"My research vision is to investigate sequential decision making, data science and signal processing to improve people's lives, establishing theoretical foundations for heuristics and offering faithful real-world solutions."

As scientific and engineering disciplines are increasingly driven by massive amounts of data, data-driven paradigms are gradually playing a critical role in expanding human life. One intrinsic bottleneck is that data acquisition can be expensive, time-consuming, or high-stakes due to safety and privacy concerns. As a result, *sample efficiency* becomes an imperative principle in method design, particularly in a multitude of *sample-starved* application scenarios such as clinics, robotics, and healthcare.

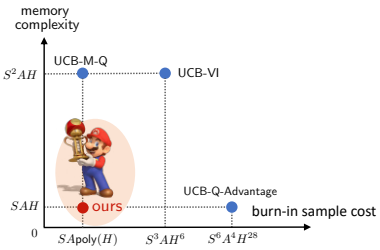
Apart from improving *sample efficiency*, designing effective solutions for real-world applications also requires circumventing other disparate challenges, such as data quality and task-specific constraints. Relying on the tools of high-dimensional statistics, large-scale optimization and machine learning, my research [1–13] thrives at tackling the aforementioned challenges to advance both theory and practice as follows:

1) **Sample-efficient reinforcement learning (RL):** breaking the sample complexity barriers provably and developing sample-efficient algorithms for a variety of RL formulations, including online, offline, robust, and curriculum RL.
2) **Data science and signal processing:** seeking theory-inspired, data-driven, and physics-driven algorithms for diverse signal and data processing real-world applications, in collaboration with civil engineering, high-performance computing, mechanical engineering and industry.

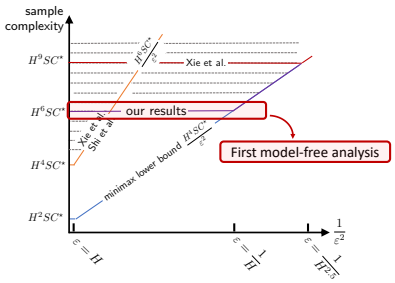
1 Sample-Efficient Reinforcement Learning

Reinforcement learning (RL) has witnessed a surge of practical success recently, with widespread applications in games, robotics, and control, financial investment, autonomous driving, etc [14]. Contemporary RL usually requires to sample from the environments with unprecedentedly large dimensionality and is eager to *sample efficiency*. Seeking provable sample-efficient algorithms is garnering a flurry of theoretical interest, following the finite-sample analysis statistical framework appearing very recently [15, 16]. Although researchers have made remarkable strides in advancing sample efficiency, there is still a sizable gap between the current state and the theoretical optimality. Extracting the underlying principles of RL, it is usually formulated as a finite-horizon (resp. infinite-horizon) Markov decision process with S states, A actions, and effective horizon length H (resp. $\frac{1}{1-\gamma}$), where S and A are finite but can be large.

Online RL: sample & memory efficiency with optimal regret [2]. Online RL refers to the problem of minimizing the metric *regret* by learning from the samples *sequentially* generated by interacting with the environment. While several competing solutions have achieved minimax-optimal regret, which scale on the order of $\sqrt{H^2SAT}$ with T the total number of samples, they are either memory-inefficient (with space complexity $O(S^2AH)$) [17, 18], or sample-inefficient (fall short of optimality unless the sample size exceeds an enormous burn-in cost (e.g., $S^6A^4\text{poly}(H)$) [19]. My research [2] is the first to achieve both memory efficiency and burn-in sample efficiency for this problem. Our model-free method with memory cost $O(SAH)$ yields optimal regret as soon as the sample size exceeds the order of $SA\text{poly}(H)$, which significantly narrows the sample complexity gap— by at least a factor of S^5A^3 — upon any prior memory-efficient algorithm that is asymptotically regret-optimal.

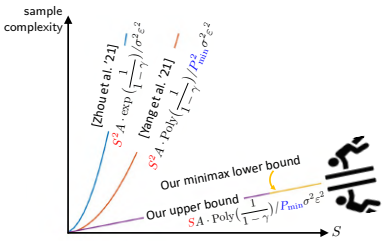


Offline RL: provable sample-efficient RL [3, 4] & practical solutions [13]. Offline/batch RL holds tremendous promise in learning a near-optimal (ε-optimal) decision making policy, resorting only to the existing history dataset without further exploration in the environment.



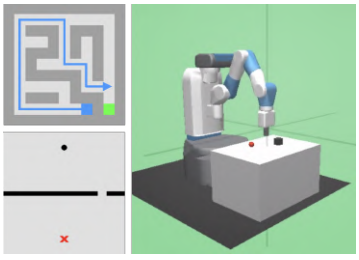
On the theory side, the current provable algorithm investigations have been confined almost exclusively to *model-based* approaches [20, 21] and only achieve optimal sample complexity performance in restricted cases when the desired accuracy level is extremely small $\varepsilon \in (0, 1/H^{2.5}]$. Note that *model-free* counterparts without requiring explicit model estimation are more flexible than *model-based* methods. My research [3] provides the first provable *model-free (Q-learning)* algorithm that achieves optimal sample complexity — in a much larger accuracy range $\varepsilon \in (0, 1/H]$ — comparing to the prior art (model-based RL [21] with $\varepsilon \in (0, 1/H^{2.5}]$) . In addition, by introducing a two-fold sampling technique, we also provide model-based methods [4] for both episodic/infinite-horizon offline RL which offer minimax optimal sample complexity for the entire accuracy range $\varepsilon \in (0, H]$.

On the practice side, the critical challenge of offline RL is the (potentially catastrophic) extrapolation error induced by the distribution shift between the history dataset and the desired policy. A large portion of prior work deals with this challenge by implicitly/explicitly regularizing the learning policy towards the behavior policy, which is difficult to estimate reliably in practice. Working with Google Brain, my research [13] regularizes towards the Q-function of the behavior policy instead of the behavior policy itself, under the premise that the Q-function can be estimated more reliably and easily and handles the extrapolation error more straightforwardly. Taking advantage of the estimated Q-function through slight regularizations, our methods exhibit strong performance on the offline D4RL benchmarks.



Robust RL: sample efficiency for robustness [1]. Test environments may deviate from the training environment induced by innumerable factors in the real world. In the face of uncertainty and high-risk, the requirement of robustness w.r.t. a huge group of potential test environments yields a pressing challenge of sample efficiency. Towards this, robust RL has attracted considerable amount of attention [22, 23], with the goal to learn a secure decision-maker that works well in the worst case of an uncertainty set (a ball) of different environments around an *ideal* nominal one. Focusing on offline robust RL using Kullback-Leibler (KL) divergence (widely used in practice) as the 'distance' for the set, existing competing solutions are extremely sample-inefficient: 1) **quality:** require a history dataset with full coverage of the entire nominal environment; 2) **quantity:** need sample size at least in the order of $O(S^2A\mathrm{poly}\left(\frac{1}{1-\gamma}\right))$ [24] and even with exponential dependency on the effective horizon length $\frac{1}{1-\gamma}$ (i.e., $O(S^2A\exp\left(\frac{1}{1-\gamma}\right))$ [25, 26].

My research [1] proposes a *pessimistic* algorithm [3, 4] to address the sample scarcity challenge which is further compounded by model uncertainty in robust RL. Our algorithms require only a partial coverage history dataset and significantly improve the sample complexity to $O(SA\mathrm{poly}\left(\frac{1}{1-\gamma}\right))$, which is only linear to the state space S. In addition, we provide the first theoretical limit of this problem — a minimax lower bound on sample complexity. The matching of the lower bound and upper bound up to a polynomial factor of the horizon length indicates that our algorithm is nearly optimal.

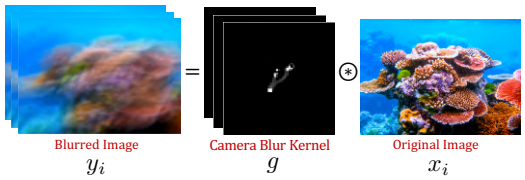


Curriculum RL: sample efficiency for generalization [12]. Sparse reward in numerous practical tasks (e.g., robotic manipulation) presents a daunting challenging environment for sample-efficient exploration. To address this, curriculum reinforcement learning (CRL) aims to create a sequence of tasks, generalizing from easy ones and gradually learning towards difficult tasks to improve sample efficiency. Prior works usually interpret the curriculum as shifting distributions with some heuristic 'distance' which poorly characterizes the inherent distance metric over the underlying manifold space of the tasks. Through collaboration with the mechanical engineering department, my research [12]

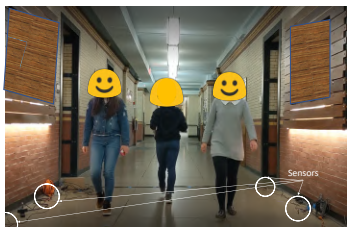
formulates CRL as an optimal transport problem to enable the usage of the inherent distance metric of the task. Armed with this, our method shows superior sample efficiency in practice and outperforms prior art, along with a theoretical guarantee analysis of empirical success.

2 Data Science and Signal Processing

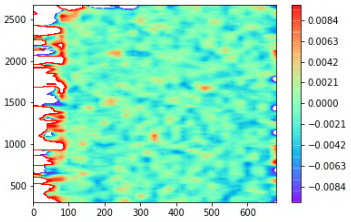
Nonconvex optimization for provable signal processing. Multi-channel sparse blind deconvolution (MSBD) aims to simultaneously recover a filter (g) and the unknown sparse inputs ($\{x_i\}$) to the filter from the observations of their convolution ($\{y_i\}$). This problem finds applications in understanding neural or seismic recordings [27, 28] as well as image debluring. which is notoriously challenging due to *shift and scaling ambiguities* and large dimensionality. My research [5] proposes a computationally efficient nonconvex optimization approach for MSBD based on simple manifold gradient descent (MGD) and provides theoretical guarantees for its global convergence. Based on the benign geometry of the problem, our proposed nonconvex approach provably solves this bilinear inverse problem with a significantly reduced sample complexity compared to prior art [29], both empirically and theoretically.



Vibration sensing for multiple occupant localization. Multiple occupant localization is a core problem in smart buildings, with various applications including elderly healthcare and energy management. While various approaches (e.g. mobile, vision, pressure sensors, etc.) have been explored for indoor occupant localization, we alternatively resort to vibration signals — simultaneously meet the desiderata of being sparsely-deployed, device-free, privacy-preserving, and robust to absence of line-of-sight. In collaboration with civil engineering — Prof. Pei Zhang (Umich), Prof. Hae-Young Noh (Stanford), and Prof. Shijia Pan (UC Merced), my research [30, 31] develops efficient algorithms for multiple occupants localization and tracking through footstep-induced floor vibrations. We conduct real-world experiments to evaluate the system and achieve sub-meter level accuracy, which enables potential indoor applications.



Mechanical strain measurement. Strain measurement of materials subjected to loadings or mechanical damages is an essential task in various industrial applications [32]. Digital image correlation (DIC) — a non-contact and non-interferometric optical technique for this problem — attracts a lot of attention. However, existing works still only support 2D planar objects/small surfaces due to the extremely high pixel resolution requirement of images. Together with Mitsubishi Electric Research Laboratories, my research [10, 11] provides an end-to-end fusion-based DIC framework that extends applicable scenarios to the curved surface of 3D objects in large scale, using a single moving camera.



Strain map

3 Future Plan

The significance of theoretical understanding and improving sample efficiency in RL, data science, and signal processing is far beyond what I studied here. To advance the fundamental theoretical study as well as empirical applications, my future research will include the following thrusts.

Towards principled and provable robust RL. As robustness is becoming a key principle in ubiquitous RL problems [33], there is a pressing need to enhance robustness in RL. While substantial progress has been made

in robust RL aforementioned, the current stage in both theory and practice is still far from maturity and theoretically optimality. Our techniques that achieved near-optimal sample complexity in offline robust RL [1] for KL divergence show tremendous promise to design optimal sample-efficient algorithms in numerous robust RL problems.

Provable sample efficiency in broader RL. Inspired by practical applications in computer science, machine learning and hardware, numerous meaningful RL problems are raised without solid theoretical underpinnings [14]. It is of interest to explore faithful/provable solutions for expansive RL problems: *risk-sensitive RL:* Risk-sensitive RL" is an appealing formulation for achieving safety and privacy in neuroscience, finance, and optimal control, referring to learning to act safely with different risks defined by different risk-measures [34]. Existing provable algorithms [35] for the risk-measure — exponential function are sample inefficient in online cases, calling for sample-efficient algorithm design for this problem and investigating various sampling mechanisms and risk-measures.

Empowering real-world applications In light of data's penetration into contemporary applications, my research aims to design principled data-driven methods for a more convenient and more appealing social life. I am eager to collaborate and explore promising application scenarios of RL, such as recommendation systems for social media, chip design for the electronics industry, and resource allocation for system design, etc. Moreover, my ongoing plan also involves the exploration of more application scenarios such as robot manipulation and the Internet of things (IoT).

References

Part A: My Preprints & Publications

[1] L. Shi and Y. Chi, "Distributionally robust model-based offline reinforcement learning with near-optimal sample complexity," *arXiv preprint arXiv:2208.05767*, 2022.

[2] G. Li, L. Shi, Y. Chen, Y. Gu, and Y. Chi, "Breaking the sample complexity barrier to regret-optimal model-free reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[3] L. Shi, G. Li, Y. Wei, Y. Chen, and Y. Chi, "Pessimistic Q-learning for offline reinforcement learning: Towards optimal sample complexity," *arXiv preprint arXiv:2202.13890*, 2022.

[4] G. Li, L. Shi, Y. Chen, Y. Chi, and Y. Wei, "Settling the sample complexity of model-based offline reinforcement learning," *arXiv preprint arXiv:2204.05275*, 2022.

[5] L. Shi and Y. Chi, "Manifold gradient descent solves multi-channel sparse blind deconvolution provably and efficiently," *IEEE Transactions on Information Theory*, vol. 67, no. 7, pp. 4784–4811, 2021.

[6] R. Chen, P. Huang, and L. Shi, "Latent goal allocation for multi-agent goal-conditioned self-supervised imitation learning," *NeurIPS Workshop on Bayesian Deep Learning*, 2021.

[7] Y. Sang, L. Shi, and Y. Liu, "Micro hand gesture recognition system using ultrasonic active sensing," *IEEE Access*, vol. 6, pp. 49 339–49 347, 2018.

[8] L. Shi, M. Mirshekari, J. Fagert, Y. Chi, H. Y. Noh, P. Zhang, and S. Pan, "Device-free multiple people localization through floor vibration," in *Proceedings of the 1st ACM International Workshop on Device-Free Human Sensing*. ACM, 2019, pp. 57–61.

[9] L. Shi, Y. Zhang, S. Pan, and Y. Chi, "Data quality-informed multiple occupant localization using floor vibration sensing," in *Proceedings of the Twenty-first International Workshop on Mobile Computing Systems and Applications*. ACM, 2020.

[10] L. Shi, D. Liu, M. Umeda, and N. Hana, "Fusion-based digital image correlation framework for strain measurement," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1400–1404.

[11] L. Shi, D. Liu, and J. Thornton, "Robust camera pose estimation for image stitching," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 2838–2842.

[12] P. Huang, M. Xu, J. Zhu, L. Shi, F. Fang, and D. Zhao, "Curriculum reinforcement learning using optimal transport via gradual domain adaptation," *Advances in Neural Information Processing Systems*, 2022.

[13] L. Shi, R. Dadashi, Y. Chi, P. S. Castro, and M. Geist, "Offline reinforcement learning with on-policy Q-function regularization," *preprint*, 2023.

Part B: Other Publications

[14] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A brief survey of deep reinforcement learning," *arXiv preprint arXiv:1708.05866*, 2017.

[15] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018, vol. 47.

[16] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press, 2019, vol. 48.

[17] P. Ménard, O. D. Domingues, X. Shang, and M. Valko, "Ucb momentum q-learning: Correcting the bias without forgetting," in *International Conference on Machine Learning*. PMLR, 2021, pp. 7609–7618.

[18] M. G. Azar, I. Osband, and R. Munos, "Minimax regret bounds for reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2017, pp. 263–272.

[19] Z. Zhang, Y. Zhou, and X. Ji, "Almost optimal model-free reinforcement learningvia reference-advantage decomposition," *Advances in Neural Information Processing Systems*, vol. 33, pp. 15 198–15 207, 2020.

[20] P. Rashidinejad, B. Zhu, C. Ma, J. Jiao, and S. Russell, "Bridging offline reinforcement learning and imitation learning: A tale of pessimism," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[21] T. Xie, N. Jiang, H. Wang, C. Xiong, and Y. Bai, "Policy finetuning: Bridging sample-efficient offline and online reinforcement learning," *Advances in neural information processing systems*, vol. 34, 2021.

[22] G. N. Iyengar, "Robust dynamic programming," *Mathematics of Operations Research*, vol. 30, no. 2, pp. 257–280, 2005.

[23] H. Xu and S. Mannor, "Distributionally robust Markov decision processes," *Mathematics of Operations Research*, vol. 37, no. 2, pp. 288–300, 2012.

[24] W. Yang, L. Zhang, and Z. Zhang, "Towards theoretical understandings of robust markov decision processes: Sample complexity and asymptotics," *arXiv preprint arXiv:2105.03863*, 2021.

[25] Z. Zhou, Q. Bai, Z. Zhou, L. Qiu, J. Blanchet, and P. Glynn, "Finite-sample regret bound for distributionally robust offline tabular reinforcement learning," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 3331–3339.

[26] K. Panaganti and D. Kalathil, "Sample complexity of robust reinforcement learning with a generative model," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 9582–9602.

[27] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE transactions on signal processing*, vol. 59, no. 10, pp. 4735–4744, 2011.

[28] D. Donoho, "On minimum entropy deconvolution," in *Applied time series analysis II*. Elsevier, 1981, pp. 565–608.

[29] Y. Li and Y. Bresler, "Global geometry of multichannel sparse blind deconvolution on the sphere," in *Advances in Neural Information Processing Systems*, 2018, pp. 1132–1143.

[30] S. Pan, K. Lyons, M. Mirshekari, H. Y. Noh, and P. Zhang, "Multiple pedestrian tracking through ambient structural vibration sensing," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. ACM, 2016, pp. 366–367.

[31] M. Mirshekari, S. Pan, J. Fagert, E. M. Schooler, P. Zhang, and H. Y. Noh, "Occupant localization using footstep-induced structural vibration," *Mechanical Systems and Signal Processing*, vol. 112, pp. 77–97, 2018.

[32] B. Pan, K. Qian, H. Xie, and A. Asundi, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review," *Measurement science and technology*, vol. 20, no. 6, p. 062001, 2009.

[33] J. Moos, K. Hansel, H. Abdulsamad, S. Stark, D. Clever, and J. Peters, "Robust reinforcement learning: A review of foundations and recent advances," *Machine Learning and Knowledge Extraction*, vol. 4, no. 1, pp. 276–315, 2022.

[34] O. Mihatsch and R. Neuneier, "Risk-sensitive reinforcement learning," *Machine learning*, vol. 49, no. 2, pp. 267–290, 2002.

[35] L. Hou, L. Pang, X. Hong, Y. Lan, Z. Ma, and D. Yin, "Robust reinforcement learning with wasserstein constraint," *arXiv preprint arXiv:2006.00945*, 2020.