WACQT | Wallenberg Center for Quantum Technology

XANADU

VECTOR INSTITUTE

Implicit differentiation of variational quantum algorithms

CHALMERS
UNIVERSITY OF TECHNOLOGY

Shahnawaz Ahmed
Ph.D. student, Chalmers

Mitacs ACCELERATE

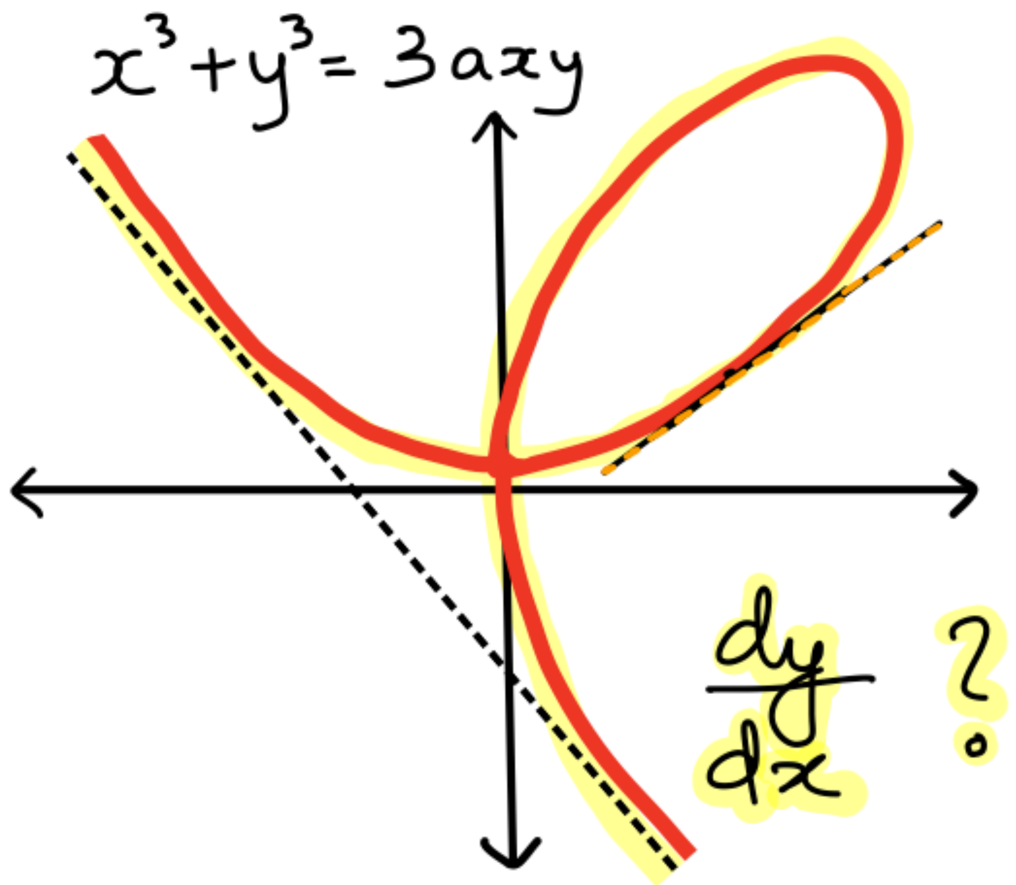
Juan Felipe Carrasquilla Álvarez
Vector Institute for AI, UWaterloo

Nathan Killoran
Xanadu

Astra Zeneca, Gothenburg, March 22

Outline

1. **Implicit differentiation**

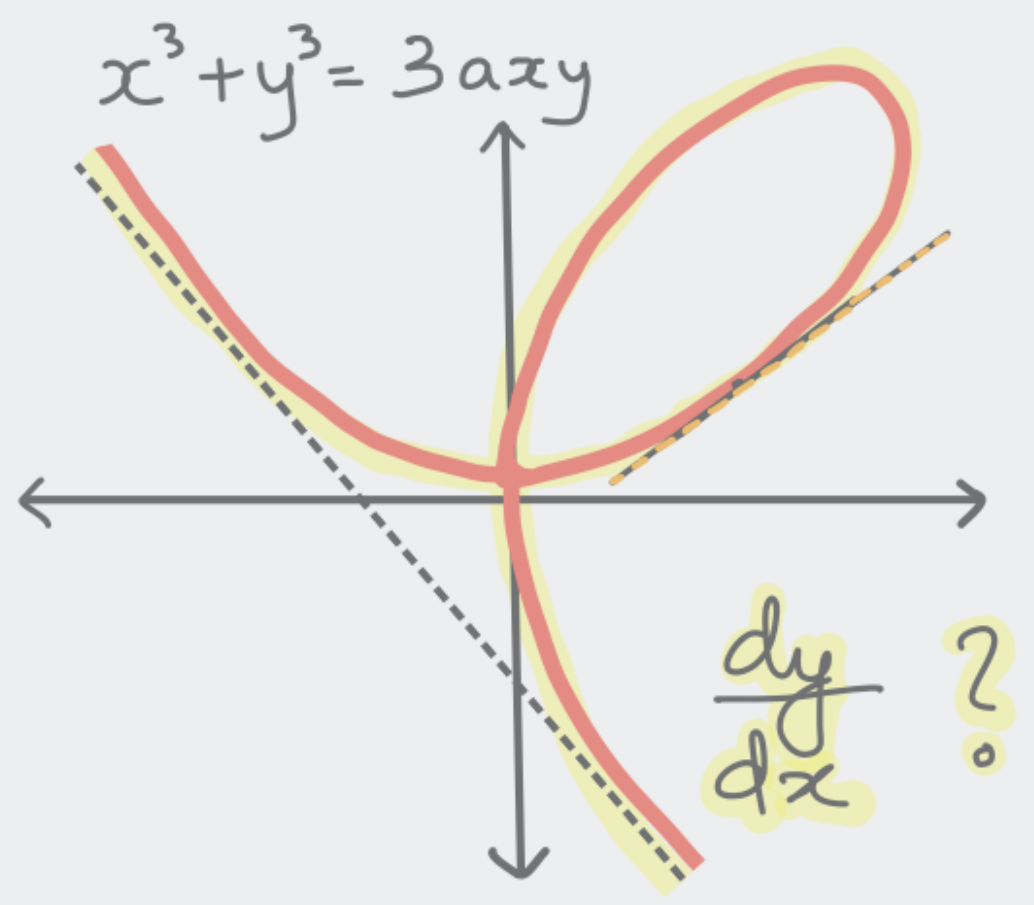


Descartes

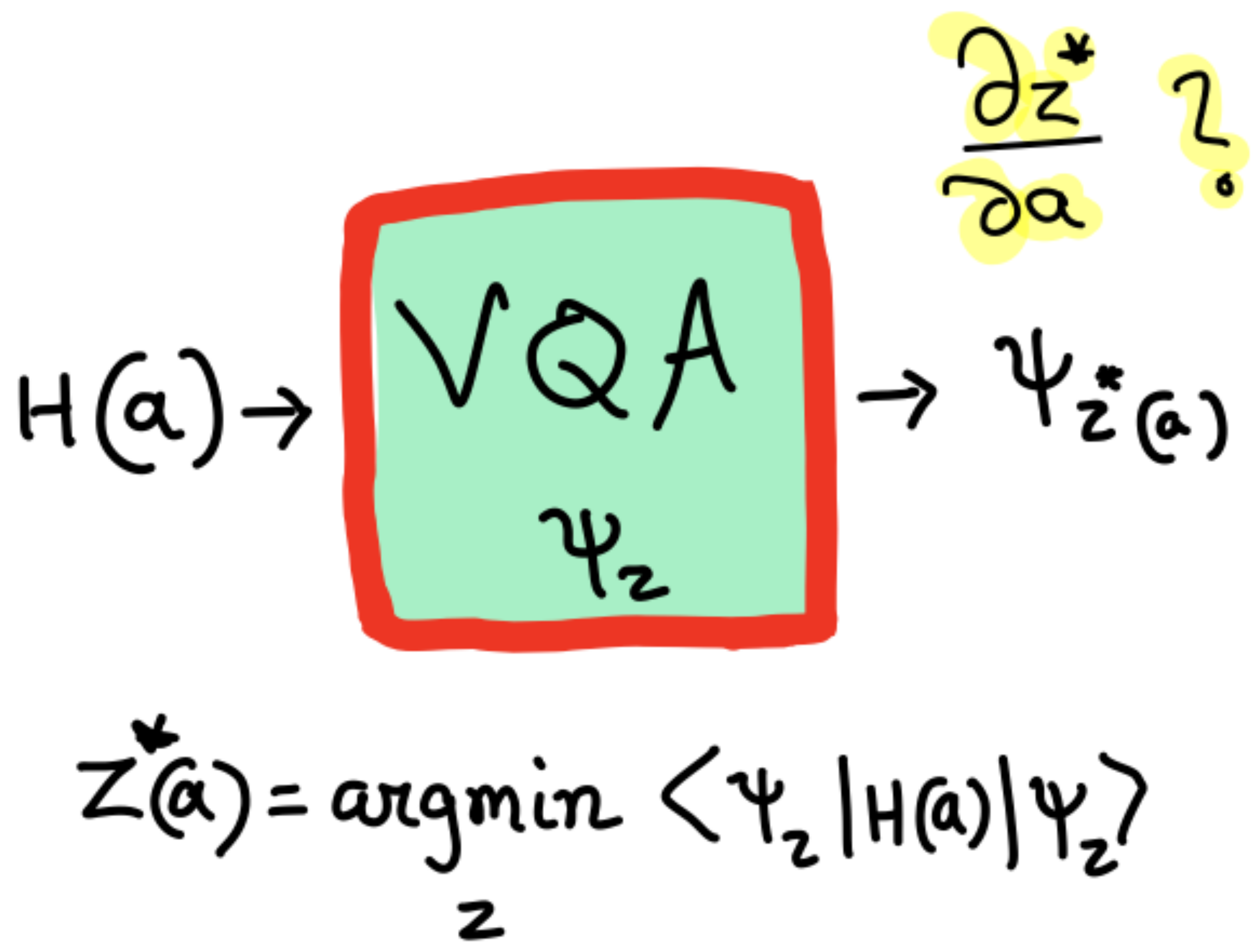
Fermat

Outline

1. Implicit differentiation

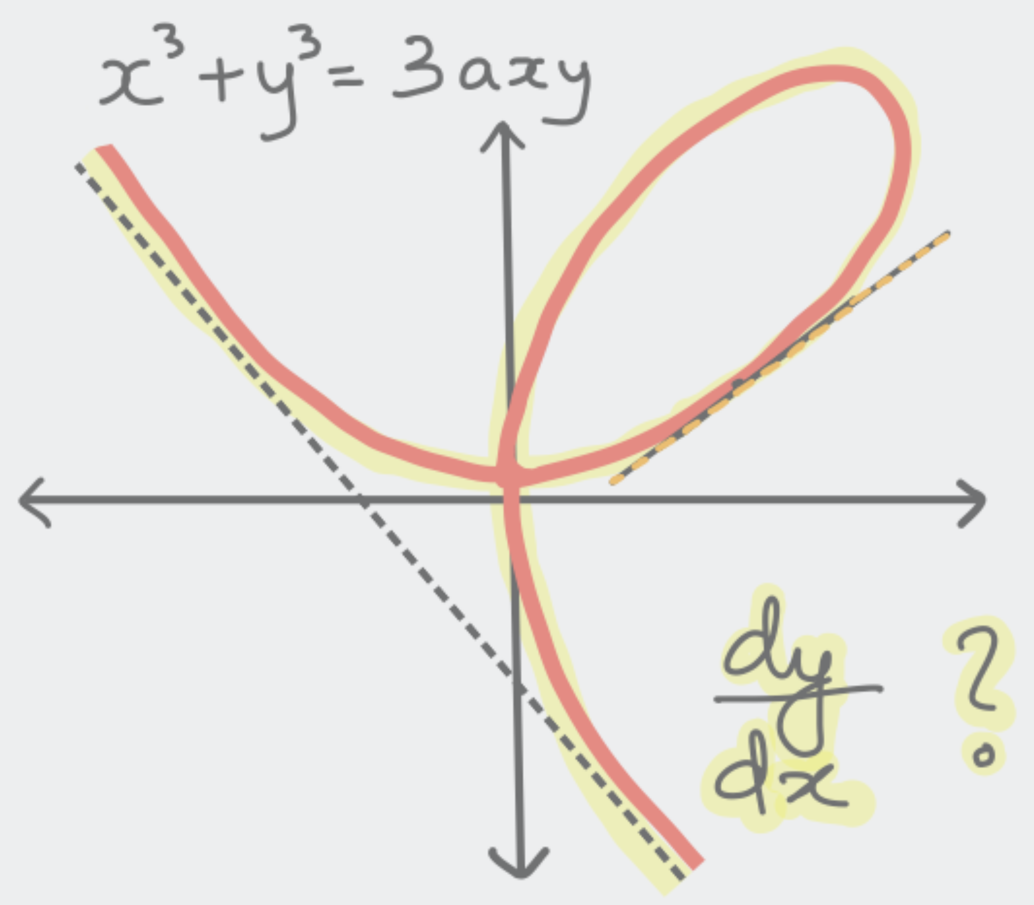


2. Gradients of solutions to variational quantum algorithms

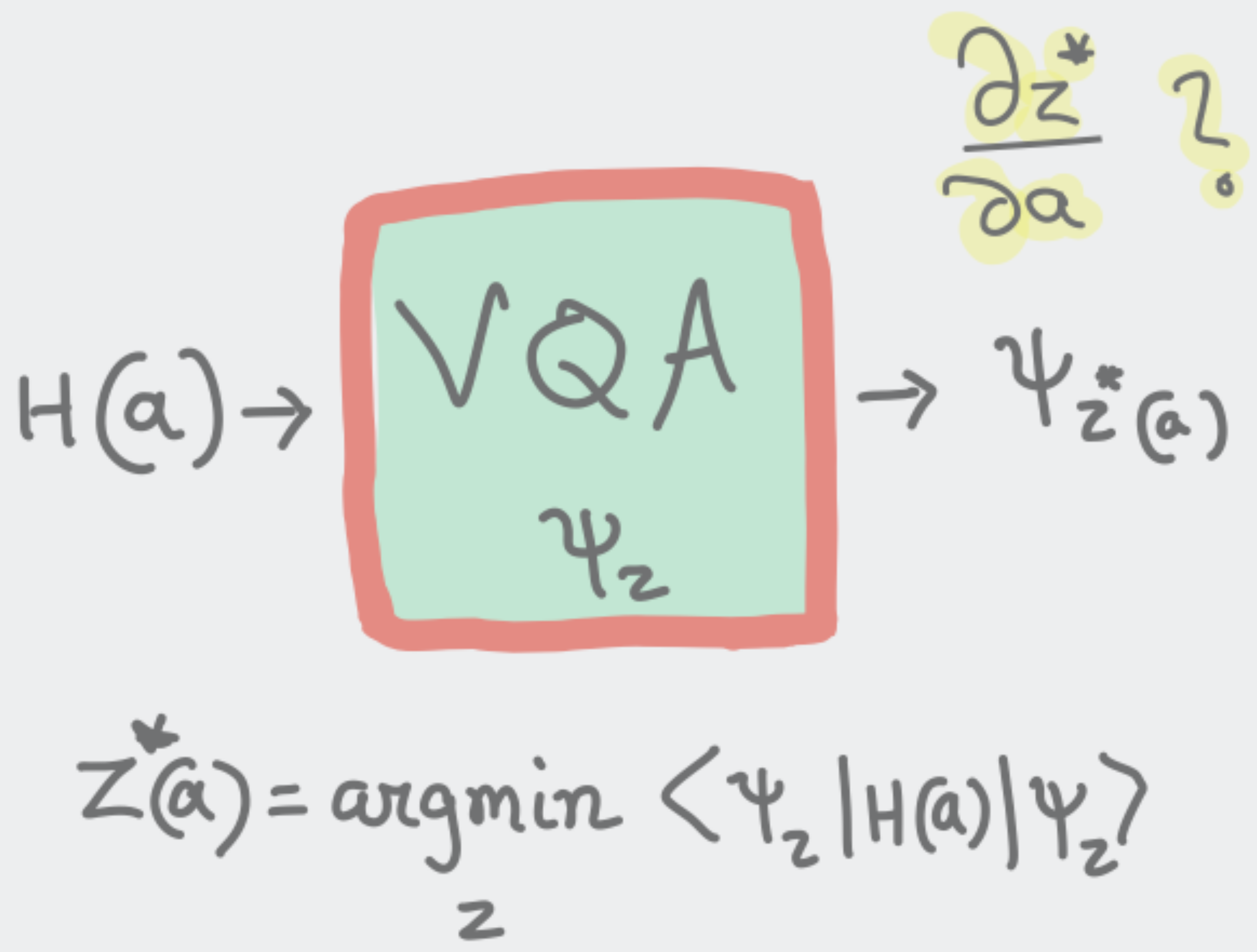


Outline

1. Implicit differentiation

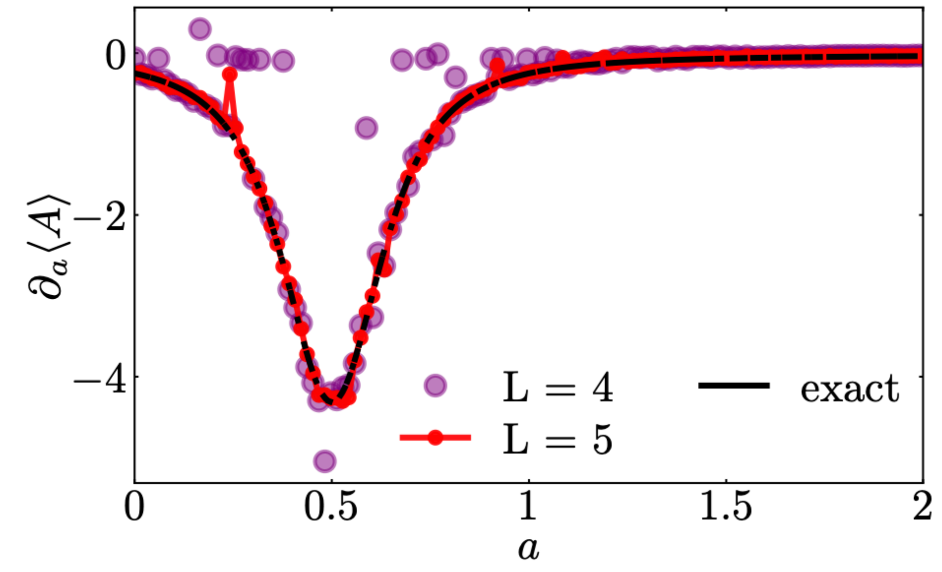


2. Gradients of solutions to variational quantum algorithms

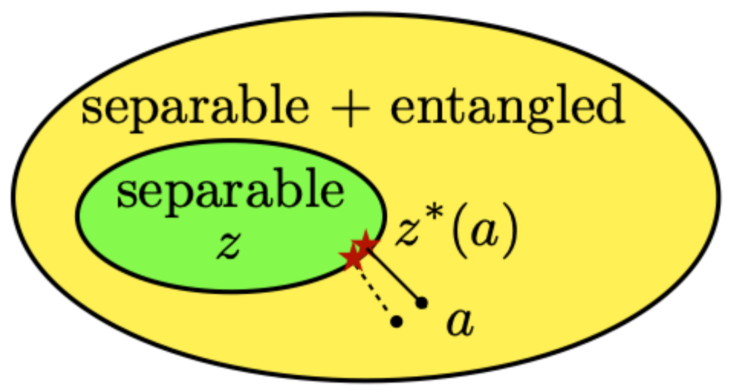


3. Applications

Ground-state gradients

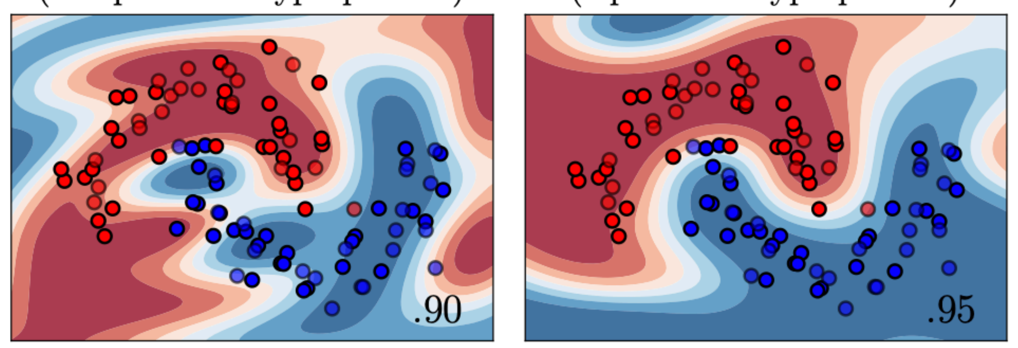


Entanglement generation



$$z^*(a) = \underbrace{\arg \max_z ||\langle \psi_z | \psi_a \rangle||}_{\Lambda_{\max}(a)}$$

(Unoptimized hyperparams) (Optimized hyperparams)



Paper and code

arXiv:2211.13765

pennylane.ai/qml/demos/tutorial_implicit_diff_susceptibility.html

Implicit differentiation of variational quantum algorithms

Shahnawaz Ahmed[1,2], Nathan Killoran[2], and Juan Felipe Carrasquilla Álvarez[3,4]

[1]Department of Microtechnology and Nanoscience, Chalmers University of Technology, 412 96 Gothenburg, Sweden
[2]Xanadu, Toronto, ON, M5G 2C8, Canada
[3]Vector Institute for Artificial Intelligence, MaRS Centre, Toronto, ON, Canada M5G 1M1
[4]Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Several quantities important in condensed matter physics, quantum information, and quantum chemistry, as well as quantities required in meta-optimization of machine learning algorithms, can be expressed as gradients of implicitly defined functions of the parameters characterizing the system. Here, we show how to leverage implicit differentiation for gradient computation through variational quantum algorithms and explore applications in condensed matter physics, quantum machine learning, and quantum information. A function defined implicitly as the solution of a quantum algorithm, e.g., a variationally obtained ground- or steady-state,

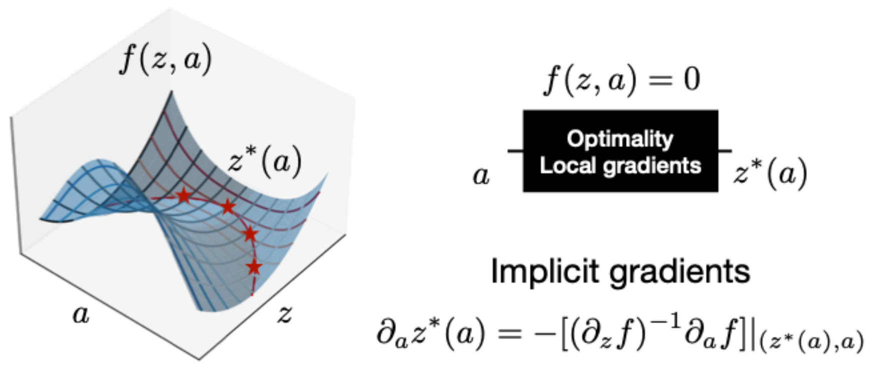


Figure 1: Automatic differentiation through a black-box algorithm that implicitly defines a solution function $z^*(a)$ to some problem using an optimality condition $f(z,a)=0$. Solution points (shown with the red stars) can be obtained using an iterative black-box solver without having access to the explicit solution. However, if we have access to local gradients $(\partial_z f, \partial_a f)$ at the solution points, implicit differentiation can compute the gradient of the solution function $\partial_a z^*(a)$ without differentiating

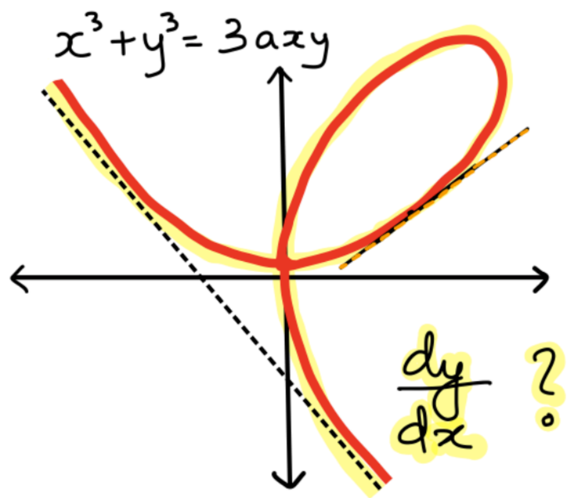
PENNYLANE Learn Demos Install Plugins Documentation Blog

Implicit differentiation of variational quantum algorithms

Authors: Shahnawaz Ahmed and Juan Felipe Carrasquilla Álvarez — Posted: 28 November 2022. Last updated: 28 November 2022.

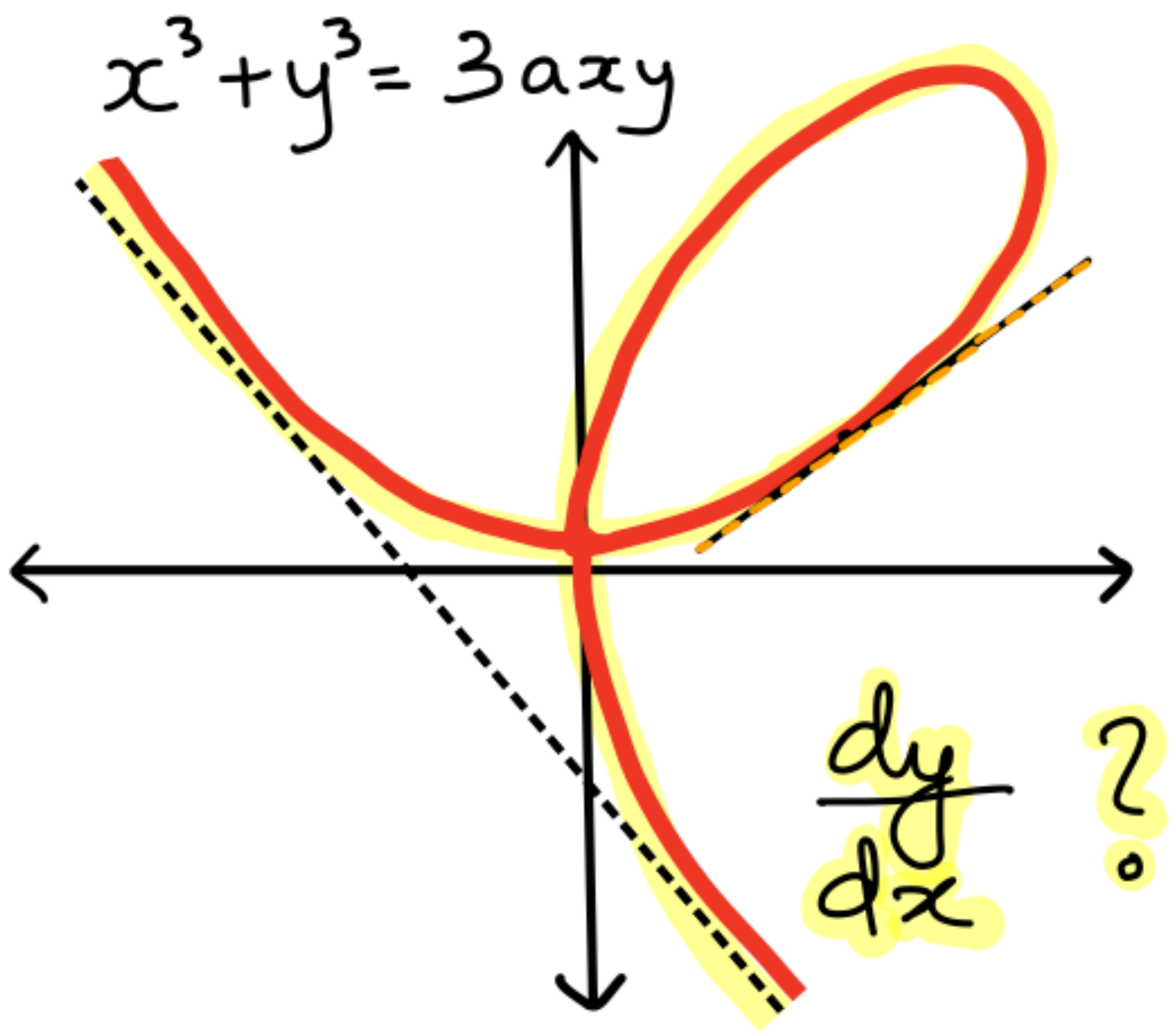
In 1638, René Descartes, intrigued by (then amateur) Pierre de Fermat's method of computing tangents, challenged Fermat to find the tangent to a complicated curve — now called the folium of Descartes:

$$x^3 + y^3 = 3axy.$$



Implicit differentiation

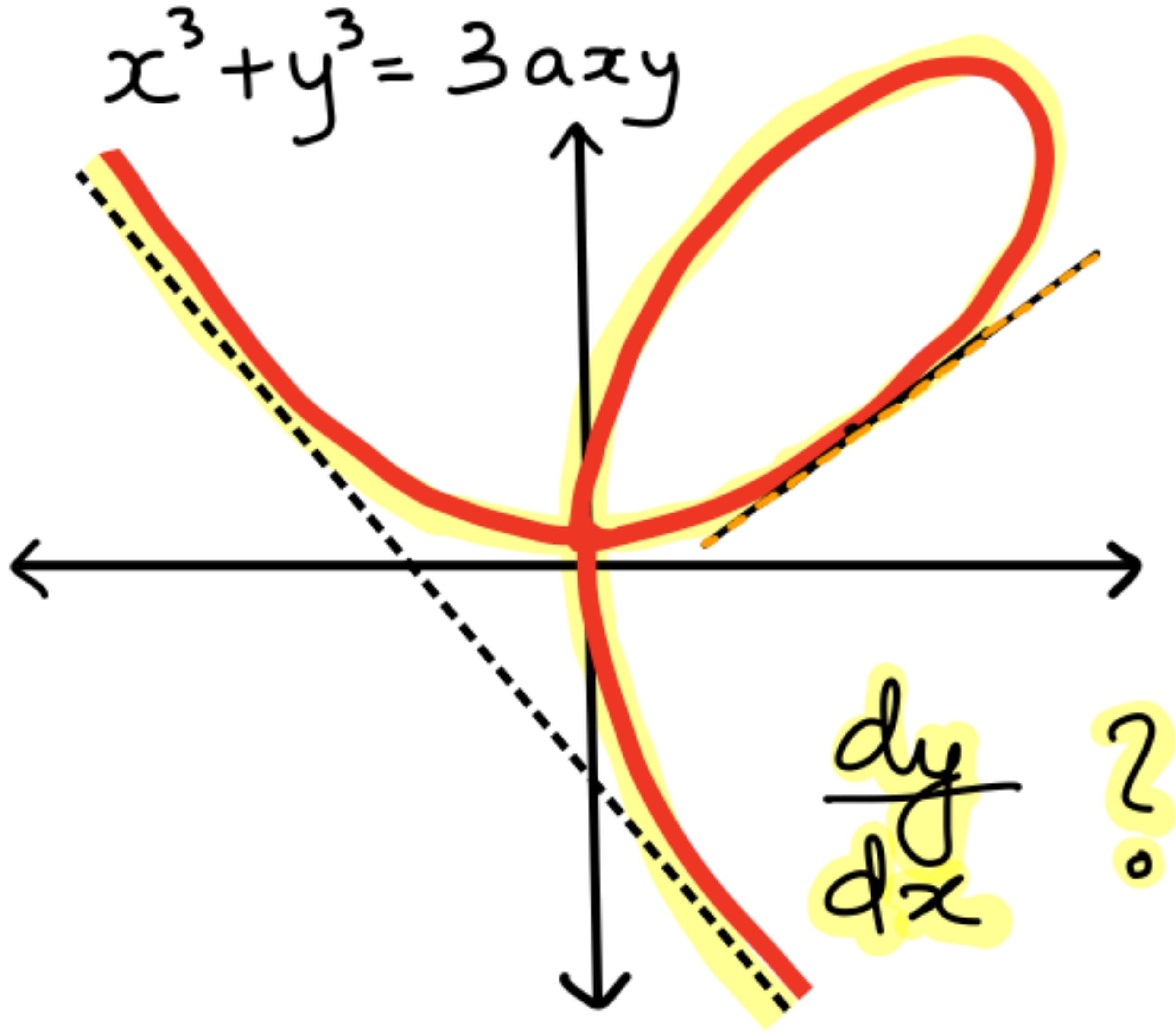
Descartes



Fermat

Implicit differentiation

Descartes



Folium of Descartes (1638)

$$\frac{dy}{dx} = -\frac{x^2 - ay}{y^2 - ax}$$

Fermat

[49] Jaume Paradís et al. Fermat and the Quadrature of the Folium of Descartes. The American Mathematical Monthly, 111:(3):216-229, 2004.

Implicit differentiation of solutions to variational quantum algorithms (VQAs)

Ground-state gradients

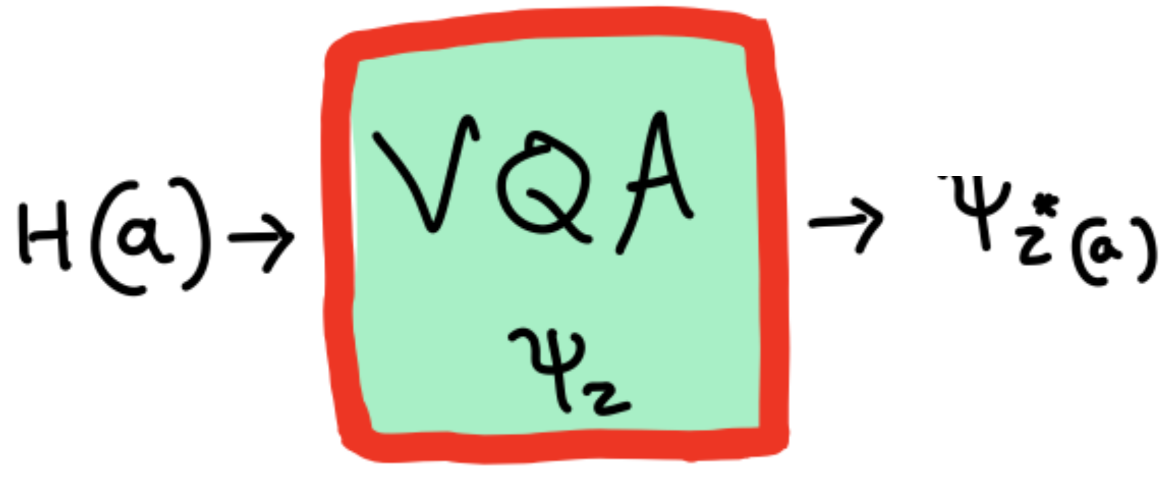
$$H(a) \rightarrow \boxed{\begin{matrix} \text{VQA} \\ \psi_z \end{matrix}} \rightarrow \psi_{z^*(a)}$$

$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} \; ?$$

Why is implicit differentiation interesting for quantum physics?

Ground-state gradients



$$\frac{\partial z^*}{\partial a} \; ?$$

$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

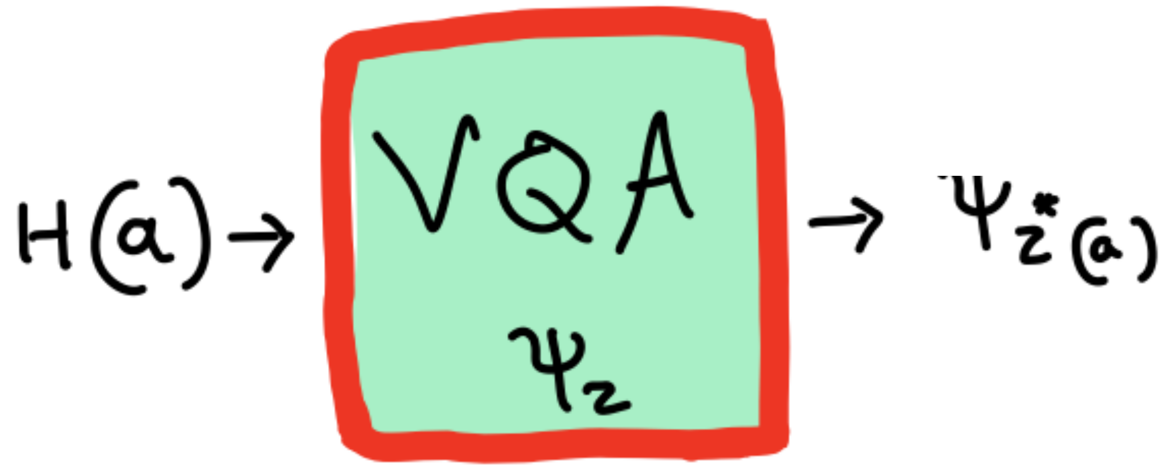
[26] Gergely Barcza, et al. **Ground-state properties** of the symmetric single-impurity anderson model on a ring from density- matrix renormalization group, hartree-fock, and gutzwiller theory. Phys. Rev. B, 99:165130, 2019.

[5] O'Brien et al. **Calculating energy derivatives for quantum chemistry** on a quantum computer. Npj Quantum Inf., 5(1):1–12, 2019.

[29] Juan Carrasquilla et al. **Scaling of the gap, fidelity susceptibility,** and bloch oscillations across the superfluid-to-mott-insulator transition in the one-dimensional bose-hubbard model. Phys. Rev. A, 87:043606, 2013.

Why is implicit differentiation interesting for quantum physics?

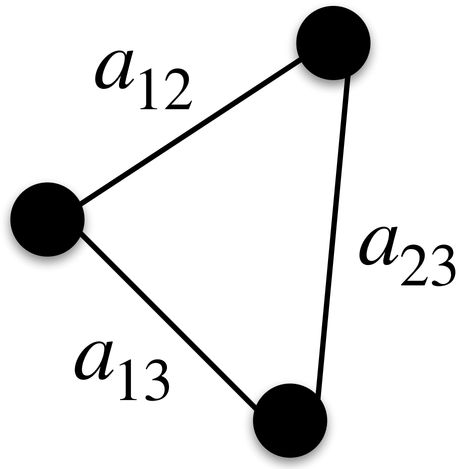
Ground-state gradients



$\frac{\partial z^*}{\partial a}$?

$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$

Geometry optimization



$H(\vec{a})$

[26] Gergely Barcza, et al. **Ground-state properties** of the symmetric single-impurity anderson model on a ring from density- matrix renormalization group, hartree-fock, and gutzwiller theory. Phys. Rev. B, 99:165130, 2019.

[5] O'Brien et al. **Calculating energy derivatives for quantum chemistry** on a quantum computer. Npj Quantum Inf., 5(1):1–12, 2019.

[29] Juan Carrasquilla et al. **Scaling of the gap, fidelity susceptibility,** and bloch oscillations across the superfluid-to-mott-insulator transition in the one-dimensional bose-hubbard model. Phys. Rev. A, 87:043606, 2013.

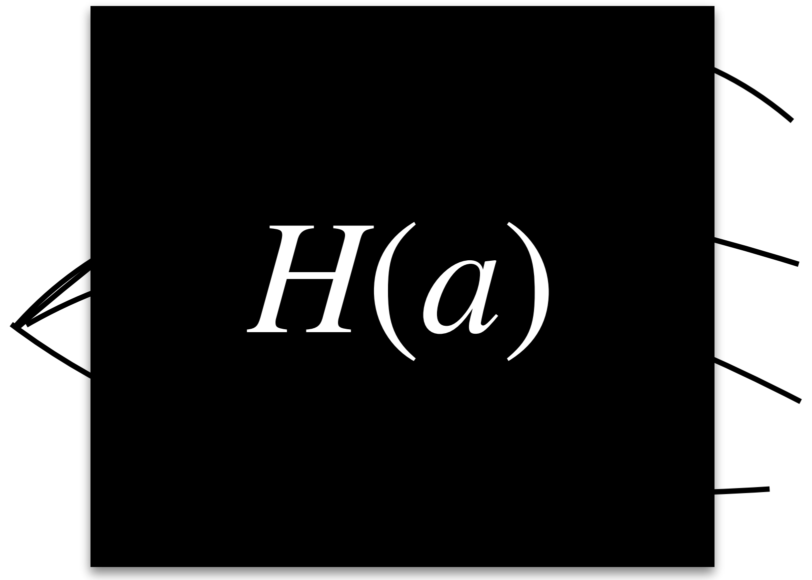
[6] P. Pulay. **Ab initio calculation of force constants and equilibrium** geometries in poly- atomic molecules. Mol. Phys., 17(2):197– 204, 1969.

[7] Kosuke Mitarai et al. **Theory of analytical energy derivatives for the variational quantum eigensolver**. Phys. Rev. Research, 2:013129, 2020.

[8] Ivan Kassal and Alán Aspuru-Guzik. **Quantum algorithm for molecular properties and geometry optimization**. J. Chem. Phys., 131(22):224102, 2009.

Why is implicit differentiation interesting for quantum physics?

Steady-state optimization, inverse design



$\rho_{ss}(a)$

a

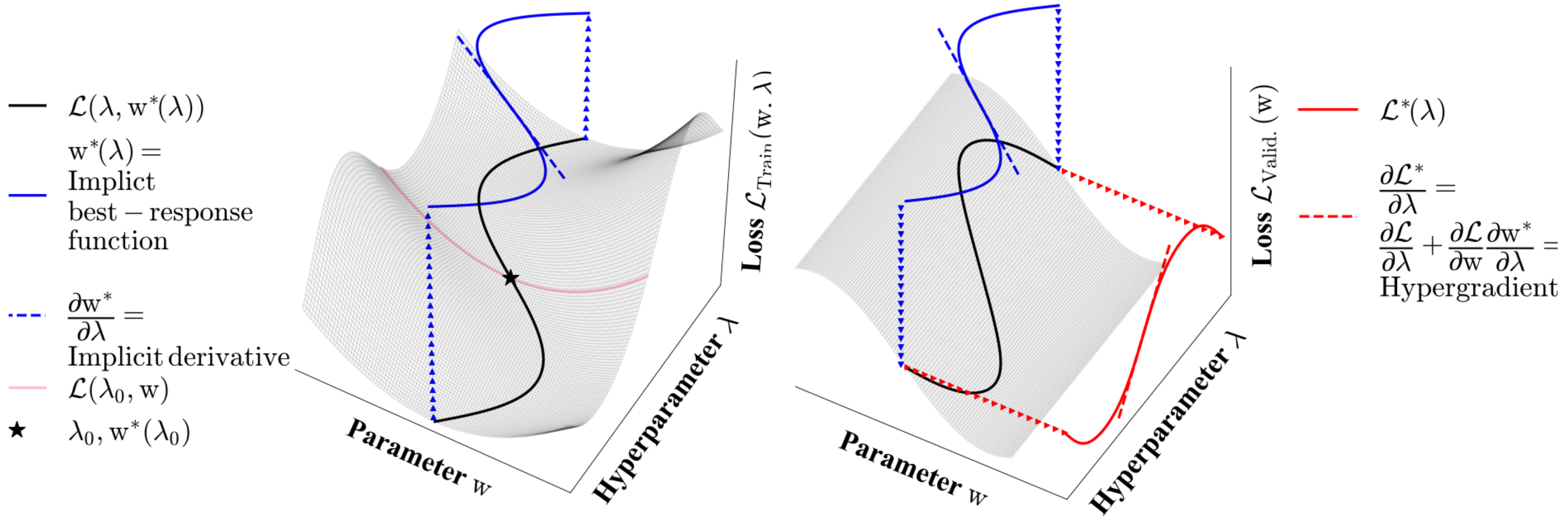
How to tune a to get desired steady state (ss)?

*did not work for control problems where there is no steady state naturally

[50] Rodrigo A. Vargas-Hernández, Ricky T. Q. Chen, Kenneth A. Jung, and Paul Brumer. **Fully differentiable optimization protocols for non-equilibrium steady states**. New J. Phys., 23(12):123006, 2021.

Why is implicit differentiation interesting for machine learning?

Machine learning, automatic differentiation and quantum circuits



[21] Jonathan Lorraine, Paul Vicol, and David Duvenaud. **Optimizing Millions of Hyper-parameters by Implicit Differentiation**. In Proceedings of the Twenty Third Interna- tional Conference on Artificial Intelligence and Statistics, pages 1540–1552. PMLR, 2020.

PENNYLANE

A cross-platform Python library for differentiable programming of quantum computers. Train a quantum computer the same way as a neural network.

JAX: High-Performance Array Computing

JAX is Autograd and XLA, brought together for high-performance numerical computing.

JAXopt

Hardware accelerated, batchable and differentiable optimizers in JAX.

[16] Mathieu Blondel et al. **Efficient and Modular Implicit Differentiation** (arXiv:2105.15183), 2021.

Can’t we just use autodifferentiation ?

Ground-state gradients

$$H(a) \rightarrow \boxed{\text{VQA} \quad \psi_z} \rightarrow \psi_{z^*(a)}$$

$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} \ ?$$

High memory cost through unrolling of the VQAs

JAX, M.D.
A Framework for Differentiable Physics

Samuel S. Schoenholz
Google Research: Brain Team
schsam@google.com

Ekin D. Cubuk
Google Research: Brain Team
cubuk@google.com

```
RuntimeError: RESOURCE_EXHAUSTED: Out of memory while trying to allocate 7935373600 bytes.
```

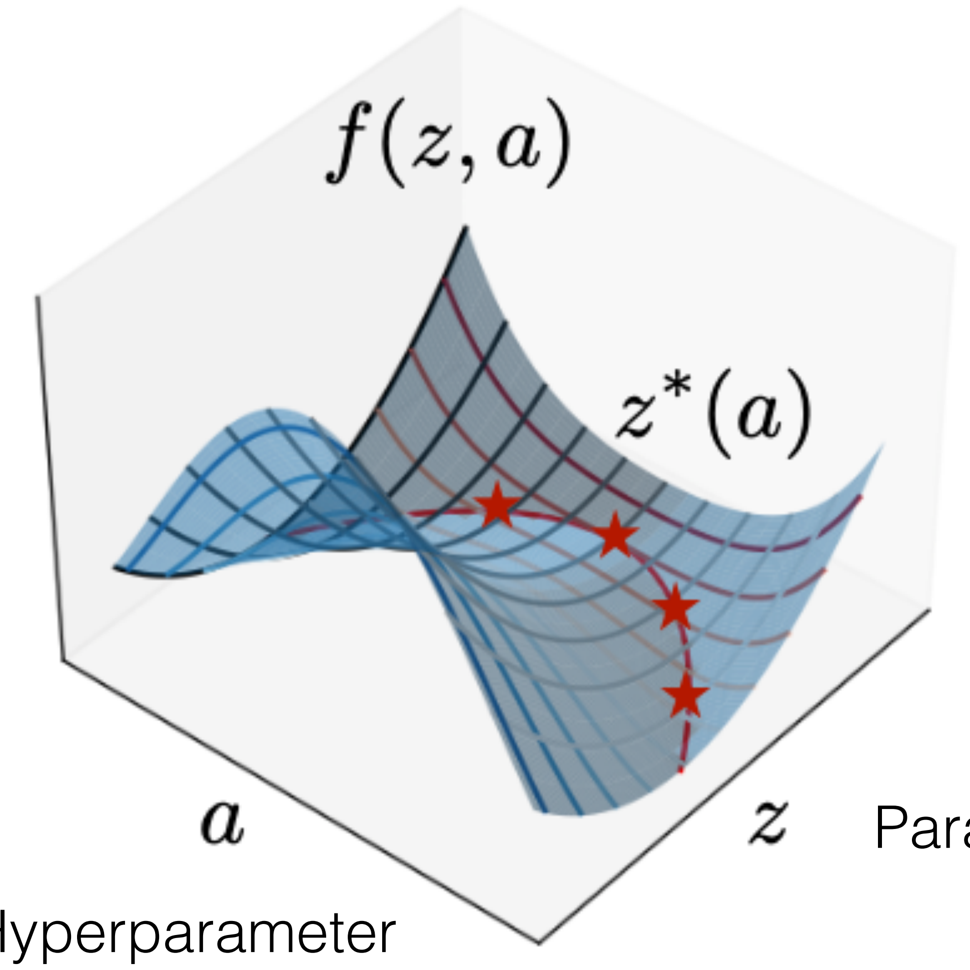
SEARCH STACK OVERFLOW

As you can see our computation fails due to running out of memory when trying to compute the gradient. The reason for that is that for reverse mode differentiation (e.g. `jax.grad`) the memory consumption grows linearly with respect to the number of optimization steps we have to take since reverse mode differentiation needs to store the whole forward pass in order to compute the gradient in the backwards pass.

We could reduce the memory requirements by using a technique called gradient rematerialization/checkpointing for a corresponding increase in computation time. While this strategy works, there exists a better solution for our problem called implicit differentiation.

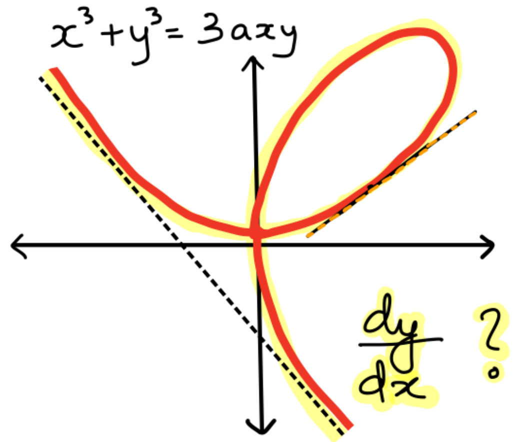
Implicit Differentiation

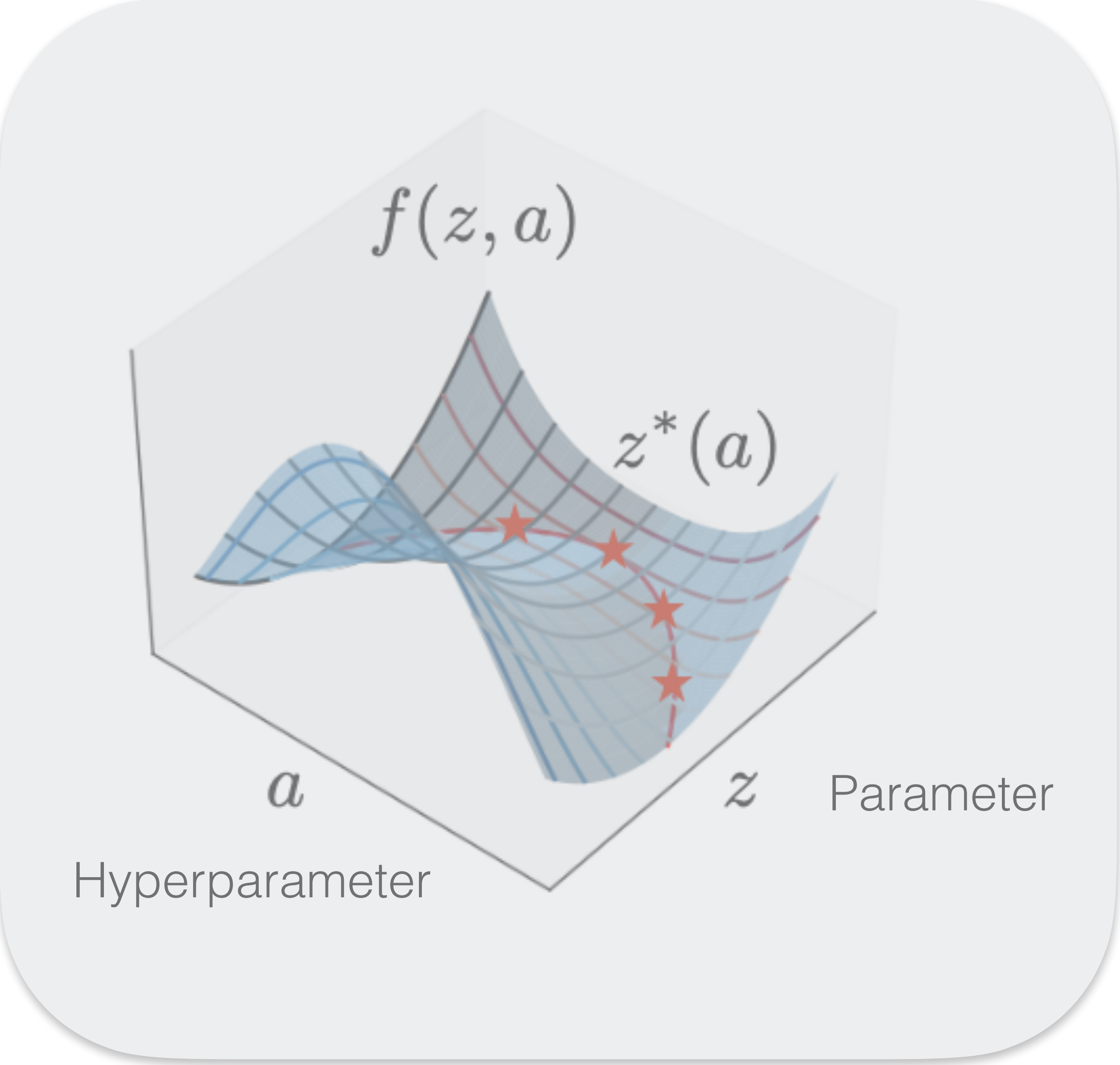
Implicit differentiation: the math



Implicit function theorem (informal):
Solutions $z^*(a)$ satisfying $f(z^*(a), a) = 0$ are locally analytic (therefore differentiable)

Implicit differentiation: the math





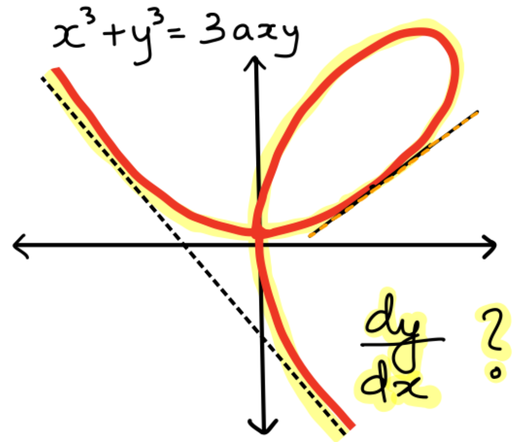
Some optimality condition

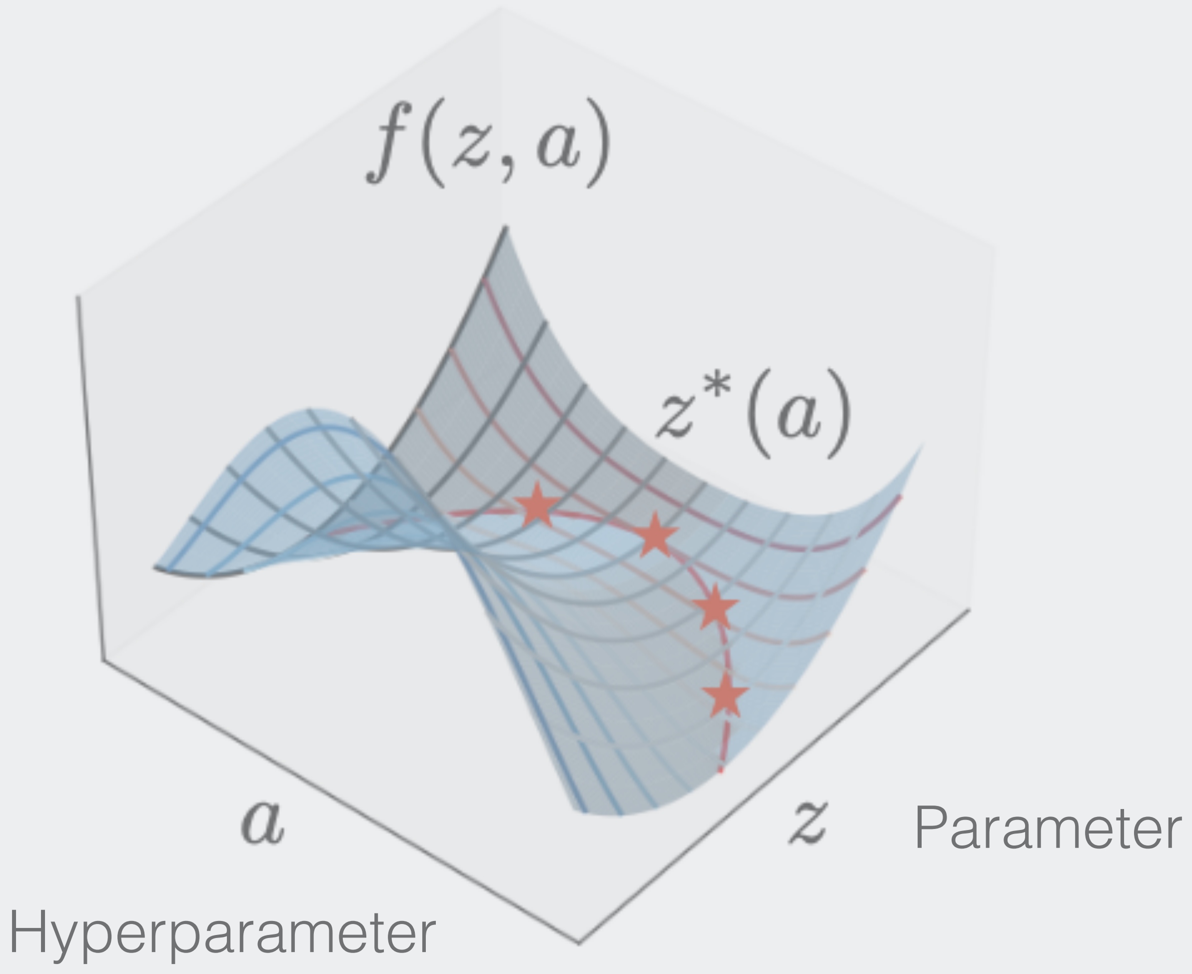
$$f(z,a) = 0$$

Implicitly defined solution function (cannot be written down analytically)

Implicit function theorem (informal):
Solutions z*(a) satisfying f(z*(a), a) = 0 are locally analytic (therefore differentiable)

Implicit differentiation: the math





Some optimality condition

$$f(z, a) = 0$$

$$a \longrightarrow \boxed{\text{Optimality Local gradients}} \longrightarrow z^*(a)$$

Implicitly defined solution function (cannot be written down analytically)

Implicit gradients

$$\partial_a z^*(a) = -[(\partial_z f)^{-1} \partial_a f]|_{(z^*(a), a)}$$

Local gradients at solution z*(a)

Also works for fixed-point equations:

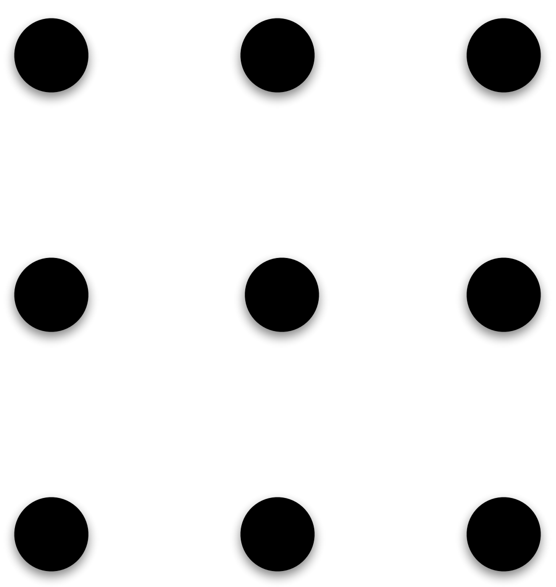
$$f'(z, a) = z$$
$$f'(z, a) - z = 0$$
$$f(z, a) = 0$$

Implicit function theorem (informal):
Solutions z*(a) satisfying f(z*(a), a) = 0 are locally analytic (therefore differentiable)

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

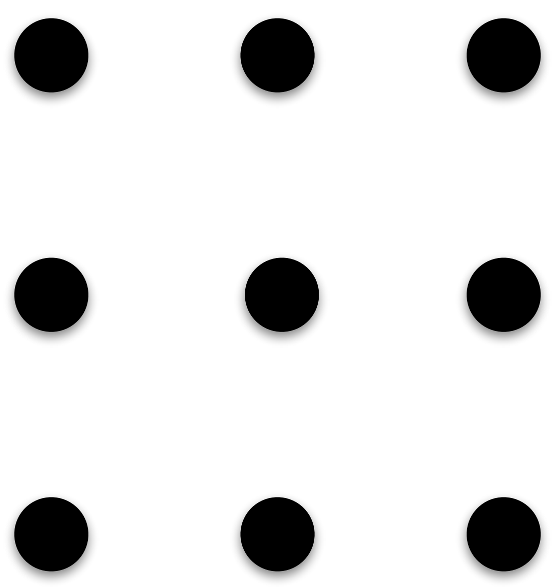
$$H(a) = -J\sum_i \sigma_i^z \sigma_{i+1}^z - \gamma \sum_i \sigma_i^x + \delta \sum_i \sigma_i^z - aA$$



Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

$$H(a) = -J\sum_i \sigma_i^z \sigma_{i+1}^z - \gamma \sum_i \sigma_i^x + \delta \sum_i \sigma_i^z - aA$$



$$H(a) \rightarrow \boxed{\begin{matrix}\text{VQA} \\ \psi_z\end{matrix}} \rightarrow \psi_{z^*(a)}$$

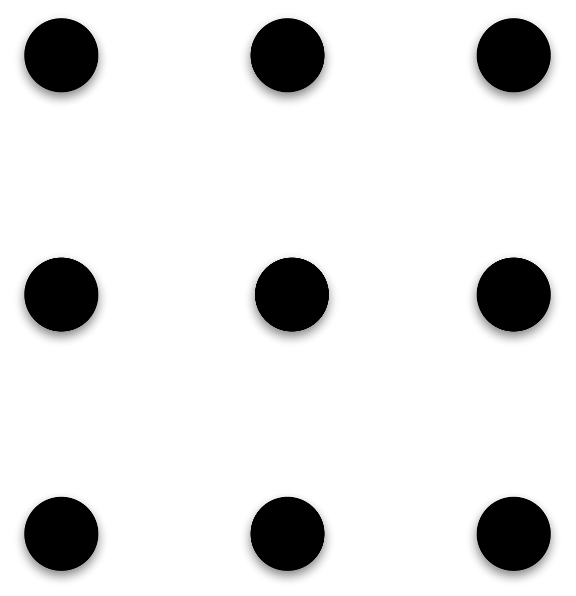
$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

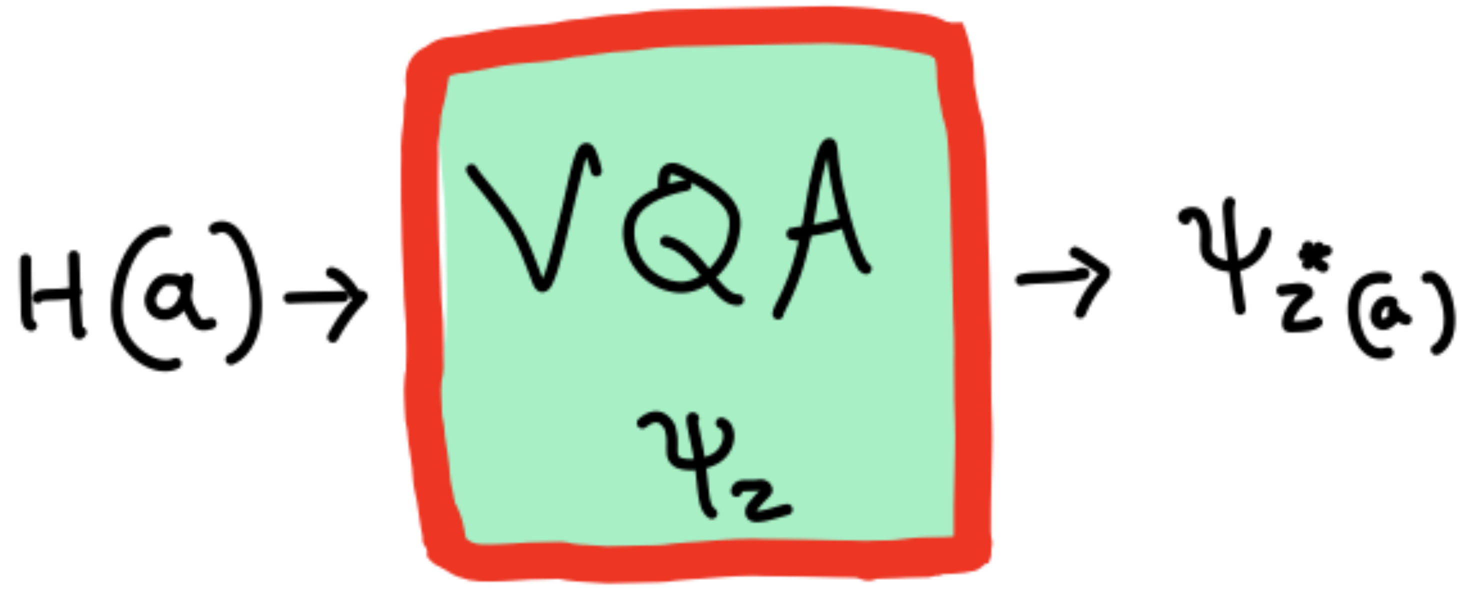
$$\frac{\partial z^*}{\partial a} \text{ ?}$$

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

$$H(a) = -J\sum_i \sigma_i^z\sigma_{i+1}^z - \gamma\sum_i \sigma_i^x + \delta\sum_i \sigma_i^z - aA$$





$$z^*(a) = \underset{z}{\mathrm{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} \, ?$$

- How do general ground-state properties change w.r.t the Hamiltonian params?

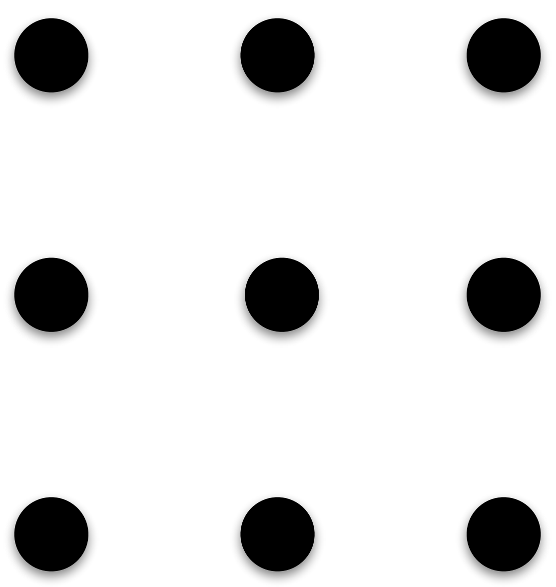
$$\langle A \rangle = \langle \psi_{z^*(a)} | A | \psi_{z^*(a)} \rangle$$

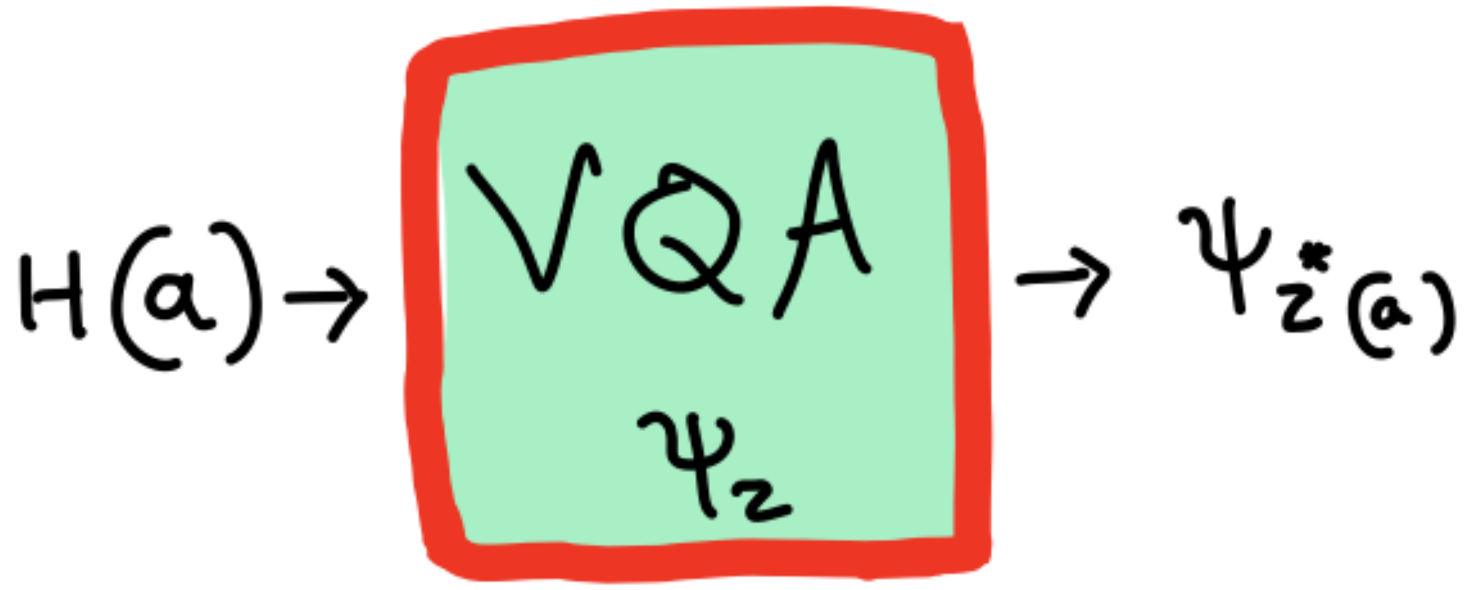
- What is $\partial_a \langle A \rangle$?

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

$$H(a) = -J\sum_i \sigma_i^z\sigma_{i+1}^z - \gamma\sum_i \sigma_i^x + \delta\sum_i \sigma_i^z - aA$$





$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} \; ?$$

- How do general ground-state properties change w.r.t the Hamiltonian params?

$$\langle A \rangle = \langle \psi_{z^*(a)} | A | \psi_{z^*(a)} \rangle$$

- What is $\partial_a \langle A \rangle$?

*If $A = H(a)$, Hellman-Feynmann theorem works: $\partial_a \langle A \rangle = \cancel{\partial_a z^*(a) \cdot \partial_z \langle A \rangle}$

$$+ \left\langle \psi_{z^*(a)} | \partial_a A | \psi_{z^*(a)} \right\rangle$$

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

$$\partial_a z^*(a) = -[(\partial_z f)^{-1} \partial_a f]|_{(z^*(a),a)}$$

$$H(a) \rightarrow \boxed{\begin{matrix}\text{VQA} \\ \psi_z\end{matrix}} \rightarrow \psi_{z^*(a)}$$

$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} \; ?$$

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

$$\partial_a z^*(a) = -[(\partial_z f)^{-1} \partial_a f]|_{(z^*(a),a)}$$

$$H(a) \to \boxed{\begin{matrix} \text{VQA} \\ \psi_z \end{matrix}} \to \psi_{z^*(a)}$$

$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} \ ?$$

What do we need?

- f(z, a) = 0:

$$f(z, a) = \partial_z \langle \psi_z | H(a) | \psi_z \rangle = 0$$

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

$$\partial_a z^*(a) = -[(\partial_z f)^{-1} \partial_a f]|_{(z^*(a),a)}$$

$$H(a) \rightarrow \boxed{\begin{matrix} \text{VQA} \\ \psi_z \end{matrix}} \rightarrow \psi_{z^*(a)}$$

$$z^*(a) = \underset{z}{\operatorname{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} \; ?$$

What do we need?

- f(z, a) = 0:

$$f(z,a) = \partial_z \langle \psi_z | H(a) | \psi_z \rangle = 0$$

- $z^*(a)$, local gradients

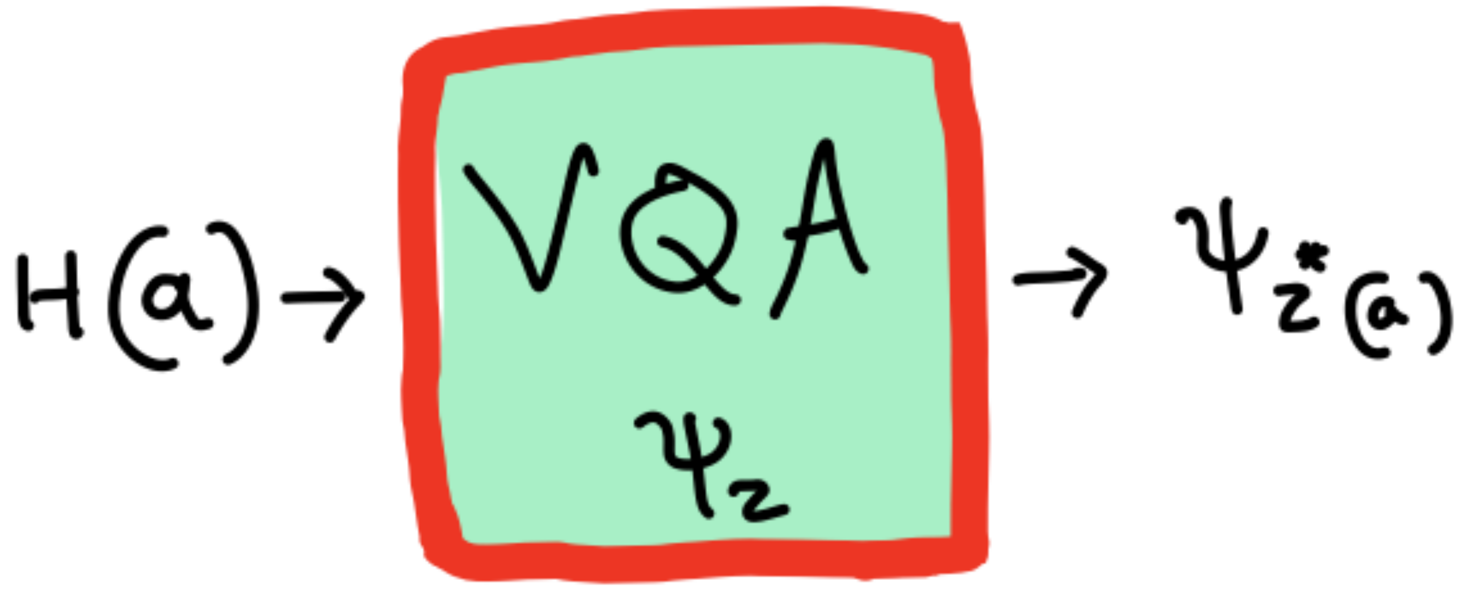
$$(\partial_z f, \partial_a f)\,|_{(z^*(a),a)}$$

XANADU VECTOR INSTITUTE CHALMERS UNIVERSITY OF TECHNOLOGY WACQT Wallenberg Center for Quantum Technology

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

$$\partial_a z^*(a) = -[(\partial_z f)^{-1}\partial_a f]|_{(z^*(a),a)}$$



$$z^*(a) = \underset{z}{\mathrm{argmin}} \langle \psi_z | H(a) | \psi_z \rangle$$

$$\frac{\partial z^*}{\partial a} ?$$

What do we need?

- f(z, a) = 0:

$$f(z,a) = \partial_z \langle \psi_z | H(a) | \psi_z \rangle = 0$$

- $z^*(a)$, local gradients

$$(\partial_z f, \partial_a f)\,|_{(z^*(a),a)}$$

- Inverting $\partial_z f$ efficiently (arXiv:2211.13765)

Implicit differentiation of VQAs:

ground-state gradients, generalized susceptibilities

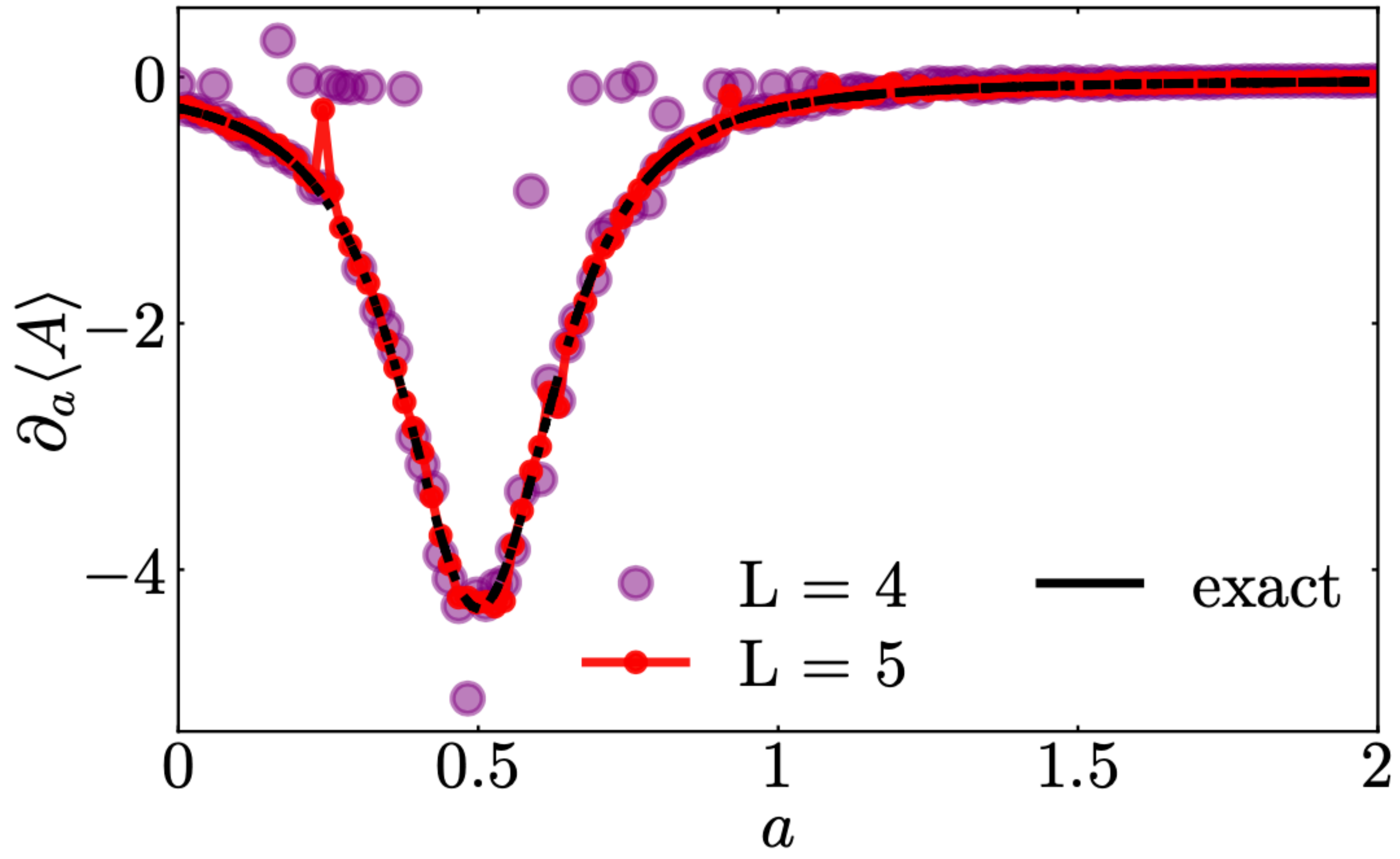


Figure 3: Susceptibility computation for a $N = 5$ spin

arXiv:2211.13765

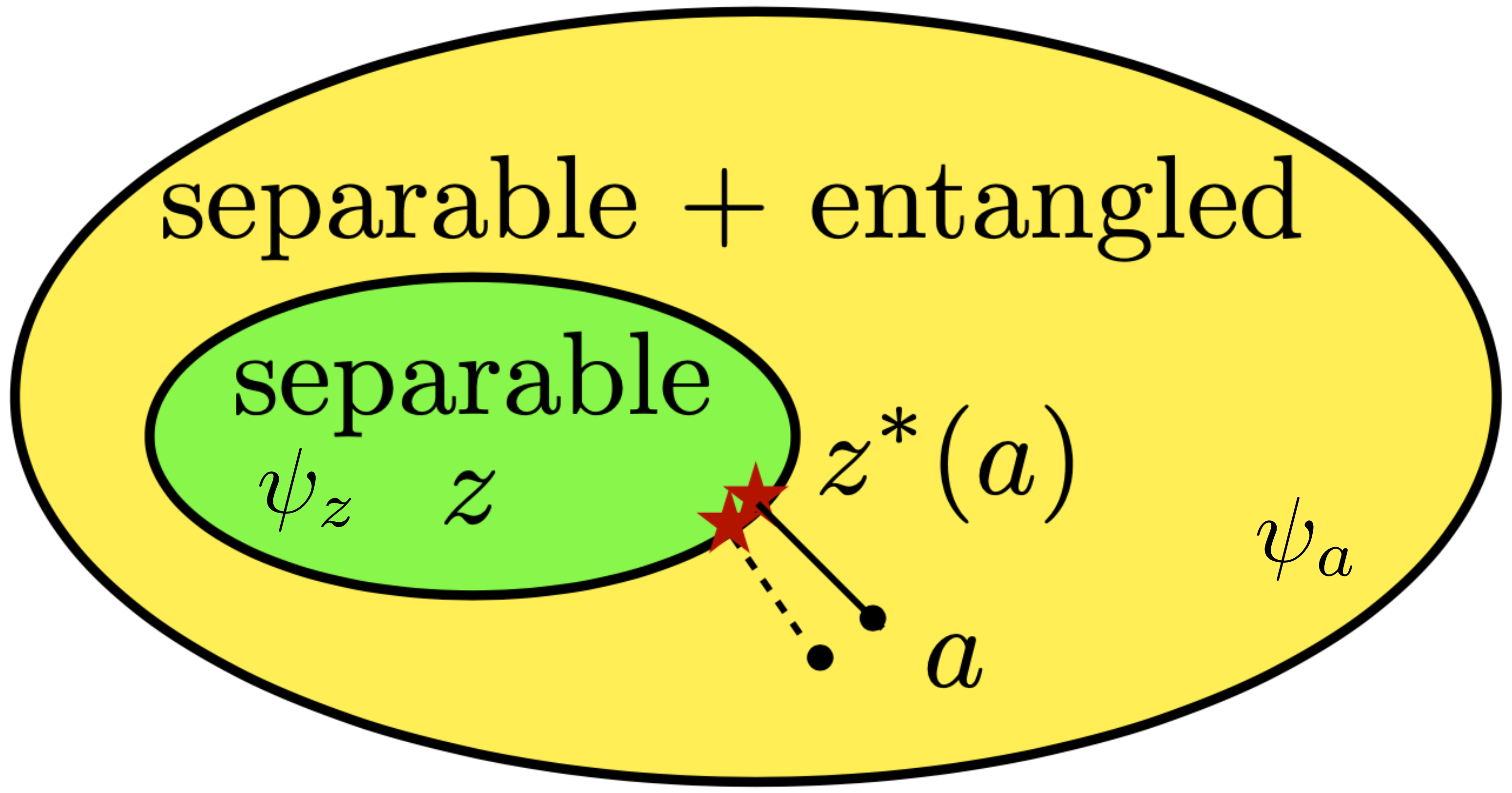
$$\partial_a z^*(a) = -[(\partial_z f)^{-1} \partial_a f]|_{(z^*(a),a)}$$

Quantum computing fidelity susceptibility using automatic differentiation

Olivia Di Matteo and R. M. Woloshyn
Phys. Rev. A **106**, 052429 – Published 28 November 2022

Implicit differentiation of VQAs:

creating entangled states by optimising a geometric measure of entanglement



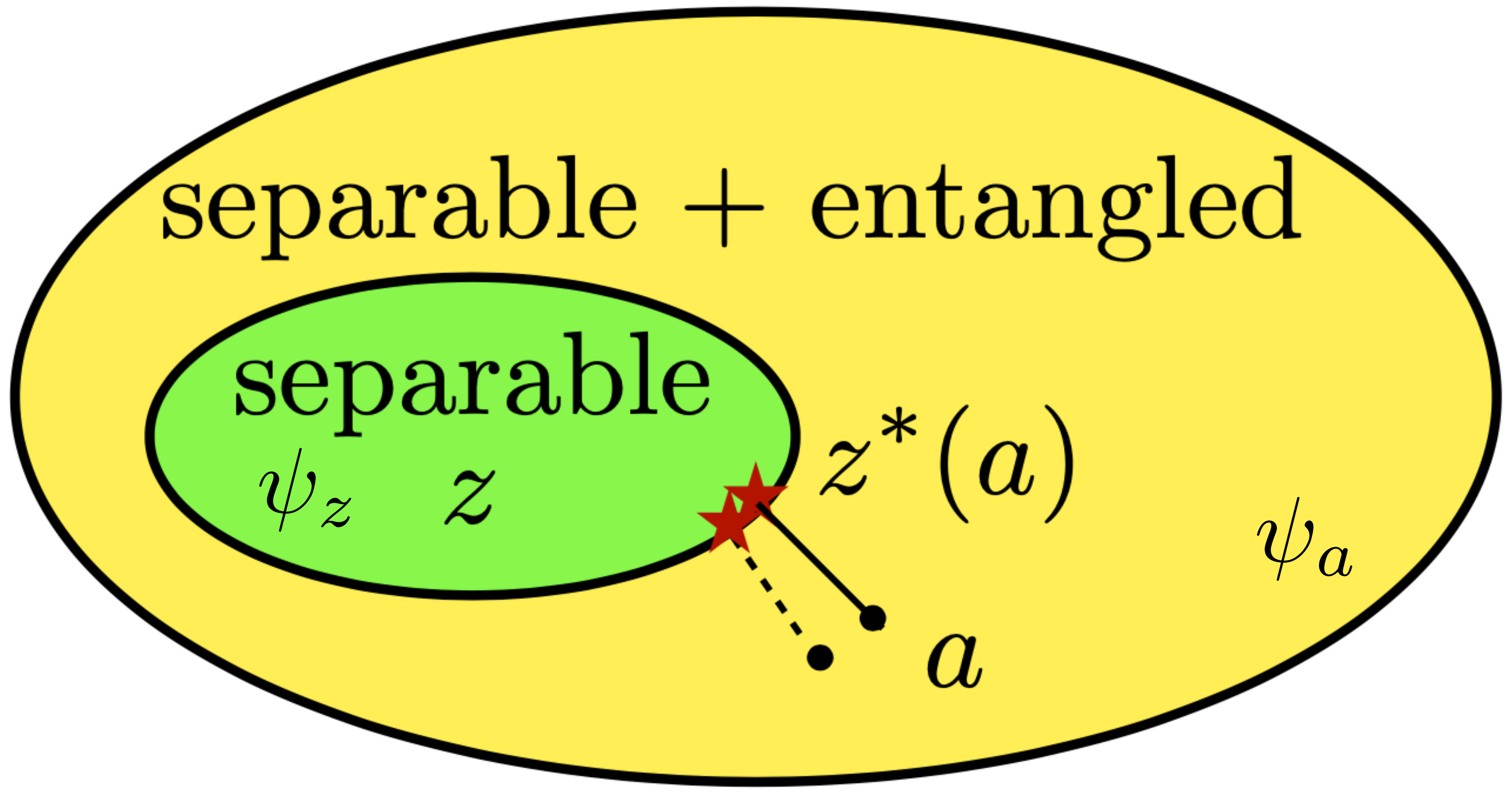
$$\partial_a z^*(a)$$

What is the separable state farthest away from our state?

arXiv:2211.13765

Implicit differentiation of VQAs:

creating entangled states by optimising a geometric measure of entanglement



$$\partial_a z^*(a)$$

What is the separable state farthest away from our state?

$$z^*(a) = \arg \underbrace{\max_z ||\langle \psi_z | \psi_a \rangle||}_{\Lambda_{\max}(a)}$$

Entanglement

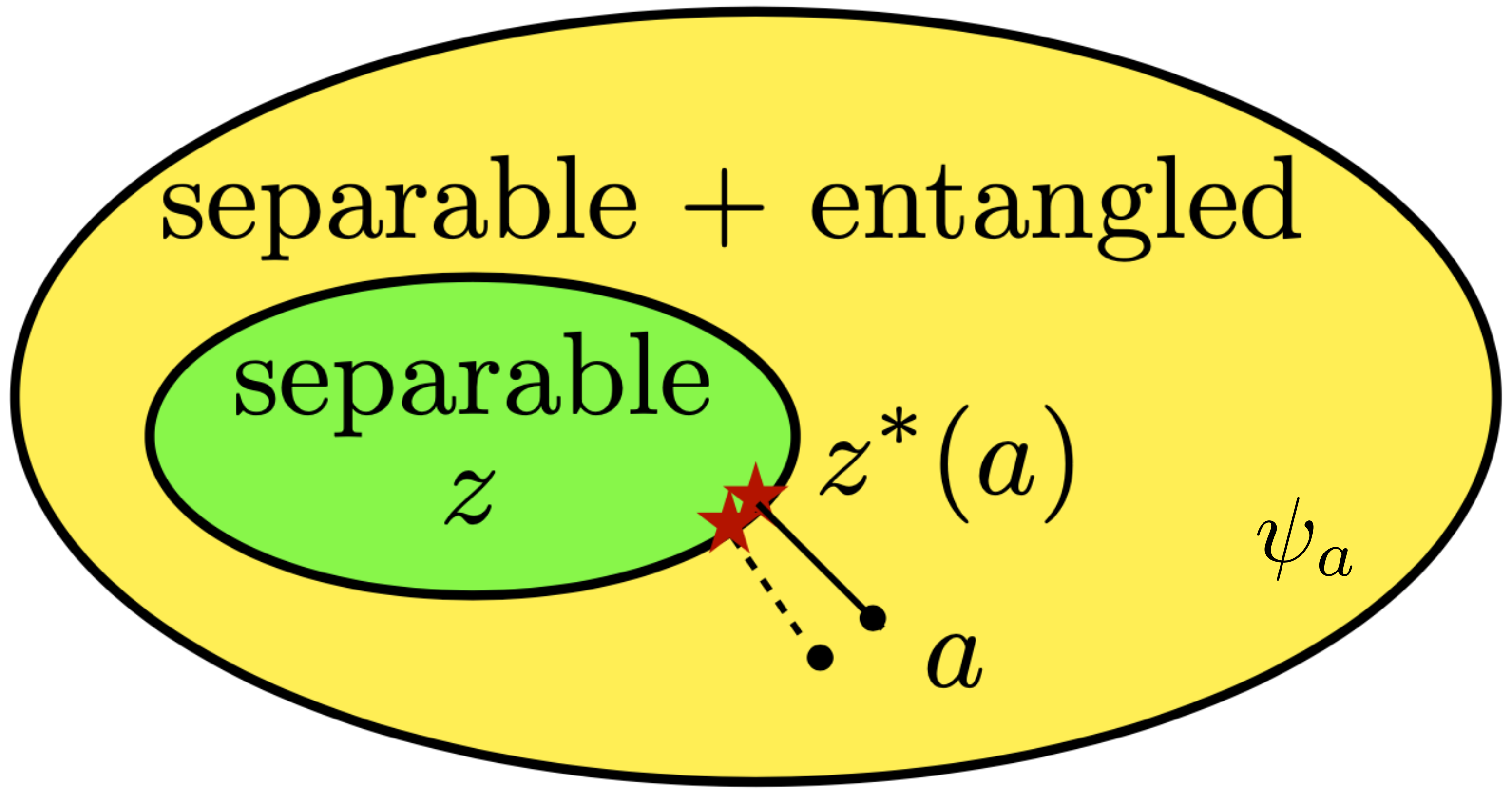
$$E(\psi_a) = 1 - \Lambda^2_{\max}(a)$$

arXiv:2211.13765

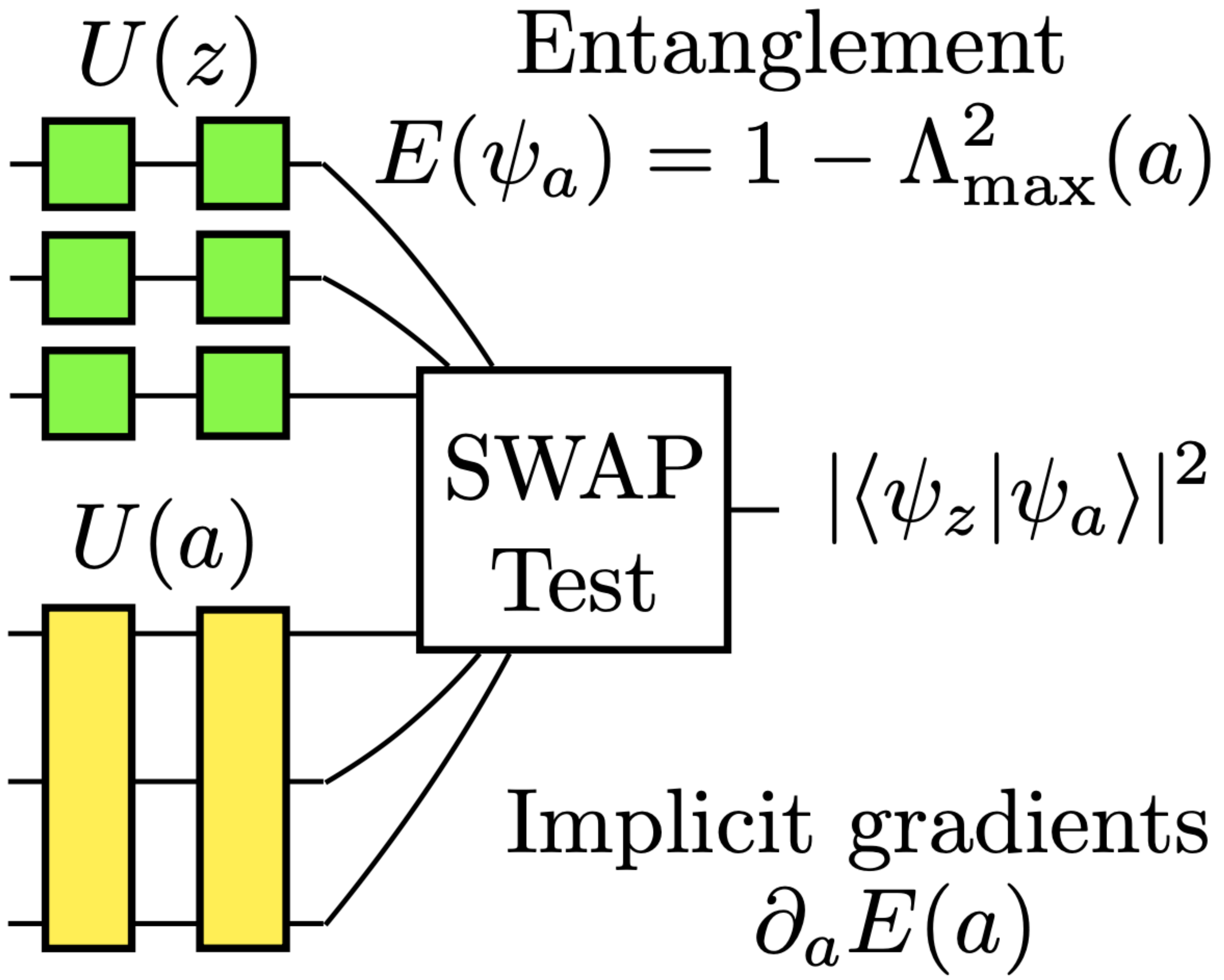
XANADU VECTOR INSTITUTE CHALMERS UNIVERSITY OF TECHNOLOGY WACQT Wallenberg Center for Quantum Technology

Implicit differentiation of VQAs:

creating entangled states by optimising a geometric measure of entanglement



$\partial_a z^*(a)$



$$z^*(a) = \arg\max_z \underbrace{||\langle\psi_z|\psi_a\rangle||}_{\Lambda_{\max}(a)}$$

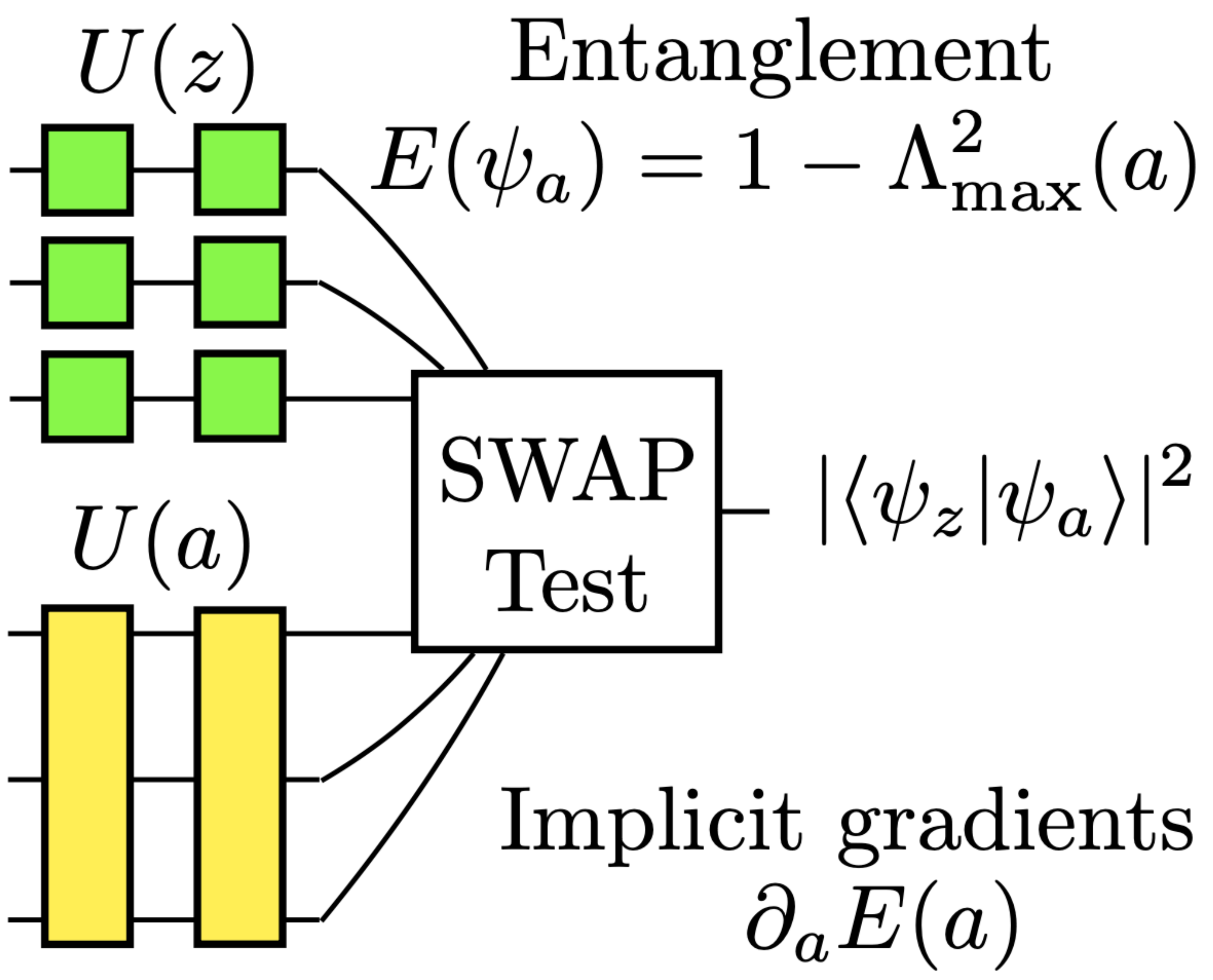
How does the entanglement measure change as we change a ?

$\partial_a E(\psi_a)$

arXiv:2211.13765

Implicit differentiation of variational quantum algorithms:

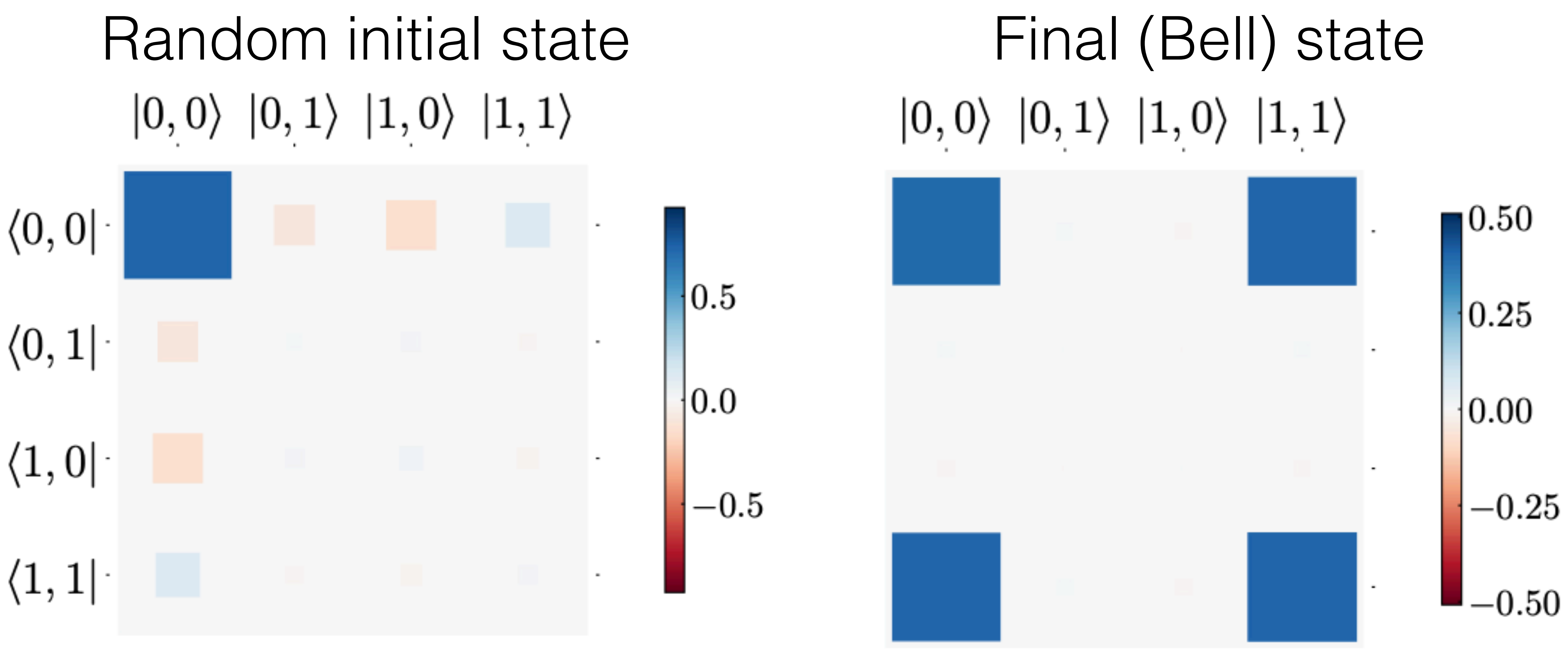
creating entangled states by optimizing a geometric measure of entanglement



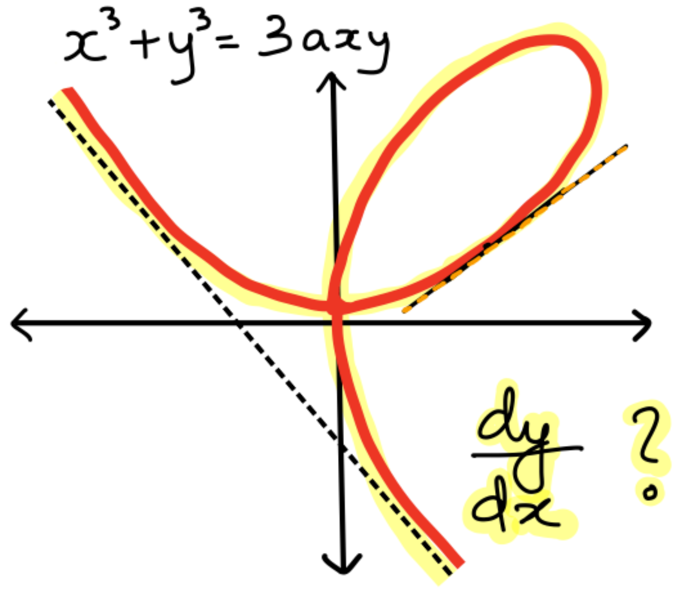
$$z^*(a) = \underbrace{\arg \max_z ||\langle \psi_z | \psi_a \rangle||}_{\Lambda_{\max}(a)}$$

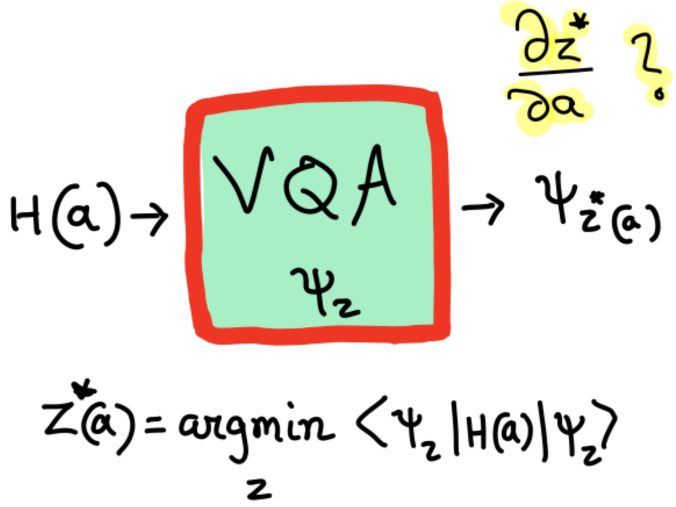
arXiv:2211.13765

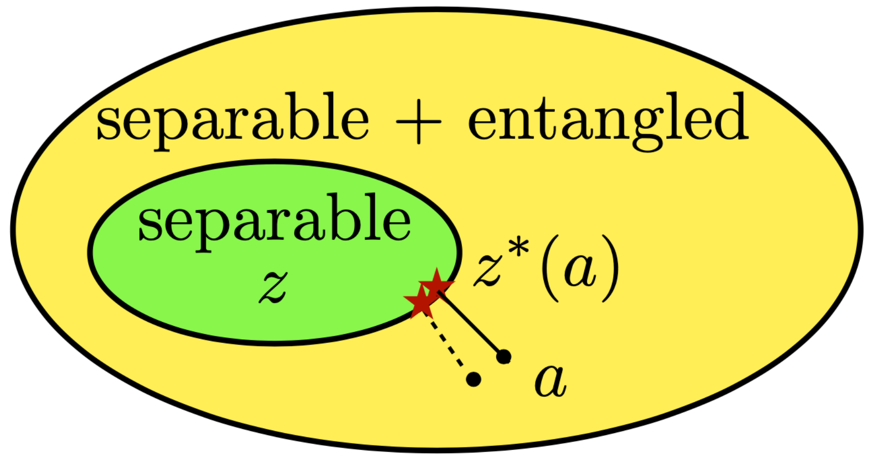
With the implicit gradients, we can simply apply gradient-based optiimzation to create entanglement variationally!



Conclusions







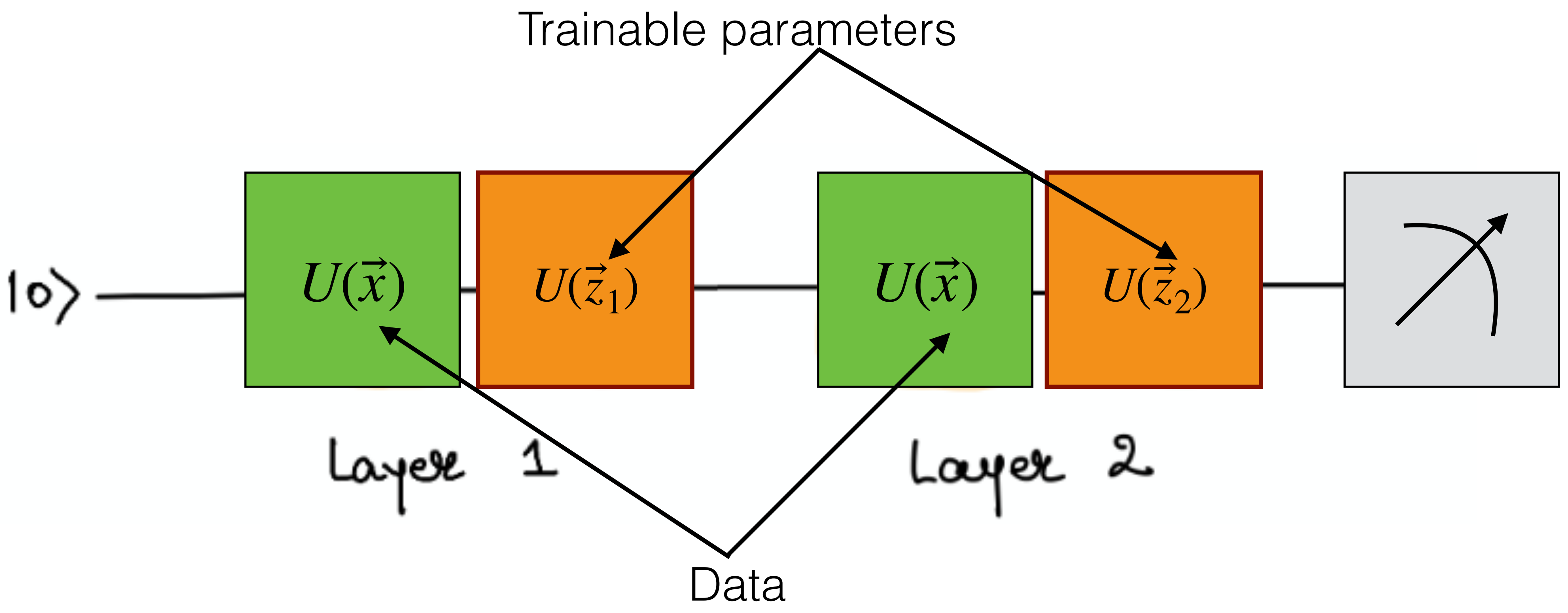
- Implicit differentiation unlocks new possibilities for VQAs:
  - Ground-state gradients
  - Entanglement generation: a novel application
  - Hyperparameter optimization in quantum machine learning algorithms

- Caveats:
  - solution points and their gradients are necessary
  - gradients could be noisy

Extra slides

Implicit differentiation of variational quantum algorithms:

Hyperparameter optmization in QML

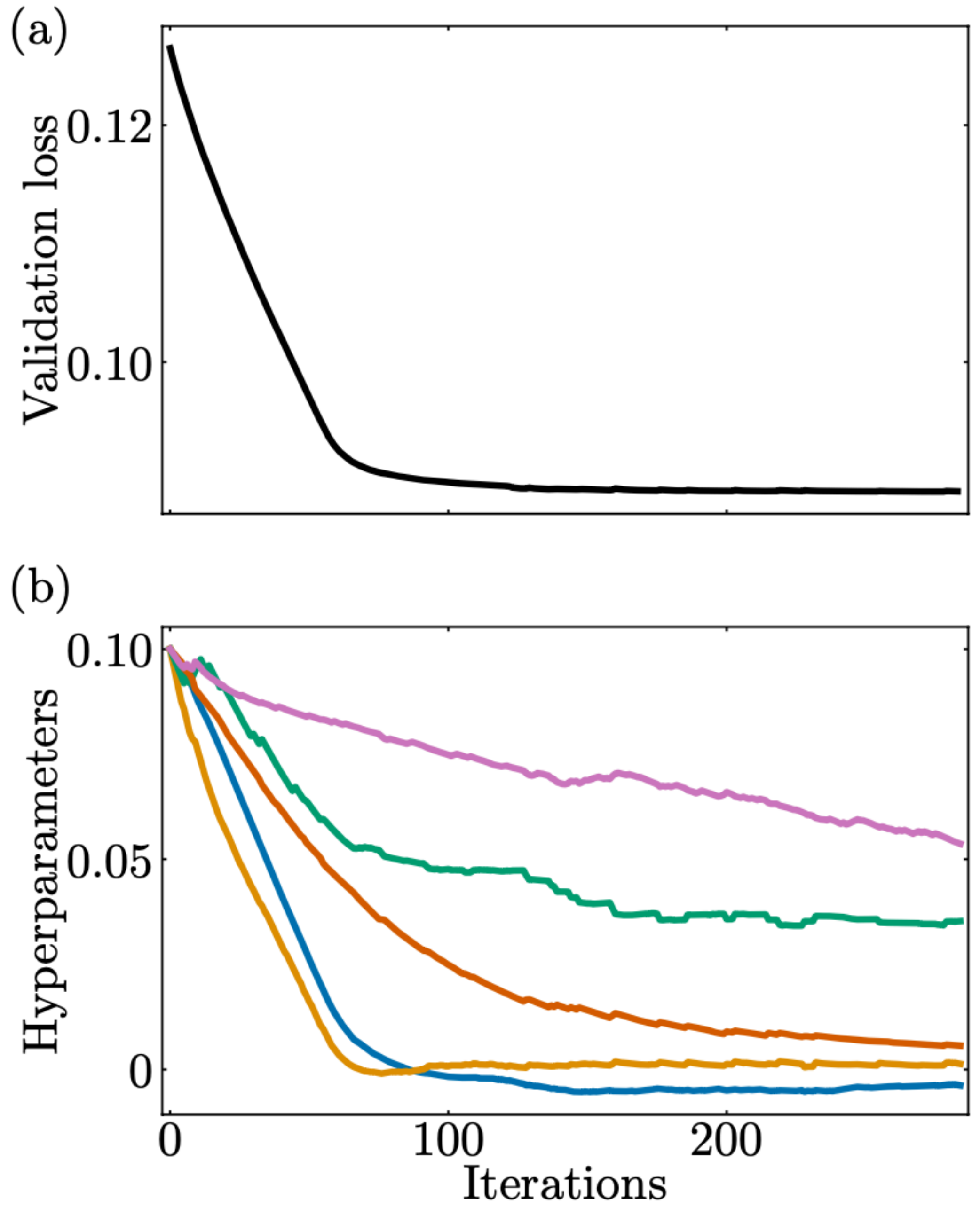


$$\mathcal{L}_T(z,a) = \mathcal{L}_{\text{bincross}}(z) + \sum_l a_l ||z_l||_2$$

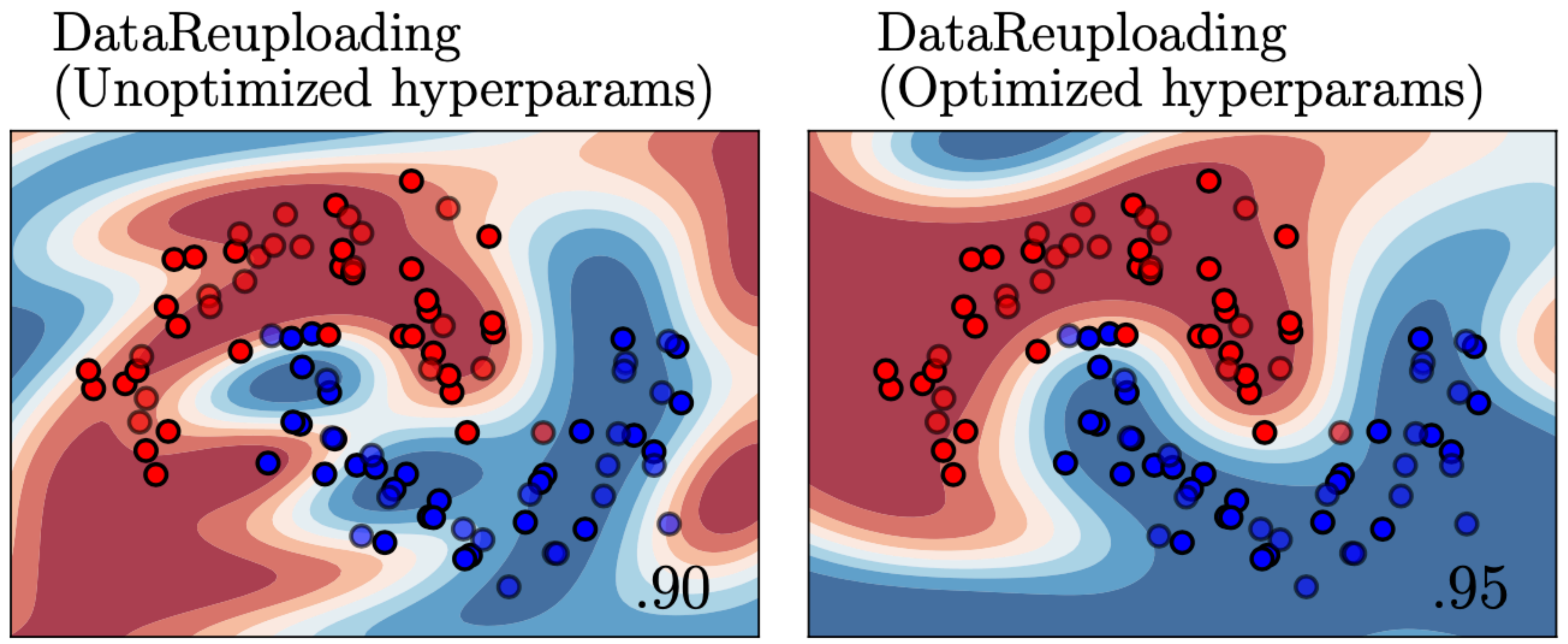
Train parameters with gradient-descent with some hyperparameter

$$L_V(a) = L_{\text{bincross}}(z^*(a))$$

Hyperparameters



Train hyperparameters also with gradient-descent!
optimize validation loss



[38] Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I. Latorre. Data re-uploading for a universal quantum classifier. Quantum, 4:226, 2020.