Extrapolation

Eesh Gupta

July 6, 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:

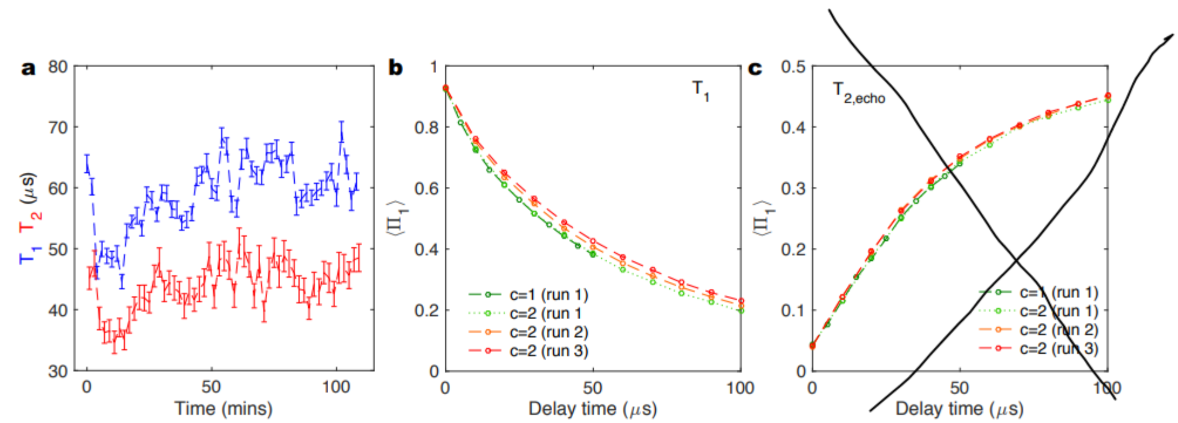

FIG. S2. **Coherence fluctuations in superconducting qubits a** Repeated measurements of T_1 and T_2 show fluctuations that are larger than the error bars to the fits. Decay of excited state projector $\langle \Pi_1 \rangle$ for a standard **b** T_1 sequence and **c** T_2 echo sequence, for stretch factors 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

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

($***$ 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 $j = 1, \dots, n$ 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. These errors are also termed as coherent noise.

3.2 Origins of Pauli Error Simulation

Single qubit gate errors are calculated by measuring the average gate fidelity i.e. how well a noisy gate $\mathcal{N}U$ preserves quantum information (Here, $\mathcal{N}$ is some noisy superoperator and U is the ideal unitary operation without errors). These error rates are pre-computed by Qiskit through a process called quantum gate set tomography. So after acting on a qubit with the gate $\mathcal{N}U$ of error value ϵ, then

- With probability ϵ, we will end up losing all information about the qubit. In other words, the resulting state of the qubit will be the completely mixed state $\frac{\mathcal{I}}{2}$.
- With probability $1-\epsilon$, we will end up with the information of the qubit intact or preserved.

Suppose the state of the qubit after the 'ideal' gate operation U is ρ. Then Qiskit is telling us

$$\widetilde{\mathcal{N}}(\rho) = (1-\epsilon)\rho + \epsilon\frac{\mathcal{I}}{2} \tag{12}$$

Here $\widetilde{\mathcal{N}}$ is an approximation of the true noise channel $\mathcal{N}$. Again, $\mathcal{N}$ may or may not be stochastic but $\widetilde{\mathcal{N}}$ looks very much stochastic. This 'approximate' noise channel is also known as the depolarizing noise channel. In fact, it has another interpretation in terms of pauli gates. To get there, we must first recognize that

$$\frac{I}{2} = \frac{I\rho I + X\rho X + Y\rho Y + Z\rho Z}{4} \tag{13}$$

for arbitrary ρ. This can be verified using the bloch vector interpreation of a qubit (i.e. $\rho = \frac{I+\vec{a}\cdot\vec{\sigma}}{2}$). Given such equation then, we can re-interpret equation 12 as

- With probability ϵ, we will act on the qubit with a random pauli gate (after ideal gate operation U).
- With probability $1-\epsilon$, we will do nothing, preserving the state of the qubit as ρ.

Note that the set of pauli gates includes identity and acting this identity on the state p will also preserve the state of the qubit. So the state of qubit will then be preserved with a probability more than $1-\epsilon$. But I chose to ignore this point so the explanation is a bit simpler.

Such reformulation makes amplifying noise very simple: tuning up the probability of adding random pauli gates. For example, suppose the noise rate of some gate $\widetilde{\mathcal{N}}U$ is given to us as λ and we want to rescale the error rate to be $c\lambda$ where c is a real number. To do so, after applying the noisy gate $\widetilde{\mathcal{N}}U$, we will add a random pauli gate σ with probability $(c-1)\lambda$. This would tune up the total error rate to be $c\lambda$.

$$\lambda + (c-1)\lambda = c\lambda$$

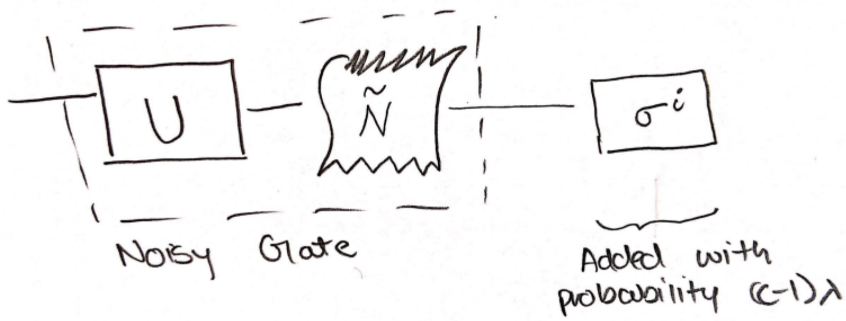

Figure 2: Simplified noise amplification procedure. In reality, you would have to twirl the noisy gate first which is explained in the next section.

3.3 Pauli Twirling

Where does twirling come in? Twirling is a technique that converts any noise channel $\mathcal{N}$ to a pauli channel i.e.

$$\mathcal{N} \longrightarrow \sum_i p_i[\sigma_i]$$

Here p_i is the probability of pauli error σ_i happening (or applying a random pauli gate σ_i).

> **Notation Check:** If U is some unitary operation, then $[U]$ is its corresponding superoperator. Why is this necessary? Suppose U acts on a quantum system whose initial state is given by some density matrix ρ. Then the evolved state will be $U\rho U^\dagger$. This notation tends to be slightly incovenient so we make
>
> $$[U]\rho = U\rho U^\dagger.$$

Why *do we need to twirl the noisy gate? Can't we just amplify the noise by adding random pauli gates and leave the noisy gate $\mathcal{N}U$ alone?* The problem lies in us amplifying a particular type of quantum error: pauli error. If $\mathcal{N}$ is not a pauli error channel, then adding pauli errors on top of it will not help in zero error extrapolation. Why so? Consider the following equation:

$$\hat{E}(\lambda) = E^\star + a_1\lambda + a_2\lambda^2 + ... + a_n\lambda^n + ... \tag{14}$$

which comes directly from equation 9. This represents the situation where the original noise $\mathcal{N}$ is not tampered with (not scaled up or down). If we then double the noise by adding in pauli errors, the expansion of $\hat{E}(2\lambda)$ would have different a_i's, since a_i's encode the noise operator and we have changed the noise operator by adding in a different type of noise. Then the extrapolation scheme discussed in section 2.3 will not work. The only way in which we can ensure that a_i's are constants when we rescale noise is by making sure $\mathcal{N}$ is a pauli error channel. To do so, we need to use twirling.

***How** is twirling done?*

Suppose we are twirling a CNOT gate U on qubits q_c, q_t ('c' for control and 't' for target). Since the gate will have some noise, our noise gate becomes $\mathcal{N}U$. Let $\sigma_c^a\sigma_t^b$ and $\sigma_c^c\sigma_t^d$ be 2 qubit pauli gates such that $\sigma_c^c\sigma_t^d = U\sigma_c^a\sigma_t^bU^\dagger$. Now $\sigma_c^a\sigma_t^b$ acts before the noisy gate $\mathcal{N}U$ and $\sigma_c^c\sigma_t^d$ acts after it. Then the effective gate $\mathcal{P}_U$ can be expressed as :

$$\mathcal{P}_U = \sigma_c^c\sigma_t^d\mathcal{N}U\sigma_c^a\sigma_t^b \tag{15}$$

$$\mathcal{P}_U = U\sigma_c^a\sigma_t^bU^\dagger\mathcal{N}U\sigma_c^a\sigma_t^b \tag{16}$$

If there was no noise originally i.e. $\mathcal{N} = I$, then the twirled CNOT gate $\mathcal{P}_U$ would just be U which is expected. The claim now is that $\mathcal{P}_U$ can be expressed as U followed by 2 qubit pauli error.

$$\mathcal{P}_U = \epsilon_{0,0}\sigma_c^0\sigma_t^0U + \sum_{(i,j)\neq(0,0)}^{(3,3)} \epsilon_{i,j}\sigma_c^i\sigma_t^jU \tag{17}$$

Here $\sigma^0, \sigma^1, \sigma^2, \sigma^3$ correspond to I, X, Y, Z respectively and $\epsilon_{i,j}$ is the probability with which pauli gate $\sigma_c^i\sigma_t^j$ acts on the qubit after ideal gate operation U. And thus we have converted the original noisy gate $\mathcal{N}U$ into a pauli-infected noisy gate $\mathcal{P}_U$. To make notation in the above discussion a little nicer, let

$$\bar{\mathcal{N}} = \epsilon_{0,0}[\sigma_c^0\sigma_t^0] + \sum_{(i,j)\neq(0,0)}^{(3,3)} \epsilon_{i,j}[\sigma_c^i\sigma_t^j.] \tag{18}$$

Then the twirled gate $\mathcal{P}_U$ is really $\bar{\mathcal{N}}U$. So, throught twirling, we have just changed the form of noise affecting the ideal gate U.

$$\mathcal{N} \xrightarrow{\text{Pauli Twirling}} \bar{\mathcal{N}}$$

For more detailed analysis, see (see Li and Benjamin 2017, VII A).

3.4 Noise Amplification

Since amplifying noise in every single gate in the circuit will be too cumbersome, we restict our attention to 2 qubit gates. This is a good assumption because generally, error rates of single qubit gates are smaller than those of 2 qubit gates by 2 orders of magnitude.

Now, continuing on from our last example, to amplify the noise of the twirled CNOT gate $\mathcal{P}_U$, we need to first find the 2 qubit gate error for associated qubits q_c, q_t. For IBMQ devices, this data can be imported from the Qiskit Noise Model

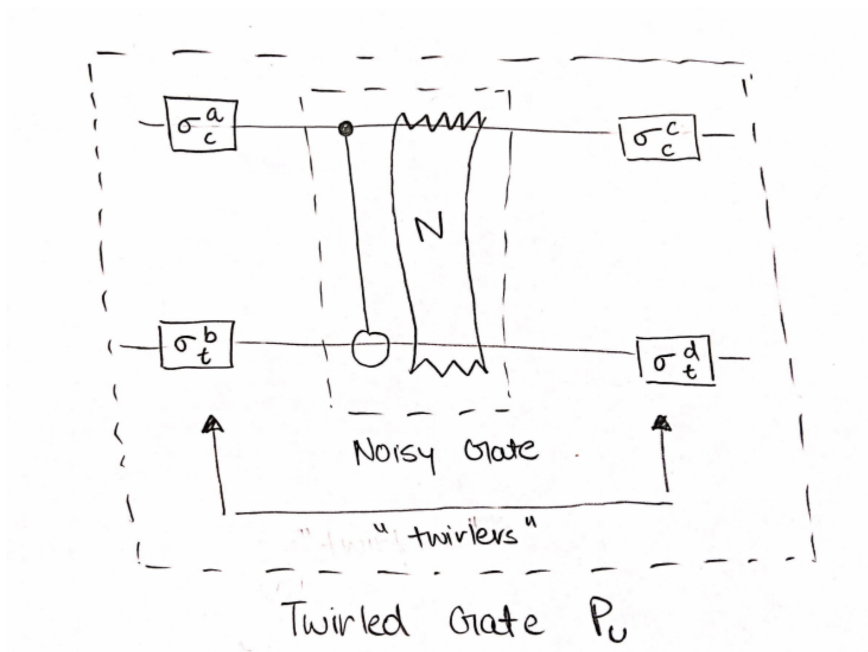

Figure 3: Twirling the noisy CNOT gate

library. Once we have the error rate λ, we will choose a random 2 qubit pauli gate from the set $\{\sigma_c^0\sigma_t^1, \sigma_c^0\sigma_t^2, \ldots, \sigma_c^3\sigma_t^3\}$ (ignore $\sigma_c^0\sigma_t^0 = I_c I_t$) and then apply that gate with probability $(c-1)\lambda$. This would amplify the noise by a factor of c.

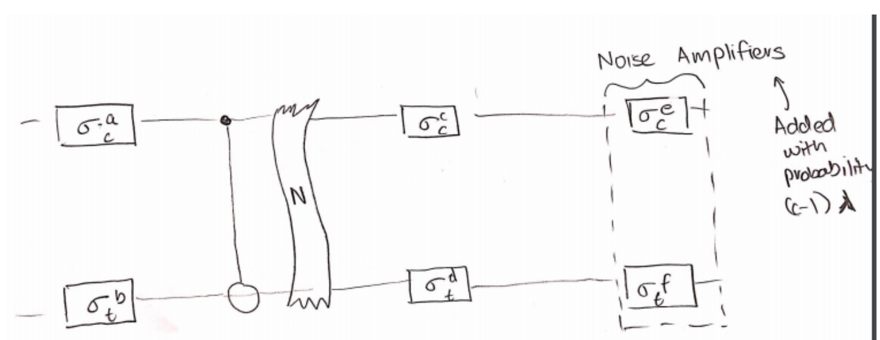

Figure 4: Amplifying noise of a CNOT gate

Note that there might be multiple 2 qubit gates with different error values λ_i. The amplification procedure is however the same for all these gates.

4 Random Corner

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:

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

4.3 Random Equations

$$\hat{E}(\lambda) = E^\star + a_1\lambda + a_2\lambda^2 + ... + a_n\lambda^n + ...$$
$$\hat{E}(2\lambda) = E^\star + 2a_1\lambda + 4a_2\lambda^2 + ... + 2^n a_n\lambda^n + ...$$
$$\hat{E}^1 = 2\hat{E}(\lambda) - \hat{E}(2\lambda) = E^\star + -2a_2\lambda^2 - 6a_3\lambda^3 + ...$$

Put another way,

$$E^\star = \hat{E}^1 + 2a_2\lambda^2 + 6a_3\lambda^3 + ...$$

Before:

$$E^\star = \hat{E}(\lambda) + O(\lambda)$$

After:

$$E^\star = \hat{E}^1 + O(\lambda^2)$$

4.4 Random Random Random

$$T_1\Phi_{\text{HF}}$$

$$T_2\Phi_{\text{HF}}$$

4.5 Reformulating Series Expansion

In section 2.1, we were concerned with obtaining a "power-series-like"expansion of expectation value of some observable as a function of some noise rate λ. In this section, we want to get a similar expansion but we will assume that all noise is stochastic. This means that if U is some ideal unitary operator, then its noisy counterpart is

$$\mathcal{N}U = ((1 - r\epsilon)\mathcal{I} + r\epsilon\mathcal{E})U, \tag{19}$$

which we get from equation 11. Here we have assumed that it is possible to rescale the error probability ϵ by some factor r. Suppose there are L such operations that act on some system. Then if the initial state of the system is ρ_i, then the final state ρ_f will be expressed as

$$\rho_f = \mathcal{N}_L U_L \dots \mathcal{N}_l U_l \dots \mathcal{N}_1 U_1 \rho_i \tag{20}$$

Now, the expectation value of some observable X with respect to ρ_f is

$$\langle X \rangle = Tr(X\rho_f).$$

Using equations 19 and 20, we will obtain

$$\langle X \rangle (r) = (1 - r\sum_l \epsilon_l) \langle X \rangle^{(0)} + r \langle X \rangle^{(1)} + r^2 \langle X \rangle^{(2)} + O(r^3) \tag{21}$$

Properly explaining this equation will require some "algebra clutter" and I will leave that to (see Li and Benjamin 2017, VII A). I will give a summary of sorts here:

- Each $\langle X \rangle^{(n)}$ describes the expectation value when n out of L operations turn out to be noisy ones i.e. $\mathcal{E}_l U_l$. Note that there are a lot of ways to choose n such operations so $\langle X \rangle^{(n)}$ will involve some summation over all combinations.
- Then $\langle X \rangle^{(0)}$ is the expectation value when all operations are ideal. This will the expectation value we hope to extrapolate!
- Another key point is that ϵ_l is error probability of some operation U_l being noisy and is assumed small. So $1 - \epsilon_l \approx 1$ and some terms with coefficients $\epsilon_1\epsilon_2 \dots \epsilon_n$ are ignored.
- The more the noisy operations, the less likely they are. This is a direct result of the coefficients of the form $\epsilon_1\epsilon_2 \dots \epsilon_n$ being really small.

- A convenient assumption made here is that the probability of each noisy operation occuring can be scaled by the same factor r. So $\langle X \rangle$ is then a function of r.

To sum up, assuming that all noise is stochastic, we have derived an equation (21) which expresses expectation value of some observable X as a function of some scaling parameter r.

References

Li, Ying and Simon C. Benjamin (June 2017). "Efficient Variational Quantum Simulator Incorporating Active Error Minimization". In: *Physical Review X* 7.2. ISSN: 2160-3308. DOI: `10.1103/physrevx.7.021050`. URL: `http://dx.doi.org/10.1103/PhysRevX.7.021050`.