \hat{\Sigma}_{B_t})\| U)$
    $u_t = \hat{K}_t x_t + e_t$
    $x_{t+1} = Ax_t + Bu_t + w_t$
  end if
end for

4.1. Least Squares Estimation for the Nominal Model

The first component of the algorithm is a standard least-squares estimator for the unknown system matrices from state-input trajectory data. In particular, at time t from data $(x_{0:t}, u_{0:t-1})$, we form the estimate

$$(\hat{A}_t, \hat{B}_t) = \text{argmin}_{(A,B)} \left\{ \sum_{\tau=0}^{t-1} \|x_{t+1} - (Ax_t + Bu_t)\|_2^2 \right\}. \tag{5}$$

More explicitly, defining the data matrices $X_t^\intercal = [x_1 \;\; x_2 \;\; \cdots \;\; x_t]$, $Z_t^\intercal = \begin{bmatrix} x_0 & x_1 & \cdots & x_{t-1} \\ u_0 & u_1 & \cdots & u_{t-1} \end{bmatrix}$ then the least squares estimate can be written as

$$[\hat{A}_t, \; \hat{B}_t] = X_t^\intercal Z_t (Z_t^\intercal Z_t)^{-1}. \tag{6}$$

A non-degenerate model estimate is obtained only when $Z_t^\intercal Z_t$ is invertible, so learning is divided into a pure exploration phase until a user-specified time $t_{\text{explore}} > n + m$ and subsequently an exploration-exploitation phase, where the estimated model is used to design a control policy. The exploration component of the input signal is iid Gaussian noise with user-specified covariance matrix U; it is well known that for any $U \succ 0$ the least-squares estimator is consistent under our modeling assumptions. The exploration noise is designed to fade out with the bootstrap-estimated model uncertainty, which yields asymptotic optimality. Under mild assumptions, the least-squares estimator can be implemented recursively to significantly simplify the repeated computation in (6) as data arrives (e.g. Simon (2006)).

4.2. Bootstrap Resampling to Quantify Non-Asymptotic Model Uncertainty

There are inevitably errors in the least-squares estimate obtained from any finite data record, due to the process noise affecting the system dynamics. Due to dependence in the time series data,

unfortunately it is not straightforward to analytically characterize non-asymptotic uncertainty in the least-squares estimate using standard statistical techniques. Therefore, to quantify non-asymptotic uncertainty in the model estimate, we propose a time series bootstrap resampling procedure. There are three broad bootstrap methods for time series (Härdle et al. (2003)): (1) Parametric bootstrap; (2) Semi-parametric bootstrap with resampled residuals; (3) Non-parametric bootstrap with block resampling. In parametric and semi-parametric methods, bootstrap data are simulated from the nominal model with the process noise sampled iid with replacement either from an assumed distribution or from residuals calculated with the nominal model. Dependence in the data is preserved by construction. In non-parametric methods, overlapping time blocks of consecutive data are sampled from the original data to preserve dependence. Generally speaking, the semi- and nonparametric bootstraps are less sensitive to assumptions about the model and the noise distribution, while the semi- and pure parametric bootstraps have better small sample performance when the model is correctly specified. For definiteness, a semi-parametric bootstrap with resampled residuals for the least-squares estimator discussed above is summarized in Algorithm 2.

Algorithm 2 Semi-Parametric Bootstrap

Input: trajectory data $(x_{0:t}, u_{0:t-1})$, nominal model estimate $(\hat{A}_t, \hat{B}_t)$, residuals $\hat{w}_\tau = x_{\tau+1} - (\hat{A}_t x_\tau + \hat{B}_t u_\tau)$, $\tau = 0, ..., t-1$, number of bootstrap resamples N_b

1: $\bar{x}_0 = x_0$
2: $\bar{u}_{0:t-1} = u_{0:t-1}$
3: **for** $k = 1, \ldots, N_b$ **do**
4: Generate data $\bar{x}_{\tau+1} = \hat{A}_t \bar{x}_\tau + \hat{B}_t \bar{u}_\tau + \tilde{w}_\tau$, $\tau = 0, ..., t-1$, where $\{\tilde{w}_\tau\}_{\tau=0}^{t-1}$ is an iid resample with replacement from residuals $\{\hat{w}_\tau\}_{\tau=0}^{t-1}$
5: $(\hat{A}_t^k, \hat{B}_t^k) = \text{argmin}_{(A,B)} \left\{ \sum_{\tau=0}^{t-1} \|\bar{x}_{\tau+1} - (A\bar{x}_\tau + B\bar{u}_\tau)\|_2^2 \right\}$
6: **end for**

Output: Bootstrap sample covariance $\hat{\Sigma}_{A_t} = \frac{1}{N_b - 1} \sum_{k=1}^{N_b} \text{vec}(\hat{A}_t^k - \hat{A}_t)\text{vec}(\hat{A}_t^k - \hat{A}_t)^\intercal$
Bootstrap sample covariance $\hat{\Sigma}_{B_t} = \frac{1}{N_b - 1} \sum_{k=1}^{N_b} \text{vec}(\hat{B}_t^k - \hat{B}_t)\text{vec}(\hat{B}_t^k - \hat{B}_t)^\intercal$

4.3. Multiplicative Noise LQR

The least squares estimator and bootstrap provide both a nominal estimate of the system model and an estimate of the covariance of the nominal model error. These quantities provide precisely the input data needed to compute an optimal policy for the LQR problem with multiplicative noise from the generalized Riccati equation (4). This policy is known to provide robustness to uncertainties in the parameters of the nominal model (Bernstein and Greeley (1986)). Furthermore, the uncertainty in the nominal model estimate used in this control design method is richly structured and derived directly from the finite available data.

We introduce a parameter γ which provides a fixed scaling of the model uncertainty. Note that $\gamma = 0$ corresponds to certainty equivalent adaptive control, and as γ increases, more weight is placed on uncertainty in the nominal model. Existence of a solution to the generalized Riccati equation (4) depends not just on stabilizability of the nominal system (A, B), but also on the *mean-square stabilizability* of the multiplicative noise system. When the multiplicative noise variances are too large, it may be impossible to stabilize the system in the mean-square sense. In this case, we scale down the model variances at each time step if necessary to compute a mean-square stabilizing

control gain via bisection; see Algorithm 3. In particular, we verify the system with specified γ is mean-square stabilizable by checking whether the generalized Riccati equation in (4) admits a positive semidefinite solution; if not, we find the upper limit $\gamma_{\max} = c_\gamma \gamma$ via bisection (e.g. Burden et al. (1978)) on a scaling $c_\gamma \in [0, 1]$.

Algorithm 3 Multiplicative Noise LQR

Input: Nominal model matrices A, B, cost matrices Q, R, multiplicative noise scaling γ and covariances $\hat{\Sigma}_A$, $\hat{\Sigma}_B$, bisection tolerance $\epsilon > 0$

1: Find largest $c_\gamma \in [0, 1]$ via bisection such that there exists a feasible solution to (4)
$(P, K) = \text{GDARE}(A, B, Q, R, c_\gamma \gamma \Sigma_A, c_\gamma \gamma \Sigma_B)$

Output: Cost matrix P, gain matrix K

5. Numerical Experiments

For brevity, we abbreviate "certainty-equivalent" as "CE" and "robustness via multiplicative noise" as "RMN". To evaluate the performance of the proposed RMN algorithm relative to CE control, we performed Monte Carlo sampling to estimate the distribution of several key quantities: instantaneous regret, model error, and multiplicative noise variances.

The instantaneous regret is heavy-tailed in the sense that the effect of outliers with non-negligible probability is significant: some exceptionally poor sequences of model estimates induce extremely high costs relative to the median. For this reason, to facilitate the most direct comparison of CE and the proposed RMN approach, we train the models using both control schemes on identical offline training data. This way the effect of outlier model estimates is applied uniformly to both algorithms, since at each time the algorithms are faced with exactly the same model estimates. In an online adaptive control setting, where the training data are the actual state trajectories experienced under adaptive control, a direct comparison of the approaches is more difficult. While we observe qualitatively similar benefits of the proposed approach, our future work will study this setting with a greater number of Monte Carlo samples to reduce the effect of outliers.

The training data are generated by initializing the state at the origin, applying random controls distributed according to a standard Gaussian distribution (zero-mean, identity-covariance), and simulating the evolution of the state with the additive process noise specified by the problem data W. The resulting training data are a set of state trajectories $x_{0:t}^{(k)}$ and input trajectories $u_{0:t-1}^{(k)}$. Model and uncertainty estimates are generated at time t according to Algorithm 1 using training data only up to time t. The optimal cost is empirically calculated by averaging over all Monte Carlo samples the cost incurred by trajectories under optimal control $u_t = K^* x_t$, $K^* = \text{DARE}(A, B, Q, R)$ for each additive noise realization, i.e.

$$c_t^* = \frac{1}{N_s} \sum_{k=0}^{N_s} x_t^{*,(k)\intercal} Q x_t^{*,(k)} + u_t^{*,(k)\intercal} R u_t^{*,(k)}$$

$$\text{where } x_{t+1}^{*,(k)} = A x_t^{*,(k)} + B u_t^{*,(k)} + w_t^{(k)},\ u_t^{*,(k)} = K^* x_t^{*,(k)}$$

The empirical cost under adaptive control $c_t^{(k)}$ is calculated similarly without averaging over Monte Carlo samples. Instantaneous regret is calculated by subtracting the empirical cost under optimal control from the empirical cost under adaptive control using each scheme i.e. $r_t^{(k)} = c_t^{(k)} - c_t^*$.

We evaluated the CE and RMN algorithms on a scalar system with true system and cost parameters $A = 1$, $B = 1$, $Q = 1$, $R = 0$, $W = 1$. The level of additive process noise is significant enough that an appreciable number of model estimates remain poor for many timesteps; this is necessary to observe a difference between CE and RMN control. We simulated the system over a time horizon of $T = 100$ steps. We drew $N_s = 100,000$ independent Monte Carlo samples and $N_b = 100$ bootstrap samples at each time step for uncertainty estimation. We used unity scaling of the multiplicative noise ($\gamma = 1$) and a tolerance of $\epsilon = 0.01$ for bisection to find the largest scaling c_γ of multiplicative noise variance in the multiplicative noise LQR algorithm. We used an exploration time of $t_{\text{explore}} = 5$ which ensures the least-squares estimate is non-degenerate. The figures have x-axis limits truncated to $[t_{\text{explore}}, T]$.

In Figure 1 we plot statistics of instantaneous regret using CE control and using RMN control. We are chiefly interested in performance in terms of expected regret and upper quantiles, which correspond to regret risk. Figure 1 demonstrates that the multiplicative noise control achieves much lower instantaneous regret in terms of both the mean and upper quantiles. In particular, we see that the performance of the multiplicative noise control is clearly many orders of magnitude better between the start and $t = 60$, $t = 40$, $t = 20$ for the 99.9th, 99th and 95th quantiles, respectively. After several time steps the model estimates improve and uncertainty estimates become sufficiently small that the difference between CE and RMN control becomes insignificant. The heaviness of regret tails is shown by the massive difference between the mean and median.

In Figure 2 we plot statistics of the nominal model estimate errors, which are applicable to both control schemes. This shows that the least-squares estimator provides models of increasing accuracy as time goes on, as expected. In Figure 3 we plot statistics of the multiplicative noise variances using RMN control. This shows that the multiplicative noise variances accurately reflect the true model error, i.e., the bootstrap model uncertainty estimator gives reasonable estimates.

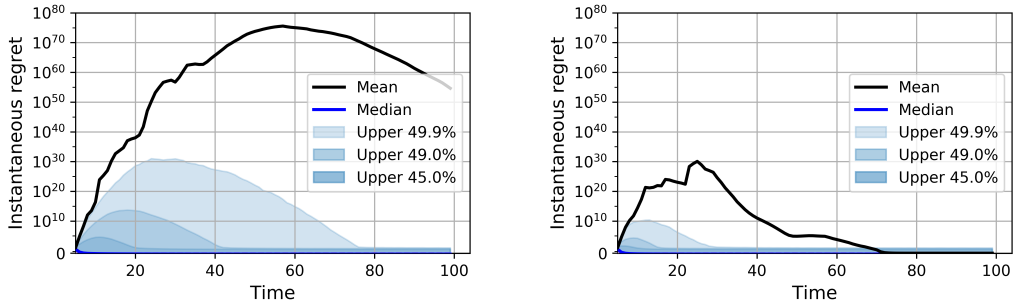
The benefits of RMN control over CE control more generally obviously cannot be inferred from this single instance. Indeed other preliminary numerical results on higher dimensional examples indicate that on some problem data (A, B, Q, R, W) the benefits of RMN control and how best to select the algorithm parameters are unclear, especially in initial stages when there is very high uncertainty around the nominal model estimates. However, there are at least some systems, like the one shown here, that are controlled with significantly lower risk using RMN control. We expect that by explicitly incorporating model uncertainty into the adaptive control design, it should be possible to realize the observed robustness benefits more broadly, which motivates further theoretical study. Code which realizes the algorithms of this paper and generates the reported results is available from https://github.com/TSummersLab/robust-adaptive-control-multinoise.

6. Conclusions

We proposed a robust adaptive control algorithm that uses the bootstrap to estimate model estimate covariances and a non-conventional multiplicative noise LQR robust control method. Ongoing and future work will go towards providing finite-time theoretical performance guarantees using tools from high-dimensional statistics, finding algorithm parameters that ensure uniform improvements over certainty-equivalent control for any system, and implementing model uncertainty estimates using recursive least-squares to alleviate computational burden.

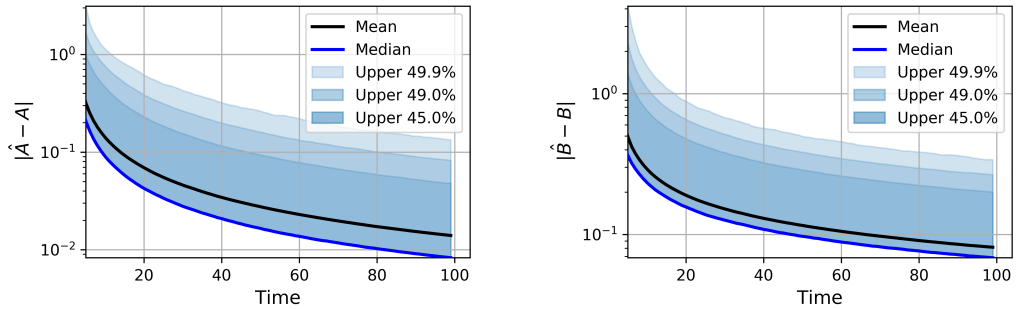
Acknowledgments

This material is based on work supported by the United States Air Force Office of Scientific Research under award number FA2386-19-1-4073.



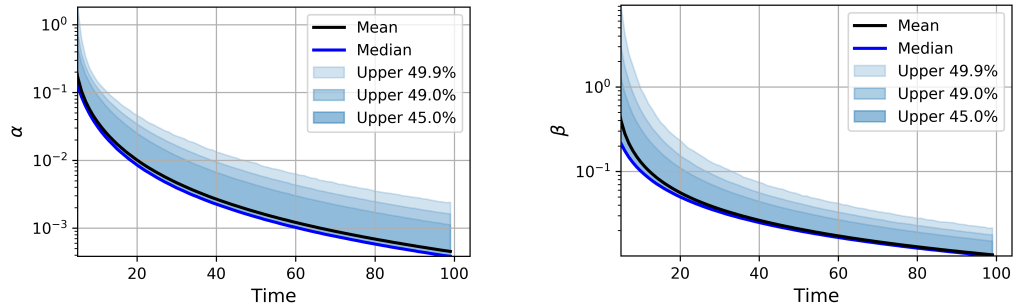
(*a*) CE (*b*) RMN

Figure 1: Instantaneous regret vs time for the example system using using certainty-equivalent (a) and multiplicative noise (b) control.



(*a*) A (*b*) B

Figure 2: Absolute error in estimated (a) A and (b) B matrices vs time for the example system.



(*a*) α (*b*) β

Figure 3: (a) State-dependent and (b) control-dependent multiplicative noise variances vs time for the example system using multiplicative noise control.

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.

Karl J Åström and Björn Wittenmark. *Adaptive control*. Courier Corporation, 2013.

Richard E Bellman. *Adaptive control processes: A guided tour*. Princeton University Press, 1961.

D Bernstein and S Greeley. Robust controller synthesis using the maximum entropy design equations. *IEEE Transactions on Automatic Control*, 31(4):362–364, 1986.

Richard L Burden, J Douglas Faires, and Albert C Reynolds. *Numerical analysis*. PWS publishing company, 1978.

Han-Fu Chen and Lei Guo. *Identification and stochastic adaptive control*. Springer Science & Business Media, 2012.

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 4188–4197, 2018.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, Aug 2019.

Laurent El Ghaoui. State-feedback control of systems with multiplicative noise via linear matrix inequalities. *Systems & Control Letters*, 24(3):223–228, 1995.

Benjamin Gravell, Peyman Mohajerin Esfahani, and Tyler Summers. Learning robust control for LQR systems with multiplicative noise via policy gradient. *CoRR*, 2019. URL http://arxiv.org/abs/1905.13547.

Wolfgang Härdle, Joel Horowitz, and Jens-Peter Kreiss. Bootstrap methods for time series. *International Statistical Review*, 71(2):435–459, 2003.

Panqanamala Ramana Kumar and Pravin Varaiya. *Stochastic systems: Estimation, identification, and adaptive control*, volume 75. SIAM, 2015.

Lennart Ljung. *System Identification: Theory for the User*. Pearson Education, 1998.

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.

Mario Milanese and Antonio Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, 27(6):997–1009, 1991.

Dan Simon. *Optimal state estimation*. John Wiley & Sons, 2006.

W Murray Wonham. Optimal stationary control of a linear system with state-dependent noise. *SIAM Journal on Control*, 5(3):486–500, 1967.