



Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows

Charles Hadfield¹, Sergey Bravyi¹, Rudy Raymond^{2,3}, Antonio Mezzacapo¹

¹ IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA.

E-mail: charles.hadfield@gmail.com; mezzacapo@ibm.com

² IBM Quantum, IBM Research – Tokyo, 19-21 Nihonbashi Chuo-ku, Tokyo 103-8510, Japan.

³ Quantum Computing Center, Keio University, Hiyoshi 3-14-1, Kohoku, Yokohama 223-8522, Japan.

Received: 17 August 2021 / Accepted: 27 September 2021

Published online: 30 March 2022 – © The Author(s) 2022

Abstract: Obtaining precise estimates of quantum observables is a crucial step of variational quantum algorithms. We consider the problem of estimating expectation values of quantum Hamiltonians, obtained on states prepared on a quantum computer. We propose a novel estimator for this task, which is locally optimised with knowledge of the Hamiltonian and a classical approximation to the underlying quantum state. Our estimator is based on the concept of classical shadows of a quantum state, and has the important property of not adding to the circuit depth for the state preparation. We test its performance numerically for molecular Hamiltonians of increasing size, finding a sizable reduction in variance with respect to current measurement protocols that do not increase circuit depths.

1. Introduction

Estimating observables of interest for quantum states prepared on a quantum processor is a central subroutine in a variety of quantum algorithms. Improving the precision of the measurement process is a pressing need, considering the fast-paced increase in size of current quantum devices. One key application is the energy estimation of complex molecular Hamiltonians, a staple of variational quantum eigensolvers (VQE) [1–4]. Readout of quantum information on quantum processors is available only through single-qubit projective measurements. The outcomes of these single-qubit measurements are combined to estimate quantum observables described by linear combinations of Pauli operators. Naively, each Pauli operator can be estimated independently by appending a quantum circuit composed of one layer of single-qubit gates at the end of state preparation, before readout.

A series of recent efforts [5–12] has shown that savings in the number of measurements can be obtained for the estimation of complex observables, at the expense of increasing circuit depths. This increase in circuit depth can defy the purpose of variational quantum algorithms, which aim to keep gate counts low [13].

Other strategies, more amenable to execution on near-term devices, have considered reducing the number of measurements while simultaneously not increasing circuit depth. These strategies are based on grouping together Pauli operators that can be measured in the same single-qubit basis. This Pauli grouping approach was introduced in [3] and explored thoroughly in [14] for chemistry systems. Machine learning techniques have also recently been used to tackle the measurement problem [15], with no increase in circuit depth. The machine learning approach is based upon the assumption that fermionic neural-network states can capture quantum correlations in ground states of molecular systems [16].

The measurement problem has been considered in the context of predicting collections of generic observables on reduced density matrices [17–21]. The best asymptotic scalings up to poly-logarithmic factors are obtained in [17], where it is proposed to characterise a quantum state through random measurements; the measurement outcomes are later used to evaluate expectation values of arbitrary observables.

In this article we introduce an estimator that recovers, in expectation, mean values of observables on quantum states prepared on quantum computers. The protocol is based on *classical shadows using random Pauli measurements* introduced in [17] and referred to as *classical shadows* in this present article. We show how sampling from random measurement bases in the original protocol can be locally biased towards certain bases on each individual qubit. We name this technique *locally-biased classical shadows*. We show how to optimise the estimator’s local bias on each qubit based on the knowledge of a target observable and a classical approximation of the quantum state, named *reference state*. We also prove that this optimisation has a convex cost function in certain regimes. We benchmark our optimisation procedure in the setting of quantum chemistry Hamiltonians, where reference states can be obtained from the Hartree–Fock solution, or multi-reference states, obtained with perturbation theory. We finally compare the variance of our estimator to previous methods for estimating average values that do not increase circuit depth, obtaining consistent improvements.

Outline of the paper. Section 2 reviews classical shadows using random Pauli measurements in a notation convenient to this current article. Section 3 provides the construction of the locally-biased classical shadows and calculates the expectation and variance associated with the estimator introduced. Section 4 shows how to optimise the estimator. Section 5 benchmarks our estimator for molecular energies on molecules of increasing sizes. Section 6 finishes with closing remarks. Appendix A reviews the methods for molecular energy estimation to which we compare our estimator.

2. Classical Shadows Using Random Pauli Measurements

Classical shadows using random Pauli measurements has been introduced in [17]. This section reproduces the procedure in a different style. Since we are only concerned with estimating one specific observable, we do not mention the aspect of a snapshot, nor the efficient description using the symplectic representation, nor the notion of median of means.

The problem that we want to address is the estimation of $\text{tr}(\rho O)$ for a given n -qubit state ρ and an observable O decomposed as a linear combination of Pauli terms:

$$O = \sum_{Q \in \{I, X, Y, Z\}^{\otimes n}} \alpha_Q Q \quad (1)$$

where $\alpha_Q \in \mathbb{R}$. Notationally, for a Pauli operator Q as above and for a given qubit $i \in \{1, 2, \dots, n\}$ we shall write Q_i for the i^{th} single-qubit Pauli operator so that $Q = \otimes_i Q_i$. We denote the support of such an operator $\text{supp}(Q) = \{i | Q_i \neq I\}$ and its weight $\text{wt}(Q) = |\text{supp}(Q)|$. An n -qubit Pauli operator Q is said to be full-weight if $\text{wt}(Q) = n$.

This task of estimating $\text{tr}(\rho O)$ is accomplished with classical shadows of [17] as described in Algorithm 1. Briefly, one randomly selects a Pauli basis for each of the n -qubits in which to measure the quantum state; this is irrespective of the operator O . Then, after measurement, non-zero estimates can be provided for all Pauli operators which qubit-wise commute with the measurement bases. All other Pauli operators are implicitly provided with the zero estimator for their expectation values.

We introduce the function from [17, Eq. E28]. For two n -qubit Pauli operators P, Q define, for each qubit i ,

$$f_i(P, Q) = \begin{cases} 1 & \text{if } P_i = I \text{ or } Q_i = I; \\ 3 & \text{if } P_i = Q_i \neq I; \\ 0 & \text{else.} \end{cases} \tag{2}$$

and extend this to the multi-qubit setting by declaring $f(P, Q) = \prod_{i=1}^n f_i(P, Q)$. Also, given a full-weight Pauli operator P , we let $\mu(P, i) \in \{\pm 1\}$ denote the eigenvalue measurement when qubit i is measured in the P_i basis. For a subset $A \subseteq \{1, 2, \dots, n\}$ declare

$$\mu(P, A) = \prod_{i \in A} \mu(P, i) \tag{3}$$

with the convention that $\mu(P, \emptyset) = 1$.

Algorithm 1 Estimation of observable via (uniform) classical shadows

```

for sample  $s \in \{1, 2, \dots, S\}$  do
    Prepare  $\rho$ ;
    Uniformly at random pick  $P \in \{X, Y, Z\}^{\otimes n}$ ;
    for qubit  $i \in \{1, 2, \dots, n\}$  do
        Measure qubit  $i$  in  $P_i$  basis providing eigenvalue measurement  $\mu(P, i) \in \{\pm 1\}$ ;
    Estimate observable expectation
    
```

$$v^{(s)} = \sum_Q \alpha_Q f(P, Q) \mu(P, \text{supp}(Q))$$

```

return  $v = \frac{1}{S} \sum_s v^{(s)}$ .
    
```

As shown in [17], the output of this algorithm is an unbiased estimator of the desired expectation value, that is, $\mathbb{E}(v) = \text{tr}(\rho O)$.

3. Locally-Biased Classical Shadows

In this section we generalise classical shadows by observing that the randomisation procedure of the Pauli measurements can be biased in the measurement basis for each

qubit. We build an estimator based on biased measurements, which in expectation recovers $\text{tr}(\rho O)$. We then proceed to calculate its variance.

As in Sect. 2, we wish to estimate $\text{tr}(\rho O)$ for a given state ρ and an observable $O = \sum_Q \alpha_Q Q$. For each qubit $i \in \{1, 2, \dots, n\}$, consider a probability distribution β_i over $\{X, Y, Z\}$ and denote by $\beta_i(P_i)$ the probability associated with each Pauli $P_i \in \{X, Y, Z\}$. We write β for the collection $\{\beta_i\}_{i=1}^n$ and note that β may be considered a probability distribution on full-weight Pauli operators by associating with $P \in \{X, Y, Z\}^{\otimes n}$ the probability $\beta(P) = \prod_i \beta_i(P_i)$.

We generalise the function introduced in Eq. (2). For two n -qubit Pauli operators P, Q and a product probability distribution β , define, for each qubit i ,

$$f_i(P, Q, \beta) = \begin{cases} 1 & \text{if } P_i = I \text{ or } Q_i = I; \\ (\beta_i(P_i))^{-1} & \text{if } P_i = Q_i \neq I; \\ 0 & \text{else.} \end{cases} \tag{4}$$

In this article, we will assume that $(\beta_i(P_i))^{-1} > 0$ in order to avoid any division-by-zero edge cases. (Note that such cases are irrelevant when we consider Sect. 4.) We extend this to the multi-qubit setting by declaring

$$f(P, Q, \beta) = \prod_{i=1}^n f_i(P, Q, \beta). \tag{5}$$

Algorithm 3 describes an estimator via locally-biased classical shadows. Note that the (uniform) classical shadows case is retrieved when $\beta_i(P_i) = \frac{1}{3}$ for every qubit $i \in \{1, 2, \dots, n\}$ and every Pauli term $P_i \in \{X, Y, Z\}$.

Algorithm 2 Estimation of observable via locally-biased classical shadows

```

for sample  $s \in \{1, 2, \dots, S\}$  do
  Prepare  $\rho$ ;
  for qubit  $i \in \{1, 2, \dots, n\}$  do
    Randomly pick  $P_i \in \{X, Y, Z\}$  from  $\beta_i$ -distribution;
    Measure qubit  $i$  in  $P_i$  basis providing eigenvalue measurement  $\mu(P, i) \in \{\pm 1\}$ ;
  Set  $P = \otimes_{i=1}^n P_i$ ;
  Estimate observable expectation
    
```

$$v^{(s)} = \sum_Q \alpha_Q f(P, Q, \beta) \mu(P, \text{supp}(Q))$$

```

return  $v = \frac{1}{S} \sum_s v^{(s)}$ .
    
```

Algorithm 3 recovers the expectation $\text{tr}(\rho O)$, as shown in the following lemma.

Lemma 1. *The estimator v from Algorithm 3 with a single sample ($S = 1$) satisfies*

$$\mathbb{E}(v) = \sum_Q \alpha_Q \text{tr}(\rho Q) \quad \text{and} \quad \mathbb{E}(v^2) = \sum_{Q,R} f(Q, R, \beta) \alpha_Q \alpha_R \text{tr}(\rho QR). \tag{6}$$

Proof. Let \mathbb{E}_P denote the expected value over the distribution $\beta(P)$. Let $\mathbb{E}_{\mu(P)}$ denote the expected value over the measurement outcomes for a fixed Pauli basis P . Using the fact that $\beta(P)$ is a product distribution one can easily check that

$$\mathbb{E}_P f(P, Q, \beta) = 1, \tag{7}$$

$$\mathbb{E}_P f(P, Q, \beta) f(P, R, \beta) = f(Q, R, \beta) \tag{8}$$

for any $Q, R \in \{I, X, Y, Z\}^{\otimes n}$.

Let us say that an n -qubit Pauli operator Q agrees with a basis $P \in \{X, Y, Z\}^{\otimes n}$ iff $Q_i \in \{I, P_i\}$ for any qubit i . Note that $f(P, Q, \beta) = 0$ unless Q agrees with P . For any n -qubit Pauli operators Q, R that agree with a basis P one has

$$\mathbb{E}_{\mu(P)} \mu(P, \text{supp}(Q)) = \text{tr}(\rho Q) \tag{9}$$

and

$$\mathbb{E}_{\mu(P)} \mu(P, \text{supp}(Q)) \mu(P, \text{supp}(R)) = \text{tr}(\rho QR). \tag{10}$$

To get the last equality, observe that $\mu(P, A) \mu(P, A') = \mu(P, A \ominus A')$ for any subsets of qubits A, A' , where $A \ominus A'$ is the symmetric difference of A and A' . The assumption that both Q and R agree with the same basis P implies that $\text{supp}(Q) \ominus \text{supp}(R) = \text{supp}(QR)$. Now Eq. (10) follows from Eq. (9).

By definition, the expected value in Eq. (6) is a composition of the expected values over a Pauli basis P and over the measurement outcomes $\mu(P)$, that is, $\mathbb{E} = \mathbb{E}_P \mathbb{E}_{\mu(P)}$. Using the above identities one gets

$$\mathbb{E}(v) = \mathbb{E}_P \mathbb{E}_{\mu(P)} v = \sum_Q \alpha_Q \text{tr}(\rho Q) \mathbb{E}_P f(P, Q, \beta) = \sum_Q \alpha_Q \text{tr}(\rho Q)$$

Here the second equality is obtained using Eq. (9) and the linearity of expected values. The third equality follows from Eq. (7). Likewise,

$$\begin{aligned} \mathbb{E}(v^2) &= \mathbb{E}_P \mathbb{E}_{\mu(P)} v^2 = \sum_{Q,R} \alpha_Q \alpha_R \text{tr}(\rho QR) \mathbb{E}_P f(P, Q, \beta) f(P, R, \beta) \\ &= \sum_{Q,R} \alpha_Q \alpha_R f(Q, R, \beta) \text{tr}(\rho QR). \end{aligned}$$

Here the second equality is obtained using Eq. (10) and observing that that $f(P, Q, \beta) f(P, R, \beta) = 0$ unless both Q and R agree with P . The third equality follows from Eq. (8). \square

Recall that in the context of using a quantum processor, we aim to use the random variable v to estimate $\text{tr}(\rho O)$ to some (additive) precision ε . This dictates the number of samples S required. Specifically, for fixed ρ, O , we require $S = O(\varepsilon^{-2} \text{Var}(v^{(s)}))$ where $\text{Var}(v^{(s)})$ is obviously independent of the specific sample s . For future reference, we record explicitly the variance of v for a single sample ($S = 1$). Lemma 1 establishes

$$\text{Var}(v) = \left(\sum_{Q,R} f(Q, R, \beta) \alpha_Q \alpha_R \text{tr}(\rho QR) \right) - (\text{tr}(\rho O))^2. \tag{11}$$

Remark 1. In [17, Proposition 3], the authors aim to upper-bound this variance independently of the state ρ . In the uniform setting, this is achieved with an application of Cauchy-Schwarz and it leads to a bound of $4^k \|O\|_\infty^2$ where k is the weight of the operator.

4. Optimised Locally-Biased Classical Shadows

In this section we show how the locally-biased classical shadows introduced in Sect. 3 can be optimised when one has partial knowledge about the underlying quantum state. This partial information is obtained efficiently with a classical computation. This is the case of VQE for molecular Hamiltonians if one initialises the variational procedure from a reference state, which can be, for example, the Hartree–Fock solution, a generic fermionic Gaussian state [22], or perturbative Møller–Plesset solutions. Our method can be also extended to generic many-body Hamiltonians, considering for example the generation of reference states with semidefinite programming [23]. On a more general note, the existence of a good reference state is the assumption of all algorithms that target ground state properties of interacting many-body problems, including quantum phase estimation.

In this setting, we optimise the probability distributions $\beta = \{\beta_i\}_{i=1}^n$ to obtain the smallest variance on a given reference state. To do this, we consider the variance calculated in Eq. (11) and extract from it the component which, associated with the reference state, explicitly depends on the distributions β . We proceed to optimise this cost function, thereby minimising the variance, noting that a negligible restriction of the cost function that we use leads to a convex optimisation problem. Finally, we use the optimised distributions β^* to build molecular energy estimators as defined in Algorithm 3.

To set notation, we introduce a molecular Hamiltonian, H , acting on n qubits. We write

$$H = \sum_{P \in \{I, X, Y, Z\}^{\otimes n}} \alpha_P P \quad (12)$$

and denote by H_0 the traceless part of H .

4.1. Single-reference optimisation. We first consider the case in which the reference state is a product state of the form $\frac{1}{2^n} \otimes_{i=1}^n (I + m_i P_i)$ where $m_i \in \{\pm 1\}$ and $P_i \in \{X, Y, Z\}$. This is the case if a VQE targeting a molecular Hamiltonian in the molecular basis is initialised with the Hartree–Fock state. In this case $P_i = Z$. Motivated by this, we use the label “HF” and, for ease of reading, we assume that $P_i = Z$. However we remark that the results here can be generalised out of the quantum chemistry domain.

We are given a reference product state $\rho_{\text{HF}} = \frac{1}{2^n} \otimes_{i=1}^n (I + m_i Z)$ where $m_i \in \{\pm 1\}$. The variance of the estimator v is independent of the constant term $H - H_0$. Writing the variance from Eq. (11) for the state ρ_{HF} upon explicit removal of the constant term reads

$$\text{Var}(v|\rho_{\text{HF}}) = \sum_{Q \neq I} \sum_{R \neq I} f(Q, R, \beta) \alpha_Q \alpha_R \text{tr}(\rho_{\text{HF}} QR) - \text{tr}(\rho_{\text{HF}} H_0)^2. \quad (13)$$

Our objective is to find probability distributions β so as to minimise Eq. (13). The following proposition explicits the relevant cost function appropriate to this task.

Proposition 1. *Given a reference product state ρ_{HF} , represented by the logical basis element $\{m_i\}_{i=1}^n$, the variance associated with Algorithm 3 is minimised upon choosing β so as to minimise*

$$\text{cost}(\beta) = \sum_{(Q, R) \in \mathcal{I}_{Z^{\otimes n}}} \alpha_Q \alpha_R \prod_{i|Q_i=R_i \neq I} (\beta_i(Q_i))^{-1} \prod_{i|Q_i \neq R_i} m_i \quad (14)$$

subject to $\beta_{i,P} \geq 0$ and $\beta_{i,X} + \beta_{i,Y} + \beta_{i,Z} = 1$ for all i . In the above, the sum is taken over “influential pairs”:

$$\mathcal{I}_{Z^{\otimes n}} = \{(Q, R) \mid Q, R \neq I^{\otimes n} \text{ and for all } i, \text{ either } Q_i = R_i, \text{ or } \{Q_i, R_i\} = \{I, Z\}\} \quad (15)$$

Proof. We must pay attention to only β -dependent terms in $\text{Var}(v \mid \rho_{\text{HF}})$. The simple structure of ρ_{HF} implies, for n -qubit non-identity Pauli operators Q, R ,

$$\begin{aligned} f(Q, R, \beta) \text{tr}(\rho_{\text{HF}} Q R) &= \prod_{i=1}^n f_i(Q, R, \beta) \text{tr} \left(\frac{1}{2} (I + m_i Z) Q_i R_i \right) \\ &= \prod_{i=1}^n f_i(Q, R, \beta) \delta_{Q_i, R_i} + m_i (\delta_{Q_i, Z} \delta_{R_i, I} + \delta_{Q_i, I} \delta_{R_i, Z}) \\ &= \prod_{i=1}^n \delta_{Q_i, R_i} \left(\delta_{Q_i, I} + (1 - \delta_{Q_i, I}) (\beta_i(Q_i))^{-1} \right) \\ &\quad + m_i (\delta_{Q_i, Z} \delta_{R_i, I} + \delta_{Q_i, I} \delta_{R_i, Z}) \end{aligned}$$

The preceding display is independent of β whenever $(Q, R) \notin \mathcal{I}_{Z^{\otimes n}}$. Hence the cost function captures precisely the component of the variance (when estimating the reference product state) which is dependent on the probability distributions β . \square

Some remarks are in order.

Remark 2. The cost function of Eq. (14) is *not* convex. If we however restrict to diagonal terms from the set of influential pairs, then we obtain the following alternative cost function, which we refer to as the *diagonal* cost function:

$$\text{cost}_{\text{diag}}(\beta) = \sum_Q \alpha_Q^2 \prod_{i \in \text{supp}(Q)} (\beta_i(Q_i))^{-1} \quad (16)$$

This diagonal cost function is convex: For fixed Q , the function $-\log \beta_i(Q_i)$ is convex, hence so too is $\sum_{i \in \text{supp}(Q)} (-\log(\beta_i(Q_i)))$. Exponentiating this result implies $\prod_{i \in \text{supp}(Q)} \beta_i(Q_i)^{-1}$ is convex. The positive linear combination over Pauli operators Q preserves convexity. In this convex case we are assured that the minimised collection of distributions provides a global minimum (of the diagonal cost function).

This diagonal cost function makes no reference to the specific single-reference state, and therefore can be used to find β_i which are independent of the underlying quantum state ρ_{HF} . In fact, the diagonal cost function can be derived from Eq. (11) when ρ is the maximally mixed state. Since the diagonal cost function is convex, we may be confident that the values for β found numerically are optimal with respect to the cost function. In the following section, the numerical experiments indicate surprisingly that the diagonal cost function leads to probability distributions which give very satisfying results irrespective of the diagonal cost function’s lack of reference to the ground state, or a state near the ground state. We explain this in Remark 4.

Remark 3. The diagonal cost function can be formulated in the language of geometric programming [24], while the original cost function is an example of signomial geometric programming. They can be numerically optimised using standard techniques such as iterative updates incorporating closed-form solutions obtained from Lagrange multipliers. See Appendix B for more details.

Remark 4. We provide a bound on the original cost function from that on the diagonal cost function. Set $\Gamma(Q, R) = |\alpha_Q| |\alpha_R| \prod_i |\alpha_{Q_i=R_i}(\beta_i(Q_i))|^{-1}$. The diagonal cost function is $\sum_Q \Gamma(Q, Q)$, while the original cost function is at most $\sum_{Q,R} \Gamma(Q, R)$, (the sum is over only influential pairs in the original cost function). Using that $2xy \leq x^2 + y^2$ for real x, y , the following inequality holds

$$\Gamma(Q, R) \leq \frac{\Gamma(Q, Q) + \Gamma(R, R)}{2}. \quad (17)$$

Summation over all pairs (Q, R) on both sides leads to

$$\sum_{Q,R} \Gamma(Q, R) \leq |H_0| \sum_Q \Gamma(Q, Q) \quad (18)$$

where $|H_0|$ is the number of traceless terms in the Hamiltonian.

It is perhaps less surprising than one might first expect that the diagonal cost function performs well for quantum chemistry systems. This is because of the structure of the problem at hand. For chemistry systems one has that Hamiltonian coefficients corresponding to integrals with fewer orbitals are larger than ones corresponding to many orbitals. This is because the overlapping spatial region between orbitals decreases with their number. Considering that in our cost function the Hamiltonian coefficients are squared, it is not surprising that coefficients corresponding to one-orbital integrals and two-orbital integrals dominate the total cost.

Remark 5. It is possible to obtain a rough confidence bound on the number of shots required to get an accurate estimate of $\text{tr}(\rho H)$ given an unknown ρ in the context of quantum chemistry. Recall Algorithm 3 samples the state ρ a total of S times and returns an estimate $v = \frac{1}{S} \sum_s v^{(s)}$. Consider an optimised β^* according to the diagonal cost function. Then the single-shot variance of the energy estimator $v^{(s)}$, given ρ and β^* , may be bounded

$$\begin{aligned} \text{Var}(v^{(s)} | \rho, \beta^*) &\leq \sum_{Q,R} f(Q, R, \beta^*) \alpha_Q \alpha_R \text{tr}(\rho QR) \leq \sum_{Q,R} f(Q, R, \beta^*) |\alpha_Q| |\alpha_R| \\ &\leq \sum_{Q,R} \Gamma(Q, R) \leq |H_0| \cdot \text{cost}_{\text{diag}}(\beta^*) \end{aligned} \quad (19)$$

where we have reintroduced the notation of the preceding remark and applied Eq. (18). Using Chebyshev's inequality, we can bound the probability that our final estimate v is worse than additive-error ε from $\text{tr}(\rho H)$

$$\mathbb{P}(|v - \text{tr}(\rho H)| \geq \varepsilon) \leq \frac{1}{S\varepsilon^2} |H_0| \cdot \text{cost}_{\text{diag}}(\beta^*) \quad (20)$$

and we recall that $|H_0| = O(n^4)$ for the quantum chemistry setting described in the following section.

4.2. *Multi-reference optimisation.* We finish this section by observing that the technique of optimising the probability distributions also works for multi-reference frame states such as fermionic Gaussian states, or perturbative solutions. Such reference states are important for quantum chemistry simulations far from equilibrium (unlike the Hartree–Fock reference state which is more appropriate near equilibrium). Specifically, consider a multi-reference state $|\psi\rangle_{\text{MR}}$, written in the logical basis

$$|\psi\rangle_{\text{MR}} = \sum_{k=1}^K \lambda_k |\psi^{(k)}\rangle, \quad |\psi^{(k)}\rangle = |b_1^{(k)} \dots b_n^{(k)}\rangle \quad (21)$$

where $b_i^{(k)} \in \{0, 1\}$ are associated with Z -eigenvalues $m_i^{(k)} = (-1)^{b_i^{(k)}}$ and $\lambda_k \in \mathbb{C}$ are amplitudes such that $|\psi\rangle_{\text{MR}}$ is normalised. The associated density now reads

$$\rho_{\text{MR}} = \sum_{k,\ell} \lambda_k \bar{\lambda}_\ell \rho^{(k,\ell)}, \quad \rho^{(k,\ell)} = \otimes_{i=1}^n |b_i^{(k)}\rangle \langle b_i^{(\ell)}| \quad (22)$$

In the following paragraphs, we calculate an appropriate cost function for this case.

Let us restrict ourselves to the single-qubit setting briefly: $\rho^{(k,\ell)} = |b^{(k)}\rangle \langle b^{(\ell)}|$. There are two cases for $\rho^{(k,\ell)}$ dependent on whether $b^{(k)}$, $b^{(\ell)}$ agree or not. If they agree then $\rho^{(k,\ell)} = \frac{1}{2}(I + (-1)^{b^{(k)}} Z)$. If they disagree, then $\rho^{(k,\ell)} = \frac{1}{2}(X + (-1)^{b^{(k)}} iY)$. In a similar way to the single-reference setting, we need to calculate $f(Q, R, \beta) \text{tr}(\rho^{(k,\ell)} QR)$. This is best done by considering the two cases: We introduce the function g when $b^{(k)} = b^{(\ell)}$ and obtain

$$\begin{aligned} f(Q, R, \beta) \text{tr}(\rho^{(k,\ell)} QR) &= \delta_{Q,R} \left(\delta_{Q,I} + (1 - \delta_{Q,I}) \beta_Q^{-1} \right) + (-1)^{b^{(k)}} (\delta_{Q,Z} \delta_{R,I} + \delta_{Q,I} \delta_{R,Z}) \\ &= g(Q, R, \beta, b^{(k)}); \end{aligned}$$

We introduce the function h when $b^{(k)} \neq b^{(\ell)}$ and obtain

$$\begin{aligned} f(Q, R, \beta) \text{tr}(\rho^{(k,\ell)} QR) &= (\delta_{Q,X} \delta_{R,I} + \delta_{Q,I} \delta_{R,X}) + (-1)^{b^{(k)}} i (\delta_{Q,Y} \delta_{R,I} + \delta_{Q,I} \delta_{R,Y}) \\ &= h(Q, R, b^{(k)}). \end{aligned}$$

We can now return to the multi-qubit setting to write down a cost function which ought be minimised:

$$\begin{aligned} \text{cost}_{\text{multi-ref}}(\beta) &= \sum_{k,\ell} \lambda_k \bar{\lambda}_\ell \sum_{Q,R} \alpha_Q \alpha_R \left(\prod_{i|b_i^{(k)}=b_i^{(\ell)}} g(Q_i, R_i, \beta_i, b_i^{(k)}) \prod_{i|b_i^{(k)} \neq b_i^{(\ell)}} h(Q_i, R_i, b_i^{(k)}) \right) \quad (23) \end{aligned}$$

We leave as future work a thorough investigation of cost functions associated with non-equilibrium quantum chemistry.

5. Numerical Experiments on Molecular Hamiltonians

In this section we test numerically the locally-biased classical shadows (LBCS) estimator defined in Algorithm 3 for molecular Hamiltonians. We consider six Hamiltonians corresponding to different molecules, represented in a minimal STO-3G basis, ranging from 4 to 16 spin orbitals. (The 8 qubit H_2 example uses a 6-31G basis.) We map the molecular Hamiltonians to qubit ones, using three encodings detailed in [25]. The result is qubit Hamiltonians defined on up to 16 qubits. The molecular Hamiltonians are defined in the molecular basis. In this basis, the Hartree–Fock state is a computational basis state. We choose the Hartree–Fock state as our single-reference state, and optimise the distributions β according to Eqs. (14) and (16) separately. We call the optimisation procedure of the β according to Eq. (16) *diagonal*. We then use the optimised β^* to compute the variance Eq. (11) on the ground state of the molecular Hamiltonians; the ground state and the ground energy are obtained by the Lanczos method for sparse matrices. Recall that, for a fixed accuracy with which one aims to estimate the energy, the variance is proportional to the number of preparations and measurements of the state one is required to perform. It may therefore be viewed as a measuring stick for the run-time of a VQE algorithm. We report the results in Table 1. In this table, we compare variances obtained with our LBCS estimator against other previously known observable estimators that do not increase circuit depth:

- An estimator based on ℓ^1 sampling of the Hamiltonian, detailed in [26,27].
- An estimator which measures together collections of qubit-wise commuting Pauli operators. To find the collections of Pauli operators, we use a largest degree first (LDF) heuristic [28]. The collections are then sampled according to their Hamiltonian ℓ^1 weights.
- Classical shadows as given in [17], which corresponds to the case $\beta_i(P_i) = \frac{1}{3}$ for any qubit i and Pauli term $P_i \in \{X, Y, Z\}$.

Details of the first two estimators may be found in Appendix A.¹ For all the estimators, we report variances exactly computed on the ground states of the Hamiltonians considered.

In all but one experiment of Table 1, we observe that the LBCS estimator outperforms the other estimators. The one case where the LDF decomposition provides a lower variance— H_2 on a minimal basis—should be considered a curiosity due to the small qubit count.

The two different cost functions used to optimise the β -distributions provide very similar variance. This is remarkable considering that the diagonal cost function defined in Eq. (16) is convex. For any given molecule, our numerical analysis indicated that the non-convex cost function Eq. (14) always converged to a single collection of distributions, irrespective of the initialised values for the distributions.

Next, we plot in Fig. 1 an optimised distribution β^* . Specifically, we take the example of H_2O on 14 qubits in the Jordan–Wigner encoding. Due to the symmetry [25] where the first 7 qubits correspond to spin-up orbitals, and the last 7 qubits correspond to spin-down orbitals, we observe that $\beta_i^* = \beta_{i+7}^*$ for $i \in \{1, 2, \dots, 7\}$. Note also that the probabilities are symmetric in X and Y (which is not the case for the Bravyi–Kitaev encoding).

Finally, we analyse the role played by the specific fermionic encoding used. For a restricted set of Hamiltonians, Table 2 reports variances for the three estimators: LDF grouping; classical shadows; and LBCS, with three different fermion-to-qubit encodings:

¹ Code is available upon request.

Table 1. Variance for various estimators considered in this work

| Molecule | Qubits | Estimator | Variance |
|------------------|--------|-------------------------------|----------|
| H ₂ | 4 | ℓ^1 sampling | 2.49 |
| | | LDF grouping | 0.402 |
| | | Classical shadows | 1.97 |
| | | LBCS | 1.86 |
| | | LBCS (diagonal cost function) | 1.86 |
| H ₂ | 8 | ℓ^1 sampling | 120 |
| | | LDF grouping | 22.3 |
| | | Classical shadows | 51.4 |
| | | LBCS | 17.5 |
| | | LBCS (diagonal cost function) | 17.7 |
| LiH | 12 | ℓ^1 sampling | 138 |
| | | LDF grouping | 54.2 |
| | | Classical shadows | 266 |
| | | LBCS | 14.8 |
| | | LBCS (diagonal cost function) | 14.8 |
| BeH ₂ | 14 | ℓ^1 sampling | 418 |
| | | LDF grouping | 135 |
| | | Classical shadows | 1670 |
| | | LBCS | 67.6 |
| | | LBCS (diagonal cost function) | 67.6 |
| H ₂ O | 14 | ℓ^1 sampling | 4360 |
| | | LDF grouping | 1040 |
| | | Classical shadows | 2840 |
| | | LBCS | 257 |
| | | LBCS (diagonal cost function) | 257 |
| NH ₃ | 16 | ℓ^1 sampling | 3930 |
| | | LDF grouping | 891 |
| | | Classical shadows | 14400 |
| | | LBCS | 353 |
| | | LBCS (diagonal cost function) | 353 |

LBCS is optimised according to Eq. (14), while the diagonal cost function is defined in Eq. (16)

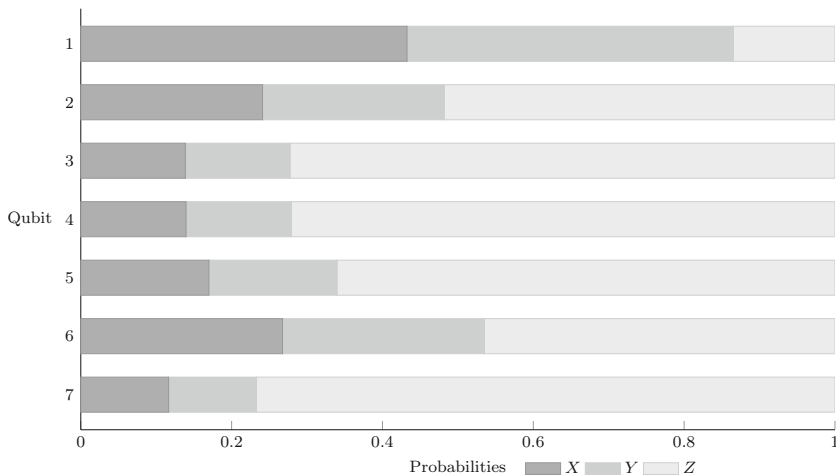


Fig. 1. Probability distributions over the first 7 of 14 qubits for H₂O Hamiltonian using the Jordan–Wigner encoding. The probability distributions have been optimised according to Eq. (16)

Table 2. Variance for different estimators computed on the ground states of the molecules indicated

| Molecule | Encoding | Variance | | |
|---------------------------|----------|--------------|-------------------|-------|
| | | LDF grouping | Classical shadows | LBCS |
| H ₂ (4 qubits) | JW | 0.402 | 1.97 | 1.86 |
| | Parity | 0.193 | 4.00 | 0.541 |
| | BK | 0.193 | 10.0 | 0.541 |
| H ₂ (8 qubits) | JW | 22.3 | 51.4 | 17.7 |
| | Parity | 38.0 | 70.8 | 18.9 |
| | BK | 38.4 | 169 | 19.5 |
| LiH | JW | 54.2 | 266 | 14.8 |
| | Parity | 85.8 | 760 | 26.5 |
| | BK | 75.5 | 163 | 68.0 |
| BeH ₂ | JW | 135 | 1670 | 67.6 |
| | Parity | 239 | 3160 | 130 |
| | BK | 197 | 947 | 238 |
| H ₂ O | JW | 1040 | 2840 | 258 |
| | Parity | 2670 | 6380 | 429 |
| | BK | 2090 | 10,600 | 1360 |

LBCS is optimized with the cost function defined in Eq. (16)

Jordan–Wigner; parity; and Bravyi–Kitaev. Note that the variances for parity and Bravyi–Kitaev mappings are higher because those mappings generate Pauli distributions that tend to have more X and Y operators, as opposed to the linear tail of Z operators of the Jordan–Wigner, against which the distributions β can be easily biased. We do not report ℓ^1 sampling in Table 2, as it is invariant under choice of encoding. Irrespective of the encoding, the locally biased classical shadows shows a reduction in variance over the LDF grouping whose collections are sampled according to their 1-norm.

6. Conclusion

This article has considered the measurement problem associated with molecular energy estimation on quantum computers and has proposed a new algorithm for that problem. Investigating the principal subroutine present in *classical shadows using random Pauli measurements*, we are able to produce a non-uniform version of these shadows, termed *locally-biased classical shadows*. These locally biased classical shadows require probability distributions for each qubit. By solving a convex optimisation problem for a given molecular Hamiltonian, we find appropriate probability distributions for measuring states which are close to the true ground state of the molecular Hamiltonian. We benchmark the proposed algorithm on systems up to 16 qubits in size and observe significant and consistent improvement over Pauli grouping heuristic algorithms. To claim this improvement over Pauli grouping heuristic algorithms we have benchmarked our algorithm against the LDF heuristic. Note that Ref. [14] finds that other heuristics produce a number of qubit-wise commuting sets that only differ by 10% and are expected to offer similar performance to the LDF heuristic. Moreover standard Pauli grouping techniques are based on heuristically solving a computationally hard problem, whereas our method requires simply solving a convex optimisation problem whose number of variables is linear in the number of qubits.

Finally, the introduction of such a domain-specific cost function is, to the authors’ knowledge, novel. It is sufficiently general that applications of this idea will also be relevant in fields unrelated to quantum chemistry.

Acknowledgements We thank Giacomo Nannicini for useful discussions regarding the convexity of the cost functions introduced here. SB acknowledges the support of the IBM Research Frontiers Institute. We thank the anonymous referees for their helpful comments which have improved the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A: Comparative Algorithms for Estimating Molecular Hamiltonians

This appendix provides details for the two algorithms against which we benchmark locally-biased classical shadows. Recall that we assume that the molecular Hamiltonian H acts on n -qubits and that a given state ρ is provided and whose energy we aim to estimate. We write

$$H = \sum_{P \in \{I, X, Y, Z\}^{\otimes n}} \alpha_P P \quad (24)$$

and denote by H_0 the traceless part of H . Denote by $\|\alpha\|_{\ell^1}$ the ℓ^1 -norm of the traceless coefficients, and associate with this norm the following ℓ^1 -distribution γ over the Pauli operators:

$$\|\alpha\|_{\ell^1} = \sum_{P \in \{I, X, Y, Z\}^{\otimes n} \setminus \{I^{\otimes n}\}} |\alpha_P| \quad \gamma(P) = \frac{1}{\|\alpha\|_{\ell^1}} |\alpha_P| \quad (25)$$

We expose the dependence of the algorithms on the identity coefficient $\alpha_{I^{\otimes n}}$. This is because in the practical setting of molecular Hamiltonians considered in this text, the identity coefficient can be on the order of 10% of $\|\alpha\|_{\ell^1}$. For the ℓ^1 sampling, it would be unwise to prepare ρ only to subsequently measure no qubits. For the largest degree first setting, it would be unwise to arbitrarily associate the identity operator to one of the collections of qubit-wise commuting Pauli operators thereby associating the identity operator's weight $|\alpha_{I^{\otimes n}}|$ to the corresponding collection's weight and overly favouring the sampling of said collection.

Recall the notation from Sect. 2. Given a Pauli operator P , we let $\mu(P, i) \in \{\pm 1\}$ denote the eigenvalue measurement when qubit i is measured in the P_i basis. For a subset $A \subseteq \{1, 2, \dots, n\}$ we write $\mu(P, A) = \prod_{i \in A} \mu(P, i)$.

A.1. Ell-1 algorithm. This algorithm was the first algorithm proposed for estimating energies in the context of variational quantum algorithms [26]. The ℓ^1 -norm of the traceless coefficients provides the probability distribution γ . We may use this probability distribution to select a Pauli operator P which dictates the Pauli basis in which to measure the state ρ , thereby providing an estimate for $\text{tr}(\rho P)$. Algorithm 4 describes this procedure precisely.

Algorithm 3 Estimation of observable via locally-biased classical shadows

for sample $s \in \{1, 2, \dots, S\}$ **do**
 Prepare ρ ;
for qubit $i \in \{1, 2, \dots, n\}$ **do**
 Randomly pick $P_i \in \{X, Y, Z\}$ from β_i -distribution;
 Measure qubit i in P_i basis providing eigenvalue measurement $\mu(P, i) \in \{\pm 1\}$;
 Set $P = \otimes_{i=1}^n P_i$;
 Estimate observable expectation

$$v^{(s)} = \sum_Q \alpha_Q f(P, Q, \beta) \mu(P, \text{supp}(Q))$$

return $v = \frac{1}{S} \sum_s v^{(s)}$.

For completeness, we record calculations for the expectation and variance of this estimator. Consider a single shot giving v . Let \mathbb{E}_P denote the expected value over the distribution $\gamma(P)$ and let $\mathbb{E}_{\mu(P)}$ denote the expected value over the measurement outcomes for a fixed Pauli operator P . Without loss of generality, we may assume $\alpha_{I^{\otimes n}} = 0$. Now $\mathbb{E}_{\mu(P)} \mu(P, \text{supp}(P)) = \text{tr}(\rho P)$ whence

$$\mathbb{E}(v) = \mathbb{E}_P \mathbb{E}_{\mu(P)} v = \mathbb{E}_P \|\alpha\|_{\ell^1} \text{sgn}(\alpha_P) \text{tr}(\rho P) = \sum_P \alpha_P \text{tr}(\rho P) = \text{tr}(\rho H). \quad (26)$$

The variance (for a single sample) can also be calculated:

$$\text{Var}(v) = \sum_{P \neq I^{\otimes n}} (\gamma_P \cdot \|\alpha\|_{\ell^1}^2) - \text{tr}(\rho H_0)^2 = \|\alpha\|_{\ell^1}^2 - \text{tr}(\rho H_0)^2. \quad (27)$$

A.2. Largest degree first. Consider a Hamiltonian decomposed into K collections $\{C^{(k)}\}_{k=1}^K$ excluding the identity term: $H = \alpha_{I^{\otimes n}} I^{\otimes n} + \sum_{k=1}^K H_k$ where $H_k = \sum_{Q \in C^{(k)}} \alpha_Q Q$. (Recalling the notation H_0 for the traceless part of the Hamiltonian, we note that $H_0 = \sum_k H_k$.) Suppose that for each collection $C^{(k)}$, the Pauli terms commute qubit-wise: for all $Q, R \in C^{(k)}$ and all qubits i , we have $[Q_i, R_i] = 0$. In this case, there exists a Pauli operator $P^{(k)}$ of weight n which commutes qubit-wise with each Pauli in $C^{(k)}$.

Consider also a probability distribution κ over the collections $\{C^{(k)}\}_{k=1}^K$. Sampling from this distribution provides Algorithm 4.

Consider a single sample giving an estimator v . Similar to the ℓ^1 algorithm we observe that v recovers $\text{tr}(\rho H)$ in expectation. Specifically, for a fixed collection $C^{(k)}$ and hence a fixed full-weight Pauli operator $P^{(k)}$, let $\mathbb{E}_{\mu(P^{(k)})}$ denote the expected value over the measurement outcomes associated with $P^{(k)}$. Now $\mathbb{E}_{\mu(P^{(k)})} \mu(P^{(k)}, \text{supp}(Q)) = \text{tr}(\rho Q)$

Algorithm 4 Energy estimation via ℓ^1 -distribution over Pauli bases

for sample $s \in \{1, \dots, S\}$ **do**
 Prepare ρ ;
 Randomly pick P from γ -distribution;
for qubit $i \in \text{supp}(P)$ **do**
 Measure qubit i in P_i basis providing eigenvalue measurement $\mu(P, i) \in \{\pm 1\}$;
 Estimate observable expectation

$$v^{(s)} = \alpha_{I^{\otimes n}} + \|\alpha\|_{\ell^1} \cdot \text{sgn}(\alpha_P) \cdot \mu(P, \text{supp}(P))$$

return $v = \frac{1}{S} \sum_s v^{(s)}$.

whenever $Q \in C^{(k)}$ and if we let $\mathbb{E}_{C^{(k)}}$ denote the expected value over the distribution $\kappa(C^{(k)})$ we conclude

$$\mathbb{E}(v) = \mathbb{E}_{C^{(k)}} \mathbb{E}_{\mu(P^{(k)})} v = \mathbb{E}_{C^{(k)}} \frac{1}{\kappa(C^{(k)})} \sum_{Q \in C^{(k)}} \alpha_Q \text{tr}(\rho Q) = \text{tr}(\rho H) \quad (28)$$

Again, we have assumed without loss of generality that H is traceless.

The variance may be calculated as

$$\text{Var}(v) = \left(\sum_{k=1}^K \frac{1}{\kappa(C^{(k)})} \sum_{Q, R \in C^{(k)}} \alpha_Q \alpha_R \prod_{i \in \text{supp}(QR)} \text{tr}(\rho QR) \right) - \text{tr}(\rho H_0)^2 \quad (29)$$

An alternative formula reads [3, Appendix A]

$$\text{Var}(v) = \sum_{k=1}^K \frac{1}{\kappa(C^{(k)})} \sum_{Q, R \in C^{(k)}} \alpha_Q \alpha_R (\text{tr}(\rho QR) - \text{tr}(\rho Q) \text{tr}(\rho R)) \quad (30)$$

Our analysis uses the LDF heuristics in order to obtain such a decomposition. Various heuristics for building decompositions are investigated in [14] for systems up to 36 qubits. The heuristics give numbers of groups that differ by 10% and they conclude LDF is attractive due to its short runtime. For the LDF decomposition, we first construct a graph $G = (V, E)$ where:

- $v_Q \in V$ for all $Q \neq I^{\otimes n}$ such that $\alpha_Q \neq 0$;
- $e_{Q,R} \in E$ if $\{Q_i, R_i\} = 0$ for some qubit i .

Second, the vertices of the graph are sorted in decreasing order of their degrees, and the smallest available colour is then progressively assigned to each ordered vertex. Colours correspond to collections in which Pauli operators commute qubit-wise. The LDF heuristics guarantee the number of colours of the graph is at most one plus the degree of the graph: $K \leq 1 + \Delta(G)$. With this decomposition constructed, our analysis is done with the following choice for κ :

$$\kappa(C^{(k)}) = \frac{\|\alpha|_{C^{(k)}}\|_{\ell^1}}{\|\alpha\|_{\ell^1}} \qquad \|\alpha|_{C^{(k)}}\|_{\ell^1} = \sum_{Q \in C^{(k)}} |\alpha_Q| \quad (31)$$

Qiskit [29] provides an implementation of the decomposition procedure.²

² https://qiskit.org/documentation/_modules/qiskit/aqua/operators/legacy/pauli_graph.html#PauliGraph. Last accessed on June 6, 2020

Appendix B: Optimisation of Cost Functions

In Sect. 4 we introduced two cost functions. Specifically, a non-convex cost function in Eq. (14) which requires knowledge of the Hamiltonian H and an arbitrary computational basis state, and a convex “diagonal” cost function in Eq. (16) which requires knowledge of the Hamiltonian only.

We solve these optimisation problems using the method of iterative updates incorporating closed-form solutions obtained with Lagrange multipliers. Specifically given current values $\beta^{(t)}(P)$ and update step-size $\Delta \in (0, 1)$, we may update iteratively:

$$\beta^{(t+1)}(P) = (1 - \Delta) \beta^{(t)}(P) + \Delta \beta^{\text{closed}}(P) \quad (32)$$

where the closed-form equations (detailed below) are obtained by solving the optimization of the cost function with Lagrange multipliers to obtain conditions for all $\beta_i(P_i)$ which must hold at optimality. The closed-form equations for the diagonal cost function of Eq. (16) are

$$\beta_i(P_i) = \frac{\sum_{Q|Q_i=P_i} \alpha_Q^2 \prod_{j \in \text{supp}(Q)} \beta_j(Q_j)^{-1}}{\sum_{Q|Q_i \neq I} \alpha_Q^2 \prod_{j \in \text{supp}(Q)} \beta_j(Q_j)^{-1}} \quad (33)$$

while for the original cost function of Eq. (14), they read

$$\beta_i(P_i) = \frac{\sum_{(Q,R) \in \mathcal{I}_{Z^{\otimes n}} | Q_i=R_i=P_i} \alpha_Q \alpha_R \prod_{j|Q_j=R_j \neq I} \beta_j(Q_j)^{-1} \prod_{j|Q_j \neq R_j} m_j}{\sum_{(Q,R) \in \mathcal{I}_{Z^{\otimes n}} | Q_i=R_i \neq I} \alpha_Q \alpha_R \prod_{j|Q_j=R_j \neq I} \beta_j(Q_j)^{-1} \prod_{j|Q_j \neq R_j} m_j}. \quad (34)$$

We can confirm that if Eq. (32) converges for some $t = T$ and $\Delta \in (0, 1)$, then the values $\beta^{(T)}(P)$ at convergence are the closed-form equations. This is because by arranging the terms of the equation, we have

$$\left| \beta^{(T+1)}(P) - \beta^{(T)}(P) \right| = \Delta \left| \beta^{(T)}(P) - \beta^{\text{closed}}(P) \right|, \quad (35)$$

where the left-hand side of the above equation approaches zero at the convergence. Also, at every iterative update the constraints (that β is a product probability distribution) are always satisfied whenever initialisation occurs with a random collection of probability distributions. Finally, we point out the use of the update step-size Δ in Eq. (32) was introduced in [30] where it is called smoothed updating.

References

1. Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.-H., Zhou, X.-Q., Love, P.J., Aspuru-Guzik, A., O’Brien, J.L.: A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014)
2. O’Malley, P.J.J., Babbush, R., Kivlichan, I.D., Romero, J., McClean, J.R., Barends, R., Kelly, J., Roushan, P., Tranter, A., Ding, N., Campbell, B., Chen, Y., Chen, Z., Chiaro, B., Dunsworth, A., Fowler, A.G., Jeffrey, E., Lucero, E., Megrant, A., Mutus, J.Y., Neeley, M., Neill, C., Quintana, C., Sank, D., Vainsencher, A., Wenner, J., White, T.C., Coveney, P.V., Love, P.J., Neven, H., Aspuru-Guzik, A., Martinis, J.M.: Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016)
3. Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J.M., Gambetta, J.M.: Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017)
4. Hempel, C., Maier, C., Romero, J., McClean, J., Monz, T., Shen, H., Jurcevic, P., Lanyon, B.P., Love, P., Babbush, R., et al.: Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X* **8**(3), 031022 (2018)
5. Jena, A., Genin, S., Mosca, M.: Pauli partitioning with respect to gate sets (2019). [arXiv:1907.07859](https://arxiv.org/abs/1907.07859)

6. Yen, T.-C., Verteletskyi, V., Izmaylov, A.F.: Measuring all compatible operators in one series of single-qubit measurements using unitary transformations. *J. Chem. Theory Comput.* **16**(4), 2400–2409 (2020)
7. Huggins, W.J., McClean, J., Rubin, N., Jiang, Z., Wiebe, N., Whaley, K.B., Babbush, R.: Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers (2019). [arXiv:1907.13117](https://arxiv.org/abs/1907.13117)
8. Gokhale, P., Angiuli, O., Ding, Y., Gui, K., Tomesh, T., Suchara, M., Martonosi, M., Chong, F.T.: Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families (2019). [arXiv:1907.13623](https://arxiv.org/abs/1907.13623)
9. Zhao, A., Tranter, A., Kirby, W.M., Ung, S.F., Miyake, A., Love, P.: Measurement reduction in variational quantum algorithms (2019). [arXiv:1908.08067](https://arxiv.org/abs/1908.08067)
10. Ryabinkin, I.G., Lang, R.A., Genin, S.N., Izmaylov, A.F.: Iterative qubit coupled cluster approach with efficient screening of generators. *J. Chem. Theory Comput.* **16**(2), 1055–1063 (2020)
11. Crawford, O., van Straaten, B., Wang, D., Parks, T., Campbell, E., Brierley, S.: Efficient quantum measurement of Pauli operators in the presence of finite sampling error (2019). [arXiv:1908.06942](https://arxiv.org/abs/1908.06942)
12. Hamamura, I., Imamichi, T.: Efficient evaluation of quantum observables using entangled measurements. *npj Quantum Inf.* **6**(1), 56 (2020)
13. Preskill, J.: Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018)
14. Verteletskyi, V., Yen, T.-C., Izmaylov, A.F.: Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *J. Chem. Phys.* **152**(12), 124114 (2020)
15. Torlai, G., Mazzola, G., Carleo, G., Mezzacapo, A.: Precise measurement of quantum observables with neural-network estimators. *Phys. Rev. Res.* **2**, 022060 (2020)
16. Choo, K., Mezzacapo, A., Carleo, G.: Fermionic neural-network states for ab-initio electronic structure. *Nat. Commun.* **11**(1), 1–7 (2020)
17. Huang, H.-Y., Kueng, R., Preskill, J.: Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020)
18. Bonet-Monroig, X., Babbush, R., O’Brien, T.E.: Nearly optimal measurement scheduling for partial tomography of quantum states (2019). [arXiv:1908.05628](https://arxiv.org/abs/1908.05628)
19. Cotler, J., Wilczek, F.: Quantum overlapping tomography. *Phys. Rev. Lett.* **124**(10), 100401 (2020)
20. Evans, T.J., Harper, R., Flammia, S.T.: Scalable Bayesian Hamiltonian learning (2019). [arXiv:1912.07636](https://arxiv.org/abs/1912.07636)
21. Paini, M., Kalev, A.: An approximate description of quantum states (2019). [arXiv:1910.10543](https://arxiv.org/abs/1910.10543)
22. Dallaire-Demers, P.-L., Romero, J., Veis, L., Sim, S., Aspuru-Guzik, A.: Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer. *Quantum Sci. Technol.* **4**(4), 045005 (2019)
23. Bravyi, S., Gosset, D., König, R., Temme, K.: Approximation algorithms for quantum many-body problems. *J. Math. Phys.* **60**(3), 032203 (2019)
24. Boyd, S., Kim, S.-J., Vandenberghe, L., Hassibi, A.: A tutorial on geometric programming. *Optim. Eng.* **8**(1), 67 (2007)
25. Bravyi, S., Gambetta, J.M., Mezzacapo, A., Temme, K.: Tapering off qubits to simulate fermionic Hamiltonians (2017). [arXiv:1701.08213](https://arxiv.org/abs/1701.08213)
26. Wecker, D., Hastings, M.B., Troyer, M.: Progress towards practical quantum variational algorithms. *Mol. Opt. Phys. Phys. Rev. A At.* **92**(4), 042303 (2015)
27. Arrasmith, A., Cincio, L., Somma, R.D., Coles, P.J.: Operator sampling for shot-frugal optimization in variational algorithms (2020). [arXiv:2004.06252](https://arxiv.org/abs/2004.06252)
28. Welsh, D.J.A., Powell, M.B.: An upper bound for the chromatic number of a graph and its application to timetabling problems. *Comput. J.* **10**(1), 85–86, 01 (1967)
29. Aleksandrowicz, G., Alexander, T., Barkoutsos, P., Bello, L., Ben-Haim, Y., Bucher, D., Cabrera-Hernández, F.J., Carballo-Franquis, J., Chen, A., Chen, C.F. et al.: Qiskit: an open-source framework for quantum computing. *16* (2019)
30. de Boer, P.-T., Kroese, D.P., Mannor, S., Rubinstein, R.Y.: A tutorial on the cross-entropy method. *Ann. Oper. Res.* **134**(1), 19–67 (2005)

Communicated by M. Christandl