



HAL
open science

Tabular and Deep Learning of Whittle Index

Francisco Robledo, Urtzi Ayesta, Konstantin Avrachenkov, Vivek S Borkar

► **To cite this version:**

Francisco Robledo, Urtzi Ayesta, Konstantin Avrachenkov, Vivek S Borkar. Tabular and Deep Learning of Whittle Index. EWRL 2022 - 15th European Workshop on Reinforcement Learning, Sep 2022, Milan, Italy. hal-03810695

HAL Id: hal-03810695

<https://univ-pau.hal.science/hal-03810695>

Submitted on 11 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introduction

- Whittle index policy is an asymptotically optimal heuristic for solving Restless Multi-Armed Bandit Problems (RMBAP).
- We propose two algorithms, QWI and QWINN, for the computation of such indices.
- Both employ a two time-scale system for the computation of the indices and the Q-values of each state/action.

Motivation

Asymptotically Optimal Heuristics for Restless Multi-Armed Bandit Problems (RMABP)

- Load-balancing problems
- Machine maintenance
- Health-care systems

QWI/QWINN Algorithm

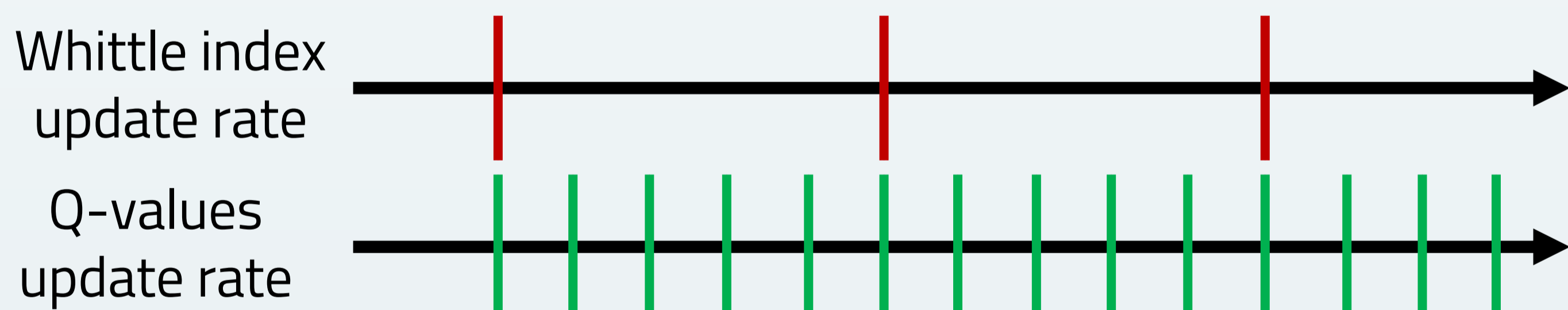
Algorithm:

Evaluate the **Q-values** for a given state/action (s_n, a_n) using the index for all states $x \in \mathcal{S}$

$$\begin{aligned} Q_{n+1}^x(s_n, a_n) &= (1 - \alpha(n))Q_{n+1}^x(s_n, a_n) \\ &+ \alpha(n) \left((1 - a_n)(r_0(s_n) + \lambda_n(x)) + a_n r_1(s_n) \right) \\ &+ \gamma \max_{v \in \{0,1\}} Q_n^x(s_{n+1}, v) \end{aligned}$$

Update **Whittle index** for all states $x \in \mathcal{S}$

$$\lambda_{n+1}(x) = \lambda_n(x) + \beta(n) + (Q_n^x(x, 1) - Q_n^x(x, 0))$$



Two time-scales set through learning rate

- **Fast time-scale:** Q-values

$$\alpha(n): \sum_n \alpha(n) = \infty, \quad \sum_n \alpha(n)^2 < \infty$$

- **Slow time-scale:** Whittle index

$$\beta(n): \sum_n \beta(n) = \infty, \quad \sum_n \beta(n)^2 < \infty, \quad \beta(n) = o(\alpha(n))$$

Neural Network implementation (QWINN):

- Hybrid system:
 - Neural network to approximate Q-values

$$\begin{aligned} Q_{target}^x(s_n, a_n) &= (1 - a_n)(r_0(s_n) + \lambda_n(x)) + a_n r_1(s_n) \\ &+ \gamma \max_{v \in \mathcal{A}} Q_{\theta'}^x(s_{n+1}, v) \end{aligned}$$

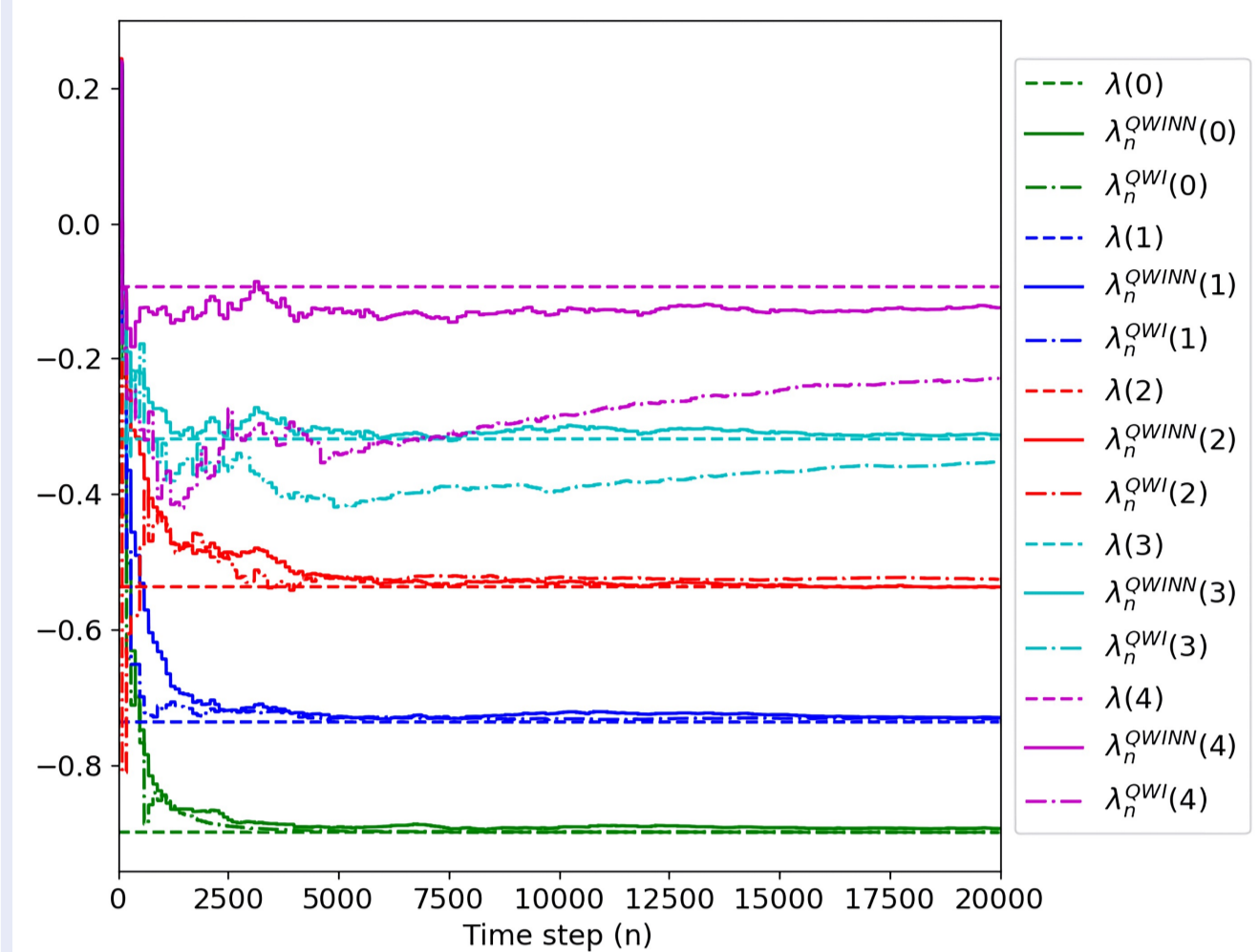
- Tabular computation for Whittle index

$$\lambda_{\theta, n+1}(x) = \lambda_{\theta, n}(x) + \beta(n) \left(Q_{\theta, n}^x(x, 1) - Q_{\theta, n}^x(x, 0) \right)$$

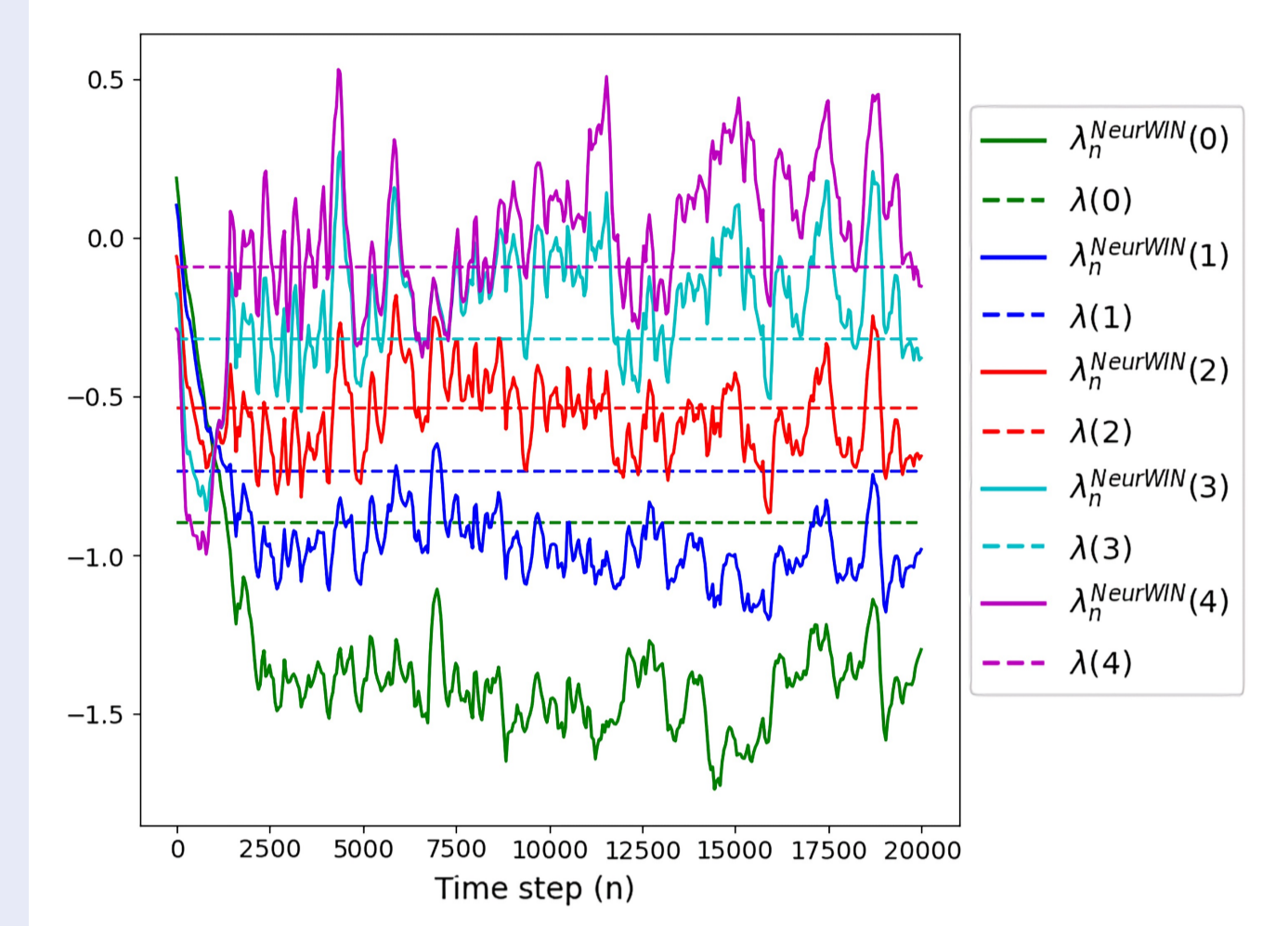
Results

Restart problem

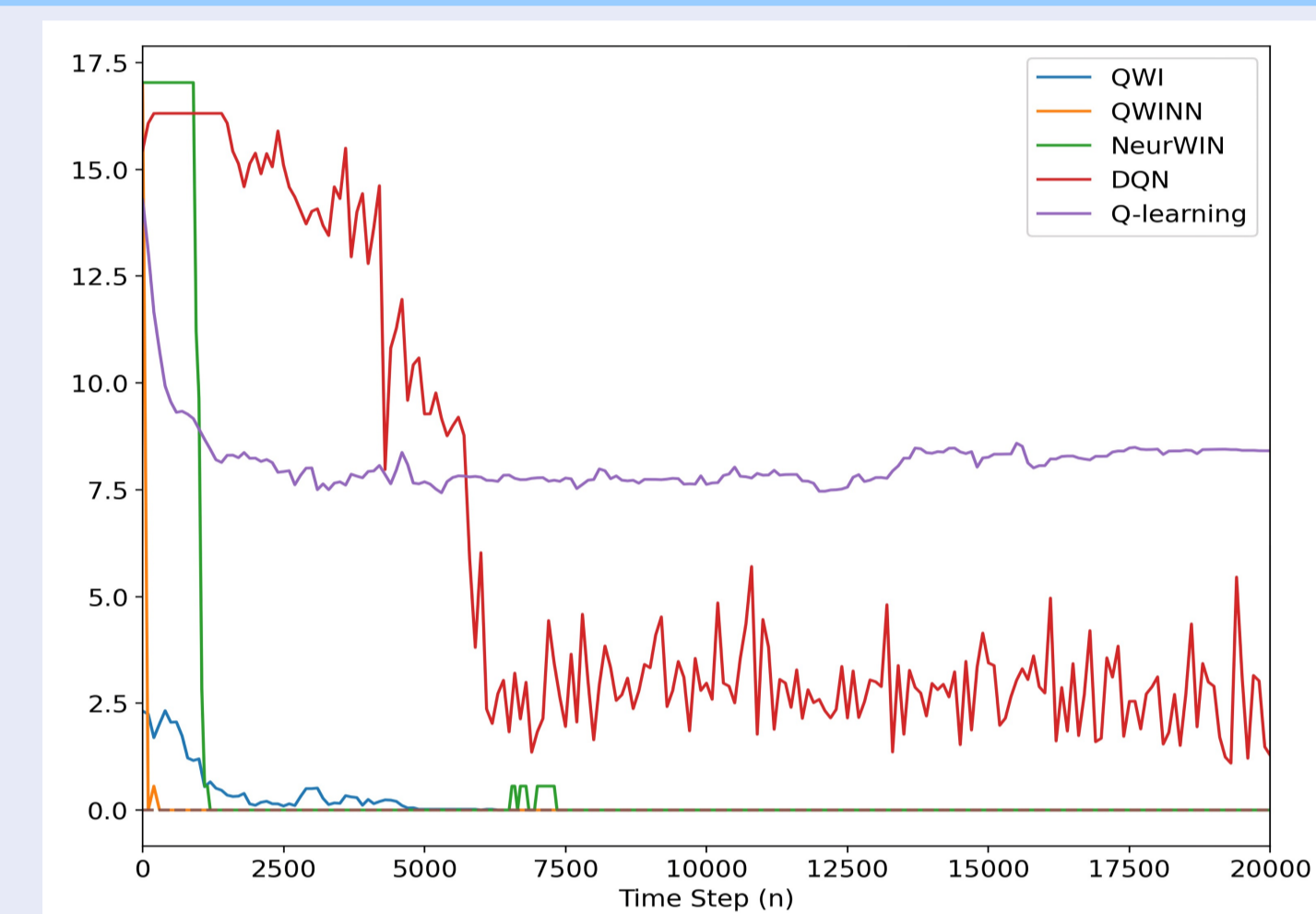
QWI/QWINN indices



NeurWIN indices [3]

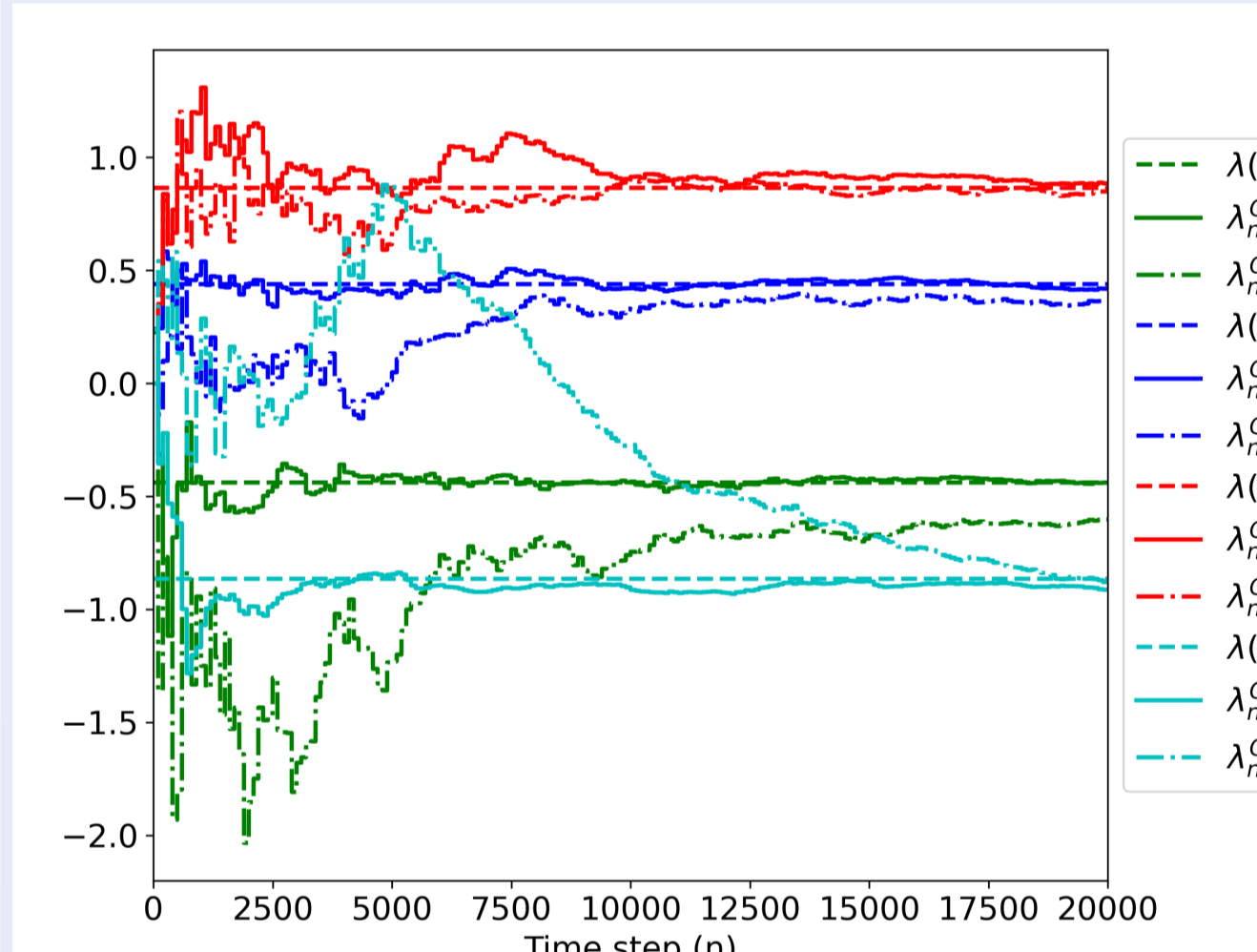


Bellman Relative Error (lower is better)

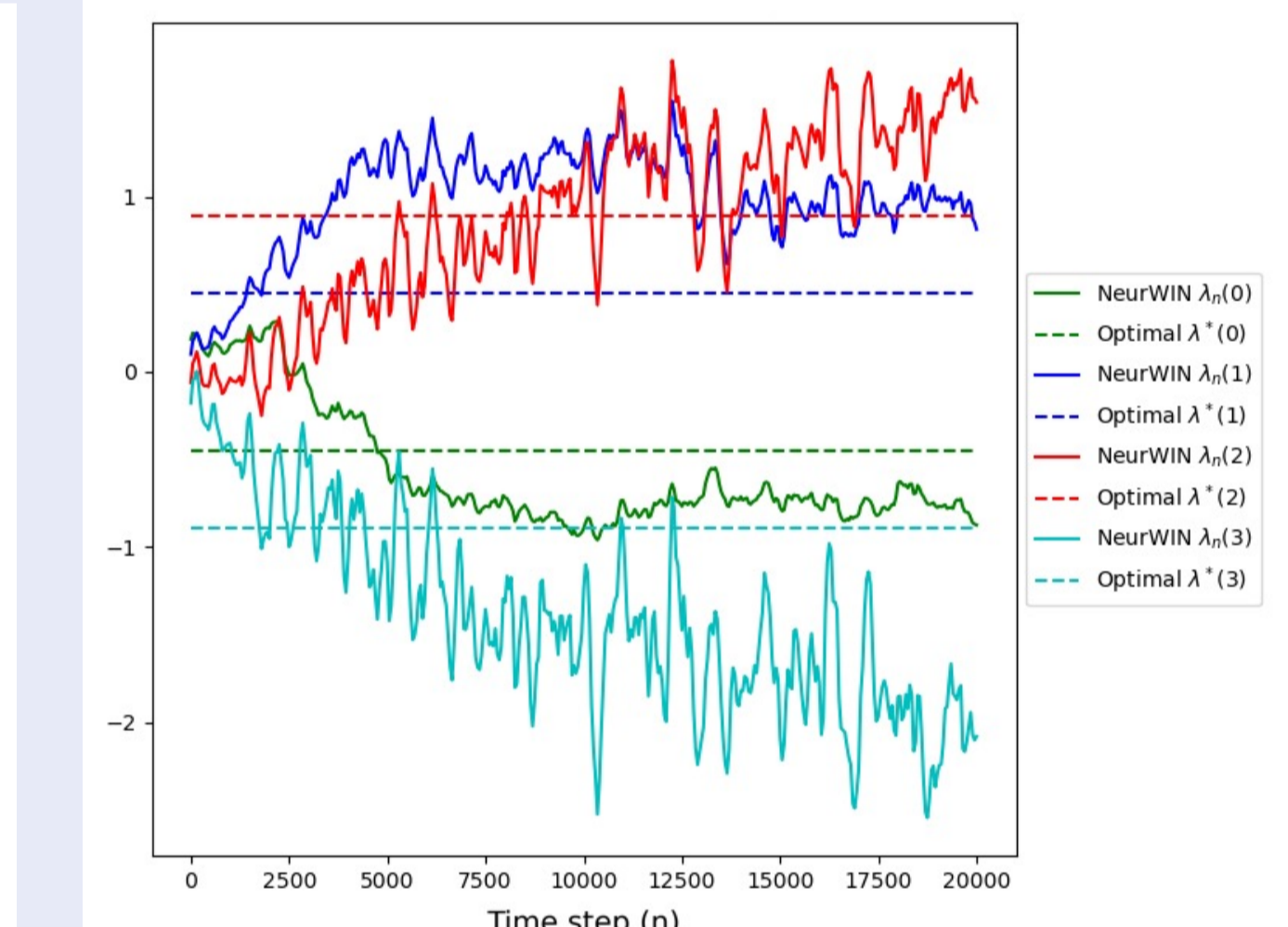


Circular problem

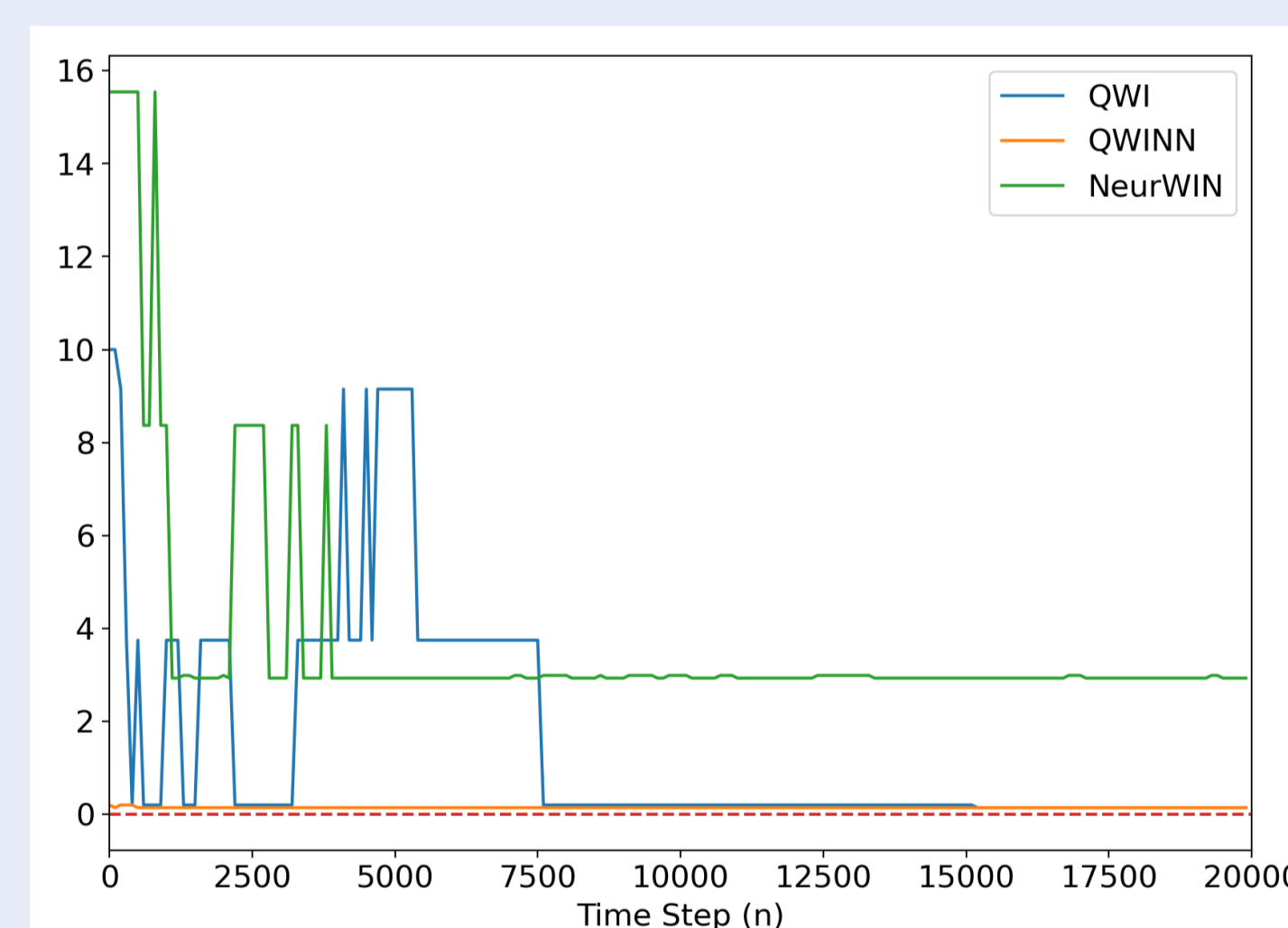
QWI/QWINN indices



NeurWIN indices [3]



Bellman Relative Error (lower is better)



Conclusion

- Two time-scale implementation to decouple Q-learning and Whittle index updates
- Fast and stable convergence to theoretical Whittle index
- Neural Network implementation improves performance in underexplored states

References

1. Abounadi, Jinane / Bertsekas, Dimitris / Borkar, Vivek S., Learning algorithms for Markov decision processes with average cost, 2001
2. Avrachenkov, Konstantin E. / Borkar, Vivek S., Whittle index based Q-learning for restless bandits with average reward, 2022-05, Automatica, Vol. 139, p. 110186
3. Nakhleh, Khaled / Ganji, Santosh / Hsieh, Ping-Chun / Hou, I.-Hong / Shakkottai, Srinivas, NeurWIN: Neural Whittle Index Network For Restless Bandits Via Deep RL, 2021