RESEARCH STATEMENT

Gautam Goel

The goal of my research is to combine machine learning with signal processing and control to obtain novel data-driven algorithms for inference and control in dynamical systems. This area is currently undergoing a renaissance due to the broad range of potential applications, which include control of robotic systems such as drones and autonomous vehicles; analysis of biological signals, such as EKGs and EEGs; decentralized control of cyberphysical systems, such as the smart grid; and climate modeling and weather forecasting. My work focuses on answering two complementary questions:

Q1: How do we incorporate learning and adaptivity into signal processing and control? Traditional algorithms for signal processing and control in dynamical systems posit a specific model for the environment encountered by the algorithm. For example, the Kalman filter is derived under an i.i.d. stochastic noise model; this model may be too optimistic in practice. Similarly, robust controllers assume a worst-case model of the disturbance; this model is often too pessimistic. A natural idea is to use ideas from machine learning to design adaptive algorithms which achieve the Best of Both Worlds: in stochastic settings they should perform as well as algorithms specifically designed for stochastic environments, but they should also retain strong performance guarantees even in adversarial environments. In a similar vein, it is usually assumed that the algorithm has a perfect model of the system dynamics; this assumption rarely holds in practice. How do we design algorithms which improve their own internal model of the dynamics over time to make better decisions?

Q2: How do we extend online learning to settings with non-stationarity and dynamics? Classical online learning algorithms are designed to perform well in stationary environments, where the rewards associated to actions do not change over time. However, many real-world applications call for online decision-making in dynamic, non-stationary environments; one such application is the smart grid, which must constantly supply enough power to meet demand while dealing with the uncertainty caused by intermittent energy resources, such as solar and wind. Classical online learning theory also assumes that the rewards are not coupled across rounds, i.e., the actions of the learner do not affect the rewards available in the future. In control this assumption does not hold; the actions selected by the algorithm directly affect the state of the system, which in turn affects future rewards. The application of online learning in such domains hence calls for new approaches to algorithm design.

These questions are both aimed at bridging several different kinds of online decision-making, namely filtering, control and online learning. My research agenda is to develop a cross-cutting theory of online decision-making which pushes past standard assumptions such as linearity and stationarity to tackle challenges such as partial observability, robustness, decentralization, and safety. While my work has a strong theoretical flavor, my ultimate aim is to have real-world impact in diverse application domains including robotics & autonomy, aerospace systems, power systems, and signal processing. I enjoy interacting with researchers across engineering, computer science, and applied mathematics, and am actively seeking interdisciplinary collaborations.

Learning-Inspired Filtering and Control

Motivation

Optimal control frames control as an optimization problem, where the goal of the controller is to minimize a cost function which measures how far away the system is from the desired trajectory and how much energy the controller is expending. In many real-world settings, the controller is unable to directly observe the state but must instead estimate the state indirectly given noisy measurements of the system. The problem of constructing this estimate is called *state estimation* or *filtering*. This problem has been widely studied in the signal processing community, leading to well-known algorithms like the Kalman filter and the particle filter.

Filtering and control have traditionally been studied in two distinct settings. In the H_2 setting, the disturbances and measurement noise are both assumed to be generated independently by zero-mean stochastic processes, and the goal of the controller is to minimize its expected cost. The Kalman filter is the H_2-optimal filter and the well-known LQG controller is the H_2-optimal measurement-feedback controller. In the H_∞ setting, control is formulated as a minimax game played between the controller and an adversary which selects the disturbances and measurement noise; the goal of the controller is to minimize its own cost, while the goal of the adversary is to force the controller to incur high cost. Intuitively, this formulation leads to "robust" controllers with bounded worst-case performance.

Both of these design methodologies suffer from an obvious drawback: a controller which encounters a disturbance which falls outside of the class it was designed to handle may perform poorly. Indeed, it has been observed that H_∞ controllers tend to be over-conservative, whereas H_2 controllers have poor performance outside of stochastic settings. This observation naturally motivates the design of adaptive controllers which dynamically adjust their control strategy as they observe the disturbance instead of blindly following a prescribed strategy. Inspired by online learning, I have explored how *regret* can be used as a design criterion for filters and controllers. Regret is the standard performance metric in online learning; it is the difference between the costs incurred by the online algorithm and the costs which could have counterfactually been incurred given the benefit of hindsight. Intuitively, a learning algorithm with low regret is one which is resilient to uncertainty - it performs almost as well as a hypothetical algorithm which could make the optimal decisions with perfect knowledge of the relationship between actions and rewards.

My contributions

I have developed a new "Regret-Optimal" approach to filtering and control. In Regret-Optimal control, we adopt a very different perspective from classical H_2 and H_∞ control: instead of explicitly modeling the disturbances and designing the controller accordingly, we instead design a controller whose cost is as close as possible to the cost which could be obtained by a hypothetical "noncausal" controller which knew the disturbance sequence in advance and selected the cost-minimizing sequence of actions accordingly. Similarly, in Regret-Optimal filtering, we design a filter whose cost is as close as possible to the cost which could be obtained by a hypothetical noncausal filter which received the whole sequence of measurements at once and selected the optimal (in a least-squares sense) sequence of state estimates.

My main technical contribution has been to show that Regret-Optimal controllers can be efficiently derived using a reduction to H_∞ control [10, 11, 13, 14]. This is surprising, because H_∞ theory was originally developed with the aim of ensuring *robustness*; my work shows that it can be repurposed to ensure *adaptivity*. Specifically, I have shown that a controller with optimal regret in a given linear dynamical system can be viewed as an H_∞ controller in a synthetic system of larger dimension (a

similar reduction allows one to construct Regret-Optimal filters [12, 13]). The extra states of the synthetic system can be viewed as a "sufficient statistic" that summarizes which information about the past the algorithm needs to store to improve its decisions in the future. Regret-Optimal control can also be extended to include the alternative metric of competitive ratio, which is the worst-case ratio between the cost incurred by the online controller and the cost incurred by the noncausal controller. Competitive ratio is the multiplicative analog of regret and the most commonly studied performance metric in the online algorithms community; over the past decade it has also attracted attention in the online learning community (e.g., [2, 15, 16]).

Prior work in learning and control (e.g., [1, 5, 8, 32]) modifies preexisting algorithms for online learning to obtain asymptotic regret bounds (e.g., $O(\log T), O(\sqrt{T})$, etc.) against finite-dimensional, parametric classes of controllers, such as state-feedback policies. In contrast, my work shows that controllers with optimal regret can be constructed directly using control-theoretic methods, without relying on algorithms from online learning. The resulting Regret-Optimal controller is the controller with *optimal* regret out of *all possible controllers*, including nonlinear controllers. My work also establishes that a Regret-Optimal controller exists whenever an H_2 or H_∞-optimal controller exists; the construction of Regret-Optimal controllers does not rely on any non-standard assumptions. In other words, a Regret-Optimal controller is a drop-in replacement for an H_2 or H_∞ controller, and can be used anywhere these controllers are used. Numerical simulations show that Regret-Optimal controllers do in fact achieve the Best of Both Worlds, i.e., they perform almost as well as an H_2 controller when the noise is stochastic but have bounded worst-case performance, like an H_∞ controller.

Although I proposed Regret-Optimal control quite recently, Regret-Optimal control has already become a very active research area; subsequent papers by several different research groups extend the Regret-Optimal control framework to include safety constraints (e.g., [6, 7, 23, 36]) or integrate predictions to improve the regret bound (e.g., [26, 27]). Another line of work shows a strong equivalence between bounded regret and closed-loop stability (e.g., [17, 28]).

Extensions & Future Directions

Nonlinear/Unknown Dynamics. Although Regret-Optimal controllers are derived in the context of control in linear systems, they can also be applied to systems with nonlinear dynamics using Model Predictive Control (MPC). In the MPC framework, we iteratively linearize the model dynamics around the current state, compute the optimal control signal in the linearized system, and then update the state in the original nonlinear system using this control signal. Numerical experiments using MPC in the nonlinear inverted pendulum system show that Regret-Optimal controllers outperform standard H_2 and H_∞ controllers, often by orders of magnitude; one possible explanation is that Regret-Optimal controllers integrate learning and are hence better able to adapt to nonlinear dynamics.

In future work, I would like to study nonlinear adaptive control through the lens of *residual learning*. In residual learning, we use deep learning to obtain a control policy which corrects for the "residual", i.e., the error between the true dynamics and the controller's internal model of the dynamics. Intuitively, this combines the strengths of classical optimal control and modern machine learning; a similar idea could be used to construct a nonlinear filter. While residual learning works very well in practice [30], it currently lacks a rigorous theory to explain its empirical successes. A key technical challenge is to understand how the performance of the policy learned with a residual network scales with the amount of training data and the curvature of the dynamics; intuitively, controllers in systems which are "almost linear" should need much less training data than controllers in highly nonlinear systems.

Safety. A central concern in autonomous systems such as drones and self-driving cars is integrating safety constraints such as collision avoidance. A recent paper directly extends my Regret-Optimal

control framework to incorporate safety constraints [23]; unfortunately the controller proposed in that paper is the solution of a high-dimensional semidefinite program and is hence not efficiently computable in practice. An important research direction is to find tractable convex formulations of safe control which retain the adaptive guarantees of Regret-Optimal control.

Timescale separation. Many control systems consist of of interacting processes which evolve at multiple timescales. The fast layer is often used for disturbance rejection, while the slow layer aggregates global information to select the optimal trajectory for the overall system. One example is sensorimotor control, which integrates a fast, reflexive layer with a slow, high-level planning layer. Another example is the smart grid, which couples fast timescale frequency regulation with slow timescale economic dispatch. In [9] I started exploring a simplified model of timescale separation; the co-design of fast and slow controllers remains an exciting direction for future work.

Online Learning 2.0

Another major thrust of my research is to understand how online algorithms should be designed for learning in non-stationary environments and in settings where the rewards obtained by the algorithm in each round do not depend solely on the decision made by the algorithm in that round, but potentially on all previous choices made by the algorithm.

In one line of work [4, 15, 16], I studied online learning in non-stationary environments, specifically in the context of Online Convex Optimization (OCO) with switching costs. While most works in online learning study how an online algorithm can learn the best fixed action in a stationary environment, this problem studies how an online algorithm should adapt its decisions in the face of a changing environment. Furthermore, in this problem the online algorithm incurs a penalty every time it changes its action; the rewards obtained by the algorithm are hence coupled across rounds. This models many applications where it is expensive to change your decision, such as dynamic right-sizing in datacenters [20] and smooth camera tracking [19]. I obtained the first online algorithm for this problem which attained a constant competitive ratio in metric spaces of arbitrary dimension, thus resolving an open question posed by Lin and Wierman [20]; this algorithm is called Online Balanced Descent (OBD) since it balances the rewards it obtains from selecting good actions with the cost of switching actions. This style of algorithm has formed the basis for several subsequent works by other researchers, e.g.,[3, 34]. I also showed that OBD achieves bounded competitive ratio when applied to online control, thus identifying a surprising connection between OCO and control which was extended by subsequent works from several different research groups, e.g. [1, 8, 31, 32].

More recently, I have explored how we can combine Regret-Optimal control with online learning algorithms to obtain novel "hybrid" algorithms. In forthcoming work with several collaborators, I showed that Regret-Optimal controllers are closely approximated by a class of control policies called disturbance-action-control (DAC) policies. These policies have attracted much recent attention in the learning and control literature, because the convexity of this class means that it can be efficiently optimized online using standard gradient-descent based algorithms from online learning. These results imply that we can combine a Regret-Optimal controller with the recently proposed algorithm in [24] to obtain a controller with optimal competitive ratio (up to a sublinear correction) even in systems with unknown dynamics; this is the first result to obtain a competitive ratio guarantee in systems with unknown dynamics. I believe that this hybrid approach to controller design, which augments an optimal controller with an iterative, gradient-based online learning layer, holds much promise in the area of learning and control; similar ideas could be used for adaptive filtering.

Extensions & Future Directions

Integrating Predictions into Online-Decision-Making. In previous work, I have studied how predictions can be leveraged to improve the performance of online algorithms [21]. Specifically, this work obtained an algorithm for OCO with switching costs whose regret decreases linearly in the prediction window, i.e., doubling the number of predictions available to the algorithm cuts its regret in half. Surprisingly, it is possible to achieve the same improvement in performance even if the costs are nonconvex, provided they satisfy a certain "order-of-growth" condition. This result highlights that leveraging predictions is a simple way to dramatically improve the performance of online learning algorithms. In future work I would like to study how online learning algorithms can be augmented with predictions forecast by ML models; this is a very active area of research (e.g., [22, 25, 29]) .

Integrating Deep Learning into Online Decision-Making. A recent line of work [18, 35] studies how standard contextual bandit algorithms like UCB can be modified to use deep learning to learn the relationship between features and rewards; the advantage of this approach is that it captures settings where the rewards are a nonlinear function of the actions and contexts, unlike standard UCB which assumes a linear relationship. These "NeuralUCB" algorithms empirically outperform vanilla UCB, but there is little theoretical understanding of this phenomenon. More generally, there is a huge opportunity to explore how the successes of deep learning can be used to improve online decision-making.

Longer-term Directions

Looking beyond my medium-term research agenda to study connections between online learning, signal processing, and control, I am also interested in more speculative long-term directions:

Best of Both Worlds in Reinforcement Learning. Reinforcement learning (RL) and optimal control are intimately related; the goal of both fields is to determine how an agent should steer the trajectory of an evolving system so as to maximize its cumulative rewards. While RL has historically focused on settings where the state and action spaces are finite (tabular RL), over the past decade there has been a surge in interest in settings where the state and action spaces are exponentially large (as in games such as Go or StarCraft) or continuous (as in robotics and control). Unfortunately, the high dimensionality of these spaces means that it is computationally infeasible to use standard tabular RL techniques to learn an optimal policy; state-of-the-art RL systems hence resort to heuristics like function approximation and deep Q-learning, which lack a sound theoretical justification. At the same time, there is growing interest in moving beyond worst-case analysis of RL to obtain algorithms with optimal instance-dependent performance, i.e., algorithms which retain worst-case guarantees but can take advantage of problem structure to perform better on easier instances. Unfortunately, prior work (e.g., [32, 33]) is restricted to the tabular setting. A major open problem is thus to obtain instance-optimal RL algorithms for high-dimensional, continuous action spaces.

Shannon meets Hamming: Best of Both Worlds in Coding Theory. Another speculative direction for future work is to study how online learning can be integrated into coding theory. Coding theory is often studied under two error models; a stochastic model (the Shannon model) and an adversarial model (the Hamming model). Consider two parties sequentially transmitting data to each other over a noisy channel - is there a transmission protocol that quickly converges to the optimal rate under both models? In other words, is there a protocol which achieves the Best of Both Worlds - efficient data transfer in both stochastic and adversarial settings? Such a protocol would be of great practical utility, since it would be able to take advantage of favorable conditions for information exchange (for example, periods of low interference) while also being robust to occasional disruptions.

References

[1] Naman Agarwal et al. "Online control with adversarial disturbances". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 111–119.

[2] Lachlan Andrew et al. "A tale of two metrics: Simultaneous bounds on competitiveness and regret". In: *Conference on Learning Theory*. PMLR. 2013, pp. 741–763.

[3] CJ Argue, Anupam Gupta, and Guru Guruganesh. "Dimension-free bounds for chasing convex functions". In: *Conference on Learning Theory*. PMLR. 2020, pp. 219–241.

[4] Niangjun Chen, Gautam Goel, and Adam Wierman. "Smoothed online convex optimization in high dimensions via online balanced descent". In: *Conference On Learning Theory*. PMLR. 2018, pp. 1574–1594.

[5] Alon Cohen, Tomer Koren, and Yishay Mansour. "Learning Linear-Quadratic Regulators Efficiently with only $\sqrt{T}$ Regret". In: *arXiv preprint arXiv:1902.06223* (2019).

[6] Alexandre Didier, Jerome Sieber, and Melanie N Zeilinger. "A system level approach to regret optimal control". In: *IEEE Control Systems Letters* (2022).

[7] Alexandre Didier and Melanie N Zeilinger. "Generalised Regret Optimal Controller Synthesis for Constrained Systems". In: *arXiv preprint arXiv:2211.08101* (2022).

[8] Dylan J Foster and Max Simchowitz. "Logarithmic regret for adversarial online control". In: *arXiv preprint arXiv:2003.00189* (2020).

[9] Gautam Goel, Niangjun Chen, and Adam Wierman. "Thinking fast and slow: Optimization decomposition across timescales". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE. 2017, pp. 1291–1298.

[10] Gautam Goel and Babak Hassibi. "Competitive Control". In: *arXiv preprint arXiv:2107.13657* (2021).

[11] Gautam Goel and Babak Hassibi. "Measurement-Feedback Control with Optimal Data-Dependent Regret". In: *arXiv preprint arXiv:2209.06425* (2022).

[12] Gautam Goel and Babak Hassibi. "Online estimation and control with optimal pathlength regret". In: *arXiv preprint arXiv:2110.12544* (2021).

[13] Gautam Goel and Babak Hassibi. "Regret-optimal Estimation and Control". In: *arXiv preprint arXiv:2106.12097* (2021).

[14] Gautam Goel and Babak Hassibi. "Regret-optimal measurement-feedback control". In: *Learning for Dynamics and Control*. PMLR. 2021, pp. 1270–1280.

[15] Gautam Goel and Adam Wierman. "An online algorithm for smoothed regression and lqr control". In: *Proceedings of Machine Learning Research* 89 (2019), pp. 2504–2513.

[16] Gautam Goel et al. "Beyond online balanced descent: An optimal algorithm for smoothed online optimization". In: *Advances in Neural Information Processing Systems*. 2019, pp. 1875–1885.

[17] Aren Karapetyan et al. "Implications of Regret on Stability of Linear Dynamical Systems". In: *arXiv preprint arXiv:2211.07411* (2022).

[18] Parnian Kassraie and Andreas Krause. "Neural contextual bandits without regret". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, pp. 240–278.

[19] Taehwan Kim et al. "A decision tree framework for spatiotemporal sequence prediction". In: *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 2015, pp. 577–586.

[20] Minghong Lin et al. "Dynamic right-sizing for power-proportional data centers". In: *IEEE/ACM Transactions on Networking* 21.5 (2012), pp. 1378–1391.

[21] Yiheng Lin, Gautam Goel, and Adam Wierman. "Online optimization with predictions and non-convex losses". In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4.1 (2020), pp. 1–32.

[22] Thodoris Lykouris and Sergei Vassilvitskii. "Competitive caching with machine learned advice". In: *Journal of the ACM (JACM)* 68.4 (2021), pp. 1–25.

[23] Andrea Martin et al. "Safe control with minimal regret". In: *Learning for Dynamics and Control Conference*. PMLR. 2022, pp. 726–738.

[24] Edgar Minasyan et al. "Online control of unknown time-varying dynamical systems". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 15934–15945.

[25] Michael Mitzenmacher and Sergei Vassilvitskii. "Algorithms with predictions". In: *Communications of the ACM* 65.7 (2022), pp. 33–35.

[26] Deepan Muthirayan, Dileep Kalathil, and Pramod P Khargonekar. "Online Robust Control of Linear Dynamical Systems with Prediction". In: *arXiv preprint arXiv:2111.15063* (2021).

[27] Deepan Muthirayan and Pramod P Khargonekar. "Online Learning Robust Control of Nonlinear Dynamical Systems". In: *arXiv preprint arXiv:2106.04092* (2021).

[28] Marko Nonhoff and Matthias A Müller. "On the relation between dynamic regret and closed-loop stability". In: *arXiv preprint arXiv:2209.05964* (2022).

[29] Manish Purohit, Zoya Svitkina, and Ravi Kumar. "Improving online algorithms via ML predictions". In: *Advances in Neural Information Processing Systems* 31 (2018).

[30] Guanya Shi et al. "Neural lander: Stable drone landing control using learned dynamics". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 2019, pp. 9784–9790.

[31] Guanya Shi et al. "Online Optimization with Memory and Competitive Control". In: *arXiv e-prints* (2020), arXiv–2002.

[32] Max Simchowitz and Kevin G Jamieson. "Non-asymptotic gap-dependent regret bounds for tabular mdps". In: *Advances in Neural Information Processing Systems* 32 (2019).

[33] Andrew J Wagenmaker, Max Simchowitz, and Kevin Jamieson. "Beyond no regret: Instance-dependent pac reinforcement learning". In: *Conference on Learning Theory*. PMLR. 2022, pp. 358–418.

[34] Lijun Zhang et al. "Revisiting smoothed online learning". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 13599–13612.

[35] Dongruo Zhou, Lihong Li, and Quanquan Gu. "Neural contextual bandits with ucb-based exploration". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 11492–11502.

[36] Hongyu Zhou and Vasileios Tzoumas. "Safe Perception-Based Control with Minimal Worst-Case Dynamic Regret". In: *arXiv preprint arXiv:2208.08929* (2022).