Bayesian Hamiltonian Learning

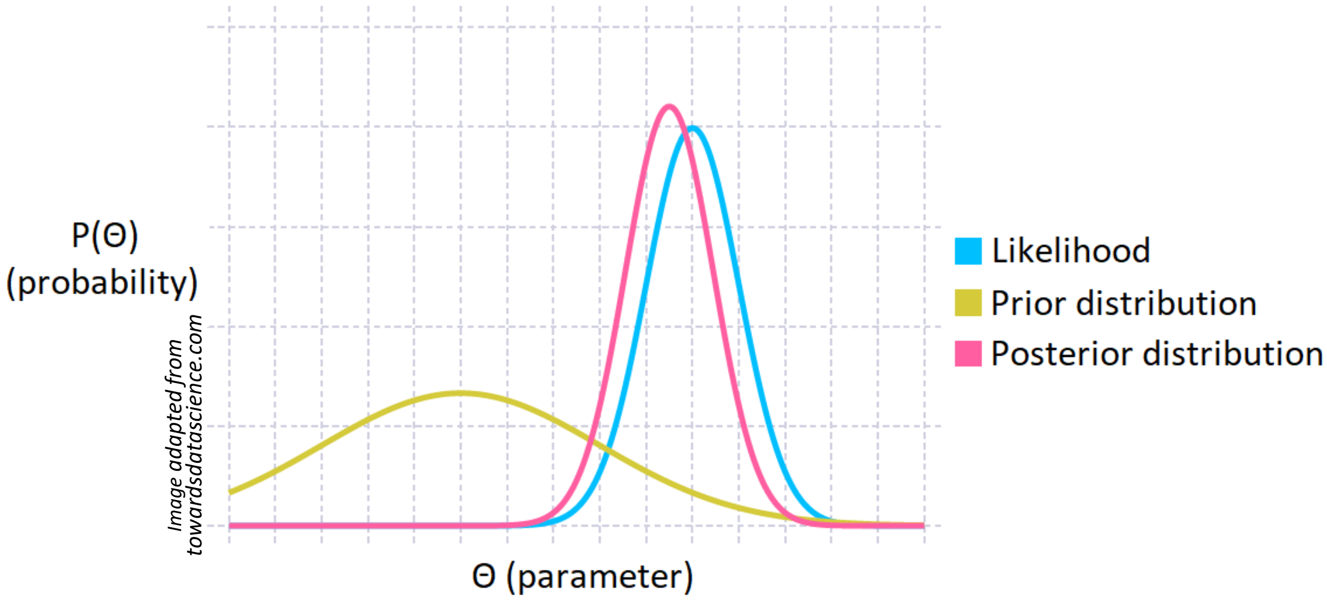
Alexandra Alves
University of Minho
QLOC talk, 11 November 2020

Universidade do Minho
Escola de Engenharia

Outline

- Bayesian parameter estimation
- Application to quantum systems
- Bayesian experimental design
- Offline vs. adaptive parameter estimation
- Implementation for a simple example and numerical results
- Characterizing open quantum systems

Bayesian Parameter Estimation



likelihood

prior distribution

$$\boxed{P(\Theta|D) = \frac{P(D|\Theta)P(\Theta)}{P(D)}}$$ **Bayes' rule**

"evidence"

$$\left[P(D) = \int P(D|\Theta)P(\Theta)d\Theta \right]$$

posterior distribution

Θ: parameters

D: data

Update our previous beliefs according to the *likelihood*
that they would have predicted the experimental data

Example: Estimating the Fairness of a Coin

For a single coin flip:

$P(Heads) = \ominus$

$P(Tails) = 1 - \ominus$

$\ominus$=?

The probability of heads is either ¼ or ¾.

We have to take a guess with 3 coin flips.

Which do we pick?

$P(Heads) = \ominus$

$P(Tails) = 1 - \ominus$

$\ominus = \frac{1}{4} \vee \frac{3}{4}$

Results: Head, Tail, Tail

$$P(\Theta|D,E)\ \alpha\ P(D|\Theta,E)P(\Theta)$$

(with $P(D|E)$ the proportionality constant, independent of Θ)

E: experiment *(flipping the coin N=3 times)*

D: experimental data *(the outcomes: heads or tails)*

- Our **prior** is $\boldsymbol{P(\ominus) = \frac{1}{2}\delta\left(\frac{1}{4}\right) + \frac{1}{2}\delta\left(\frac{3}{4}\right)}$.
- For this experiment and outcomes, the **likelihoods** for each $\ominus$ are $\boldsymbol{P(D|\Theta,E)}$= $=\ominus\cdot(\mathbf{1}-\ominus)\cdot(\mathbf{1}-\ominus)$.
- All that is left is the divisor, which can be thought as a **normalization factor**. We have $\boldsymbol{P(D|E)}$= $\boldsymbol{\sum_{\Theta} P(D|\Theta,E)P(\Theta)}$, where the sum is taken over our two thetas: $\frac{1}{4}$ and $\frac{3}{4}$.

$$P(Heads) = \ominus$$

$$P(Tails) = 1 - \ominus$$

$$\mathbf{P}(\ominus) = \frac{1}{2}\delta\left(\frac{1}{4}\right) + \frac{1}{2}\delta\left(\frac{3}{4}\right)$$

Results: Head, Tail, Tail

$$P(\Theta|D,E)\ \alpha\ P(D|\Theta,E)P(\Theta)$$

(with $P(D|E)$ the proportionality constant, independent of Θ)

E: experiment *(flipping the coin N=3 times)*
D: experimental data *(the outcomes: heads or tails)*

We then have:

$$P\left(\ominus = \frac{1}{4}\middle|\{Head, Tail, Tail\}, 3\ flips\right)\ \alpha\ P(\{Head, Tail, Tail\}|3\ flips)\cdot\frac{1}{2} = \left(\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{3}{4}\right)\cdot\frac{1}{2} = \frac{9}{128}$$

$$P\left(\ominus = \frac{3}{4}\middle|\{Head, Tail, Tail\}, 3\ flips\right)\ \alpha\ P(\{Head, Tail, Tail\}|3\ flips)\cdot\frac{1}{2} = \left(\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\right)\cdot\frac{1}{4} = \frac{3}{128}$$

And get, upon normalizing:

$$\boldsymbol{P}\left(\ominus = \mathbf{\frac{1}{4}}\right) \equiv P\left(\ominus = \frac{1}{4}\middle|\{Heads, Tails, Tails\}, 3\ flips\right) = \mathbf{\frac{3}{4}} \quad ; \quad \boldsymbol{P}\left(\ominus = \mathbf{\frac{3}{4}}\right) \equiv P\left(\ominus = \frac{1}{4}\middle|\{Heads, Tails, Tails\}, 3\ flips\right) = \mathbf{\frac{1}{4}}$$

$$P(Head) = \ominus$$

$$P(Tail) = 1 - \ominus$$

$$P_0\left(\ominus = \frac{1}{4}\right) = P_0\left(\ominus = \frac{3}{4}\right) = 50\%$$

Results: Head, Tail, Tail

$$P(\Theta|D,E)\ \alpha\ P(D|\Theta,E)P(\Theta)$$

(with $P(D|E)$ the proportionality constant, independent of Θ)

E: experiment *(flipping the coin 3times)*
D: experimental data *(the outcomes: heads or tails)*

The **Bayesian** update has changed our beliefs to better match the experimental evidence:

$$\boldsymbol{P}\left(\ominus = \frac{\mathbf{1}}{\mathbf{4}}\right) \equiv P\left(\ominus = \frac{1}{4}\middle|\{Head, Tail, Tail\}, 3\ flips\right) = \frac{\mathbf{3}}{\mathbf{4}}$$

$$\boldsymbol{P}\left(\ominus = \frac{\mathbf{3}}{\mathbf{4}}\right) \equiv P\left(\ominus = \frac{1}{4}\middle|\{Head, Tail, Tail\}, 3\ flips\right) = \frac{\mathbf{1}}{\mathbf{4}}$$

We would based on this say that $\ominus = \frac{\mathbf{1}}{\mathbf{4}}$ is the *most likely* value for $\ominus$.

Whereas the **frequentist** approach would have been to compute

$$\boldsymbol{P}(\boldsymbol{Head}) \approx \frac{Number\ of\ heads}{Number\ of\ experiments} = \frac{\mathbf{1}}{\mathbf{3}}$$

$$\boldsymbol{P}(\boldsymbol{Tail}) \approx 1 - P(Head) = \frac{\mathbf{2}}{\mathbf{3}}$$

, without directly estimating the underlying parameter $\ominus$ that explains the system's behaviour, or incorporating our prior knowledge into the calculations.

Application to Quantum Systems

Likelihood $\longleftrightarrow$ Born rule

$P(D|\Theta)$ given by, for $D \in \{0,1\}$:

$$\boldsymbol{P(0) = a}$$

$$\boldsymbol{P(1) = b = 1 - a}$$

or condensing this into a single expression,

$P(D|\Theta) = a^{1-D}(1-a)^D$

In the coin example:

$P(Heads) = \ominus$

$P(Tails) = 1 - \ominus$

Head $\leftrightarrow$ 0 ; a $\leftrightarrow$ $\ominus$; Tail $\leftrightarrow$ 1 ; b $\leftrightarrow$ $(1 - \ominus)$

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

$$\boldsymbol{P(|0\rangle) = |\alpha|^2}$$

$$\boldsymbol{P(|1\rangle) = |\beta|^2 = 1 - |\alpha|^2}$$

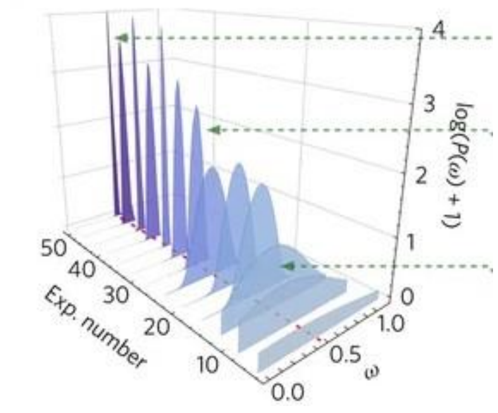
or again,

$P(|D\rangle|\Theta) = (|\alpha|^2)^{1-D}(1-|\alpha|^2)^D$

Given the *probabilistic nature* of quantum mechanics, we can use Bayesian learning to **infer the parameters of quantum systems**

Bayesian Hamiltonian Learning

- A lack of knowledge can be expressed by a flat prior over the plausible region in the parameter space
- We can then apply Bayes' rule repeatedly as we perform several experiments
- In the end we will have a probability distribution, rather than a single point estimate

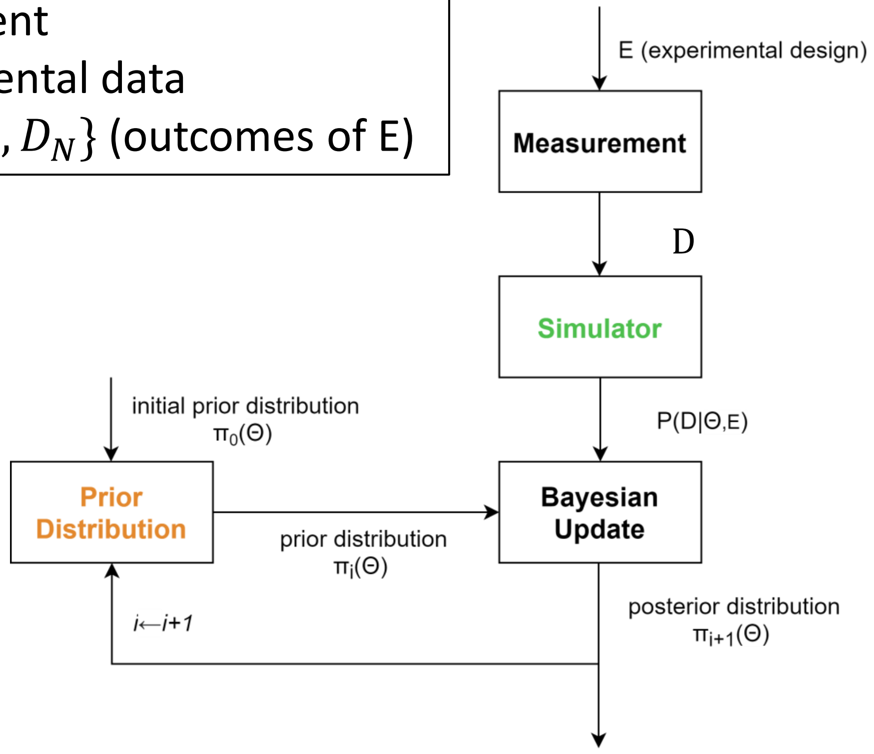


Wang, J., Paesani, S., Santagati, R. et al. (2017). Experimental quantum Hamiltonian learning. Nature Phys 13, 551–555.

Granade, Christopher & Ferrie, Chris & Wiebe, Nathan & Cory, D. (2012). Robust Online Hamiltonian Learning. New Journal of Physics. 14. 10.1088/1367-2630/14/10/103013.

Bayesian Hamiltonian Learning

E: experiment
D: experimental data
$D \in \{D_1, \ldots, D_N\}$ (outcomes of E)



$$\mathrm{D} \sim P(D|\Theta_{real}, E)$$

$$P(\Theta|D,E) = \frac{P(D|\Theta,E)P(\Theta)}{P(D|E)}$$

$$\boldsymbol{P_{i+1}(\Theta) \propto P(D|\Theta,E)P_i\,(\Theta)}$$

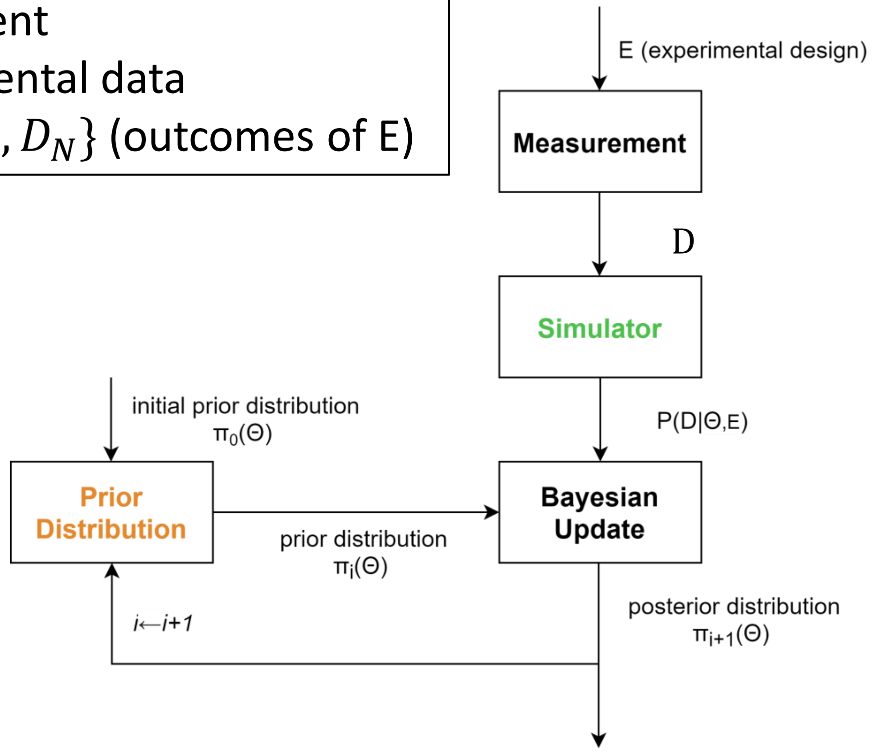
$$(apply\ sequentially)$$

Ferrie, Chris & Granade, Christopher & Cory, D. (2011). Adaptive Hamiltonian Estimation Using Bayesian Experimental Design. AIP Conference Proceedings. 1443. 10.1063/1.3703632.

Granade, Christopher & Ferrie, Chris & Wiebe, Nathan & Cory, D. (2012). Robust Online Hamiltonian Learning. New Journal of Physics. 14. 10.1088/1367-2630/14/10/103013.

Bayesian Hamiltonian Learning

E: experiment
D: experimental data
$D \in \{D_1, \ldots, D_N\}$ (outcomes of E)



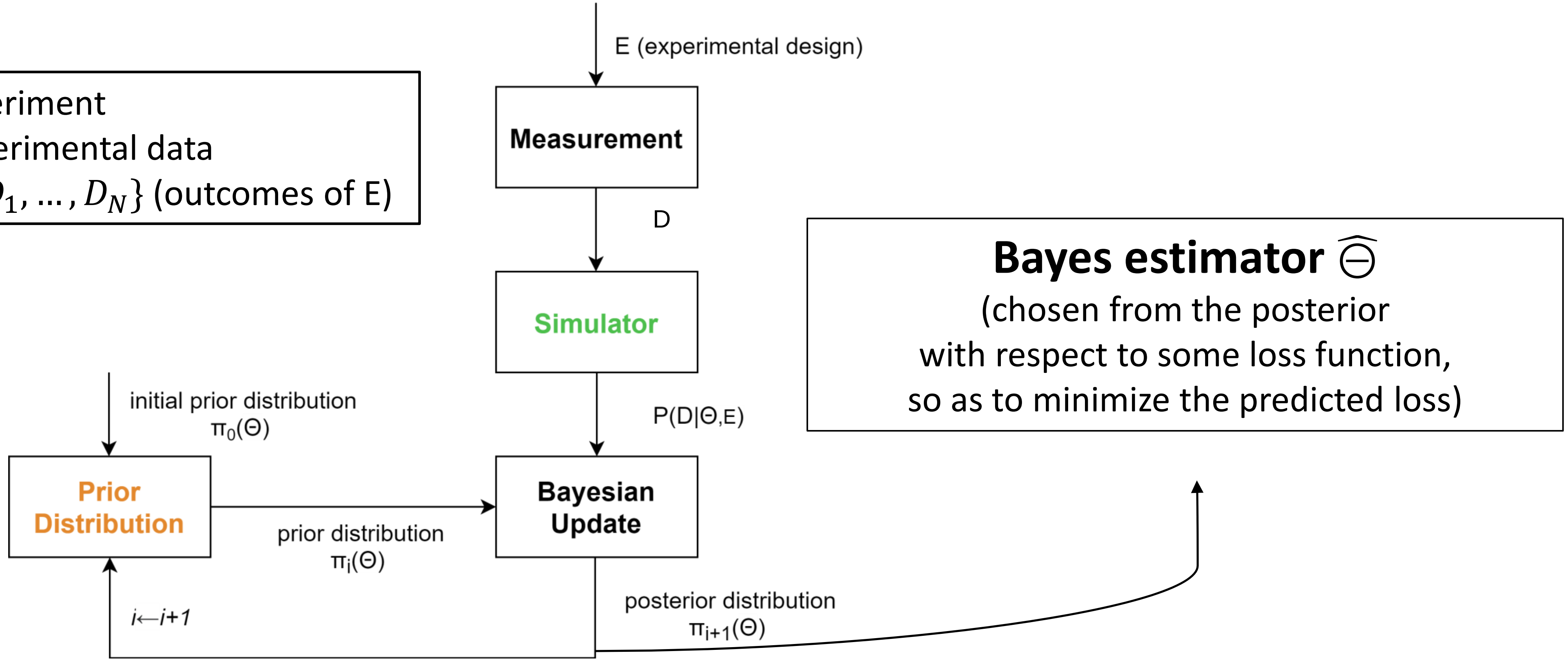
The likelihoods $P(D|\Theta, E)$ are obtained from a simulator; for complex enough systems, the simulation becomes intractable for classical computers, requiring the use of a **quantum simulator**

Wiebe, Nathan & Granade, Christopher & Ferrie, Chris & Cory, D. (2013). Hamiltonian Learning and Certification Using Quantum Resources. Physical review letters. 112. 10.1103/PhysRevLett.112.190501.

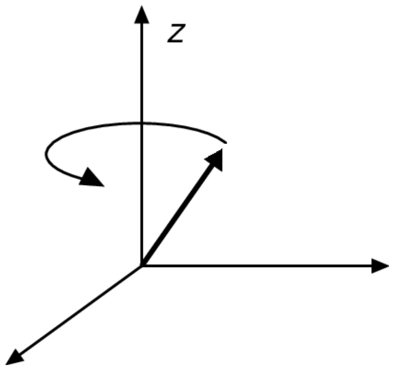
Granade, Christopher & Ferrie, Chris & Wiebe, Nathan & Cory, D. (2012). Robust Online Hamiltonian Learning. New Journal of Physics. 14. 10.1088/1367-2630/14/10/103013.

Bayesian Hamiltonian Learning

E: experiment
D: experimental data
$D \in \{D_1, \dots, D_N\}$ (outcomes of E)



Example: estimating a precession frequency



We have a single qubit evolving under an internal Hamiltonian for which we assume a fixed form:

$$H = \frac{\boldsymbol{\omega}}{2}\Omega_Z$$

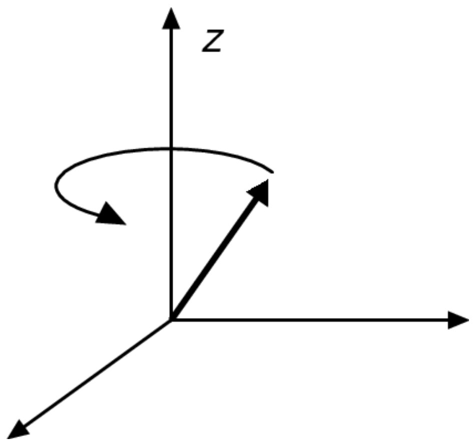
, where $\boldsymbol{\omega}$ is the parameter we mean to estimate ("Θ").

To get information about the system, we can measure the ***x*** (or y) component of the spin at a controllable time ***t*** (after initialization).

We can then **characterize** the system (parametrized by $\boldsymbol{\omega}$) using these **experimental data** along with **simulations** of the system (which provide the likelihoods for the Bayesian updates).

Ferrie, Chris & Granade, Christopher & Cory, D. (2011). Adaptive Hamiltonian Estimation Using Bayesian Experimental Design. AIP Conference Proceedings. 1443. 10.1063/1.3703632.
Granade, Christopher & Ferrie, Chris & Wiebe, Nathan & Cory, D. (2012). Robust Online Hamiltonian Learning. New Journal of Physics. 14. 10.1088/1367-2630/14/10/103013.
Wang, J., Paesani, S., Santagati, R. et al. (2017). Experimental quantum Hamiltonian learning. Nature Phys 13, 551–555.

Example: estimating a precession frequency



$$\boxed{H = \frac{\boldsymbol{\omega}}{2}\Omega_z} \;\rightarrow\; \hat{U} = e^{i\hat{H}t} = \cos(\frac{\omega t}{2}) \cdot \hat{I} + i\sin(\frac{\omega t}{2})\hat{\Omega}_z$$

With $\Psi_0 = |+\rangle$, we have the time evolution given by:

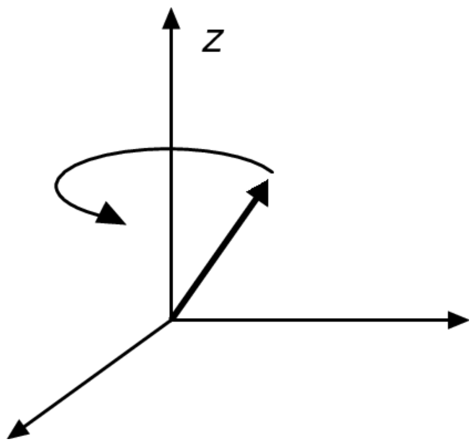
$$\Psi(t) = e^{i\hat{H}t}|+\rangle = \cos\left(\frac{\omega t}{2}\right) \cdot |+\rangle - i\sin(\frac{\omega t}{2})|-\rangle$$

So if we measure the qubit at a time $\boldsymbol{t}$ in the x-eigenbasis, we get the eigenstates $|+\rangle$ or $|-\rangle$ with likelihood:

$$\boxed{\boldsymbol{P}(|+\rangle) = \cos^2\left(\frac{\omega t}{2}\right)} \qquad \boxed{\boldsymbol{P}(|-\rangle) = \sin^2\left(\frac{\omega t}{2}\right)}$$

, given $\boldsymbol{\omega}$.

Example: estimating a precession frequency

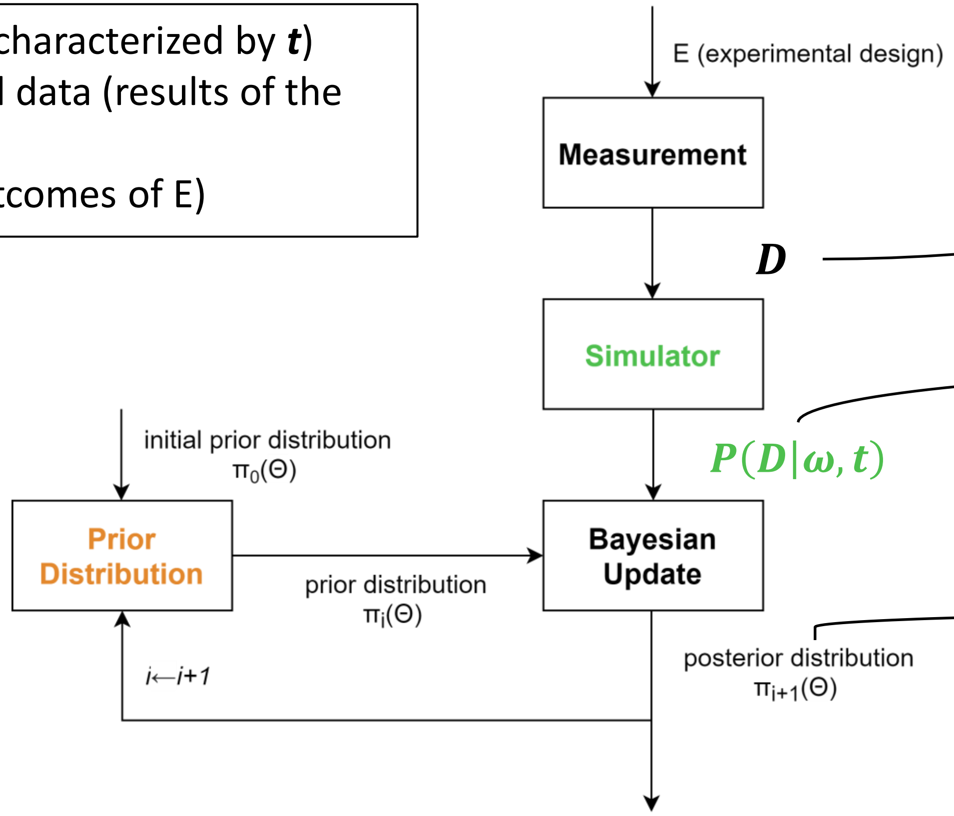


E: experiment (measuring after a controlled evolution time ***t*** after initialization)
D: experimental data (results of the measurements)
$D \in \{+, -\}$ (outcomes of E)

We determine a prior, and iteratively Bayes-update it according to the experimental data, replacing it at each step with the freshly updated distribution.
When we are satisfied with the sharpness of the distribution, we can **choose an estimator** from the final posterior and **quantify our belief** in it (e.g. through the variance)

The Parameter Estimation Algorithm

E: experiment (characterized by ***t***)
D: experimental data (results of the measurements)
$D \in \{+,-\}$ (outcomes of E)

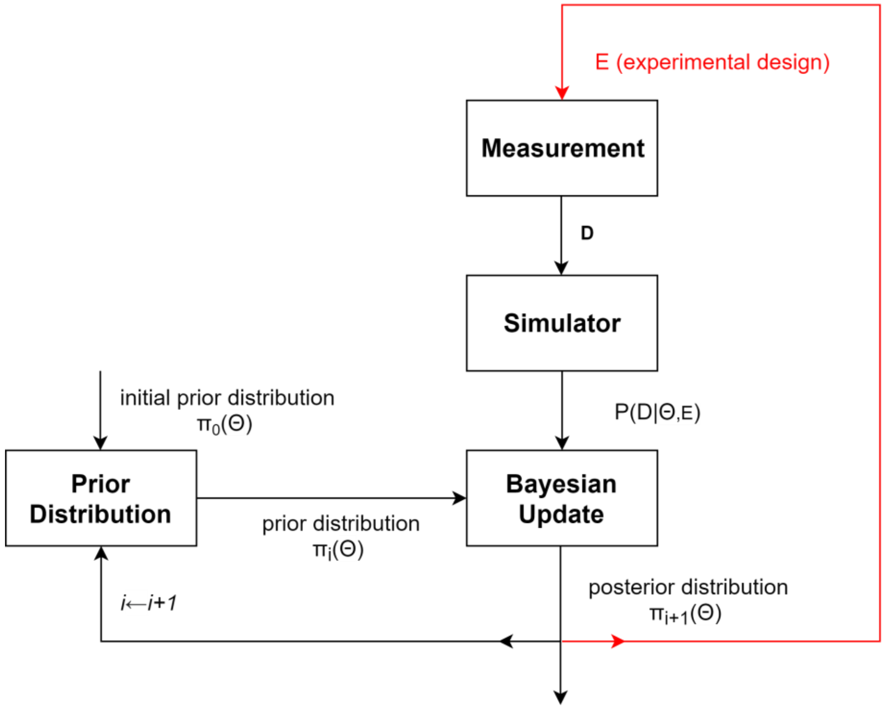


$$\mathrm{D} \sim P(D|\omega_{real}, t)$$

$$\text{e.g. } P(+|\omega, t) = \cos^2\left(\frac{\omega}{2}t\right)$$

$$\boxed{\boldsymbol{P_{i+1}(\omega) \propto P(D|\omega, t) P_i\ (\omega)}}$$

Bayesian Experimental Design



We can *adaptively* design each experiment (e.g. choose the time of the measurement) based on the results gotten so far (i.e. the current distribution).

This can be done for instance to **maximize the expectation value of a utility function** (e.g. the *negative variance*, or the *information gain*), which we can compute given the prior (and again access to a simulator)

Ferrie, Chris & Granade, Christopher & Cory, D. (2011). Adaptive Hamiltonian Estimation Using Bayesian Experimental Design. AIP Conference Proceedings. 1443. 10.1063/1.3703632.
Granade, Christopher & Ferrie, Chris & Wiebe, Nathan & Cory, D. (2012). Robust Online Hamiltonian Learning. New Journal of Physics. 14. 10.1088/1367-2630/14/10/103013.
Wang, J., Paesani, S., Santagati, R. et al. (2017). Experimental quantum Hamiltonian learning. Nature Phys 13, 551–555.

Offline vs. Adaptive Strategies for Experimental Design

Offline

- Choose the experiments in advance
- Process the data only after performing all the experiments (*offline*)

Adaptive

- Choose the experiments at each step, based on the results from the previous ones
- Process the data online, as the experiments are performed
- Locally optimize the next experiments' controls

Adaptive Hamiltonian Learning

An alternative to utility optimization is the use of heuristics, such as the following particle guess heuristic:

$$t_{i+1} \propto \frac{1}{\sigma_i}$$, with σ_i the standard deviation of the distribution $\mathrm{P}(\theta)$ at step *i*

or

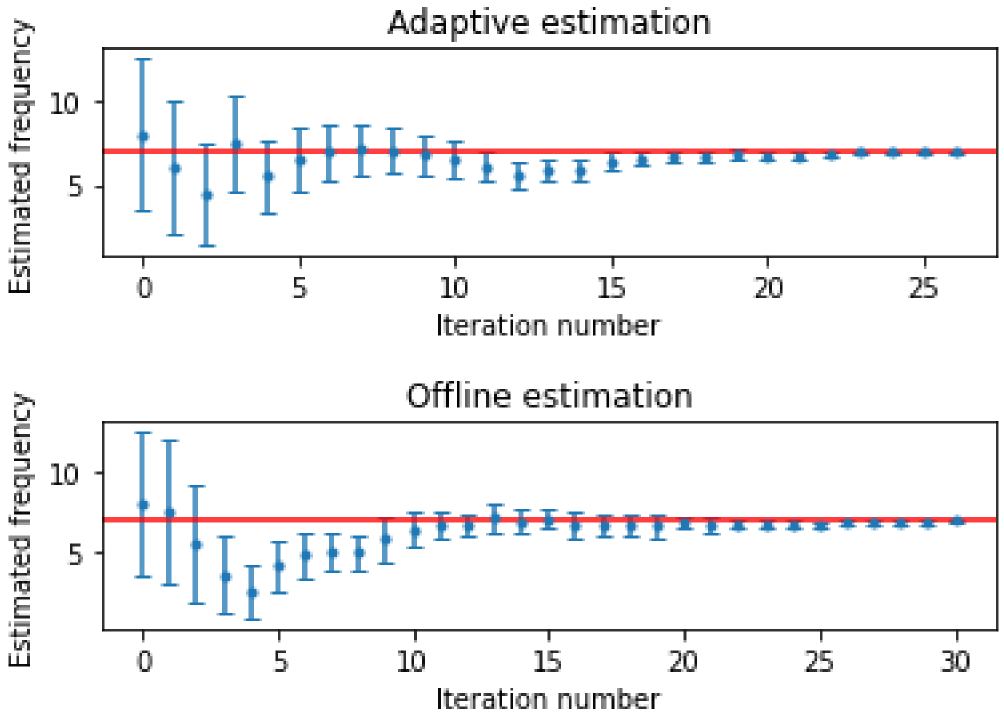
$$t_{i+1} \propto \frac{1}{|\theta_2 - \theta_1|}$$, with θ_1 , θ_2 sampled from the distribution $\mathrm{P}(\theta)$ at step *i*

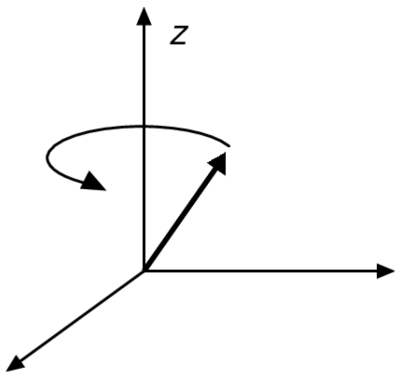
Intuitively, we want to choose *short* evolution times when we know little about the system, and *longer* ones when our distribution is already more sharply peaked, so that **the guesses continue to be informative as we learn more about the system**

Wiebe, Nathan & Granade, Christopher & Ferrie, Chris & Cory, D. (2013). Hamiltonian Learning and Certification Using Quantum Resources. Physical review letters. 112. 10.1103/PhysRevLett.112.190501.

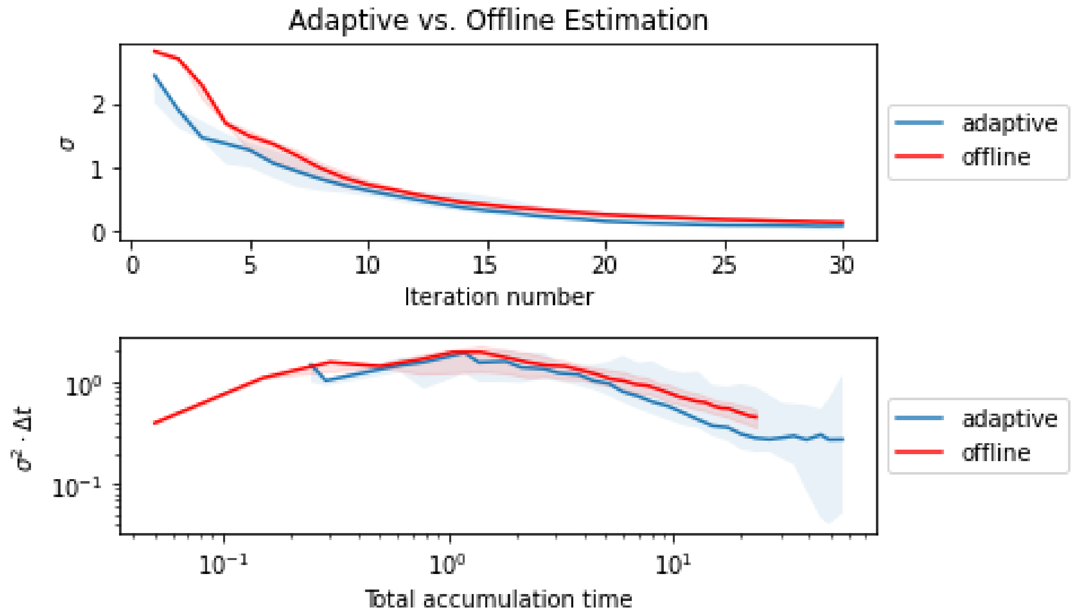
Hamiltonian Parameter Estimation

Numerical Results for the Precession Example





Hamiltonian Parameter Estimation
Numerical Results for the Precession Example



n=50; N=30; $\omega_{max} = 10;\ T_c = \infty$

Adaptive strategy:
- Standard deviation: *0.08*
- Error: *0.06*
- Final precision: *0.27*

Offline strategy :
- Standard deviation: *0.14*
- Error: *0.09*
- Final precision: *0.46*

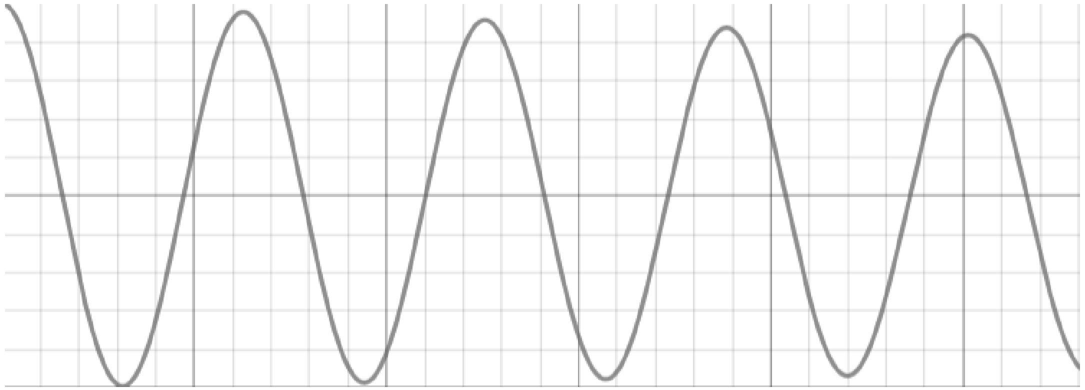
Learning the Dynamics of *Open* Quantum Systems

Open quantum systems require that we account for

the interaction with an external environment, by

adapting the protocol for a finite coherence time

Learning the Dynamics of *Open* Quantum Systems



In the precession example:

$$\boldsymbol{P}(|+\rangle) = e^{-\frac{t}{T_c}}\cos^2\left(\frac{\omega t}{2}\right) + \frac{1 - e^{-\frac{t}{T_c}}}{2}$$

$$\boldsymbol{P}(|-\rangle) = \mathbf{1} - \boldsymbol{P}(|+\rangle)$$

$$\left(= e^{-\frac{t}{T_c}}\sin^2\left(\frac{\omega t}{2}\right) + \frac{1 - e^{-\frac{t}{T_c}}}{2}\right)$$

We now have a 2 dymensional parameter space, $\vec{\theta} = (\theta_1, \theta_2) \equiv (\omega, T_C)$.

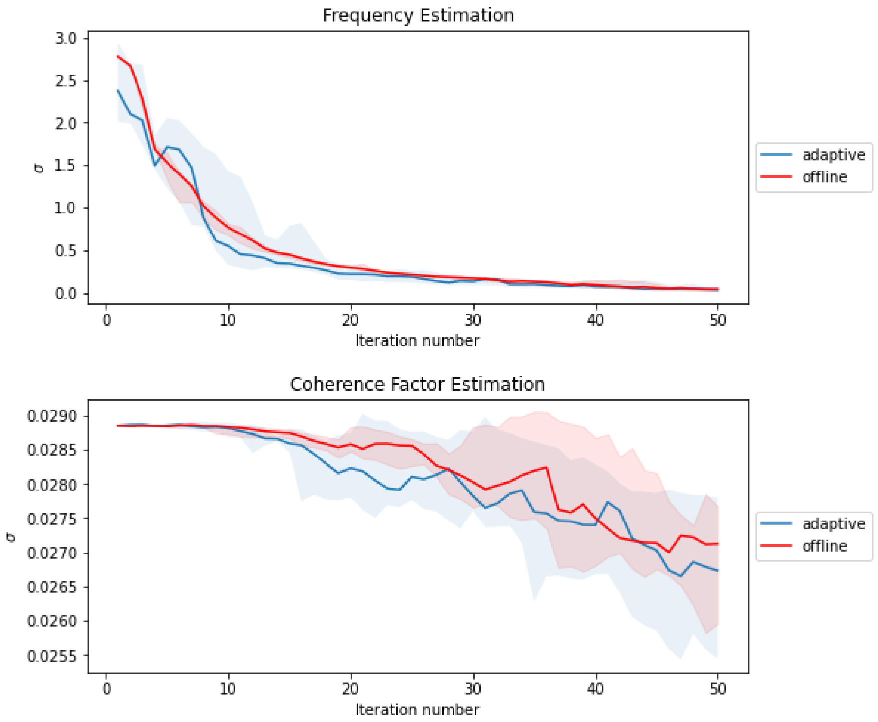
We can **estimate these parameters *simultaneously***

(or (ω, α), where $\alpha \equiv \frac{1}{T_C}$)

Granade, Christopher & Ferrie, Chris & Wiebe, Nathan & Cory, D. (2012). Robust Online Hamiltonian Learning. New Journal of Physics. 14. 10.1088/1367-2630/14/10/103013.

Hamiltonian Parameter Estimation

Numerical Results for the Precession Example in the Presence of Decoherence



n=50; N=900 (30 each); $\omega_{max} = 10$; $T_{c_{min}} = 10$

Adaptive strategy:
- Standard deviation: 0.03; 0.0267
- Error: 0.04; 0.0275

Offline strategy:
- Standard deviation: 0.04; 0.0271
- Error: 0.04; 0.0386

(Results for ω; $\alpha \equiv \frac{1}{T_C}$)

Thank you for your attention!