# Article information:

[2008.03559] Convex Q-Learning, Part 1: Deterministic Optimal Control  
<https://arxiv.org/abs/2008.03559>

# Article summary:

1. The extension of Watkins' algorithm to general function approximation settings is challenging, and the projected Bellman equation may not have a solution that generates a good policy.

2. A class of convex Q-learning algorithms is introduced based on the convex relaxation of the Bellman equation, with convergence established under general conditions including linear function approximation for the Q-function.

3. The batch implementation of convex Q-learning appears similar to DQN, but the two algorithms are very different in their approach and theoretical foundations.

# Article rating:

Appears moderately imbalanced: The article provides some useful information, but is missing several important points or pieces of evidence that would be required to present the discussed topics in a balanced and reliable way. You are encouraged to seek a more balanced perspective on the presented issues by exploring the provided research topics and looking at different information sources.

# Article analysis:

The article "Convex Q-Learning, Part 1: Deterministic Optimal Control" presents a new class of convex Q-learning algorithms based on the convex relaxation of the Bellman equation. The paper begins with a brief survey of linear programming approaches to optimal control, leading to a particular over parameterization that lends itself to applications in reinforcement learning.

The article is well-written and provides a clear overview of the problem at hand. The authors acknowledge the challenges associated with extending Watkins' algorithm to general function approximation settings and highlight the paradoxical nature of this challenge given the long history of convex analytic approaches to dynamic programming.

The main conclusions drawn from the study are that (i) convergence is established under general conditions, including a linear function approximation for the Q-function, and (ii) a batch implementation appears similar to the famed DQN algorithm. However, it is shown that these two algorithms are very different in terms of their theoretical foundations.

One potential bias in this article is that it focuses solely on deterministic nonlinear systems with total cost criterion. While many extensions are proposed, including kernel implementation and extension to MDP models, there is no discussion of how these results might apply in other contexts or domains.

Another potential issue with this article is that it does not present both sides equally. While the authors acknowledge some of the challenges associated with extending Watkins' algorithm to general function approximation settings, they do not provide an in-depth analysis of why these challenges exist or what alternative approaches might be available.

Overall, "Convex Q-Learning, Part 1: Deterministic Optimal Control" provides valuable insights into a challenging problem in reinforcement learning. However, readers should be aware of its potential biases and limitations when interpreting its findings.

# Topics for further research:

* Nonlinear systems with stochastic dynamics in reinforcement learning
* Challenges of extending Watkins' algorithm to general function approximation settings
* Alternative approaches to Q-learning in reinforcement learning
* Convex analytic approaches to dynamic programming in reinforcement learning
* Applications of convex relaxation in reinforcement learning beyond optimal control
* Comparison of batch implementation of convex Q-learning and DQN algorithm in reinforcement learning

# Report location:

<https://www.fullpicture.app/item/cf32e6f153713143757adf8fd1696521>