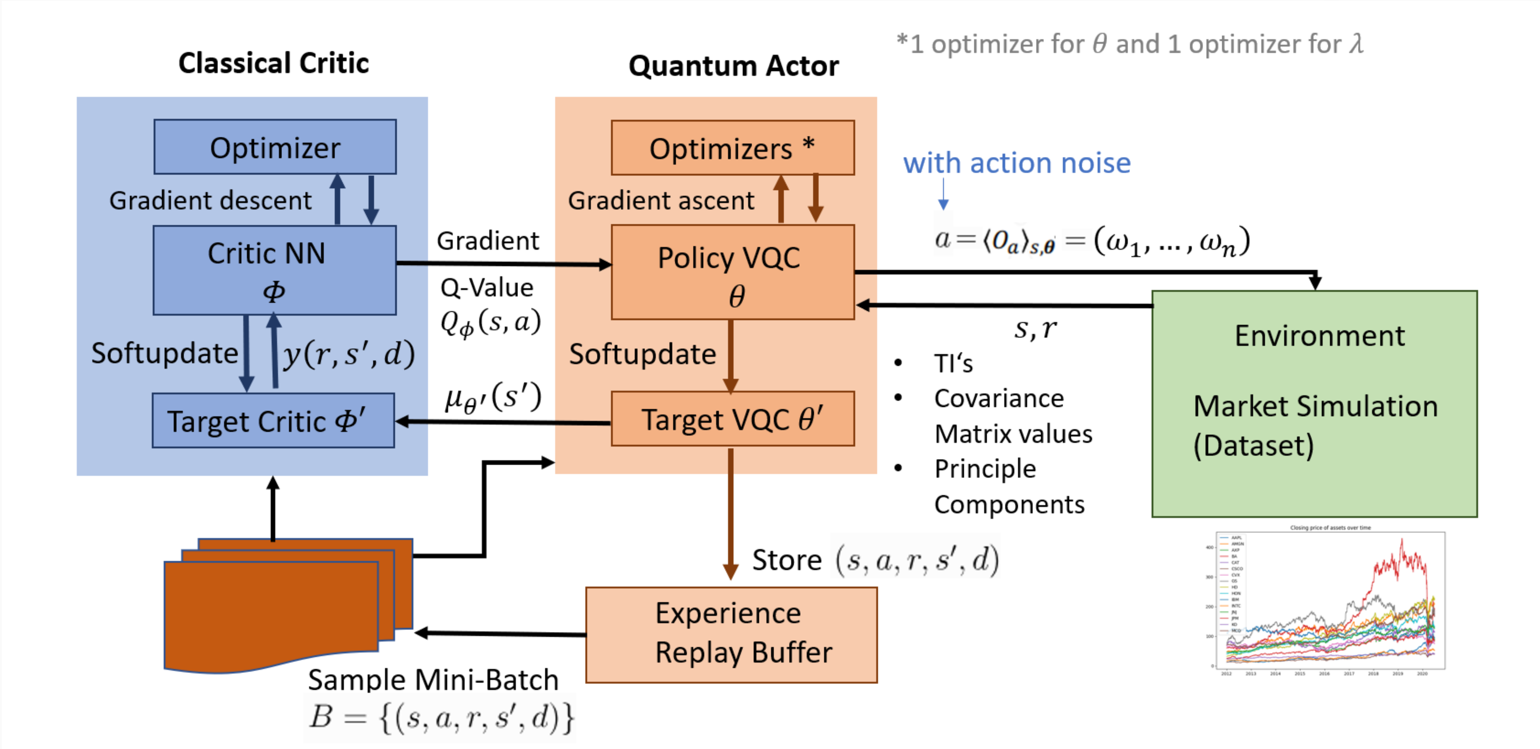
Exploring the impact of noise on Quantum DDPG in portfolio allocation

Annette Zapf, Sabine Wölk
Institute of Quantum Technologies, German Aerospace Center (DLR), Ulm

Motivation

In the era of NISQ devices, Variational Quantum Circuits in Quantum Machine Learning are gaining attention, advancing towards practical quantum computing applications on NISQ devices. Reinforcement Learning (RL), known for its human-like, trial-and-error learning, is inherently suited for dynamic financial applications that require adaptability [1,2]. Classical deep RL models like DDPG and PPO show promise, while emerging quantum neural networks offer potential for improved function approximation, better generalization capabilities and reduced parameters [3]. In light of these advancements, we explored a quantum-enhanced version of the DDPG agent, aiming to leverage these quantum capabilities for more efficient financial decision-making processes. Our objective is to explore the practicality and potential benefits of QRL in finance, aiming to realize viable quantum computing applications on NISQ devices.

Quantum DDPG Agent

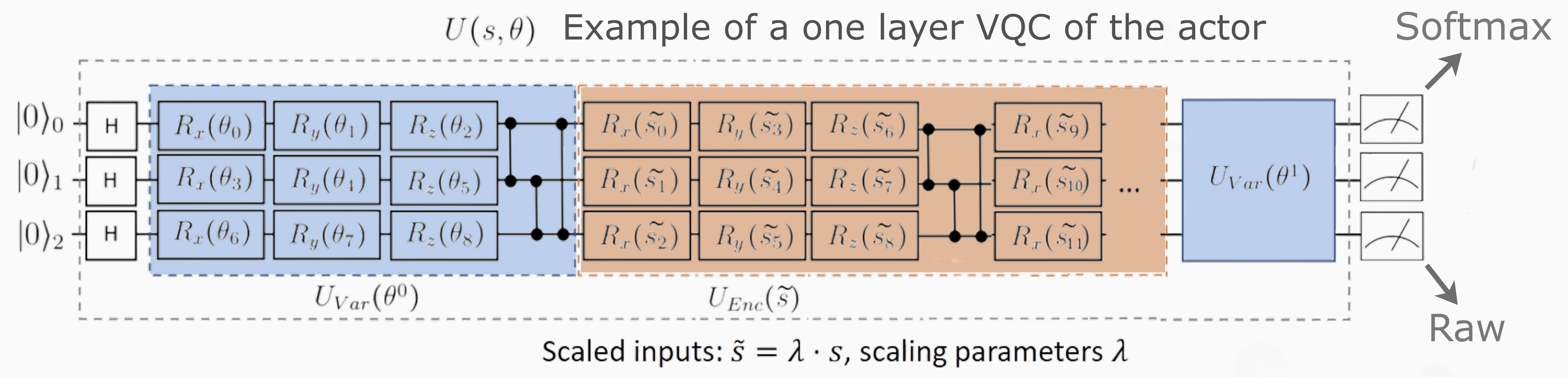


Actor-Critic model (Q-Learning) for continous actions and percepts with tartet networks to stabilize training. Off-policy model with replay buffer.

Update routine of DDPG [4]:

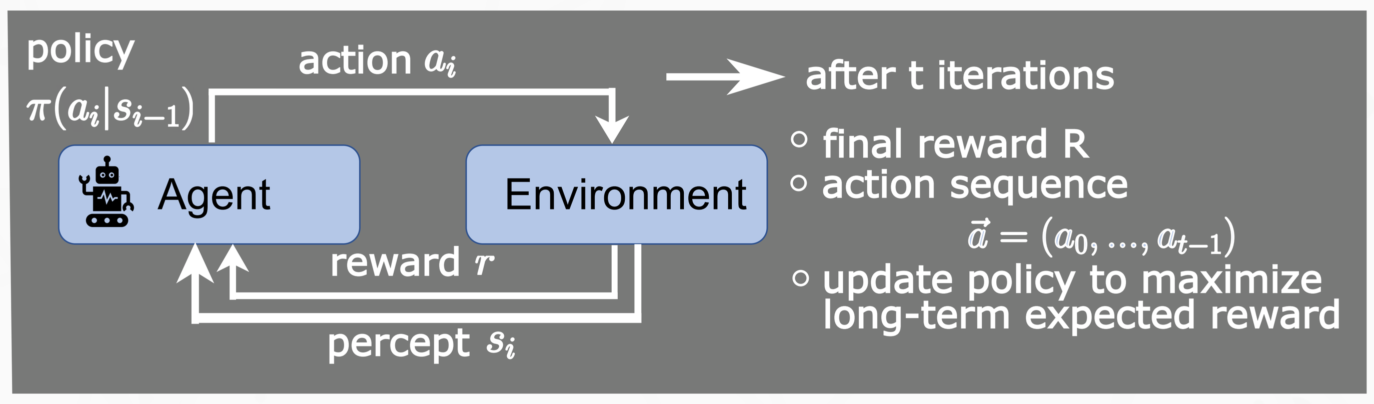
1. Randomly sample batch B from buffer
2. Compute targets $y(r,s',d) = r + \gamma(1-d)Q_{\phi_{targ}(s',\mu_{\Theta_{targ}}(s'))}$
3. Update Q-function (gradient descent)
$$\frac{\nabla_\phi}{|\mathcal{B}|}\sum_{(s,a,r,s',d)\in\mathcal{B}}\left(Q_\phi(s,a) - y(r,s',d)\right)^2$$
4. Update policy (gradient ascent)
$$\frac{\nabla_\theta}{|\mathcal{B}|}\sum_{s\in\mathcal{B}} Q_\phi(s,\mu_\theta(s))$$
5. Update target networks $\phi_{\text{targ}} \leftarrow \rho\phi_{\text{targ}} + (1-\rho)\phi$
$\theta_{\text{targ}} \leftarrow \rho\theta_{\text{targ}} + (1-\rho)\theta$

$U(s,\theta)$ Example of a one layer VQC of the actor



Scaled inputs: $\tilde{s} = \lambda \cdot s$, scaling parameters λ

Reinforcement Learning



Variational Quantum Policies

Softmax policy

$$\pi_\theta(a|s) = \frac{e^{\beta\langle O_a\rangle_{s,\theta}}}{\sum_{a'} e^{\beta\langle O_{a'}\rangle_{s,\theta}}} \text{ with}$$

$$\langle O_a\rangle_{s,\theta} = \left\langle \psi_{s,\theta}\middle|\sum_i w_{a,i}H_{a,i}\middle|\psi_{s,\theta}\right\rangle.$$

$w_{a,i}H_{a,i}$ weighted Hermitian operators associated to action a.

Raw policy

$$\pi_\theta(a|s) = \langle P_a\rangle_{s,\theta}$$

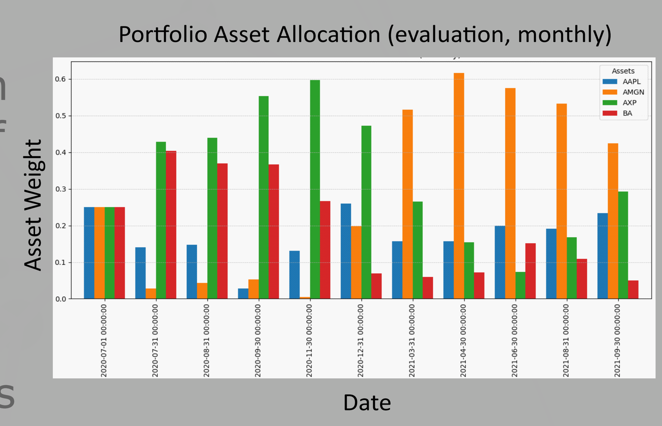
with projection P_a associated to action a, $\sum_a P_a = I$ and $P_a P_{a'} = \delta_{a,a'}$.
θ all trainable parameters.

[5]

Portfolio Allocation

A strategic approach to distribute investment capital across various assets to maximize returns and minimize risk.
The QRL agent dynamically adjusts the optimal allocation weight vector daily, responding to market fluctuations for effective time series optimization and meeting investment objectives.

- **Variations (constraints):** Risk minimization, transaction costs, diversity, restrictions of weight distribution.
- **Task Complexity:** From linear problem to NP-complete based on constraints and number of assets.



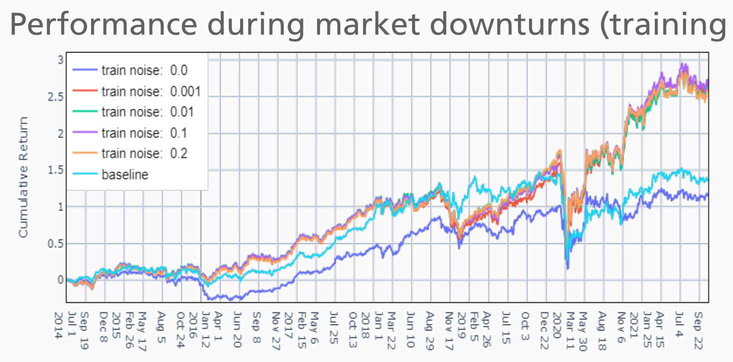
Methodology:

- Environment: market simulation from historical data
- Split of training period and evaluation period
- Features: Technical Indicators, values of covariance matrix or principle components of PCA
- Reward function: Portfolio value

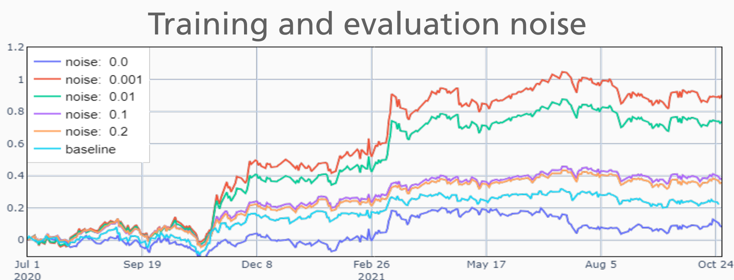
Results

Softmax (noisy)

a.) Performance during market downturns (training noise)



b.) Training and evaluation noise



c.) Training noise

Raw (noiseless)

d.)

d.) Average cumulative returns of 15 Q-DDPG agents with raw policy on an evaluation dataset are compared to a baseline resembling the underlying market dynamics and the Min-Variance strategy. Sharpe ratios: Q-DDPG 1.20, Min-Variance 0.96, baseline 0.81, indicating superior balance of returns and risks.

a.)-c.) Exemplary performance of Q-DDPG agents with softmax policy under depolarizing noise (seed 43). Demonstrating how noise during training can prevent suboptimal policies, while evaluation noise converges to market baseline with increasing noise levels.

Conclusion

Performance Benefits: Achieves favorable return-to-risk ratios compared to conventional Min-Variance strategy and baseline.

Reduced Parameters: Requires fewer trainable parameters than traditional deep RL methods.

Scalability and Efficiency: Demonstrates scalability in asset counts $N_{Assets} = 2^{N_{Qubits}}$ with gate counts $N_{\text{Gates}} \propto (\log_2(N_{Assets}))^2$, supporting realistic portfolio sizes.

Noise Resilience: Evaluation shows robustness to depolarizing, amplitude damping, phase damping, measurement, and shot noise while training. However, noise during evaluation can be detrimental, particularly at higher noise levels.
Specific noise types, such as depolarizing, have similar effects to classical hyperparameters, enhancing learning in particular scenarios.

[1] AVinay Singh et al., "How are reinforcement learning and deep learning algorithms used for big data based decision making", International Journal of Information Management Data Insights,vol. 2, 100094 (2022).
[2] Xiao-Yang Liu et al., "FinRL: deep reinforcement learning framework to automate trading in quantitative finance", In Proceedings of the Second ACM International Conference on AI in Finance (ICAIF '21), New York, 1-9 (2022).
[3] Abbas, A. et al., "The power of quantum neural networks", Nat Comput Sci 1, 403–409 (2021).
[4] Lillicrap, T.P. et al., "Continuous control with deep reinforcement learning", Proceedings of the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico (2016).
[5] Jerbi, S. et al., "Parametrized Quantum Policies for Reinforcement Learning", Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS), virtual (2021).

This work contributes to the QuGov project of the Federal Ministry of Finance (BMF) and the Bundesdruckerei in cooperation with the University of Ulm and DLR-QT.

DLR Deutsches Zentrum für Luft- und Raumfahrt
Institut für Quantentechnologien
annette.mueller@dlr.de