Extrapolation

Eesh Gupta

June 17, 2020

Contents

Extrapolation is a technique in which we deliberately make noise rate worse in order to get a more accurate result.

1 Background

1.1 Richardson Extrapolation

Suppose you have approximated $A^\star$ as $A(h)$ (assume h is the error rate) but you want to make a better approximation. Fortunately, you are also given a series

$$A^\star = A(h) + a_0 h^{k_0} + a_1 h^{k_1} + a_2 h^{k_2} + \dots \tag{1}$$

where a's are unknown constants and k's are known constants such that $h^{k_i} > h^{k_{i+1}}$. That is, the sequence on RHS of equation 1 is decreasing. If we write the same equation in terms of Big Oh notation, we have

$$A^\star = A(h) + O(h^{k_0}) \tag{2}$$

The current error is on the order of $O(h^{k_0})$. Our need is to get a better approximation of $A^\star$, let's say with a higher order error $O(h^{k_1})$ estimate. What do we do?

1. *Big Oh Notation* Rewrite the equation 1 as

$$A^\star = A(h) + a_0 h^{k_0} + O(h^{k_1}) \tag{3}$$

   In the following steps, our objective is to get rid of the term $a_0 h^{k_0}$ so we end up with an improved error bound of $O(h^{k_1})$.

2. *Rescaling the Parameter* h as $\frac{h}{t}$ where t is a constant (hopefully $t \neq 1$!).

$$A^\star = A(\frac{h}{t}) + a_0(\frac{h}{t})^{k_0} + O(h^{k_1}) \tag{4}$$

3. *Ridding of a term from equation 1:* Multiply equation 4 by t^{k_0}, subtract equation 3 from the modified 4 and divide both sides of the resulting equation by $t^{k_0} - 1$. Then we have

$$A^\star = \frac{t^{k_0} A(\frac{h}{t}) - A(h)}{t^{k_0} - 1} + O(h^{k_1}) \tag{5}$$

   We can repeat this process many times for different error rates to improve the aproximation further.

2 Extrapolating Noise via Rescaling Time

2.1 Series Expansion in Noise Parameter

Our starting point is the following eqution

$$\frac{\partial \rho}{\partial t} = -i[K(t), \rho] + \lambda \mathcal{L}(\rho) \tag{6}$$

where the first term on RHS corresponds the time dependent qubit hamiltonian (set of pauli gates acting on qubits with time dependent coupling coefficients) and the second term accounts for noise. Here, λ is the streingth of noise, assumed weak because of short depth circuits. While using a lindbladian superoperator may imply that we are considering only markovian noise, we can also account for non markovian noise by taking $\mathcal{L}(\rho)$ as

$$\lambda \mathcal{L}(\rho) = -i[V, \rho] \tag{7}$$

where V is some hamiltonian.

Now, we can then remove the time evolution due to qubit hamiltonian $K(t)$ to zoom in on the noisy evolution in the interation picture of $K(t)$. This involves "heisenburg-ing" our noisy superoperator $\mathcal{L}$ and "shrodinger-ing" our density matrix ρ with respect to $K(t)$ to get the following partial differential equation:

$$\frac{\partial}{\partial t} \rho_I(t) = +\lambda \mathcal{L}_{I,t}(\rho_I(t)) \tag{8}$$

From here on, we use some tricks relating to perturbative expansion and go back to the schrodinger picture. After this process, we obtain the expectation value of some observable A as

$$E_K(\lambda) = E^\star + \sum_{k=1}^{n} a_k \lambda^k + R_{n+1}(\lambda, \mathcal{L}, T) \tag{9}$$

Here, $E_K(\lambda)$ is the expectation value of A given noise rate of λ, E^* is the noiseless expectation value of A, a_k hides huge integrals involving the noisy superoperator $\mathcal{L}$, R is the "remainder" of the infinite series and T is the time for which noisy evolution lasts.

Fortunately, R is bounded so the series is decreasing. Hence, we can apply richarson extrapolati to improve upon our approximation of noiseless expectation value $E^\star$.

2.2 Experimental Rescaling of the noise parameter

The chief concern in this section is how to control the noise parameters λ in an experimental setting. One possible solution is to rescale the time T for which the noisy evolution lasts. This may result in us either "waiting longer" to measure the results, "slowing down" the gate operations or a combination of both. Using some clever substitution techniques inside integrals, we can show that rescaling time T results in a rescaled noise rate λ.

Note that in this case, noise is assumed constant in time. This is not always true. For example, consider the figure below:

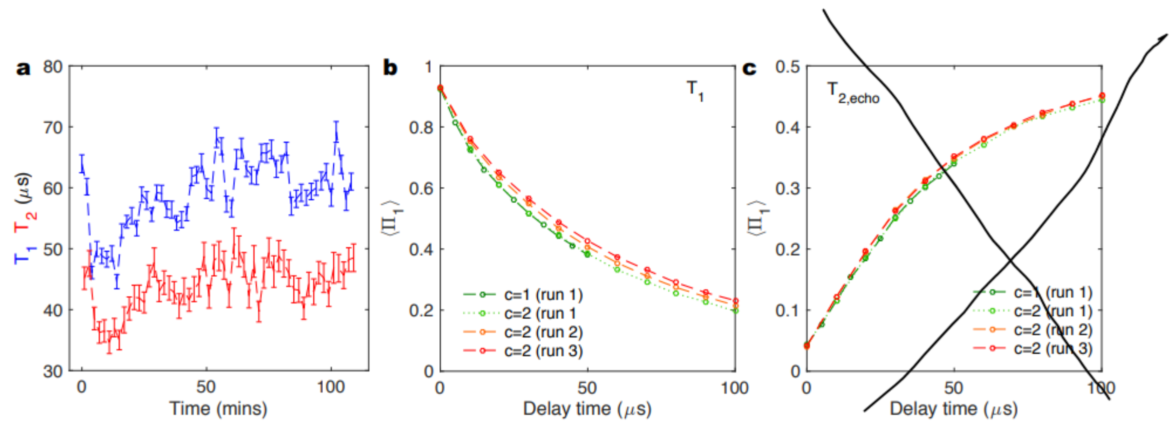

FIG. S2. **Coherence fluctuations in superconducting qubits a** Repeated measurements of T_1 and T_2 show fluctuations that are larger than the error bars to the fits. Decay of excited state projector $\langle \Pi_1 \rangle$ for a standard **b** T_1 sequence and **c** T_2 echo sequence, for stretch factors c_i=1,2 obtained using 10^5 samples. In both cases, the average decay over the total sampling duration for the different stretch factors is the same, when their measurements are grouped together (run 1), in contrast to when sampled separately (run 2, run 3).

Figure 1: Figure from https://arxiv.org/pdf/1805.04492.pdf

Figure 1 shows decoherence times T_1, T_2 on a 5 qubit superconducting processor at IBM. T_1 is a decay constant which measures how long it takes for a qubit in excited state $|1\rangle$ to relax to state $|0\rangle$. On the other hand, T_2 measures the time it takes for a qubit in state $|+\rangle = \frac{|0\rangle+|1\rangle}{2}$ to turn into mixture of $|+\rangle$ and $|-\rangle$. Simply put, T_1 is the relaxation time constant and T_2 is the dephasing time constant.

In 1 a, we see how the decoherence times fluctuate over 2 hours. These fluctuations shows that the noise rate is not constant in time. Futher 1 b (which I do not understand completely) tells us that we need to perform "rescaled" experiments shortly after one another. If we wait too long between experiments, our results would be affected and extrapolated results would be less accurate.

2.3 Error bounds on the noise free parameter

Once we can rescale our noise rate λ, we are ready to perform Richardson Extrapolation. In our earlier discussion on this trick, we rescaled our noise just once to improve our error bound from $O(h^{k_0})$ to $O(h^{k_1})$. However, in general, we can perform multiple rescalings and improve our error bounds by multiple orders with only a few equations. So in this scenario, we will scale our noise rate n times and hence cancel n terms from RHS of equation 9. To do so, we need to combine expectation values with distinct noise rates as

$$E_K^n(\lambda) = \sum_{j=0}^{n} \gamma_j \hat{E}_K(c_j\lambda) \tag{10}$$

where γ_j is the weight of the expectation value with noise rate scaled as $c_j\lambda$. To do n steps at once, we have to abide by 2 constraints:

1. $E^\star$ has to end up normalized at the end. So

$$\sum_{j=0}^{n} \gamma_j = 1$$

2. We need to cancel out n terms from RHS of equation 9 to improve our aprroximation. So

$$\sum_{j=0}^{n} \gamma_j c_j^k = 0 \text{ for } k = 1 \ldots n$$

($\ast\ast\ast$ It's not clear how such constraints result in the equations presented in the above text. But I am just writing out the function of those equations without showing all the math. That can be found however in equation 36 of the Temme paper)

This results in

$$E^\star = E_K^n(\lambda) + O(\lambda^{n+1})$$

Note that strecth factor $c_j > 1$ for $j = 1, \ldots, n$ i.e. we are scaling up the noise rate. In the numerical literature, it is common to scale down the noise rate. However, the result is just the same so this point is not so important.

3 Extrapolating Noise via Pauli Gates

3.1 Stochastic vs Non stochastic noise

In the pauli error model, errors are *stochastic* if the noise superoperator $\mathcal{N}$ can be written in the form

$$\mathcal{N} = (1-\epsilon)\mathcal{I} + \epsilon\mathcal{E} \tag{11}$$

where $\mathcal{I}$ is the identity operation and $\mathcal{E}$ is some error that occurs with probabilty ϵ. Note that $\mathcal{E}$ is a valid quantum operation i.e. it is TPCP (Trace-preserving-completely-positive) map.

Then, stochastic or "probabilitic" error models are those in which the system remains unchanged with some probability $1-\epsilon$ and system changes due to error with probability ϵ. Then what does *non-stochastic* error mean? A non-stochastic error is simply one which cannot be expressed in such a form as 11. They are also known as coherent noise.

3.2 Reformulating Extrapolation

3.3 Pauli Twirling

3.4 Noise Amplification

Let $_{i,j}$ be the the 2 qubit error rate between qubits q_i and q_j. Because of Pauli Twirling, there exists another interpretation of $_{i,j}$: the probability of randomly applying 2 qubit pauli gates on qubits q_i, q_j. Then to increase this error rate by some factor let's say r, we would need to apply 2 pauli operations σ_c^e, σ_t^f randomly with probability $(r-1)_{i,j}$. Why so? Because this gives us a new error rate of $_new =_{i,j} +(r-1)_{i,j}$. Here, the first term on RHS is due to the error rate of the controlled phase gate (or the CNOT gate). And the second term on RHS is due to the additional pauli gates σ_c^e, σ_t^f.

4 Extrapolation equations

4.1 First Order

$$\gamma_1 + \gamma_2 = 1$$

$$\gamma_1 + 2\gamma_2 = 0$$

Solution to the system:

$$\gamma_1 = 2$$

$$\gamma_2 = -1$$

Combining these solutions to get result:

$$\hat{E}^1(\lambda) = 2\hat{E}(\lambda) - \hat{E}(2\lambda)$$

4.2 Second Order

$$\gamma_1 + \gamma_2 + \gamma_3 = 1$$

$$\gamma_1 + 2\gamma_2 + 3\gamma_3 = 0$$

$$\gamma_1 + 4\gamma_2 + 9\gamma_3 = 0$$

Solution to the system:

$$\gamma_1 = 3$$

$$\gamma_2 = -3$$

$$\gamma_3 = 1$$

Combining these solutions to get result:

$$\hat{E}^2(\lambda) = 3\hat{E}(\lambda) - 3\hat{E}(2\lambda) + \hat{E}(3\lambda)$$