



# Enhanced observable estimation through classical optimization of informationally overcomplete measurement data: Beyond classical shadows

Joonas Malmi <sup>\*</sup>, Keijo Korhonen , Daniel Cavalcanti, and Guillermo García-Pérez  
Algorithmiq Limited, Kanavakatu 3 C, FI-00160 Helsinki, Finland



(Received 23 February 2024; accepted 13 May 2024; published 10 June 2024)

In recent years, informationally complete measurements have attracted considerable attention, especially in the context of classical shadows. In the particular case of informationally overcomplete measurements, for which the number of possible outcomes exceeds the dimension of the space of linear operators in Hilbert space, the dual positive operator-valued measure operators used to interpret the measurement outcomes are not uniquely defined. In this paper, we propose a method to optimize the dual operators after the measurements have been carried out in order to produce sharper, unbiased estimations of observables of interest. We discuss how this procedure can produce zero-variance estimations in cases where the classical shadows formalism, which relies on so-called canonical duals, incurs exponentially large measurement overheads. We also analyze the algorithm in the context of quantum simulation with randomized Pauli measurements, and show that it can significantly reduce statistical errors with respect to canonical duals on multiple observable estimations.

DOI: [10.1103/PhysRevA.109.062412](https://doi.org/10.1103/PhysRevA.109.062412)

## I. INTRODUCTION

The study of quantum physics and the development of quantum technologies are based on our ability to extract useful information from quantum systems. In particular, quantum simulation on quantum computers typically requires performing measurements from which we can infer physical properties such as energy, magnetization, entropy, or correlations. A standard procedure is to perform quantum state tomography and obtain a description of the quantum state, from which we can estimate any observable or quantum-information-theoretical quantity. However, the number of parameters needed for such a task generally grows exponentially with the number of constituents of the system. This implies that the measurement cost (either in terms of measurement settings or shots) required to reach a certain precision, and the classical memory to store the data, become unattainable even for small system sizes.

An efficient way to estimate the mean value of different observables without the need to reconstruct the full quantum state is to apply an informationally complete measurement, given by informationally complete positive operator-valued measures (IC-POVMs), and classically postprocess the data using the dual effects of the measurement [1,2]. This idea recently attracted significant attention after the realization that particular choices of IC-POVMs and dual effects (called *classical snapshots* in the framework of *shadow estimation* [3]) can lead to efficient estimations in the number of measurement shots and qubits [3]. Several works have proposed other classes of IC-POVMs [4–9], and even on-the-fly optimization procedures [9–11].

While previous works on this type of estimation focused on proposing different IC-POVMs (which correspond to different measurement setups), Refs. [1,2,12] recognized that the accuracy of the estimations depends on the dual POVM effects used in the estimator as well. Here, we make this idea practical and introduce a method to optimize the dual effects of a POVM in a practical scenario where one has access to a finite sample of measurement data. We show that this method can provide exponential advantages in measurement overhead with respect to local shadow estimation, and apply it to physically relevant problems ranging from spin chain dynamics to quantum chemistry calculations.

## II. MULTIPLE OBSERVABLE ESTIMATION WITH INFORMATIONALLY COMPLETE MEASUREMENTS

Let us first review the idea of observable estimation through informationally complete POVMs. A POVM is described by positive operators (also called POVM effects)  $\Pi_i \geq 0$  ( $i = 0, \dots, r - 1$ ) that add up to identity, i.e.,  $\sum_{i=0}^{r-1} \Pi_i = \mathbb{I}$ . Upon measuring a state  $\rho$  with a POVM we obtain an outcome  $i$  with probability  $p_i = \text{Tr}[\Pi_i \rho]$ , so that we have  $r$  possible results, or outcomes. A particularly important set of POVMs is informationally complete POVMs (IC-POVMs), for which the POVM effects span the space of linear operators in the Hilbert space,  $\mathcal{L}(\mathcal{H})$ . This means that an IC-POVM needs to have  $r \geq d^2$  effects, where  $d^2$  effects are linearly independent. Thus, we can write any operator  $O \in \mathcal{L}(\mathcal{H})$  as  $O = \sum_i c_i \Pi_i$ . If an IC-POVM is IC and has exactly  $r = d^2$  linearly independent effects, it is called a minimal IC-POVM.

As mentioned above, IC-POVMs can be used to estimate the mean value of different observables while bypassing the explicit reconstruction of the quantum state [1]. This is done by first noticing that every IC-POVM can be associated with a set of *dual effects*  $D_i$  ( $i = 0, \dots, r - 1$ ) that are defined by

<sup>\*</sup>joonas@algorithmiq.fi

operators satisfying

$$O = \sum_i \text{Tr}[OD_i]\Pi_i = \sum_i \text{Tr}[O\Pi_i]D_i \quad (1)$$

for every operator  $O$ . This means that  $\{D_i\}_{i=1}^r$  also form the dual operator basis in  $\mathcal{L}(\mathcal{H})$  (consequently, they are sometimes called *dual frames*). If we choose  $O$  to be a quantum state  $\rho$  the second equality in Eq. (1) gives  $\rho = \sum_i p_i D_i$ . This means that  $\rho$  can be recovered as an average over dual POVMs, where the average is taken with respect to the measurement outcome probabilities  $p_i = \text{Tr}[\rho\Pi_i]$ . Thus, in a real experiment where we have access to  $S$  experimental shots of an IC-POVM, we can construct an unbiased estimator of  $\rho$  as  $\bar{\rho} = \sum_i f_i D_i$ , where  $f_i$  is the observed frequency of outcome  $i$  [13]. Furthermore, we can also define unbiased estimators of different observables as

$$\bar{O} = \text{Tr}[O\bar{\rho}] = \sum_i f_i \text{Tr}[OD_i]. \quad (2)$$

These estimators are also consistent, that is, they converge to  $\langle O \rangle$  in the limit of  $S \rightarrow \infty$  since  $f_i \rightarrow p_i$ .

*Classical shadows* [3] can be seen as a special case of the above estimation method where the dual effects are defined as

$$D_i = \Lambda^{-1}(\Pi_i), \quad (3)$$

with

$$\Lambda(\cdot) = \sum_i \text{Tr}[\cdot\Pi_i]\Pi_i. \quad (4)$$

Such dual effects are also called *canonical dual frames* or *covariant dual frames* [2].

### III. PARAMETRIZING DUAL EFFECTS

We can also construct a POVM that is IC, but have it be composed of more than  $d^2$  effects, so that it would be *overcomplete* (an OC-POVM for short). Consequently, some of the effects are linearly dependent on the others in a POVM of this form. As we will show next, the duals of OC-POVMs are not uniquely determined, and this freedom of choice can be used to improve the estimation.

To find the duals of an  $r$ -outcome OC-POVM, let us first choose  $d^2$  linearly independent effects among the ones of the POVM (the chosen effects are arbitrary for the functionality of our method). We will call these linearly independent effects the *basis effects*, and denote them by  $\{\bar{\Pi}_i\}_{i=0}^{d^2-1}$ . The remaining  $r - d^2$  effects will be called the *redundant effects*, and will be denoted by  $\{\tilde{\Pi}_j\}_{j=d^2}^{r-1}$ . Similarly, to these effects we will associate the *basis dual effects*  $\{\bar{D}_i\}_{i=0}^{d^2-1}$  and the *redundant dual effects*  $\{\tilde{D}_j\}_{j=d^2}^{r-1}$ . Using this notation, Eq. (1) reads

$$O = \sum_{i=0}^{d^2-1} \text{Tr}[O\bar{D}_i]\bar{\Pi}_i + \sum_{j=d^2}^{r-1} \text{Tr}[O\tilde{D}_j]\tilde{\Pi}_j. \quad (5)$$

Notice that the basis effects form an (unnormalized) minimal IC-POVM, so we can write  $\tilde{\Pi}_j = \sum_{i=0}^{d^2-1} \text{Tr}[D_i^* \tilde{\Pi}_j]\bar{\Pi}_i$ , where we have used the symbol  $\star$  to denote the unique duals to said

minimal basis. Thus, we get

$$\begin{aligned} O &= \sum_{i=0}^{d^2-1} \left( \text{Tr}[O\bar{D}_i] + \sum_{j=d^2}^{r-1} \text{Tr}[O\tilde{D}_j]\text{Tr}[D_i^* \tilde{\Pi}_j] \right) \bar{\Pi}_i \\ &= \sum_{i=0}^{d^2-1} \text{Tr}[OD_i^*]\bar{\Pi}_i. \end{aligned} \quad (6)$$

The last term in the expression is the unique decomposition of  $O$  in terms of the basis dual effects. Since this equality must hold for any operator  $O$ , the dual effects must fulfill

$$\bar{D}_i + \sum_{j=d^2}^{r-1} \tilde{D}_j \text{Tr}[D_i^* \tilde{\Pi}_j] = D_i^*. \quad (7)$$

This constraint between basis and redundant dual effects can be automatically satisfied by writing the former in terms of the latter, so we can parametrize the full set of duals of an OC-POVM as

$$\begin{aligned} D_i^* &= \sum_{j=d^2}^{r-1} \text{Tr}[D_i^* \tilde{\Pi}_j]\tilde{D}_j, \quad i = 0, \dots, d^2 - 1, \\ \tilde{D}_j &, \quad j = d^2, \dots, r - 1, \end{aligned} \quad (8)$$

where  $\tilde{D}_j$  are Hermitian matrices that can be chosen freely.

### IV. REDUCING STATISTICAL ERROR BY OPTIMIZING DUAL EFFECTS

We may exploit the freedom of choice presented in the parametrization in Eq. (8) to find the dual effects in such a way that the *variance* of the estimator of an observable  $O$  is minimized. The per-shot variance when estimating  $O$  from the POVM outcomes is given by

$$\text{Var}[O] = \sum_i p_i (\text{Tr}[OD_i])^2 - \left( \sum_i p_i \text{Tr}[OD_i] \right)^2, \quad (9)$$

where  $p_i$  is the probability of obtaining the  $i$ th outcome. While the second term does not depend on the choice of dual effects, as long as these satisfy Eq. (7), the first term—the second moment of  $\text{Tr}[OD_i]$  with respect to the probability distribution of the outcomes—does. Therefore, by choosing dual effects that minimize the second moment, the statistical errors in the estimation of the expectation value of the operator  $O$  may be significantly smaller than with, e.g., the canonical duals, even when using the same measurement outcome data.

As a simple example of how optimizing the duals of OC-POVMs can provide an advantage over the canonical duals, consider the task of estimating the average value of the Pauli-Z observable for the state  $|0\rangle$  using a random measurement on the Pauli basis, i.e., measuring  $Z$ ,  $X$ , or  $Y$  with equal probability. The POVM effects are given by

$$\begin{aligned} \Pi_0 &= \frac{1}{3}|0\rangle\langle 0|, & \Pi_1 &= \frac{1}{3}|1\rangle\langle 1|, & \Pi_2 &= \frac{1}{3}|+\rangle\langle +| \\ \Pi_3 &= \frac{1}{3}|-\rangle\langle -|, & \Pi_4 &= \frac{1}{3}|+i\rangle\langle +i|, & \Pi_5 &= \frac{1}{3}|-i\rangle\langle -i|, \end{aligned} \quad (10)$$

where  $|\pm\rangle = 1/\sqrt{2}(|0\rangle \pm |1\rangle)$  and  $|\pm i\rangle = 1/\sqrt{2}(|0\rangle \pm i|1\rangle)$ . One can readily see that the probabilities to obtain the different outcomes are  $p_0 = 1/3$ ,  $p_1 = 0$ ,  $p_2 = p_3 = p_4 = p_5 = 1/6$ . The canonical duals read

$$\begin{aligned} D_0 &= \frac{1}{2}(I + 3Z), & D_1 &= \frac{1}{2}(I - 3Z), & D_2 &= \frac{1}{2}(I + 3X), \\ D_3 &= \frac{1}{2}(I - 3X), & D_4 &= \frac{1}{2}(I + 3Y), & D_5 &= \frac{1}{2}(I - 3Y). \end{aligned} \quad (11)$$

This choice of duals provides a variance  $\text{Var}[Z] = \sum_i p_i (\text{Tr}[D_i Z])^2 - (\sum_i p_i \text{Tr}[D_i Z])^2 = 2$ . We can now follow the previously discussion to minimize the variance with respect to the choice of dual effect. In order to do so we first choose the basis effects to be the set  $\{\Pi_0, \Pi_1, \Pi_2, \Pi_4\}$ . The corresponding basis duals are then fixed to be  $\{D_0, D_1, D_2, D_4\}$ . If we now optimize the variance duals to minimize the variance in Eq. (9) using the parametrization in Eq. (8), we find  $\text{Var}[Z] = 0$  with the choice of duals

$$\begin{aligned} D_0^{\text{opt}} &= \frac{1}{2}(I + Z), & D_1^{\text{opt}} &= \frac{1}{2}(I - 5Z), \\ D_2^{\text{opt}} &= \frac{1}{2}(I + 3X + Z), & D_3^{\text{opt}} &= \frac{1}{2}(I - 3X + Z), \\ D_4^{\text{opt}} &= \frac{1}{2}(I + 3Y + Z), & D_5^{\text{opt}} &= \frac{1}{2}(I - 3Y + Z). \end{aligned}$$

Notice that in the case of a multiqubit state  $|0\rangle^{\otimes N}$  and an observable  $Z^{\otimes N}$ , the canonical duals in Eq. (11) in Ref. [3] result in an exponential overhead, with variance  $\text{Var}[Z] = 3^N - 1$ , while the optimal duals result in zero-variance estimations. This shows that optimizing the dual effects can provide an exponential advantage in terms of measurement overhead with respect to canonical duals.

In this particular example one could claim that it is better to estimate the observable directly instead of using an OC-POVM. Our main point here is to illustrate the power of duals and show that we can still use IC measurements and recover the optimal variance. Moreover, dual optimization is a general-purpose approach that can be used with more complex OC-POVMs as well, for which no simple and efficient data postprocessing strategy may be obvious. This is particularly relevant in schemes in which one optimizes overcomplete measurement setups, as in Ref. [11].

## V. DUAL OPTIMIZATION IN THE CASE OF FINITE STATISTICS

While the above discussion introduces the basic idea of dual optimization and its potential, its practical implementation poses additional challenges. In general, the probability distribution of the outcomes  $\{p_i\}$  is not knowable since its characterization is exponentially hard, and we only have access to the experimental frequencies  $\{f_i\}$  obtained after repeating the measurement a finite number of times. In this case we can estimate the second moment of the estimator as

$$\overline{\text{Tr}[OD_i]^2} = \sum_i f_i (\text{Tr}[OD_i])^2. \quad (12)$$

We can then minimize this quantity using the parametrization in Eq. (8).

Notice, however, doing so may introduce statistical biases: If the dual effects  $\{D_i\}$  are modified as to minimize  $\sum_i f_i (\text{Tr}[OD_i])^2$ , then the duals and the measurement data  $f_i$  are no longer statistically independent, which means that we cannot guarantee that the estimator of the mean  $\sum_i f_i \text{Tr}[OD_i]$  is unbiased. In order to prevent this, our procedure consists in splitting the measurement data evenly into two disjoint sets,  $A$  and  $B$ , so that for a total of  $S$  measurements each subset will have  $S/2$  measurement outcomes (the way in which the data are split is somewhat arbitrary and we leave as a point for further investigation how to optimally choose the sizes of sets  $A$  and  $B$ ). We first optimize the duals using a *training data set*  $A$ , and then evaluate the mean and the variance using the optimized duals in an *estimation data set*  $B$ . Since  $A$  and  $B$  are disjoint, the optimized duals and the frequencies used in the final estimation are statistically independent. We then repeat the procedure swapping the roles of  $A$  and  $B$ , hence producing another pair of estimations of the mean and the variance that are then combined with the former ones, so no data are left unused.

In our simulations the optimization of the duals is performed in the following way: We start by considering local random Pauli measurements (10) from which the measurement frequencies are obtained. We then fixed the canonical duals (11) for  $N - 1$  qubits and use the parametrization in Eq. (8) to optimize the variance of the estimator as a function of the free parameters for the remaining qubit. Finally, we sweep this procedure over all qubits until we find convergence of the variance. Each single-qubit optimization is carried out using the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimizer. Notice that this qubit-by-qubit procedure scales linearly with the number of qubits.

## VI. EXAMPLES

In order to test the approach, we consider two physically motivated problems: the estimation of the real part of 2-reduced density matrix (2-RDM) operator elements of the  $\text{H}_2\text{O}$  and  $\text{LiH}$  molecules, and the quantum simulation of a spin chain performed in Ref. [14]. Notice that while we optimize the second moment (12), the first moment  $\bar{O} = \sum_i f_i \text{Tr}[OD_i]$  can also get affected (contrary to the infinite statistics where  $\bar{O} = \langle O \rangle$  is independent of the choice of duals). Since the variance of the estimator decreases upon the optimization, we expect that the absolute error of the estimation  $\epsilon = |\langle O \rangle - \bar{O}|$  also decreases. In what follows we explicitly show that this is the case by comparing  $\epsilon$  with and without dual optimization for various physically motivated examples.

We first consider the problem of estimating the real part of fermionic 2-reduced density matrix (2-RDM) operator elements  $a_i^\dagger a_j^\dagger a_k a_l$ , where  $a_i$  are fermionic annihilation operators. To that end, we consider the  $\text{H}_2\text{O}$  and  $\text{LiH}$  molecules in a minimal basis set. Their second quantized Hamiltonians and 2-RDM elements are mapped to qubit space using the Jiang-Kalev-Mruczkiewicz-Neven (JKMN) fermion-to-qubit mapping [15], resulting in 12-qubit operators. We then use pretrained hardware-efficient variational quantum eigensolver (VQE) states [16] to approximate their ground states, and

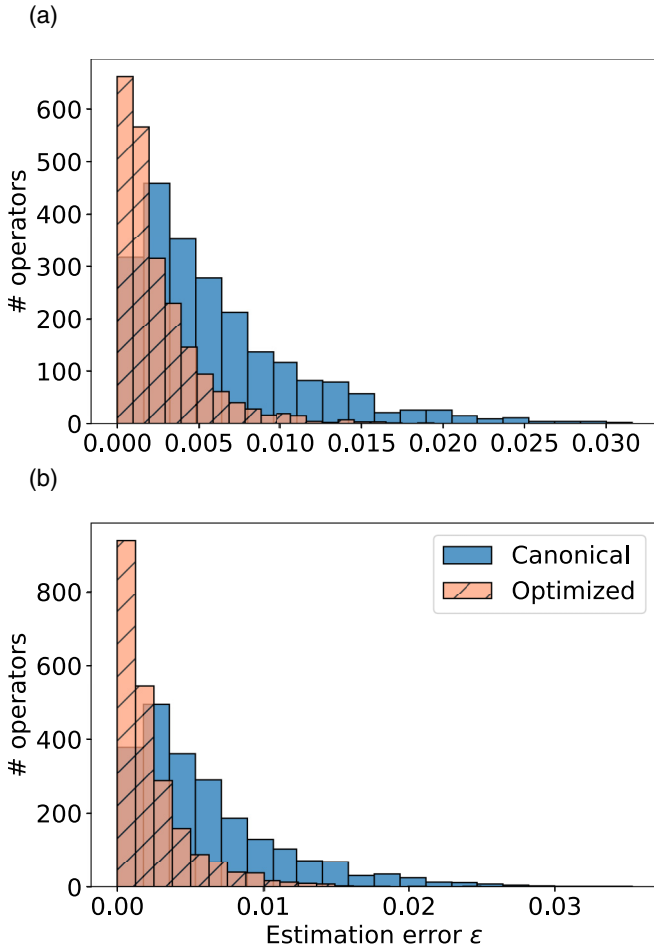


FIG. 1. Distribution of absolute errors  $\epsilon = |\langle O \rangle - \bar{O}|$  over the real parts of fermionic 2-RDM elements for the (a) H<sub>2</sub>O and (b) LiH molecules. Pretrained VQE states are measured using the POVM in Eq. (10). The resulting  $2 \times 10^6$  shot data set is then used to evaluate all the 2-RDM expectation values with the canonical duals, as well as to optimize the duals using the protocol introduced in the main text and produce the corresponding estimations.

measure these states with the randomized Pauli measurements in Eq. (10) to obtain  $2 \times 10^6$  shots. The resulting data set is then used to evaluate all the 2-RDM elements with canonical duals, as well as to optimize them and produce the corresponding estimations. As it can be seen in Fig. 1, where we depict the distribution of  $\epsilon$  over RDM elements, the optimization has the overall effect of significantly reducing statistical errors.

As a second application, in Fig. 2, we show the results for the simulation of the Trotter evolution of a 10-spin transverse-field Ising model performed in Ref. [14]. The Hamiltonian that drives the evolution reads

$$H = -J \sum_i Z_i \otimes Z_{i+1} + h \sum_i X_i, \quad (13)$$

with  $J = 0.5236$  and  $h = 1$ , with initial state  $|0\rangle^{\otimes 10}$ . The estimated observable is  $O = Z^{\otimes 10}$ . For each Trotter step, we repeat the simulation 1000 times, sampling  $2 \times 10^6$  shots per simulation.

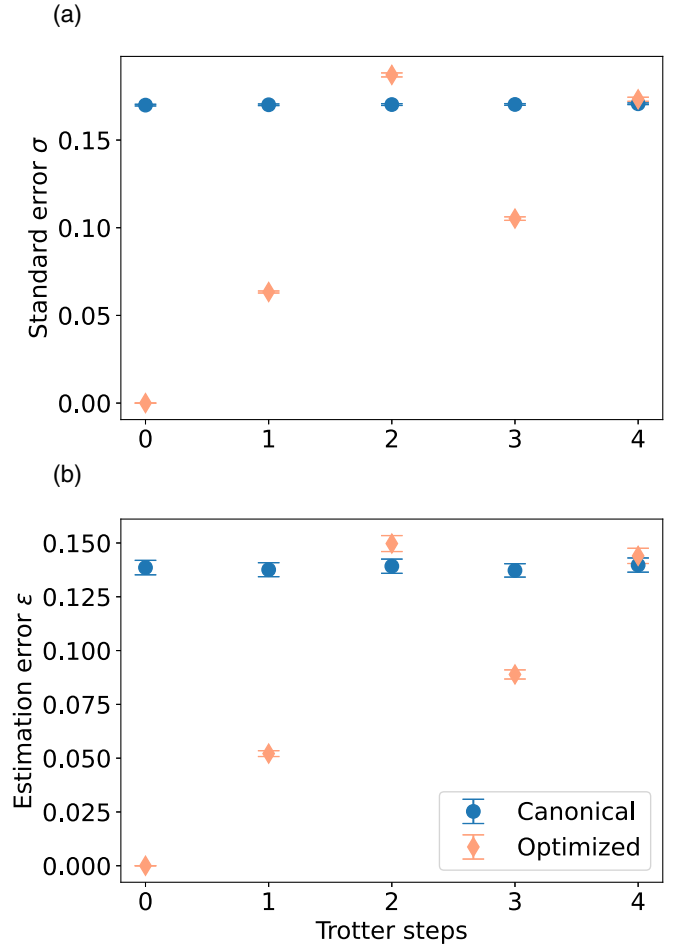


FIG. 2. (a) Estimated standard error  $\sigma = \sqrt{\text{Var}[O]/S}$ , where  $S$  is the number of shots used in the estimation, and (b) absolute error  $\epsilon = |\langle O \rangle - \bar{O}|$  of the observable  $O = Z^{\otimes 10}$  for different Trotter steps in a 10-qubit system. At each time step, the state is measured 1000 times using 2 million measurement shots per repetition. In each repetition, the data set is split in two 1-million-shot data sets  $A$  and  $B$ . First, data set  $A$  is used for optimization and data set  $B$  for estimation, and then vice versa. The values shown here are averaged over the estimation values between the two data sets. For these results, we performed 20 sweeps over the qubits during the optimization. The average and standard deviation over repetitions.

As we can see, dual optimization provides a reduced error for steps 0, 1, and 3. In particular, at step 0, the procedure consistently achieves a zero-variance estimation. This is possible because the state and observable are  $|0\rangle^{\otimes N}$  and  $Z^{\otimes N}$ , respectively, which is precisely the aforementioned example of an exponential overhead with canonical duals. However, notice that the algorithm consistently finds optimal duals based only on the provided measurement data, assuming the canonical duals as a starting point. For steps 2 and 4, the optimization procedure leads to higher  $\sigma$  and  $\epsilon$ . This is a direct consequence of overfitting: While the optimized duals reduce the second moment Eq. (12) for the training data set, they increase it for the estimation data set. In order to mitigate this overfitting, one can simply monitor the value for the estimation set and stop the optimization as soon as the training and estimation data-set values differ significantly. In any case, notice that



the estimated standard error  $\sigma$  is an observable quantity that does not require previous knowledge of  $\langle O \rangle$  (as opposed to the absolute error  $\epsilon$ ) and, as Fig. 2 shows,  $\sigma$  is a good proxy of the actual estimation error  $\epsilon$ . Therefore, one can always use the estimation provided by the duals with the smallest standard error on the estimation data set.

In this paper, we have shown that the estimation of physical properties by means of informationally overcomplete POVMs can be greatly improved by the optimization of the dual effects used in the estimator. This improvement comes in the form of a reduced estimation variance, which in turn provides an advantage in terms of measurement overhead (that is, the number of shots required to achieve a certain precision). Furthermore, this optimization is purely classical, and can be performed in postprocessing, thus not requiring any modification of the physical setup.

We have demonstrated the feasibility of our optimization procedure in physically relevant problems, such as the Trotter evolution of a spin system, and different physical properties of molecular systems. We believe that our method can be of great help not only for the estimation of physical properties, but also for subroutines of quantum computing and simulation protocols. For instance, in some VQE approaches, one needs to estimate a great number of commutators [17], which typically results in a prohibitive measurement cost. Another example is the estimation of stabilizer observables for error correction codes and other applications such as one-way quantum computation. This method can also be

used in conjunction with adaptive POVMs [10], where the POVM is first optimized with respect to some state and operators, and the duals are then optimized in postprocessing as well.

Finally, the numerical simulations presented here only considered as starting points local POVMs and local canonical duals. We leave as a further investigation the possibility of starting from more general initial conditions and going beyond product duals.

*Note added.* Recently, we became aware of a related work by Fischer *et al.*, in which they also observe estimation improvements using dual optimization for randomly chosen states and observables [18]. We also became aware of a related work by Caprotti *et al.*, in which they demonstrate that dual frame optimization for local measurements can achieve the same efficiency as global Clifford shadows [19].

## ACKNOWLEDGMENTS

We thank Sergey Filippov and Boris Sokolov for helpful discussions. Guillermo García-Pérez conceived the idea for this method, Daniel Cavalcanti and Guillermo García-Pérez designed and directed the research, Joonas Malmi implemented the algorithms and ran the simulations, and all authors contributed to scientific discussions and to the writing of the manuscript. The authors declare no competing interests.

- 
- [1] G. M. D'Ariano and P. Perinotti, Optimal data processing for quantum measurements, *Phys. Rev. Lett.* **98**, 020403 (2007).
  - [2] L. Innocenti, S. Lorenzo, I. Palmisano, F. Albarelli, A. Ferraro, M. Paternostro, and G. M. Palma, Shadow tomography on general measurement frames, *PRX Quantum* **4**, 040328 (2023).
  - [3] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
  - [4] N. Guo, F. Pan, and P. Reberntrost, Estimating properties of a quantum state by importance-sampled operator shadows, [arXiv:2305.09374](https://arxiv.org/abs/2305.09374).
  - [5] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, *Commun. Math. Phys.* **391**, 951 (2022).
  - [6] M. Ippoliti, Y. Li, T. Rakovszky, and V. Khemani, Operator relaxation and the optimal depth of classical shadows, *Phys. Rev. Lett.* **130**, 230403 (2023).
  - [7] C. Bertoni, J. Haferkamp, M. Hinsche, M. Ioannou, J. Eisert, and H. Pashayan, Shallow shadows: Expectation estimation using low-depth random Clifford circuits, [arXiv:2209.12924](https://arxiv.org/abs/2209.12924).
  - [8] M. Arienzo, M. Heinrich, I. Roth, and M. Kliesch, Closed-form analytic expressions for shadow estimation with brickwork circuits, *Quantum Inf. Comput.* **23**, 961 (2022).
  - [9] H. Chau Nguyen, J. Lennart Bönsel, J. Steinberg, and O. Gühne, Optimizing shadow tomography with generalized measurements, *Phys. Rev. Lett.* **129**, 220502 (2022).
  - [10] G. García-Pérez, M. A. C. Rossi, B. Sokolov, F. Tacchino, P. K. Barkoutsos, G. Mazzola, I. Tavernelli, and S. Maniscalco, Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms, *PRX Quantum* **2**, 040342 (2021).
  - [11] A. Glos, A. Nykänen, E.-M. Borrelli, S. Maniscalco, M. A. C. Rossi, Z. Zimborás, and G. García-Pérez, Adaptive POVM implementations and measurement error mitigation strategies for near-term quantum devices, [arXiv:2208.07817](https://arxiv.org/abs/2208.07817).
  - [12] L. Guerini, R. Wiersema, J. F. Carrasquilla, and L. Aolita, Quasiprobabilistic state-overlap estimator for NISQ devices, [arXiv:2112.11618](https://arxiv.org/abs/2112.11618).
  - [13] To see that  $\bar{\rho}$  is unbiased,  $\mathbb{E}[\bar{\rho}] = \rho$ , notice that  $\mathbb{E}[\bar{\rho}] = \sum_i \mathbb{E}[f_i] D_i = \sum_i p_i D_i$ , since  $f_i = s_i/S$ , where  $s_i$  is the number of times outcome  $i$  was obtained out of the  $S$  experiments. The random variables  $s_i$  follow a multinomial distribution with mean  $S p_i$ .
  - [14] E. van den Berg, Z. K. Mineev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors, *Nat. Phys.* **19**, 1116 (2023).
  - [15] Z. Jiang, A. Kalev, W. Mruzckiewicz, and H. Neven, Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning, *Quantum* **4**, 276 (2020).
  - [16] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules

- and quantum magnets, *Nature (London)* **549**, 242 (2017).
- [17] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, *Nat. Commun.* **10**, 3007 (2019).
- [18] L. E. Fischer, T. Dao, I. Tavernelli, and F. Tacchino, Dual frame optimization for informationally complete quantum measurements, [arXiv:2401.18071](https://arxiv.org/abs/2401.18071).
- [19] A. Caprotti, J. Morris, and B. Dakić, Optimising quantum tomography via shadow inversion, [arXiv:2402.06727](https://arxiv.org/abs/2402.06727).