Quantum Physics

[Submitted on 22 Nov 2023]

A Unified Framework for Trace-induced Quantum Kernels

Beng Yee Gan, Daniel Leykam, Supanut Thanasilp

~~Expressivity and Generalization error of Trace-induced Quantum Kernels~~

Quantum Techniques in Machine Learning (QTML) 2023

Beng Yee Gan, Daniel Leykam, Supanut Thanasilp

arXiv:2311.13552

Centre for Quantum Technologies

EPFL

CHICS

Big picture of the work

- Global fidelity quantum kernels
  Nature 567, 209 (2019).
  PRL 122 (4), 040504.
- Linear projected quantum kernels
  Nat. Commun. 12, 2631 (2021)
- Quantum neural tangent kernels
  PRX Quantum 3, 030323 (2022).
  arXiv:2111.02951 (2021)
- Gaussian projected quantum kernels
  Nat. Commun. 12, 2631 (2021)
- Quantum topological kernel
  arXiv:2307.07383 (2023)
- Quantum Fisher kernels
  arXiv:2210.16581 (2022)

How are some of the existing quantum kernels related to each other?

Our work: A unified framework for trace-induced quantum kernels.

Insight on how to choose the quantum kernels for a given task.

Part 1

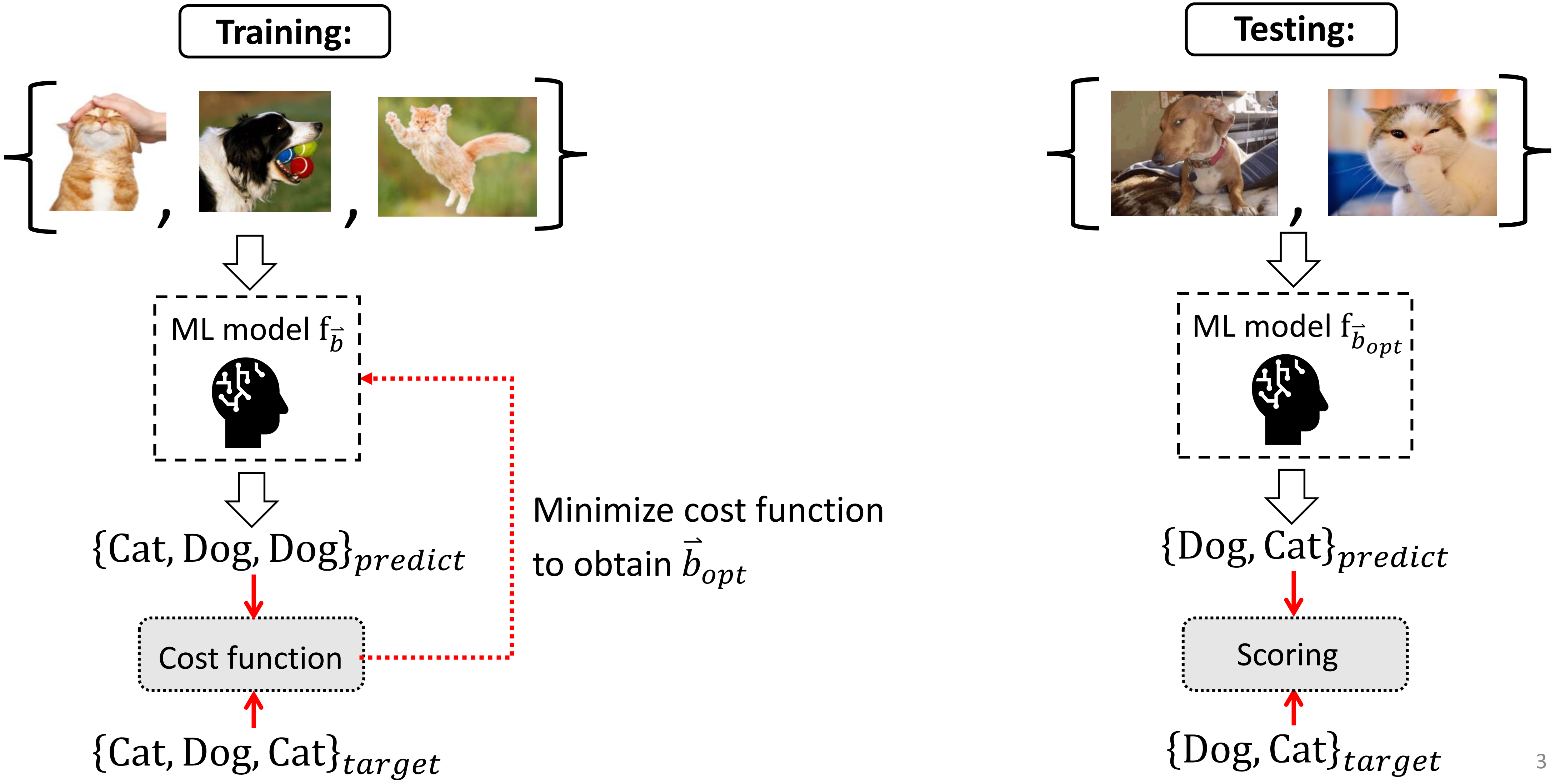
- (Quantum) kernel methods
- Expressivity
- Trainability
- Generalization ability

Part 2

- Unified framework
  - Expressivity
  - Generalization ability
- Practicality of the framework

Supervised Machine Learning

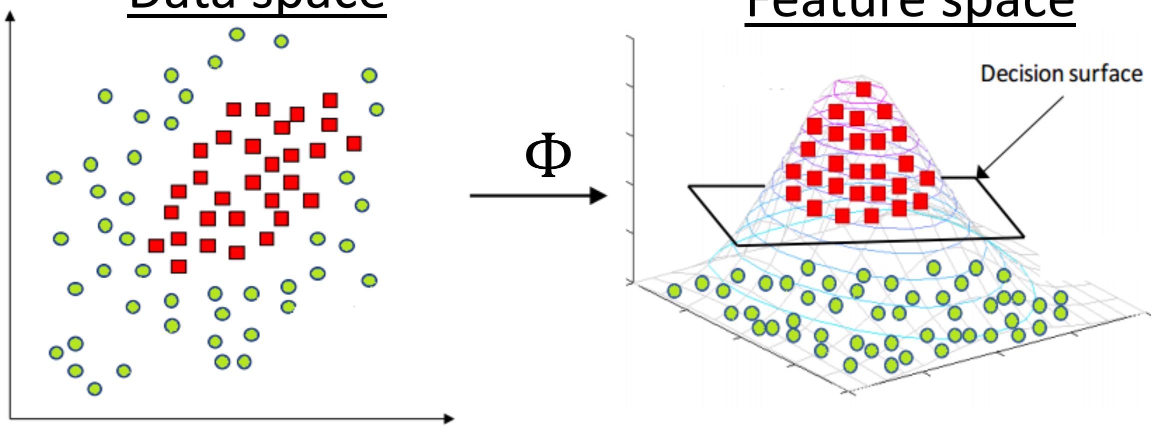
Goal: Extract patterns from **labelled dataset** to make accurate predictions on unknown and unseen data.



Classical Kernel Methods

Mapping to feature space:

Data space | Feature space



$f(\boldsymbol{x}) = \boldsymbol{p} \cdot \boldsymbol{x}$ ⟶ $\mathrm{g}(\boldsymbol{x}) = \boldsymbol{q} \cdot \boldsymbol{\Phi}(\boldsymbol{x})$

Kernel trick: $k(\boldsymbol{x}, \boldsymbol{x}') = \langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{x}') \rangle$

- Symmetric for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$
  $$k(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x}', \boldsymbol{x})$$
- Positive semi-definite
  $$\iint_{\mathcal{X}\times\mathcal{X}} c(\boldsymbol{x}) x(\boldsymbol{x}') k(\boldsymbol{x}, \boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \geq 0$$

Space of functions:

$$\mathcal{H} = \left\{ f(\boldsymbol{x}) = \sum_{i=1}^{\infty} \alpha_i\, k(\boldsymbol{x}_i, \boldsymbol{x}), \alpha_i \in \mathbb{R} \right\}$$

Representer theorem: ⟶

$$f_{\vec{a}}(\boldsymbol{x}) = \sum_{i=1}^{N} a_i\, k(\boldsymbol{x}_i, \boldsymbol{x})$$

Only depends on the number of training data points, N.

Eigen-decomposition of kernels

One well-known construction is to express a feature map using its eigenbasis functions.

Eigen-decomposition: $k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{j=1}^{\infty} \gamma_j \phi_j(\boldsymbol{x}) \phi_j(\boldsymbol{x}') = \sum_{j=1}^{\infty} \sqrt{\gamma_j} \phi_j(\boldsymbol{x}) \cdot \sqrt{\gamma_j} \phi_j(\boldsymbol{x}') = \langle \boldsymbol{\Psi}(\boldsymbol{x}), \boldsymbol{\Psi}(\boldsymbol{x}') \rangle$

Relative importance of the functions

Base functions for the models

$$\mathcal{H} = \left\{ f(\boldsymbol{x}) = \sum_{j=1}^{\infty} \beta_j \sqrt{\gamma_j} \phi_j(\boldsymbol{x}) \right\}$$

Generalized kernel: $k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{j=1}^{\infty} w_j \cdot \gamma_j \phi_j(\boldsymbol{x}) \phi_j(\boldsymbol{x}')$

$w_j \geq 0, \forall j$

Utilise multiple kernel learning algorithms to find optimal w_j

Multiple kernel learning:

Given $\mathcal{K} = \{k_i(\vec{x}, \vec{x}')\}$, construct

$$k_{tot} = \sum_{i=1}^{|\mathcal{K}|} w_i k_i(\vec{x}, \vec{x}')$$

with $w_i \geq 0$.

Gönen, M., et al (2011). Multiple kernel learning algorithms. *The J. Mach. Learn. Res.*, *12*, 2211-2268.

Eigen-decomposition of kernels

One well-known construction is to express a feature map using its eigenbasis functions.

Eigen-decomposition: $k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{j=1}^{\infty} \gamma_j \phi_j(\boldsymbol{x}) \phi_j(\boldsymbol{x}') = \sum_{j=1}^{\infty} \sqrt{\gamma_j} \phi_j(\boldsymbol{x}) \cdot \sqrt{\gamma_j} \phi_j(\boldsymbol{x}') = \langle \boldsymbol{\Psi}(\boldsymbol{x}), \boldsymbol{\Psi}(\boldsymbol{x}') \rangle$

$$\mathcal{H} = \left\{ f(\boldsymbol{x}) = \sum_{j=1}^{\infty} \beta_j \sqrt{w_j \gamma_j} \phi_j(\boldsymbol{x}) \right\}$$

Generalized kernel: $k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{j=1}^{\infty} w_j \cdot \gamma_j \phi_j(\boldsymbol{x}) \phi_j(\boldsymbol{x}')$

$w_j \geq 0, \forall j$

Utilise multiple kernel learning algorithms to find optimal w_j

Multiple kernel learning:

Given $\mathcal{K} = \{k_i(\vec{x}, \vec{x}')\}$, construct

$$k_{tot} = \sum_{i=1}^{|\mathcal{K}|} w_i k_i(\vec{x}, \vec{x}')$$

with $w_i \geq 0$.

Gönen, M., et al (2011). Multiple kernel learning algorithms. *The J. Mach. Learn. Res.*, *12*, 2211-2268.

Expressivity, Trainability, Generalization error

Expressivity

- Informs the complexity of the model class.

$$\mathcal{H} = \left\{ f(\boldsymbol{x}) = \sum_{i=1}^{\infty} \alpha_i\, k(\boldsymbol{x}_i, \boldsymbol{x}) \right\}$$

- Dependent on the kernel choice
- Study kernel's eigenbasis

$$\mathcal{H} = \left\{ f(\boldsymbol{x}) = \sum_{j=1}^{\infty} \beta_j \sqrt{\gamma_j} \phi_j(\boldsymbol{x}) \right\}$$

Trainability

- How easy to reach the optimal solution in the optimization process

$$\vec{a}_{(opt)} := \operatorname{argmin}_{\vec{a}} \mathcal{L}_{\vec{a}}(S)$$

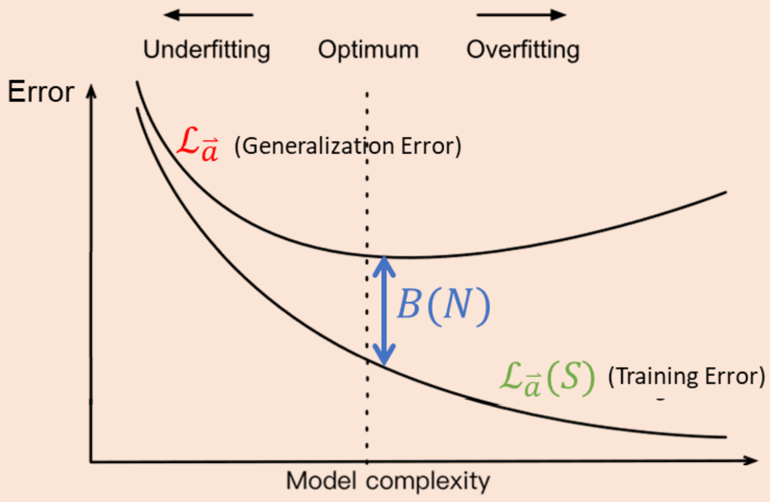
to find

$$f_{\vec{a}_{opt}}(\boldsymbol{x}) = \sum_{i=1}^{N} a_i^{(opt)}\, k(\boldsymbol{x}_i, \boldsymbol{x})$$

- Optimization become convex if $\mathcal{L}_{\vec{a}}$ is properly chosen
- Kernel ridge regression
  - ➢ Square loss function
- Support vector machines
  - ➢ Hinge loss function

Generalization error

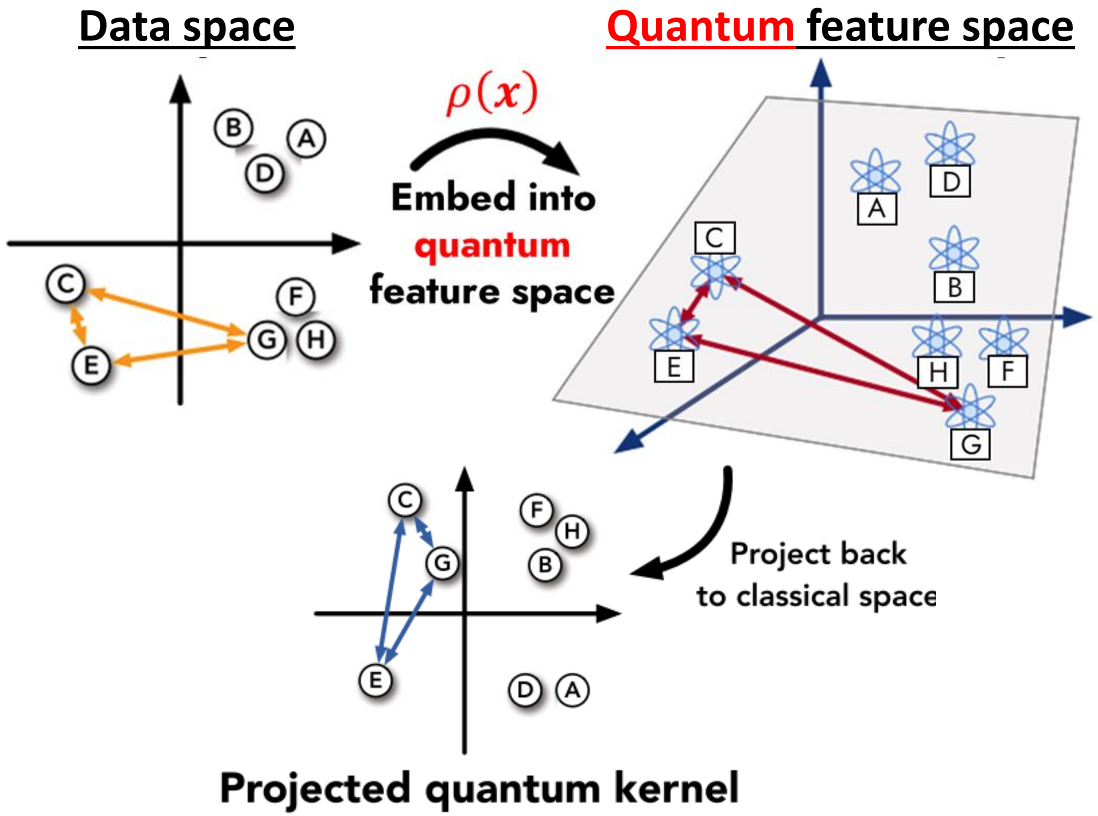
- How well the trained model predicts on the unseen dataset.



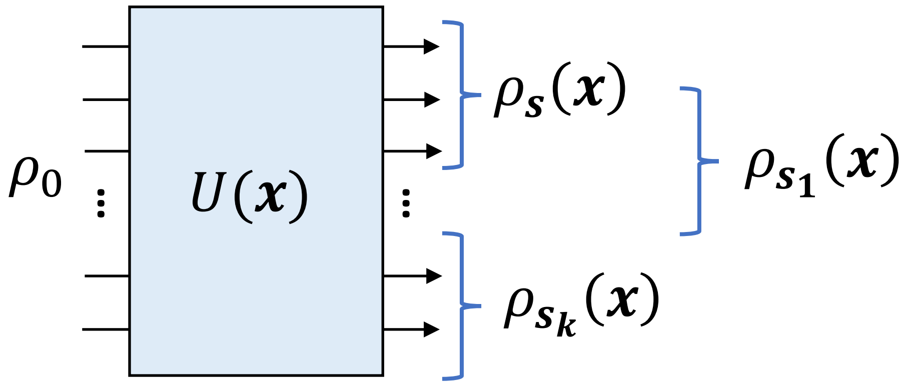
$$\mathcal{L}_{\vec{a}} - \mathcal{L}_{\vec{a}}(S) \leq B(N)$$

Quantum Kernel Methods

Mapping to quantum feature space:



- Quantum model: $g(\boldsymbol{x}) = \mathrm{tr}\big(M\rho(\boldsymbol{x})\big)$
- Quantum kernel: $k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{tr}(\rho(\boldsymbol{x})\rho(\boldsymbol{x}'))$ **(Global fidelity quantum kernels)**



s-linear projected quantum kernels (LPQK)

$$k_{\boldsymbol{s}}(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{tr}_{\boldsymbol{s}}(\rho_{\boldsymbol{s}}(\boldsymbol{x})\rho_{\boldsymbol{s}}(\boldsymbol{x}')), \; |\boldsymbol{s}| = S$$

S-linear projected quantum kernels (LPQK)

$$k_{\boldsymbol{s}}(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{\sqrt{\{\boldsymbol{s}\}}} \sum_{\{\boldsymbol{s}\}} \mathrm{tr}_{\boldsymbol{s}}(\rho_{\boldsymbol{s}}(\boldsymbol{x})\rho_{\boldsymbol{s}}(\boldsymbol{x}')), \; \forall |\boldsymbol{s}| = S$$

Huang, H. Y., et al (2021). Power of data in quantum machine learning. *Nat. Commun.*, *12*(1), 2631.

What’s going on in the field?

Complexity theory-based expressivity

nature physics
ARTICLES
https://doi.org/10.1038/s41567-021-01287-z
Check for updates

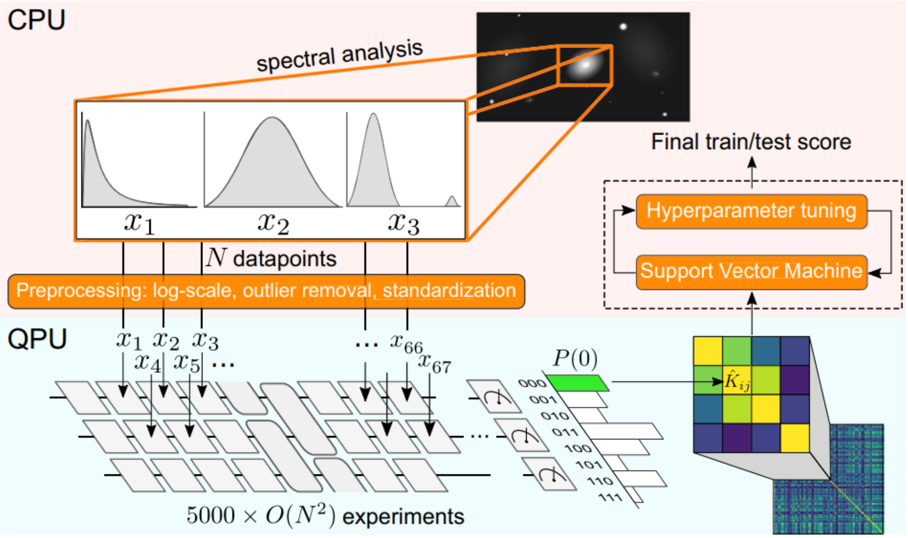
A rigorous and robust quantum speed-up in supervised machine learning

Yunchao Liu[1,2], Srinivasan Arunachalam[2] and Kristan Temme[2]

nature communications
Article
https://doi.org/10.1038/s41467-023-36144-5

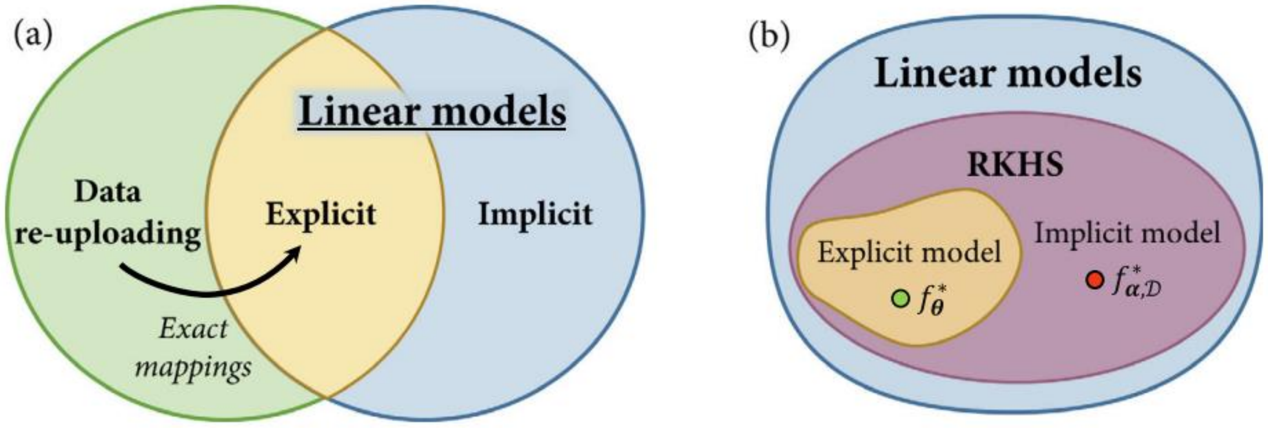
Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines

Quantum kernel for real world dataset



E. Peters, et. al. npj Quantum Infor. 7, 1 (2021)

Connections to other QML candidates:

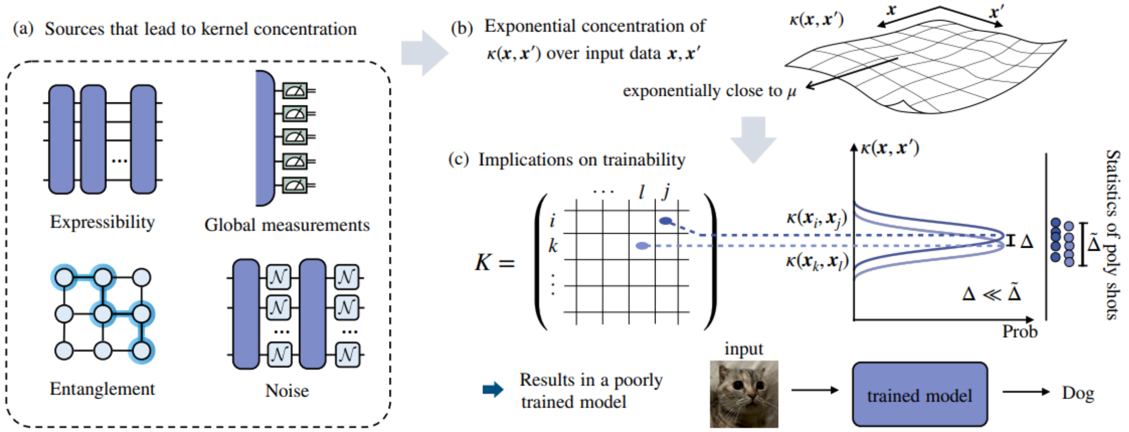


S. Jerbi, et. al. Nat. Commun. 14, 517 (2023).

What's going on in the field? (cont'd)

Kernel concentrations:

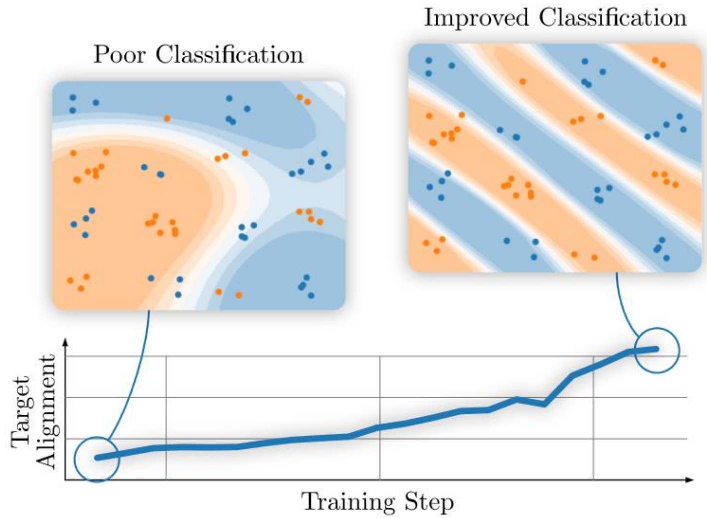
S. Thanasilp, et. al. arXiv:2208.11060 (2022)



The Inductive Bias of Quantum Kernels

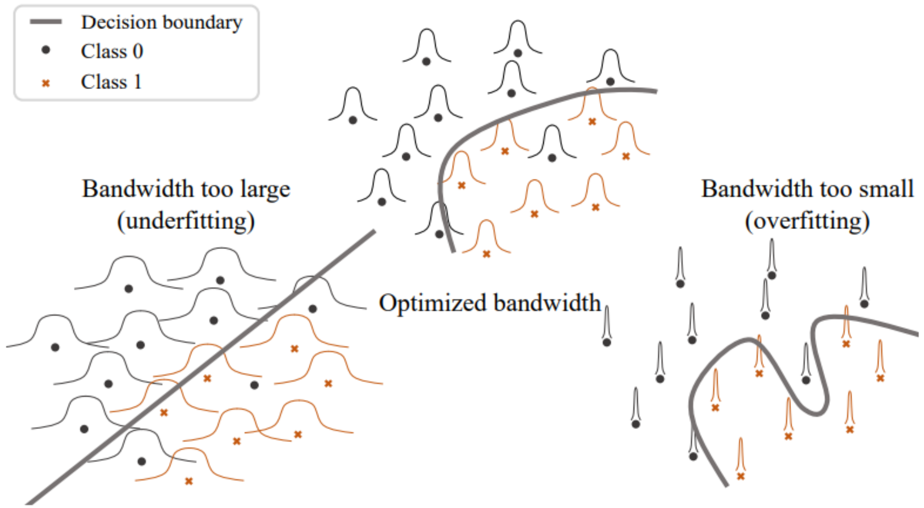
Jonas M. Kübler* Simon Buchholz* Bernhard Schölkopf
Max Planck Institute for Intelligent Systems
Tübingen, Germany
{jmkuebler, sbuchholz, bs}@tue.mpg.de

Training parameterized kernels:



T. Hubregtsen, et. al. Phys. Rev. A 106, 042431 (2022)

Kernel bandwidth:



R. Shaydulin, et. al., Phys. Re. A 106, 042407 (2022)

What’s going on in the field? (cont’d)

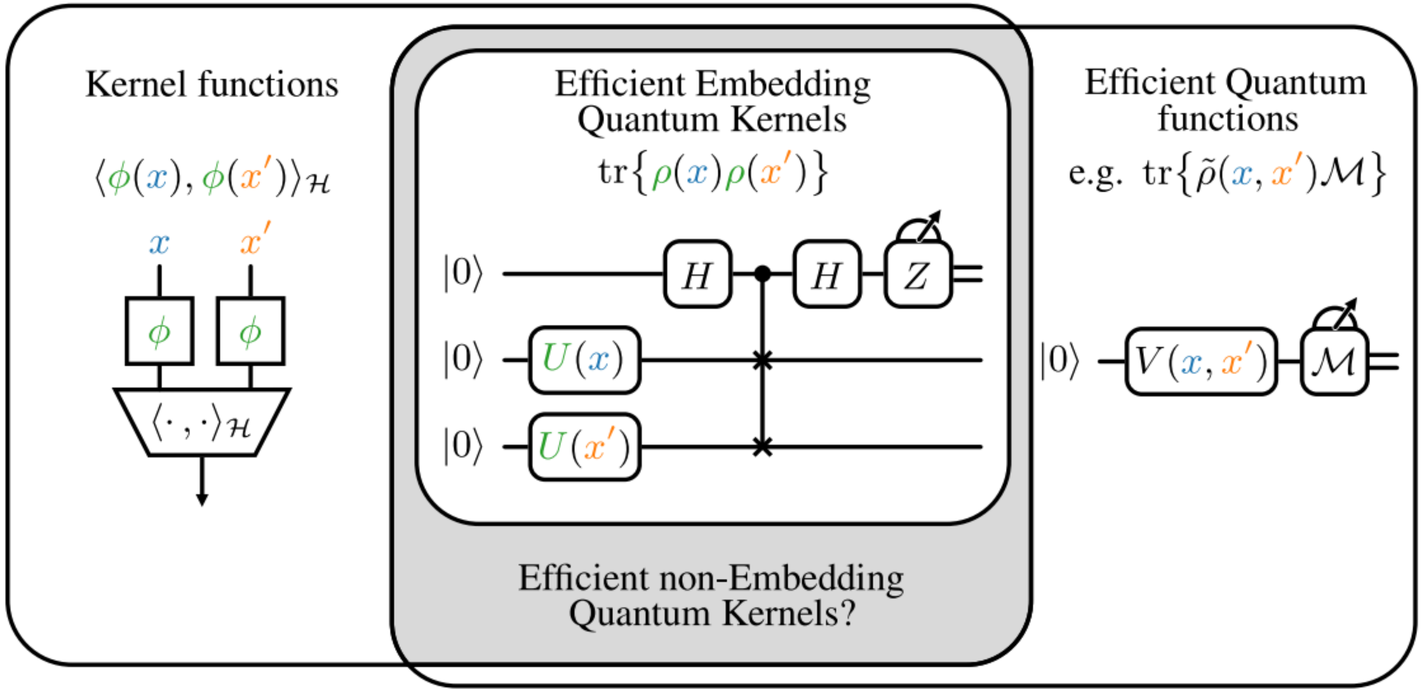
arXiv > quant-ph > arXiv:2309.14419

Quantum Physics

[Submitted on 25 Sep 2023]

On the expressivity of embedding quantum kernels

Elies Gil-Fuster, Jens Eisert, Vedran Dunjko



Unified framework for Trace-induced Quantum Kernels

Generalized Trace-induced Quantum Kernels (GTQKs)

Lego kernels:

Given $\mathcal{A} = \{A_i\}_{i=1}^{4^n}$ with $\mathrm{tr}(A_i A_j) = \delta_{ij}\ \forall\, i,j$ → $k_i(\vec{x},\vec{x}') = \mathrm{tr}(\rho(\vec{x})A_i)\,\mathrm{tr}(\rho(\vec{x}')A_i)$

Arbitrary orthonormal Hermitian basis

Compare $\rho(\vec{x})$ and $\rho(\vec{x}')$ in the direction of A_i.

Building up expressive power:

$$k(\boldsymbol{x},\boldsymbol{x}') = \sum_{i=1}^{4^n} 2^n w_i \cdot \mathrm{tr}(\rho(\boldsymbol{x})A_i)\,\mathrm{tr}(\rho(\boldsymbol{x}')A_i) = \mathrm{tr}(\tilde{\rho}(\boldsymbol{x})\tilde{\rho}(\boldsymbol{x}'))$$

(Generalized trace-induced quantum kernels)

$$\sum_{i=1}^{4^n} w_i^2 = 1$$

- $w_i = \frac{1}{2^n}\ \forall i \rightarrow \mathrm{tr}(\rho(\boldsymbol{x})\rho(\boldsymbol{x}'))$ (GFQK) — Regardless of basis
- $w_i = \begin{cases} \frac{1}{2^S} & \text{, if } \mathcal{T} = \text{True} \\ 0 & \text{, otherwise} \end{cases} \rightarrow \mathrm{tr}_s(\rho_s(\boldsymbol{x})\rho_s(\boldsymbol{x}'))$ (**s**-LPQKs)

With Pauli basis

- Encompass GFQKs and LPQKs as subsets
- Study expressivity and generalizability under the same unified framework

Generalized Trace-induced Quantum Kernels (GTQKs)

$$k(\boldsymbol{x},\boldsymbol{x}') = \sum_{i=1}^{4^n} 2^n w_i \cdot \mathrm{tr}(\rho(\boldsymbol{x})A_i)\,\mathrm{tr}(\rho(\boldsymbol{x}')A_i)$$

$$= \sum_{i=1}^{4^n} \sqrt{2^n w_i}\,\mathrm{tr}(\rho(\boldsymbol{x})A_i) \cdot \sqrt{2^n w_i}\,\mathrm{tr}(\rho(\boldsymbol{x}')A_i)$$

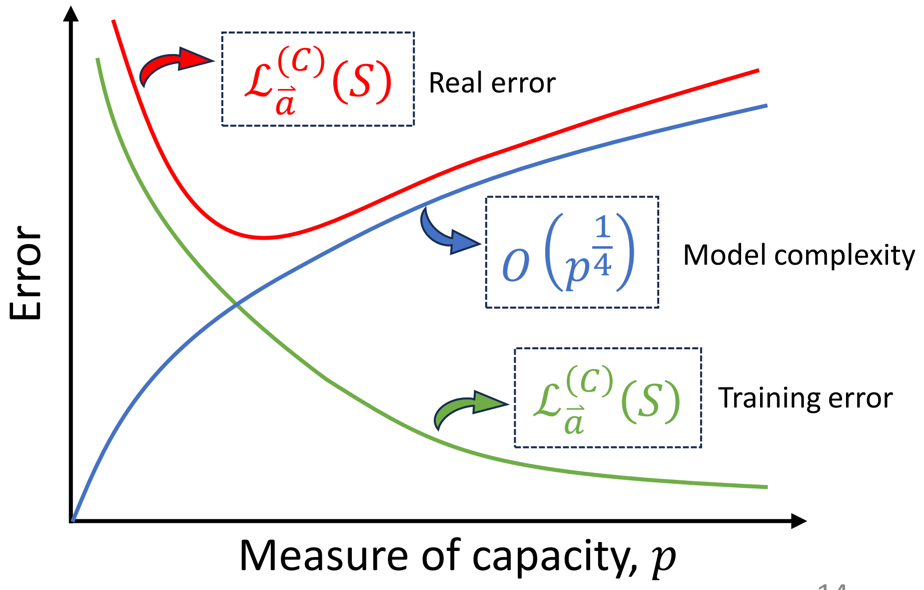
$$= \sum_{i=1}^{4^n} \psi_i(\boldsymbol{x}) \cdot \psi_i(\boldsymbol{x}')$$

$$\int_{\mathcal{X}} \mathrm{tr}(\rho(\boldsymbol{x})A_i)\,\mathrm{tr}\big(\rho(\boldsymbol{x})A_j\big)\,\mu(dx) \neq \delta_{ij}$$

Space of functions:

$$\mathcal{H}_G = \left\{ f(\boldsymbol{x}) = \sum_{i=1}^{4^n} \alpha_i \sqrt{2^n w_i}\,\mathrm{tr}(\rho(\boldsymbol{x})A_i) \right\}$$

- p : Control model complexity (# of non-zero weights)
- w_i: Control the inductive bias
- Generalization bound: $\mathcal{L}_{\vec{a}}^{(C)} \leq \mathcal{L}_{\vec{a}}^{(C)}(S) + O\left(p^{\frac{1}{4}}\right)$



Eigenbasis for GTQKs

Diagonalize $\int_{\mathcal{X}} \rho(\boldsymbol{x}) \otimes \rho(\boldsymbol{x})\, \mu(d\boldsymbol{x})$

$$\mathcal{A}_{U_\mathcal{X}} = \left\{A_i^{(U_\mathcal{X})}\right\}_{i=1}^{4^n} \text{ with eigenvalues } \{\gamma_i\}_{i=1}^{4^n}$$

(Mercer basis)

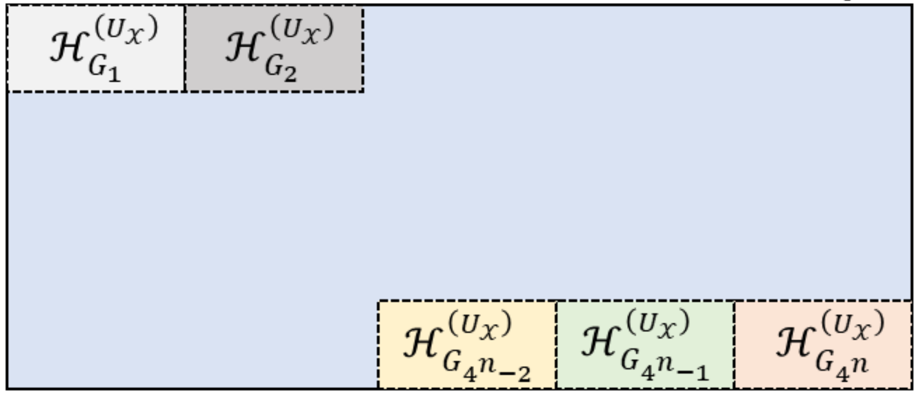
GTQKs in Mercer basis:

$$k^{(U_\mathcal{X})}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{4^n} 2^n w_i \cdot \underbrace{\operatorname{tr}\left(\rho(\boldsymbol{x}) A_i^{(U_\mathcal{X})}\right) \operatorname{tr}\left(\rho(\boldsymbol{x}') A_i^{(U_\mathcal{X})}\right)}_{k_i^{(U_\mathcal{X})} \text{ with space } \mathcal{H}_{G_i}^{(U_\mathcal{X})}}$$

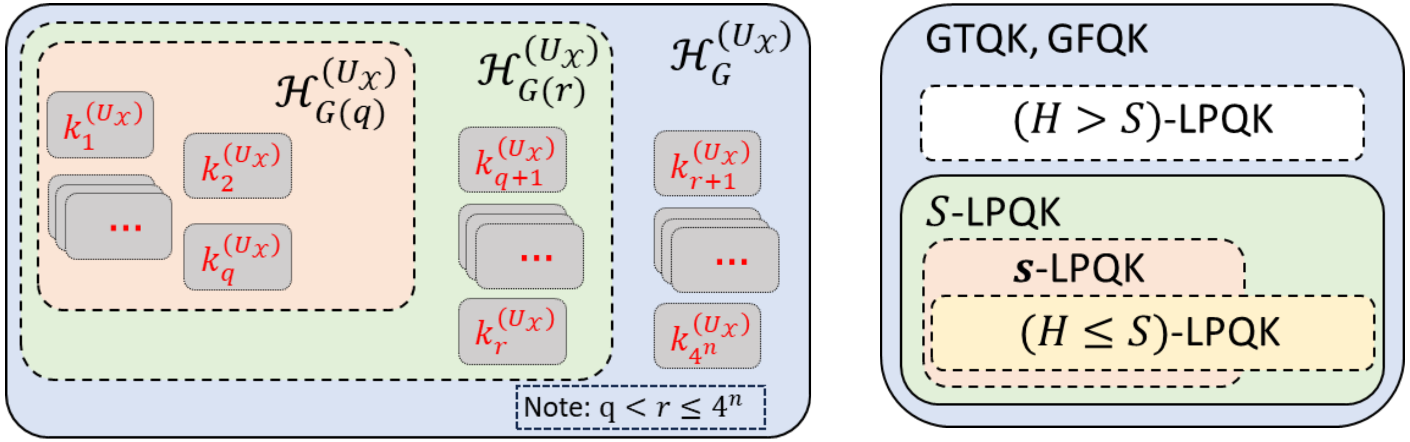
$\mathcal{H}_{G_i}^{(U_\mathcal{X})}$ and $\mathcal{H}_{G_i}^{(U_\mathcal{X})}$ are orthogonal to each other

Decomposition of space:

Space of $GTQK$: $\mathcal{H}_G^{(U_\mathcal{X})} = \mathcal{H}_{G_1}^{(U_\mathcal{X})} \oplus \cdots \oplus \mathcal{H}_{G_{4^n}}^{(U_\mathcal{X})}$



Expressivity hierarchy:



Practical implications of the unified framework

Practical strategy to select GTQKs

1

Choose the basis as Pauli basis: $P_i \in \{1, X, Y, Z\}^{\otimes n}$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{4^n} w_i \cdot \underbrace{\mathrm{tr}(\rho(\boldsymbol{x}) P_i)\, \mathrm{tr}(\rho(\boldsymbol{x}') P_i)}$$

- w_i: Classical post-processing
- $\mathrm{tr}(\rho(\boldsymbol{x}) P_i)\, \mathrm{tr}(\rho(\boldsymbol{x}') P_i)$: Measure the Pauli expectation value

2

Systematic way to set non-zero weights, p

One Pauli: $\{X \otimes 1^{\otimes n-1}, Y \otimes 1^{\otimes n-1}, Z \otimes 1^{\otimes n-1}, \cdots, 1^{\otimes n-1} \otimes Z\}$

Two Pauli: $\{XX \otimes 1^{\otimes n-2}, XY \otimes 1^{\otimes n-2}, \cdots, 1^{\otimes n-2} \otimes ZZ\}$

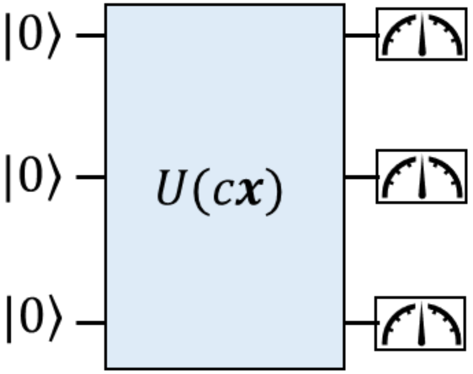
⋮

H Pauli: $\{\underbrace{XX \dots XX}_{H} \otimes 1^{\otimes n-2}, \underbrace{XX \dots XY}_{H} \otimes 1^{\otimes n-2}, \cdots, 1^{\otimes n-2} \otimes \underbrace{ZZ \dots ZZ}_{H}\}$ ⇨ ***H*-body LPQKs**

Practical strategy to select GTQKs (cont'd)

<u>A simple example: 8 qubits, $H = 2$</u>

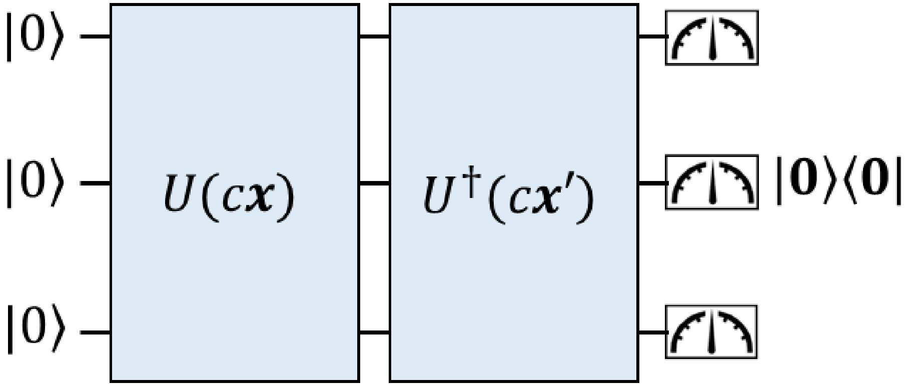
2 Pauli: $\mathcal{P}_2 = \{X_1X_2 \otimes 1^{\otimes 6}, X_1Y_2 \otimes 1^{\otimes 6}, \cdots, 1^{\otimes 6} \otimes Z_7Z_8\}$ ($|\mathcal{P}_2|$ = 252)



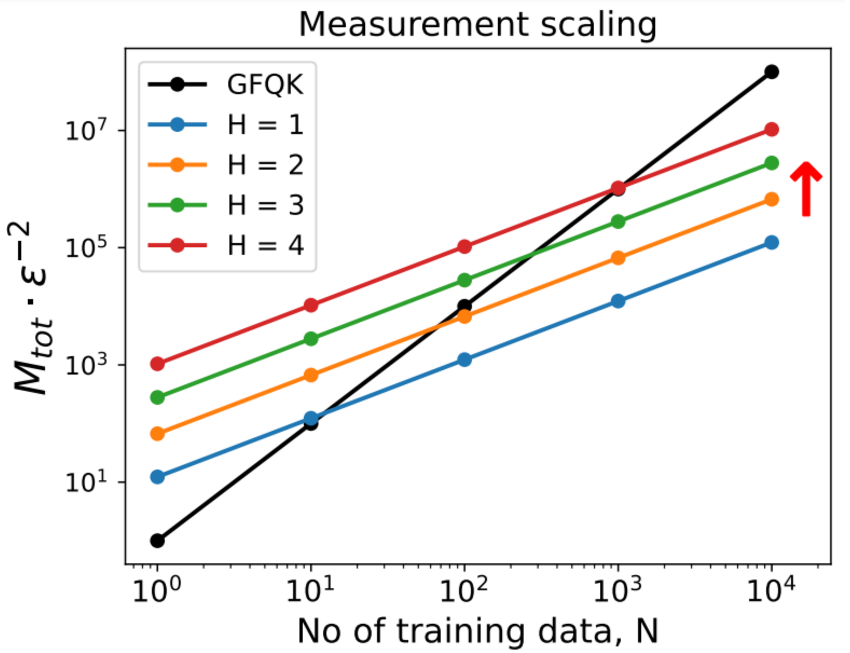
$$\left.\begin{matrix} \mathrm{tr}(\rho(c\boldsymbol{x}) \cdot X_1X_2 \otimes I^{\otimes 6}) \\ \mathrm{tr}(\rho(c\boldsymbol{x}) \cdot X_1Y_2 \otimes I^{\otimes 6}) \\ \vdots \\ \mathrm{tr}(\rho(c\boldsymbol{x}) \cdot I^{\otimes 6} \otimes Z_7Z_8) \end{matrix}\right\} |\mathcal{P}_2|$$

$w_i = \frac{1}{\sqrt{p}}$

$$k^{(2)}(c\boldsymbol{x}, c\boldsymbol{x}') = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} \mathrm{tr}\left(\rho(c\boldsymbol{x}) P_i^{(2)}\right) \mathrm{tr}\left(\rho(c\boldsymbol{x}') P_i^{(2)}\right)$$

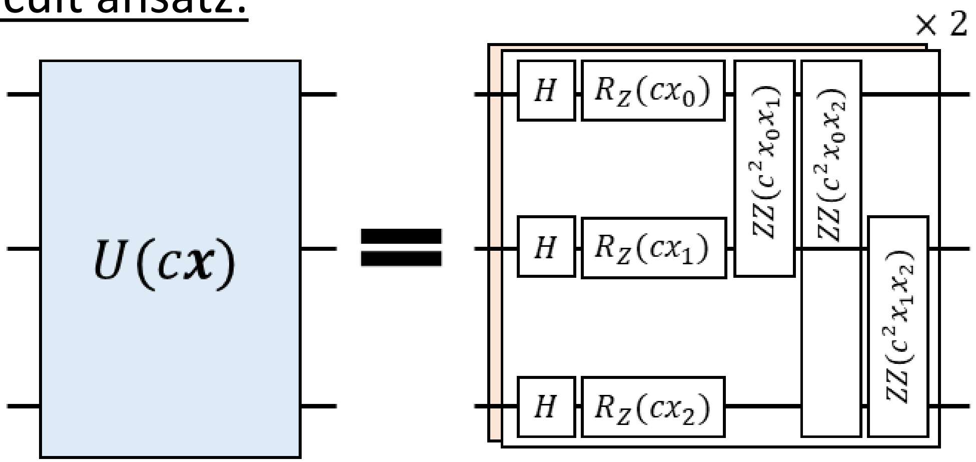


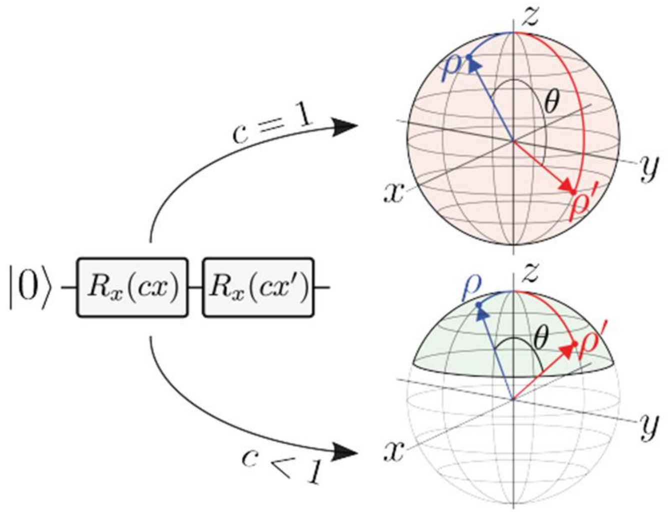
$$\mathrm{tr}(\rho(c\boldsymbol{x})\rho(c\boldsymbol{x}')) = |\langle\vec{0}|U^\dagger(c\boldsymbol{x})U(c\boldsymbol{x}')|\vec{0}\rangle|^2$$



Empirical study of H-body LPQK's performance

Circuit ansatz:



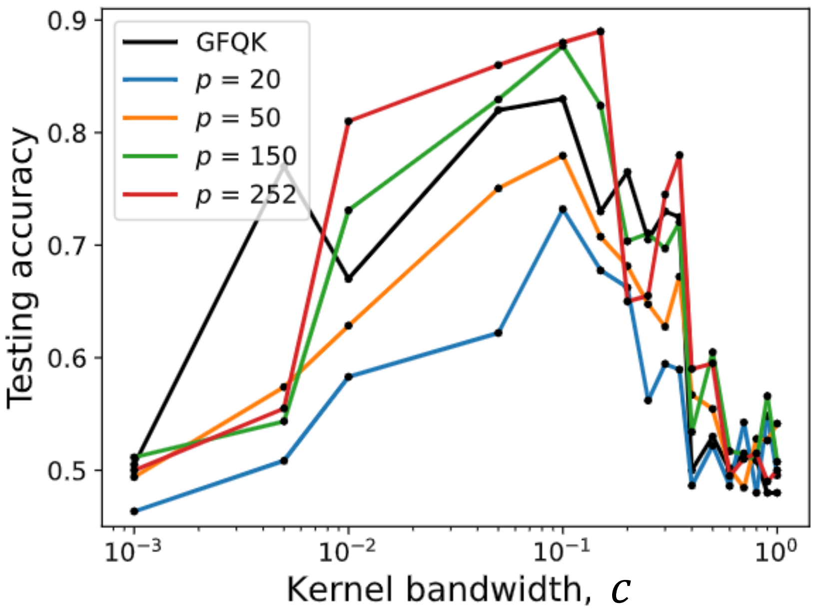


8 qubits, binary classify fashion-mnist

Linear projected quantum kernels:

$$k^{(2)}(c\boldsymbol{x}, c\boldsymbol{x}') = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} \mathrm{tr}\left(\rho(c\boldsymbol{x}) P_i^{(2)}\right) \mathrm{tr}\left(\rho(c\boldsymbol{x}') P_i^{(2)}\right)$$

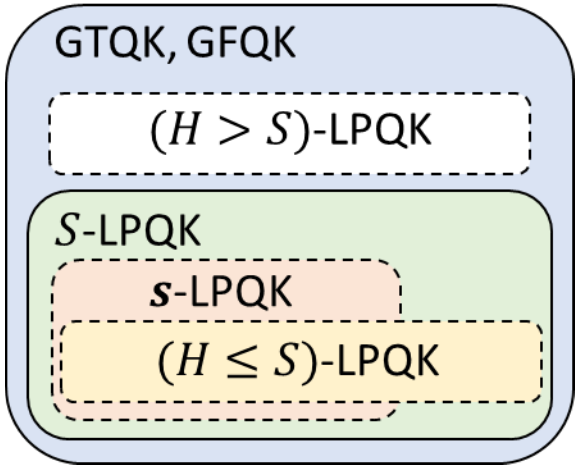
$$P_i^{(2)} \in \{X_1X_2 \otimes 1^{\otimes 6}, X_1Y_2 \otimes 1^{\otimes 6}, \cdots, 1^{\otimes 6} \otimes Z_7Z_8\} \ (|\mathcal{P}_2| = 252)$$



Shaydulin, R. et al (2022). Importance of kernel bandwidth in quantum machine learning. Physical Review A, 106(4), 042407.
Canatar, A., et. al. (2022). Bandwidth enables generalization in quantum kernel models. arXiv:2206.06686 [quant-ph].

Conclusion

- Global fidelity quantum kernels
  Nature 567, 209 (2019).
  PRL 122 (4), 040504.
- Linear projected quantum kernels
  Nat. Commun. 12, 2631 (2021)
- Quantum neural tangent kernels
  PRX Quantum 3, 030323 (2022).
  arXiv:2111.02951 (2021)
- Gaussian projected quantum kernels
  Nat. Commun. 12, 2631 (2021)
- Quantum topological kernel
  arXiv:2307.07383 (2023).
- Quantum Fisher kernels
  arXiv:2210.16581 (2022)



arXiv:2311.13552

How are some of the existing quantum kernels related to each other?

Generalized Trace-induced Quantum Kernels!

Our work: A unified framework for trace-induced quantum kernels.

Insight on how to choose the quantum kernels for a given task.