Unitary Fund

Richardson Extrapolation

Quantum Wednesday

Nate Stemen

Feb 8, 2023

Today's Paper

Optimization of Richardson extrapolation for quantum error mitigation

Michael Krebsbach,[1,*] Björn Trauzettel,[1,2] and Alessio Calzona[1,2]

[1] *Institute for Theoretical Physics and Astrophysics, University of Würzburg, D-97074 Würzburg, Germany*
[2] *Würzburg-Dresden Cluster of Excellence ct.qmat, Germany*

(Dated: August 26, 2022)

Quantum error mitigation is a key concept for the development of practical applications based on current noisy intermediate scale quantum (NISQ) devices. One of the most promising methods is Richardson extrapolation to the zero noise limit. While its main idea is rather simple, the full potential of Richardson extrapolation has not been completely uncovered yet. We give an in-depth analysis of the relevant parameters of Richardson extrapolation and propose an optimized protocol for its implementation. This protocol allows for a precise control of the increase in statistical uncertainty and lays the foundation for a significant improvement of the mitigation performance achieved by increasing the number of nodes. Furthermore, we present a novel set of nodes that, on average, outperforms the linear, exponential or Chebyshev nodes frequently used for Richardson extrapolation without requiring any additional resources.

https://arxiv.org/abs/2201.08080

Overview

1. What is Richardson Extrapolation?
2. How do we *already* use it?
3. What sort of optimizations do the authors find?

What is extrapolation?

How do we use Richardson Extrapolation?

```
class RichardsonFactory():
  def extrapolate(
    scale_factors: Sequence[float],
    exp_values: Sequence[float],
    full_output: bool = False,
  ) -> ExtrapolationResult:
    # Richardson extrapolation is a particular case of a
    # polynomial fit with order equal to the number of
    # data points minus 1.
    order = len(scale_factors) - 1
    return PolyFactory.extrapolate(
      scale_factors, exp_values, order, full_output
    )
```

Problem setup

- $E(\lambda) = \mathrm{tr}(\rho_\lambda A)$
- $\lim_{\lambda \to 0} E(\lambda) = E^*$
- Define $\lambda = \lambda_0 x$, and assume we can access any $x \geq 1$
- Take $1 = x_0 < x_1 < \cdots < x_n$, and collect $E(x_i)$
- Interpolate a polynomial of degree n to $(x_i, E(x_i))$
- Evaluate polynomial at $\lambda = 0$

$$E(0) \approx \sum_{i=0}^{n} E(x_i)\gamma_i =: R_n$$

$$\gamma_i = \prod_{k \neq j} \frac{x_k}{x_k - x_j}$$

$$E(x) = E^* + \sum_{k=0}^{\infty} a_k \lambda_0^k x^k$$

$$\begin{aligned} E(0) &\approx \sum_{i=0}^{n} \left[E^* + \sum_{k=0}^{\infty} a_k \lambda_0^k x_i^k \right] \gamma_i \\ &= E^* \sum_{i=0}^{n} \gamma_i + \sum_{i=0}^{n} \sum_{k=0}^{\infty} a_k \lambda_0^k \gamma_i x_i^k \\ &= E^* + \sum_{k=n+1}^{\infty} a_k \lambda_0^k \sum_{i=0}^{n} \gamma_i x_i^k \end{aligned}$$

Bias and Variance

$$\mathrm{Bias}[\hat{R}_n] = (-1)^n E^{(n+1)}(\xi)\frac{C_n}{(n+1)!}$$

$$\mathrm{Var}[\hat{E}(x_i)] = \frac{\sigma_i^2}{N_i}$$

$$\mathrm{Var}[\hat{R}_n] = \sum_{i=0}^{n} \gamma_i^2 \frac{\sigma^2}{N_i}$$

$$C_n = \prod_{i=0}^{n} x_i$$

$$\xi \in [0, x_n]$$

$$\sigma_i = \sqrt{\mathrm{tr}(\rho_{\lambda_0 x_i} A^2) - E(x_i)^2}$$

$$N_i = \#\text{ times } A \text{ is measured at } x_i$$

What could we optimize?

1. Is there a way to reduce the increase in variance?
2. Is there a scaling method for the x_i's that optimizes the bias/variance tradeoff?
3. Is there an optimal number of samples n?

Variance reduction

A **simple** calculation shows $\mathrm{Var}[\hat{R}_n] = \sum_{i=0}^{n} \gamma_i^2 \frac{\sigma^2}{N_i}$ is minimized by

$$N_i = N_{\text{total}} \frac{|\gamma_i|}{\sum_{k=0}^{n} |\gamma_k|}$$

Defining $\Lambda = \sum_{i=0}^{n} |\gamma_i|$ and $N_{\text{eff}} = N_{\text{total}}/\Lambda^2$:

$$\mathrm{Var}[\hat{R}_n] = \frac{\sigma^2}{N_{\text{total}}} \Lambda^2 = \frac{\sigma^2}{N_{\text{eff}}}$$

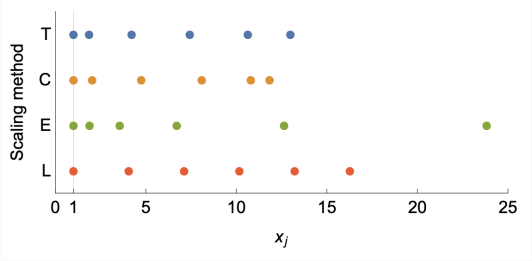
Node Spacing

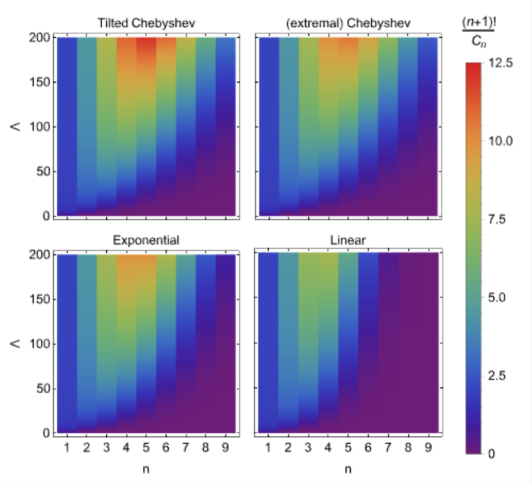
$$x_j^{\mathrm{L}} = 1 + j(x_1 - 1)$$

$$x_j^{\mathrm{E}} = x_1^j$$

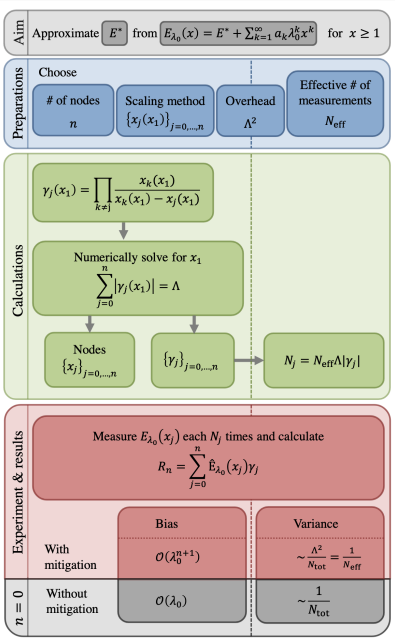
$$x_j^{\mathrm{C}} = 1 + \frac{\sin^2(\frac{j}{n}\frac{\pi}{2})}{\sin^2(\frac{1}{n}\frac{\pi}{2})}(x_1 - 1)$$

$$x_j^{\mathrm{T}} = 1 + \frac{\sin^2(\frac{j}{n+1}\frac{\pi}{2})}{\sin^2(\frac{1}{n+1}\frac{\pi}{2})}(x_1 - 1)$$

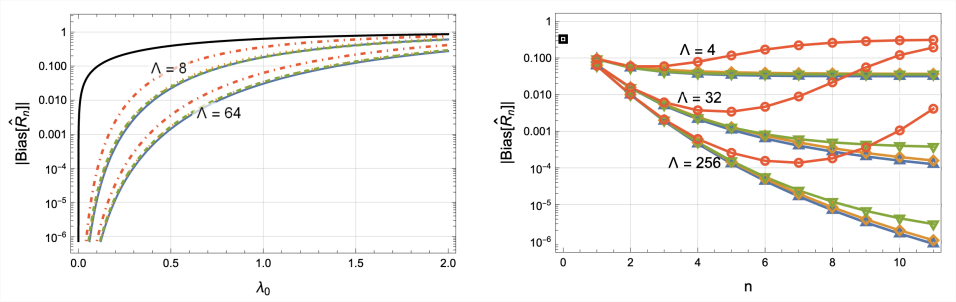


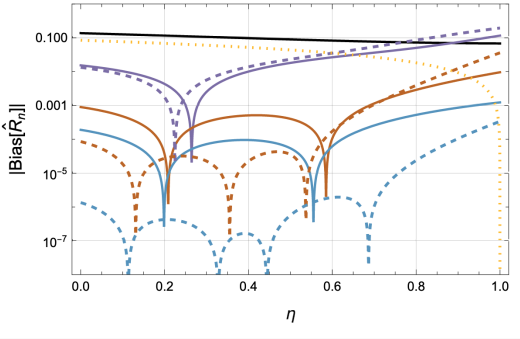


Overview


Aim
Approximate E^* from $E_{\lambda_0}(x) = E^* + \sum_{k=1}^{\infty} a_k \lambda_0^k x^k$ for $x \geq 1$
Preparations
Choose
# of nodes n
Scaling method $\{x_j(x_1)\}_{j=0,\dots,n}$
Overhead Λ^2
Effective # of measurements N_{eff}
Calculations
$\gamma_j(x_1) = \prod_{k \neq j} \frac{x_k(x_1)}{x_k(x_1) - x_j(x_1)}$
Numerically solve for x_1
$\sum_{j=0}^{n} |\gamma_j(x_1)| = \Lambda$
Nodes $\{x_j\}_{j=0,\dots,n}$
$\{\gamma_j\}_{j=0,\dots,n}$
$N_j = N_{\text{eff}} \Lambda |\gamma_j|$
Experiment & results
Measure $E_{\lambda_0}(x_j)$ each N_j times and calculate
$R_n = \sum_{j=0}^{n} \hat{E}_{\lambda_0}(x_j) \gamma_j$
Bias
Variance
With mitigation
$\mathcal{O}(\lambda_0^{n+1})$
$\sim \frac{\Lambda^2}{N_{\text{tot}}} = \frac{1}{N_{\text{eff}}}$
$n = 0$
Without mitigation
$\mathcal{O}(\lambda_0)$
$\sim \frac{1}{N_{\text{tot}}}$

Results





Thank you!