Zero-noise extrapolation (ZNE)

David D.W. Ren

March 2024

1 The QOSF screening task challenge

Build a simple ZNE function.

1. Build a simple noise model with depolarising noise
2. Create different circuits to test your noise model. Choose an observable to measure.
3. Apply the unitary folding method
4. Apply an extrapolation method to get the zero-noise limit. Try linear, polynomial, and exponential
5. Compare your noise mitigated and unmitigated results
6. Run your ZNE function on real quantum hardware

2 Density operator

It is possible to describe quantum computing in terms of state vectors. A mathematically equivalent description comes from using the *density operator* or density matrix but is more convenient [3], especially for describing subsystems of a composite quantum system and for describing quantum systems whose state is not completely known. Due to their mathematical equivalence and due to fact that the same quantum mechanics can be represented using state vectors or the density matrix, one should select the representation that best solves the problem of interest.

Suppose a quantum system is one of a set of states $|\psi_i\rangle$ with probability p_i. Then, we can label $p_i, |\psi_i\rangle$ an ensemble of **pure states**. A pure state is a quantum state is exactly known i.e., $\rho = |\psi\rangle\langle\psi|$. We will discuss more about pure and mixed states later. For example, the quantum system could be $|\psi_1\rangle, |\psi_2\rangle, \dots$ with probability $p_1, p_2, \dots$. The density operator for the system is defined by,

$$\rho \equiv \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{1}$$

An explicit example of this superposition of outer products with $|\psi_1\rangle = \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}$ and $|\psi_2\rangle = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$ is,

$$\rho = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2| = p_1 \begin{pmatrix}\alpha_1^2 & \alpha_1\alpha_2^* \\ \alpha_2\alpha_1^* & \alpha_2^2\end{pmatrix} + p_2 \begin{pmatrix}\beta_1^2 & \beta_1\beta_2^* \\ \beta_2\beta_1^* & \beta_2^2\end{pmatrix} \tag{2}$$

where the example demonstrates why the density operator is interchangeably named the density *matrix*. As an example of how the state vector and density operator representations are equivalent, suppose that a closed (i.e., noise-free) quantum system initially at $|\psi_i\rangle$ with probability p_i is acted upon by a unitary operator U. The final system will be in state $U|\psi_i\rangle$ with probability p_i. This final state can also be described by the density operator as,

$$\sum_i p_i U|\psi_i\rangle\langle\psi_i|U^\dagger = U\rho U^\dagger, \tag{3}$$

that is obtained using the transformation $|\psi_i\rangle \to U|\psi_i\rangle$.

The probability of measuring an outcome m using the measurement operator M_m given a quantum state $|\psi_i\rangle$ is,

$$p(m|i) = \langle\psi_i|M_m^\dagger M_m|\psi_i\rangle = \mathrm{tr}(M_m^\dagger M_m|\psi_i\rangle\langle\psi_i|), \tag{4}$$

where the first equality comes from the third postulate of quantum mechanics [3] and the second equality comes from a Gram-Schmidt orthonormalisation procedure. The probability of measuring outcome m across all the possible $|\psi_i\rangle$ states is,

$$p(m) = \sum_i p(m|i)p_i = \sum_i p_i \mathrm{tr}(M_m^\dagger M_m|\psi_i\rangle\langle\psi_i|) = \mathrm{tr}(M_m^\dagger M_m\rho) \tag{5}$$

2.1 Pure states and mixed states

Below are equivalent statements that define a pure state

- A pure state is one where the quantum system's state is known exactly i.e., the quantum system is unaffected by noise
- A pure state has a density operator $\rho = |\psi_i\rangle\langle\psi_i|$
- A pure state satisfies $\mathrm{tr}(\rho^2) = 1$

It is probably easiest to think about pure states as when a quantum system is unaffected by noise.

For example, $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|-\rangle$ are all examples of pure states. Any state vector lying on the surface of the Bloch sphere is a pure state. If the quantum state can be written as $|\psi\rangle = |+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$, then the probability of the qubit being at state $|+\rangle$ is 100% and the probability of the qubit being in state $|0\rangle$ is 0%. Prior to measurement, the probability of the qubit being in state $|0\rangle$ or state $|1\rangle$ is exactly 0%. In other words, a pure state is a quantum state where we have complete knowledge of the system and what state it is in. After measurement, of course there is a 50% probability of the system being $|0\rangle$ and 50% probability of being $|1\rangle$.

A mixed state is one a quantum state where it is possible that the quantum system *could* be $|\psi_1\rangle$ or $|\psi_2\rangle$ and so on due to noise or improper state preparation. Therefore, a mixed state is described as a probabilistic combination of all of the pure states it could have been prepared as.

For example, if it is desired to prepare a state $|\psi_1\rangle = |0\rangle$, but errors occur, we must account for all possible errors. Suppose that the system produces $|\psi_2\rangle = |1\rangle$ and $|\psi_3\rangle = |+\rangle$ erroneously at half the rate of producing the desired state.

We **cannot** describe the above mixed state situation using *state vectors* because state vectors can only describe pure states. This statistical mixture can only be conveniently described by a *density operator* as:

$$\rho = \sum_{i=1}^{3} p_i|\psi_i\rangle\langle\psi_i| = \frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{4}|\psi_2\rangle\langle\psi_2| + \frac{1}{4}|\psi_3\rangle\langle\psi_3| \tag{6}$$

A mixed state is **not** the same as a superposition state. One way to clarify the language is a little is to refer to superpositions as *coherent* superpositions and mixed states as *incoherent* mixtures of states.

Generally, mixed states are composed of *convex combinations* of pure states. Convex combinations are linear combinations $p_1 f_1 + p_2 f_2 + \ldots p_N f_N$ such that all the coefficients are non-negative $0 \geq p_i \geq 1 \;\; \forall i$ and the sum of all coefficients in unity i.e., $p_1 + p_2 + \cdots + p_N = 1$. A density operator is not unique. Pure states are the extreme points of the convex set of states.

The expectation value of measuring the quantum state to be $|\phi\rangle$ is $\langle\phi|\rho|\phi\rangle$.

2.2 Properties of the density operator

$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$ is the density operator for an ensemble of quantum states $p_i, |\psi_i\rangle$ iff:

1. $\mathrm{tr}(\rho) = 1$
2. ρ is positive i.e., all components of the density matrix are positive, meaning that $\langle\psi|\rho|\psi\rangle \geq 0$

3 Noise theory

Now that we have covered the language of density operators, we can move on to discussing how to use them to describe the effect of noise on a quantum system.

3.1 Quantum channels and the Kraus representation

A quantum channel is a general operation on a quantum state. It could be a desired operation, such as the application of a gate, or an undesired operation, such as the introduction of noise.

Quantum channels are modelled as completely-positive, trace-preserving transformations of density matrices. A quantum channel $\mathcal{N}$ acting on an existing density matrix ρ can be written as $\mathcal{N}(\rho) = \sum_i s_i K_i \rho K_i^\dagger$, where K_i are the Kraus operators that describe the quantum channel and s_i are the probabilities that the operation associated with K_i occurs.

3.1.1 Example 1: Application of a unitary gate U on a pure state $|\psi\rangle$.

In the language of state vectors, this operation is $|\phi\rangle = U|\psi\rangle$. The original density matrix is $\rho = |\psi\rangle\langle\psi|$. The new density matrix after the gate has been applied is $\rho' = |\phi\rangle\langle\phi|$. Substituting in the definition of $|\phi\rangle$, we obtain: $\rho' = U|\psi\rangle\langle\psi|U^\dagger = U\rho U^\dagger$.

The quantum channel, then, is $\rho' = \mathcal{N}(\rho) = U\rho U^\dagger$. There is only one Kraus operator, $K = U$, and there is a $s = 1$ probability of that K being applied.

3.1.2 Example 2: Bit flip error of $|\psi\rangle$.

Suppose that our initial state if $|\psi\rangle = |0\rangle$. This single qubit could unfortunately undergo undesired bit-flipping with probability s. The initial density operator is $\rho = |0\rangle\langle 0|$ and the final density operator is $\rho' = (1-s)|0\rangle\langle 0| + s|1\rangle\langle 1|$. To expound, the initial state $|0\rangle \to X|0\rangle$ with probability s, where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the Pauli-X matrix (i.e., the bit flip matrix).

By substitution, the final density operator can be written as $\rho' = (1-s)|0\rangle\langle 0| + sX|0\rangle\langle 0|X^\dagger$. After using the definition of the initial ρ here, we obtain $\rho' = \mathcal{N}(\rho) = (1-s)I\rho I^\dagger + sX\rho X^\dagger$.

Clearly, $K_1 = I$ with probability $(1-s)$ of occurring and $K_2 = X$ with probability s of occurring.

3.1.3 Example 3: Phase flip error of $|\psi\rangle$

Similarly to Example 2, the initial state is $|\psi\rangle = |0\rangle$. A phase flip error rotates the qubit by $\Phi = \pi$ as in $|\psi\rangle = a|0\rangle + b|1\rangle \to a|0\rangle + be^{i\Phi}|1\rangle$. With $\Phi = \pi$, we have $|\psi\rangle = a|0\rangle + b|1\rangle \to a|0\rangle - b|1\rangle$. Recall that the Pauli-Z gate is $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

By the same reasoning as in Example 2, a phase flip error model is $\rho' = \mathcal{N}(\rho) = (1-s)\rho + sZ\rho Z^\dagger$, with s as the probability of a phase flip occurring.

3.2 Depolarising noise

Depolarisation is when a qubit loses all quantum information and decays into a completely incoherent mixture of the computational basis states $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. The loss of information is complete and so is called a completely mixed state, $I/2$. The state of a single qubit after the noise is introduced is,

$$\rho' = \mathcal{N}(\rho) = (1-s)\rho + s\frac{I}{2}. \tag{7}$$

where ρ is the density matrix prior to the introduction of noise. $\mathcal{N}(\rho)$ represents a quantum operation on the initial state ρ. s is the probability of total decay occurring and $1-s$ is the probability of the single qubit state being left untouched. Graphically, the effect of depolarisation on the qubit represented by the Bloch sphere is a uniform contraction i.e., radius of the sphere decreases about the origin [3].

A completely or maximally mixed state has a density matrix that is proportional to the identity matrix. The reason for that can be derived as follows. If we assume that $\rho = \sum_i p_i|\psi\rangle\langle\psi|$, where $|\psi_i\rangle$ forms an orthonormal basis and p_i is the probability of observing $|\psi_i\rangle$, then a maximally-mixed state is a quantum state where it is equally probable to observe any of $|\psi_i\rangle$. In other words, $p_i = 1/n$ - a uniform probability distribution where n is the dimension of the state. This state is maximally mixed because it is a mixture where all states have equal probability. For finite N, this is equivalent to saying that $\rho \propto I$. To see why, consider a 1-qubit system. The bases are $|0\rangle$ and $|1\rangle$. If $\rho = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$, then we can see that ρ is the scaled identity matrix with trace $= 1$. For a general number of qubits n, $\rho' = \mathcal{N}(\rho) = (1-s)\rho + s\frac{I}{2^n}$ [3].

An equivalent description of depolarising noise is,

$$\rho' = \mathcal{N}(\rho) = (1-s)\rho + s\frac{X\rho X + Y\rho Y + Z\rho Z}{3}, \tag{8}$$

which may be interpreted as applying the X, Y, or Z gates with probability $s/3$.

4 Zero-noise extrapolation theory

In a nutshell, we deliberately introduce noise or allow more noise to contaminate the the quantum system. From a set of systematically higher-noise results, we can extrapolate to obtain the zero noise result.

4.1 Unitary Folding

Unitary folding makes the circuit longer so as to permit more time for noise to contaminate the system [2]. Unitary folding is performed by replacing the operator G with,

$$G \to G(G^\dagger G)^n \tag{9}$$

where G represents a gate or a circuit and n is a positive integer [1]. Note that, because G is unitary, $G^\dagger G = I$, so the expected result without the extra terms should be obtained on an noise-free quantum computer. On real quantum hardware, the noise should increase because the total number of operations increases by a factor of $1 + 2n$.

There are two forms of unitary folding: Local (i.e., gate) folding and Global (i.e., circuit) folding.

4.1.1 Local unitary folding

Individual gates or unitary layers in a circuit C undergo the unitary folding transformation of Equation (9) in-place in the circuit. By folding n times, the total number of gates increases by a factor of $2n + 1$

4.1.2 Global unitary folding

The substitution in Equation (9) is made for the entire circuit. That is, if a circuit C is composed of d unitary layers as,

$$C = G_d \dots G_3 G_2 G_1, \tag{10}$$

then the final circuit after global unitary folding has been performed n times is,

$$C = [G_d \dots G_3 G_2 G_1](G_1^\dagger G_2^\dagger G_3^\dagger \dots G_d^\dagger \; G_d \dots G_3 G_2 G_1)^n \tag{11}$$

This results in a circuit that is an odd integer number of times deeper and thus theoretically scales the non-unitary effect of the noise by an odd integer factor $\lambda = 1, 3, 5, \dots$.

A more resolved folding process occurs by folding the entire circuit n times and then append unitary folding of the final s layers as in,

$$C = U(U^\dagger U)^n (L_d^\dagger \dots L_s^\dagger \; L_s \dots L_d) \tag{12}$$

From a given value of λ and d, the procedure to determine the value of s is:

1. Determine the value of $k = \text{round}(d(\lambda - 1)/2)$
2. Perform integer division of k by d. The quotient is n and the remainder is s
3. Perform unitary folding n times and then s partial folds

4.1.3 When will unitary folding work?

Unitary folding is not likely to scale the noise amplitude resulting from systematic or coherent errors because applying the inverse operation is likely to undo the error [1]. Also, unitary folding will not work for state preparation and measurement (SPAM) noise because this type of noise is invariant to circuit depth.

Unitary folding is likely to amplify incoherent noise that result from the application of gates or from the time taken to complete a computation.

4.1.4 Theory of unitary folding

To model the noise associated with each gate L_j, suppose that each gate had a probability $(1 - p_j)$ of introducing decoherent noise. That is, for any input $|\psi\rangle$, the output could be the desired output $L_j|\psi\rangle$ with a probability of p_j or it could be an undesired decoherent noise I/D with probability $(1 - p_j)$. Using density matrix formalism, that is,

$$\rho' = \mathcal{N}_j(\rho) = p_j L_j \rho L_j^\dagger + (1 - p_j)\frac{I}{D} \tag{13}$$

where $D = 2^N$ is the dimension of the Hilbert space associated with all N qubits in the quantum circuit. By chaining a sequence of gate operations until all gates in the circuit have been applied, one obtains the circuit-level noise model of,

$$\rho' = \mathcal{N}(\rho) = pC\rho C^\dagger + (1 - p)\frac{I}{D} \tag{14}$$

where $p = \prod_i p_i$. You can prove this to yourself with just two gates. Obtain $\rho_1(\rho_0$ and then insert ρ_1 into the functional form of ρ_2. Another way to obtain this is from realising that the depolarising channel commutes with unitary operators and therefore we can postpone all noise channels to the end [1].

Now, if we apply unitary folding by a factor of $\lambda = 1 + 2n$, where n is an odd integer, we obtain a density operator of,

$$\rho' = \mathcal{N}(\rho) = p^\lambda C\rho C^\dagger + (1 - p^\lambda)\frac{I}{D}. \tag{15}$$

The expectation value of any observable A is $\langle\phi|A|\phi\rangle = \mathrm{tr}(\rho A)$ and should depend on λ in an exponential manner: $E(\lambda) = a + bp^\lambda$.

References

[1] Tudor Giurgica-Tiron, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J. Zeng. Digital zero noise extrapolation for quantum error mitigation. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, October 2020.

[2] Ying Li and Simon C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X*, 7:021050, Jun 2017.

[3] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, 2010.