On Dequantization of Supervised Quantum Machine Leaning via Random Fourier Features

Mehrad Sahebi, Alice Barthe, Yudai Suzuki, Zoe Holmes, Michele Grossi

TEAM

On Dequantization of Supervised Quantum Machine Learning via Random Fourier Features

Mehrad Sahebi,[1,2] Alice Barthe,[3,4] Yudai Suzuki,[1,2,5] Zoë Holmes,[1,2] and Michele Grossi[4]

[1] *Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

[2] *Center for Quantum Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

[3] $\langle aQa^L \rangle$ *Applied Quantum Algorithms, Universiteit Leiden, Leiden, the Netherlands*

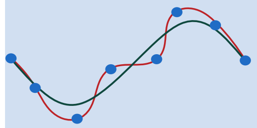
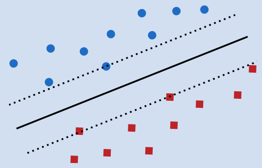
[4] *Quantum Technology Initiative, CERN, Geneva, Switzerland*

[5] *Quantum Computing Center, Keio University, Hiyoshi 3-14-1, Kohoku-ku, Yokohama 223-8522, Japan*

Are Quantum Computers Useful for ML?

Dequantization: To find efficient classical ML algorithms that work as well as quantum algorithms (in terms of true risk)

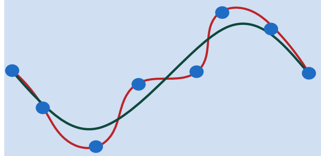
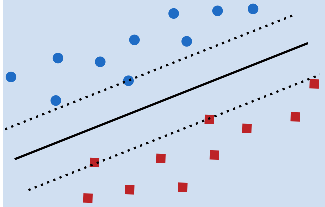
QML models have Fourier representations! → RFF (Landman et al.)

Problem	QNN	QK
Regression 	(Sweke et al.)	??
SVM 	??	??

Landman, Jonas, et al. "Classically approximating variational quantum machine learning with random Fourier features." *arXiv preprint arXiv:2210.13200* (2022).

Sweke, Ryan, et al. "Potential and limitations of random Fourier features for dequantizing quantum machine learning." *Quantum* 9 (2025): 1640.

Summary of Results

Problem	**QNN**	**QK**
Regression 	▫ **Alignment:** $p_\omega \propto \|c^\star_\omega\|$. ● **Concentration:** $p_{max} \in \mathcal{O}(poly(d))$	▫ **Alignment:** $\left\Vert \sqrt{P}^{-1}\sqrt{Q} \right\Vert_\infty \in \mathcal{O}(poly(d))$ ● **Concentration:** $p_{max} \in \mathcal{O}(poly(d))$ ■ **Bounded RKHS norm:** $\left\Vert f_q \right\Vert_{k_Q} \in \mathcal{O}(poly(d))$
SVM 	▫ **Alignment:** $p_\omega \propto \|c^\star_\omega\|$ ■ **Bounded Fourier Sum:** $\sum\|c_\omega\| \in \mathcal{O}(poly(d))$	▫ **Alignment:** $p_\omega = \frac{1}{C}\sqrt{q_\omega}$ ● **Concentration:** $C = \sum_{\omega \in \Omega} \sqrt{q_\omega} \in \mathcal{O}(poly(d))$ ■ **Bounded RKHS norm:** $\left\Vert f_q \right\Vert_{k_Q} \in \mathcal{O}(poly(d))$

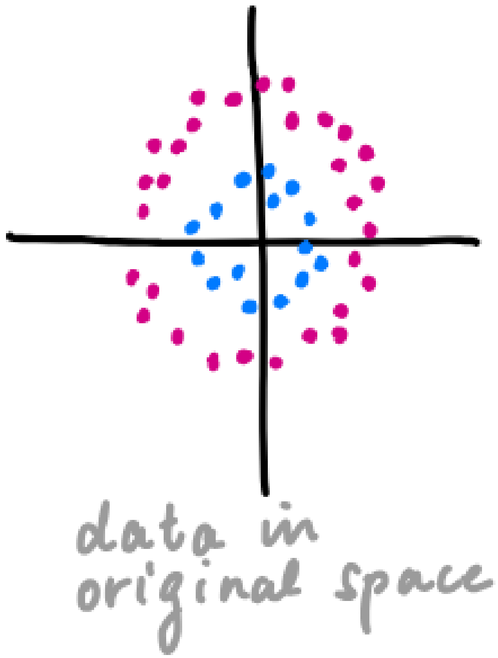
Outline

- Background
- Main Results
  - Approximation of Quantum Kernels
  - RFF-Dequantization QML
- Numerical Experiments
- Conclusion

[1] Sweke, Ryan, et al. "Potential and limitations of random Fourier features for dequantizing quantum machine learning." *Quantum* 9 (2025): 1640.

Feature Maps

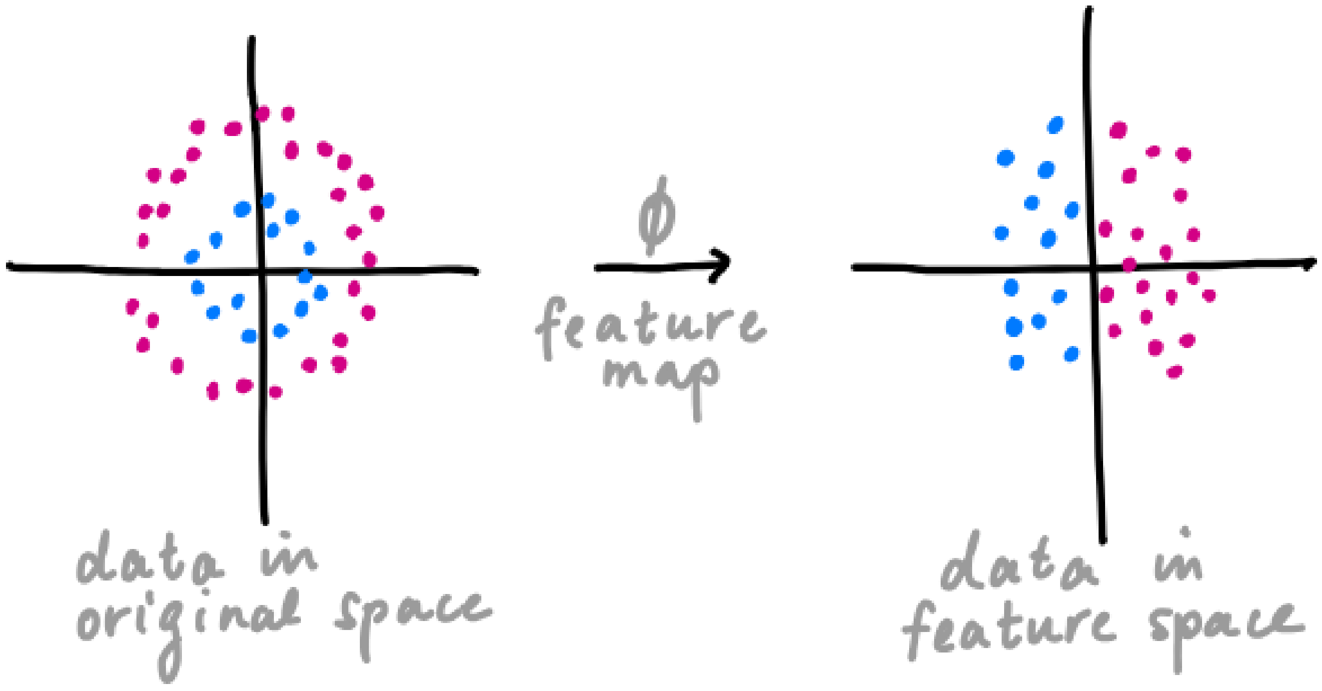
Feature maps are used to make data easier to work with.



Quantum Feature Maps, PennyLane tutorials.

Feature Maps

Feature maps are used to make data easier to work with.



Quantum Feature Maps, PennyLane tutorials.

Kernel Method

- Some ML algorithm can be reframed to only require inner product of feature map for pairs of data i.e. $\langle \phi(x), \phi(y) \rangle$.

> ***Definition:*** A kernel is a function that can be written as the inner product of feature maps i.e. $k(x, y) = \langle \phi(x), \phi(y) \rangle$

- **The kernel trick**: Using a kernel we already know has a complicated feature map e.g. the Gaussian kernel.

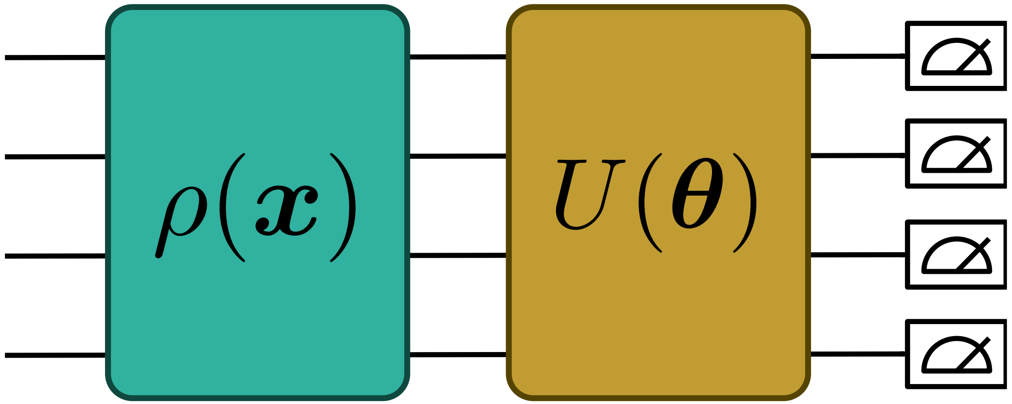
$$k(x, y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right)$$

Quantum Neural Networks (QNNs)

- Layers of encoding and parametrized gates

$$f_\theta(x) = \mathrm{Tr}[U(x,\theta)|0\rangle\langle 0|U^\dagger(x,\theta)\ \hat{O}]$$

$$U(x,\theta) = \prod_{l=1}^{L} W_l(\theta)V_l(x)$$

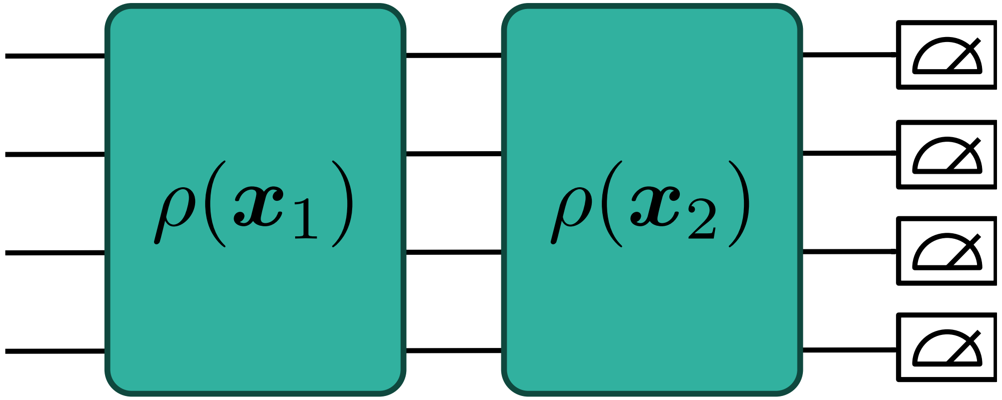


Quantum Kernels (QKs)

- Fidelity Quantum Kernels (with pure state encoding)

$$k_Q(x, y) = \mathrm{Tr}[\rho(x)\rho(y)]$$

$$\rho(x) = U(x)|0\rangle\langle 0|U^\dagger(x)$$



Hamiltonian Encoding

- Encoding gates $U(x) = \exp(ixH)$, H a Hermitian operator.

  - For QNNs: $f_\theta(x) = \sum_{\omega \in \Omega} c_\omega e^{i\omega x}$
  - For QKs: $k_Q(x, y) = \sum_{\omega, \nu \in \Omega} Q_{\omega\nu} e^{i(\omega x - \nu y)}$

$|\Omega|$ is exponential in the dimension of input data! ☹

What if we only consider some of these frequencies?

Schuld, Maria. "Supervised quantum machine learning models are kernel methods." *arXiv preprint arXiv:2101.11020* (2021).

Schuld, Maria, Ryan Sweke, and Johannes Jakob Meyer. "Effect of data encoding on the expressive power of variational quantum-machine-learning models." *Physical Review A* 103.3 (2021): 032430.

Random Fourier Features

Rahimi, Ali, and Benjamin Recht. "Random features for large-scale kernel machines." *Advances in neural information processing systems* 20 (2007).

- The goal is to find an approximate feature map for a kernel.
- Shift-invariant kernels: $k(x, y) = g(x - y)$
- **Bochner's theorem**: if p is the Fourier transform of g,

$$k(x, y) = \mathbb{E}_{\omega \backsim p}[\cos(\omega \cdot (x - y))]$$

i.e. the Fourier transform of a SI kernel, is *positive*.

Random Fourier Features

$$k(x, y) = \mathbb{E}_{\omega \backsim p}[\cos \omega(x - y)]$$

- Replace expectation with sample mean:

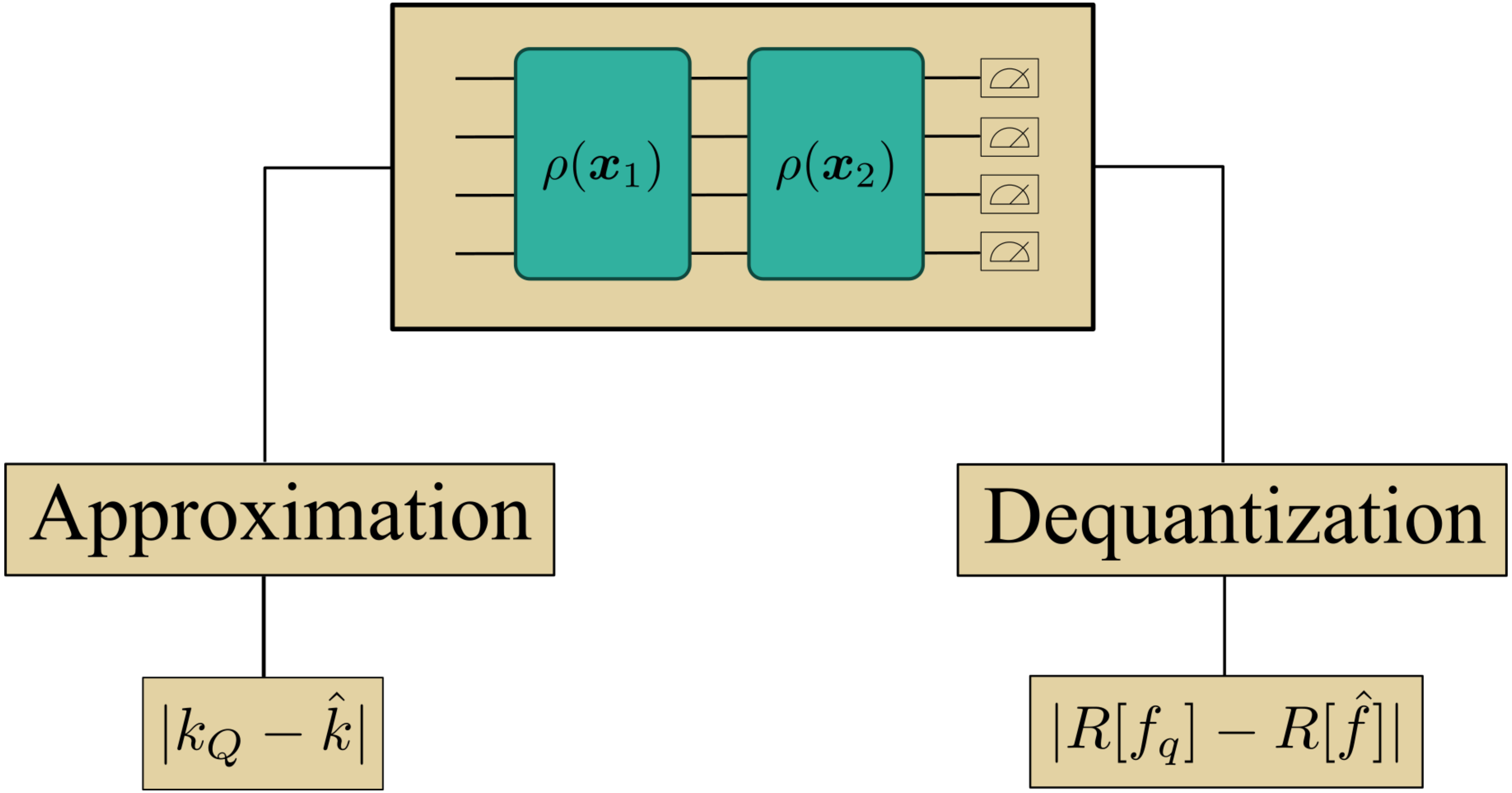
$$\hat{k}(x, y) = \phi^{\dagger}_{D,p}(x)\ \phi_{D,p}(y)$$

ω_i's are i.i.d samples of p,

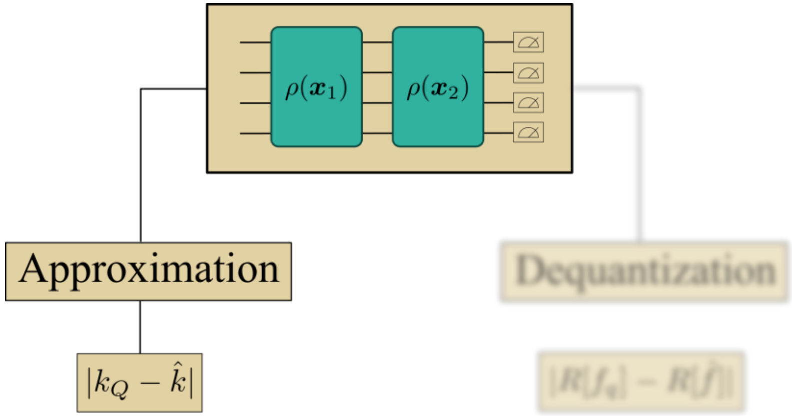
$$\phi_{D,p}(x) = \frac{1}{\sqrt{D}}[\cos(\omega_1 \cdot x), \sin(\omega_1 \cdot x), \cdots, \cos(\omega_D \cdot x), \sin(\omega_D \cdot x)]^T$$

- $D \in \Omega\left(\frac{d}{\epsilon^2} \log\left(\frac{\mathbb{E}[\|\omega\|_2^2]}{\epsilon}\right)\right)$ samples are enough for ϵ point-wise error

How to apply this to QKs?



QKs are not shift invariant!



$$k_Q(x, y) = \sum_{\omega,\nu\in\Omega} Q_{\omega\nu} e^{i(\omega\cdot x-\nu\cdot y)} = z(x)^\dagger Q z(y)$$

$$z(x) = \left[e^{i\omega_1\cdot x}, \cdots, e^{i\omega_{|\Omega|}\cdot x}\right]^T$$

Diagonal Q → SI kernel

Q is generally not diagonal!
RFF Approximation Does NOT WORK!

Random Features for QKs

$k_Q(x, y) = z(x)^\dagger Q z(y)$

Key Idea: Write $k_Q(x, y) = \mathbb{E}_N\left[\psi^\dagger(x, N)\psi(y, N)\right]$ and estimate with sample mean

Q is Hermitian, unit-trace and positive semi-definite

Diagonal elements of Q form a distribution q	Eigenvalues of Q form a distribution v

$$k_Q(x, y) = \sum_i v_i z(x)^\dagger u_i \underbrace{u_i^\dagger z(y)}_{\psi(y, N)}$$

Error Bounds

Error Bound: $D \in O\left(\frac{d|\Omega|^2}{\epsilon^2}\log\left(\frac{\mathbb{E}_N[\zeta_N^2]}{\epsilon}\right)\right)$ samples are enough for ϵ point-wise error

Computational Complexity of this approach

- FFT takes $\mathcal{O}(|\Omega| \log |\Omega|)$
- EVD takes $\mathcal{O}(poly(|\Omega|))$
- Number of samples scale with $|\Omega|^2$

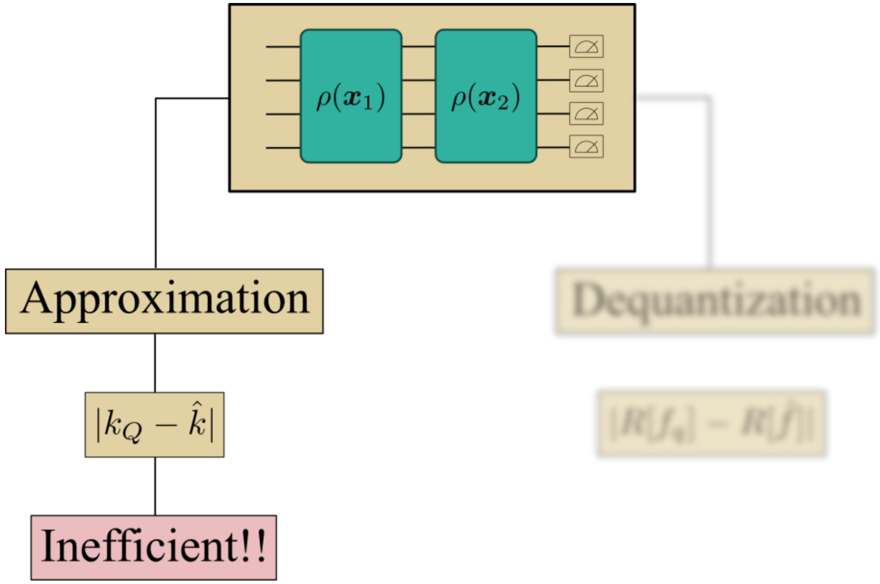
Bad News!!

Error Bounds

Error Bound: $D \in O\left(\frac{d|\Omega|^2}{\epsilon^2}\log\left(\frac{\mathbb{E}_N[\zeta_N^2]}{\epsilon}\right)\right)$ samples are enough for ϵ point-wise error

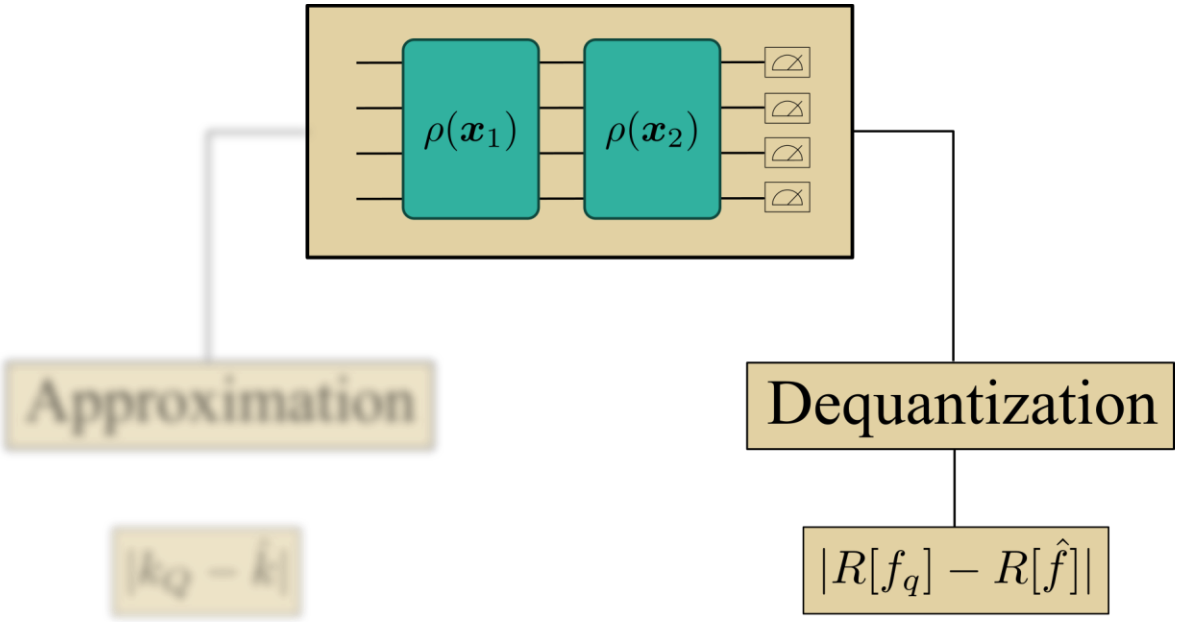
Computational Complexity of this approach

- FFT takes $\mathcal{O}(|\Omega| \log |\Omega|)$
- EVD takes $\mathcal{O}(poly(|\Omega|))$
- Number of samples scale with $|\Omega|^2$

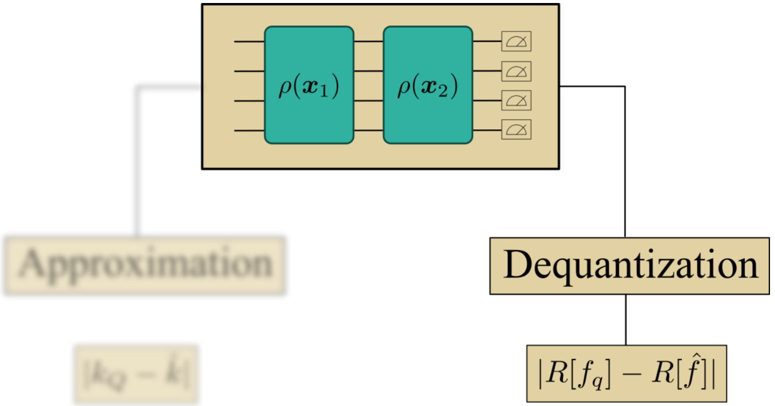


RFF-Dequantization

Can RFF feature map **perform as well as** quantum models in learning tasks?



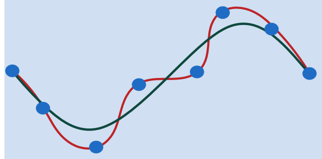
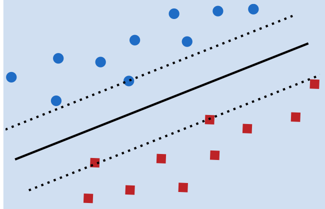
RFF-Dequantization

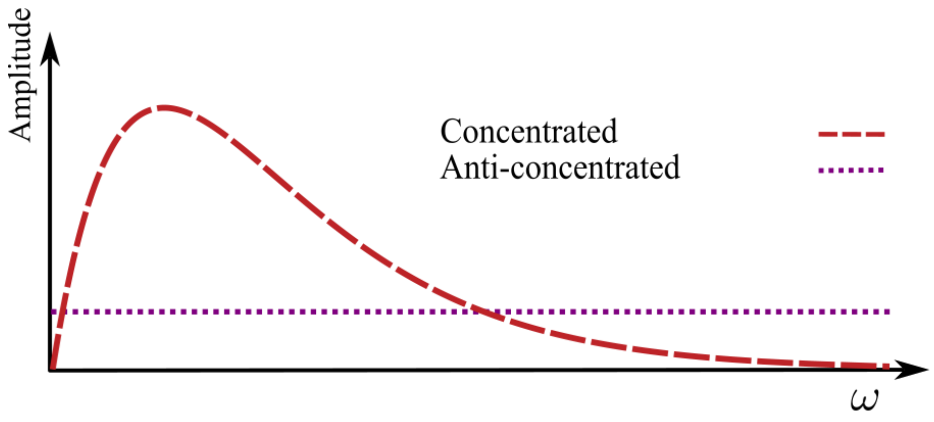


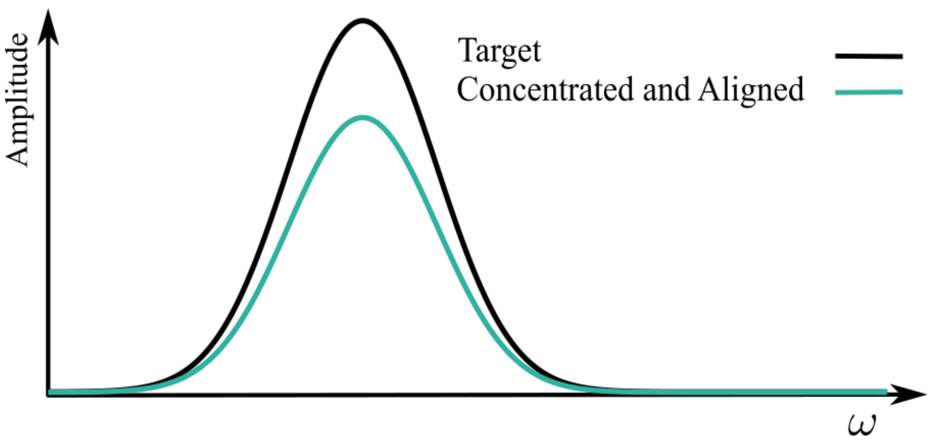
Definition: *A* QML task is **RFF-dequantized** if there exists a distribution p such that $\phi_{D,p}(x)$ reaches true risks at most ϵ greater than optimum true risk of the QML model, with $\mathcal{O}\big(poly(d, \epsilon^{-1})\big)$ frequency and data samples.

$$\phi_{D,p}(x) = \frac{1}{\sqrt{D}}[\cos(\omega_1 \cdot x), \sin(\omega_1 \cdot x), \cdots, \cos(\omega_D \cdot x), \sin(\omega_D \cdot x)]^T$$

Summary

Problem	QNN	QK
Regression 	▫ **Alignment:** $p_\omega \propto \lvert c^\star_\omega \rvert$. ● **Concentration:** $p_{max} \in \mathcal{O}(poly(d))$	▫ **Alignment:** $\left\lVert \sqrt{P}^{-1}\sqrt{Q} \right\rVert_\infty \in \mathcal{O}(poly(d))$ ● **Concentration:** $p_{max} \in \mathcal{O}(poly(d))$ ■ **Bounded RKHS norm:** $\lVert f_q \rVert_{k_Q} \in \mathcal{O}(poly(d))$
SVM 	▫ **Alignment:** $p_\omega \propto \lvert c^\star_\omega \rvert$ ■ **Bounded Fourier Sum:** $\sum \lvert c_\omega \rvert \in \mathcal{O}(poly(d))$	▫ **Alignment:** $p_\omega = \frac{1}{C}\sqrt{q_\omega}$ ● **Concentration:** $C = \sum_{\omega \in \Omega} \sqrt{q_\omega} \in \mathcal{O}(poly(d))$ ■ **Bounded RKHS norm:** $\lVert f_q \rVert_{k_Q} \in \mathcal{O}(poly(d))$





Limitations of Theoretical Approach

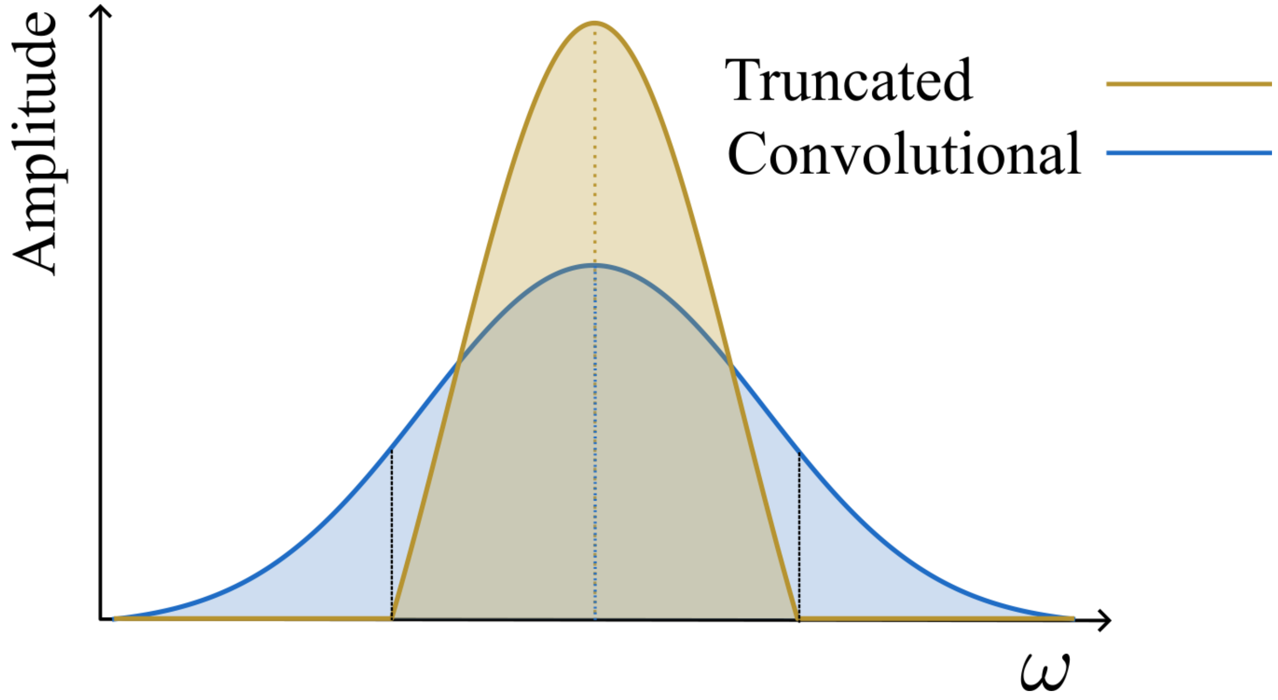
- The optimal QNN function, is not obtainable.
- Computational complexity of Fourier transform is high.
- Optimal sampling distribution may be hard to sample from.

Theoretically dequantizable does not mean you can actually find the proper distribution.

Does RFF with simple sampling strategies work as well?

Numerical Experiments

- RFF-SVM: Sampling Strategies
  - Uniform
  - Convolutional
  - Truncated



Numerical Experiments

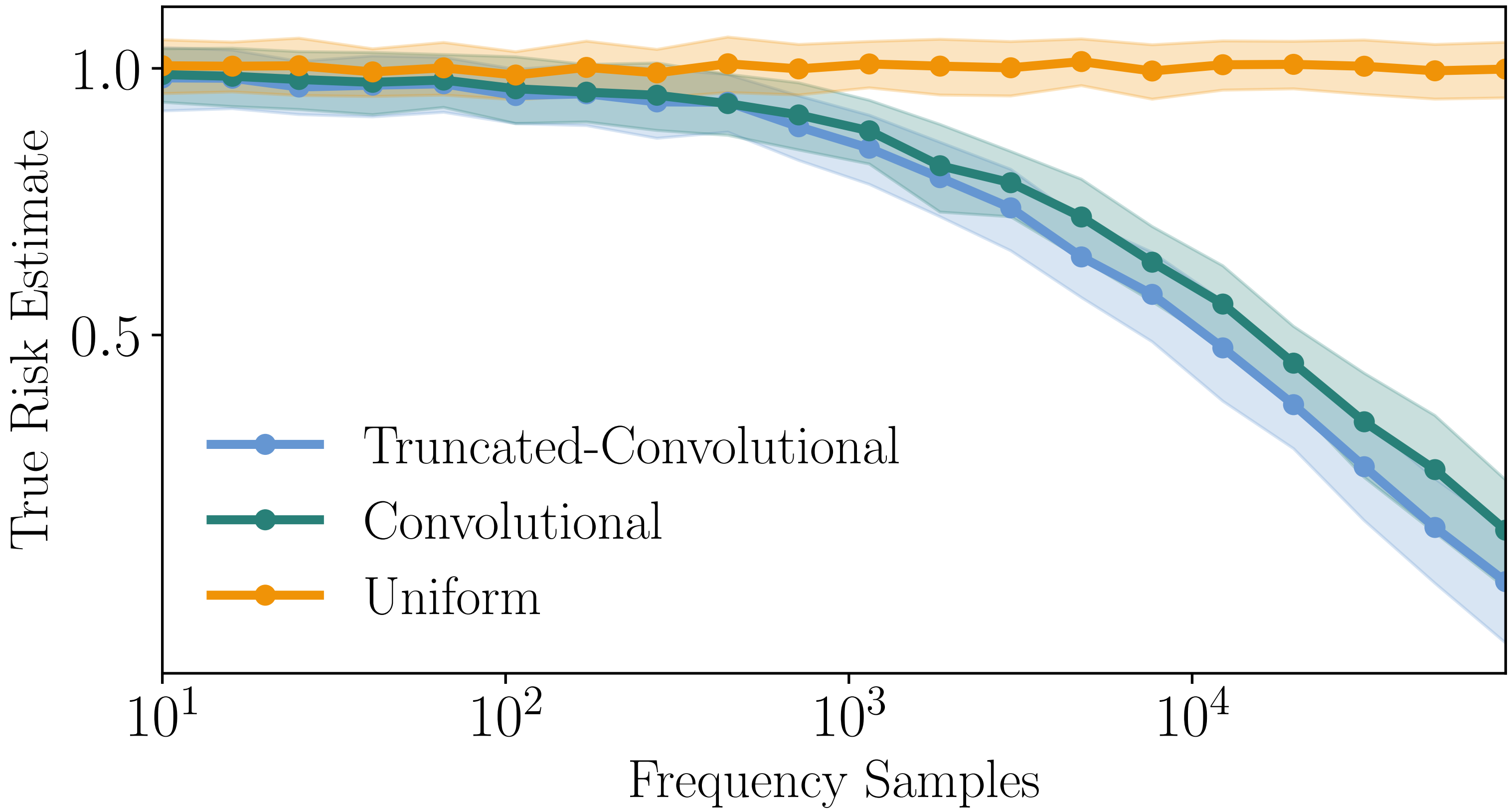
- Data set from particle collisions
- Dimension 64
- Comparison of sampling methods

Belis, Vasilis, et al. "Quantum anomaly detection in the latent space of proton collision events at the LHC." *Communications Physics* 7.1 (2024): 334.

Numerical Experiments

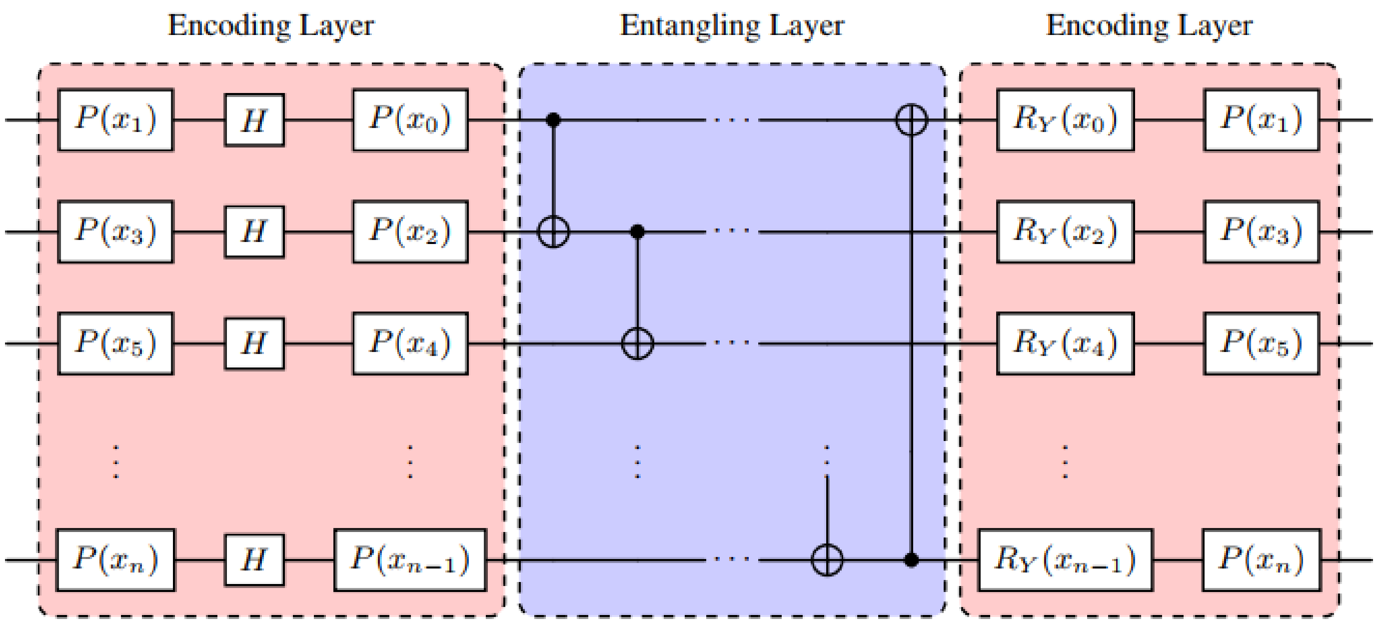
- Data set from particle collisions
- Dimension 64
- Comparison of samplir

Belis, Vasilis, et al. "Quantum anomaly detection in the latent space of proton collision events at the LHC." *Communications Physics* 7.1 (2024): 334.



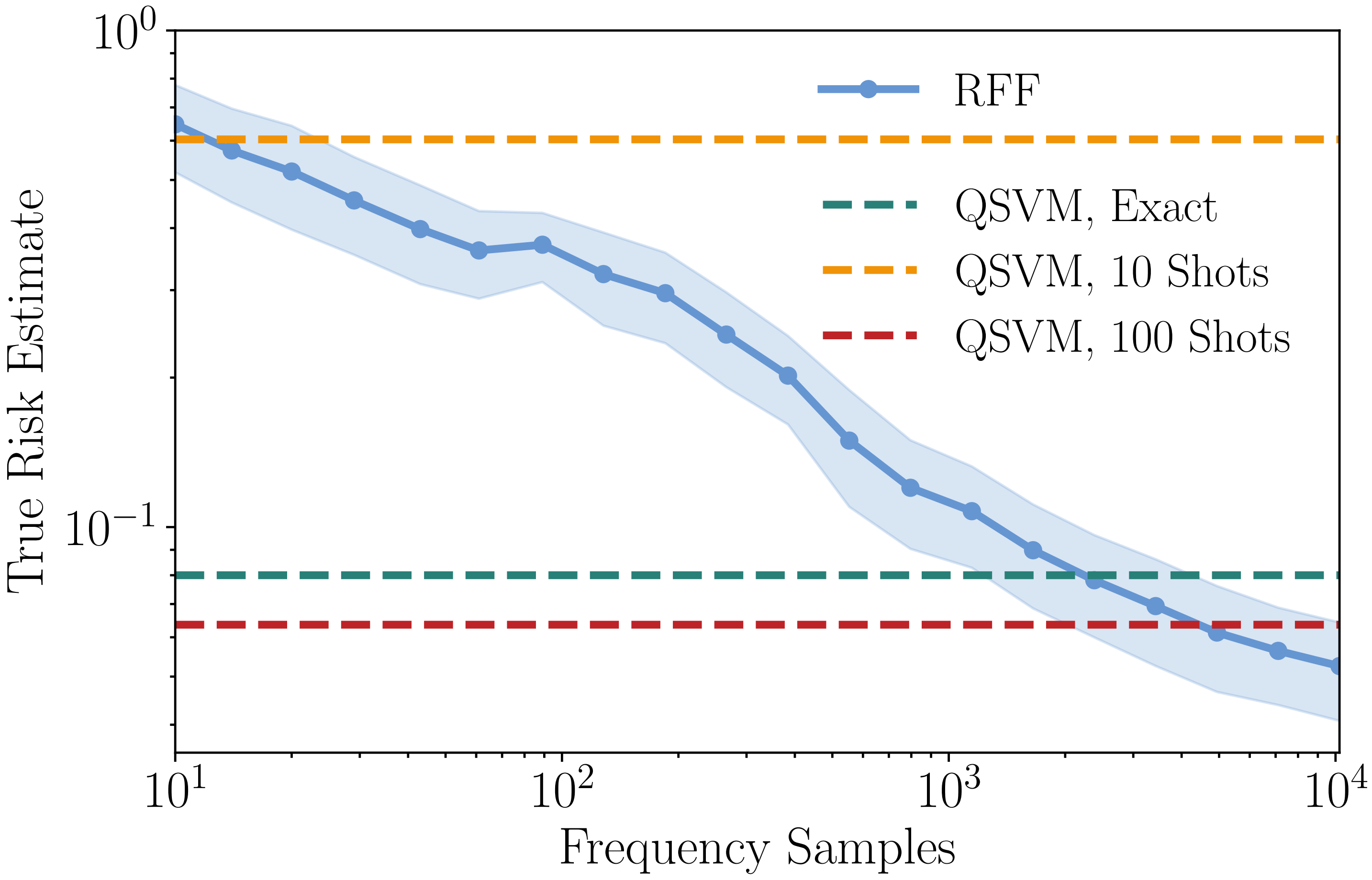
Numerical Experiment

- QK-SVM Settings:
  - Feature map
  - QK evaluated for pairs of data
  - Shot noise added as a binomial RV



Numerical Experiment

- Dimension 32
- 1000 training
- 200 test



Conclusion

Alignment and **concentration** appear as sufficient conditions for QK methods and SVM

Obtaining the optimal RFF sampling distribution is hard.

But simple, task independent distributions such as **convolutional** and **truncated** may be good options.

Thank you! ☺

Questions?