Why Sample Size Strongly Affects Results in Sample-Based Quantum Diagonalization

Sample-based quantum diagonalization (SQD), also known as Quantum Selected Configuration Interaction (QSCI), relies on sampling quantum states to build a subspace for classical diagonalization and thus approximate eigenvalues and eigenstates of quantum systems. The accuracy and utility of this method are highly sensitive to the number of samples collected. This sensitivity arises from several fundamental aspects of the sampling process and the structure of quantum states.

Key Reasons for Sample Size Sensitivity:

- **Redundant Sampling and Diminishing Returns:**
  As the number of samples increases, the chance of repeatedly sampling the same determinants (basis states) grows, especially for those with larger weights in the quantum state. This leads to many samples providing no new information, resulting in diminishing returns for increasing the sample size. For example, a 100-fold increase in samples might yield only a 12-fold increase in unique determinants discovered, as observed in practical studies[1]. This inefficiency directly impacts the ability to construct a sufficiently rich subspace for accurate diagonalization.
- **Trade-off Between Accuracy and Compactness:**
  To achieve high-accuracy results, a large number of unique, important determinants must be found. However, sampling tends to favor determinants with higher probabilities, making it difficult to discover those with smaller weights that may still be crucial for accuracy. Adjusting the sampling distribution to find more unique determinants can lead to less compact (larger) CI expansions, which are computationally expensive to handle classically[2] [1]. Thus, increasing the sample size can improve accuracy but at the cost of efficiency and tractability.
- **Sampling Noise and Statistical Error:**
  The finite number of samples introduces statistical (sampling) noise. This noise propagates into the estimation of matrix elements and, ultimately, the calculated eigenvalues and eigenstates. The error in the results decreases only as the square root of the number of samples, so achieving high precision requires a very large number of samples[3]. In quantum Krylov subspace diagonalization, for instance, the error bound from finite sampling accurately predicts the observed experimental errors, underscoring the direct link between sample size and result quality[3].
- **Measurement Overheads and Practical Limitations:**
  Because many samples are "wasted" by repeating already-seen determinants, the measurement overhead (number of quantum measurements required) becomes immense for large systems or high-accuracy targets. This overhead can make the method impractical, especially given the current limitations of quantum hardware[1].

- **Effect on Subspace Dimension and Diagonalization:**
  The number of unique samples directly determines the dimension of the subspace used for diagonalization. Larger subspaces generally yield more accurate results but require more computational resources for classical diagonalization[4]. However, if the subspace grows too large due to non-compact sampling, the classical step can become intractable, offsetting the benefits of quantum sampling[2] [1].

Summary Table: Sample Size Effects in SQD

Effect of Increasing Sample Size	Benefit	Drawback
More unique determinants found	Higher accuracy possible	Diminishing returns, redundancy
Larger subspace for diagonalization	Better approximation of eigenstates	Increased classical cost, less compact CI expansion
Reduced statistical noise	More reliable energy estimates	Requires exponentially more samples for high precision
Sampling from low-probability determinants	More complete wavefunction	Less efficient, more memory/computation needed

Conclusion

In sample-based quantum diagonalization, the sample size critically affects results because:

- More samples are needed to discover enough unique, important determinants for accurate results.
- However, most additional samples end up redundant, leading to inefficiency and high measurement overhead.
- Statistical noise from finite sampling limits precision, and only very large sample sizes can reduce this error.
- There is a fundamental trade-off: increasing sample size can improve accuracy but at the cost of computational efficiency and practicality, especially as the CI expansion becomes less compact and more expensive to process[2] [1] [3].

These challenges highlight why sample size is such a crucial—and limiting—factor in the effectiveness of sample-based quantum diagonalization methods.

⁂

1. https://arxiv.org/html/2501.07231v3
2. https://arxiv.org/html/2501.07231v1
3. https://quantum-journal.org/papers/q-2024-09-19-1477/
4. https://qiskit.github.io/qiskit-addon-sqd/