# Deep Q-Learning on Hölder Spaces

**Qian Qi** QIQIAN@PKU.EDU.CN
*School of Computer Science, Peking University*

**Editors:** Steve Hanneke and Tor Lattimore

## Abstract

We study the operator-theoretic core of Q-learning in continuous-time stochastic control with continuous states and actions. In value-based reinforcement learning, each Q-learning or DQN update is built from a Bellman optimality target; our analysis isolates this target in a uniformly elliptic diffusion setting and studies its regularity and approximation complexity. Under Hölder-regular coefficients, we show that a Bellman update maps bounded inputs into an anisotropic regularity class: it smooths the state variable through parabolic regularization while preserving only Lipschitz dependence on the action variable. This identifies a compact family of Bellman iterates and motivates tensor-product neural-operator approximators adapted to the mixed regularity of the problem. We derive explicit approximation and resource bounds, including a stiffness–complexity trade-off as the time step $\delta \to 0$. The result is an operator-level theory for the Bellman targets underlying Q-learning in continuous stochastic control, rather than a convergence theorem for practical sampled DQN training.

**Keywords:** Q-learning theory, Bellman value iteration, neural operators, parabolic smoothing, stochastic optimal control, approximation theory.

Q-learning and its deep variants are built around repeated approximation of Bellman optimality targets (Watkins and Dayan, 1992; Mnih et al., 2015; Fan et al., 2020). In continuous state and action spaces, however, the target is not an arbitrary function: in diffusion control it is tied to a parabolic Bellman equation (Krylov, 1987; Fleming and Soner, 2006). Building on prior operator-level analyses of deep Q-network approximation and viscosity convergence (Qi, 2025, 2026), this extended abstract summarizes a regularity and approximation theory for these idealized Bellman targets. The full version contains the Schauder estimates, truncation arguments, compactness proof, and neural-operator approximation theorem.

We consider a finite-horizon controlled diffusion on $\mathbb{R}^d$ with compact action space $\mathcal{A}$. For a time step $\delta$, the Bellman update $\mathcal{T}_\delta$ combines a maximization over actions with an expectation over a short stochastic transition. The maximization step can create nonsmooth policy-switching boundaries, while the nondegenerate diffusion step smooths the resulting value function. Under uniform ellipticity and Hölder-regular coefficients, we show that a Bellman step maps bounded inputs into an anisotropic class, informally $C^{2,\alpha}_x C^{0,1}_a$: smooth in state, but only Lipschitz in action.

This mixed regularity yields a compact family of Bellman iterates and motivates tensor-product DeepONet-style approximators (Lu et al., 2021; Yarotsky, 2017). The resulting bounds separate the smoothed state dimension from the action dimension and track the dependence on accuracy $\varepsilon$, state dimension $d$, action dimension $d_a$, and time step $\delta$. As $\delta \to 0$, the Bellman operator approaches the identity, so parabolic smoothing weakens and the approximation problem becomes stiff. Our analysis identifies this stiffness–complexity trade-off and the associated diffusion-length-scale resolution requirement. The result is an operator-level theory for ideal Bellman targets, not a finite-sample or optimization theorem for practical DQN with exploration, replay buffers, target networks, and stochastic gradient updates.

∗. Extended abstract. Full version appears as arXiv:2606.16846v1.

## References

Jianqing Fan, Zhaoran Wang, Yuchen Xie, and Zhuoran Yang. A theoretical analysis of deep q-learning. In *Learning for dynamics and control*, pages 486–489. PMLR, 2020.

Wendell H Fleming and Halil M Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.

Nikolai Vladimirovich Krylov. *Nonlinear elliptic and parabolic equations of the second order*. Springer, 1987.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

Qian Qi. Universal approximation theorem of deep q-networks. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 50024–50046. PMLR, 2025.

Qian Qi. Viscosity convergence analysis for deep q-networks. Revision submitted to the Journal of Machine Learning Research, 2026.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

Dmitry Yarotsky. Error bounds for approximations with deep relu networks. *Neural networks*, 94: 103–114, 2017.