University of Stuttgart
Germany

SimTech

Quantum Techniques in Machine Learning, CERN, November 2023

Training robust quantum classifiers based on Lipschitz bounds





Julian Berberich[1] Daniel Fink[2] Daniel Pranjić[3] Christian Tutschku[3] Christian Holm[2]

[1] Institute for Systems Theory and Automatic Control, University of Stuttgart

[2] Institute for Computational Physics, University of Stuttgart

[3] Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering, Stuttgart

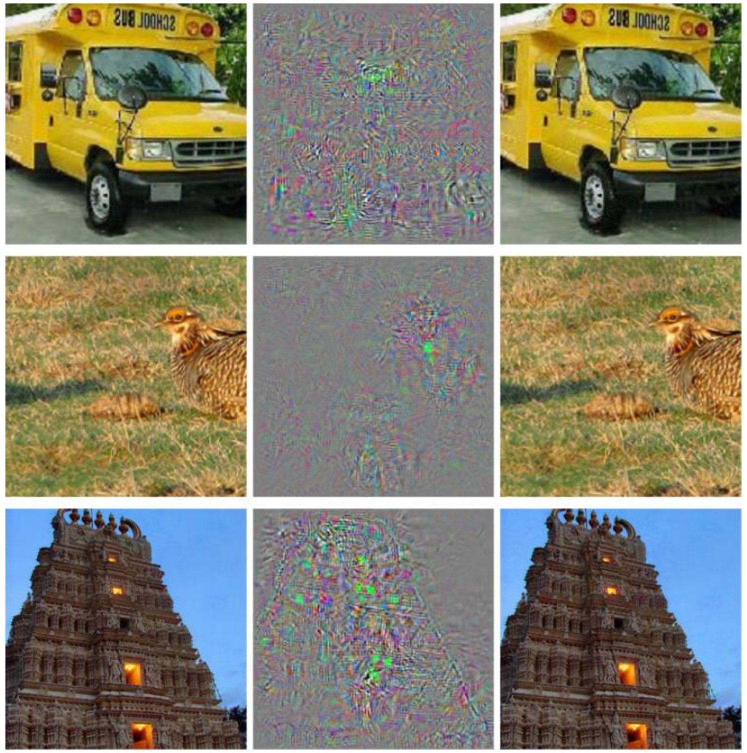
Adversarial attacks

correct

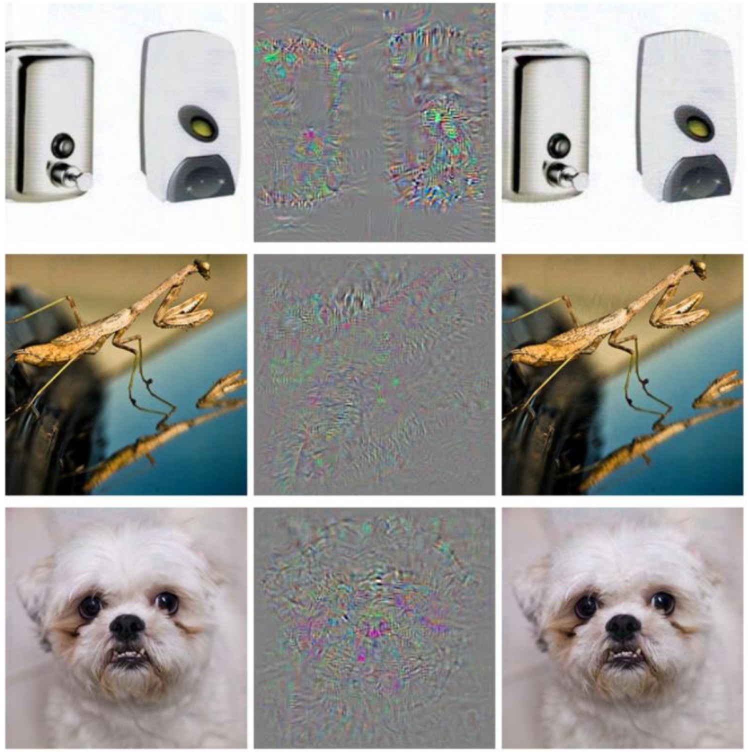
correct

"Intriguing properties of neural networks", C. Szegedy et al., arXiv:1312.6199, 2013
"EAD: Elastic-net attacks to deep neural networks via adversarial examples", P.-Y. Chen et al., arXiv:1709.04114, 2017

Adversarial attacks



correct +distort ostrich



correct +distort ostrich

“Intriguing properties of neural networks“, C. Szegedy et al., arXiv:1312.6199, 2013
“EAD: Elastic-net attacks to deep neural networks via adversarial examples“, P.-Y. Chen et al., arXiv:1709.04114, 2017

Adversarial attacks

correct +distort ostrich correct +distort ostrich

“Intriguing properties of neural networks“, C. Szegedy et al., arXiv:1312.6199, 2013
“EAD: Elastic-net attacks to deep neural networks via adversarial examples“, P.-Y. Chen et al., arXiv:1709.04114, 2017

Adversarial attacks

correct +distort ostrich correct +distort ostrich

Adversarial attacks can cause misclassification!

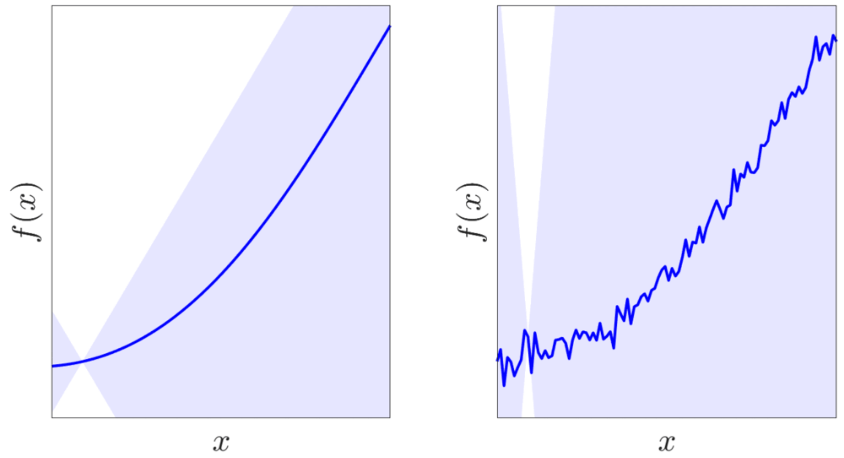
“Intriguing properties of neural networks“, C. Szegedy et al., arXiv:1312.6199, 2013
“EAD: Elastic-net attacks to deep neural networks via adversarial examples“, P.-Y. Chen et al., arXiv:1709.04114, 2017

Lipschitz bounds

$L > 0$ is a **Lipschitz bound** of a model f if

$$\|f(x) - f(y)\| \leq L\|x - y\|$$

for all x, y.



L is a simple **measure of robustness:**

- describes **sensitivity** to data perturbations:
  $\|f(x + \varepsilon) - f(x)\| \leq L\|\varepsilon\|$
- quantifies **robustness against adversarial attacks**

small Lipschitz bound
= **high robustness**

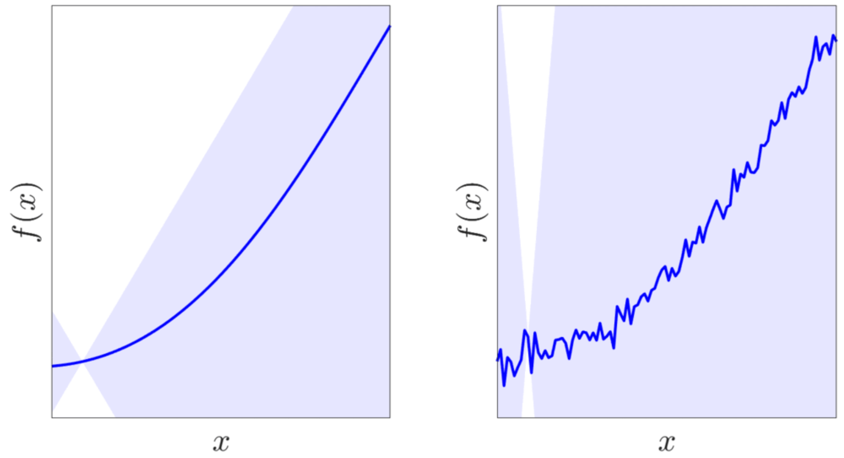
"Intriguing properties of neural networks", C. Szegedy et al., arXiv:1312.6199, 2013
"Evaluating the robustness of neural networks: an extreme value theory approach", T.-W. Weng et al., ICLR, 2018

Lipschitz bounds

$L > 0$ is a **Lipschitz bound** of a model f if

$$\|f(x) - f(y)\| \leq L\|x - y\|$$

for all x, y.



L is also **connected to generalization:**

- small variability ➔ **less overfitting**
- Lipschitz-based **generalization bounds**

small Lipschitz bound
= **good generalization**

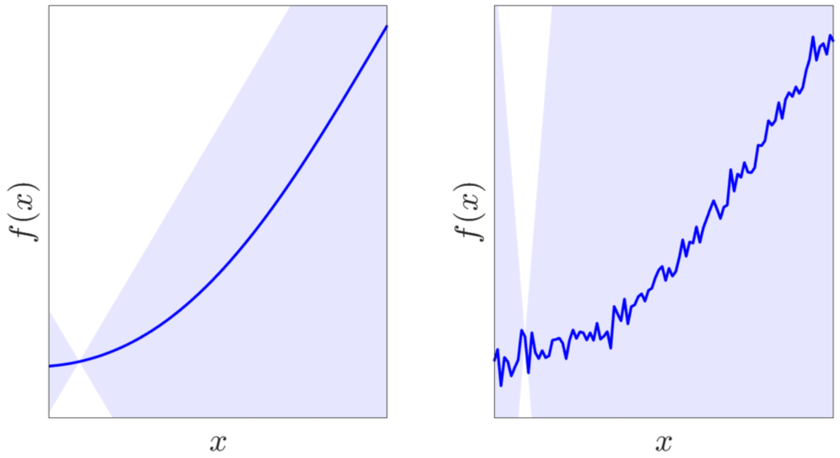
"Distance-based classification with Lipschitz functions", U. von Luxburg and O. Bousquet, JMLR, 2004
"Spectrally-normalized margin bounds for neural networks", P. Bartlett et al., NeurIPS, 2017
"Exploring generalization in deep learning", B. Neyshabur et al., NeurIPS, 2017
"Robustness and generalization", H. Xu and S. Mannor, Mach Learn, 2012

Lipschitz bounds

$L > 0$ is a **Lipschitz bound** of a model f if

$$\|f(x) - f(y)\| \leq L\|x - y\|$$

for all x, y.



Lipschitz bound regularization **improves robustness and generalization!**

"A simple weight decay can improve generalization", A. Krogh and J. Hertz, NeurIPS, 1991
"Regularisation of neural networks by enforcing Lipschitz continuity", H. Gouk et al., ML, 2021
"Training robust neural networks using Lipschitz bounds", P. Pauli, A. Koch, JB et al., IEEE LCSS, 2022

Contribution

Our Contribution

Use **Lipschitz bounds** to

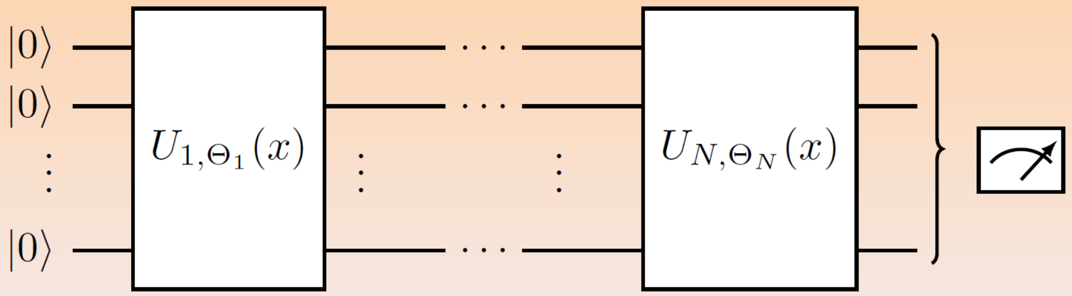
- **study robustness and generalization** of quantum models,
- **train robust and generalizable quantum models** via regularization,
- demonstrate benefits of **trainable encodings**.

Quantum models and their Lipschitz bounds

Quantum model

Variational quantum circuit:

$$\Theta_j = \{w_j, \theta_j\}$$



$$U_{j,\Theta_j}(x) = e^{-i(w_j^\top x + \theta_j) H_j}$$

- x: data
- w_j, θ_j: trainable parameters
- H_j: Hermitian generators

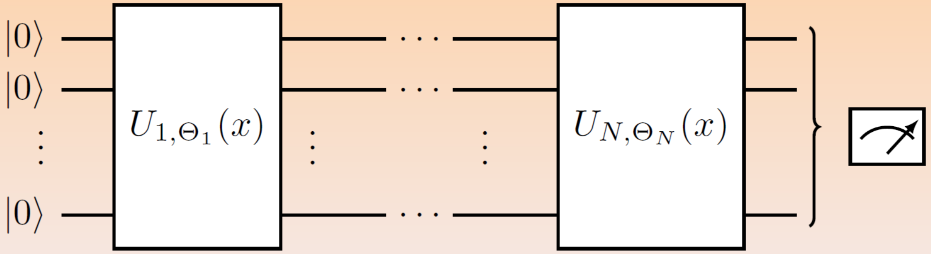
Output: $f_\Theta(x) = \langle 0 | U_\Theta(x)^\dagger M U_\Theta(x) | 0 \rangle$

→ Quantum model with **data re-uploading** and **trainable encoding**

"Data re-uploading for a universal quantum classifier", A. Pérez-Salinas et al., Quantum, 2020

Lipschitz bounds of quantum models

$$\Theta_j = \{w_j, \theta_j\}$$



$$U_{j,\Theta_j}(x) = e^{-i(w_j^\top x + \theta_j)H_j}$$

The quantum model $f_\Theta(x)$ admits the Lipschitz bound

$$L_\Theta = 2\|M\| \sum_{j=1}^{N} \|w_j\| \|H_j\|$$

- Can be easily computed
- Depends on the **observable** M and on the **encoding** w_j, H_j
- Does **NOT** depend on the parameters θ_j

Robustness

Robustness of quantum models

- **Robustness against hardware errors** is critical, especially in the NISQ era
  - can be studied based on Lipschitz bounds
  - **NOT the focus of our work!**

“Quantum Computing in the NISQ era and beyond”, J. Preskill, Quantum, 2018
“Robustness of quantum algorithms against coherent control errors”, **JB** et al., arXiv:2303.00618, 2023

Robustness of quantum models

- **Robustness against hardware errors** is critical, especially in the NISQ era
  - can be studied based on Lipschitz bounds
  - **NOT the focus of our work!**

We focus on **(adversarial) robustness against data perturbations**

- Quantum models are **also vulnerable to adversarial attacks**
- **Lipschitz bounds** quantify robustness: For an adversarial attack ε

$$\|f_\Theta(x+\varepsilon) - f_\Theta(x)\| \leq L_\Theta \|\varepsilon\|$$

"Quantum Computing in the NISQ era and beyond", J. Preskill, Quantum, 2018
"Robustness of quantum algorithms against coherent control errors", JB et al., arXiv:2303.00618, 2023
"Quantum adversarial machine learning", S. Lu et al., Phys. Rev. Res., 2020
"Towards quantum enhanced adversarial robustness in machine learning", M. T. West et al., Nature Machine Intelligence, 2023

Lipschitz bound regularization

Recall: Lipschitz bound $L_\Theta = 2\|M\| \sum_{j=1}^{N} \|w_j\| \|H_j\|$

Training with Lipschitz bound regularization

- **Setup:** Supervised learning with loss ℓ and data (x_k, y_k) of length n
- We solve the **training problem**

$$\min_{\Theta=\{\theta_j, w_j\}_j} \quad \frac{1}{n}\sum_{k=1}^{n} \ell(f_\Theta(x_k), y_k)$$

Lipschitz bound regularization

Recall: Lipschitz bound $L_\Theta = 2\|M\| \sum_{j=1}^{N} \|w_j\| \|H_j\|$

Training with Lipschitz bound regularization

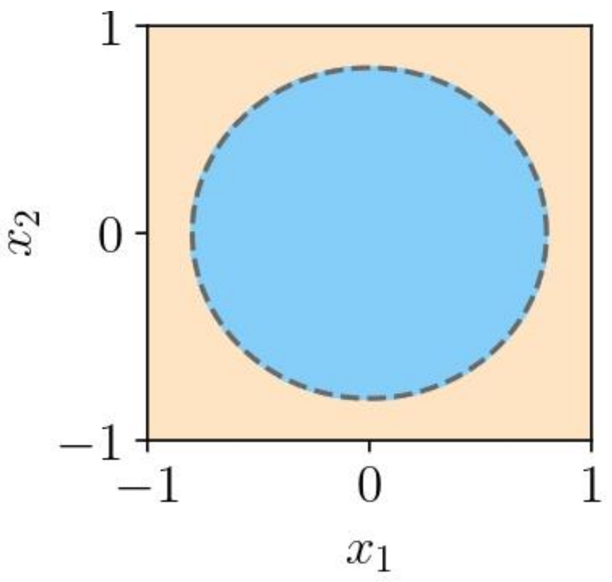
- **Setup:** Supervised learning with loss ℓ and data (x_k, y_k) of length n
- We solve the regularized **training problem**

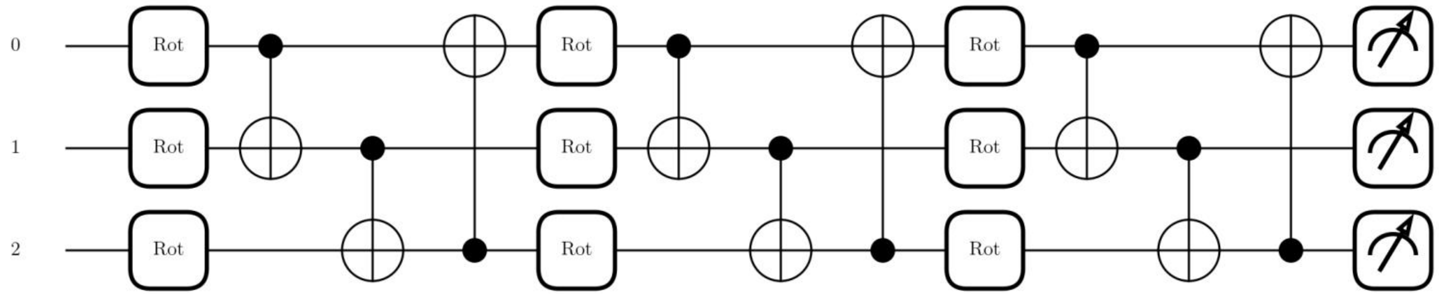
$$\min_{\Theta=\{\theta_j, w_j\}_j} \quad \frac{1}{n} \sum_{k=1}^{n} \ell(f_\Theta(x_k), y_k) + \lambda \sum_{j=1}^{N} \|w_j\|^2 \|H_j\|^2$$

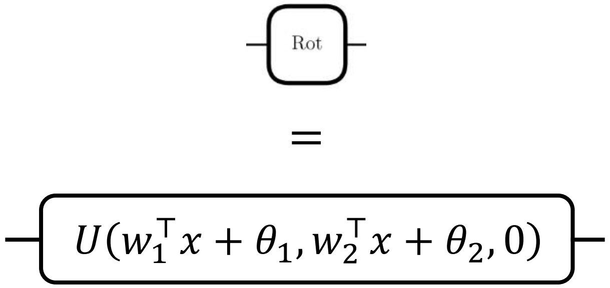
- Encourages model with **small Lipschitz bound** ⇒ improved robustness
- **Hyperparameter** λ: trading off training error and robustness

Numerical results: robustness

- Binary 2D Classification problem
- Training data: $n = 200$ samples drawn uniformly from $[-1,1]^2$
- Optimization with ADAM

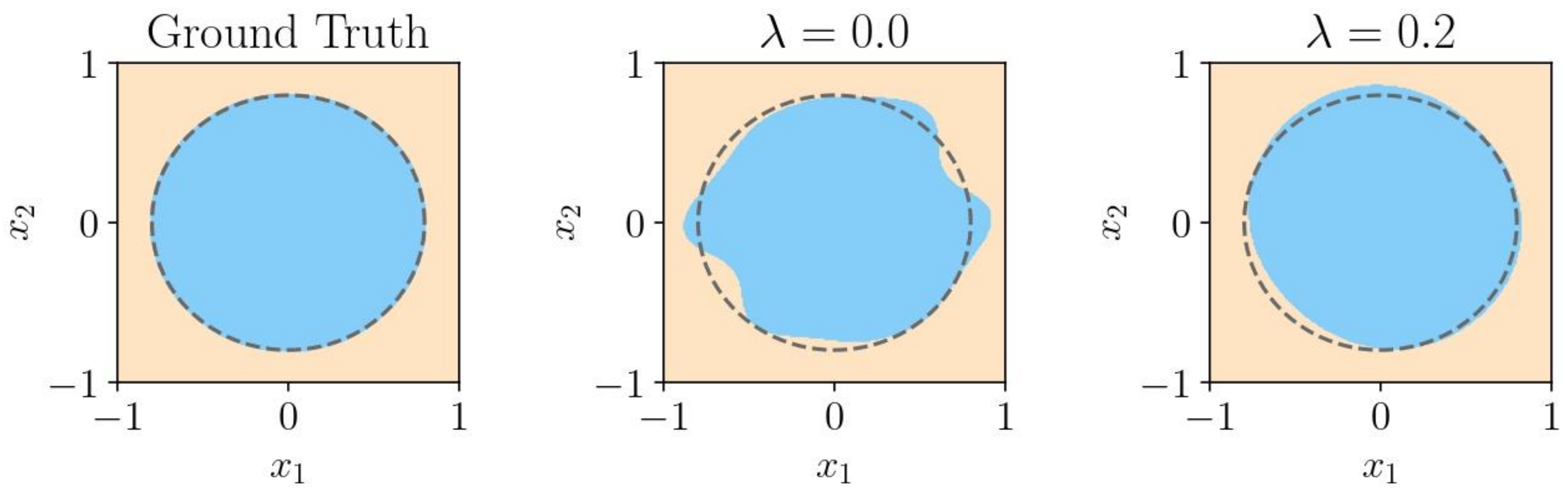






"Data re-uploading for a universal quantum classifier", A. Pérez-Salinas et al., Quantum, 2020
https://pennylane.ai/qml/demos/tutorial_data_reuploading_classifier/

Numerical results: robustness



Lipschitz bounds:	$L = 34.4$	$L = 4.0$

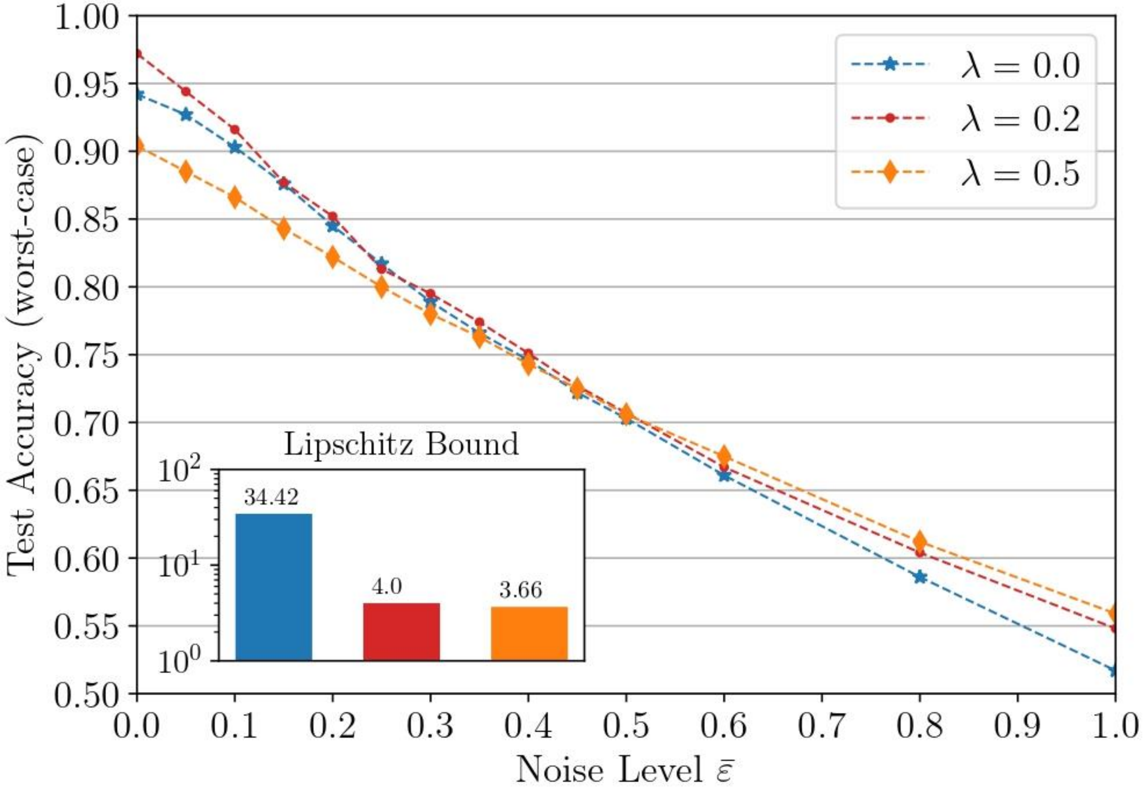
- Regularization ➔ smaller Lipschitz bound ➔ smoother decision boundary
- **Expectation:** better robustness against data perturbations

Numerical results: robustness

- **Test data:** **1,000** data points drawn uniformly from $[-1,1]^2$

- **Noise:** perturbs each test data point by ε drawn uniformly from $\varepsilon \in [-\bar{\varepsilon}, \bar{\varepsilon}]$

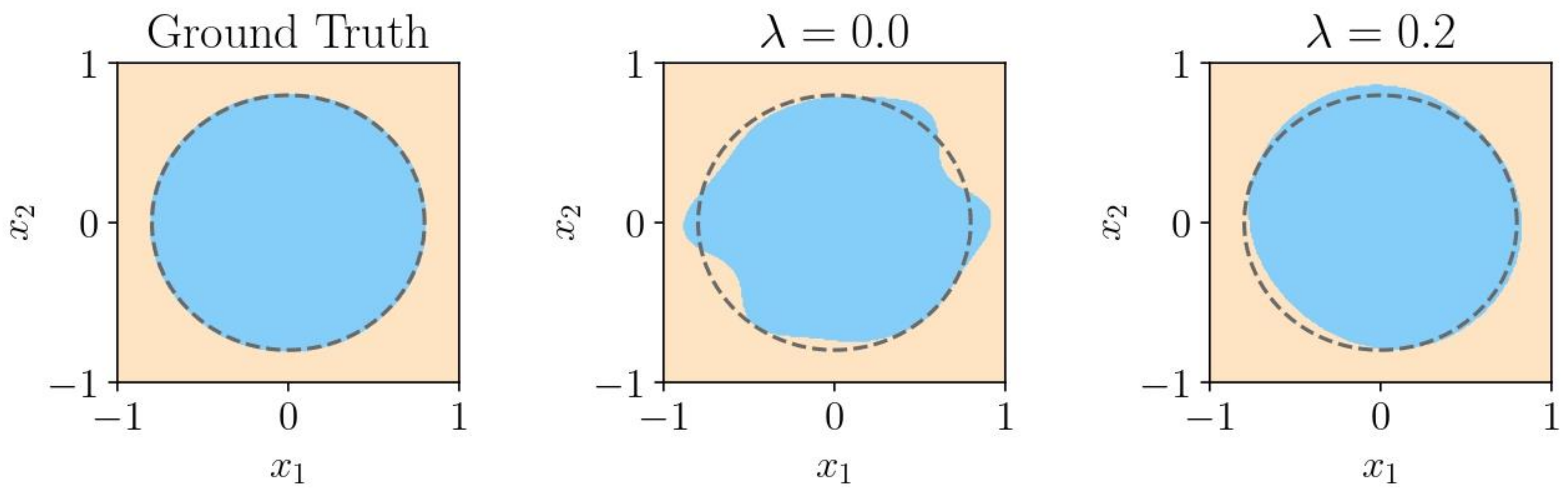
→ Worst case over **200** noise samples



Regularization improves robustness

Generalization

Numerical results: robustness



Lipschitz bounds:	$L = 34.4$	$L = 4.0$

- Regularization ➔ smaller Lipschitz bound ➔ smoother decision boundary

Regularization should also **improve generalization**!

Generalization bound

- Expected risk $R(f_\Theta) = \int_{X \times Y} \ell\big(y, f_\Theta(x)\big) \mathrm{d}P(x,y)$
- Empirical risk $R_n(f_\Theta) = \frac{1}{n}\sum_k \ell\big(y_k, f_\Theta(x_k)\big)$

Theorem (informal)

The generalization error of f_Θ is bounded as

$$|R(f_\Theta) - R_n(f_\Theta)| \leq C_1 \|M\| \sum_j \|w_j\| \|H_j\| + \frac{C_2}{\sqrt{n}}$$

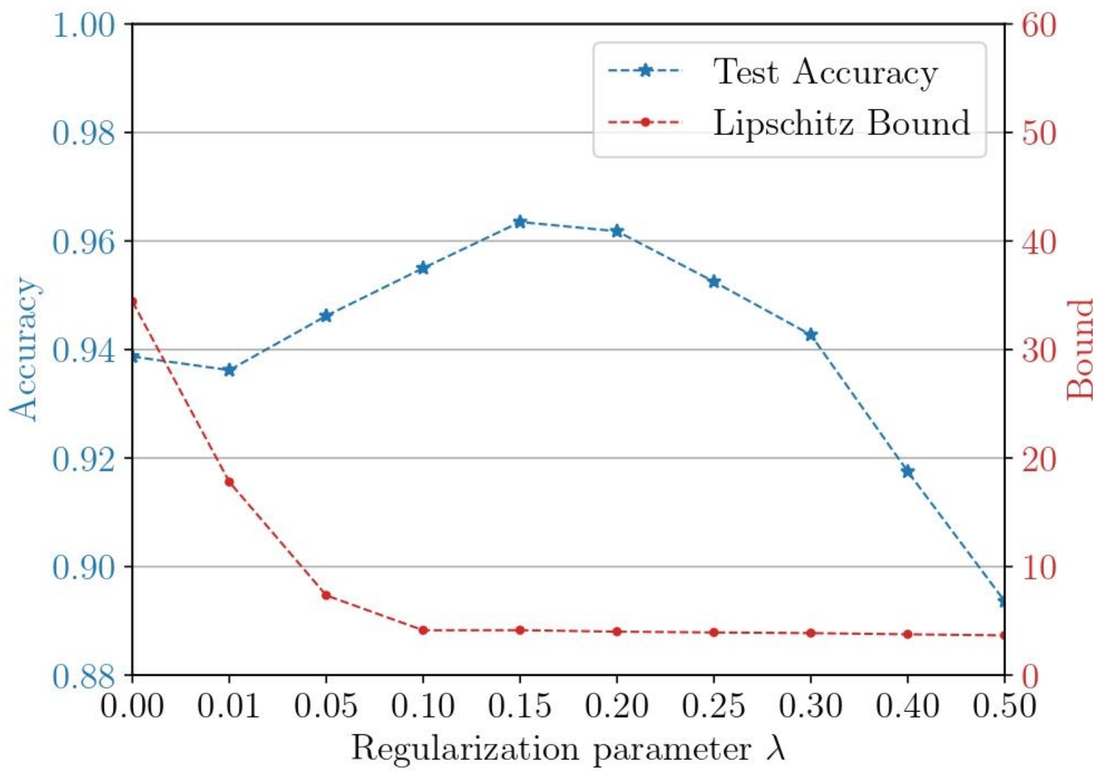
for some $C_1, C_2 > 0$.

- Proof uses **classical ML techniques**[1]
- Explicitly involves the **Lipschitz bound** (observable & data encoding)
- **Trade off:** Training error vs. Lipschitz bound

[1] "Robustness and generalization", H. Xu and S. Mannor, Mach Learn, 2012

Numerical results: generalization

- **Training as before** for different hyperparameters λ
- **Test data:** 10,000 data points drawn uniformly from $[-1,1]^2$

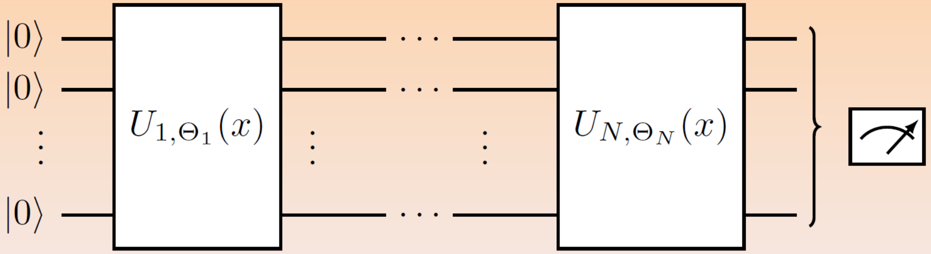


Regularization improves generalization

Benefits of trainable encodings

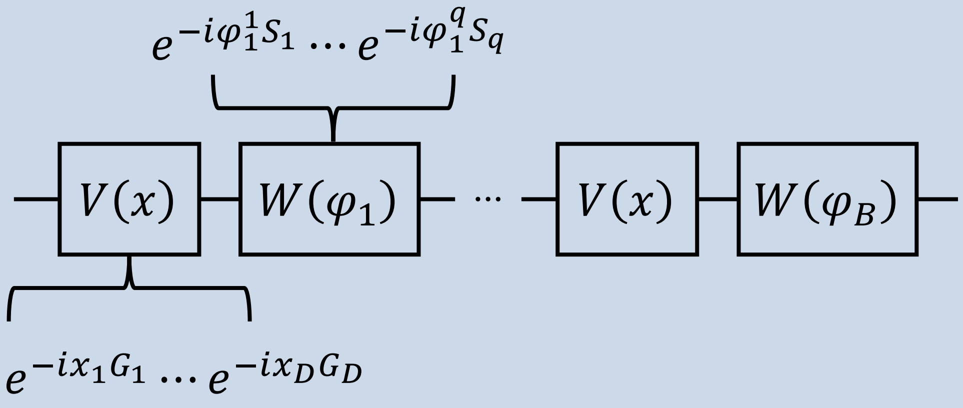
Quantum models with fixed encoding

$$\Theta_j = \{w_j, \theta_j\}$$



$$U_{j,\Theta_j}(x) = e^{-i\left(w_j^\top x+\theta_j\right)H_j}$$

Special case: fixed encoding



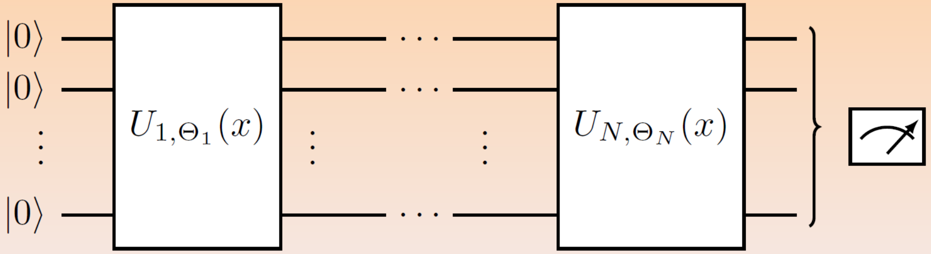
- **Parametrized gates**: $w_j = 0$ and $\theta_j = \varphi_j^i$
- **Data-dependent** gates: w_j is j-th unit vector, $\theta_j = 0$

→ Adapting w_j provides improved expressivity

“Data re-uploading for a universal quantum classifier”, A. Pérez-Salinas et al., Quantum, 2020
“Let quantum neural networks choose their own frequencies”, B. Jaderberg et al., arXiv:2309.03279, 2023

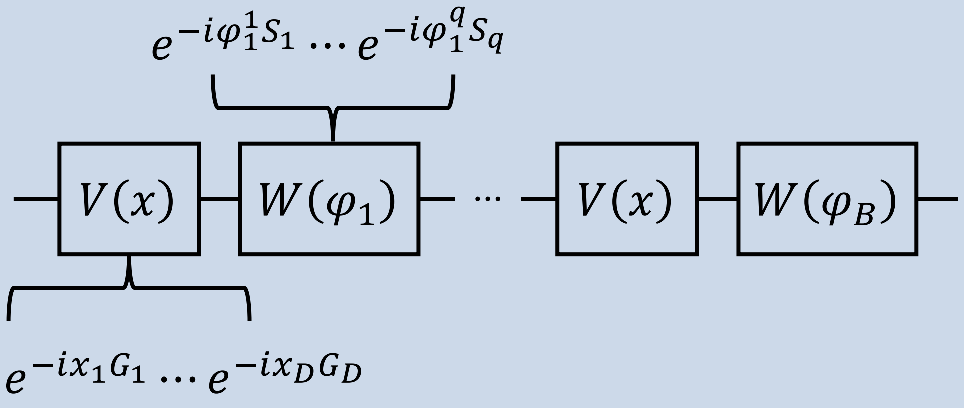
Benefits of trainable encodings

$$\Theta_j = \{w_j, \theta_j\}$$



$$U_{j,\Theta_j}(x) = e^{-i\left(w_j^\top x+\theta_j\right)H_j}$$

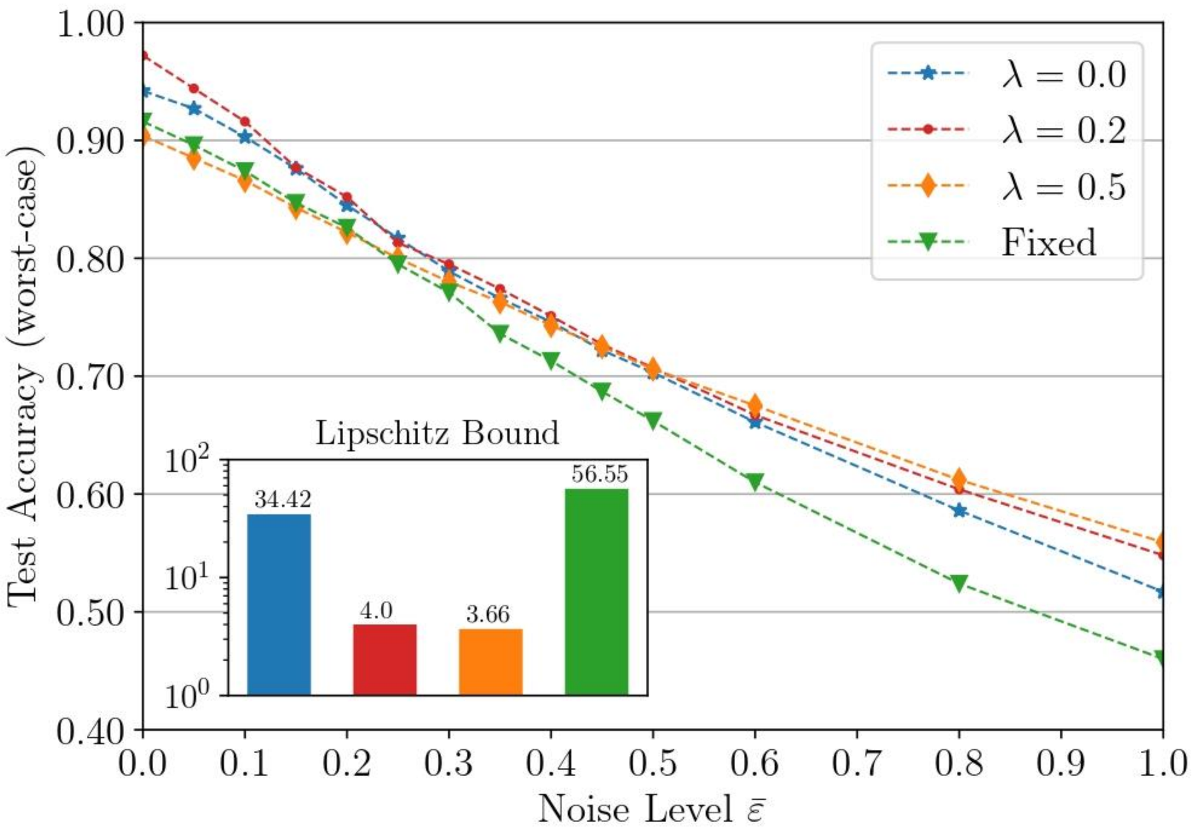
Special case: fixed encoding



- **Recall:** Lipschitz bound only depends on $w_j, H_j, M \Rightarrow$ **independent of θ_j**
- Fixed-encoding quantum models: **Lipschitz bound** $= 2\|M\| B \sum_j \|G_j\|$
  → **cannot be changed** during training
  → **limits influence of training** on robustness and generalization

Numerical results: robustness

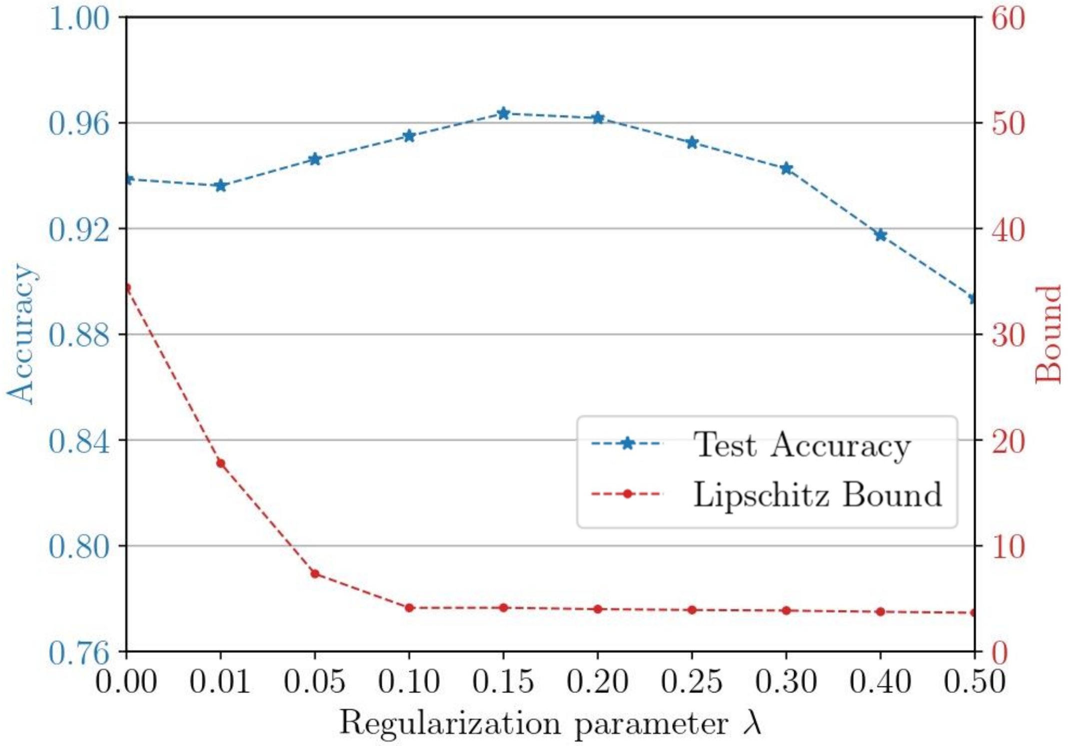
- Test data + noise from $\varepsilon \in [-\bar{\varepsilon}, \bar{\varepsilon}]$ (as before)
- Fixed-encoding quantum model with comparable circuit structure



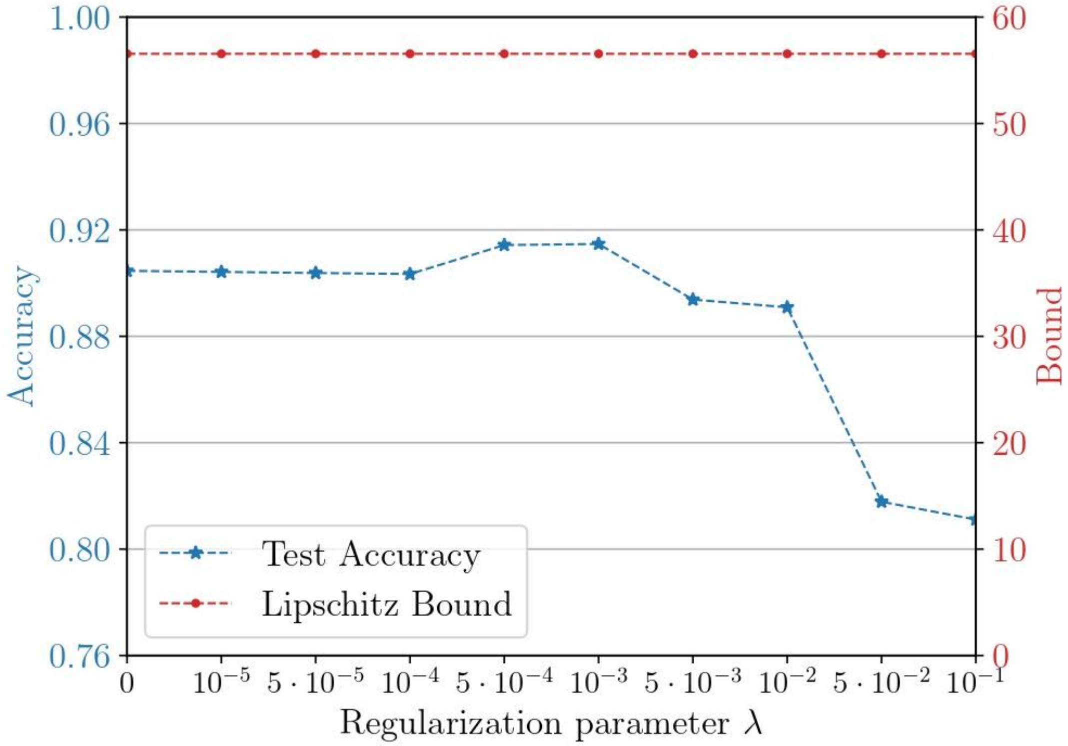
Trainable encoding + regularization improves robustness

Numerical results: generalization

Trainable encoding



Fixed encoding



Trainable encoding + regularization improves generalization

Conclusion

University of Stuttgart
Germany

SimTech

ist

Lipschitz bounds of quantum models

- Robustness
- Generalization
- Benefits of trainable encodings

Outlook:

- Tighter Lipschitz bounds
- Different quantum models
- Nonlinear data encodings

Details: arXiv:2311.11871

Julian Berberich
Institute for Systems Theory and Automatic Control
University of Stuttgart
julian.berberich@ist.uni-stuttgart.de

We acknowledge the support by the DFG under Germany's Excellence Strategy – EXC 2075 – 390740016, by the Stuttgart Center for Simulation Science (SimTech), and by the German Federal Ministry of Economic Affairs and Climate Action (grant no. 01MQ22002A).