Title: Quantum algorithms for classical causal learning

Speakers: Sally Shrapnel

Series: Quantum Foundations, Quantum Information

Date: September 18, 2024 - 9:00 AM

URL: https://pirsa.org/24090089

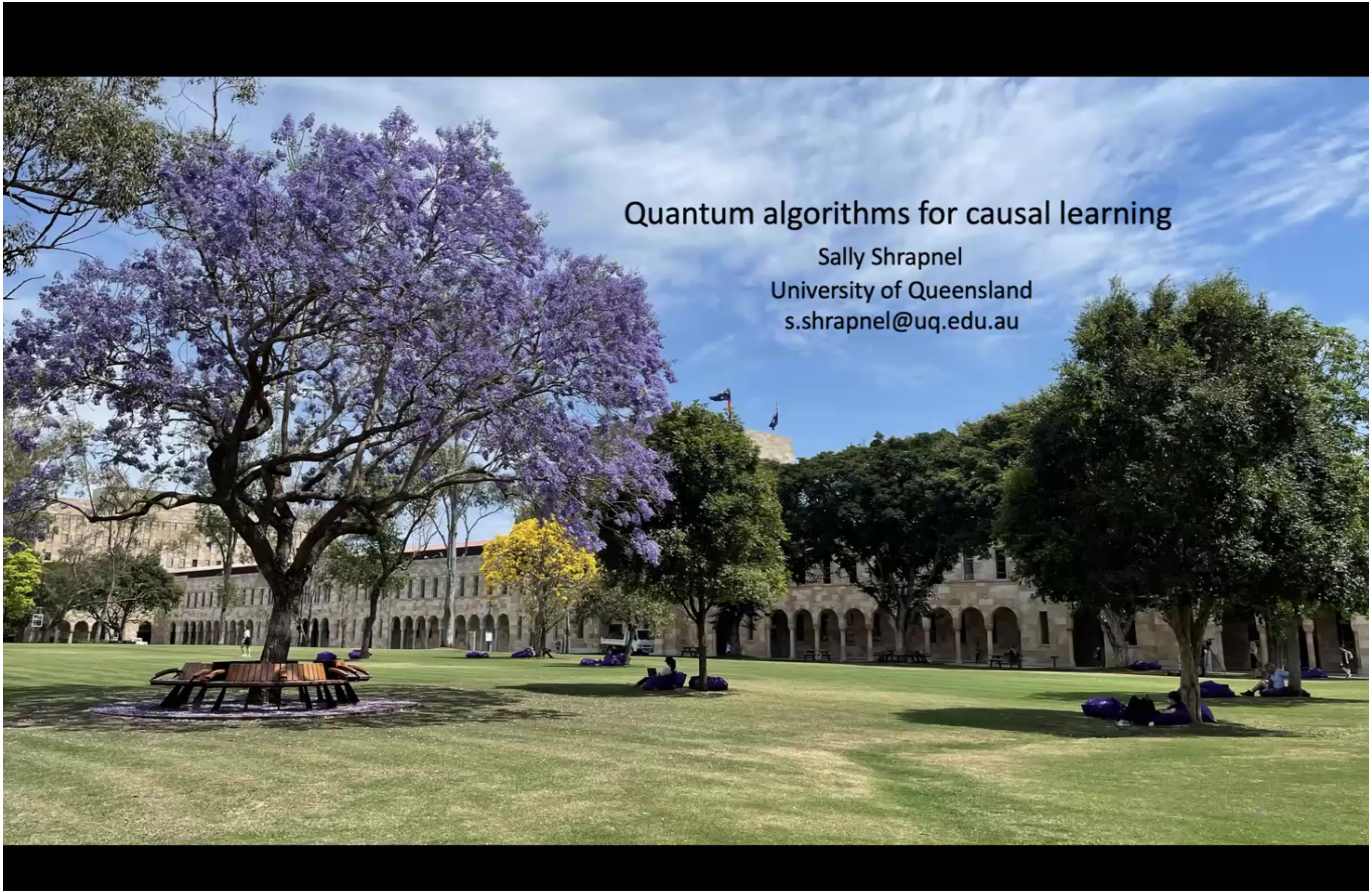
Abstract: Given the large number of proposed quantum machine learning (QML) algorithms, it is somewhat surprising that ideas from this field have not yet been extended to causal learning. While deep learning and generative machine learning models have taken centre stage in the industrial application of automated learning on classical data, it is nonetheless well known that these techniques don't reliably capture causal concepts, leading to significant performance vulnerabilities. Increasingly, classical ML experts are taking ideas from causal inference, a field traditionally limited to small data sets of low dimensionality, and injecting modern ML elements to create new algorithms that benefit from the best of both worlds. These hybrid classical approaches provide new opportunity to search for potential quantum advantage. In this talk I explore this new research direction and propose several new quantum algorithms for classical causal inference.

Quantum algorithms for causal learning

Sally Shrapnel

18 Sept 2024

1


Quantum algorithms for causal learning
Sally Shrapnel
University of Queensland
s.shrapnel@uq.edu.au

Interested in causality?

- Bell violations? 😱 arXiv:1512.07106
- Is quantum causal structure epistemic or ontic? arXiv:1702.01845, arXiv:1708.00137
- What are physical “interventions”? arXiv:1809.03191
- Can we use AI to learn quantum causal structure? arXiv:1901.05158
- Are quantum causal agents more efficient learners? arXiv:207.04426
- Causal perspectivalism from thermodynamics? arXiv:200904121
- Casual, thermodynamic, temporal arrows at zero T? arXiv:210901998
- Can we have “unknown” quantum causal structure? arxiv:2403.10316

Learning about classical causes

Talk outline:

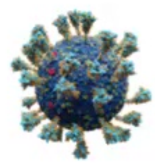
1. Introduce problem and describe family of classical non-parametric causal estimators[1]

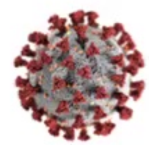
2. Quantize.[2]

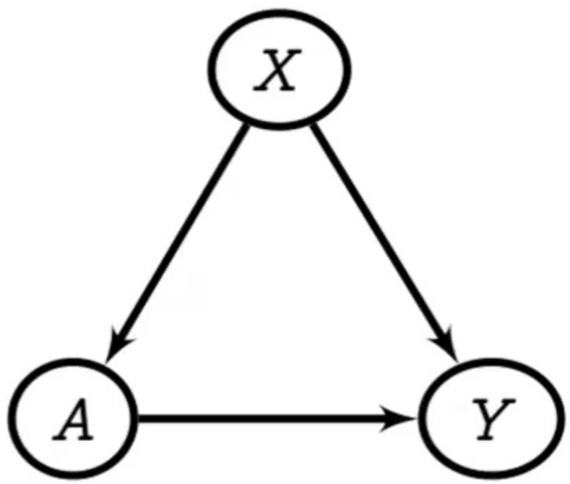
Rishi Goel

1. arXiv:2010.04855. 2. arXiv any day now...

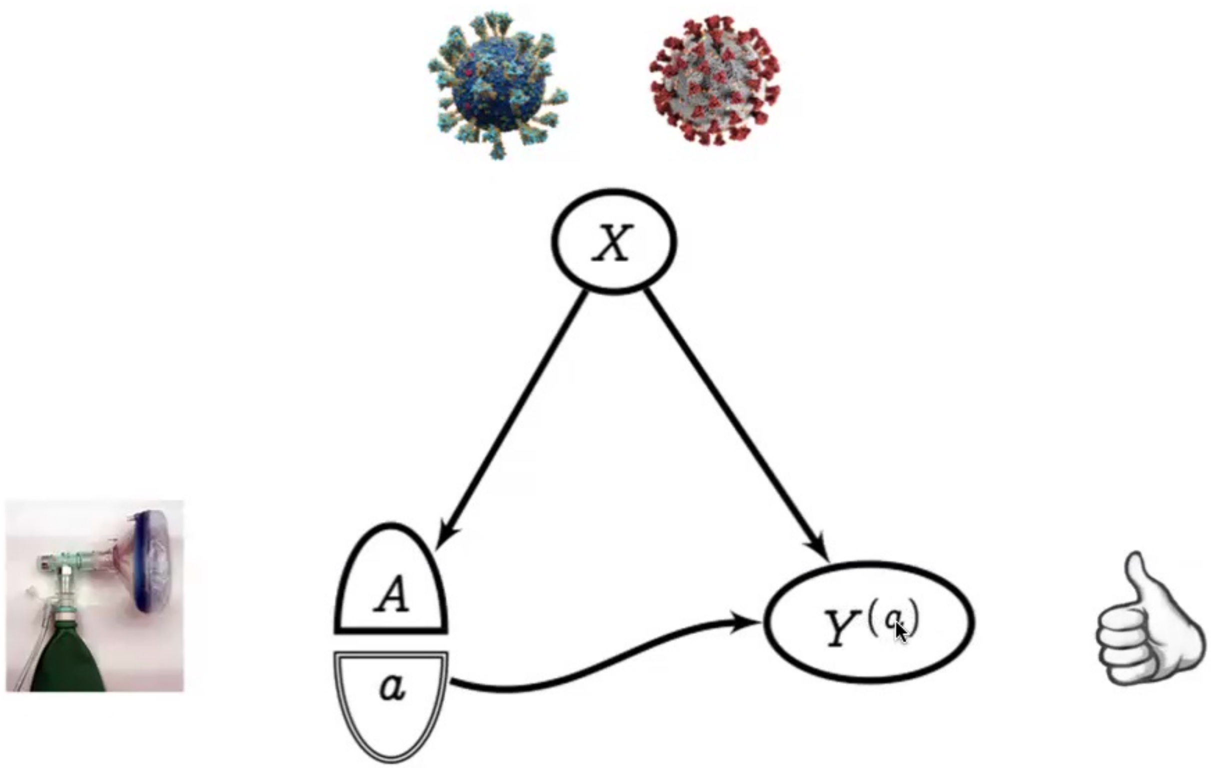
Observation vs intervention







Observation vs intervention



6

Causal estimation

- Average treatment effect (dose response)

$$\theta_0^{ATE}(a) = \mathbb{E}\{Y^{(a)}\}$$

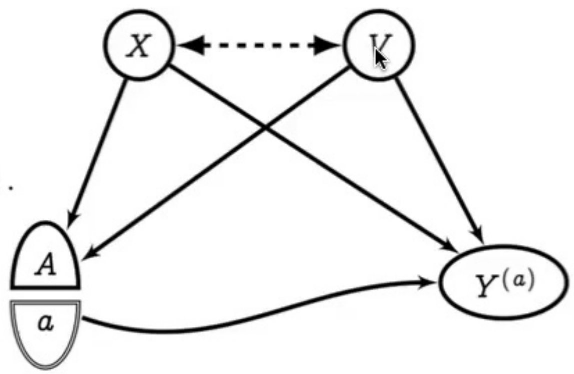
Causal estimation

- Conditional average treatment effect (heterogenous response)

$$\theta_0^{CATE}(a, v) = \mathbb{E}\{Y^{(a)} \mid V = v\}$$

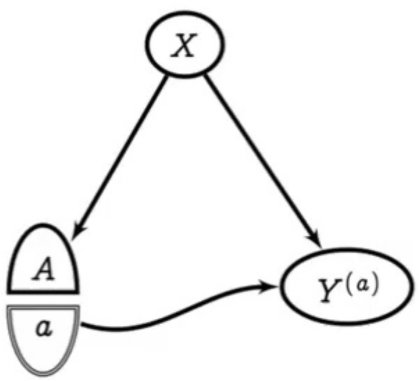
Causal estimation

- Conditional average treatment effect (heterogenous response)



Non-parametric, continuous estimators

- Dose response $$\theta_0^{ATE}(a) = \int \gamma_0(Y|A = a, X = x)d\mathbf{P}(x)$$
- Heterogenous resp. $$\theta_0^{CATE}(a, v) = \int \gamma_0(Y|A = a, V = v, X = x)d\mathbf{P}(x|v)$$

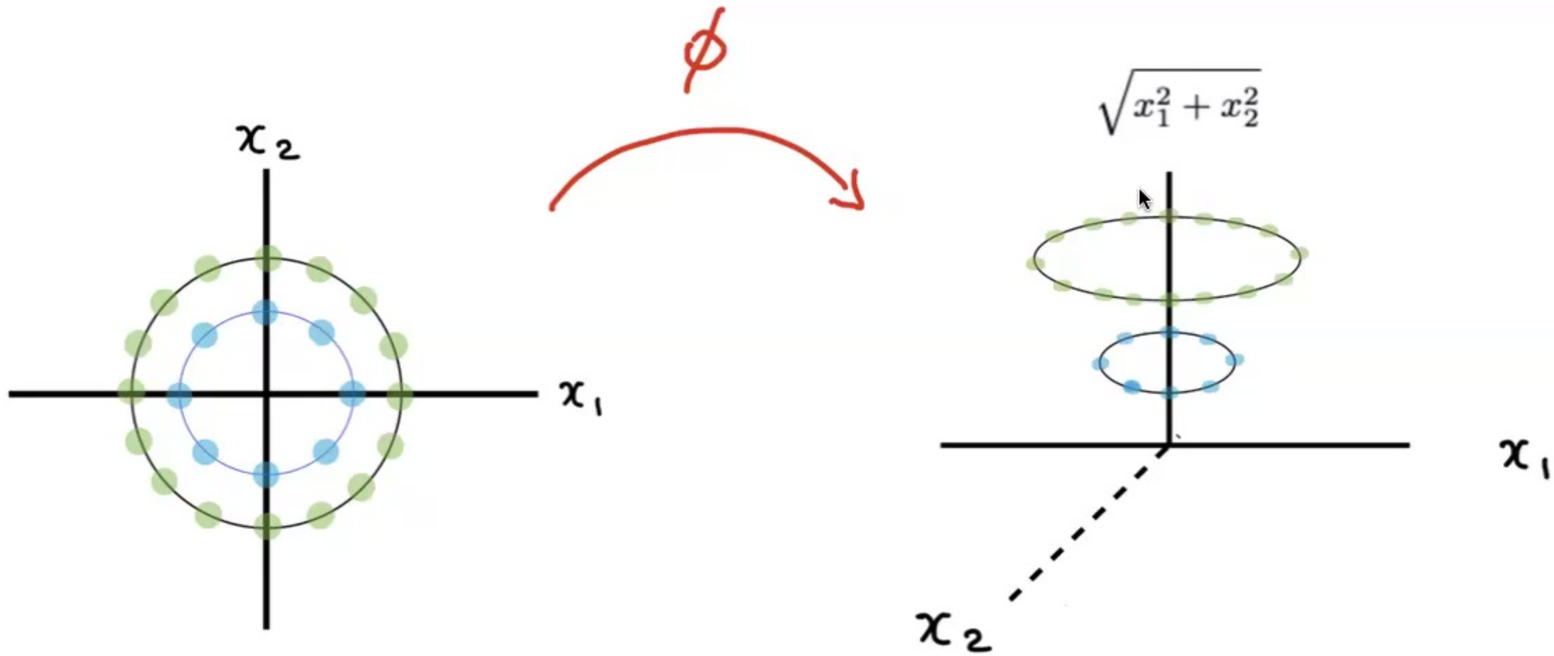


Assumptions:

1. No interference
2. Conditional exchangeability
3. Overlap

Kernel ML

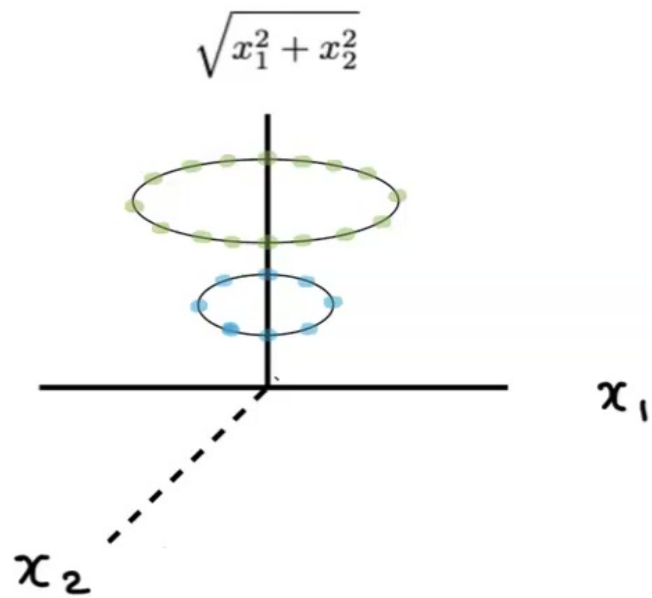
$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$



Kernel ML

$$\phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ \sqrt{x_1^2 + x_2^2} \end{bmatrix}$$

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$



12

RKHS

- Defined as the closure of the $\mathrm{span}\{\phi(x)\}$, such that $\{\phi(x)\}_{x\in\mathcal{X}}$ form the basis function of the RKHS

- The kernel is defined as the symmetric, continuous, p.d. function:

$$k(x,x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

- Reproducing property:

$$\hat{\gamma}(x) = \langle \hat{\gamma}, \phi(x) \rangle_{\mathcal{H}}$$

Kernel ridge regression

Goal is learn $\gamma_0(x) := \mathbb{E}[Y \mid X = x]$ from features $\phi(x_i)$ and outcomes y_i

$$\hat{\gamma}_0 = \underset{\gamma_0 \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \langle \gamma_0, \phi(x_i) \rangle_{\mathcal{H}}\right)^2 + \lambda \|\gamma_0\|_{\mathcal{H}}^2$$

$$\gamma = \sum_{i=1}^{n} \alpha_i \phi(x_i)$$

Kernel ridge regression

$$\hat{\gamma}_0 = \underset{\gamma_0 \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \langle \gamma_0, \phi(x_i) \rangle_{\mathcal{H}}\right)^2 + \lambda \|\gamma_0\|_{\mathcal{H}}^2$$

$$\langle \gamma_0, \phi(x_i) \rangle = \sum_j \alpha_j \langle \phi(x_j), \phi(x_i) \rangle \qquad \|\gamma\|^2 = \sum_i \sum_j \alpha_i \alpha_j \langle \phi(x_j), \phi(x_i) \rangle$$

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}^d}{\operatorname{argmin}} \|y - K\alpha\|^2 + \lambda \alpha^\top K \alpha$$

$$= (K + \lambda I_n)^{-1} y$$

$$\hat{\gamma}(x) = Y^\top (K_{XX} + n\lambda I)^{-1} K_{Xx}$$

Dose response estimation

$$\gamma_0(a, x) := \mathbb{E}[Y \mid a, x]$$

$$\phi(x, a) = \phi(a) \otimes \phi(x)$$

$$\gamma_0(a, x) := \langle \gamma_0, \phi(a) \otimes \phi(x) \rangle$$

$$\hat{\theta}^{ATE}(a) = \frac{1}{n} \sum_{i=1}^{n} Y^T (K_{AA} \odot K_{XX} + n\lambda I)^{-1} (K_{Aa} \odot K_{Xx})$$

Dose response estimation

$$\hat{\theta}^{ATE}(a) = \frac{1}{n}\sum_{i=1}^{n} Y^T (K_{AA} \odot K_{XX} + n\lambda I)^{-1}(K_{Aa} \odot K_{Xx})$$

Convergence guarantees $\mathcal{O}(n^{-1/4})$

Two complexity bottlenecks:

1. Exact inversion scales as $O(N^3)$
2. Evaluating each kernel matrix scales as $O(N^2)$

Linear solvers

$$\hat{\theta}^{ATE}(a) = \frac{1}{n}\sum_{i=1}^{n} Y^T \underline{(K_{AA} \odot K_{XX} + n\lambda I)}^{-1}(K_{Aa} \odot K_{Xx})$$

Quantum linear system solvers: $\mathcal{O}(poly \log(N))$

Algorithm	Complexity
HHL	$\mathcal{O}(\kappa^2/\epsilon)$
Variable Time Amplitude Amplification	$\mathcal{O}(\kappa(\log(\kappa)/\epsilon)^3)$
Fourier/Chebyshev fitting using LCU	$\mathcal{O}(\kappa(\text{polylog}(\kappa/\epsilon))$
Adiabatic randomization	$\mathcal{O}(\kappa \log(\kappa)/\epsilon)$
Time optical adiabatic method	$\mathcal{O}(\kappa(\text{polylog}(\kappa/\epsilon))$
Zeno eigenstate filtering	$\mathcal{O}(\kappa \log(\kappa/\epsilon))$
Discrete adiabatic theorem	$\mathcal{O}(\kappa \log(1/\epsilon))$

Quantum dose response

Algorithm 1 Quantum Dose Response $\hat{\theta}_0^{ATE}(a)$ $(\hat{\theta}_0^{DS}(a))$

Input: Kernel entries: $K_{AA}, K_{XX}, K_{Aa}, \sum_{x_i} K_{Xx_i}$ $(\sum_{\tilde{x_i}} K_{X\tilde{x_i}})$, regularization parameter λ, number of data points n $(\tilde{n})$, vector Y.

Output: Estimate of $\hat{\theta}_0^{ATE}(a)$ $(\hat{\theta}_0^{DS}(a))$

1: Form matrix $A = K_{AA} \odot K_{XX} + n\lambda I$.
2: Compute vector $b = \frac{1}{n} K_{Aa} \odot \sum_{x_i} K_{Xx_i}$ $(\frac{1}{\tilde{n}} K_{Aa} \odot \sum_{\tilde{x_i}} K_{X\tilde{x_i}})$.
3: Normalize matrix A, vectors b, and Y by their respective norms.
4: Encode matrix A into a block encoding.
5: Prepare quantum states $|b\rangle$ and $|Y\rangle$ for vectors b and Y.
6: Use a quantum adiabatic solver to obtain $|A^{-1}b\rangle$.
7: Measure overlap $\langle Y|A^{-1}b\rangle$.
8: Reintroduce the normalisation factors for A_1, A_2, b, Y.

Quantum dose response convergence rates

Classically, convergence scales as

- Assuming smoothness and fast spectra decay $\mathcal{O}(n^{-1/4})$

Discrete adiabatic algorithm: $\mathcal{O}(\text{poly}\log(N))$; $\mathcal{O}(\kappa\log(1/\epsilon))$

- Estimation of inner product: $1/\epsilon_k^2$

Quantum kernels?

Algorithm 1 Quantum Dose Response $\hat{\theta}_0^{ATE}(a)$ $(\hat{\theta}_0^{DS}(a))$

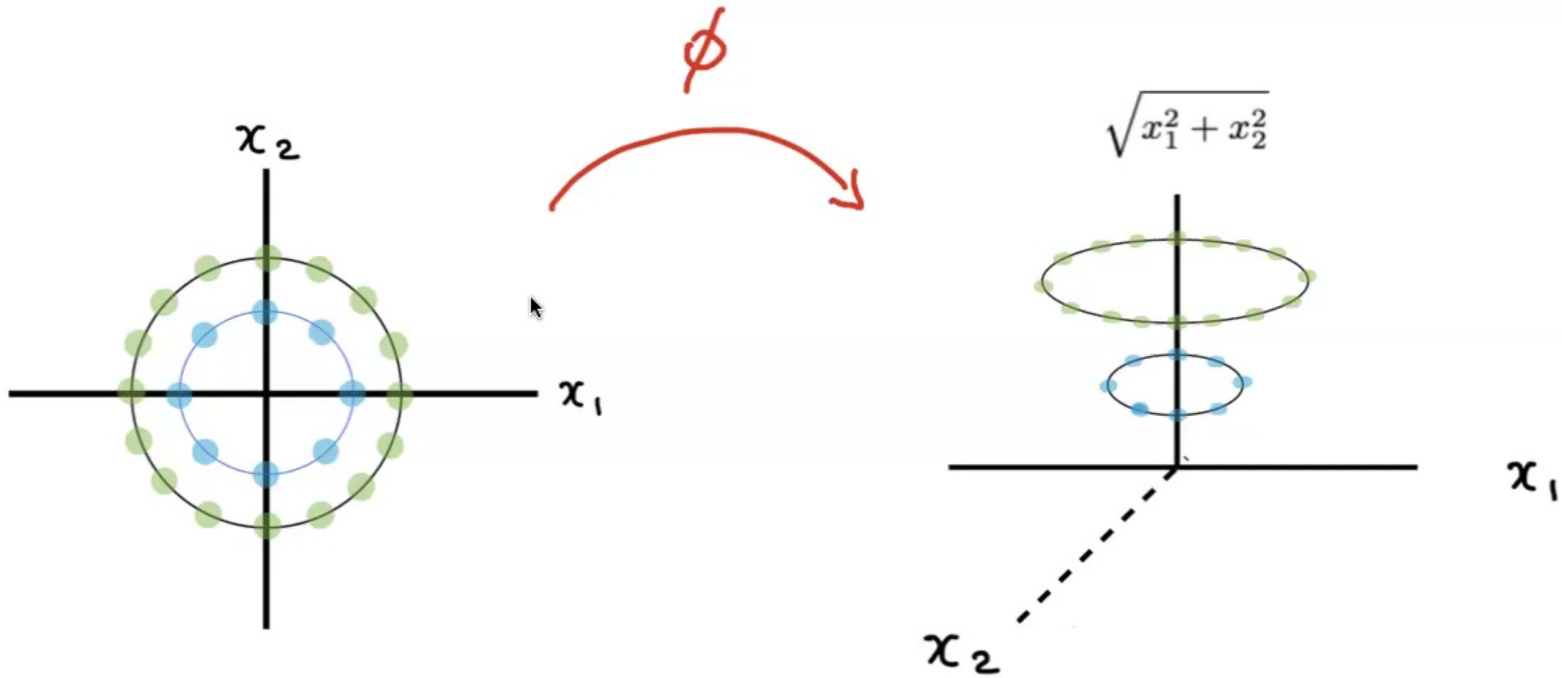
Input: Kernel entries: $K_{AA}, K_{XX}, K_{Aa}, \sum_{x_i} K_{Xx_i}$ $(\sum_{\tilde{x_i}} K_{X\tilde{x_i}})$, regularization parameter λ, number of data points n $(\tilde{n})$, vector Y.

Output: Estimate of $\hat{\theta}_0^{ATE}(a)$ $(\hat{\theta}_0^{DS}(a))$

1: Form matrix $A = K_{AA} \odot K_{XX} + n\lambda I$.
2: Compute vector $b = \frac{1}{n} K_{Aa} \odot \sum_{x_i} K_{Xx_i}$ $(\frac{1}{\tilde{n}} K_{Aa} \odot \sum_{\tilde{x_i}} K_{X\tilde{x_i}})$.
3: Normalize matrix A, vectors b, and Y by their respective norms.
4: Encode matrix A into a block encoding.
5: Prepare quantum states $|b\rangle$ and $|Y\rangle$ for vectors b and Y.
6: Use a quantum adiabatic solver to obtain $|A^{-1}b\rangle$.
7: Measure overlap $\langle Y|A^{-1}b\rangle$.
8: Reintroduce the normalisation factors for A_1, A_2, b, Y.

Quantum kernels?

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$



Need characteristic kernels: arXiv:2401.05647

Other causal estimators?

Heterogeneous dose response:

$$\hat{\theta}_0^{CATE}(d,v) = Y^T(K_{DD} \odot K_{VV} \odot K_{XX} + n\lambda I)^{-1}[K_{Dd} \odot K_{Vv} \odot \{K_{XX}(K_{VV} + n\lambda_2 I)^{-1}K_{Vv}\}]$$

Query complexity

$$\hat{\theta}^{CATE}(a,v) \quad \mathcal{O}(\kappa_1 \kappa_2 \log(1/\epsilon_1) \log(1/\epsilon_2))$$

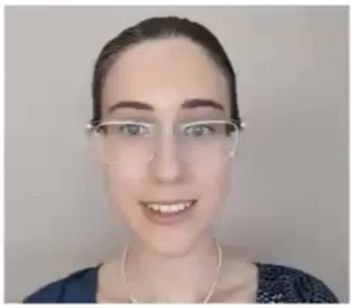
Other causal estimators?

1. Dose response with distribution shift
2. Average treatment on treated
3. Front door and back door identification
4. Causal discovery using KCIT

Lots of opportunities for further research!

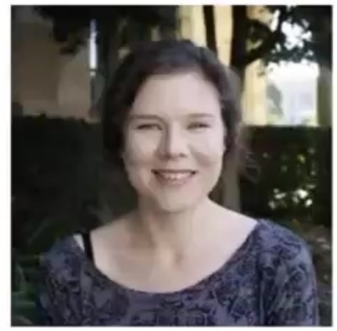
Algorithm	Quantum Query Complexity
$\hat{\theta}^{ATE}(a)$	$\mathcal{O}(\kappa \log(1/\epsilon))$
$\hat{\theta}^{DS}(a)$	$\mathcal{O}(\kappa \log(1/\epsilon))$
$\hat{\theta}^{ATT}(a, a')$	$\mathcal{O}(\kappa_1 \kappa_2 \log(1/\epsilon_1) \log(1/\epsilon_2))$
$\hat{\theta}^{CATE}(a, v)$	$\mathcal{O}(\kappa_1 \kappa_2 \log(1/\epsilon_1) \log(1/\epsilon_2))$

Group



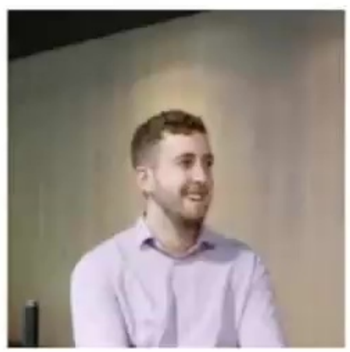
Dr Laura Henderson

Rishi Goel

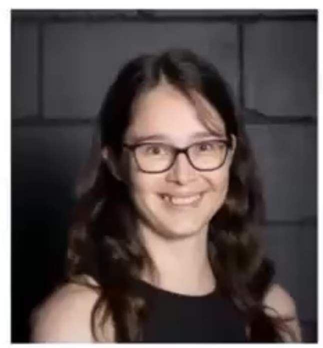


Dr Carolyn Wood

Dr Riddhi Gupta



Connor Van Rossum



Aleesha Isaacs


Thank you for your attention!
26