Scalably learning quantum many-body Hamiltonians from dynamical data

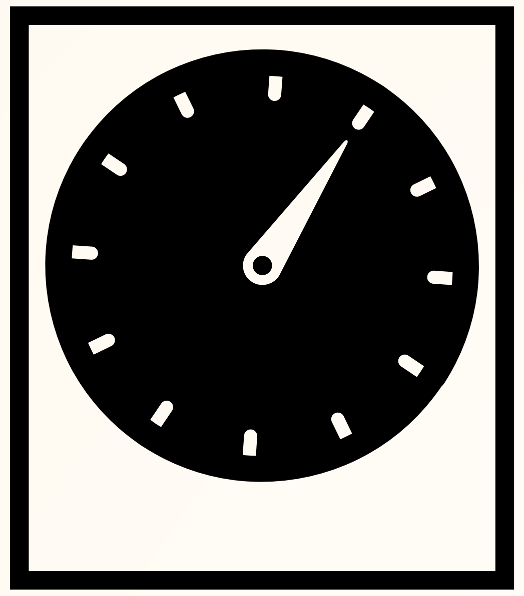
Frederik Wilde, Augustine Kshetrimayum, Ingo Roth, Dominik Hangleiter, Ryan Sweke, Jens Eisert

2022-12-12

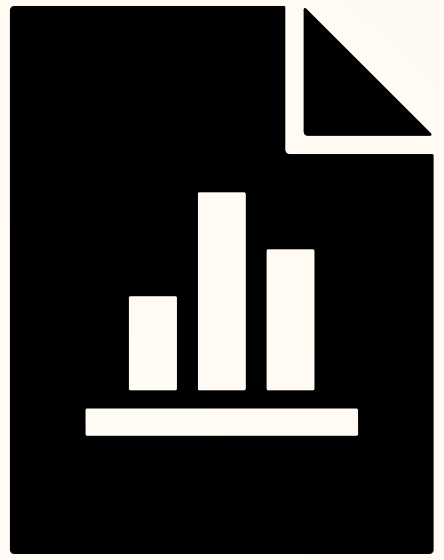
arXiv:2209.14328

frederikwil.de/eru2022

Freie Universität Berlin

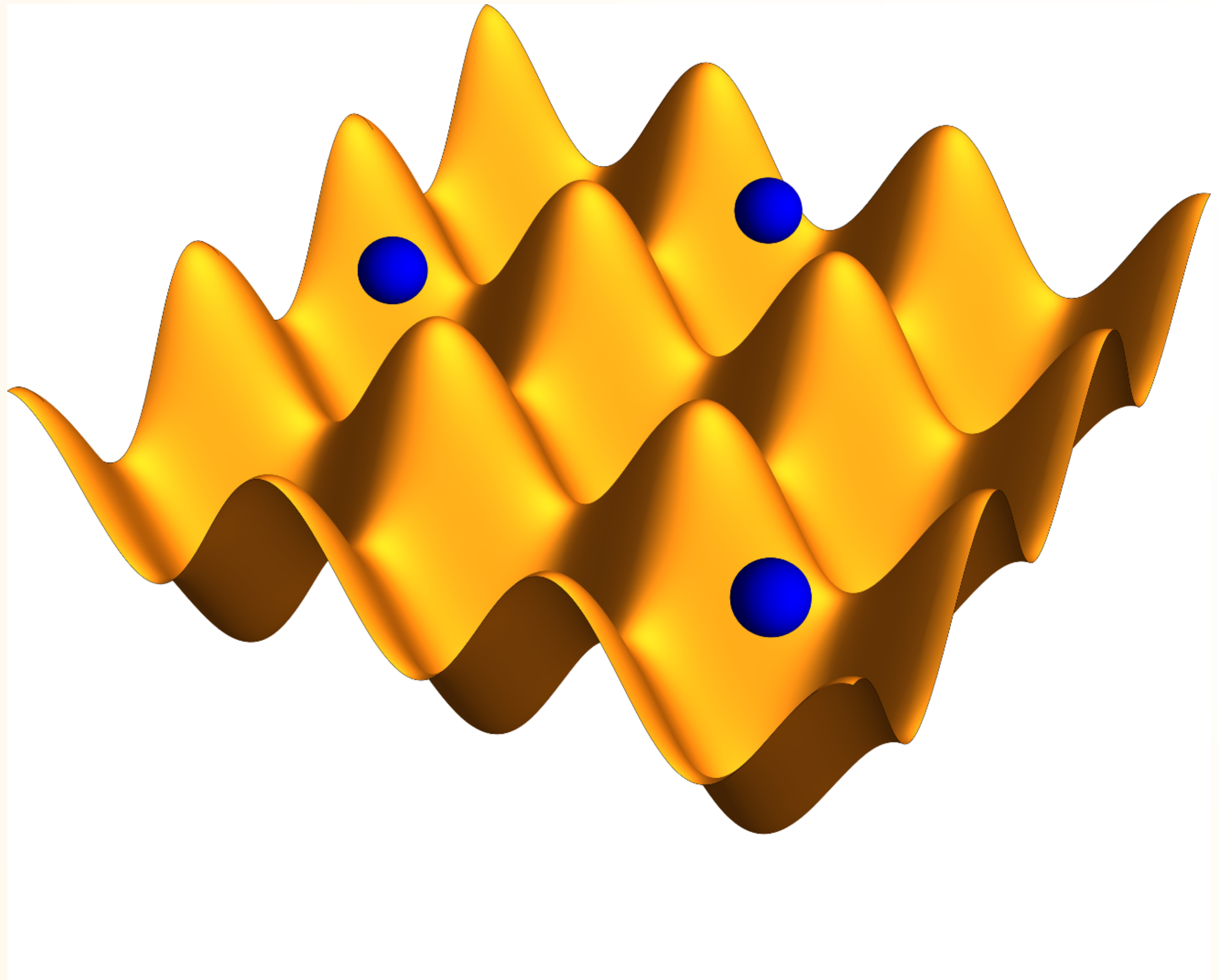


Data



Measure

$U(t, t_0)$



What is the Hamiltonian of the system?

Previous methods

- Thermal and ground states
- (Process) tomography (on subsystems)
- Bayesian estimation via MC integration
- "Infinitesimal" time evolution
- Energy conservation
- Classical (SOTA) ML methods
- Quantum assisted
- Time traces/stamps

Lots of methods available.

Ours is very **experimentally friendly** and can handle **large quantum many-body systems** with **many unknown parameters**.

$|0\rangle^{\otimes n}$
Time evolution in the lab under H^*
$|\psi(t)\rangle$
Measurement
$(p_k,\ s_i) = (ZZXZ, 0100)$
Time-Evolving Block Decimation (TEBD)
$|0\rangle^{\otimes n}$
Simulated time evolution under $H(\theta)$
$|\psi_\theta(t)\rangle$
Probabilities
$\mathbb{P}\left(s_i\,|\,t, p_k, \theta\right)$
$|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$
$e^{-i\Delta t\, h_{12}}$
$e^{-i\Delta t\, h_{34}}$
$e^{-i\Delta t\, h_{23}}$
$|0\rangle$ $|1\rangle$ $|+\rangle$ $|0\rangle$
$\left|\langle\phi_{ik}|e^{-iH(\theta)t}|\psi_{\text{ini}}\rangle\right|^2$
Minimize $-\sum_{i,k} \log \mathbb{P}(s_i\,|\,t, p_k, \theta)$
θ
1
2
3
Epochs
Negative log-likelihood loss function
$\mathscr{L}^{(d)}(\theta) = -\frac{1}{d}\sum_{ijk} \log \mathbb{P}(\hat{s}_i\,|\,p_k, t_j, \theta)$
Minimizer
$\hat{\theta}^{(d)} = \underset{\theta}{\text{argmin}}\ \mathscr{L}^{(d)}(\theta)$
Relative error
$\hat{\epsilon}^{(d)} = \frac{\|\hat{\theta}^{(d)} - \theta^*\|_2}{\|\theta^*\|_2}$
$i = 1,\ldots,M \quad j = 1,\ldots,J \quad k = 1,\ldots,K$

Strategy

Inspired by machine learning, use **automatic differentiation** and mini-batch **stochastic gradients**. (as well as just-in-time compilation and vectorization)

But what's the derivative of the SVD? $A \mapsto U, S, V^\dagger$

$$\mathrm{d}U = \frac{U}{2}\left[F \circ (\mathrm{d}\tilde{A}S + S\mathrm{d}\tilde{A}^\dagger) + S^{-1} \circ (\mathrm{d}\tilde{A} - \mathrm{d}\tilde{A}^\dagger)\right], \quad \text{(D12)}$$

$$\mathrm{d}V = \frac{V}{2}\left[F \circ (S\mathrm{d}\tilde{A} + \mathrm{d}\tilde{A}^\dagger S)\right]. \quad \text{(D13)}$$

$$\frac{1}{s_j^2 - s_i^2}$$

github.com/google/jax/pull/5225

$$H(J_x, J_y, J_z, h_1, \ldots, h_n) = \sum_i J_x X_i X_{i+1} + J_y Y_i Y_{i+1} + J_z Z_i Z_{i+1} + h_i X_i$$

- 1D Heisenberg model
- Generate synthetic data
- Run the optimizer starting from multiple random initial points θ_{ini}

$d^{-1/2}$ - scaling of the error

Theorem 1 (Asymptotic error (informal)). *Suppose the class of parameterized Hamiltonians $\mathcal{C}$ is well conditioned, such that the estimator $\hat{\theta}^{(d)}$ is consistent and the Hessian of the loss is non-singular. Then for any $\delta \in (0,1]$ there exists a function $f(d) = \mathcal{O}(d^{-1/2})$ such that $\mathbb{P}\left[\hat{\epsilon}^{(d)} > f(d)\right] < \delta$ when d is sufficiently large.*

Statement about the global minimizer.
No guarantee that this can be found efficiently!

Not trivial to show!
Obvious obstacles: eigenstates, commensurate times, dense $2^n \times 2^n$ Hamiltonians.
But in typically one time stamp is *in theory* sufficient.

Li, et al. 2020

Thank you for your attention!

arXiv:2209.14328

frederikwil.de/eru2022

Formal statement of the theorem

Theorem 3 (Asymptotic error). *Let $\hat{\mathcal{L}}^{(d)}$ and $\hat{\theta}^{(d)}$ be defined as above. Let $\Theta \subset \mathbb{R}^{\nu}$ be compact and suppose the true parameters θ^* are contained in the interior of Θ. Suppose the class*

$$\mathcal{C} = \{H(\theta) | \theta \in \Theta\}$$

of parametrized Hamiltonians is well conditioned, such that the estimator $\hat{\theta}^{(d)}$ is consistent and the Hessian at the true parameter value is non-singular. Furthermore, let the parametrization $\theta \mapsto H(\theta)$ be twice continuously differentiable. Then for any $\delta \in (0, 1]$ there exists a function $f(d) = \mathcal{O}(d^{-1/2})$ and some value $D \in \mathbb{N}$ such that $\mathbb{P}\left[\hat{\epsilon}^{(d)} > f(d)\right] < \delta$ for all $d \geq D$.

Proof idea

$$0 = \nabla \mathscr{L}^{\mathscr{D}}(\theta^*)$$ first-order condition

$$0 = \nabla \mathscr{L}^{D}(\theta^{D}) \; + \; \nabla^2 \mathscr{L}^{D}(\bar{\theta}) \; (\theta^{D} - \theta^*)$$ mean-value theorem (element-wise)

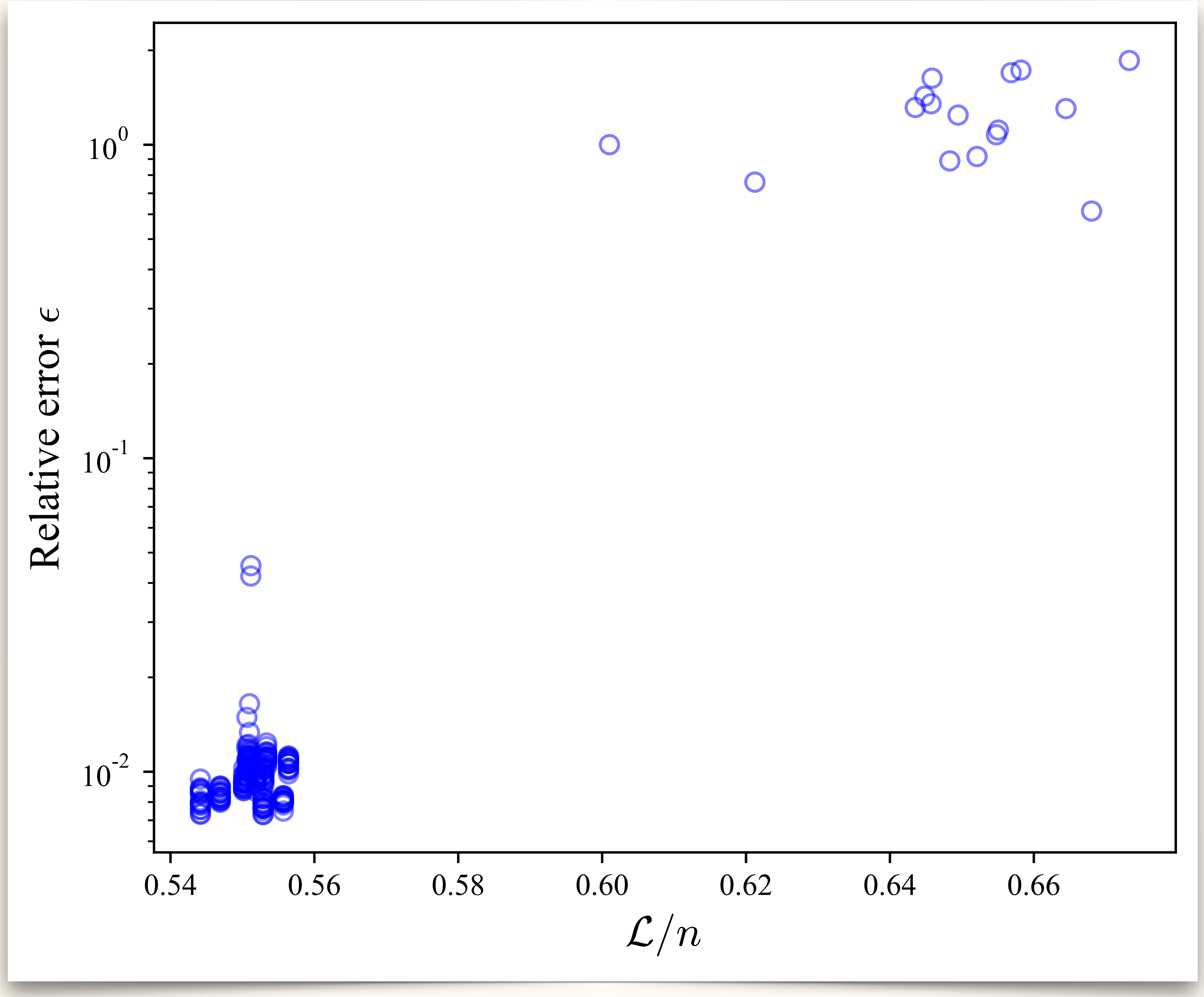
$$\sqrt{d} \; (\theta^{D} - \theta^*) = -\frac{1}{d} \left[\sum_{s \in D} \nabla^2 \log \mathbb{P}(s \,|\, \bar{\theta}) \right]^{-1} \frac{1}{\sqrt{d}} \sum_{s \in D} \nabla \log \mathbb{P}(s \,|\, \theta^*) \xrightarrow{\text{dist.}} \mathscr{N}(0, \tilde{\Sigma})$$

Newey, McFadden 1994

inverse Hessian

central-limit theorem

Loss vs. error



Loss landscape