Parameter Estimation for Quantum Information

Christopher Granade

www.cgranade.com • *cgranade@cgranade.com*

Joint work with:
Christopher Ferrie   Nathan Wiebe   D. G. Cory

Institute for Quantum Computing
University of Waterloo, Ontario, Canada

June 18, 2013

LFQIS 2013, Denali Park

- Characterizing unknown quantum systems is critical for design of control.

- Characterizing unknown quantum systems is critical for design of control.
- Enabling adaptive measurement allows for large reductions in data collection costs.

- Characterizing unknown quantum systems is critical for design of control.
- Enabling adaptive measurement allows for large reductions in data collection costs.
- Want accurate reporting of errors incurred by estimate, and of smallest credible regions.

Online adaptive characterization of quantum systems can improve and enable experimental practice, including in nitrogen vacancy centers, neutron interferometers, and in superconducting qubit circuits.

Counting Statistics

Nitrogen vacancy centers in diamond are measured by optical readout. The number of photons emitted by a center in an interval Δt depends on the m_s quantum number of the center.

Manson et al. 2006

Counting Statistics

Nitrogen vacancy centers in diamond are measured by optical readout. The number of photons emitted by a center in an interval Δt depends on the m_s quantum number of the center.

Dark counts, quantum efficiency, stray flourescences all affect statistics of photon detection.

Manson et al. 2006

Counting Statistics

Nitrogen vacancy centers in diamond are measured by optical readout. The number of photons emitted by a center in an interval Δt depends on the m_s quantum number of the center.

Dark counts, quantum efficiency, stray flourescences all affect statistics of photon detection.

Estimation Problem

Given that n_d photons were observed, what state was the NV center in?

Manson et al. 2006

Precise Magnetometry

$$H = \Delta S_z^2 + \gamma \underline{B} \cdot \underline{S}$$

Energy levels in an NV center are split by magnetic fields. By preparing, evolving and measuring different states, we thus gain information about $\underline{B}$.

Said et al. 2011

Precise Magnetometry

$$H = \Delta S_z^2 + \gamma \underline{B} \cdot \underline{S}$$

Energy levels in an NV center are split by magnetic fields. By preparing, evolving and measuring different states, we thus gain information about $\underline{B}$.

Said et al. 2011

Precise Magnetometry

$$H = \Delta S_z^2 + \gamma \underline{B} \cdot \underline{S}$$

Energy levels in an NV center are split by magnetic fields. By preparing, evolving and measuring different states, we thus gain information about $\underline{B}$.

Estimation Problem

Given a set of observed photon counts, what is the strength and direction of the magnetic field $\underline{B}$?

Said et al. 2011

Precise Magnetometry

$$H = \Delta S_z^2 + \gamma \underline{B} \cdot \underline{S}$$

Energy levels in an NV center are split by magnetic fields. By preparing, evolving and measuring different states, we thus gain information about $\underline{B}$.

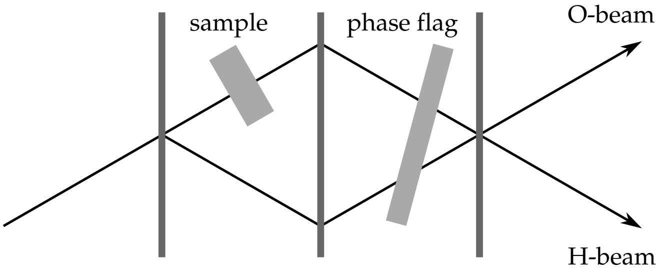
Estimation Problem

Given a set of observed photon counts, what is the strength and direction of the magnetic field $\underline{B}$?

By applying a magnetic field gradient such that $\underline{B} = \underline{B}(\underline{r})$, measurement of the NV center reveals information about its location in the diamond.

Said et al. 2011

Neutron Interferometry Geometry

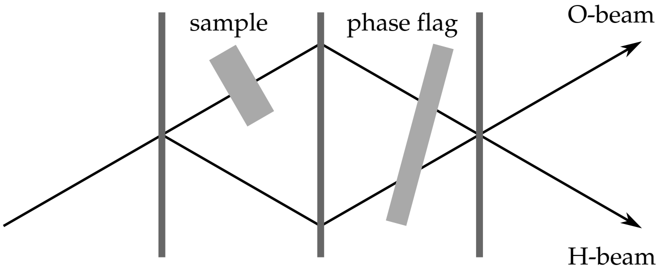


The sample introduces a phase difference of ϕ between the two paths. By rotating the phase flag, an additional phase of θ can be introduced, so that the ideal probabilty of a neutron reaching the O-beam detector is

$$\Pr(\text{O-beam}) = \cos^2(\phi + \theta).$$

Pushin 2007

Neutron Interferometry Geometry



The sample introduces a phase difference of ϕ between the two paths. By rotating the phase flag, an additional phase of θ can be introduced, so that the ideal probabilty of a neutron reaching the O-beam detector is

$$\Pr(\text{O-beam}) = \cos^2(\phi + \theta).$$

In practice, there is a limited *contrast* between the two beams, related to the *visibility*.

Pushin 2007

Improved Contrast

Due to interaction with the environment, a phase difference $\Delta\phi(\underline{\epsilon})$ is introduced for a state $\underline{\epsilon}$ of the environment. Averaging over this random phase costs contrast in the final signal.

Improved Contrast

Due to interaction with the environment, a phase difference $\Delta\phi(\underline{\epsilon})$ is introduced for a state $\underline{\epsilon}$ of the environment. Averaging over this random phase costs contrast in the final signal.

Estimation Problem

By measuring the temperature, humidity, etc., as well as the neutron count, can we improve contrast and measure the static phase difference with better accuracy?

Spectral Density Estimation

In order to characterize a superconducting qubit circuit, we must know what decoherence mechanisms the system is subject to. As such, we would like to know the power spectral density $S(\omega)$ of the environment.

Yan et al. 2012

Spectral Density Estimation

In order to characterize a superconducting qubit circuit, we must know what decoherence mechanisms the system is subject to. As such, we would like to know the power spectral density $S(\omega)$ of the environment.

By measuring the circuit with a variety of dynamical decoupling pulse trains, we gain information about S.

Yan et al. 2012

Spectral Density Estimation

In order to characterize a superconducting qubit circuit, we must know what decoherence mechanisms the system is subject to. As such, we would like to know the power spectral density $S(\omega)$ of the environment.

By measuring the circuit with a variety of dynamical decoupling pulse trains, we gain information about S.

Estimation Problem

Given measurements of the superconducting circuit, what is the power spectral density of its environment?

Yan et al. 2012

Modeling Data Collection

Model data collection as a probability distribution, called a *likelihood function*

$$\Pr(d|\underline{x};\underline{e}).$$

d: data, $\underline{x}$: model, $\underline{e}$: experiment

Modeling Data Collection

Model data collection as a probability distribution, called a *likelihood function*

$$\Pr(d|\underline{x};\underline{e}).$$

d: data, $\underline{x}$: model, $\underline{e}$: experiment

Example

Consider a single qubit undergoing Larmor precession at an unknown frequency ω, with unknown dephasing time T_2:

$$H(\omega) = \frac{\omega}{2}\sigma_z, \quad |\psi_{\text{in}}\rangle = |+\rangle, \quad M = \{|+\rangle\langle+|, |-\rangle\langle-|\}$$

$$\Pr(d = 0|\underline{x} = (\omega, T_2); \underline{e} = (t)) = \frac{1}{2}(1 - e^{-t/T_2}) + e^{-t/T_2}\cos^2(\omega t/2)$$

Updating Knowledge

Once we have a likelihood function for our model, we can reason about

$$\Pr(\underline{x}|d, \underline{e}),$$

what we know about our model having seen some data.

Updating Knowledge

Once we have a likelihood function for our model, we can reason about

$$\Pr(\underline{x}|d, \underline{e}),$$

what we know about our model having seen some data.
By Bayes' rule,

$$\Pr(\underline{x}|d, \underline{e}) = \frac{\Pr(d|\underline{x}; \underline{e})}{\Pr(d|\underline{e})} \Pr(\underline{x}),$$

telling us that our knowledge is intimately connected to our ability to simulate.

Updating Knowledge

Once we have a likelihood function for our model, we can reason about

$$\Pr(\underline{x}|d, \underline{e}),$$

what we know about our model having seen some data.
By Bayes' rule,

$$\Pr(\underline{x}|d, \underline{e}) = \frac{\Pr(d|\underline{x}; \underline{e})}{\Pr(d|\underline{e})} \Pr(\underline{x}),$$

telling us that our knowledge is intimately connected to our ability to simulate.

Report as estimate of $\underline{x}$ the expectation value over $\underline{x}$,

$$\hat{\underline{x}} = \mathbb{E}[\underline{x}] = \int \underline{x} \Pr(\underline{x}) \, \mathrm{d}\underline{x}.$$

Loss

We require a figure of merit how how well we have learned a model. Thus, we assign to each estimate $\hat{\underline{x}}$ of a "true" model $\underline{x}$ a *loss*, describing how bad $\hat{\underline{x}}$ does at estimating $\underline{x}$.

Loss

We require a figure of merit how how well we have learned a model. Thus, we assign to each estimate $\hat{\underline{x}}$ of a "true" model $\underline{x}$ a *loss*, describing how bad $\hat{\underline{x}}$ does at estimating $\underline{x}$.

Definition (Quadratic Loss)

$$L_{\underline{\underline{Q}}}(\hat{\underline{x}}, \underline{x}) = (\hat{\underline{x}} - \underline{x})^{\mathrm{T}} \underline{\underline{Q}} (\hat{\underline{x}} - \underline{x}),$$

where $\underline{\underline{Q}}$ is a positive semidefinite matrix that establishes the scale between the various model parameters.

Loss

We require a figure of merit how how well we have learned a model. Thus, we assign to each estimate $\hat{\underline{x}}$ of a "true" model $\underline{x}$ a *loss*, describing how bad $\hat{\underline{x}}$ does at estimating $\underline{x}$.

Definition (Quadratic Loss)

$$L_{\underline{\underline{Q}}}(\hat{\underline{x}}, \underline{x}) = (\hat{\underline{x}} - \underline{x})^{\mathrm{T}} \underline{\underline{Q}} (\hat{\underline{x}} - \underline{x}),$$

where $\underline{\underline{Q}}$ is a positive semidefinite matrix that establishes the scale between the various model parameters.

The quadratic loss generalizes the mean-squared error for the case of multiple parameters.

Risk and Bayes Risk

Thinking of an estimator as a function from data records D to estimates $\hat{x}(D)$, we can reason about what the loss will be on average.

Risk and Bayes Risk

Thinking of an estimator as a function from data records D to estimates $\hat{x}(D)$, we can reason about what the loss will be on average.

Definition (Risk)

$$R(\hat{\underline{x}}, \underline{x}) = \mathbb{E}_D[L(\hat{\underline{x}}(D), \underline{x})]$$

Risk and Bayes Risk

Thinking of an estimator as a function from data records D to estimates $\hat{x}(D)$, we can reason about what the loss will be on average.

Definition (Risk)

$$R(\hat{\underline{x}}, \underline{x}) = \mathbb{E}_D[L(\hat{\underline{x}}(D), \underline{x})]$$

Since we don't know the true model *a priori*, we average again to obtain the Bayes risk.

Risk and Bayes Risk

Thinking of an estimator as a function from data records D to estimates $\hat{x}(D)$, we can reason about what the loss will be on average.

Definition (Risk)

$$R(\hat{\underline{x}}, \underline{x}) = \mathbb{E}_D[L(\hat{\underline{x}}(D), \underline{x})]$$

Since we don't know the true model *a priori*, we average again to obtain the Bayes risk.

Definition (Bayes Risk)

$$r(\hat{\underline{x}}, \pi) = \mathbb{E}_{\underline{x}\sim\pi}[R(\hat{\underline{x}}, \underline{x})]$$

Cramér-Rao Bound

The Fisher information

$$\underline{\underline{I}}(\underline{x}) = \mathbb{E}_D[(\underline{\nabla}_{\underline{x}} \log \Pr(D|\underline{x}))(\underline{\nabla}_{\underline{x}} \log \Pr(D|\underline{x}))^{\mathrm{T}}]$$

describes how much information about $\underline{x}$ is obtained by sampling data.

Cramér-Rao Bound

The Fisher information

$$\underline{\underline{I}}(\underline{x}) = \mathbb{E}_D[(\underline{\nabla}_{\underline{x}} \log \Pr(D|\underline{x}))(\underline{\nabla}_{\underline{x}} \log \Pr(D|\underline{x}))^{\mathrm{T}}]$$

describes how much information about $\underline{x}$ is obtained by sampling data.

The Cramér-Rao bound then tells us how well any unbiased estimator can perform. If $\underline{\underline{Q}} = \mathbb{1}$, then

$$R(\hat{\underline{x}}, \underline{x}) = \mathrm{Tr}(\mathrm{Cov}(\hat{\underline{x}})) \geq \mathrm{Tr}(\underline{\underline{I}}(\underline{x})^{-1}).$$

Cramér-Rao Bound

The Fisher information

$$\underline{\underline{I}}(\underline{x}) = \mathbb{E}_D[(\underline{\nabla}_{\underline{x}} \log \Pr(D|\underline{x}))(\underline{\nabla}_{\underline{x}} \log \Pr(D|\underline{x}))^{\mathrm{T}}]$$

describes how much information about $\underline{x}$ is obtained by sampling data.

The Cramér-Rao bound then tells us how well any unbiased estimator can perform. If $\underline{\underline{Q}} = \mathbb{1}$, then

$$R(\hat{\underline{x}}, \underline{x}) = \mathrm{Tr}(\mathrm{Cov}(\hat{\underline{x}})) \geq \mathrm{Tr}(\underline{\underline{I}}(\underline{x})^{-1}).$$

Compare to the quantum Cramér-Rao bound, which corresponds to the Heisenberg limit, and represents quantum mechanical limits rather than practical limits in specific scenarios.

Bayesian Cramér-Rao Bound

As before, integrating the Fisher information over the prior distribution π results in a Bayesian analog, the Bayesian Cramér-Rao bound:

$$r(\pi) \geq \left(\mathbb{E}_{\underline{x}}[\underline{\underline{I}}(\underline{x})]\right)^{-1}.$$

The BCRB can be computed iteratively, making it useful for tracking optimality in an online fashion.

Gill and Levit 1995

Sequential Monte Carlo

To implement our approach on a computer, we approximate distributions by a sum over weighted delta functions,

$$\Pr(\underline{x}) = \sum_i w_i \cdot \delta(\underline{x} - \underline{x}_i).$$

Doucet and Johansen 2011, Granade et al 2012

Sequential Monte Carlo

To implement our approach on a computer, we approximate distributions by a sum over weighted delta functions,

$$\Pr(\underline{x}) = \sum_i w_i \cdot \delta(\underline{x} - \underline{x}_i).$$

Each term in this sum is called a *particle*, and is described by the weight w_i and the model $\underline{x}_i$.

Doucet and Johansen 2011, Granade et al 2012

Sequential Monte Carlo

To implement our approach on a computer, we approximate distributions by a sum over weighted delta functions,

$$\Pr(\underline{x}) = \sum_i w_i \cdot \delta(\underline{x} - \underline{x}_i).$$

Each term in this sum is called a *particle*, and is described by the weight w_i and the model $\underline{x}_i$.

Updates to distributions now require evaluation the model $\Pr(d|\underline{x};\underline{e})$ at a finite number of points. Integrals over distributions are now represented by finite sums.

Doucet and Johansen 2011, Granade et al 2012

Numerical Stability and Resampling

With large amounts of data, the likelihood of any particular data record becomes very small, such that the normalization of the weights introduces numerical instabilities.

Liu and West 2001

Numerical Stability and Resampling

With large amounts of data, the likelihood of any particular data record becomes very small, such that the normalization of the weights introduces numerical instabilities.

Can mitigate by *resampling*: moving information from the weights to the density of SMC particles.

Liu and West 2001

Numerical Stability and Resampling

With large amounts of data, the likelihood of any particular data record becomes very small, such that the normalization of the weights introduces numerical instabilities.

Can mitigate by *resampling*: moving information from the weights to the density of SMC particles.

Sample new particle locations from mixed-normal distribution, with each normal centered on an old particle.

$$\Pr(\underline{x}') \propto \sum_i w_i \exp\left((\underline{x}' - \underline{\mu}_i)^{\mathrm{T}} \underline{\underline{\Sigma}} (\underline{x}' - \underline{\mu}_i)\right)$$

$$\underline{\mu}_i = a\underline{x}_i + (1-a)\mathbb{E}[\underline{x}]$$

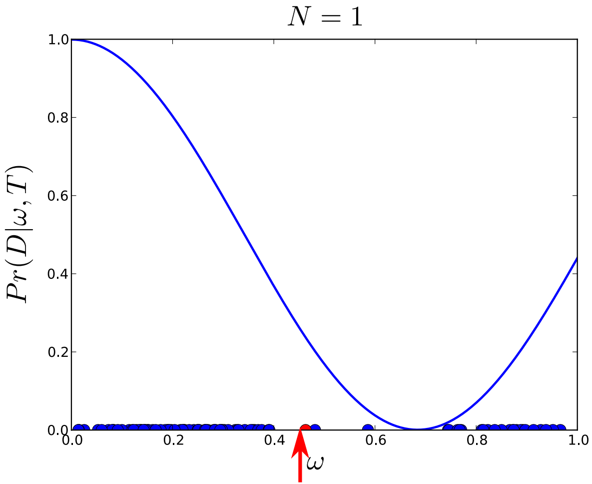
$$\underline{\underline{\Sigma}} = (1-a^2)\,\mathrm{Cov}[\underline{x}]$$

Set new weights to be uniform, hence resetting numerical stability.

Liu and West 2001

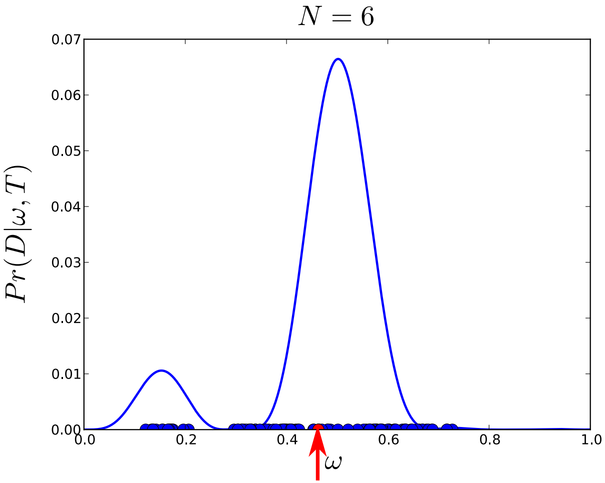
Sequential Monte Carlo

With SMC and resampling, particles move towards the true model as data is collected.



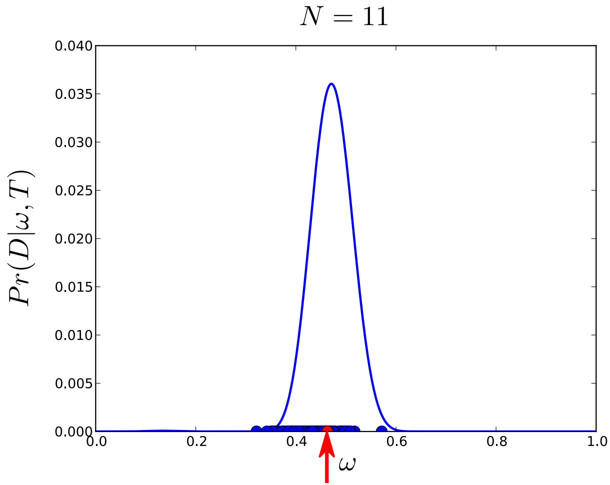
Sequential Monte Carlo

With SMC and resampling, particles move towards the true model as data is collected.



Sequential Monte Carlo

With SMC and resampling, particles move towards the true model as data is collected.



Experiment Design

Given the *utility* for an experiment $\underline{e}$, optimization algorithms can be used to find the most useful next experiment.

Experiment Design

Given the *utility* for an experiment $\underline{e}$, optimization algorithms can be used to find the most useful next experiment.

Since utilities are often multimodal for highly periodic models, *locally optimizing* each of many initial guesses works well. In our work, we use the Newton Conjugate-Gradient method (NCG) for locally optimizing each guess.

Experiment Design

Given the *utility* for an experiment $\underline{e}$, optimization algorithms can be used to find the most useful next experiment.

Since utilities are often multimodal for highly periodic models, *locally optimizing* each of many initial guesses works well. In our work, we use the Newton Conjugate-Gradient method (NCG) for locally optimizing each guess.

The quality of experiment design depends on having good heuristics for generating guesses. For Larmor precession model, exponentially sparse heuristics work well.

Cost and Utility

In order to select experiments, we assign a *cost* function, describing how hard it is to perform an experiment. For instance, $\$(\underline{e}) = at + b$ describes that there is a cost to evolving for a time t, as well as a constant per-experiment cost.

Cost and Utility

In order to select experiments, we assign a *cost* function, describing how hard it is to perform an experiment. For instance, $\$(\underline{e}) = at + b$ describes that there is a cost to evolving for a time t, as well as a constant per-experiment cost.

The *utility* can then be defined as the reduction in Bayes risk per cost,

$$U(\underline{e}) = \frac{\Delta r(\pi; \underline{e})}{\$(\underline{e})}.$$

Cost and Utility

In order to select experiments, we assign a *cost* function, describing how hard it is to perform an experiment. For instance, $\$(\underline{e}) = at + b$ describes that there is a cost to evolving for a time t, as well as a constant per-experiment cost.

The *utility* can then be defined as the reduction in Bayes risk per cost,

$$U(\underline{e}) = \frac{\Delta r(\pi; \underline{e})}{\$(\underline{e})}.$$

If $\$(\underline{e})$ is constant, then we can use the *negative variance* utility instead, as both functions are optimized at the same experiment:

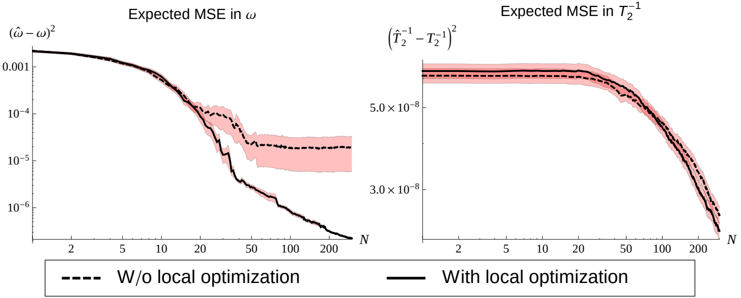
$$U_{\mathrm{NV}}(\underline{e}) = -\mathbb{E}_d[\mathrm{Tr}(\underline{\underline{Q}} \cdot \mathrm{Cov}_{\underline{x}|d;\underline{e}}(\underline{x}))].$$

Numerical Results: Unknown T_2

Our method allows for ω and T_2 to be learned with very few measurements.

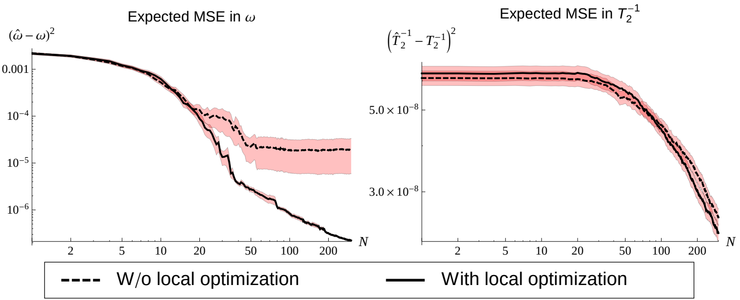
Numerical Results: Unknown T_2

Our method allows for ω and T_2 to be learned with very few measurements: only 200 bits of data!



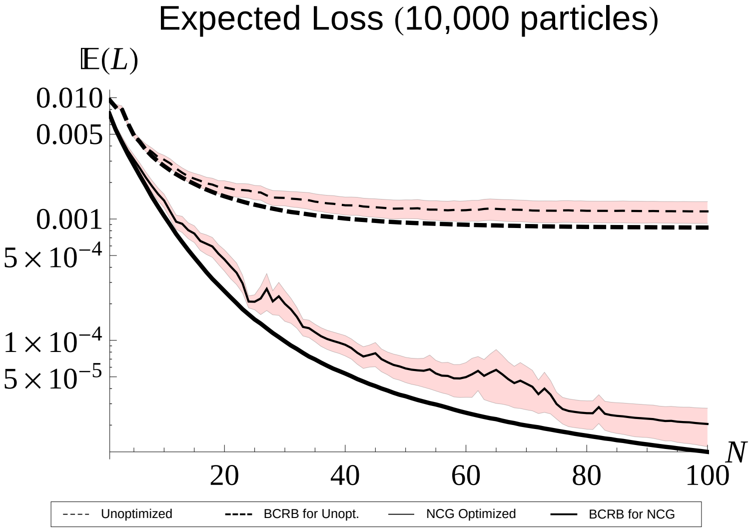
Numerical Results: Unknown T_2

Our method allows for ω and T_2 to be learned with very few measurements: only 200 bits of data!



Moreover, using our best available knowledge to optimize experiment designs, we can continue to learn about ω even in the presence of unknown T_2.

Numerical Results: Adaptive Experiment Design



Simulation and learning are intimately connected: if we can simulate a model under different hypotheses, then we can determine which hypotheses describe an unknown system.

Simulation and learning are intimately connected: if we can simulate a model under different hypotheses, then we can determine which hypotheses describe an unknown system.

Big Idea

Use quantum simulation to learn about unknown quantum systems.

Weak and Strong Simulation

Van den Nest defines as *strong* simulation the calculation of a likelihood function, and *weak* simulation as the sampling of data from a likelihood function.

Aaronson and Arkhipov 2010, Josza and Miyake 2008, Van den Nest 2008 and 2011, Josza and Van den Nest 2013

Weak and Strong Simulation

Van den Nest defines as *strong* simulation the calculation of a likelihood function, and *weak* simulation as the sampling of data from a likelihood function.

Much of quantum simulation is concerned with the latter: producing data according to a desired distribution.

Problem

Sequential Monte Carlo seems to require strong simulation.

Aaronson and Arkhipov 2010, Josza and Miyake 2008, Van den Nest 2008 and 2011, Josza and Van den Nest 2013

Adaptive Likelihood Estimation

Solution

Treat estimating the likelihood as a secondary estimation problem.

Ferrie and Blume-Kohout 2012, Ferrie and Granade 2013

Adaptive Likelihood Estimation

Solution

Treat estimating the likelihood as a secondary estimation problem.

For a two-outcome model, a hedged binomial estimator finds the probability p_0 of a “0” outcome by repeatedly sampling a weak simulator.

Ferrie and Blume-Kohout 2012, Ferrie and Granade 2013

Adaptive Likelihood Estimation

Solution

Treat estimating the likelihood as a secondary estimation problem.

For a two-outcome model, a hedged binomial estimator finds the probability p_0 of a "0" outcome by repeatedly sampling a weak simulator.

The variance of such estimators is well-known, so repeatedly collect until a fixed *tolerance* is reached.

We will show later that SMC is robust to likelihood estimation errors.

Ferrie and Blume-Kohout 2012, Ferrie and Granade 2013

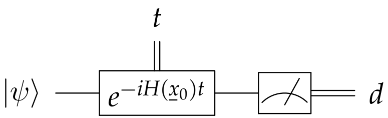
No Inversion Model

If we wish to compare the classical outcomes of an unknown quantum system to a simulator, then we can do so by preparing an input state $|\psi\rangle$, evolving under either the "true" model $\underline{x}_0$ or a simulated model $\underline{x}_e$, and then measuring.

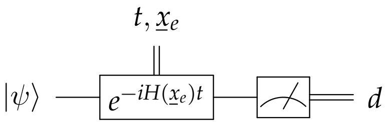
No Inversion Model

If we wish to compare the classical outcomes of an unknown quantum system to a simulator, then we can do so by preparing an input state $|\psi\rangle$, evolving under either the "true" model $\underline{x}_0$ or a simulated model $\underline{x}_e$, and then measuring.

Unknown System



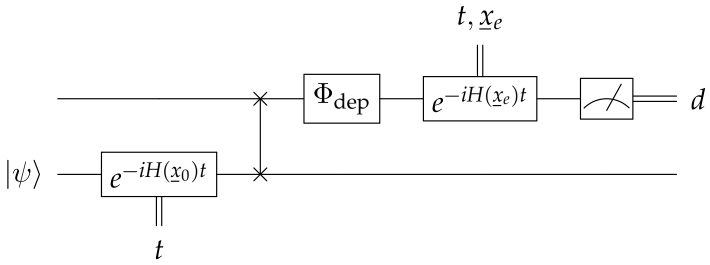
Simulator



Inversion Model

On the other hand, if the unknown system is coherently coupled to a quantum simulator, then we can transfer the state produced by a quantum device and attempt to invert the evolution by simulating according to a hypothesis.

We model noise in the coupling by a depolarizing channel Φ_{dep}.



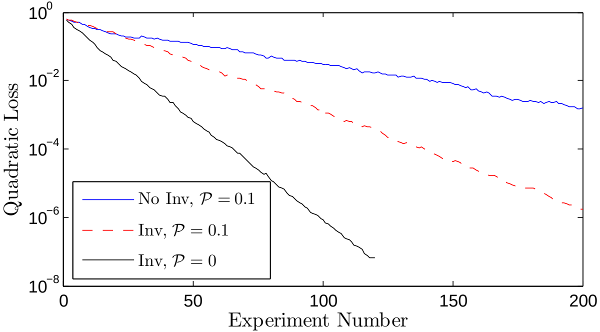
Posterior Particle Heuristic

The inversion model connects the model and experiment parameter spaces. Use this connection to come up with a heuristic for experiment designs.

- Choose $\underline{x}_e, \underline{x}_e' \sim \Pr(\underline{x})$, the most recent posterior.
- Choose $t = 1/\|\underline{x}_e - \underline{x}_e'\|$.
- Return $\underline{e} = (\underline{x}_e, t)$.

Ising Model on Spin Chains

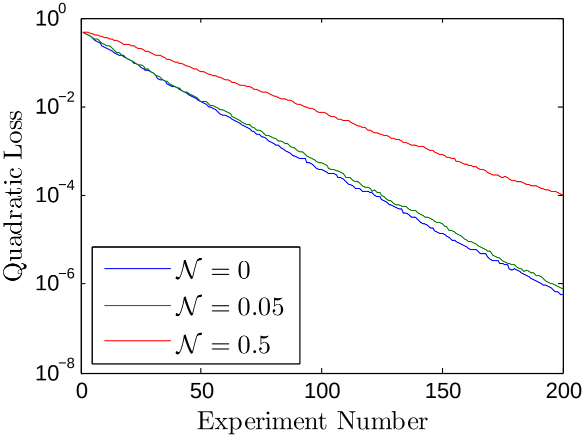
For a Hamiltonian representing nearest-neighbor Ising models on a chain of nine qubits, the inversion model allows for dramatic improvements over comparing classical data.



Here, $\mathcal{P}$ is the adaptive likelihood estimation tolerance.

Ising Model on the Complete Graph

Using the inversion model, our algorithm also works when the interaction graph is *complete*. Here, we show the performance as a function of the depolarization strength $\mathcal{N}$.

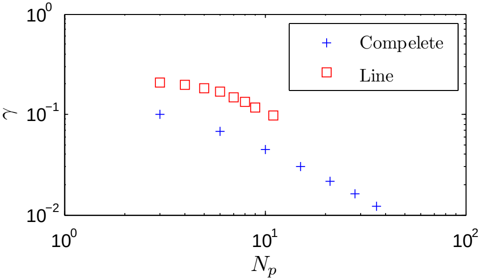


Scaling and Dimensionality

In both the spin-chain and complete graph cases, the quadratic loss on average decays exponentially, $L_{\underline{\underline{Q}}} \propto e^{-\gamma N}$, for some rate constant γ.

Scaling and Dimensionality

In both the spin-chain and complete graph cases, the quadratic loss on average decays exponentially, $L_{\underline{\underline{Q}}} \propto e^{-\gamma N}$, for some rate constant γ. Consider γ as a function of the model dimension N_p:



This suggests that, with access to a quantum simulator, learning may scale efficiently.

- Statistical inference can be used to characterize unknown quantum systems.

- Statistical inference can be used to characterize unknown quantum systems.
- Sequential Monte Carlo allows for Bayesian updates to be efficiently implemented on a classical computer.

- Statistical inference can be used to characterize unknown quantum systems.
- Sequential Monte Carlo allows for Bayesian updates to be efficiently implemented on a classical computer.
- Current best knowledge can be applied to adaptively design new experiments and measurements.

- Statistical inference can be used to characterize unknown quantum systems.
- Sequential Monte Carlo allows for Bayesian updates to be efficiently implemented on a classical computer.
- Current best knowledge can be applied to adaptively design new experiments and measurements.
- Our approach is *generic*, treating simulation as a resource for learning.

- Statistical inference can be used to characterize unknown quantum systems.
- Sequential Monte Carlo allows for Bayesian updates to be efficiently implemented on a classical computer.
- Current best knowledge can be applied to adaptively design new experiments and measurements.
- Our approach is *generic*, treating simulation as a resource for learning.
- We can apply quantum resources to characterize quantum systems.

Special Thanks

Thanks to Ian Hincks, Osama Moussa and Dimitri Pushin for their work and discussions in developing the motivating physical examples.

Further Information

Slides, a journal reference for this work, a full bibliography and a software implementation can be found at *http://www.cgranade.com/research/talks/lfqis2013/*.

Thank you for your kind attention!

Method of Hyperparameters

Suppose the "true" model $\underline{x}$ changes between experiments according to a distribution $\Pr(\underline{x}|\underline{y})$, for some vector of parameters $\underline{y}$, called *hyperparameters*.

Method of Hyperparameters

Suppose the "true" model $\underline{x}$ changes between experiments according to a distribution $\Pr(\underline{x}|\underline{y})$, for some vector of parameters $\underline{y}$, called *hyperparameters*.

We can estimate the hyperparameters instead of the "true" model,

$$\Pr(d|\underline{y};\underline{e}) = \int \Pr(d|\underline{x},\underline{y};\underline{e}) \Pr(\underline{x}|\underline{y};\underline{e})\, \mathrm{d}\underline{x}.$$

Method of Hyperparameters

Suppose the "true" model $\underline{x}$ changes between experiments according to a distribution $\Pr(\underline{x}|\underline{y})$, for some vector of parameters $\underline{y}$, called *hyperparameters*.

We can estimate the hyperparameters instead of the "true" model,

$$\Pr(d|\underline{y};\underline{e}) = \int \Pr(d|\underline{x},\underline{y};\underline{e}) \Pr(\underline{x}|\underline{y};\underline{e}) \, \mathrm{d}\underline{x}.$$

In many cases, hyperparameters represent an *epistimic decoherence*. For instance, the unknown T_2 model corresponds to ω being drawn from a Cauchy (Lorentz) distribution.