Lecture 17: Error mitigation
Reading: Temme, Bravyi, and Gambetta (IBM team) (2019).

1 Introduction

Fully fault-tolerant quantum computers are the device that we all wish we had. However, present devices are not yet at the error thresholds needed to implement error correction, and even if they were, we still are working out the engineering to include sufficient numbers of qubits to create a single logical qubit. For near-term devices, our calculations will thus be exposed to errors, and the effect of these errors is already obvious in estimates of observables in the papers we have seen in this class.

The term "error mitigation" refers to schemes that seek to reduce the effect of the errors on the measurement of observables rather than eliminate errors in a fault-tolerant manner. Several such schemes have been proposed [1–7]. In this lecture we will focus on one of these schemes that has been implemented experimentally and seems to perform well.

This scheme is based on extrapolation of an observable to the zero-noise limit based on measurements with various amounts of noise. In this simulations, the noise is actually increased, by means that we will see shortly, and the observables obtained with these different noise values are used to extrapolate to zero noise.

2 Error extrapolation

In more detail, our general goal in quantum simulation is to estimate an observable $\langle A \rangle$ for an evolved state $\rho_\lambda(T)$ after time T subject to noise characterized by λ. We would like the observable in the limit $\lambda \to 0$. We can cancel the lower order contributions of the noise to the observable using Richardson's deferred approach.

To see how, let's model the circuits not as gates, but rather as time-dependent Hamiltonian dynamics specified by a term $K(t)$. This term can generally be expanded as:

$$K(t) = \sum_\alpha J_\alpha(t) P_\alpha \tag{2.1}$$

where P_α is a Pauli string.

The evolution of an open system with initial state ρ_0 is given by a master equation:

$$\frac{\partial \rho}{\partial t} = -[K, \rho] + \lambda L(\rho(t)) \tag{2.2}$$

for time $t \in [0, T]$. For the scheme to work, we now have to make the following assumptions for L: (1) it is invariant under time-rescaling and (2) it is independent of the time-varying terms $J_\alpha(t)$. In words, these assumptions require that the noise is independent of the evolution of our system of qubits. Generally, such an assumption does not seem to be so bad for qubits coupled to an environment.

The goal of our simulation is to compute $E_K(\lambda) = tr(A\rho_\lambda(T)$. It can be proved that there exists an asymptotic expansion for $E_K(\lambda)$ in terms of λ. The proof sketch is to move to the interaction picture, generate an iterative series expansion for the density matrix at time T, from which the above statement follows. I will do the derivation in class, and it is in the SI of the Temme paper. The upshot is that we can write:

$$E_K(\lambda) = E^* + \sum_{k=1}^{n} a_k \lambda^k + R_{n+1} \tag{2.3}$$

where $E^* = tr(A\rho_0(T))$ is the observable without noise, a_k are expansion coefficients, and R_{n+1} is a remainder term that can be bounded.

To perform Richardson extrapolation, we need to run a quantum device at different noise rates λ_j to obtain:

$$E'_K(\lambda_j) = E_K(\lambda_j) + \delta_j \tag{2.4}$$

where $\lambda_j = c_j\lambda$ are the rescaled experimental noise rates and δ_j is a deviation term. In the Richardson extrapolation scheme, our estimate of E^* can be improved by computing:

$$\begin{aligned} E'^n_K(\lambda) &= \sum_{j=0} \gamma_j E'_K(c_j\lambda) && (2.5) \\ \sum_{j=0}^{n} \gamma_j &= 1 && (2.6) \\ \sum_{j=0}^{n} \gamma_j (c_j)^k &= 0 && (2.7) \\ & && (2.8) \end{aligned}$$

The result is an approximation to E^* accurate to $O(\lambda^{n+1})$. These equations follow from inserting the expression for $E'_K(c_j\lambda)$ as:

$$E'^n_K(\lambda) = \sum_{j=0} \gamma_j E'_K(c_j\lambda) = E^* \sum_{j=0}^{n} \gamma_j \sum_{k=1}^{n} a_k \lambda^k \left(\sum_{j=0}^{n} \gamma_j (c_j)^k \right) + ... \tag{2.9}$$

The main question now is how to obtain rescaled noise parameters? Having such a degree of control would appear to be difficult because it requires control over the environmental coupling to the qubits, which is a process that we are unable to even characterize completely.

Fortunately, it turns out that there is a trick to obtain the same effect without controlling the noise explicitly. We can run the same circuit $n+1$ times with rescaled parameters in the evolution term $K(t)$, and that has exactly the same effect as if we had rescaled the noise parameter λ. The rescaling is defined as $T \to T' = cT$, $J_\alpha(t) \to J'_\alpha(t) = c^{-1}J_\alpha(c^{-1}t)$, implying that $\rho(t) \to \rho(t') = \rho(c^{-1}t)$.

Under this mapping, the claim is that $\rho'_\lambda(T') = \rho_{c\lambda}(T)$ if L is independent of the evolution of the system and time. If true, then the rescaled reduced density matrix $E'_K(\lambda) \to E_K(c\lambda)$ will provide our observables with different noise values, which is what we need for our observable.

The above claim can be proved by considering the original master equation, substituting in the rescaling, and simplifying the expression. I will do it in class and it is in the SI of the Temme paper.

3 Protocol

The method for the scheme now becomes:

1. For $j = 0...n$, choose $c_j > 1$, taking $c_0 = 1$ as the increased noise parameter compared to the usual computation. Evolve ρ_0 with $K^j(t) = \sum_\alpha J_\alpha^j(t) P_\alpha$, where J_α^j is the rescaled J term and the final time is $T_j = c_j T$. Thus each gate applied in to the qubits is progressively stretched in time.
2. Estimate $\langle A \rangle$ to get $E'_K(c_j \lambda)$.
3. Solve the equations for γ and obtain $E'^n_K(\lambda)$ as an improved estimator for E^*.

This scheme was originally proposed [3] and later implemented [7] in superconducting qubits. For the hardware implementation, the length of the microwave pulses implementing the gates were systematically increased, taking consideration of pulse length, rise/fall time, and buffer times between pulses. The IBM device uses fixed frequency transmon qubits with coplanar waveguide resonators, frequencies around 5 GHz, with dispersive readout. T_1 and T_2 were around 40-70 μs and the readout error was < 0.05 for a 2 μs integration time. The shortest duration of the microwave pulse was 83.3 ns with a buffer of 6.7 ns which was stretched by $c_j > 1$.

References

[1] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, "Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states", Physical Review A **95**, 042308 (2017).

[2] S. McArdle, X. Yuan, and S. Benjamin, "Error mitigated digital quantum simulation", arXiv (2018).

[3] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits.", Physical Review Letters **119**, 180509 (2017).

[4] S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications", Physical Review X **8**, 031027 (2018).

[5] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O'Brien, and L. DiCarlo, "Error mitigation by symmetry verification on a variational quantum eigensolver", arXiv (2019).

[6] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, "Decoding quantum errors with subspace expansions", arXiv (2019).

[7] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor.", Nature **567**, 491–495 (2019).