Analysis of variational quantum algorithms for differential equations in the presence of quantum noise : application to the stationary Gross-Pitaevskii equation

Michel Fabrice Serret

Sorbonne Université
LJLL - Eviden Quantum Lab

École doctorale **École Doctorale Sciences Mathématiques de Paris Centre**

Laboratoire **Laboratoire Jacques-Louis Lions**

Thèse présentée par **Michel Fabrice Serret**

Soutenue le **8 novembre 2024**

En vue de l'obtention du grade de docteur de Sorbonne Université

Discipline **Applied Mathematics/Mathématiques appliquées**

Spécialité **Numerical Analysis/Analyse numérique**

Analysis of variational quantum algorithms for differential equations in the presence of quantum noise,

application to the stationary Gross-Pitaevskii equation

Thèse dirigée par

Thomas Ayral	co-directeur
Laurent Boudin	directeur
Yvon Maday	co-directeur

Composition du jury

Rapporteurs	Nana Liu	professeure à Shanghai Jiao Tong University
	Anthony Nouy	professeur à l'École Centrale de Nantes
Examinatrices	Virginie Ehrlacher	professeure à l'École Nationale des Ponts et Chaussées, présidente
	Marie Postel	professeure à Sorbonne Université
	Elvira Shishenina	Quantinuum
Directeurs de thèse	Thomas Ayral	Eviden Quantum Lab
	Laurent Boudin	MCF HDR à Sorbonne Université
	Yvon Maday	professeur à Sorbonne Université

Colophon

Mémoire de thèse intitulé « Analysis of variational quantum algorithms for differential equations in the presence of quantum noise, », écrit par Michel Fabrice Serret, achevé le 21 novembre 2024, composé au moyen du système de préparation de document LaTeX et de la classe yathesis dédiée aux thèses préparées en France.

Mots clés : quantum computing, differential equations, variational algorithms, gross-pitaevskii equation

Keywords: informatique quantique, équations différentielles, algorithmes variationels, équation de gross-pitaevskii

Cette thèse a été préparée dans les laboratoires suivants.

Laboratoire Jacques-Louis Lions
Sorbonne Université
Campus Pierre et Marie Curie
4 place Jussieu
75005 Paris
France

☎ +33 1 44 27 42 98
Site `https://www.ljll.fr/`

Eviden Quantum Lab
Avenue Jean-Jaurès
78340 Les Clayes-sous-bois
France

Site `https://www.eviden.com/`

EVIDEN

Analysis of variational quantum algorithms for differential equations in the presence of quantum noise,
application to the stationary Gross-Pitaevskii equation

Résumé

Variational quantum algorithms (VQAs) have been proposed for solving partial differential equations on quantum computers. This thesis focuses on analyzing VQAs for the stationary Gross-Pitaevskii Equation (GPE) both under ideal (noiseless) conditions and in the presence of quantum noise, providing error bounds, convergence properties, and estimates for the number of samples required.
A central concept, make use of a relationship between the representation of functions and functional operators on dyadic rationals, through the Walsh basis, and the encoding of functions and operators for N-qubit quantum systems through the Pauli operators and their eigenstates.
In chapter 1, we link Pauli operators of N-qubit quantum systems with the Walsh basis on N-bit dyadic rationals, presenting new error bounds for the convergence of the N-bit Walsh series for functions in $H^1(0,1)$ and presenting some results on the representation of Fourier basis functions in the Walsh basis.
In chapter 2 we analyse VQAs for the GPE without noise, detailing the mathematical setting, discretization, and a-priori analysis. We introduce new energy estimators, either based on the Walsh decomposition of operators or obtained through inductive methods, and compare them to direct sampling, in the diagonal basis of the operators, and the Hadamard-test method. Our results show that in the absence of noise the most promising methods for energy estimation is direct sampling in the diagonal basis, yielding the lowest variance and sample requirements.
In chapter 3, we further examine the impact of quantum noise on energy estimation. Depolarizing noise introduces bias and shifts the variance of estimators. We show that the Pauli estimators proves least affected by noise, due to their lower circuit size requirement, outperforming others both without mitigation, due to a lower bias, and with mitigation, as its sample efficiency is less affected.
This research provides a foundation for the use of VQAs to solve partial differential equations on quantum computers, offering realistic estimates of computational costs of the energy estimation part of the algorithms. Furthermore, the insights gained may aid in the development of practical quantum algorithms for block encoding of differential operators, and their associated evolution operators.

Mots clés : quantum computing, differential equations, variational algorithms, gross-pitaevskii equation

Laboratoire Jacques-Louis Lions – Sorbonne Université – Campus Pierre et Marie Curie – 4 place Jussieu – 75005 Paris – France
Eviden Quantum Lab – Avenue Jean-Jaurès – 78340 Les Clayes-sous-bois – France

Remerciements

Avant tout, je souhaite exprimer ma profonde gratitude à mes directeurs de thèse pour leur accompagnement précieux au cours de ces trois années. Merci à Laurent Boudin, pour son humour, sa bienveillance, pour avoir réussi à me rendre moins approximatif et à me cadrer, et pour m'avoir toujours sagement conseillé dans tous les domaines de la vie, de la diplomatie au meilleur restaurant du 13ème ; merci à Thomas Ayral, qui m'a fait prendre goût à la recherche et sans qui je n'aurais probablement jamais tenté un doctorat ; et merci à Yvon Maday, qui a toujours trouvé du temps pour moi et m'a toujours poussé à aller plus loin. Ce fut un plaisir de travailler avec vous ; j'ai tellement appris et je ne saurais jamais assez vous remercier pour tout ce que vous m'avez apportés !

Je remercie mes rapporteurs, Nana Liu et Anthony Nouy, pour leur relecture assidue de mon manuscrit, ainsi que les membres de mon jury de thèse, Virginie Ehrlacher, Marie Postel et Elvira Shishenina, d'avoir accepté de faire partie du jury.

I would like to thank my reviewers, Nana Liu and Anthony Nouy, for their thorough review of my manuscript, as well as the members of my thesis committee, Virginie Ehrlacher, Marie Postel, and Elvira Shishenina, for agreeing to be part of the jury.

Merci aussi à Mark Asch, Henri Calandra et Frédérique Charles, mon comité de suivi de thèse, pour leur bienveillance.

Merci aussi à Corentin Lacombe et Jean-François Venuti de l'École doctorale pour toute leur aide durant cette thèse.

Merci à Nicolas Treps, qui m'a donné l'occasion d'enseigner l'informatique quantique et avec lequel j'ai énormément appris sur l'enseignement.

Je tiens à remercier ceux qui ont marqué ma vie au sein du LJLL : merci au directeur du laboratoire, Emmanuel Trélat, pour sa bienveillance ; merci à Malika, Erika, Salima et Corentin pour leur bonne humeur constante et leur efficacité remarquable grâce à laquelle la vie est si agréable au LJLL. Merci aussi à Khashayar pour toute son aide.

Merci à mes adelphes de thèse, Thomas B., pour tous ces moments au tableau à imaginer de nouveaux mondes mathématiques, Aleksandra, pour sa jovialité permanente, et à Alicia ! Merci aussi à Federica pour sa gaieté, son amitié et sa présence, et à mes autres co-bureaux, Ruikang, Zhe et Pablo ! Merci aussi à leurs prédécesseurs, Emma, Jesus, Eugenio et Matthieu, qui m'ont accueilli au LJLL ! Merci aussi à Valentine pour sa bonne humeur contagieuse !

Merci à Mi-Song de m'avoir supporté pendant ces trois années ; ce fut un plaisir d'organiser NumerIQ à tes côtés, merci pour tout ! Merci à Cindy, Fabien, Guillaume, Frédérique Noël, Chloé et Marie pour tout votre soutien et votre bienveillance durant ces trois années ! Merci à Nicolaï, Lucas P., Charles, Robin et Anatole pour toutes nos conversations et moments de convivialité aux Sciences et au labo ! Merci à Igor et Hassan ! Merci à Yvonne, Roxane, Elena, Guillaume, Pauline, Edouard, Pierre, Lucas J., Siwar et Yipeng ! Merci à Lise, Rémi, Jules, Alexandre P. et Antoine ! Merci à Aloïs, Siguang, Ludovic, Lucia et Sebastian ! Merci à Julien, Luis, Nathalie et Fabrice ! Merci à Julien, César et tous les membres de NumerIQ ! Merci à tous ceux qui ont

participé, de près ou de loin, à ma vie au LJLL durant ces années !

Je tiens aussi à remercier ceux qui ont partagé ma vie au sein d'Eviden durant ces trois années. Tout d'abord, merci à Cyril, qui a fait en sorte que cette thèse soit possible. Merci à mes compères doctorants, Pauline et Maxime R., avec qui l'aventure a commencé en 2019 et sur qui j'ai toujours pu compter ; merci à Océane pour son énergie inépuisable et sa joie de vivre ! Merci aussi à Arthur et à Vivien ! Merci aussi à Tristan et à Tom pour tous ces moments privilégiés où nous participions aux quiz en équipe avec ce cher Weston ! Merci aussi à tous les permanents de l'équipe : Arnaud, qui a toujours su mettre une bonne ambiance et qui est toujours présent quand on a besoin de lui ; Corentin pour sa sagesse et sa bienveillance constante ; merci à Cyprien et Stanley, sans qui l'équipe quantique ne serait pas la même ! Merci à Grigori et à Maxime O. pour toutes nos discussions. Merci à Thomas T. et Baptiste ; le peu de temps que j'ai passé à vos côtés fut un plaisir ! Merci aussi à Simon pour ses conseils avisés et sa bienveillance !

Je remercie aussi mes amis et ma famille pour leur soutien durant cette thèse. Merci à Haris, mon compère, mon poto, qui est toujours là pour moi ! Merci aux docteurs Nico et Tangui pour tous ces déjeuners aux arènes (et alentours) à partager nos expériences de thèse, ce fut un plaisir de partager ces années avec vous ! Merci aux potos de la PM & compagnie Nico et Thomas, Alexis et Ashley, Tangui et Edwige ainsi que Maxime et Claudia, qui me supportent depuis maintenant onze ans et sans qui ma vie n'aurait pas été la même ! Merci du fond du cœur à mes amis Haris, Paulo, Alex, Marie, Ombeline et Maeva, qui ont toujours su me redonner le sourire et m'offrir des moments de pause bienvenus, loin de la thèse ! Merci à Charly et Gaetan pour leur accueil à Lausanne et la découverte de la fondue suisse ! Merci à Nils et Vanessa, Johnny et Bejan, Yacine et Julien, Alexandra et Quentin ainsi que Morgane et Eric, Caroline, Virginie et Arthur. Merci à tous mes autres amis que je n'ai pas cités ici, mais que je n'oublie pas dans mon cœur !

Merci à mes beaux-parents, Guillaume et Béatrice, pour tout leur soutien ! A Titi, Clémence et David, ainsi qu'aux petiots, Louise et Constant.

Merci à Robert et Françoise, à Elisabeth et Fred, ma famille adoptive en France ! Merci à mes parents, Michèle et Yves, qui m'ont encouragé à développer ma curiosité et à trouver mon propre chemin depuis ma tendre enfance, à mon frère Ludovic, qui a toujours été un exemple pour moi, ainsi qu'à ma belle-sœur Sonia, et à mes neveux Quentin et Ethan !

Enfin, merci à ma douce Clotilde, l'amour de ma vie, pour son soutien indéfectible durant ces années de thèse !

Table des matières

Introduction

Solving differential equations on quantum computers: expectation and challenges on current devices

Differential equations are ubiquitous across the fields of science, engineering, and technology, ranging from modeling physical systems in quantum mechanics to predicting financial markets. Classical methods for solving partial differential equations, either for high-dimensional or nonlinear problems, often run into computational bottlenecks. For linear high-dimensional problems, the curse of dimensionality causes the size of the linear systems to grow prohibitively fast. A simple example is to study a quantum N-body problem, where the quantum state of the complete system is a function of $\mathbb{R}^{3N}$ and thus even for relatively small number of particles the dimension of the space being discretized becomes prohibitively large. This leads to two problems, a runtime problem, since, if we consider N_{disc} discretization points per dimension, the numerical solving of the linear system has a cost of N_{disc}^{9N}, and a memory problem due to insufficient memory and data transfer speeds.

For nonlinear problems, the solving of the equations cannot be achieved through single matrix inversions at each time step. Indeed, in most cases, nonlinear programming [8] or linearization and fixed-point methods have to be used to achieve convergence which induce considerable overheads.

Quantum computing may provide alternatives by leveraging quantum mechanics to perform specific computations which cannot be handled by classical computers. This thesis investigates quantum computational approaches to solving nonlinear differential equations, with a focus on noisy intermediate-scale quantum (NISQ) devices, *i.e.* devices with no error correction, and thus not suited to most algorithms with known computational advantages.

Why we study what we study

In this thesis, we aim to ascertain to which extent variational quantum algorithms can be used in practice to solve differential equations and to determine the computational cost of such algorithms. We studied one of the nonlinear problems present in the literature, the Gross-Pitaevskii equation studied in [68], and restricted ourselves to the study of one step of VQA, the energy estimation.

Practicality of solving DEs with NISQ devices efficiently

Our methodology to obtain estimates on the algorithm computational cost can be simplified as follows.

In Chapter 2, we first consider the performance of the algorithm in the absence of noise.

- First, we determine how the convergence rate of the discretized solution approaches the exact solution as the number of qubits increases. This enables us to set an upper bound on the precision required in energy estimation for a given number of qubits, as additional precision does not further reduce the distance of the minimizer from the exact solution.
- Second, we explore alternative estimators with lower circuit depths than the estimators proposed in [68] for performance comparison, due to the their high sensitivity to quantum noise.
- Third, we calculate the variance of the estimators and the number of samples needed to achieve a specified error, which is dependent on the number of qubits.
- Finally, we analyze the number of samples necessary to achieve the required precision in energy for convergence and compare the performance of the different estimators.

In Chapter 3, we aim to determine the effect of noise on the performance of the algorithm as compared to the noiseless case.

- Our methodology in this chapter is to first determine the effect of noise on the estimators. We show that quantum noise introduces a bias and changes the variance of the estimators which depends on the number of gates in the associated quantum circuit.
- Then we estimate the number of gates contained in the circuits associated with each estimator and finally we determine the effect of the noise on the estimators.
- We then provide an analysis of the performance of the algorithm both without, and with, quantum error mitigation.

Chapter 1 is divided into two sections, one on the basics of quantum computing required for Chapters 2 and 3 and one on the Walsh functional basis in which we define some new results on the representation of the Fourier basis functions in the Walsh basis, and on convergence rates for the Walsh series of periodic H^1 functions. The results on the Walsh series is used in Chapter 2, thanks to its link to the Pauli operators, to obtain alternative Pauli estimators for the estimation of the energy.

Summary of Chapter 1

Key concepts and formalism in quantum computing

In order to understand the inner workings of variational quantum algorithms, a few prerequisites are mandatory, and first, and foremost, the nature of the underlying structure of quantum systems. Quantum systems can be described as normalised vectors, denoted by "kets" $|\cdot\rangle$, in a Hilbert space $\mathcal{H}$, while the vectors of the dual $\mathcal{H}^*$ are denoted by a "bra" $\langle\cdot|$. The kets represent the wave functions, or probability amplitudes, which describe the state of the quantum system.

For a quantum system composed of N qubits, the state space $\mathcal{H}$ is the tensor product of N two-dimensional Hilbert spaces, each associated to a single qubit. The computational basis $\mathscr{B}_{\mathrm{C}}(\mathcal{H})$ of $\mathcal{H}$ is given by the set of kets $\{|i_1\rangle \otimes \cdots \otimes |i_N\rangle \ , \ i_1, \dots i_N \in \{0,1\}\}$ where each $|i_k\rangle$ is the state of the k-th qubit, $1 \leq k \leq N$. In this computational basis, we denote a state by $|i\rangle$ and its Hermitian conjugate by $\langle i|$, where we set $i = \sum_{k=1}^{N} i_k 2^{N-k+1}$, a writing of the integer i following the so-called most-significant-bit-first convention.

In the context of quantum computation, the operations allowed on a quantum systems are modelled through unitary operators called gates, which act upon a subset of qubits. In practical implementations, quantum circuits consists of a restricted set of 1 and 2-qubit gates, as any unitary operation can be approximated by only involving such gates, according to the Solovay-Kitaev theorem [73]. Then a set of ordered quantum gates applied to specific qubits is called a quantum circuit.

Observables in quantum systems are operators corresponding to physical quantities which can be measured, and which are represented by Hermitian operators on $\mathcal{H}$. The Pauli operators are traceless and "local" operators, in the sense that they are tensors of 1-qubit Pauli operators, and that, when combined with the identity matrix, form a basis of $\mathscr{L}(\mathcal{H})$. One of the postulates of quantum physics is that the quantum measurement of an observable O on a state $|\psi\rangle$ yields an eigenvalue of O and collapses the system into the corresponding eigenstate of O. In the case of quantum computers, due to the physical constraints, particularly in qubit systems, direct measurement is typically restricted to the Pauli-Z operators of the individual qubits. This implies that direct measurement of operators requires the application of a basis change before measurement, which may induce long circuits.

The expectation value of a generic observable O can be obtained by diagonalizing O and sampling it in its eigenvector basis. Alternatively, it can be obtained by decomposing O into the Pauli basis, sampling the associated Pauli operators and summing the contributions of each component. Finally, the mean value of unitary operators for a state $|\psi\rangle$ can be obtained through circuits known as Hadamard-test circuits through the controlled application of the unitary operator on the state $|\psi\rangle$.

As we shall see, the ability to manipulate the state of the quantum system and measure the associated mean value of operators will allow to encode discretized values of functions and obtain estimators for the discretized energy functionals. For the Pauli operators, the link between the functionals and the quantum states can be seen through the Walsh functions.

Walsh functions, analog of Pauli operators for functions on the dyadics

The definition of the Walsh functions is intricately linked to the notion of dyadic rationals, which we first describe. The dyadic decomposition of a real number $x \in [0,1]$ consists in the unique binary sequence $(x_k)_{k\geq 1} \in \{0,1\}^{\mathbb{N}^*}$ such that

$$x = \sum_{k=1}^{+\infty} x_k 2^{-k}$$

and that the sequence is not stationary at the value 1. A dyadic rational is a real number for which there exists a dyadic decomposition with a finite number of non-zero terms, *i.e.* it is a finite sum of negative powers of 2. In the same way, any integer $m \geq 1$ admits a unique finite sequence $(m_k)_{k\geq 1} \in \{0,1\}^{\mathbb{N}^*}$ such that

$$m = \sum_{k=1}^{+\infty} m_k 2^{k-1}.$$

Furthermore, we define σ_m as the highest non-zero index of m. The Walsh functions provide a link between the integers and dyadic rationals. Indeed, given $m \in \mathbb{N}$, and $x \in [0,1]$, the m-th Walsh function is given by

$$w_m(x) = (-1)^{\langle m,x\rangle_{\text{bin}}},$$

where $\langle m,x\rangle_{\text{bin}} = \bigoplus_{k=1}^{\sigma_m} m_k x_k$ and $\bigoplus$ denotes the sum modulo 2 in $\mathbb{N}$.

Another useful property of the Walsh functions is that they form a Hilbert basis of $L^2(0,1)$ [85]. The reason why we are interested in this functional basis is that it allows the representation of arbitrary functions based on the dyadic decompositions of their input variables. For instance, as we shall see, this is useful when discretizing a function on a regular subdivision of $[0,1]$ with 2^N elements.

For now, let us focus on the representation of $H^1_\#(0,1)$ functions, that is L^2 functions on a periodic unit lattice with L^2 weak derivatives. We aim to prove the L^2 convergence of the N-bit truncated Walsh series, the representation of the function on the set of N-bit Walsh functions.

To do so, we first decompose, in Theorem 1, the Fourier basis functions $\mathrm{e}_k : x \mapsto \exp(i2\pi kx)$, for any $k \in \mathbb{Z}$, into the Walsh basis. That theorem states

Theorem (Walsh series of the Fourier basis functions)**.** *Let $k \in \mathbb{Z}^*$ and $K \in \mathbb{N}$ the multiplicity of 2 in the prime factorization of k. Then, for any $x \in \mathbb{R}$, we have*

$$\exp(i2\pi kx) = \sum_{m \in 2\mathbb{N}+1} [\operatorname{sign} k]^{|m|_{\mathrm{Ham}}} \gamma_{m,|k/2^K|} w_{2^K m}(x),$$

where $\operatorname{sign} k$ *is the sign of k, $|m|_{\mathrm{Ham}}$ is the so-called Hamming weight of m and the coefficients $\gamma_{m,k'}$ can be explicitly computed with respect to m and k'.*

Since the Fourier basis diagonalises the differential operators, this allows to bound, for any $v = \sum_{m\in\mathbb{N}} \tilde{v}_m w_m \in H^1_\#(0,1)$, the L^2 norm of the associated N-bit truncated Walsh series

$$\Theta_N v = \sum_{m=0}^{2^N-1} \tilde{v}_m w_m,$$

by its H^1 norm $\|v\|_{H^1}$. Indeed, as shown in Theorem 3, the following result holds.

Theorem (Walsh series convergence rate)**.** *For any $v \in H^1_\#(0,1)$ and any $N \geq 1$, the N-bit truncated Walsh series associated to v converges towards v in $L^2(0,1)$, with the error estimate*

$$\|\Theta_N v - v\|_{L^2} \leq \left(\frac{1}{\pi} + \frac{1}{2\sqrt{3}}\right) \frac{1}{2^N} \|v'\|_{L^2}.$$

To prove this result, we use the explicit formulation of the Walsh series of the Fourier basis functions to bound their L^2 norms and get the result through the Fourier series of v. This convergence result is then the first step to obtain a new Pauli decomposition for our diagonal operators in Chapter 2.

Summary of Chapter 2

This thesis focuses on the one-dimensional stationary Gross-Pitaevskii equation, a nonlinear differential equation that has been previously explored within the framework of variational quantum algorithms [68]. In this chapter, we determine the performance of variational quantum algorithms in the absence of quantum noise.

Problem discretization

The variational quantum algorithm we consider requires us to calculate the energy of a test function v at each optimization step. To do so, we must first encode the function on a quantum computer. In order to provide a solution to this problem, we are led to consider a Fourier subspace $\mathcal{X}_N = \operatorname{Span}_{\mathbb{C}}\big(\{\mathrm{e}_k, -2^{N-1} - 1 \leq k \leq 2^{N-1} - 1\} \cup \{c_{2^{N-1}}\}\big)$ of $\mathcal{X} = H^1_\#(0,1)$ where $c_{2^{N-1}} : x \mapsto \cos\big(2\pi 2^{N-1} x\big)$. For any $N \in \mathbb{N}^*$, we consider the regular subdivision

$$\Omega_N = \{\mathrm{x}_k = kh_N \mid 0 \leq k \leq 2^N\}$$

of Ω, of step $h_N = 1/2^N$.

For a function $v \in \mathcal{X}_N$, we denote by v^N the vector representing v at each point x_k of the subdivision. To represent v on a quantum system, the values of the vector v^N are encoded in the amplitudes of the quantum state in the following way, known as an amplitude encoding,

$$|\psi\rangle = \sqrt{h_N} v^N = \sqrt{h_N} \sum_{k=0}^{2^N-1} v_k^N |k\rangle ,$$

of v^N, where $\psi_k = \sqrt{h_N} v_k^N$ for any k, $0 \leq k < 2^N$.

For the sake of comparison, we introduce the discretized energies based on the finite difference approximation used in [68], the discretized minimization problem then reads

$$\left| \begin{array}{l} \text{Find } u_N \in \mathcal{X}_N \text{ such that} \\ E_N(u_N) = \min\limits_{v_N \in \mathcal{X}_N} E_N(v_N), \\ \|u_N\|_{L^2_N(\Omega)} = 1, \end{array} \right.$$

with

$$E_N(v_N) = \mathcal{K}_N(v_N) + \mathcal{P}_N(v_N) + \mathcal{I}_N(v_N) = -\frac{1}{2}\langle v, \Delta_N v\rangle_{L^2_N} + \langle V, v^2\rangle_{L^2_N} + \frac{\kappa}{2}\langle v^2, v^2\rangle_{L^2_N},$$

and where $\langle \cdot, \cdot \rangle_{L^2_N}$ is the so-called discrete L^2 scalar product and $\Delta_N = \frac{1}{h_N^2}(\tau_{h_N} + \tau_{-h_N} - 2I)$, with τ_a is the translation operator by the constant $a \in \mathbb{R}$, is the discretized finite-difference Laplace operator. Note that the amplitude encoding enforces the L^2_N normalisation constraint by default, *i.e.* $\langle \psi | \psi \rangle = h_N \sum_{k=0}^{2^N-1} |v_k^N|^2 = 1$.

Using a methodology similar to the one of [28], we show, in Appendix A.1, that the solution u_N to the discretized problem converges to the solution in H^1 norm as $O(1/M)$, where $M = 2^N$ is the number of points in Ω_N. Furthermore, the distance between any $v \in \mathcal{X}_N$ and the solution is bounded by the energy difference, *i.e.*

$$\|v - u_N\|^2_{H_1} \leq C \left(E_N(v) - E_N(u_N) \right).$$

This allows to determine the precision which will be required on the energy to achieve convergence of the solution to be of order of $1/M^2$.

For the energy measurements, we denote, for any operator $A \in \mathscr{L}(\mathcal{H})$, by $\langle A \rangle_\psi = \langle \psi | A | \psi \rangle$ the mean value of the operator for the state $|\psi\rangle$. This allows to write

$$\langle \mathcal{K}^N \rangle_\psi = \mathcal{K}_N(v_N), \qquad \langle \mathcal{P}^N \rangle_\psi = \mathcal{P}_N(v_N), \qquad \langle \mathcal{I}^N \rangle_\psi = \mathcal{I}_N(v_N).$$

Using this description based on $|\psi\rangle$, we propose new methods for the decomposition of such operators in the Pauli basis. These decompositions will allow to define Pauli estimators which we will use to recover the energies. For each term, we provide methods to obtain the Pauli decomposition of the discretized operator.

Pauli decomposition of the kinetic operator

For the kinetic term, while the Pauli decomposition was already obtained in [19, 54] we provide the Pauli decomposition of a larger class of matrices the Laplacian of which is a particular case. We introduce the N-qubit k-bidiagonal matrix B_k^N and provide its Pauli decomposition through

a geometrical argument. That leads to rewrite the kinetic term as

$$\mathcal{K}^N = \frac{1}{h_N^2} I^N - \frac{1}{2h_N^2} \left(B_1^N + B_{2^N-1}^N \right).$$

Then, using the Pauli decomposition of the bidiagonal matrices, we obtain in Proposition 3 the Pauli decomposition of $\mathcal{K}^N$.

Proposition (Pauli decomposition of the kinetic operator)**.** *The Pauli decomposition of $\mathcal{K}^N$ is given by*

$$\mathcal{K}^N = \alpha_0 I^N + \sum_{k=1}^{N_\mathcal{K}} \alpha_k P_k,$$

with $N_\mathcal{K} = 3 \cdot 2^{N-2}$, we have $\alpha_0 = \frac{1}{h_N^2}$ and, for any k, $1 \leq k \leq N_\mathcal{K}$,

$$|\alpha_k| = \begin{cases} \frac{1}{2h_N^2} 2^{-(|\operatorname{Supp}(P_k)|-2)} & \text{if } P_k|_1 = I, \\ \frac{1}{2h_N^2} 2^{-(N-2)} & \text{else.} \end{cases}$$

Pauli decomposition of the harmonic potential operator

The potential term in the finite-difference approximation context is obtained as the mean value of an N-qubit diagonal operator, as shown in [68],

$$\tilde{\mathcal{P}}^N = \operatorname{diag} \begin{bmatrix} V(\mathrm{x}_0) \\ \vdots \\ V(\mathrm{x}_{2^N-1}) \end{bmatrix} = \operatorname{diag} V^N.$$

The Pauli decomposition of such an operator can be obtained through a Walsh-Hadamard Transform (FWHT) for a $O(N \log N)$ classical cost [19]. In order to link the Pauli decomposition to the Walsh series of the function V, we are led to establish the relationship between Walsh functions and Pauli operators through the Hadamard vectors. We set, for any N-bit positive integer k, $P_k = Z_1^{k_1} \dots Z_N^{k_N}$, and notice that $P_k = \operatorname{diag} w_k^N$. Thiat leads to Proposition 6.

Proposition (Relationship between Pauli operators and Walsh functions)**.** *Given a nonnegative integer $k < 2^N$, let $|\psi\rangle = \sqrt{h_N} \sum_{j=0}^{2^N-1} v(\mathrm{x}_j) |j\rangle$, then,*

$$\langle P_k \rangle_\psi = \langle Z_1^{k_1} \dots Z_N^{k_N} \rangle_\psi = \langle v^2, w_k \rangle_{L_N^2}.$$

Finally, this proposition allows us to identify the Pauli decomposition of the operator with its Walsh interpolant, from Definition 8,

$$\tilde{\mathcal{P}}^N = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L_N^2} P_k, \qquad \mathrm{I}_N^\mathcal{W} V = \sum_{m=0}^{2^N-1} \langle V, w_m \rangle_{L_N^2} w_m.$$

We then introduce an analogous Pauli decomposition which we call the Walsh-Pauli decomposition, based, this time, on the N-bit truncated Walsh series of V,

$$\mathcal{P}^N = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L^2} P_k, \qquad \Theta_N V = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L^2} w_k.$$

This Walsh-Pauli decomposition has a reduced number of non-zero Pauli terms, and has the advantage of being obtainable directly from the Walsh series of the function. For $V = V_0(x - 1/2)^2$, we obtain the following proposition using the explicit Walsh series for monomials given in [25].

Proposition (Walsh-Pauli decomposition of the potential operator)**.** *For $V(x) = (x - \frac{1}{2})^2$, we have*

$$\mathcal{P}^N = \frac{V_0}{12}\left(I^N + 3\sum_{\ell_1=1,\ell_2>\ell_1}^{N} 2^{-(\ell_1+\ell_2-1)} Z_{\ell_1} Z_{\ell_2}\right),$$

and $N_{\mathcal{P}} = N(N-1)/2$.

We also prove, using some other convergence results, that the Walsh-Pauli decomposition converges fast enough to allow the required accuracy for convergence.

Pauli decomposition of the interaction operator

For the interaction term, using the methods devised for the potential operator we show that the interaction term admits a "squared Pauli" decomposition. Indeed, recall that, given $v \in \mathcal{X}$,

$$\mathcal{I}_N(v^N) = \frac{\kappa}{2}\langle v^2, v^2\rangle_{L^2_N} = \frac{\kappa}{2h_N}\sum_{i=0}^{2^N-1} |\psi_i|^4 = \langle \mathcal{I}^N\rangle_\psi.$$

Using the Walsh series of $v^2 = \sum_{k=0}^{\infty}\langle v^2, w_k\rangle_{L^2} w_k$, we obtain

$$\mathcal{I}(v) = \frac{\kappa}{2}\langle v^2, v^2\rangle_{L^2} = \frac{\kappa}{2}\sum_{k=0}^{+\infty} |\langle v^2, w_k\rangle_{L^2}|^2.$$

and by identifying the $\langle v^2, w_k\rangle_{L^2}$ terms with $\langle v^2, w_k\rangle_{L^2_N}$, we recover the Walsh-Pauli decomposition of the interaction term in Proposition 9.

Proposition (Squared Pauli decomposition of the interaction operator)**.** *The interaction energy term can be calculated as the following sum of squared Pauli means,*

$$\mathcal{I}_N(v_N) = \frac{\kappa}{2}\sum_{k=0}^{2^N-1} \langle P_k\rangle_\psi^2,$$

where, for any k, $0 \le k < 2^N$, $P_k = Z_1^{k_1} \dots Z_N^{k_N}$.

We also notice that, as it turns out, in the case of the interaction term, the Walsh-Pauli and Pauli decomposition coincide.

Pauli estimators

Endowed with the Pauli decompositions of the different energy operators, we now need to clarify how we define our Pauli estimators. To do so, we introduce a new framework, sampling strategies based on complete set of commuting observables (CSCO), which allows us to formalize the reconstruction of our three energy terms from samples in different bases.

Indeed, we defined sampling strategies to be general enough to define all our estimators, including the one associated to the interaction term, which is nonlinear.

We consider in this work two main types of sampling strategies for our Pauli estimators. On the one hand, the diagonal sampling consists in a change of basis and a measurement of the Z Pauli observable for each qubit from which we reconstruct eigenvalues of the observable for each sample. And on the other hand, for the importance sampling, where we estimate an observable using its Pauli decomposition, we more specifically sample each Pauli term proportionally to its coefficient in the decomposition. An important thing to note is that the change of basis requires an additional circuit. This implies that for non-diagonal operators, a consequential overhead in the number of gates, and thus a higher sensitivity to quantum noise, may be required for such a strategy, as we shall see in Chapter 3.

We then provide results concerning the general properties, most notably, the variance, of Pauli operators, this allows us to define Pauli estimators for each energy term, and determine their associated variance. For the kinetic term, given $N_{\mathrm{sh}} > 0$ samples, we consider an importance-sampling estimator,

$$\overline{\mathcal{K}^N}^{N_{\mathrm{sh}}} = \alpha_0 + \sum_{i=1}^{N_{\mathcal{K}}} \alpha_i \overline{P_i}^{N_{\mathrm{sh}}^i},$$

where each $\overline{P_i}^{N_{\mathrm{sh}}^i}$ is the empirical mean of the N_{sh}^i Pauli samples attributed to the operator P_i. and obtain its variance as a direct application of the variance of an importance-sampling Pauli estimator Proposition 11 to obtain Theorem 5.

Theorem (Kinetic term Pauli estimator variance)**.** *The sampling variance is upper-bounded by*

$$\operatorname{Var} \overline{\mathcal{K}^N}^{N_{\mathrm{sh}}} = \|\alpha_{1:}\|_1 \sum_{i=1}^{N_{\mathcal{K}}} \frac{|\alpha_i| \left(1 - \langle P_i \rangle_\psi^2\right)}{N_{\mathrm{sh}}} \leq \frac{\|\alpha_{1:}\|_1^2}{N_{\mathrm{sh}}} = \frac{N^2}{4h_N^4 N_{\mathrm{sh}}} = \frac{N^2 16^N}{4N_{\mathrm{sh}}}.$$

For the potential term, as the operator is diagonal for both Pauli decompositions, given $N_{\mathrm{sh}} > 0$, we consider a diagonal-sampling Pauli estimator

$$\overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} = \alpha_0 + \sum_{i=1}^{N_{\mathcal{P}}} \alpha_i \overline{P_i}^{N_{\mathrm{sh}}},$$

where each $\overline{P_i}^{N_{\mathrm{sh}}}$ is the empirical mean of the N_{sh} samples. Similarly, we obtain its variance, using Proposition 12, in Theorem 6.

Theorem (Potential term Pauli estimator variance)**.** *The variance of the Walsh-Pauli sampling estimator for the potential term satisfies, for any $N \geq 1$,*

$$\operatorname{Var} \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} = \frac{\langle (\Theta_N V)^2, v^2 \rangle_{L^2_N} - (\langle \Theta_N V, v^2 \rangle_{L^2_N})^2}{N_{\mathrm{sh}}} \leq \frac{\|\Theta_N V\|_{L^2}^2 - \|V\|_{L^1}^2}{N_{\mathrm{sh}}}.$$

Finally, for the interaction term, we need to introduce a new estimator $\overline{P_i^2}^{N_{\mathrm{sh}}}$ for the squared Pauli terms in the diagonal basis of the P_i which is the computational basis. This allows to introduce the interaction term estimator

$$\overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} = \alpha_0 + \sum_{i=1}^{N_{\mathcal{I}}} \alpha_0 \overline{P_i^2}^{N_{\mathrm{sh}}},$$

with $\alpha_0 = \kappa/2$. We then provide, in Theorem 7, the variance for the nonlinear interaction term estimator.

Theorem (Interaction term squared Pauli estimator variance)**.** *The Pauli sampling estimator of the interaction term has a variance given by*

$$\operatorname{Var}\overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} = \frac{\alpha_0^2}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)}\left[(6-4N_{\mathrm{sh}})\langle v^2, v^2\rangle_{L_N^2}^2 + 4(N_{\mathrm{sh}}-2)\langle v^4, v^2\rangle_{L_N^2} + 2^{N+1}\langle v^2, v^2\rangle_{L_N^2}\right].$$

Direct-sampling estimators

We compare the Pauli estimators to the estimators used in the 'hardware-reduced' algorithm defined in [68] which we denote as 'direct-sampling' estimators. We notice that, in the case of linear operators, these estimators coincide with the diagonal-sampling Pauli estimators. Furthermore, we show in Theorem 9 that the variance of diagonal/direct-sampling estimators for operators that come from discretized integrals can be expressed as follows.

Theorem (Diagonal/direct-sampling estimator bias and variance)**.** *Given a diagonal operator D such that, for any i, $0 \leq i < 2^N$, $D_i = d(i/2^N)$ for a function $d \in H^1_{\#}(0,1)$, the mean of the sampling estimator for D is*

$$\langle D\rangle_\psi = \langle v^2, d\rangle_{L_N^2},$$

and the variance of the estimator is given by

$$\operatorname{Var}\check{D}^{N_{\mathrm{sh}}} = \frac{\langle d^2, v^2\rangle_{L_N^2} - \langle d, v^2\rangle_{L_N^2}^2}{N_{\mathrm{sh}}}.$$

The variance is bounded by the difference of the discrete L^2 and L^1 norms of d

$$\operatorname{Var}\check{D}^{N_{\mathrm{sh}}} \leq \frac{1}{N_{\mathrm{sh}}}\left(\|d\|_{L_N^2}^2 - \|d\|_{L_N^1}^2\right),$$

where the discrete L^1 norm is defined, for any $v \in H^1$, by $\|v\|_{L_N^1} = h_N \sum_{k=0}^{2^N-1} |v(\mathrm{x}_k)|$.

For the nonlinear interaction term, we provide an unbiased version of the estimator given in [68] and show in Theorem 10, that the new unbiased estimator is equivalent to the Pauli estimator we introduced earlier.

Hadamard-test estimators

Lastly, we introduce the Hadamard-test estimators used in [68]. A Hadamard test is a type of circuit allowing the estimation of the mean value of unitaries through Z Pauli observable of the circuit's output qubit. Using this, given N_{sh} samples, we can define a Hadamard test estimator for an observable A as

$$\widehat{A}^{N_{\mathrm{sh}}} = c_{\mathrm{H}} + \alpha_{\mathrm{H}}\frac{1}{N_{\mathrm{sh}}}\sum_{k=1}^{N_{\mathrm{sh}}}\mathbf{h}^k,$$

where $\mathbf{h}^k$ is a sample of the associated Hadamard-test such that $\langle A\rangle_\psi = c_{\mathrm{H}} + \alpha_{\mathrm{H}}\langle Z_1\rangle_{\psi,\text{H-test}}$. Using such decomposition, [68] provides Hadamard-test estimators for each energy term. We provide variance estimates for these operators and show that there is a slight error in the variance of the potential term Hadamard-test estimator provided in [68], which we correct.

Estimator comparison and algorithmic performance

Using the results presented above, we are now able to provide a comparison between the different algorithms in the case of noiseless quantum hardware. In order to compare the performance of the estimators in practice, introduce a metric $N_{\mathrm{sh}}^{\epsilon}$, which is the number of samples required for the estimator to achieve an error $\epsilon > 0$ in the the worst case. For the linear estimators, this metric reduces to

$$N_{\mathrm{sh}}^{\epsilon} = \frac{\mathrm{Var}\,\mathcal{A}}{\epsilon^2},$$

while for the Pauli estimator of the interaction term operator, we obtain that the number of samples required is given by

$$N_{\mathrm{sh}}^{\epsilon} = \frac{1}{2}\left(1 + \frac{B_V}{\epsilon^2} + \sqrt{(1 - \frac{B_V}{\epsilon^2})^2 + 4\frac{B_W}{\epsilon^2}}\right).$$

We provide formulae, as well as upper bounds, for the "sampleless" variance $N_{\mathrm{sh}}^{\epsilon} \cdot \epsilon^2$ for each energy term. Using this, we show that the direct-sampling/Pauli diagonal-sampling methods outperform both other methods in all cases, while the Pauli importance-sampling method for the kinetic term induces a quadratic overhead in N^2.

We consider this metric for $\epsilon = \epsilon_{\mathrm{conv}} = C/M$, where $M = 2^N$ is the number of discretization points, *i.e.* $\mathrm{Card}(\Omega_N)$, the accuracy required for convergence in order to estimate practically how many samples would be required to obtain convergence in the worst case which is shown in Figure 1. We also provide on the right side of Figure 1 $N_{\mathrm{sh}}^{\epsilon}$ for a different discretization where, instead of considering a finite-difference discretization, we consider a spectral discretization on $\mathcal{X}_N$ which allows a $O(1/M^{3/2})$ convergence of the discretized solution.

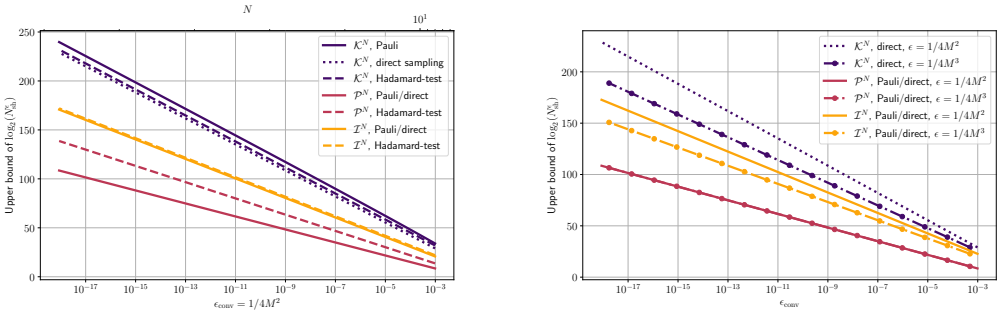


Figure 1 – Minimum number of shots to achieve convergence accuracy, $\epsilon = \epsilon_{\mathrm{conv}}$. For illustration purposes, we consider the case $V_0 = 1$, $\kappa = 1$. Left: finite-difference discretization, right: spectral discretization.

Summary of Chapter 3

In the third chapter, we study how quantum noise affects our algorithms. To do so, we provide some background on density matrices

$$\rho = \sum_s \mathrm{p}_s \left|\psi_s\right\rangle \left\langle\psi_s\right|,$$

which represent statistical distributions of quantum states and are hence able to represent the statistical uncertainty induced by quantum noise, and quantum channels with which we can describe the noise effect on a density matrix. We consider in this work a depolarizing noise model in which a depolarizing channel $\mathcal{D}_{p_1}$ is added after each 1-qubit gate, and a depolarizing channel $\mathcal{D}_{p_2}$ is added after each 2−qubit gate, where $0 \leq p_1 \leq p_2 \leq 1$ are depolarization probabilities. Using the current hardware fidelity, we estimate these probabilites to be approximately

$$p_1 = 0.0002, \qquad p_2 = 0.001.$$

How noise affects estimators

The effect of the noise model on a quantum circuit is described in Lemma 6 and allows to determine how the noise affects the mean value of an observable, as we develop in Proposition 14 it introduces a bias.

Proposition (Noise effect on observables)**.** *Let a Hermitian operator O, given a unitary U_ψ such that $U_\psi |0\rangle = |\psi\rangle$ and a circuit implementing U_ψ with N_{G_1} 1-qubit gates and N_{G_2} 2-qubit gates. We obtain, for the density matrix ρ,*

$$\langle O\rangle_\rho = q\langle O\rangle_\psi + (1-q)\frac{\operatorname{Tr} O}{2^N} = \langle O\rangle_\psi - (1-q)\langle O\rangle_\psi + (1-q)\lambda_0.$$

The noise thus introduces the nonnegative bias

$$b^O = (1-q)|\langle O\rangle_\psi - \lambda_0|,$$

where

$$q = (1-p_1)^{N_{G_1}}(1-p_2)^{N_{G_2}}.$$

The effect of noise on the mean value of observable can be seen as a contraction with Lipschitz constant $q \in (0,1)$. This first allows to determine the noise effects on individual Pauli samples, and eventually the one on Pauli estimators for arbitrary sampling strategies in Proposition 16.

Proposition (Noisy Pauli estimator mean and variance)**.** *The mean of a Pauli sampling estimator $\overline{A}^{N_{\mathrm{sh}}}$ of an operator $A = \alpha_0 I + \sum_{i=1}^{N_A} \alpha_i P_i$ under global depolarizing noise is given by*

$$\mathbb{E}_\rho\left[\overline{A}^{N_{\mathrm{sh}}}\right] = (1-q)\alpha_0 + q\langle A\rangle_\psi,$$

and its variance is given by

$$\operatorname{Var}_\rho \overline{A}^{N_{\mathrm{sh}}} = q^2 \operatorname{Var}_\psi \overline{A}^{N_{\mathrm{sh}}} + (1-q^2)\sum_{i=1}^{N_A}\frac{\alpha_i^2}{|S_i|} + (1-q)q\sum_{i=1}^{N_A}\sum_{j=1,j\neq i}^{N_A}\frac{|S_i\cap S_j|}{|S_i||S_j|}\alpha_i\alpha_j\langle P_iP_j\rangle_\psi.$$

This allows to determine how the kinetic and potential term estimators are affected. Using the same methods, we find the noise effect on the interaction term Pauli estimator in Proposition 19.

Proposition (Noisy interaction term Pauli estimator variance and bounds)**.** *The bias of the interaction term Pauli estimator using a diagonal-sampling strategy is given by*

$$b_{\mathrm{P},\mathcal{I}} = (1-q^2)|\langle \mathcal{I} - \alpha_0 I^N\rangle_\psi|,$$

and its variance by

$$\begin{aligned}\mathrm{Var}_\rho \overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} &= q^4 \,\mathrm{Var}_\psi \overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} + (1-q^4)\frac{2(2^N-1)\alpha_0^2}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)} \\ &\quad + q^3(1-q)\frac{4(N_{\mathrm{sh}}-2)}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)}\alpha_0^2\left(\langle v^4, v^2\rangle_{L_N^2} - 3\langle v^2, v^2\rangle_{L_N^2} + 2\langle v^2, 1\rangle_{L_N^2}^2\right) \\ &\quad + q^2(1-q^2)\frac{4(N_{\mathrm{sh}}-2)+2(2^N-2)}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)}\alpha_0^2\left(\langle v^2, v^2\rangle_{L_N^2} - \langle v^2, 1\rangle_{L_N^2}^2\right).\end{aligned}$$

We also have the upper bound

$$\mathrm{Var}_\rho \overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} \leq \frac{B_{V,q}}{N_{\mathrm{sh}}} + \frac{B_{W,q}}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)},$$

for some positive constants $B_{V,q}$ (3.22), $B_{W,q}$ (3.23), only depending on α_0, q and N.

Finally, by applying the previous proposition to the Hadamard-test estimators, we describe how noise affects them in Proposition 20.

Circuit size estimation

As we have seen however, the Lipschitz constant of the noise contraction depends on the number of gates in the associated circuit of the estimator. We thus provide estimates on the number of gates required for the estimation circuits. To do so, we consider two ansatz models, the U2-ansatz based on disentangler methods from [83, 84] which allow to relate the number of ansatz layers N_{lay} with the maximum bond dimension of the tensor train of the represented function v, and the Hardware-Efficient (HE) ansatz which is a staple of VQAs. We provide estimates of the scaling of the number of one-qubit and two-qubit gates for each ansatz and then go on to estimate the circuit size of the additional subroutines from the Hadamard-test methods, as well as that of the QFT required for the diagonal sampling of the kinetic term. These are shown in Table 1 for the disentrangler ansatz.

Table 1 – Circuit size estimation

Disentangler ansatz	$\mathcal{K}$	$\mathcal{P}$	$\mathcal{I}$
Pauli - D N_{G_1}	$15N_{\mathrm{lay}}(N-1)+N$	$15N_{\mathrm{lay}}(N-1)$,	$15N_{\mathrm{lay}}(N-1)$
Pauli - D N_{G_2}	$3N_{\mathrm{lay}}(N-1)$	$3N_{\mathrm{lay}}(N-1)$	$3N_{\mathrm{lay}}(N-1)$
Direct - D N_{G_1}	$15N_{\mathrm{lay}}(N-1)+N$	$15N_{\mathrm{lay}}(N-1)$	$15N_{\mathrm{lay}}(N-1)$
Direct - D N_{G_2}	$(3N_{\mathrm{lay}}+N/2)(N-1)$	$3N_{\mathrm{lay}}(N-1)$	$3N_{\mathrm{lay}}(N-1)$
H-test - D N_{G_1}	$15N_{\mathrm{lay}}(N-1)+2$	$15N_{\mathrm{lay}}(N-1)+2$	$15N_{\mathrm{lay}}(N-1)+2$
H-test - D N_{G_2}	$(3N_{\mathrm{lay}}+5N)(N-1)+3N$	$(21N_{\mathrm{lay}})(N-1)+N$	$(39N_{\mathrm{lay}})(N-1)+2N$

Bias-induced maximum number of gates

The mean squared error of the estimation can be divided into two parts, a fixed one due to the bias term and one due to the variance, which decreases with the number of samples. The fixed error due to the bias implies that we can relate, for a given error ϵ, the bias to a maximum number of gates in the circuit. This is exemplified in the Figure 2 where we can see that the number of gates to achieve convergence error decreases dramatically as the number of qubits increase. As can be seen in Table 1 the number of usable gates is insufficient and prevents the use of the algorithm. To overcome this difficulty, we explain how to resort to quantum error

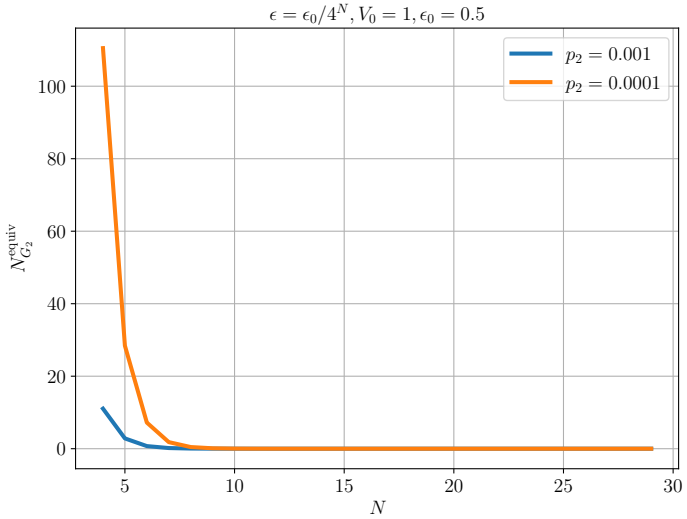


Figure 2 – Maximum number of 2-qubit gates usable under depolarization noise to achieve a bias under $\epsilon_{\text{conv}} = 1/2^N$.

mitigation to correct the bias term.

Error mitigation

Quantum error mitigation [14] allows to correct the bias of our estimators at the cost of an increased variance. We show that in the case of depolarization noise increase in variance is proportional to the squared inverse of the Lipschitz constant q and as such is exponential in the number of gates in the estimation circuits. This is illustrated in Figure 3, the red line shown in the figure represents $\frac{1}{\epsilon_{\text{conv}}^2} = M^4 = 16^N$ which is the number of samples required for ϵ_{conv} in the case of a variance of 1. This illustrates that to obtain an effect of noise on the number of samples of the same order as the increase due to precision, upwards of 200 qubits are required. And, as we have observed in Figure 1, the number of samples is already prohibitive for $N < 30$, requiring more than the age of the universe to estimate the potential energy, the term requiring the least number of samples.

To summarize, Chapter 1 introduces fundamental quantum computing concepts and the Walsh functional basis, with new results on Fourier-Walsh representation and convergence rates. Chapter 2 examines the algorithm's performance without noise, focusing on convergence rates, alternative estimators with lower circuit depths, and the required number of samples for energy precision.

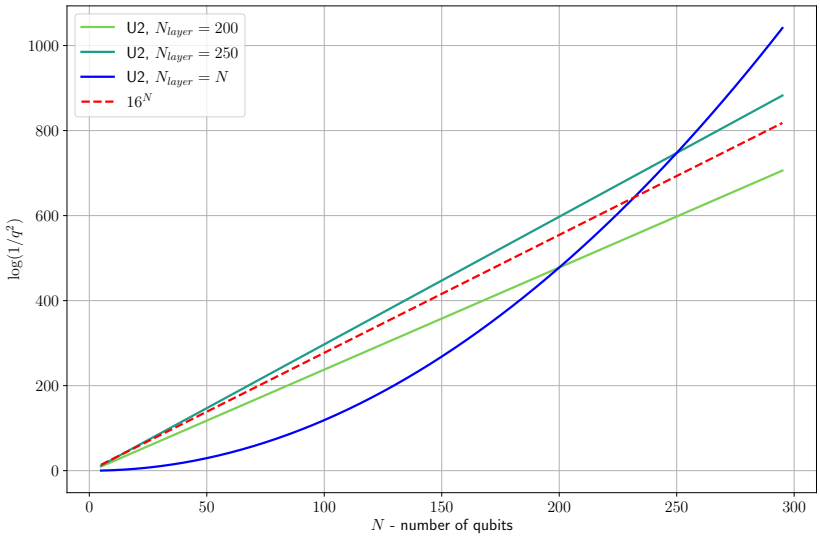


Figure 3 – Plot of the mitigation overhead q^{-2} for different number of layers of the U2 ansatz for $p_2 = 0.001$.

Chapter 3 assesses the impact of noise, analyzing how it affects estimator bias, variance, and circuit performance, with and without error mitigation.

Chapter 1

Quantum computing context, Walsh series

Quantum computation is a field that holds much promise for solving complex problems more efficiently than classical computers. In this manuscript, we explore a resolution method for differential equations on noisy quantum computers, a domain at the intersection of quantum mechanics, computer science, and applied mathematics.

In order to provide a solid foundation for this, the first chapter is dedicated to presenting key definitions and results which are crucial to understand the topics and methods discussed in the subsequent chapters. This chapter is divided into two sections, one on the basics of quantum computation and another one on the Walsh functional basis which is intricately linked to the Pauli matrices. In the latter part, we focus on our novel contributions in two primary areas: the representation of Fourier functions within the Walsh basis and an analysis of the convergence rate of truncated Walsh series for H^1 functions.

1.1 Quantum computing for differential equations

Quantum computing (QC) is built upon the principles of quantum mechanics, where quantum systems serve as the fundamental units of information. Unlike their classical counterparts, quantum systems can exist in superpositions of states, enabling a form of parallelism in computations through the interference of superposed states. This difference allows quantum algorithms with known quantum computational advantage [73] with respect to their classical counterpart, but at the cost of requiring noiseless or fault-tolerant quantum computers. However, by their very nature, quantum systems are extremely sensitive and thus require quantum error correction in order to palliate the effects of quantum noise. Quantum error correction is a field attracting significant interest and in the last few years has seen steady progress, most notably a first physical demonstration of error reduction and new methods reducing the number of physical qubits required to implement an error-corrected logical qubit. While error-correction on a large-scale will probably not be available in the near future, they offer well defined advantages. Let us, thus, first consider the case of quantum algorithms for error-corrected hardware and then go into the heart of our subject, near-term quantum algorithms. In the context of quantum algorithms with error correction, several avenues have been explored for the resolution of partial differential equations on quantum computing devices. Quantum algorithms dedicated to solving partial differential equations can be divided into two types.

The first method consists in using the quantum advantage offered for solving linear systems of equations. The second consists in encoding the differential equation into a Hamiltonian form and using the ability of quantum computers to efficiently simulate Hamiltonian evolution.

Quantum linear systems algorithms Quantum linear systems algorithms (QLSA), such as the HHL algorithm [42], [4, 5, 23, 36], are able to solve linear systems of equations exponentially faster than classical methods under certain conditions of implementability in terms of circuit. These algorithms require what is known as Quantum Random Access Memory (QRAM) which allows the storage of the data over long periods of time. QRAM is for now a theoretical device as no implementation method currently exists. QLSA-based differential equation solvers [7, 22, 57, 62] consist in solving, thanks to QLSAs, large linear systems resulting from the discretization of differential equations, generally both in space and time variables. One caveat of this method is the fact that the resulting solution for all times is encoded in a quantum state and extracting the information as classical data is an extremely inefficient process, and thus the algorithms are most useful for the calculation of observables [50].

Hamiltonian simulation Hamiltonian simulation based methods [6, 29, 48, 51, 52] involves simulating the time evolution of quantum systems as described by the Schrödinger equation

$$i\partial_t \psi = \hat{H}\psi,$$

with potentially non-Hermitian Hamiltonians. It is of note however that if the Hamiltonian is not Hermitian, the evolution operator is not unitary and block-encoding methods [61] are required to approximate the evolution operator. While QLSA-based algorithms are restricted to digital quantum computers, there have been proposals to port Hamiltonian simulation methods to analog quantum computers [49]. Generally, for Hamiltonian simulation methods on digital quantum computers, the space variables are discretized and the evolution operator is approximated through Trotterization methods, while in the analog case the space variables do not need to be discretized.

Nonlinear DEs Both previous methods allow to handle nonlinearities by mapping nonlinear differential equations into linear forms. This is done either approximately through linearization procedures, which are limited in their ability to represent nonlinearities, or through exactly equivalent higher-dimensional linear representation of the equations, which are adapted to specific equations, such as the Hamilton-Jacobi equations through the level set method [50] or the Koopman-von Neumann method [53] which allows to map systems of nonlinear ordinary differential equations to linear transport PDEs. Nonlinearities have been treated, in the case of linearization procedures, through Carleman linearization methods [64, 65], by solving a truncation of the resulting infinite-dimensional linear systems of equations, or through "mean-field" methods [66, 90] consisting in encoding the nonlinearity as the mean-field limit of a linear higher-dimensional system of interacting particles.

Noisy Intermediate-Scale Quantum paradigm The era of error-corrected quantum computing is, however, still far off as even the least qubit-greedy error correcting codes still require hundreds of physical qubits to encode a single logical qubit, while quantum algorithms often require on the same order of logical qubits. Current quantum systems are composed of at most several hundred physical qubits and are thus not suited to error correction or Fault-Tolerant algorithms.

Noisy Intermediate-Scale Quantum (NISQ) devices, characterized by their limited number of qubits and susceptibility to noise [81], are the current state of quantum hardware. Despite their limitations, they present unique opportunities for solving specific problems through innovative algorithmic techniques tailored to their constraints.

The NISQ paradigm is concerned with such quantum algorithms that may be able to provide a quantum advantage even in the presence of noise, either directly or through what is known as error mitigation.

NISQ algorithms [9] typically involve hybrid quantum-classical approaches, where quantum circuits are used to perform the core computations, and classical optimization techniques are employed to refine the results. Examples include variational quantum algorithms (VQAs) such as the Variational Quantum Eigensolver (VQE) [79] and the Quantum Approximate Optimization Algorithm (QAOA) [30]. These methods leverage the strengths of both quantum and classical computing to try to achieve results that neither could efficiently accomplish alone.

Error mitigation techniques [14], such as zero-noise extrapolation [37, 55, 56, 59, 89], quasi-probabilistic error cancellation [24, 67, 89] and subspace-expansion methods [70, 100], are also crucial in the NISQ era, as they allow the use of certain algorithms even under the influence of quantum noise. These methods aim to reduce the impact of noise without the overhead of full error correction, making it feasible [14, 98] to achieve meaningful proof-of-concept results with current quantum hardware through classical post-processing.

In summary, the NISQ paradigm focuses on maximizing the utility of existing quantum devices, exploring the boundaries of what can be achieved with imperfect quantum systems. This approach not only attempts to determine the immediate benefits current hardware could provide but also paves the way for future advancements in quantum computing as hardware continues to improve.

NISQ algorithms for differential equations Several types of algorithms have been proposed to solve differential equations, including Hamiltonian simulation [19, 86], Quantum walks [101] and Quantum machine learning methods [58, 76, 77] based on quantum circuit learning [72]. However, in this thesis we will take a closer look at a specific type of NISQ algorithms, Variational quantum algorithms, which originate and have had some success, in quantum chemistry [31, 39, 47, 74, 79, 88].

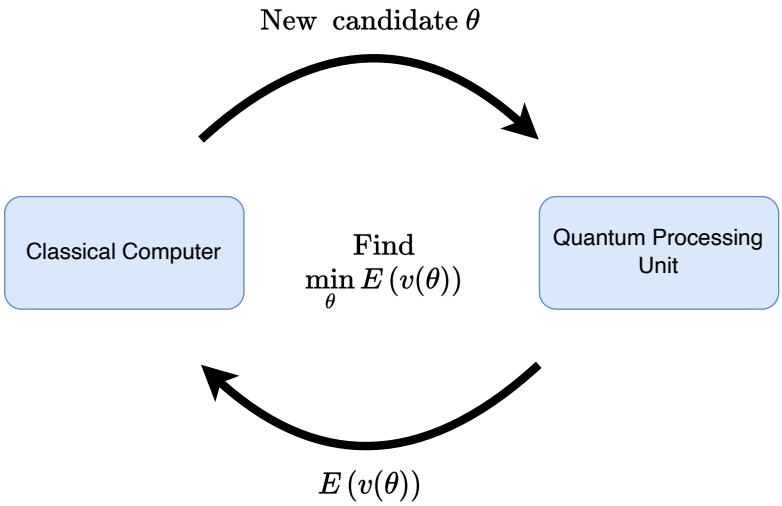


Figure 1.1 – Variational quantum algorithms use classical optimization techniques to minimize an energy estimated on quantum devices.

Variational Quantum Algorithms Variational Quantum Algorithms (VQAs) [18, 71, 79, 91] are hybrid quantum-classical algorithms which consist in using classical optimisation loops to minimize an energy $E(\psi(\theta))$ which depends on a parametric ansatz $\psi(\theta)$. The quantum part of the algorithm consists in using parametric quantum circuits to construct an ansatz state $|\psi(\theta)\rangle$ and measure observables allowing the estimation of the energy $E(\psi(\theta))$ through a quantum processor. Based on the energy calculated on the quantum processor, classical methods are used to update the parameters θ in order to minimize the energy, as illustrated in Figure 1.1. As the circuits required to obtain those expected values are generally shallower than algorithms designed for error-corrected QC, requiring only a state preparation circuit and the circuit associated with the measurement method, this computational paradigm enables the utilization of near-term NISQ hardware. We can decompose the quantum side of such algorithms into two parts, ansatz representativity and trainability on the one hand, and energy estimation on the other hand. First, ansatz representativity is the ability of an ansatz circuit to represent with sufficient accuracy the manifold of states near the target solution. Ansatz trainability refers to the problem of being able to find efficiently the set of parameters optimising a given functional. These two points are intricately related and it has been shown that there is a tradeoff between the two, namely ansatz with high representativity tend to have 'barren plateaus' [43, 82] in their gradient landscape where the magnitude of the gradient goes to zero in all directions. Furthermore, recent work [17] hints at the fact that ansatz with good trainability may be efficiently simulated classically and thus putting into question whether VQAs using such ansatz may offer any computational advantage. Two things are to be noted however, first while an ansatz may not be efficiently trainable over its complete set of parameters, starting from a carefully selected set of parameters constructed using a-priori knowledge of the problem being solved may allow to dodge barren plateaus[2]. Second, using classical simulations to obtain a first approximation and then continuing the optimization using VQAs with non-efficiently trainable ansatz might allow for some advantage. Energy estimation for variational quantum algorithms is generally obtained through the empirical mean of observables measured on the quantum system, although some recent alternatives such as classical shadows [41, 44, 45] and statistical phase estimation [10, 27, 96] methods may lead to

better convergence rates.

PDEs can often be reformulated as the Euler-Lagrange equation associated to an energy functional, *i.e.* a zero of its functional derivative in the Gâteaux sense. Due to the ubiquity of variational methods for partial differential equations, several variational quantum algorithms have been proposed for solving differential equations [2, 34, 63, 68, 80]. Such algorithms appear to be most promising for nonlinear problems, as for linear ones it is less clear that a quantum advantage could be obtained without fault-tolerant QLSA. Indeed, the main advantage of quantum variational methods for solving nonlinear differential equations is that they do not require any form of linearisation. Their only requirement is the ability to efficiently estimate the energy functional.

The advantages of VQAs can thus be summarised as follows:

- potential memory advantage;
- no linearization requirement;
- less susceptible to quantum noise.

The main disadvantages on the other hand are the following:

- optimisation procedure and trainability of ansatz;
- representativity of ansatz;
- unknown computational complexity.

In the context of this thesis, we shall study the performance of such variational quantum algorithms in practice, namely, we shall specifically answer the question of computational complexity. To do so, we now provide some definitions and mathematical context of quantum computation.

1.1.1 Definitions

The state of quantum systems are represented through their wavefunctions or probability amplitudes which are vectors in a Hilbert space. In the context of quantum computing, we consider abstractions of quantum systems with two-dimensional Hilbert spaces called qubits. Let us now give a description of the mathematical formalism behind systems composed of multiple qubits.

We consider a *quantum system* composed of $N \in \mathbb{N}^*$ qubits. For $1 \leq n \leq N$, we denote by $\mathcal{H}_n$ the Hilbert space associated to the n-th qubit. Each Hilbert space $\mathcal{H}_n$ can be seen as the complex projective line $\mathbb{CP}^1$, also known as the Bloch sphere [73]. Then we name $\mathcal{H} = \otimes_{n=1}^{N} \mathcal{H}_n$ the tensor product of the individual Hilbert spaces $\mathcal{H}_n$, $1 \leq n \leq N$. The *states of the quantum system* are the elements of $\mathcal{H}$. We also denote by $\mathcal{H}^*$ the dual space of the finite-dimensional space $\mathcal{H}$, and by $\mathscr{L}(\mathcal{H})$ the space of linear applications on $\mathcal{H}$, which is isomorphic to $\mathcal{H} \otimes \mathcal{H}^*$.

The notation $|\cdot\rangle$ is used to represent a (column) vector in the Hilbert space $\mathcal{H}$. Similarly, $\langle\cdot|$ is used to represent a (row) vector in $\mathcal{H}^*$.

The notations $|i\rangle\,|j\rangle$, respectively $\langle i|\,\langle j|$, are shorthands for the tensor product of $|i\rangle$ and $|j\rangle$, respectively $\langle i|$ and $\langle j|$. We denote the canonical basis of $\mathbb{CP}^1$, respectively $(\mathbb{CP}^1)^*$ as

$$\mathscr{B}_{\mathrm{C}}(\mathbb{CP}^1) = (|0\rangle\,, |1\rangle), \quad \text{respectively } \mathscr{B}_{\mathrm{C}}((\mathbb{CP}^1)^*) = (\langle 0|\,, \langle 1|).$$

and the canonical basis of $\mathcal{H}$, respectively $\mathcal{H}^*$, as

$$\mathscr{B}_{\mathrm{C}}(\mathcal{H}) = \mathscr{B}_{\mathrm{C}}(\mathcal{H}_1) \otimes \cdots \otimes \mathscr{B}_{\mathrm{C}}(\mathcal{H}_N), \quad \text{respectively } \mathscr{B}_{\mathrm{C}}(\mathcal{H}^*) = \mathscr{B}_{\mathrm{C}}(\mathcal{H}_1^*) \otimes \cdots \otimes \mathscr{B}_{\mathrm{C}}(\mathcal{H}_N^*).$$

The states in the computational basis are denoted as $|i\rangle$ and their Hermitian conjugate $\langle i|$, for $0 \leq i \leq 2^N - 1$. They correspond to the product state of their associated representation in base 2

using the most-significant-bit-first convention. In other words, writing

$$i = \sum_{n=1}^{N} i_n \cdot 2^{N-n}, \tag{1.1}$$

with i_1, ...,$i_N \in \{0, 1\}$, we have

$$|i\rangle = |i_1\rangle \otimes \cdots \otimes |i_N\rangle = |i_1\rangle \dots |i_N\rangle .$$

A useful implication of this notation is that we can also handle linear applications on the Hilbert space $\mathcal{H}$ through their decomposition on the canonical basis of $\mathscr{L}(\mathcal{H})$. For instance, the canonical basis of $\mathscr{L}(\mathbb{CP}^1)$ is given by

$$\mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathbb{CP}^1)) = \mathscr{B}_{\mathrm{C}}(\mathbb{CP}^1) \otimes \mathscr{B}_{\mathrm{C}}((\mathbb{CP}^1)^*) = (|i\rangle \langle j|)_{i,j\in\{0,1\}} .$$

And, more generally, the canonical basis of $\mathscr{L}(\mathcal{H})$ is

$$\mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathcal{H})) = \mathscr{B}_{\mathrm{C}}(\mathcal{H}) \otimes \mathscr{B}_{\mathrm{C}}(\mathcal{H}^*) = (|i\rangle \langle j|)_{i,j\in\{0,\dots,2^N-1\}} .$$

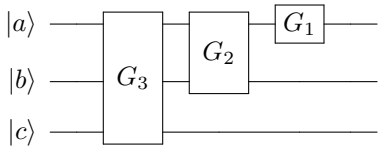


Figure 1.2 – Simple quantum circuit with three qubits starting in the state $|a\rangle \otimes |b\rangle \otimes |c\rangle$ and being applied, in succession, gates G_3, G_2 and G_1 with respective arity 3, 2 and 1.

A *quantum gate* G of arity γ, $1 \leq \gamma \leq N$, is defined as a couple $(U, \mathcal{S})$, where $\mathcal{S}$ is a γ-tuple of ascending-ordered elements of $\{1, \dots, N\}$ without repetition, and $U \in \mathbb{C}^{2^\gamma \times 2^\gamma}$ is a unitary matrix. The *gate support* $\mathcal{S}$ is the subset of qubits of the quantum system which the gate acts on. Then we denote by $\overline{\mathcal{S}}$ the $(N-\gamma)$-tuple of the ascending-ordered elements of $\{1, \dots, N\}$ which do not appear in $\mathcal{S}$. The subspaces of $\mathcal{H}$ defined by

$$\mathcal{H}_{\mathcal{S}} = \bigotimes_{j=1}^{\gamma} \mathcal{H}_{\mathcal{S}_j}, \qquad \mathcal{H}_{\overline{\mathcal{S}}} = \bigotimes_{j=1}^{N-\gamma} \mathcal{H}_{\overline{\mathcal{S}}_j}$$

are respectively the Hilbert spaces of the (ordered) subsystem composed of the qubits in $\mathcal{S}$ and $\overline{\mathcal{S}}$. We observe that $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\overline{\mathcal{S}}} \simeq \mathcal{H}$.

Besides, the canonical basis $\mathscr{B}_{\mathrm{C}}(\mathcal{H}_{\mathcal{S}})$ of $\mathcal{H}_{\mathcal{S}}$ is the tensor product of the canonical bases of its component Hilbert spaces in the tensoring order, *i.e.*

$$\mathscr{B}_{\mathrm{C}}(\mathcal{H}_{\mathcal{S}}) = \mathscr{B}_{\mathrm{C}}(\mathcal{H}_{\mathcal{S}_1}) \otimes \cdots \otimes \mathscr{B}_{\mathrm{C}}(\mathcal{H}_{\mathcal{S}_\gamma}) = (|i_1\rangle \otimes \cdots \otimes |i_\gamma\rangle)_{i\in\{0,\dots,2^\gamma-1\}}$$

where, in the previous equality, the indices i_n, $1 \leq n \leq \gamma$, provide the binary decomposition of $i \in \{0, 2^\gamma - 1\}$ as $i = \sum_{n=1}^{\gamma} i_n \cdot 2^{\gamma-n}$, following the most-significant-bit first convention (1.1) for the ordered bits in $\mathcal{S}$. The canonical basis for $\mathcal{H}_{\mathcal{S}}^*$ is defined similarly. In the same way, the canonical basis of $\mathscr{L}(\mathcal{H}_{\mathcal{S}})$ is given by $\mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathcal{H}_{\mathcal{S}})) = \mathscr{B}_{\mathrm{C}}(\mathcal{H}_{\mathcal{S}}) \otimes \mathscr{B}_{\mathrm{C}}(\mathcal{H}_{\mathcal{S}}^*)$, and similarly for $\overline{\mathcal{S}}$.

In $\mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathcal{H}_{\mathcal{S}}))$, the gate action on the subsystem is given by the unitary matrix U. Thus the action of gate G on $\mathcal{H}$ in the basis $\mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathcal{H}_{\mathcal{S}})) \otimes \mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathcal{H}_{\overline{\mathcal{S}}}))$ becomes $U \otimes I_{\mathcal{H}_{\overline{\mathcal{S}}}}$. We denote by $U_G \in \mathbb{C}^{2^N \times 2^N}$ the unitary matrix representing the action of gate G in the canonical basis $\mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathcal{H}))$, which coincides with $U \otimes I_{\mathcal{H}_{\overline{\mathcal{S}}}}$ up to a qubit-reordering basis change.

A *quantum circuit* is defined as an ordered N_G-tuple $C = (G_{C,k})_{1 \leq k \leq N_G}$ of gates, where $N_G \in \mathbb{N}^*$ is the number of gates in the circuit. Whenever there is no ambiguity, we shall name U_k the representative of $G_{C,k}$ in $\mathscr{B}_{\mathrm{C}}(\mathscr{L}(\mathcal{H}))$, and $\mathcal{S}_k$ its support. Then the unitary matrix associated to the circuit is

$$U_C = U_{N_G} \dots U_1. \tag{1.2}$$

A set of common quantum gates is illustrated in Table 1.1.

Furthermore, as current quantum hardware is only able to implement one or two-qubit gates, we limit ourselves to the cases where we only have 1 or 2-qubit gates, i.e. $|\mathcal{S}_k| \in \{1, 2\}$, we thus set N_{G_1} (resp N_{G_2}) the number of 1 (resp. 2) qubit gates in the circuit. This gate restriction is not a restriction on the unitaries that we are able to implement. Indeed, the Solovay-Kitaev theorem [73] states that any unitary matrix can be approximated to an arbitrary precision by a quantum circuit with a universal set of 1 and 2-qubit gates.

Pauli basis

The set of local observables that can be directly measured on a quantum system are given by the Pauli operators. The strategy we develop in this thesis involves the Pauli basis of $\mathscr{L}(\mathcal{H})$. It is built from the 1-qubit Pauli basis $\mathscr{B}_{\mathrm{P}}(\mathscr{L}(\mathbb{CP}_1)) = (I, X, Y, Z)$ of $\mathscr{L}(\mathbb{CP}_1)$, where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Recall that $\mathbb{CP}_1$ is the complex projective line, which is isomorphic to the Hilbert space associated to a 2-state quantum system. The Pauli basis $\mathscr{B}_{\mathrm{P}}(\mathscr{L}(\mathcal{H}))$ of $\mathcal{H}$ is composed of the tensor products of the Pauli matrices on the qubits composing $\mathcal{H}$. For any $Q \in \{X, Y, Z\}$, we set

$$Q_j = I \otimes I \cdots \otimes \underbrace{Q}_{j\text{-th qubit}} \otimes I \otimes \cdots \otimes I,$$

which only acts on the j-th qubit, $1 \leq j \leq N$.

For a Pauli matrix $P \in \mathscr{B}_{\mathrm{P}}(\mathscr{L}(\mathcal{H}))$, its component in the Pauli basis of the qubit i, $1 \leq i \leq N$, is written $P|_i$, and we denote the support of a Pauli observable

$$\operatorname{Supp} P = \{i \in \{1, \dots, N\} \mid P|_i \neq I\}.$$

1.1.2 Quantum measurements

Measuring an observable, *i.e.* a Hermitian operator O, on a quantum system differs from classical measurements. Indeed, instead of measuring the value $\langle\psi| O |\psi\rangle$ of the operator for a specific state $|\psi\rangle$, the measure of an observable O returns an eigenvalue λ of the observable O and projects the system into the associated eigenstate $|\phi\rangle \in \mathcal{H}$. More precisely, let us denote by Λ the ordered spectrum of O with multiplicity. For any $\lambda \in \Lambda$, with multiplicity $\mathrm{mult}(\lambda)$, the probability of measuring λ for a given state $|\psi\rangle$ is given by the sum of squared overlaps $\sum_{k=1}^{\mathrm{mult}(\lambda)} |\langle\phi_\lambda|\psi\rangle|^2$ of $|\psi\rangle$ for the set of orthogonal eigenstates $\{|\phi_\lambda^k\rangle\}_{1 \leq k \leq \mathrm{mult}(\lambda)}$ associated to λ. Since we only have access to samples of the eigenvalues distributed following the law given by the squared overlap

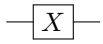
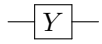
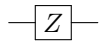
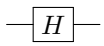
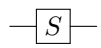
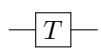
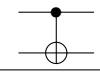
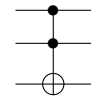
Gate	Symbol	Matrix Representation	Circuit Representation
Identity	I	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	
Pauli-X (NOT)	X	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	
Pauli-Y	Y	$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$	
Pauli-Z	Z	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	
Hadamard	H	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	
Phase	S	$\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$	
$\pi/8$	T	$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$	
CNOT	CX	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	
Swap	SWAP	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	
Toffoli	CCX	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	

Table 1.1 – Common quantum gates, their matrix Representations, and circuit representations

with the eigenstates, we cannot directly measure the mean value of the operator $\langle\psi| O |\psi\rangle$ for $|\psi\rangle$. However, we can obtain an approximation of $\langle\psi| O |\psi\rangle$ by using $N_s \in \mathbb{N}^*$ measurement samples $\left(s_k^O\right)_{1\leq k\leq N_s} \in \Lambda^{N_s}$ and considering the empirical mean estimator

$$\overline{O}^{N_s} = \frac{1}{N_s}\sum_{k=1}^{N_s} s_k^O.$$

Quantum systems are generally restricted to a certain subset of possible observables to measure due to physical constraints. For qubits, the Pauli observable for which a direct measurement is possible is defined to be the Z Pauli operator. In practice, we can indeed only measure the value of the local Pauli Z operator Z_k for each individual qubit k, $1 \leq k \leq N$. This means that for a N-qubit system, for each $1 \leq k \leq N$, we measure $z_k \in \{1, -1\}$ which corresponds to the eigenvalues of Z for the k-th qubit. Equivalently, we can consider also the binary samples $s_k \in \{0, 1\}$, for $1 \leq k \leq N$, associated with the local eigenvectors $|0\rangle$, and $|1\rangle$ on the k-th qubit

associated to the eigenvalues of Z_k.

To measure the values of the non Z Pauli matrices $P \in \{X, Y\}$, we apply the suitable local rotations $Q_X = H, Q_Y = HS$ which send their respective operators onto the Z axis, where H is the π rotation about the $X + Z$ axis and $S = \sqrt{Z}$.

Single-qubit observable measurements

To measure the expectation $\langle\psi| O |\psi\rangle$ on a single-qubit system for a generic observable O we have two options. The first option is to decompose the observable on a basis which we can readily measure, the Pauli basis. In the simple case of a single qubit, we can compute the projection of the operator on each Pauli matrix and recover its decomposition as

$$O = \alpha_0 I + \alpha_1 X + \alpha_2 Y + \alpha_3 Z,$$

with $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}$. Then the expectation value of the operator writes

$$\langle\psi| O |\psi\rangle = \alpha_0 \langle\psi| I |\psi\rangle + \alpha_1 \langle\psi| X |\psi\rangle + \alpha_2 \langle\psi| Y |\psi\rangle + \alpha_3 \langle\psi| Z |\psi\rangle$$

Since $\langle\psi| I |\psi\rangle = \langle\psi|\psi\rangle = 1$, we only need to measure the three remaining terms. For $\langle\psi| Z |\psi\rangle$, we can directly measure the value of the operator after the application of the the state-preparation of $|\psi\rangle$, U_ψ. For the other terms, we measure the system after the additional application of the $Q_X = H$ and $Q_Y = HS$ rotations.

The alternative to the Pauli measurements is to consider the basis change that diagonalises the operator O and allows a projection onto one of its eigenvectors through the measurements of the Z Pauli matrix. Indeed, since O is Hermitian there is a unitary U such that $O = U^\dagger D U$, where

$$D = \lambda_0 |0\rangle\langle 0| + \lambda_1 |1\rangle\langle 1| = \frac{\lambda_0 + \lambda_1}{2} I + \frac{\lambda_0 - \lambda_1}{2} Z.$$

In the diagonal basis, the quantity we physically measure is thus the observable Z for the state $U |\psi\rangle$. The projection of the state of the system onto the Pauli Z eigenvectors is equivalently a projection onto an eigenvector of D. We can thus reconstruct the eigenvalue of O as

$$\omega_O = \lambda_0 \frac{(1 + \omega_Z)}{2} + \lambda_1 \frac{(1 - \omega_Z)}{2}.$$

where ω_O is the eigenvalue of O associated to ω_Z, the measured eigenvalue of Z.

Equivalently, we can rewrite the mean value of O as

$$\langle\psi| O |\psi\rangle = \frac{\lambda_0}{2}(1 + \langle\psi| U^\dagger Z U |\psi\rangle) + \frac{\lambda_1}{2}(1 - \langle\psi| U^\dagger Z U |\psi\rangle).$$

We are thus able to reduce the number of terms to measure from 3, the X, Y and Z Pauli matrices, to 1, the Z Pauli matrix, by considering a diagonal basis of the operator O. This showcases that to measure any given observable directly, a basis change into its diagonalization basis is required to allow the direct sampling on its eigenvalues. Then we can reconstruct the mean value of the observable for the state $|\psi\rangle$ by considering the empirical mean of the eigenvalues sampled in this way.

If we take another look at the Pauli measurements with that in mind, the described Q_X and Q_Y rotations diagonalize their respective Pauli term and allow the sampling of their eigenvalues and the reconstruction of their mean value for the state $|\psi\rangle$ through an empirical mean estimator.

Pauli measurements

For a single qubit system, as we have seen, the problem of diagonalising the observable is simple, since the dimension of $\mathscr{L}(\mathcal{H})$ is 4. For quantum systems with multiple qubits, the problem of diagonalising an operator O becomes much harder as the dimension of the space of linear applications $\mathscr{L}(\mathcal{H})$ increases exponentially with the number of qubits.

In practice, when dealing with N qubits, the only operators we are able to measure are the Z_k operator corresponding to each qubit k, $1 \leq k \leq N$. Then we can measure all Pauli operators with only Z terms, as they all commute with each other. Let us consider a generic N-qubit Pauli observable $P = \otimes_{k=1}^{N} P^k$, where $P^k \in \{I, X, Y, Z\}$ for any k, $1 \leq k \leq N$. Similarly to the 1-qubit case, the measurement of P only requires a change of basis affecting each qubit separately, *i.e.* a tensor product of local 1-qubit basis changes. The basis change unitary matrix can thus be expressed as $Q = \otimes_{k=1}^{N} Q^k$ where $Q^k \in \{I, Q_X = H, Q_Y = HS\}$ for any k, $1 \leq k \leq N$. The observable P can thus be measured on a quantum state $|\psi(\theta)\rangle$ with the circuit shown in Figure 1.3.

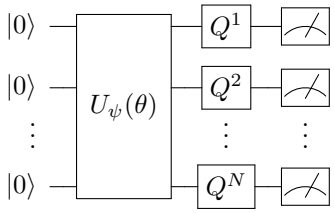


Figure 1.3 – Pauli-observable measurement.

In systems with multiple qubits, Pauli operators can commute and thus be measured simultaneously in a basis that diagonalises them simultaneously. Indeed, this is a general property of quantum systems. Given two observables O_1 and O_2, they can be simultaneously measured if they commute, i.e. $O_1 O_2 = O_2 O_1$. This is due to the projection of the quantum system onto an eigenstate of the measured operator. Indeed, if O_1 and O_2 commute, their eigenspaces are stable with respect to both operators and thus share a common eigenbasis and thus can be reconstructed from Z-Pauli measurements in that basis.

Observable estimation

Now, we can use this decomposition of a generic observable O for such a system in the Pauli basis $\mathscr{B}_{\mathrm{P}}(\mathscr{L}(\mathcal{H}))$

$$O = \sum_{k=0}^{4^N-1} \alpha_i P_i,$$

in order to estimate its mean value$\langle\psi| O |\psi\rangle$. That raises two problems. First, the projection of an arbitrary operator on the Pauli basis becomes prohibitive exponentially fast as N increases, since the number of operators on which to project is 4^N. The cost of this projection can be decreased either through analytic methods or by using the structure of the operator. Second, we have to compute the mean value of each non-zero Pauli term in the decomposition through sampling methods. It results in considering a number of samples proportional to the number of non-zero Pauli terms, which is exponential in the worst case. To overcome this issue, two methods are at hand. We can either reduce the number of samples needed per non-zero Pauli by grouping and

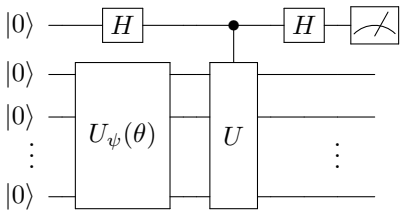


Figure 1.4 – Unitary measurement.

measuring directly commuting terms together in the same local basis. Or we can change the basis to merge groups of non-commuting terms together, thus reducing the number of Pauli matrices and increasing the proportion of commuting terms. That can be seen as a partial diagonalisation. Alternatively, we can also perform, as before, the direct measurement of the observable through a basis change with unitary U into its diagonal basis such that $D = U^\dagger O U$. The caveat of this direct measurement is that for such a basis change on a multi-qubit system, deep circuits are required, which increases sensitivity of the measurements to noise.

Unitary measurement: Hadamard test

An alternative exists for the measurement of unitary operators, it turns out that these measurements can be done through their implementation as a quantum circuit. Indeed, the Hadamard test is a quantum circuit (see Figure 1.4) which allows the computation of the real and imaginary parts of expected values of a unitary operator U for a state $|\psi\rangle = U_\psi |0\rangle$. This is done through an additional qubit, which we index by 0, by considering the mean value of the Z operator for this qubit, used as the output of a 1-qubit phase estimation circuit of U on the state $|\psi\rangle$. The unitary matrix associated to the quantum circuit is denoted by U_H, and the mean value of the Z-Pauli for the first qubit after the circuit is

$$\begin{aligned}
\langle Z_1 \rangle_{\psi,\text{H-test}} &= \langle 0| U_\mathrm{H}^\dagger Z_1 U_\mathrm{H} |0\rangle \\
&= \frac{1}{4} \langle 0| Z |0\rangle \langle \psi| (I + U^\dagger)(I + U) |\psi\rangle + \frac{1}{4} |1\rangle Z \langle 1| \langle \psi| (I - U^\dagger)(I - U) |\psi\rangle \\
&= \frac{1}{4} \langle \psi| (I + U^\dagger)(I + U) |\psi\rangle - \frac{1}{4} \langle \psi| (I - U^\dagger)(I - U) |\psi\rangle = \frac{1}{2} \langle \psi| \left(U + U^\dagger\right) |\psi\rangle ,
\end{aligned}$$

which gives

$$\langle Z_1 \rangle_{H_\text{test}(\psi)} = \mathrm{Re}(\langle \psi| U |\psi\rangle). \tag{1.3}$$

The imaginary part can be computed by addition of a phase gate after the control, making the interference cancel out the real part instead. However, by its very design, the Hadamard-test circuit requires the unitaries to be controlled, and thus results in a deeper circuit, more subject to noise. The main advantage of this method is that it enables the computation of the expected value of any unitary that can be implemented on a quantum computer providing more freedom in the decomposition of our target operator than in the observable-basis measurement method, albeit at the cost of deeper and noisier circuits. Note that the method can be made more general by considering not only the action of a unitary on $|\psi\rangle$, but on a larger system with additional ancillary qubits.

1.2 Walsh series

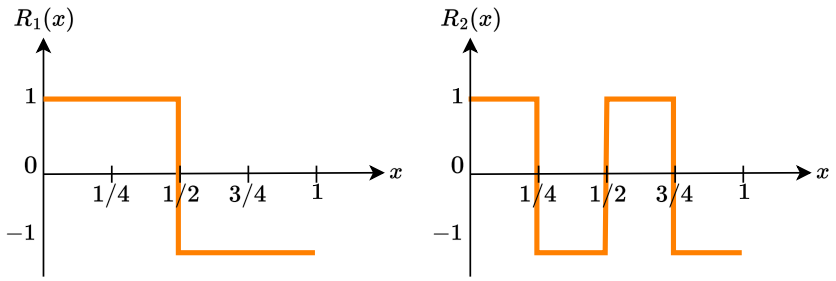


Figure 1.5 – First two Rademacher functions

As we shall see in Chapter 2, the Pauli decomposition of diagonal operators can be obtained through their representation in the Walsh functional basis.

Indeed, consider the action of the single-qubit Pauli Z operator acting on its eigenvectors, $Z|0\rangle = |0\rangle$ and $Z|1\rangle = -|1\rangle$. This can be written as

$$Z|i\rangle = (-1)^i |i\rangle ,$$

which is very similar to the function shown in Figure 1.5 from which the Walsh basis is constructed. This similarity allows to relate the Walsh series behind operators acting on functions to the Pauli decomposition of operators acting on the quantum amplitude encoding of functions on qubits, which is the main motivation behind this section.

The periodic nature of the problem we shall study in Chapter 2 also leads us to consider the relationship between the Fourier and Walsh bases. In this section we thus provide some mathematical context on the Walsh basis, with a focus on the description of the Fourier basis elements in the Walsh basis.

The Walsh functions, which belong to the space $L^2((0,1),\mathbb{C})$, which we denote $L^2(0,1)$, are closely related with the dyadic decomposition of a real number in the interval $(0,1)$. These functions form an orthonormal basis for $L^2(0,1)$ and own several properties that are particularly advantageous for computational applications due to their relation to binary numbers. Since their introduction by Walsh in the 1920s [95], the representation of functions as Walsh series has garnered significant attention and study[32, 33, 85].

Noteworthy contributions to the understanding of Walsh functions in Sobolev spaces include results on the decay of coefficients and the convergence of integration schemes for generalized p-adic Walsh bases, as discussed in [25, 26]. Walsh-based methods for solving differential equations have been proposed, using variational approaches, in [20], and the numerical analysis of these methods were presented, more recently, in [92] and [35], showing the uniform convergence of the discretized solutions.

In this thesis, we focus exclusively on the Walsh basis defined on dyadic rationals. We present new results on the L^2 convergence of Walsh series through their Sobolev norms, using a novel Walsh series decomposition of Fourier basis elements. These convergence results will allow the determination of the precision of Walsh-based Pauli estimators we shall define in Chapter 2. Additionally, we develop formulations for Walsh interpolation and truncation operators that are analogous to those defined for the Fourier basis, as detailed in [15, 16].

1.2.1 Dyadic/binary rationals and the Walsh basis

First, let us define dyadic decompositions, or binary fraction representation, of a real $x \in [0,1)$.

Definition 1. *We say that $x \in [0,1)$ admits a dyadic decomposition if there exists $(x_k)_{k\geq 1} \in \{0,1\}^{\mathbb{N}^*}$ such that*

$$x = \sum_{k=1}^{+\infty} x_k 2^{-k}, \tag{1.4}$$

the last equality standing as the decomposition itself.

Note that the decomposition exists, and is unique if and only if (x_k) does not have an infinite trail of 1. Second, let us recall the binary decomposition of any integer $m \geq 1$.

Definition 2. *For any $m \geq 1$, there exists a unique sequence $(m_k)_{k\geq 1} \in \{0,1\}^{\mathbb{N}^*}$ such that*

$$m = \sum_{k=1}^{+\infty} m_k 2^{k-1}. \tag{1.5}$$

This sum is in fact finite, up to $\sigma_m = \max\{j \in \mathbb{N}^* \mid m_j = 1\} = 1 + \lfloor \log_2 m \rfloor$, so that $m_{\sigma_m} = 1$ and $m_k = 0$ for any $k > \sigma_m$.

Equality (1.5) stands for the binary decomposition of m. Then we need several useful notions and notations.

- For any a, $b \in \{0,1\}$, we denote by $a \oplus b$ the representative of $a + b$ modulo 2 in $\{0,1\}$. Obviously, we have $0 \oplus 0 = 1 \oplus 1 = 0$ and $0 \oplus 1 = 1 \oplus 0 = 1$.
- As an extension, for any m, $n \geq 1$, using their binary decomposition (1.5), we denote by $m \oplus n$ the bitwise sum of m and n, *i.e.*

$$m \oplus n = \sum_{k=1}^{+\infty} (m_k \oplus n_k) 2^{k-1},$$

  where the previous sum is finite, up to $\max(\sigma_m, \sigma_n)$ at most.
- For any $x \in [0,1)$, and $m \in \mathbb{N}^*$, their binary product is given, in terms of their binary and dyadic decompositions, by

$$\langle m, x \rangle_{\text{bin}} = \bigoplus_{k=1}^{+\infty} m_k x_k,$$

  where the sum is in fact finite, up to σ_m.
- For any m, $n \in \mathbb{N}^*$, their binary product is given, in terms of their binary decompositions, by

$$\langle m, n \rangle_{\text{bin}} = \bigoplus_{k=1}^{+\infty} m_k n_k,$$

  where we use the same notation $\langle \cdot, \cdot \rangle_{\text{bin}}$ for the sake of simplicity, and the sum is finite up to $\min\{\sigma_n, \sigma_m\}$.
- The Hamming weight $|m|_{\text{Ham}}$ of an integer $m \geq 1$, defined by

$$|m|_{\text{Ham}} = \langle m, m \rangle_{\text{bin}},$$

  which allows to count the number of non-zero terms in the binary decomposition of m.

To define the Walsh functions, we first need the Rademacher functions

Definition 3. *For any $n \geq 1$, the n-th Rademacher function is defined by,*

$$R_n(x) = (-1)^{x_n} = \text{sign}(\sin(2^n \pi x)), \qquad x \in [0,1)$$

where x_n is the n-th coefficient in the dyadic decomposition (1.4) of x. That function can then be extended on $\mathbb{R}$ by 1-periodicity.

That allows to define the Walsh functions $(w_m)_{m\in\mathbb{N}}$.

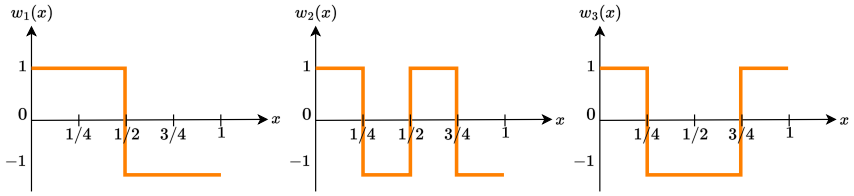


Figure 1.6 – First three Walsh functions

Definition 4. *For any $m \geq 1$, the m-th Walsh function is given, for any $x \in [0,1)$, by*

$$w_m(x) = (-1)^{\langle m,x\rangle_{\text{bin}}} = \prod_{k=1}^{\sigma_m} R_k(x)^{m_k}, \tag{1.6}$$

and is extended to $\mathbb{R}$ by 1-periodicity.

Additionally, we set $w_0(x) = 1$ for any $x \in \mathbb{R}$. It is worth noticing that equality (1.6) between w_m and the Rademacher functions also holds for any $x \in \mathbb{R}$, because of the 1-periodicity of all the functions involved.

We now go into several useful properties that the Walsh functions own.

Proposition 1 ([85] p. 245)**.** *The Walsh family of functions $\mathcal{W} = (w_m)_{m\in\mathbb{N}}$ is a Hilbert basis of $L^2(0,1)$.*

The series $\sum \langle w_m, f\rangle_{L^2} w_m$ is called the Walsh series associated to $f \in L^2(0,1)$. Besides, the product of two Walsh functions remains a Walsh function. More precisely, we have the following

Lemma 1. *Let m, $n \in \mathbb{N}^*$, then we have, for any $x \in \mathbb{R}$,*

$$w_m(x)w_n(x) = w_{m\oplus n}(x).$$

Proof. For any $x \in [0,1)$, we can write, using (1.6),

$$w_m(x)w_n(x) = (-1)^{\langle m,x\rangle_{\text{bin}}}(-1)^{\langle n,x\rangle_{\text{bin}}} = (-1)^{\sum_\ell x_\ell(m_\ell \oplus n_\ell)} = (-1)^{\langle m\oplus n,x\rangle_{\text{bin}}} = w_{m\oplus n}(x),$$

which is extended for $x \in \mathbb{R}$ by 1-periodicity. □

Effect of a power-of-two homothety on the Walsh functions

Let us now consider how scaling the input variable by a power-of-two affects the Walsh functions. First, we consider the effect of such an homothety on the Rademacher functions.

Lemma 2. *Let $n \geq 1$. For any $m \in \mathbb{N}$ and $x \in \mathbb{R}$, we have*

$$R_n(2^m x) = R_{m+n}(x). \tag{1.7}$$

Proof. Let $m \geq 1$ and $x \in [0,1)$. Using the dyadic and binary decompositions (1.4) and (1.5) of m and x, we write

$$2^m x = \sum_{k=1}^{+\infty} x_k 2^{m-k} = \lfloor 2^m x \rfloor + \sum_{k'=1}^{+\infty} x_{k'+m} 2^{-k'}, \tag{1.8}$$

the latter term in the right-hand side being an element of $[0,1)$, which we denote by y. Consequently, by 1-periodicity, we have, for any $n \geq 1$,

$$R_n(2^m x) = R_n(y).$$

It is then clear, thanks to (1.8), that

$$R_n(y) = (-1)^{y_n} = (-1)^{x_{m+n}},$$

which allows to obtain (1.7). □

Lemma 3. *For any $m, n \in \mathbb{N}^*$, and any $x \in \mathbb{R}$, we have*

$$w_n(2^m x) = w_{2^m n}(x). \tag{1.9}$$

Proof. Let m, $n \in \mathbb{N}^*$, and $x \in \mathbb{R}$. On the one hand, we have, thanks to (1.6) and (1.7),

$$w_n(2^m x) = \prod_{k=1}^{+\infty} R_k(2^m x)^{n_k} = \prod_{\ell=1}^{\sigma_n} R_{k+m}(x)^{n_k}, \tag{1.10}$$

remembering that the products are in fact finite, as $n_k = 0$ when $k > \sigma_n$. On the other hand, we can write the binary decomposition of $2^m n$

$$2^m n = \sum_{k=m+1}^{+\infty} n_{k-m} 2^{k-1},$$

so that, again with (1.6),

$$w_{2^m n}(x) = \prod_{k=1}^{+\infty} [R_k(x)]^{(2^m n)_k} = \prod_{k=m+1}^{+\infty} R_k(x)^{n_{k-m}},$$

which, together with (1.10), ensures (1.9) □

The properties we explored allow to discuss consider the link between the Fourier and the Walsh series associated to function.

1.2.2 Walsh decomposition of $H^1_\#$ functions through the Fourier basis

In this subsection, we aim to describe the link between the Walsh and Fourier series of a specific class of functions.

First, let us define the functional spaces $H^m(0,1)$ which consists of square integrable functions whose derivatives up to the k-th order are also square integrable.

Definition 5 ([12])**.** *We define the Sobolev space $H^1(0,1)$ on $\mathbb{C}$ as the complex Hilbert space*

$$H^1(0,1) = \{v \in L^2(0,1) \mid \exists g \in L^2(0,1)\ \langle \phi', v\rangle_{L^2} = \langle \phi, g\rangle_{L^2},\ \forall \phi \in C_c^\infty((0,1),\mathbb{C})\},$$

where $C_c^\infty((0,1),\mathbb{C})$ is the set of infinitely differentiable, compactly-supported functions. We also denote by v' the weak derivative $g \in L^2(0,1)$ associated to $v \in H^1(0,1)$.

Setting $H^0(0,1) = L^2(0,1)$ and for $m > 0$, the Sobolev space $H^m(0,1)$ is inductively defined as the complex Hilbert space

$$H^m(0,1) = \left\{v \in H^{m-1}(0,1) \mid v' \in H^{m-1}(0,1)\right\}.$$

The H^m norm, for any $v \in H^m(0,1)$, is given by

$$\|v\|_{H^m} = \left(\sum_{\ell=0}^{m-1} \|v^{(\ell)}\|_{L^2}^2\right)^{\frac{1}{2}}.$$

with $v^{(\ell)}$ the ℓ-th weak derivative of v.

Similarly, we define the equivalent space in the periodic setting.

Definition 6. *The Sobolev space $H^1_\#(0,1)$ of periodic functions is defined as*

$$H^1_\#(0,1) = \left\{v \in H^1(0,1) \mid v(0) = v(1)\right\}. \tag{1.11}$$

We then inductively define, for $m > 1$,

$$H^m_\#(0,1) = \left\{v \in H^{m-1}_\#(0,1) \mid v' \in H^{m-1}_\#(0,1)\right\}.$$

Note that (1.11) makes sense because $H^1(0,1)$ is compactly injected in $C^0([0,1])$. We aim to describe a function $v \in H^1_\#(0,1)$ in the Walsh basis through its Fourier coefficients.

To do so, we need to decompose each $\mathrm{e}_k : x \mapsto \exp(i2\pi kx)$, $k \in \mathbb{Z}$, which belongs to $H^1_\#(0,1)$, in the Walsh basis.

Fourier functions in the Walsh basis

Now, to determine the decomposition of each Fourier basis functions e_k, let us first find the coefficients of the Walsh series for an odd k.

Lemma 4. *Let $k \in 2\mathbb{N}+1$. The Walsh series associated to $\mathrm{e}_k : x \mapsto \exp(i2\pi kx)$ is given by*

$$\mathrm{e}_k = \sum_{m \in 2\mathbb{N}+1} \gamma_{m,k} w_m,$$

where we set, for any $m \in 2\mathbb{N}+1$,

$$\gamma_{m,k} = -\frac{2}{k\pi}(-i)^{|m|_{\mathrm{Ham}}} \prod_{\ell=2}^{\sigma_m} \tan\left(\frac{k\pi}{2^\ell}\right)^{m_\ell}.$$

Proof. First, note that, given $k \in 2\mathbb{N}+1$, for any $x \in [0,1)$,

$$\exp(i2\pi kx) = \prod_{\ell=1}^{+\infty} \exp\Big(i2\pi k \frac{x_\ell}{2^\ell}\Big) = \prod_{\ell=1}^{+\infty} \left(1 - x_\ell + x_\ell \exp\left(\frac{i2\pi k}{2^\ell}\right)\right).$$

Using the identity $a = \frac{1-(-1)^a}{2}$ for any $a \in \{0,1\}$, the previous equality becomes

$$\begin{aligned}\exp(i2\pi kx) &= \prod_{\ell=1}^{+\infty} \left[1 - \frac{1-(-1)^{x_\ell}}{2} + \frac{1-(-1)^{x_\ell}}{2} \exp\left(\frac{i2\pi k}{2^\ell}\right)\right] \\ &= \prod_{\ell=1}^{+\infty} \left[\frac{1+\exp\big(\frac{i2\pi k}{2^\ell}\big)}{2} + (-1)^{x_\ell} \frac{1-\exp\big(\frac{i2\pi k}{2^\ell}\big)}{2}\right].\end{aligned}$$

Now, introducing a binary index m_ℓ for each element ℓ of the product, we obtain

$$\exp(i2\pi kx) = \prod_{\ell=1}^{+\infty} \sum_{m_\ell=0}^{1} (-1)^{x_\ell m_\ell} \left(\frac{1+(-1)^{m_\ell}\exp\big(\frac{i2\pi k}{2^\ell}\big)}{2}\right), \tag{1.12}$$

Then, involving an integer index m associated to the binary one, and inverting the product and sum, we get from (1.12)

$$\begin{aligned}\exp(i2\pi kx) &= \sum_{m=1}^{+\infty} \prod_{\ell=1}^{+\infty} (-1)^{x_\ell m_\ell} \prod_{\ell=1}^{+\infty} \frac{1+(-1)^{m_\ell}\exp\big(\frac{i2\pi k}{2^\ell}\big)}{2} \\ &= \sum_{m=1}^{+\infty} (-1)^{\langle m,x\rangle_{\mathrm{bin}}} \prod_{\ell=1}^{+\infty} \frac{1+(-1)^{m_\ell}\exp\big(\frac{i2\pi k}{2^\ell}\big)}{2},\end{aligned} \tag{1.13}$$

noticing that the term corresponding to $m=0$ is zero, since $\frac{1+e^{i\pi}}{2}$ (obtained for $\ell=1$) is zero. Let us set, for any $m \geq 1$,

$$\gamma_{m,k} = \prod_{\ell=1}^{+\infty} \frac{1+(-1)^{m_\ell}\exp\big(\frac{i2\pi k}{2^\ell}\big)}{2}, \tag{1.14}$$

so that (1.13), thanks to (1.6), for any $x \in [0,1)$,

$$\exp(i2\pi kx) = \sum_{m=1}^{+\infty} \gamma_m (-1)^{\langle m,x\rangle_{\mathrm{bin}}} = \sum_{m=1}^{+\infty} \gamma_{m,k} w_m(x).$$

Let us now obtain an expression of γ_m, $m \geq 1$, handier than (1.14), starting with

$$\gamma_{m,k} = \prod_{\ell=1}^{+\infty} \exp\left(\frac{i2\pi k}{2^{\ell+1}}\right) \frac{\exp\big(\frac{-i2\pi k}{2^{\ell+1}}\big) + (-1)^{m_\ell}\exp\big(\frac{i2\pi k}{2^{\ell+1}}\big)}{2}$$

$$= \prod_{\ell=1}^{+\infty} \exp\left(\frac{i\pi k}{2^\ell}\right) \times \prod_{\ell=1}^{+\infty} \left[\cos\big(2\pi k/2^{\ell+1}\big)(1-m_\ell) - im_\ell \sin\big(2\pi k/2^{\ell+1}\big)\right]$$
$$= \exp(i\pi k) \cdot \prod_{\ell=1}^{+\infty} \left[\cos\big(\pi k/2^\ell\big)(1-m_\ell) - im_\ell \sin\big(\pi k/2^\ell\big)\right]$$
$$= (-1)^k \prod_{\ell=\sigma_m+1}^{+\infty} \cos\big(\pi k/2^\ell\big) \cdot \prod_{\ell=1}^{\sigma_m} \left(\cos\big(\pi k/2^\ell\big)(1-m_\ell) - im_\ell \sin\big(\pi k/2^\ell\big)\right)$$

Then we notice that, for any $K \geq 1$, and any θ for which it makes sense,

$$\prod_{k=1}^{K} \cos\left(\frac{\theta}{2^k}\right) = \prod_{k=1}^{K} \frac{\sin\left(\theta/2^{k-1}\right)}{2\sin\left(\theta/2^k\right)} = \frac{\sin\theta}{\theta} \times \frac{\theta/2^K}{\sin\left(\theta/2^K\right)} \xrightarrow{K\to+\infty} \frac{\sin\theta}{\theta}, \tag{1.15}$$

which we can definitely use for $\theta = \pi k/2$ since k is odd here. Eventually, we get, for any $m \geq 1$,

$$\gamma_{m,k} = im_1 \prod_{\ell=2}^{+\infty} \cos\left(\frac{\pi k}{2^\ell}\right) \times \prod_{\ell=2}^{\sigma_m} \left[1 - m_\ell - im_\ell \tan\left(\frac{\pi k}{2^\ell}\right)\right]$$
$$= -\frac{2m_1}{\pi k}(-i)^{|m|_{\mathrm{Ham}}} \prod_{\ell=2}^{\sigma_m} \tan\left(\frac{\pi k}{2^\ell}\right)^{m_\ell},$$

since $1 - m_\ell - im_\ell \tan\frac{\pi k}{2^\ell}$ equals 1 when $m_\ell = 0$ and $-i\tan\left(\frac{\pi k}{2^\ell}\right)$ when $m_\ell = 1$. Hence, for any even m, $\gamma_m = 0$ as $m_1 = 0$, and we can limit the indices in the Walsh series to the odd ones, for which $m_1 = 1$. □

The following proposition is then a straightforward consequence of Lemmas 3–4.

Proposition 2. *Let $k \in \mathbb{N}^*$ and denote by K the multiplicity of 2 in its prime factorization. For any $x \in \mathbb{R}$, we can write*

$$\exp(i2\pi kx) = \sum_{m\in 2\mathbb{N}+1} \gamma_{m,k/2^K} w_{2^K m}(x).$$

Remark 1. *Notice that, for any $k' \geq 0$, the non-zero Walsh coefficients of $\mathrm{e}_{2^K k'}$ have indices of the form $(2p+1)2^K$. Hence, 2 is always a prime factor of the indices of the non-zero Walsh coefficients with multiplicity k'. That implies two things:*

- *all odd-indexed e_k share the same set of indices of non-zero Walsh coefficients,*
- *when m and ℓ do not have the same power of 2 in their factorization, then their associated Walsh series do not share any Walsh function.*

Theorem 1. *Let $k \in \mathbb{Z}^*$, with $K \in \mathbb{N}$ the multiplicity of 2 in its prime factorization, for any $x \in \mathbb{R}$, we have*

$$\exp(i2\pi kx) = \sum_{m\in 2\mathbb{N}+1} [\operatorname{sign} k]^{|m|_{\mathrm{Ham}}} \gamma_{m,|k/2^K|} w_{2^K m}(x), \tag{1.16}$$

where $\operatorname{sign} k$ is the sign of k, and for any $k' \in 2\mathbb{N}+1$, $m \in 2\mathbb{N}+1$, $\gamma_{m,k'}$ is given by

$$\gamma_{m,k'} = -\frac{2m_1}{\pi k'}(-i)^{|m|_{\mathrm{Ham}}} \prod_{\ell=2}^{\sigma_m} \tan\left(\frac{\pi k'}{2^\ell}\right)^{m_\ell}.$$

Similarly, when $k \in \mathbb{N}^*$, we have, for any $x \in \mathbb{R}$,

$$\sin(2\pi kx) = \sum_{\substack{m\in 2\mathbb{N}+1,\\ |m|_{\mathrm{Ham}}\in 2\mathbb{N}+1}} (-1)^{\frac{|m|_{\mathrm{Ham}}+1}{2}+1} |\gamma_{m,k/2^K}| w_{2^K m}(x), \tag{1.17}$$
$$\cos(2\pi kx) = \sum_{\substack{m\in 2\mathbb{N}+1,\\ |m|_{\mathrm{Ham}}\in 2\mathbb{N}}} (-1)^{\frac{|m|_{\mathrm{Ham}}}{2}+1} |\gamma_{m,k/2^K}| w_{2^K m}(x).$$

Proof. For $k \in \mathbb{N}$, we recover equation (1.16) directly from Proposition 2, for negative frequencies, we use the fact that $\mathrm{e}_k = \overline{\mathrm{e}_{-k}}$ to obtain

$$\mathrm{e}_{-k} = \overline{\mathrm{e}_k} = \sum_{m\in 2\mathbb{N}+1} \overline{\gamma_{m,k'}} w_{2^K m}(x).$$

As $\gamma_{m,k'}$ is real for even m and imaginary for odd m, we obtain

$$\mathrm{e}_{-k} = \sum_{m\in 2\mathbb{N}+1} (-1)^{|m|_{\mathrm{Ham}}} \gamma_{m,k'} w_{2^K m}(x),$$

which gives us the result for the remaining $k \in \mathbb{Z}$.

For the Walsh series of the sines and cosines, we use the Euler formula and equation (1.16) to obtain, for any $k \in \mathbb{N}$,

$$\exp(i2\pi kx) = \cos(2\pi kx) + i\sin(2\pi kx) = \sum_{m\in 2\mathbb{N}+1} \gamma_{m,|k/2^K|} w_{2^K m}(x).$$

Again using the fact that $\gamma_{m,k}$ is real for even m and imaginary for odd m, we obtain

$$\sin(2\pi kx) = \sum_{\substack{m\in 2\mathbb{N}+1,\\ |m|_{\mathrm{Ham}}\in 2\mathbb{N}+1}} \frac{1}{i}\gamma_{m,k/2^K} w_{2^K m}(x), \tag{1.18}$$
$$\cos(2\pi kx) = \sum_{\substack{m\in 2\mathbb{N}+1,\\ |m|_{\mathrm{Ham}}\in 2\mathbb{N}}} \gamma_{m,k/2^K} w_{2^K m}(x).$$

For $|m|_{\mathrm{Ham}} \in 2\mathbb{N}$, we have

$$\mathrm{sign}(\gamma_{m,k/2^K}) = -(-i)^{|m|_{\mathrm{Ham}}} = (-1)^{\frac{|m|_{\mathrm{Ham}}}{2}+1},$$

and for $|m|_{\mathrm{Ham}} \in 2\mathbb{N}+1$,

$$\mathrm{sign}\left[\frac{1}{i}\gamma_{m,k/2^K}\right] = -(-i)^{|m|_{\mathrm{Ham}}} = (-1)^{|m|_{\mathrm{Ham}}+1}\, i^{|m|_{\mathrm{Ham}}-1} = (-1)^{\frac{|m|_{\mathrm{Ham}}-1}{2}},$$

which finally allows to rewrite (1.18) into (1.17). □

From the Fourier decomposition to the Walsh decomposition

In this subsubsection, we describe the relationship between the Fourier and Walsh series of functions in $L^2(0,1)$.

Theorem 2. *Let $v \in L^2(0,1)$ and its associated Fourier series*

$$\sum_{k\in\mathbb{Z}} \hat{v}_k \mathrm{e}_k.$$

The Walsh series associated to v,

$$\sum_{m\in\mathbb{N}} \tilde{v}_m w_m,$$

can be recovered in terms of the Fourier coefficients of v. Indeed, we have $\tilde{v}_0 = \hat{v}_0$ and for any $m \in \mathbb{N}^*$,

$$\tilde{v}_m = \sum_{k\in 2^M(2\mathbb{N}+1)} \left(\hat{v}_k + (-1)^{|m|_{\mathrm{Ham}}} \hat{v}_{-k}\right) \gamma_{m/2^M,k/2^M}, \tag{1.19}$$

where $M \in \mathbb{N}$ is the multiplicity of 2 in the prime factorization of m.

Proof. We can successively write

$$\begin{aligned} v &= \hat{v}_0 \mathrm{e}_0 + \sum_{k\in\mathbb{N}^*} (\hat{v}_k \mathrm{e}_k + \hat{v}_{-k}\mathrm{e}_{-k}) = \hat{v}_0 w_0 + \sum_{K=0}^{+\infty} \sum_{k\in 2^K(2\mathbb{N}+1)} (\hat{v}_k \mathrm{e}_k + \hat{v}_{-k}\mathrm{e}_{-k}) \\ &= \hat{v}_0 w_0 + \sum_{K=0}^{+\infty} \sum_{k\in 2^K(2\mathbb{N}+1)} \sum_{m\in(2\mathbb{N}+1)} \left(\hat{v}_k + (-1)^{|m|_{\mathrm{Ham}}} \hat{v}_{-k}\right) \gamma_{m,k/2^K} w_{2^K m} \\ &= \hat{v}_0 w_0 + \sum_{K=0}^{+\infty} \sum_{m\in 2\mathbb{N}+1} \sum_{k\in 2^K(2\mathbb{N}+1)} \left(\hat{v}_k + (-1)^{|m|_{\mathrm{Ham}}} \hat{v}_{-k}\right) \gamma_{m,k/2^K} w_{2^K m}. \end{aligned}$$

In the previous equality, we perform the index change $m' = 2^K m$ to obtain

$$v = \hat{v}_0 w_0 + \sum_{K=0}^{+\infty} \sum_{m\in 2^K(2\mathbb{N}+1)} \left(\sum_{k\in 2^K(2\mathbb{N}+1)} \left(\hat{v}_k + (-1)^{|m'|_{\mathrm{Ham}}} \hat{v}_{-k}\right) \gamma_{m'/2^K,k/2^K} \right) w_{m'}.$$

The first two $\sum$ symbols eventually provide a unique sum symbol on $m \in \mathbb{N}^*$, and, together with the zero-indexed term, allow to get the Walsh series and the values of the Walsh coefficients. □

Remark 2. *It is important to note that the only Fourier indices that appear in the expression for the Walsh coefficients* (1.19) *are such that, given $M > 0$, for any $m \in 2^M(2\mathbb{N}+1)$ the only Fourier coefficients in the decomposition of $\tilde{v}_M$ are multiples of 2^M, this gives us a glimpse at the link between the structures of both series.*

1.2.3 Walsh series truncation

In this subsection, thanks to the above results on the Walsh series of the Fourier basis functions, we can provide a new result on the convergence rate of Walsh series for H^1 functions through their Fourier series.

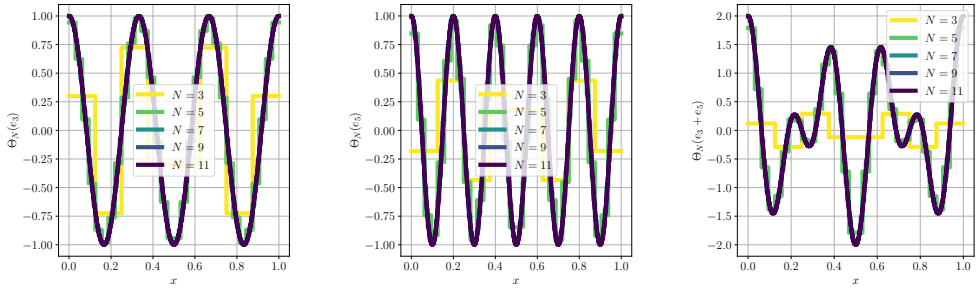


Figure 1.7 – Truncated Walsh series of e_3, e_5 and $\mathrm{e}_3 + \mathrm{e}_5$. The truncated Walsh series are obtained through (1.16) for e_3 and e_5 and these are summed to obtain $\Theta_N(\mathrm{e}_3 + \mathrm{e}_5)$.

Theorem 3. *For any $v \in H^1_{\#}(0,1)$ and any $N \geq 1$, the N-bit truncated Walsh series associated to v given by*

$$\Theta_N v = \sum_{m=0}^{2^N-1} \tilde{v}_m w_m, \tag{1.20}$$

converges towards v in $L^2(0,1)$ as

$$\|\Theta_N v - v\|_{L^2} \leq \left(\frac{1}{\pi} + \frac{1}{2\sqrt{3}}\right) \frac{1}{2^N} \|v'\|_{L^2}. \tag{1.21}$$

Proof. To study the convergence rate, when N goes to $+\infty$, in L^2 of the N-bit truncated Walsh series of $v \in H^1_{\#}(0,1)$, we start again with its Fourier series

$$v = \sum_{k \in \mathbb{Z}} \hat{v}_k \mathrm{e}_k.$$

The proof is then divided into four steps.

- Step 1. Behaviour with respect to N of $\Theta_N \mathrm{e}_k$ for $k \in \mathbb{Z}$.
- Step 2. Orthogonality of $\Theta_N \mathrm{e}_k$ and $\Theta_N \mathrm{e}_j$ when $k \neq j$.
- Step 3. Use truncated Fourier series and provide convergence results for its associated truncated Walsh series.
- Step 4. Conclusion

Step 1 Thanks to Theorem 1, we can write

$$\Theta_N \mathrm{e}_k = \sum_{m=0}^{2^{N-K}-1} \gamma_{m,k/2^K} w_{2^K m}.$$

Hence, by orthogonality of $\mathrm{e}_k - \Theta_N \mathrm{e}_k$ and $\Theta_N \mathrm{e}_k$, we can write

$$\|\mathrm{e}_k - \Theta_N \mathrm{e}_k\|^2_{L^2} = \|\mathrm{e}_k\|^2_{L^2} - \|\Theta_N \mathrm{e}_k\|^2_{L^2} = 1 - \|\Theta_N \mathrm{e}_k\|^2_{L^2}. \tag{1.22}$$

The norm of $\Theta_N \mathrm{e}_k$ is given by

$$\|\Theta_N \mathrm{e}_k\|_{L^2}^2 = \sum_{m=0}^{2^{N-K}-1} |\gamma_{m,k/2^K}|^2. \tag{1.23}$$

By setting, for any $k' \in 2\mathbb{N}+1$ and $n>0$, $S_{n,k'} = \sum_{m=0}^{2^n-1} |\gamma_{m,k'}|^2$, we have

$$\|\Theta_N \mathrm{e}_k\|_{L^2}^2 = S_{N-K,k/2^K}.$$

Let us now focus on of $S_{n,k'}$. Note that, for any $k' \in 2\mathbb{N}+1$,

$$|\gamma_{m,k'}|^2 = \prod_{\ell=\sigma_m+1}^{+\infty} \cos^2(\pi k'/2^\ell) \cdot \prod_{\ell=1}^{\sigma_m} \left(\cos^2(\pi k'/2^\ell)(1-m_\ell) + m_\ell \cdot \sin^2(\pi k'/2^\ell)\right).$$

Consequently, we have, for any $n \geq 1$,

$$\begin{aligned} S_{n,k'} &= \sum_{m=0}^{2^n-1} \prod_{\ell=n+1}^{+\infty} \cos^2(\pi k'/2^\ell) \cdot \prod_{\ell=1}^{n} \left(\cos^2(\pi k'/2^\ell)(1-m_\ell) + m_\ell \sin^2(\pi k'/2^\ell)\right) \\ &= \prod_{\ell=n+1}^{+\infty} \cos^2(\pi k'/2^\ell) \sum_{m=0}^{2^n-1} \prod_{\ell=1}^{n} \left(\cos^2(\pi k'/2^\ell)(1-m_\ell) + m_\ell \sin^2(\pi k'/2^\ell)\right). \end{aligned}$$

In the same way as in the proof of 4, the previous equation can be rewritten as a sum over the binary coefficients of m, which ensures

$$S_{n,k'} = \prod_{\ell=n+1}^{+\infty} \cos^2(\pi k'/2^\ell) \sum_{m_1,\dots m_n=0}^{1} \prod_{\ell=1}^{n} \left(\cos^2(\pi k'/2^\ell)(1-m_\ell) + m_\ell \cdot \sin^2(\pi k'/2^\ell)\right).$$

Then, inverting the sums on m_ℓ with the product, we have

$$\begin{aligned} S_{n,k'} = \prod_{\ell=n+1}^{+\infty} \cos^2(\pi k'/2^\ell) \prod_{\ell=1}^{n} \left(\cos^2(\pi k'/2^\ell) + \sin^2(\pi k'/2^\ell)\right) \\ = \prod_{\ell=n+1}^{+\infty} \cos^2(\pi k'/2^\ell) = \sin^2(\frac{\pi k'}{2^n}) \left(\frac{4^n}{\pi^2 k'^2}\right), \end{aligned} \tag{1.24}$$

thanks to (1.15). From (1.24),we obtain the value of (1.23),

$$\|\Theta_N \mathrm{e}_k\|_{L^2}^2 = \sin^2(\frac{\pi k}{2^{(N-K)+K}}) \cdot \frac{4^{N-K}}{\pi^2 k^2/4^K} \geq 1 - \frac{1}{3}\left(\frac{\pi k}{2^N}\right)^2. \tag{1.25}$$

Finally, we bound the convergence rate using (1.25)and (1.22), *i.e.*

$$\|\Theta_N \mathrm{e}_k - \mathrm{e}_k\|_{L^2}^2 \leq \frac{1}{3}\frac{\pi^2 k^2}{4^N}. \tag{1.26}$$

Step 2 We now show orthogonality results on the truncated series of e_j and e_k for $-2^{N-1} < k < 2^{N-1}$ and $-2^{N-1} < j < 2^{N-1}$. Their scalar product is given by

$$\langle \Theta_N \mathrm{e}_j, \Theta_N \mathrm{e}_k \rangle_{L^2} = \sum_{m=0}^{\min(2^{N-K}-1, 2^{N-J}-1)} \operatorname{sign}(jk)^{|m|_{\mathrm{Ham}}} \overline{\gamma_{m,|j|/2^J}} \gamma_{m,|k|/2^K} \langle w_{2^J m}, w_{2^K m} \rangle_{L^2},$$

where J, $K \in \{0, \dots N-2\}$ are the multiplicity of 2 in the prime factorisation of j and k respectively. Note that if $J \neq K$, $\langle \Theta_N \mathrm{e}_j, \Theta_N \mathrm{e}_k \rangle_{L^2} = 0$ as they share no indices with non-zero coefficients. Now, let us focus on the terms such that $J = K$ and $j \neq k$. For such terms, using (1.16), we have

$$\begin{aligned}\langle \Theta_N e_j, \Theta_N e_k \rangle_{L^2} = \sum_{m=0}^{2^N-1} \operatorname{sign}(kj)^{|m|_{\mathrm{Ham}}} \frac{4m_1}{\pi^2 |kj|} (-1)^{2|m|_{\mathrm{Ham}}} \\ \times \prod_{\ell=2}^{\sigma_m} \tan\left(\frac{\pi |k|/2^K}{2^\ell}\right)^{m_\ell} \prod_{\ell=2}^{\sigma_m} \tan\left(\frac{\pi |j|/2^J}{2^\ell}\right)^{m_\ell} \\ = \sum_{m=0}^{2^N-1} \operatorname{sign}(kj)^{|m|_{\mathrm{Ham}}} \frac{4}{\pi^2 |kj|} \tau_{m,|k/2^K|} \tau_{m,|j|/2^J},\end{aligned}$$

where, for any m, $n \in 2\mathbb{N}+1$ we denote by $\tau_{m,n} = \prod_{\ell=2}^{\sigma_n} \tan\left(\frac{\pi n}{2^\ell}\right)^{m_\ell}$. For simplicity we set $J = K = 1$ for the following calculations as they do not depend of the value of J or K. We now show that the above scalar product is equal to zero as for each positive term in the above sum, an opposite term exists with equal magnitude and opposite sign. Let us now prove that there exists an $n < N$, which depends only on j and k, such that for each $m \in 2\mathbb{N}+1$ the associated $m \oplus 2^n$ is such that

$$\overline{\gamma_{m,|j|}} \gamma_{m,|k|} + \operatorname{sign}(kj) \overline{\gamma_{m\oplus 2^n,|j|}} \gamma_{m\oplus 2^n,|k|} = 0. \tag{1.27}$$

Let us then proceed by equivalence: (1.27) implies

$$\overline{\gamma_{m,|j|}} \gamma_{m,|k|} + \operatorname{sign}(kj) \overline{\gamma_{m\oplus 2^n,|j|}} \gamma_{m\oplus 2^n,|k|} = \frac{4}{\pi^2 kj} \left(\tau_{m,|k|} \tau_{m,|j|} + \operatorname{sign}(kj) \tau_{m\oplus 2^n,|k|} \tau_{m\oplus 2^n,|j|}\right).$$

That gives

$$\tau_{m,|k|} \tau_{m,|j|} = -\operatorname{sign}(kj) \tau_{m\oplus 2^n,|k|} \tau_{m\oplus 2^n,|j|}.$$

As m and $m \oplus 2^n$ only differ in their n-th binary coefficient, we can write

$$\left(\tan\left(\frac{\pi|k|}{2^n}\right) \tan\left(\frac{\pi|j|}{2^n}\right)\right)^{m_n} + \operatorname{sign}(kj) \left(\tan\left(\frac{\pi|k|}{2^n}\right) \tan\left(\frac{\pi|j|}{2^n}\right)\right)^{1\oplus m_n} = 0.$$

Since either m_n or $1 \oplus m_n$ equals 1, we necessarily have

$$\tan\left(\frac{\pi|k|}{2^n}\right) \tan\left(\frac{\pi|j|}{2^n}\right) = -\operatorname{sign}(kj).$$

We rewrite the previous equation as

$$\sin\left(\frac{\pi|k|}{2^n}\right) \sin\left(\frac{\pi|j|}{2^n}\right) + \operatorname{sign}(kj) \cos\left(\frac{\pi|k|}{2^n}\right) \cos\left(\frac{\pi|j|}{2^n}\right) = \cos\left(\pi \frac{(|k| - \operatorname{sign}(kj)|j|)}{2^n}\right) = 0$$

which is possible if and only if

$$\frac{|k-j|}{2^n}\pi = \frac{\pi}{2}[\pi],$$

where we obtain the last line as $|k| - \mathrm{sign}(kj)|j| = |k-j|$ as it is equal to $|k| - |j|$ if they share a sign and $|k| + |j|$ if they do not. Note that $|k-j| \in 2\mathbb{N}^*$, as both terms are odd, and that $2 \leq |k-j| < 2^N$ as $k \neq j$ and $|k|, |j| < 2^{N-1}$. Hence, the result holds for the unique $2 \leq n \leq N$ such that $\frac{|k-j|}{2^n} \in 2\mathbb{N}+1$. That ensures that, for any j, $k \in \{-2^N+1, \dots 2^{N-1}-1\}$ such that $j \neq k$,

$$\begin{aligned}\langle \Theta_N e_j, \Theta_N e_k \rangle_{L^2} &= \sum_{m \in 2\mathbb{N}+1} \overline{\gamma_{m,|j|/2^J}} \gamma_{m,|k|/2^K} \\ &= \sum_{m \in 2\mathbb{N}+1, m_n = 0} \overline{\gamma_{m,|j|/2^J}} \gamma_{m,|k|/2^K} + \mathrm{sign}(kj) \overline{\gamma_{m\oplus 2^n,|j|/2^J}} \gamma_{m\oplus 2^n,|k|/2^K} = 0.\end{aligned}$$

Step 3 Let us now introduce, for any $n \in \mathbb{N}^*$, the truncated Fourier series of any $v \in H^1_{\#}(0,1)$,

$$\Pi_n v = \sum_{|k| \leq n} \hat{v}_k \mathrm{e}_k.$$

We focus on

$$\|\Pi_{2^{N-1}-1} v - \Theta_N(\Pi_{2^{N-1}-1} v)\|^2_{L^2} = \left\| \sum_{|k|<2^{N-1}} \hat{v}_k (\Theta_N \mathrm{e}_k - \mathrm{e}_k) \right\|^2_{L^2}.$$

Using Step 2 for $|k| < 2^{N-1}$ and (1.26), we obtain

$$\begin{aligned}\|\Pi_{2^{N-1}-1} v - \Theta_N(\Pi_{2^{N-1}-1} v)\|^2_{L^2} &= \sum_{|k|<2^{N-1}} |\hat{v}_k|^2 \|\Theta_N \mathrm{e}_k - \mathrm{e}_k\|^2_{L^2} \\ &\leq \sum_{|k|<2^{N-1}} \frac{1}{12 \cdot 4^N} 4\pi^2 k^2 |\hat{v}_k|^2.\end{aligned}$$

Thus, bounding $\sum_{|k|<2^{N-1}} \frac{1}{12\cdot 4^N} 4\pi^2 k^2 |\hat{v}_k|^2$ by $\|v'\|^2_{L^2}$, we get

$$\|\Pi_{2^{N-1}-1} v - \Theta_N(\Pi_{2^{N-1}-1} v)\|^2_{L^2} = \frac{1}{12 \cdot 4^N} \|v'\|^2_{L^2}.$$

Step 4 We eventually use Step 3 to obtain the convergence rate. It is clear that, for any $v \in H^1_{\#}(0,1)$,

$$\begin{aligned}\|v - \Theta_N v\|_{L^2} &= \|v - \Pi_{2^{N-1}-1} v + \Pi_{2^{N-1}-1} v - \Theta_N(v - \Pi_{2^{N-1}-1} v + \Pi_{2^{N-1}-1} v)\|_{L^2} \\ &\leq \|\Pi_{2^{N-1}-1} v - \Theta_N(\Pi_{2^{N-1}-1} v)\|_{L^2} + \|v - \Pi_{2^{N-1}-1} v - \Theta_N(v - \Pi_{2^{N-1}-1} v)\|_{L^2}.\end{aligned}$$

Thanks to the orthogonality of the Walsh basis, we have

$$\|v - \Pi_{2^{N-1}-1} v - \Theta_N(v - \Pi_{2^{N-1}-1} v)\|_{L^2} \leq \|v - \Pi_{2^{N-1}-1} v\|_{L^2}.$$

Furthermore, we have the following bound

$$\|v - \Pi_{2^{N-1}-1} v\|_{L^2}^2 = \sum_{|k| \geq 2^{N-1}} |\hat{v}_k|^2 = \sum_{|k| \geq 2^{N-1}} \frac{4\pi^2 k^2}{4\pi^2 k^2} |\hat{v}_k|^2 \leq \frac{1}{\pi^2 4^N} \|v'\|_{L^2}^2,$$

which allows us to get

$$\|v - \Theta_N v\|_{L^2} \leq \frac{1}{\pi 2^N} \|v'\|_{L^2} + \frac{1}{2^{N+1} \cdot \sqrt{3}} \|v'\|_{L^2},$$

and finally (1.21). □

1.2.4 Discussion

In this section, we have described the Walsh functions and provided some results on the explicit decomposition of the Fourier basis functions as Walsh series. These results allowed to provide some new results on the convergence of Walsh series for H^1 functions through their Fourier series. As we shall see in the following chapter, the Walsh series have an important role in the representation of functions on quantum systems and the truncated Walsh series enable the construction of new operator decompositions in the Pauli basis, with a reduced number of non-zero terms. The $O\left(\frac{1}{2^N}\right)$ convergence rate further allows to show that the energy estimation methods based on such series provide sufficient accuracy for variational algorithms.

Chapter 2

Error analysis of VQAs for the Gross-Pitaevskii equation

Chapter description

In this chapter, we aim to determine the performance in practice of the variational algorithm for the resolution of the Gross-Pitaevskii equation defined in [68]. We choose the Gross-Pitaevskii equation because it allows a clear proof of concept for quantum algorithms. It adds a quadratic nonlinearity to the Schrödinger equation, which is well-adapted to our setting, as it governs the evolution of quantum devices.

To do so, we first provide a mathematical framework for the study of the convergence of the discretized solution towards the exact one. This allows to determine the precision required on the energy terms so as to achieve convergence.

We then exhibit the decomposition of the various energy operators in the Pauli basis. Three energy terms appear in the variational formulation of the Gross-Pitaevskii equation. A kinetic energy related to the Laplace operator, a potential one and an interaction one. For the kinetic term, we propose a new decomposition method for sums of 'bi-diagonal' operators, which the Laplacian is a special case of, and recover the Pauli decomposition obtained in [19, 54](up to a basis change). For the potential term, we define a new decomposition method through a Walsh series truncation of the potential function. Finally, for the interaction energy, we provide an original 'squared Pauli' decomposition.

We proceed to define an abstract framework for estimators called sampling strategies, which is sufficiently broad to allow the description of Pauli estimators based on the above decompositions as well as the other estimators found in [68]. We go on to define Pauli estimators for two sampling strategies, namely a diagonal-sampling strategy in which sampling is done in the diagonal basis of the operator, and the importance-sampling strategy in which each term in the operator's Pauli decomposition is sampled separately, through a local basis change, with a number of samples proportional to the absolute value of its associated coefficient. We then provide explicit formulae for the variance of estimators following such strategies.

Using this framework, we define the Pauli sampling estimators associated with the energy terms and we use the above results to obtain their variance explicitly. For the kinetic term, we consider an importance-sampling strategy which allows for shorter circuits than the methods in [68] as only local basis changes are required for the sampling of Pauli operators. For the potential term estimator, we consider the Walsh truncated series of the potential function with

a diagonal-sampling strategy, as it is the only logical choice as the operator is diagonal in the computational basis. We then provide similar results for the interaction term Pauli estimator.

We then consider the 'hardware reduced' estimators of [68], which we denote as direct-sampling estimators. We remark that for the potential term, the estimator is expressed as the diagonal-sampling Pauli estimator of the Walsh interpolant. For the interaction term, we notice that the estimator as defined in the article is biased and provide a modified, unbiased, version of the estimator which we show to be strictly equivalent to our Pauli estimator.

The QNPU Hadamard-test estimators are then briefly Finally, some results on the variance of these estimators are given and a comparison of the estimators, the metric used is the number of samples required to achieve the precision necessary for the convergence of the model for a given number of qubits.

We show that, while the direct/sampling estimators outperform the other methods, the number of samples increase exponentially fast and become prohibitive for even dozens of qubits.

2.1 Introduction: Gross-Pitaevskii equation

In this chapter, we are interested in the numerical resolution of the one-dimensional stationary Gross-Pitaevskii equation (GPE) with periodic boundary condition, denoting by $\Omega = (0, 1)$ the unit cell of a periodic lattice of $\mathbb{R}$. We considera harmonic potential $V : \Omega \to \mathbb{R}$, $x \mapsto V_0(x - 1/2)^2$ with $V_0 > 0$. The unknown is the wave function $u := u(x)$ describing the ground state of the GPE.

The stationary Gross-Pitaevskii equation then consists in a nonlinear eigenvalue problem on u. More precisely, we look for the solutions in the space $\mathcal{X} = H^1_\#(\Omega)$ and where $H^1_\#(\Omega)$ is the Sobolev space of L^2 integrable functions with L^2 integrable first derivatives that are 1-periodic, as already stated in Definition 6. The nonlinear eigenvalue problem reads

$$\left|\begin{array}{l} \text{Find } (\mu, u) \in \mathbb{R} \times \mathcal{X} \text{ such that} \\ A_u u = \mu u, \\ \|u\|_{L^2(\Omega)} = 1, \end{array}\right. \tag{2.1}$$

where the operator A_u in (2.1) is defined, for any wave function v, by

$$A_v = -\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V + f(v),$$

and the nonlinear interaction function f is defined by $f(\xi) = \kappa|\xi|^2$ for any $\xi \in \mathbb{C}$, $\kappa > 0$. The Gross-Pitaevskii equation physically represents the ground state of a system of interacting bosons sharing the same quantum state. The nonlinear interaction function f arises from the fact that the bosons interact with each other attractively. The nonlinear term makes solving the equation harder as linear methods are not directly usable, and finer techniques are required to solve the nonlinear problem. One method consists in linearizing the equation allows its resolution through fixed points methods but requires solving the linearized problem multiple times until convergence is achieved. Another one consists in treating the variational formulation of the equation as a nonlinear optimisation problem, this is the path we shall follow as it is more suited to near-term quantum computing and because of its natural link with VQAs.

In order to solve the equation (2.1), we consider its variational formulation and solve (2.1)

through the following constrained minimization problem

$$\left|\begin{array}{l} \text{Find } u \in \mathcal{X} \text{ such that} \\ E(u) = \min\limits_{v \in \mathcal{X}} E(v), \\ \|u\|_{L^2(\Omega)} = 1, \end{array}\right. \tag{2.2}$$

where E, the associated energy functional has the following form,

$$E(v) = \mathcal{K}(v) + \mathcal{P}(v) + \mathcal{I}(v), \tag{2.3}$$

for any v for which it makes sense. The kinetic, potential and interaction contributions are respectively given by

$$\mathcal{K}(v) = \frac{1}{2}\int_\Omega \left|\frac{\mathrm{d}v(x)}{\mathrm{d}x}\right|^2 \mathrm{d}x, \qquad \mathcal{P}(v) = \int_\Omega V(x)|v(x)|^2\,\mathrm{d}x, \qquad \mathcal{I}(v) = \frac{1}{2}\int_\Omega f(|v(x)|)|v(x)|^2\,\mathrm{d}x.$$

In a way similar to the one explored in [28], we are led to consider the variational formulation on a Fourier subspace $\mathcal{X}_N$ of $\mathcal{X}$.

In our case, we take

$$\mathcal{X}_N = \mathrm{Span}_{\mathbb{C}}\left(\{\mathrm{e}_k : x \mapsto e^{i2\pi kx}, -2^{N-1}-1 \leq k \leq 2^{N-1}-1\} \cup \{c_{2^{N-1}} : x \mapsto \cos\big(2\pi 2^{N-1}x\big)\}\right),$$

and seek the solution of

$$\left|\begin{array}{l} \text{Find } u_N \in \mathcal{X}_N \text{ such that} \\ E_N(u_N) = \min\limits_{v \in \mathcal{X}_N} E_N(v), \\ \|u_N\|_{L^2_N(\Omega)} = 1, \end{array}\right. \tag{2.4}$$

where E_N is a discretization of E which arises as we consider discretized energy estimates of functions $v \in \mathcal{X}_N$ calculated on a quantum computer, and $\|.\|_{L^2_N(\Omega)}$ is a discretization of the L^2-norm, the definitions of which will be given later.

As the quantum state a system of $N \geq 1$ qubits can be represented by a 2^N-sized vector, it can be used to represent functions in subspaces $\mathcal{X}_N$ of $\mathcal{X}$, with exponentially increasing dimensions in N.

A first attempt in using quantum systems for energy estimates to solve the Gross-Pitaevskii equation was given by [68]. In the article, the authors compute the energy associated to a variational state $|\psi\rangle$ using deep circuits called *quantum nonlinear processing units* (QNPUs). A N-qubit variational state $|\psi\rangle$ is used to represent a discretized version of $v \in \mathcal{X}$ through its values on a uniform grid using a finite-difference scheme, analogously, $|\phi\rangle$ is a discretized version of the solution $u \in \mathcal{X}$. The QNPU circuits involve multiple N-qubit registers as inputs to a Hadamard-test circuit in order to compute the discretized versions of $\mathcal{K}$, $\mathcal{P}$, $\mathcal{I}$ separately. To be more accurate, the kinetic term QNPU circuit requires a single N-qubit register, the potential one requires two registers of the same kind, and the nonlinear interaction term three. However, the performance of these circuits is degraded by quantum noise exponentially with respect to their depth/number of gates. To remedy this shortcoming, the authors of [68] also propose direct-sampling estimators for a highly noisy setting consisting of shorter circuits. These direct-sampling estimators require the implementation of the ansatz circuit for the potential and interaction term, but also an additional Quantum Fourier Transform (QFT) circuit to diagonalize the kinetic operator enabling direct measurement.

In this work, we propose alternative estimators to compute the discretized energy functional and provide variance comparisons with the existing estimators and provide an a-priori error analysis of the discretization (2.4) of (2.2).

Our estimators are used to compute the discretized energy functional by involving measurements of Pauli operators on a single register, similarly to the direct-sampling estimators. More precisely, we decompose each energy term operator into a sum of Pauli operators. For the kinetic term, we use a well-known finite-induction relationship, similarly to the methods developed in [19, 54]. For the potential and interaction operators, we propose new estimators obtained by considering, before discretization, the decomposition of the operators in the Walsh functional basis which allow their representation on the binary/dyadic rationals, that is, rationals described using sums of negative powers of 2. Thanks to this representation, we can identify the equivalent Pauli operators for a N-qubit amplitude encoding by truncating the Walsh basis decomposition to N-bit binary/dyadic rationals. Thus, instead of discretizing the operators and projecting them, as in [19], to determine the Pauli decomposition of the discretized potential operator, we consider a functional projection and truncation method. Indeed, as can be seen in Figure 2.1, our methodology consists in analytically describing the operator through its Walsh series, to truncate it, and to identify Walsh functions therein with their Pauli counterpart.

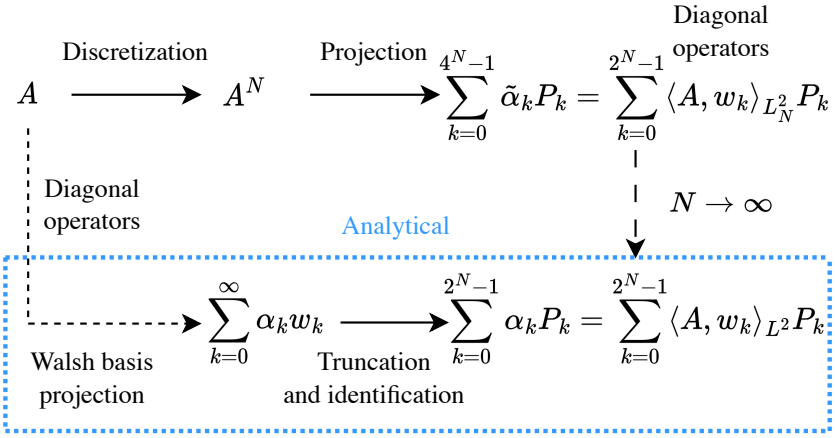


Figure 2.1 – Diagonal operator decomposition in the Pauli basis.
Top: The discretization and projection method has a computational cost $O(N2^N)$ through a Fast Walsh-Hadamard Transform (FWHT) [19]. Bottom: Our method (outlined in blue), through a functional decomposition, achieves a $O(1)$ cost and reduced numbers of non-zero Pauli terms.

The comparison metric we considered is based on the sampling error associated with each estimator. For each method, we consider the number of samples N_{sh}^{ϵ} required to achieve a given absolute error ϵ for each term. We provide upper bounds on the variance of our estimators as well as the one of the estimators described in [68]. For the kinetic term, we show that our method provides reduced circuit depth while only adding a quadratic overhead to the exponential number of samples required to achieve a given error. In the case of the potential term, we show that our method performs similarly to the sampling estimator and that both outperform the Hadamard-test estimator. Finally, we show that in the case of the interaction term, the unbiased direct-sampling and Pauli estimators are the same and that they consistently outperform the Hadamard-test estimator.

In Section 2, we provide a description of the problem we want to solve, the discretization method and present an a-priori convergence analysis. Then, in Section 3, we define how to

encode the solution into a quantum system through amplitude encoding and provide methods to decompose the energy term operators for such an encoding onto the Pauli basis. Using the Pauli decompositions, we describe in Section 4 the Pauli estimators and give formulae and bounds on their variance and estimation error. In Section 5, we define our comparison metrics and compare the different estimators. Finally, in the last section, we discuss our results in the context of variational quantum algorithms for the resolution of differential equations.

2.2 Problem definition

To solve Problem (2.1) through energy estimates obtained by using Noisy Intermediate Scale Quantum (NISQ) processors, we consider a variational quantum algorithm (VQA). VQAs consist in using classical minimisation methods to minimize a cost function estimated through measurements on a quantum system. To obtain those energies, an ansatz quantum state is constructed through a parametric quantum circuit. The depth and structure of these circuits can be fine-tuned so as to decrease the total circuit depth of the algorithm albeit at the cost of the ansatz representativity. The circuit depth of the algorithm can thus be modulated in order to palliate the noise effects, which are be studied in the second part of this work in Chapter 3. In order to encode Problem (2.2) onto a quantum system, we first must discretize it. In this section, we consider the finite-difference approximation, similar to the one defined in [68], of the energy terms in the context of functions in $\mathcal{X}_N$ and provide an a-priori analysis of the convergence of the discretized model.

2.2.1 Discretized problem

Recall that the three operators we wish to discretize are given by

$$\mathcal{K}(v) = \frac{1}{2}\int_\Omega |v'(x)|^2 \,\mathrm{d}x, \qquad \mathcal{P}(v) = \int_\Omega V(x)|v(x)|^2 \,\mathrm{d}x, \qquad \mathcal{I}(v) = \frac{1}{2}\int_\Omega f\left(v(x)\right)|v(x)|^2 \,\mathrm{d}x.$$

To perform a finite-difference discretization of our problem on Ω we introduce a regular subdivision of $\bar{\Omega} = [0,1]$. Our minimisation problem is then discretized as follows. For any $N \in \mathbb{N}^*$, we consider the regular subdivision

$$\Omega_N = \{\mathrm{x}_k = kh_N \mid 0 \leq k \leq 2^N\} \tag{2.5}$$

of Ω, of step $h_N = 1/2^N$ (note that by definition of the periodic condition $x_0 = x_{2^N}$).

Remark 3. *We note here that, given $0 \leq k \leq 2^N$ and $\ell \in \mathbb{N}^*$, the k-th element, x_k, of Ω_N should not be confused with the ℓ-th bit, x_ℓ of the dyadic decomposition of an arbitrary $x \in \Omega$.*

We then denote by v^N the vector representing a function $v \in \mathcal{X}_N$ at each point x_k of the subdivision, *i.e.* for any $0 \leq k \leq 2^N - 1$,

$$v^N = \begin{bmatrix} v(\mathrm{x}_0) \\ \vdots \\ v(\mathrm{x}_{2^N-1}) \end{bmatrix}, \tag{2.6}$$

and for the sake of convenience, we also set $v^N_{2^N} = v^N_0$. We now define the discrete L^2 scalar

product $\langle \cdot, \cdot \rangle_{L^2_N}$ such that, for any $v, w \in \mathcal{X}$,

$$\langle v, w \rangle_{L^2_N} = h_N \sum_{k=0}^{2^N-1} v(\mathrm{x}_k)\overline{w(\mathrm{x}_k)}.$$

The associated norm $\| \cdot \|_{L^2_N}$ on $\mathcal{X}$ is known as the discrete L^2 norm. Furthermore, for any a, $b \in \mathbb{C}^{2^N}$, we define

$$\langle a, b \rangle_{\mathbb{C}^{2^N}} = \sum_{k=0}^{2^N-1} a_k \overline{b_k},$$

the usual Hermitian scalar product of $\mathbb{C}^{2^N}$, and its associated norm $|a|_2^2 = \langle a, a \rangle_{\mathbb{C}^{2^N}}$. Note that, given v, $w \in \mathcal{X}$, the two scalar products are related through v^N and w^N by the following equation

$$\langle v, w \rangle_{L^2_N} = h_N \langle v^N, w^N \rangle_{\mathbb{C}^{2^N}}.$$

Remark 4. *We note that the discrete scalar product result equals the left Riemann sum of the associated integral, in the periodic case, this is equivalent to a Gaussian quadrature. This result will be useful for the a-priori error estimation, as we shall see later.*

As we have access to the values of v_N, we consider a finite-difference approximation of the energy terms on Ω_N.

The discretized energy terms thus read, for any $v \in \mathcal{X}$,

$$\mathcal{K}_N(v) = \frac{1}{h_N} \sum_{k=0}^{2^N-1} \left| v_k^N \right|^2 - \frac{1}{h_N} \sum_{k=0}^{2^N-1} \mathrm{Re}\left(\overline{v_{k+1}^N} v_k^N \right), \tag{2.7}$$

$$\mathcal{P}_N(v) = h_N \sum_{k=0}^{2^N-1} V(\mathrm{x}_k) |v_k^N|^2,$$

$$\mathcal{I}_N(v) = \frac{\kappa h_N}{2} \sum_{k=0}^{2^N-1} |v_k^N|^4. \tag{2.8}$$

The problem we aim to solve is to find the (discretized) solution $u_N \in \mathcal{X}_N$ to the discretized energy problem, *i.e.* the unique function in $\mathcal{X}_N$ satisfying

$$\left| \begin{array}{l} \text{Find } u_N \in \mathcal{X}_N \text{ such that} \\ E_N(u_N) = \min\limits_{v_N \in \mathcal{X}_N} E_N(v_N), \\ \|u_N\|_{L^2_N(\Omega)} = 1, \end{array} \right. \tag{2.9}$$

with

$$E_N(v_N) = \mathcal{K}_N(v_N) + \mathcal{P}_N(v_N) + \mathcal{I}_N(v_N) = -\frac{1}{2}\langle v, \Delta_N v \rangle_{L^2_N} + \langle V, v^2 \rangle_{L^2_N} + \frac{\kappa}{2}\langle v^2, v^2 \rangle_{L^2_N}, \tag{2.10}$$

where

$$\Delta_N v : x \mapsto \frac{v(x + h_N) + v(x - h_N) - 2v(x)}{h_N^2} = \frac{\tau_{h_N} + \tau_{-h_N} - 2I}{h_N^2} v, \tag{2.11}$$

where τ_a is the translation operator by the constant $a \in \mathbb{R}$ (with the proper identification for $x + a$ in Ω whenever $x + a > 1$ or < 0.

2.2.2 Convergence results

From Appendix A.1, which extends the a-priori analysis of the Gross-Pitaevskii equation problem in [28] to our discretized case, we have the convergence rate of u_N to u in H^1 norm is bounded as follows

$$\|u - u_N\|_{H^1} \leq \frac{C}{M},$$

where we set $M = 2^N$. Furthermore, we have, there exists a constant $C > 0$ such that, for any $v \in \mathcal{X}_N$,

$$\|v - u_N\|_{H_1}^2 \leq C \left(E_N(v) - E_N(u_N)\right).$$

This result implies that a precision on the energy of the order of $\frac{1}{M^2}$ is required. Indeed, we have that

2.3 Amplitude encoding and Pauli decomposition

Since we know that our discretized model converges, let us now explain how to obtain the energy estimates using a quantum system. Amplitude encoding is a method through which evaluations of functions can be encoded on a quantum system, the implementation of probability distributions inside quantum states by Grover [40] is probably one of the first instances of amplitude encoding. It has since been used extensively in quantum computing, in fault tolerant algorithms such as Quantum Linear systems algorithms [21, 42], as well as in the NISQ context, for example in the Variational Quantum Linear Solver (VQLS) algorithm [11].

2.3.1 Amplitude encoding

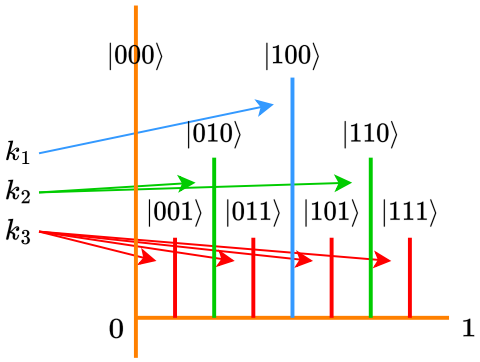


Figure 2.2 – Amplitude encoding on 3-bit dyadic rationals

To estimate E_N for $v \in \mathcal{X}_N$, we choose to represent our functions on a quantum computer, through the amplitude encoding method. The amplitude encoding consists in representing the vector v^N associated to v in some basis, in the vector of amplitudes associated to the basis states of our quantum system. To do so, we have to construct a 2^N-dimensional vector $|\psi\rangle$

representation of our function indexed by N bits. Let us now consider the amplitude encoding of v^N on a N-qubit system, which we denote $|\psi\rangle$ (we drop the N exponent in $|\psi\rangle$ for the sake of simplicity). Given a quantum system composed of N qubits, we define the quantum states of the computational basis using the most-significant-bit-first convention, *i.e.*

$$|k\rangle = |k_1\rangle \cdots |k_N\rangle$$

where $k = \sum_{j=1}^{N} 2^{N-j} k_j$. We consider the following encoding,

$$|\psi\rangle = \sqrt{h_N} v^N = \sqrt{h_N} \sum_{k=0}^{2^N-1} v_k^N |k\rangle , \tag{2.12}$$

of v^N, where $\psi_k = \sqrt{h_N} v_k^N$ for any k, $0 \leq k < 2^N$.

Remark 5. *By construction of Ω_N, we notice that x_k also writes, for any $0 \leq k < 2^N$,*

$$\mathrm{x}_k = k \cdot h_N = 0.k_1 \dots k_N = \sum_{j=1}^{N} \frac{k_j}{2^j}.$$

This corresponds to its representation as a dyadic rational, i.e. a sum of negative powers of 2. This structure implies that the different scales of the Ω are encoded through the qubits, indeed each qubit determines where the dyadic rational is located with respect to each scale. More precisely, the first qubit determines if x_k is located in $[0, 2^{-1})$ or $[2^{-1}, 1)$, the second determines if x_k is located in $[0.k_1, 0.k_1 + 2^{-2})$ or $[0.k_1 + 2^{-2}, 0.k_1 + 2^{-1})$ and so on, as shown on Figure 2.2. As we shall see later, this dyadic encoding will enable to relate operators acting on functions to their equivalent operators acting on the amplitude encoding of functions.

We note that the L_N^2 normalisation constraint is hence enforced by design by our choice of amplitude encoding such that $h_N |v^N|_2^2 = 1$.

Finally, we can rewrite (2.7)–(2.8) as

$$\langle \mathcal{K}^N \rangle_\psi = \mathcal{K}_N(v) = \frac{1}{h_N^2} \sum_{k=0}^{2^N-1} |\psi_k|^2 - \frac{1}{h_N^2} \sum_{k=0}^{2^N-1} \mathrm{Re}\left(\bar{\psi}_{k+1} \psi_k\right), \tag{2.13}$$

$$\langle \mathcal{P}^N \rangle_\psi = \mathcal{P}_N(v) = \sum_{k=0}^{2^N-1} V(\mathrm{x}_k) |\psi_k|^2,$$

$$\langle \mathcal{I}^N \rangle_\psi = \mathcal{I}_N(v) = \frac{\kappa}{2h_N} \sum_{k=0}^{2^N-1} |\psi_k|^4. \tag{2.14}$$

We use the above notation $\langle \cdot \rangle_\psi$, to consider the mean value of Pauli observables on the quantum state $|\psi\rangle$ and estimate the values of the integrals $\mathcal{K}, \mathcal{P}$ and $\mathcal{I}$.

2.3.2 Pauli decomposition of operators

In this subsection, we investigate the Pauli decomposition of each operator, in (2.13)–(2.14). In order to obtain estimators only requiring local Pauli measurements. Let us write the Pauli

decomposition of an operator O as

$$O = \alpha_0 I^N + \sum_{i=1}^{N_O} \alpha_i P_i, \tag{2.15}$$

where I^N is the N-qubit identity operator, N_O is an integer which is non-zero if O is not proportional to I^N, $(\alpha_i)_{0\leq i\leq N_O}$ are complex numbers, and $(P_i)_{1\leq i\leq N_O}$ is a subfamily of $\mathscr{B}_{\mathrm{P}}(\mathscr{L}(\mathcal{H}))$. We first investigate the kinetic term by inductively determining the Pauli decomposition. We then delve into the decomposition of the potential and nonlinear interaction terms by considering their representation on the dyadic rationals through the Walsh basis. That allows to identify Pauli decompositions as the truncated representations of the operators on dyadic rationals, from Definition 1 of Chapter 1, of at most N-bits.

Kinetic operator decomposition

The discretized kinetic energy term (2.7) for the one-dimensional Gross-Pitaevskii equation can be written as

$$\mathcal{K}_N(v^N) = \langle\psi|\, \mathcal{K}^N \,|\psi\rangle\,,$$

where we set

$$\mathcal{K}^N = \frac{1}{2h_N^2}\begin{bmatrix} 2 & -1 & 0 & \dots & 0 & -1 \\ -1 & 2 & -1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -1 & 2 & -1 & 0 \\ 0 & \dots & 0 & -1 & 2 & -1 \\ -1 & 0 & \dots & 0 & -1 & 2 \end{bmatrix}.$$

The Pauli decomposition of the kinetic energy term is then obtained by induction on N. For a given N-qubit system, we decompose $\mathcal{K}^N$ as

$$\mathcal{K}^N = \frac{1}{h_N^2} I^N - \frac{1}{2h_N^2}\left(B_1^N + B_{2^N-1}^N\right). \tag{2.16}$$

with

$$B_1^N = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad B_{2^N-1}^N = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & \ddots & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \ddots & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

More generally, we can define, for $1 \leq k < 2^N$, B_k^N the $2^N \times 2^N$ k-bidiagonal matrix and C_k^N the k-antibidiagonal matrix, the matrix B_k^N, resp. C_k^N, consisting of 1 on its k-th superdiagonal and 1, resp. -1, on its k-th subdiagonal and 0 elsewhere.

We have the following result.

Theorem 4. *For $N \geq 2$, the N-qubit one-dimensional periodic Laplace operator can be decomposed as a linear combination of at most $3 \times 2^{N-2}$ Pauli matrices only involving tensor products of I, X and Y.*

For example,

$$B_3^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \qquad C_6^3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Example 1 – Bi-diagonal matrices

Proof. This decomposition can be obtained from inductive computation of the Pauli decomposition of B_k^n and C_k^n for $1 \leq n \leq N$ and $1 \leq k \leq 2^n - 1$. Indeed, in order to prove Theorem 4, we relate matrices (B_k^n) and (C_k^n), to Pauli matrices. That is the topic of the following lemma, whose proof is a straightforward checking and geometrically intuitive.

Lemma 5. *For any $n \in \mathbb{N}^*$, matrices (B_k^n) and (C_k^n) satisfy*

$$\begin{aligned}
B_0^n &= I^n, \\
B_k^n &= I \otimes B_k^{n-1} + \frac{1}{2}\left(-iY \otimes C_{2^{n-1}-k}^{n-1} + X \otimes B_{2^{n-1}-k}^{n-1}\right), && 1 \leq k \leq 2^{n-1} - 1, \\
B_{2^{n-1}}^n &= X \otimes I^{n-1}, \\
B_k^n &= \frac{1}{2}\left(iY \otimes C_{k-2^{n-1}}^{n-1} + X \otimes B_{k-2^{n-1}}^{n-1}\right), && 2^{n-1} + 1 \leq k \leq 2^n - 1, \\
C_0^n &= 0, \\
C_k^n &= I \otimes C_k^{n-1} + \frac{1}{2}\left(iY \otimes B_{2^{n-1}-k}^{n-1} - X \otimes C_{2^{n-1}-k}^{n-1}\right), && 1 \leq k \leq 2^{n-1} - 1, \\
C_{2^{n-1}}^n &= iY \otimes I^{n-1}, \\
C_k^n &= \frac{1}{2}\left(iY \otimes B_{k-2^{n-1}}^{n-1} + X \otimes C_{k-2^{n-1}}^{n-1}\right), && 2^{n-1} + 1 \leq k \leq 2^n - 1.
\end{aligned}$$

Thanks to Lemma 5, we get, for any $n \geq 1$,

$$\begin{aligned}
B_1^n &= I \otimes B_1^{n-1} + \frac{1}{2}\left(-iY \otimes C_{2^{n-1}-1}^{n-1} + X \otimes B_{2^{n-1}-1}^{n-1}\right), && (2.17) \\
C_1^n &= I \otimes C_1^{n-1} + \frac{1}{2}\left(iY \otimes B_{2^{n-1}-1}^{n-1} - X \otimes C_{2^{n-1}-1}^{n-1}\right), \\
B_{2^n-1}^n &= \frac{1}{2}\left(iY \otimes C_{2^{n-1}-1}^{n-1} + X \otimes B_{2^{n-1}-1}^{n-1}\right), \\
C_{2^n-1}^n &= \frac{1}{2}\left(iY \otimes B_{2^{n-1}-1}^{n-1} + X \otimes C_{2^{n-1}-1}^{n-1}\right), && (2.18)
\end{aligned}$$

which, given the initial conditions $B_0^1 = I, B_1^1 = X$ and $C_0^1 = 0, C_1^1 = iY$, then provides inductive relationships on the sequences $(B_1^n)_{n\geq 1}$, $(B_{2^n-1}^n)_{n\geq 1}$ and $(C_{2^n-1}^n)_{n\geq 1}$.

Thanks to a simple reasoning on (2.17)–(2.18), we obtain that, for any $n \in \mathbb{N}^*$, $B_{2^n-1}^n$ and $C_{2^n-1}^n$ are both linear combinations of at most 2^{n-1} Pauli matrices, and B_1^n is a linear combination of at most $2^n - 1$ Pauli matrices. Going back to (2.16) and then, noticing that, in fact,

$$\mathcal{K}^N = \frac{1}{2h_N^2}(2I^N - I \otimes B_1^{N-1} - X \otimes B_{2^{N-1}-1}^{N-1}), \tag{2.19}$$

it is clear that $\mathcal{K}^N$ is a linear combination of $1+(2^{N-1}-1)+2^{N-2} = 3\times 2^{N-2}$ Pauli matrices. □

From Theorem 4, we can thus write

Proposition 3. *We write the Pauli decomposition of $\mathcal{K}^N$ obtained from* (2.19) *as*

$$\mathcal{K}^N = \alpha_0 I^N + \sum_{k=1}^{N_\mathcal{K}} \alpha_k P_k,$$

with $N_\mathcal{K} = 3 \cdot 2^{N-2}$, we have $\alpha_0 = \frac{1}{h_N^2}$ and, for any k, $1 \leq k \leq N_\mathcal{K}$,

$$|\alpha_k| = \begin{cases} \frac{1}{2h_N^2} 2^{-(|\operatorname{Supp}(P_k)|-2)} & \text{if } P_k|_1 = I, \\ \frac{1}{2h_N^2} 2^{-(N-2)} & \text{else.} \end{cases}$$

Furthermore, the following equalities hold,

$$\sum_{i=1}^{N_\mathcal{K}} |\alpha_i| = \frac{N}{2h_N^2}, \qquad \sum_{i=1}^{N_\mathcal{K}} |\alpha_i|^2 = \frac{1}{2h_N^4}. \tag{2.20}$$

Proof. From (2.19), we notice that all Pauli terms in $B_{2^{N-1}-1}^{N-1}$ have an absolute value of $\frac{1}{h_N^2} 2^{-(N-2)}$, and, in the same way, we observe that Pauli terms in B_1^{N-1} are divided by two for each non-identity component. Hence, the absolute values of the coefficients α_i of the Pauli decomposition for the kinetic term are of the form

$$c_k = \frac{1}{2h_N^2} \cdot 2^{-k}, \qquad 0 \leq k \leq N-2,$$

and the number of terms n_k of Pauli coefficients α_i equal to c_k for each value of k is

$$n_k = \begin{cases} 2^k & \text{if } 0 \leq k \leq N-3, \\ 2^{N-1} & \text{if } k = N-2. \end{cases}$$

Consequently, we have

$$\sum_{i=1}^{N_\mathcal{K}} \alpha_i^2 = \frac{1}{4h_N^4} \sum_{k=0}^{N-2} 2^{-2k} n_k = \frac{1}{4h_N^4} \left(\sum_{k=0}^{N-3} 2^{-k} + 2^{-N+3} \right) = \frac{1}{2h_N^4}$$

and, in the same way,

$$\sum_{i=1}^{N_{\mathcal{K}}} |\alpha_i| = \frac{N}{2h_N^2},$$

which completes the proof. □

Potential operator decomposition

The potential term in our finite-difference approximation context is obtained as the mean value of an N-qubit diagonal operator

$$\tilde{\mathcal{P}}^N = \text{diag}\left(\begin{bmatrix} V(\mathrm{x}_0) \\ \vdots \\ V(\mathrm{x}_{2^N-1}) \end{bmatrix}\right) = \text{diag}(V^N) \tag{2.21}$$

Indeed, the mean value of $\tilde{\mathcal{P}}^N$ is the left-Riemann sum, $\sum_{k=0}^{2^N-1} v^2(\mathrm{x}_k)V(\mathrm{x}_k)h_N$, associated to the integral $\langle v^2, V\rangle_{L^2}$ we wish to estimate, *i.e.*

$$\langle\psi|\,\tilde{\mathcal{P}}^N\,|\psi\rangle = \langle v^2, V\rangle_{L^2_N}$$

with v a L^2_N normalized function. Thus in order to obtain a Pauli estimator for this term, we are led to decompose $\mathcal{P}^N$ in the Pauli basis.

Walsh basis

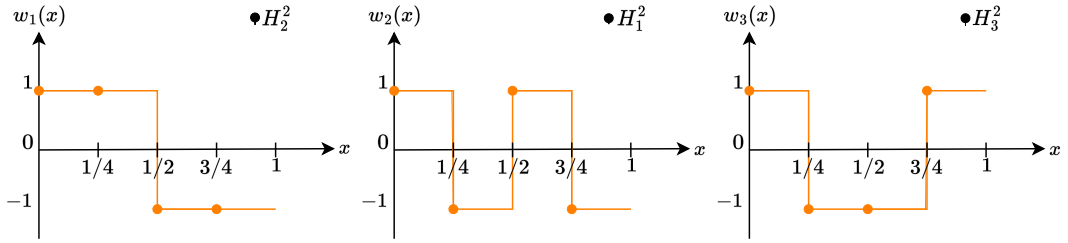


Figure 2.3 – First three Walsh functions and their associated Hadamard-Walsh vector for $N = 2$. As can be seen in the diagram, the N-bit Hadamard-Walsh vector associated with the k-th Walsh function is the $k^\ddagger$-th, with $k^\ddagger$ the N-bit integer whose binary decomposition is in reverse order of k, e.g. w_1 is associated to H_2^2 as 1 is 01 in 2-bit binary and 2 is 10.

The discrete analog of the Walsh basis in $\mathbb{C}^{2^N}$ is given by the Hadamard-Walsh vectors, i.e. the column vectors of the N-qubit Hadamard matrix, H^N. These can be seen as the evaluation of their associated N-bit Walsh functions on Ω_N (as can be seen in Figure 2.3). Given the binary decomposition of the integer

$$k = \sum_{i=1}^{N} k_i 2^{i-1} \in [0, 2^N),$$

$$H^1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H^2 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix},$$

$$H^3 = \frac{1}{2\sqrt{2}} \left(\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ \hline 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{array}\right)$$

Example 2 – First three Hadamard matrices.

and the k-th Walsh function w_k, the column vector of H^N corresponding to the evaluation of w_k on the N-bit dyadic rationals is the column whose index is

$$\sum_{i=1}^{N} k_{N-i+1} 2^{i-1}.$$

This N-bit integer with the previous binary representation is

Definition 7. *Let $N > 0$, and $k = \sum_{i=1}^{N} k_i 2^{i-1} < 2^N$ (i.e. $\sigma_k \leq N$). We define the N-bit integer $k^\ddagger$ as*

$$k^\ddagger = \sum_{i=1}^{N} k_{N-i+1} 2^{i-1},$$

the the N-bit integer whose binary representation is the reverse of k.

Remark 6. *This property arises from the order the elements of Ω_N are counted in. Indeed, given $j = \sum_{i=1}^{N} j_i 2^{i-1}$,*

$$\mathrm{x}_j = j/2^N = \sum_{i=1}^{N} j_i 2^{i-1-N} = \sum_{i=1}^{N} j_{N-i+1} 2^{-i} = 0.j_N \dots j_1.$$

This natural way of counting the elements of Ω_N works only for fixed N and does not represent the same point for different values of N. For example, colorbluelet us consider the index of $\frac{1}{4} = 0.01$ for different values of N, for $N = 2$ its index is $j = 2$, i.e. $j = 01$ in 2-bit binary, in $\Omega_2 = (0, 1/4, 2/4, 3/4)$, while for $N = 3$ its index is 2, i.e. $j = 010$ in 3-bit binary, in $\Omega_3 = (0, 1/8, 1/4, 3/8, 1/2, 5/8, 3/4, 7/8)$. This method of counting the N-bit dyadic rational is shown graphically in Figure 2.4.

The dyadic decomposition offers an alternative way of counting the elements of Ω_N which allows the same indexing of elements for different values of N. Indeed, given a N-bit dyadic rational $y = \sum_{\ell=1}^{N} y_\ell 2^{-\ell} = 0.y_1 \dots y_N$, we can describe it uniquely through the integer $\eta = y_N \dots y_1$,

whereas its index in Ω_N *would be given by* $\eta^\ddagger = y_1 \dots y_N$. *This is shown in Figure 2.5. Indeed,*

$$y = \sum_{i=1}^{N} y_i 2^{-i} = 0.y_1 \dots y_N = \eta^\ddagger / 2^N = \mathrm{x}_{\eta^\ddagger}.$$

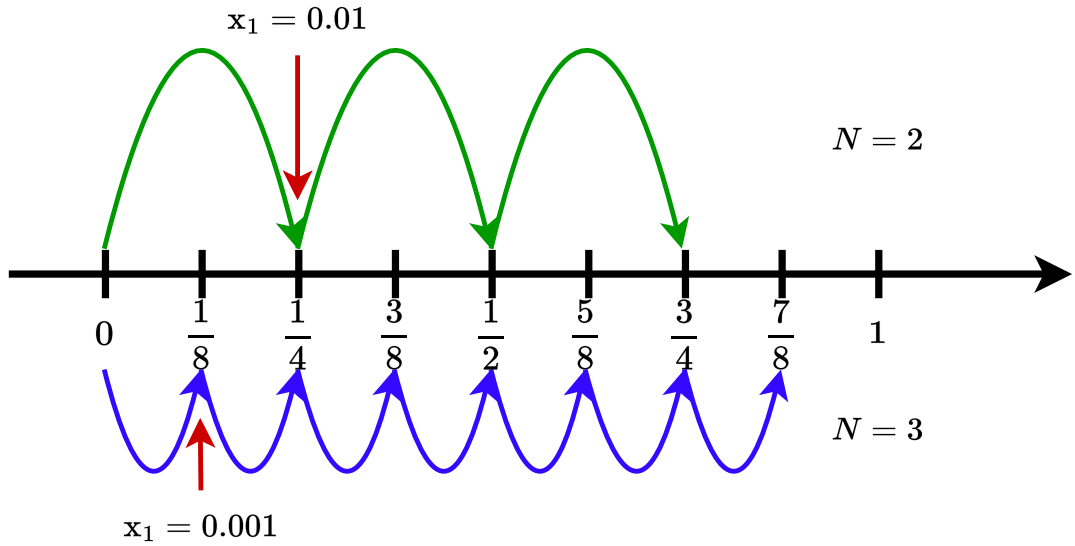


Figure 2.4 – Elements of Ω_N for $N = 2$ (green) and $N = 3$ (blue), the same points are indexed differently for different values of N.

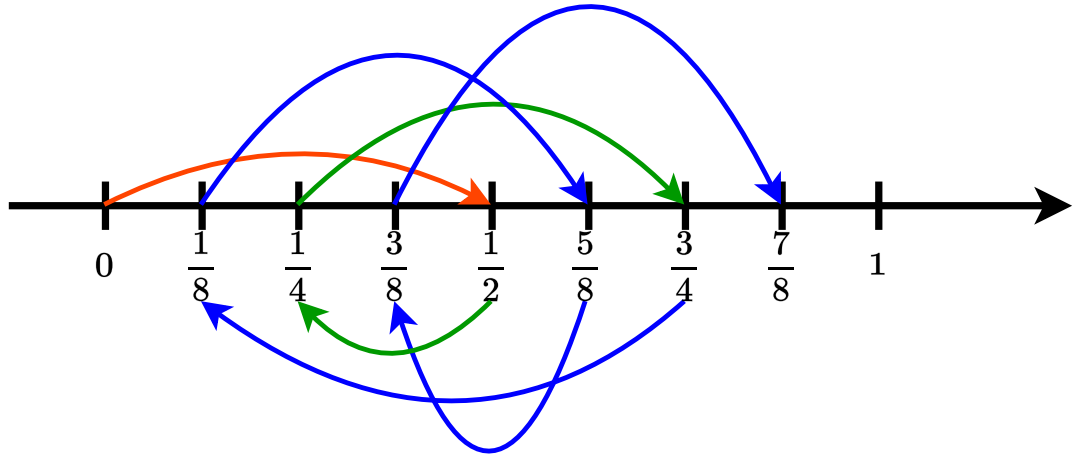


Figure 2.5 – Dyadic rationals of Ω_3 in dyadic ordering, we see that each dyadic rational is indexed only once. Each arrow color corresponds to the number of bits in the next dyadic, orange for 1 bit, green for 2 bits and blue for 3 bits.

We can thus formalize this property.

Proposition 4. *Consider two nonnegative integers $N \geq 1$ and $k \leq 2^N$. The N-bit Hadamard-Walsh vector corresponding to the k-th Walsh function w_k is the $k^\ddagger$-th column vector of H^N, i.e.*

$$H^N_{k^\ddagger} = \sqrt{h_N} w^N_k = \frac{1}{2^{N/2}} \begin{bmatrix} w_k(0) \\ \vdots \\ w_k(\frac{j}{2^N}) \\ \vdots \\ w_k(1 - \frac{1}{2^N}) \end{bmatrix}.$$

In order to relate the scalar product in C^{2^N} with the L^2_N scalar product, we introduce below two useful properties of this relationship between w_k and $H^N_{k^\ddagger}$.

Proposition 5. *Given $N > 0$ and $g \in L^2(\Omega)$, the left-Riemann sum of the integral of $g \in L^2(\Omega)$ and w_k (for $k \in \mathbb{N}^*$) is given by the scalar product of the vector g^N of evaluations of g on Ω_N with the Hadamard-Walsh vector $H^N_{k^\ddagger}$, i.e.*

$$\langle g, w_k \rangle_{L^2_N} = h_N \langle g^N, w^N_k \rangle_{\mathbb{C}^{2^N}} = \sqrt{h_N} \langle g^N, H^N_{k^\ddagger} \rangle_{\mathbb{C}^{2^N}}.$$

We can hence set, in the Hadamard-Walsh basis,

$$g^N = \frac{1}{\sqrt{h_N}} \sum_{k=0}^{2^N-1} \langle g, w_k \rangle_{L^2_N} H^N_{k^\ddagger}. \tag{2.22}$$

With this description of g^N in the Hadamard-Walsh basis, we can show the following property.

Corollary 1. *Given g_1, $g_2 \in L^2(\Omega)$, we can express the Riemann sum of their product as*

$$\langle g_1, g_2 \rangle_{L^2_N} = \sum_{k=0}^{2^N-1} \langle g_1, w_k \rangle_{L^2_N} \langle g_2, w_k \rangle_{L^2_N}. \tag{2.23}$$

Proof. Since the N-bit Hadamard matrix is unitary, we have

$$\begin{aligned} \langle g_1, g_2 \rangle_{L^2_N} &= h_N \langle g^N_1, g^N_2 \rangle_{\mathbb{C}^{2^N}} \\ &= h_N \langle g^N_1, H^N (H^N)^\dagger g^N_2 \rangle_{\mathbb{C}^{2^N}} \\ &= h_N \langle (H^N)^\dagger g^N_1, (H^N)^\dagger g^N_2 \rangle_{\mathbb{C}^{2^N}} \\ &= h_N \sum_{k=0}^{2^N-1} \langle g^N_1, H^N_{k^\ddagger} \rangle_{\mathbb{C}^{2^N}} \langle g^N_2, H^N_{k^\ddagger} \rangle_{\mathbb{C}^{2^N}}, \end{aligned}$$

where we obtain the last line by using (2.22) and using the unitarity of H, *i.e.* by using $\langle H^N_{k^\ddagger}, H^N_{(k')^\ddagger} \rangle_{\mathbb{C}^{2^N}} = \delta_{k,k'}$. This allows to recover (2.23). □

Equation (2.23) allows to relate the representation of g^N on the Hadamard-Walsh basis with the Walsh interpolants

$$\mathrm{I}^{\mathcal{W}}_N g_1 = \sum_{k=0}^{2^N-1} \langle g_1, w_k \rangle_{L^2_N} w_k, \qquad \mathrm{I}^{\mathcal{W}}_N g_2 = \sum_{k=0}^{2^N-1} \langle g_2, w_k \rangle_{L^2_N} w_k$$

of g_1 and g_2 respectively. The Walsh interpolation operator is discussed in more details in Appendix A.2.

Discretized Walsh-Hadamard projection

The decomposition can be obtained by using a fast Hadamard-Walsh transform at the cost of $O(N2^N)$ classical operations, this was used in [19] to obtain the Pauli decomposition of the potential term operator for the linear Schrodinger equation. Consider the Hadamard-Walsh transform of the vector

$$V^N = \begin{bmatrix} V(\mathrm{x}_0) \\ \vdots \\ V(\mathrm{x}_{2^N-1}) \end{bmatrix} = 2^{N/2} \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L_N^2} H_{k^\ddagger}^N.$$

By noticing that the diagonal of the Pauli matrix Z is the column vector $\sqrt{2}H_1^1$, we can identify the N-qubit Hadamard-Walsh vectors $H_{k^\ddagger}^N$ with their associated Pauli operators

$$P_k = Z_1^{k_1} \dots Z_N^{k_N} = 2^{N/2} \operatorname{diag}(H_{k^\ddagger}^N).$$

Hence, we recover the decomposition of our operator in the Pauli basis

$$\tilde{\mathcal{P}}^N = \operatorname{diag}(V^N) = 2^{N/2} \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L_N^2} \operatorname{diag}(H_{k^\ddagger}^N) = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L_N^2} P_k. \tag{2.24}$$

Indeed, using (2.23), we can write $\langle |v^2|, V \rangle_{L_N^2} = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L_N^2} \langle v^2, w_k \rangle_{L_N^2}$, the result is immediate since

$$\langle \psi | P_k | \psi \rangle = 2^{N/2} \sum_{i=0}^{2^N-1} |\psi_i|^2 (H_{k^\ddagger})_i = \sum_{i=0}^{2^N-1} |v(x_i)|^2 w_k(x_i) h_N = \langle v^2, w_k \rangle_{L_N^2}. \tag{2.25}$$

It is formalized in the following proposition.

Proposition 6. *Given a nonnegative integer $k < 2^N$, let $|\psi\rangle = \sqrt{h_N} \sum_{j=0}^{2^N-1} v(\mathrm{x}_j) |j\rangle$, then,*

$$\langle P_k \rangle_\psi = \langle Z_1^{k_1} \dots Z_N^{k_N} \rangle_\psi = \langle v^2, w_k \rangle_{L_N^2}.$$

Unfortunately, the Fast Hadamard-Walsh Transform (FHWT) allows to recover the coefficients $\langle V, w_k \rangle_{L_N^2}$ of $\tilde{\mathcal{P}}^N$ in the Pauli basis at a prohibitive classical cost.

Using the parallel between the Walsh basis and the Hadamard-Walsh transform, we can rewrite the potential term, using (2.24)–(2.25), as

$$\langle \tilde{\mathcal{P}}^N \rangle_\psi = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L_N^2} \langle v^2, w_k \rangle_{L_N^2} = \langle v^2, V \rangle_{L_N^2}.$$

We are led to propose another method, the Walsh-Pauli decomposition. It consists in decomposing the diagonal operator directly in the Walsh basis before discretizing the problem (and not after), and identifying the Walsh functions with their associated Pauli matrices (instead of projecting them numerically). We link this to the method proposed in [19] thanks to the relationship between the Walsh basis and the Hadamard-Walsh vectors. Thanks to this method,

we bypass the costly computations required for the decomposition and to provide an energy term estimator with less terms at a $O(1)$ classical cost when N goes to $+\infty$.

Functional Walsh-Pauli decomposition

Let us now consider the decomposition of a generic potential function V in the Walsh basis and define a new estimator based on its N-bit truncation. Writing V in the Walsh basis gives us

$$V = \sum_{k=0}^{\infty} \langle w_k, V \rangle_{L^2} w_k.$$

In a N-qubit system, the Walsh-Pauli decomposition $\mathcal{P}^N$ of $\mathcal{P}$ is obtained through the identification of the Walsh functions w_k with the associated Pauli matrices $P_k = Z_1^{k_1} \cdots Z_N^{k_N}$ in the truncation

$$\Theta_N V = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L^2} w_k.$$

Hence, the Walsh-Pauli decomposition of the potential term is given by

$$\mathcal{P}^N = \sum_{k=0}^{2^N-1} \langle V, w_k \rangle_{L^2} P_k.$$

Let us now define the Walsh interpolation operator which is explored in more details in Appendix A.2.

Definition 8. *Set $\mathcal{W}_N = \mathrm{Span}_{\mathbb{C}}\{w_k \mid 0 \leq k \leq 2^N - 1\}$. The Walsh interpolation operator $\mathrm{I}_N^{\mathcal{W}} : L^2(0,1) \to \mathcal{W}_N$ is defined, for any $v \in L^2(0,1)$, by*

$$\langle \mathrm{I}_N^{\mathcal{W}} v, g \rangle_{L_N^2} = \langle v, g \rangle_{L_N^2}, \, \forall g \in L^2(0,1).$$

Equivalently, $\mathrm{I}_N^{\mathcal{W}} v$ is the N-bit Walsh series in $\mathcal{W}_N$ that takes the same values as v on Ω_N. From Proposition 25, the interpolant can be written as

$$\mathrm{I}_N^{\mathcal{W}} v = \sum_{m=0}^{2^N-1} \langle v, w_m \rangle_{L_N^2} w_m = \sum_{m=0}^{2^N-1} \left(\tilde{v}_m + \sum_{\ell=1}^{+\infty} \tilde{v}_{m \oplus 2^N \ell} \right) w_m = \Theta_N v + \mathrm{R}_N^{\mathcal{W}} v,$$

where $\mathrm{R}_N^{\mathcal{W}}$ is known as the Walsh aliasing operator.

Hence, for the Pauli decomposition of the potential operator, we have, on the one hand,

$$\langle \psi | \, \tilde{\mathcal{P}}^N \, | \psi \rangle = \sum_{k=0}^{2^N-1} \langle w_k, V \rangle_{L_N^2} \langle v^2, w_k \rangle_{L_N^2} = \langle V, v^2 \rangle_{L_N^2} = \langle \mathrm{I}_N^{\mathcal{W}} V, v^2 \rangle_{L_N^2}.$$

which corresponds to the L_N^2 scalar product with the Walsh interpolant $\mathrm{I}_N^{\mathcal{W}} V$ of V, and on the other hand,

$$\langle \psi | \, \mathcal{P}^N \, | \psi \rangle = \sum_{k=0}^{2^N-1} \langle w_k, V \rangle_{L^2} \langle v^2, w_k \rangle_{L_N^2} = \langle \Theta_N V, v^2 \rangle_{L_N^2}, \tag{2.26}$$

which is associated with its truncation $\Theta_N V$, see Chapter 1 Section 1.2.3 for more details.

We thus obtain the main advantage of our decomposition method, namely that it does not require any computation, compared with a $O(N2^N)$ cost in the case of the previous method requiring a Walsh-Hadamard transform. By using the L^2-scalar product instead of the L^2_N one, we recover the truncation of the Walsh series instead of the interpolation which contains additional non-zero terms due to aliasing caused by frequencies higher than 2^N in the Walsh series of V. Hence, we can reduce the number of terms in the Pauli decomposition. Indeed, for instance , we have the following result when dealing with monomial functions [25].

Proposition 7. *Let $r \in \mathbb{N}$ and consider the monomial x^r. The only non-zero Hadamard-Walsh coefficients are those whose indices in their binary decomposition contain at most r occurrences of 1. Moreover, for any $m \in \mathbb{N}^*$, the m-th Walsh coefficient of the Walsh series of x^r is inductively given by*

$$\mu_{r,m} = \begin{cases} 0 & \text{if } |m|_{\text{Ham}} > r, \\ \dfrac{(-1)^r r!}{2^r} \displaystyle\prod_{\ell=1}^{\sigma_m} \left[1 - m_\ell + m_\ell 2^{-\ell}\right] & \text{if } |m|_{\text{Ham}} = r, \\ \dfrac{(-1)^{r-1} r!}{2^r} \displaystyle\prod_{\ell=1}^{\sigma_m} \left[1 - m_\ell + m_\ell 2^{-\ell}\right] & \text{if } |m|_{\text{Ham}} = r-1, \\ \dfrac{-r}{2^{\sigma_m+1}} \mu_{r-1,m-2^{\sigma_m}} + \displaystyle\sum_{\ell=1}^{+\infty} \frac{r}{2^{\sigma_m+1+\ell}} \mu_{r-1,m+2^{\sigma_m+\ell}} & \text{if } 1 \le |m|_{\text{Ham}} < r-1, \end{cases}$$

and the zeroth coefficient is $\mu_{r,0} = \dfrac{1}{r+1}$.

Applying the previous proposition and (2.26) to our harmonic potential lead to the Walsh-Pauli decomposition of the potential term.

Proposition 8. *For $V(x) = (x - \frac{1}{2})^2$, we have*

$$\mathcal{P}^N = \frac{V_0}{12} \left(I^N + 3 \sum_{\ell_1=1, \ell_2 > \ell_1}^{N} 2^{-(\ell_1+\ell_2-1)} Z_{\ell_1} Z_{\ell_2} \right), \tag{2.27}$$

and $N_{\mathcal{P}} = N(N-1)/2$.

Walsh truncation convergence The truncated Walsh-Pauli estimator converges in L^2 norm as $O(1/2^N)$ for H^1 functions as is shown in Theorem 3. In the case of the potential term estimator, we successively have

$$\begin{aligned} |\langle \tilde{\mathcal{P}}_N \rangle_\psi - \langle \mathcal{P}_N \rangle_\psi| = |\langle V - \Theta_N V, v^2 \rangle_{L^2_N}| &= |\langle V - \Theta_N V, v^2 - \Theta_N(v^2) \rangle_{L^2_N}| \\ &\le \|V - \Theta_N V\|_{L^2_N} \|v^2 - \Theta_N(v^2)\|_{L^2_N} \\ &\le C \|V - \Theta_N V\|_{L^2} \|v^2 - \Theta_N(v^2)\|_{L^2} \le \frac{C}{4^N} \|V\|_{H^1} \|v\|^2_{H^1}, \end{aligned}$$

where we applied Theorem 3 and used the fact that H^1 is a Banach algebra. From Section 2.2.2, the precision required on the energy estimates is of the order of $O(\frac{1}{4^N})$ and hence the estimator can be used without loss of performance.

Remark 7. *If better convergence rates are needed, we can recover the Walsh interpolation of any function analytically from its Walsh series. Indeed, the Walsh interpolant, from Definition 8, coincides with the Fourier interpolant for the discrete L^2_N scalar product and hence the convergence rate of the resulting estimator only depends on the regularity of the function considered. Note that this would be at the cost of an increased number of Pauli terms and would be strictly equivalent to the direct-sampling methods devised in [68].*

Interaction operator decomposition

The finite-difference approximation of the interaction term can be obtained as the sum of the squared probability distribution of $|\psi\rangle$. Indeed, given $v \in \mathcal{X}$, we have, for any $N \geq 2$,

$$\mathcal{I}_N(v^N) = \frac{\kappa}{2}\langle v^2, v^2\rangle_{L^2_N} = \frac{\kappa}{2h_N}\sum_{i=0}^{2^N-1}|\psi_i|^4 = \langle \mathcal{I}^N\rangle_\psi,$$

where we recall that v, v^N and $|\psi\rangle$ are related through (2.12) and (2.6). By considering the functional representation of the interaction term (before discretization) in the Walsh basis and truncating the exact representation of the functional operator, we obtain the decomposition of this operator as a weighted sum of squared Pauli values

$$\langle \mathcal{I}^N\rangle_\psi = \frac{\kappa}{2}\sum_{k=0}^{2^N-1}\langle\psi| Z_1^{k_1}\dots Z_N^{k_N}|\psi\rangle^2 .$$

Let us write the decomposition of v^2 on the Walsh functional basis

$$v^2 = \sum_{k=0}^{\infty}\langle v^2, w_k\rangle_{L^2} w_k.$$

From the previous equality, we compute

$$\mathcal{I}(v) = \frac{\kappa}{2}\langle v^2, v^2\rangle_{L^2} = \frac{\kappa}{2}\sum_{k,k'=0}^{+\infty}\langle v^2, w_k\rangle_{L^2}\overline{\langle v^2, w'_k\rangle_{L^2}}\langle w_{k'}, w_k\rangle_{L^2} = \frac{\kappa}{2}\sum_{k=0}^{+\infty}|\langle v^2, w_k\rangle_{L^2}|^2.$$

By truncating and identifying the Walsh function w_k with $P_k = Z_1^{k_1}\dots Z_N^{k_N}$, using equation (2.25), we obtain the Pauli decomposition of the interaction operator

$$\langle \mathcal{I}^N\rangle_\psi = \frac{\kappa}{2}\sum_{k=0}^{2^N-1}\langle v^2, w_k\rangle^2_{L^2_N} = \frac{\kappa}{2}\sum_{k=0}^{2^N-1}\langle\psi| P_k|\psi\rangle^2 . \tag{2.28}$$

In this case, note that the projection of the discretized operator and the truncation of the functional operator results in the same Pauli decomposition. Indeed, using (2.23) we obtain

$$\mathcal{I}_N(v) = \frac{\kappa}{2}\sum_{k=0}^{2^N-1}\langle w_k, v^2\rangle^2_{L^2_N}.$$

We thus have a Pauli decomposition of $N_{\mathcal{I}} = 2^N - 1$ squared Pauli terms for the interaction operator.

Proposition 9. *The interaction energy term can be calculated as the following sum of squared Pauli means,*

$$\mathcal{I}_N(v) = \frac{\kappa}{2} \sum_{k=0}^{2^N-1} \langle P_k \rangle_\psi^2,$$

where, for any k,$0 \leq k < 2^N$, $P_k = Z_1^{k_1} \dots Z_N^{k_N}$.

Remark 8. *Alternatively, we can see this as a decomposition of a linear term on a* $2N$ *qubit system, indeed,* $|\psi_i|^4$ *terms can only be obtained as the mean value of two registers in the state* $|\psi\rangle$*. We decompose*

$$|\psi_i|^4 = \langle\psi| \langle\psi| \left(|i\rangle\langle i| \otimes |i\rangle\langle i|\right) |\psi\rangle |\psi\rangle ,$$

thus we have

$$\mathcal{I}^N = \frac{\kappa}{2h_N} \sum_{i=0}^{2^N-1} |i\rangle \langle i| \otimes |i\rangle \langle i| = \sum_{i=0}^{2^N-1} |i_1\rangle \langle i_1| \otimes \dots \otimes |i_N\rangle \langle i_N| \otimes |i_1\rangle \langle i_1| \otimes \dots \otimes |i_N\rangle \langle i_N|$$

for a system with two N*-qubit registers in the state* $|\psi\rangle$ *. We can rewrite* $|i_k\rangle\langle i_k| = \frac{I_k + (-1)^{i_k} Z_k}{2}$ *and we obtain*

$$\mathcal{I}^N = \frac{\kappa}{2} \prod_{n=1}^{N} (1 + Z_n Z_{N+n}) = \frac{\kappa}{2} \sum_{k=0}^{2^N-1} Z_1^{k_1} \dots Z_N^{k_N} Z_{1+N}^{k_1} \dots Z_{2N}^{k_N},$$

thus recovering (2.28)*.*

Having obtained the Pauli decomposition for each of our energy terms, let us now define the energy estimators based on those decompositions.

2.4 Pauli estimators

In this section, we define Pauli estimators based on the Pauli decomposition of operators. We provide a framework, sampling strategies, which formalises the method through which the operators are sampled and reconstructed and allows us to consider different sampling methods based on the Pauli decomposition of operators. We then obtain formulae for the variance of Pauli estimators following specific sampling strategies, namely importance-sampling and diagonal-sampling, and apply the results to our energy terms. Let us start by defining the concept of a sampling strategy.

2.4.1 Sampling strategy

A complete set of commuting observables (CSCO) of $\mathscr{L}(\mathcal{H})$ allows the reconstruction of any observable commuting with it. Measuring a CSCO is equivalent to measuring in the unique (up to a permutation) basis where all the terms in the CSCO are diagonal. Given an observable O and a CSCO $\mathcal{O}_{\text{ex}}$, if O commutes with every $O' \in \mathcal{O}_{\text{ex}}$, then O is diagonal in the same unique basis and can be reconstructed as a linear combination of elements of $\mathcal{O}_{\text{ex}}$. In practice, we can only measure quantum systems through their Z-Pauli operators. Thus, measuring a CSCO $\mathcal{O}_{\text{ex}}$ corresponds to

For example, consider the CSCO composed of the Z-Pauli matrices for a N-qubit system:

$$\mathcal{O}_Z = \{Z_k \mid 1 \leq k \leq N\}.$$

As each term in the CSCO is diagonal in the computational basis, the measurement of $\mathcal{O}_Z$ is equivalent to the measurement in the computational basis, and any diagonal operator can be reconstructed from the Z-Pauli matrices.

Example 3 – CSCO example

Consider the interaction term $\mathcal{I}_N$ from (2.28), we can choose

$$\mathcal{A} = \{A_k = Z_1^{k_1} \cdots Z_N^{k_N} \mid 1 \leq k \leq 2^N - 1\}.$$

The associated nonlinear function g is

$$g(y_1, \dots y_{2^N-1}) = \frac{\kappa}{2} + \frac{\kappa}{2} \sum_{k=1}^{2^N-1} |y_k|^2.$$

Example 4 – Nonlinear reconstruction

measuring the Z-Pauli operators after the application of a unitary $U_{\mathcal{O}_{\text{ex}}}$ that sends the system into the unique diagonalization basis of $\mathcal{O}_{\text{ex}}$. Let us denote by $\Lambda_{\hat{O}}$ the set of eigenvectors of an observable $\hat{O}$. Then we define $\mathit{\Lambda}_{\mathcal{O}_{\text{ex}}}$ the set of possible samples from the CSCO $\mathcal{O}_{\text{ex}}$, *i.e.*

$$\mathit{\Lambda}_{\mathcal{O}_{\text{ex}}} = \Lambda_{(\mathcal{O}_{\text{ex}})_1} \times \cdots \times \Lambda_{(\mathcal{O}_{\text{ex}})_{|\mathcal{O}_{\text{ex}}|}},$$

where we denote by $(\mathcal{O}_{\text{ex}})_k$ the k-th observable of $\mathcal{O}_{\text{ex}}$. A sample of the CSCO $\mathcal{O}_{\text{ex}}$ is then vector of $\mathit{\Lambda}_{\mathcal{O}_{\text{ex}}}$.

Given a budget of N_{sh} samples, $N_{\text{sh}} \geq 1$, to estimate the mean value of a potentially nonlinear operator A for a quantum state $|\psi\rangle$, we proceed in the following way. We first identify an indexed set of N_A observables $\mathcal{A} = \{\mathcal{A}_i \mid 1 \leq i \leq N_A\}$, $N_A \geq 1$, such that the mean value of A can be expressed as a possibly nonlinear function g of the mean values of the $\mathcal{A}_i$, i.e. we have

$$\langle A \rangle_\psi = g(\langle \mathcal{A}_1 \rangle_\psi, \dots, \langle \mathcal{A}_{N_A} \rangle_\psi).$$

Next, we select an indexed set of N_O CSCOs, $\mathcal{O} = \{\mathcal{O}_j \mid 1 \leq j \leq N_O\}$, with $1 \leq N_O \leq N_A$, and the associated basis change circuits $(U_{\mathcal{O}_j})$. Each observable $\mathcal{A}_k$ must be reconstructible from at least one $\mathcal{O}_j$ as a linear combination of the observables in $\mathcal{O}_j$. We define the reconstruction set $\mathcal{R} \subset \{1, \dots, N_A\} \times \{1, \dots, N_O\}$, such that, for any $(i, j) \in \mathcal{R}$, the following equivalent properties hold:

- $\mathcal{A}_i$ can be reconstructed as a linear combination of observables in $\mathcal{O}_j$,
- every observable in $\mathcal{O}_j$ commutes with $\mathcal{A}_i$,
- $\mathcal{A}_i$ is co-diagonal with every observable in $\mathcal{O}_j$.

Thus we can write

$$\mathcal{R} = \{(i, j) \in \{1, \dots, N_A\} \times \{1, \dots, N_O\} \mid \mathcal{A}_i \hat{O} = \hat{O} \mathcal{A}_i \ \forall \hat{O} \in \mathcal{O}_j\}.$$

Definition 9. *A sampling strategy is described by the tuple* (Ξ, Υ)*, where*

$$\Xi = (\Xi_k)_{1\leq k\leq N_O} \in [0,1]^{N_O}$$

is the vector representing the proportion of samples allocated to each CSCO in $\mathcal{O}$*, and*

$$\Upsilon \subset \mathcal{R}$$

indicates the subset of reconstructions which were used. The elements of Υ *are the couples of indices* (i,j) *specifying whether the samples of* $\mathcal{O}_j$ *are used to reconstruct the term* $\mathcal{A}_i$ *in the strategy.*

Given a sampling strategy (Ξ, Υ) *and* N_{sh} *samples, we define* $\xi^{(N_{\mathrm{sh}})} \in \{1,\dots,N_O\}^{N_{\mathrm{sh}}}$ *the vector of CSCO assignments. The measured samples*

$$s = \left\{ s_k \in \Lambda_{\mathcal{O}_{\xi_k^{(N_{\mathrm{sh}})}}} \mid 1 \leq k \leq N_{\mathrm{sh}} \right\}$$

are then used to reconstruct each $\mathcal{A}_i$ *according to* Υ*. For a given* i*,* $1 \leq i \leq N_A$*, the set of samples used to reconstruct* $\mathcal{A}_i$ *is denoted by*

$$S_i = \left\{ k \in \{1,\dots,N_{\mathrm{sh}}\} \mid \left(i, \xi_k^{(N_{\mathrm{sh}})}\right) \in \Upsilon \right\}.$$

2.4.2 Pauli estimator

In what follows, the observables will be the Pauli matrices as we can write the Pauli decomposition (2.15) of any operator A, *i.e.*

$$A = \alpha_0 I + \sum_{i=1}^{N_A} \alpha_i P_i, \tag{2.29}$$

where $N_A \geq 1$ if A is not a homothethy, $(\alpha_i)_{0\leq i\leq N_A}$ are (a priori) complex numbers and $(P_i)_{1\leq i\leq N_A}$ is a subfamily of $\mathscr{B}_{\mathrm{P}}(\mathscr{L}(\mathcal{H}))$. We recall that, for any i, S_i is the set of samples used to reconstruct P_i, and we denote by N_{sh}^i the cardinality of S_i. The reconstructed value of each P_i is given by

$$\overline{P}^{N_{\mathrm{sh}}^i} = \frac{1}{N_{\mathrm{sh}}^i} \sum_{k=1}^{N_{\mathrm{sh}}^i} \mathbf{p}_i^k, \tag{2.30}$$

where $\mathbf{p}_i^k \in \{-1,1\}$ is the reconstructed value of the P_i observable for the k-th sample in S_i. Then (2.29)–(2.30) imply that the Pauli sampling estimator $\overline{A}^{N_{\mathrm{sh}}}$ for A is

$$\overline{A}^{N_{\mathrm{sh}}} = \alpha_0 + \sum_{i=1}^{N_A} \alpha_i \overline{P_i}^{N_{\mathrm{sh}}^i}. \tag{2.31}$$

Let us now state a general result of the expected value and the variance of such an estimator.

Proposition 10. *The expected value of* $\overline{A}^{N_{\mathrm{sh}}}$ *satisfies*

$$\mathbb{E}\left[\overline{A}^{N_{\mathrm{sh}}}\right] = \langle A \rangle_\psi,$$

and its variance is given by

$$\operatorname{Var}\overline{A}^{N_{\mathrm{sh}}} = \sum_{i=1}^{N_A} \frac{\alpha_i^2}{N_{\mathrm{sh}}^i}\left(1 - \langle P_i\rangle_\psi^2\right) + \sum_{i=1}^{N_A}\sum_{j\neq i} \alpha_i\alpha_j \frac{|S_i \cap S_j|}{N_{\mathrm{sh}}^i N_{\mathrm{sh}}^j}\left(\langle P_iP_j\rangle - \langle P_i\rangle_\psi\langle P_j\rangle_\psi\right). \tag{2.32}$$

Proof. We first have, from (2.31),

$$\mathbb{E}\left[\overline{A}^{N_{\mathrm{sh}}}\right] = \alpha_0 + \sum_{i=1}^{N_A} \alpha_i \mathbb{E}\left[\overline{P_i}^{N_{\mathrm{sh}}^i}\right]. \tag{2.33}$$

Next, for any i, $1 \leq i \leq N_A$, we can write

$$\mathbb{E}\left[\overline{P_i}^{N_{\mathrm{sh}}^i}\right] = \frac{1}{N_{\mathrm{sh}}^i}\sum_{k=1}^{N_{\mathrm{sh}}^i} \mathbb{E}\left[\mathbf{p}_i^k\right] = \langle P_i\rangle_\psi.$$

Plugging (2.29) in (2.33), implies that

$$\mathbb{E}\left[\overline{A}^{N_{\mathrm{sh}}}\right] = \alpha_0 + \sum_{i=1}^{N_A} \alpha_i\langle P_i\rangle_\psi = \langle A\rangle_\psi.$$

Next, the variance of the estimator is given by

$$\begin{aligned}\operatorname{Var}\overline{A}^{N_{\mathrm{sh}}} &= \mathbb{E}\left[\left(\overline{A}^{N_{\mathrm{sh}}}\right)^2\right] - \mathbb{E}\left[\overline{A}^{N_{\mathrm{sh}}}\right]^2 \\ &= \sum_{i=1}^{N_A}\sum_{i=1}^{N_A} \alpha_i\alpha_j\left(\mathbb{E}\left[\overline{P_i}^{N_{\mathrm{sh}}^i}\overline{P_j}^{N_{\mathrm{sh}}^j}\right] - \mathbb{E}\left[\overline{P_i}^{N_{\mathrm{sh}}^i}\right]\mathbb{E}\left[\overline{P_j}^{N_{\mathrm{sh}}^j}\right]\right) \\ &= \sum_{i=1}^{N_A} \alpha_i^2 \operatorname{Var}\left(\overline{P_i}^{N_{\mathrm{sh}}^i}\right) + \sum_{i=1}^{N_A}\sum_{j\neq i} \alpha_i\alpha_j \operatorname{Cov}(\overline{P_i}^{N_{\mathrm{sh}}^i}, \overline{P_j}^{N_{\mathrm{sh}}^j}).\end{aligned} \tag{2.34}$$

The variance terms in the right-hand side of (2.34) are given, for each i, by

$$\begin{aligned}\operatorname{Var}\overline{P_i}^{N_{\mathrm{sh}}^i} &= \left(\frac{1}{\left(N_{\mathrm{sh}}^i\right)^2}\sum_{k=1}^{N_{\mathrm{sh}}^i}\sum_{\ell=1}^{N_{\mathrm{sh}}^i} \mathbb{E}\left[\mathbf{p}_i^k\mathbf{p}_i^\ell\right] - \mathbb{E}\left[\mathbf{p}_i^k\right]\mathbb{E}\left[\mathbf{p}_i^\ell\right]\right) \\ &= \frac{1}{(N_{\mathrm{sh}}^i)^2}\sum_{k=1}^{N_{\mathrm{sh}}^i} \mathbb{E}\left[(\mathbf{p}_i^k)^2\right] - \mathbb{E}\left[\mathbf{p}_i^k\right]^2 = \frac{1}{N_{\mathrm{sh}}^i}\left(1 - \langle P_i\rangle_\psi^2\right).\end{aligned} \tag{2.35}$$

The covariance term in (2.34) between the estimators of the i-th and j-th Pauli matrices depends on the shared samples between both. Indeed, it writes

$$\operatorname{Cov}\left(\overline{P_i}^{N_{\mathrm{sh}}^i}, \overline{P_j}^{N_{\mathrm{sh}}^j}\right) = \frac{1}{N_{\mathrm{sh}}^i N_{\mathrm{sh}}^j}\sum_{k=1}^{N_{\mathrm{sh}}^i}\sum_{\ell=1}^{N_{\mathrm{sh}}^j}\left(\mathbb{E}\left[\mathbf{p}_i^k\mathbf{p}_j^\ell\right] - \mathbb{E}\left[\mathbf{p}_i^k\right]\mathbb{E}\left[\mathbf{p}_j^\ell\right]\right). \tag{2.36}$$

Each contribution $\operatorname{Cov}\left(\mathbf{p}_i^k, \mathbf{p}_j^\ell\right) = \mathbb{E}\left[\mathbf{p}_i^k\mathbf{p}_j^\ell\right] - \mathbb{E}\left[\mathbf{p}_i^k\right]\mathbb{E}\left[\mathbf{p}_j^\ell\right]$ is non-zero if and only if $\mathbf{p}_i^k$ and $\mathbf{p}_j^\ell$

are not independent. Then the only non-zero terms are the ones where the k-th sample of P_i is the same as the ℓ-th sample of P_j, in which case, we have

$$\mathrm{Cov}\left(\mathbf{p}_i^k, \mathbf{p}_j^\ell\right) = \langle P_i P_j \rangle_\psi - \langle P_i \rangle_\psi \langle P_j \rangle_\psi.$$

Hence, (2.36) becomes

$$\mathrm{Cov}\left(\overline{P_i}^{N_{\mathrm{sh}}^i}, \overline{P_j}^{N_{\mathrm{sh}}^j}\right) = \frac{|S_i \cap S_j|}{|S_i||S_j|} \left(\langle P_i P_j \rangle_\psi - \langle P_i \rangle_\psi \langle P_j \rangle_\psi\right). \tag{2.37}$$

Using (2.35) and (2.37) in (2.34) allows to conclude. □

We now define the two sampling strategies which we consider in the rest of this chapter. First, we focus on at the importance-sampling strategy, which consists in sampling the Pauli terms in the decomposition of an operator proportionally to the absolute value of their coefficient. Second, we discuss the diagonal-sampling strategy which consists in sampling in the diagonal basis of the operator. These two strategies are at opposite ends of sampling strategies in terms of the basis-change circuits requirements and thus in terms of sensitivity to noise, as we shall see in Chapter 3.

Definition 10 (Importance-sampling strategy [99])**.** *Given $A = \alpha_0 I + \sum_{i=1}^{N_A} \alpha_i P_i$, $N_A \geq 1$. We set*

$$\alpha_{1:} = (\alpha_i)_{1 \leq i \leq N_A} \in \mathbb{C}^{N_A}, \qquad \mathcal{A} = \{P_i \mid 1 \leq i \leq N_A\}. \tag{2.38}$$

For each P_i, $1 \leq i \leq N_A$, we define a CSCO composed of its local operators and completed by Z-Pauli operators

$$\mathcal{O}_i = \{P_i|^k \text{ if } P_i|_k \neq I \text{ else } Z_k \mid 1 \leq k \leq N\},$$

where we set $P_i|^k = \bigotimes_{i=1}^N (P_i|_k)^{\delta_{i,k}}$. This allows to define the set of CSCOs to be sampled

$$\mathcal{O}_{\mathrm{Imp}} = \{\mathcal{O}_i \mid 1 \leq i \leq N_A\}.$$

The importance-sampling strategy for A is the tuple $(\Xi_{\mathrm{Imp}}, \Upsilon_{\mathrm{Imp}})$, where

$$\Xi_{\mathrm{Imp}} = \{|\alpha_i|/\|\alpha_{1:}\|_1 \mid 1 \leq i \leq N_A\},$$

with $\|\cdot\|_1$ the 1-norm on $\mathbb{C}^{N_A}$, and

$$\Upsilon_{\mathrm{Imp}} = \{(i,i) \mid 1 \leq i \leq N_A\}.$$

Proposition 11. *The variance of $\overline{A}^{N_{\mathrm{sh}}}$ associated to an importance-sampling strategy is given by*

$$\mathrm{Var}\,\overline{A}^{N_{\mathrm{sh}}} = \frac{\|\alpha_{1:}\|_1}{N_{\mathrm{sh}}} \sum_{i=1}^{N_A} |\alpha_i| \left(1 - \langle P_i \rangle_\psi^2\right). \tag{2.39}$$

Proof. We notice that, for any i, $j \in \{1, \dots N_A\}$ such that $i \neq j$, $S_i \cap S_j = \emptyset$ and that $N_{\mathrm{sh}}^i = \frac{|\alpha_i|}{\|\alpha_{1:}\|_1} N_{\mathrm{sh}}$. Using (2.32), it is easy to recover (2.39). □

Definition 11 (Diagonal-sampling strategy)**.** *Let us denote by U_A the basis-change operator into*

the diagonalization basis of A. We set

$$\mathcal{O}_1 = \{U_A^\dagger Z_k U_A \mid 1 \leq k \leq N\}, \qquad \mathcal{O}_{\text{diag}} = \{\mathcal{O}_1\}\,.$$

The diagonal-sampling strategy is given by the tuple $(\Xi_{\text{diag}}, \Upsilon_{\text{diag}})$, where

$$\Xi_{\text{diag}} = \{1\}, \qquad \Upsilon_{\text{diag}} = \{(i,1) \mid 1 \leq i \leq N_A\}.$$

Proposition 12. *The variance of $\overline{A}^{N_{\text{sh}}}$ associated to a diagonal-sampling strategy is given by*

$$\operatorname{Var} \overline{A}^{N_{\text{sh}}} = \frac{1}{N_{\text{sh}}} \left(\langle A^2 \rangle_\psi - \langle A \rangle_\psi^2 \right). \tag{2.40}$$

Proof. Notice that, for any $i,\ j \in \{1, \dots N_A\}$ such that $i \neq j$, $S_i \cap S_j = S_i = S_j$ and that $|S_i| = |S_j| = N_{\text{sh}}$. Hence, using (2.34), we obtain

$$\begin{aligned}
\operatorname{Var} \overline{A}^{N_{\text{sh}}} &= \sum_{i=1}^{N_A} \sum_{j=1}^{N_A} \frac{\alpha_i \alpha_j}{N_{\text{sh}}} \left(\langle P_i P_j \rangle_\psi - \langle P_i \rangle_\psi \langle P_j \rangle_\psi \right) \\
&= \frac{1}{N_{\text{sh}}} \left[\left\langle \left(\sum_{i=1}^{N_A} \alpha_i P_i \right)^2 \right\rangle_\psi - \left(\sum_{i=1}^{N_A} \alpha_i \langle P_i \rangle_\psi \right)^2 \right] = \frac{1}{N_{\text{sh}}} \left[\langle (A - \alpha_0 I)^2 \rangle_\psi - \langle A - \alpha_0 I \rangle_\psi^2 \right],
\end{aligned}$$

which allows to recover (2.40). □

In the next three subsections, we apply Proposition 10 successively to the kinetic, potential and interaction operators defined in (2.13)–(2.14).

2.4.3 Kinetic energy

Let us consider the Pauli sampling estimator of $\mathcal{K}^N$

$$\overline{\mathcal{K}^N}^{N_{\text{sh}}} = \alpha_0 + \sum_{i=1}^{N_\mathcal{K}} \alpha_i \overline{P_i}^{N_{\text{sh}}^i}, \tag{2.41}$$

using the Pauli decomposition of $\mathcal{K}^N$ from Proposition 3. We consider for this energy term a naive importance-sampling strategy as the Pauli terms do not commute trivially. That allows to obtain the following result on the estimator variance.

Theorem 5. *The sampling variance is upper bounded by*

$$\operatorname{Var} \overline{\mathcal{K}^N}^{N_{\text{sh}}} = \|\alpha_{1:}\|_1 \sum_{i=1}^{N_\mathcal{K}} \frac{|\alpha_i| \left(1 - \langle P_i \rangle_\psi^2\right)}{N_{\text{sh}}} \leq \frac{\|\alpha_{1:}\|_1^2}{N_{\text{sh}}} = \frac{N^2}{4 h_N^4 N_{\text{sh}}} = \frac{N^2 16^N}{4 N_{\text{sh}}}. \tag{2.42}$$

Proof. As an importance-sampling strategy is used, from (2.39) we have,

$$\operatorname{Var} \overline{\mathcal{K}^N}^{N_{\mathrm{sh}}} = \|\alpha_{1:}\|_1 \sum_{i=1}^{N_{\mathcal{K}}} \frac{|\alpha_i| \left(1 - \langle P_i \rangle_\psi^2\right)}{N_{\mathrm{sh}}}.$$

We recover the inequality in (2.42) by using the lower bound, for all $1 \leq i \leq N_{\mathcal{K}}$, $\langle P_i \rangle_\psi^2 \geq 0$ and the value of $\|\alpha_{1:}\|_1$ from (2.20). □

Remark 9. *In fact, some terms do commute in the Pauli decomposition of $\mathcal{K}^N$, and thus a more relevant sampling scheme may provide a better variance than our naive importance sampling method. For example, the Gray-encoding basis change as in [19] would allow the sampling of commuting terms at the cost of a CNOT wall.*

2.4.4 Potential energy

We define the Walsh-Pauli sampling estimator of $\langle \mathcal{P}^N \rangle_\psi$,

$$\overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} = \alpha_0 + \sum_{i=1}^{N_{\mathcal{P}}} \alpha_i \overline{P_i}^{N_{\mathrm{sh}}^i}, \tag{2.43}$$

where we have, for any $1 \leq i \leq N_{\mathcal{P}}$, α_i and P_i from (2.27).

As the operator $\mathcal{P}$ is diagonal, we consider a diagonal-sampling strategy with $U_{\mathcal{P}} = I^N$ for which we obtain the following result.

Theorem 6. *The variance of the Walsh-Pauli sampling estimator for the potential term satisfies, for any $N \geq 1$,*

$$\operatorname{Var} \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} = \frac{\langle (\Theta_N V)^2, v^2 \rangle_{L^2_N} - (\langle \Theta_N V, v^2 \rangle_{L^2_N})^2}{N_{\mathrm{sh}}} \leq \frac{\|\Theta_N V\|_{L^2}^2 - \|V\|_{L^1}^2}{N_{\mathrm{sh}}}. \tag{2.44}$$

Proof. Indeed, we have, from (2.40),

$$\begin{aligned} N_{\mathrm{sh}} \operatorname{Var} \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} &= \sum_{i=1}^{N_{\mathcal{P}}} \sum_{j=1}^{N_{\mathcal{P}}} \left[\alpha_i \alpha_j \langle P_i P_j \rangle_\psi - \alpha_i \alpha_j \langle P_i \rangle_\psi \langle P_j \rangle_\psi\right] \\ &= \sum_{i=1}^{N_{\mathcal{P}}} \sum_{j=1}^{N_{\mathcal{P}}} \left[\alpha_i \alpha_j \langle P_i P_j \rangle_\psi \langle I \rangle_\psi - \alpha_i \alpha_j \langle P_i \rangle_\psi \langle P_j \rangle_\psi\right]. \end{aligned} \tag{2.45}$$

For any j, k such that $0 \leq j, k < 2^N$, we set $p_k = |\psi_k|^2$ and $P_j^k = \langle k | P_j | k \rangle$, the k-th term on the diagonal of P_j. As the involved Pauli terms are diagonal, we can further rewrite the mean value

of each term $\langle P_i P_j\rangle_\psi = \sum_{k=0}^{2^N-1} P_i^k P_j^k p_k$ and similarly for $\langle P_i\rangle_\psi$ and $\langle I\rangle_\psi$. Then (2.45) becomes

$$\begin{aligned} N_{\mathrm{sh}} \operatorname{Var} \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} &= \sum_{i=1}^{N_{\mathcal{P}}}\sum_{j=1}^{N_{\mathcal{P}}} \alpha_i\alpha_j \left(\sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} P_i^k P_j^k p_k p_\ell - \sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} P_i^k P_j^\ell p_k p_\ell \right) \\ &= \sum_{i=1}^{N_{\mathcal{P}}}\sum_{j=1}^{N_{\mathcal{P}}} \alpha_i\alpha_j \left(\sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} P_i^k p_k p_\ell \left(P_j^k - P_j^\ell\right) \right) \\ &= \sum_{i=1}^{N_{\mathcal{P}}}\sum_{j=1}^{N_{\mathcal{P}}} \alpha_i\alpha_j \left(\sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} P_i^\ell p_\ell p_k \left(P_j^\ell - P_j^k\right) \right). \end{aligned}$$

Noticing that the previous sum is antisymmetric with respect to ℓ and k, we can rewrite it as

$$\begin{aligned} N_{\mathrm{sh}} \operatorname{Var} \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} &= \frac{1}{2}\sum_{i=1}^{N_{\mathcal{P}}}\sum_{j=1}^{N_{\mathcal{P}}} \alpha_i\alpha_j \sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} p_\ell p_k \left(P_i^k - P_i^\ell\right)\left(P_j^k - P_j^\ell\right) \\ &= \frac{1}{2}\sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} p_\ell p_k \sum_{i=1}^{N_{\mathcal{P}}} \alpha_i \left(P_i^k - P_i^\ell\right) \sum_{j=1}^{N_{\mathcal{P}}} \alpha_j \left(P_j^k - P_j^\ell\right). \end{aligned} \tag{2.46}$$

The Pauli decomposition of the potential term $\mathcal{P}_N$ gives

$$\frac{1}{2}\sum_{i=1}^{N_{\mathcal{P}}} \alpha_i P_i^k = \Theta_N V(k/2^N) - \alpha_0 = \Theta_N V(k/2^N) - \|V\|_{L^1}.$$

Consequently, (2.46) becomes

$$\begin{aligned} N_{\mathrm{sh}} \operatorname{Var} \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} &= \frac{1}{2}\sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} p_\ell p_k \left[\left(\Theta_N V(k/2^N) - \|V\|_{L^1}\right) - \left(\Theta_N V(\ell/2^N) - \|V\|_{L^1}\right)\right]^2 \\ &= \frac{1}{2}\sum_{k=0}^{2^N-1}\sum_{\ell=0}^{2^N-1} p_\ell p_k \left[\left(\Theta_N V(k/2^N)\right) - \left(\Theta_N V(\ell/2^N)\right)\right]^2. \end{aligned} \tag{2.47}$$

The last term in (2.47) is the left-Riemann sum associated to the double integral

$$\int_0^1 \left[\Theta_N V(x) - \Theta_N V(y)\right]^2 v(x)^2 v(y)^2 \mathrm{d}x\mathrm{d}y,$$

where the sum on ℓ is related to the integration on x and the one on k to the integration on y. Hence, we rewrite (2.47) as

$$\begin{aligned} N_{\mathrm{sh}} \operatorname{Var} \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} &= \frac{1}{2}\langle \left(\Theta_N V(x) - \Theta_N V(y)\right)^2, v(x)^2 v(y)^2 \rangle_{L^2_N} \\ &\leq \frac{1}{2}\left(2\|\Theta_N V\|^2_{L^2_N} - 2\|\Theta_N V\|^2_{L^1_N}\right)\|v\|^2_{L^2_N} = \|\Theta_N V\|^2_{L^2} - \|V\|^2_{L^1}, \end{aligned}$$

which allows to conclude. □

2.4.5 Interaction term

Recall from (2.28) that the interaction term operator is given by

$$\langle \mathcal{I}^N \rangle_\psi = \frac{\kappa}{2} \sum_{k=0}^{N_\mathcal{I}} \langle \psi | \, P_k \, | \psi \rangle^2 ,$$

where, for any $0 \leq k \leq N_\mathcal{I}$, $P_k = Z_1^{k_1} \dots Z_N^{k_N}$. In order to construct an estimator for the interaction term based on its 'squared Pauli' decomposition, we first need to estimate, for $0 \leq k < 2^N$, the terms $\langle \psi | \, P_k \, | \psi \rangle^2$. We are thus led to define the unbiased empirical mean estimator of $\langle P_i \rangle_\psi^2$, given N_{sh} samples, for any $1 \leq i \leq N_\mathcal{I}$,

$$\overline{P_i^2}^{N_{\text{sh}}} = \frac{1}{N_{\text{sh}}^i \left(N_{\text{sh}}^i - 1 \right)} \sum_{k=1}^{N_{\text{sh}}^i} \sum_{\substack{k'=1 \\ k' \neq k}}^{N_{\text{sh}}^i} \mathbf{p}_i^k \mathbf{p}_i^{k'},$$

where, for $1 \leq k \leq N_{\text{sh}}^i$, $\mathbf{p}_i^k \in \{-1, 1\}$ is the reconstructed value of the P_i observable for the k-th sample in S_i. As the Pauli operators in the decomposition are all diagonal, we consider a diagonal-sampling strategy. Let us now look at the covariance between two of these estimators.

Proposition 13. *Let $1 \leq i, j \leq N_\mathcal{I}$. We have*

$$\begin{aligned} \operatorname{Cov}\left(\overline{P_i^2}^{N_{\text{sh}}}, \overline{P_j^2}^{N_{\text{sh}}} \right) = {} & \frac{6 - 4N_{\text{sh}}}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi^2 \langle P_j \rangle_\psi^2 \\ & + \frac{4(N_{\text{sh}} - 2)}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi \langle P_j \rangle_\psi \langle P_i P_j \rangle_\psi + \frac{2}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i P_j \rangle_\psi^2, \end{aligned} \tag{2.48}$$

and

$$\operatorname{Var} \overline{P_i^2}^{N_{\text{sh}}} = \frac{6 - 4N_{\text{sh}}}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi^4 + \frac{4(N_{\text{sh}} - 2)}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi^2 + \frac{2}{N_{\text{sh}}(N_{\text{sh}} - 1)}. \tag{2.49}$$

Proof. Let $1 \leq i, j \leq N_\mathcal{I}$. We first write

$$\operatorname{Cov}\left(\overline{P_i^2}^{N_{\text{sh}}}, \overline{P_j^2}^{N_{\text{sh}}} \right) = \mathbb{E}\left[\overline{P_i^2}^{N_{\text{sh}}} \overline{P_j^2}^{N_{\text{sh}}} \right] - \mathbb{E}\left[\overline{P_i^2}^{N_{\text{sh}}} \right] \mathbb{E}\left[\overline{P_j^2}^{N_{\text{sh}}} \right]. \tag{2.50}$$

Let us look more closely at the first term $\mathbb{E}\left[\overline{P_i^2}^{N_{\text{sh}}} \overline{P_j^2}^{N_{\text{sh}}} \right]$ over the individual samples

$$\mathbb{E}\left[\overline{P_i^2}^{N_{\text{sh}}} \overline{P_j^2}^{N_{\text{sh}}} \right] = \frac{1}{\left(N_{\text{sh}}(N_{\text{sh}} - 1) \right)^2} \sum_{k=1}^{N_{\text{sh}}} \sum_{\substack{k'=1 \\ k' \neq k}}^{N_{\text{sh}}} \sum_{\ell=1}^{N_{\text{sh}}} \sum_{\substack{\ell'=1 \\ \ell' \neq \ell}}^{N_{\text{sh}}} \mathbb{E}\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'} \right]. \tag{2.51}$$

The value of $\mathbb{E}\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'} \right]$ depends on the sample associated to each Pauli term, i.e. the indices k, ℓ, k', ℓ'. Indeed, if the samples are shared, the terms depend on each other and the expectation value of the product of the terms is not the product of expectation values. For each case, we now detail the different possible values of $\mathbb{E}\left[p_i^k p_i^{k'} p_j^\ell p_j^{\ell'} \right]$ depending on the indices.

- If both k and k' do not equal ℓ or ℓ', *i.e.* for $N_{\text{sh}}(N_{\text{sh}} - 1)(N_{\text{sh}} - 2)(N_{\text{sh}} - 3)$ terms of the sum in (2.51),

$$\mathbb{E}\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'} \right] = \langle P_i \rangle_\psi^2 \langle P_j \rangle_\psi^2.$$

- If ($k \in \{1, \dots, N_{\rm sh}\} \setminus \{\ell, \ell'\}$ and $k' \in \{\ell, \ell'\}$), or ($k \in \{\ell, \ell'\}$ and $k' \in \{1, \dots, N_{\rm sh}\} \setminus \{\ell, \ell'\}$), *i.e.* for $4N_{\rm sh}(N_{\rm sh}-1)(N_{\rm sh}-2)$ terms of the sum in (2.51),

$$\mathbb{E}\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right] = \langle P_i \rangle_\psi \langle P_j \rangle_\psi \langle P_i P_j \rangle_\psi.$$

- If ($k = \ell$ and $k' = \ell'$), or ($k = \ell'$ and $k' = \ell$), *i.e.* for $2N_{\rm sh}(N_{\rm sh}-1)$ terms of the sum in (2.51),

$$\mathbb{E}\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right] = \langle P_i P_j \rangle_\psi^2.$$

Consequently, the second term in (2.50) becomes

$$\mathbb{E}\left[\overline{P_i^2}^{N_{\rm sh}}\right] \mathbb{E}\left[\overline{P_j^2}^{N_{\rm sh}}\right] = \frac{1}{\left(N_{\rm sh}(N_{\rm sh}-1)\right)^2} \sum_{k=1}^{N_{\rm sh}} \sum_{\substack{k'=1 \\ k' \neq k}}^{N_{\rm sh}} \sum_{\ell=1}^{N_{\rm sh}} \sum_{\substack{\ell'=1 \\ \ell' \neq \ell}}^{N_{\rm sh}} \mathbb{E}\left[\mathbf{p}_i^k \mathbf{p}_i^{k'}\right] \mathbb{E}\left[\mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right] = \langle P_i \rangle_\psi^2 \langle P_j \rangle_\psi^2,$$

so that we obtain (2.48). We similarly get (2.49), using the additional fact that, for any $1 \leq i \leq N_{\mathcal{I}}$, $\langle P_i^2 \rangle_\psi = 1$. □

Using the squared Pauli decomposition (2.28) of $\mathcal{I}^N$ and these new estimators, with $N_{\rm sh}$ samples, we define the Pauli sampling estimator of $\langle \mathcal{I}^N \rangle_\psi$,

$$\overline{\mathcal{I}^N}^{N_{\rm sh}} = \alpha_0 + \sum_{i=1}^{N_{\mathcal{I}}} \alpha_0 \overline{P_i^2}^{N_{\rm sh}^i}, \tag{2.52}$$

with $\alpha_0 = \frac{\kappa}{2}$. We note that, as we estimate the mean of the squared Pauli operators, part of the variance will decrease quadratically with $N_{\rm sh}$ as we now explain. Indeed, as can be observed in the following theorem, the variance of the estimator is equal to a polynomial of order one divided by a polynomial of order two, and thus the remainder term will decrease quadratically in $N_{\rm sh}$.

Theorem 7. *The Pauli sampling estimator of the interaction term has a variance given by*

$$\operatorname{Var} \overline{\mathcal{I}^N}^{N_{\rm sh}} = \frac{\alpha_0^2}{N_{\rm sh}(N_{\rm sh}-1)} \Big[(6 - 4N_{\rm sh}) \langle v^2, v^2 \rangle_{L_N^2}^2 + 4(N_{\rm sh}-2) \langle v^4, v^2 \rangle_{L_N^2} + 2 \cdot 2^N \langle v^2, v^2 \rangle_{L_N^2} \Big]$$

The dependence of the interaction term on $N_{\rm sh}(N_{\rm sh}-1)$ comes from the fact that we estimate the squared value of the Pauli operator.

Proof. We first write

$$\operatorname{Var} \overline{\mathcal{I}^N}^{N_{\rm sh}} = \mathbb{E}\left[(\overline{\mathcal{I}^N}^{N_{\rm sh}})^2\right] - \mathbb{E}\left[\overline{\mathcal{I}^N}^{N_{\rm sh}}\right]^2.$$

Thanks to (2.52), the previous equality becomes

$$\operatorname{Var} \overline{\mathcal{I}^N}^{N_{\rm sh}} = \alpha_0^2 \sum_{i=1}^{N_{\mathcal{I}}} \sum_{j=1}^{N_{\mathcal{I}}} \left(\mathbb{E}\left[\overline{P_i^2}^{N_{\rm sh}} \overline{P_j^2}^{N_{\rm sh}}\right] - \alpha_0^2 \langle P_i \rangle_\psi^2 \langle P_j \rangle_\psi^2 \right).$$

Using (2.49) and (2.48), we obtain

$$\operatorname{Var}\overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} = \frac{\alpha_0^2}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)}\Bigg[(6-4N_{\mathrm{sh}})\sum_{i=1}^{N_{\mathcal{I}}}\sum_{j=1}^{N_{\mathcal{I}}}\langle P_i\rangle_\psi^2\langle P_j\rangle_\psi^2 + 4(N_{\mathrm{sh}}-2)\sum_{i=1}^{N_{\mathcal{I}}}\langle P_i\rangle_\psi^2 + 4(N_{\mathrm{sh}}-2)\sum_{i=1}^{N_{\mathcal{I}}}\sum_{\substack{j=1\\ j\neq i}}^{N_{\mathcal{I}}}\langle P_i\rangle_\psi\langle P_j\rangle_\psi\langle P_iP_j\rangle_\psi + 2\sum_{i=1}^{N_{\mathcal{I}}}\sum_{\substack{j=1\\ j\neq i}}^{N_{\mathcal{I}}}\langle P_iP_j\rangle_\psi^2 + 2N_{\mathcal{I}}\Bigg]. \quad (2.53)$$

Using (2.28), we simplify the above sums into L^2_N scalar products of v^2. First, we have

$$\alpha_0\sum_{i=1}^{N_{\mathcal{I}}}\langle P_i\rangle_\psi^2 = \alpha_0\left(\langle v^2,v^2\rangle_{L^2_N} - \langle v^2,1\rangle^2_{L^2_N}\right) = \langle\mathcal{I}^N\rangle_\psi - \alpha_0.$$

This allows to obtain, for the sum of terms $\langle P_i\rangle_\psi^2\langle P_j\rangle_\psi^2$,

$$\alpha_0^2\sum_{i=1}^{N_{\mathcal{I}}}\sum_{j=1}^{N_{\mathcal{I}}}\langle P_i\rangle_\psi^2\langle P_j\rangle_\psi^2 = \alpha_0^2\left(\langle v^2,v^2\rangle_{L^2_N} - \langle v^2,1\rangle^2_{L^2_N}\right)^2 \quad (2.54)$$

Furthermore, for terms $\langle P_iP_j\rangle_\psi^2$, as all the Z-Pauli operators are contained in the Pauli decomposition of $\mathcal{I}$, we can write

$$\begin{aligned}\alpha_0\sum_{i=1}^{N_{\mathcal{I}}}\sum_{\substack{j=1\\ j\neq i}}^{N_{\mathcal{I}}}\langle P_iP_j\rangle_\psi^2 &= \alpha_0\sum_{i=1}^{N_{\mathcal{I}}}\left(\sum_{k=1}^{N_{\mathcal{I}}}\langle P_k\rangle_\psi^2 - \langle P_i\rangle_\psi^2\right)\\ &= \alpha_0 N_{\mathcal{I}}\sum_{k=1}^{N_{\mathcal{I}}}\langle P_k\rangle_\psi^2 - \alpha_0\sum_{i=1}^{N_{\mathcal{I}}}\langle P_i\rangle_\psi^2 = (2^N-2)\alpha_0\left(\langle v^2,v^2\rangle_{L^2_N} - \langle v^2,1\rangle^2_{L^2_N}\right). \quad (2.55)\end{aligned}$$

For terms $\langle P_i\rangle_\psi\langle P_j\rangle_\psi\langle P_iP_j\rangle_\psi$, we are led to consider the Walsh interpolant of v^2

$$\mathrm{I}_N^{\mathcal{W}}(v^2) = \sum_{k=0}^{2^N-1}\langle v^2,w_k\rangle_{L^2_N}w_k.$$

We now use Theorem 16, from Appendix A.2, ensures that

$$\mathrm{I}_N^{\mathcal{W}}(v^4) = \left[\mathrm{I}_N^{\mathcal{W}}(v^2)\right]^2.$$

Using this, we obtain

$$\begin{aligned}\mathrm{I}_N^{\mathcal{W}}\left(v^2-\|v\|^2_{L^2}\right)^2 &= \sum_{i=1}^{2^N-1}\sum_{j=1}^{2^N-1}\langle v^2,w_i\rangle_{L^2_N}\langle v^2,w_j\rangle_{L^2_N}w_{i\oplus j}\\ &= \sum_{i=1}^{2^N-1}\sum_{\substack{k=0\\ k\neq i}}^{2^N-1}\langle v^2,w_i\rangle_{L^2_N}\langle v^2,w_{i\oplus k}\rangle_{L^2_N}w_k.\end{aligned}$$

Finally, we notice that

$$\left\langle \left(v^2 - \|v\|_{L^2}^2\right)^2, v^2 - \|v\|_{L^2}^2 \right\rangle_{L_N^2} = \sum_{j=1}^{2^N-1} \sum_{i=1}^{2^N-1} \sum_{\substack{k=0 \\ k\neq i}}^{2^N-1} \langle v^2, w_i\rangle_{L_N^2} \langle v^2, w_j\rangle_{L_N^2} \langle v^2, w_{i\oplus k}\rangle_{L_N^2} \langle w_k, w_j\rangle_{L_N^2}$$

$$= \sum_{i=1}^{2^N-1} \sum_{\substack{j=1 \\ j\neq i}}^{2^N-1} \langle v^2, w_i\rangle_{L_N^2} \langle v^2, w_j\rangle_{L_N^2} \langle v^2, w_{i\oplus j}\rangle_{L_N^2} = \sum_{i=1}^{2^N-1} \sum_{\substack{j=1 \\ j\neq i}}^{2^N-1} \langle P_i\rangle_\psi \langle P_j\rangle_\psi \langle P_i P_j\rangle_\psi,$$

which gives

$$\sum_{i=1}^{2^N-1} \sum_{\substack{j=1 \\ j\neq i}}^{2^N-1} \langle P_i\rangle_\psi \langle P_j\rangle_\psi \langle P_i P_j\rangle_\psi = \left\langle \left(v^4 - 2v^2\|v\|_{L^2}^2 + \left(\|v\|_{L^2}^2\right)^2\right), \left(v^2 - \|v\|_{L^2}^2\right) \right\rangle_{L_N^2}$$

$$= \langle v^4, v^2\rangle_{L_N^2} - 3\langle v^2, v^2\rangle_{L_N^2} + 2. \quad (2.56)$$

Replacing (2.54), (2.55) and (2.56) in (2.53) provides the result. □

Since the term $\langle v^4, v^2\rangle_{L_N^2}$ is not uniquely determined by the interaction energy, we consider the following bound, obtained thanks to the Cauchy-Schwarz inequality.

Theorem 8. *Set*

$$B_V = 4\alpha_0^2 \langle v^2, v^2\rangle_{L_N^2}^{3/2} \left(2^{N/2} - \langle v^2, v^2\rangle_{L_N^2}^{1/2}\right), \qquad B_W = 2\alpha_0^2 \langle v^2, v^2\rangle_{L_N^2} \left(2^{N/2} - \langle v^2, v^2\rangle_{L_N^2}^{1/2}\right)^2, \quad (2.57)$$

which are nonnegative. Then the variance of the interaction Pauli sampling estimator is upper-bounded by

$$\operatorname{Var} \overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} \leq \frac{B_V}{N_{\mathrm{sh}}} + \frac{B_W}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)}.$$

Proof. Consider again the Walsh interpolant of v^4, which, using Theorem 16, can be written as

$$\mathrm{I}_N^{\mathcal{W}}(v^4) = \sum_{i=0}^{2^N-1} \sum_{k=0}^{2^N-1} \langle v^2, w_i\rangle_{L_N^2} \langle v^2, w_{i\oplus k}\rangle_{L_N^2} w_k.$$

This allows to write, by definition of the interpolant,

$$\langle v^4, v^2\rangle_{L_N^2} = \langle \mathrm{I}_N^{\mathcal{W}}(v^4), \mathrm{I}_N^{\mathcal{W}}(v^2)\rangle_{L_N^2}$$

and we thus obtain

$$\langle v^4, v^2\rangle_{L_N^2} = \sum_{i=0}^{2^N-1} \sum_{j=0}^{2^N-1} \langle v^2, w_i\rangle_{L_N^2} \langle v^2, w_{i\oplus j}\rangle_{L_N^2} \langle v^2, w_j\rangle_{L_N^2}.$$

Since we have

$$2\langle v^2, w_i\rangle_{L_N^2}\langle v^2, w_{i\oplus j}\rangle_{L_N^2}\langle v^2, w_j\rangle_{L_N^2} \leq |\langle v^2, w_i\rangle_{L_N^2}|\left(|\langle v^2, w_{i\oplus j}\rangle_{L_N^2}|^2 + |\langle v^2, w_j\rangle_{L_N^2}|^2\right),$$

we get

$$\begin{aligned}\langle v^4, v^2\rangle_{L_N^2} \leq \frac{1}{2}\sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} |\langle v^2, w_i\rangle_{L_N^2}|(|\langle v^2, w_{i\oplus j}\rangle_{L_N^2}|^2 + |\langle v^2, w_j\rangle_{L_N^2}|^2) \\ = \sum_{i=0}^{2^N-1} |\langle v^2, w_i\rangle_{L_N^2}| \sum_{j=0}^{2^N-1} \langle v^2, w_j\rangle_{L_N^2}^2 = |\beta|_1|\beta|_2^2\end{aligned}$$

where we define the 2^N-dimensional vector of the Walsh coefficients of v^2, $\beta = \left(\langle v^2, w_i\rangle_{L_N^2}\right)_{0\leq i<2^N}$. From the Cauchy-Schwarz inequality, and since β is 2^N-dimensional, we can write

$$|\beta|_1 \leq 2^{N/2}|\beta|_2 = 2^{N/2}\langle v^2, v^2\rangle_{L_N^2}^{1/2}.$$

Thus we have

$$\langle v^4, v^2\rangle_{L_N^2} \leq 2^{N/2}\langle v^2, v^2\rangle_{L_N^2}^{3/2} \tag{2.58}$$

Plugging (2.58) in (2.53), we obtain

$$\begin{aligned}\operatorname{Var}\overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} &\leq \alpha_0^2\frac{-4(N_{\mathrm{sh}}-6)\langle v^2, v^2\rangle_{L_N^2}^2 + \left(4(N_{\mathrm{sh}}-2)2^{N/2}\langle v^2, v^2\rangle_{L_N^2}^{3/2} + 2\cdot 2^N\langle v^2, v^2\rangle_{L_N^2}\right)}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)} \\ &= 4\alpha_0^2\langle v^2, v^2\rangle_{L_N^2}^{3/2}\frac{2^{N/2}-\langle v^2, v^2\rangle_{L_N^2}^{1/2}}{N_{\mathrm{sh}}} + \alpha_0^2\langle v^2, v^2\rangle_{L_N^2}\frac{2\left(\langle v^2, v^2\rangle_{L_N^2}^{1/2} - 2^{N/2}\right)^2}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)},\end{aligned}$$

which gives us the result.

$$\operatorname{Var}\overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} \leq \frac{4\alpha_0^2}{N_{\mathrm{sh}}}\langle v^2, v^2\rangle_{L_N^2}^{3/2}\left(2^{N/2} - \langle v^2, v^2\rangle_{L_N^2}^{1/2}\right) + \frac{2\alpha_0^2}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)}\langle v^2, v^2\rangle_{L_N^2}\left(2^{N/2} - \langle v^2, v^2\rangle_{L_N^2}^{1/2}\right)^2$$

□

2.5 Estimators

We now compare our method with the two types of estimators defined in [68], the direct-sampling estimator and the Hadamard-test estimator.

2.5.1 Direct-sampling estimators

In this subsection, we consider the direct-sampling estimators which are used in the 'hardware-reduced' version of the algorithm developed in [68]. Given N_{sh} samples, we denote the direct-

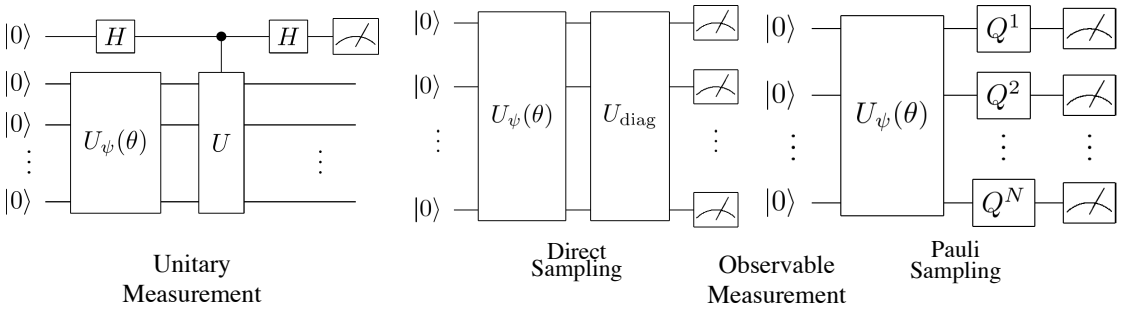


Figure 2.6 – Operator measurement methods

sampling estimator of an operator A by

$$\check{A}^{N_{\text{sh}}} = \frac{1}{N_{\text{sh}}} \sum_{k=0}^{N_{\text{sh}}} D_{s_k}, \tag{2.59}$$

where $s_k \in \{0, \dots, 2^N - 1\}$ corresponds to the k-th sample in the diagonal basis of A, D is the diagonal matrix of A in the basis, and for any j, $0 \leq j < 2^N$, we denote by D_j the j-th term of the diagonal of D for simplicity. By definition, the direct-sampling estimator is in fact exactly the diagonal-sampling estimator. This implies that the method requires knowing the diagonal basis of the operator and implementing a basis change at the end of the circuit.

In our case, the direct-sampling estimator is used for operators with diagonals corresponding to function evaluations. This allows to obtain explicit formulations of its mean and variance.

Theorem 9. *Given a diagonal operator D such that, for any i, $0 \leq i < 2^N$, $D_i = d(i/2^N)$ for a function $d \in H^1_{\#}(0,1)$, the mean of the sampling estimator for D is*

$$\langle D \rangle_\psi = \langle v^2, d \rangle_{L^2_N},$$

and the variance of the estimator is given by

$$\operatorname{Var} \check{D}^{N_{\text{sh}}} = \frac{\langle d^2, v^2 \rangle_{L^2_N} - \langle d, v^2 \rangle^2_{L^2_N}}{N_{\text{sh}}}. \tag{2.60}$$

The variance is bounded by the difference of the discrete L^2 and L^1 norms of d

$$\operatorname{Var} \check{D}^{N_{\text{sh}}} \leq \frac{1}{N_{\text{sh}}} \left(\|d\|^2_{L^2_N} - \|d\|^2_{L^1_N} \right), \tag{2.61}$$

where the discrete L^1 norm is defined, for any $v \in H^1$, by $\|v\|_{L^1_N} = h_N \sum_{k=0}^{2^N-1} |v(\mathrm{x}_k)|$.

Proof. Given a positive integer $k < N_{\text{sh}}$, we can write

$$D_{s_k} = \sum_{i=0}^{2^N-1} D_i \delta_{s_k, i}, \tag{2.62}$$

where δ denotes the Kronecker delta. The mean value of the estimator is then given by

$$\mathbb{E}\left[\check{D}^{N_{\mathrm{sh}}}\right] = \frac{1}{N_{\mathrm{sh}}} \sum_{k=1}^{N_{\mathrm{sh}}} \mathbb{E}\left[D_{s_k}\right].$$

For each sample s_k, we thus have

$$\mathbb{E}\left[D_{s_k}\right] = \sum_{i=0}^{2^N-1} D_i p_i = \sum_{i=0}^{2^N-1} d(i/2^N) v^2(i/2^N) h_N = \langle v^2, d\rangle_{L^2_N},$$

where we denote $p_i = v(i/2^N)^2 h_N$, the probability of a sample being equal to i. Thus

$$\mathbb{E}\left[\check{D}^{N_{\mathrm{sh}}}\right] = \frac{N_{\mathrm{sh}}}{N_{\mathrm{sh}}} \langle v^2, d\rangle_{L^2_N} = \langle D\rangle_\psi.$$

The variance on the other hand is given by

$$\begin{aligned}\mathrm{Var}\left(\check{D}^{N_{\mathrm{sh}}}\right) &= \mathbb{E}\left[(\check{D}^{N_{\mathrm{sh}}})^2\right] - \mathbb{E}\left[\check{D}^{N_{\mathrm{sh}}}\right]^2 \\ &= \frac{1}{N_{\mathrm{sh}}^2} \sum_{k=1}^{N_{\mathrm{sh}}} \sum_{\ell=1}^{N_{\mathrm{sh}}} \mathbb{E}\left[D_{s_k} D_{s_\ell}\right] - \mathbb{E}\left[D_{s_k}\right]\mathbb{E}\left[D_{s_\ell}\right] = \frac{1}{N_{\mathrm{sh}}^2} \sum_{k=1}^{N_{\mathrm{sh}}} \sum_{\ell=1}^{N_{\mathrm{sh}}} \mathrm{Cov}(D_{s_k}, D_{s_\ell}).\end{aligned}$$

As the samples are independent, we only have the resulting variance terms for each D_{s_k}. Using the decomposition (2.62) for D_{s_k}, we write

$$\mathrm{Var}\, D_{s_k} = \sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} D_i D_j \mathbb{E}\left[\delta_{s_k i}\delta_{s_k,j}\right] - \sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} D_i D_j p_i p_j = \sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} D_i D_j \left(\delta_{ij} p_i - p_i p_j\right).$$

That ensures

$$\mathrm{Var}\, \check{D}^{N_{\mathrm{sh}}} = \frac{1}{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} d_i d_j \delta_{i,j} p_i - p_i p_j,$$

that is (2.60). To obtain the bound (2.61), we write

$$\begin{aligned}\mathrm{Var}\, \check{D}^{N_{\mathrm{sh}}} &= \frac{1}{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1} d_i^2 p_i - \frac{1}{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} d_i d_j p_i p_j = \frac{1}{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1} d_i p_i \left(d_i - \sum_{j=0}^{2^N-1} d_j p_j\right) \\ &= \frac{1}{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1} d_i p_i \left(d_i \sum_{j=0}^{2^N-1} p_j - \sum_{j=0}^{2^N-1} d_j p_j\right) = \frac{1}{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} d_i p_i p_j \left(d_i - d_j\right).\end{aligned}$$

Then, we notice that

$$\sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} d_i p_i p_j \left(d_i - d_j\right) = -\sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} d_i p_i p_j \left(d_j - d_i\right) = -\sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} d_j p_i p_j \left(d_i - d_j\right).$$

This anti-symmetry ensures that

$$\mathrm{Var}\left(\check{D}^{N_{\mathrm{sh}}}\right) = \frac{1}{2N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1} \sum_{j=0}^{2^N-1} \left[d_i p_i p_j \left(d_i - d_j\right) - d_j p_i p_j \left(d_i - d_j\right)\right]$$
$$= \frac{1}{2N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1} \sum_{j=0}^{2^N-1} p_i p_j \left(d_i - d_j\right)^2 .$$

This previous sum can be identified to the discrete L^2 product $\langle v^2(x)v^2(y), (d(x)-d(y))^2\rangle_{L^2_N}$, which allows to obtain

$$\mathrm{Var}\left(\check{D}^{N_{\mathrm{sh}}}\right) = \frac{1}{2N_{\mathrm{sh}}} \langle v^2(x)v^2(y), (d(x)-d(y))^2\rangle_{L^2_N}$$
$$\leq \frac{1}{2N_{\mathrm{sh}}} \left(\|v\|^2_{L^2_N}\right)^2 \|d(x)-d(y)\|^2_{L^2_N},$$

and the required bound (2.61). □

Remark 10. *It is important to note that, in the case of a diagonal operator, we have an equivalence between the direct-sampling estimator and the Walsh-Pauli estimator if the discretization and truncation approaches provide the same coefficients, i.e. if the truncation and interpolation of the function coincide. This is the case for example in the context of the interaction term. Furthermore, the variance of the estimator is optimal in the sense that it is equal to the variance of the observable divided by the number of samples.*

Unbiased direct-sampling interaction term

In the case of the nonlinear interaction term, the estimator defined in [68] can be rewritten as

$$\mathring{\mathcal{I}}_N^{N_{\mathrm{sh}}} = \frac{\kappa}{2N_{\mathrm{sh}}} \sum_{k=0}^{N_{\mathrm{sh}}} \check{Q}_{s_k}^{N_{\mathrm{sh}}} = \frac{\kappa}{2N_{\mathrm{sh}}^2} \sum_{k=0}^{N_{\mathrm{sh}}} \sum_{\ell=0}^{N_{\mathrm{sh}}} \delta_{s_\ell, s_k},$$

where $\check{Q}_i^{N_{\mathrm{sh}}} = \frac{1}{N_{\mathrm{sh}}} \sum_{\ell=1}^{N_{\mathrm{sh}}} \delta_{s_\ell, i}$ is the unbiased estimator of the probability of measuring the computational state i, *i.e.* $\mathbb{E}\left[\check{Q}_i^{N_{\mathrm{sh}}}\right] = \langle |i\rangle\langle i| \rangle_\psi = |\psi_i|^2 = p_i$. If we now look at the mean value of the above estimator, we have,

$$\frac{1}{\alpha_0}\mathbb{E}\left[\mathring{\mathcal{I}}_N^{N_{\mathrm{sh}}}\right] = \frac{1}{N_{\mathrm{sh}}^2} \sum_{k=0}^{N_{\mathrm{sh}}} \sum_{\ell=0}^{N_{\mathrm{sh}}} \mathbb{E}\left[\delta_{s_\ell, s_k}\right] = \frac{1}{N_{\mathrm{sh}}^2} \sum_{k=0}^{N_{\mathrm{sh}}} \sum_{\ell=0}^{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1} \mathbb{E}\left[\delta_{s_\ell, i}\delta_{s_k, i}\right],$$

Now we separate the cases when $k = \ell$ and $k \neq \ell$,

$$\frac{1}{\alpha_0}\mathbb{E}\left[\mathring{\mathcal{I}}_N^{N_{\mathrm{sh}}}\right] = \frac{1}{N_{\mathrm{sh}}^2} \left(\sum_{k=0}^{N_{\mathrm{sh}}} \sum_{i=0}^{2^N-1} \mathbb{E}\left[\delta_{s_k, i}\right] + \sum_{k=0}^{N_{\mathrm{sh}}} \sum_{\ell \neq k} \sum_{i=0}^{2^N-1} \mathbb{E}\left[\delta_{s_\ell, i}\delta_{s_k, i}\right] \right)$$

As k and ℓ are different samples in the second term, they are independent and thus the previous equality becomes

$$\frac{1}{\alpha_0}\mathbb{E}\left[\mathring{\mathcal{I}_N}^{N_{\rm sh}}\right] = \frac{1}{N_{\rm sh}^2}\left(N_{\rm sh}\sum_{i=0}^{2^N-1} p_i + N_{\rm sh}(N_{\rm sh}-1)\sum_{i=0}^{2^N-1} p_i^2\right) = \langle \mathcal{I}_N\rangle_\psi + \frac{1}{N_{\rm sh}}\left(1-\langle \mathcal{I}_N\rangle_\psi\right).$$

This shows that the estimator is thus biased, we are thus led to introduce an unbiased estimator for the interaction term

$$\check{\mathcal{I}_N}^{N_{\rm sh}} = \frac{\kappa}{2}\sum_{i=0}^{2^N-1}\check{Q}_{i,2}^{N_{\rm sh}},$$

where the unbiased estimator of p_i^2 is denoted by

$$\check{Q}_{i,2}^{N_{\rm sh}} = \frac{1}{N_{\rm sh}(N_{\rm sh}-1)}\sum_{k=0}^{N_{\rm sh}}\sum_{\ell=0,\ell\neq k}^{N_{\rm sh}}\left(\delta_{s_k,i}\delta_{s_\ell,i}\right).$$

As mentioned above, the unbiased direct-sampling estimator is equivalent to the unbiased Pauli-sampling estimator, an equivalence which we now demonstrate.

Theorem 10. *The unbiased direct-sampling estimator $\check{\mathcal{I}_N}^{N_{\rm sh}}$ for the interaction term equals the unbiased Pauli sampling estimator $\overline{\mathcal{I}_N}^{N_{\rm sh}}$.*

Proof. For any given sample s_k, if we denote the associated Z-Pauli values for each qubit as $z_\ell^k \in \{-1,1\}$ for any ℓ, $1\leq \ell\leq N$, we have, for any value in the computational basis i, $0\leq i<2^N$, the equalities

$$\delta_{s_k,i} = \prod_{\ell=1}^{N}\frac{1+(-1)^{i_\ell}z_\ell^{i_\ell}}{2} = \frac{1}{2^N}\sum_{n=0}^{2^N-1}(-1)^{i_1n_1+\cdots+i_Nn_N}\mathbf{p}_n^k.$$

Hence we can write

$$\begin{aligned}
\check{\mathcal{I}_N}^{N_{\rm sh}} &= \frac{\kappa}{2^{N+1}N_{\rm sh}(N_{\rm sh}-1)}\sum_{i=0}^{2^N-1}\sum_{k=0}^{N_{\rm sh}}\sum_{\ell=0,\ell\neq k}^{N_{\rm sh}}\left(\delta_{s_k i}\delta_{s_\ell,i}\right)\\
&= \frac{\kappa}{2^{N+1}N_{\rm sh}(N_{\rm sh}-1)}\sum_{i=0}^{2^N-1}\sum_{k=0}^{N_{\rm sh}}\sum_{\ell=0,\ell\neq k}^{N_{\rm sh}}\sum_{n=0}^{2^N-1}\sum_{n'=0}^{2^N-1}(-1)^{\langle i,n\rangle_{\rm bin}}\mathbf{p}_n^k\cdot(-1)^{\langle i,n'\rangle_{\rm bin}}\mathbf{p}_{n'}^\ell\\
&= \frac{\kappa}{2^{N+1}N_{\rm sh}(N_{\rm sh}-1)}\sum_{k=0}^{N_{\rm sh}}\sum_{\ell=0,\ell\neq k}^{N_{\rm sh}}\sum_{n=0}^{2^N-1}\sum_{n'=0}^{2^N-1}\left(\sum_{i=0}^{2^N-1}(-1)^{\langle i,n+n'\rangle_{\rm bin}}\right)\mathbf{p}_n^k\mathbf{p}_{n'}^\ell. \qquad (2.63)
\end{aligned}$$

Note that

$$\sum_{i=0}^{2^N-1}(-1)^{\langle i,n+n'\rangle_{\rm bin}} = \prod_{\ell=1}^{N}\left(1+(-1)^{k_\ell(n_\ell+n'_\ell)}\right) = \begin{cases}0 & \text{if } n\neq n',\\ (1+1)^N = 2^N & \text{else.}\end{cases}$$

Indeed, if $n\neq n'$, then there exists ℓ', $1\leq\ell'\leq N$, such that $k_\ell(n_{\ell'}+n'_{\ell'})=1$ and thus one of the terms of the above product is equal to zero. Hence (2.63) becomes

$$\check{\mathcal{I}_N}^{N_{\rm sh}} = \frac{\kappa}{2N_{\rm sh}(N_{\rm sh}-1)} \sum_{k=0}^{N_{\rm sh}} \sum_{\ell=0,\ell\neq k}^{N_{\rm sh}} \sum_{n=0}^{2^N-1} \mathbf{p}_n^k \mathbf{p}_n^\ell = \overline{\mathcal{I}_N}^{N_{\rm sh}}.$$

This gives the equality between the estimators. □

2.5.2 Hadamard-test estimators

A Hadamard-test consists in determining the mean value of a unitary operator U for a given state $|\psi\rangle$ via the mean value of a 1-bit phase estimation circuit as in Figure 2.6. As shown in Subsection 1.1.2, we obtain from (1.3), for the Hadamard-test circuit associated to the unitary matrix U, that

$$\langle Z_1 \rangle_{\psi,\text{H-test}} = \langle\psi| \operatorname{Re}(U) |\psi\rangle .$$

Let us now define estimators based on this measurement method for arbitrary operators. Given an operator A, an associated quantum state ψ and a Hadamard-test circuit with unitary $U_{\rm H}$ such that

$$\langle A \rangle_\psi = c_{\rm H} + \alpha_{\rm H} \langle Z_1 \rangle_{\psi,\text{H-test}}, \tag{2.64}$$

where $c_{\rm H}$ and $\alpha_{\rm H}$ are constants. We define the Hadamard-test estimator of $\langle A \rangle_\psi$ as

$$\widehat{A}^{N_{\rm sh}} = c_{\rm H} + \alpha_{\rm H} \frac{1}{N_{\rm sh}} \sum_{k=1}^{N_{\rm sh}} \mathbf{h}^k$$

where, for any $1 \leq k \leq N_{\rm sh}$, $\mathbf{h}^k \in \{-1, 1\}$ is the k-th sample of the Hadamard test. By construction, $\langle A \rangle_\psi$ is encoded as the mean value of a one-qubit Pauli operator. This induces a 'Paulification' of the variance, as the variance of the estimator behaves in the same way as the one of a Pauli operator.

Theorem 11. *The variance of the Hadamard-test estimator is given by*

$$\operatorname{Var} \widehat{A}^{N_{\rm sh}} = \frac{\alpha_{\rm H}^2}{N_{\rm sh}} \left(1 - \langle Z_1 \rangle^2_{\psi,\text{H-test}}\right), \tag{2.65}$$

with $\alpha_{\rm H}$ *and* $\langle Z_1 \rangle_{\psi,\text{H-test}}$ *defined in* (2.64).

Proof. By definition,

$$\operatorname{Var} \widehat{A}^{N_{\rm sh}} = \frac{\alpha_{\rm H}^2}{N_{\rm sh}^2} \sum_{k=1}^{N_{\rm sh}} \sum_{\ell=1}^{N_{\rm sh}} \left(\mathbb{E}\left[\mathbf{h}_k \mathbf{h}_\ell\right] - \mathbb{E}\left[\mathbf{h}_k\right] \mathbb{E}\left[\mathbf{h}_\ell\right]\right).$$

Furthermore, we have

$$\mathbb{E}\left[\mathbf{h}_k\right] = \langle 0| U_{\rm H}^\dagger Z_1 U_{\rm H} |0\rangle = \langle A \rangle_\psi$$

and that

$$\mathbb{E}\left[\mathbf{h}_k^2\right] = \langle 0| U_{\rm H}^\dagger Z_1^2 U_{\rm H} |0\rangle = 1,$$

which gives (2.65). □

Let us now look at the specific estimators used for each energy term.

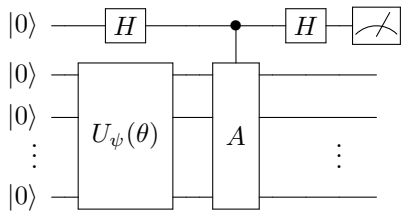


Figure 2.7 – Hadamard-test circuit for the kinetic term as defined in [68].

Kinetic term

In order to describe the Hadamard-test estimator for the kinetic term, we must first define the N-qubit 1-adder A. It is a unitary operator such that, for any i, $0 \leq i < 2^N$,

$$A\,|i\rangle = \left|(i+1) \mod 2^N\right\rangle, \qquad A^\dagger\,|i\rangle = \left|(i-1) \mod 2^N\right\rangle.$$

In matrix form, this gives

$$A = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 & 1 \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & 1 & \ddots & & \vdots \\ \vdots & & \ddots & 1 & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

By noticing that $A + A^\dagger = B_1^N + B_{2^N-1}^N$, the kinetic term, as given in (2.16), can then be written as

$$\mathcal{K}^N = \frac{1}{h_N^2} I^N - \frac{1}{2h_N^2}\left(B_1^N + B_{2^N-1}^N\right) = \frac{1}{h_N^2} I^N - \frac{1}{2h_N^2}\left(A + A^\dagger\right). \tag{2.66}$$

From (2.66), and using the fact that $A + A^\dagger = 2\,\mathrm{Re}\,A$, the kinetic term writes

$$\langle\psi|\,\mathcal{K}^N\,|\psi\rangle = c_\mathrm{H} + \alpha_\mathrm{H}\,\langle\psi|\,\mathrm{Re}\,A\,|\psi\rangle,$$

with $c_\mathrm{H} = \alpha_\mathrm{H} = \frac{1}{h_N^2}$.

Hence, we obtain that the Hadamard-test of an 1-adder circuit, shown in Figure 2.7, can be used to reconstruct $\langle\psi|\,\mathrm{Re}(A)\,|\psi\rangle$. We thus define the Hadamard-test sampling estimator as

$$\widehat{\mathcal{K}^N}^{N_\mathrm{sh}} = c_\mathrm{H} + \frac{\alpha_\mathrm{H}}{N_\mathrm{sh}} \sum_{k=1}^{N_\mathrm{sh}} \mathbf{h}^k,$$

with $\mathbf{h}^k$ the k-th sample of the Hadamard-test of A. The variance of the estimator thus satisfies

$$\mathrm{Var}(\widehat{\mathcal{K}^N}^{N_\mathrm{sh}}) = \frac{|\alpha_\mathrm{H}|^2 - |\alpha_\mathrm{H}|^2 \langle Z_1\rangle^2_{\psi,\text{H-test}}}{N_\mathrm{sh}} \leq \frac{|\alpha_\mathrm{H}|^2}{N_\mathrm{sh}} = \frac{\alpha_0^2}{N_\mathrm{sh}}.$$

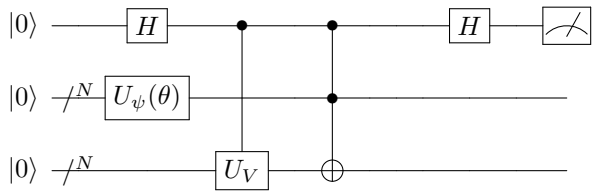


Figure 2.8 – Hadamard-test circuit for the potential term as defined in [68].

Potential term

For this term, let us first consider the Hadamard-test circuit shown in Figure 2.8 which uses two registers of N qubits. Let us denote $|\phi\rangle = U_V |0\rangle$, then we have, for this circuit,

$$\langle Z_1 \rangle_{\psi,\text{H-test}} = \langle\psi| \operatorname{diag}(|\phi\rangle) |\psi\rangle ,$$

where $\operatorname{diag}(|\phi\rangle) = \operatorname{diag}\left(\begin{bmatrix} \phi_0 \\ \vdots \\ \phi_{2^N-1} \end{bmatrix}\right)$ is a $2^N \times 2^N$ diagonal matrix. Now, the potential term can be rewritten, using (2.21), as

$$\tilde{\mathcal{P}}^N = \operatorname{diag}(V^N) = |V^N|_2 \operatorname{diag}\left(\frac{V^N}{|V^N|_2}\right). \tag{2.67}$$

As $\frac{V^N}{|V^N|_2}$ is normalised, it can be constructed as a quantum state using a unitary U_V, and thus we can write

$$|V\rangle = \frac{V^N}{|V^N|_2} = U_V |0\rangle .$$

This peculiar form used in in [68] allows the use of the Hadamard-test to calculate the potential term. Indeed, using (2.67), we can write the potential energy as

$$\langle\psi| \tilde{\mathcal{P}}^N |\psi\rangle = \alpha_{\text{H}} \langle Z_1 \rangle_{\psi,\text{H-test}},$$

where $\alpha_{\text{H}} = |V^N|_2$.

The Hadamard-test estimator for the potential term is hence given by

$$\widehat{\mathcal{P}^N}^{N_{\text{sh}}} = \frac{\alpha_{\text{H}}}{N_{\text{sh}}} \sum_{k=1}^{N_{\text{sh}}} \mathbf{h}^k$$

with $\mathbf{h}^k$ the k-th sample of the associated Hadamard-test of $\operatorname{diag}\left(\frac{V^N}{|V^N|_2}\right)$. The variance of the estimator is given by

$$\operatorname{Var} \widehat{\mathcal{P}^N}^{N_{\text{sh}}} = \frac{|V^N|_2^2 \left(1 - \langle Z_1 \rangle^2_{\psi,\text{H-test}}\right)}{N_{\text{sh}}}.$$

Since the Hermitian norm of V^N writes

$$|V^N|_2^2 = \sum_{k=0}^{2^N-1} V\left(\frac{k}{2^N}\right)^2 = 2^N |V|_{L_N^2}^2,$$

then the mean value of the operator is bounded by

$$|\langle Z_1\rangle_{\psi,\text{H-test}}| = \frac{|\langle v^2, V\rangle_{L_N^2}|}{2^{N/2}\|V\|_{L_N^2}} \leq \frac{\|V\|_{L^\infty}\|v\|_{L_N^2}^2}{2^{N/2}\|V\|_{L_N^2}} \leq \frac{C}{2^{N/2}}.$$

The range of the expected value of the Z_1 Pauli operator, $\text{Range}(Z_1) \subseteq [-\frac{C}{2^{N/2}}, \frac{C}{2^{N/2}}]$. Thus its measure exponentially decreases with the number of qubits. This allows to obtain the following bounds for the variance, using $\|V\|_{L^\infty}^2 = V_0^2/16$, and $\|V\|_{L_N^2}^2 = V_0^2/80 + O(1/4^N)$,

$$\left[1 - O\left(\frac{1}{2^N}\right)\right] 2^N \frac{V_0^2}{80} + O(\frac{1}{2^N}) \leq N_{\text{sh}} \operatorname{Var} \widehat{\mathcal{P}^N}^{N_{\text{sh}}} \leq 2^N \frac{V_0^2}{80} + O\left(\frac{1}{2^N}\right).$$

Hence, we have for the potential term an exponentially increasing number of samples for the Hadamard-test estimator even in the best-case scenarii.

Interaction term

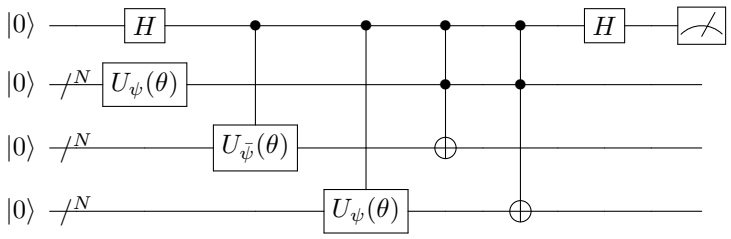


Figure 2.9 – Hadamard-test circuit for the interaction term as defined in [68].

The interaction term estimator works on the same principle as the potential term, as illustrated by their similar circuits, shown in Figures 2.8 and 2.9 respectively. Indeed, since

$$\sum_{k=0}^{2^N-1} |\psi_k|^4 = \langle\psi| \operatorname{diag}\left(\begin{bmatrix}\psi_0\\ \vdots \\ \psi_{2^N-1}\end{bmatrix}\right) \operatorname{diag}\left(\begin{bmatrix}\bar\psi_0\\ \vdots \\ \bar\psi_{2^N-1}\end{bmatrix}\right) |\psi\rangle = \langle\psi| \operatorname{diag}(|\psi\rangle) \operatorname{diag}(\langle\psi|)\, |\psi\rangle\,,$$

we can use a circuit similar to the one of the potential term to estimate this value. The circuit shown in Figure 2.9 can be used to obtain

$$\langle Z_1\rangle_{\psi,\text{H-test}} = \langle\psi| \operatorname{diag}(|\psi\rangle) \operatorname{diag}(\langle\psi|)\, |\psi\rangle\,.$$

And hence, using (2.14) and the previous equality, we can write

$$\langle \mathcal{I}_N\rangle_\psi = \frac{\kappa}{2h_N} \sum_{k=0}^{2^N-1} |\psi_k|^4 = \frac{\kappa}{2h_N} \langle Z_1\rangle_{\psi,\text{H-test}}.$$

We thus have the Hadamard-test estimator for the interaction term,

$$\widehat{\mathcal{I}^N}^{N_{\text{sh}}} = \frac{\alpha_{\text{H}}}{N_{\text{sh}}} \sum_{k=1}^{N_{\text{sh}}} \mathbf{h}^k,$$

with $\alpha_{\text{H}} = \frac{\kappa}{2h_N}$. For the variance, we have

$$\operatorname{Var} \widehat{\mathcal{I}^N}^{N_{\text{sh}}} = \frac{4^N \kappa^2}{4N_{\text{sh}}} \left(1 - \langle Z_1 \rangle^2_{\psi,\text{H-test}}\right).$$

Remark 11. *Note that, while we use the same method for the potential term, as the ket and bra are normalised vectors for the Hermitian norm $|\cdot|_2$ of $\mathbb{C}^{2^N}$, we do not have a contraction of the range, as we did for the potential term.*

Using the above results, we now define the estimation error for our estimators. As all our estimators are unbiased, the root-mean-squared-error of our estimators only consists of the square root of the variance and is just simply the sampling error, *i.e.* given an estimator $\mathcal{A}^{N_{\text{sh}}}$,

$$\text{RMSE}(\mathcal{A}^{N_{\text{sh}}}) = \sqrt{\operatorname{Var} \mathcal{A}^{N_{\text{sh}}}}.$$

Given $\epsilon > 0$, let us now determine the number of samples, N^ϵ_{sh}, required for the sampling error to be upper bounded by ϵ. To achieve

$$\text{RMSE}(\mathcal{A}^{N_{\text{sh}}}) \leq \epsilon,$$

we need

$$N_{\text{sh}} \geq \frac{\operatorname{Var} \mathcal{A}}{\epsilon^2} = N^\epsilon_{\text{sh}}.$$

As we shall see later, this N^ϵ_{sh} metric needs to be slightly modified for the interaction term due to the additional term in $\frac{1}{N_{\text{sh}}(N_{\text{sh}}-1)}$.

In order to compare our estimators, we consider the upper bounds so as to determine the worst-case scenario in terms of N^ϵ_{sh}. Now, we study how the different estimators compare for each energy term.

2.5.3 Estimator comparison

Kinetic term

For the kinetic term, as shown in Table 2.1, the Pauli method with importance sampling is outperformed by the two other methods, as expected, due to the fact that each sample is used for the estimation of a single Pauli operator. However, while the number of Pauli terms is exponential in N, it is interesting to note that the difference in performance is only quadratic in N^2. Furthermore, the circuit size required to implement the Pauli estimator consists in the implementation of only the ansatz, while for the direct-sampling estimator the implementation of an additional QFT circuit is required.

Potential term

For the potential term, as can be seen in Table 2.2, the Pauli and direct-sampling estimators perform similarly, converging to the same variance, whereas the Hadamard-test method suffers

Table 2.1 – Kinetic term - number of samples to achieve ϵ sampling error, with $d = (1-\cos(2\pi x))^2$

	Pauli	Direct-sampling	Hadamard-test
$\epsilon^2 \cdot N_{\text{sh}}^{\epsilon}$	$\lvert\alpha_1\rvert_1 \sum_{i=1}^{N_P} \lvert\alpha_i\rvert \left(1 - \langle P_i\rangle_\psi^2\right)$	$\langle v, \Delta_N^2 v\rangle_{L_N^2} - \langle v, \Delta_N v\rangle_{L_N^2}^2$	$16^N - \langle v, \Delta_N v\rangle_{L_N^2}^2$
Upper bound	$4N^2 16^{N-1}$	$2 \cdot 16^{N-1}$	$16 \cdot 16^{N-1}$
Circuit size	Ansatz	Ansatz + QFT	2H + Ansatz + controlled Adder

from an exponential overhead in N as the effective support of the expected value of the Pauli decreases exponentially with the number of qubits.

In Figure 2.10, we provide a comparison of the variance bounds of the different estimators. On the left-hand side, we display the limit of bounds for the Pauli/direct-sampling estimator, which are simply the bounds of the undiscretized operator. We display two bounds, the upper bound in Table 2.2 and a parametric bound obtained by considering $V^2 \leq \|V\|_{L^\infty} V$. On the second graph, we compare these bounds with those of the Hadamard-test operators for increasing numbers of qubits. An important note is that the potential term is the only energy term for which the variance and thus number of samples required for a fixed ϵ converges with N.

Table 2.2 – Potential term - number of samples to achieve ϵ sampling error

	Pauli	Direct-sampling	Hadamard-test
$\epsilon^2 \cdot N_{\text{sh}}^{\epsilon}$	$\langle(\Theta_N V)^2, v^2\rangle_{L_N^2} - \langle\Theta_N V, v^2\rangle_{L_N^2}^2$	$\langle V^2, v^2\rangle_{L_N^2} - \langle V, v^2\rangle_{L_N^2}^2$	$\|V^N\|_2^2 \left(1 - \lvert\langle Z_0^H\rangle\rvert^2\right)$
Upper bound	$\frac{V_0^2}{180} + O(1/4^N)$	$\frac{V_0^2}{180} + O(1/4^N)$	$2^N \frac{V_0^2}{80} + O(1)$
Circuit size	Ansatz	Ansatz	2H + Ansatz + C-Ansatz + N CNOTs

Interaction term

Recall that the direct-sampling and Pauli estimators are equivalent in the case of the interaction term and that their variance can be bounded, using Theorem 8, as

$$\operatorname{Var} \overline{\mathcal{I}^N}^{N_{\text{sh}}} \leq \frac{B_V}{N_{\text{sh}}} + \frac{B_W}{N_{\text{sh}}(N_{\text{sh}} - 1)}.$$

To bound the sampling error by $\epsilon > 0$, *i.e.* $\operatorname{RMSE}\left(\overline{\mathcal{I}^N}^{N_{\text{sh}}}\right) = \sqrt{\operatorname{Var} \overline{\mathcal{I}^N}^{N_{\text{sh}}}} \leq \epsilon$, we consider an upper bound of N_{sh}^{ϵ} large enough such that

$$-\epsilon^2 N_{\text{sh}}^2 + (B_V + \epsilon^2) N_{\text{sh}} + B_W - B_V \leq 0$$

Given that the squared term is negative and that the maximum $\frac{(\epsilon^2 - B_V)^2}{4\epsilon^2} + B_W$, attained in $N_{\text{sh}}^{\epsilon} = \frac{B_V + \epsilon^2}{2\epsilon^2}$, is positive as $B_W > 0$, there always exists real roots to the above polynomial which

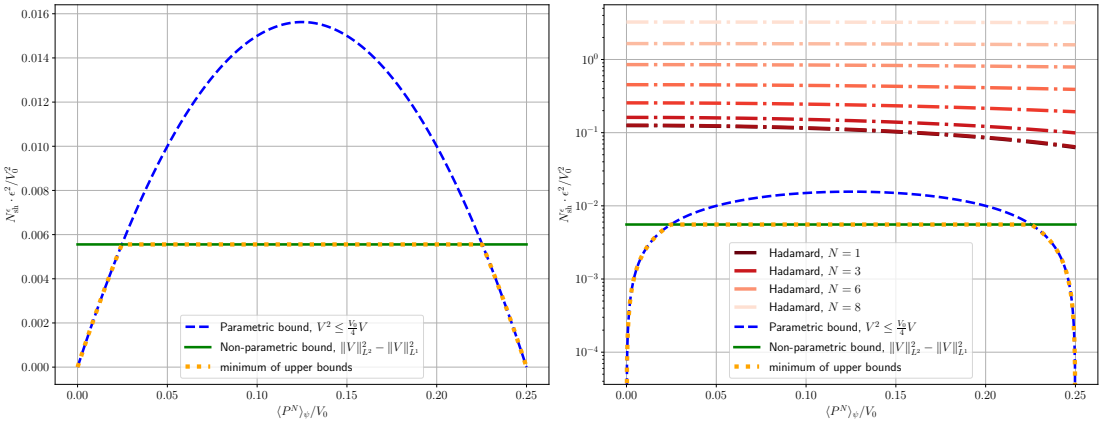


Figure 2.10 – Minimum number of shots to achieve an error ϵ for the potential term. The second graph shows the relative difference between the exponential dependence of the number of samples required for the Hadamard-test on the number of qubits.

are given by

$$r_\pm = \frac{1}{2}\left(1 + \frac{B_V}{\epsilon^2} \pm \sqrt{(1 - \frac{B_V}{\epsilon^2})^2 + 4\frac{B_W}{\epsilon^2}}\right). \tag{2.68}$$

The largest root, which we note $N_{\text{sh}}^\epsilon = r_+$ for consistency, is shown in Figure 2.11 compared to the maximum value attained in the Hadamard-test case, $\alpha_0^2\gamma$ with $\gamma = 4^N/\epsilon^2$. Indeed, by factoring out the $\alpha_0^2\gamma$ from the root, we obtain that

$$N_{\text{sh}}^\epsilon \alpha_0^{-2}\gamma^{-1} = \frac{1}{2}\left(\gamma^{-1} + 4z^3(1-z) + \sqrt{(\gamma^{-1} - 4z^3(1-z))^2 + \gamma^{-1}8z^2(1-z)^2}\right)$$

with $z = \sqrt{\frac{\langle \mathcal{I}^N \rangle_\psi}{2^N}} \in [2^{-N/2}, 1]$. We note that, as γ becomes large, either through a large N or small ϵ, the root converges to the solution $B_V = \alpha_0^2\gamma z^3(1-z)$ which is bounded by $\frac{27}{64}\alpha_0^2\gamma$. A summary of the results is found in Table 2.3 and bounds of N_{sh}^ϵ are represented in Figure 2.11, displaying the fact that in all cases, the Pauli/direct-sampling estimator outperform the Hadamard-test. It is interesting to note that here the 'Paulification' of the variance of the Hadamard-test estimator results in severely reduced performance in the case where z is far from 1.

Table 2.3 – Interaction term - number of samples to achieve ϵ sampling error with $z = \sqrt{\langle \mathcal{I}^N \rangle_\psi / (\alpha_0 2^N)} = \|v^2\|_{L_N^2}/2^N$

	Pauli/ direct-sampling	Hadamard-test
$\epsilon^2 \cdot N_{\text{sh}}^\epsilon$	$\frac{\alpha_0^2 4^N}{2}\left(\frac{\epsilon^2}{4^N} + 4z^3(1-z) + \sqrt{\left(\frac{\epsilon^2}{4^N} - 4z^3(1-z)\right)^2 + \frac{8\epsilon^2}{4^N}z^2(1-z)^2}\right)$	$\alpha_0^2 4^N\left(1 - z^4\right)$
Upper bound	$\frac{27\alpha_0^2}{64}4^N$	$\alpha_0^2 4^N$
Circuit Size	Ansatz	2H + Ansatz + 2 C-Ansatz + $2N$ CNOTs

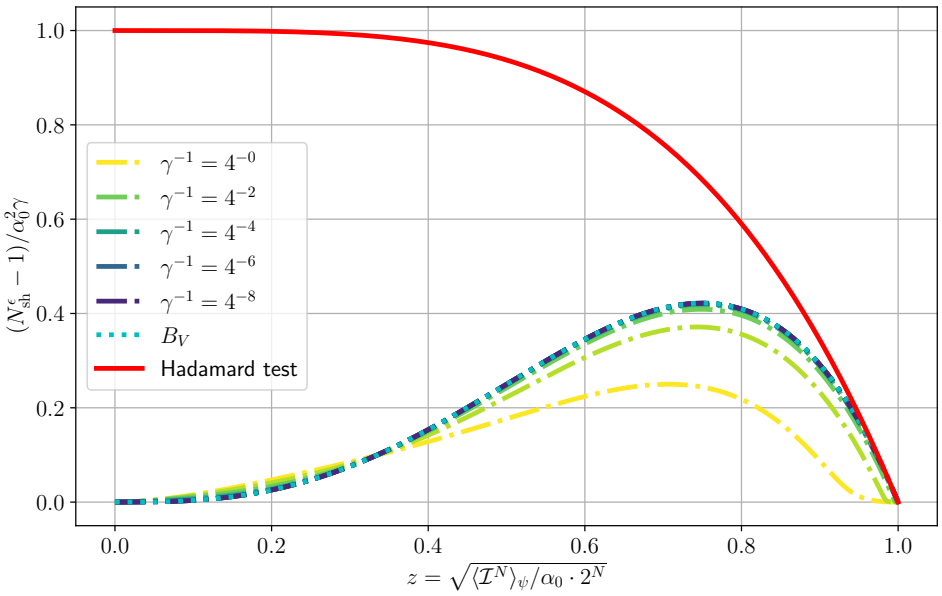


Figure 2.11 – Minimum number of shots to achieve an error ϵ for the interaction term. Interaction term number of shots to achieve sampling error ϵ for a given number of qubits N, $\gamma = \frac{4^N}{\epsilon^2}$. As can be readily observed, the direct-sampling/Pauli estimators consistently outperform the Hadamard-test estimator.

Number of samples for convergence

As shown in Appendix A.1, the minimizer of the discretized model converges in H^1 norm in $O(1/2^N)$. Furthermore, using the same reasoning as in Lemma 13, and the fact that C_{λ_N} is uniformly bounded with respect to N, there exists $C > 0$, which does not depend on N, such that, for any $v \in \mathcal{X}_N$,

$$\|v - u_N\|_{H^1}^2 \leq C \left(E_N(v) - E_N(u_N)\right).$$

This implies the following bound through the triangle inequality

$$\|v - u\|_{H^1} \leq \|u - u_N\|_{H^1} + \|v - u_N\|_{H^1} \leq \frac{C}{2^N} + C \left(E_N(v) - E_N(u_N)\right)^{1/2}.$$

Hence, we need to estimate our energy terms with precision $\epsilon \leq \epsilon_{\text{conv}} = C^2/4^N$. The effect of this choice of precision is shown in Figure 2.12 where we set $\epsilon = \epsilon_{\text{conv}}$. As expected from the previous tables, we observe that for the interaction and potential terms, the direct-sampling and Pauli estimators have better performances, and that, in the case of the kinetic term, the direct-sampling estimator outperforms both other methods. Figure 2.12 also shows that even with the best methods, the number of samples required to obtain an estimate of the energy with the required precision becomes prohibitive very fast and that the kinetic term dominates both others in terms of number of samples needed. Indeed, for example, if we target an error of $\epsilon \approx 10^{-8}$, we can see that the number of samples required to estimate the kinetic term are of order $2^{100} \approx 10^{30}$ samples. Let us now estimate the amount of time needed for a single current-era quantum computer to

achieve this number of samples. To obtain a lower bound, let us consider only the measurement time, 500 ns and reset time 160 ns of one of the Google superconducting quantum computers [1]. This gives us a repetition rate upper bounded by ≈ 1.5 Mhz, to achieve 10^{30} samples, the time required is thus greater than $2/3 \cdot 10^2 4s \approx 2 \cdot 10^{16}$ years, *i.e.* about 10^6 times the age of the universe.

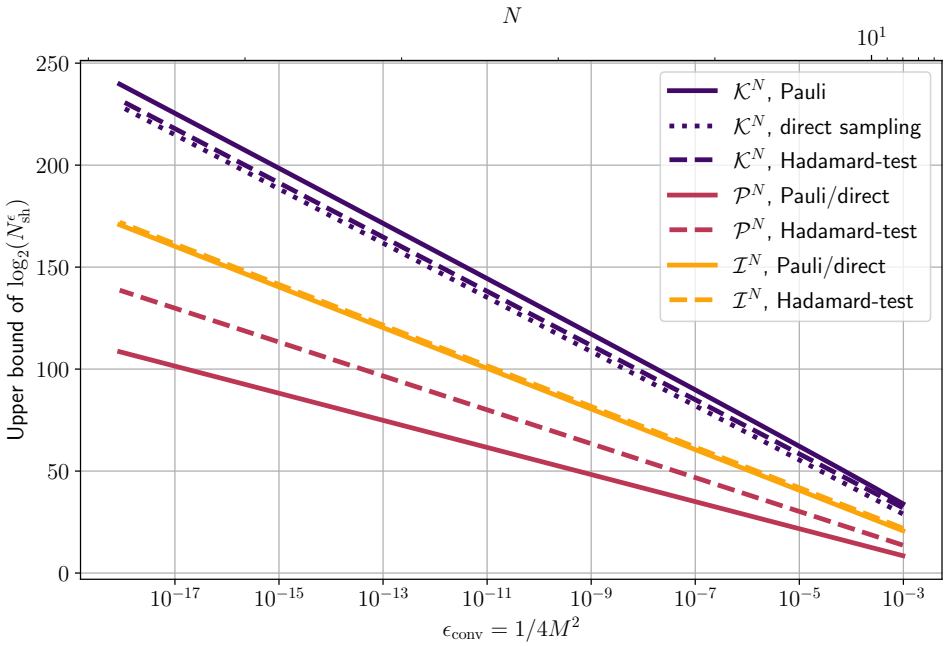


Figure 2.12 – Minimum number of shots to achieve convergence accuracy, $\epsilon = \epsilon_{\text{conv}}$. For illustration purposes, we consider the case $V_0 = 1$, $\kappa = 1$.
In this figure we compare the number of shots required to obtain a precision $\epsilon_{\text{conv}} = C/2^N$ on each energy term for each estimator. The number of qubits displayed for illustration purposes are in the range $\{4, \ldots, 30\}$.

2.6 Discussion

In this chapter, we have defined new operators based on the truncation of the analytical Walsh series of functions, and highlighted a convergence rate of the potential term estimator to the discretized energy of order $1/M^2$. We also provided an a-priori convergence analysis of the discretized model and shown that, due to the ill-discretized kinetic term, we only have a $O(1/M)$ convergence in H^1 norm of the discretized solution u_N to the exact solution u. We then compared the variance of our estimators and each required number of samples to achieve a fixed error. Through this comparison, we emphasized that our estimators perform similarly to the direct-sampling one in the diagonal case, and provide the exact same estimator in the case of the unbiased interaction energy estimator. For the kinetic energy, the corresponding operator being non-diagonal, a quadratic overhead of the number of samples is induced for the Pauli estimator, but with the advantage of a shorter circuit, since it does not require the implementation of a

QFT circuit to measure in the diagonalization basis of the operator. For the potential energy, the number of samples required to reach an accuracy of ϵ converges in the case of the Pauli and direct-sampling estimators, while an exponential overhead is induced for the Hadamard-test estimator due to the restriction of the support of the Pauli term being measured. We have provided estimates for the number of samples required in order to achieve a sufficient precision on the energy so as to be able to converge to the discretized solution. These estimates show that the number of samples required to estimate the energy of a single state are extreme, even in the case of 30 qubits, and are thus extremely prohibitive in the context of an optimization procedure even for lower numbers of qubits.

One way to reduce the number of samples required to achieve a given precision on the solution would be to implement the spectral Laplace operator. Indeed, this would reduce the overhead for the kinetic and interaction term, significantly, as the number of qubits required for a given error would be lowered. In the case of the direct-sampling estimator, the use of the spectral Laplace operator is very simple to implement: it would simply amount to considering a different diagonal operator in the QFT basis. A comparison of the number of samples for the direct-sampling estimator for the correct discretization is given in Figure 2.13 and shows that it can heavily reduce the number of samples required, but not enough to make it viable for higher precisions.

The exorbitant number of samples required even with the spectral Laplace operator can be attributed to three things. First and most importantly, in the case of unbounded operators on $\mathcal{X}$, the norm of the discretized operator diverges with N and so does its variance. For the Laplace operator, as it is in H^{-2}, it diverges as M^2, while for the interaction operator, it only diverges as M. Thus, the norms of the kinetic and interaction operators diverge exponentially fast with respect to N and so will the number of samples required to achieve an accuracy ϵ. Secondly, but to a lesser degree, the statistical nature of the energy estimation implies that the number of samples required will always be proportional to ϵ^{-2}. Finally, we note that, while the convergence of the minimizer u_N to u in H^1 norm is bounded by $\frac{1}{M}$, the accuracy ϵ required on the energy is of order $O(\frac{1}{M^2})$ as $\|v - u_N\|^2_{H^1}$ is bounded by the square-root of the energy. Hence, in the end, we need a number of samples proportional to M^4 to achieve a $1/M$-order convergence.

There is however one positive take-away from our analysis, the energy estimation in the case of the potential operator shows that the number of samples required for a given precision on the energy, even in the large N limit, remains constant.

Finally, we note that, in practice, the additional problem of quantum noise adds a bias to the estimators which depends exponentially in the number of gates. In this context, the relative performance of each estimator needs to be reevaluated as the circuit depth of the estimators heavily impacts their performance. The Pauli-sampling estimators, which only requires the implementation of an ansatz thus will be the least affected, and, in the large N limit, may outperform the other estimators. While, for our current case, this might not have much of an effect, on the overall performance of the algorithm, for the sake of thoroughness, let us now study the effect of noise on the algorithm. In Chapter 3, we shall explore the additional effects of depolarization noise and the effects of mitigation techniques on our estimators and their performance in practice.

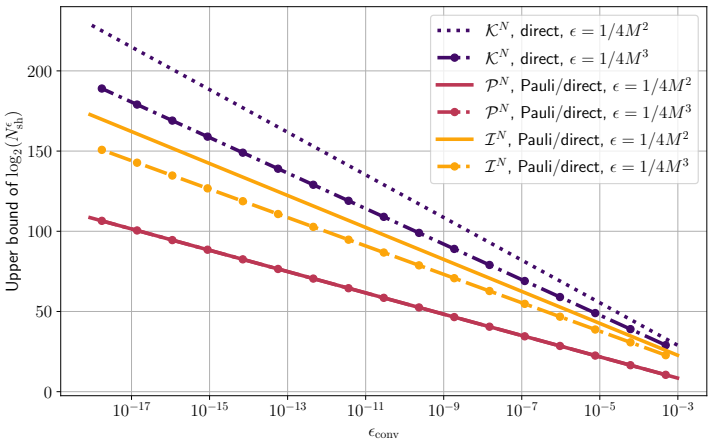


Figure 2.13 – Minimum number of shots to achieve convergence accuracy, $\epsilon = \epsilon_{\mathrm{conv}}$. Comparison with spectral Laplace operator discretization. The graph displays the case when $V_0 = 1$ and $\kappa = 1$.

Chapter 3

Error analysis under depolarizing noise

3.1 Introduction

In this chapter, we look at the effects of quantum noise on the estimation of energy terms in the case of the amplitude encoding of a function v and determine the change in performance induced by such effects on the estimation methods defined in the previous chapter. We consider a depolarization noise model and the effects of such a noise model on a quantum circuit. We then determine the effects of this noise on the mean values of operators and thus on our different estimators. Finally, we consider the total effect on each energy term estimator and consider the impact in terms of performance in two cases, without mitigation techniques and with error mitigation techniques. We show that, in the unmitigated case, the bias induced by quantum noise limits severely the number of gates usable while achieving the precision $\epsilon_{\text{conv}} > 0$ required for convergence. For the mitigated case, we demonstrate that the number of samples is increased exponentially with respect to the number of gates, but at a much slower rate than $N_{\text{sh}}^{\epsilon} \approx \epsilon_{\text{conv}}^{-2}$ increases with N. Furthermore, we show that, for a given discretization method, there always exists a number of qubits after which the Pauli estimator requires fewer samples to achieve a specified precision on the energy.

3.2 Quantum noise

In order to determine the effect of noise on our energy estimators, let us first provide an overview on how quantum noise can be modelled and how it affects a quantum system. In this section, we define the density matrix, the noisy analog of a quantum state, describe how it is affected by quantum channels, the noisy analog of quantum gates, and provide the noise model we consider in the rest of this chapter.

3.2.1 Density matrices and quantum channels

Until now, we only worked with noiseless quantum systems and were able to represent the state of the quantum system using kets $|\cdot\rangle$. As quantum noise introduces classical uncertainty, we are required to introduce a new tool which allows to describe a statistical distribution of quantum states.

Definition 12. *A statistical distribution of states can be described through a density matrix $\rho \in \mathscr{L}(\mathcal{H})$, a positive semi-definite operator with trace equal to 1. Given a distribution $(p_s, |\psi_s\rangle)_{1\leq s\leq N_{\text{st}}}$ of $N_{\text{st}} \in \mathbb{N}^*$ states, where $p_s \in [0,1]$ and $|\psi_s\rangle \in \mathcal{H}$ for any s, $1 \leq s \leq N_{\text{st}}$, and such that $\sum_{s=1}^{N_{\text{st}}} p_s = 1$, the associated density matrix is given by*

$$\rho = \sum_{s=1}^{N_{\text{st}}} p_s \, |\psi_s\rangle \langle\psi_s| \, .$$

The action of noise on a quantum system can be described as linear maps on $\mathscr{L}(\mathcal{H})$, with some additional properties, namely, the maps are completely positive and trace-preserving, which we now define. Let id_k be the identity operator on $\mathbb{C}^{k\times k}$, then

Definition 13 ([73, 97])**.** *Given A and B two C^*-algebras, a map $\phi : A \to B$ is said to be positive if any positive element of A is mapped to a positive element of B. A map $\phi : A \to B$ is said to be completely positive if, for any $k \in \mathbb{N}^*$, $id_k \otimes \phi : \mathbb{C}^{k\times k} \otimes A \to \mathbb{C}^{k\times k} \otimes B$ is a positive map.*

The physical implication of complete positivity is that such a map preserves the fact that the partial trace of a density matrix on any subsystem is also a positive semi-definite matrix. The other physical requirement for describing quantum noise is that it must be trace-preserving. These two properties allow to define a quantum channel.

Definition 14. *A quantum channel $\mathcal{E}$ on $\mathscr{L}(\mathcal{H})$ is a completely positive and trace-preserving linear map from $\mathscr{L}(\mathcal{H})$ onto itself.*

A quantum channel is able to describe the effects of noise, indeed by construction, it sends a density operator onto a density operator, it can hence be seen as the generalization for density matrices of unitary operators which send quantum states onto quantum states.

The application of a quantum gate with matrix $U_G \in \mathscr{L}(\mathcal{H})$ in the density matrix formalism is given by the quantum channel $\mathcal{E}_{U_G} : \rho \mapsto U_G \rho U_G^\dagger$. This channel preserves the purity of a density matrix and thus sends the density matrix of a single quantum state onto the density matrix of another single quantum state.

For gates under the influence of noise, we shall model the effect of noise by considering that, after the application of the gate, an additional noise channel is applied. Given a gate G associated to U_G, and a noisy channel $\mathcal{E}_{\text{Noise}}$, then the quantum channel of the gate under the effect of noise is given by

$$\mathcal{E}_G = \mathcal{E}_{\text{Noise}} \circ \mathcal{E}_{U_G}.$$

In order to consider the effect of a noisy implementation of a circuit C of size $N_G \geq 1$ we define the quantum channel $\mathcal{E}^C$ associated to the circuit as

$$\mathcal{E}^C = \mathcal{E}^C_{N_G} \circ \cdots \circ \mathcal{E}^C_1,$$

where, for any $1 \leq k \leq N_G$, $\mathcal{E}^C_k$ is the quantum channel associated to the k-th gate of the circuit C. In the case of a perfect circuit with no noise, we recover $\mathcal{E}^C_k(\rho) = U^C_k \rho U^{C\dagger}_k$, and thus $\mathcal{E}^C(\rho) = U^C \rho U^{C\dagger}$, where

$$U^C = U^C_{N_G} \dots U^C_1 \tag{3.1}$$

is the matrix of the circuit in the canonical basis $\mathscr{B}_{\text{C}}(\mathcal{H})$. This is equivalent to the unitary of a circuit (1.2) as defined in Chapter 1.

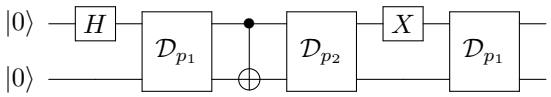


Figure 3.1 – Noisy gate: insertion of a depolarizing channel after each gate.

3.2.2 Noise model

In order to approximate the effect of noise on quantum gates, we consider the addition of a depolarizing channel [73, 97] after each gate. The choice of depolarizing channels to model noise is based on two facts. First, its effects can be seen as a uniform contraction in the Pauli basis towards the maximally-mixed state $\frac{I_N}{2^N}$, which allows to obtain analytical results. Second, for any quantum hardware under noise, the noise can be modified to follow a depolarisation model through Pauli twirling methods[69, 94] wherein random Pauli gates are inserted into the circuit. A depolarizing channel $\mathcal{D}_p$ of probability $p \in [0, 1]$ can be defined as a quantum channel sending, with probability p, the density matrix ρ of the system into the uniform distribution on all quantum states of the system, *i.e.*

$$\mathcal{D}_p(\rho) = (1-p)\rho + \frac{p}{2^N} I_{\mathcal{H}}.$$

We consider a noise model where a global depolarization noise channel is added after each gate, with a probability $p_1 \in [0, 1]$ for 1-qubit gates and $p_2 \in [0, 1]$ for 2-qubit gates. More precisely, for any gate G of unitary U_G and support $\mathcal{S}_G$, we consider the following channel

$$\mathcal{E}_G : \rho \mapsto \mathcal{D}_{p_{|\mathcal{S}_G|}} \left(U_G \rho U_G^\dagger \right).$$

This is illustrated in Figure 3.1 for the simple circuit $C = (H \otimes I, \text{CNOT}, X \otimes I)$.

We now consider the effect of such a noise model on a circuit C with $N_G \geq 1$ gates with a system initialised in the state $\rho_0 = |0\rangle\langle 0|$, where $N_{G_1} \geq 0$ is the number of 1-qubit gates, $N_{G_2} \geq 0$ the number of 2-qubit gates and $N_G = N_{G_1} + N_{G_2}$.

This noisy implementation of our circuit corresponds to the noisy circuit channel

$$\begin{aligned}
\mathcal{E}^C(\rho_0) &\equiv \mathcal{E}_{N_G} \left(U_{N_G} \mathcal{E}_{N_G - 1} \left(U_{N_G - 1} \dots \mathcal{E}_1 \left(U_1 \rho_0 U_1^\dagger \right) \dots U_{N_G-1}^\dagger \right) U_{N_G}^\dagger \right) \\
&= \mathcal{E}_{N_G} \circ \dots \circ \mathcal{E}_1 \left(U_{N_G} \dots U_1 \rho_0 U_1^\dagger \dots U_{N_G}^\dagger \right) \\
&= (1-p_1)^{N_{G_1}} (1-p_2)^{N_{G_2}} U_{N_G} \dots U_1 \rho_0 U_1^\dagger \dots U_{N_G}^\dagger + \left[1 - (1-p_1)^{N_{G_1}} (1-p_2)^{N_{G_2}} \right] \frac{I_{\mathcal{H}}}{2^N},
\end{aligned}$$

where we define $\mathcal{E}_k$ to be either $\mathcal{D}_{p_1}$ if the k-th gate is a 1-qubit gate, or $\mathcal{D}_{p_2}$ if it is a 2-qubit gate. We hence obtain the following lemma.

Lemma 6. *For a circuit C of N_{G_1} 1-qubit gates and N_{G_2} 2-qubit gates under a global depolarizing noise model with probabilities p_1 for 1-qubit gates and p_2 for 2-qubit gates, the resulting density matrix at the end of the circuit is*

$$\mathcal{E}^C(\rho_0) = qU^C\rho_0 U^{C\dagger} + (1-q)\,\frac{I_{\mathcal{H}}}{2^{N_q}},$$

with U^C as defined in (3.1) *and*

$$q = (1-p_1)^{N_{G_1}}(1-p_2)^{N_{G_2}}. \tag{3.2}$$

Note that q decreases to 0 exponentially with the number of gates in the circuit. Let us now investigate the effects of such a noise model in practice.

3.3 Quantum noise-induced errors

To determine the effects of our noise model on our algorithm's performance, we need to determine its effects on our kinetic, potential and interaction energy estimators. We now consider the effect of the quantum noise on the different energy estimation methods under study here. Given an arbitrary operator $A \in \mathscr{L}(\mathcal{H})$ and a quantum state $|\psi\rangle$, as we have seen in the previous chapter, in the noiseless case, the direct-sampling, Pauli-sampling and Hadamard-test estimators provide unbiased estimators of the mean value $\langle A\rangle_\psi$. Let us now consider the circuit C of N_{G_1} 1-qubit gates and N_{G_2} 2-qubit gates implementing the above estimators. Under global depolarizing noise, with respective probabilities p_1 and p_2 for 1 and 2 qubit gates, we have, as stated in Lemma 6, the density matrix at the end of the circuit is

$$\rho = \mathcal{E}^C(\rho_0) = qU^C\rho_0 U^{C\dagger} + (1-q)\frac{I_{\mathcal{H}}}{2^N}, \tag{3.3}$$

where $\rho_0 = |\psi\rangle\langle\psi|$ For any observable O, we thus have $\langle O\rangle_\rho = q\langle O\rangle_\psi + (1-q)\frac{\operatorname{Tr} O}{2^N}$. Additionally, we can rewrite

$$\langle O\rangle_\psi = \left\langle O - \frac{\operatorname{Tr} O}{2^N} I \right\rangle_\psi + \frac{\operatorname{Tr} O}{2^N}\langle I\rangle_\psi = \left\langle O - \frac{\operatorname{Tr} O}{2^N} I \right\rangle_\psi + \frac{\operatorname{Tr} O}{2^N}.$$

Hence, setting

$$\lambda_0 = \frac{\operatorname{Tr} O}{2^N},$$

we obtain the following proposition.

Proposition 14. *Let a Hermitian operator O, given a unitary U_ψ such that $U_\psi|0\rangle = |\psi\rangle$ and a circuit implementing U_ψ with N_{G_1} 1-qubit gates and N_{G_2} 2-qubit gates. We obtain, for the density matrix ρ from* (3.3)*,*

$$\langle O\rangle_\rho = q\langle O\rangle_\psi + (1-q)\frac{\operatorname{Tr} O}{2^N} = \langle O\rangle_\psi - (1-q)\langle O\rangle_\psi + (1-q)\lambda_0. \tag{3.4}$$

The noise thus introduces a bias

$$b^O = (1-q)|\langle O\rangle_\psi - \lambda_0| \geq 0.$$

Since the Pauli operators are all traceless, apart from the identity, we obtain the following simplification of Proposition 14.

Proposition 15. *Let $P \in \mathscr{B}_{\mathrm{P}}(\mathcal{H})$, given a unitary U_ψ such that $U_\psi |0\rangle = |\psi\rangle$ and a circuit implementing U_ψ with N_{G_1} 1-qubit gates and N_{G_2} 2-qubit gates, we have*

$$\langle P \rangle_\rho = q \langle P \rangle_\psi. \tag{3.5}$$

This can be seen as a contraction, with Lipschitz constant q (3.2), of the mean value of the Pauli operator towards zero, exponentially fast with respect to the number of gates. Hence, the main effect of noise is to contract the range of the expected value of operators. Endowed with these properties, we can now investigate the noise effects on our estimators.

3.3.1 Pauli-sampling estimator

Recall, from (2.31), that the the Pauli-sampling estimator for the operator $A = \alpha_0 I^N + \sum_{i=1}^{N_A} \alpha_i P_i$ is given by

$$\overline{A}^{N_{\mathrm{sh}}} = \alpha_0 + \sum_{i=1}^{N_A} \alpha_i \overline{P_i}^{N_{\mathrm{sh}}^i}$$

with $\overline{P_i}^{N_{\mathrm{sh}}^i} = \dfrac{1}{N_{\mathrm{sh}}^i} \sum\limits_{k=1}^{N_{\mathrm{sh}}^i} \mathbf{p}_i^k$. In order to determine the noise effect on this estimator, we first look at how the noise affects the expectancy of a single sample.

Lemma 7. *For a sample $\mathbf{p}$ of a Pauli operator P, we have, from Proposition 15,*

$$\mathbb{E}_\rho [\mathbf{p}] = q \mathbb{E}_\psi [\mathbf{p}],$$

and

$$\mathrm{Var}_\rho \, \mathbf{p} = 1 - q^2 \mathbb{E}_\psi [\mathbf{p}]^2 .$$

Proof. Indeed, from Proposition 15, we obtain

$$\mathbb{E}_\rho [\mathbf{p}] = \langle P \rangle_\rho = q \, \mathrm{Tr}(P \rho_0),$$

and hence the first equality. For the variance, we have

$$\mathrm{Var}_\rho \, \mathbf{p} = \mathbb{E}_\rho \left[\mathbf{p}^2\right] - \mathbb{E}_\rho [\mathbf{p}]^2 = \langle P^2 \rangle_\rho - \langle P \rangle_\rho^2$$

and thus, using the first equality and the fact that $P^2 = 1$, we obtain the second equality. □

We can then compute the noisy mean and variance of $\overline{A}^{N_{\mathrm{sh}}}$.

Proposition 16. *The mean of a Pauli sampling estimator $\overline{A}^{N_{\mathrm{sh}}}$ of an operator $A = \alpha_0 I + \sum_{i=1}^{N_A} \alpha_i P_i$ under global depolarizing noise is given by*

$$\mathbb{E}_\rho \left[\overline{A}^{N_{\mathrm{sh}}}\right] = (1-q)\alpha_0 + q \langle A \rangle_\psi,$$

and its variance is given by

$$\operatorname{Var}_\rho \overline{A}^{N_{\mathrm{sh}}} = q^2 \operatorname{Var}_\psi \overline{A}^{N_{\mathrm{sh}}} + (1-q^2)\sum_{i=1}^{N_A} \frac{\alpha_i^2}{|S_i|} + (1-q)q \sum_{i=1}^{N_A} \sum_{j=1, j\neq i}^{N_A} \frac{|S_i \cap S_j|}{|S_i||S_j|} \alpha_i \alpha_j \langle P_i P_j \rangle_\psi. \quad (3.6)$$

Proof. For the mean, we obtain using 7,

$$\mathbb{E}_\rho \left[\overline{A}^{N_{\mathrm{sh}}} \right] = \alpha_0 + \sum_{i=1}^{N_A} \frac{\alpha_i}{N_{\mathrm{sh}}^i} \sum_{k=1}^{N_{\mathrm{sh}}^i} \mathbb{E}_\rho[\mathbf{p}_i^k] = \alpha_0 + \sum_{i=1}^{N_A} \frac{\alpha_i}{N_{\mathrm{sh}}^i} N_{\mathrm{sh}}^i q \langle P_i \rangle_\psi$$

$$= \alpha_0 + q\langle A - \alpha_0 I^N \rangle_\psi = (1-q)\alpha_0 + q\langle A \rangle_\psi$$

For the variance we have

$$\operatorname{Var}_\rho \overline{A}^{N_{\mathrm{sh}}} = \sum_{i=1}^{N_A}\sum_{j=1}^{N_A} \frac{\alpha_i \alpha_j}{N_{\mathrm{sh}}^i N_{\mathrm{sh}}^j} \left(\sum_{k\in S_i} \sum_{\ell \in S_j} \mathbb{E}_\rho \left[\mathbf{p}_i^k \mathbf{p}_j^\ell\right] - \mathbb{E}_\rho \left[\mathbf{p}_i^k\right] \mathbb{E}_\rho \left[\mathbf{p}_j^\ell\right] \right)$$

$$= \sum_{i=1}^{N_A} \left(\frac{\alpha_i^2}{|S_i|} \left(1 - \langle P_i \rangle_\rho^2\right) + \sum_{j=1, j\neq i}^{N_A} \frac{|S_i \cap S_j|}{|S_i||S_j|} \alpha_i \alpha_j \left(\langle P_i P_j \rangle_\rho - \langle P_i \rangle_\rho \langle P_j \rangle_\rho \right) \right).$$

Rewriting the previous equality, we make the noiseless variance appear

$$\operatorname{Var}_\rho \overline{A}^{N_{\mathrm{sh}}} = \sum_{i=1}^{N_A} \frac{q^2 \alpha_i^2}{|S_i|} \left(1 - \langle P_i \rangle_\psi^2\right) + \frac{(1-q^2)\alpha_i^2}{|S_i|}$$
$$+ \sum_{i=1}^{N_A} \sum_{j=1, j\neq i}^{N_A} \frac{|S_i \cap S_j|}{|S_i||S_j|} \alpha_i \alpha_j \left(q^2 \langle P_i P_j \rangle_\psi - q^2 \langle P_i \rangle_\psi \langle P_j \rangle_\psi \right) + \frac{|S_i \cap S_j|}{|S_i||S_j|} \alpha_i \alpha_j (1-q) q \langle P_i P_j \rangle_\psi,$$

which, using (2.32), allows to conclude. □

Let us now recall that the estimators we considered in the previous chapter follow two specific sampling strategies, the diagonal-sampling and importance-sampling ones, from Definitions 11 and 10 respectively. Let us apply Proposition 16 to those cases.

Corollary 2. *For an importance-sampling strategy, the variance of a Pauli sampling estimator* $\overline{A}^{N_{\mathrm{sh}}}$ *of an operator* $A = \alpha_0 I + \sum_{i=1}^{N_A} \alpha_i P_i$ *under global depolarizing noise is given by*

$$\operatorname{Var}_\rho \overline{A}^{N_{\mathrm{sh}}} = q^2 \operatorname{Var}_\psi \overline{A}^{N_{\mathrm{sh}}} + (1-q^2)\frac{|\alpha_{1:}|_1^2}{N_{\mathrm{sh}}},$$

recalling that $\alpha_{1:}$ *is given by* (2.38) *and where we denote the noiseless variance* (2.39) *by* $\operatorname{Var}_\psi \overline{A}^{N_{\mathrm{sh}}}$. *We also have the upper bound*

$$\operatorname{Var}_\rho \overline{A}^{N_{\mathrm{sh}}} \leq \frac{|\alpha_{1:}|_1^2}{N_{\mathrm{sh}}}.$$

Proof. The result is straightforward using (3.6) and the fact that, for any $i, j \in \{1, \dots, N_A\}$ such that $i \neq j$, $S_i \cap S_j = \emptyset$ and that $N_{\text{sh}}^i = \frac{|\alpha_i|}{\|\alpha_{1:}\|_1} N_{\text{sh}}$. The bound comes from (2.39). □

Corollary 3. *For a diagonal-sampling strategy, the variance of a Pauli sampling estimator $\overline{A}^{N_{\text{sh}}}$ of an operator $A = \alpha_0 I + \sum_{i=1}^{N_A} \alpha_i P_i$ under global depolarizing noise is given by*

$$\operatorname{Var}_\rho \overline{A}^{N_{\text{sh}}} = q^2 \operatorname{Var}_\psi \overline{A}^{N_{\text{sh}}} + \frac{(1-q)}{N_{\text{sh}}} |\alpha_{1:}|_2^2 + \frac{(1-q)q}{N_{\text{sh}}} \langle (A - \alpha_0 I^N)^2 \rangle_\psi \tag{3.7}$$

where we denote the noiseless variance (2.40) *by* $\operatorname{Var}_\psi \overline{A}^{N_{\text{sh}}}$. *Furthermore, we have the bound*

$$\operatorname{Var}_\rho \overline{A}^{N_{\text{sh}}} \leq q^2 \operatorname{Var}_\psi \overline{A}^{N_{\text{sh}}} + \frac{(1-q)}{N_{\text{sh}}} \left(|\alpha_{1:}|_2^2 + q(\|A\|_\sigma - \alpha_0)^2 \right). \tag{3.8}$$

Proof. Notice that, for any i, $j \in \{1, \dots N_A\}$ such that $i \neq j$, $S_i \cap S_j = S_i = S_j$ and that $|S_i| = |S_j| = N_{\text{sh}}$. Hence, using (3.6), we obtain

$$\operatorname{Var}_\rho \overline{A}^{N_{\text{sh}}} = q^2 \operatorname{Var}_\psi \overline{A}^{N_{\text{sh}}} + \frac{1-q^2}{N_{\text{sh}}} \sum_{i=1}^{N_A} \alpha_i^2 + (1-q)q \frac{1}{N_{\text{sh}}} \sum_{i=1}^{N_A} \left[\sum_{j=1}^{N_A} \alpha_i \alpha_j \langle P_i P_j \rangle_\psi - \alpha_i^2 \right],$$

which we rewrite as

$$\operatorname{Var}_\rho \overline{A}^{N_{\text{sh}}} = q^2 \operatorname{Var}_\psi \overline{A}^{N_{\text{sh}}} + \frac{1-q}{N_{\text{sh}}} \|\alpha_{1:}\|_2^2 + \frac{(1-q)q}{N_{\text{sh}}} \langle (A - \alpha_0 I^N)^2 \rangle_\psi.$$

This allows to recover (3.7), and (3.8), thanks to the inequality $\langle (A - \alpha_0 I^N)^2 \rangle_\psi \leq (\|A\|_\sigma - \alpha_0)^2$. □

Note that, for the importance sampling, we obtain the same bound as for the noiseless case (2.39), while for the diagonal-sampling, we obtain a variance which is an interpolation between the noiseless variance and the variance for a uniform distribution.

Thanks to these results, we can determine the noise effect on the kinetic and potential energy Pauli estimators. For the interaction term, however, we first need to study the noise effect on the squared Pauli estimator from Definition 2.4.5. Let us now provide some results on the variance, similar to Proposition 13 for the noiseless case, and bias of such estimators.

Proposition 17. *Let $1 \leq i \leq N_{\mathcal{I}}$. With N_{sh} samples following a diagonal-sampling strategy, we have*

$$\mathbb{E}_\rho \left[\overline{P_i^2}^{N_{\text{sh}}} \right] = q^2 \langle P_i \rangle_\psi^2. \tag{3.9}$$

Moreover, let $1 \leq j \leq N_{\mathcal{I}}$ such that $i \neq j$. We have

$$\begin{aligned} \operatorname{Cov}_\rho \left(\overline{P_i^2}^{N_{\text{sh}}}, \overline{P_j^2}^{N_{\text{sh}}} \right) &= q^4 \frac{6 - 4N_{\text{sh}}}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi^2 \langle P_j \rangle_\psi^2 \\ &\quad + q^3 \frac{4(N_{\text{sh}} - 2)}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi \langle P_j \rangle_\psi \langle P_i P_j \rangle_\psi + q^2 \frac{2}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i P_j \rangle_\psi^2. \end{aligned} \tag{3.10}$$

Furthermore, we have

$$\operatorname{Var}_\rho \overline{P_i^2}^{N_{\text{sh}}} = q^4 \frac{6 - 4N_{\text{sh}}}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi^4 + q^2 \frac{4(N_{\text{sh}} - 2)}{N_{\text{sh}}(N_{\text{sh}} - 1)} \langle P_i \rangle_\psi^2 + \frac{2}{N_{\text{sh}}(N_{\text{sh}} - 1)}. \tag{3.11}$$

Proof. The mean is given by

$$\mathbb{E}_\rho\left[\overline{P_i^2}^{N_{\mathrm{sh}}}\right] = \sum_{k=1}^{N_{\mathrm{sh}}} \sum_{\ell=1,\ell\neq k}^{N_{\mathrm{sh}}} \mathbb{E}_\rho[\mathbf{p}_k \mathbf{p}_\ell]$$

by independence of $\mathbf{p}_k$ and $\mathbf{p}_\ell$, and then using Proposition 7, we obtain (3.9).

For the variance, we follow the exact same methodology as the proof of Proposition 13 but with the noisy expected values of the Pauli samples. Let $1 \leq i, j \leq N_{\mathcal{I}}$, we have

$$\mathrm{Cov}_\rho\left(\overline{P_i^2}^{N_{\mathrm{sh}}}, \overline{P_j^2}^{N_{\mathrm{sh}}}\right) = \mathbb{E}_\rho\left[\overline{P_i^2}^{N_{\mathrm{sh}}} \overline{P_j^2}^{N_{\mathrm{sh}}}\right] - \mathbb{E}_\rho\left[\overline{P_i^2}^{N_{\mathrm{sh}}}\right] \mathbb{E}\left[\overline{P_j^2}^{N_{\mathrm{sh}}}\right]. \tag{3.12}$$

Let us look more closely at the first term $\mathbb{E}_\rho\left[\overline{P_i^2}^{N_{\mathrm{sh}}} \overline{P_j^2}^{N_{\mathrm{sh}}}\right]$ over the individual samples, *i.e.*

$$\mathbb{E}_\rho\left[\overline{P_i^2}^{N_{\mathrm{sh}}} \overline{P_j^2}^{N_{\mathrm{sh}}}\right] = \frac{1}{\left(N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)\right)^2} \sum_{k=1}^{N_{\mathrm{sh}}} \sum_{\substack{k'=1 \\ k'\neq k}}^{N_{\mathrm{sh}}} \sum_{\ell=1}^{N_{\mathrm{sh}}} \sum_{\substack{\ell'=1 \\ \ell'\neq \ell}}^{N_{\mathrm{sh}}} \mathbb{E}_\rho\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right]. \tag{3.13}$$

The value of $\mathbb{E}_\rho\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right]$ depends on the sample associated to each Pauli term, *i.e.* the indices k, ℓ, k', ℓ'. Indeed, if the samples are shared, the terms are not independent of each other and the expectation value of the product of the terms is not the product of expectation values. For each case, we now detail the different possible values of $\mathbb{E}_\rho\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right]$ depending on the indices.

- If both k and k' do not equal ℓ or ℓ', *i.e.* for $N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)(N_{\mathrm{sh}}-2)(N_{\mathrm{sh}}-3)$ terms of the sum in (3.13),
$$\mathbb{E}_\rho\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right] = q^4 \langle P_i\rangle_\psi^2 \langle P_j\rangle_\psi^2.$$
- If $k \in \{1, \ldots, N_{\mathrm{sh}}\} \setminus \{\ell, \ell'\}$ and $k' \in \{\ell, \ell'\}$ or $k \in \{\ell, \ell'\}$ and $k' \in \{1, \ldots, N_{\mathrm{sh}}\} \setminus \{\ell, \ell'\}$, *i.e.* for $4N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)(N_{\mathrm{sh}}-2)$ terms of the sum in (3.13),
$$\mathbb{E}_\rho\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right] = q^3 \langle P_i\rangle_\psi \langle P_j\rangle_\psi \langle P_i P_j\rangle_\psi.$$
- If $k = \ell$ and $k' = \ell'$ or $k = \ell'$ and $k' = \ell$, *i.e.* for $2N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)$ terms of the sum in (3.13),
$$\mathbb{E}_\rho\left[\mathbf{p}_i^k \mathbf{p}_i^{k'} \mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right] = q^2 \langle P_i P_j\rangle_\psi^2.$$

Besides, the second term in (3.12) becomes

$$\begin{aligned}
&\mathbb{E}_\rho\left[\overline{P_i^2}^{N_{\mathrm{sh}}}\right] \mathbb{E}_\rho\left[\overline{P_j^2}^{N_{\mathrm{sh}}}\right] \\
&\qquad = \frac{1}{\left(N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)\right)^2} \sum_{k=1}^{N_{\mathrm{sh}}} \sum_{\substack{k'=1 \\ k'\neq k}}^{N_{\mathrm{sh}}} \sum_{\ell=1}^{N_{\mathrm{sh}}} \sum_{\substack{\ell'=1 \\ \ell'\neq \ell}}^{N_{\mathrm{sh}}} \mathbb{E}_\rho\left[\mathbf{p}_i^k \mathbf{p}_i^{k'}\right] \mathbb{E}_\rho\left[\mathbf{p}_j^\ell \mathbf{p}_j^{\ell'}\right] = q^4 \langle P_i\rangle_\psi^2 \langle P_j\rangle_\psi^2.
\end{aligned}$$

All in all, we eventually obtain (3.10). We get (3.11) similarly, using the additional fact that, for any $1 \leq i \leq N_{\mathcal{I}}$, $\langle P_i^2\rangle_\psi = 1$. □

3.3.2 Direct-sampling operator estimator

Let us now focus on the noise effect on the direct-sampling estimator. Consider $d \in H^1_{\#}(0,1)$ and the associated diagonal matrix $D \in \mathbb{C}^{2^N} \times \mathbb{C}^{2^N}$ such that the i-th diagonal coefficient D_i equals $d(i/2^N)$ for any i, $0 \leq i < 2^N$. Recall, from (2.59), that, given N_{sh} samples, the direct-sampling estimator $\check{D}^{N_{\text{sh}}}$ of D is given by

$$\check{D}^{N_{\text{sh}}} = \frac{1}{N_{\text{sh}}} \sum_{k=1}^{N_{\text{sh}}} D_{s_k} = \frac{1}{N_{\text{sh}}} \sum_{k=1}^{N_{\text{sh}}} \sum_{j=0}^{2^N-1} D_j \delta_{s_k,j}, \tag{3.14}$$

where $s_k \in \{0, \ldots, 2^N - 1\}$ corresponds to the k-th sample in the diagonal basis of A. Furthermore, in the noiseless case, the expected value of the estimator is given by

$$\mathbb{E}\left[\check{D}^{N_{\text{sh}}}\right]_\psi = \langle D \rangle_\psi = \langle v^2, d \rangle_{L^2_N},$$

and the variance of the estimator is given by

$$\operatorname{Var}_\psi \check{D}^{N_{\text{sh}}} = \frac{\langle d^2, v^2 \rangle_{L^2_N} - \langle d, v^2 \rangle^2_{L^2_N}}{N_{\text{sh}}}.$$

To determine the effects of noise on this estimator, let us now look at its effect on a single sample in the computational basis.

Proposition 18. *For a sample s in the computational basis, we have, for any j such that $0 \leq j < 2^N$, the mean of obtaining j for the sample is given by*

$$\mathbb{E}_\rho \left[\delta_{s,j}\right] = q p_j + \frac{1-q}{2^N}. \tag{3.15}$$

Proof. Using (3.4), we obtain

$$\mathbb{E}_\rho \left[\delta_{s,j}\right] = \operatorname{Tr}(|j\rangle\langle j| \, \rho) = q \operatorname{Tr}(|j\rangle\langle j| \, \rho_0) + (1-q) \operatorname{Tr}\left(|j\rangle\langle j| \, \frac{I}{2^N}\right) = q \mathbb{E}_\psi \left[\delta_{s,j}\right] + (1-q)\frac{1}{2^N},$$

and thus the result. □

These properties allow to investigate the noise effect on $\check{D}^{N_{\text{sh}}}$.

Lemma 8. *The mean of a direct-sampling operator $\check{D}^{N_{\text{sh}}}$ of a diagonal operator D under global depolarizing noise is given by*

$$\mathbb{E}_\rho \check{D}^{N_{\text{sh}}} = q \langle v^2, d \rangle_{L^2_N} + (1-q) \langle 1, d \rangle_{L^2_N},$$

and its variance by

$$\begin{aligned} \operatorname{Var}_\rho \check{D}^{N_{\text{sh}}} = \frac{q^2}{N_{\text{sh}}} \left[\langle v^2, d^2 \rangle_{L^2_N} - \langle v^2, d \rangle^2_{L^2_N}\right] + \frac{(1-q)^2}{N_{\text{sh}}} \left[\langle 1, d^2 \rangle_{L^2_N} - \langle 1, d \rangle^2_{L^2_N}\right] \\ + \frac{q(1-q)}{N_{\text{sh}}} \left[\langle v^2, d^2 \rangle_{L^2_N} + \langle 1, d^2 \rangle_{L^2_N} - 2 \langle v^2, d \rangle_{L^2_N} \langle 1, d \rangle_{L^2_N}\right]. \end{aligned} \tag{3.16}$$

Proof. From (3.14), and using (3.15), we obtain the mean value of the estimator

$$\mathbb{E}_\rho\left[\check{D}^{N_{\rm sh}}\right] = \frac{1}{N_{\rm sh}}\sum_{k=1}^{N_{\rm sh}}\sum_{j=0}^{2^N-1} D_j \mathbb{E}_\rho[\delta_{s_k,j}] = \frac{1}{N_{\rm sh}}\sum_{k=1}^{N_{\rm sh}}\sum_{j=0}^{2^N-1} D_j\left(qp_j + (1-q)\frac{1}{2^N}\right)$$
$$= q\langle v^2, d\rangle_{L^2_N} + (1-q)\langle 1, d\rangle_{L^2_N},$$

.

The variance is given by

$$\mathrm{Var}_\rho\,\check{D}^{N_{\rm sh}} = \frac{1}{N_{\rm sh}^2}\sum_{k=1}^{N_{\rm sh}}\sum_{\ell=1}^{N_{\rm sh}}\sum_{i=0}^{2^N-1}\sum_{j=0}^{2^N-1} D_iD_j\left(\mathbb{E}_\rho[\delta_{s_k,j}\delta_{s_\ell,i}] - \mathbb{E}_\rho[\delta_{s_k,j}]\mathbb{E}_\rho[\delta_{s_\ell,i}]\right)$$
$$= \frac{1}{N_{\rm sh}^2}\sum_{k=1}^{N_{\rm sh}}\sum_{j=0}^{2^N-1}\left(D_j^2\mathbb{E}_\rho[\delta_{s_k,j}] - \sum_{i=0}^{2^N-1} D_iD_j\mathbb{E}_\rho[\delta_{s_k,j}]\mathbb{E}_\rho[\delta_{s_\ell,i}]\right),$$

where we used the fact that, for any i, j such that $0 \le i, j \le 2^N-1$ and k such that $1 \le k \le N_{\rm sh}$, if $i \neq j$, then $\delta_{s_k,i}\delta_{s_k,j} = 0$. We rewrite this using (3.15), which gives

$$\mathrm{Var}_\rho\,\check{D}^{N_{\rm sh}} = \frac{1}{N_{\rm sh}}\sum_{j=0}^{2^N-1} D_j^2\left(qp_j + (1-q)h_N\right)$$
$$- \frac{1}{N_{\rm sh}}\sum_{j=0}^{2^N-1}\sum_{i=0}^{2^N-1} D_iD_j\left(q^2p_ip_j + q(1-q)(p_i+p_j)h_N + (1-q)^2h_N^2\right)$$

Noticing that, for $n \in \{1,2\}$, $\langle v^2, d^n\rangle_{L^2_N} = \sum_{i=0}^{2^N-1} p_iD_i^n$, and $\langle 1, d^n\rangle_{L^2_N} = \sum_{i=0}^{2^N-1} h_ND_i^n$, we obtain

$$\mathrm{Var}_\rho\,\check{D}^{N_{\rm sh}} = \frac{q^2}{N_{\rm sh}}\left(\langle v^2, d^2\rangle_{L^2_N} - \langle v^2, d\rangle^2_{L^2_N}\right)$$
$$+ \frac{(1-q)^2}{N_{\rm sh}}\left(\langle 1, d^2\rangle_{L^2_N} - \langle 1, d\rangle^2_{L^2_N}\right)$$
$$+ \frac{q(1-q)}{N_{\rm sh}}\left(\langle v^2, d^2\rangle_{L^2_N} + \langle 1, d^2\rangle_{L^2_N} - 2\langle v^2, d\rangle_{L^2_N}\langle 1, d\rangle_{L^2_N}\right),$$

and hence (3.16). □

With these results, we now consider the noise effects on the different energy terms of our method.

3.4 Errors for the Gross-Pitaevskii energy terms

Having determined the noise effects on generic estimators, let us now apply those pieces of information to our problem (2.9). For each energy term (2.7)–(2.8), the noise effect implies a bias in our estimators and changes the variance as a function of the number of gates in the associated

circuit. We now determine and compare this effect on the bias and variance of the different estimators.

3.4.1 Kinetic term

Let us first recall, from (2.20), that

$$\|\alpha_{1:}\|_1 = \frac{N}{2h_N^2}, \qquad \|\alpha_{1:}\|_2^2 = \frac{1}{2h_N^4}.$$

Using these, and the results of Proposition 16, let us investigate the effect of noise on our estimators.

For the Pauli-sampling estimator (2.41) of the kinetic term, recall that we consider an importance-sampling strategy. Hence, using Corollary 14 and the fact that

$$\mathrm{Tr}(\mathcal{K}^N)/2^N = 4^N, \tag{3.17}$$

from (2.16), we have the bias

$$b_{\mathrm{P}}^{\mathcal{K}} = \mathbb{E}\left[\overline{\mathcal{K}^N}^{N_{\mathrm{sh}}}\right] - \mathbb{E}_\rho\left[\overline{\mathcal{K}^N}^{N_{\mathrm{sh}}}\right] = (1-q)(\langle \mathcal{K}^N\rangle_\psi - 4^N),$$

and, for the variance, using (2.42) and Corollary 2, we get

$$\mathrm{Var}_\rho\, \overline{\mathcal{K}^N}^{N_{\mathrm{sh}}} = q^2\, \mathrm{Var}_\psi\, \overline{\mathcal{K}^N}^{N_{\mathrm{sh}}} + (1-q^2)\frac{16^N N^2}{2N_{\mathrm{sh}}}.$$

For the direct/diagonal-sampling method, we obtain, through Corollary 3,

$$b_{\mathrm{D}}^{\mathcal{K}} = (1-q)(\langle \mathcal{K}^N\rangle_\psi - 4^N).$$

And for the variance, using (2.60), we obtain

$$\mathrm{Var}_\rho\, \check{\mathcal{K}^N}^{N_{\mathrm{sh}}} = q^2\, \mathrm{Var}_\psi\, \check{\mathcal{K}^N}^{N_{\mathrm{sh}}} + (1-q)\frac{16^N}{2N_{\mathrm{sh}}} + (1-q)q\frac{\langle (\mathcal{K}^N - \alpha_0 I^N)^2\rangle_\psi}{N_{\mathrm{sh}}}.$$

and the associated bound, using the fact that

$$\|\mathcal{K}^N\|_\sigma = 4^{N+1},$$

we have

$$\mathrm{Var}_\rho\, \check{\mathcal{K}^N}^{N_{\mathrm{sh}}} \leq q^2\, \mathrm{Var}_\psi\, \check{\mathcal{K}^N}^{N_{\mathrm{sh}}} + (1-q)\frac{16^N}{2N_{\mathrm{sh}}} + (1-q)q\frac{9\cdot 16^N}{N_{\mathrm{sh}}}.$$

Finally for the Hadamard-test method, for the bias we have, using Proposition 20,

$$b_{\mathrm{H}}^{\mathcal{K}} = \mathbb{E}[\widehat{\mathcal{K}^N}^{N_{\mathrm{sh}}}] - \mathbb{E}_\rho[\widehat{\mathcal{K}^N}^{N_{\mathrm{sh}}}] = (1-q)(\langle \mathcal{K}^N\rangle_\psi - 4^N),$$

and, for the variance, we obtain, through (2.65),

$$\mathrm{Var}_\rho\, \widehat{\mathcal{K}^N}^{N_{\mathrm{sh}}} \leq q^2\, \mathrm{Var}_\psi\, \widehat{\mathcal{K}^N}^{N_{\mathrm{sh}}} + (1-q^2)\frac{16^N}{N_{\mathrm{sh}}}.$$

3.4.2 Potential term

For the potential term, we first look at the Walsh-Pauli truncation estimator (2.43), with a diagonal-sampling strategy. Using Proposition 14, we obtain that the bias term is equal to

$$b_{\mathrm{P}}^{\mathcal{P}} = \mathbb{E}[\overline{\mathcal{P}^N}^{N_{\mathrm{sh}}}] - \mathbb{E}_\rho[\overline{\mathcal{P}^N}^{N_{\mathrm{sh}}}] = (1-q)(\langle \mathcal{P}^N \rangle_\psi - \|V\|_{L^1_N}),$$

where

$$\mathrm{Tr}(\mathcal{P}^N)/2^N = \|V\|_{L^1_N} = \frac{V_0}{12}. \tag{3.18}$$

And, for the variance, using (2.44) and Corollary 3, we have

$$\mathrm{Var}_\rho\, \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} = q^2\, \mathrm{Var}_\psi\, \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} + (1-q)\frac{V_0^2 + O(\frac{1}{4^N})}{180 N_{\mathrm{sh}}} + (1-q)q\frac{\langle (V-\alpha_0 I^N)^2\rangle_\psi}{N_{\mathrm{sh}}}.$$

Furthermore, as $\|V\|_\infty - \frac{V_0}{12} = \frac{V_0}{6}$, we obtain

$$\mathrm{Var}_\rho\, \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} \leq q^2\, \mathrm{Var}_\psi\, \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} + (1-q)\frac{V_0^2}{180 N_{\mathrm{sh}}} + (1-q)q\frac{V_0^2}{36 N_{\mathrm{sh}}}.$$

Similarly, for the direct-sampling, or Walsh interpolation estimator, we obtain the bias

$$b_{\mathrm{P}}^{\mathcal{P}} = \mathbb{E}[\overline{\mathcal{P}^N}^{N_{\mathrm{sh}}}] - \mathbb{E}_\rho[\overline{\mathcal{P}^N}^{N_{\mathrm{sh}}}] = (1-q)(\langle \mathcal{P}^N \rangle_\psi - 4^N).$$

And, for the variance, using (2.60), we have

$$\mathrm{Var}_\rho\, \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} = q^2\, \mathrm{Var}_\psi\, \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} + (1-q)\frac{V_0^2 + O(\frac{1}{4^N})}{180 N_{\mathrm{sh}}} + (1-q)q\frac{\langle (V-\alpha_0 I^N)^2\rangle_\psi}{N_{\mathrm{sh}}}.$$

Similarly, as $\|V\|_\infty - \frac{V_0}{12} = \frac{V_0}{6}$, we obtain

$$\mathrm{Var}_\rho\, \overline{\mathcal{P}^N}^{N_{\mathrm{sh}}} \leq (1-q+q^2)(\frac{V_0^2}{180}) + (1-q)q\frac{V_0^2}{36} + O(\frac{1}{4^N}).$$

3.4.3 Interaction term

Due to the non-linearity, the Pauli/direct-sampling estimator (2.52) of the interaction term is not affected in the same way as the linear operators. Let us now determine how depolarization noise affects it.

Proposition 19. *The bias of the interaction term Pauli estimator using a diagonal-sampling strategy is given by*

$$b_{\mathrm{P},\mathcal{I}} = (1-q^2)|\langle \mathcal{I} - \alpha_0 I^N \rangle_\psi|, \tag{3.19}$$

and its variance by

$$\begin{aligned}\mathrm{Var}_\rho\, \overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} &= q^4\, \mathrm{Var}_\psi\, \overline{\mathcal{I}^N}^{N_{\mathrm{sh}}} + (1-q^4)\frac{2(2^N-1)\alpha_0^2}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)} \\ &\quad + q^3(1-q)\frac{4(N_{\mathrm{sh}}-2)}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)}\alpha_0^2\left(\langle v^4, v^2\rangle_{L^2_N} - 3\langle v^2, v^2\rangle_{L^2_N} + 2\langle v^2, 1\rangle^2_{L^2_N}\right)\end{aligned}$$

$$+ q^2(1-q^2)\frac{4(N_{\rm sh}-2)+2(2^N-2)}{N_{\rm sh}(N_{\rm sh}-1)}\alpha_0^2\left(\langle v^2,v^2\rangle_{L^2_N}-\langle v^2,1\rangle^2_{L^2_N}\right), \quad (3.20)$$

which are both positive terms. We also have the upper bound

$$\mathrm{Var}_\rho\,\overline{\mathcal{I}^N}^{N_{\rm sh}} \leq \frac{B_{V,q}}{N_{\rm sh}}+\frac{B_{W,q}}{N_{\rm sh}(N_{\rm sh}-1)}, \quad (3.21)$$

where we set

$$B_{V,q} = q^4\frac{27}{64}\alpha_0^2 4^N + 4\alpha_0^2q^2(2^N-1)\left(1+(2^N-2)q-q^2(2^N-3)\right), \quad (3.22)$$

$$B_{W,q} = q^4\frac{\alpha_0^2 4^N}{8} + 2\alpha_0^2q^2(2^N-1)(2^N-2)(1-q)^2 + 2\alpha_0^2(2^N-1)(1-q^2)^2, \quad (3.23)$$

which are both positive terms as well.

Proof. The bias is given by

$$b_{\mathrm{P},\mathcal{I}} = \left|\mathbb{E}_\rho\left[\overline{\mathcal{I}^N}^{N_{\rm sh}}\right]-\mathbb{E}\left[\overline{\mathcal{I}^N}^{N_{\rm sh}}\right]\right| = \left|\sum_{i=1}^{N_\mathcal{I}}\mathbb{E}_\rho\left[\overline{P_i^2}^{N_{\rm sh}}\right]-\mathbb{E}\left[\overline{P_i^2}^{N_{\rm sh}}\right]\right|$$

and thus the result on the bias term is straightforward using (3.9). For the variance, the results are almost immediate using the same methods as in Proposition 13. We have

$$\begin{aligned}\mathrm{Var}_\rho\,\overline{\mathcal{I}^N}^{N_{\rm sh}} &= \sum_{i=1}^{N_\mathcal{I}}\mathrm{Var}_\rho\,\overline{P_i^2}^{N_{\rm sh}} + \sum_{i=1}^{N_\mathcal{I}}\sum_{j\neq i}\mathrm{Cov}_\rho(\overline{P_i^2}^{N_{\rm sh}},\overline{P_j^2}^{N_{\rm sh}})\\ &= q^4\frac{6-4N_{\rm sh}}{N_{\rm sh}(N_{\rm sh}-1)}\sum_{i=1}^{N_\mathcal{I}}\sum_{j\neq i}\alpha_i\alpha_j\langle P_i\rangle_\psi^2\langle P_j\rangle_\psi^2 + q^4\frac{6-4N_{\rm sh}}{N_{\rm sh}(N_{\rm sh}-1)}\sum_{i=1}^{N_\mathcal{I}}\alpha_i^2\langle P_i\rangle_\psi^4\\ &+ q^3\frac{4(N_{\rm sh}-2)}{N_{\rm sh}(N_{\rm sh}-1)}\sum_{i=1}^{N_\mathcal{I}}\sum_{j\neq i}\alpha_i\alpha_j\langle P_i\rangle_\psi\langle P_j\rangle_\psi\langle P_iP_j\rangle_\psi + q^2\frac{2}{N_{\rm sh}(N_{\rm sh}-1)}\sum_{i=1}^{N_\mathcal{I}}\sum_{j\neq i}\alpha_i\alpha_j\langle P_iP_j\rangle_\psi^2\\ &+ q^2\frac{4(N_{\rm sh}-2)}{N_{\rm sh}(N_{\rm sh}-1)}\sum_{i=1}^{N_\mathcal{I}}\alpha_i^2\langle P_i\rangle_\psi^2 + \frac{2N_\mathcal{I}\alpha_0^2}{N_{\rm sh}(N_{\rm sh}-1)}.\end{aligned}$$

using (3.10) and (3.11) for the second equation. Using (2.54)–(2.56), from the proof of the variance formula (2.53) in the noiseless case, we obtain

$$\begin{aligned}\mathrm{Var}_\rho\,\overline{\mathcal{I}^N}^{N_{\rm sh}} &= q^4\frac{6-4N_{\rm sh}}{N_{\rm sh}(N_{\rm sh}-1)}\alpha_0^2\left(\langle v^2,v^2\rangle^2_{L^2_N}-2\langle v^2,v^2\rangle_{L^2_N}\langle v^2,1\rangle^2_{L^2_N}+\langle v^2,1\rangle^4_{L^2_N}\right)\\ &+ q^3\frac{4(N_{\rm sh}-2)}{N_{\rm sh}(N_{\rm sh}-1)}\alpha_0^2\left(\langle v^4,v^2\rangle_{L^2_N}-3\langle v^2,v^2\rangle_{L^2_N}+2\langle v^2,1\rangle^2_{L^2_N}\right)\\ &+ q^2\frac{4(N_{\rm sh}-2)+2(2^N-2)}{N_{\rm sh}(N_{\rm sh}-1)}\alpha_0^2\left(\langle v^2,v^2\rangle_{L^2_N}-\langle v^2,1\rangle^2_{L^2_N}\right)+\frac{2(2^N-1)\alpha_0^2}{N_{\rm sh}(N_{\rm sh}-1)}.\end{aligned}$$

By rewriting the equation to make the noiseless variance (2.53) appear, we obtain (3.20).

To obtain the bound (3.21), we use (2.58) and the fact that $\|v^2\|^2_{L^2}\leq 2^N\|v\|^2_{L^2}$, and get

$$\mathrm{Var}_\rho(\overline{\mathcal{I}^N}^{N_{\mathrm{sh}}}) \leq \frac{q^4 B_V + 4\alpha_0^2(2^N-1)\left(q^2+(2^N-2)q^3-q^4(2^N-3)\right)}{N_{\mathrm{sh}}} + \frac{q^4 B_W + \alpha_0^2(2^N-1)(2^N-2)(2q^2-4q^3+2q^4)+\alpha_0^2(2^N-1)(2-4q^2+2q^4)}{N_{\mathrm{sh}}(N_{\mathrm{sh}}-1)},$$

with B_V and B_W from the noiseless variance (2.57) and where the two terms are shown to be positive using the fact that $q < 1$. With the additional bounds

$$B_V \leq \frac{27}{64}\alpha_0^2 4^N \qquad B_W \leq \frac{1}{8}\alpha_0^2 4^N,$$

we get the result. □

To obtain a bound on the number of samples required to achieve an error ϵ, using (2.68), we can choose

$$N_{\mathrm{sh}}^{\epsilon,q} = \frac{1}{2}\left(1 + \frac{B_{V,q}}{\epsilon^2} + \sqrt{(1-\frac{B_{V,q}}{\epsilon^2})^2 + 4\frac{B_{W,q}}{\epsilon^2}}\right).$$

3.4.4 Hadamard-test estimator

As a Hadamard-test consists in the measurement of a Pauli operator, we have using Proposition 15, for a single Hadamard-test sample $\mathbf{h}$,

$$\mathbb{E}_\rho[\mathbf{h}] = \langle Z_1\rangle_{\rho,\text{H-test}} = q\langle Z_1\rangle_{\psi,\text{H-test}},$$

where N_{G_1} and N_{G_2} the number of 1-qubit and 2-qubit gates respectively, in the Hadamard-test circuit. This gives us the equivalent of Theorem 11 in the presence of noise.

Proposition 20. *The mean value of the Hadamard-test estimator $\widehat{A}^{N_{\mathrm{sh}}}$ for an operator A under the effect of noise is given by*

$$\mathbb{E}_\rho\left[\widehat{A}^{N_{\mathrm{sh}}}\right] = c_{\mathrm{H}} + \alpha_{\mathrm{H}}\sum_{k=1}^{N_{\mathrm{sh}}}\mathbb{E}_\rho[\mathbf{h}_k] = (1-q)c_{\mathrm{H}} + q\langle A\rangle_\psi. \tag{3.24}$$

Its variance is given by

$$\mathrm{Var}_\rho\,\widehat{A}^{N_{\mathrm{sh}}} = q^2\,\mathrm{Var}_\psi\,\widehat{A}^{N_{\mathrm{sh}}} + (1-q^2)\frac{\alpha_{\mathrm{H}}^2}{N_{\mathrm{sh}}}, \tag{3.25}$$

where we denote the noiseless variance (2.65) *by* $\mathrm{Var}_\psi\,\widehat{A}^{N_{\mathrm{sh}}}$. *The variance is then bounded by*

$$\mathrm{Var}_\rho\,\widehat{A}^{N_{\mathrm{sh}}} \leq \frac{\alpha_{\mathrm{H}}^2}{N_{\mathrm{sh}}}. \tag{3.26}$$

Proof. Equation (3.24) is straightforward from (3.5). For (3.25), we write

$$\mathrm{Var}_\rho\,\widehat{A}^{N_{\mathrm{sh}}} = \frac{\alpha_{\mathrm{H}}^2}{N_{\mathrm{sh}}^2}\sum_{k=1}^{N_{\mathrm{sh}}}\sum_{\ell=1}^{N_{\mathrm{sh}}}\mathbb{E}_\rho[\mathbf{h}_k\mathbf{h}_\ell] - \mathbb{E}_\rho[\mathbf{h}_k]\mathbb{E}[\mathbf{h}_\ell] = \frac{\alpha_{\mathrm{H}}^2}{N_{\mathrm{sh}}}\left(1-q^2\langle Z_1\rangle_{\psi,\text{H-test}}^2\right),$$

where we notice that, for any $1 \leq k \leq N_{\mathrm{sh}}$, $\mathbb{E}_\rho[\mathbf{h}_k^2] = 1$ in the second equality. Using (2.65), the previous equation gives (3.25) and (3.26). □

Table 3.1 – Variance bound of estimators under depolarizing noise

$\epsilon^2 N_{\mathrm{sh}}^{\epsilon}$ Var	Pauli sampling estimator	Direct sampling estimator	Hadamard-test estimator
$\mathcal{K}$	$\frac{N^2}{4} \cdot 16^N$	$\left(\frac{1}{8}q^2 + (1-q)\left(\frac{1}{2} + 9q\right)\right) 16^N$	16^N
$\mathcal{P}$	$(1 + 4q - 4q^2)\frac{V_0^2}{180} + O(\frac{1}{4^N})$		$2^N \frac{V_0^2}{80} + O(\frac{1}{2^N})$
$\mathcal{I}$	$\frac{1}{2}\left(1 + \frac{B_{V,q}}{\epsilon^2} + \sqrt{(1 - \frac{B_{V,q}}{\epsilon^2})^2 + 4\frac{B_{W,q}}{\epsilon^2}}\right)$		$\alpha_0^2 4^N$

We now have the mandatory properties to determine the effect of noise on our energy estimators. However, the value of q depends on the circuit considered. Hence, we now go into more details on the ansatz models used to represent quantum states, and the number of gates in the circuits associated to them and to the estimation circuits.

3.5 Circuit size estimation

In this section, we provide rough estimates of the circuit size for the various energy estimation methods. We consider two ansatz models, one based on tensor trains and able to approximate any quantum state, and one constructed for its ease of implementation and ability to achieve fast entanglement in the system. We then estimate the size of the circuits associated to the Quantum Fourier Transform and the Hadamard-test circuits.

3.5.1 Ansatz models and related circuits

A quantum state $|\psi\rangle = \sum_{i=0}^{2^N-1} \psi_i |i\rangle$ in a N-qubit system can be represented as a tensor on each component tensor space as

$$|\psi\rangle = \Psi_{i_1,\dots,i_N} |i_1\rangle \dots |i_N\rangle .$$

As for any tensor, $\Psi_{i_1,\dots,i_N}$ can be represented as a tensor train (TT) [13], also known as a matrix product state (MPS),

$$\Psi_{i_1,\dots,i_N} = \sum_{\alpha_1=1}^{\chi_1} \cdots \sum_{\alpha_{N-1}=1}^{\chi_{N-1}} (\Psi^{(1)})_{i_1}^{\alpha_1} (\Psi^{(2)})_{\alpha_1,i_2}^{\alpha_2} \dots (\Psi^{(N-1)})_{i_{N-1},\alpha_{N-2}}^{\alpha_{N-1}} (\Psi^{(N)})_{\alpha_{N-1},i_N},$$

where, for $1 \leq k < N$, $\chi_k \in \mathbb{N}^*$ and, for $1 < \ell < N$, $\Psi^{(k)} \in \mathbb{C}^{2\times\chi_\ell\times\chi_{\ell-1}}$ and $\Psi^{(1)} \in \mathbb{C}^{2\times\chi_1}$, $\Psi^{(N)} \in \mathbb{C}^{2\times\chi_{N-1}}$.

The minimum bond dimensions attainable for a given state $|\psi\rangle$ depends on the entanglement of the system for this state. Furthermore, it can be shown [78], by an inductive argument on the rank of each tensor $\Psi^{(k)}$ starting from both ends and stopping in the middle, that, for any state, we can bound the bond dimension χ_k, for any $1 \leq k < N$, by $\chi_k \leq 2^{\lceil \frac{N}{2} - k \rceil}$. In the context of quantum computing, 1-qubit gates are local and do not affect the bond dimension of a quantum state, whereas 2-qubit gates can at most double the bond dimension of a quantum state on the subsystem on which they act. We now describe two methods that may be used to represent states achieving a bond dimension $\chi_{\max}$ at most.

U2 Disentangler ansatz The first one, proposed in [83] and refined in [84], can approximately represent any tensor train by applying N_{lay} successive layers of parametrized $(N-1)$ universal 2-qubit gates (see Figure 3.3) in a staircase structure as in Figure 3.2. Each layer is parametrized to maximize the fidelity $|\langle 0| U_D^\dagger |\psi\rangle|^2$, hence the moniker disentangler, as it sends $|\psi\rangle$ to $|0\rangle$. By looking at the circuit as an ansatz, we see that each layer is able, at most, to double the bond dimension of the resulting state. Consequently, the resulting bond dimension is at most $2^{N_{\text{lay}}}$. The entire circuit is thus composed of N_{lay} of $(N-1)$ universal gates. Each gate being composed of 15 1-qubit gates and 3 2-qubit gates, we have

$$N_{G_1}^D = 15N_{\text{lay}}(N-1), \qquad N_{G_2}^D = 3N_{\text{lay}}(N-1).$$

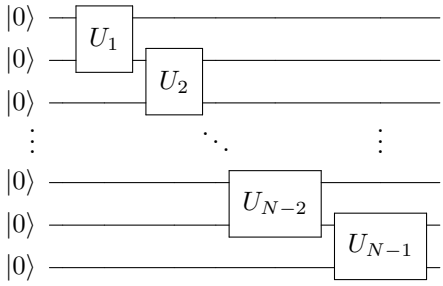


Figure 3.2 – Single layer of universal 2-qubit gates.

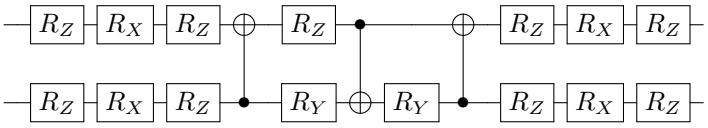


Figure 3.3 – Universal 2-qubit gate from [93].

Hardware-efficient ansatz Another popular choice of ansatz, the hardware-efficient ansatz consists of layers of alternating CNOT gates and Pauli rotation gates as in Figure 3.4. For this ansatz model, we have $N_{\text{lay}}+1$ layers of Pauli rotation gates, and N_{lay} layers of CNOT walls. This gives us

$$N_{G_1}^{\text{HE}} = (N_{\text{lay}}+1)N, \qquad N_{G_2}^{\text{HE}} = N_{\text{lay}}(N-1).$$

Controlled ansatz For the controlled ansatz, which appears in the Hadamard-test circuits of the potential and interaction terms, see Figures 2.8 and 2.9, only the 1-qubit gates need to be controlled, as the ansatz is applied to $|0\rangle$, and thus CNOT leaves the state unchanged if the control is not set to 1. We shall thus consider that, in the case of controlled ansatz, the number of 2-qubit gates is the sum of 1-qubit and 2-qubit gates in the regular ansatz.

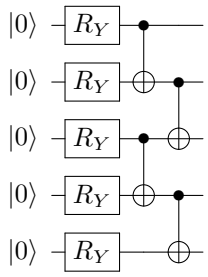


Figure 3.4 – Single layer of Hardware-efficient ansatz for a 5-qubit system.

3.5.2 Energy estimation subroutine circuits

Now, let us provide the number of gates used for the quantum routines involved in the energy estimation circuits, namely the Quantum Fourier Transform circuit, used for the diagonal-sampling Pauli estimator of the kinetic term, and Adder circuit, which appears in the Hadamard-test for the kinetic term.

Quantum Fourier transform circuit

The basis change used to measure the kinetic term in the direct-sampling/diagonal-sampling strategy is obtained through the quantum Fourier transform. The QFT circuit consists of $N(N-1)/2$ 2-qubit gates and N 1-qubit gates, see [73] for more details. An example for 4 qubits is given in Figure 3.5.

$$N_{G_1}^{\mathrm{QFT}} = N, \qquad N_{G_2}^{\mathrm{QFT}} = N(N-1)/2$$

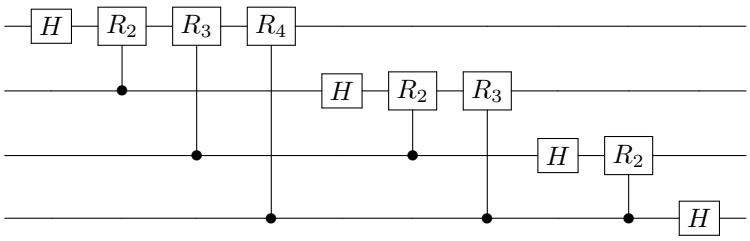


Figure 3.5 – Four qubit quantum Fourier transform. We set, for $k \geq 2$, $R_k = RZ(\pi/2^k)$ with $RZ(\theta) = \exp(-i\theta Z/2)$ the Pauli Z rotation of angle θ.

Adder circuit

The kinetic term Hadamard-test circuit consists of a controlled 1-adder. While some ripple carry adders may boast of a number of gates increasing linearly with N, see [87], the prefactor of N is prohibitively large. We instead consider here a QFT-based adder which consists, in our case, in using a QFT to change basis, applying N 1-qubit Pauli rotations in the Fourier basis and

applying an inverse QFT to come back to the computational basis. Then, the total number of gates is

$$N_{G_1}^{\text{Add}} = 3N, \qquad N_{G_2}^{\text{Add}} = N(N-1).$$

For a rudimentary implementation of a controlled adder the 1-qubit gates become 2-qubit gates and the 2-qubit gates become twice-controlled Pauli rotations. This specific type of twice controlled rotations can be implemented from two CNOT gates and three controlled rotation gates using half the angle of the original rotation, see [73, pp. 181–182]. Hence, we have

$$N_{G_1}^{C\text{-Add}} = 0, \qquad N_{G_2}^{C\text{-Add}} = 5N(N-1) + 3N.$$

Using these circuits, we now estimate the total number of gates for each energy term, ansatz and estimation method.

3.5.3 Circuit size estimates

For the ansatz circuits, as our quantum states represent functions, we can bound the maximum bond dimension achieved in our U2 ansatz, and thus limit its number of layers. Indeed, in the one-dimensional case [3, 38, 75], functions regular enough can be represented through matrix product states with relatively low bond dimensions. We consider four layers for the U2 ansatz and eight for the HE ansatz. Using the number of gates estimated in the previous subsections for each circuit, we show in Table 3.2, the number of estimated gates for each method and ansatz, for each energy term.

3.6 Minimum achievable error and maximum number of usable gates

Having obtained the number of gates in each case, we now determine, for a noisy system, the maximum number of gates usable without mitigation. As we have seen in the previous sections, the quantum noise adds a bias to our estimators as well as changing their variance.

In this subsection, let us focus on the bias term since, without mitigation, it provides the lowest bound for the accuracy of the energy estimation methods.

For estimators using observable measurements, in our case the direct/diagonal-sampling and importance-sampling estimators, from Proposition 18 and Corollary 3 for direct-sampling, and Corollary 2 for importance-sampling, we can see that the effect of noise on the mean value of the estimator is to contract its traceless part by the factor

$$q = (1-p_1)^{N_{G_1}}(1-p_2)^{N_{G_2}}.$$

Thus the noise-induced bias for the linear energy operators, *i.e.* if $O = \mathcal{K}^N$ or $O = \mathcal{P}^N$, has the form

$$b_{O,\mathrm{P}} = \left|\mathbb{E}_\rho\left[\overline{O}^{N_{\text{sh}}}\right] - \mathbb{E}_\psi\left[\overline{O}^{N_{\text{sh}}}\right]\right| = (1-q)\left|\langle O\rangle_\psi - \frac{\operatorname{Tr} O}{2^N}\right|. \tag{3.27}$$

Let us now denote by $\|O\|_\sigma$ the spectral norm of O. Since, in these cases, O only has positive eigenvalues, including $\|O\|_\sigma$. Obviously, $\operatorname{Tr} O$ and $\langle O\rangle_\psi$ are subsequently positive too. And we also have

$$\|O\|_\sigma - \frac{\operatorname{Tr} O}{2^N} \geq 0,$$

because of the definition of $\|O\|_\sigma$.

Table 3.2 – Circuit size estimation

Disentangler ansatz	$\mathcal{K}$	$\mathcal{P}$	$\mathcal{I}$
Pauli - D N_{G_1}	$15N_{\text{lay}}(N-1)+N$	$15N_{\text{lay}}(N-1)$,	$15N_{\text{lay}}(N-1)$
Pauli - D N_{G_2}	$3N_{\text{lay}}(N-1)$	$3N_{\text{lay}}(N-1)$	$3N_{\text{lay}}(N-1)$
Direct - D N_{G_1}	$15N_{\text{lay}}(N-1)+N$	$15N_{\text{lay}}(N-1)$	$15N_{\text{lay}}(N-1)$
Direct - D N_{G_2}	$(3N_{\text{lay}}+N/2)(N-1)$	$3N_{\text{lay}}(N-1)$	$3N_{\text{lay}}(N-1)$
H-test - D N_{G_1}	$15N_{\text{lay}}(N-1)+2$	$15N_{\text{lay}}(N-1)+2$	$15N_{\text{lay}}(N-1)+2$
H-test - D N_{G_2}	$(3N_{\text{lay}}+5N)(N-1)+3N$	$(21N_{\text{lay}})(N-1)+N$	$(39N_{\text{lay}})(N-1)+2N$
Hardware-efficient ansatz	$\mathcal{K}$	$\mathcal{P}$	$\mathcal{I}$
Pauli - HE N_{G_1}	$(N_{\text{lay}}+2)N$	$(N_{\text{lay}}+1)N$	$(N_{\text{lay}}+1)N$
Pauli - HE N_{G_2}	$N_{\text{lay}}(N-1)$	$N_{\text{lay}}(N-1)$	$N_{\text{lay}}(N-1)$
Direct - HE N_{G_1}	$(N_{\text{lay}}+2)N$	$(N_{\text{lay}}+1)N$	$(N_{\text{lay}}+1)N$
Direct - HE N_{G_2}	$(N_{\text{lay}}+N/2)(N-1)$	$N_{\text{lay}}(N-1)$	$N_{\text{lay}}(N-1)$
H-test - HE N_{G_1}	$(N_{\text{lay}}+1)N+2$	$N_{\text{lay}}N+2$	$(N_{\text{lay}}+1)N+2$
H-test - HE N_{G_2}	$(N_{\text{lay}}+5N)(N-1)+3N$	$N_{\text{lay}}(3N-2)+3N$	$N_{\text{lay}}(5N-3)+5N$

All those remarks imply that

$$\left|\langle O\rangle_\psi - \frac{\operatorname{Tr} O}{2^N}\right| \leq \max\left(\|O\|_\sigma - \frac{\operatorname{Tr} O}{2^N}, \frac{\operatorname{Tr} O}{2^N}\right).$$

Then, thanks to (3.17) and (3.18), we notice that

$$\frac{\operatorname{Tr} O}{2^N} \leq \|O\|_\sigma - \frac{\operatorname{Tr} O}{2^N},$$

which eventually leads to the upper bound

$$b_{O,\mathrm{P}} \leq (1-q)\left(\|O\|_\sigma - \frac{\operatorname{Tr} O}{2^N}\right). \tag{3.28}$$

Hence, in the worst case, the bias is obtained through the spectral norm, or Schatten-∞ norm [97], of O, *i.e.* its largest singular value.

Lastly, for the interaction term, using (3.19), we get a similar result

$$b_{\mathcal{I},\mathrm{P}} \leq (1-q^2)\left|\max_{|\psi\rangle}(\langle \mathcal{I}^N\rangle_\psi) - \frac{\kappa}{2}\right| = \frac{\kappa}{2}(1-q^2)(2^N-1).$$

For unitary measurements, for the Hadamard-test estimator, the noise contraction does not occur on the traceless part of the observable. Instead, the mean value of the measured Pauli operator is contracted and the mean value of the unitary itself is contracted in the estimator, as we can see in Proposition 20. This implies that the bias for these estimators is instead given by

$$b_{O,\mathrm{H}} = \left|\mathbb{E}_\rho[\widehat{O}^{N_{\mathrm{sh}}}] - \mathbb{E}_\psi[\widehat{O}^{N_{\mathrm{sh}}}]\right| = (1-q)\left|\langle O\rangle_\psi - c_{\mathrm{H}}\right|,$$

which depending on the value of $\langle O\rangle_\psi$ may provide a lower or higher bias.

The mean squared error of our estimators is given by the sum of the squared bias and the variance of the estimators

$$\mathrm{MSE}_\rho(\overline{O}^{N_{\mathrm{sh}}}) = b_{O,\mathrm{P}}^2 + \mathrm{Var}_\rho\, \overline{O}^{N_{\mathrm{sh}}}.$$

Thus, a hard limit is set on our MSE by the bias term if we do not mitigate its effects (as we shall see in the next section). We now show that, for a given maximum acceptable error $\epsilon > 0$ and a given depolarizing probabilities p_1 and p_2, the bias term allows to quantify the maximum number of gates we can use.

First, we upper bound our mean squared error by ϵ^2,

$$b_{O,\mathrm{P}}^2 + \mathrm{Var}_\rho\, \overline{O}^{N_{\mathrm{sh}}} \leq \epsilon^2.$$

Since the variance is always positive, we thus have $b_{O,\mathrm{P}} \leq \epsilon$ and, using the formula (3.27) for the bias

$$q \geq 1 - \frac{\epsilon}{\left|\langle O\rangle_\psi - \frac{\mathrm{Tr}\, O}{2^N}\right|},$$

and thanks to (3.28),

$$q \geq 1 - \frac{\epsilon}{\|O\|_\sigma - \frac{\mathrm{Tr}\, O}{2^N}}. \tag{3.29}$$

Let us rewrite the bias term as a function of p_2 only by setting, for a constant $c > 0$, $p_1 = cp_2$. We have

$$N_{G_1} \ln(1-p_1) = N_{G_1} \frac{\ln(1-cp_2)}{\ln(1-p_2)} \ln(1-p_2).$$

We hence set $\zeta_c(p_2) = \frac{\ln(1-cp_2)}{\ln(1-p_2)}$, which allows to write

$$q = (1-p_1)^{N_{G_1}} (1-p_2)^{N_{G_2}} = (1-p_2)^{\zeta_c(p_2) N_{G_1} + N_{G_2}}.$$

We set $N_{G_2}^{\mathrm{equiv}} = \zeta_c(p_2) N_{G_1} + N_{G_2}$, and rewrite inequality (3.29) in terms of number of gates

$$N_{G_2}^{\mathrm{equiv}} \leq \frac{\ln\left(1 - \dfrac{\epsilon}{\|O\|_\sigma - 2^{-N}\,\mathrm{Tr}\, O}\right)}{\ln(1-p_2)}.$$

Hence, for a given observable, the depolarizing noise probability p_2 allows to upper-bound the number of gates usable in our circuit. Let us now estimate the number of 2-qubit gates usable for near-term devices. For current devices, the best error rates for 1-qubit gates are of the order of $\mathfrak{e}_1 = 0.02\%$ and for 2-qubit gates of the order of $\mathfrak{e}_2 = 0.1$ [98]. The gate fidelity is defined as the average fidelity, over input states, of the output of the implemented gate versus its expected output. For depolarization noise models, all states are equally affected, this implies that the fidelity of a gate is equal to its fidelity for any input state. For a N-qubit system, a gate with

depolarizing probability p has a fidelity

$$F(p) = \langle\psi|\,\rho\,|\psi\rangle = \langle\psi|\left((1-p)\,|\psi\rangle\langle\psi| + p\frac{I_N}{2^N}\right)|\psi\rangle = (1-p) + \frac{p}{2^N} = 1 - p\frac{2^N-1}{2^N},$$

where $|\psi\rangle$ is the ideal state and ρ is the output state of the gate. As gate error rates are defined as $1 - F(p)$, we thus have our relationship betwen p and the error rate $\mathfrak{e}$

$$p = \frac{2^N}{2^N - 1}\mathfrak{e}.$$

The depolarization probability p thus quickly converges to $\mathfrak{e}$ when N increases. By considering

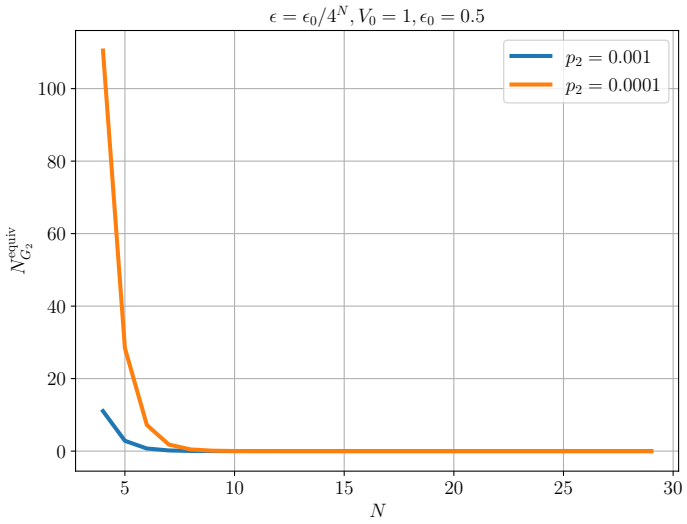


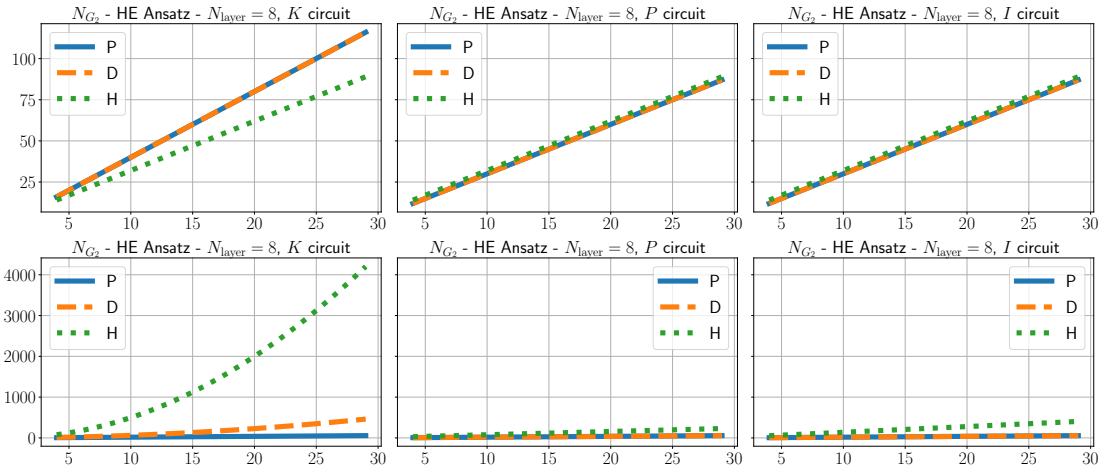
Figure 3.6 – Maximum number of 2-qubit gates usable under depolarization noise to achieve a bias under $\epsilon_N = \epsilon/2^N$.

the potential energy term whose norm is the only one to converge with respect to N, we obtain estimates of the number of usable gates for current hardware and for future hardware with a tenth of the error rates which are shown in Figure 3.6. We also show the number of required gates simply for the energy estimation circuits for each ansatz in Figure 3.7.

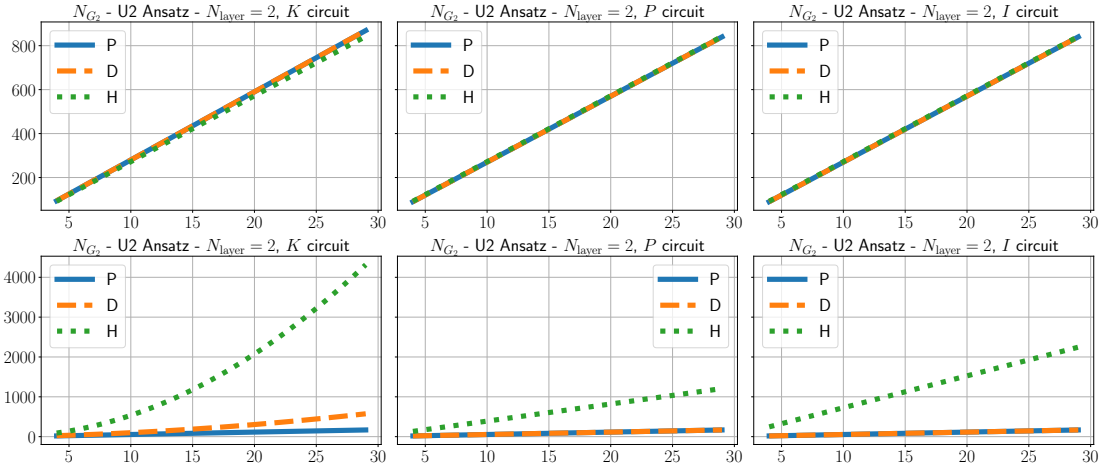
Obviously, even for the potential term, the number of gates is prohibitively low. Hence, the only resort is quantum error mitigation.

3.7 Mitigation

Quantum error mitigation [14] is the process of recovering information from the measurement of noisy quantum circuits through classical post-processing. In the case of zero-noise extrapolation [37, 55, 56, 59, 89], the aim is to reconstruct a unbiased estimator from noisy samples using knowledge of the effects of noise on the quantum system, and of the considered circuits. This mitigated estimator however requires a higher number of samples to achieve an error ϵ. This can



(a) Hardware efficient ansatz with 8 layers



(b) 2-qubit universal ansatz with 2 layers

Figure 3.7 – Number of gates in the energy estimation circuits.

be seen as the fact that the effect of noise is to contract the range of our estimators, and thus an error ϵ on the noisy estimator will be multiplied by the inverse of the contraction coefficient in the mitigated estimator.

In the general case, as shown in [14] using Hoeffding's inequality, the sampling overhead required to compensate the effect of noise is given by

$$C_q = \frac{\text{Var}_\psi(O)}{\text{Var}_\rho(O)}.$$

This overhead can be upper-bounded by the squared ratio of ranges for the expected values of O

in the absence and presence of noise, *i.e.*

$$C_q \leq \frac{R[\langle O\rangle_\psi]^2}{R[\langle O\rangle_\rho]^2}. \tag{3.30}$$

3.7.1 Mitigated estimators under depolarizing noise

In our case, as we have a noise model and explicit formulae of the variance of our estimators under noise, we can recover a finer bound. Let us start by noticing that to reconstruct $\langle O\rangle_\psi$ from $\langle O\rangle_\rho$, we can proceed as follows. Recall that

$$\langle O\rangle_\rho = \alpha_0 + q(\langle O\rangle_\psi - \alpha_0).$$

The effect of noise can be mitigated by reconstructing $\langle O\rangle_\psi$ from $\langle O\rangle_\rho$, from the above equality, *i.e.*

$$\langle O\rangle_\psi = \frac{1}{q}(\langle O\rangle_\rho - \alpha_0) + \alpha_0.$$

Our mitigated Pauli estimators can thus be constructed as

$$\overline{O}_q^{N_{\mathrm{sh}}} = \sum_{j=1}^{N_O} \frac{\alpha_i \overline{P}_i^{N_{\mathrm{sh}}}}{q} + \alpha_0 = \frac{1}{q}\left(\overline{O}^{N_{\mathrm{sh}}} - \alpha_0\right) + \alpha_0.$$

The effect of this mitigation on the variance of the Pauli estimators is thus

$$\mathrm{Var}_\rho\, \overline{A}_q^{N_{\mathrm{sh}}} = \frac{1}{q^2}\,\mathrm{Var}_\rho\, \overline{A}^{N_{\mathrm{sh}}},$$

and hence, using (3.6), we obtain the variance of the mitigated estimator

$$\mathrm{Var}_\rho\, \overline{A}_q^{N_{\mathrm{sh}}} = \mathrm{Var}_\psi\, \overline{A}^{N_{\mathrm{sh}}} + \frac{1-q^2}{q^2}\sum_{i=1}^{N_A}\frac{\alpha_i^2}{|S_i|} + \frac{1-q}{q}\sum_{i=1}^{N_A}\sum_{j=1,j\neq i}^{N_A}\frac{|S_i\cap S_j|}{|S_i||S_j|}\alpha_i\alpha_j\langle P_iP_j\rangle_\psi.$$

On the one hand, in the case of an importance-sampling strategy, we obtain

$$\begin{aligned}
\mathrm{Var}_\rho\, \overline{A}_q^{N_{\mathrm{sh}}} &= \mathrm{Var}_\psi\, \overline{A}^{N_{\mathrm{sh}}} + \frac{1-q^2}{q^2}\sum_{i=1}^{N_A}\frac{|\alpha_i|\|\alpha_{1:}\|_1}{N_{\mathrm{sh}}} \\
&= \mathrm{Var}_\psi\, \overline{A}^{N_{\mathrm{sh}}} + \frac{1-q^2}{q^2}\frac{\|\alpha_{1:}\|_1^2}{N_{\mathrm{sh}}} \\
&\leq \left(1 + \frac{1-q^2}{q^2}\right)\frac{\|\alpha_{1:}\|_1^2}{N_{\mathrm{sh}}} = \frac{\|\alpha_{1:}\|_1^2}{q^2 N_{\mathrm{sh}}}.
\end{aligned}$$

On the other hand, for a diagonal-sampling strategy, we have

$$\begin{aligned}
\mathrm{Var}_\rho\, \overline{A}_q^{N_{\mathrm{sh}}} &= \mathrm{Var}_\psi\, \overline{A}^{N_{\mathrm{sh}}} + \frac{1-q}{q^2 N_{\mathrm{sh}}}\sum_{i=1}^{N_A}\alpha_i^2 + \frac{1-q}{qN_{\mathrm{sh}}}\sum_{i=1}^{N_A}\sum_{j=1}^{N_A}\alpha_i\alpha_j\langle P_iP_j\rangle_\psi \\
&\leq \frac{(1+\frac{1-q}{q^2})\|\alpha_{1:}\|_2^2}{N_{\mathrm{sh}}} + \frac{\frac{(1-q)}{q}(\|A\|_\sigma - \alpha_0)^2}{N_{\mathrm{sh}}}.
\end{aligned}$$

For the squared Pauli terms, for any i, $1 \leq i < 2^N$, we consider the following mitigated estimator

$$\overline{\overline{P_i^2}_q}^{N_{\mathrm{sh}}} = \frac{1}{q^2}\overline{P_i^2}^{N_{\mathrm{sh}}}.$$

Its variance is thus given by

$$\mathrm{Var}_\rho\, \overline{\overline{P_i^2}_q}^{N_{\mathrm{sh}}} = \frac{1}{q^4}\,\mathrm{Var}_\rho\, \overline{P_i^2}^{N_{\mathrm{sh}}},$$

and, for any $j \neq i$ such that $0 \leq j < 2^N$, we have

$$\mathrm{Cov}_\rho(\overline{\overline{P_i^2}_q}^{N_{\mathrm{sh}}}, \overline{\overline{P_j^2}_q}^{N_{\mathrm{sh}}}) = \mathrm{Cov}_\rho(\overline{P_i^2}^{N_{\mathrm{sh}}}, \overline{P_j^2}^{N_{\mathrm{sh}}}).$$

Finally, we set $\overline{\mathcal{I}}_q^{N_{\mathrm{sh}}} = \frac{1}{q^2}\overline{\mathcal{I}}^{N_{\mathrm{sh}}}$ and we obtain that the variance of the mitigated Pauli interaction estimator is given by

$$\mathrm{Var}_\rho\, \overline{\mathcal{I}}_q^{N_{\mathrm{sh}}} = \frac{1}{q^4}\,\mathrm{Var}_\rho\, \overline{\mathcal{I}}^{N_{\mathrm{sh}}}.$$

Using these results, let us now study the effects of mitigation on the variance of estimators in practice.

3.7.2 Numerical results for mitigated estimators

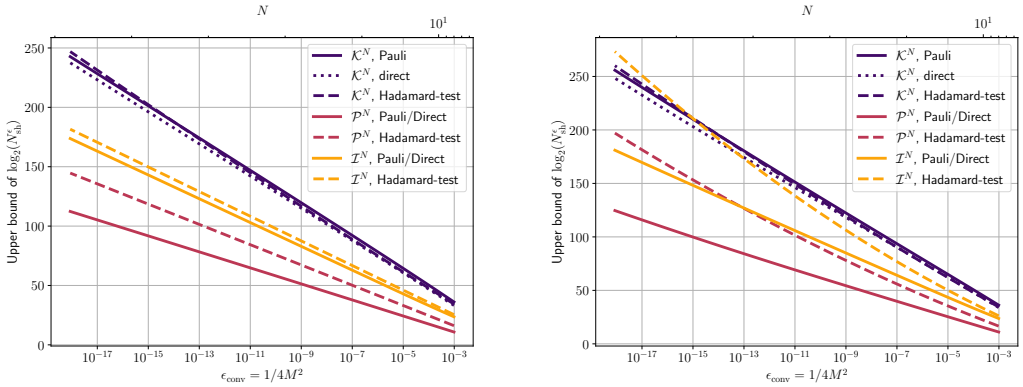


Figure 3.8 – Minimum number of samples to achieve sufficient precision for convergence, $\epsilon < \epsilon_{\mathrm{conv}}$ under depolarization noise for $V_0 = 1$, $\kappa = 1$. Left: $N_{\mathrm{lay}} = 2$. Right: $N_{\mathrm{lay}} = N$.

Using the results on mitigated estimators obtained in the previous section, the bounds obtained for the effects of noise, summarized in Table 3.1, and the gate estimates in Table 3.2, we are able to provide estimates of the number of samples N_{sh}^ϵ required to achieve a precision ϵ on the energy for convergence. This is illustrated in Figure 3.8 in the case of the U2 ansatz for $p_2 = 0.001$. On the left, we consider a fixed number of layers equal to 2 and on the right a number of layers equal to the number of qubits N. As expected, the Pauli method for the kinetic term outperforms the Hadamard-test for a sufficiently large number of qubits. Recall that the Pauli estimator of the kinetic term in the noiseless case had a variance which scaled as N^2 with respect to the other estimators, this overhead is compensated in the noisy case due to the number of additional gates in the circuits of the other estimators scaling with respect to N. Indeed, as all methods share the state preparation circuit the additional gates imply an exponential growth of the variance and

thus of N_{sh}^{ϵ} when compard to the Pauli estimator. This implies that, while not visible in Figure 3.8, the same is true for the diagonal-sampling estimator due to the basis change circuit required being a QFT, indeed if we look at the ratio of mitigated variances, we get

$$\frac{\operatorname{Var} \overline{\mathcal{K}^N}^{N_{\text{sh}}}}{\operatorname{Var} \check{\mathcal{K}^N}^{N_{\text{sh}}}} \approx \frac{N^2}{(1-p_2)^{N(N-1)}}$$

When $p_2 = 0.001$, as we can see in Figure 3.9, the number of qubits required for the Pauli importance-sampling estimator to outperform the diagonal estimator is around 100 qubits.

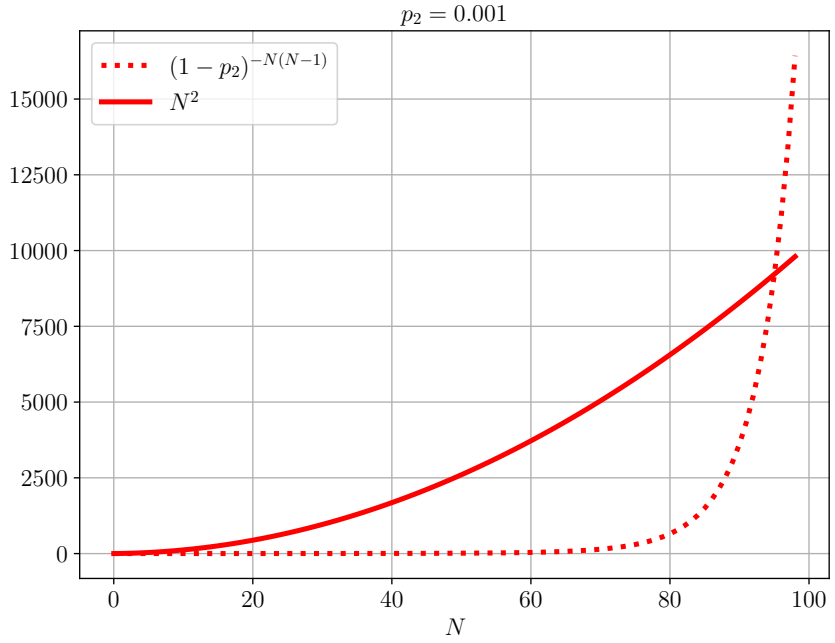


Figure 3.9 – Plot of the terms in the ratio of mitigated variances for $\overline{\mathcal{K}^N}^{N_{\text{sh}}}$ and $\check{\mathcal{K}^N}^{N_{\text{sh}}}$.

For the other energy terms, as the Hadamard-test circuits require additional controlled state preparations, their performance is severely reduced, even more so in the case where the number of layers is proportional to N. Now, as we have seen in Chapter 2, Figure 2.13, using the spectral Laplacian for the diagonal-sampling estimator vastly outperforms the other methods as it allows to obtain better solution convergence for a given number of qubits. This remains true in the presence of noise as can be seen in Figure 3.10, and even more so, since q explicitly depends on N even in the case of the potential term.

Another noteworthy result is that, given the exponential dependence of q on the number of gates (3.2), and the fact that the number of gates in our ansatz is proportional to NN_{lay}, there exists a number of layers such that the effect of noise is comparable to the variances of the energy terms. If we set $q^2 = 16^N$, *i.e.* such that the effect of noise is comparable to the variance of the kinetic term. We obtain that in the case of the U2 ansatz and for $p_2 = 0.001$, this amounts to $N_{\text{lay}} = -\frac{4 \log 2}{12 \log(1-p_2)} \approx 231$, as is illustrated in Figure 3.11.

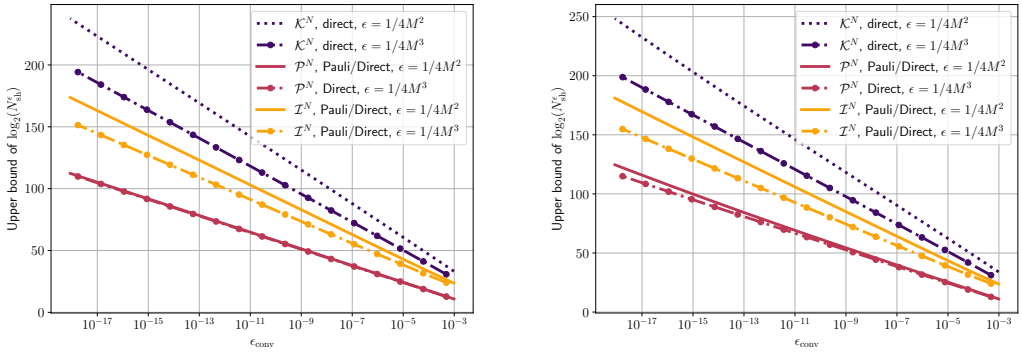


Figure 3.10 – Minimum number of samples to achieve sufficient precision for convergence, $\epsilon < \epsilon_{\rm conv}$ under depolarization noise for $V_0 = 1$, $\kappa = 1$. Comparison with spectral Laplacian discretization and with classical calculation of matrix vector product and L^4 norm calculation. Left: $N_{\rm lay} = 2$. Right: $N_{\rm lay} = N$.

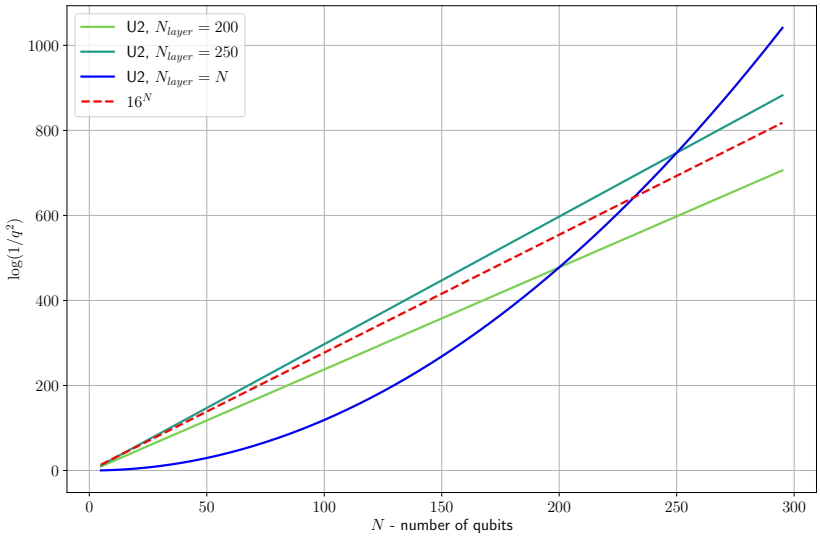


Figure 3.11 – Plot of the mitigation overhead q^2 for different number of layers for $p_2 = 0.001$.

3.7.3 Discussion

As is readily apparent in the numerical results shown above, when compared to the increase in the variance of the discretized unbounded operators, the effect of noise on $N_{\rm sh}^{\epsilon}$ is negligible for values of N less than a hundred.

Furthermore, as the increased precision required to achieve convergence is such that $\epsilon \leq O\left(\frac{1}{M^2}\right)$, the increase in the number of samples due to the increase in precision is vastly greater

than the increase due to mitigation. Note that the results obtained for any noise model would be similar to the analytical ones we got in this chapter with a very simple noise model. Indeed, as shown in [14], the overhead induced by mitigation is related to the contraction of the range of values of the measured observable (3.30). Any noise model, such as amplitude damping or de-phasing [73], which induces such a contraction would thus lead to similar results as they also induce such contractions. In practice such mitigation methods would only allow the enhancement of precision up to a certain point, mainly due to the fact that the effect of noise cannot be perfectly captured. These results showcase that the problems found in Chapter 2 still dominate even in the presence of noise and that even with perfect mitigation, energy estimation is too costly for any practical use.

Conclusion

Throughout this thesis, we have explored the potential and challenges of solving differential equations on noisy quantum computers, with a particular focus on energy estimation, a cornerstone of many NISQ algorithms. While the energy estimation methods we proposed through the Walsh-Pauli decomposition are promising, they may not scale well to very large quantum systems due to the following issues.

Indeed, the results we have obtained imply that NISQ algorithms for the solving of differential equations need to be engineered differently in order to have a hope at quantum advantage due to the following caveats of variational quantum algorithms:

1. **Sampling error and convergence accuracy**: As discussed in Chapter 2, the number of samples needed to achieve an error ϵ for an energy term A scales as follows:

$$N_{\text{sh}}^{\epsilon} \propto \frac{\operatorname{Var} A}{\epsilon^2}.$$

   To ensure convergence, the required accuracy for the energy terms must be of the order $\epsilon \propto \frac{1}{M^2} = 4^{-N}$, where M is the number of discretization points. Consequently, the number of samples needed scales as $N_{\text{sh}}^{\epsilon} \propto M^4 = 16^N$. In comparison, for classical calculations of energy terms, the computational cost does not directly scale with precision. Specifically, for matrix-vector multiplication, the cost is generally $O(M^2)$ in serial computations and $O(M)$ in parallel ones. However, in our case, the energy operators' specific structure allows us to perform vector-vector operations, reducing the cost to $O(M)$. Additionally, since our operators can be represented as low-rank matrix product operators (MPOs), further optimization is likely achievable. Nevertheless, regardless of the computational approach, the main constraint for classical algorithms is the memory required to store a vector of size 2^N with sufficient precision. For instance, representing a 32-qubit vector with quadruple precision (128-bit floating point) necessitates 64 terabytes of RAM.

   Table 3.3 compares classical and quantum energy calculations, focusing on the matrix-vector case for broader applicability. As anticipated, the number of samples required for quantum algorithms is prohibitively high due to the steep increase in cost associated with precision requirements.

2. **Unbounded variance**: In theory, the upper bound of the variance with respect to the spectral norm $\|A\|_\sigma$ of an operator should not impact the algorithm's performance. If the ansatz states are selected such that their H^1 norms remain controlled, the variance stays bounded, and the exponential scaling due to the diverging norms of the kinetic and interaction operators should not adversely affect the algorithm. However, current ansatz circuits often do not constrain the exploration to such controlled manifolds. Consequently, minor adjustments in ansatz parameters can lead to increased kinetic energies and a correspondingly spiky energy landscape. Although this issue does not affect the energy

	Quantum memory requirement: $N = 32$ qubits		Classical memory requirement: 2^N float128 = 64 GB of RAM	
	Number of samples N_{sh}^{ϵ}	**Quantum computation time**	**Classical number of FLOPs**	**Classical computation time**
FD $\epsilon_{\text{conv}} \approx 10^{-19}$	$O(M^4) \approx 10^{38}$	$10^{17}\times$ aotu	$O(M^2) \approx 10^{19}$	≈ 10 years
Spectral $\epsilon_{\text{conv}} \approx 10^{-29}$	$O(M^6) \approx 10^{58}$	$10^{36}\times$ aotu	$O(M^2) \approx 10^{19}$	≈ 10 years

Table 3.3 – Computational cost of estimating the energy of a single vector when the variance is equal to 1 for $N = 32$. We consider as in Chapter 2 a sample frequency of 1.5 Mhz for the quantum system and we consider a classical computer with a 1 teraflop/s performance(consumer GPU-level performance) with 1 terabyte of RAM using float128 to represent the vectors, a sufficient precision for the energy estimation. 'aotu' is used as a shorthand for 'age of the universe' $\approx 10^{10}$ years.

estimation for the final solution state, it significantly impacts the states encountered during optimization, often resulting in a dramatic increase in computational cost. A potential remedy is to develop tailored ansätze that confine the exploration to specific manifolds. This approach could be facilitated by leveraging the connections between tensor trains and function representations, as suggested by recent studies on matrix product states and tensor decompositions [83, 84].

3. **Noise-induced bias**: As we saw in Chapter 3, the main effect of noise is to introduce a bias (3.27) which upper-bounds the error of the energy estimation. This severely hinders the maximum number of gates useable to achieve a minimum error of ϵ as shown in Figure 3.6 and thus imposes the use of error mitigation methods.
4. **Mitigation-induced overhead**: Error mitigation introduces a sampling overhead that grows exponentially with the number of gates in the quantum circuit,

$$N_{\text{sh}} \approx \frac{\operatorname{Var} A}{(1-p_2)^{2N_{G_2}^{\text{equiv}}}\epsilon^2}.$$

Given that the circuit size for the ansatz is proportional to $N \times N_{\text{lay}}$, this means the number of samples needed for mitigated energy estimation increases as $(1-p_2)^{2NN_{\text{lay}}}$, resulting in a sampling overhead exponentially proportional to the number of gates in the quantum circuit. However, as demonstrated by the numerical results in Chapter 3, particularly in Figures 3.8 and 3.10, the impact of noise on Pauli estimators remains minimal compared to other sources of error, even for systems with up to 100 qubits.

Take-away message While it is readily apparent that variational quantum algorithms for differential equations have limited prospects of achieving any quantum advantage in the high-precision setting due to the challenges outlined above, there remains hope for such an advantage within the NISQ regime. For VQAs, this hope may reside in the high-dimensional, fixed-precision setting, which suffers to a lesser degree from the above argument on sampling error. However, the other limitations of VQAs remain valid in this setting. Indeed, the fact that multivariate functions tend to require higher bond dimensions imply that much deeper ansatz would be needed. Furthermore, the increased depth also induces a need for better hardware to reduce the

noise-induced overhead. Finally, the problem of ansatz trainability remains an open question in this case. Further work would thus be needed to investigate this use-case more thoroughly.

By changing the lens through which we represent differential equations and employing innovative techniques, Hamiltonian simulation is another method which offers a glimmer of promise. Indeed, in the classical context, due to memory considerations it has generally been preferable to reduce the number of dimensions of a linear problem in order to obtain simulations, even at the cost of introducing nonlinearities. In the quantum case, it appears that the reverse seems to apply, encoding nonlinear problems as high-dimensional linear ones may allow for their efficient quantum simulation [66, 90]. Furthermore, by encoding a nonlinear problem as a Schrödinger equation, for a larger number of qubits and with a potentially non-Hermitian operator[48, 51, 52], may provide a quantum advantage for observable measurements, so long as the observable is not unbounded.

These types of algorithms, however, make use of quantum phase estimation and generally require fault-tolerant quantum computers to achieve an advantage. To achieve an advantage in the NISQ context, one promising approach to solve this problem is the statistical phase estimation algorithm [10, 27, 96] which may allow phase estimation for non-error-corrected quantum devices.

The work presented in this thesis, while showing the problems inherent to VQAs for differential equations, may allow for new methods for Hamiltonian simulation. Indeed, while the results we obtained concerning the Walsh series do not lead to an advantage concerning VQAs, they may allow for an explicit Pauli decomposition of differential and evolution operators through the representation of the Fourier basis functions. By enabling the exact representation of evolution operators in the Pauli basis, our results could lead to more efficient implementations of Trotterization schemes for the evolution of differential equations.

Furthermore, an intriguing avenue for further exploration lies in the encoding of nonlinear differential equations into higher-dimensional, but finite, linear ones more akin to be solved on quantum systems. The Gross-Pitaevskii equation, for instance, can be viewed as a 1-body representation of the ground state of a complex N-body bosonic system. Notably, it has been established that the ground-state energy of the Gross-Pitaevskii equation for an N-body system can be reformulated in terms of a $k < N$ body system with rescaled parameters [60]. This raises interesting questions about the possibility of encoding the wavefunction in the associated k-body system as well. Moreover, it would be valuable to investigate whether this approach can be extended to excited states of N-body systems or even to fermionic equivalents.

Appendix A

Appendix

A.1 A-priori Analysis

In this section we will provide an a-priori error analysis of 2.4. To do so, we first introduce some context and provide several results which will allow us to determine the convergence rate of u_N to u.

A.1.1 Context

Let us first set $M = 2^N = \frac{1}{h_N}$ and introduce for any $n \geq 0$ the following functional spaces

$$X_n = \mathrm{Span}_{\mathbb{C}}\left(\{\mathrm{e}_k : x \mapsto e^{i2\pi kx}, -n \leq k \leq n\}\right)$$

Note that $\dim X_n = 2n+1$, furthermore, we have that $\mathcal{X}_N = X_{M/2-1} \times \mathrm{Span}_{\mathbb{C}}(c_{M/2})$ where we recall that $c_{M/2} : x \mapsto \cos\left(2\pi\frac{M}{2}x\right)$. Hence we have $X_{M/2-1} \subset \mathcal{X}_N \subset X_{M/2}$. Let us now introduce a reduced Fourier subspace $X_{M_{\mathrm{red}}}$, with $M_{\mathrm{red}} = M/4 - 1$ which will allow us to make a link between the solutions of the (2.2) and our solutions in the discretized case.

Fourier Interpolation and Truncation

First, we define the truncated Fourier series.

Definition 15. *For any $v \in L^2(0,1)$, its Fourier series is $\sum_{k\in\mathbb{Z}} \hat{v}_k \mathrm{e}_k$. Let $n > 0$, we define the projection operator $\Pi_n : L^2(0,1) \to X_n$ which to any $v \in L^2(0,1)$ associates its the truncated Fourier series $\Pi_n v \in X_n$,*

$$\Pi_n v = \sum_{|k|\leq n} \hat{v}_k \mathrm{e}_k.$$

We note the following property:

Proposition 21 ([15, 16])**.** *Let $n > 0$, given $v \in H^m(0,1)$ for $m > 0$, we have that*

$$\|\Pi_n v - v\|_{L^2} \leq CN^{-m}\|v^{(m)}\|_{L^2}$$

We now give the rational behind the introduction of our special interpolation operator, $\mathrm{I}_N : L^2(0,1) \to \mathcal{X}_N$. Given $T_n = \mathrm{Span}_{\mathbb{R}}\{\mathrm{c}_k | 0 \leq k \leq n\} \cup \{\mathrm{s}_k | 1 \leq k \leq n\}$, where for $1 \leq k \leq M/2$ $c_k : x \mapsto \cos(2\pi kx)$ and $s_k : x \mapsto \sin(2\pi kx)$ and where $c_0 : x \mapsto 1$. It is readily evident that T_n is the restriction of real-valued functions of X_n, indeed, as for any $1 \leq k \leq n$ e_k and e_{-k} are elements of X_n, c_k and s_k can be reconstructed using Euler's formula. The same however is not true for the space of trigonometric polynomials $S_N = \{\mathrm{e}_k \mid -N \leq k < N\}$, indeed the absence of e_N implies that $\mathcal{X}_N \not\subset S_N$. The trigonometric interpolation operator is usually defined on S_N[15]. In that case, it corresponds to the operator I_N^S sending v to the unique function of S_N taking the same values as the function v on Ω_N defined in (2.5). As we aim to describe functions that may be strictly real-valued, we are led to consider the trigonometric interpolant on $\mathcal{X}_N$.

Definition 16. *Let $n > 0$, given $v \in \mathcal{C}^0([0,1])$ we define the trigonometric interpolation operator $\mathrm{I}_N : \mathcal{C}^0([0,1]) \to \mathcal{X}_N$ such that*

$$\mathrm{I}_N v(\mathrm{x}_k) = v(\mathrm{x}_k), \quad 0 \leq k < 2^N.$$

For functions of $X_{M/2-1}$, it is straightforward to note that our interpolation operator and the interpolation operator on S_N coincide.

First let us note that we have, for any $\ell, k, |\ell| \leq M/2-1, |k| \leq M/2-1$,

$$\langle e_\ell, e_k \rangle_{L^2_N} = \sum_{\ell=0}^{2^N-1} \exp(i2\pi(\ell-k)x_k)h_N = \delta_{\ell,k} h_N \cdot 2^N = \delta_{\ell,k} = \langle e_\ell, e_k \rangle_{L^2}.$$

Similarly

$$\langle e_\ell, c_{M/2} \rangle_{L^2_N} = 0 = \langle e_\ell, c_{M/2} \rangle_{L^2}$$

For the norm of the cosine term however, we have that $\|c_{M/2}\|^2_{L^2} = \frac{1}{2}$ while its discrete norm is equal to 1, indeed,

$$\|c_{M/2}\|^2_{L^2_N} = h_N \cdot \sum_{k=0}^{2^N-1} (-1)^{2k} = 1 = 2\|c_{M/2}\|^2_{L^2}.$$

This implies that given the decomposition $v = a_{M/2}\mathrm{c}_{M/2} + \sum_{k=-M/2-1}^{M/2-1} a_k \mathrm{e}_k$ of an element in $\mathcal{X}_N$, then

$$\|v\|^2_{L^2_N} = |a_{M/2}|^2 + \sum_{k=-M/2-1}^{M/2-1} |a_k|^2 = \|v\|^2_{L^2} + \frac{|a_{M/2}|^2}{2}$$

and thus

$$\forall v \in \mathcal{X}_N, \quad \|v\|_{L^2} \leq \|v\|_{L^2_N} \leq \sqrt{2}\|v\|_{L^2}$$

Proposition 22. *Let $N > 0$, given $v \in \mathcal{X}$, for any $w \in X_{M/2-1}$ we have*

$$\langle v, w \rangle_{L^2_N} = \langle \mathrm{I}_N v, w \rangle_{L^2_N} = \langle \mathrm{I}_N v, w \rangle_{L^2}.$$

Furthermore, we have that for any v, $w \in X_{M/2-1}$,

$$\langle v, w \rangle_{L^2_N} = \langle v, w \rangle_{L^2}.$$

Definition 17. *Given $v \in H^1(0,1)$, we denote the Fourier series of its trigonometric interpolant as*

$$\mathrm{I}_N v = \sum_{|k|<M/2} \hat{v}_k \mathrm{e}_k + \hat{v}_{M/2}\mathrm{c}_{M/2}$$

with $\hat{v}_k = \frac{\langle v, \mathrm{e}_k \rangle_{L^2_N}}{\|\mathrm{e}_k\|^2_{L^2_N}}$ and $\hat{v}_{M/2} = \frac{\langle v, \mathrm{c}_{M/2} \rangle_{L^2_N}}{\|\mathrm{c}_{M/2}\|^2_{L^2_N}}$ and

We now give two lemmas for functions of $\mathcal{X}_N$.

Lemma 9. *Let $N > 0$, given $v \in \mathcal{X}$, for any $w \in \mathcal{X}_N$ we have*

$$\langle v, w \rangle_{L^2_N} = \langle \mathrm{I}_N v, w \rangle_{L^2} + \frac{\hat{v}_{M/2}\hat{w}_{M/2}}{2}$$

and equivalently, $\langle v, \mathrm{c}_{M/2} \rangle_{L^2_N} = 2\langle v, \mathrm{c}_{M/2} \rangle_{L^2}$

Furthermore, we can recover a convergence result for the interpolation operator through the convergence result on the S_N interpolation.

Proposition 23. *Let $N > 0$, given $v \in H^k(0,1)$ for $k > 0$, then we have*

$$\|\mathrm{I}_N v - v\|_{L^2} \leq C\frac{1}{2^{Nk}}\|v^{(k)}\|_{L^2} = C\frac{1}{M^k}\|v^{(k)}\|_{L^2}.$$

Proof. Let $N > 0$, given $v \in H^k(0,1)$ for $k > 0$, from [16] p. 272, we have the results for I_N^S, namely, there exists constant $C > 0$ such that,

$$\|\mathrm{I}_N^S v - v\|_{L^2} \leq C\frac{1}{2^{Nk}}\|v^{(k)}\|_{L^2} = C\frac{1}{M^k}\|v^{(k)}\|_{L^2}.$$

Let us rewrite these norms using Parseval identity,

$$\begin{aligned}\|\mathrm{I}_N v - v\|_{L^2}^2 = &\sum_{|k|<M/2} |\hat{\hat{v}}_k - \hat{v}_k|^2 + \sum_{|k|>M/2} |\hat{v}_k|^2 \\ &+ |\hat{\hat{v}}_{M/2} - (\hat{v}_{M/2} + \hat{v}_{-M/2})|^2 \|\mathrm{c}_{M/2}\|_{L^2}^2 + |\hat{v}_{M/2} - \hat{v}_{-M/2}|^2 \|\mathrm{s}_{M/2}\|_{L^2}^2 \\ = &\sum_{|k|<M/2} |\hat{\hat{v}}_k - \hat{v}_k|^2 + \sum_{|k|>M/2} |\hat{v}_k|^2 \\ &+ \frac{|\hat{\hat{v}}_{M/2} - (\hat{v}_{M/2} + \hat{v}_{-M/2})|^2 + |\hat{v}_{M/2} - \hat{v}_{-M/2}|^2}{2},\end{aligned}$$

on the other hand, for I_N^S, we have

$$\begin{aligned}\|\mathrm{I}_N^S v - v\|_{L^2}^2 = &\sum_{|k|<M/2} |\hat{\hat{v}}_k - \hat{v}_k|^2 + \sum_{|k|>M/2} |\hat{v}_k|^2 \\ &+ |\hat{\hat{v}}_{M/2} - \hat{v}_{-M/2}|^2 + |\hat{v}_{M/2}|^2.\end{aligned}$$

Notice that

$$\begin{aligned}&\frac{|\hat{\hat{v}}_{M/2} - (\hat{v}_{M/2} + \hat{v}_{-M/2})|^2 + |\hat{v}_{M/2} - \hat{v}_{-M/2}|^2}{2} \\ &\leq \frac{|\hat{\hat{v}}_{M/2} - \hat{v}_{-M/2}|^2 + |\hat{v}_{M/2}|^2 + |\hat{v}_{M/2}|^2 + |\hat{v}_{-M/2}|^2}{2} \\ &\leq |\hat{\hat{v}}_{M/2} - \hat{v}_{-M/2}|^2 + |\hat{v}_{M/2}|^2 + \frac{|\hat{v}_{M/2}|^2 + |\hat{v}_{-M/2}|^2}{2}.\end{aligned}$$

This allows us to obtain the following inequality,

$$\begin{aligned}\|\mathrm{I}_N v - v\|_{L^2}^2 &\leq \|\mathrm{I}_N^S v - v\|_{L^2}^2 + \frac{|\hat{v}_{M/2}|^2 + |\hat{v}_{-M/2}|^2}{2} \\ &= \|\mathrm{I}_N^S v - v\|_{L^2}^2 + \frac{(|\hat{v}_{M/2}|^2 + |\hat{v}_{-M/2}|^2)(\pi^2 M^2)^k}{2(\pi^2 M^2)^k} \\ &\leq (C + \frac{1}{2\pi^2})\frac{1}{M^{2k}}\|v^{(k)}\|_{L^2}^2\end{aligned}$$

which gives us the result for the L^2 norm.

□

We also have a similar result for the L^∞ norm.

Proposition 24. *Let $N > 0$, given $v \in H^k(0,1)$ for $k > 0$, then we have*

$$\|\mathrm{I}_N v - v\|_{L^\infty} \leq C\frac{N}{2^{Nk}}\|v^{(k)}\|_{L^2} = C\frac{\log M}{M^k}\|v^{(k)}\|_{L^2}.$$

Proof. Let us denote generically $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, we denote by $\mathcal{T}_N^{\mathbb{K}} = \mathrm{Span}_{\mathbb{K}}\{\mathrm{c}_k \mid 0 \leq k \leq M/2\} \cup \{\mathrm{s}_k \mid 0 < k < M/2\}$. As shown in [46], we have the above result for $\mathcal{T}_N^{\mathbb{R}}$ which can be immediately extended to any $\mathcal{T}_N^{\mathbb{K}}$. Specifically, we extend this result to $\mathcal{T}_N^{\mathbb{C}}$ which is identical to $\mathcal{X}_N$. Alternatively, we can use a similar methodology as the previous proof for the L^2 norm bound. Indeed, let $N > 0$, given $v \in H^k(0,1)$ for $k > 0$, from [16] p. 272, we have the results for I_N^S, namely, there exists constant $C > 0$ such that,

$$\|\mathrm{I}_N^S v - v\|_{L^\infty} \leq C\frac{N}{2^{Nk}}\|v^{(k)}\|_{L^\infty} = C\frac{\log M}{M^k}\|v^{(k)}\|_{L^\infty}.$$

We have

$$\|\mathrm{I}_N v - v\|_{L^\infty}^2 = \|\mathrm{I}_N^S v - v - i\hat{\hat{v}}_{M/2}\mathrm{s}_k\|_{L^\infty}^2 \leq \|\mathrm{I}_N^S v - v\|_{L^\infty}^2 + |\hat{\hat{v}}_{M/2}|\|\mathrm{s}_{M/2}\|_{L^\infty}$$

The value of $\hat{\hat{v}}_{M/2}$ can be recovered from the Fourier series of v, indeed,

$$\hat{\hat{v}}_{M/2} = \langle v, \mathrm{e}_{M/2}\rangle_{L_N^2} = \sum_{\ell\in\mathbb{Z}} \hat{v}_\ell \langle \mathrm{e}_\ell, \mathrm{e}_{M/2}\rangle_{L_N^2} = \sum_{\ell\in\mathbb{Z}} \hat{v}_{M/2+M\ell},$$

the higher frequencies obtained being those with the same value as $\mathrm{e}_{M/2}$ on Ω_N. This gives us the bound

$$|\hat{\hat{v}}_{M/2}| \leq \frac{|\hat{v}_{M/2}|(M^2\pi^2)^k}{(M^2\pi^2)^k} + \|v - \pi_{M/2}v\|_{L^2} \leq \frac{C}{\pi^2 M^2}\|v^{(k)}\|_{L^2} \leq \frac{C}{\pi^2 M^2}\|v^{(k)}\|_{L^\infty}.$$

Finally,

$$\|\mathrm{I}_N v - v\|_{L^\infty}^2 \leq \frac{C(\log M + C)}{M^k}\|v^{(k)}\|_{L^\infty}.$$

□

Results on discretized Laplacian

For convenience's sake, we define the following bilinear functions.

Definition 18. *Let $v, w \in \mathcal{X}$, we define*

$$a(v,w) = \langle v', w'\rangle_{L^2}, \qquad a_N(v,w) = \langle v, -\Delta_N w\rangle_{L_N^2},$$

where $\Delta_N = \tau_{h_N} + \tau_{-h_N} - 2I$ is the discretized Laplacian operator defined in (2.11).

We also rewrite the energy terms (2.3) and (2.10) as

$$E(v) = \frac{1}{2}a(v,v) + \langle V, v^2\rangle_{L^2} + \frac{\kappa}{2}\langle v^2, v^2\rangle_{L^2}$$
$$E_N(v) = \frac{1}{2}a_N(v,v) + \langle V, v^2\rangle_{L_N^2} + \frac{\kappa}{2}\langle v^2, v^2\rangle_{L_N^2}$$

We also introduce the discretized operator

$$A_v^N = -\frac{1}{2}\Delta_N + V + \kappa v^2.$$

Let us first state

Lemma 10. *For $v_N, w_N \in \mathcal{X}_N$, we have*

$$a_N(v_N, w_N) = \sum_{|k| \leq M/2-1} 4\hat{v}_k \hat{w}_k^* \sin\left(\pi \frac{k}{M}\right)^2 M^2 + 8\hat{v}_{M/2}\hat{w}_{M/2}^* M^2$$

Proof. We readily note that $h_N^2 \Delta_N = \tau_{h_N} + \tau_{-h_N} - 2I = -(\tau_{h_N} - I)(\tau_{-h_N} - I)$. Then, for any $v_N, w_N \in X_{M/2-1}$, we have that $\frac{I-\tau_{h_N}}{h} v_N \in X_{M/2-1}$, and thus that:

$$\begin{aligned}
a_N(v_N, w_N) &= \langle \frac{I-\tau_{h_N}}{h} v_N, \frac{I-\tau_{h_N}}{h} w_N \rangle_{L_N^2} \\
&= \sum_{|k| \leq M/2-1} \hat{v}_k \hat{w}_k^* \left| \frac{e^{i2\pi k h_N} - 1}{2\pi k h_N} \right|^2 4\pi^2 k^2 \\
&= \sum_{|k| \leq M/2-1} \hat{v}_k \hat{w}_k^* \left| e^{i\pi k h_N} \frac{e^{i\pi \frac{k}{M}} - e^{-i\pi \frac{k}{M}}}{2\pi \frac{k}{M}} \right|^2 4\pi^2 k^2 \\
&= \sum_{|k| \leq M/2-1} \hat{v}_k \hat{w}_k^* \left| i e^{i\pi k h_N} \frac{\sin\left(\pi \frac{k}{M}\right)}{\pi \frac{k}{M}} \right|^2 4\pi^2 k^2 \\
&= \sum_{|k| \leq M/2-1} \hat{v}_k \hat{w}_k^* \frac{\sin\left(\pi \frac{k}{M}\right)^2}{\pi^2 \frac{k^2}{M^2}} 4\pi^2 k^2 \\
&= \sum_{|k| \leq M/2-1} 4\hat{v}_k \hat{w}_k^* \sin\left(\pi \frac{k}{M}\right)^2 M^2
\end{aligned}$$

For $v_N, w_N \in \mathcal{X}_N$, we have an additional term

$$a_N(v_N, w_N) = \sum_{|k| \leq M/2-1} 4\hat{v}_k \hat{w}_k^* \sin\left(\pi \frac{k}{M}\right)^2 M^2 + 8\hat{v}_{M/2}\hat{w}_{M/2}^* M^2$$

where we the additional term is

$$\begin{aligned}
4\langle c_{M/2}, v_N \rangle_{L_N^2} \langle w_N, c_{M/2} \rangle_{L_N^2} \sin(\pi/2)^2 M^2 &= 4(2 \cdot \langle c_{M/2}, v_N \rangle_{L^2})(2 \cdot \langle w_N, c_{M/2} \rangle_{L^2}) \sin(\pi/2)^2 M^2 \\
&= 8\hat{v}_{M/2}\hat{w}_{M/2}^* M^2
\end{aligned}$$

□

Lemma 11. *For any $v_N \in \mathcal{X}_N$, we have that*

$$\frac{1}{2\pi^2} a(v_N, v_N) \leq a_N(v_N, v_N) \leq a(v_N, v_N).$$

Furthemore, the discretization error $a - a_N$ for v_N, w_N is given by

$$(a - a_N)(v_N, w_N) = \sum_{|k| \leq M/2-1} \hat{v}_k \hat{w}_k^* 4\pi^2 k^2 \left(1 - \frac{\sin\left(\pi \frac{k}{M}\right)^2}{\left(\pi \frac{k}{M}\right)^2}\right) + 4\pi^2 (M/2)^2 \hat{v}_{M/2}\hat{w}_{M/2}^* \left(1 - \frac{8}{\pi^2}\right)$$

Proof. For $v_N, w_N \in X_{M/2-1}$, we have from Lemma 10,

$$a(v_N, w_N) = \sum_{|k| \leq M/2-1} \hat{v}_k \hat{w}_k^* 4\pi^2 k^2$$

This gives us :

$$(a - a_N)(v_N, w_N) = \sum_{|k| \leq M/2-1} \hat{v}_k \hat{w}_k^* 4\pi^2 k^2 \left(1 - \frac{\sin\left(\pi \frac{k}{M}\right)^2}{\left(\pi \frac{k}{M}\right)^2}\right)$$

For any $v_N, w_N \in \mathcal{X}_N$, we have the additional term

$$4\hat{v}_{M/2}\hat{w}_{M/2}^* \pi^2 (M/2)^2 - 8\hat{v}_{M/2}\hat{w}_{M/2}^* M^2 = 4\pi^2 (M/2)^2 \hat{v}_{M/2}\hat{w}_{M/2}^* \left(1 - \frac{8}{\pi^2}\right)$$

Finally, for any $v_N \in \mathcal{X}_N$

$$(a - a_N)(v_N, v_N) = \sum_{|k| \leq M/2-1} |\hat{v}_k|^2 4\pi^2 k^2 \left(1 - \frac{\sin\left(\pi \frac{k}{M}\right)^2}{\left(\pi \frac{k}{M}\right)^2}\right) + 4\pi^2 (M/2)^2 |\hat{v}_{M/2}|^2 \left(1 - \frac{8}{\pi^2}\right)$$

we have that each term $\left\{ \left(1 - \frac{\sin\left(\pi \frac{k}{M}\right)^2}{\left(\pi \frac{k}{M}\right)^2}\right) \mid \forall |k| \leq M/2 - 1 \right\} \cup \left\{ \left(1 - \frac{8}{\pi^2}\right) \right\} \subset [\frac{1}{2\pi^2}, 1]$, hence the result. □

Lemma 12. *The discretization error $(a - a_N)(v_N, v_N)$ for $v_N \in \mathcal{X}_N$ can bounded as follows*

$$(a - a_N)(v_N, v_N) \leq \frac{C}{M^2} \|v_N\|_{H^2}^2$$

Furthermore, we also have that for any $v_N, w_N \in \mathcal{X}_N$:

$$|(a - a_N)(v_N, w_N)| \leq \frac{C}{M^2} \|v_N\|_{H^2} \|w_N\|_{H^2}$$

Proof. If we consider the discretization error from Lemma 11 for $v_N \in X_{M/2-1}$, we have that

$$\begin{aligned}
(a - a_N)(v_N, v_N) &= \sum_{|k| < M/2} |\hat{v}_k|^2 4\pi^2 k^2 \left(1 - \frac{\sin\left(\pi \frac{k}{M}\right)^2}{\left(\pi \frac{k}{M}\right)^2}\right) \\
&= \sum_{|k| < M/2} |\hat{v}_k|^2 4 \left(\pi^2 k^2 - M^2 \left(\frac{\pi^2 k^2}{M^2} - 2\frac{\pi^4 k^4}{6 \cdot M^4} + O\left(\frac{k^6}{M^6}\right)\right)\right) \\
&= \sum_{|k| < M/2} |\hat{v}_k|^2 \left(\frac{4\pi^4 k^4}{3 \cdot M^2} - O\left(\frac{k^6}{M^4}\right)\right) \\
&\leq C \frac{\|v_N\|_{H^2}^2}{M^2}
\end{aligned}$$

The second inequality for $v_N, w_N \in X_{M/2-1}$is given by

$$\begin{aligned}|(a-a_N)(v_N,w_N)| &\leq \sum_{|k|<M/2} |\hat{v}_k||\hat{w}_k| \frac{4\pi^4 k^4}{3\cdot M^2} \\ &\leq C \frac{\|v_N''\|_{L^2}\|w_N''\|_{L^2}}{M^2}\end{aligned}$$

For any $v_N, w_N \in \mathcal{X}_N$, we bound the additional terms as follows:

$$\begin{aligned}|(a-a_N)(v_N,w_N)| &\leq \sum_{|k|\leq M/2} |\hat{v}_k||\hat{w}_k| \frac{4\pi^4 k^4}{3\cdot M^2} \\ &\quad + \pi^2 M^2 \left(|\hat{v}_{M/2}||\hat{w}_{M/2}| - \langle v_N, c_{M/2}\rangle_{L^2_N} \langle w_N, c_{M/2}\rangle_{L^2_N} \frac{\sin(\pi/2)^2}{\pi^2 (1/2)^2} \right) \\ &= \sum_{|k|\leq M/2} |\hat{v}_k||\hat{w}_k| \frac{4\pi^4 k^4}{3\cdot M^2} + 4\pi^2 (M/2)^2 |\hat{v}_{M/2}||\hat{w}_{M/2}| \left(1 - \frac{8}{\pi^2}\right) \\ &\leq \sum_{|k|\leq M/2} |\hat{v}_k||\hat{w}_k| \frac{4\pi^4 k^4}{3\cdot M^2} + \frac{4\pi^4 (M/2)^4}{\pi^2 (M/2)^2} |\hat{v}_{M/2}||\hat{w}_{M/2}| \left(1 - \frac{8}{\pi^2}\right) \\ &\leq C \frac{\|v_N''\|_{L^2}\|w_N''\|_{L^2}}{M^2}\end{aligned}$$

□

A.1.2 Bounding the convergence rate

Let us denote by

$$u = \operatorname*{arg\,min}_{v\in\mathcal{X}, \|v\|_{L^2}=1} E(v) \tag{A.1}$$

$$u_N = \operatorname*{arg\,min}_{v_N\in\mathcal{X}_N, \|v\|_{L^2_N}=1} E_N(v_N) \tag{A.2}$$

$$u_{M_{\text{red}}} = \operatorname*{arg\,min}_{v_{M_{\text{red}}}\in X_{M_{\text{red}}}, \|v\|_{L^2_N}=\|v\|_{L^2}=1} E_N(v_{M_{\text{red}}}) \tag{A.3}$$

$$\tilde{u}_{M_{\text{red}}} = \operatorname*{arg\,min}_{v_{M_{\text{red}}}\in X_{M_{\text{red}}}, \|v\|_{L^2}=\|v\|_{L^2_N}=1} E(v_{M_{\text{red}}}) \tag{A.4}$$

To bound the convergence rate of $\|u_N - u\|_{H^1}$, we consider the following decomposition

$$\|u_N - u\|_{H^1} \leq \underbrace{\|u - \tilde{u}_{M_{\text{red}}}\|_{H^1}}_{(1)} + \underbrace{\|\tilde{u}_{M_{\text{red}}} - u_{M_{\text{red}}}\|_{H^1}}_{(2)} + \underbrace{\|u_{M_{\text{red}}} - u_N\|_{H^1}}_{(3)}.$$

To obtain convergence rates, we show that each term converges to zero as N increases. For the first term, as shown in [28] we have that $\tilde{u}_{M_{\text{red}}}$, the solution to the non-discretized problem on the subspace $X_{M_{\text{red}}} = X_{M/4-1}$ converges strongly to the solution u. For the second term, we have that the distance of the two functions can be bounded by their difference in energy (following again the same lines as in [28]). We then show that the discretization error on $X_{M_{\text{red}}}$ for the solutions converges to zero and use this to bound the difference of the energies of the solutions.

For the third term, we develop a discretized analog to [28] and bound the term by the difference of the discrete energies. As we consider two minimizers, we are able to bound this by the difference of energies of u_N and its normalised projector on $X_{M_{\text{red}}}$.

Term (1)

First, let us look at the term (1): From [28] Theorem 2.3 on page 11, we have that

Theorem 12. *Let u be the solution to* (A.1) *and $\tilde{u}_{M_{\text{red}}}$ the solution to* (A.4)*, the following upper bound between the two solutions holds*

$$\|u - \tilde{u}_{M_{\text{red}}}\|_{H^1}^2 \leq CE(\tilde{u}_{M_{\text{red}}}) - E(u)$$

in addition there exists a $N_0 > 0, C_1$ and C_2 such that if $M_{\text{red}} \geq N_0$:

$$\|u - \tilde{u}_{M_{\text{red}}}\|_{H^1} \leq \min_{v_{M_{\text{red}}} \in X_{M_{\text{red}}}} C_1 \|v_{M_{\text{red}}} - u\|_{H^1} \leq C_2 \frac{\|u^{(k)}\|_{L^2}}{M_{\text{red}}^{(k-1)}}.$$

Term (3)

For the last term, we use the same idea as for term (1) shown in [28] now in the discrete setting. Under the assumption that the discrete operator $A_{u_N}^N$ admits λ_N as a simple lowest eigenvalue and we denote by $\lambda_{N,2}$ the next one, we show that

Lemma 13. *Let u_N be the solution to* (A.2) *and $u_{M_{\text{red}}}$ the solution to* (A.3)*, under the above assumption, the H^1 norm of $u_{M_{\text{red}}} - u_N$ is bounded by their difference in discretized energies, i.e.*

$$\|u_{M_{\text{red}}} - u_N\|_{H^1}^2 \leq C_{\lambda_N} \left(E_N(u_{M_{\text{red}}}) - E_N(u_N)\right),$$

with $C_{\lambda_N} = \frac{\lambda_{N,2} - \lambda_N}{2C_c + \lambda_{N,2} + \lambda_N}$.

Proof. We proceed in the exact same manner as in the proof of Theorem 3 in [28] pg.12. Indeed we have

$$\begin{aligned}
E_N(u_{M_{\text{red}}}) - E_N(u_N) &= \frac{1}{2} a_N(u_{M_{\text{red}}}, u_{M_{\text{red}}}) + \langle V, u_{M_{\text{red}}}^2 \rangle_{L_N^2} + \frac{\kappa}{2} \langle u_{M_{\text{red}}}^2, u_{M_{\text{red}}}^2 \rangle_{L_N^2} \\
&\quad - \left(\frac{1}{2} a_N(u_N, u_N) + \langle V, u_N^2 \rangle_{L_N^2} + \frac{\kappa}{2} \langle u_N^2, u_N^2 \rangle_{L_N^2} \right) \\
&\quad + \kappa \langle u_{M_{\text{red}}}^2, u_N^2 \rangle_{L_N^2} - \kappa \langle u_{M_{\text{red}}}^2, u_N^2 \rangle_{L_N^2} \\
&\quad - \frac{\kappa}{2} \langle u_N^2, u_N^2 \rangle_{L_N^2} + \frac{\kappa}{2} \langle u_N^2, u_N^2 \rangle_{L_N^2} \\
&= \frac{1}{2} a_N(u_{M_{\text{red}}}, u_{M_{\text{red}}}) + \langle V, u_{M_{\text{red}}}^2 \rangle_{L_N^2} + \kappa \langle u_N^2, u_{M_{\text{red}}}^2 \rangle_{L_N^2} \\
&\quad - \left(\frac{1}{2} a_N(u_N, u_N) + \langle V, u_N^2 \rangle_{L_N^2} + \kappa \langle u_N^2, u_N^2 \rangle_{L_N^2} \right) \\
&\quad + \frac{\kappa}{2} \langle u_N^2, u_N^2 \rangle_{L_N^2} + \frac{\kappa}{2} \langle u_{M_{\text{red}}}^2, u_{M_{\text{red}}}^2 \rangle_{L_N^2} - \kappa \langle u_{M_{\text{red}}}^2, u_N^2 \rangle_{L_N^2} \\
&= \langle A_{u_N}^N u_{M_{\text{red}}}, u_{M_{\text{red}}} \rangle_{L_N^2} - \langle A_{u_N}^N u_N, u_N \rangle_{L_N^2} + \frac{\kappa}{2} \|u_{M_{\text{red}}}^2 - u_N^2\|_{L_N^2}^2
\end{aligned}$$

Then we have

$$\begin{aligned}\langle A^N_{u_N} u_{M_{\text{red}}}, u_{M_{\text{red}}}\rangle_{L^2_N} - \langle A^N_{u_N} u_N, u_N\rangle_{L^2_N} &= \langle A^N_{u_N} u_{M_{\text{red}}}, u_{M_{\text{red}}} - u_N\rangle_{L^2_N} \\ &\quad + \langle A^N_{u_N} u_{M_{\text{red}}}, u_N\rangle_{L^2_N} - \langle A^N_{u_N} u_N, u_N\rangle_{L^2_N} \\ &= \langle A^N_{u_N} u_{M_{\text{red}}}, u_{M_{\text{red}}} - u_N\rangle_{L^2_N} + \langle A^N_{u_N} u_{M_{\text{red}}} - \lambda_N u_N, u_N\rangle_{L^2_N} \\ &= \langle (A^N_{u_N} - \lambda_N) u_{M_{\text{red}}}, u_{M_{\text{red}}} - u_N\rangle_{L^2_N} + \langle \lambda_N u_{M_{\text{red}}}, u_{M_{\text{red}}} - u_N\rangle_{L^2_N} \\ &\quad + \langle A^N_{u_N} u_{M_{\text{red}}} - \lambda_N u_N, u_N\rangle_{L^2_N} \\ &= \langle (A^N_{u_N} - \lambda_N) u_{M_{\text{red}}}, u_{M_{\text{red}}} - u_N\rangle_{L^2_N} \\ &= \langle (A^N_{u_N} - \lambda_N)(u_{M_{\text{red}}} - u_N), u_{M_{\text{red}}} - u_N\rangle_{L^2_N}\end{aligned}$$

where we have used the fact that $\|u_{M_{\text{red}}}\|_{L^2_N} = \|u_N\|_{L^2_N} = 1$ and $\forall v_N \in \mathcal{X}_N$,

$$\langle A^N_{u_N} u_N, v_N\rangle_{L^2_N} = \lambda_N \langle u_N, v_N\rangle_{L^2_N},$$

for the the penultimate equality and the last line.

This gives us :

$$E_N(u_{M_{\text{red}}}) - E_N(u_N) = \langle (A^N_{u_N} - \lambda_N)(u_{M_{\text{red}}} - u_N), u_{M_{\text{red}}} - u_N\rangle_{L^2_N} + \frac{\kappa}{2}\|u^2_{M_{\text{red}}} - u^2_N\|_{L^2_N}$$

Now, using the same reasoning as in [28], as $A^N_{u_N} - \lambda_N I$ is a positive linear operator with kernel u_N and lowest nonzero eigenvalue $\lambda_{N,2} - \lambda_N$, we obtain that for any $v_N \in \mathcal{X}_N$:

$$\begin{aligned}\langle (A^N_{u_N} - \lambda_N) v_N, v_N\rangle_{L^2_N} &\geq (\lambda_{N,2} - \lambda_N)\|v_N - \langle v_N, u_N\rangle_{L^2_N} u_N\|^2_{L^2_N} \\ &= (\lambda_{N,2} - \lambda_N)(\|v_N\|^2_{L^2_N} - |\langle v_N, u_N\rangle_{L^2_N}|^2)\end{aligned}$$

as a direct result of Parseval's Theorem. We denote $\eta_N = \lambda_{N,2} - \lambda_N$.

On the other hand, we obtain through the coercivity of the discretized Laplacian(from Lemma11) and the positivity of V and of the L^4 norm that for any $v_N \in \mathcal{X}_N$

$$\langle (A^N_{u_N} - \lambda_N) v_N, v_N\rangle_{L^2_N} \geq \frac{1}{4}\|v_N\|^2_{H^1} - (\lambda_N + \frac{1}{2})\|v_N\|^2_{L^2_N}$$

We now apply the first inequality to $v_N = u_{M_{\text{red}}} - u_N$, we have that :

$$\begin{aligned}\langle \left(A^N_{u_N} - \lambda_N\right)\left(u_{M_{\text{red}}} - u_N\right), u_{M_{\text{red}}} - u_N\rangle_{L^2_N} &= \langle \left(A^N_{u_N} - \lambda_N\right) u_{M_{\text{red}}}, u_{M_{\text{red}}}\rangle_{L^2_N} \\ &\geq \eta_N \left(\|u_{M_{\text{red}}}\|^2_{L^2_N} - |\langle u_{M_{\text{red}}}, u_N\rangle_{L^2_N}|^2\right) \\ &= \frac{\eta_N}{2}\left(\|u_{M_{\text{red}}}\|^2_{L^2_N} - 2|\langle u_{M_{\text{red}}}, u_N\rangle_{L^2_N}|^2 + \|u_N\|^2_{L^2_N}\right) \\ &= \frac{\eta_N}{2}\|u_{M_{\text{red}}} - u_N\|^2_{L^2_N}\end{aligned}$$

If we now consider the second inequality for $v_N = u_{M_{\text{red}}} - u_N$, we have :

$$\langle (A^N_{u_N} - \lambda_N) u_{M_{\text{red}}} - u_N, u_{M_{\text{red}}} - u_N \rangle_{L^2_N} \geq \frac{1}{2} C \|u_{M_{\text{red}}} - u_N\|^2_{H^1} - (\lambda_N + C_c) \|u_{M_{\text{red}}} - u_N\|^2_{L^2_N}$$

where $C_c = \frac{1}{2\pi^2}$ is the coercivity constant from Lemma 11. Combining these we obtain:

$$\begin{aligned}
\langle (A^N_{u_N} - \lambda_N)(u_{M_{\text{red}}} - u_N), u_{M_{\text{red}}} - u_N \rangle_{L^2_N} &\geq (1-t) \frac{C_{\text{c}}}{2} \|u_{M_{\text{red}}} - u_N\|^2_{H^1} \\
&\quad + \left(\frac{t\eta_N}{2} - (1-t)(\lambda_N + C_c) \right) \|u_{M_{\text{red}}} - u_N\|^2_{L^2_N} \\
&= (1-t) \frac{C_c}{2} \|u_{M_{\text{red}}} - u_N\|^2_{H^1} \\
&\quad + (t(\lambda_N + C_c + \frac{\eta_N}{2}) - \lambda_N - C_c) \|u_{M_{\text{red}}} - u_N\|^2_{L^2_N}
\end{aligned}$$

Notably, for $t = \frac{\lambda_N + C_c}{\lambda_N + C_c + \eta_N/2}$, we have

$$\langle (A^N_{u_N} - \lambda_N)(u_{M_{\text{red}}} - u_N), u_{M_{\text{red}}} - u_N \rangle_{L^2_N} \geq \frac{\eta_N}{2\lambda_N + 2C_c + \eta_N} \|u_N - u_{M_{\text{red}}}\|^2_{H^1}$$

We thus have

$$\begin{aligned}
E_N(u_{M_{\text{red}}}) - E_N(u_N) &\geq C_{\lambda_N} \|u_N - u_{M_{\text{red}}}\|^2_{H^1} + \frac{\kappa}{2} \|u^2_{M_{\text{red}}} - u^2_N\|_{L^2_N} \\
&\geq C_{\lambda_N} \|u_N - u_{M_{\text{red}}}\|^2_{H^1}
\end{aligned}$$

with $C_{\lambda_N} = \frac{\lambda_{N,2} - \lambda_N}{2C_c + \lambda_{N,2} + \lambda_N} > 0$. □

Using this result, we show that

Theorem 13. *There exists a constant $C > 0$, independant on N, such that, for any $k \geq 1$*

$$C_{\lambda_N} \|u_{M_{\text{red}}} - u_N\|^2_{H^1} \leq 2C \frac{1}{M^{2k}_{\text{red}}} \|V\|_{L^\infty} \|u_N\|^2_{H^k} + \frac{4C^2\kappa}{M^{2k}_{\text{red}}} \|u_N\|^4_{H^k}$$

Proof. We have for any normalized $v_{M_{\text{red}}} \in X_{M_{\text{red}}}$.

$$\begin{aligned}
C_{\lambda_N} \|u_{M_{\text{red}}} - u_N\|^2_{H^1} &\leq E_N(u_{M_{\text{red}}}) - E_N(u_N) \\
&\leq E_N(v_{M_{\text{red}}}) - E_N(u_N) \\
&\leq E_N(\frac{\Pi_{M_{\text{red}}} u_N}{\|\Pi_{M_{\text{red}}} u_N\|_{L^2_N}}) - E_N(u_N)
\end{aligned}$$

Let us now obtain the convergence rate.

$$C_{\lambda_N} \|u_{M_{\text{red}}} - u_N\|^2_{H^1} \leq E_N(\frac{\Pi_{M_{\text{red}}} u_N}{\|\Pi_{M_{\text{red}}} u_N\|_{L^2_N}}) - E_N(u_N)$$

$$
\begin{aligned}
&\leq a_N\left(\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}, \frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}\right) - a_N(u_N, u_N)\\
&+ \langle V, \left(\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}\right)^2 - u_N^2\rangle_{L_N^2}\\
&+ \kappa/2\langle\left(\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}\right)^4 - u_N^4, 1\rangle_{L_N^2}
\end{aligned}
$$

For the discretized kinetic term error, $\delta_{\mathcal{K},3} = \frac{a_N\left(\Pi_{M_{\text{red}}}u_N, \Pi_{M_{\text{red}}}u_N\right)}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}^2} - a_N(u_N, u_N)$ we use Lemma 11 to bound

$$
\begin{aligned}
\delta_{\mathcal{K},3} &= \sum_{|k|\leq M/4-1} |\hat{u}_k|^2 \frac{1 - \sum_{|k'|\leq M/4-1} |\hat{u}_{k'}|^2}{\sum_{|k'|\leq M/4-1} 4|\hat{u}_{k'}|^2} \sin\left(\pi\frac{k}{M}\right)^2 M^2\\
&- \sum_{|k|>M/4-1} 4|\hat{u}_k|^2 \sin\left(\pi\frac{k}{M}\right)^2 M^2\\
&\leq \left(1 - \|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}^2\right) 4\pi^2(M/4-1)^2\\
&- \|u_N - \Pi_{M_{\text{red}}}u_N\|_{L_N^2}^2 4\pi^2(M/4)^2\\
&\leq 0
\end{aligned}
$$

For the potential term error, $\delta_{\mathcal{P},3} = \left\langle V, \left(\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}\right)^2 - u_N^2 \right\rangle_{L_N^2}$ we have

$$
\begin{aligned}
\delta_{\mathcal{P},3} &= \langle V, \left(\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}} - u_N\right)\left(\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}} + u_N\right)\rangle_{L_N^2}\\
&\leq \|V\|_{L^\infty} \left\langle \left|\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}} - u_N\right|, \left|\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}} + u_N\right| \right\rangle_{L_N^2}\\
&= \|V\|_{L^\infty} \left\langle \left|\frac{\Pi_{M_{\text{red}}}u_N(1 - \|\Pi_{M_{\text{red}}}u_N\|_{L_N^2})}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}} - (u_N - \Pi_{M_{\text{red}}}u_N)\right|, \right.\\
&\qquad \left. \left|\frac{\Pi_{M_{\text{red}}}u_N(1 + \|\Pi_{M_{\text{red}}}u_N\|_{L_N^2})}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}} + (u_N - \Pi_{M_{\text{red}}}u_N)\right| \right\rangle_{L_N^2}\\
&\leq \|V\|_{L^\infty} \left\langle \left|\frac{\Pi_{M_{\text{red}}}u_N(1 - \|\Pi_{M_{\text{red}}}u_N\|_{L_N^2})}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}\right|, \right.\\
&\qquad \left. \left|\frac{\Pi_{M_{\text{red}}}u_N(1 + \|\Pi_{M_{\text{red}}}u_N\|_{L_N^2})}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}\right| + |(u_N - \Pi_{M_{\text{red}}}u_N)| \right\rangle_{L_N^2}\\
&+ \|V\|_{L^\infty} \left\langle |u_N - \Pi_{M_{\text{red}}}u_N|, \left|\frac{\Pi_{M_{\text{red}}}u_N(1 + \|\Pi_{M_{\text{red}}}u_N\|_{L_N^2})}{\|\Pi_{M_{\text{red}}}u_N\|_{L_N^2}}\right| + |(u_N - \Pi_{M_{\text{red}}}u_N)| \right\rangle_{L_N^2}
\end{aligned}
$$

$$
\begin{aligned}
&= \|V\|_{L^\infty}\left(1 - \|\Pi_{M_{\text{red}}}u_N\|^2_{L^2_N} + \|u_N - \Pi_{M_{\text{red}}}u_N\|^2_{L^2_N}\right)\\
&= 2\|V\|_{L^\infty}\|u_N - \Pi_{M_{\text{red}}}u_N\|^2_{L^2_N}\\
&\leq 2C\frac{1}{M^{2k}_{\text{red}}}\|V\|_{L^\infty}\|u^{(k)}_N\|^2_{L^2}
\end{aligned}
$$

For the interaction term error, $\delta_{\mathcal{I},3}$ we obtain that

$$
\begin{aligned}
\delta_{\mathcal{I},3} &= \langle(\Pi_{M_{\text{red}}}u_N)^4, 1\rangle_{L^2_N} - \langle u^4_N, 1\rangle_{L^2_N}\\
&= \left\langle \frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L^2_N}} - u_N,\right.\\
&\qquad \left.(\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L^2_N}})^3 + (\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L^2_N}})u^2_N + (\frac{\Pi_{M_{\text{red}}}u_N}{\|\Pi_{M_{\text{red}}}u_N\|_{L^2_N}})^2u_N + u^3_N\right\rangle_{L^2_N}\\
&= \frac{1}{\|\Pi_{M_{\text{red}}}u_N\|^3_{L^2_N}}\langle\Pi_{M_{\text{red}}}u_N - u_N, (\Pi_{M_{\text{red}}}u_N)^3 - \Pi_{M_{\text{red}}}\left(\Pi_{M_{\text{red}}}u_N\right)^3)\rangle_{L^2_N}\\
&+ \frac{1}{\|\Pi_{M_{\text{red}}}u_N\|^2_{L^2_N}}\langle\Pi_{M_{\text{red}}}u_N - u_N, (\Pi_{M_{\text{red}}}u_N)u^2_N - \Pi_{M_{\text{red}}}\left(\Pi_{M_{\text{red}}}u_N\right)u^2_N)\rangle_{L^2_N}\\
&+ \frac{1}{\|\Pi_{M_{\text{red}}}u_N\|_{L^2_N}}\langle\Pi_{M_{\text{red}}}u_N - u_N, (\Pi_{M_{\text{red}}}u_N)^2u_N - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)^2u_N\right)\rangle_{L^2_N}\\
&+ \langle\Pi_{M_{\text{red}}}u_N - u_N, u^3_N - \Pi_{M_{\text{red}}}(u^3_N)\rangle_{L^2_N}\\
&\leq \|\Pi_{M_{\text{red}}}u_N - u_N\|_{L^2_N}\Big(\|(\Pi_{M_{\text{red}}}u_N)^3 - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)^3\right)\|_{L^2_N}/\|\Pi_{M_{\text{red}}}u_N\|^3_{L^2_N}\\
&+ \|(\Pi_{M_{\text{red}}}u_N)u^2_N - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)u^2_N\right)\|_{L^2_N}/\|\Pi_{M_{\text{red}}}u_N\|^2_{L^2_N}\\
&+ \|(\Pi_{M_{\text{red}}}u_N)^2u_N - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)^2u_N\right)\|_{L^2_N}/\|\Pi_{M_{\text{red}}}u_N\|_{L^2_N}\\
&+ \|u^3_N - \Pi_{M_{\text{red}}}\left(u^3_N\right)\|_{L^2_N}\Big)\\
&\leq C\|\Pi_{M_{\text{red}}}u_N - u_N\|_{L^2}\Big(\|(\Pi_{M_{\text{red}}}u_N)^3 - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)^3\right)\|_{L^2}/\|\Pi_{M_{\text{red}}}u_N\|^3_{L^2}\\
&+ \|(\Pi_{M_{\text{red}}}u_N)u^2_N - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)u^2_N\right)\|_{L^2}/\|\Pi_{M_{\text{red}}}u_N\|^2_{L^2}\\
&+ \|(\Pi_{M_{\text{red}}}u_N)^2u_N - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)^2u_N\right)\|_{L^2}/\|\Pi_{M_{\text{red}}}u_N\|_{L^2}\\
&+ \|u^3_N - \Pi_{M_{\text{red}}}\left(u^3_N\right)\|_{L^2}\Big)
\end{aligned}
$$

where we obtain the last inequality from the fact that for any $v \in X_{3\cdot M/2+1}$

$$
\begin{aligned}
\|v\|_{L^2_N} = \|\mathrm{I}_N v\|_{L^2_N} &\leq 2\|\mathrm{I}_N v\|_{L^2} \leq 2\|v - \mathrm{I}_N v\|_{L^2} + 2\|v\|_{L^2}\\
&\leq CM^{-1}\|v\|_{H^1} + 2\|v\|_{L^2} = C\|v\|_{L^2}
\end{aligned}
$$

(where the last inequality holds from an inverse inequality between the H^1 and the L^2 norms for trigonometric polynomials and the fact that for any $v \in X_{M_{\text{red}}}$, $\|v_{M_{\text{red}}}\|_{L^2} = \|v_{M_{\text{red}}}\|_{L^2_N}$. We note that for any $k > 1/2$, and for $\alpha \in \{1, 2, 3\}$

$$
T(\alpha) = \|(\Pi_{M_{\text{red}}}u_N)^\alpha u^{3-\alpha}_N - \Pi_{M_{\text{red}}}\left((\Pi_{M_{\text{red}}}u_N)^\alpha u^{3-\alpha}_N\right)\|_{L^2}/\|\Pi_{M_{\text{red}}}u_N\|^\alpha_{L^2}
$$

$$
\begin{aligned}
&\leq M_{\text{red}}^{-k}\|((\Pi_{M_{\text{red}}}u_N)^{\alpha}u_N^{3-\alpha})^{(k)}\|_{L^2}/\|\Pi_{M_{\text{red}}}u_N\|_{L^2}^{\alpha}\\
&\leq CM_{\text{red}}^{-k}\|u_N\|_{H^k}^{3-\alpha}\|\Pi_{M_{\text{red}}}u_N\|_{H^k}^{\alpha}/\|\Pi_{M_{\text{red}}}u_N\|_{L^2}^{\alpha}\\
&\leq CM_{\text{red}}^{-k}\|u_N\|_{H^k}^{3}
\end{aligned}
$$

where we use the fact that

$$\|\Pi_{M_{\text{red}}}u_N - u_N\|_{L^2} \leq CM_{\text{red}}^{-k}\|u_N^{(k)}\|_{L^2} \leq CM_{\text{red}}^{-k}\|u_N\|_{H^k}$$

for the first inequality. The second inequality arises from the fact that H^k is an algebra for any $k > 1/2$. And we use the following lemma for the last inequality:

Lemma 14. *Given the normalised truncation of the first $n > 0$ terms of the Fourier series of a normalised function $v_N \in \mathcal{X}_N$, we have that*

$$\frac{\|\Pi_n v_N\|_{H^k}}{\|\Pi_n v_N\|_{L^2}} \leq \frac{\|v_N\|_{H^k}}{\|v_N\|_{L^2}}$$

Proof. This result comes from the fact that the ratio of H^k to L^2 norms is strictly increasing, which we now show. We have that :

$$\frac{\|\Pi_n v_N\|_{H^k}}{\|\Pi_n v_N\|_{L^2}} = \sum_{k=-n}^{n} \frac{|\hat{v}_k|^2}{\sum_{k'=-n}^{n}|\hat{v}_k|^2}(1+k^2)^k$$

Hence :

$$
\begin{aligned}
\frac{\|\Pi_{n+1} v_N\|_{H^k}}{\|\Pi_{n+1} v_N\|_{L^2}} - \frac{\|\Pi_n v_N\|_{H^k}}{\|\Pi_n v_N\|_{L^2}} &= \sum_{k=-n}^{n}|\hat{v}_k|^2(1+k^2)^k\left(\frac{1}{\sum_{k'=-n-1}^{n+1}|\hat{v}_k|^2} - \frac{1}{\sum_{k'=-n}^{n}|\hat{v}_k|^2}\right)\\
&\quad + \frac{|\hat{v}_{n+1}|^2 + |\hat{v}_{-(n+1)}|^2}{\sum_{k'=-n-1}^{n+1}|\hat{v}_k|^2}(1+(n+1)^2)^k\\
&= -\sum_{k=-n}^{n}|\hat{v}_k|^2(1+k^2)^k\frac{|\hat{v}_{n+1}|^2 + |\hat{v}_{-(n+1)}|^2}{\sum_{k'=-n-1}^{n+1}|\hat{v}_k|^2\sum_{k''=-n}^{n}|\hat{v}_k|^2}\\
&\quad + \frac{|\hat{v}_{n+1}|^2 + |\hat{v}_{-(n+1)}|^2}{\sum_{k'=-n-1}^{n+1}|\hat{v}_k|^2}(1+(n+1)^2)^k\\
&= \frac{|\hat{v}_{n+1}|^2 + |\hat{v}_{-(n+1)}|^2}{\sum_{k'=-n-1}^{n+1}|\hat{v}_k|^2}\Bigg((1+(n+1)^2)^k\\
&\qquad\qquad - \sum_{k=-n}^{n}\frac{|\hat{v}_k|^2}{\sum_{k'=-n}^{n}|\hat{v}_k|^2}(1+k^2)^k\Bigg)\\
&> 0
\end{aligned}
$$

□

We are thus able to bound the interaction term error as follows

$$\delta_{\mathcal{I},3} \leq 4C^2M_{\epsilon}^{-2k}\|u_N\|_{H^k}^4$$

Combining the three results, we have

$$C_{\lambda_N}\|u_{M_{\text{red}}} - u_N\|_{H^1}^2 \leq 2C\frac{1}{M_{\text{red}}^{2k}}\|V\|_{L^\infty}\|u_N\|_{H^k}^2 + \frac{4C^2\kappa}{M_{\text{red}}^{2k}}\|u_N\|_{H^k}^4$$

□

Term (2)

For the term (2), we have that

Theorem 14. *Let $u_{M_{\text{red}}}$ be the solution to* (A.3) *and $\tilde{u}_{M_{\text{red}}}$ the solution to* (A.4)*, then, the H^2 norm of $u_{M_{\text{red}}}$ is uniformly bounded and the H^1 norm of $\tilde{u}_{M_{\text{red}}} - u_{M_{\text{red}}}$ converges to zero as*

$$\|\tilde{u}_{M_{\text{red}}} - u_{M_{\text{red}}}\|_{H^1} \leq \frac{C}{M}(\|u_{M_{\text{red}}}\|_{H^2} + \|\tilde{u}_{M_{\text{red}}}\|_{H^2} + \|V\|_{H^1}).$$

Proof. Starting again by the bound of the norm of the difference with the difference in energies

$$\begin{aligned}
C\|\tilde{u}_{M_{\text{red}}} - u_{M_{\text{red}}}\|_{H^1}^2 &\leq E(u_{M_{\text{red}}}) - E(\tilde{u}_{M_{\text{red}}}) \\
&= E_N(u_{M_{\text{red}}}) + (E(u_{M_{\text{red}}}) - E_N(u_{M_{\text{red}}})) - E(\tilde{u}_{M_{\text{red}}}) \\
&\leq (E_N - E)(\tilde{u}_{M_{\text{red}}}) + (E - E_N)(u_{M_{\text{red}}}) \\
&= (a - a_N)(u_{M_{\text{red}}}, u_{M_{\text{red}}}) - (a - a_N)(\tilde{u}_{M_{\text{red}}}, \tilde{u}_{M_{\text{red}}}) \\
&+ \langle V, u_{M_{\text{red}}}^2 - \tilde{u}_{M_{\text{red}}}^2\rangle_{L^2} - \langle \mathrm{I}_N V, \mathrm{I}_N u_{M_{\text{red}}}^2 - \mathrm{I}_N \tilde{u}_{M_{\text{red}}}^2\rangle_{L^2} \\
&+ \kappa/2\langle u_{M_{\text{red}}}^4 - \tilde{u}_{M_{\text{red}}}^4, 1\rangle_{L^2} - \kappa/2\langle \mathrm{I}_N u_{M_{\text{red}}}^4 - \mathrm{I}_N \tilde{u}_{M_{\text{red}}}^4, 1\rangle_{L^2}
\end{aligned}$$

The choice of M_{red} such that $4 \cdot M_{\text{red}} \leq M/2 - 1$, gives us that $\mathrm{I}_N \tilde{u}_{M_{\text{red}}}^4 = \tilde{u}_{M_{\text{red}}}^4$ and $\mathrm{I}_N u_{M_{\text{red}}}^4 = u_{M_{\text{red}}}^4$ as well as $\mathrm{I}_N \tilde{u}_{M_{\text{red}}}^2 = \tilde{u}_{M_{\text{red}}}^2$ and $\mathrm{I}_N u_{M_{\text{red}}}^2 = u_{M_{\text{red}}}^2$. We thus have that

$$\begin{aligned}
\|\tilde{u}_{M_{\text{red}}} - u_{M_{\text{red}}}\|_{H^1}^2 &\leq (a - a_N)(u_{M_{\text{red}}} - \tilde{u}_{M_{\text{red}}}, u_{M_{\text{red}}} + \tilde{u}_{M_{\text{red}}}) \\
&+ \langle V - \mathrm{I}_N V, u_{M_{\text{red}}}^2 - \tilde{u}_{M_{\text{red}}}^2\rangle_{L^2}
\end{aligned}$$

Let us consider the discretization error of the Laplacian term.

$$\begin{aligned}
(a - a_N)(u_{M_{\text{red}}} - \tilde{u}_{M_{\text{red}}}, u_{M_{\text{red}}} + \tilde{u}_{M_{\text{red}}}) &\leq \frac{C}{M^2}\|u_{M_{\text{red}}} - \tilde{u}_{M_{\text{red}}}\|_{H^2}\|u_{M_{\text{red}}} + \tilde{u}_{M_{\text{red}}}\|_{H^2} \\
&\leq \frac{C}{M}\|u_{M_{\text{red}}} - \tilde{u}_{M_{\text{red}}}\|_{H^1}\|u_{M_{\text{red}}} + \tilde{u}_{M_{\text{red}}}\|_{H^2}
\end{aligned}$$

the last bound being derived from an inverse inequality between H^2 and H^1-norms. For the potential term, we have :

$$\begin{aligned}\langle V - \mathrm{I}_N V, u_{M_{\mathrm{red}}}^2 - \tilde{u}_{M_{\mathrm{red}}}^2 \rangle_{L^2} &\leq \|V - \mathrm{I}_N V\|_{L^2} \|u_{M_{\mathrm{red}}} - \tilde{u}_{M_{\mathrm{red}}}\|_{L^\infty} \|u_{M_{\mathrm{red}}} + \tilde{u}_{M_{\mathrm{red}}}\|_{L^2} \\ &\leq 2C \|V - \mathrm{I}_N V\|_{L^2} \|u_{M_{\mathrm{red}}} - \tilde{u}_{M_{\mathrm{red}}}\|_{H^1}^{1/2} \|u_{M_{\mathrm{red}}} - \tilde{u}_{M_{\mathrm{red}}}\|_{L^2}^{1/2} \\ &\leq 2C \|V - \mathrm{I}_N V\|_{L^2} \|u_{M_{\mathrm{red}}} - \tilde{u}_{M_{\mathrm{red}}}\|_{H^1} \\ &\leq \frac{2C}{M} \|V\|_{H^1} \|u_{M_{\mathrm{red}}} - \tilde{u}_{M_{\mathrm{red}}}\|_{H^1}\end{aligned}$$

Hence we have :

$$\|\tilde{u}_{M_{\mathrm{red}}} - u_{M_{\mathrm{red}}}\|_{H^1} \leq \frac{C}{M} (\|u_{M_{\mathrm{red}}}\|_{H^2} + \|\tilde{u}_{M_{\mathrm{red}}}\|_{H^2} + \|V\|_{H^1})$$

Note that in the above estimate, the H^2-norm of $u_{M_{\mathrm{red}}}$ and $\tilde{u}_{M_{\mathrm{red}}}$ are well defined, as norms of elements in X_M, but may depend on M.

Lemma 15. *The H^2 norm of u_N is bounded uniformly in N.*

Proof. Given that for any $v_N \in \mathcal{X}_N$,

$$\langle A_{u_N} u_N, v_N \rangle_{L_N^2} = \lambda_N \langle u_N, v_N \rangle_{L_N^2}.$$

We have in particular that :

$$\langle \Delta_N u_N, e_k \rangle_{L_N^2} = 2 \underbrace{\langle (V + \frac{\kappa}{2} u_N^2 - \lambda_N) u_N}_{f_N}, e_k \rangle_{L_N^2}$$

$$\begin{aligned}4 \cdot M^2 \sin\left(\frac{\pi k}{M}\right)^2 \hat{u}_k &= 2 \hat{f}_k \\ |\hat{u}_k| &= \frac{1}{2 \cdot M^2 \sin\left(\frac{\pi k}{M}\right)^2} |\hat{f}_k| \\ &\leq \frac{1}{\pi^2 k^2} |\hat{f}_k|\end{aligned}$$

Where the last inequality is obtained as $\sin\left(\frac{\pi k}{M}\right)^2 \geq \frac{\pi^2 k^2}{2 \cdot M^2}$. We note that this is also true for $\hat{u}_{M/2} = \langle u_N, c_{M/2} \rangle_{L_N^2}$ as the multiplicative error in the scalar product cancels on both sides.

As we know that $u_N, V \in H^1(\Omega)$, we have that $V u_N$ is in H^1 as H^s is an algebra for $s > 1/2$ in one dimension. Furthermore, as $u_N \in H^1(\Omega)$, we have that $u_N^3 \in H^1(\Omega)$ for the same reason. Finally, we have that $\lambda_N > 0$ and is uniformly bounded. Hence we have that for any N, $f_N \in H^1$ and thus that its L^2 norm can be uniformly bounded. Thus we have that

$$\|u_N\|_{H^2}^2 = \sum_k |\hat{u}_k|^2 (1 + 4\pi^2 k^2)^2 \leq \sum_k |\hat{f}_k|^2 \frac{(1 + 4\pi^2 k^2)^2}{\pi^4 k^4} \leq C \|f_N\|_{L^2}^2$$

We now show that $\|f_N\|_{L^2}$ is uniformly bounded. Indeed we have that

$$\begin{aligned}\|f_N\|_{L^2}^2 &= \langle |u_N|^2, \left(V^2 + \kappa u_N^2 V + \frac{\kappa^2}{4}u_N^4 + \lambda_N^2 - 2\lambda_N(V + \frac{\kappa}{2}u_N^2)\right)\rangle_{L^2} \\ &\leq \|V^2\|_{L^\infty}\|u_N\|_{L^2}^2 + \kappa\|V\|_{L^\infty}\|u_N\|_{L^4}^4 + \frac{\kappa^2}{4}\|u_N\|_{L^6}^6 + \lambda_N^2\|u_N\|_{L^2} \\ &- 2\lambda_N\|V\|_{L^\infty}\|u_N\|_{L^2}^2 - 2\lambda_N\|u_N\|_{L^4}^4 \\ &\leq \|V^2\|_{L^\infty}\|u_N\|_{L^2}^2 + \kappa\|V\|_{L^\infty}\|u_N\|_{L^4}^4 + \frac{\kappa^2}{4}\|u_N\|_{L^6}^6 + \lambda_N^2\|u_N\|_{L^2}\end{aligned}$$

The u_N function is L^2 normalised and the Gagliardo–Nirenberg inequality for $p = 6, j = 0, r = 2q = 2$ gives us :

$$\|v\|_{L^6} \leq C\|u'\|_{L^2}^{1/3}\|u\|_{L^2}^{2/3}$$

And for $p = 4, j = 0, r = 2, q = 2$:

$$\|v\|_{L^4} \leq C\|u'\|_{L^2}^{1/4}\|u\|_{L^2}^{2/4}$$

This gives us:

$$\|f_N\|_{L^2}^2 \leq \|V^2\|_{L^\infty} + \kappa\|V\|_{L^\infty}\|u_N\|_{H^1} + \frac{\kappa^2}{4}\|u_N\|_{H^1}^2 + \lambda_N^2$$

We now show that

Lemma 16. *$E_N(u_N)$ and λ_N are uniformly bounded, namely, there exists a $C > 0$ such that $\forall N > 2$,*

$$E_N(u_N) \leq CE(u),\ \lambda_N \leq 2CE(u)$$

Proof. To show the above bound, we consider the truncation $\Pi_{M_{\text{red}}}u$ of u, i.e. its truncation on $X_{M_{\text{red}}}$, this gives us:

$$\begin{aligned}E_N(u_N) &= \operatorname*{arg\,min}_{v_N \in \mathcal{X}_N} E_N(v_N) \\ &\leq E_N(\Pi_{M_{\text{red}}}u) \\ &= E(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) + (E_N - E)(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) \\ &\leq E(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) \\ &- \frac{1}{2\|\Pi_{M_{\text{red}}}u\|_{L^2}^2}\langle\Delta_N\Pi_{M_{\text{red}}}u, \Pi_{M_{\text{red}}}u\rangle_{L_N^2} + \frac{1}{2\|\Pi_{M_{\text{red}}}u\|_{L^2}^2}\langle\Delta\Pi_{M_{\text{red}}}u, \Pi_{M_{\text{red}}}u\rangle_{L^2} \\ &+ \frac{1}{\|\Pi_{M_{\text{red}}}u\|_{L^2}^2}\langle V, |\Pi_{M_{\text{red}}}u|^2\rangle_{L_N^2} - \frac{1}{\|\Pi_{M_{\text{red}}}u\|_{L^2}^2}\langle V, |\Pi_{M_{\text{red}}}u|^2\rangle_{L^2} \\ &+ \frac{\kappa}{2\|\Pi_{M_{\text{red}}}u\|_{L^2}^4}\langle|\Pi_{M_{\text{red}}}u|^4, 1\rangle_{L_N^2} - \frac{\kappa}{2\|\Pi_{M_{\text{red}}}u\|_{L^2}^4}\langle|\Pi_{M_{\text{red}}}u|^4, 1\rangle_{L^2} \\ &\leq E(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) + \frac{\|\Pi_{M_{\text{red}}}u\|_{H^1}^2}{2\|\Pi_{M_{\text{red}}}u\|_{L^2}^2} + \frac{1}{\|\Pi_{M_{\text{red}}}u\|_{L^2}^2}\langle \mathrm{I}_N V - \langle V, |\Pi_{M_{\text{red}}}u|^2\rangle_{L^2} \\ &\leq E(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) + \frac{\|\Pi_{M_{\text{red}}}u\|_{H^1}^2}{2\|\Pi_{M_{\text{red}}}u\|_{L^2}^2} + \frac{1}{\|\Pi_{M_{\text{red}}}u\|_{L^2}^2}C\langle V, |\Pi_{M_{\text{red}}}u|^2\rangle_{L^2}\end{aligned}$$

$$
\begin{aligned}
&\leq CE(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) \\
&\leq CE(u) + C(E(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) - E(u)) \\
&\leq CE(u) + C(2(\|u\|_{H^1}+1)\|u\|_{H^1}) \\
&\leq CE(u) + 4C\|u\|_{H^1}^2 \\
&\leq CE(u)
\end{aligned}
$$

where we use the bound $E(\frac{\Pi_{M_{\text{red}}}u}{\|\Pi_{M_{\text{red}}}u\|_{L^2}}) - E(u) \leq (2(\|u\|_{H^1}+1)\|u\|_{H^1}$ from [28] eq. 2.23. The bound on λ_N is obtained immediately as

$$\lambda_N = E(u_N) + \frac{\kappa}{2}\langle u_N^4, 1\rangle_{L_N^4} \leq 2E(u_N).$$

□

Hence, we have from the lemma that $\|u_N\|_{H^1}$ and λ_N are uniformly bounded, and so is $\|f_N\|_{L^2}$. We thus obtain that $\|u_N\|_{H^2}$ is uniformly bounded. □

and using the exact same argument, we have that

Corollary 4. *The H^2 norm of $u_{M_{\text{red}}}$ is bounded uniformly in N.*

□

The convergence rate of u_N can thus be obtained by combining the three theorems above, this gives us :

Theorem 15.

$$
\begin{aligned}
\|u - u_N\|_{H^1} \leq C\Bigg(&\frac{\|u\|_{H^{k_1}}}{M^{(k_1-1)}} + \frac{\|V\|_{H^1}}{M} + \frac{1}{M}(\|u_{M_{\text{red}}}\|_{H^2} + \|\tilde{u}_{M_{\text{red}}}\|_{H^2}) \\
&+ \frac{1}{M^{k_2}}\sqrt{\|V\|_{L^\infty}\|u_N\|_{H^{k_2}}^2} + \frac{\|u_N\|_{H^{k_2}}}{M^{k_2}}\Bigg),
\end{aligned}
$$

with $0 \leq k_1 < \infty, 0 \leq k_2 \leq 2$.

Proof. We have that

$$\frac{1}{M_{\text{red}}} = \frac{4}{M-4} \leq \frac{4M}{(M-4)M} = \frac{4}{M} + \frac{4}{M(M-4)} \leq C\frac{1}{M}$$

We thus bound the $\frac{1}{M_{\text{red}}}$ from Theorem 12 and 13 by $\frac{1}{M}$, and hence the result. □

From this theorem, we see that the discretized Laplacian prohibits us from obtaining a better convergence rate than $\frac{1}{M}$. This discretization is not "standard" in the spectral community and can be eliminated by taking an alternative definition for a_N based on the use of $\tilde{a}_N(v_N, w_N) = \langle v_N', w_N'\rangle_N$. Nevertheless, with this choice of potential V with a discontinuity of the derivative on the border, the maximum convergence rate is $M^{-3/2}$.

A.2 Walsh interpolation

Similarly to the interpolation operator used in the context of a discretized Fourier basis (see Chapter 5.1 of [15]), we can define an interpolation operator for the Walsh basis. Let $N \geq 1$, and set $\mathcal{W}_N = \mathrm{Span}_{\mathbb{C}}\{w_k \mid 0 \leq k \leq 2^N - 1\}$.

Definition 19. *The Walsh interpolation operator* $\mathrm{I}_N^{\mathcal{W}} : L^2(0,1) \to \mathcal{W}_N$ *is defined, for any* $v \in L^2(0,1)$*, by*

$$\langle \mathrm{I}_N^{\mathcal{W}} v, g \rangle_{L_N^2} = \langle v, g \rangle_{L_N^2}, \forall g \in L^2(0,1).$$

Equivalently, it is the N*-bit Walsh series in* $\mathcal{W}_N$ *that takes the same values as* v *on* Ω_N*.*

Similarly to the Fourier interpolation, we can decompose the Walsh interpolation operator as the sum of the truncation operator Θ_N and an aliasing operator R_N. Indeed, we have the following

Proposition 25. *For any* $v \in L^2(0,1)$*, the Walsh interpolant of* v *is*

$$\mathrm{I}_N^{\mathcal{W}} v = \sum_{m=0}^{2^N-1} \langle v, w_m \rangle_{L_N^2} w_m = \sum_{m=0}^{2^N-1} \left(\tilde{v}_m + \sum_{\ell=1}^{+\infty} \tilde{v}_{m \oplus 2^N \ell} \right) w_m \tag{A.5}$$

Furthermore, we can write the N*-bit Walsh interpolation operator as*

$$\mathrm{I}_N^{\mathcal{W}} = \Theta_N + \mathrm{R}_N^{\mathcal{W}}, \tag{A.6}$$

where Θ_N *is the* N*-bit Walsh truncation operator* (1.20)*, and* $\mathrm{R}_N^{\mathcal{W}}$ *the* N*-bit Walsh aliasing operator such that, for any* $v \in L^2(0,1)$*,*

$$\mathrm{R}_N^{\mathcal{W}} v = \sum_{m=0}^{2^N-1} \left(\sum_{\ell=1}^{+\infty} \tilde{v}_{m \oplus 2^N \ell} \right) w_m.$$

Proof. Let $k \in \mathbb{N}$ such that $0 \leq k < 2^N$, we have

$$\langle v, w_k \rangle_{L_N^2} = \sum_{\ell=0}^{+\infty} \tilde{v}_\ell \langle w_\ell, w_k \rangle_{L_N^2}.$$

If $\ell \neq k$ and $\ell \neq 0$, for any $x \in \Omega_N$, we write

$$w_\ell(x) = \prod_{n=1}^{N} R_n(x)^{\ell_n + k_n} \cdot \prod_{n=N+1}^{+\infty} R_n(x)^{\ell_n} = \prod_{n=1}^{N} R_n(x)^{\ell_n + k_n},$$

where the infinite product equals 1 as each of its terms is 1 on Ω_N. Hence, the L_N^2 scalar product reduces to

$$\langle w_\ell, w_k \rangle_{L_N^2} = \sum_{j=0}^{2^N-1} R_1^{\ell_1 + k_1} \dots R_N^{\ell_N + k_N}(j/2^N) = \langle w_{\mathrm{Trunc}_N(\ell)}, w_k \rangle_{L_N^2},$$

where, for any $j \in \mathbb{N}$, $\mathrm{Trunc}_N(j) = \sum_{n=1}^{N} j_n 2^{n-1}$ is the N-bit truncation of the integer j. We thus obtain

$$\langle w_\ell, w_k \rangle_{L_N^2} = \delta_{\mathrm{Trunc}_N(\ell) = k},$$

which implies that $\langle w_\ell, w_k\rangle_{L^2_N} = 1$ for all terms of the form $\ell = k + 2^N\ell'$, with $\ell' \in \mathbb{N}$, and is equal to zero otherwise. Hence, we obtain

$$\langle v, w_k\rangle_{L^2_N} = \sum_{\ell\in\mathbb{N}} \langle v, w_{k+2^N\ell}\rangle_{L^2} = \langle v, w_k\rangle_{L^2} + \sum_{\ell>0} \langle v, w_{k\oplus 2^N\ell}\rangle_{L^2} = \tilde{v}_k + \sum_{\ell=1}^{+\infty} \tilde{v}_{k\oplus 2^N\ell}$$

which gives (A.5), and (A.6) by identification of the aliasing of v. □

We now provide a useful result on the Walsh interpolant of the square of a function.

Theorem 16. *For any function $v \in L^2(0,1)$, we have the interpolant of the square of v is the square of the interpolant of v, i.e.*

$$\mathrm{I}_N^{\mathcal{W}}(v^2) = (\mathrm{I}_N^{\mathcal{W}} v)^2$$

Proof. On the one hand, recall that

$$\mathrm{I}_N^{\mathcal{W}} v = \sum_{i=0}^{2^N-1} \tilde{v}_i^N w_i.$$

This gives

$$(\mathrm{I}_N^{\mathcal{W}} v)^2 = \sum_{i=0}^{2^N-1}\sum_{k=0}^{2^N-1} \tilde{v}_i^N \tilde{v}_{i\oplus k}^N w_k$$

where we denote $\tilde{v}_k^N = \langle v, w_k\rangle_{L^2_N}$. By replacing the interpolant coefficients $\tilde{v}_k^N$ with their value in terms of $\tilde{v}_k$, namely

$$\tilde{v}_k^N = \sum_{\ell=0}^{\infty} \tilde{v}_{k\oplus 2^N\ell},$$

we obtain

$$(\mathrm{I}_N^{\mathcal{W}} v)^2 = \sum_{k=0}^{2^N-1} \left(\sum_{i=0}^{2^N-1}\sum_{\ell=0}^{\infty}\sum_{\ell'=0}^{\infty} \tilde{v}_{i\oplus 2^N\ell'} \tilde{v}_{i\oplus k\oplus 2^N\ell} \right) w_k. \tag{A.7}$$

On the other hand, using the definition of $v = \sum_{i=0}^{\infty} \tilde{v}_i w_i$, and the property that for a pair $i, j \in \mathbb{N}$, there exists a unique $k \in \mathbb{N}$ such that $i \oplus j = k$, we have

$$v^2 = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \tilde{v}_i\tilde{v}_j w_{i\oplus j} = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{\infty} \tilde{v}_i \tilde{v}_{i\oplus k} \right) w_k.$$

Hence, the interpolant of v^2 is given by

$$\begin{aligned}\mathrm{I}_N^{\mathcal{W}}(v^2) &= \sum_{k=0}^{2^N-1} \left(\sum_{\ell=0}^{\infty}\sum_{i=0}^{\infty} \tilde{v}_i \tilde{v}_{i\oplus k\oplus 2^N\ell} \right) w_k \\ &= \sum_{k=0}^{2^N-1} \left(\sum_{i'=0}^{2^N-1}\sum_{\ell'=0}^{\infty}\sum_{\ell=0}^{\infty} \tilde{v}_{i'\oplus 2^N\ell'} \tilde{v}_{i'\oplus k\oplus 2^N\ell\oplus 2^N\ell'} \right) w_k,\end{aligned}$$

where we obtain the second line by setting $i = i' + 2^N\ell$.

Using the fact that, for fixed values of $\ell, \ell' \in \mathbb{N}$, there exists a unique $\ell'' = \ell \oplus \ell'$, we replace $2^N \ell \oplus 2^N \ell'$ by $2^N \ell''$ giving us

$$\mathrm{I}_N^{\mathcal{W}}(v^2) = \sum_{k=0}^{2^N-1} \left(\sum_{i'=0}^{2^N-1} \sum_{\ell'=0}^{\infty} \sum_{\ell''=0}^{\infty} \tilde{v}_{i' \oplus 2^N \ell'} \tilde{v}_{i' \oplus k \oplus 2^N \ell''} \right) w_k \tag{A.8}$$

and we obtain the result as (A.7) and (A.8) are equal. □

Bibliography

[1] R. Acharya et al. « Suppressing Quantum Errors by Scaling a Surface Code Logical Qubit ». In: *Nature* 614.7949 (Feb. 2023), pp. 676–681.

[2] A. S. Albino, L. C. Jardim, D. C. Knupp, A. J. S. Neto, O. M. Pires, and E. G. S. Nascimento. *Solving Partial Differential Equations on Near-Term Quantum Computers*. Aug. 2022. arXiv: `2208.05805 [physics, physics:quant-ph]`.

[3] M. Ali and A. Nouy. « Approximation theory of tree tensor networks: tensorized univariate functions ». In: *Constr. Approx.* 58.2 (2023), pp. 463–544.

[4] A. Ambainis. « Variable time amplitude amplification and quantum algorithms for linear algebra problems ». In: *29th International Symposium on Theoretical Aspects of Computer Science*. Vol. 14. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2012, pp. 636–647.

[5] D. An and L. Lin. « Quantum linear system solver based on time-optimal adiabatic quantum computing and quantum approximate optimization algorithm ». In: *ACM Trans. Quantum Comput.* 3.2 (2022), Art. No. 5, 28.

[6] D. An, J.-P. Liu, and L. Lin. « Linear combination of Hamiltonian simulation for nonunitary dynamics with optimal state preparation cost ». In: *Phys. Rev. Lett.* 131.15 (2023), Paper No. 150603, 6.

[7] D. W. Berry, A. M. Childs, A. Ostrander, and G. Wang. « Quantum algorithm for linear differential equations with exponentially improved dependence on precision ». In: *Comm. Math. Phys.* 356.3 (2017), pp. 1057–1081.

[8] D. P. Bertsekas. *Nonlinear programming*. Third. Athena Scientific Optimization and Computation Series. Athena Scientific, Belmont, MA, 2016, pp. xviii+861.

[9] K. Bharti et al. « Noisy intermediate-scale quantum algorithms ». In: *Rev. Mod. Phys.* 94 (1 2022), p. 015004.

[10] N. S. Blunt, L. Caune, R. Izsák, E. T. Campbell, and N. Holzmann. « Statistical Phase Estimation and Error Mitigation on a Superconducting Quantum Processor ». In: *PRX Quantum* 4 (4 2023), p. 040341.

[11] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles. « Variational Quantum Linear Solver ». In: *Quantum* 7 (Nov. 2023), p. 1188.

[12] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011, pp. xiv+599.

[13] J. C. Bridgeman and C. T. Chubb. « Hand-waving and interpretive dance: an introductory course on tensor networks ». In: *J. Phys. A* 50.22 (2017), pp. 223001, 61.

[14] Z. Cai et al. « Quantum error mitigation ». In: *Rev. Modern Phys.* 95.4 (2023), Paper No. 045005, 37.

[15] C. Canuto, ed. *Spectral Methods: Fundamentals in Single Domains*. Scientific Computation. Berlin ; New York: Springer-Verlag, 2006. 563 pp.

[16] C. Canuto, M. Y. Hussaini, and A. Quarteroni. *Spectral Methods: Evolution to Complex Geometries and Applications to Fluid Dynamics*. Scientific Computation. Berlin Heidelberg: Springer e-books, 2007.

[17] M. Cerezo et al. « Does Provable Absence of Barren Plateaus Imply Classical Simulability? Or, Why We Need to Rethink Variational Quantum Computing ». In: arXiv:2312.09121 (Mar. 2024). arXiv: `2312.09121 [quant-ph, stat]`.

[18] M. Cerezo et al. « Variational Quantum Algorithms ». In: *Nature Reviews Physics* 3.9 (Sept. 2021), pp. 625–644.

[19] C. C. Chang, K. S. McElvain, E. Rrapaj, and Y. Wu. « Improving Schrödinger Equation Implementations with Gray Code for Adiabatic Quantum Computers ». In: *PRX Quantum* 3 (2 2022), p. 020356.

[20] C. F. Chen and C. H. Hsiao. « A Walsh series direct method for solving variational problems ». In: *J. Franklin Inst.* 300.4 (1975), pp. 265–280.

[21] A. M. Childs, R. Kothari, and R. D. Somma. « Quantum algorithm for systems of linear equations with exponentially improved dependence on precision ». In: *SIAM J. Comput.* 46.6 (2017), pp. 1920–1950.

[22] A. M. Childs and J.-P. Liu. « Quantum spectral methods for differential equations ». In: *Comm. Math. Phys.* 375.2 (2020), pp. 1427–1457.

[23] A. M. Childs and N. Wiebe. « Hamiltonian simulation using linear combinations of unitary operations ». In: *arXiv preprint arXiv:1202.5822* (2012).

[24] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio. « Error mitigation with Clifford quantum-circuit data ». In: *Quantum* 5 (Nov. 2021), p. 592.

[25] J. Dick. « The Decay of the Walsh Coefficients of Smooth Functions ». In: *Bull. Aust. Math. Soc.* 80.3 (Dec. 2009), pp. 430–453.

[26] J. Dick. « Walsh spaces containing smooth functions and quasi-Monte Carlo rules of arbitrary high order ». In: *SIAM J. Numer. Anal.* 46.3 (2008), pp. 1519–1553.

[27] Z. Ding and L. Lin. « Even Shorter Quantum Circuit for Phase Estimation on Early Fault-Tolerant Quantum Computers with Applications to Ground-State Energy Estimation ». In: *PRX Quantum* 4 (2 2023), p. 020331.

[28] G. Dusson and Y. Maday. « A *Posteriori* Analysis of a Nonlinear Gross–Pitaevskii-type Eigenvalue Problem ». In: *IMA J Numer Anal* 37.1 (Jan. 2017), pp. 94–137.

[29] D. Fang, L. Lin, and Y. Tong. « Time-marching based quantum solvers for time-dependent linear differential equations ». In: *Quantum* 7 (Mar. 2023), p. 955.

[30] E. Farhi, J. Goldstone, and S. Gutmann. « A Quantum Approximate Optimization Algorithm ». In: *arXiv: Quantum Physics* (2014), null.

[31] C. Feniou, M. Hassan, D. Traoré, E. Giner, Y. Maday, and J.-P. Piquemal. « Overlap-ADAPT-VQE: Practical Quantum Chemistry on Quantum Computers via Overlap-Guided Compact Ansätze ». In: *Communications Physics* 6.1 (July 2023), p. 192.

[32] N. J. Fine. « On the Walsh Functions ». In: *Transactions of the American Mathematical Society* 65.3 (1949), pp. 372–414.

[33] N. J. Fine. « The Generalized Walsh Functions ». In: *Transactions of the American Mathematical Society* 69 (1950), pp. 66–77.

[34] P. García-Molina, J. Rodríguez-Mediavilla, and J. J. García-Ripoll. « Quantum Fourier analysis for multivariate functions and applications to a class of Schrödinger-type partial differential equations ». In: *Phys. Rev. A* 105 (1 2022), p. 012433.

[35] G. Gát and R. Toledo. « Numerical solution of linear differential equations by Walsh polynomials approach ». In: *Studia Sci. Math. Hungar.* 57.2 (2020), pp. 217–254.

[36] O. Gervasi et al., eds. *Computational science and its applications—ICCSA 2016. Part I.* Vol. 9786. Lecture Notes in Computer Science. Springer, [Cham], 2016, pp. xxvii+650.

[37] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng. « Digital zero noise extrapolation for quantum error mitigation ». In: (2020), pp. 306–316.

[38] M. Griebel and H. Harbrecht. « Analysis of Tensor Approximation Schemes for Continuous Functions ». In: *Foundations of Computational Mathematics* 23.1 (Feb. 2023), pp. 219–240.

[39] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall. « An Adaptive Variational Algorithm for Exact Molecular Simulations on a Quantum Computer ». In: *Nature Communications* 10.1 (July 2019), p. 3007.

[40] L. Grover and T. Rudolph. *Creating Superpositions That Correspond to Efficiently Integrable Probability Distributions*. Aug. 2002. arXiv: `quant-ph/0208112`.

[41] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo. « Measurements of quantum Hamiltonians with locally-biased classical shadows ». In: *Comm. Math. Phys.* 391.3 (2022), pp. 951–967.

[42] A. W. Harrow, A. Hassidim, and S. Lloyd. « Quantum algorithm for linear systems of equations ». In: *Phys. Rev. Lett.* 103.15 (2009), pp. 150502, 4.

[43] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles. « Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus ». In: *PRX Quantum* 3 (1 2022), p. 010313.

[44] H.-Y. Huang, R. Kueng, and J. Preskill. « Predicting many properties of a quantum system from very few measurements ». In: *Nature Physics* 16.10 (2020), pp. 1050–1057.

[45] M. Ippoliti. *Classical Shadows Based on Locally-Entangled Measurements*. May 18, 2023. arXiv: `2305.10723 [cond-mat, physics:quant-ph]`. URL: `http://arxiv.org/abs/2305.10723` (visited on 09/11/2023). preprint.

[46] D. Jackson. *The theory of approximation*. Vol. 11. American Mathematical Society Colloquium Publications. Reprint of the 1930 original. American Mathematical Society, Providence, RI, 1994, pp. viii+178.

[47] J. B. James D. Whitfield and A. Aspuru-Guzik. « Simulation of electronic structure Hamiltonians using quantum computers ». In: *Molecular Physics* 109.5 (2011), pp. 735–750. eprint: `https://doi.org/10.1080/00268976.2011.552441`.

[48] S. Jin, X. Li, N. Liu, and Y. Yu. « Quantum simulation for partial differential equations with physical boundary or interface conditions ». In: *J. Comput. Phys.* 498 (2024), Paper No. 112707, 18.

[49] S. Jin and N. Liu. « Analog quantum simulation of partial differential equations ». In: *Quantum Science and Technology* 9.3 (2024), p. 035047.

[50] S. Jin and N. Liu. « Quantum algorithms for computing observables of nonlinear partial differential equations ». In: (2022).

[51] S. Jin, N. Liu, and Y. Yu. « Quantum simulation of partial differential equations: applications and detailed analysis ». In: *Phys. Rev. A* 108.3 (2023), Paper No. 032603, 20.

[52] S. Jin, N. Liu, and Y. Yu. *Quantum Simulation of Partial Differential Equations via Schrodingerisation*. Dec. 28, 2022. arXiv: 2212.13969 [quant-ph]. URL: http://arxiv.org/abs/2212.13969 (visited on 02/23/2023). preprint.

[53] I. Joseph. « Koopman–von Neumann approach to quantum simulation of nonlinear classical dynamics ». In: *Phys. Rev. Res.* 2 (4 2020), p. 043102.

[54] B. N. Khoromskij. « $O(d \log N)$-Quantics Approximation of $N - d$ Tensors in High-Dimensional Numerical Modeling ». In: *Constr Approx* 34.2 (Oct. 2011), pp. 257–280.

[55] Y. Kim et al. « Scalable Error Mitigation for Noisy Quantum Circuits Produces Competitive Expectation Values ». In: *Nature Physics* 19.5 (May 2023), pp. 752–759.

[56] M. Krebsbach, B. Trauzettel, and A. Calzona. « Optimization of Richardson extrapolation for quantum error mitigation ». In: *Phys. Rev. A* 106 (6 2022), p. 062436.

[57] H. Krovi. « Improved quantum algorithms for linear and nonlinear differential equations ». In: *Quantum* 7 (Feb. 2023), p. 913.

[58] O. Kyriienko, A. E. Paine, and V. E. Elfving. « Solving Nonlinear Differential Equations with Differentiable Quantum Circuits ». In: *Physical Review A* 103.5 (May 2021), p. 052416.

[59] Y. Li and S. C. Benjamin. « Efficient Variational Quantum Simulator Incorporating Active Error Minimization ». In: *Phys. Rev. X* 7 (2 2017), p. 021050.

[60] E. H. Lieb, R. Seiringer, and J. Yngvason. « Bosons in a trap: A rigorous derivation of the Gross-Pitaevskii energy functional ». In: *Phys. Rev. A* 61 (4 2000), p. 043602.

[61] L. Lin. *Lecture Notes on Quantum Algorithms for Scientific Computation*. 2022. arXiv: 2201.08309 [quant-ph].

[62] N. Linden, A. Montanaro, and C. Shao. « Quantum vs. classical algorithms for solving the heat equation ». In: *Comm. Math. Phys.* 395.2 (2022), pp. 601–641.

[63] H.-L. Liu et al. « Variational quantum algorithm for the Poisson equation ». In: *Phys. Rev. A* 104 (2 2021), p. 022418.

[64] J.-P. Liu, D. An, D. Fang, J. Wang, G. H. Low, and S. Jordan. « Efficient quantum algorithm for nonlinear reaction-diffusion equations and energy estimation ». In: *Comm. Math. Phys.* 404.2 (2023), pp. 963–1020.

[65] J.-P. Liu, H. O. i. Kolden, H. K. Krovi, N. F. Loureiro, K. Trivisa, and A. M. Childs. « Efficient quantum algorithm for dissipative nonlinear differential equations ». In: *Proc. Natl. Acad. Sci. USA* 118.35 (2021), Paper No. e2026805118, 6.

[66] S. Lloyd et al. « Quantum algorithm for nonlinear differential equations ». In: *arXiv e-prints*, arXiv:2011.06571 (Nov. 2020), arXiv:2011.06571. arXiv: 2011.06571 [quant-ph].

[67] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio. « Unified approach to data-driven quantum error mitigation ». In: *Phys. Rev. Res.* 3 (3 2021), p. 033098.

[68] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch. « Variational Quantum Algorithms for Nonlinear Problems ». In: *Physical Review A* 101.1 (Jan. 2020), p. 010301. arXiv: 1907.09032.

[69] E. Magesan, J. M. Gambetta, and J. Emerson. « Characterizing quantum gates via randomized benchmarking ». In: *Phys. Rev. A* 85 (4 2012), p. 042311.

[70] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven. « Decoding Quantum Errors with Subspace Expansions ». In: *Nature Communications* 11.1 (Jan. 2020), p. 636.

[71] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik. « The Theory of Variational Hybrid Quantum-Classical Algorithms ». In: *New Journal of Physics* 18.2 (Feb. 2016), p. 023023.

[72] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. « Quantum circuit learning ». In: *Phys. Rev. A* 98 (3 2018), p. 032309.

[73] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. 10th anniversary ed. Cambridge ; New York: Cambridge University Press, 2010. 676 pp.

[74] P. J. J. O'Malley et al. « Scalable Quantum Simulation of Molecular Energies ». In: *Phys. Rev. X* 6 (3 2016), p. 031007.

[75] I. V. Oseledets. « Constructive Representation of Functions in Low-Rank Tensor Formats ». In: *Constr Approx* 37.1 (Feb. 2013), pp. 1–18.

[76] A. E. Paine, V. E. Elfving, and O. Kyriienko. « Physics-Informed Quantum Machine Learning: Solving Nonlinear Differential Equations in Latent Spaces without Costly Grid Evaluations ». In: arXiv:2308.01827 (Aug. 2023). arXiv: `2308.01827 [cond-mat, physics:quant-ph]`.

[77] A. E. Paine, V. E. Elfving, and O. Kyriienko. « Quantum kernel methods for solving regression problems and differential equations ». In: *Phys. Rev. A* 107 (3 2023), p. 032428.

[78] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. « Matrix product state representations ». In: *Quantum Inf. Comput.* 7.5-6 (2007), pp. 401–430.

[79] A. Peruzzo et al. « A Variational Eigenvalue Solver on a Photonic Quantum Processor ». In: *Nature Communications* 5.1 (Sept. 2014), p. 4213.

[80] A. J. Pool, A. D. Somoza, C. M. Keever, M. Lubasch, and B. Horstmann. « Nonlinear Dynamics as a Ground-State Solution on Quantum Computers ». In: arXiv:2403.16791 (Mar. 2024). arXiv: `2403.16791 [physics, physics:quant-ph]`.

[81] J. Preskill. « Quantum Computing in the NISQ era and beyond ». In: *Quantum* 2 (Aug. 2018), p. 79.

[82] M. Ragone et al. « A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits ». In: arXiv:2309.09342 (Sept. 2023). arXiv: `2309.09342 [quant-ph]`.

[83] S.-J. Ran. « Encoding of matrix product states into quantum circuits of one- and two-qubit gates ». In: *Phys. Rev. A* 101 (3 2020), p. 032310.

[84] M. S. Rudolph, J. Chen, J. Miller, A. Acharya, and A. Perdomo-Ortiz. *Decomposition of Matrix Product States into Shallow Quantum Circuits*. Sept. 1, 2022. arXiv: `2209.00595 [quant-ph]`. URL: `http://arxiv.org/abs/2209.00595` (visited on 09/07/2022). preprint.

[85] F. Schipp, W. R. Wade, and P. Simon. *Walsh Series: An Introduction to Dyadic Harmonic Analysis*. Bristol [England] ; New York: Adam Hilger, 1990. 560 pp.

[86] A. Suau, G. Staffelbach, and H. Calandra. « Practical Quantum Computing: Solving the Wave Equation Using a Quantum Approach ». In: *ACM Transactions on Quantum Computing* 2.1 (2021).

[87] Y. Takahashi, S. Tani, and N. Kunihiro. « Quantum addition circuits and unbounded fan-out ». In: *Quantum Inf. Comput.* 10.9-10 (2010), pp. 872–890.

[88] H. L. Tang et al. « Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient Ansätze on a Quantum Processor ». In: *PRX Quantum* 2 (2 2021), p. 020310.

[89] K. Temme, S. Bravyi, and J. M. Gambetta. « Error mitigation for short-depth quantum circuits ». In: *Phys. Rev. Lett.* 119.18 (2017), pp. 180509, 5.

[90] F. Tennie, S. Laizet, S. Lloyd, and L. Magri. « Quantum Computing for nonlinear differential equations and turbulence ». In: (June 2024). arXiv: `2406.04826 [physics.flu-dyn]`.

[91] J. Tilly et al. « The Variational Quantum Eigensolver: A Review of Methods and Best Practices ». In: *Physics Reports* 986 (2022), pp. 1–128.

[92] R. Toledo. « Solving systems of linear differential equations by Walsh polynomials approach ». In: *Ann. Univ. Sci. Budapest. Sect. Comput.* 52 (2021), pp. 313–348.

[93] F. Vatan and C. Williams. « Optimal Quantum Circuits for General Two-Qubit Gates ». In: *Phys. Rev. A* 69.3 (Mar. 22, 2004), p. 032315. arXiv: `quant-ph/0308006`.

[94] J. J. Wallman and J. Emerson. « Noise tailoring for scalable quantum computation via randomized compiling ». In: *Phys. Rev. A* 94 (5 2016), p. 052325.

[95] J. L. Walsh. « A Closed Set of Normal Orthogonal Functions ». In: *Amer. J. Math.* 45.1 (1923), pp. 5–24.

[96] K. Wan, M. Berta, and E. T. Campbell. « Randomized Quantum Algorithm for Statistical Phase Estimation ». In: *Phys. Rev. Lett.* 129 (3 2022), p. 030503.

[97] J. Watrous. *The Theory of Quantum Information*. 1st ed. Cambridge University Press, Apr. 26, 2018.

[98] T. Weaving, A. Ralli, W. M. Kirby, P. J. Love, S. Succi, and P. V. Coveney. « Benchmarking noisy intermediate scale quantum error mitigation strategies for ground state preparation of the HCl molecule ». In: *Phys. Rev. Res.* 5 (4 2023), p. 043054.

[99] B. Wu, J. Sun, Q. Huang, and X. Yuan. « Overlapped Grouping Measurement: A Unified Framework for Measuring Quantum States ». In: *Quantum* 7 (Jan. 13, 2023), p. 896. arXiv: `2105.13091 [quant-ph]`.

[100] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo. « Generalized Quantum Subspace Expansion ». In: *Phys. Rev. Lett.* 129 (2 2022), p. 020502.

[101] J. Zylberman, G. Di Molfetta, M. Brachet, N. F. Loureiro, and F. Debbasch. *Hybrid Quantum-Classical Algorithm for Hydrodynamics*. Feb. 2022. arXiv: `2202.00918 [physics, physics:quant-ph]`.

Abstract

Variational quantum algorithms (VQAs) have been proposed for solving partial differential equations on quantum computers. This thesis focuses on analyzing VQAs for the stationary Gross-Pitaevskii Equation (GPE) both under ideal (noiseless) conditions and in the presence of quantum noise, providing error bounds, convergence properties, and estimates for the number of samples required.
A central concept, make use of a relationship between the representation of functions and functional operators on dyadic rationals, through the Walsh basis, and the encoding of functions and operators for N-qubit quantum systems through the Pauli operators and their eigenstates.
In chapter 1, we link Pauli operators of N-qubit quantum systems with the Walsh basis on N-bit dyadic rationals, presenting new error bounds for the convergence of the N-bit Walsh series for functions in $H^1(0,1)$ and presenting some results on the representation of Fourier basis functions in the Walsh basis.
In chapter 2 we analyse VQAs for the GPE without noise, detailing the mathematical setting, discretization, and a-priori analysis. We introduce new energy estimators, either based on the Walsh decomposition of operators or obtained through inductive methods, and compare them to direct sampling, in the diagonal basis of the operators, and the Hadamard-test method. Our results show that in the absence of noise the most promising methods for energy estimation is direct sampling in the diagonal basis, yielding the lowest variance and sample requirements.
In chapter 3, we further examine the impact of quantum noise on energy estimation. Depolarizing noise introduces bias and shifts the variance of estimators. We show that the Pauli estimators proves least affected by noise, due to their lower circuit size requirement, outperforming others both without mitigation, due to a lower bias, and with mitigation, as its sample efficiency is less affected.
This research provides a foundation for the use of VQAs to solve partial differential equations on quantum computers, offering realistic estimates of computational costs of the energy estimation part of the algorithms. Furthermore, the insights gained may aid in the development of practical quantum algorithms for block encoding of differential operators, and their associated evolution operators.

Keywords: informatique quantique, équations différentielles, algorithmes variationels, équation de gross-pitaevskii

Laboratoire Jacques-Louis Lions – Sorbonne Université – Campus Pierre et Marie Curie – 4 place Jussieu – 75005 Paris – France
Eviden Quantum Lab – Avenue Jean-Jaurès – 78340 Les Clayes-sous-bois – France