Online estimation and control with optimal pathlength regret

Gautam Goel GGOEL@CALTECH.EDU
Caltech

Babak Hassibi HASSIBI@CALTECH.EDU
Caltech

Editors: R. Firoozi, N. Mehr, E. Yel, R. Antonova, J. Bohg, M. Schwager, M. Kochenderfer

Abstract

A natural goal when designing online learning algorithms for non-stationary environments is to bound the regret of the algorithm in terms of the temporal variation of the input sequence. Intuitively, when the variation is small, it should be easier for the algorithm to achieve low regret, since past observations are predictive of future inputs. Such data-dependent "pathlength" regret bounds have recently been obtained for a wide variety of online learning problems, including online convex optimization (OCO) and bandits. We obtain the first pathlength regret bounds for online control and estimation (e.g. Kalman filtering) in linear dynamical systems. The key idea in our derivation is to reduce pathlength-optimal filtering and control to certain variational problems in robust estimation and control; these reductions may be of independent interest. Numerical simulations confirm that our pathlength-optimal algorithms outperform traditional H_2 and H_∞ algorithms when the environment varies over time.

1. Introduction

Online learning has traditionally focused on minimizing regret against static policies. For example, in the multi-armed bandit problem, we compare the rewards obtained by the online learner to the rewards that the learner could have counterfactually obtained if they had simply pulled the same arm in each round. However, many real-world environments are non-stationary, e.g. the rewards associated to actions vary over time. In such settings, it is more natural to compare the actions of the online learner against a time-varying sequence of actions, rather than a single fixed action. It is also natural to bound the regret of the online algorithm in terms of the temporal variation in the reward sequence. Intuitively, when the variation is small, it should be easier for the algorithm to achieve low regret, since past observations are predictive of future rewards. Such data-dependent "pathlength" regret bounds have recently been obtained for a wide variety of online learning problems, including online convex optimization (OCO) and bandits, e.g. Zinkevich (2003); Wei and Luo (2018); Bubeck et al. (2019); Goel and Wierman (2019).

A particularly challenging problem not considered in these prior works is obtaining pathlength regret bounds for online algorithms in environments with underlying dynamics. Intuitively, learning is harder in such settings because the rewards and actions are tightly coupled across rounds via the state; selecting a poor action in one round affects the rewards obtained in all subsequent rounds. This is in stark contrast to classical online learning problems (e.g. bandits), where a single poor decision results in a low reward for the algorithm in that round, but does not directly affect the rewards in future rounds. We ask: is it possible to design algorithms for online decision-making in dynamical systems which have regret bounded by pathlength? Answering this question has taken on a newfound urgency with the recent deployment of autonomous control systems such as drones and self-driving cars; these systems invariably encounter non-stationarity in their environments (e.g. changing weather, shifting traffic patterns, etc.)

1.1. Contributions of this paper

In this paper, we obtain the first pathlength regret bounds for online control and estimation (e.g. Kalman filtering) in linear dynamical systems. Our results show that it possible to design controllers and filters which dynamically adapt to regularity in the disturbance sequence, even though the rewards are strongly coupled across time by dynamics. We construct a controller whose regret against a clairvoyant offline optimal controller has optimal dependence on the pathlength of the driving disturbance. Similarly, we construct a filter whose regret against the optimal smoothed estimator has optimal dependence on the pathlength of the measurement disturbance. The key idea underpinning our results is to reduce pathlength-optimal filtering and control to certain variational problems widely studied in the context of robust estimation and control during the 1980s. In Section 3, we show that the problem of obtaining a pathlength-optimal controller in the original dynamical system can be reduced to the problem of obtaining the H_∞-optimal controller in a specially constructed synthetic system. Similarly, in Section 4 we reduce pathlength-optimal filtering to the classical Nehari problem from robust control. This problem, first introduced in 1957 in an operator theoretic context by Zeev Nehari Nehari (1957), asks how closely a noncausal function can be approximated by a causal one. We describe the pathlength-optimal filter in terms of a computationally efficient state-space solution to the Nehari problem. In Section 5, we present numerical experiments which confirm that our pathlength-optimal algorithms outperform traditional H_2 and H_∞ algorithms when the environment varies over time. Our results show that classical techniques from H_∞ estimation and control, which were originally designed with the aim of ensuring *robustness*, can instead be used to obtain *adaptivity*.

To comply with the page limit requirements, in this short version of our paper we focus on simply describing the state-space solutions of the pathlength-optimal controller and pathlength-optimal filter; we present a detailed derivation of all of our results, as well as additional experiments, in the full version of our paper, available at `https://arxiv.org/abs/2110.12544`.

1.2. Related work

The idea of bounding dynamic regret by pathlength was introduced in the seminal work of Zinkevich Zinkevich (2003); this idea was further explored in Chiang et al. (2012). Dynamic regret bounds in terms of pathlength were obtained for the multi-armed bandit problem in Besbes et al. (2015); Wei et al. (2016). More recently, static regret bounds (i.e. regret against the best fixed arm) in terms of pathlength were obtained in Wei and Luo (2018); Bubeck et al. (2019). Pathlength regret bounds for online convex optimization with switching costs were obtained in Goel and Wierman (2019).

The problem of designing controllers with optimal dynamic regret was studied in the finite-horizon, time-varying setting in Goel and Hassibi (2021c), in the infinite-horizon LTI setting in Sabag et al. (2021), and in the measurement-feedback setting in Goel and Hassibi (2021b). These works all bounded regret by the energy in the disturbances; the pathlength regret bounds we obtain in this paper also imply energy regret bounds which are optimal up to a factor of 4. Filtering algorithms with energy regret bounds were obtained in the finite-horizon setting in Goel and Hassibi (2021c) and the infinite-horizon setting in Sabag and Hassibi (2021). Gradient-based control algorithms with low dynamic regret against the class of disturbance-action policies were obtained in Zhao et al. (2021); the stronger metric of adaptive regret was studied in Gradu et al. (2020). We also note recent work Goel and Wierman (2019); Goel and Hassibi (2021a), which considered control through the lens of competitive ratio, a multiplicative analog of dynamic regret.

2. Preliminaries

2.1. Control setting

We focus on the linear-quadratic (LQ) control setting, where a linear time-invariant (LTI) system evolves according to the dynamics

$$x_{t+1} = Ax_t + B_u u_t + B_w w_t,$$

where $x_t \in \mathbb{R}^n$ is the state, $u_t \in \mathbb{R}^m$ is the control and $w_t \in \mathbb{R}^p$ is an external disturbance. We incur a quadratic cost

$$x_t^* Q x_t + u_t^* R u_t$$

in each round, where $Q \succeq 0, R \succ 0$. Naturally, our goal is to select the control actions u so as to minimize the aggregate cost across rounds (note that cost minimization can be reframed as reward maximization). As is standard in infinite-horizon control, we assume that (A, B_u) is stabilizable and $(A, Q^{1/2})$ is detectable. We consider both the causal setting, where the control u_t is selected after observing the state x_t and the disturbance w_t, and the strictly causal setting, where the control u_t is selected after observing x_t but before observing w_t.

It is convenient to reparameterize the system dynamics and costs as follows. Let L be a square-root of Q (e.g. $L^*L = Q$) and let $R^{1/2}$ denote a square-root of R. Define $s_t = Lx_t$, $v_t = R^{1/2}u_t$; note that we can easily recover u_t from v_t by setting $u_t = R^{-1/2}v_t$. With this notation, the cost incurred by an online algorithm which selects v in response to w is

$$ALG(w) = \|s\|_2^2 + \|v\|_2^2.$$

The dynamics can also be rewritten in terms of s_t and v_t:

$$x_{t+1} = Ax_t + B_u R^{-1/2} v_t + B_w w_t, \quad s_t = Lx_t.$$

We can cleanly capture the dynamics in terms of transfer matrices as

$$s = Fv + Gw$$

where F, G are strictly causal transfer matrices encoding $\{A, B_u R^{-1/2}, B_w, L\}$.

Our goal is to design a controller which minimizes regret against a clairvoyant offline optimal controller K_0 which selects the optimal control given perfect noncausal knowledge of w. One can show (Theorem 11.2.1 in Hassibi et al. (1999)) that the offline optimal control policy is

$$K_0(w) = -(I + F^*F)^{-1}F^*Gw, \tag{1}$$

and the corresponding offline optimal cost is

$$OPT(w) = w^*G^*(I + FF^*)^{-1}Gw. \tag{2}$$

Formally, we seek to solve the following problem:

Problem 1 (Pathlength-optimal control at level γ)**.** *Given $\gamma > 0$, find a causal (or strictly causal) controller K whose cost, $ALG(w)$, satisfies*

$$ALG(w) - OPT(w) < \gamma^2 \cdot \text{PATHLENGTH}(w), \tag{3}$$

or determine whether no such controller exists.

Once this feasibility problem is solved, it is easy to recover the optimal value of γ via bisection; we define the *pathlength-optimal controller* to be the controller which satisfies (3) with the smallest possible γ.

Our proof technique is to reduce the problem of finding a controller satisfying (3) to an H_∞ control problem:

Problem 2 (H_∞-optimal control at level γ)**.** *Given $\gamma > 0$, find a causal (or strictly causal) controller K whose cost, $ALG(w)$, satisfies*

$$ALG(w) < \gamma^2 \cdot \text{ENERGY}(w),$$

or determine whether no such controller exists.

2.2. Filtering setting

We consider a linear time-invariant (LTI) system that evolves according to the dynamics

$$x_{t+1} = Ax_t + Bw_t, \quad y_t = Cx_t + v_t$$

where $x_t \in \mathbb{R}^n$ is the state, y_t is a noisy linear measurement of the state, and $w_t \in \mathbb{R}^m$ and $v_t \in \mathbb{R}^p$ are external disturbances. Our goal is to estimate a linear function of the states:

$$s_t = Lx_t,$$

given the measurements, where $L \in \mathbb{R}^r$. The special case where $r = n$ and $L = I$ corresponds to estimating the state itself. In each round, we output an estimate $\hat{s}_t$ based on the current observation y_t and all previous observations and incur a squared-error loss

$$\|\hat{s}_t - s_t\|_2^2.$$

Naturally, our goal is to select the estimates so as to minimize the aggregate loss across all rounds. As is standard in infinite-horizon estimation, we assume that (A, B) is stabilizable and (A, C) is detectable.

In this paper, we obtain the pathlength-optimal filter using frequency domain analysis. We let $J(z)$ be the transfer matrix mapping the disturbance w to the signal s:

$$J(z) = L(zI - A)^{-1}B$$

and let $H(z)$ be the transfer matrix mapping the disturbance w to the observations y:

$$H(z) = C(zI - A)^{-1}B.$$

Any filter K naturally induces a transfer matrix T_K which maps the disturbances w and v to the estimation error $\hat{s} - s$ incurred by K:

$$T_K(z) = \begin{bmatrix} J(z) - K(z)H(z) & -K(z) \end{bmatrix}.$$

Our goal is to design a filter K which minimizes regret against the optimal smoothed estimator, i.e. the estimator which minimizes estimation error given noncausal access to the observations y. One can show (Theorem 10.3.1 in Hassibi et al. (1999)) that the optimal smoothed estimator in the z-domain is

$$K_0(z) = J(z)H^*(z^{-*})(I + H(z)H^*(z^{-*}))^{-1}. \tag{4}$$

In this paper, we focus on bounding the regret by the pathlength of the measurement disturbance v and the energy of w; we leave the problem of bounding regret by the pathlength of both w and v for future work. Formally, we seek to solve the following problem:

Problem 3 (Pathlength-optimal filtering at level γ)**.** *Given $\gamma > 0$, find a filter K such that the regret*

$$\begin{bmatrix} w^* & v^* \end{bmatrix} (T_K^* T_K - T_{K_0}^* T_{K_0}) \begin{bmatrix} w \\ v \end{bmatrix}$$

is at most

$$\gamma^2 \cdot (\text{ENERGY}(w) + \text{PATHLENGTH}(v)), \tag{5}$$

or determine whether no such filter exists.

Once this feasibility problem is solved, it is easy to recover the optimal value of γ via bisection; we define the *pathlength-optimal filter* to be the filter which satisfies (5) with the smallest possible value of γ.

Our proof technique reduces the problem of finding a filter satisfying (5) to Nehari problem:

Problem 4 (Nehari problem at level γ)**.** *Given $\gamma > 0$ and a strictly anticausal function $T(z)$, find a causal and bounded function $K(z)$s such that*

$$\|K(z) - T(z)\|_\infty < \gamma, \tag{6}$$

or determine whether no such $K(z)$ exists.

Recall that

$$\|K(z) - T(z)\|_\infty = \sup_{\theta \in [0, 2\pi]} \bar{\sigma}\left(K(e^{i\theta}) - T(e^{i\theta})\right).$$

Informally, the Nehari problem seeks to find a causal function $K(z)$ which is γ-close to the anticausal function $T(z)$ at every frequency $\theta \in [0, 2\pi]$.

3. Pathlength-optimal control

Our main control-theoretic result is a computationally efficient state-space description of a controller satisfying condition (3):

theorempathlength-optimal-controller-thm-ih Suppose (A, B_u) is stabilizable and $(A, Q^{1/2})$ is detectable. A causal infinite-horizon controller satisfying (3) exists if and only if there exists a solution to the Ricatti equation

$$\hat{P} = \hat{L}^* \hat{L} + \hat{A}^* \hat{P} \hat{A} - \hat{A}^* \hat{P} \tilde{B} \tilde{H}^{-1} \tilde{B}^* \hat{P} \hat{A}$$

where

$$\hat{L} = \begin{bmatrix} L & 0 \end{bmatrix}, \quad \hat{A} = \begin{bmatrix} A & -B_w K_2 \\ 0 & \tilde{A} - \tilde{B}_w K_2 \end{bmatrix} \quad \hat{B}_w = \begin{bmatrix} B_w \Sigma_2^{-1/2} \\ \tilde{B}_w \Sigma_2^{-1/2} \end{bmatrix}, \quad \hat{B}_u = \begin{bmatrix} B_u \\ 0 \end{bmatrix},$$

$$\tilde{B} = \begin{bmatrix} \hat{B}_u & \hat{B}_w \end{bmatrix}, \quad \tilde{R} = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \quad \tilde{H} = \tilde{R} + \tilde{B}^* P \tilde{B},$$

where $\tilde{A}, \tilde{B}_w, K_2, \Sigma_2$ are defined in the full paper, such that

- 1. $\hat{A} - \tilde{B}\tilde{H}^{-1}\tilde{B}^*\hat{P}\hat{A}$ is stable;

2. $\tilde{R}$ and $\tilde{H}$ have the same inertia;

3. $\hat{P} \succeq 0$.

In this case, a causal infinite-horizon H_∞ controller satisfying (3) is given by

$$u_t = -R^{-1/2} \hat{H}^{-1} \hat{B}_u^* \hat{P} (\hat{A} \xi_t + \hat{B}_w w_t'),$$

where $\hat{H} = I + \hat{B}_u^* \hat{P} \hat{B}_u$ and the dynamics of ξ are

$$\xi_{t+1} = \hat{A}\xi_t + \hat{B}_u u_t + \hat{B}_w w_t'$$

and we initialize $\xi_0 = 0$. The synthetic disturbance w' can be computed using the recursion

$$\nu_{t+1} = \tilde{A}\nu_t + \tilde{B}_w w_t, \quad w_t' = \Sigma_2^{1/2}(K_2 \nu_t + w_t),$$

A strictly causal infinite-horizon controller satisfying (3) exists if and only if conditions 1 and 3 hold, and additionally

$$\hat{B}_u^* \hat{P} \hat{B}_u \prec \gamma^2 I, \quad I + \hat{B}_w^* \hat{P} (I - \hat{B}_u (-\gamma^2 I + \hat{B}_u^* \hat{P} \hat{B}_u)^{-1} \hat{B}_u^* \hat{P}) \hat{B}_w \succ 0.$$

In this case, a strictly causal controller satisfying (3) is given by

$$u_t = -R^{-1/2} \hat{H}^{-1} \hat{B}_u^* \hat{P} \hat{A} \xi_t.$$

Proof We refer to the full version of our paper, available at https://arxiv.org/abs/2110.12544.

4. Pathlength-optimal filtering

We obtain a computationally efficient state-space description of a filter satisfying condition (5): theoremfilter satisfying (5) exists if and only if

$$\bar{\sigma}(Z\Pi) \leq 1,$$

where Z and Π are solutions of the Lyapunov equations

$$Z = F^*ZF + H^*H, \quad \Pi = F\Pi F^* + GG^*$$

and

$$H = \hat{L}W_2A_2^*(I - A_2^*)^{-1}, \quad F = A_2^*, \quad G = C^*\Sigma_2^{-1/2}. \tag{7}$$

In this case, a filtered estimator satisfying (5) is given by

$$\hat{s}_t = \Sigma_3^{-1/2}\beta_t - K_3\pi_t,$$

where β, π have the dynamics

$$\pi_{t+1} = (A_1 - A_1W_1L^*K_3)\pi_t + A_1W_1L^*\Sigma_3^{-1/2}\beta_t, \quad \beta_t = \Delta_3(1)Q(1)z_t + \alpha_t - \alpha_{t-1}.$$

The variable α is given by

$$\alpha_t = H(\Pi\xi_{t+1}^1 + \xi_t^2),$$

where ξ^1, ξ^2 have dynamics

$$\xi_{t+1}^1 = F_\gamma\xi_t^1 + K_\gamma z_t, \quad \xi_{t+1}^2 = A_2^*\xi_t^2 + C^*\Sigma_2^{-1/2}z_t$$

where

$$z_t = \Sigma_2^{-1/2}y_t - Ce_t, \quad e_{t+1} = A_2e_t + K_2y_t.$$

The matrices $\Delta_3(1)Q(1), A_1, A_2, \Sigma_2, \Sigma_3, K_2, K_3, W_1, F_\gamma, K_\gamma$ are defined in the full paper.

Proof We refer to the full version of our paper, available at https://arxiv.org/abs/2110.12544.

5. Experiments

5.1. Pathlength-optimal control

We benchmark our pathlength-optimal controller in a nonlinear inverted pendulum system. The inverted pendulum system has two scalar states, θ and $\dot{\theta}$, representing angular position and angular velocity, respectively, and a single scalar control input u. The state $(\theta, \dot{\theta})$ evolves according to the nonlinear evolution equation

$$\frac{d}{dt}\begin{bmatrix}\theta \\ \dot{\theta}\end{bmatrix} = \begin{bmatrix}\dot{\theta} \\ \frac{mg\ell}{J}\sin\theta + \frac{\ell}{J}u\cos\theta + \frac{\ell}{J}w\cos\theta\end{bmatrix},$$

where w is an external scalar disturbance, and m, ℓ, g, J are physical parameters describing the system; we assume that units are scaled so that these parameters are 1. Although these dynamics are

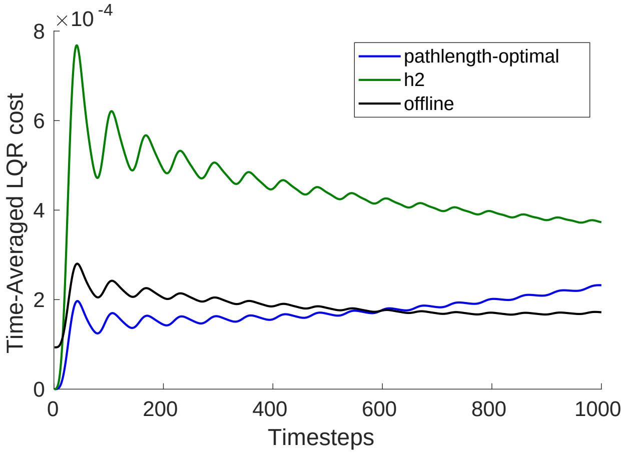

Figure 1: Relative performance of LQ controllers in an inverted pendulum system driven by sinusoidal noise.

nonlinear, we can benchmark the regret-optimal controller against the H_2-optimal, H_∞-optimal, and clairvoyant offline optimal controllers using Model Predictive Control (MPC). In the MPC framework, we iteratively linearize the model dynamics around the current state, compute the optimal control action in the linearized system, and then update the state in the original nonlinear system using this control action. In all of our experiments we take $Q, R = I$ and initialize θ and $\dot{\theta}$ to zero. We set the discretization parameter $\delta_t = 0.001$ and sample the dynamics at intervals of δ_t.

In Figures 1 and 2, we plot plot the relative performance of the pathlength-optimal, H_2-optimal, and offline optimal controllers across different input disturbances. In our first experiment the cost incurred by the H_∞-optimal controller is orders of magnitude higher than that of the other controllers and is not shown in the corresponding figure. We first consider a sinusoidal disturbance with amplitude 1 and period 20π; this disturbance is temporally 0.1-Lipschitz, and hence has low pathlength. In our second experiment, the disturbance is a step-function: it is $+1$ for 500 timesteps and -1 for the next 500 timesteps. While this disturbance has high energy, its pathlength is only 4. In both of these experiments, the pathlength-optimal controller outperforms the H_2-optimal and H_∞-optimal controllers and closely tracks the clairvoyant offline optimal controller. For more experiments, we refer to the full version of our paper, available at https://arxiv.org/abs/2110.12544.

5.2. Pathlength-optimal filtering

We consider a one-dimensional tracking problem, where the goal is to estimate an object's position given noisy observations of its trajectory. We consider the state-space model

$$\begin{bmatrix} x_{t+1} \\ \nu_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & \delta_t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \nu_t \end{bmatrix} + \begin{bmatrix} 0 \\ \delta_t \end{bmatrix} \alpha_t, \quad y_t = x_t + v_t, \quad s_t = x_t,$$

where x_t is the object's position, ν_t is the object's velocity, α_t is the object's instantaneous acceleration due to external forces, and v_t is measurement noise. We take $\delta_t = 0.01$ and initialize $x_0 = 0$. In this system, the optimal value of γ is $\gamma^\star \approx 35.64$.

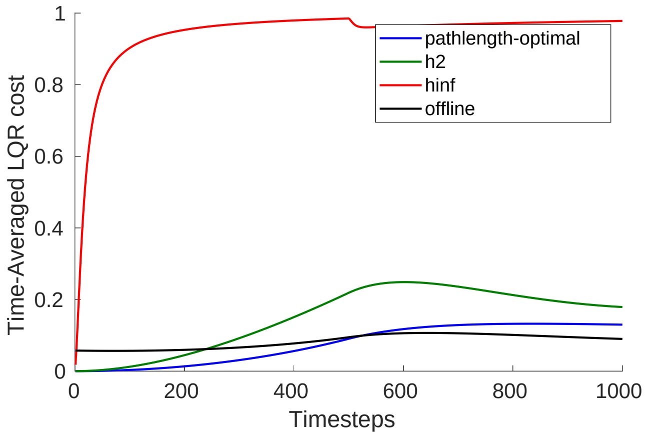

Figure 2: Relative performance of LQ controllers in an inverted pendulum system driven by step-function noise.

We benchmark the performance of the pathlength-optimal filter against that of the Kalman filter (e.g. the H_2-optimal filter). Recall that the key innovation of the pathlength-optimal filter relative to standard filters is that the pathlength-optimal filter is designed to achieve low regret when the measurement disturbance v has low pathlength; for this reason, we focus on measuring how the performance varies across many different values of v. For simplicity, we take the driving disturbance α to be picked i.i.d from a standard Gaussian across all of our experiments. In Figure 3, we plot the relative performance of the pathlength-optimal filter against that of the Kalman filter when the measurement disturbance is constant, i.e. $v_t = 1$ for all t. This disturbance has high energy but zero pathlength - as expected, the pathlength-optimal filter easily beats the Kalman filter, achieving orders-of-magnitude less estimation error. In Figure 4, v varies sinusoidally with period 200π. This disturbance has very low pathlength, and in fact is temporally 0.01-Lipshitz, e.g. $\|v_{t+1} - v_t\| \leq 0.01$ for all t. Again, the pathlength-optimal filter outperforms the Kalman filter by orders of magnitude. Together, these plots are consistent with the message of this paper: when the pathlength of the measurement disturbance is small, the pathlength-optimal filter produces very accurate estimates of the state and outperforms standard filtering algorithms. For more experiments, we refer to the full version of our paper, available at https://arxiv.org/abs/2110.12544.

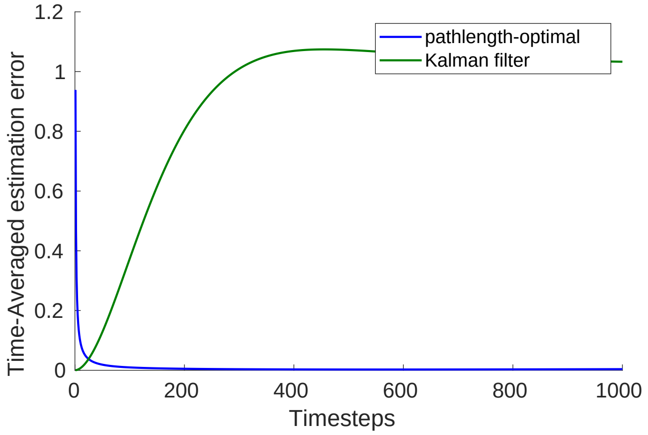

Figure 3: The driving disturbance is drawn i.i.d from a standard Gaussian and the measurement disturbance is constant ($v_t = 1$ for all t). The pathlength of the measurement disturbance is zero, though its energy is large.

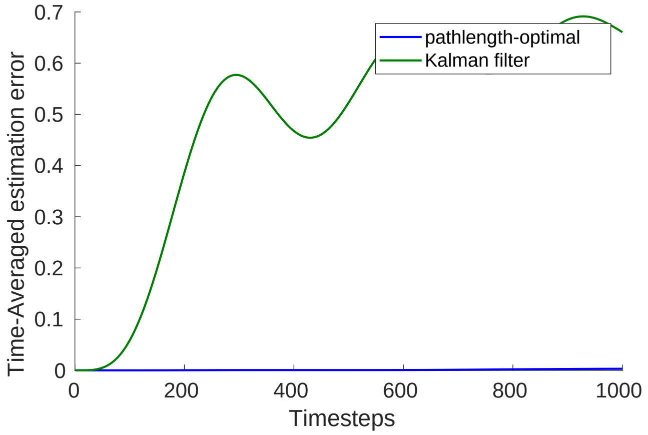

Figure 4: The driving disturbance is drawn i.i.d from a standard Gaussian and the measurement disturbance varies sinusoidally with period 200π. The pathlength of the measurement disturbance is small, relative to its energy.

References

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

Sébastien Bubeck, Yuanzhi Li, Haipeng Luo, and Chen-Yu Wei. Improved path-length regret bounds for bandits. In *Conference On Learning Theory*, pages 508–528. PMLR, 2019.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, pages 6–1. JMLR Workshop and Conference Proceedings, 2012.

Gautam Goel and Babak Hassibi. Competitive control. *arXiv preprint arXiv:2107.13657*, 2021a.

Gautam Goel and Babak Hassibi. Regret-optimal measurement-feedback control. In *Learning for Dynamics and Control*, pages 1270–1280. PMLR, 2021b.

Gautam Goel and Babak Hassibi. Regret-optimal estimation and control. *arXiv preprint arXiv:2106.12097*, 2021c.

Gautam Goel and Adam Wierman. An online algorithm for smoothed regression and lqr control. *Proceedings of Machine Learning Research*, 89:2504–2513, 2019.

Paula Gradu, Elad Hazan, and Edgar Minasyan. Adaptive regret for control of time-varying dynamics. *arXiv preprint arXiv:2007.04393*, 2020.

Babak Hassibi, Ali H Sayed, and Thomas Kailath. *Indefinite-quadratic estimation and control: a unified approach to H 2 and H-infinity theories*. SIAM, 1999.

Zeev Nehari. On bounded bilinear forms. *Annals of Mathematics*, pages 153–162, 1957.

Oron Sabag and Babak Hassibi. Regret-optimal filtering. In *International Conference on Artificial Intelligence and Statistics*, pages 2629–2637. PMLR, 2021.

Oron Sabag, Gautam Goel, Sahin Lale, and Babak Hassibi. Regret-optimal controller for the full-information problem. In *2021 American Control Conference (ACC)*, pages 4777–4782. IEEE, 2021.

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Conference On Learning Theory*, pages 1263–1291. PMLR, 2018.

Chen-Yu Wei, Yi-Te Hong, and Chi-Jen Lu. Tracking the best expert in non-stationary stochastic environments. *Advances in neural information processing systems*, 29:3972–3980, 2016.

Peng Zhao, Yu-Xiang Wang, and Zhi-Hua Zhou. Non-stationary online learning with memory and non-stochastic control. *arXiv preprint arXiv:2102.03758*, 2021.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936, 2003.