Reducing Measurement Costs by Recycling the Hessian in Adaptive Variational Quantum Algorithms

Mafalda Ramôa,[1,2,3,4,5] Luis Paulo Santos,[3,4,5] Nicholas J. Mayhall,[2,6] Edwin Barnes,[1,2] and Sophia E. Economou[1,2]

[1] *Department of Physics, Virginia Tech, Blacksburg, VA, 24061, USA*
[2] *Virginia Tech Center for Quantum Information Science and Engineering, Blacksburg, VA 24061, USA*
[3] *International Iberian Nanotechnology Laboratory (INL), Portugal*
[4] *High-Assurance Software Laboratory (HASLab), Portugal*
[5] *Department of Computer Science, University of Minho, Portugal*
[6] *Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061, USA*

Simulating strongly correlated quantum many-body systems is one of the main tasks envisioned for quantum computers. While current devices are limited by high error rates and low qubit counts, there is still hope that they may offer quantum advantage in these problems. As such, the development of algorithms suitable for these near-term machines has gathered remarkable interest in recent years. The proposal of the variational quantum eigensolver (VQE) [1] was the first step in this direction. In contrast with algorithms designed for the fault-tolerant quantum computing era, VQE employs shallow state preparation circuits (ansätze) and undertakes a naturally noise-resilient learning strategy, where a classical optimizer is used to tune parameters of quantum gates.

While remarkable progress was made with VQE, ansätze remain deep beyond viability, and the most hardware-efficient options have been shown not to be trainable [2]. One promising alternative is to build the ansatz adaptively, such that its structure is dictated by the problem at hand. The first algorithm to use such a strategy was the Adaptive Derivative-Assembled Problem-Tailored VQE (ADAPT-VQE), proposed in Ref. [3] and extended in Refs. [4–9].

ADAPT-VQE produces shallower and more trainable circuits at the expense of a significant measurement overhead [10]. Starting from a classical reference state, it creates a circuit block-by-block using information available on the fly. Each iteration selects one anti-Hermitian generator from an operator pool, based on the associated gradient magnitude. This operator is multiplied by a variational parameter and exponentiated to create a parameterized unitary operator, which is appended to the ansatz. A VQE subroutine is then employed to optimize all of the parameters. The initial value for the new parameter is set to zero, while old parameters are recycled.

The optimizer in the VQE subroutine is usually regarded as a black box outputting nothing but optimal parameters. In contrast, we propose a quasi-Newton optimizer for ADAPT-VQE where second-order information gathered by the optimizer naturally flows between optimizations. Our optimizer is an adaptation of BFGS [11–14], a quasi-Newton numerical optimization method which approximates an inverse Hessian matrix H based on evaluations of the cost function and its gradient. H is typically initialized to the identity matrix, under the assumption that we have no prior information regarding the optimization landscape. However, in the context of ADAPT-VQE, this means that each optimization will re-approximate all second derivatives from scratch and discard them in the end, in spite of the fact that the final state for the $(n-1)$th ADAPT-VQE optimization is the initial state for the nth. Since all second derivatives concerning the former $n-1$ parameters should remain as accurate as before, we propose to set the initial inverse Hessian for the nth ADAPT-VQE optimization as

$$H_0 \leftarrow \begin{pmatrix} H^*_{n-1\times n-1} & 0 \\ 0 & 1 \end{pmatrix}, \tag{1}$$

where $H^*_{n-1\times n-1}$ is the final inverse Hessian from the $(n-1)$th ADAPT-VQE optimization. We refer to this strategy as 'Hessian recycling'.

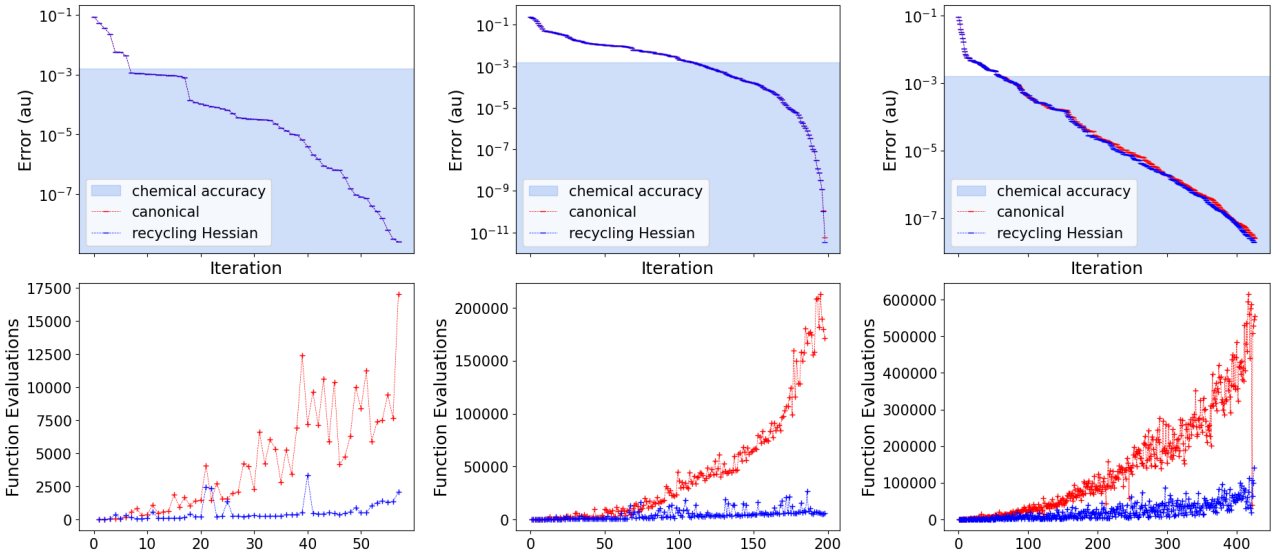
Table I shows the impact of this Hessian recycling strategy on QEB-ADAPT-VQE [5]. In this work, we used the STO-3G basis set for all molecules and set the ADAPT-VQE convergence threshold to 10^{-6} for the 12 qubit molecules (LiH, H_6) and 10^{-5} for BeH_2 (14 qubits). Expectation values were calculated via matrix algebra. The table shows that the cost bottleneck is the VQE optimization—precisely what our protocol addresses. We observe savings between 80 and 90% in the *total* measurement costs of ADAPT-VQE for these small molecules.

In Fig. 1, we observe the evolution of the error and function evaluations per ADAPT-VQE iteration for different systems and pools. We confirm that recycling the Hessian preserves solution quality: The error matches the error of the original protocol in all cases, and the resulting ansätze are the same. Despite preserving the accuracy, our strategy significantly decreases the total number of function evaluations. Its impact is the greatest for the largest (BeH_2) and most strongly correlated (H_6) molecules. Within the same system, it is greater for larger iteration numbers (higher dimensional optimizations). Thus, we can expect even greater savings for classically intractable systems.

In Fig. 2 we look into the optimization process associated with one particular ADAPT-VQE iteration. In line with previous results, Fig. 2a shows that recycling the Hessian results in a similar final error for fewer iterations. In Fig. 2b, we compare the search direction with Newton's direction. Newton's method requires explicitly evaluating the Hessian, which is then used to

			Molecule					
			LiH		H_6		BeH_2	
			1.5Å	3Å	1Å	3Å	1.3Å	3Å
Step		Gradient Measurement	5.2×10^3	5.4×10^3	1.9×10^4	1.9×10^4	1.2×10^4	1.3×10^4
	VQE	Canonical	2.5×10^5	2.4×10^5	1.3×10^7	2.2×10^7	4.2×10^6	3.6×10^6
		Recycling Hessian	6.0×10^4	3.2×10^4	1.7×10^6	7.9×10^6	9.4×10^5	5.6×10^5

TABLE I: Measurement costs incurred by QEB-ADAPT-VQE [5] for several systems. The costs are given as multipliers for the cost of a naive energy evaluation. We consider the gradient measurement strategy in [15].



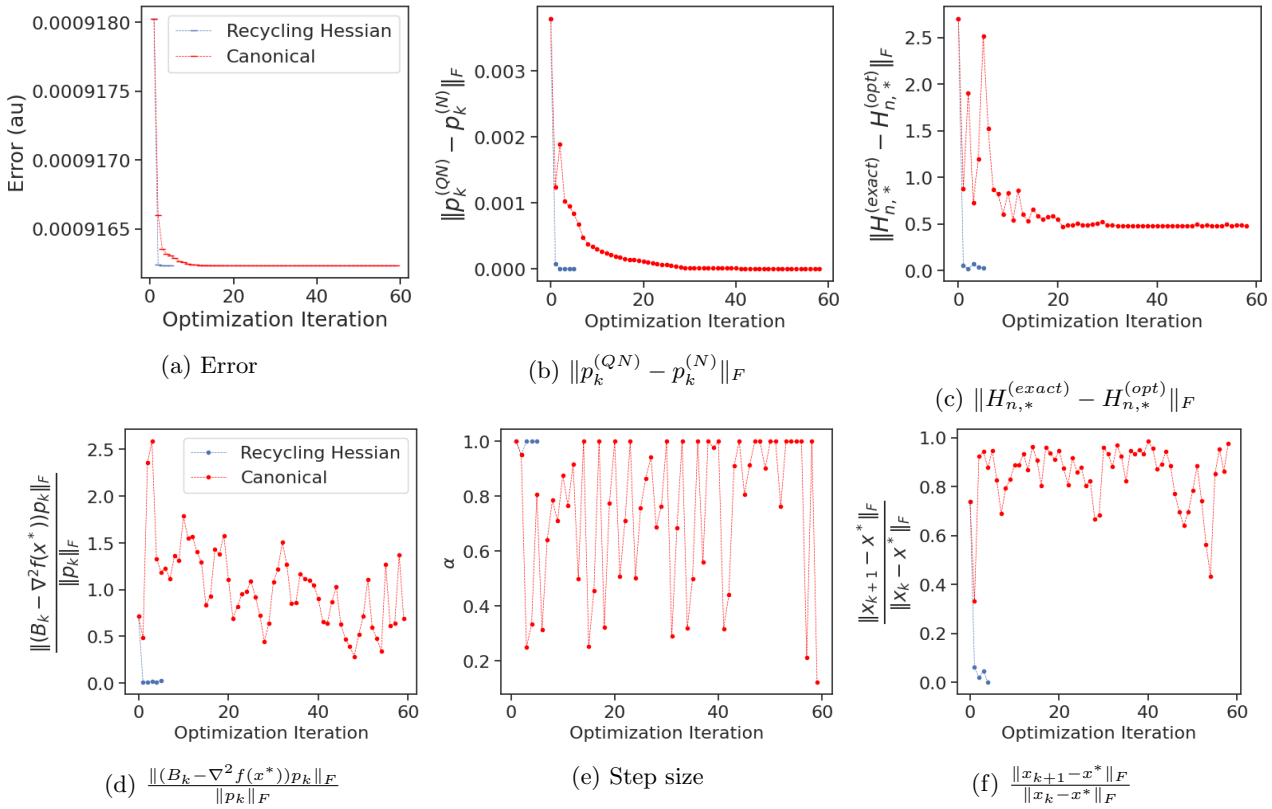
(a) LiH at 3Å, QE pool [5] (b) H_6 at 1.5Å, QE pool [5] (c) BeH_2 at 2Å, Pauli pool [4]

FIG. 1: Energy error (top) and number of function evaluations (bottom) per ADAPT-VQE iteration, with and without Hessian recycling, for several molecules at various bond distances.

set the search direction. This allows Newton's method to converge in fewer iterations, at significant additional measurement cost per iteration. We can see that the search direction aligns much faster with Newton's direction when the Hessian is recycled. This is remarkable: Unlike Newton's method, our optimization method never evaluates the Hessian. Yet, it aligns with the Newton search direction in a fraction of the iterations needed for the canonical BFGS implementation. Figure 2c shows the distance between the last row of the approximate and exact inverse Hessians. Surprisingly, recycling the Hessian allows the second derivatives describing correlations between the new parameter and old ones to converge faster to the true values, even though they were initialized to zero just as in canonical BFGS. Even more surprisingly, the final error is reduced despite the lower number of iterations.

Finally, we focus on the convergence rate of the optimization. The relevant quantities are plotted in the lower panels of Fig. 2. In Fig. 2d, we see that the difference between the approximate and exact Hessians along the search direction goes to zero almost immediately if we recycle the Hessian, but this does not happen when we do not. Further, the step size (Fig. 2e) saturates to unity after only two iterations when we recycle the Hessian, but oscillates instead of stabilizing when we do not. Together these results indicate that recycling the Hessian results in a superlinear convergence that would otherwise not be achieved [14]. Figure 2f confirms this. BFGS with Hessian recycling enjoys superlinear convergence (the optimal scenario for quasi-Newton methods), while canonical BFGS converges linearly with a convergence constant r close to 1—the worst possible scenario [14]. Such a convergence rate is expected of gradient descent in ill-conditioned problems and constitutes an underwhelming performance for a quasi-Newton optimizer.

To conclude, the new technique we propose brings forth significant improvements in the cost-efficiency of ADAPT-VQE, the leading gate-based algorithm for molecular simulations in near-term devices. This is an impactful advance in the race to practical quantum advantage. We believe this will be of interest to the quantum information community and motivate researchers to explore new directions for VQE, related to the development of algorithm-specific optimization protocols.



(a) Error (b) $\|p_k^{(QN)} - p_k^{(N)}\|_F$ (c) $\|H_{n,*}^{(exact)} - H_{n,*}^{(opt)}\|_F$

(d) $\frac{\|(B_k - \nabla^2 f(x^*))p_k\|_F}{\|p_k\|_F}$ (e) Step size (f) $\frac{\|x_{k+1} - x^*\|_F}{\|x_k - x^*\|_F}$

FIG. 2: Evolution of the 75th QEB-ADAPT-VQE optimization with and without Hessian recycling for H_6 at 1Å.

[1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nature Communications **5**, 10.1038/ncomms5213 (2014).

[2] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications **9**, 10.1038/s41467-018-07090-4 (2018).

[3] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nature Communications **10**, 10.1038/s41467-019-10988-2 (2019).

[4] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor, PRX Quantum **2**, 020310 (2021).

[5] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Qubit-excitation-based adaptive variational quantum eigensolver, Communications Physics **4**, 10.1038/s42005-021-00730-0 (2021).

[6] P. G. Anastasiou, Y. Chen, N. J. Mayhall, E. Barnes, and S. E. Economou, Tetris-adapt-vqe: An adaptive algorithm that yields shallower, denser circuit ansätze (2022).

[7] V. Shkolnikov, N. J. Mayhall, S. E. Economou, and E. Barnes, Avoiding symmetry roadblocks and minimizing the measurement overhead of adaptive variational quantum eigensolvers (2021), arXiv:2109.05340 [quant-ph].

[8] L. W. Bertels, H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, Symmetry breaking slows convergence of the adapt variational quantum eigensolver (2022).

[9] C. K. Long, K. Dalton, C. H. W. Barnes, D. R. M. Arvidsson-Shukur, and N. Mertig, Layering and subpool exploration for adaptive variational quantum eigensolvers: Reducing circuit depth, runtime, and susceptibility to noise (2023), arXiv:2308.11708 [quant-ph].

[10] H. R. Grimsley, G. S. Barron, E. Barnes, S. E. Economou, and N. J. Mayhall, Adapt-vqe is insensitive to rough parameter landscapes and barren plateaus (2022).

[11] J. J. E. Dennis and J. J. Moré, Quasi-newton methods, motivation and theory, SIAM Review **19**, 46 (1977).

[12] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Society for Industrial and Applied Mathematics, 1996).

[13] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. (John Wiley & Sons, New York, NY, USA, 1987).

[14] J. Nocedal and S. J. Wright, eds., *Numerical Optimization* (Springer-Verlag, 1999).

[15] P. G. Anastasiou, N. J. Mayhall, E. Barnes, and S. E. Economou, How to really measure operator gradients in adapt-vqe (2023), arXiv:2306.03227 [quant-ph].