8.2 Quantum Error Mitigation

Quantum Error Mitigation (QEM) wants to translate the improvements of quantum hardware in those of quantum information and computation. Namely, it is an algorithmic scheme that reduces noise-induced bias in the expectation value of an observable of interest by post-processing outputs from an ensemble of circuit runs. To do this, it employs circuits at the same depth as the original unmitigated circuit or above. QEM applies post-processing directly from the hardware outputs. Thus, if the circuit size, being the product of the circuit depth times the number of qubits, becomes too large then QEM loses its usefulness.

A good QEM approach should employ a limited number of qubits, while still providing a guaranteed accuracy. This converts in a formal error bound, which indicates how well the QEM code works. Moreover, it should employ only a few (or better none) assumptions about the final state. For example, assuming that the final state is factorised is not a good assumption. Indeed, it would strongly limit the applicability of the corresponding QEM algorithm.

Before dwelling in two, among various, algorithms in the QEM context, we provide the general idea of the QEM approach. We defined the primary circuit as that process that would ideally produce the perfect output state $\hat{\rho}_0$. Due to the presence of noises, the primary circuit produces the noisy state $\hat{\rho}$. To account how a circuit works, we consider an observable of interests $\hat{O}$ whose expectation value is the output information we seek. To compute this, we will run the circuit N_{sample} times, which is the number of circuit executions. Also in the noiseless case, a finite value of N_{sample} implies a finite inaccuracy of the estimated average. This is the so-called shot noise. However, in such a case, there will be no systematic shift, i.e. bias, in the expectation value of $\hat{O}$ due to the noise. QEM aims to reduce such a bias. Often, this implies that the corresponding variace increases. Then, one needs to increase the number of circuit runs $N > N_{\text{sample}}$ to compensate. The sampling overhead is the cost, in terms of number of repetitions, of the QEM method when compared to the noiseless circuit.

We underline that, conversely to QEC, in QEM there is no monitoring of the errors occurring during the run of the circuit.

8.2.1 Zero noise extrapolation

The Zero noise extrapolation (ZNE) method extracts the zero-noise expectations from a fitting of the circuit run at different values of the noise. We define a time dependent Hamiltonian $\hat{H}(t)$ that embeds action of the noiseless circuit. It can be written as

$$\hat{H}(t) = \sum_\alpha J_\alpha(t)\hat{P}_\alpha, \tag{8.80}$$

where $J_\alpha(t)$ are some time dependent couplings that switch on and off the gates of the circuit, which are implemented by the corresponding N-qubit Pauli operators $\hat{P}_\alpha$. The full dynamics, including the action of the noise, is given by the following master equation

$$\frac{\mathrm{d}\hat{\rho}_\lambda(t)}{\mathrm{d}t} = -\frac{i}{\hbar}\left[\hat{H}(t), \hat{\rho}_\lambda(t)\right] + \lambda\mathcal{L}[\hat{\rho}_\lambda(t)], \tag{8.81}$$

where $t \in [0, T]$, with T being the time at which the circuit ends. We assume here that the noise coupling λ is small. Moreover, we assume that the noise dissipator $\mathcal{L}$ is invariant under time rescaling and it is independent of $J_\alpha(t)$.

Now, given the observable of interest $\hat{A}$, we compute the corresponding expectation value on the noisy circuit as $E(\lambda) = \mathrm{Tr}\left[\hat{A}\hat{\rho}_\lambda(T)\right]$, where $\hat{\rho}_\lambda(T)$ is the solution of Eq. (8.81). What we want to do is to estimate $E(\lambda)$ for $\lambda \to 0$. Since one cannot reduce the value of λ, to construct a series of measurement from where extrapolate the estimate $E(0)$, we increase the value of λ. This can be done by considering the following rescaling. We dilate the time T at which the circuit is ran and, due to the time invariance of $\mathcal{L}$, this is equivalent to let the noise act more on the circuit. Then, one applies this idea with different values of λ and can perform a fit and deduce the value of E for $\lambda \to 0$. Practically, we perform the circuit $N_{\rm cir}$ times, at different values of the noise rate $\lambda_j = c_j\lambda$, $j = 0, \ldots, N_{\rm cir} - 1$ with $c_0 = 1 < c_1 < \cdots < c_{N_{\rm cir}-1}$. For each value of λ_j, we run the circuit with the following rescaled Hamiltonian

$$\hat{H}^{(j)}(t) = \sum_\alpha J_\alpha^{(j)}(t)\hat{P}_\alpha, \quad \text{where} \quad J_\alpha^{(j)}(t) = c_j^{-1} J_\alpha(c_j^{-1}t), \tag{8.82}$$

for a time $T_j = c_j T$. The rescaled dynamics gives

$$\frac{\mathrm{d}\hat{\rho}_\lambda^{(j)}(t)}{\mathrm{d}t} = -\frac{i}{\hbar}\left[\hat{H}^{(j)}(t), \hat{\rho}_\lambda^{(j)}(t)\right] + \lambda\mathcal{L}[\hat{\rho}_\lambda^{(j)}(t)]. \tag{8.83}$$

By merging the latter with Eq. (8.82), we obtain

$$\frac{\mathrm{d}\hat{\rho}_\lambda^{(j)}(t)}{\mathrm{d}t} = -\frac{i}{\hbar}\sum_\alpha c_j^{-1} J_\alpha(c_j^{-1}t)\left[\hat{P}_\alpha, \hat{\rho}_\lambda^{(j)}(t)\right] + \lambda\mathcal{L}[\hat{\rho}_\lambda^{(j)}(t)]. \tag{8.84}$$

By defining $s = c_j^{-1}t$, which runs in the interval $s \in [0, T]$ since $t \in [0, T_j]$, we rewrite the above master equation as

$$\frac{\mathrm{d}\hat{\rho}_\lambda^{(j)}(t)}{\mathrm{d}t} = \frac{\mathrm{d}\hat{\rho}_\lambda^{(j)}(c_j s)}{c_j \mathrm{d}s} = -\frac{i}{\hbar}\sum_\alpha c_j^{-1} J_\alpha(s)\left[\hat{P}_\alpha, \hat{\rho}_\lambda^{(j)}(c_j s)\right] + \lambda\mathcal{L}[\hat{\rho}_\lambda^{(j)}(c_j s)]. \tag{8.85}$$

By multiplying the left and right hand side by c_j we obtain

$$\frac{\mathrm{d}\hat{\rho}_\lambda^{(j)}(c_j s)}{\mathrm{d}s} = -\frac{i}{\hbar}\left[\hat{H}(s), \hat{\rho}_\lambda^{(j)}(c_j s)\right] + c_j\lambda\mathcal{L}[\hat{\rho}_\lambda^{(j)}(c_j s)], \tag{8.86}$$

which is Eq. (8.81) with λ substituted with $c_j\lambda$. Its solution at time $s = T$ is given by $\hat{\rho}_{c_j\lambda}(T) = \hat{\rho}_\lambda^{(j)}(T_j)$. Correspondingly, we compute the expectation value $E(\lambda_j) = \mathrm{Tr}\left[\hat{A}\hat{\rho}_\lambda^{(j)}(T_j)\right] = \mathrm{Tr}\left[\hat{A}\hat{\rho}_{c_j\lambda}(T)\right]$. Experimentally, for each c_j, one performs $N_{\rm sample}$ runs of the circuit and obtains an estimator $\tilde{E}(\lambda_j)$, which converges to the true value $E(\lambda_j)$ only in the asymptotic limit $N_{\rm sample} \to \infty$. Specifically, one has

$$\tilde{E}(\lambda_j) = E(\lambda_j) + \tilde{\delta}, \tag{8.87}$$

where $\tilde{\delta}$ is a random variable with zero mean and variance $\mathbb{E}[\tilde{\delta}^2] = \sigma_0^2/N_{\rm sample}$, with σ_0^2 corresponding to the single-shot variance. Here, $\mathbb{E}$ is to the mean over the sampling.

Now, the ZNE problem is to construct a good estimator $\tilde{E}(0)$ for the expectation value $E(\lambda = 0) = \mathrm{Tr}\left[\hat{A}\hat{\rho}_0(T)\right]$ from the set of estimators $\tilde{E}(\lambda_j)$. Figure 8.2 represents the problem. To be a good estimator, we want that its bias

$$\mathrm{Bias}(\tilde{E}(0)) = \mathbb{E}[\tilde{E}(0) - E(0)], \tag{8.88}$$

and its variance

$$\mathrm{Var}(\tilde{E}(0)) = \mathbb{E}[\tilde{E}(0)^2] - \mathbb{E}[\tilde{E}(0)]^2, \tag{8.89}$$

are both small. We employ the mean squared error (MSE) are a figure of merit with respect to the true unknown parameter

$$\begin{aligned} \mathrm{MSE}(\tilde{E}(0)) &= \mathbb{E}[(\tilde{E}(0) - E(0))^2], \\ &= \mathrm{Var}(\tilde{E}(0)) + (\mathrm{Bias}(\tilde{E}(0)))^2. \end{aligned} \tag{8.90}$$

If the expectation value $E(\lambda)$ can be an arbitrary function of λ without any regularity assumption, then ZNE is impossible. However, from physical considerations, it is reasonable to have a model for it, for example we can assume a linear, a polynomial or an exponential dependence with respect to λ.

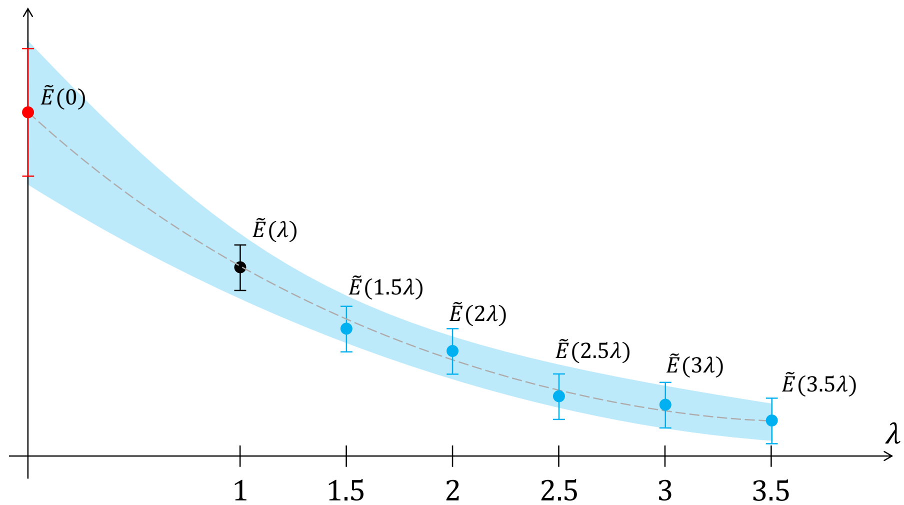

Fig. 8.2: Graphical representation of the Zero Error Extrapolation. Given the set of estimators $\tilde{E}(\lambda)$ at different values of the noise (black and blue dots and error bars), one performs a fit assuming a specific model (gray dashed line) with corresponding confidence region (light blue region). In such a way, one extrapolates the value of $\tilde{E}(0)$ with its corresponding error bar (red point and error bar).

1 If we assume a linear dependence on λ, the corresponding linear model is given by

$$E_{\text{linear}}(\lambda) = a_0 + a_1\lambda. \tag{8.91}$$

In such a case, a simple analytic solution exists, which is that of the ordinary least squared estimator of the intercept parameter. Namely, we have

$$\tilde{E}_{\text{linear}}(0) = \bar{E}(\lambda) - \frac{S_{\lambda E}}{S_{\lambda\lambda}}\bar{\lambda}, \tag{8.92}$$

where

$$\begin{aligned} \bar{\lambda} &= \frac{1}{N_{\text{cir}}}\sum_{j=0}^{N_{\text{cir}}-1}\lambda_j, \\ \bar{E}(\lambda) &= \frac{1}{N_{\text{cir}}}\sum_{j=0}^{N_{\text{cir}}-1}\tilde{E}(\lambda_j), \\ S_{\lambda E} &= \sum_{j=0}^{N_{\text{cir}}-1}(\lambda_j - \bar{\lambda})(\tilde{E}(\lambda_j) - \bar{E}(\lambda)), \\ S_{\lambda\lambda} &= \sum_{j=0}^{N_{\text{cir}}-1}(\lambda_j - \bar{\lambda})^2. \end{aligned} \tag{8.93}$$

With respect to the zero noise value $E_{\text{linear}}(0)$, the estimator $\tilde{E}_{\text{linear}}(0)$ is unbiased. Its variance, under the assumption that the statistical uncertainty is the same for each λ_j, reads

$$\text{Var}(\tilde{E}_{\text{linear}}(0)) = \frac{\sigma_0^2}{N_{\text{sample}}}\left(\frac{1}{N_{\text{cir}}} + \frac{\bar{\lambda}^2}{S_{\lambda\lambda}}\right). \tag{8.94}$$

2 The Richardson's extrapolation is a special case of the polynomial extrapolation, which is limited at order $N_{\text{cir}} - 1$. The corresponding model is given by

$$E_{\text{Rich}}(\lambda) = a_0 + a_1\lambda + \cdots + c_{N_{\text{cir}}-1}\lambda^{N_{\text{cir}}-1}. \tag{8.95}$$

This is the only case in which the fitted polynomial perfectly interpolates the N_{cir} data points such that, in the ideal limit of an infinite number of samples $N_{\text{sample}} \to \infty$, the error with respect to the true expectation value is by construction $\mathcal{O}(N_{\text{cir}})$. Using the Lagrange polynomial, the estimator can be expressed explicitly as

$$\tilde{E}_{\text{Rich}}(0) = \sum_{j=0}^{N_{\text{cir}}-1} \tilde{E}(\lambda_j)\gamma_j, \tag{8.96}$$

where

$$\gamma_j = \prod_{m \neq j} \frac{c_m}{c_j - c_m}. \tag{8.97}$$

The error of the estimator is $\mathcal{O}(N_{\text{cir}})$ only in the asymptotic limit $N_{\text{sample}} \to \infty$. In other words, $\mathcal{O}(N_{\text{cir}})$ corresponds to the bias term in Eq. (8.88). In a real scenario, N_{sample} is finite, and the variance term in Eq. (8.88) grows exponentially as we increase N_{cir}.