Technische Universität München

Fakultät für Mathematik

The Gaussian Wave Packet Transform and its Application in Quantum Dynamics

Paul Ferdinand Bergold

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:	Prof. Dr. Silke Rolles
Prüfer*innen der Dissertation:	1. Prof. Dr. Caroline Lasser
	2. Prof. Arieh Iserles, Ph.D. University of Cambridge, United Kingdom
	3. Prof. Olof Runborg, Ph.D. KTH Royal Institute of Technology, Sweden

Die Dissertation wurde am 11.05.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 04.08.2022 angenommen.

Abstract

Gaussian wave packets are used in numerous numerical methods to solve the time-dependent Schrödinger equation. In particular, superpositions of Gaussian wave packets are often used, because they have useful analytical properties and allow wave functions outside the class of Gaussian functions to be approximated.

In this dissertation we investigate superpositions of Gaussian wave packets resulting from discretisations of the continuous wave packet transform in phase space. Based on a rigorous analysis of the underlying approximation properties for different quadrature rules, we focus on the so-called "*Time-Sliced Thawed Gaussian Propagation Method*", a numerical method recently proposed for solving the Schrödinger equation, in which the wave packet transform appears as an important ingredient. Finally, after a detailed mathematical investigation supported by numerical experiments, I present the latest results from the field of tensor-train approximations, which can be used to simulate high-dimensional quantum systems.

Zusammenfassung

Gaußsche Wellenpakete werden in zahlreichen numerischen Methoden zum Lösen der zeitabhängigen Schrödingergleichung eingesetzt. Dabei kommen insbesondere oftmals Überlagerungen Gaußscher Wellenpakete zum Einsatz, da diese nützliche analytische Eigenschaften besitzen und es ermöglichen, auch Wellenfunktionen außerhalb der Klasse von Gauß-Funktionen zu approximieren.

In dieser Dissertation untersuchen wir Überlagerungen Gaußscher Wellenpakete, die sich aus Diskretisierungen der stetigen Wellenpaket-Transformation im Phasenraum ergeben. Aufbauend auf einer rigorosen Analyse der zugrundeliegenden Approximationseigenschaften für verschiedene Quadraturregeln beschäftigen wir uns anschließend mit der sogenannten "*Time-Sliced Thawed Gaussian Propagation Method*", einer kürzlich vorgeschlagenen numerischen Methode zum Lösen der Schrödingergleichung, in der die Wellenpaket-Transformation als wichtiger Bestandteil auftritt. Nach einer ausführlichen mathematischen Untersuchung gestützt durch numerische Experimente präsentiere ich zuletzt neuste Ergebnisse aus dem Bereich der Tensor-Train Approximationen, die zur Simulation hochdimensionaler Quantensysteme genutzt werden können.

Contents

Acknowledgements

As a student of my advisor Caroline Lasser, I have gained many valuable impressions and been able to develop my mathematical competencies over the past five years. I am very grateful for her constant support, her guidance and for giving me the opportunity to meet inspiring researchers during this time. It is a great pleasure for me to learn from her on a professional and personal level. Thank you!

I am also very grateful to Victor S. Batista for the many fruitful discussions in which I learned to use mathematics with an eye for applications in chemistry. After we met at a workshop in Oberwolfach in April 2019, I spent three months in his group at Yale University in New Haven in the summer. I would especially like to thank him for his warm hospitality during this time, for his interest in my work and for collaborating with me on so many exciting projects.

I would also like to thank Micheline B. Soley for sharing discussions on quantum algorithms, tensor-train implementations and boson sampling. Her work on the joint projects, especially on the implementation of the tensor trains for the Chebyshev method, was a great support.

Special thanks go to my colleagues and friends at TUM. I would like to mention Maximilian Engel for interesting discussions on philosophical interpretations of probabilities, Mi-Song Dupuy for his nice introduction to the theory of tensor trains, Chunmei Su for discussions on the Gaussian wave packet transform, Stephanie Troppmann, who warmly welcomed me to the chair and supported me at the beginning of my doctoral studies, Fabian Flassig for discussions on window functions, and Isabella Wiegand for many enriching discussions at the interface of mathematics and literary studies.

I have recently become a Research Fellow in Coupled Quantum-Classical Dynamics at the University of Surrey. I would like to thank Cesare Tronci, who warmly welcomed me and made it possible for me to complete this dissertation while moving to his group.

Finally, I would like to thank all those who have not yet been named but who have contributed in some way to the success of this work.

Notation

Some important notations that occur repeatedly in the thesis should be introduced right at the beginning. For a complex-valued function $\psi\colon \mathbb{R}^d \times \mathbb{R} \to \mathbb{C}$, defined for $x \in \mathbb{R}^d$ (position) and $t \in \mathbb{R}$ (time), we use the following conventions:

$$\partial_t \psi(x,t) := \frac{\partial}{\partial t}\psi(x,t) \quad \text{and} \quad \Delta\psi(x,t) = \Delta_x \psi(x,t) := \sum_{k=1}^{d} \frac{\partial^2}{\partial x_k^2}\psi(x,t).$$

We denote by $L^2(\mathbb{R}^d)$ the Hilbert space of square-integrable functions (wave functions) and work with the inner product

$$\langle f \mid g \rangle = \langle f \mid g \rangle_{L^2(\mathbb{R}^d)} := \int_{\mathbb{R}^d} \overline{f(x)} g(x) \,\mathrm{d}x, \quad f, g \in L^2(\mathbb{R}^d),$$

which is taken antilinear in its first argument. Furthermore, $\mathcal{S}(\mathbb{R}^d)$ denotes the Schwartz space of rapidly decaying smooth functions defined by

$$\mathcal{S}(\mathbb{R}^d) := \Big\{ f \in C^\infty(\mathbb{R}^d) \;:\; \sup_{x \in \mathbb{R}^d} |x^\alpha \partial^\beta f(x)| < \infty \;\; \forall \alpha, \beta \in \mathbb{N}_0^d \Big\},$$

where $C^\infty(\mathbb{R}^d)$ is the space of complex-valued smooth functions on $\mathbb{R}^d$ and $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ and $\partial^\beta = \partial_{x_1}^{\beta_1} \cdots \partial_{x_d}^{\beta_d}$. In particular, recall that $\mathcal{S}(\mathbb{R}^d)$ is a dense subset of $L^2(\mathbb{R}^d)$.

It should also be noted that we work with a rescaled version of the Fourier transform, which for $\varepsilon > 0$ (semiclassical parameter) is defined by

$$\mathcal{F}_\varepsilon \psi(p) := (2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} \psi(x) e^{-ip\cdot x/\varepsilon} \,\mathrm{d}x, \quad \psi \in \mathcal{S}(\mathbb{R}^d),$$

where i is the imaginary unit and $p \cdot x = p^T x$ is the dot product in $\mathbb{R}^d$. Recall that since $\mathcal{F}_\varepsilon(\mathcal{S}(\mathbb{R}^d)) = \mathcal{S}(\mathbb{R}^d)$ and $\|\psi\|_{L^2(\mathbb{R}^d)} = \|\mathcal{F}_\varepsilon\psi\|_{L^2(\mathbb{R}^d)}$ for all $\psi \in \mathcal{S}(\mathbb{R}^d)$ (Plancherel), the rescaled Fourier transform is a surjective isometry on $L^2(\mathbb{R}^d)$ and thus uniquely extends to a mapping from $L^2(\mathbb{R}^d)$ onto $L^2(\mathbb{R}^d)$, see e.g. [RS75, Theorem IX.6].

Finally, we note that we use bold letters for multi-indices that number entries of a tensor, *i.e.*, elements of $\mathbb{R}^\Gamma$ or $\mathbb{C}^\Gamma$, where Γ is a finite set. For example, for $\Gamma \subset \mathbb{N}^{2d}$ we denote by $c_{\mathbf{n}} \in \mathbb{C}$ the entry of a tensor $c \in \mathbb{C}^\Gamma$ corresponding to the multi-index $\mathbf{n} = (n_1, \ldots, n_{2d}) \in \Gamma$. Occasionally we will replace $c_{\mathbf{n}}$ with $c(n_1, \ldots, n_{2d})$. In addition to tensors, we use bold indices for grid points in phase space. For example, we denote by $z_{\mathbf{n}} \in \mathbb{R}^{2d}$ the grid point corresponding to the multi-index $\mathbf{n} = (n_1, \ldots, n_{2d})$.

1 Introduction

The development of algorithms for efficient simulations of quantum dynamics plays a central role in numerical analysis, as these methods contribute to a deeper understanding of many physical and chemical models. The time-dependent Schrödinger equation in semiclassical scaling given by

$$i\varepsilon\partial_t\psi(x,t) = -\frac{\varepsilon^2}{2}\Delta_x\psi(x,t) + V(x)\psi(x,t), \quad \psi(\bullet,0) = \psi_0 \in L^2(\mathbb{R}^d), \tag{1.1}$$

where $0 < \varepsilon \ll 1$ is a small positive parameter and $V\colon \mathbb{R}^d \to \mathbb{R}$ is a smooth potential, has been shown to be fundamental to molecular quantum dynamics and will be the central equation in this dissertation, which essentially deals with the question of how well solutions to (1.1) can be approximated by concatenations of Gaussian superpositions. The right-hand side of (1.1) is given by the action of the operator

$$H = H^\varepsilon := -\frac{\varepsilon^2}{2}\Delta_x + V$$

as it results from the celebrated Born–Oppenheimer approximation, see e.g. [LL20, Section 2], where the dimensionless semiclassical parameter ε represents the square root of a mass ratio of nuclei and electrons, typically on the order of 10^{-2} to 10^{-3}, and must be formally distinguished from $\hbar \approx 1.055 \cdot 10^{-34}$ Js, known as the reduced Planck constant.

Motivated by various problems in physics and chemistry, a large number of numerical algorithms for solving (1.1) have been developed in the last decades. Recently, Kong *et al.* have proposed the so-called *Time-Sliced Thawed Gaussian (TSTG) Propagation Method*, see [KMB16], in which Gaussian wave packets are decomposed into linear combinations of Gaussian basis functions without the need for multidimensional numerical integration. It turns out that the approximations of wave packets used by Kong *et al.* can be obtained by discretising the *FBI inversion formula*, according to which any function $\psi \in L^2(\mathbb{R}^d)$ can be decomposed as

$$\psi = (2\pi\varepsilon)^{-d}\int_{\mathbb{R}^{2d}} \langle g_z \mid \psi\rangle\, g_z\, \mathrm{d}z, \tag{1.2}$$

where the semiclassically scaled wave packet $g_z \in \mathcal{S}(\mathbb{R}^d)$ is defined for a given Schwartz function $g\colon \mathbb{R}^d \to \mathbb{C}$ of unit L^2-norm, which may or may not be a Gaussian, and a phase space centre $z = (q,p) \in \mathbb{R}^{2d}$ by

$$g_z(x) := \varepsilon^{-d/4} g\left(\frac{x-q}{\sqrt{\varepsilon}}\right) e^{ip\cdot(x-q)/\varepsilon}. \tag{1.3}$$

A direct discretisation of the phase space integral in (1.2) using a multidimensional quadrature rule in phase space leads to an approximation of the form

$$\psi \approx \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi)\, g_{\mathbf{n}}, \tag{1.4}$$

where $\Gamma \subset \mathbb{N}^{2d}$ is a given finite multi-index set, the representation coefficients $c_{\mathbf{n}}(\psi) \in \mathbb{C}$ are complex numbers depending on ψ and the underlying quadrature rule, and the functions $g_{\mathbf{n}} := g_{z_{\mathbf{n}}}$ are wave packets centred in the grid points $z_{\mathbf{n}} \in \mathbb{R}^{2d}$. In particular, if both the function ψ and the basis functions $g_{\mathbf{n}}$ are Gaussian wave packets, the coefficients $c_{\mathbf{n}}(\psi)$, which in this case are essentially given by the inner products $\langle g_{\mathbf{n}} \mid \psi\rangle$ (multiplied by a weight), can be calculated by hand, which Kong *et al.* used for the design of the TSTG method, since it allows to express time-evolved Gaussian basis functions in the original Gaussian basis without multidimensional numerical integration.

Starting from the representation of the initial wave function ψ_0 according to (1.4), the solution to the Schrödinger equation (1.1) is approximated after a short propagation time $\tau > 0$ by the linear combination of time-evolved basis functions as follows

$$\psi(\tau) = U(\tau)\psi_0 \approx \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\, U(\tau) g_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\, g_{\mathbf{n}}(\tau),$$

where $U(t) := e^{-iHt/\varepsilon}$ denotes the unitary propagator and we have introduced the abbreviations $\psi(\tau)$ and $g_{\mathbf{n}}(\tau)$ for $\psi(\bullet, \tau)$ and $g_{\mathbf{n}}(\bullet, \tau)$ respectively.

Remark 1. *If the Hamiltonian H is self-adjoint, the existence and uniqueness of the strongly continuous group of unitary operators $U(t)$, $t \in \mathbb{R}$, on $L^2(\mathbb{R}^d)$ is guaranteed by Stone's theorem, see [Sto32]. Moreover, for all initial states $\psi_0 \in \mathcal{D}(H) \subset L^2(\mathbb{R}^d)$ (we denote by $\mathcal{D}(H)$ the domain of H), the solution to the time-dependent Schrödinger equation in semiclassical scaling (1.1) is given for all times $t \in \mathbb{R}$ by*

$$\psi(t) = e^{-itH/\varepsilon}\psi_0, \quad \|\psi(t)\| = \|\psi_0\| = 1,$$

where we let $\| \bullet \|$ denote the L^2-norm. In particular, H is self-adjoint if the potential V is of sub-quadratic growth and we refer to [Lub08, Chapter I.3.2] for other conditions on the potential that yield self-adjoint operators.

Using thawed Gaussian approximations for the time evolution of the individual basis functions $g_{\mathbf{n}}$, the discretisation of the wave packet transform according to (1.4) is then brought into play again, this time to represent the individual thawed Gaussian approximations $u_{\mathbf{n}}^{\tau} \approx g_{\mathbf{n}}(\tau)$ for the time-evolved basis functions as follows

$$u_{\mathbf{n}}^{\tau} \approx \sum_{\mathbf{n}'\in\Gamma} c_{\mathbf{n}'}(u_{\mathbf{n}}^{\tau})\, g_{\mathbf{n}'},$$

which allows the exact solution $\psi(\tau)$ to be approximated directly in the original basis in terms of updated coefficients $c_{\mathbf{n}}^{1,\tau}$ as

$$\psi(\tau) \approx \psi^{1,\tau} := \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{1,\tau} g_{\mathbf{n}}, \quad \text{where} \quad c_{\mathbf{n}}^{1,\tau} := \sum_{\mathbf{n}'\in\Gamma} c_{\mathbf{n}'}(\psi_0) c_{\mathbf{n}}(u_{\mathbf{n}'}^{\tau}).$$

The concatenation of these steps leads to approximations for longer times $2\tau, 3\tau, \ldots$, which are obtained (without additional time integration) by computing corresponding update coefficients $c_{\mathbf{n}}^{2,\tau}, c_{\mathbf{n}}^{3,\tau}, \ldots$. Since all these coefficients can be expressed analytically, multidimensional numerical quadrature can be avoided completely, which means that the total error of the TSTG method is essentially produced by three different sources:

i) the discretisation of the continuous wave packet transform

ii) the thawed Gaussian approximations for the propagation of the basis functions

iii) the numerical integration of the thawed equations of motion

The precise mathematical description and a complete error representation for both the discretisation of the wave packet transform and the TSTG method is the subject of this dissertation. Furthermore, we focus on the connection to other state-of-the-art methods and use our mathematical analysis to show where the original method introduced by Kong *et al.* can be further improved.

Remark 2. *I will refrain from repeating the basics of quantum mechanics, especially the statistical interpretation of wave functions due to Born, which does not play a major role in this thesis anyway. Instead, I refer to the book of Hall, see [Hal13], for a comprehensive introduction to the theory of quantum mechanics from the perspective of a pure mathematician, as well as to the book of Lubich, see [Lub08], which gives a general overview of numerical methods for the time-dependent Schrödinger equation. In addition to these rich sources and the methods described therein, I would like to explicitly mention some numerical methods based on Gaussian wave packets. For example, reduced models via variational approximations have been investigated, which include the variational multi-configuration Gaussian wave packet (vMCG) method [WRB04] and the variational Gaussian wave packets [Hel76, CK90]. Furthermore, semiclassical approaches such as Hagedorn wave packets [FGL09, GH14], Gaussian beams [LQ09, Zhe14, KKR15, LRT16], or the Herman–Kluk propagator [HK84, LS17] have been developed to include quantum effects especially for high-dimensional systems, for which standard grid-based numerical methods are infeasible.*

1.1 Main results

The main contributions in this dissertation have been developed over the past five years and most of them have already been published. In chronological order, these are:

1. with C. Lasser: "Fourier Series Windowed by a Bump Function",
   appeared in *Journal of Fourier Analysis and Applications*, 26(4):65, 2020,
   e-print at arXiv:1901.04365

2. with C. Lasser: "The Gaussian Wave Packet Transform via Quadrature Rules",
   submitted to *IMA Journal of Numerical Analysis* on 15/12/2021,
   e-print at arXiv:2010.03478

3. with C. Lasser: "An Error Representation for the Time-Sliced Thawed Gaussian Propagation Method",
submitted to *Numerische Mathematik* on 27/08/2021,
e-print at arXiv:2108.12182

4. with M. B. Soley, A. A. Gorodetsky and V. S. Batista: "Functional Tensor-Train Chebyshev Method for Multidimensional Quantum Dynamics Simulations",
appeared in *Journal of Chemical Theory and Computation*, 18(1):25–36, 01 2022,
e-print at arXiv:2109.08985

While article no. 3 can be seen as a follow-up article to no. 2, at first glance there seems to be no connection between the other publications. In order to bring the results together into an overall picture in this thesis, parts of the above-mentioned articles have been rearranged and connected with the help of new sections that have not been published anywhere before. In particular, the newly added sections explain connections that have already been mentioned by other authors but, to the best of my knowledge, have not yet been stringently presented. For example, in Section 3.3, the relationship between the Gaussian wave packet transform based on plain Gaussians and compactly supported basis functions is explained using bump windows, and it is shown that the FBI formula can be identified as the common origin for the expansions used in the TSTG method and the fast Gaussian wave packet transform previously introduced by Qian and Ying in [QY10].

The main results of this dissertation are formulated as theorems. Theorem 17 presents the approximation errors for the discretisation of the FBI formula for different quadrature rules in momentum space and shows the superiority of a new variant of the Gaussian wave packet transform based on Gauss–Hermit quadrature. The proof of Theorem 17, the preceding results in Section 3.1, the corresponding numerical results in Section 3.1.4 as well as Appendix 7.2 were taken from the joint preprint [BL20b] with C. Lasser. Furthermore, the numerical results in Section 3.2.2 and Appendix 7.1 were taken from the joint preprint [BL21] with C. Lasser.

Theorem 50 and Theorem 57 present results related to windowed Fourier series. The first shows that pointwise multiplication by a window with plateau yields smaller reconstruction errors in the interior of the plateau as compared to those without plateau, while the second connects the decay rate of windowed Fourier coefficients to a new bound for the variation of windowed functions. The corresponding Sections 4.1-4.4 on windowed Fourier series as well as Appendix 7.3, Appendix 7.4 and Appendix 7.5 were taken from the joint publication [BL20a] with C. Lasser.

Theorem 78 represents the first rigorous error representation for the TSTG method, both for the original version introduced by Kong *et al.* with non-variationally evolving Gaussian basis functions and for the new variant with variational Gaussians. In addition, Section 5.1 provides the first mathematical formulation of the TSTG method, which in particular allows comparison with other methods and indicates possible approaches for future research. The proof of Theorem 78, the preceding results in Sections 5.1-5.3, the corresponding numerical results in Section 5.4 as well as Appendix 7.6 were taken from the joint preprint [BL21] with C. Lasser. Furthermore, the results on the *Tensor-Train Chebyshev (TTC) Method* show how tensor trains can be used to perform multidimensional quantum dynamics simulations, motivating a promising approach to make the TSTG method applicable to high-dimensional systems. The description of the TTC method in Section 6.2, including the numerical experiments, as well as Appendix 7.8 were taken from the joint publication [SBGB22] with M. B. Soley, A. A. Gorodetsky and V. S. Batista.

1.2 Outline

The thesis is organised as follows: In the next chapter we focus on the essential “tool” of this dissertation, which is of course the Gaussian wave packet transform. The aim is to recall some important properties of Gaussian wave packets that will be used repeatedly in the subsequent chapters. Chapter 3 deals with the discretisation of the continuous wave packet transform, working our way step by step to the different discrete variants. First we derive a semi-discrete representation in which the integral over position space is replaced by a Gaussian summation curve. We then focus on the discretisation of the remaining integral over momentum space via different quadrature rules and summarise the underlying approximation errors in the main result Theorem 17. The remainder of the chapter is divided into two sections. In Section 3.2 we study the direct discretisation of the phase space integral, whereas in Section 3.3 we present the connection to the variants of the Gaussian wave packet transform used by other authors. Chapter 4 is dedicated to windowed Fourier series. Here we will follow up on some results on bump windows that have already been briefly discussed in Chapter 3 in the context of windowed Gaussian wave packets. In Chapter 5 we then analyse the TSTG method. The mathematical description will reveal that the underlying approximation of wave packets is an application of the Gaussian wave packet transform, and the main result Theorem 78 provides the first error representation of the TSTG method. Finally, in Chapter 6 we will look at tensor-train approximations and show how they can be used to overcome the curse of dimensionality that we usually face in grid-based methods for simulating high-dimensional quantum dynamics.

2 The Gaussian wave packet transform

Gaussian functions are used in many models today and occur in almost all fields of study. For example, in numerical analysis, where Gaussians are used in the form of radial basis function interpolations to construct solutions to partial differential equations, see e.g. [LF03], in widely used applications of statistics, in which probability densities are approximated by Gaussian mixtures, see e.g. [TSM85], or in seismology, where Gaussian functions appear in connection with the famous Gabor transform and are used for the decomposition of seismic waves, see e.g. [ML01]. In this chapter we recall some important properties of Gaussian wave packets, which we will refer to several times later in the thesis. After introducing a precise definition of Gaussian wave packets, we present Hagedorn's parametrisation in preparation of a favourable representation of the semiclassical equations of motion for the propagation of Gaussian wave packets. Furthermore, we introduce an approximation manifold that is used for the variational propagation based on the Dirac–Frenkel time-dependent variational principle and present a useful formula for inner products. Section 2.2 deals with the FBI formula (1.2) and establishes the connection with the inversion formula of the short-time Fourier transform.

2.1 Gaussian wave packets and their properties

In the following we work with d-dimensional complex Gaussians whose width matrix is contained in a special subset of matrices that goes back to Siegel, see [Sie39].

Definition 3. *The Siegel upper half-space of degree $d \geq 1$, denoted by $\mathfrak{S}^+(d)$, is the set of complex symmetric $d \times d$ matrices with positive definite imaginary part, i.e.,*

$$\mathfrak{S}^+(d) := \left\{ C \in \mathbb{C}^{d\times d} \,:\, C = C^T,\ \mathrm{Im}(C) \text{ is positive definite} \right\}.$$

We will also use the common shorthand notation "$\mathrm{Im}(C) > 0$" to express that the matrix $\mathrm{Im}(C)$ is positive definite.

Now, for a given matrix $C \in \mathfrak{S}^+(d)$, consider the complex-valued Gaussian

$$g(x) := \pi^{-d/4} \det(\mathrm{Im}\, C)^{1/4} \exp\left(\frac{i}{2} x^T C x\right), \quad x \in \mathbb{R}^d. \tag{2.1}$$

First, we note that it follows from the construction of the prefactor that g is normalised with respect to the L^2 norm, *i.e.*,

$$\|g\| = \left(\int_{\mathbb{R}^d} |g(x)|^2 \,\mathrm{d}x \right)^{1/2} = 1.$$

Furthermore, with this choice, the wave packet g_z defined in (1.3) has the form

$$g_z(x) = g_z^{C,\varepsilon}(x) := (\pi\varepsilon)^{-d/4} \det(\operatorname{Im} C)^{1/4} \cdots \exp\left[\frac{i}{\varepsilon}\left(\frac{1}{2}(x-q)^T C(x-q) + p^T(x-q)\right)\right], \tag{2.2}$$

which is typically referred to as *Gaussian wave packet* in this dissertation, and from now on, whenever we use the notation $g_z^{C,\varepsilon}$, we assume that the amplitude function g corresponds exactly to the Gaussian function defined in (2.1). In particular, the dependence on the matrix C and the semiclassical parameter ε is always implicitly assumed in the shorthand notation g_z, and in the one-dimensional case we always write $\gamma = \gamma_r + i\gamma_i$ instead of C.

Remark 4. *The Gaussian wave packets introduced above with an arbitrary width matrix $C \in \mathfrak{S}^+(d)$ generalise the ground states of the harmonic oscillator. To understand this statement better, let us consider the one-dimensional annihilation operator given by*

$$\hat{a} := \frac{1}{\sqrt{2\varepsilon}}\left(\hat{q} + i\hat{p}\right),$$

where $\hat{q}$ is the position operator, i.e., $(\hat{q}\varphi)(x) := x\varphi(x)$, and $\hat{p}$ the momentum operator, i.e., $(\hat{p}\varphi)(x) := -i\varepsilon\frac{\mathrm{d}}{\mathrm{d}x}\varphi(x)$, for all Schwartz functions $\varphi \in \mathcal{S}(\mathbb{R})$. Then, every element of the kernel of the annihilation operator, i.e., every element of

$$\left\{\varphi \in \mathcal{S}(\mathbb{R}) \,:\, \hat{a}\varphi = 0\right\},$$

is of the form $\varphi(x) = c \cdot e^{-x^2/2\varepsilon}$, where $c \in \mathbb{C}$ is a complex constant, see [Tro17, Chapter 3.2]. Therefore, the ground state of the harmonic oscillator Hamiltonian

$$H_{\mathrm{ho}} = -\frac{\varepsilon^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}x^2 = \frac{1}{2}\left(\hat{q}^2 + \hat{p}^2\right) = \varepsilon\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right)$$

is given by the Gaussian function $\varphi_0(x) = (\pi\varepsilon)^{-1/4}e^{-x^2/2}$, which in the context of quantum dynamics is usually called the "coherent ground state of the harmonic oscillator", see e.g. [CR12, Chapter 1]. Accordingly, some authors refer to the function g_z in (2.2) as coherent states, although in Glauber's original use of the term, see [Gla63], "wave function of a coherent state" would be the correct description.

2.1.1 Hagedorn's parametrisation

We will see later in Section 5.2.1 that the time-dependent Schrödinger equation (1.1) leaves the class of Gaussian wave packets invariant for quadratic potentials and the equations of motion for the width matrix C result in a form comparable to those for the phase space parameters q and p if we use a special factorisation $C = PQ^{-1}$, which goes back to the work of Hagedorn, see [Hag80, Hag98].

The following lemma was taken from [LL20, Lemma 3.16] and provides a useful connection between the Siegel upper half-space and symplectic matrices. It also guarantees the existence of Hagedorn's parametrisation for elements in $\mathfrak{S}^+(d)$.

Lemma 5. *Let Q and P be complex $d \times d$ matrices such that the real matrix*

$$Y = \begin{pmatrix} \operatorname{Re} Q & \operatorname{Im} Q \\ \operatorname{Re} P & \operatorname{Im} P \end{pmatrix} \in \mathbb{R}^{2d\times 2d}$$

is symplectic, i.e.,

$$Y = Y^T J Y = J \quad \textit{with} \quad J = \begin{pmatrix} 0 & -\operatorname{Id}_d \\ \operatorname{Id}_d & 0 \end{pmatrix},$$

or equivalently (in the following, $Q^* = \overline{Q}^T$ denotes the Hermitian adjoint of Q),

$$Q^T P - P^T Q = 0 \tag{2.3}$$

$$Q^* P - P^* Q = 2i \operatorname{Id}_d . \tag{2.4}$$

Then, Q and P are invertible, and

$$C = PQ^{-1}$$

is an element of the Siegel upper half-space $\mathfrak{S}^+(d)$ with imaginary part

$$\operatorname{Im}(C) = (QQ^*)^{-1}. \tag{2.5}$$

Conversely, every $C \in \mathfrak{S}^+(d)$ can be written as $C = PQ^{-1}$ with matrices Q and P satisfying (2.3), (2.4) and (2.5).

For the proof of Lemma 5 we refer to [LL20, Lemma 3.16]. □

The factorisation $C = PQ^{-1}$ provides an alternative way of representing Gaussian wave packets, namely

$$\begin{aligned} g_z(x) = g_z^{Z,\varepsilon}(x) := (\pi\varepsilon)^{-d/4} |\det(Q)|^{-1/2} \cdots \\ \exp\left[\frac{i}{\varepsilon}\left(\frac{1}{2}(x-q)^T PQ^{-1}(x-q) + p^T(x-q)\right)\right], \end{aligned} \tag{2.6}$$

where we introduced the matrix $Z = (Q, P) \in \mathbb{C}^{2d\times d}$ and used that

$$\det(\operatorname{Im} C)^{1/4} = \det(QQ^*)^{-1/4} = |\det(Q)|^{-1/2}.$$

The complex normalisation factor $\det(Q)^{-1/2}$ without the absolute value is usually used for numerical implementations where the branch of the square root must be chosen appropriately, see e.g. [Lub08, Chapter V.1].

Gaussian wave packets, parameterised as in (2.6), were originally developed by Hagedorn to construct an orthonormal basis of $L^2(\mathbb{R}^d)$ that generalises Hermite functions. We note that Hagedorn wave packets have a wide range of applications. For example, Hagedorn wave packets with complex phase space centres have recently been considered for non-self-adjoint evolution problems, see [LST18].

2.1.2 Gaussian approximation manifold

For the variational propagation of Gaussian wave packets, we will exploit the so-called "Dirac–Frenkel time-dependent variational approximation principle" later in Section 5.2. We therefore identify the approximation space

$$\mathcal{M} = \Big\{ u \in L^2(\mathbb{R}^d) \, : \, u(x) = g_z^{C,\varepsilon}(x) e^{iS/\varepsilon}, \, z \in \mathbb{R}^{2d}, \, C \in \mathfrak{S}^+(d), \, S \in \mathbb{R} \Big\}, \tag{2.7}$$

consisting of Gaussian wave packets multiplied by a phase factor $e^{iS/\varepsilon}$, as a manifold. We will use variationally evolving Gaussians to approximate the time evolution of the basis functions in the TSTG method and in Section 5.2.1 we will give a more precise meaning to the phase factor $e^{iS/\varepsilon}$, but for the time being we consider the parameter $S \in \mathbb{R}$ only as an additional degree of freedom of the Gaussian function u.

Since the Dirac–Frenkel variational principle works with the orthogonal projection onto the tangent spaces of $\mathcal{M}$, let us give an exact characterisation of tangent vectors. The following lemma was taken from [LL20, Lemma 3.1].

Lemma 6. *At every Gaussian function $u \in \mathcal{M}$, the tangent space equals*

$$\mathcal{T}_u\mathcal{M} = \big\{ \varphi u \, : \, \varphi \text{ is a complex } d\text{-variate polynomial of degree at most } 2. \big\}$$

In particular, the tangent space is a complex-linear subspace of $L^2(\mathbb{R}^d)$, in the sense that $v \in \mathcal{T}_u\mathcal{M}$ implies $iv \in \mathcal{T}_u\mathcal{M}$. Moreover, for all differential operators A of order ≤ 2 with constant coefficients we have $Au \in \mathcal{T}_u\mathcal{M}$.

For the proof of Lemma 6 we refer to [LL20, Lemma 3.1]. □

The fact that the tangent spaces of $\mathcal{M}$ arise from multiplication by polynomials of degree ≤ 2 is used to prove the exactness of variational approximations for quadratic potentials, see Proposition 70.

2.1.3 Inner products

The discretisation of the FBI formula leads to representation coefficients that sample weighted inner products of Gaussians. The next lemma was taken from the joint preprint [BL21] (see Lemma 3.2) and presents an analytical expression for these inner products, which shows that the inner products can be written as Gaussians in phase space.

Lemma 7. *For $C_1, C_2 \in \mathfrak{S}^+(d)$ and $z_1, z_2 \in \mathbb{R}^{2d}$, we have*

$$\langle g_{z_1}^{C_1,\varepsilon} \mid g_{z_2}^{C_2,\varepsilon} \rangle = \beta(z) \exp\left(\frac{i}{2\varepsilon} (z_2 - z_1)^T M (z_2 - z_1) \right), \tag{2.8}$$

where the matrix

$$M := \begin{pmatrix} \big(C_2^{-1} - \bar{C}_1^{-1}\big)^{-1} & 0 \\ 0 & -(C_2 - \bar{C}_1)^{-1} \end{pmatrix} \in \mathbb{C}^{2d \times 2d} \tag{2.9}$$

is an element of the Siegel upper half-space $\mathfrak{S}^+(2d)$ *of* $2d \times 2d$ *matrices, and for* $B = C_2 - \bar{C}_1$ *the complex constant* $\beta(z) \in \mathbb{C}$ *is given by*

$$\beta(z) := \frac{2^{d/2} \det(\operatorname{Im} C_1 \operatorname{Im} C_2)^{1/4}}{\sqrt{\det(-iB)}} \exp\left(\frac{i}{2\varepsilon}(p_1 + p_2)^T (q_1 - q_2)\right) \cdots$$
$$\exp\left(\frac{i}{2\varepsilon}(p_2 - p_1)^T B^{-1}(C_2 + \bar{C}_1)(q_2 - q_1)\right).$$

Moreover, if the eigenvalues of the positive definite matrices $\operatorname{Im}(C_k)$ *and* $\operatorname{Im}(-C_k^{-1})$*,* $k = 1, 2$*, are bounded from below by a constant* $\theta > 0$*, then the absolute value of the inner product is bounded by*

$$\left|\langle g_{z_1}^{C_1,\varepsilon} \mid g_{z_2}^{C_2,\varepsilon}\rangle\right|^2 \leq \zeta \exp\left(-\frac{\theta}{4\varepsilon}\|z_2 - z_1\|_2^2\right), \tag{2.10}$$

where the positive constant $\zeta > 0$ *depends on* θ *and an upper bound on the eigenvalues of the matrices* $\operatorname{Im}(C_k)$ *and* $\operatorname{Im}(-C_k^{-1})$*, but is independent of* ε*.*

The crucial ingredient for the proof uses a formula for integrals of complex-valued Gaussians and is presented in Appendix 7.1. We also find that the bound in (2.10) can be easily improved if the lower bound for the eigenvalues of $\operatorname{Im}(C_k)$ and $\operatorname{Im}(-C_k^{-1})$ is not chosen uniformly. We also refer to the proof for the dependence of the constant ζ on the spectral parameters.

2.2 Continuous superpositions of Gaussian wave packets

In the previous section we introduced Gaussian wave packets and presented properties that are important for the next chapters. This section deals with the FBI formula (1.2), which we will discretise in Chapter 3 to approximate wave functions by discrete superpositions of Gaussian basis functions. Let us start by proving the following representation, which was taken from [LL20, Proposition 5.1].

Proposition 8. *For every Schwartz function* $\psi \in \mathcal{S}(\mathbb{R}^d)$ *we have*

$$\psi(x) = (2\pi\varepsilon)^{-d} \int_{\mathbb{R}^{2d}} \langle g_z \mid \psi\rangle \, g_z(x) \, \mathrm{d}z, \quad x \in \mathbb{R}^d.$$

We present the proof of Lasser and Lubich based on the Fourier inversion theorem.

Proof. We use the inversion formula for the Fourier transform: a Schwartz function f is reconstructed from its scaled Fourier transform

$$\mathcal{F}_\varepsilon f(\xi) = (2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i\xi \cdot x/\varepsilon} \, \mathrm{d}x$$

by the inversion formula

$$f(x) = (2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} \mathcal{F}_\varepsilon f(\xi) e^{i\xi\cdot x/\varepsilon} \, \mathrm{d}\xi.$$

For a Schwartz function ψ and for $x \in \mathbb{R}^d$, this yields

$$\begin{aligned}
&(2\pi\varepsilon)^{-d} \int_{\mathbb{R}^{2d}} \langle g_z \mid \psi \rangle \; g_z(x) \, \mathrm{d}z \\
&= (2\pi\varepsilon)^{-d} \varepsilon^{-d/2} \int_{\mathbb{R}^{3d}} \overline{g}\left(\frac{y-q}{\sqrt{\varepsilon}}\right) g\left(\frac{x-q}{\sqrt{\varepsilon}}\right) e^{ip\cdot(x-y)/\varepsilon} \psi(y) \, \mathrm{d}(y,q,p) \\
&= \varepsilon^{-d/2} \int_{\mathbb{R}^d} (2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} \mathcal{F}_\varepsilon \left(\overline{g}\left(\frac{\bullet - q}{\sqrt{\varepsilon}}\right) \psi \right)(p) e^{ip\cdot x/\varepsilon} \, \mathrm{d}p \, g\left(\frac{x-q}{\sqrt{\varepsilon}}\right) \mathrm{d}q \\
&= \varepsilon^{-d/2} \int_{\mathbb{R}^d} \left| g\left(\frac{x-q}{\sqrt{\varepsilon}}\right) \right|^2 \psi(x) \, \mathrm{d}q = \psi(x),
\end{aligned}$$

since the normalisation of the function g implies

$$\varepsilon^{-d/2} \int_{\mathbb{R}^d} \left| g\left(\frac{x-q}{\sqrt{\varepsilon}}\right) \right|^2 \mathrm{d}q = \int_{\mathbb{R}^d} |g(y)|^2 \, \mathrm{d}y = 1.$$

□

Since we have already referred to (1.2) as "FBI formula" several times, let us catch up with the exact reference to the FBI transform. The following definition was taken from [LS17, Definition 1].

Definition 9. *For a point $z = (q,p) \in \mathbb{R}^{2d}$ in phase space and $C = i\,\mathrm{Id}_d$, consider the Gaussian wave packet $g_z = g_z^{C,\varepsilon}$. Then, the mapping*

$$T^\varepsilon \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^{2d}), \; (T^\varepsilon \psi)(z) := (2\pi\varepsilon)^{-d/2} \langle g_z \mid \psi \rangle \tag{2.11}$$

is called the FBI (Fourier–Bros–Iagolnitzer) transform.

We note that the transform can be extended to an isometry from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^{2d})$, see e.g. [Mar02, Proposition 3.1.1], and the formula in Proposition 8 extends directly to functions $\psi \in L^2(\mathbb{R}^d)$ by density.

The above definition shows that the FBI transform works with Gaussian wave packets of unit width (*i.e.*, $C = i\,\mathrm{Id}_d$) and can be seen as a natural generalisation of the plain Fourier transform using a Gaussian window. Although in this dissertation we use Gaussian functions equipped with an arbitrary width matrix $C \in \mathfrak{S}^+(d)$ and this does not correspond to the classical definition, we follow the convention and continue to use the name "FBI formula" or "FBI inversion formula" for (1.2). We also note that the FBI transform for matrices of arbitrary width is sometimes called "Fourier–Bargmann transform", see e.g. [CR12, Chapter 1.2.3].

2.2.1 Connection to the short-time Fourier transform

We realise that the FBI transform in (2.11) can be seen as a windowed Fourier transform using a Gaussian window, which is known in time-frequency analysis as Gabor transform, see e.g. [FS98]. Moreover, the Gabor transform is itself a special variant of the so-called "short-time Fourier transform" (STFT), and thus we can use the inversion formula of the STFT to obtain a more general representation than in Proposition 8. Since the semiclassical parameter is typically not considered in the definition of the STFT, let us introduce a rescaled version:

Definition 10. *Let $\varepsilon > 0$ and $g \neq 0$ be a function in $L^2(\mathbb{R}^d)$ (window). For $\psi \in L^2(\mathbb{R}^d)$ we define the* <u>ε-rescaled short-time Fourier transform</u> *by*

$$\mathrm{STFT}_g^\varepsilon(\psi)(z) := (2\pi\varepsilon)^{-d} \int_{\mathbb{R}^d} \psi(x) g\left(\frac{x-q}{\sqrt{\varepsilon}}\right) e^{-ip\cdot x/\varepsilon}\,\mathrm{d}x, \quad z \in \mathbb{R}^{2d}.$$

We note that $\mathrm{STFT}_g^\varepsilon(\psi)$ is uniformly continuous on $\mathbb{R}^{2d}$, see e.g. [Grö01, Lemma 3.1.1], and by introducing the rescaled translation and modulation operators

$$T_q^\varepsilon g(x) := g\left(\frac{x-q}{\sqrt{\varepsilon}}\right) \quad \text{and} \quad M_p^\varepsilon g(x) := g(x)e^{ip\cdot x/\varepsilon},$$

we can write

$$\mathrm{STFT}_g^\varepsilon(\psi)(z) = (2\pi\varepsilon)^{-d}\langle M_p^\varepsilon T_q^\varepsilon g \mid \psi\rangle.$$

It is easy to see that if the window is localised in a neighbourhood of q, the same is true for the windowed function and therefore the spectrum of the (rescaled) STFT is associated with the windowed domain.

For the rescaled STFT we have the following inversion formula, which generalises the FBI formula in Proposition 8. The following result is an extension (inclusion of the semiclassical parameter) of [Grö01, Corollary 3.2.3].

Proposition 11. *Suppose that $g, h \in L^2(\mathbb{R}^d)$ and $\langle g \mid h\rangle \neq 0$. Then, for all $\psi \in L^2(\mathbb{R}^d)$,*

$$\psi = (2\pi\varepsilon)^{-d} \frac{1}{\langle g \mid h\rangle} \int_{\mathbb{R}^{2d}} \mathrm{STFT}_g^\varepsilon(\psi)(q,p) M_p^\varepsilon T_q^\varepsilon h\,\mathrm{d}p\,\mathrm{d}q.$$

For the proof we refer to [Grö01, Corollary 3.2.3]. □

Based on the representation in Proposition 11, the choice $h = g$ for a normalised window g implies that

$$\psi = (2\pi\varepsilon)^{-d} \int_{\mathbb{R}^{2d}} \langle M_p^\varepsilon T_q^\varepsilon g \mid \psi\rangle\, M_p^\varepsilon T_q^\varepsilon g\,\mathrm{d}p\,\mathrm{d}q,$$

which gives us the FBI formula (1.2) for

$$g(x) = \pi^{-d/4} \exp\left(-\frac{1}{2}|x|^2\right).$$

Remark 12. *We would like to note that the inversion formula does not imply that the STFT is invertible, see [Grö01, Corollary 3.2.3 (Remarks)]: Let $A_h \colon L^2(\mathbb{R}^{2d}) \to L^2(\mathbb{R}^d)$ be the bounded linear operator defined by*

$$A_h(F) := (2\pi\varepsilon)^{-d} \int_{\mathbb{R}^{2d}} F(q,p) M_p^\varepsilon T_q^\varepsilon h \, \mathrm{d}p \, \mathrm{d}q.$$

Since for all $h \in L^2(\mathbb{R}^d)$ and $F \in L^2(\mathbb{R}^{2d})$ we have

$$\langle h \mid A_h(F) \rangle = \big\langle \mathrm{STFT}_h^\varepsilon(h) \mid F \big\rangle = \big\langle h \mid (\mathrm{STFT}_h^\varepsilon)^* \, F \big\rangle,$$

it follows that A_h is the adjoint operator of $\mathrm{STFT}_h^\varepsilon$ and the inversion formula is

$$\frac{1}{\langle g \mid h \rangle} (\mathrm{STFT}_h^\varepsilon)^* \, \mathrm{STFT}_g^\varepsilon = \mathrm{Id}\,.$$

We also note that the concept of windowed Fourier atoms was introduced by Gabor, see [Gab46], who studied Gaussian windows in terms of the uncertainty principle to obtain "optimal" windows. In many applications, windows are discussed in terms of data weighting and spectral leakage, and numerous windows have been developed depending on the type of signal, see for example [Har78, Section IV (Table 1)]. In particular, we will elaborate further on the discussion of windows and their properties in Chapter 4.

2.3 Summary of this chapter

We have presented some important properties of Gaussian wave packets and proved that the Gaussian wave packet g_z defined in (2.2) can be used to represent arbitrary wave functions via the FBI formula (1.2), which itself is as a special variant of the inversion formula of the short-time Fourier transform based on Gaussian windows. In particular, there seems to be no uniform definition of the term

"Gaussian wave packet transform"

but in this dissertation we associate the term with decompositions of a given wave function $\psi \in L^2(\mathbb{R}^d)$ into (not necessarily continuous) superpositions of Gaussian wave packets (or at least wave packets with a similar profile) according to the FBI formula. Finally, let us summarise in a small table the various names for representations that appear in the literature in connection with the FBI formula, see also [LL20, 5.9 Notes]:

	Semiclassical Analysis	Time-Frequency Analysis
arbitrary windows	*Fourier–Bargmann*	*Short-Time Fourier* *Windowed Fourier* *(Continuous) Wavelet*
Gaussian windows	*FBI (unit width)*	*(Continuous) Gabor*

Table 2.1: Different names for the FBI formula.

3 The Gaussian wave packet transform via quadrature rules

In this chapter we investigate different approaches to discretise the FBI formula (1.2). Using a uniform grid $\{q_k\}_{k\in\Gamma_q}$ in position space, we start by deriving a semi-discrete representation of the form

$$\psi(x) = (2\pi\varepsilon)^{-d}\left[\frac{1}{S(x)}\sum_{k\in\Gamma_q}\int_{\mathbb{R}^d}\langle g_{(q_k,p)} \mid \psi\rangle\, g_{(q_k,p)}(x)\,\mathrm{d}p\right], \quad x\in\mathbb{R}^d,$$

where the function

$$S(x) := \sum_{k\in\Gamma_q}|g_0(x-q_k)|^2 \tag{3.1}$$

is a summation curve in position space and $g_0 := g_{(0,0)}$. Afterwards, by discretising the remaining integral over momentum space via different quadrature rules, we obtain a rescaled superposition of the Gaussian basis functions $g_{j,k} := g_{(q_k,p_j)}$, as follows

$$\psi(x) \approx \frac{1}{S(x)}\sum_{k\in\Gamma_q}\sum_{j\in\Gamma_p^{(\text{rule})}} r_{j,k}^{(\text{rule})}\, g_{j,k}(x), \tag{3.2}$$

where the representation coefficients $r_{j,k}^{(\text{rule})}$ are complex numbers depending on ψ and the underlying quadrature rule in momentum space. This discretisation of the wave packet transform based on uniform Riemann sums in both position and momentum space was used in particular by Kong *et al.* for the TSTG method, who used sufficiently dense grids to approximate the Gaussian summation curve $S(x)$ by a constant value $S>0$, since this has the consequence that the wave function ψ can be approximated by a pure superposition of Gaussians as follows:

$$\psi(x) \approx \sum_{k\in\Gamma_q}\sum_{j\in\Gamma_p^{(\text{rule})}} c_{j,k}^{(\text{rule})}\, g_{j,k}(x), \quad \text{where} \quad c_{j,k}^{(\text{rule})} = \frac{1}{S} r_{j,k}^{(\text{rule})}. \tag{3.3}$$

We present a rigorous error analysis and numerical experiments for the approximations described above and extend the results of Kong *et al.* by a new representation based on Gauss–Hermite quadrature, which significantly reduces the number of grid points. Furthermore, we show that (3.3) corresponds to a direct discretisation of the phase space integral in the FBI formula. Finally, we use bump windows and windowed Fourier series to relate our results to the work of Qian and Ying on the fast Gaussian wave packet transform, see [QY10], who derived a similar representation based on compactly supported basis functions.

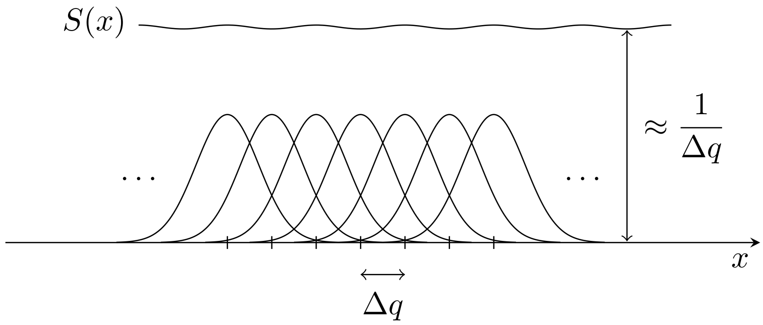


Figure 3.1: Gaussians $|g_0(x - q_k)|^2$ and summation curve on a uniform grid for $d = 1$. According to Lemma 13, $S(x)$ can be approximated by $1/\Delta q$.

3.1 Semi-discrete approximations

As described in detail in Section 1.1, parts of the present section (Sec. 3.1) overlap to a large extent with the joint preprint "The Gaussian Wave Packet Transform via Quadrature Rules" with C. Lasser submitted to *IMA Journal of Numerical Analysis* on 15/12/2021, e-print available at arXiv:2010.03478.

For a non-empty index set $\Gamma_q \subseteq \mathbb{Z}^d$ and a uniform grid $\{q_k\}_{k\in\Gamma_q}$ in position space, recall the definition of the summation curve in (3.1). A quick look at the one-dimensional situation in Figure 3.1 makes it plausible that for a sufficiently small grid spacing $\Delta q > 0$ the summation curve can be approximated by a constant value. The next lemma tells us that this constant is equal to $1/\Delta q$.

Lemma 13. *For $d = 1$ consider the Gaussian g defined in (2.1) with width parameter $\gamma = \gamma_r + i\gamma_i \in \mathbb{C}$, $\gamma_i > 0$. Then, for $\Gamma_q = \mathbb{Z}$ and the uniform grid points $q_k = k\Delta q$ with distance $\Delta q > 0$, the one-dimensional summation curve has the expansion*

$$S(x) = \frac{1}{\Delta q} + \frac{2}{\Delta q}\sum_{n=1}^{\infty} \cos\left(\frac{2\pi n x}{\Delta q}\right) \exp\left(-\frac{\pi^2 n^2 \varepsilon}{\gamma_i (\Delta q)^2}\right), \quad x \in \mathbb{R}, \tag{3.4}$$

where the convergence is uniform in x. In particular, we obtain spectral convergence of the summation curve to $1/\Delta q$ as $\Delta q \to 0$, i.e., for all $s \in \mathbb{N}$, there exists a positive constant $C_s > 0$, depending on s, ε and γ, such that

$$\left| S(x) - \frac{1}{\Delta q} \right| < C_s (\Delta q)^{2s-1} \quad \textit{for all } x \in \mathbb{R},$$

where the constant C_s can be chosen as

$$C_s = \frac{2s!\gamma_i^s}{\pi^{2s}\varepsilon^s}.$$

Moreover, the summation curve is Δq-periodic and infinitely differentiable.

Proof. Using a convolution of an unshifted Gaussian with a Dirac comb, it is proven in [ML01, Appendix A] that for $\Delta q > 0, T > 0$ and all $x \in \mathbb{R}$ we have

$$\sum_{k\in\mathbb{Z}} \frac{\Delta q}{T\sqrt{\pi}} \exp\left(-\frac{(x-k\Delta q)^2}{T^2}\right) = 1 + 2\sum_{n=1}^{\infty} \cos\left(\frac{2\pi n x}{\Delta q}\right) \exp\left(-\left(\frac{\pi n T}{\Delta q}\right)^2\right). \tag{3.5}$$

Hence, using (3.5) for $T = \sqrt{\varepsilon/\gamma_i}$, we obtain

$$\sum_{k\in\mathbb{Z}} \frac{\sqrt{\gamma_i}}{\sqrt{\pi\varepsilon}} \exp\Big(-\frac{\gamma_i}{\varepsilon}(x-q_k)^2\Big) = \frac{1}{\Delta q}\left(1 + 2\sum_{n=1}^{\infty} \cos\Big(\frac{2\pi n x}{\Delta q}\Big) \exp\Big(-\frac{\pi^2 n^2 \varepsilon}{\gamma_i(\Delta q)^2}\Big)\right),$$

which establishes (3.4). For the following calculations let us introduce the parameter $\eta = \exp(-\pi^2\varepsilon/\gamma_i(\Delta q)^2) < 1$. Then we get

$$\begin{aligned}\left|S(x) - \frac{1}{\Delta q}\right| &\le \frac{2}{\Delta q}\sum_{n=1}^{\infty}\left|\cos\left(\frac{2\pi n x}{\Delta q}\right)\exp\left(-\frac{\pi^2 n^2\varepsilon}{\gamma_i(\Delta q)^2}\right)\right| \\ &\le \frac{2}{\Delta q}\sum_{n=1}^{\infty}\exp\left(-\frac{\pi^2 n\varepsilon}{\gamma_i(\Delta q)^2}\right) = \frac{2}{\Delta q}\sum_{n=1}^{\infty}\eta^n,\end{aligned}$$

and since $e^y - 1 > y^s/s!$ for all $y > 0$ and all $s \in \mathbb{N}$, we finally conclude that

$$\sum_{n=1}^{\infty}\eta^n = \frac{\eta}{1-\eta} = \frac{1}{\exp(\pi^2\varepsilon/\gamma_i(\Delta q)^2) - 1} < \frac{s!\gamma_i^s}{\pi^{2s}\varepsilon^s}(\Delta q)^{2s}.$$

Moreover, a short calculation proves that the summation curve is Δq-periodic,

$$S(x+\Delta q) = \sum_{k\in\mathbb{Z}} |g_0(x-(k-1)\Delta q)|^2 = S(x), \quad x \in \mathbb{R},$$

and by the Weierstrass test, see for example [WW96, Chapter 3.34], the infinite sum converges absolutely and uniformly on any set. It therefore follows from periodicity that $S \in C^\infty(\mathbb{R})$. □

Provided that the position grid is aligned with the eigenvectors of the symmetric and positive definite matrix $\operatorname{Im} C > 0$, multidimensional summation curves can be written as the product of one-dimensional summation curves, which itself can be expanded according to (3.4). More precisely, if $\operatorname{Im} C = UDU^T$ is an eigendecomposition with corresponding eigenvalues $\lambda_1, \dots, \lambda_d > 0$ and Γ_q can be decomposed as

$$\Gamma_q = \Gamma_q^{(1)} \times \dots \times \Gamma_q^{(d)},$$

then the multidimensional summation curve can be decomposed for all $x \in \mathbb{R}^d$ as

$$S(x) = \prod_{n=1}^{d} S_n(x) := \prod_{n=1}^{d}\left(\sum_{k_n\in\Gamma_q^{(n)}} g_n(x^T u_n - k_n\Delta q)^2\right), \tag{3.6}$$

where $u_n \in \mathbb{R}^d$ is the nth column of U and the one-dimensional functions g_n are

$$g_n(y) := (\pi\varepsilon)^{-1/4}\lambda_n^{1/4}\exp\left(-\frac{\lambda_n}{2\varepsilon}y^2\right), \quad y \in \mathbb{R}. \tag{3.7}$$

Remark 14. *The reason why we restrict ourselves to uniform grids in position space is that the variants of the wave packet transform, which we derive later in Section 3.1.2, depend only on the grid in momentum space. However, the following results and estimates do not depend on this specific choice unless we indicate otherwise.*

The definition of the summation curve $S(x) > 0$ in (3.1) allows to construct the functions $\chi_k(x) := |g_0(x - q_k)|^2/S(x)$ for all $k \in \Gamma_q$ that satisfy the two conditions

$$0 < \chi_k(x) \leq 1 \quad \text{and} \quad \sum_{k\in\Gamma_q} \chi_k(x) = 1 \quad \text{for all } x \in \mathbb{R}^d. \tag{3.8}$$

The family $\{\chi_k\}_{k\in\Gamma_q}$ thus forms a so-called "partition of unity". Partitions of unity typically occur in the theory of manifolds, see e.g. [Tu11, Chapter 13.1], but also in numerical applications, for example to construct solutions of differential equations, see [GS00, Section 4.1.2]. In the next step, we combine the partition $\{\chi_k\}_{k\in\Gamma_q}$ from above with the Fourier inversion formula in momentum space to obtain a semi-discrete decomposition of square-integrable functions.

Proposition 15. *Let $\psi \in \mathcal{S}(\mathbb{R}^d)$. For $C \in \mathfrak{S}^+(d)$ and $z = (q,p) \in \mathbb{R}^{2d}$, recall the definition of the Gaussian wave packet g_z in (2.2) and let us introduce the map*

$$x \mapsto \mathcal{I}_q(x) := (2\pi\varepsilon)^{-d}\int_{\mathbb{R}^d} \langle g_z \mid \psi\rangle\, g_z(x)\,\mathrm{d}p.$$

Then, for all $x \in \mathbb{R}^d$, we have

$$\psi(x) = \frac{1}{S(x)}\sum_{k\in\Gamma_q} \mathcal{I}_{q_k}(x) \quad \textit{and} \tag{3.9}$$

$$\psi(x) = \int_{\mathbb{R}^d} \mathcal{I}_q(x)\,\mathrm{d}q. \tag{3.10}$$

For convenience we have taken $\psi \in \mathcal{S}(\mathbb{R}^d)$, but just as with the FBI formula, the above representations apply directly to $L^2(\mathbb{R}^d)$. With respect to the approximation of Schrödinger dynamics, we would also like to point out that semi-discrete representations with a summation in position space such as in (3.9) are also used in the construction of higher-order Gaussian beam approximations, see e.g. [LRT13, Section 2.1].

Proof. We start by proving (3.9). Let $\psi \in \mathcal{S}(\mathbb{R}^d)$ and $g_k(x) := g_0(x - q_k)$, where the Gaussian amplitude g is defined in (2.1). According to the properties of the partition $\{\chi_k\}_{k\in\Gamma_q}$ in (3.8), the function ψ can be decomposed as follows:

$$\psi = \psi\sum_{k\in\Gamma_q}\chi_k = \psi\left(\frac{1}{S}\sum_{k\in\Gamma_q}|g_k|^2\right) = \frac{1}{S}\sum_{k\in\Gamma_q}\left(\psi\overline{g_k}\right)g_k \tag{3.11}$$

In particular, since $g_k \in \mathcal{S}(\mathbb{R}^d)$ for all $k \in \Gamma_q$, we conclude that $\psi\overline{g_k} \in \mathcal{S}(\mathbb{R}^d)$ and therefore, using the Fourier inversion theorem, we obtain

$$(\psi\overline{g_k})(x) = (2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} \mathcal{F}_\varepsilon\left[\psi\overline{g_k}\right](p)\, e^{ix\cdot p/\varepsilon}\,\mathrm{d}p, \quad \text{for all } x \in \mathbb{R}^d, \tag{3.12}$$

where the ε-rescaled Fourier transform is given by

$$\mathcal{F}_\varepsilon\left[\psi\overline{g_k}\right](p) = (2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} \psi(x)\overline{g_k(x)}e^{-ip\cdot x/\varepsilon}\,\mathrm{d}x.$$

Furthermore, for all $p \in \mathbb{R}^d$, we get

$$\begin{aligned} &\mathcal{F}_\varepsilon\left[\psi\overline{g_k}\right](p)\, e^{ix\cdot p/\varepsilon} g_k(x) \\ &\quad = (2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} \psi(y)\overline{g_k(y)}e^{-ip\cdot(y-q_k)/\varepsilon}\,\mathrm{d}y\, g_k(x)e^{ip\cdot(x-q_k)/\varepsilon} \\ &\quad = (2\pi\varepsilon)^{-d/2}\langle g_{(q_k,p)} \mid \psi\rangle\, g_{(q_k,p)}(x). \end{aligned} \tag{3.13}$$

Consequently, by inserting (3.12) and (3.13) into (3.11), we conclude that

$$\psi(x) = \frac{1}{S(x)} \sum_{k\in\Gamma_q} \mathcal{I}_{q_k}(x).$$

For proving (3.10) we use the fact that g_0 is of unit norm. Hence, for all $x \in \mathbb{R}^d$ we get

$$\psi(x) = \psi(x)\int_{\mathbb{R}^d} |g_0(x-q)|^2\,\mathrm{d}q = \int_{\mathbb{R}^d} \left(\psi(x)\overline{g_0(x-q)}\right) g_0(x-q)\,\mathrm{d}q$$

and thus, again by the Fourier inversion formula, we obtain

$$\begin{aligned} \psi(x) &= \int_{\mathbb{R}^d} \left((2\pi\varepsilon)^{-d/2} \int_{\mathbb{R}^d} \mathcal{F}_\varepsilon[\psi\overline{g_0(\bullet - q)}](p)\, e^{ix\cdot p/\varepsilon}\,\mathrm{d}p \right) g_0(x-q)\,\mathrm{d}q \\ &= \int_{\mathbb{R}^d} \left((2\pi\varepsilon)^{-d} \int_{\mathbb{R}^d} \langle g_z \mid \psi\rangle\, g_z(x)\,\mathrm{d}p \right)\,\mathrm{d}q = \int_{\mathbb{R}^d} \mathcal{I}_q(x)\,\mathrm{d}q, \end{aligned}$$

which makes the proof complete. □

The representation in (3.9) can be seen as a semi-discrete version of the FBI formula. In particular, we emphasise that this is an exact representation and not an approximation, as we would obtain, for example, by a direct discretisation. Indeed, starting from (3.10) in the one-dimensional setting and discretising the integral over $\mathcal{I}_q(x)$ using a uniform grid of size Δq, we obtain the approximation (not an exact representation)

$$\psi(x) \approx \Delta q \sum_{k\in\mathbb{Z}} \mathcal{I}_{q_k}(x). \tag{3.14}$$

The relation between (3.9) and (3.14) is then obtained via Lemma 13, according to which $\Delta q \approx 1/S(x)$ for a sufficiently small grid spacing Δq.

3.1.1 A new representation for Gaussian wave packets

In the next step we focus on the special case that the wave function of interest ψ is a Gaussian wave packet and therefore the inner products $\langle g_z \mid \psi\rangle$ as well as $\mathcal{I}_q(x)$ can be expressed analytically. These representations are of particular interest for the TSTG method, where we need to approximate the time-evolved Gaussian basis functions $g_{j,k}(\tau)$ in terms of the original Gaussian basis functions $g_{j,k}$.

Lemma 16. *For $C, C_0 \in \mathfrak{S}^+(d)$ and $z, z_0 \in \mathbb{R}^{2d}$ let $g_z = g_z^{C,\varepsilon}$ and $\psi_0 := g_{z_0}^{C_0,\varepsilon}$. Moreover, let us introduce the parameters*

$$A := i(C_0 - \overline{C})^{-1}, \quad b(x) := x - q - iAC_0(q - q_0), \quad \textit{and}$$

$$c(x) := \frac{\det(\operatorname{Im} C \operatorname{Im} C_0)^{1/4}}{(\pi\varepsilon)^d \sqrt{2^d \det(A^{-1})}} \cdots \exp\left(-\frac{1}{2\varepsilon}(q - q_0)^T \overline{C} A C_0 (q - q_0) + \frac{i}{\varepsilon} p_0^T (x - q_0)\right).$$

Then, for all $x \in \mathbb{R}^d$, we have

$$\mathcal{I}_q(x) = g_0(x - q)c(x) \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2\varepsilon} p^T A p + \frac{i}{\varepsilon} b(x)^T p\right) \mathrm{d}p. \tag{3.15}$$

In particular, the integral in (3.15) exists because $\operatorname{Re}(A) > 0$.

Proof. Using the formula for inner products in Lemma 7, we obtain

$$\langle g_z \mid \psi_0\rangle = \beta(z) \exp\left(\frac{i}{2\varepsilon}(z - z_0)^T M (z - z_0)\right), \tag{3.16}$$

where the matrix

$$M = \begin{pmatrix} \left(C_0^{-1} - \overline{C}^{-1}\right)^{-1} & 0 \\ 0 & -(C_0 - \overline{C})^{-1} \end{pmatrix} \in \mathbb{C}^{2d \times 2d}$$

is an element of the Siegel upper half-space of $2d \times 2d$ matrices and the complex constant $\beta(z) \in \mathbb{C}$ is given by

$$\beta(z) = \frac{2^{d/2} \det(\operatorname{Im} C \operatorname{Im} C_0)^{1/4}}{\sqrt{\det(A^{-1})}} \exp\left(\frac{i}{2\varepsilon}(p + p_0)^T (q - q_0)\right) \cdots \exp\left(\frac{1}{2\varepsilon}(p - p_0)^T A (C_0 + \overline{C})(q - q_0)\right).$$

Moreover, according to Proposition 15, the function $\mathcal{I}_q$ is given for all $x \in \mathbb{R}^d$ by

$$\mathcal{I}_q(x) = (2\pi\varepsilon)^{-d} \int_{\mathbb{R}^d} \langle g_z \mid \psi_0\rangle \, g_z(x) \, \mathrm{d}p.$$

Consequently, the formula for the inner product in (3.16) yields

$$
\begin{aligned}
&(2\pi\varepsilon)^{-d}\langle g_z \mid \psi_0\rangle \, g_z(x) \\
&\quad = (2\pi\varepsilon)^{-d}\beta(z)\, g_0(x-q) \exp\left(\frac{i}{2\varepsilon}(z-z_0)^T M(z-z_0) + \frac{i}{\varepsilon}p^T(x-q)\right) \\
&\quad = g_0(x-q)c(x) \exp\left(-\frac{1}{2\varepsilon}(p-p_0)^T A(p-p_0) + \frac{i}{\varepsilon}b(x)^T(p-p_0)\right),
\end{aligned} \tag{3.17}
$$

where we rearranged terms only by simple algebraic manipulations and we used the fact that $(C_0^{-1} - \overline{C}^{-1})^{-1} = i\overline{C}AC_0$. The representation in (3.15) therefore follows from a linear transformation of the integral. Finally, since $Z \in \mathfrak{S}^+(d)$ implies $-Z^{-1} \in \mathfrak{S}^+(d)$, see e.g. [Fol89, Theorem 4.64], we conclude that $\mathrm{Re}(A) > 0$. □

The combination of Proposition 15 with Lemma 16 gives the exact representation

$$
\psi_0(x) = \frac{1}{S(x)} \sum_{k\in\Gamma_q} \left(g_0(x-q_k)c_k(x) \int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p \right), \tag{3.18}
$$

where we introduced the Gaussian function

$$
f_{k,x}(p) := \exp\left(-\frac{1}{2\varepsilon}p^T A p + \frac{i}{\varepsilon}b_k(x)^T p\right), \quad p \in \mathbb{R}^d, \tag{3.19}
$$

and we write b_k and c_k to indicate that we have replaced the variable q with q_k in the definition of b and c. At first glance, the representation in (3.18) may seem unfinished, since it still contains a Gaussian integral that could be solved by hand. Let us briefly discuss how (3.18) can be used for quadrature: Consider the one-dimensional situation for a moment. Since $f_{k,x}$ is a Gaussian centred at $p = 0$ and therefore decays rapidly relative to its width matrix, we can use a uniform grid $\{p_j\}_{j=1}^J$ on a given finite interval $[p_0 - L_p, p_0 + L_p]$ (where $L_p > 0$ depends on the width matrix) to discretise

$$
\int_{-\infty}^{\infty} f_{k,x}(p)\,\mathrm{d}p \approx \int_{p_0-L_p}^{p_0+L_p} f_{k,x}(p-p_0)\,\mathrm{d}p \approx \sum_{j=1}^J f_{k,x}(p_j - p_0)\,\Delta p \tag{3.20}
$$

and therefore, by inserting (3.20) into (3.18), the formula in (3.17) yields

$$
\psi_0(x) \approx \frac{1}{S(x)} \frac{\Delta p}{2\pi\varepsilon} \sum_{k\in\Gamma_q} \sum_{j=1}^J \langle g_{j,k} \mid \psi_0\rangle \, g_{j,k}(x).
$$

This reveals that discretisations of the multidimensional momentum integral

$$
\int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p \tag{3.21}
$$

lead to discrete variants of the FBI formula which can be used to approximate Gaussian wave packets by rescaled (prefactor $1/S(x)$) Gaussian superpositions.

The next section deals with the discretisation of the integral in (3.21), where the aim is to find a quadrature rule to keep the number of grid points small. We will see that Gauss–Hermite quadrature is a good candidate because $f_{k,x}$ has a Gaussian envelope.

3.1.2 Discretisation of the momentum integral

In the following we derive error bounds for discretisations based on

- truncations of the integral combined with the (composite) midpoint rule (TcM)
- infinite Riemann sums on uniform grids (RS)
- Gauss–Hermite quadrature (GH)

The rules (TcM) and (RS) were chosen because they can be used for the analysis of the TSTG method later in Chapter 5. To show that the approach of Kong *et al.* can be further improved by using more efficient rules, we investigate (GH), and to the best of my knowledge this is the first time that Gauss–Hermite quadrature is used to derive a new variant of the Gaussian wave packet transform.

Depending on the underlying quadrature rule, we choose for

(TcM) the finite uniform grid defined by

$$p_{j,n} = p_{0,n} - L_p + \frac{2j_n - 1}{2}\Delta p, \quad j \in \Gamma_p^{(\mathrm{TcM})} = \{1, \dots, J\}^d, \tag{3.22}$$

with grid size $\Delta p = 2L_p/J$, which discretises the box

$$\Lambda_p := \prod_{n=1}^{d} [p_{0,n} - L_p, p_{0,n} + L_p] \subset \mathbb{R}^d \tag{3.23}$$

in momentum space of length $2L_q$ in each coordinate direction.

(RS) the infinite uniform grid of size $\Delta p > 0$ defined by

$$p_{j,n} = p_{0,n} + j_n \Delta p, \quad j \in \Gamma_p^{(\mathrm{RS})} = \mathbb{Z}^d. \tag{3.24}$$

(GH) the finite (non-uniform) grid

$$p_{j,n} = p_{0,n} + s_{j_n}\sqrt{2\varepsilon}, \quad j \in \Gamma_p^{(\mathrm{GH})} = \{1, \dots, J\}^d,$$

depending on the zeros s_{j_n} of the Jth Hermite polynomial, see Section 3.1.2.

As discussed in the previous subsection, the different discretisations then lead to rescaled superpositions of the form

$$\psi_{\mathrm{rec}}^{(\mathrm{rule})}(x) := \frac{1}{S(x)} \sum_{k \in \Gamma_q} \sum_{j \in \Gamma_p^{(\mathrm{rule})}} r_{j,k}^{(\mathrm{rule})}\, g_{j,k}(x), \tag{3.25}$$

where the corresponding coefficients can be calculated analytically as follows:

(TcM) and (RS)

$$r_{j,k}^{(\mathrm{TcM})} = r_{j,k}^{(\mathrm{RS})} = \frac{(\Delta p)^d}{(2\pi\varepsilon)^d} \langle g_{j,k} \mid \psi_0 \rangle. \tag{3.26}$$

(GH) Depending on the weights w_{j_n} of the Gauss–Hermite rule, see Section 3.1.2, as

$$r_{j,k}^{(\mathrm{GH})} = \frac{w_{j_1} \cdots w_{j_d}}{(2\pi\varepsilon)^d} \langle g_{j,k} \mid \psi_0 \rangle.$$

The difference between the representation coefficients based on Riemann sums and Gauss–Hermite quadrature lies, on the one hand, in the spacing of the grid points p_j and, on the other hand, in the different weighting of the inner product $\langle g_{j,k} \mid \psi_0 \rangle$. Furthermore, it should be noted that for (TcM) a suitable truncation constant $L_p > 0$ must be chosen, while for (GH) the grid points are in a sense "optimally" distributed even without an additional truncation constant.

Equipped with the different grids in momentum space, we are left with the choice of the grid in position space. Since the wave function of interest $\psi_0 = g_{z_0}^{C_0,\varepsilon}$ is a Gaussian, it is plausible to assume that we are only interested in an approximation for values x in a certain neighbourhood of the centre q_0, e.g. the box

$$\Lambda_q := \prod_{n=1}^{d} [q_{0,n} - L_q, q_{0,n} + L_q] \subset \mathbb{R}^d$$

of length $2L_q$ in each coordinate direction (where $L_q > 0$ depends on the width of ψ_0). For a multi-index $k \in \Gamma_q = \{1, \dots, K\}^d$ we consider the uniform grid

$$q_{k,n} = q_{0,n} - L_q + \frac{2k_n - 1}{2} \Delta q, \tag{3.27}$$

where $\Delta q = 2L_q/K$. As mentioned earlier, for simplicity, in position space we focus on uniform grids aligned with the eigenvectors of the width matrix of the basis functions. More precisely, again using the eigendecomposition $\operatorname{Im} C = UDU^T$, we work on $U\Lambda_q$ with corresponding grid points Uq_k. However, to keep the notation simple, we write Λ_q and q_k and implicitly assume the action of the matrix U.

We are now ready to present the different approximation errors:

Theorem 17 (Gaussian wave packet transform via quadrature rules)**.**
Recall the definition of $\psi_{\mathrm{rec}}^{(\mathrm{rule})} \in L^2(\mathbb{R}^d)$ in (3.25) and assume that the grid points in position space are chosen according to (3.27). The approximation errors

$$E^{(\mathrm{rule})} := \sup_{x \in \Lambda_q} \left| \psi_0(x) - \psi_{\mathrm{rec}}^{(\mathrm{rule})}(x) \right|$$

for the rules (TcM), (RS) *and* (GH) *satisfy the following bounds:*

Truncation and composite midpoint rule (TcM)
There exist positive constants $C^{(\mathrm{T})} > 0$ *and* $C^{(\mathrm{cM})} > 0$ *such that*

$$E^{(\mathrm{TcM})} < C^{(\mathrm{T})} + C^{(\mathrm{cM})} J^{-2}.$$

Infinite Riemann sums (RS)
For all $s \geq 1$*, there exists a positive constant* $C_s^{(\mathrm{RS})} > 0$ *such that*

$$E^{(\mathrm{RS})} < C_s^{(\mathrm{RS})} (\Delta p)^{2s+1}.$$

Gauss–Hermite quadrature (GH)
For all $s \geq 1$*, there exists a positive constant* $C_s^{(\mathrm{GH})} > 0$ *such that*

$$E^{(\mathrm{GH})} < C_s^{(\mathrm{GH})} J^{-s/2}.$$

In particular, the constants $C^{(\mathrm{T})}, C^{(\mathrm{cM})}, C_s^{(\mathrm{RS})}$ *and* $C_s^{(\mathrm{GH})}$ *can be chosen independently of the number* K *of grid points in position space.*

Remark 18. *The total number of grid points in momentum space for fully tensorised quadrature rules is given by* $J_d = J^d$ *and thus we have* $J^{-2} = J_d^{-2/d}$ *and* $J^{-s/2} = J_d^{-s/2d}$.

The proof is presented later in Section 3.1.3 and is based on the error estimates for the individual rules. For a detailed analysis of the constants $C^{(\mathrm{T})}, C^{(\mathrm{cM})}, C_s^{(\mathrm{RS})}$ and $C_s^{(\mathrm{GH})}$ and their dependence on the semiclassical parameter ε we also refer to the proof.

As we can see, approximations based on infinite Riemann sums and Gauss–Hermite quadrature lead to spectral convergence, while the composite midpoint rule only gives order $\mathcal{O}(J^{-2})$. However, since the midpoint rule achieves spectral accuracy for smooth and periodic integrands, see e.g. [SI88, Theorem 8], and $f_{k,x}$ is a Gaussian which can be viewed as a periodic function vanishing at infinity, faster convergence is to be expected in practical applications, as confirmed by our numerical examples in Section 3.1.4.

We now go on to analyse the different discretisation errors.

Truncation error

Recall the definition of the Gaussian integrand $f_{k,x}$ in (3.19). First, we note that if the matrices C and C_0 are purely imaginary, then both the matrix $A = i(C_0 - \overline{C})^{-1}$ and the vector $b_k(x)$ are real-valued. In this case, using the Cholesky decomposition $A = LL^T$, where $L \in \mathbb{R}^{d\times d}$ is a lower triangular matrix with positive diagonal entries, yields

$$\int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p = \det(A)^{-1/2} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2\varepsilon}|y|^2 + \frac{i}{\varepsilon}(L^{-1}b_k(x))^T y \right) \mathrm{d}y.$$

In the general case where $C, C_0 \in \mathfrak{S}^+(d)$ and A and b_k are complex-valued, a linear transformation of the momentum integral leads to a similar result, but with additional transformations for $\operatorname{Im} A$ and $\operatorname{Im} b_k(x)$. However, to keep the calculations simple, in the following we assume that the integrand has the form

$$f_{k,x}(p) = \exp\left(-\frac{1}{2\varepsilon}|p|^2 + \frac{i}{\varepsilon} b_k(x)^T p\right), \tag{3.28}$$

with a real-valued vector $b_k(x) \in \mathbb{R}^d$. All of the following estimates can be extended to the general case and lead to similar, albeit technically more involved, calculations.

From a numerical point of view, the Gaussian decay of the integrand in (3.28) allows the approximation of the improper integral (3.21) by the truncated integral over $\Lambda_p \subset \mathbb{R}^d$. The next lemma provides the corresponding truncation error.

Lemma 19. *Let $\Lambda_p \subset \mathbb{R}^d$ be the box in (3.23). Then, for all $k \in \Gamma_q$, we have*

$$\sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p - \int_{\Lambda_p} f_{k,x}(p)\,\mathrm{d}p \right| \le (2\pi\varepsilon)^{d/2} \exp\left(-\frac{d}{2\varepsilon} L_p^2\right). \tag{3.29}$$

Proof. Let $k \in \Gamma_q$. The triangle inequality for integrals gives us

$$\sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p - \int_{\Lambda_p} f_{k,x}(p)\,\mathrm{d}p \right| \le \int_{\mathbb{R}^d\setminus\Lambda_p} \exp\left(-\frac{1}{2\varepsilon}|p|^2\right)\,\mathrm{d}p.$$

Hence, the symmetry of the integral and Fubini's theorem yields

$$\int_{\mathbb{R}^d\setminus\Lambda_p} \exp\left(-\frac{1}{2\varepsilon}|p|^2\right)\,\mathrm{d}p = \prod_{n=1}^{d} 2\int_{L_p}^{\infty} \exp\left(-\frac{1}{2\varepsilon}z^2\right)\,\mathrm{d}z.$$

Using the exponential-type bound $\operatorname{erfc}(y) \le e^{-y^2}$, $y > 0$, for the complementary error function, see e.g. [CDS03, Equation 5], we conclude that

$$2\int_{L_p}^{\infty} \exp\left(-\frac{1}{2\varepsilon}z^2\right)\,\mathrm{d}z = \sqrt{2\pi\varepsilon}\operatorname{erfc}\left(L_p/\sqrt{2\varepsilon}\right) \le \sqrt{2\pi\varepsilon}\exp\left(-\frac{1}{2\varepsilon}L_p^2\right),$$

which finally yields the following upper bound:

$$\prod_{n=1}^{d} 2\int_{L_p}^{\infty} \exp\left(-\frac{1}{2\varepsilon}z^2\right)\,\mathrm{d}z \le (2\pi\varepsilon)^{d/2}\exp\left(-\frac{d}{2\varepsilon}L_p^2\right).$$

□

Error bounds for fully tensorised quadrature rules

Error bounds for fully tensorised quadrature rules can be derived from one-dimensional theory by applying a given quadrature formula to each variable individually:

Lemma 20. *Consider the one-dimensional quadrature formula*

$$Q_J f := \sum_{j=1}^{J} w_j f(p_j) \approx \int_0^1 f(p)\,\mathrm{d}p$$

using the non-negative weights $w_j \geq 0$, $\sum_{j=1}^{J} w_j = 1$, and abscissas $p_j \in [0,1]$. Moreover, consider the corresponding "Cartesian product" formula

$$Q_J^d f = (Q_J \otimes \cdots \otimes Q_J)\, f := \sum_{j_1=1}^{J} \cdots \sum_{j_d=1}^{J} w_{j_1} \cdots w_{j_d} f(p_{1,j_1}, \ldots, p_{d,j_d})$$

and assume that for all $n \in \{1, \ldots, d\}$ there exists a constant $E_n \geq 0$ such that

$$\left| \int_0^1 f(p_1, \ldots, p_d)\,\mathrm{d}p_n - Q_J(f; p_n) \right| \leq E_n$$

for all values of $p_m \in [0,1]$, $m \neq n$. Then,

$$\left| \int_{[0,1]^d} f(p)\,\mathrm{d}p - Q_J^d f \right| \leq \sum_{n=1}^{d} E_n. \tag{3.30}$$

We have formulated Lemma 20 as a special variant of a more general result that can be found in [Hab70, Section 3]. In the next step we will apply the estimate in (3.30) to the different discretisation schemes (TcM), (RS) and (GH).

Composite midpoint rule

Since the error bounds E_n of the one-dimensional quadrature rules depend crucially on the smoothness of the integrand (in our case $f_{k,x}$), it will be useful to have a formula for the bounds of the higher-order derivatives:

Lemma 21. *For all $k \in \Gamma_q$ and $(x,p) \in \Lambda_q \times \Lambda_p \subset \mathbb{R}^{2d}$, the second-order partial derivatives of the Gaussian integrand $f_{k,x}$ are bounded as follows:*

$$\left|\partial_n^2 f_{k,x}(p)\right| \leq 4d\left(L_p + L_q\right)^2 \varepsilon^{-2} + \varepsilon^{-1}, \quad n = 1, \ldots, d \tag{3.31}$$

Moreover, for all $s \geq 1$, there exists a constant $C_s > 0$, depending on s, L_q and ε, such that for all $k \in \Gamma_q$ and $x \in \Lambda_q$ we have

$$\int_{-\infty}^{\infty} \left|\partial_n^s f_{k,x}(p)\right|\,\mathrm{d}p_n \leq C_s \tag{3.32}$$

for all values of $p_m \in \mathbb{R}$, $m \neq n$.

Proof. For $k \in \Gamma_q$, $x \in \Lambda_q$ and $n \in \{1, \dots, d\}$, let us introduce the complex-valued univariate functions

$$f_{k,x,n}(\xi) := \exp\left(-\frac{1}{2\varepsilon}\xi^2 + \frac{i}{\varepsilon} b_{k,n}(x)\xi\right), \quad \xi \in \mathbb{R},$$

where $b_{k,n}(x)$ denotes the nth component of the real-valued vector $b_k(x)$. Since

$$f_{k,x}(p) = \prod_{n=1}^{d} f_{k,x,n}(p_n),$$

we conclude that for all $s \geq 1$ the derivatives of $f_{k,x}$ can be bounded as follows

$$|\partial_n^s f_{k,x}(p)| \leq |f_{k,x,n}^{(s)}(p_n)|, \quad p \in \mathbb{R}^d,$$

and therefore it suffices to find an upper bound for the derivatives of the Gaussian $f_{k,x,n}$, uniformly in k, x and n. As presented in [DLCS00, Equation 13], for $\alpha, \beta \in \mathbb{C}, \alpha \neq 0$, the sth derivative of the exponential function

$$g(\xi) := \exp\left(\alpha\xi^2 + \beta\xi\right)$$

can be expressed in terms of the second order Kampé de Fériét polynomial as

$$g^{(s)}(\xi) = g(\xi)\, s! \sum_{m=0}^{\lfloor s/2 \rfloor} \frac{\alpha^m (2\alpha\xi + \beta)^{s-2m}}{m!\,(s-2m)!}.$$

Hence, by choosing $\alpha = -1/2\varepsilon$ and $\beta = ib_{k,n}(x)/\varepsilon$, we conclude that $g(\xi) = f_{k,x,n}(\xi)$ and thus we get

$$|f_{k,x,n}^{(s)}(\xi)| \leq e^{-\xi^2/2\varepsilon}\, s! \sum_{m=0}^{\lfloor s/2 \rfloor} \frac{\left(|\xi| + 2L_q\sqrt{d}\right)^{s-2m}}{2^m \varepsilon^{s-m} m!\,(s-2m)!},$$

where we used that

$$\sup_{x \in \Lambda_q} |b_{k,n}(x)| = \frac{1}{2} \sup_{x \in \Lambda_q} \left(|x - q_0)| + |x - q_k|\right) \leq 2L_q\sqrt{d}. \tag{3.33}$$

Consequently, the estimate in (3.31) follows for $s = 2$ and therefore it remains to prove the bound in (3.32). Using the binomial theorem, we further obtain

$$\int_{-\infty}^{\infty} |f_{k,x,n}^{(s)}(\xi)|\, \mathrm{d}\xi \leq \frac{s!}{\varepsilon^s} \sum_{m=0}^{\lfloor s/2 \rfloor} \frac{\varepsilon^m}{2^m m!} \sum_{r=0}^{s-2m} \frac{\left(2L_q\sqrt{d}\right)^{s-2m-r}}{r!\,(s-2m-r)!} \int_{-\infty}^{\infty} |\xi|^r e^{-\xi^2/2\varepsilon}\, \mathrm{d}\xi,$$

where the last integral can be transformed as

$$\int_{-\infty}^{\infty} |\xi|^r e^{-\xi^2/2\varepsilon}\, \mathrm{d}\xi = \varepsilon^{(r+1)/2} \int_{-\infty}^{\infty} |t|^r e^{-t^2/2}\, \mathrm{d}t.$$

In particular, using a formula for moments of the normal distribution, see e.g. [PP02, Equation 5-73], we conclude that

$$M_r := \int_{-\infty}^{\infty} |t|^r e^{-t^2/2}\,\mathrm{d}t = \begin{cases} \sqrt{2\pi}(r-1)!!, & \text{if } r = 2k, \\ 2^{k+1}k!, & \text{if } r = 2k+1. \end{cases}$$

Note that M_r is an increasing function in r and $M_r/r! \leq 2$ for all $r \geq 1$. Hence, for all values of $p_m \in \mathbb{R}$, $m \neq n$, we finally get

$$\begin{aligned} \int_{-\infty}^{\infty} |\partial_n^s f_{k,x}(p)|\,\mathrm{d}p_n &\leq \frac{s!}{\varepsilon^s} \sum_{m=0}^{\lfloor s/2 \rfloor} \frac{\varepsilon^{m+1/2}}{2^m m!} \sum_{r=0}^{s-2m} \frac{\varepsilon^{r/2}\left(2L_q\sqrt{d}\right)^{s-2m-r}}{r!\,(s-2m-r)!} M_r \\ &\leq \frac{s!}{\varepsilon^s} \sum_{m=0}^{\lfloor s/2 \rfloor} \frac{\varepsilon^{m+1/2}}{2^m m!} \frac{M_{s-2m}}{(s-2m)!} \left(\sqrt{\varepsilon} + 2L_q\sqrt{d}\right)^{s-2m} \\ &\leq \frac{s!}{\varepsilon^s} \sum_{m=0}^{\lfloor s/2 \rfloor} \frac{\varepsilon^{m+1/2}}{2^{m-1} m!} \left(\sqrt{\varepsilon} + 2L_q\sqrt{d}\right)^{s-2m} =: C_s. \end{aligned}$$

□

In the next step, we approximate the truncated integral over Λ_p in Lemma 19 using the Cartesian product formula for the one-dimensional composite midpoint rule.

Lemma 22. *Consider the uniform grid on $\Lambda_p \subset \mathbb{R}^d$ defined in (3.22) and let $Q_J^{(\mathrm{cM})}$ denote the one-dimensional quadrature formula*

$$Q_J^{(\mathrm{cM})} f := \Delta p \sum_{j=1}^{J} f(p_j - p_0) \approx \int_{-L_p}^{L_p} f(p)\,\mathrm{d}p.$$

There exists a positive constant $C^{(\mathrm{cM})} > 0$, depending on L_p, L_q and ε, such that for all $k \in \Gamma_q$ and $x \in \Lambda_q$ we have

$$\left| \int_{\Lambda_p} f_{k,x}(p)\,\mathrm{d}p - \left(Q_J^{(\mathrm{cM})}\right)^d f_{k,x} \right| \leq C^{(\mathrm{cM})} J^{-2}. \tag{3.34}$$

Proof. Let $k \in \Gamma_q$ and $x \in \Lambda_q$. As it can be found e.g. in [DR07, Chapter 2.1], the error of the one-dimensional composite midpoint rule is bounded by

$$\left| \int_{-L_p}^{L_p} f_{k,x}(p)\,\mathrm{d}p_n - Q_J^{(\mathrm{cM})}(f_{k,x}; p_n) \right| \leq \frac{L_p^3}{3J^2} \sup_{p \in \Lambda_p} \left|\partial_n^2 f_{k,x}(p)\right|,$$

$n = 1, \ldots, d$, and therefore, using the bound in (3.31), we conclude that

$$\frac{L_p^3}{3J^2} \sup_{p \in \Lambda_p} |\partial_n^s f_{k,x}(p)| \leq \frac{L_p^3}{3J^2} \left(4d\,(L_p + L_q)^2\,\varepsilon^{-2} + \varepsilon^{-1}\right) =: E_n^{(\mathrm{cM})}.$$

Hence, Lemma 20 implies that

$$\left|\int_{\Lambda_p} f_{k,x}(p)\,\mathrm{d}p - \left(Q_J^{(\mathrm{cM})}\right)^d f_{k,x}\right| \le \sum_{n=1}^{d} E_n^{(\mathrm{cM})} = c^{(\mathrm{cM})} J^{-2},$$

where the constant $C^{(\mathrm{cM})}$ is given by

$$C^{(\mathrm{cM})} = d \cdot \frac{L_p^3}{3}\left(4d\left(L_p + L_q\right)^2 \varepsilon^{-2} + \varepsilon^{-1}\right).$$

□

Infinite Riemann sums

In addition to approximating the truncated integral with the midpoint rule as presented in Lemma 22, we can also use infinite Riemann sums to directly approximate the improper integral (3.21). As we will see, the error estimate for this approximation is based on the Euler–Maclaurin formula:

Lemma 23. *Consider the uniform momentum grid defined in (3.24) and let $Q^{(\mathrm{RS})}$ denote the one-dimensional quadrature formula*

$$Q^{(\mathrm{RS})} f := \Delta p \sum_{j\in\mathbb{Z}} f(p_j - p_0) \approx \int_{-\infty}^{\infty} f(p)\,\mathrm{d}p.$$

For all $s \ge 1$, there exists a positive constant $C_s^{(\mathrm{RS})} > 0$, depending on s, L_q and ε, such that for all $k \in \Gamma_q$ and $x \in \Lambda_q$ we have

$$\left|\int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p - \left(Q^{(\mathrm{RS})}\right)^d f_{k,x}\right| \le C_s^{(\mathrm{RS})} (\Delta p)^{2s+1}.$$

Proof. Let $k \in \Gamma_q, x \in \Lambda_q$ and $s \ge 1$. Since the integrand $f_{k,x}$ is a smooth function that vanishes at infinity, we use the Euler–Maclaurin formula, see e.g. [KU98, Theorem 7.2.1], to obtain

$$\left|\int_{-\infty}^{\infty} f_{k,x}(p)\,\mathrm{d}p_n - Q^{(\mathrm{RS})}(f_{k,x}; p_n)\right| \le \int_{-\infty}^{\infty} \left|\partial_n^{2s+1} f_{k,x}(p)\right|\,\mathrm{d}p_n \cdot \frac{(\Delta p)^{2s+1}}{(2\pi)^{2s+1}},$$

for all values of $p_m \in \mathbb{R}$, $m \neq n$. Furthermore, the bound in (3.32) yields

$$\int_{-\infty}^{\infty} \left|\partial_n^{2s+1} f_{k,x}(p)\right|\,\mathrm{d}p_n \cdot \frac{(\Delta p)^{2s+1}}{(2\pi)^{2s+1}} \le \frac{C_{2s+1}(\Delta p)^{2s+1}}{(2\pi)^{2s+1}} =: E_n^{(\mathrm{RS})},$$

and thus the claim follows again by Lemma 20 for $C_s^{(\mathrm{RS})} = d \cdot C_{2s+1}/(2\pi)^{2s+1}$. □

In the last step we use Gauss–Hermite quadrature, which is a special form of Gaussian quadrature on the real line for a Gaussian weight function, see e.g. [DR07, Chapter 1.12].

Gauss–Hermite quadrature

Consider the one-dimensional formula

$$\sum_{j=1}^{J} w_j h(s_j) \approx \int_{-\infty}^{\infty} e^{-p^2} h(p)\,\mathrm{d}p,$$

where the nodes $s_1, \dots, s_J$ are chosen as the zeros of the Jth Hermite polynomial and the positive numbers $w_j > 0$ are the corresponding quadrature weights. In particular, the Jth Hermite polynomial and the weights are given by

$$H_J(x) = (-1)^J e^{x^2} \frac{\mathrm{d}^J}{\mathrm{d}x^J} e^{-x^2} \quad \text{and} \quad w_j = \frac{2^{J+1} J! \sqrt{\pi}}{[H_{J+1}(s_j)]^2}.$$

Note that both the weights and the nodes depend on J, although we do not express this dependence in our notation. We obtain the following error bound:

Lemma 24. *Consider the transformed Gauss–Hermite formula defined by*

$$Q_J^{(\mathrm{GH})} f := \sum_{j=1}^{J} \omega_j f(p_j - p_0) \approx \int_{-\infty}^{\infty} f(p)\,\mathrm{d}p,$$

where the transformed nodes and weights are defined by

$$p_j := p_0 + s_j \sqrt{2\varepsilon} \quad \textit{and} \quad \omega_j := e^{s_j^2} w_j \sqrt{2\varepsilon}.$$

For all $s \geq 1$, there exists a positive constant $C_s^{(\mathrm{GH})} > 0$, depending on s, L_q and ε, such that for all $k \in \Gamma_q$ and $x \in \Lambda_q$ we have

$$\left| \int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p - \left(Q_J^{(\mathrm{GH})}\right)^d f_{k,x} \right| \leq C_s^{(\mathrm{GH})} J^{-s/2}.$$

Proof. Let $k \in \Gamma_q$ and $x \in \Lambda_q$. A linear transformation of the integral yields

$$\int_{\mathbb{R}^d} f_{k,x}(p)\,\mathrm{d}p = (2\varepsilon)^{d/2} \int_{\mathbb{R}^d} e^{-|p|^2} h_{k,x}(p)\,\mathrm{d}p,$$

where the complex-valued function $h_{k,x} \colon \mathbb{R}^d \to \mathbb{C}$ is given for all $p \in \mathbb{R}^d$ by

$$h_{k,x}(p) := \exp\left(i b_k(x)^T p \sqrt{2/\varepsilon} \right).$$

In particular, the partial derivatives of order $s \geq 1$ are bounded for all $p \in \mathbb{R}^d$ by

$$|\partial_n^s h_{k,x}(p)| \leq \left(\sqrt{2/\varepsilon} |b_k(x)| \right)^s \leq \left(2\sqrt{2d/\varepsilon} L_q \right)^s, \quad n = 1, \dots, d,$$

where we used the estimate in (3.33). In [MM94, Theorem 2], the authors prove that if the $(s-1)$th derivative of a function $h\colon \mathbb{R} \to \mathbb{C}$ is locally absolutely continuous and $h^{(s)}(p)e^{-(1-\delta)p^2} \in L^1(\mathbb{R})$ for some $0 < \delta < 1$, then

$$\left| \int_{-\infty}^{\infty} e^{-p^2} h(p) \, \mathrm{d}p - \sum_{j=1}^{J} w_j h(s_j) \right| \leq C J^{-s/2} \| h^{(s)}(p) e^{-(1-\delta)p^2} \|_{L^1(\mathbb{R})},$$

where $C > 0$ is a constant that is independent of J and h. Since for all $0 < \delta < 1$ and $n \in \{1, \dots, d\}$ we have

$$\int_{-\infty}^{\infty} \left| \partial_n^s h_{k,x}(p) e^{-(1-\delta)p_n^2} \right| \mathrm{d}p_n \leq \left(2\sqrt{2d/\varepsilon} L_q \right)^s \sqrt{\frac{\pi}{1-\delta}},$$

we thus obtain the following bound

$$\begin{aligned} &\left| \int_{-\infty}^{\infty} e^{-p_n^2} h_{k,x}(p) \, \mathrm{d}p_n - \sum_{j=1}^{J} w_j \, h_{k,x}(p_1, \dots, p_{n-1}, s_j, p_{n+1}, \dots, p_d) \right| \\ &\quad \leq C J^{-s/2} \left(2\sqrt{2d/\varepsilon} L_q \right)^s \sqrt{\pi} =: E_n^{(\mathrm{GH})} \end{aligned}$$

for all values of $p_m \in \mathbb{R}$, $m \neq n$, and therefore the claim follows again by Lemma 20 for the constant

$$C_s^{(\mathrm{GH})} = C\sqrt{\pi} \cdot 2^{(3s+d)/2} d^{s/2+1} \varepsilon^{(d-s)/2} L_q^s.$$

□

Now that we have derived the corresponding discretisation errors, we can catch up with the proof of Theorem 17.

3.1.3 Proof (Gaussian wave packet transform via quadrature rules)

Proof. Using the representation in (3.9) for the Gaussian wave packet $\psi_0 = g_{z_0}^{C_0,\varepsilon}$, the triangle inequality yields

$$E^{(\mathrm{rule})} \leq \sup_{x \in \Lambda_q} \frac{1}{S(x)} \sum_{k \in \Gamma_q} \left| \mathcal{I}_{q_k}(x) - \sum_{j \in \Gamma_p^{(\mathrm{rule})}} r_{j,k}^{(\mathrm{rule})} \, g_{j,k}(x) \right|.$$

In the following, let us consider rule = TcM. Combining the representation of $\mathcal{I}_q(x)$ according to Lemma 16 with the truncation and the composite midpoint rule, we obtain

$$\begin{aligned} \mathcal{I}_{q_k}(x) &\overset{(3.15)}{=} g_0(x - q_k) c_k(x) \int_{\mathbb{R}^d} f_{k,x}(p) \, \mathrm{d}p \\ &\overset{(3.29)}{\approx} g_0(x - q_k) c_k(x) \int_{\Lambda_p} f_{k,x}(p) \, \mathrm{d}p \quad \text{(truncation)} \end{aligned}$$

$$\overset{(3.34)}{\approx} \sum_{j_1=1}^{J} \cdots \sum_{j_d=1}^{J} r_{j,k}^{(\mathrm{TcM})}\, g_{j,k}(x) \quad \text{(composite midpoint rule)},$$

where the representation coefficients $r_{j,k}^{(\mathrm{TcM})} \in \mathbb{C}$ are given by (cf. Equation (3.26)):

$$r_{j,k}^{(\mathrm{TcM})} = \frac{(\Delta p)^d}{(2\pi\varepsilon)^d} \langle g_{j,k} \mid \psi_0 \rangle = (\Delta p)^d c_k(x) f_{k,x}(p_j - p_0) e^{-ip_j \cdot (x-q_k)/\varepsilon}.$$

Hence, for all $j \in \{1, \ldots, J\}^d$ and $k \in \Gamma_q = \{1, \ldots, K\}^d$, we get

$$\begin{aligned}&\left| \mathcal{I}_{q_k}(x) - \sum_{j_1=1}^{J} \cdots \sum_{j_d=1}^{J} r_{j,k}^{(\mathrm{TcM})}\, g_{j,k}(x) \right| \\ &\qquad = |c_k(x)|\, |g_0(x - q_k)| \left| \int_{\mathbb{R}^d} f_{k,x}(p)\, \mathrm{d}p - \left(Q_J^{(\mathrm{cM})}\right)^d f_{k,x} \right|,\end{aligned}$$

and by definition of the numbers $c_k(x)$ in Lemma 16 it follows that

$$\sup_{x \in \Lambda_q} |c_k(x)| = (2\pi\varepsilon)^{-d} \exp\left(-\frac{1}{8\varepsilon} |q_k - q_0|^2 \right) \leq (2\pi\varepsilon)^{-d}.$$

Moreover, by Lemma 19 (truncation) and Lemma 22 (composite midpoint rule), we conclude that there are positive constants $\tilde{C}^{(\mathrm{T})} > 0$ and $\tilde{C}^{(\mathrm{cM})} > 0$ such that

$$\sup_{x \in \Lambda_q} \left| \int_{\mathbb{R}^d} f_{k,x}(p)\, \mathrm{d}p - \left(Q_J^{(\mathrm{cM})}\right)^d f_{k,x} \right| \leq \tilde{C}^{(\mathrm{T})} + \tilde{C}^{(\mathrm{cM})} J^{-2}.$$

Consequently, since there exists a constant $C_{\Gamma_q} > 0$ (see Appendix 7.2), depending on the width matrix of the basis functions, ε and L_q, but independent of K (number of grid points in position space), such that

$$\sup_{x \in \Lambda_q} \frac{1}{S(x)} \sum_{k_1=1}^{K} \cdots \sum_{k_d=1}^{K} |g_0(x - q_k)| < C_{\Gamma_q}^d,$$

we finally conclude that

$$\begin{aligned}&\sup_{x \in \Lambda_q} \frac{1}{S(x)} \sum_{k \in \Gamma_q} \left| \mathcal{I}_{q_k}(x) - \sum_{j \in \Gamma_p^{(\mathrm{TcM})}} r_{j,k}^{(\mathrm{TcM})}\, g_{j,k}(x) \right| \\ &\qquad \leq (2\pi\varepsilon)^{-d} \left(\tilde{C}^{(\mathrm{T})} + \tilde{C}^{(\mathrm{cM})} J^{-2} \right) \sup_{x \in \Lambda_q} \frac{1}{S(x)} \sum_{k_1=1}^{K} \cdots \sum_{k_d=1}^{K} |g_0(x - q_k)| \\ &\qquad < C^{(\mathrm{T})} + C^{(\mathrm{cM})} J^{-2},\end{aligned}$$

where the positive constants $C^{(\mathrm{T})}, C^{(\mathrm{cM})} > 0$ are given by

$$C^{(\mathrm{T})} = (2\pi\varepsilon)^{-d} \tilde{C}^{(\mathrm{T})} C_{\Gamma_q}^d \quad \text{and} \quad C^{(\mathrm{cM})} = (2\pi\varepsilon)^{-d} \tilde{C}^{(\mathrm{cM})} C_{\Gamma_q}^d.$$

The corresponding estimates for infinite Riemann sums and Gauss–Hermite quadrature follow the same arguments as presented for (TcM), but using Lemma 23 and Lemma 24.

□

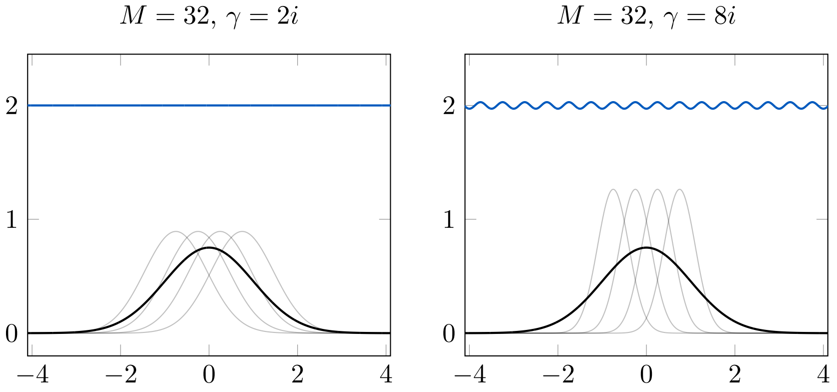


Figure 3.2: Gaussian wave packet ψ_0 (black), four basis functions around the centre $q_0 = 0$ (grey) and summation curve (blue) for two choices of the width parameter γ. The larger width (left) leads to a better approximation to the summation curve, which is a consequence of the larger overlap and can be derived in particular from the analytical representation in Lemma 13.

3.1.4 Numerical experiments

We present experiments for the reconstruction of one-dimensional Gaussian wave packets according to Theorem 17, which illustrate the superiority of our new representation by Gauss–Hermite quadrature. Two examples are used to visualise the dependence of the errors on the various parameters involved. The first example deals with the interplay of γ (width of the basis functions) and K (number of grid points in position space), while the second investigates the dependence on the semiclassical parameter ε, which controls the oscillations of the wave packets.

Example 1

We consider the wave packet

$$\psi_0(x) = \pi^{-1/4} \exp\left(-\frac{1}{2}x^2\right), \qquad \Big[\varepsilon = 1,\, \gamma_0 = i,\, (q_0, p_0) = (0, 0)\Big],$$

on $\Lambda_q = [-8, 8]$. Note that this corresponds to $L_q = 8$. The plots in Figure 3.2 show the wave packet ψ_0 together with four basis functions around the centre $q_0 = 0$ (grey) and the summation curve (blue) for $\gamma = 2i$ (left) and $\gamma = 8i$ (right). In both plots, the summation curve is built on a uniform grid with $K = 32$ grid points, giving a distance of $\Delta q = 0.5$ for the basis functions. In particular, smaller values of $\mathrm{Im}(\gamma)$ cause the spread (and hence the overlap) of the basis functions to increase, giving a faster convergence of the summation curve. This can also be seen in the plots: For the summation curve on the right-hand side we see the typical oscillations as we know them from the cosine function, while on the left-hand side no oscillations can be seen because the summation curve has approached the predicted value $1/\Delta q = 2$.

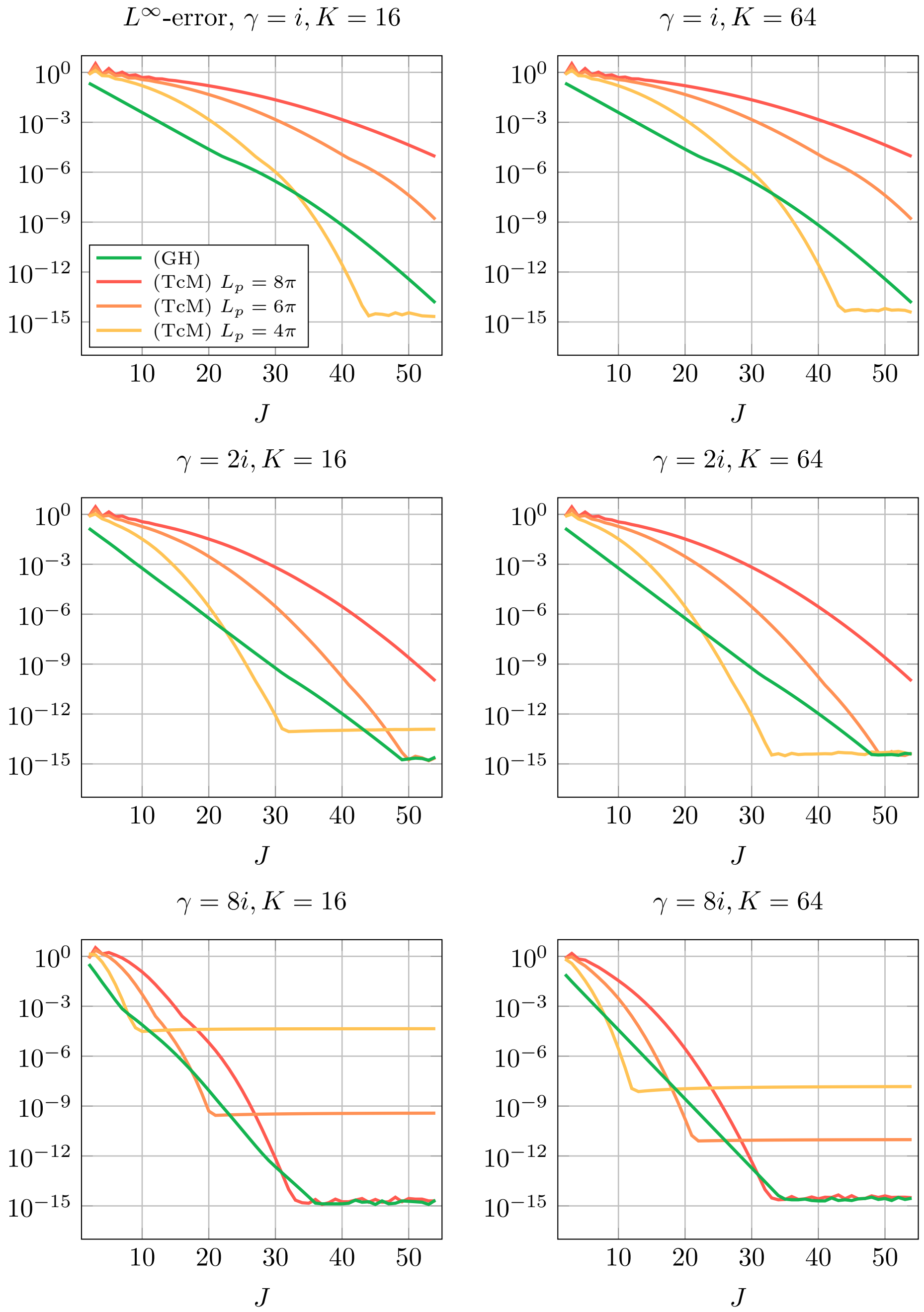


Figure 3.3: Reconstruction error for different combinations of γ and K. The approximations based on (TcM) show a fast initial decay. For (GH), all plots initially show an exponential decay (green lines).

The plots in Figure 3.3 show the reconstruction errors in the supremum norm on Λ_q for different combinations of γ and K. For each of the three rows (γ is fixed here) we compare $K = 16$ (left) and $K = 64$ (right) and three choices for the truncation parameter in momentum space ($L_p = 4\pi, 6\pi, 8\pi$). For (TcM), we observe that larger values of L_p lead to a worse decay of the errors, which is consistent with our theoretical result in Lemma 22. Furthermore, fast initial decay is observed for (TcM), which can be explained by the fact that the composite midpoint rule achieves spectral accuracy. The plots also show, for example, that for the smallest box (yellow) the truncation error is reached at about 10 grid points (plateau for $\gamma = 8i, K = 16$). A comparison of the two columns in Figure 3.3 shows that the error is only slightly affected by the number of grid points in position space (left: $K = 16$, right: $K = 64$), which can be explained by the fact that all error constants are independent of K (cf. in Theorem 17). In the reconstructions based on Gauss–Hermite quadrature (green lines), the errors initially show an exponential decay (Lemma 24 predicts spectral convergence). In particular, all plots show the superiority of (GH) for small values of J. As we will see in the next example, the discrepancy between (TcM) and (GH) becomes even more pronounced if the underlying wave packet is oscillating.

Example 2

For $\varepsilon \in \big\{0.1, 0.05\big\}$ we consider the wave packet

$$\psi_0(x) = (\pi\varepsilon)^{-1/4} \exp\left(-\frac{1}{2\varepsilon}(x-1)^2 + \frac{2i}{\varepsilon}(x-1)\right), \quad \Big[\gamma_0 = i,\ (q_0, p_0) = (1, 2)\Big],$$

on $\Lambda_q = [-8, 8]$. For small values of ε, it follows from the presence of the complex phase factor that the wave packet is oscillatory, see Figure 3.4. For all computations we used $K = 128$, which corresponds to a uniform spacing of $\Delta q = 1/8$, and two values for the width of the basis functions ($\gamma = 16i$ and $\gamma = 32i$). For $\varepsilon = 0.1$, a plot of four basis functions and the summation curve can be found in Figure 3.5. As we have already discussed in the previous example, the smaller value of $\mathrm{Im}(\gamma)$ (left) gives a better approximation to the constant value $1/\Delta q = 8$. The errors in the reconstruction of the wave packets can be seen in Figure 3.6. All plots underline the superiority of (GH) for wave packets with high oscillations, independent of the width of the basis functions. The experiment clearly shows that with the new rule the number of grid points can be significantly reduced.

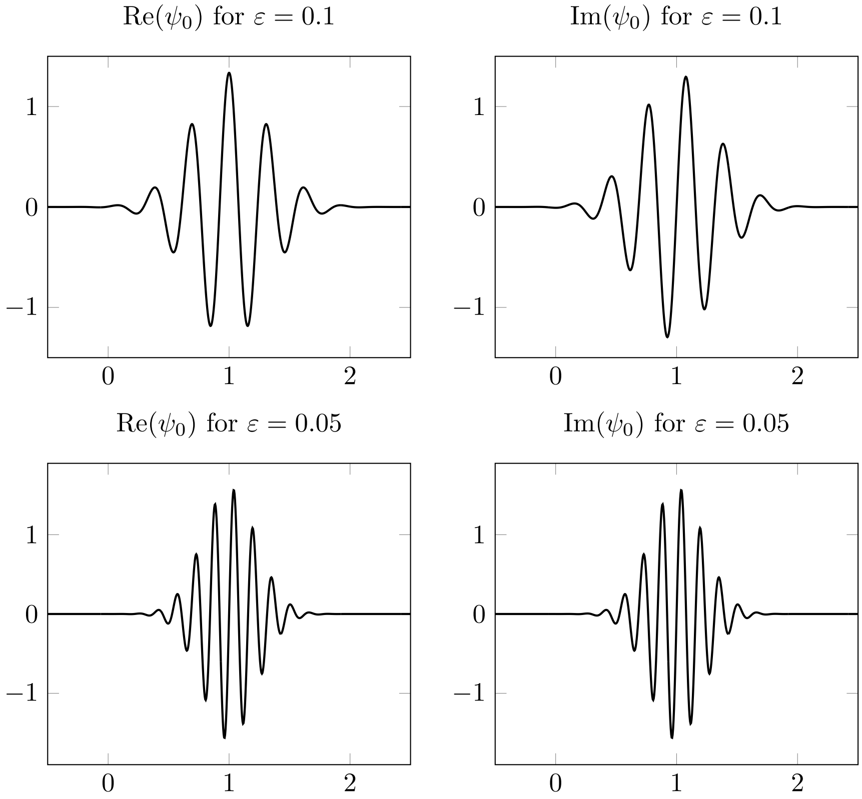


Figure 3.4: Real (left) and imaginary part (right) of the wave packet ψ_0 for different values of ε. Smaller values result in higher oscillations.

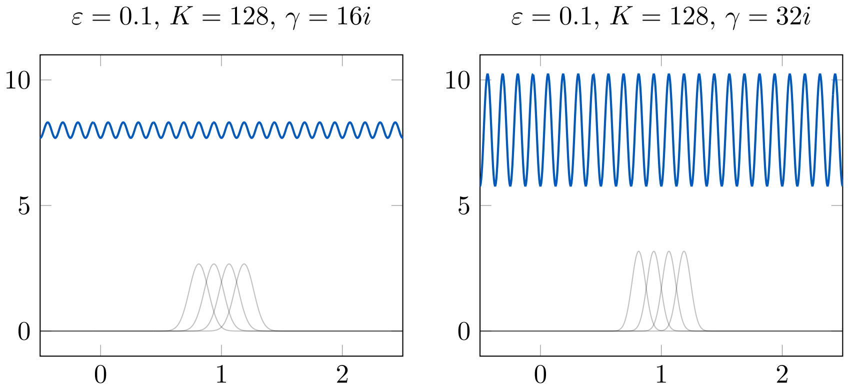


Figure 3.5: Plot of four basis functions (gray) and the summation curve (blue). For the larger value of $\mathrm{Im}(\gamma)$ (right), the summation curve shows larger oscillations.

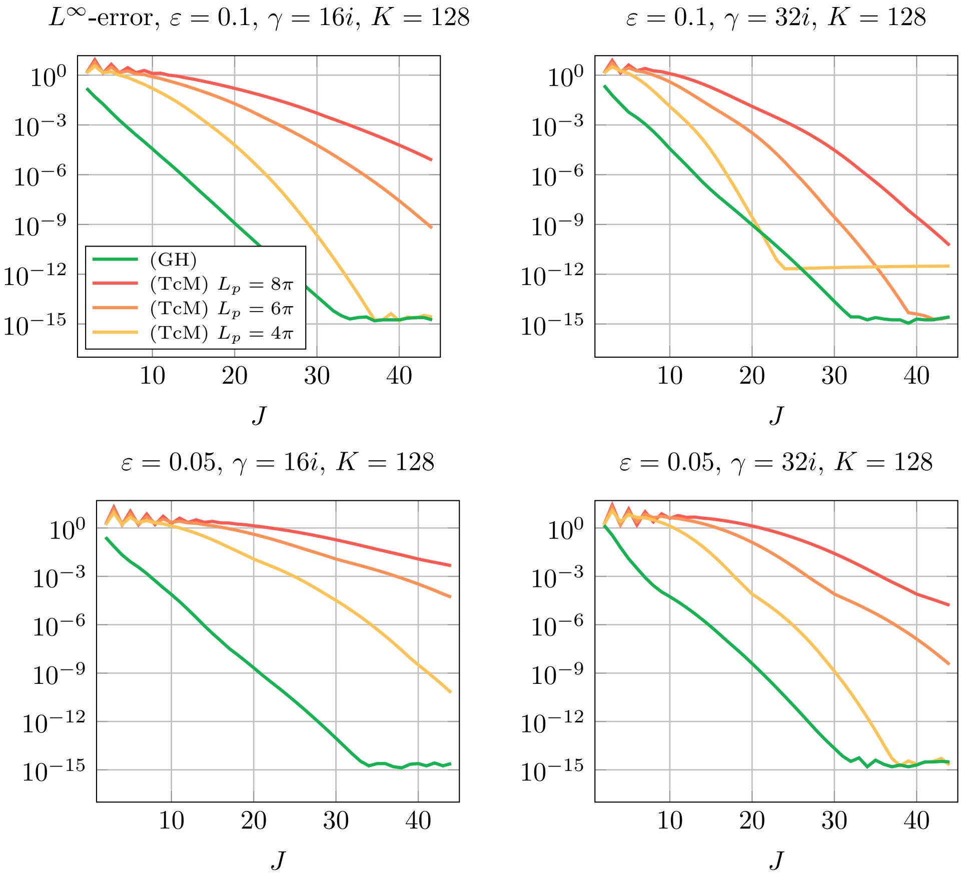


Figure 3.6: Reconstruction errors for different values of γ and ε. The approximations based on Gauss–Hermite quadrature (green) show the best decay. Compared to the midpoint rule, the number of grid points can be significantly reduced, especially for smaller values of ε (higher oscillations in the wave packet).

3.2 Full discretisation of the phase space integral

As described in detail in Section 1.1, parts of the present section (Sec. 3.2) overlap to a large extent with the joint preprint "An Error Representation for the Time-Sliced Thawed Gaussian Propagation Method" with C. Lasser submitted to *Numerische Mathematik* on 27/08/2021, e-print available at arXiv:2108.12182.

In the previous section we derived the semi-discrete representation (3.18) and showed that discretisations of the momentum integral lead to approximations of the form

$$\psi \approx \psi_{\mathrm{rec}}^{(\mathrm{rule})}(x) = \frac{1}{S(x)} \sum_{k\in\Gamma_q} \sum_{j\in\Gamma_p^{(\mathrm{rule})}} r_{j,k}^{(\mathrm{rule})}\, g_{j,k}(x). \tag{3.35}$$

In particular, using uniform Riemann sums in both position and momentum space, the formula for the corresponding coefficients $r_{j,k}^{(\mathrm{RS})}$ in (3.26) and the approximation of the summation curve $S(x)$ according to Lemma 13 yields the approximation

$$\psi(x) \approx \left(\frac{\Delta q \Delta p}{2\pi\varepsilon}\right)^d \sum_{k\in\Gamma_q} \sum_{j\in\Gamma_p^{(\mathrm{RS})}} \langle g_{j,k} \mid \psi\rangle\, g_{j,k}(x),$$

which can be seen as a direct discretisation of the phase space integral. Regarding the approximation of solutions of the time-dependent Schrödinger equation, we recognise that the latter approximation is advantageous since it allows the propagation of the wave packet ψ to be directly transferred to the Gaussian basis functions by linearity, whereas this is not possible in (3.35) due to the additional prefactor $1/S(x)$.

It is of course well known that fully tensorised quadrature rules are impractical from a numerical point of view because the number of grid points increases exponentially with dimension, but since uniform Riemann sums in each coordinate direction have been used in moderate dimensions without error estimates by Kong *et al.*, we extend our error representations from Section 3.1 to include direct discretisations of the phase space integral. We therefore adjust our notation as follows: For a given finite multi-index set $\Gamma \subset \mathbb{N}^{2d}$, e.g. a cube $\{\mathbf{n} \in \mathbb{N}^{2d} : n_j \le N\}$ or a simplex $\{\mathbf{n} \in \mathbb{N}^{2d} : \sum_{j=1}^{2d} n_j \le N\}$, the bold multi-index $\mathbf{n} \in \Gamma$ is used to specify grid points $z_{\mathbf{n}}$ in phase space, where the first d components correspond to the position space and the last d components to the momentum space. With the previously used indices $k \in \Gamma_q$ and $j \in \Gamma_p$ we could also write $\mathbf{n} = (k, j)$. Accordingly, we write the discretised FBI formula as

$$\psi \approx \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi)\, g_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} w_{\mathbf{n}} \langle g_{\mathbf{n}} \mid \psi\rangle\, g_{\mathbf{n}} \tag{3.36}$$

with basis functions $g_{\mathbf{n}} = g_{z_{\mathbf{n}}}$ and complex-valued representation coefficients

$$c_{\mathbf{n}}(\psi) := w_{\mathbf{n}} \langle g_{\mathbf{n}} \mid \psi\rangle,$$

which result from weighted point evaluations of the integrand and depend on ψ and the positive weights $w_{\mathbf{n}} > 0$. For example, the weights for uniform Riemann sums based on grid sizes $\Delta q_j > 0$ and $\Delta p_j > 0$ in each coordinate direction $j = 1, \dots, d$ are given by

$$w_{\mathbf{n}} = (2\pi\varepsilon)^{-d} \prod_{j=1}^{d} \Delta q_j \Delta p_j.$$

Furthermore, we introduce the following quadrature-based pair of operators.

Definition 25. *For a given finite multi-index set $\Gamma \subset \mathbb{N}^{2d}$, let $\{z_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ be a grid in phase space and $w_{\mathbf{n}} > 0$ positive weights. The operator*

$$\mathcal{A}_\Gamma \colon L^2(\mathbb{R}^d) \to \mathbb{C}^\Gamma,\ (\mathcal{A}_\Gamma \psi)_{\mathbf{n}} := \langle g_{\mathbf{n}} \mid \psi \rangle \quad \text{for all } \mathbf{n} \in \Gamma, \tag{3.37}$$

is called the analysis operator*. Moreover, the operator*

$$\mathcal{S}_\Gamma \colon \mathbb{C}^\Gamma \to L^2(\mathbb{R}^d),\ (s_{\mathbf{n}}) \mapsto \psi_\Gamma := \sum_{\mathbf{n}\in\Gamma} w_{\mathbf{n}} s_{\mathbf{n}}\, g_{\mathbf{n}},$$

which maps a given coefficient tensor $(s_{\mathbf{n}})$ to the weighted Gaussian superposition ψ_Γ, is called the (weighted) synthesis operator.

The next lemma shows that the operators $\mathcal{A}_\Gamma$ and $\mathcal{S}_\Gamma$ are formally adjoint with respect to weighted inner products and therefore we can write $\mathcal{S}_\Gamma = \mathcal{A}_\Gamma^*$.

Lemma 26. *Let $\Gamma \subset \mathbb{N}^{2d}$ be a finite multi-index set. Moreover, for all $\mathbf{n} \in \Gamma$, let $w_{\mathbf{n}} > 0$ be positive weights. For $x, y \in \mathbb{C}^\Gamma$ we define the weighted inner product*

$$\langle x, y \rangle_w := \sum_{\mathbf{n}\in\Gamma} w_{\mathbf{n}} \overline{x_{\mathbf{n}}}\, y_{\mathbf{n}}.$$

Then, for all $\psi \in L^2(\mathbb{R}^d)$ and $s \in \mathbb{C}^\Gamma$, we have

$$\langle \mathcal{S}_\Gamma s \mid \psi \rangle = \langle s, \mathcal{A}_\Gamma \psi \rangle_w.$$

Proof. Let $\psi \in L^2(\mathbb{R}^d)$ and $s \in \mathbb{C}^\Gamma$. By definition of the operators $\mathcal{A}_\Gamma$ and $\mathcal{S}_\Gamma$ we have

$$\langle \mathcal{S}_\Gamma s \mid \psi \rangle = \int_{\mathbb{R}^d} \sum_{\mathbf{n}\in\Gamma} \overline{w_{\mathbf{n}} s_{\mathbf{n}}\, g_{\mathbf{n}}(x)} \psi(x)\, \mathrm{d}x = \sum_{\mathbf{n}\in\Gamma} w_{\mathbf{n}} \overline{s_{\mathbf{n}}} \langle g_{\mathbf{n}} \mid \psi \rangle = \langle s, \mathcal{A}_\Gamma \psi \rangle_w.$$

□

Let us assume again that we are interested in the representation of a Gaussian wave packet $\psi_0 = g_{z_0}^{C_0,\varepsilon}$ for some $z_0 \in \mathbb{R}^{2d}$ and $C_0 \in \mathfrak{S}^+(d)$. Similar to the discretisation of the momentum integral based on (TcM) in the previous section, the total error in the discretisation of the phase space integral based on fully tensorised Riemann sums consists of two different sources: In the first step, the improper integral must be truncated with respect to a compact phase space box $\Lambda \subset \mathbb{R}^{2d}$, which should of course be located at the centre z_0 and be aligned with the width matrix of the Gaussian wave packet ψ_0. This produces a truncation error. In the second step, the truncated integral must then be approximated by a multidimensional Riemann sum, which produces a discretisation error. In particular, the total error can be written using the above analysis and synthesis operators as follows

$$E_{wp} = E_{wp}(\psi_0, \Lambda, \Gamma) := \|\psi_0 - \mathcal{A}_\Gamma^* \mathcal{A}_\Gamma \psi_0\|_{L^\infty(\Lambda_q)}, \tag{3.38}$$

where $\Lambda_q \subset \mathbb{R}^d$ now denotes the projection of the phase space box Λ onto position space, which is the domain of ψ_0 and its approximation $\mathcal{A}_\Gamma^* \mathcal{A}_\Gamma \psi_0 \in L^2(\mathbb{R}^d)$. In the next step we analyse this error in more detail.

Truncation error

Recall that the formula for inner products of Gaussians in Lemma 7 shows that the FBI transform of ψ_0 has a Gaussian envelope in phase space. Similar to Lemma 19, we therefore expect good approximations if we truncate the phase space integral using a sufficiently large hypercube.

Lemma 27. *For a given phase space centre $z_0 \in \mathbb{R}^{2d}$ and a positive parameter $L > 0$, consider the phase space box*

$$\Lambda = \prod_{j=1}^{2d} [z_{0,j} - L, z_{0,j} + L] \subset \mathbb{R}^{2d}. \tag{3.39}$$

Moreover, for $C, C_0 \in \mathfrak{S}^+(d)$ let $g_z = g_z^{C,\varepsilon}$ and $\psi_0 = g_{z_0}^{C_0,\varepsilon}$ and assume that the eigenvalues of the matrices $\operatorname{Im}(C), \operatorname{Im}(C_0)$ and $\operatorname{Im}(-C^{-1}), \operatorname{Im}(-C_0^{-1})$ are bounded from below by $\theta > 0$ and from above by $\Theta > 0$. Then, there exists a positive constant $C^{(\mathrm{T})} > 0$, depending on ε and the spectral parameters, such that

$$\sup_{x \in \mathbb{R}^d} \left| \psi_0(x) - (2\pi\varepsilon)^{-d} \int_\Lambda \langle g_z \mid \psi_0 \rangle \, g_z(x) \, \mathrm{d}z \right| \le C^{(\mathrm{T})} \exp\left(-\frac{\theta d}{4\varepsilon} L^2 \right). \tag{3.40}$$

Proof. Using the bound for the inner product of Gaussian wave packets in Lemma 7 and for the constant ζ in (7.4), we obtain the following estimate for all $x \in \mathbb{R}^d$:

$$\begin{aligned} \left| \psi_0(x) - (2\pi\varepsilon)^{-d} \int_\Lambda \langle g_z \mid \psi_0 \rangle \, g_z(x) \, \mathrm{d}z \right| &\le (2\pi\varepsilon)^{-d} \|g_z\|_\infty \int_{\mathbb{R}^{2d} \setminus \Lambda} |\langle g_z \mid \psi_0 \rangle| \, \mathrm{d}z \\ &\le (2\pi\varepsilon)^{-d} (\pi\varepsilon)^{-d/4} \theta^{-d/2} \Theta^{3d/4} \int_{\mathbb{R}^{2d} \setminus \Lambda} \exp\left(-\frac{\theta}{8\varepsilon} \|z - z_0\|_2^2 \right) \mathrm{d}z. \end{aligned}$$

Furthermore, the symmetry of the integral and Fubini's theorem yields that

$$\int_{\mathbb{R}^{2d} \setminus \Lambda} \exp\left(-\frac{\theta}{8\varepsilon} \|z - z_0\|_2^2 \right) \mathrm{d}z = 4^d \left(\int_L^\infty \exp\left(-\frac{\theta}{8\varepsilon} y^2 \right) \mathrm{d}y \right)^{2d}.$$

Hence, using again the exponential-type bound $\operatorname{erfc}(y) \leq e^{-y^2}$, $y > 0$, for the complementary error function (cf. Lemma 19), we conclude that

$$\int_L^\infty \exp\left(-\frac{\theta}{8\varepsilon}y^2\right) \mathrm{d}y = \frac{\sqrt{2\pi\varepsilon}}{\sqrt{\theta}} \operatorname{erfc}\left(L\sqrt{\theta/8\varepsilon}\right) \leq \frac{\sqrt{2\pi\varepsilon}}{\sqrt{\theta}} \exp\left(-\frac{\theta}{8\varepsilon}L^2\right),$$

and therefore we finally get

$$\int_{\mathbb{R}^{2d}\setminus\Lambda} \exp\left(-\frac{\theta}{8\varepsilon}\|z - z_0\|_2^2\right) \mathrm{d}z \leq 4^d(2\pi\varepsilon)^d\theta^{-d} \exp\left(-\frac{\theta d}{4\varepsilon}L^2\right).$$

This proves that the constant $C^{(\mathrm{T})}$ in (3.40) can be chosen as

$$C^{(\mathrm{T})} = \left(\frac{4\Theta^{3/4}}{(\pi\varepsilon)^{1/4}\theta^{3/2}}\right)^d.$$

□

Lemma 27 shows that the truncation error decreases exponentially with the length of the hypercube. Furthermore, the proof shows that the error bound could be further improved by using separate boxes $\Lambda_q \subset \mathbb{R}^d$ and $\Lambda_p \subset \mathbb{R}^d$ in position and momentum space which are aligned with the eigenvectors of the width matrix of ψ_0.

Fully tensorised uniform Riemann sums revisited

In Lemma 20 we have already shown how to derive error bounds for fully tensorised quadrature rules from the one-dimensional theory. As a special case, we obtain the following result if uniform grids are used in each coordinate direction.

Lemma 28. *Let $f \in C^\infty(\mathbb{R}^{2d})$. There exists a positive constant $C_f > 0$, depending only on the function f, such that*

$$\left| \int_{[0,1]^{2d}} f(z)\,\mathrm{d}z - N^{-2d} \sum_{\mathbf{n}\in\Gamma} f\left(\frac{n_1}{N}, \dots, \frac{n_{2d}}{N}\right) \right| \leq C_f \cdot d \cdot N^{-1},$$

where $\Gamma = \{1, 2, \dots, N\}^{2d}$. In particular, C_f can be chosen as the total variation of the function f in the sense of Hardy and Krause.

We have formulated Lemma 28 as a special variant of a more general result that can be found in [DR07, Chapter 5.5.5]. The proof uses the same techniques we used for the discretisation of the momentum integral in Section 3.1.2 and is therefore omitted.

3.2.1 Error for the full phase space discretisation

The total error for the full discretisation of the FBI formula based on uniform Riemann sums is now obtained by combining the estimates in Lemma 27 and Lemma 28:

Proposition 29. *Let $C_0 \in \mathfrak{S}^+(d)$ and $z_0 \in \mathbb{R}^{2d}$. For the discretisation of*

$$(2\pi\varepsilon)^{-d} \int_{\mathbb{R}^{2d}} \left\langle g_z \mid g_{z_0}^{C_0,\varepsilon} \right\rangle g_z \, \mathrm{d}z,$$

using the phase space box $\Lambda \subset \mathbb{R}^{2d}$ in (3.39) and uniform Riemann sums with $N \geq 1$ grid points in each coordinate direction, there exist positive constants $C^{(\mathrm{T})}, C^{(\mathrm{RS})} > 0$ such that the total reconstruction error defined in (3.38) is bounded by

$$E_{wp} \leq C^{(\mathrm{T})} + C^{(\mathrm{RS})} N^{-1}.$$

This result is not surprising, since we have already seen in Theorem 17 that the discretisation of the momentum integral via (TcM) leads to an error of order $\mathcal{O}(J^{-2})$, where J is the number of grid points in momentum space. In fact, the estimate in Lemma 28 can be improved to a bound of order $\mathcal{O}(N^{-2})$ if the composite midpoint rule is used instead of the composite rectangle rule. Nevertheless, we recognise the advantage of the semi-discrete representation (3.18), which allows the separation of position and momentum space via the summation curve and thus the more efficient Gauss–Hermite quadrature can be used for the discretisation of the momentum integral.

3.2.2 Numerical experiments

We present numerical experiments for the approximation of a Gaussian wave packet according to Proposition 29 based on uniform Riemann sums for

$$\psi_0(x) = (\pi\varepsilon)^{-1/4} \exp\left(-\frac{1}{2\varepsilon}(x + \sqrt{2\eta})^2\right), \quad \eta = 1.3544, \tag{3.41}$$

which is later used in Section 5.4 as the initial wave function for the TSTG method. Figure 3.7 shows the reconstruction errors in the supremum norm as a function of grid points for different truncation boxes $\Lambda = [-L_q, L_q] \times [-L_p, L_p]$, where we used the same number of grid points for both intervals. For each column (the width γ of the basis functions is fixed here) we compare $\varepsilon = 1$ (top) and $\varepsilon = 0.1$ (bottom). All panels show that larger boxes lead to a worse decay of the errors, which is consistent with Lemma 27. In particular, the two upper plots show that for the smallest box (yellow) the truncation error is reached after about 64 grid points (plateaus). Moreover, we see that the number of grid points needed to achieve a certain accuracy increases with decreasing ε.

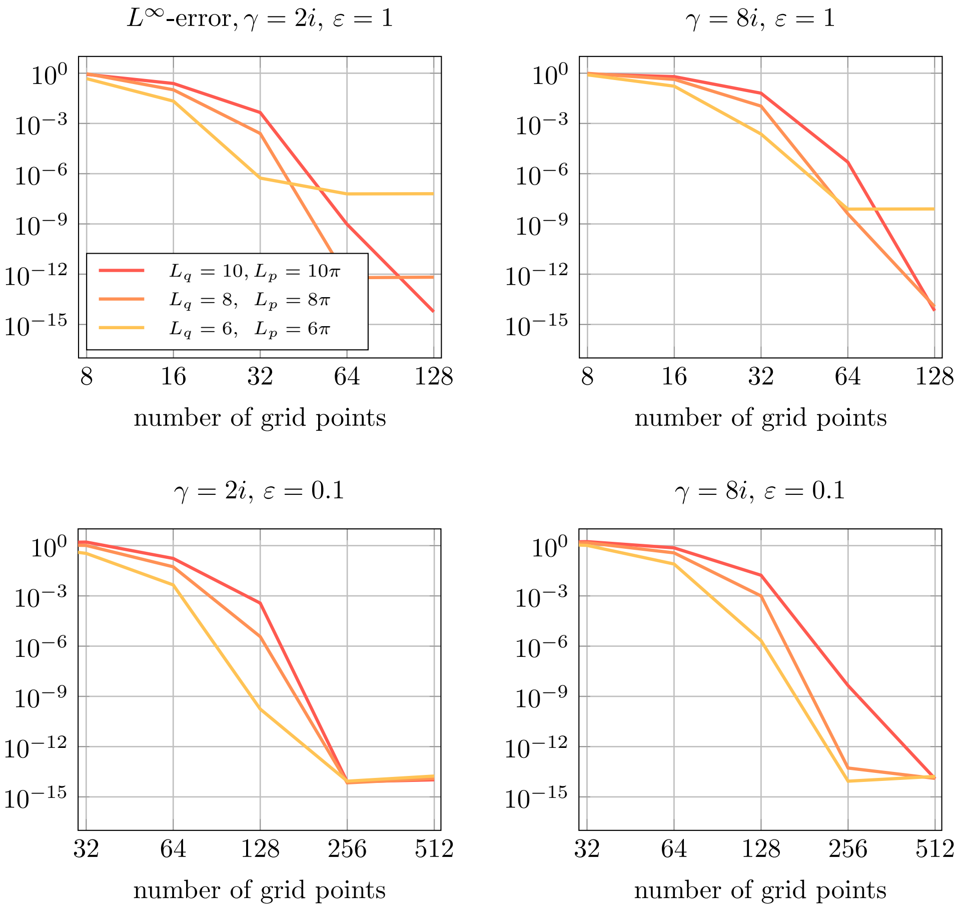


Figure 3.7: The four panels show the reconstruction errors for different combinations of γ (width of the basis functions) and ε. The number of grid points used to achieve a certain accuracy depends on ε and the truncation box.

3.2.3 A note on sparse grids and Monte Carlo integration

In Section 3.1.2 we have seen that the grid points q_k and p_j resulting from the discretisation of the FBI formula depend on the centre of the Gaussian wave packet we want to represent. This is of course a problem if many different functions have to be represented at the same time, because this means that many different grids are needed. In the TSTG method, for example, every individual time-evolved basis function must be represented in the original basis. To avoid the dependence for each individual basis function, Kong *et al.* choose sufficiently dense uniform grids in phase space, although other "function-independent" grids would also be possible. The question therefore arises whether there are alternatives to fully tensorised quadrature rules. In the following we address two possibilities that have been adopted from [LL20, Section 8].

In moderate dimensions, so-called "sparse-grid methods" offer an alternative, as they can overcome the curse of dimensionality to a certain extent. The idea of sparse grids is to rewrite the tensorised quadrature rule

$$Q_N^d f = (Q_N \otimes \cdots \otimes Q_N)\, f = \sum_{n_1=1}^{N} \cdots \sum_{n_d=1}^{N} w_{n_1} \cdots w_{n_d} f(x_{1,n_1}, \ldots, x_{d,n_d})$$

for $N = 2^L$ grid points in each coordinate direction as follows

$$Q_N^d f = \sum_{l_1=0}^{L} \cdots \sum_{l_d=0}^{L} (\Delta_{l_1} \otimes \cdots \otimes \Delta_{l_d})\, f,$$

where the one-dimensional difference formulas Δ_{l_j} are given for $N_j := 2^{l_j}$ by

$$\Delta_{l_j} := Q_{N_j} - Q_{N_{j-1}}, \quad \Delta_0 := Q_1.$$

By using only terms with $l_1 + ... + l_d \leq L$, we then obtain the sparse grid formula

$$S_L^d f := \sum_{l_1+...+l_d \leq L} (\Delta_{l_1} \otimes \cdots \otimes \Delta_{l_d})\, f.$$

These variants require less than $N(\log N)^{d-1}$ quadrature nodes, while N^d are required for the full tensor grid. Sparse grids have been used for the midpoint rule, see [BD93], as well as for Gauss–Hermite quadrature, see [LL20, Section 8.1]. In particular, an error bound for approximations based on sparse-grid Gauss–Hermite quadrature can be found in [LL20, Theorem 8.2]. We also refer to [GG98] for a comprehensive presentation of sparse grids and further developments.

If no fixed grid structure is needed, e.g. for the decomposition of the initial state in the TSTG method, (quasi-) Monte Carlo methods provide a useful alternative and have already been used for the discretisation of the FBI formula in connection with the discretisation of the Herman–Kluk propagator, see [LS17]. To recall the basic idea of Monte Carlo methods, consider the integral

$$(2\pi\varepsilon)^{-d}\int_{\mathbb{R}^{2d}}\langle g_z \mid g_{z_0}\rangle\; g_z\,\mathrm{d}z$$

for a Gaussian wave packet of unit width, *i.e.*, $C = C_0 = i\,\mathrm{Id}$, where as usual C denotes the width matrix of the basis function g_z and C_0 the width of g_{z_0}. Using the formula for the inner products of Gaussians from Lemma 7, we write

$$(2\pi\varepsilon)^{-d}\langle g_z \mid g_{z_0}\rangle = r_0(z)\mu_0(z)$$

with

$$r_0(z) := 2^d \exp\left(\frac{i}{2\varepsilon}(p+p_0)^T(q-q_0)\right)$$

and

$$\mu_0(z) := (4\pi\varepsilon)^{-d}\exp\left(-\frac{1}{4\varepsilon}|z-z_0|^2\right).$$

Thus, assuming that μ_0 defines a Gaussian probability density in phase space, we can use independent samples $z_1,\dots,z_N \in \mathbb{R}^{2d}$ of it to approximate the FBI formula by the Monte Carlo estimator

$$\psi_{0,N} := \frac{1}{N}\sum_{n=1}^{N} f_0(z_n) \in L^2(\mathbb{R}^d), \quad f_0(z) = r_0(z)g_z, \tag{3.42}$$

which converges almost surely to the expected value

$$\mathbb{E}(f_0) = \int_{\mathbb{R}^{2d}} r_0(z)g_z\,\mathrm{d}\mu_0(z) = (2\pi\varepsilon)^{-d}\int_{\mathbb{R}^{2d}}\langle g_z \mid g_{z_0}\rangle\; g_z\,\mathrm{d}z = g_{z_0}.$$

Remark 30. *We note that Markov chain Monte Carlo methods can be used to obtain samples $z_1,\dots,z_N \in \mathbb{R}^{2d}$ for the representation of non-Gaussian functions for which it is generally not known how to draw independent identically distributed samples.*

A measure of accuracy is the mean squared error, which is of order $\mathcal{O}(N^{-1})$ and does not depend on the dimension d. The following result was taken from [LL20, Theorem 8.4].

Lemma 31. *Let $\psi_0 = g_{z_0}$ and let $\psi_{0,N}$ be the Monte Carlo estimator defined in (3.42). Then, the mean squared error is given by*

$$\mathbb{E}\left(\|\psi_{0,N}-\psi_0\|^2\right) \le \frac{\mathbb{V}(f_0)}{N},$$

where the variance of f_0 is given by

$$\mathbb{V}(f_0) = \int_{\mathbb{R}^{2d}}\int_{\mathbb{R}^d}|r_0(z)g_z(x) - g_{z_0}(x)|^2\,\mathrm{d}x\,\mathrm{d}\mu(z) = 4^d - 1.$$

In particular, the error does not depend on the semiclassical parameter.

For the proof we refer to [LL20, Theorem 8.4] and [LS17, Examples 2]. □

Since classical Monte Carlo simulations based on random or pseudo-random numbers are faced with the problem that the samples are not uniformly distributed over the integration domain, Quasi-Monte Carlo quadrature can be used to speed up convergence. This approach uses so-called “low discrepancy sequences” and corresponding error estimates for the discretisation of the FBI formula can be found in [LL20, Theorem 8.3]. Furthermore, we refer to [LS17, Example 6], where the authors present numerical examples for the reconstruction of a Gaussian wave packet.

3.3 The Gaussian wave packet transform in other works

We have already mentioned several times that the representations that follow from the discretisation of the FBI formula have been used by other authors. The approximation of Gaussian wave packets based on uniform Riemann sums can be found in the TSTG method [KMB16], but unlike us, the authors work with non-normalised basis functions. More precisely, Kong *et al.* use the one-dimensional basis functions

$$\phi_{j,k}(x) = \phi(x - q_k)e^{ip_j(x-q_k)}, \quad \phi(x) = \frac{\sqrt{\Delta q}}{\sqrt{\pi}}\frac{\sigma}{2}\exp\left(-\frac{\sigma^2}{4}x^2\right), \quad x \in \mathbb{R}, \tag{3.43}$$

where $\sigma > 0$ is chosen such that $\sigma/2 = \Delta p$ in order to minimise the oscillations of the summation curve, which Kong *et al.* define in momentum space (we defined it in position space). For the basis functions in (3.43), the discretised FBI formula according to Theorem 17 then takes the equivalent form (cf. [KMB16, Equation 10])

$$\psi_0(x) \approx \frac{1}{\sqrt{2\pi}} \sum_{k=1}^{K} \sum_{j=1}^{J} \langle \phi_{j,k} \mid \psi_0 \rangle \, \phi_{j,k}(x).$$

Kong *et al.* do not use the FBI formula to derive their representation of wave packets, nor do they present an error analysis. However, they refer to the fast Gaussian wave packet transform introduced by Qian and Ying [QY10], who used compactly supported basis functions instead of simple Gaussian functions and prove a similar representation via frame theory. Let us therefore take a closer look at the definition of frames.

Remark 32. *To show that our error analysis for the discretisation of the FBI formula is directly related to approximations resulting from frame representations, we focus only on the description of the fast Gaussian wave packet transform by Qian and Ying. However, it is important to note that frame representations related to the Schrödinger equation have also been studied by other authors and we refer the interested reader to [BBCN17] and [CNR09] and the references given therein.*

3.3.1 Overcomplete sets and frames

Recall that discretisations of the FBI formula based on weighted point evaluations of the integrand can be written as (cf. (3.36))

$$\psi \approx \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi)\, g_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} w_{\mathbf{n}} \langle g_{\mathbf{n}} \mid \psi \rangle\, g_{\mathbf{n}},$$

which, at second glance, shows a certain similarity to a decomposition of the wave function $\psi \in L^2(\mathbb{R}^d)$ according to an orthonormal basis. Although the functions $g_{\mathbf{n}}$ obviously do not form an orthonormal basis (we have $\langle g_{\mathbf{n}} \mid g_{\mathbf{n}'}\rangle \neq 0$ for $\mathbf{n} \neq \mathbf{n}'$), we will see in a moment that with a suitable choice of the phase space grid we obtain an overcomplete set. This means that every function $\psi \in L^2(\mathbb{R}^d)$ can be represented as

$$\psi = \sum_{\mathbf{n}\in\Gamma} d_{\mathbf{n}}(\psi)\, g_{\mathbf{n}},$$

where the corresponding coefficients $d_{\mathbf{n}}(\psi) \in \mathbb{C}$ are no longer given by the weighted inner products $w_{\mathbf{n}}\langle g_{\mathbf{n}} \mid \psi\rangle$. Our investigations will show how the coefficients $d_{\mathbf{n}}(\psi)$ can be calculated using frames.

The following definition was taken from [Mal09, Definition 5.1].

Definition 33. *Let $\mathcal{H}$ be a Hilbert space over the complex numbers and Γ an index set that might be finite or infinite. The sequence $\{\phi_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ is a frame of $\mathcal{H}$, if there exist positive constants $B \geq A > 0$ such that*

$$\forall \psi \in \mathcal{H}, \qquad A\|\psi\|^2_{\mathcal{H}} \leq \sum_{\mathbf{n}\in\Gamma} |\langle \phi_{\mathbf{n}}, \psi\rangle_{\mathcal{H}}|^2 \leq B\|\psi\|^2_{\mathcal{H}}. \tag{3.44}$$

When $A = B$ the frame is said to be tight.

Remark 34. *Every frame spans the Hilbert space $\mathcal{H}$ and if the vectors $\phi_{\mathbf{n}}$ are linearly independent, the frame is called a "Riesz basis". Moreover, every orthonormal basis of the Hilbert space is a tight frame with $A = B = 1$. The concept of frames is of great importance in the field of signal processing and goes back to Duffin and Schaeffer [DS52]. For a general introduction to this topic, we refer to [Mal09]. We also note that frames have been generalised to the continuous domain, which allows, for example, an alternative description of different function spaces, see [FR05].*

If the frame condition is satisfied, then the operator

$$\Phi\colon \mathcal{H} \to \ell^2(\Gamma), (\Phi\psi)_{\mathbf{n}} := \langle \phi_{\mathbf{n}}, \psi\rangle_{\mathcal{H}} \quad \text{for all } n \in \Gamma,$$

is called a "frame analysis operator", where

$$\ell^2(\Gamma) := \Big\{ s \in \mathbb{C}^\Gamma \,:\, \sum_{\mathbf{n}\in\Gamma} |s_{\mathbf{n}}|^2 < \infty \Big\},$$

and the adjoint Φ^* is given by the synthesis operator

$$\Phi^*\colon \ell^2(\Gamma) \to \mathcal{H},\ \Phi^* s = \sum_{\mathbf{n}\in\Gamma} s_{\mathbf{n}} \phi_{\mathbf{n}}.$$

For the choice $\mathcal{H} = L^2(\mathbb{R}^d)$ and $\phi_{\mathbf{n}} = g_{\mathbf{n}}$ we see that the frame analysis operator Φ is equal to our analysis operator $\mathcal{A}_\Gamma$ defined in (3.37), and therefore the question arises under which conditions the sequence $\{g_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ is a frame. Without going into too much detail, it should be noted that it can be proven that the sequence $\{g_{\mathbf{n}}\}_{\mathbf{n}\in\mathbb{Z}^{2d}}$ is a frame for sufficiently dense uniform phase space grids, which is then called a "Gabor frame" or" Weyl–Heisenberg frame", see [Grö01, Chapter 5.2 and Chapter 6.5]. This reflects in particular the fact that the completeness of the basis set depends crucially on the density of the grid points. For example, in the one-dimensional case, it is known that $\{g_{j,k}\}_{j,k\in\mathbb{Z}}$ is overcomplete if and only if $\Delta q \Delta p < 2\pi\varepsilon$, see e.g. [MA01, Section III]. More precisely, the completeness of the basis set depends on the sampling density

$$D = \frac{2\pi\varepsilon}{\Delta q \Delta p}, \tag{3.45}$$

where $D < 1$ implies undercompleteness and $D > 1$ implies overcompleteness.

Remark 35. *We note that "completeness" means that $\langle g_{j,k} \mid \psi\rangle = 0$ for all $j, k \in \mathbb{Z}$ implies $\psi = 0$. The proof that we get a complete set for $D = 1$ can be found in [BGZ75], which generalises the results of Bargmann* et al. *[BBGK71] and Perelomov [Per71]. However, it should be noted that not every $\psi \in L^2(\mathbb{R})$ has an L^2-convergent expansion in terms of the Gaussian wave packets $g_{j,k}$, see [Fol89, Chapter 3.4].*

In the case of an orthonormal basis, one can use the analysis coefficients $(\Phi\psi)_{\mathbf{n}}$ to reconstruct any element ψ of the Hilbert space. The next proposition was taken from [Mal09, Theorem 5.4 and Theorem 5.5] and shows that frames have a similar property.

Proposition 36. *Let $\{\phi_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ be a frame with bounds $B \geq A > 0$. Then, the operator $\Phi^*\Phi$ is invertible and the sequence $\{\tilde{\phi}_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ defined by*

$$\tilde{\phi}_{\mathbf{n}} := (\Phi^*\Phi)^{-1} \phi_{\mathbf{n}} \quad \text{for all } \mathbf{n} \in \Gamma$$

is a frame, the dual frame, that can be used to reconstruct every $\psi \in \mathcal{H}$ as follows:

$$\psi = \sum_{\mathbf{n}\in\Gamma} \langle \phi_{\mathbf{n}}, \psi\rangle_{\mathcal{H}}\, \tilde{\phi}_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} \langle \tilde{\phi}_{\mathbf{n}}, \psi\rangle_{\mathcal{H}}\, \phi_{\mathbf{n}}$$

In particular, the dual frame satisfies

$$\forall \psi \in \mathcal{H}, \qquad \frac{1}{B}\|\psi\|_{\mathcal{H}}^2 \leq \sum_{\mathbf{n}\in\Gamma} |\langle \tilde{\phi}_{\mathbf{n}}, \psi\rangle_{\mathcal{H}}|^2 \leq \frac{1}{A}\|\psi\|_{\mathcal{H}}^2,$$

and if the frame is tight, then $\tilde{\phi}_{\mathbf{n}} = A^{-1}\phi_{\mathbf{n}}$.

For the proof we refer to [Mal09, Theorem 5.4 and Theorem 5.5]. □

Unless the frame is tight, we see that the computation of the dual frame requires the inversion of the operator $\Phi^*\Phi$, whose supremum and infimum of the spectrum are the same as of the Gram matrix $G = (\langle \phi_{\mathbf{n}}, \phi_{\mathbf{n}'} \rangle_{\mathcal{H}})_{\mathbf{n},\mathbf{n}' \in \ell^2(\Gamma)}$, see [Mal09, Theorem 5.1], which is known to become ill-conditioned if the basis functions have a large overlap. This problem has been studied extensively, see e.g. [FF15, Section 3], and several stabilisation algorithms have been proposed, see e.g. [FLF11, KLY19]. In addition, various algorithms have been proposed for the computation of the dual frame and the inversion of the Gram matrix, including, for example, the Richardson iteration (aka "frame algorithm", see [Dau92]), the acceleration methods proposed by Gröchenig, see [Grö93], as well as the iterative refinement method proposed by Andersson, see [MA01].

Remark 37. *The reciprocal of a real number $a \neq 0$ can be computed by Newton's method using that $1/a$ is the root of the function $f\colon \mathbb{R} \setminus \{0\} \to \mathbb{R}$, $f(x) = x^{-1} - a$. The corresponding iteration is given by*

$$x^{(n+1)} = x^{(n)} - \frac{f(x^{(n)})}{f'(x^{(n)})} = 2x^{(n)} - a\left(x^{(n)}\right)^2.$$

The same algorithm can be used to compute the inverse of a given invertible matrix G, which is known as Newton–Schulz iteration and reads

$$X^{(n+1)} = 2X^{(n)} - X^{(n)} G X^{(n)}.$$

In particular, it can easily be shown that this method converges quadratically if the initial datum is chosen such that $\|\operatorname{Id} - GX^{(0)}\| < 1$ for a given submultiplicative matrix norm. Using the Newton–Schulz iteration for the inversion of the Gram matrix, we get exactly the iterative refinement method introduced by Andersson, see [MA01, Equation 23], and it seems that this connection has remained undiscovered until now.

In the next step, we show that compactly supported basis functions that approximate a Gaussian profile form a frame whose dual can be explicitly specified. To transform our original Gaussian basis into a basis with compact support, we use bump windows.

3.3.2 Bump windows and windowed basis functions

As described in detail in Section 1.1, parts of the present section (Sec. 3.3.2) overlap to a large extent with

1. the joint publication "Fourier Series Windowed by a Bump Function" with C. Lasser appeared in *Journal of Fourier Analysis and Applications*, 26(4):65, 2020;

2. the joint preprint "The Gaussian Wave Packet Transform via Quadrature Rules" with C. Lasser submitted to *IMA Journal of Numerical Analysis* on 15/12/2021, e-print available at arXiv:2010.03478.

There seems to be no general definition of window functions, but most authors tend to think of a real function $w \neq 0$ that vanishes outside a certain interval. We now introduce C^s-bump windows by singling out two additional properties: On the one hand, we require that bump windows fall off smoothly at the boundary of their support, and on the other hand, to receive a faithful windowed shape of the original function, bump windows have to equal 1 on a closed subinterval of their support. The plots in Figure 3.8 show three different bump windows (left) and their action on $\psi(x) = x$ (right).

Let us summarise these properties in a definition:

Definition 38. *Let $\lambda > 0$ and $0 \leq \rho < \lambda$. For some $s \geq 1$ we say that the function $w_{\rho,\lambda} \in C_c^s(\mathbb{R})$ is a $\underline{C^s\text{-bump window}}$, if the following properties are satisfied:*

$$\begin{aligned}
&(1)\, 0 \leq w_{\rho,\lambda}(x) \leq 1, \textit{ for } x \in (-\lambda, \lambda)\\
&(2)\, w_{\rho,\lambda}(x) = 0, \textit{ for } x \in \mathbb{R} \setminus (-\lambda, \lambda)\\
&(3)\, w_{\rho,\lambda}(x) = 1, \textit{ for } x \in [-\rho, \rho]
\end{aligned}$$

If $\rho = 0$, we say that the bump is <u>degenerate</u>. Moreover, whenever $w_{\rho,\lambda} \in C_c^\infty(\mathbb{R})$, we say that $w_{\rho,\lambda}$ is a <u>smooth bump</u>.

Remark 39. *We note that the class of bump windows include the famous Hann and Tukey windows, see Section 4.3. In particular, compactly supported windows can obtain at most root exponential accuracy, see [Tad86], while smooth windows without compact support can be used for pointwise reconstructions of exponential accuracy. For examples of non-compactly supported windows we refer to the work of Boyd in [Boy96] and subsequent papers, who pioneered the concept of adaptive filters.*

Smooth bump windows typically occur when working with partitions of unity and have previously been used for data analysis of gravitational waves, see [DIS00, Equation 3.35] and [MRS10, Section 2 (Equation 7)]. An example is given by the even function

$$w_{\rho,\lambda}(x) = \begin{cases} 1 & \text{if } 0 \leq |x| \leq \rho, \\ \dfrac{1}{\exp\left(\frac{1}{\lambda - |x|} + \frac{1}{\rho - |x|}\right) + 1} & \text{if } \rho < |x| < \lambda, \\ 0 & \text{if } |x| \geq \lambda. \end{cases} \tag{3.46}$$

As we can see in the bottom two plots of Figure 3.8, the product of a non-degenerate bump $w_{\rho,\lambda}$ and a given function ψ produces a (smooth) windowed shape, matching with the original function ψ in $[-\rho, \rho]$ and tending to zero at the boundaries of $(-\lambda, \lambda)$.

For a given bump window $w_{\rho,\lambda}$ we now consider the windowed basis functions

$$g_{j,k}^w(x) := g_{j,k}(x) w_{\rho,\lambda}(x - q_k), \tag{3.47}$$

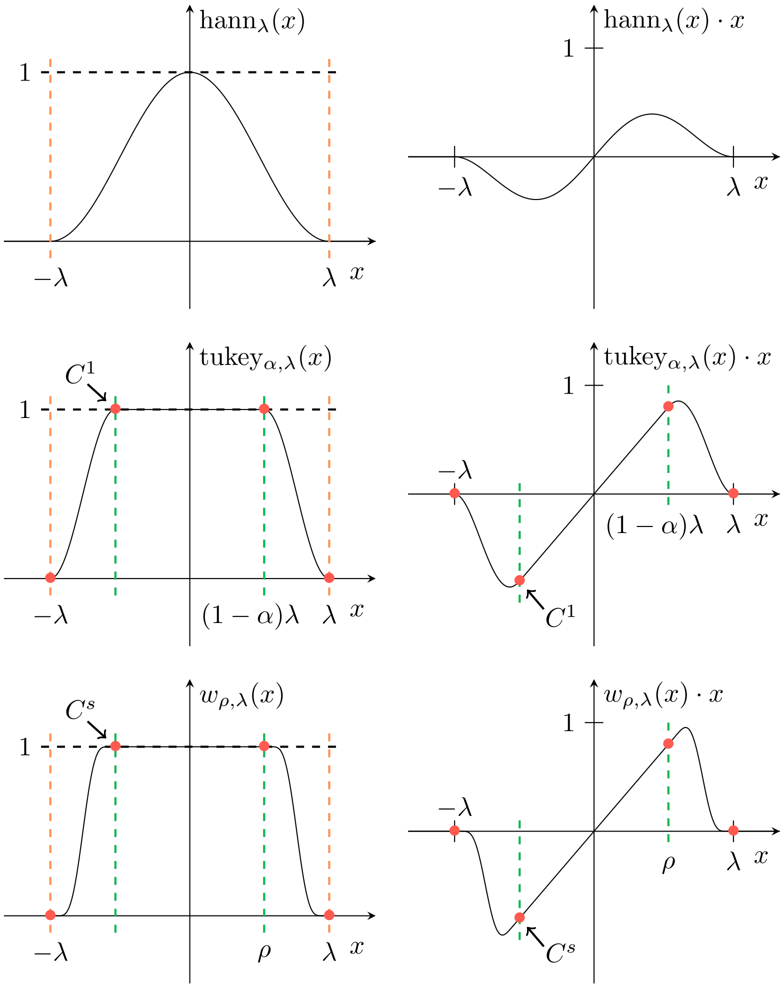


Figure 3.8: Three different bump windows (left) and their action on $\psi(x) = x$ (right). The Hann window (top) can be viewed as a degenerate C^1-bump, whereas for $0 < \alpha < 1$ the Tukey window (middle) is a non-degenerate C^1-bump. In general, the C^s-bump $w_{\rho,\lambda}$ (bottom) is s-times, but not $(s+1)$-times continuously differentiable.

which approximate our original Gaussian basis functions $g_{j,k}$ with centre q_k and are compactly supported in $[q_k - \lambda, q_k + \lambda]$. Furthermore, analogous to the definition of the Gaussian summation curve $S(x)$ in (3.1), we define the windowed summation curve

$$S^w(x) := \sum_{k\in\mathbb{Z}} |g_0^w(x - q_k)|^2, \quad x \in \mathbb{R}, \tag{3.48}$$

where $g_0^w := g_0 w_{\rho,\lambda}$. In particular, S^w is strictly positive if consecutive bump windows are sufficiently overlapping, that is, if the window parameter ρ satisfies

$$\rho \geq \frac{\Delta q}{2}. \tag{3.49}$$

Since on the one hand the original basis functions $g_{j,k}$ form an overcomplete set if the phase space density D defined in (3.45) satisfies $D > 1$, and on the other hand we have $\rho < \lambda$ by the definition of bump windows, we expect that for $\lambda = \pi\varepsilon/\Delta p$ the windowed basis functions $g_{j,k}^w$ also form an overcomplete set. The next lemma shows that we even get a frame. Furthermore, the lemma shows how the windowed summation curve can be used to construct a second frame.

Lemma 40. *Recall the definition of the windowed basis functions in* (3.47) *and the windowed summation curve in* (3.48) *for a bump that satisfies* (3.49) *and* $\lambda = \pi\varepsilon/\Delta p$. *Moreover, let* $Q^w := 1/S^w$. *Then, the following sequences are frames of* $L^2(\mathbb{R})$*:*

$$\left\{\frac{1}{\sqrt{2\lambda}} g_{j,k}^w\right\}_{j,k\in\mathbb{Z}} \quad \textit{and} \quad \left\{\frac{1}{\sqrt{2\lambda}} Q^w g_{j,k}^w\right\}_{j,k\in\mathbb{Z}} \tag{3.50}$$

The crucial ingredient for the proof is the fact that inner products with the windowed basis functions $g_{j,k}^w$ can be viewed as windowed Fourier coefficients and therefore the frame condition (3.44) follows via Parseval's equation from bounds on the windowed summation curve, see also [QY10, Lemma 3.1].

Proof. Like Qian and Ying, we only present the proof for the second family, because the proof for the first follows the same arguments. Let $k \in \mathbb{Z}$ and consider the functions

$$h_k(x) := Q^w(x) g_k^w(x), \quad \text{where } g_k^w := g_0^w(x - q_k), \quad x \in \mathbb{R}.$$

Moreover, for $\psi \in L^2(\mathbb{R})$ let us introduce the functions

$$\psi_k := \psi h_k,$$

as well as the windowed Fourier coefficients

$$\begin{aligned} c_{\psi_k}(j) &:= \frac{1}{2\lambda} \int_{q_k-\lambda}^{q_k+\lambda} \psi_k(x) e^{-\frac{i}{\varepsilon} p_j x} \, \mathrm{d}x \\ &= \frac{1}{2\lambda} \int_{-\infty}^{\infty} \psi(x) Q^w(x) g_k^w(x) \, e^{-\frac{i}{\varepsilon} p_j x} \, \mathrm{d}x \quad \text{for all } j \in \mathbb{Z}. \end{aligned}$$

To verify the frame condition (3.44), note that

$$\left|\left\langle \frac{1}{\sqrt{2\lambda}} Q^w g^w_{j,k} \mid \psi \right\rangle\right|^2 = \frac{1}{2\lambda}\left|\langle Q^w g^w_{j,k} \mid \psi\rangle\right|^2 = 2\lambda |c_{\psi_k}(j)|^2.$$

Hence, since Parseval's equation (see e.g. [Edw82, Chapter 8.2]) yields

$$\sum_{j\in\mathbb{Z}} |c_{\psi_k}(j)|^2 = \frac{1}{2\lambda}\int_{q_k-\lambda}^{q_k+\lambda} |\psi_k(x)|^2 \,\mathrm{d}x = \frac{1}{2\lambda}\int_{-\infty}^{\infty} |\psi(x)|^2 \,|h_k(x)|^2 \,\mathrm{d}x,$$

the monotone convergence theorem gives us

$$\begin{aligned}\sum_{k\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} \left|\left\langle \frac{1}{\sqrt{2\lambda}} Q^w g^w_{j,k} \mid \psi \right\rangle\right|^2 &= \sum_{k\in\mathbb{Z}} \int_{-\infty}^{\infty} |\psi(x)|^2 \,|h_k(x)|^2 \,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} |\psi(x)|^2 \sum_{k\in\mathbb{Z}} |h_k(x)|^2 \,\mathrm{d}x.\end{aligned}$$

Consequently, it suffices to prove the existence of positive numbers $B \geq A > 0$ such that

$$A < \sum_{k\in\mathbb{Z}} |h_k(x)|^2 < B \quad \text{for all } x \in \mathbb{R},$$

which then yields the frame condition

$$A\|\psi\|^2 < \sum_{j,k\in\mathbb{Z}} \left|\left\langle \frac{1}{\sqrt{2\lambda}} Q^w g^w_{j,k} \mid \psi \right\rangle\right|^2 < B\|\psi\|^2.$$

Since the windowed summation curve is Δq-periodic (cf. Lemma 13) and, due to the condition in (3.49), also strictly positive, $S^w(x)$ is bounded for all $x \in \mathbb{R}$, where the lower and upper bounds are given by

$$S_L := \min_{x\in[0,\Delta q]} S^w(x) \quad \text{and} \quad S_U := \max_{x\in[0,\Delta q]} S^w(x).$$

Consequently, the frame bounds A and B can be chosen as follows:

$$\begin{gathered}\sum_{k\in\mathbb{Z}} |h_k(x)|^2 = \sum_{k\in\mathbb{Z}} |Q^w(x)|^2 |g^w_k(x)|^2 > \frac{1}{S_U^2}\sum_{k\in\mathbb{Z}} |g^w_k(x)|^2 = \frac{S^w(x)}{S_U^2} > \frac{S_L}{S_U^2} =: A \quad \text{and} \\ \sum_{k\in\mathbb{Z}} |h_k(x)|^2 < \frac{S^w(x)}{S_L^2} < \frac{S_U}{S_L^2} =: B.\end{gathered}$$

□

Remark 41. *The above proof shows that the frame constants A and B can be expressed in terms of the bounds S_L and S_U of the windowed summation curve, and a comparison must be made with [Grö01, Theorem 6.4.1], which extends Lemma 40 to a larger class of compactly supported basis functions.*

We are now ready to formulate the wave packet representation used by Qian and Ying for the fast Gaussian wave packet transform. As shown in Proposition 36, the reconstruction of any square-integrable function ψ can be performed using dual frames. Given our results so far, it seems reasonable to expect that the sequences in (3.50) form dual frames. Indeed, we get the following result:

Proposition 42. *For any $\psi \in L^2(\mathbb{R})$, we have*

$$\psi = \sum_{j,k\in\mathbb{Z}} \left\langle \frac{1}{\sqrt{2\lambda}} g^w_{j,k} \mid \psi \right\rangle \frac{1}{\sqrt{2\lambda}} Q^w g^w_{j,k} = \frac{1}{S^w} \sum_{j,k\in\mathbb{Z}} r^w_{j,k} g^w_{j,k}, \tag{3.51}$$

where the complex-valued representation coefficients $r^w_{j,k} \in \mathbb{C}$ are given by

$$r^w_{j,k} = \frac{\Delta p}{2\pi\varepsilon} \langle g^w_{j,k} \mid \psi \rangle.$$

Recall that according to Theorem 17 (RS) a wave packet ψ can be approximated as

$$\psi \approx \frac{1}{S(x)} \sum_{j,k\in\mathbb{Z}} r^{(\mathrm{RS})}_{j,k} g_{j,k}, \quad \text{where} \quad r^{(\mathrm{RS})}_{j,k} = \frac{\Delta p}{2\pi\varepsilon} \langle g_{j,k} \mid \psi \rangle. \tag{3.52}$$

The relationship between (3.51) and (3.52) thus puts Theorem 17 in a new light, namely as an error representation for the reconstruction of wave packets based on dual frames. In addition to the proof presented by Qian and Ying, see [QY10, Lemma 3.2], we present a new proof based on windowed Fourier series.

Proof (of Proposition 42 via windowed Fourier series).
Let $\psi \in L^2(\mathbb{R})$. For $k \in \mathbb{Z}$ we introduce the windowed function

$$\psi^w_k := \psi \overline{g^w_k} \in L^2(\mathbb{R}).$$

In particular, ψ^w_k is compactly supported in $[q_k - \lambda, q_k + \lambda]$. We represent ψ^w_k almost everywhere in $[q_k - \lambda, q_k + \lambda]$ via its windowed Fourier series (cf. Proposition 49) as

$$\psi^w_k(x) = \sum_{j\in\mathbb{Z}} c^w_{\psi_k}(j) e^{\frac{i}{\varepsilon} p_j x},$$

where the grid points p_j are given by $p_j = j\pi\varepsilon/\lambda$ and the windowed Fourier coefficients $c^w_{\psi_k}(j)$ are given for all $j \in \mathbb{Z}$ by

$$c^w_{\psi_k}(j) = \frac{1}{2\lambda} \int_{q_k-\lambda}^{q_k+\lambda} \psi^w_k(x) e^{-\frac{i}{\varepsilon} p_j x} \,\mathrm{d}x = \frac{1}{2\lambda} \int_{-\infty}^{\infty} \psi(x) \overline{g^w_k(x)} e^{-\frac{i}{\varepsilon} p_j x} \,\mathrm{d}x.$$

In particular, for all $j, k \in \mathbb{Z}$ we conclude that

$$c^w_{\psi_k}(j) e^{\frac{i}{\varepsilon} p_j q_k} = \frac{1}{2\lambda} \int_{-\infty}^{\infty} \psi(x)\, \overline{g^w_{j,k}(x)} \,\mathrm{d}x = \frac{\Delta p}{2\pi\varepsilon} \langle g^w_{j,k} \mid \psi \rangle = r^w_{j,k},$$

which yields that

$$c^w_{\psi_k}(j)e^{\frac{i}{\varepsilon}p_jx} = r^w_{j,k}e^{\frac{i}{\varepsilon}p_j(x-q_k)} \quad \text{for all } x \in \mathbb{R}. \tag{3.53}$$

Consequently, using that g^w_k is compactly supported, we obtain (almost everywhere)

$$\begin{aligned} \psi(x) = \psi(x)\left(\frac{1}{S^w(x)}\sum_{k\in\mathbb{Z}}|g^w_k(x)|^2\right) &= \frac{1}{S^w(x)}\sum_{k\in\mathbb{Z}}\psi^w_k(x)g^w_k(x) \\ = \frac{1}{S^w(x)}\sum_{k\in\mathbb{Z}}\left(\sum_{j\in\mathbb{Z}}c^w_{\psi_k}(j)e^{\frac{i}{\varepsilon}p_jx}\right)g^w_k(x) &= \frac{1}{S^w(x)}\sum_{j,k\in\mathbb{Z}}r^w_{j,k}g^w_{j,k}(x). \end{aligned}$$

□

The proof of Proposition 42 shows that the coefficients $r^w_{j,k}$ are given by

$$r^w_{j,k} = c^w_{\psi_k}(j)e^{\frac{i}{\varepsilon}p_jq_k}, \quad j,k \in \mathbb{Z},$$

where $c^w_{\psi_k}(j)$ is the jth Fourier coefficients of the windowed function ψ^w_k. Consequently, using the fast Fourier transform (FFT) for the computation of the coefficients $r^w_{j,k}$ and its fast inverse (IFFT) for the synthesis, we obtain a fast algorithm for the reconstruction of arbitrary wave functions (not necessarily Gaussians). This explains why Qian and Ying use the prefix “fast”. In particular, a detailed description of how the FFT can be used to compute windowed Fourier coefficients can be found in Appendix 7.3.

Remark 43. *Qian and Ying use the fast computation of the representation coefficients for a reinitialisation algorithm and introduce a Gaussian beam method to solve the time-dependent Schrödinger equation, see [QY10, Algorithm 4.1]. Although the FFT was used for the first time in this context, we would like to point out that the fast computation of frame coefficients was already used before, see e.g. [Mal09, Chapter 5.4.1].*

Decay of the representation coefficients

For the original Gaussian basis functions $g_{j,k}$ we have already seen that the norms of the representation coefficients decrease exponentially, see Lemma 16, and therefore we expect the coefficients to decrease very fast also for compactly supported basis functions. Indeed, Qian and Ying write ([QY10, Page 7857]): *“For a typical initial function [...], most of the coefficients [...] have small norms.”*, but give no rigorous explanation for this statement. Before we close this chapter, let us justify this statement in more detail: Since it follows that

$$|r^w_{j,k}| = |c^w_{\psi_k}(j)| \quad \text{for all } j,k \in \mathbb{Z},$$

the decay of $r^w_{j,k}$ with respect to j is given by the decay rate of the Fourier coefficients of the compactly supported function ψ^w_k, which is known to be at most root exponential, cf. Remark 39. On the other hand, we can prove spectral convergence for the decay rate of $|r^w_{j,k}|$ with respect to k, provided that ψ is a Schwartz function and the bump is even:

Lemma 44. *Let $\psi \in \mathcal{S}(\mathbb{R})$ and assume that the bump window $w_{\rho,\lambda}$ is an even function. Then, for all $s \geq 1$, there exists a positive constant $C_s > 0$, depending on $s, \psi, w_{\rho,\lambda}$ and the width γ of the basis functions, such that, uniformly in $j \in \mathbb{Z}$,*

$$|r^w_{j,k}| \leq \frac{C_s}{|k|^{s+1}} \quad \textit{for all } k \neq 0.$$

Proof. The relation in (3.53) yields for all $j, k \in \mathbb{Z}$:

$$|r^w_{j,k}| \leq \frac{1}{2\lambda} \int_{-\infty}^{\infty} |\psi(x)\, g^w_0(x - q_k)| \, \mathrm{d}x.$$

Using that the window is an even function, we conclude that

$$|r^w_{j,k}| \leq \frac{\|g_0\|_\infty}{2\lambda} \int_{-\infty}^{\infty} |\psi(x)| w_{\rho,\lambda}(q_k - x) \, \mathrm{d}x = \frac{(\pi\varepsilon)^{-1/4}\gamma_i^{1/4}}{2\lambda} \left(|\psi| * w_{\rho,\lambda}\right)(q_k),$$

and by [Fol99, Proposition 8.11] it follows that $|\psi| * w_{\rho,\lambda}$ is a real and non-negative Schwartz function. Consequently, there exists a positive constant $C(s, \psi, w_{\rho,\lambda}) > 0$ such that, for all $s \geq 1$ and all $x \in \mathbb{R} \setminus \{0\}$, we have

$$\left(|\psi| * w_{\rho,\lambda}\right)(x) \leq \frac{C(s, \psi, w_{\rho,\lambda})}{|x|^{s+1}}.$$

Hence, the claim follows for the constant

$$C_s = C(s, \psi, w_{\rho,\lambda}) \frac{(\pi\varepsilon)^{-1/4}\gamma_i^{1/4}}{2\lambda^{s+2}},$$

where we used that $q_k = k\lambda$. □

Our analysis shows that the representation coefficients $r^w_{j,k}$ decrease rapidly, but we do not obtain a Gaussian decay. However, depending on the choice of bump window, the estimate in (2.10) can be used for practical applications, because it is not possible to distinguish Gaussian functions from windowed Gaussian functions due to machine precision (provided that the window parameter ρ is sufficiently large).

3.4 Summary of this chapter

Discretisations of the FBI formula lead to discrete representations of wave functions and the separation of position and momentum space can be used to apply efficient rules that go beyond simple Riemann sums. If the separation is accomplished via the Gaussian summation curve in position space, we obtain a rescaled superposition of Gaussian wave packets, where each basis function is rescaled by $S(x)$. In particular, since for sufficiently dense grids the summation curve can be approximated by a constant depending on the grid size, the resulting approximations lead to pure Gaussian superpositions and correspond to a direct discretisation of the phase space integral. The rescaled basis functions $g_{j,k}(x)/S(x)$ can be understood as an approximation to the dual frame formed by the basis functions. Furthermore, for the representation of Gaussian wave packets, the representation coefficients resulting from the discretisation of the FBI transform can be calculated analytically, without high-dimensional numerical integration or the inversion of overlap matrices. Bump windows can be used to construct a basis of compactly supported wave packets approximating a Gaussian profile and the corresponding representation coefficients are then given by windowed Fourier coefficients. The variants (3.2) and (3.3) of the discrete Gaussian wave packet transform based on uniform grids in both position and momentum space can be derived in three closely related ways:

1. via the discretisation of the phase space integral in the FBI formula.
2. via the decomposition of the wave packet using the frame that is formed by the (windowed) Gaussian basis functions.
3. via the approximation of the wave packet by windowed Fourier series.

4 Excursus: Windowed Fourier series

As described in detail in Section 1.1, parts of the present chapter overlap to a large extent with the joint publication “Fourier Series Windowed by a Bump Function” with C. Lasser appeared in *Journal of Fourier Analysis and Applications*, 26(4):65, 2020.

In this chapter we study windowed Fourier transforms and transfer Jackson's classical results on the convergence of Fourier series to windowed series of not necessarily periodic functions. We start by recalling basic properties of Fourier series for functions with bounded variation in Section 4.1. Afterwards, in Section 4.2 we present windowed transforms and estimate the corresponding reconstruction errors. In Section 4.3 the results are then applied to bump windows, which we have already used in Section 3.3.2 for the construction of compactly supported basis functions. Finally, in Section 4.4 we present numerical experiments that underline our theoretical results and illustrate the advantages of bump windows.

The theory of Fourier series plays an essential role in numerous applications of contemporary mathematics. It allows us to represent a periodic function in terms of complex exponentials. Indeed, any square-integrable function $f\colon \mathbb{R} \to \mathbb{C}$ of period 2π has a norm-convergent Fourier series such that (see e.g. [BN71, Proposition 4.2.3])

$$f(x) = \sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{ikx} \quad \text{almost everywhere,}$$

where the Fourier coefficients are defined according to

$$\widehat{f}(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} \, \mathrm{d}x, \quad k \in \mathbb{Z}.$$

By the classical results of Jackson in 1930, see [Jac94], the decay rate of the Fourier coefficients and therefore the convergence speed of the Fourier series depend on the regularity of the function. If f has a jump discontinuity, then the order of magnitude of the coefficients is $\mathcal{O}(1/|k|)$, as $|k| \to \infty$. Moreover, if f is a smooth function of period 2π, say $f \in C^{s+1}(\mathbb{R})$ for some $s \geq 1$, then the order improves to $\mathcal{O}(1/|k|^{s+1})$.

In the following we focus on the reconstruction of a not necessarily periodic function with respect to a finite interval $(-\lambda, \lambda)$. For this purpose let us think of a smooth, non-periodic real function $\psi\colon \mathbb{R} \to \mathbb{R}$, which we want to represent by a Fourier series. Therefore, we will examine its 2λ-periodic extension, see Figure 4.1. As we can see,

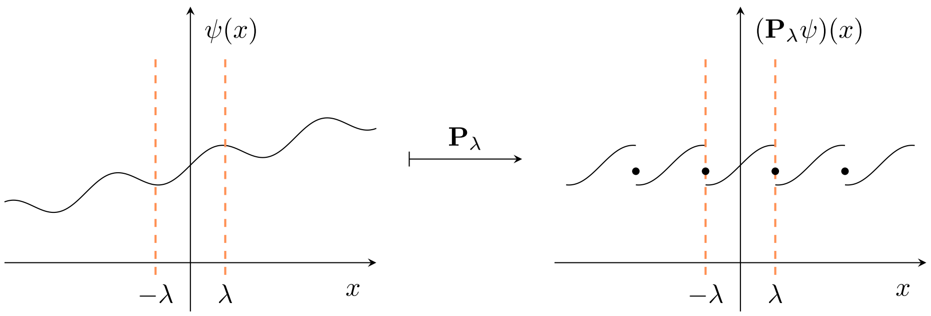


Figure 4.1: Effect of the periodisation: If $\psi(-\lambda^+) \neq \psi(\lambda^-)$, then the 2λ-periodic extension produces jump discontinuities at $\pm\lambda$. Consequently, the order of the Fourier coefficients is $\mathcal{O}(1/|k|)$.

whenever $\psi(-\lambda^+) \neq \psi(\lambda^-)$, the periodisation has a jump discontinuity at λ, and thus the Fourier coefficients are $\mathcal{O}(1/|k|)$. An easy way to eliminate these discontinuities at the boundary, is to multiply the original function by a smooth window, compactly supported in $[-\lambda, \lambda]$. The resulting periodisation has no jumps. Consequently, one expects faster convergence of the windowed Fourier sums. Therefore, we investigate the convergence speed of Fourier series windowed by compactly supported bump functions with a plateau. The properties of these bump windows will allow an effortless transfer of Jackson's classical results on the convergence of the Fourier series for smooth functions.

4.1 Functions of bounded variation and their Fourier series

Let us start by recalling basic properties of the Fourier series for functions of bounded variation. We denote by $\mathrm{BV}_{\mathrm{loc}}$ the set of functions $f\colon \mathbb{R} \to \mathbb{R}$, which are locally of bounded variation, that is of bounded variation on every finite interval. In particular, we assume that such functions are normalised for any x in the interior of the interval of definition, see [BN71, Chapter 0.6], by

$$f(x) = \frac{1}{2}\Big(f(x^+) + f(x^-)\Big) = \frac{1}{2}\left(\lim_{t\to 0+} f(x+t) + \lim_{t\to 0+} f(x-t)\right).$$

We recall that a function of bounded variation is bounded, has at most a countable set of jump discontinuities, and that the pointwise evaluation is well-defined.

4.1.1 The classical Fourier representation

Recall that any 2π-periodic function $f \in \mathrm{BV}_{\mathrm{loc}}$ has a pointwise converging Fourier series, see e.g. [BN71, Proposition 4.1.5]. Let us transfer this representation to an arbitrary interval of length 2λ:

Lemma 45. *Suppose that $\psi \in \mathrm{BV}_{\mathrm{loc}}$ as well as $\lambda > 0$ and $t \in \mathbb{R}$. Then,*

$$\psi(x) = \sum_{k\in\mathbb{Z}} c_\psi(k) e^{ik\frac{\pi}{\lambda}x}, \quad x \in (t-\lambda, t+\lambda),$$

where the coefficients $c_\psi(k)$ are given by

$$c_\psi(k) := \frac{1}{2\lambda}\int_{t-\lambda}^{t+\lambda} \psi(x) e^{-ik\frac{\pi}{\lambda}x}\,\mathrm{d}x, \quad k \in \mathbb{Z}.$$

For the proof of Lemma 45 and our subsequent analysis, we will use a translation, a scaling and a periodisation operator. For the centre $t \in \mathbb{R}$ and a scaling factor $a > 0$, we introduce:

$$\mathbf{T}_t\colon\ \mathrm{BV}_{\mathrm{loc}} \to \mathrm{BV}_{\mathrm{loc}},\ (\mathbf{T}_t\psi)(x) := \psi(x+t),$$
$$\mathbf{S}_a\colon\ \mathrm{BV}_{\mathrm{loc}} \to \mathrm{BV}_{\mathrm{loc}},\ (\mathbf{S}_a\psi)(x) := \psi(ax).$$

For the period half length $\lambda > 0$, we set

$$\mathbf{P}_\lambda\colon\ \mathrm{BV}_{\mathrm{loc}} \to \mathrm{BV}_{\mathrm{loc}},$$
$$(\mathbf{P}_\lambda\psi)(x) := \begin{cases} \psi(x), & \text{if } x \in (-\lambda,\lambda), \\ \frac{1}{2}\Big(\psi(-\lambda^+) + \psi(\lambda^-)\Big), & \text{if } x = \lambda. \end{cases}$$

Proof. Consider the 2π-periodic function $f = \mathbf{P}_\pi \mathbf{S}_{\lambda/\pi}\mathbf{T}_t\psi$. Then, it follows from Lemma 48 that $f \in \mathrm{BV}_{\mathrm{loc}}$ and therefore

$$f(x) = \sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{ikx}, \quad x \in \mathbb{R}.$$

The Fourier coefficients of f are given by

$$\begin{aligned}\widehat{f}(k) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi}\big(\mathbf{S}_{\lambda/\pi}\mathbf{T}_t\psi\big)(x)e^{-ikx}\,\mathrm{d}x \\ &= \frac{1}{2\lambda}\int_{t-\lambda}^{t+\lambda}\psi(x)e^{-ik\frac{\pi}{\lambda}(x-t)}\,\mathrm{d}x.\end{aligned}$$

Consequently, for all $x \in (t-\lambda, t+\lambda)$ we obtain

$$\psi(x) = \big(\mathbf{T}_{-t}\mathbf{S}_{\pi/\lambda}f\big)(x) = \sum_{k\in\mathbb{Z}} c_\psi(k)e^{ik\frac{\pi}{\lambda}x}.$$

□

4.1.2 The classical result of Jackson

In general, even if ψ is a smooth function, the periodic extension $f = \mathbf{P}_\pi \mathbf{S}_{\lambda/\pi} \mathbf{T}_t \psi \in \mathrm{BV}_{\mathrm{loc}}$ has jump discontinuities at $\pm\pi$. Let $V(f) < \infty$ denote the total variation of f. Then, the decay of the Fourier coefficients can be bounded (see e.g. [Edw82, Chapter 2.3.6]) as follows:

$$|k \cdot c_\psi(k)| = |k \cdot \widehat{f}(k)| \leq \frac{1}{2\pi} V(f), \quad \text{for all } k \in \mathbb{Z}.$$

Hence, the coefficients are $\mathcal{O}(1/|k|)$. As we will see in a moment, the rate of the coefficients transfers to an estimate for the reconstruction errors. For an arbitrary function $f \in \mathrm{BV}_{\mathrm{loc}}$ of period 2π let us introduce the partial Fourier sum

$$S_n f(x) := \sum_{k=-n}^{n} \widehat{f}(k) e^{ikx}, \quad n \geq 1,\, x \in \mathbb{R}.$$

Our analysis relies on the following classical result by Jackson on the convergence of the Fourier sum, see [Jac94, Chapter II.3 (Theorem IV)]:

Proposition 46. *If $f\colon \mathbb{R} \to \mathbb{R}$ is a function of period 2π, which has a sth derivative with limited variation, $s \geq 1$, and if V is the total variation of $f^{(s)}$ over a period, then, for $n > 0$,*

$$|f(x) - S_n f(x)| \leq \frac{2V}{s\pi n^s}, \quad x \in \mathbb{R}.$$

4.2 The windowed transform

In Section 3.3.2 we have introduced bump windows. Since the following results on windowed Fourier series hold for a larger class, let us define general window functions.

Definition 47. *Let $\lambda > 0$. We say that a function $w \in \mathrm{BV}_{\mathrm{loc}}$ is a window function on the interval $(-\lambda, \lambda)$, if the following properties are satisfied:*

$$(1)\, 0 \leq w(x) \leq 1, \textit{ for } x \in (-\lambda, \lambda)$$
$$(2)\, w(x) = 0, \textit{ for } x \in \mathbb{R} \setminus (-\lambda, \lambda)$$

In particular, we obtain the rectangular window, if $w(x) = 1$ for all $x \in (-\lambda, \lambda)$, and for simplicity we just write $w \equiv 1$ in this case. For $\psi \in \mathrm{BV}_{\mathrm{loc}}$, a fixed parameter $t \in \mathbb{R}$ and a window w on $(-\lambda, \lambda)$, we introduce the windowed periodisation

$$\psi_w := \mathbf{P}_\pi \mathbf{S}_{\lambda/\pi} \left[w \cdot \mathbf{T}_t \psi \right].$$

Note that ψ_w is 2π-periodic. Moreover, the next lemma shows that $\psi_w \in \mathrm{BV}_{\mathrm{loc}}$.

Lemma 48. *Let w be a window on $(-\lambda, \lambda)$. Moreover, let $\psi \in \mathrm{BV}_{\mathrm{loc}}$ and $t \in \mathbb{R}$. Then, the windowed periodisation ψ_w is of bounded variation.*

Proof. Since the window w is of bounded variation, it is a bounded function and so it is easy to see that $w \cdot \mathbf{T}_t\psi$ is also of bounded variation. Furthermore, since the variation of $w \cdot \mathbf{T}_t\psi$ on a finite interval $[a, b]$ is equal to the variation of $\mathbf{S}_{\lambda/\pi}\,[w \cdot \mathbf{T}_t\psi]$ on $[a\pi/\lambda, b\pi/\lambda]$ it suffices to show that the periodisation operator maps $\mathrm{BV}_{\mathrm{loc}}$ to $\mathrm{BV}_{\mathrm{loc}}$. For a function $f\colon [a, b] \to \mathbb{R}$ and a partition P of some finite interval $[a, b]$ we denote by $V(f, P)$ the variation of f with respect to P, and by $V(f)$ the total variation of f on $[a, b]$. Now, for $\psi \in \mathrm{BV}_{\mathrm{loc}}$ and $\lambda > 0$ consider $f := \mathbf{P}_\lambda\psi$. It remains to show that $V(f|_{[-\lambda,\lambda]})$ is a finite number. Therefore, let

$$P = \{-\lambda = x_0, x_1, \dots, x_{k-1}, x_k = \lambda\}$$

be a partition of $[-\lambda, \lambda]$. Then,

$$\begin{aligned}
V(f, P) &= \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \\
&\leq \sum_{i=1}^{k} |\psi(x_i) - \psi(x_{i-1})| + |\psi(-\lambda) - f(-\lambda)| + |f(\lambda) - \psi(\lambda)| \\
&= V(\psi, P) + |\psi(-\lambda) - f(-\lambda)| + |f(\lambda) - \psi(\lambda)|.
\end{aligned}$$

Thus, taking the supremum among such partitions, we conclude that

$$V(f|_{[-\lambda,\lambda]}) = V(\psi|_{[-\lambda,\lambda]}) + |\psi(-\lambda) - f(-\lambda)| + |f(\lambda) - \psi(\lambda)| < \infty.$$

□

4.2.1 The windowed representation

According to the classical Fourier series of the periodisation presented in Lemma 45, the windowed series allows an alternative representation with potentially faster convergence.

Proposition 49. *Let $\psi \in \mathrm{BV}_{\mathrm{loc}}$ and $\lambda > 0$ and $t \in \mathbb{R}$. If $w \in \mathrm{BV}_{\mathrm{loc}}$ is a window on $(-\lambda, \lambda)$, then,*

$$\psi(x)w(x - t) = \sum_{k\in\mathbb{Z}} c^w_\psi(k) e^{ik\frac{\pi}{\lambda}x}, \quad x \in (t - \lambda, t + \lambda),$$

where the coefficients $c^w_\psi(k)$ are given by

$$c^w_\psi(k) := \frac{1}{2\lambda} \int_{t-\lambda}^{t+\lambda} \psi(x)w(x - t)e^{-ik\frac{\pi}{\lambda}x}\,\mathrm{d}x, \quad k \in \mathbb{Z}.$$

The statement in the above Proposition follows as in Lemma 45, but this time for the Fourier series of the 2π-periodic windowed shape $\psi_w \in \mathrm{BV}_{\mathrm{loc}}$.

Suppose that $\psi_w \in C^s(\mathbb{R})$, $s \geq 1$, and that $\psi_w^{(s)}$ has bounded variation. Then, as it follows from [Jac94, Chapter II.3 (Corollary I)],

$$|c_\psi^w(k)| = |\widehat{\psi}_w(k)| \leq \frac{V(\psi_w^{(s)})}{\pi |k|^{s+1}}, \quad k \neq 0,$$

and thus the decay rate of the windowed coefficients c_ψ^w improves to $\mathcal{O}\left(1/|k|^{s+1}\right)$.

4.2.2 An error estimate for the representations

For $n \geq 1$ and $x \in \mathbb{R}$ let

$$R_n^w \psi(x) := \sum_{k=-n}^{n} c_\psi^w(k) e^{ik\frac{\pi}{\lambda}x} \quad \text{and} \quad R_n \psi(x) := R_n^{w\equiv 1}\psi(x) = \sum_{k=-n}^{n} c_\psi(k) e^{ik\frac{\pi}{\lambda}x}.$$

Note that $R_n^w \psi = \mathbf{T}_{-t}\mathbf{S}_{\pi/\lambda}(S_n \psi_w)$. We now transfer Jackson's classical result in Proposition 46 to an estimate for the windowed reconstruction errors in terms of the Lipschitz constant of $\psi_w^{(s)}$. In order not to overload the notation unnecessarily for the presentation of the main results, we assume that $\lambda = \pi$ and $t = 0$, that is, both the function ψ and ψ_w are 2π-periodic and centred at the origin. However, all results could also be formulated for an arbitrary choice of $\lambda > 0$ and $t \in \mathbb{R}$ by performing an appropriate scaling and translation.

Theorem 50 (Reconstruction, windowed series, $\lambda = \pi$ and $t = 0$)**.**
Suppose that $\psi_w \in C^{s+1}(\mathbb{R})$, $s \geq 1$ and let $L_s > 0$ denote the Lipschitz constant of $\psi_w^{(s)}$ over $[-\pi, \pi]$. Moreover, let $0 < \rho < \pi$. Then, for $n \geq 1$ the error of the reconstruction $R_n^w \psi$ in the interval $[-\rho, \rho]$ is given by

$$\left| \sup_{x \in [-\rho,\rho]} |\psi(x) - R_n^w \psi(x)| - K_\infty(\psi, w, \rho) \right| \leq \frac{4L_s}{sn^s}, \tag{4.1}$$

where the non-negative constant $K_\infty(\psi, w, \rho) \geq 0$ is given by

$$K_\infty(\psi, w, \rho) = \sup_{x \in [-\rho,\rho]} \Big(|\psi(x)| \big(1 - w(x)\big) \Big).$$

Proof. Let $V < \infty$ denote the total variation of $\psi_w^{(s)}$ over a period. In particular,

$$V = \int_{-\pi}^{\pi} |\psi_w^{(s+1)}(x)| \, \mathrm{d}x \leq 2\pi L_s.$$

Hence, for all $x \in \mathbb{R}$ the classical Jackson result in Proposition 46 yields

$$A_n(x) := |\psi_w(x) - R_n^w \psi_w(x)| = |\psi_w(x) - S_n \psi_w(x)| \leq \frac{4L_s}{sn^s}.$$

Moreover, for all $x \in [-\rho, \rho]$ we have $0 \leq w(x) \leq 1$ and thus, by the reverse triangle inequality, we obtain

$$|\psi(x) - R_n^w\psi(x)| \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \Big| |\psi(x)| \big(1 - w(x)\big) \begin{Bmatrix} + \\ - \end{Bmatrix} A_n(x) \Big|, \qquad x \in [-\rho, \rho]. \tag{4.2}$$

Taking the supremum proves (4.1). □

Note that for $w \equiv 1$ we obtain the convergence of the plain reconstruction $R_n\psi$, where $K_\infty(\psi, w, \rho) = 0$. Theorem 50 allows a calculation of the L^2-error:

Corollary 51. *The L^2-error of the reconstruction is given by*

$$\Big| \|\psi - R_n^w\psi\|^2_{L^2([-\rho,\rho])} - K_2(\psi, w, \rho) \Big| \leq \frac{16\rho L_s}{sn^s} K_\infty(\psi, w, \rho) + \frac{32\rho L_s^2}{s^2 n^{2s}}, \tag{4.3}$$

where the non-negative constant $K_2(\psi, w, \rho) \geq 0$ is given by

$$K_2(\psi, w, \rho) = \int_{-\rho}^{\rho} |\psi(x)|^2 \big(1 - w(x)\big)^2 \, \mathrm{d}x.$$

In particular, $K_2(\psi, w, \rho) = 0$, if and only if $K_\infty(\psi, w, \rho) = 0$.

Proof. For $p \in \{1, 2\}$ we introduce $N_{p,n,\rho} := \|\psi_w - R_n^w\psi\|_{L^p([-\rho,\rho])}$. Then, it follows from (4.2) that for all $x \in [-\rho, \rho]$:

$$|\psi(x) - R_n^w\psi(x)|^2 \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \Big| |\psi(x)| \big(1 - w(x)\big) \begin{Bmatrix} + \\ - \end{Bmatrix} A_n(x) \Big|^2,$$

and therefore, integration yields

$$\|\psi - R_n^w\psi\|^2_{L^2([-\rho,\rho])} \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} K_2(\psi, w, \rho) \begin{Bmatrix} + \\ - \end{Bmatrix} 2K_\infty(\psi, w, \rho) N_{1,n,\rho} + N^2_{2,n,\rho}.$$

Consequently, (4.3) follows from

$$N_{1,n,\rho} \leq \sqrt{2\rho} \cdot N_{2,n,\rho} \leq 2\rho \left(\sup_{x \in [-\rho,\rho]} A_n(x) \right) \leq \frac{8\rho L_s}{sn^s}.$$

□

In addition to the assumptions in Theorem 50, let us assume that $w(x) = 1$ for all $x \in [-\rho, \rho]$. Then, it follows that $K_\infty(\psi, w, \rho) = 0$ and therefore $K_2(\psi, w, \rho) = 0$. Hence, the reconstruction errors converge to 0 as $n \to \infty$. This motivates the investigation of bump windows.

4.3 Bump windows revisited

Recall the definition of bump windows in Section 3.3.2. A famous member of this class is the Hann window, which can be defined as follows, see [Mal09, Chapter 4.2.2]:

Definition 52. *Let $\lambda > 0$. For all $x \in \mathbb{R}$ the* <u>Hann window</u> *is given by*

$$\mathrm{hann}_\lambda(x) = \cos^2\left(\frac{\pi}{2\lambda}x\right) \cdot \mathbf{1}_{[0,\lambda]}(|x|) = \frac{1}{2}\left[1 + \cos\left(\frac{\pi}{\lambda}x\right)\right] \cdot \mathbf{1}_{[0,\lambda]}(|x|).$$

In the sense of Definition 38, the Hann window is a degenerate C^1-bump. In particular, for $0 < \rho' < \lambda$ it follows from Theorem 50 and Corollary 51, that the reconstruction errors for a function $\psi \neq 0$ on the interval $[t - \rho', t + \rho']$ are bounded from below by positive constants $K_\infty(\psi, w, \rho'), K_2(\psi, w, \rho') > 0$. This fact can also be observed in our numerical experiments, see Section 4.4.1 and Section 4.4.2.

The Hann window is a famous representative of windows specially used in signal processing. As it turns out, the Hann window arises as a special candidate of a more general class, the Tukey windows, see [Tuk67], often called "cosine-tapered windows". These windows can be seen as a cosine lobe convolved with a rectangular window:

Definition 53. *The* <u>Tukey window</u> *with parameter $\alpha \in (0, 1]$ is given by*

$$\mathrm{tukey}_{\alpha,\lambda}(x) := \mathbf{1}_{[0,(1-\alpha)\lambda)}(|x|) + \frac{1}{2}\left[1 - \cos\left(\frac{\pi|x|}{\alpha\lambda} - \frac{\pi}{\alpha}\right)\right] \cdot \mathbf{1}_{[(1-\alpha)\lambda,\lambda]}(|x|).$$

The Tukey window is a C^1-bump $w_{\rho,\lambda}$ with $\rho = (1 - \alpha)\lambda$. In particular,

$$\mathrm{tukey}_{1,\lambda} = \mathrm{hann}_\lambda = w_{0,\lambda},$$

and for $0 < \alpha < 1$ the Tukey window is not degenerate. We note that the sum of phase-shifted Hann windows creates a Tukey window:

Lemma 54. *Let $\tau > 0$ and $m \geq 0$. Then, for $\alpha = 1/(m + 1)$ and $\lambda = (m + 1)\tau$,*

$$\sum_{k=-m}^{m} \mathrm{hann}_\tau(\bullet - k\tau) = \mathrm{tukey}_{\alpha,\lambda}\,.$$

Proof. For all $x \in \mathbb{R}$ we introduce the function

$$H_{\tau,m}(x) := \sum_{k=-m}^{m} \mathrm{hann}_\tau(x - k\tau).$$

Obviously, $H_{\tau,m}$ is an even function. Thus, for all $x \in \mathbb{R}$ we obtain

$$H_{\tau,m}(x) = \sum_{k=0}^{m} \mathrm{hann}_\tau(|x| - k\tau) = \frac{1}{2}\sum_{k=0}^{m}\left[1 + \cos\left(\frac{\pi}{\tau}(|x| - k\tau)\right)\right] \cdot \mathbf{1}_{[-\tau,\tau]}(|x| - k\tau)$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{k=0}^{m}\left[1+\cos\left(\frac{\pi}{\tau}(|x|-k\tau)\right)\right]\left(\mathbf{1}_{[(k-1)\tau,k\tau)}(|x|)+\mathbf{1}_{[k\tau,(k+1)\tau)}(|x|)\right)\\
&= \frac{1}{2}\sum_{k=0}^{m}\left[1+\cos\left(\frac{\pi}{\tau}(|x|-k\tau)\right)\right]\cdot\mathbf{1}_{[(k-1)\tau,k\tau)}(|x|)\\
&\quad+\frac{1}{2}\sum_{k=1}^{m+1}\left[1+\cos\left(\frac{\pi}{\tau}(|x|-(k-1)\tau)\right)\right]\cdot\mathbf{1}_{[(k-1)\tau,k\tau)}(|x|)\\
&= \mathbf{1}_{[0,m\tau)}(|x|)+\frac{1}{2}\left[1-\cos\left(\frac{\pi}{\tau}(|x|-(m+1)\tau)\right)\right]\cdot\mathbf{1}_{[m\tau,(m+1)\tau)}(|x|)\\
&\quad+\frac{1}{2}\sum_{k=1}^{m}\left(\cos\left(\frac{\pi}{\tau}(|x|-k\tau)\right)+\cos\left(\frac{\pi}{\tau}(|x|-k\tau)+\pi\right)\right)\cdot\mathbf{1}_{[(k-1)\tau,k\tau)}(|x|)\\
&= \mathbf{1}_{[0,m\tau)}(|x|)+\frac{1}{2}\left[1-\cos\left(\frac{\pi}{\tau}(|x|-(m+1)\tau)\right)\right]\cdot\mathbf{1}_{[m\tau,(m+1)\tau)}(|x|).
\end{aligned}
$$

□

4.3.1 The representation for bump windows

The windowed Fourier series in Proposition 49 applies to bump functions and yields the following representation in the restricted interval $[t-\rho, t+\rho]$:

Corollary 55. *Suppose that $\psi \in C^{s+1}(\mathbb{R})$, $s \geq 1$, as well as $\lambda > 0$ and $0 \leq \rho < \lambda$ and $t \in \mathbb{R}$. If $w_{\rho,\lambda} \in C_c^{s+1}(\mathbb{R})$ is a C^{s+1}-bump on $(-\lambda, \lambda)$, satisfying the three conditions in Definition 38, then,*

$$\psi(x) = \sum_{k\in\mathbb{Z}} c_\psi^w(k) e^{ik\frac{\pi}{\lambda}x}, \quad x \in [t-\rho, t+\rho],$$

where the coefficients $c_\psi^w(k)$ are given by

$$c_\psi^w(k) = \frac{1}{2\lambda}\int_{t-\lambda}^{t+\lambda} \psi(x) w_{\rho,\lambda}(x-t) e^{-ik\frac{\pi}{\lambda}x}\,\mathrm{d}x, \quad k \in \mathbb{Z}.$$

In particular, if $L_s > 0$ denotes the Lipschitz constant of $\psi_w^{(s)}$ over $[-\pi,\pi]$, then,

$$|c_\psi^w(k)| \leq \frac{V(\psi_w^{(s)})}{\pi|k|^{s+1}} \leq \frac{2L_s}{|k|^{s+1}}, \quad k \neq 0. \tag{4.4}$$

We note that for $w = \mathrm{hann}_\lambda$ the representation in Corollary 55 shrinks to a pointwise representation at $x = t$. Furthermore, the bound in (4.4) depends on the choice of the bump $w_{\rho,\lambda}$, and for $\rho \approx \lambda$ the windowed transform does not lead to an improvement of the decay for low frequencies k, because in this case the action of the bump is comparable to a truncation of ψ, such that the Lipschitz constant L_s dominates. We will illustrate this fact with numerical experiments in Section 4.4.2. Moreover, we note that for a smooth bump $w_{\rho,\lambda} \in C_c^\infty(\mathbb{R})$ the coefficients $c_\psi^w(k)$ do not decay exponentially fast, since the window is compactly supported and thus not analytic, see [Tad86]. Nevertheless, the coefficients of a smooth bump have an exponential rate of fractional order and the actual rate can be classified by analysing their so-called "Gevrey regularity", see [Tad07, Equation 2.4].

Remark 56. *In [Boy06] a smooth bump is designed such that the order of the windowed Fourier coefficients is root-exponential (at least for the saw wave function), wheres in [Tan06] we find a non-compactly supported window, for which we obtain true exponential decay. We note that Boyd and Tanner focus on an optimal choice of window parameters in order to obtain the best possible approximation results.*

4.3.2 A bound for the Lipschitz constant

We now investigate the Lipschitz constant L_s in Corollary 55. Using the work of Ore in [Ore38], we crucially use an estimate on the higher-order derivatives of the product of two functions, which is developed in Section 4.3.3.

For a function $f\colon \mathbb{R} \to \mathbb{R}$, that is $(s+1)$-times differentiable, $s \geq 1$, with a $(s+1)$th derivative bounded on a finite interval (a,b), let us introduce the non-negative constant

$$C_{s,f} = \sup_{x\in(a,b)} |f(x)| + \frac{(b-a)^{s+1}}{(s+1)!} \sup_{x\in(a,b)} |f^{(s+1)}(x)| \geq 0. \tag{4.5}$$

Theorem 57 (Bound for the Lipschitz constant, $\lambda = \pi$ and $t = 0$)**.**
Let $0 \leq \rho < \pi$ and suppose that $\psi \in C^{s+1}(\mathbb{R})$ and $w_{\rho,\pi} \in C_c^{s+1}(\mathbb{R})$ for some $s \geq 1$. Assume the existence of two non-negative constants $M_\psi, M_{\psi_{s+1}} \geq 0$, such that

$$|\psi(x)| \leq M_\psi \quad \textit{and} \quad |\psi^{(s+1)}(x)| \leq M_{\psi_{s+1}} \quad \textit{for all } x \in (-\pi,\pi).$$

Then, the Lipschitz constant L_s in Corollary 55 is bounded by

$$L_s \leq M_{\psi_{s+1}} + M_\psi \|w_{\rho,\pi}^{(s+1)}\|_\infty + \frac{C_{s,\psi} C_{s,w}}{(2\pi)^{s+1}} \cdot K_s,$$

where the non-negative constants $C_{s,\psi}, C_{s,w} \geq 0$ are given by

$$C_{s,\psi} = M_\psi + \frac{(2\pi)^{s+1}}{(s+1)!} M_{\psi_{s+1}} \quad \textit{and} \quad C_{s,w} = 1 + \frac{(2\pi)^{s+1}}{(s+1)!} \|w_{\rho,\pi}^{(s+1)}\|_\infty,$$

and the constant $K_s > 0$ is given by

$$K_s = \frac{2^{2s+1} s^2 (3s)!}{(2s+1)!}. \tag{4.6}$$

Proof. According to Proposition 59 in the next section, we use the bound for the $(s+1)$th derivative of the product fg for $f = w_{\rho,\pi}$ and $g = \psi$. This results in

$$L_s = \sup_{x\in(-\pi,\pi)} \left| \frac{\mathrm{d}^{s+1}}{\mathrm{d}x^{s+1}} \Big(w_{\rho,\pi}(x)\psi(x) \Big) \right| \leq M_{\psi_{s+1}} + M_\psi \| w_{\rho,\pi}^{(s+1)} \|_\infty + \frac{C_{s,\psi} C_{s,w}}{(2\pi)^{s+1}} \cdot K_s.$$

Moreover, for the formula of the constant K_s in (4.6) we refer to Lemma 61. □

Remark 58. *Stirling's formula yields the following approximation of K_s:*

$$K_s = \frac{2s}{2s+1} \frac{2^{2s} s (3s)!}{(2s)!} \sim \frac{4^s s \sqrt{6\pi s} (3s)^{3s} e^{-3s}}{\sqrt{4\pi s} (2s)^{2s} e^{-2s}} = s \sqrt{\frac{3}{2}} \left(\frac{27s}{e} \right)^s.$$

The sign $\sim$ means that the ratio of the quantities tends to 1 as $s \to \infty$.

In [GT85, Lemma 3.2], Gottlieb and Tadmor present a bound for the largest maximum norm of a windowed Dirichlet kernel (regularisation kernel) and its first s derivatives. This bound is used to derive an error estimate for the reconstruction of a function by a discretisation of the convolution integral with an appropriate trapezoidal sum (cf. [GT85, Proposition 4.1]). Instead of working with the largest maximum norm of the first s derivatives, we are now presenting a new bound for the $(s+1)$th derivative of a product of two functions. We therefore combine the Leibniz product rule with individual bounds for intermediate derivatives, and to the best of my knowledge, this is the first time that an explicit bound has been revealed this way.

4.3.3 Estimating higher-order derivatives of a product

If f is $(s+1)$-times differentiable, and if its $(s+1)$th derivative is bounded on a finite interval (a,b), then, it follows from [Ore38, Theorem 2] that all intermediate derivatives are bounded. In particular, for all $i = 1, \dots, s$ and all $x \in (a,b)$,

$$|f^{(i)}(x)| \leq K(i,s) \cdot \frac{C_{s,f}}{(b-a)^i}, \tag{4.7}$$

where the combinatorial constant $K(i,s) > 0$ is defined according to

$$K(i,s) = \frac{2^i \cdot s^2(s^2-1^2)\cdots(s^2-(i-1)^2)}{1\cdot 3\cdot 5\cdots(2i-1)}, \qquad i \in \{1,\dots,s\}. \tag{4.8}$$

We now use the general Leibniz rule to lift this result to an explicit bound for the $(s+1)$th derivative of the product of two functions.

Proposition 59. *Let $s \geq 1$ and $f, g\colon \mathbb{R} \to \mathbb{R}$, both $(s+1)$-times differentiable in a finite interval (a,b). Assume the existence of four non-negative constants*

$$M_f, M_g, M_{f_{s+1}}, M_{g_{s+1}} \geq 0,$$

such that for all $x \in (a,b)$:

$$|f(x)| \leq M_f, \ |g(x)| \leq M_g \quad \textit{and} \quad |f^{(s+1)}(x)| \leq M_{f_{s+1}}, \ |g^{(s+1)}(x)| \leq M_{g_{s+1}}.$$

Then, for all $x \in (a,b)$ we have

$$|(fg)^{(s+1)}(x)| \leq M_f M_{g_{s+1}} + M_{f_{s+1}} M_g + \frac{C_{s,f} C_{s,g}}{(b-a)^{s+1}} \cdot K_s,$$

where the constants $C_{s,f}, C_{s,g} \geq 0$ are defined according to (4.5) and the constant $K_s > 0$, which only depends on s, is given by

$$K_s = \sum_{k=1}^{s} \binom{s+1}{k} K(s+1-k,s) K(k,s). \tag{4.9}$$

Proof. By the general Leibniz rule the $(s+1)$th derivative of fg is given by

$$(fg)^{(s+1)} = \sum_{k=0}^{s+1} \binom{s+1}{k} f^{(s+1-k)} g^{(k)}.$$

We therefore obtain the following estimate for all $x \in (a,b)$:

$$\begin{aligned} |(fg)^{(s+1)}(x)| &\leq \sum_{k=0}^{s+1} \binom{s+1}{k} |f^{(s+1-k)}(x)||g^{(k)}(x)| \\ &\leq M_f M_{g_{s+1}} + M_{f_{s+1}} M_g + \sum_{k=1}^{s} \binom{s+1}{k} |f^{(s+1-k)}(x)||g^{(k)}(x)|. \end{aligned}$$

Using (4.7) for $1 \leq k \leq s$, we conclude that

$$\begin{aligned} |f^{(s+1-k)}(x)| &\leq K(s+1-k,s) \cdot \frac{C_{s,f}}{(b-a)^{s+1-k}}, \\ |g^{(k)}(x)| &\leq K(k,s) \cdot \frac{C_{s,g}}{(b-a)^{k}}, \end{aligned}$$

and thus

$$|(fg)^{(s+1)}(x)| \leq M_f M_{g_{s+1}} + M_{f_{s+1}} M_g + \frac{C_{s,f} C_{s,g}}{(b-a)^{s+1}} \cdot K_s.$$

□

Remark 60. *The bound*

$$|f^{(i)}(x)| \leq K(i,s) \cdot \frac{M_f}{(b-a)^i}, \quad x \in (a,b),$$

for a polynomial f of degree s is due to W. Markoff (1916) and it is known that the equality sign is attained for the Chebyshev polynomials, see [MG16].

4.3.4 The combinatorial constant

Next, we will investigate the combinatorial constant K_s and derive formula (4.6) presented in Theorem 57.

Lemma 61. *Let $s \geq 1$. The combinatorial constant $K_s > 0$ in (4.9) satisfies*

$$K_s = \frac{2^{2s+1} s^2 (3s)!}{(2s+1)!}. \tag{4.10}$$

Proof. We start by rewriting the constant $K(i,s)$ that has been defined in (4.8). Let $i \in \{1, \dots, s\}$. For the numerator we obtain

$$2^i \cdot s^2 \cdot (s^2 - 1^2) \cdots (s^2 - (i-1)^2) = 2^i \cdot \frac{s}{s+i} \cdot \frac{(s+i)!}{(s-i)!}.$$

For the denominator we have

$$1 \cdot 3 \cdot 5 \cdots (2i-1) = \frac{(2i-1)!}{2^{i-1}(i-1)!} = \frac{(2i)!}{2^i i!}.$$

Hence, we can rewrite $K(i,s)$ as

$$K(i,s) = 2^i \cdot \frac{s}{s+i} \cdot \frac{(s+i)!}{(s-i)!} \cdot \frac{2^i \cdot i!}{(2i)!} = \frac{s}{s+i} \cdot 2^{2i} \cdot i! \cdot \binom{s+i}{2i},$$

and the summands that define the number K_s in (4.9) can be expressed as

$$\begin{aligned} &\binom{s+1}{k} K(s+1-k,s) \cdot K(k,s) \\ &\quad = 2^{2s} \cdot (s+1)! \cdot (2s)^2 \cdot \frac{(s+k-1)! \cdot (2s-k)!}{(2k)! \cdot (s-k)! \cdot (2s-2k+2)! \cdot (k-1)!}. \end{aligned}$$

Therefore we conclude that

$$\begin{aligned} K_s &= 2^{2s} \cdot (s+1)! \cdot \sum_{k=0}^{s-1} \frac{(s+k)! \cdot (2s-k-1)! \cdot (2s)^2}{(2k+2)! \cdot (s-k-1)! \cdot (2s-2k)! \cdot k!} \qquad (4.11) \\ &= 2^{2s} \cdot (s+1)! \cdot \sum_{k=0}^{s-1} \left[\binom{2s-k}{k} \frac{2s}{2s-k} \cdot \binom{s+k+1}{s-k-1} \frac{2s}{s+k+1} \right]. \end{aligned}$$

Finally, let us introduce

$$\kappa_j(2s) = \binom{2s-j}{j} \frac{2s}{2s-j} \quad \text{for} \quad j = 0, 1, \dots, s-1.$$

Recognising our constant K_s as a Vandermonde-type convolution and using the representation in [Gou56, Equation 4], we write

$$K_s = 2^{2s}(s+1)! \sum_{k=0}^{s-1} \kappa_k(2s) \kappa_{s-k-1}(2s) = 2^{2s}(s+1)! \kappa_{s-1}(4s)$$

$$= 2^{2s}(s+1)!\binom{3s+1}{s-1}\frac{4s}{3s+1} = \frac{2^{2s+1}s^2(3s)!}{(2s+1)!}.$$

□

In Appendix 7.4 we derive an upper bound for K_s based on binomial coefficients.

4.4 Numerical experiments

According to our results in Theorem 50 and Corollary 55 we present numerical experiments for three different functions. We investigate reconstructions with the smooth bump $w_{\rho,\lambda}$ given by (3.46), compared to those with the Hann window in Definition 52 and the Tukey window in Definition 53. Besides the reconstructions we also present the decay of the coefficients and the reconstruction errors.
In Section 4.4.1 we start with the saw wave function to demonstrate the superiority of the windowed transform with a smooth bump for a function having a high jump discontinuity. Afterwards, the experiments in Section 4.4.2 deal with a parabola function. The symmetric periodic extension has no discontinuities, and therefore the parabola is a good candidate to illustrate the limitations of bump windows. Last, in Section 4.4.3 we work with a rapidly decreasing function. As we will see in this example, for low frequencies all coefficients (plain, tukey, bump) have a rapid initial decrease, implying excellent reconstructions.

Remark 62. *In the following experiments, the dependency of the windows on the parameters λ, ρ and α are always assumed implicitly and therefore we write*

$$\mathrm{hann} = \mathrm{hann}_\lambda, \quad \mathrm{tukey} = \mathrm{tukey}_{\alpha,\lambda}, \quad \mathrm{bump} = w_{\rho,\lambda}.$$

For the numerical computation of the (windowed) coefficients we used the fast Fourier transform, see Appendix 7.3.

4.4.1 Saw wave function

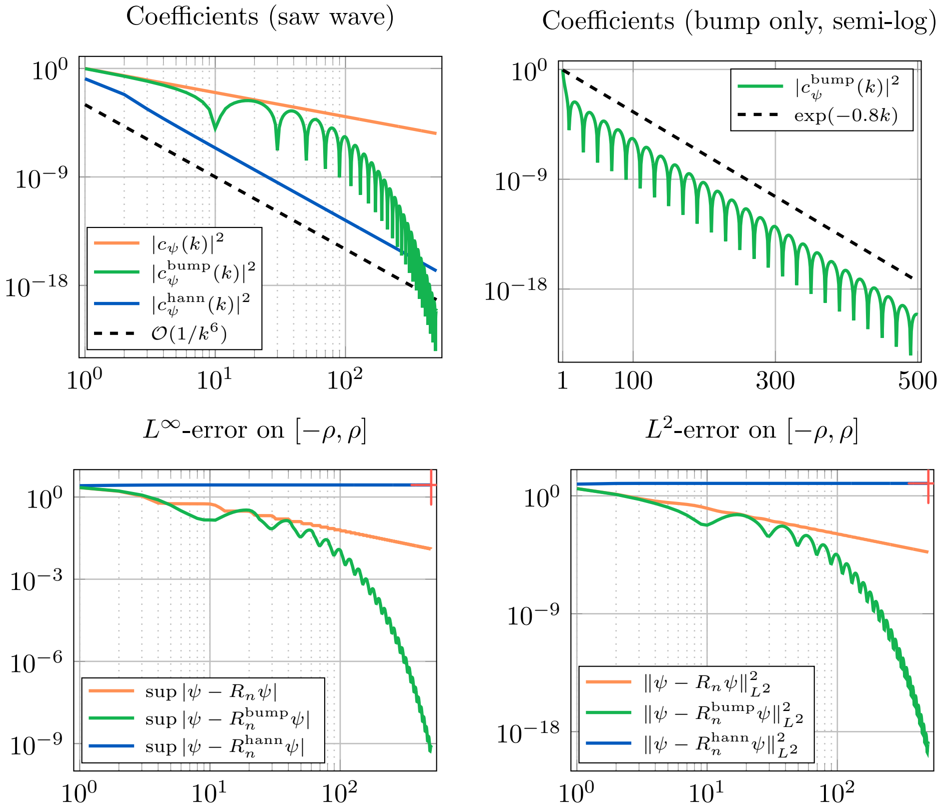
In the first example we consider the function

$$\psi(x) = x, \quad \Big[\lambda = \pi,\ \rho = 0.9\pi,\ t = 0\Big].$$

The corresponding periodic extension $\mathbf{P}_\lambda\psi$ results in a saw wave function.

We note that $c_\psi(k)$ and $c_\psi^{\mathrm{hann}}(k)$ can be evaluated analytically and are given by

$$c_\psi(k) = i\cdot\frac{(-1)^k}{k},\ k\in\mathbb{Z}\setminus\{0\}, \quad c_\psi^{\mathrm{hann}}(k) = -i\cdot\frac{(-1)^k}{2k(k^2-1)},\ k\in\mathbb{Z}\setminus\{-1,0,1\}$$



Coefficients (saw wave)

Coefficients (bump only, semi-log)

L^∞-error on $[-\rho, \rho]$

L^2-error on $[-\rho, \rho]$

Figure 4.2: Decay of the coefficients (above) and reconstruction errors (below) for the saw wave. The plain coefficients (orange) have order $\mathcal{O}(1/|k|)$, while the coefficients for the bump (green) show exponential decay (upper right side). For the Hann window the errors converge to constant values larger than 1 (red crosses).

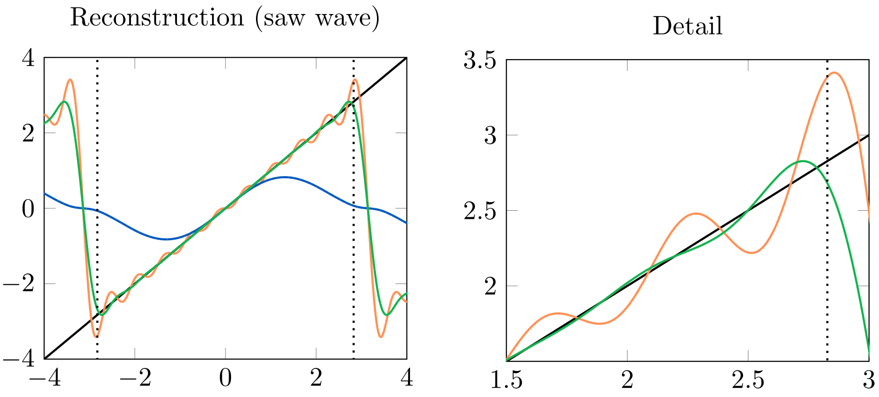


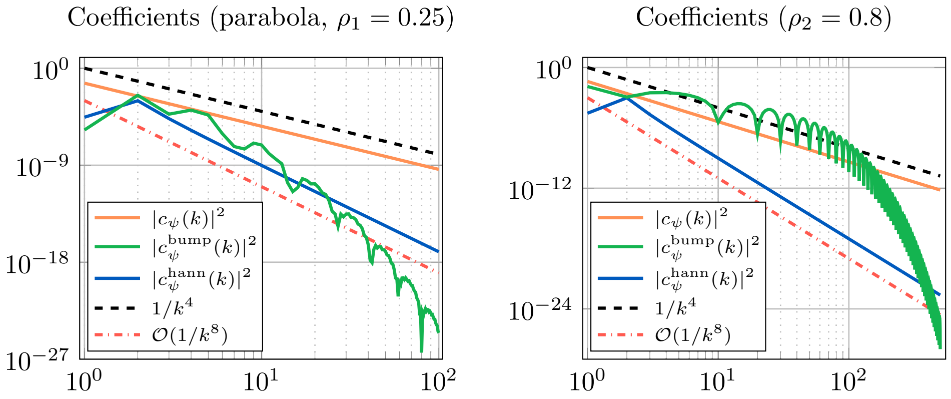
Figure 4.3: Plot of the reconstructions $R_{10}\psi$ and $R_{10}^{w}\psi$ for the saw wave. For $x \in [-\rho, \rho]$ (dotted lines) the bump-windowed reconstruction (green) matches well with the original function and the typical overshoots (Gibbs phenomenon) of the Fourier sum (orange) are dampened. The reconstruction with the Hann window (blue) is accurate only in a small neighborhood of 0.

and $c_\psi(0) = c_\psi^{\text{hann}}(0) = 0, c_\psi^{\text{hann}}(-1) = -c_\psi^{\text{hann}}(1) = 3i/8$. Moreover, since ψ is a real function, we conclude that

$$c_\psi^w(-k) = \overline{c_\psi^w(k)}, \quad k \in \mathbb{Z}.$$

The upper left-hand side of Figure 4.2 shows $|c_\psi(k)|^2 = 1/k^2$, as well as $|c_\psi^w(k)|^2$ for both windows (hann and bump). We observe that the windowed coefficients have a faster asymptotic decay than the plain Fourier coefficients. The coefficients and the reconstruction errors for the bump (green) show the best asymptotic decay. As we observe in the upper right plot of Figure 4.2, the bump-windowed coefficients show exponential initial decay. In particular, we recognise a trembling for these coefficients, while the other (plain and hann) have a smooth decay. We provide an explanation of this phenomenon in Appendix 7.5. The reconstructions $R_{10}\psi$ and $R_{10}^{w}\psi$ are visualised in Figure 4.3. For the bump we recognise a good convergence to the original function ψ in $[-\rho, \rho]$ (dotted lines), and the typical overshoots of the Fourier sum at the discontinuity (Gibbs phenomenon, see e.g. [Tad07, Section 3]) are dampened. As expected, the reconstruction with the Hann window is accurate only in a small neighborhood of the centre $t = 0$, and according to Theorem 50 and Corollary 51 the reconstruction errors converge to $K_\infty(\psi, w, \rho), K_2(\psi, w, \rho) > 0$. For the saw wave these constants can be calculated analytically in terms of λ and ρ, and their values are given by $K_\infty \approx 8.91$ and $K_2 \approx 2.76$. We have marked these values with red crosses and observe a perfect match.

Remark 63. *As we have discussed in Section 4.3.1, the coefficients of the bump do not fall exponentially fast for all k, since the bump is not analytic. However, in [Boy06] the author presents a smooth bump that is based on the erf-function, such that the Fourier coefficients for the saw wave fall exponentially fast (exponential of the square root of k). This is achieved by an optimisation of the corresponding window parameters. In view of the bump used here, this relates to an optimal choice of ρ.*


Coefficients (parabola, $\rho_1 = 0.25$)

Coefficients ($\rho_2 = 0.8$)

Figure 4.4: Decay of the representation coefficients for the parabola with $\rho_1 = 0.25$ (left) and $\rho_2 = 0.8$ (right). Again, the coefficients for the bump show a fast asymptotic decay.

4.4.2 Parabola

We consider the symmetric function

$$\psi(x) = x^2, \qquad \Big[\lambda = 1,\ \rho_1 = 0.25,\ \rho_2 = 0.8,\ t = 0\Big].$$

Note that

$$c_\psi(k) = \frac{2(-1)^k}{k^2\pi^2},\ k \in \mathbb{Z} \setminus \{0\}, \qquad c_\psi^{\mathrm{hann}}(k) = \frac{(-1)^k(1-3k^2)}{k^2(k^2-1)^2\pi^2},\ k \in \mathbb{Z} \setminus \{-1, 0, 1\},$$

as well as

$$c_\psi(0) = \frac{1}{3}, \quad c_\psi^{\mathrm{hann}}(0) = \frac{1}{6} - \frac{1}{\pi^2}, \quad c_\psi^{\mathrm{hann}}(-1) = c_\psi^{\mathrm{hann}}(1) = \frac{1}{12} - \frac{7}{8\pi^2}.$$

The plots in Figure 4.4 show the decay of the coefficients. Especially for low frequencies, the coefficients for the Hann window show the fastest decay. Nevertheless, we observe once more that the bump coefficients and errors have the best asymptotics, see Figure 4.5. As for the saw wave, the constants $K_\infty(\psi, w, \rho)$ and $K_2(\psi, w, \rho)$ can be calculated analytically and are given by

$$K_\infty \approx \begin{cases} 9.1 \cdot 10^{-3}, & \text{if } \rho = 0.25, \\ 0.58, & \text{if } \rho = 0.8, \end{cases} \quad \text{and} \quad K_2 \approx \begin{cases} 4.7 \cdot 10^{-6}, & \text{if } \rho = 0.25, \\ 0.075, & \text{if } \rho = 0.8. \end{cases}$$

We have marked these values with red crosses and verify the predicted convergence of the errors. The reconstructions $R_{50}(\psi)$ and $R_{50}^w(\psi)$ are visualised in Figure 4.6. For the first choice $\rho_1 = 0.25$ (left) the bump-windowed series approximates the original function only in the small interval $[-\rho_1, \rho_1] = [-0.25, 0.25]$. We note that the periodic extension of the parabola has no discontinuities and therefore the plain reconstruction

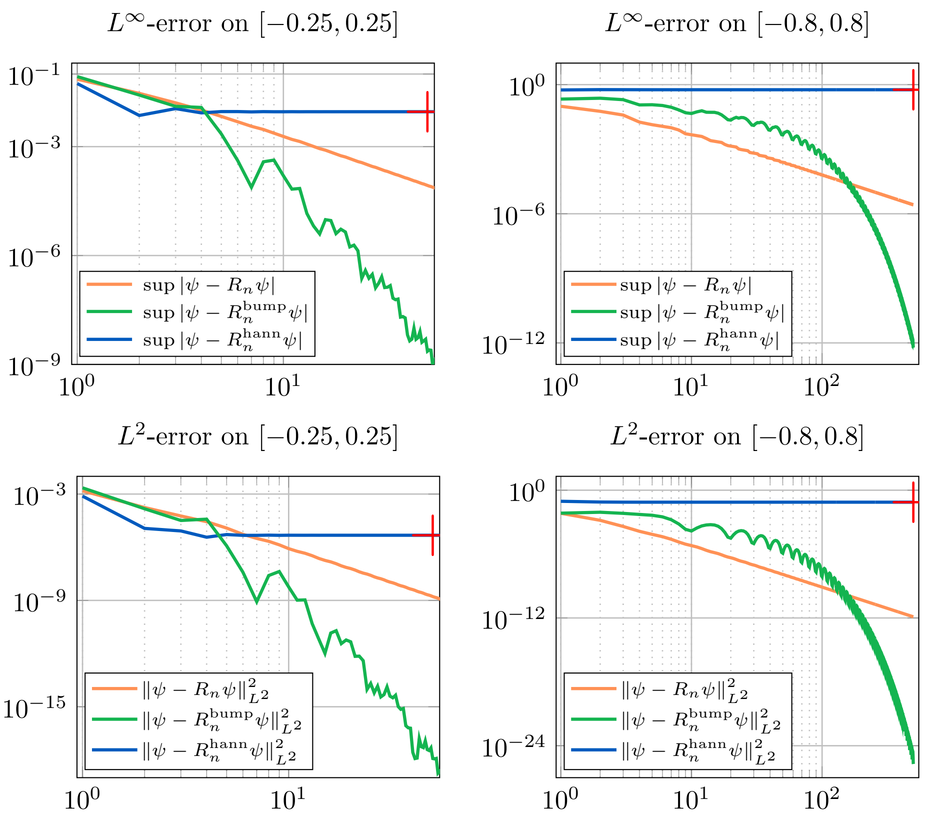


Figure 4.5: Reconstruction errors for the parabola with $\rho_1 = 0.25$ (left) and $\rho_2 = 0.8$ (right). For the second choice the smooth bump has a large derivative in the interval $(0.8, 1)$, implying a large Lipschitz constant L_s. Consequently, for low frequencies the errors are worse than for the plain coefficients.

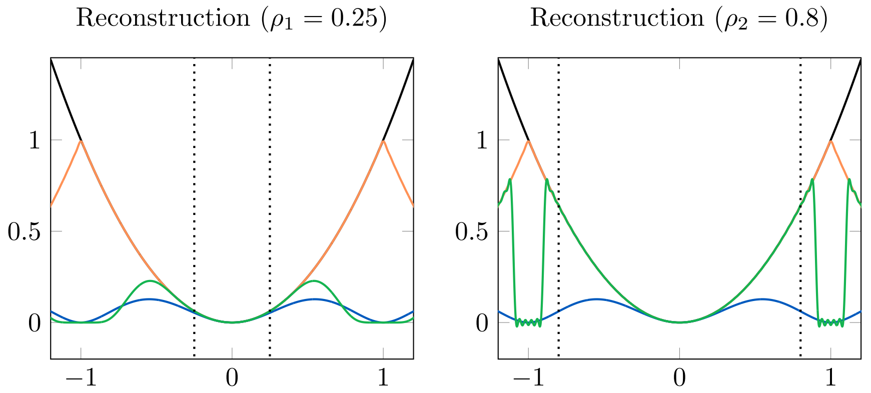


Figure 4.6: Reconstructions of the parabola. For $\rho_2 = 0.8$ (right) the bump-windowed shape (green) has a large derivative in $(0.8, 1)$, implying a slow decay of the windowed coefficients. For $\rho_1 = 0.25$ (left) the coefficients fall off much faster, but the reconstruction is faithful only in a small interval, comparable to the Hann window.

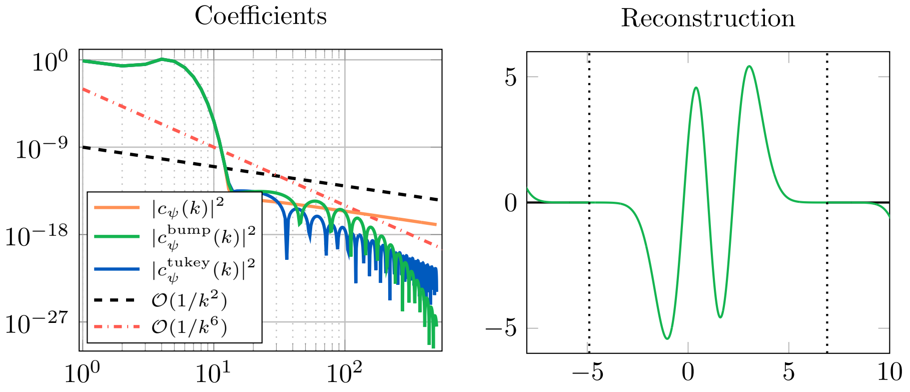


Figure 4.7: Coefficients (left) and reconstructions $R_{10}\psi$ and $R_{10}^{w}\psi$ (right) for the rescaled Hermite function. All coefficients show a rapid decrease for low frequencies and thus we obtain excellent reconstructions for all series.

gives a good approximation, even with few coefficients.

For a bad choice of the parameter ρ, the reconstruction with the bump gets worse. According to Theorem 57, the Lipschitz constant L_s is getting large as $\rho \to \lambda$, implying a slow decay for low frequencies, which can particularly be observed for the choice $\rho_2 = 0.8$. This value leads to a large derivative of the smooth bump $w_{0.8,1}$ in the interval $(0.8, 1)$. For low frequencies, the coefficients and the errors for the bump show a slow decay (right plots in Figure 4.4,4.5) and are even worse than for the plain Fourier series.

4.4.3 A function of rapid decrease

We also applied the windowed reconstructions to

$$\psi(x) = \left(8x^3 - 24x^2 + 12x + 4\right) e^{-(x-1)^2/2}, \qquad \left[\lambda = 2\pi,\ \rho = 5.9,\ t = 1\right].$$

We note that $\psi(x+1)$ is the product of the Hermite polynomial $H_3(x) = 8x^3 - 12x$ times a Gaussian, *i.e.*, a rescaled Hermite function. For the centre we chose $t = 1$. In contrast to the previous examples, we now work with the Tukey window for $\alpha = 1 - \rho/\lambda$, see Definition 53. We recall that this window is a non-degenerate C^1-bump. The 2λ-periodic extension of ψ produces discontinuities with very small jumps, which can only be resolved with high frequencies. Consequently, for low frequencies all coefficients are almost the same and fall off rapidly, see Figure 4.7. Nevertheless, the plain coefficients are $\mathcal{O}(1/|k|)$, while the coefficients for the smooth bump again show the best asymptotic decay. For the reconstructions we used $R_{10}\psi$ and $R_{10}^{w}\psi$. As we observe in the right plot of Figure 4.7, the rapid decrease of the coefficients yields excellent reconstructions and no differences can be determined to the original function.

4.5 Summary of this chapter

The periodisation of smooth functions usually creates jump discontinuities and therefore the Fourier coefficients decay slowly. Bump windows can be used to avoid this effect. The corresponding windowed Fourier coefficients have a faster decay, which implies that the pointwise reconstruction via the windowed Fourier series converges faster in the region where the bump has its plateau. In particular, the decay rate of the windowed Fourier coefficients depends on the Lipschitz constant of the windowed periodisation. In Theorem 57 we presented a new bound for the Lipschitz constant based on an explicit bound for derivatives of the product of two functions.

5 The Time-Sliced Thawed Gaussian Propagation Method

As described in detail in Section 1.1, parts of the present chapter overlap to a large extent with the joint preprint "An Error Representation for the Time-Sliced Thawed Gaussian Propagation Method" with C. Lasser submitted to *Numerische Mathematik* on 27/08/2021, e-print available at arXiv:2108.12182.

In this chapter, we study the time-sliced thawed Gaussian propagation method introduced in 2016 by Kong *et al.* for solving the time-dependent Schrödinger equation

$$i\varepsilon\partial_t\psi(x,t) = -\frac{\varepsilon^2}{2}\Delta_x\psi(x,t) + V(x)\psi(x,t), \quad \psi(x,0) = \psi_0 \in L^2(\mathbb{R}^d), \tag{5.1}$$

based on the concatenation of thawed Gaussian propagation steps. We present a detailed mathematical description of all subroutines, which allows a direct comparison with other state-of-the-art methods and the derivation of a rigorous error representation. Since the representation of Gaussian wave packets according to the discretisation of the FBI formula is the central tool of the TSTG method, our results from Chapter 3 will play an important role here.

The chapter is organised as follows: After presenting a detailed description and the connections to other methods in Section 5.1, we investigate the errors generated by the individual subroutines and their concatenation, which includes error analysis of thawed Gaussian approximations and time discretisation for both variationally and non-variationally evolving basis functions in Section 5.2. The full error representation of the method is then discussed in Section 5.3. Finally, the one-dimensional numerical experiments in Section 5.4 support our theoretical results and illustrate the applicability of the method to simulations of quantum dynamics, including tunneling dynamics in a double-well potential.

5.1 Mathematical description of the method

Let $\{z_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ be a given grid in phase space and recall that we use the notation $g_{\mathbf{n}} = g_{z_{\mathbf{n}}}$ for a multi-index $\mathbf{n} \in \Gamma \subset \mathbb{N}^{2d}$ and the Gaussian wave packet g_z in (2.2). In the following we work with time-evolved basis functions, and to make clear that we distinguish between original and time-evolved basis functions, we write $g_{\mathbf{n},0}$ for the original and $g_{\mathbf{n}}(t)$ for the time-evolved basis functions at time $t > 0$.

Based on the time-independent linear approximation space

$$\mathcal{V}_\Gamma := \operatorname{span}\left\{g_{\mathbf{n},0} \in L^2(\mathbb{R}^d) \,:\, \mathbf{n} \in \Gamma\right\} \subset L^2(\mathbb{R}^d),$$

the TSTG method approximates the solution ψ of the Schrödinger equation (5.1) with time-dependent coefficients as follows,

$$\psi(t) \approx \psi_\Gamma(t) := \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(t)\, g_{\mathbf{n},0}, \tag{5.2}$$

where the coefficients $c_{\mathbf{n}}(t)$ result from a concatenation of thawed Gaussian propagation steps for the basis functions $g_{\mathbf{n},0}$ with the reinitialisation of the time-evolved basis in the time-independent approximation space $\mathcal{V}_\Gamma$. To formulate the underlying equations of motion for the coefficients, we extend the quadrature-based pair of operators $\mathcal{A}_\Gamma$ and $\mathcal{S}_\Gamma = \mathcal{A}_\Gamma^*$ from Definition 25 by the so-called *reinitialisation operator*, which can be viewed as a multidimensional version of the matrix-vector product and is defined for a given tensor $\mathcal{C} \in \mathbb{C}^{\Gamma\times\Gamma}$ by

$$\mathcal{R}_\Gamma(\mathcal{C})\colon \mathbb{C}^\Gamma \to \mathbb{C}^\Gamma,\ (c_{\mathbf{n}}) \mapsto \sum_{\mathbf{n}'\in\Gamma} \mathcal{C}_{\mathbf{n},\mathbf{n}'} c_{\mathbf{n}'}.$$

Remark 64. *Note that the approximation space $\mathcal{V}_\Gamma$ depends on the underlying phase space grid $\{z_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$. Kong* et al. *used uniform grids, but other choices are also possible.*

With the triplet $(\mathcal{A}_\Gamma, \mathcal{S}_\Gamma, \mathcal{R}_\Gamma)$ in hand, we can now formulate the TSTG method, which starts to run through the following three subroutines:

(s1) *Representation coefficients of the initial wave function*:
The first subroutine calculates the inner products

$$\mathcal{A}_\Gamma\psi_0 = (\langle g_{\mathbf{n},0} \mid \psi_0\rangle) \in \mathbb{C}^\Gamma,$$

which can be used to reconstruct a given initial wave function ψ_0 as follows:

$$\psi_0 \approx \mathcal{A}_\Gamma^*\mathcal{A}_\Gamma\psi_0 = \sum_{\mathbf{n}\in\Gamma} w_{\mathbf{n}} \langle g_{\mathbf{n},0} \mid \psi_0\rangle\, g_{\mathbf{n},0} = \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\, g_{\mathbf{n},0} \in \mathcal{V}_\Gamma$$

(s2) *Thawed Gaussian propagation of the basis functions*:
Recall the definition of the approximation manifold $\mathcal{M}$ in (2.7). In the second subroutine of the TSTG method, each individual basis function $g_{\mathbf{n},0}$ is propagated for a short propagation period $\tau > 0$. More precisely, each time-evolved basis function $g_{\mathbf{n}}(\tau)$ is approximated by an element $u_{\mathbf{n}}(\tau)$ in the manifold $\mathcal{M}$ according to the thawed Gaussian propagation method, see [Hel75] and Section 5.2. Based on a numerical integrator for the corresponding equations of motion, we introduce the numerical propagator

$$\mathcal{U}_{\mathbf{n}}^\tau\colon \mathcal{M} \to \mathcal{M},\ g_{\mathbf{n},0} \mapsto u_{\mathbf{n}}^\tau, \tag{5.3}$$

where the superscript indicates that $u^\tau_{\mathbf{n}} \in \mathcal{M}$ is the numerical approximation to $u_{\mathbf{n}}(\tau)$ obtained by solving a system of ordinary differential equations (see [KMB16, Equation 17] and Section 5.2). Then, for all $\mathbf{n} \in \Gamma$, the second subroutine produces the numerical approximations

$$u^\tau_{\mathbf{n}} = \mathcal{U}^\tau_{\mathbf{n}}\, g_{\mathbf{n},0} \approx g_{\mathbf{n}}(\tau).$$

Remark 65. *It is known that the thawed Gaussian $u_{\mathbf{n}}(\tau)$ is an accurate approximation to the true solution $g_{\mathbf{n}}(\tau)$ only if the potential V in the Schrödinger equation can be approximated as harmonic over the entire "support" of $u_{\mathbf{n}}(\tau)$, i.e., as long as its width is not too wide, see also Lemma 72 for a precise estimate.*

(s3) *Computation of coefficients for the reinitialisation*:
The numerical approximations $u^\tau_{\mathbf{n}} \in \mathcal{M}$ obtained in subroutine (s2) are now reexpanded in $\mathcal{V}_\Gamma$. For all $\mathbf{n} \in \Gamma$ we apply the analysis operator $\mathcal{A}_\Gamma$ to the Gaussian wave packet $u^\tau_{\mathbf{n}}$, which gives us the inner products

$$\mathcal{A}_\Gamma u^\tau_{\mathbf{n}} = (\langle g_{\mathbf{n}',0} \mid u^\tau_{\mathbf{n}}\rangle) \in \mathbb{C}^\Gamma.$$

The result of the third subroutine is then a tensor $\mathcal{C}^\tau \in \mathbb{C}^{\Gamma\times\Gamma}$ containing the coefficients $\mathcal{C}^\tau_{\mathbf{n}',\mathbf{n}} := c_{\mathbf{n}'}(u^\tau_{\mathbf{n}}) = w_{\mathbf{n}'}\langle g_{\mathbf{n}',0} \mid u^\tau_{\mathbf{n}}\rangle$ for all $\mathbf{n}, \mathbf{n}' \in \Gamma$. In particular, this tensor is obtained **without numerical integration**, since all coefficients sample inner products of two Gaussians and can be calculated by hand according to Lemma 7.

Remark 66. *Note that the approximation $u^\tau_{\mathbf{n},\Gamma} \in \mathcal{V}_\Gamma$ of $u^\tau_{\mathbf{n}}$ is given by*

$$u^\tau_{\mathbf{n},\Gamma} := \mathcal{A}^*_\Gamma \mathcal{A}_\Gamma u^\tau_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} w_{\mathbf{n}'}\langle g_{\mathbf{n},0} \mid \psi_0\rangle\, g_{\mathbf{n},0} = \sum_{\mathbf{n}'\in\Gamma} c_{\mathbf{n}'}(u^\tau_{\mathbf{n}})\, g_{\mathbf{n}',0}.$$

Once we run the above subroutines, we are equipped with the tensor $\mathcal{A}_\Gamma\psi_0$ for the approximation of the initial wave function and the tensor $\mathcal{C}^\tau \in \mathbb{C}^{\Gamma\times\Gamma}$ containing the coefficients $c_{\mathbf{n}'}(u^\tau_{\mathbf{n}})$. We therefore obtain an approximation of the solution $\psi(\tau)$ to the Schrödinger equation at time τ as follows,

$$\begin{aligned}
\psi(\tau) &\overset{(s1)}{\approx} U(\tau)\mathcal{A}^*_\Gamma\mathcal{A}_\Gamma\psi_0 \overset{(s2)}{\approx} \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\,\mathcal{U}^\tau_{\mathbf{n}}\, g_{\mathbf{n},0} = \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\, u^\tau_{\mathbf{n}} \\
&\overset{(s3)}{\approx} \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\, \mathcal{A}^*_\Gamma\mathcal{A}_\Gamma u^\tau_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0) u^\tau_{\mathbf{n},\Gamma} \\
&= \sum_{\mathbf{n}\in\Gamma}\left(\sum_{\mathbf{n}'\in\Gamma} \mathcal{C}^\tau_{\mathbf{n},\mathbf{n}'} c_{\mathbf{n}'}(\psi_0)\right) g_{\mathbf{n},0} = \mathcal{A}^*_\Gamma \mathcal{R}^\tau_\Gamma \mathcal{A}_\Gamma \psi_0 =: \psi^{1,\tau}_\Gamma,
\end{aligned}$$

where we have changed the names of the indices $\mathbf{n}$ and $\mathbf{n}'$ to get to the third line and introduced the notation $\mathcal{R}_\Gamma^\tau := \mathcal{R}_\Gamma(\mathcal{C}^\tau)$. Furthermore, using that the unitary propagator $U(t) = e^{-iHt/\varepsilon}$ can be decomposed for $n > 1$ as

$$U(n\tau) = U(\tau) \cdot U(\tau) \cdots U(\tau),$$

single TSTG propagation steps can be concatenated to approximate the solution at times $2\tau, 3\tau, \dots$, where for the $(n+1)$th iteration the result $\psi_\Gamma^{n,\tau}$ of the nth iteration is used as the new initial datum. This results in the following approximation at larger times $t_n = n\tau$,

$$\psi(t_n) \approx \mathcal{A}_\Gamma^* \left(\mathcal{R}_\Gamma^\tau\right)^n \mathcal{A}_\Gamma \psi_0 =: \psi_\Gamma^{n,\tau},$$

where we have replaced the operator $\mathcal{A}_\Gamma \mathcal{A}_\Gamma^*$ in the intermediate steps with the identity, because the coefficients from a previous step can be kept in memory. In particular, reinitialising the time-evolved basis functions yields that the coefficients of $\psi_\Gamma^{n,\tau}$ are given for all $\mathbf{n} \in \Gamma$ by the following recursion formula:

$$\begin{aligned}(c_\mathbf{n}^{n,\tau}) := \left(\mathcal{R}_\Gamma^\tau\right)^n (c_\mathbf{n}(\psi_0)) &= \mathcal{R}_\Gamma^\tau \left(\left(\mathcal{R}_\Gamma^\tau\right)^{n-1} (c_\mathbf{n}(\psi_0))\right) \\ &= \sum_{\mathbf{n}' \in \Gamma} c_{\mathbf{n}'}^{n-1,\tau} c_\mathbf{n}(u_{\mathbf{n}'}^\tau), \quad c_\mathbf{n}^{0,\tau} := c_\mathbf{n}(\psi_0).\end{aligned} \tag{5.4}$$

Finally, it should be noted that the coefficients (and hence the approximation) are iteratively updated on the discrete time grid $2\tau, 3\tau, \dots$ and therefore (5.2) should be rewritten for a fixed propagation time t_n as follows:

$$\psi(t_n) \approx \psi_\Gamma^{n,\tau} = \sum_{\mathbf{n} \in \Gamma} c_\mathbf{n}^{n,\tau} \, g_{\mathbf{n},0}$$

5.1.1 Comparison with the Galerkin method

Looking at the ansatz in (5.2), one could determine the corresponding time-dependent coefficient tensor $c = (c_\mathbf{n})$ using the standard Galerkin method, which yields a linear system of ordinary differential equations and is derived from the condition that

$$\begin{aligned}&\partial_t \psi_\Gamma(t) \in \mathcal{V}_\Gamma \quad \text{is such that} \\ &\left\langle \varphi \mid -i\varepsilon \partial_t \psi_\Gamma(t) + H\psi_\Gamma(t) \right\rangle = 0 \quad \text{for all } \varphi \in \mathcal{V}_\Gamma.\end{aligned} \tag{5.5}$$

With the orthogonal projection $P\colon L^2(\mathbb{R}^d) \to \mathcal{V}_\Gamma$ onto the approximation space, the Galerkin condition (5.5) can also be written as

$$i\varepsilon \partial_t \psi_\Gamma = PH\psi_\Gamma.$$

However, in order to achieve a certain accuracy for the discretisation of the wave packet transform, the grid points $z_\mathbf{n}$ must be chosen sufficiently dense, which means that the basis functions have a large overlap and therefore, as discussed in Section 3.3.1, the

Gram matrix of the Galerkin method becomes ill-conditioned. One way around this problem would be to replace the Gaussian basis functions with an orthonormal basis, which implies that the Gram matrix becomes the identity. A comparison must be made with the Galerkin method in [Lub08, Chapter III.1.1], where the time-independent approximation space is spanned by the first $N \geq 1$ Hermite functions

$$\varphi_n(x) := \frac{1}{\sqrt{2^n n! \sqrt{\pi}}} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}, \quad n = 0, 1, \ldots, N-1,\, x \in \mathbb{R},$$

which are known to form an orthonormal set. While this choice of basis functions solves the ill-conditioned inversion problem of the Gram matrix and provides a simple representation of the orthogonal projection, namely

$$P\psi = \sum_{n=0}^{N-1} \langle \varphi_n \mid \psi \rangle \, \varphi_n, \quad \psi \in L^2(\mathbb{R}), \tag{5.6}$$

which is used in [Lub08, Chapter III.1.1 (Theorem 1.2)] to derive the approximation error of the Galerkin method, in practical applications the dimension of $\mathcal{V}_\Gamma$ must be chosen large in order to compute the time-evolution of the wave function with sufficient accuracy. For instance, for simulations of tunnelling in double-well potentials (quartic potentials with two local minima separated by energy barriers) as presented later in Section 5.4.2, the Hermite basis is expensive because the functions are localised by a Gaussian envelope and therefore the degree of the polynomial prefactors must be chosen large to capture both minima of the potential. Besides the Hermite functions, we would also like to mention the Galerkin approximation for Hagedorn functions, a generalisation of the Hermite functions based on a Gaussian amplitude with arbitrary width matrix in the Siegel upper half-space, see e.g. [LL20, Section 4.3] and [GH14, BG20].

Remark 67. *Like the Galerkin method, the TSTG method is based on a time-independent approximation space. We note that time-dependent approximation spaces have also been studied in the past. For example, linear combinations of time-evolved frozen Gaussians were proposed by Heller, see [Hel81].*

Orthogonal projections: frames revisited

As we have already mentioned above, the projection operator is typically used to estimate the approximation error of the Galerkin method, see e.g. [Lub08, Chapter III.1.1 (Theorem 1.3)] for the Hermite basis. While there is a simple representation of the projection operator for orthonormal bases, cf. (5.6), the situation is more difficult for the approximation space $\mathcal{V}_\Gamma$, which is based on the non-orthogonal Gaussian functions. Let us briefly discuss how to calculate the orthogonal projection on $\mathcal{V}_\Gamma$.

The first observation is that $\{g_\mathbf{n}\}_{\mathbf{n}\in\Gamma}$ is a frame of $\mathcal{V}_\Gamma$, which obviously follows from our assumption that Γ is a finite index set. Thus, if $\{\tilde{g}_\mathbf{n}\}_{\mathbf{n}\in\Gamma}$ denotes the dual frame, it follows from Proposition 36 that any function $\psi \in \mathcal{V}_\Gamma$ can be represented as

$$\psi = \sum_{\mathbf{n}\in\Gamma} \langle \tilde{g}_{\mathbf{n}} \mid \psi \rangle \, g_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} \langle g_{\mathbf{n}} \mid \psi \rangle \, \tilde{g}_{\mathbf{n}},$$

and therefore it is reasonable to expect that the orthogonal projection can be written as

$$P = \sum_{\mathbf{n}\in\Gamma} \langle \tilde{g}_{\mathbf{n}} \mid \bullet \rangle \, g_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} \langle g_{\mathbf{n}} \mid \bullet \rangle \, \tilde{g}_{\mathbf{n}}.$$

The next proposition was taken from [Mal09, Theorem 5.6] and shows that this representation of the projection is indeed valid.

Proposition 68. *Let $\mathcal{V}$ be a subspace of the Hilbert space $\mathcal{H}$. Moreover, let $\{\phi_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ be a frame of $\mathcal{V}$ and $\{\tilde{\phi}_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ its dual frame in $\mathcal{V}$. Then, the orthogonal projection of $\psi \in \mathcal{H}$ in $\mathcal{V}$ is given by*

$$P\psi = \sum_{\mathbf{n}\in\Gamma} \langle \tilde{\phi}_{\mathbf{n}}, \psi \rangle_{\mathcal{H}} \, \phi_{\mathbf{n}} = \sum_{\mathbf{n}\in\Gamma} \langle \phi_{\mathbf{n}}, \psi \rangle_{\mathcal{H}} \, \tilde{\phi}_{\mathbf{n}}. \tag{5.7}$$

The following proof was taken from [Mal09, Theorem 5.6]:

Proof. Since both frames are dual in $\mathcal{V}$, if $\psi \in \mathcal{V}$, then Proposition 36 proves that the operator P in (5.7) satisfies $P\psi = \psi$. To prove that it is an orthogonal projection, it is sufficient to verify that if $\psi \in \mathcal{H}$ then $\langle \phi_{\mathbf{m}}, \psi - P\psi \rangle_{\mathcal{H}} = 0$ for all $\mathbf{m} \in \Gamma$. Indeed,

$$\langle \phi_{\mathbf{m}}, \psi - P\psi \rangle_{\mathcal{H}} = \langle \phi_{\mathbf{m}}, \psi \rangle_{\mathcal{H}} - \sum_{\mathbf{m}\in\Gamma} \langle \phi_{\mathbf{n}}, \psi \rangle_{\mathcal{H}} \langle \phi_{\mathbf{m}}, \tilde{\phi}_{\mathbf{n}} \rangle_{\mathcal{H}} = 0,$$

because the dual frame property implies that

$$\sum_{\mathbf{m}\in\Gamma} \langle \phi_{\mathbf{m}}, \tilde{\phi}_{\mathbf{n}} \rangle_{\mathcal{H}} \, \phi_{\mathbf{n}} = \phi_{\mathbf{m}}.$$

□

Since we have no analytic expression of the dual frame for a finite number of Gaussian basis functions, we learn from Proposition 68 that there is no analytic representation of the orthogonal projection $P\colon L^2(\mathbb{R}^d) \to \mathcal{V}_\Gamma$. Furthermore, we observe that discretisations of the FBI formula can be understood as approximations to the projection operator, which in the limit $\mathcal{V}_\Gamma \to L^2(\mathbb{R}^d)$ converge to the identity operator

$$\mathrm{Id}_{L^2(\mathbb{R}^d)} = (2\pi\varepsilon)^{-d} \int_{\mathbb{R}^{2d}} |g_z\rangle\langle g_z| \, \mathrm{d}z := (2\pi\varepsilon)^{-d} \int_{\mathbb{R}^{2d}} \langle g_z \mid \bullet \rangle \, g_z \, \mathrm{d}z,$$

where we have used the bra-ket notation in the middle of the equation.

5.2 Gaussian wave packet dynamics

This section deals with the thawed Gaussian propagation of the basis functions $g_{\mathbf{n},0} \in \mathcal{M}$. The main result is the error representation for a single TSTG step in Proposition 75, which combines an estimate for thawed Gaussian approximations with an estimate for the numerical integration of the underlying equations of motion.

Recall that in subroutine (s2) the individual basis functions are propagated according to the (non-variational) thawed Gaussian equations for $z \in \mathbb{R}^{2d}$, $C \in \mathfrak{S}^+(d)$ and $S \in \mathbb{C}$, which combine the Hamiltonian system

$$\dot{z}(t) = J\nabla h(z), \quad h(z) = \frac{1}{2}|p|^2 + V(q), \quad J = \begin{pmatrix} 0 & \mathrm{Id}_d \\ -\mathrm{Id}_d & 0 \end{pmatrix} \in \mathbb{R}^{2d\times 2d} \tag{5.8}$$

for the motion of the centre $z(t)$ with equations for $C(t)$ and $S(t)$ which ensure that we obtain exact solutions in the presence of a quadratic potential. In addition to the work of Kong *et al.*, other propagation methods are also possible as long as the approximations $u_{\mathbf{n}}(\tau)$ lie in the Gaussian manifold $\mathcal{M}$ so that the coefficients for the re-expansion can be calculated analytically. An alternative is offered, for example, by variationally evolving Gaussian wave packets, which are employed in this dissertation and have not yet been used in connection with the TSTG method.

5.2.1 Variational approximation

Suppose that for a given time interval $[0, T]$ we want to approximate the solution of the Schrödinger equation

$$i\varepsilon\partial_t\psi(x,t) = H\psi(x,t), \quad \psi(x,0) = \psi_0 \in \mathcal{M},$$

by a Gaussian wave packet $u(t) = u(\bullet, t) \in \mathcal{M}$. By requiring that

$$\begin{aligned} &\partial_t u(t) \in \mathcal{T}_{u(t)}\mathcal{M} \quad \text{is chosen such that} \\ &\big\langle v \mid -i\varepsilon\partial_t u(t) + Hu(t)\big\rangle = 0 \quad \text{for all } v \in \mathcal{T}_{u(t)}\mathcal{M}, \end{aligned} \tag{5.9}$$

we guarantee that $\partial_t u(t)$ is given by the unique element w in the tangent space $\mathcal{T}_{u(t)}\mathcal{M}$ at $u(t)$ such that

$$\left\| w - \frac{1}{i\varepsilon}Hu \right\| \quad \text{is minimal,}$$

or, in other words, $\partial_t u(t)$ is the orthogonal projection of $\frac{1}{i\varepsilon}Hu$ onto the tangent space, which can also be written as

$$i\varepsilon\partial_t u(t) = P_u Hu.$$

The condition (5.9) is called the Dirac–Frenkel time-dependent variational approximation principle and we refer to [Lub08, Chapter II.1] for further discussion and properties of variational approximations such as norm and energy conservation.

Remark 69. *Note that the Galerkin condition (5.5) can be viewed as the time-dependent variational principle on the linear approximation space $\mathcal{V}_\Gamma$, see [Lub08, Chapter III.1.1].*

Recall Lemma 6 which states that for all differential operators A of order ≤ 2 and $u \in \mathcal{M}$ it holds that $Au \in \mathcal{T}_u\mathcal{M}$. Consequently, it follows for a quadratic potential that $Hu \in \mathcal{T}_u\mathcal{M}$ and thus the approximation $u(t)$ and the correct solution $\psi(t)$ satisfy the same differential equation. This proves that the variational approximation for quadratic potentials is exact. The following result was taken from [LL20, Proposition 3.2].

Proposition 70. *If the potential V in the time-dependent Schrödinger equation (5.1) is quadratic, then the variational approximation is exact, that is, $u(t) = \psi(t)$ for all $t \in \mathbb{R}$, provided that the initial wave function is a Gaussian, $u(0) = \psi(0) \in \mathcal{M}$.*

For the proof of Proposition 70 we refer to [LL20, Proposition 3.2]. □

We will see in a moment that under suitable conditions for the potential and the spectrum of the time-dependent width matrix $C(t)$, variational Gaussians provide approximations of order $\mathcal{O}(\sqrt{\varepsilon})$. The corresponding equations of motion for the Gaussian parameters were first derived by Coalson and Karplus, see [CK90]. Using Hagedorn's parametrisation $C = PQ^{-1}$ introduced in Section 2.1.1, these equations read

$$\begin{aligned} \dot{q} &= p \qquad \text{and} \qquad \dot{p} = -\langle \nabla_x V \rangle_u, \\ \dot{Q} &= P \qquad \text{and} \qquad \dot{P} = -\langle \nabla_x^2 V \rangle_u Q, \\ S(t) &= \int_0^t \left(\frac{1}{2}|p(s)|^2 - \langle V \rangle_{u(s)} + \frac{\varepsilon}{4} \operatorname{tr}\left(Q(s)^* \langle \nabla_x^2 V \rangle_{u(s)} Q(s) \right) \right) \mathrm{d}s, \end{aligned} \tag{5.10}$$

where we denote by $\langle W \rangle_u := \langle u \mid Wu \rangle$, $W \in \{V, \nabla_x V, \nabla_x^2 V\}$, the expected values. In particular, for the propagation of the basis function $g_{\mathbf{n},0}$ according to subroutine (s2) the initial conditions are given by

$$z_{\mathbf{n}}(0) = z_{\mathbf{n}},\ Q_{\mathbf{n}}(0) = \operatorname{Im}(C_0)^{-1/2},\ P_{\mathbf{n}}(0) = C_0 Q_{\mathbf{n}}(0) \quad \text{and} \quad S_{\mathbf{n}}(0) = 0,$$

where $\operatorname{Im}(C_0)^{1/2}$ is the unique positive definite square root of $\operatorname{Im}(C_0) > 0$.

Remark 71. *To obtain the equations of motion for the non-variational thawed Gaussians used by Kong* et al.*, we have to replace the equations in (5.10) for the parameters $(q(t), p(t), Q(t), P(t))$ by the point evaluations*

$$\begin{aligned} \dot{q} &= p \qquad \textit{and} \qquad \dot{p} = -V(q), \\ \dot{Q} &= P \qquad \textit{and} \qquad \dot{P} = -\nabla_x^2 V(q) Q, \\ S(t) &= \int_0^t \left(\frac{1}{2}|p(s)|^2 - V\big(q(s)\big) \right) \mathrm{d}s, \end{aligned}$$

which are computationally less demanding than the variational equations. In particular, this implies that the matrix $Z(t) = (Q(t), P(t))$ is a solution of the linearisation of the classical equations

$$\dot{Z}(t) = J\nabla^2 h(z(t))Z(t),$$

where the function h and the matrix J are defined according to (5.8). We also note that the matrix conditions in (2.3) and (2.4) ensure the correct normalisation of the approximation $u \in \mathcal{M}$ and that the above equations agree with those in (5.10) in the presence of a quadratic potential. Like variational Gaussians, non-variational Gaussians yield approximations of order $\mathcal{O}(\sqrt{\varepsilon})$, see Lemma 72.

5.3 Error representation

The next lemma presents the accuracy of the thawed Gaussian methods and extends the results for the L^2-error for variational Gaussians in [LL20, Theorem 3.5] to non-variational Gaussians. We note that the first L^2-error for non-variational Gaussians was proved by Hagedorn, see [Hag98, Theorem 2.9].

Lemma 72. *Assume that*

- *the eigenvalues of the positive definite width matrix $\mathrm{Im}(C(t))$ are bounded from below by a constant $\rho > 0$, for all $t \in [0, \tau]$.*
- *the potential function V is three times continuously differentiable with a polynomially bounded third derivative.*

Moreover, assume that $u(t) \in \mathcal{M}$ is an approximation to the Schrödinger equation (5.1) that results from the variational or non-variational thawed Gaussian propagation method. Then, there exists a positive constant $C^{(1)} > 0$ such that the error between the approximation $u(t)$ and the solution $\psi(t)$ is bounded in the L^2-norm by

$$\|u(t) - \psi(t)\| \leq C^{(1)}\, t\sqrt{\varepsilon}, \quad 0 \leq t \leq \tau, \tag{5.11}$$

where $C^{(1)}$ is independent of ε and t but depends on ρ.

The estimate in (5.11) shows that thawed Gaussian approximations produce errors that increase linearly in t, where a small semiclassical parameter yields an improvement by a factor $\sqrt{\varepsilon}$ for the constant $C^{(1)}$. Crucial to the proof of Lemma 72 is the fact that both the variational and non-variational approximations are exact provided the potential is quadratic, see Proposition 70. Therefore, the estimate results from a bound for the defect of the cubic part of the potential.

Proof. Let $U_q \colon \mathbb{R}^d \to \mathbb{R}$ denote the second-order Taylor polynomial of V at q and let $W_q \colon \mathbb{R}^d \to \mathbb{R}$ be the corresponding remainder, *i.e.*,

$$V = U_q + W_q.$$

Since the approximation $u(t) \in \mathcal{M}$ is the exact solution to

$$i\varepsilon\partial_t u(t) = -\frac{\varepsilon^2}{2}\Delta_x u(t) + U_{q(t)}u(t), \quad u(0) = \psi(0) = \psi_0,$$

we obtain

$$\partial_t(u - \psi) = \frac{1}{i\varepsilon}H(u - \psi) - \frac{1}{i\varepsilon}W_q u,$$

where

$$\|W_q u\| = (\pi\varepsilon)^{-d/4}\det(\operatorname{Im} C)^{1/4}\cdots \\ \left(\int_{\mathbb{R}^d}|W_q(x)|^2\exp\left(-\frac{1}{\varepsilon}(x-q)^T\operatorname{Im} C(x-q)\right)\,\mathrm{d}x\right)^{1/2}.$$

Moreover, using that $W_q(x)$ is the non-quadratic remainder at q, an estimate for moments of Gaussian functions (see [LL20, Lemma 3.8]) yields the existence of a constant $C^{(1)} > 0$, depending on ρ, such that

$$\|W_q u\| \le C^{(1)}\,\varepsilon^{3/2}.$$

Consequently, since $u - \psi$ satisfies the Schrödinger equation up to the defect

$$d(t) = -\frac{i}{\varepsilon}W_{q(t)}u(t),$$

we finally conclude that

$$\|u(t) - \psi(t)\| \le \int_0^t \|d(s)\|\,\mathrm{d}s = \frac{1}{\varepsilon}\int_0^t \|W_{q(s)}u(s)\|\,\mathrm{d}s \le C^{(1)}t\sqrt{\varepsilon}.$$

□

Remark 73. *We note that the equations of motion for the variational and non-variational thawed Gaussian methods are different and we therefore obtain individual lower bounds for the eigenvalues of the width matrix, so that, although we have omitted this dependency in our notation of Lemma 72, we obtain individual constants for the two methods. In particular, the estimate of Lasser and Lubich for Gaussian moments shows that the constant $C^{(1)}$ depends on the third derivative of V and is of order $\rho^{-3/2}$ with respect to the spectral parameter ρ. We also mention that, in contrast to the computation of the full wave function, the error in the expected value of observables improves to an order $\mathcal{O}(\varepsilon)$ accuracy, see [LL20, Theorem 3.5b].*

For the thawed Gaussian propagation of the basis functions $g_{\mathbf{n},0}$ we see that the propagation time τ must be chosen in such a way that we obtain accurate approximations for all $\mathbf{n} \in \Gamma$. With this in mind, it should be noted that small values of τ lead to more concatenation steps to approximate the solution for a fixed final time. We present numerical experiments for the dependency on τ in Section 5.4.1. Frozen Gaussian approximations would also be possible, see [Hel81]. On the one hand, this leads to simpler equations of motion, since these approximations do not require information about the second derivative of the potential, on the other hand, the frozen Gaussian method reduces the order to $\mathcal{O}(1)$ with respect to the parameter ε.

We now turn to numerical integration for the equations of motion.

5.3.1 Time discretisation

For the integration of the equations of motion we need a suitable numerical integrator. In (5.3) we have therefore introduced the numerical propagator $\mathcal{U}_{\mathbf{n}}^{\tau}\colon \mathcal{M} \to \mathcal{M}$, which has not yet been defined in detail, except that it maps a Gaussian basis function $g_{\mathbf{n},0}$ to a numerical approximation $u_{\mathbf{n}}^{\tau} \approx g_{\mathbf{n}}(\tau)$. The development of such integrators essentially uses exponential operator splitting methods such as the first-order Lie splitting or the second-order Strang splitting, where the integrator is said to have order $s \geq 1$, if there exists a constant $C^{(2)} > 0$ such that the error between the approximation $u_{\mathbf{n}}^{\tau}$ obtained after $m \geq 1$ steps of size $h_\tau = \tau/m$ and the true solution $u_{\mathbf{n}}(\tau)$ is bounded by

$$\|u_{\mathbf{n}}^{\tau} - u_{\mathbf{n}}(\tau)\| \leq C^{(2)}\tau\frac{h_\tau^s}{\varepsilon}. \tag{5.12}$$

For example, the L^2-error of Strang splitting is $\mathcal{O}(h_\tau^2/\varepsilon)$, which implies that the step size h_τ must be sufficiently smaller than $\sqrt{\varepsilon}$. We refer to [DT10] for rigorous error bounds in the semiclassical scaling $\varepsilon \ll 1$.

Equipped with a numerical integrator, we obtain the following error estimate:

Proposition 74. *For $\tau > 0$ and a uniform time grid of step size $h_\tau > 0$ let*

$$E_{\mathbf{n}}^{\tau} = E_{\mathbf{n}}^{\tau}(h_\tau) := \|u_{\mathbf{n}}^{\tau} - g_{\mathbf{n}}(\tau)\|, \quad \mathbf{n} \in \Gamma. \tag{5.13}$$

Moreover, assume that $\mathcal{U}_{\mathbf{n}}^{\tau}\colon \mathcal{M} \to \mathcal{M}$ is a numerical integrator of order $s \geq 1$. Then, under the hypotheses of Lemma 72, for all $\mathbf{n} \in \Gamma$ there exists a positive constant $C_{\mathbf{n}} > 0$ such that

$$E_{\mathbf{n}}^{\tau} \leq C_{\mathbf{n}}\tau\left(\frac{h_\tau^s}{\varepsilon} + \sqrt{\varepsilon}\right). \tag{5.14}$$

Proof. Let $\tau > 0$ and $h_\tau > 0$. For all $\mathbf{n} \in \Gamma$, we combine the estimate in (5.11) with the estimate in (5.12) to obtain

$$E_{\mathbf{n}}^{\tau} \leq \|u_{\mathbf{n}}^{\tau} - u_{\mathbf{n}}(\tau)\| + \|u_{\mathbf{n}}(\tau) - g_{\mathbf{n}}(\tau)\| \leq C_{\mathbf{n}}^{(2)}\tau\frac{h_\tau^s}{\varepsilon} + C_{\mathbf{n}}^{(1)}\tau\sqrt{\varepsilon}$$

with corresponding positive constants $C_{\mathbf{n}}^{(1)}, C_{\mathbf{n}}^{(2)} > 0$. Consequently, the bound in (5.14) follows for the constant

$$C_{\mathbf{n}} = \max\left(C_{\mathbf{n}}^{(1)}, C_{\mathbf{n}}^{(2)}\right).$$

□

A second-order algorithm of the variational splitting was proposed and studied by Faou and Lubich, see [FL06]. In particular, it preserves the norm and the symplecticity relations of the matrices Q and P in (2.3) and (2.4). There are several other higher-order splittings for the unitary propagator that can also be implemented, and we refer the interested reader to [MQ02] and [HLW06, Chapter III]. Furthermore, we note that the symmetric Zassenhaus splitting, see [BIKS14], is an alternative to splitting methods that can also be used to increase the order of the time discretisation. In particular, in [BIKS16] the symmetric Zassenhaus splitting was combined with the Magnus expansion of the time-dependent Hamiltonian, see e.g. [IN99, IMKNZ00].

We are now equipped with an error estimate for thawed Gaussian approximations and the numerical integration of the thawed equations of motion. Together with the error representations for the discretisation of the FBI formula in Chapter 3, we are therefore able to analyse the total error introduced by a single TSTG step. Afterwards, in Theorem 78 we lift this error estimate to a global one.

5.3.2 Error after a single TSTG step

Recall that a single TSTG step consists of the following approximations:

- the approximation of the initial wave function ψ_0 in the approximation space $\mathcal{V}_\Gamma$ according to subroutine (s1)
- the thawed Gaussian approximations for the propagation of the basis functions and the numerical integration of the thawed equations of motion according to (s2)
- the re-expansion of the time-evolved basis functions in $\mathcal{V}_\Gamma$ according to (s3)

Let us introduce the following notation for the 1-norm of a tensor $(c_{\mathbf{n}}) \in \mathbb{C}^\Gamma$:

$$\|c_{\mathbf{n}}\|_1 := \sum_{\mathbf{n}\in\Gamma} |c_{\mathbf{n}}|.$$

The next proposition presents an error bound for a single TSTG step:

Proposition 75. *For a given phase space box $\Lambda \subset \mathbb{R}^{2d}$, a finite multi-index set $\Gamma \subset \mathbb{N}^{2d}$, grid points $z_{\mathbf{n}} \in \mathbb{R}^{2d}$ and positive weights $w_{\mathbf{n}} > 0$, $\mathbf{n} \in \Gamma$, recall the definition of the spatial discretisation error E_{wp} defined in (3.38). Moreover, for $\tau > 0$ and $h_\tau > 0$, recall the definition of the time discretisation error $E_{\mathbf{n}}^\tau$ in (5.13), produced by a numerical propagator for the thawed equations of motion of order $s \geq 1$. Then, there exists a positive constant $C > 0$ such that*

$$\|\psi(\tau) - \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{1,\tau} g_{\mathbf{n},0}\|_{L^2(\Lambda_q)} \leq C\tau\left(\frac{h_\tau^s}{\varepsilon} + \sqrt{\varepsilon}\right) + E^{1,\tau}, \tag{5.15}$$

where $E^{1,\tau} > 0$ denotes the total spatial discretisation error

$$E^{1,\tau} := E_{wp}(\psi_0) + C \cdot \max_{\mathbf{n}\in\Gamma} E_{wp}(u_{\mathbf{n}}^\tau).$$

Note that the bound in (5.15) depends on the spatial error E_{wp} resulting from the discretisation of the FBI formula. For example, if we use a direct discretisation of the phase space integral based on fully tensorised uniform Riemann sums with $N \geq 1$ grid points in each coordinate direction, Proposition 29 guarantees the existence of positive constants $C^{(\mathrm{T})}, C^{(\mathrm{RS})} > 0$ such that

$$E_{wp} \leq C^{(\mathrm{T})} + C^{(\mathrm{RS})} N^{-1}.$$

Proof. In the following, let $\| \bullet \|$ denote the L^2-norm on the projected box $\Lambda_q \subset \mathbb{R}^d$ in position space. Using that the evolution operator $U(\tau) = e^{-iH\tau/\varepsilon}$ is unitary, we get

$$\begin{aligned} \|\psi(\tau) - \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{1,\tau}\, g_{\mathbf{n},0}\| &= \|U(\tau)\psi_0 - \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{1,\tau}\, g_{\mathbf{n},0}\| \\ &\leq \|U(\tau)\,(\psi_0 - \mathcal{A}_\Gamma^* \mathcal{A}_\Gamma \psi_0) + \mathcal{A}_\Gamma^* \mathcal{A}_\Gamma U(\tau)\psi_0 - \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{1,\tau}\, g_{\mathbf{n},0}\| \\ &\leq E_{wp}(\psi_0) + \| \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\, g_{\mathbf{n}}(\tau) - \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{1,\tau}\, g_{\mathbf{n},0}\|. \end{aligned}$$

Moreover, for the second summand the definition of the coefficients $c_{\mathbf{n}}^{1,\tau}$ in (5.4) yields

$$\begin{aligned} &\| \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}(\psi_0)\, g_{\mathbf{n}}(\tau) - \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{1,\tau}\, g_{\mathbf{n},0}\| \\ &\qquad \leq \sum_{\mathbf{n}\in\Gamma} |c_{\mathbf{n}}(\psi_0)| \Big(\|\, g_{\mathbf{n}}(\tau) - u_{\mathbf{n}}^\tau\| + \|u_{\mathbf{n}}^\tau - \sum_{\mathbf{n}'\in\Gamma} c_{\mathbf{n}'}(u_{\mathbf{n}}^\tau)\, g_{\mathbf{n}',0}\| \Big) \\ &\qquad \leq \sum_{\mathbf{n}\in\Gamma} |c_{\mathbf{n}}(\psi_0)| \Big(E_{\mathbf{n}}^\tau + E_{wp}(u_{\mathbf{n}}^\tau) \Big). \end{aligned} \tag{5.16}$$

Consequently, using the bound for $E_{\mathbf{n}}^\tau$ in (5.14) with the constant $C_{\mathbf{n}} > 0$, the estimate in (5.15) follows for the choice

$$C = \|c_{\mathbf{n}}(\psi_0)\|_1 \cdot \max\left(1, \max_{\mathbf{n}\in\Gamma} C_{\mathbf{n}}\right).$$

□

The estimate for the sum in (5.16) combines the 1-norm with the maximum norm. However, since the spatial errors $E_{wp}(u_{\mathbf{n}}^\tau)$ increase at the boundary of the grid $\{z_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$, but the coefficients $c_{\mathbf{n}}(\psi_0)$ decrease exponentially with the distance $\|z_{\mathbf{n}} - z_0\|_2$, other Hölder conjugate exponents that reflect this grid-dependent interplay more accurately could also be chosen.

In the next step, we investigate the error resulting from the concatenation of the individual TSTG steps.

5.3.3 Estimate for the update-coefficients

As discussed in Section 5.1, the approximations for larger times $2\tau, 3\tau, \ldots$ are based on the update-coefficients $c_{\mathbf{n}}^{2,\tau}, c_{\mathbf{n}}^{3,\tau}, \ldots$ given by the recursion formula in (5.4). Let us take a closer look at the magnitude of these coefficients. Recall that

$$c_{\mathbf{n}}^{1,\tau} = \mathcal{R}_\Gamma^\tau \left(c_{\mathbf{n}}(\psi_0)\right) = \sum_{\mathbf{n}'\in\Gamma} c_{\mathbf{n}'}(\psi_0) c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau),$$

where both the factors $c_{\mathbf{n}'}(\psi_0)$ and $c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau)$ are Gaussian wave packets in phase space. Hence, as a sum of Gaussian wave packets, the update-coefficients $c_{\mathbf{n}}^{1,\tau}$ can be bounded by a Gaussian envelope. Furthermore, by induction on n, Gaussian bounds can be derived for all update coefficients $c_{\mathbf{n}}^{n,\tau}, n > 1$:

Proposition 76. *For $z_0 \in \mathbb{R}^{2d}$ and $C_0 \in \mathfrak{S}^+(d)$ let $\psi_0 = g_{z_0}^{C_0,\varepsilon}$. Moreover, let $\{z_{\mathbf{n}}\}_{\mathbf{n}\in\Gamma}$ be an arbitrary grid in phase space. Then, for all $n \geq 0$ and $\tau > 0$, there exist positive constants ζ_n^τ and $\theta_n^\tau > 0$ such that for all $\mathbf{n} \in \Gamma$ we have*

$$|c_{\mathbf{n}}^{n,\tau}| \leq \zeta_n^\tau \exp\left(-\frac{\theta_n^\tau}{8\varepsilon}\|z_{\mathbf{n}} - z_0\|_2^2\right). \tag{5.17}$$

For the proof of Proposition 76 we first derive an auxiliary result that allows us to bound the coefficients $c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau)$ of $u_{\mathbf{n}'}^\tau$ by a Gaussian envelope centred at $z_{\mathbf{n}'}$.

Lemma 77. *Under the assumptions of Proposition 76, for all $\mathbf{n}' \in \Gamma$, there exist positive constants $\zeta_{\mathbf{n}'}^\tau > 0$ and $\theta_{\mathbf{n}'}^\tau > 0$ such that for all $\mathbf{n} \in \Gamma$ we have*

$$|c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau)| \leq \zeta_{\mathbf{n}'}^\tau \exp\left(-\frac{\theta_{\mathbf{n}'}^\tau}{8\varepsilon}\|z_{\mathbf{n}} - z_{\mathbf{n}'}\|^2\right). \tag{5.18}$$

Proof. Let $\mathbf{n}' \in \Gamma$ and $\tau > 0$. The definition of the coefficients $c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau)$ implies

$$|c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau)| = w_{\mathbf{n}} |\langle g_{\mathbf{n}} \mid u_{\mathbf{n}'}^\tau \rangle| \quad \text{for all } \mathbf{n} \in \Gamma,$$

where the non-negative weights $w_{\mathbf{n}} \geq 0$ depend on the underlying quadrature rule. Therefore, using the bounds for inner products in Lemma 7, we find constants $\beta_{\mathbf{n}'}^\tau, \theta_{\mathbf{n}'}^\tau > 0$ such that

$$|c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau)| \leq \beta_{\mathbf{n}'}^\tau \exp\left(-\frac{\theta_{\mathbf{n}'}^\tau}{8\varepsilon}\|z_{\mathbf{n}} - z_{\mathbf{n}'}(\tau)\|_2^2\right),$$

where $z_{\mathbf{n}'}(\tau)$ is the centre of the time-evolved Gaussian $u_{\mathbf{n}'}^\tau \in \mathcal{M}$. To bound $|c_{\mathbf{n}}(u_{\mathbf{n}'}^\tau)|$ by a Gaussian centred at the original grid point $z_{\mathbf{n}'} = z_{\mathbf{n}'}(0)$, we write the time-evolved centres in terms of the original grid points as

$$z_{\mathbf{n}'}(\tau) = z_{\mathbf{n}'} + \delta_{\mathbf{n}'}(\tau)$$

and introduce the maximal phase space shift

$$\delta(\tau) := \max_{\mathbf{n}'\in\Gamma} \|\delta_{\mathbf{n}'}(\tau)\|_2.$$

Using the Cauchy–Schwarz inequality in $\mathbb{R}^d$, it then follows that

$$\begin{aligned}\exp\left(-\frac{\theta^\tau_{\mathbf{n}'}}{8\varepsilon}\|z_{\mathbf{n}} - z_{\mathbf{n}'}(\tau)\|_2^2\right) &= \exp\left(-\frac{\theta^\tau_{\mathbf{n}'}}{8\varepsilon}\|z_{\mathbf{n}} - z_{\mathbf{n}'} - \delta_{\mathbf{n}'}(\tau)\|_2^2\right)\\ &= \exp\left(-\frac{\theta^\tau_{\mathbf{n}'}}{8\varepsilon}\|z_{\mathbf{n}} - z_{\mathbf{n}'}\|_2^2\right)\exp\left(\frac{\theta^\tau_{\mathbf{n}'}}{4\varepsilon}\delta_{\mathbf{n}'}(\tau)^T(z_{\mathbf{n}} - z_{\mathbf{n}'})\right)\exp\left(-\frac{\theta^\tau_{\mathbf{n}'}}{8\varepsilon}\|\delta_{\mathbf{n}'}(\tau)\|_2^2\right)\\ &\leq \exp\left(-\frac{\theta^\tau_{\mathbf{n}'}}{8\varepsilon}\|z_{\mathbf{n}} - z_{\mathbf{n}'}\|_2^2\right)\exp\left(\frac{\theta^\tau_{\mathbf{n}'}}{4\varepsilon}\delta(\tau)\|z_{\mathbf{n}} - z_{\mathbf{n}'}\|_2\right).\end{aligned}$$

Hence, if we denote by $D_{\max} > 0$ the maximal distance $\|z_{\mathbf{n}} - z_{\mathbf{n}'}\|_2$ between two grid points in phase space and

$$\beta^\tau := \max_{\mathbf{n}'\in\Gamma} \exp\left(\frac{\theta^\tau_{\mathbf{n}'}}{4\varepsilon}\delta(\tau)D_{\max}\right),$$

the bound in (5.18) follows for $\zeta^\tau_{\mathbf{n}'} = \beta^\tau_{\mathbf{n}'}\beta^\tau$. □

Proof (of Proposition 76). We present a proof by induction on $n \geq 0$. For $n = 0$, the bound in (5.17) follows from Lemma 77 by replacing $u^\tau_{\mathbf{n}'}$ by ψ_0. In particular, for this special case the constants ζ^τ_0 and θ^τ_0 do not depend on either ε or τ and thus we could also write ζ_0 and θ_0. Now, let $n > 1$ and assume that the bound in (5.17) holds for $n-1$. The recursion formula (5.4) yields

$$|c^{n,\tau}_{\mathbf{n}}| \leq \sum_{\mathbf{n}'\in\Gamma} |c^{n-1,\tau}_{\mathbf{n}'}||c_{\mathbf{n}}(u^\tau_{\mathbf{n}'})| \quad \text{for all } \mathbf{n}\in\Gamma,$$

where the factor $|c^{n-1,\tau}_{\mathbf{n}'}|$ can be estimated by the induction hypothesis and the second factor $|c_{\mathbf{n}}(u^\tau_{\mathbf{n}'})|$ by Lemma 77. This implies that we find constants $\zeta^\tau_{n-1}, \theta^\tau_{n-1} > 0$ and $\zeta^\tau_{\mathbf{n}'}, \theta^\tau_{\mathbf{n}'} > 0$ such that

$$\begin{aligned}|c^{n-1,\tau}_{\mathbf{n}'}| &\leq \zeta^\tau_{n-1}\exp\left(-\frac{\theta^\tau_{n-1}}{8\varepsilon}\|z_{\mathbf{n}'} - z_0\|_2^2\right) \quad \text{and}\\ |c_{\mathbf{n}}(u^\tau_{\mathbf{n}'})| &\leq \zeta^\tau_{\mathbf{n}'}\exp\left(-\frac{\theta^\tau_{\mathbf{n}'}}{8\varepsilon}\|z_{\mathbf{n}} - z_{\mathbf{n}'}\|^2\right).\end{aligned}$$

Therefore, we conclude that

$$\sum_{\mathbf{n}'\in\Gamma} |c^{n-1,\tau}_{\mathbf{n}'}||c_{\mathbf{n}}(u^\tau_{\mathbf{n}'})| \leq \zeta^\tau_{n-1}\zeta^\tau \sum_{\mathbf{n}'\in\Gamma} \exp\left(-\frac{\theta^\tau_{n-1}}{8\varepsilon}\|\tilde{z}_{\mathbf{n}'}\|_2^2\right)\exp\left(-\frac{\theta^\tau}{8\varepsilon}\|\tilde{z}_{\mathbf{n}} - \tilde{z}_{\mathbf{n}'}\|_2^2\right),$$

where we have introduced

$$\zeta^\tau := \max_{\mathbf{n}'\in\Gamma}\zeta^\tau_{\mathbf{n}'} > 0 \quad \text{and} \quad \theta^\tau := \min_{\mathbf{n}'\in\Gamma}\theta^\tau_{\mathbf{n}'} > 0,$$

as well as the shifted grid points $\tilde{z}_{\mathbf{n}} := z_{\mathbf{n}} - z_0$. In Appendix 7.6 we show that there exists a positive constant $C > 0$, depending on $\theta^\tau_{n-1}, \theta^\tau, \varepsilon$ and the phase space grid, such that for all components $j = 1, \ldots, 2d$ we have

$$\begin{aligned}\sum_{\mathbf{n}'\in\Gamma} \exp\left(-\frac{\theta^\tau_{n-1}}{8\varepsilon}\left(z^{(j)}_{\mathbf{n}}\right)^2\right)\exp\left(-\frac{\theta^\tau}{8\varepsilon}\left(\tilde{z}^{(j)}_{\mathbf{n}} - \tilde{z}^{(j)}_{\mathbf{n}'}\right)^2\right) \\ \leq C\exp\left(-\frac{1}{8\varepsilon}\frac{\theta^\tau_{n-1}\theta^\tau}{\theta^\tau_{n-1}+\theta^\tau}\left(\tilde{z}^{(j)}_{\mathbf{n}}\right)^2\right).\end{aligned}$$

Consequently, using the definition of the shifted grid $\tilde{z}_{\mathbf{n}} = z_{\mathbf{n}} - z_0$, we finally get

$$\begin{aligned}\sum_{\mathbf{n}'\in\Gamma} \exp\left(-\frac{\theta^\tau_{n-1}}{8\varepsilon}\|\tilde{z}_{\mathbf{n}'}\|_2^2\right)\exp\left(-\frac{\theta^\tau}{8\varepsilon}\|\tilde{z}_{\mathbf{n}} - \tilde{z}_{\mathbf{n}'}\|_2^2\right) \\ \leq C^{2d}\exp\left(-\frac{1}{8\varepsilon}\frac{\theta^\tau_{n-1}\theta^\tau}{\theta^\tau_{n-1}+\theta^\tau}\|z_{\mathbf{n}} - z_0\|_2^2\right),\end{aligned}$$

which proves the bound in (5.17) for

$$\zeta^\tau_n = \zeta^\tau_{n-1}\zeta^\tau C^{2d} \quad \text{and} \quad \theta^\tau_n = \frac{\theta^\tau_{n-1}\theta^\tau}{\theta^\tau_{n-1}+\theta^\tau}.$$

□

The above proposition provides a bound for the magnitude of the coefficients $|c^{n,\tau}_{\mathbf{n}}|$. Together with the error representation for a single TSTG step in Proposition 75, we are now ready to present the overall error representation for the concatenation.

5.3.4 Global error estimate for the concatenation

From Proposition 75 we learn that the total error of a single TSTG propagation step can be decomposed into a time and a spatial component. In particular, the time error consists of the error for the thawed Gaussian approximation of order $\mathcal{O}(\sqrt{\varepsilon})$ and the error for the numerical integration of order $\mathcal{O}(h^s_\tau/\varepsilon)$, while the spatial error consists of the error for the approximation of the initial datum ψ_0 in $\mathcal{V}_\Gamma$ and the error for re-expansion of the time-evolved approximation $u^\tau_{\mathbf{n}}$ in $\mathcal{V}_\Gamma$. Our final result generalises this result for the concatenation of $n > 1$ TSTG steps:

Theorem 78 (Global error estimate of the TSTG method)**.**
Under the hypotheses of Proposition 75, there exists a positive constant $C > 0$ such that the global error of the TSTG propagation method with $n \geq 1$ concatenated steps at time $t_n = n\tau$ can be bounded as

$$\|\psi(t_n) - \sum_{\mathbf{n}\in\Gamma} c^{n,\tau}_{\mathbf{n}}\, g_{\mathbf{n},0}\|_{L^2(\Lambda_q)} \leq Ct_n\left(\frac{h^s_\tau}{\varepsilon} + \sqrt{\varepsilon}\right) + E^{n,\tau}, \tag{5.19}$$

where $E^{n,\tau} > 0$ denotes the total spatial discretisation error

$$E^{n,\tau} = E_{wp}(\psi_0) + Cn \cdot \max_{\mathbf{n}\in\Gamma} E_{wp}(u^\tau_{\mathbf{n}}).$$

Proof. Again, let $\| \bullet \|$ denote the L^2-norm on Λ_q. For $n \geq 1$ we define

$$e_{n,\tau} := \|\psi(n\tau) - \mathcal{A}_\Gamma^* \left(\mathcal{R}_\Gamma^\tau\right)^n \mathcal{A}_\Gamma \psi_0\| = \|\psi(n\tau) - \psi^{n,\tau}\|.$$

Using that $U(\tau)$ is unitary, we obtain the recursion

$$\begin{aligned} e_{n+1,\tau} &= \|U(\tau)\psi(n\tau) - \psi^{n+1,\tau}\| = \|U(\tau)\big(\psi(n\tau) - \psi^{n,\tau} + \psi^{n,\tau}\big) - \psi^{n+1,\tau}\| \\ &\leq \|\psi(n\tau) - \psi^{n,\tau}\| + \|U(\tau)\psi^{n,\tau} - \psi^{n+1,\tau}\| = e_{n,\tau} + \|U(\tau)\psi^{n,\tau} - \psi^{n+1,\tau}\|, \end{aligned}$$

where the second summand is the local error of the nth step. Hence, the global error $e_{n,\tau}$ after n steps can be expressed in terms of the local errors as

$$e_{n,\tau} = e_{1,\tau} + \sum_{l=1}^{n-1} \|U(\tau)\psi^{l,\tau} - \psi^{l+1,\tau}\|.$$

We note that $e_{1,\tau}$ is the error after a single propagation step in Proposition 75. Furthermore, for $1 \leq l \leq n-1$, the definition of the coefficients $c_{\mathbf{n}}^{l,\tau}$ in (5.4) yields

$$\begin{aligned} \|U(\tau)\psi^{l,\tau} - \psi^{l+1,\tau}\| &= \| \sum_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}^{l,\tau}\, g_{\mathbf{n}}(\tau) - \psi^{l+1,\tau} \| \\ &\leq \sum_{\mathbf{n}\in\Gamma} |c_{\mathbf{n}}^{l,\tau}| \Big(\|\, g_{\mathbf{n}}(\tau) - u_{\mathbf{n}}^\tau\| + \|u_{\mathbf{n}}^\tau - \sum_{\mathbf{n}'\in\Gamma} c_{\mathbf{n}'}(u_{\mathbf{n}}^\tau)\, g_{\mathbf{n}',0}\| \Big) \\ &\leq \sum_{\mathbf{n}\in\Gamma} |c_{\mathbf{n}}^{l,\tau}| \Big(E_{\mathbf{n}}^\tau + E_{wp}(u_{\mathbf{n}}^\tau) \Big). \end{aligned}$$

Using once more the bound for $E_{\mathbf{n}}^\tau$ in (5.14) and defining

$$c^{max} := \max\left(1, \max_{\mathbf{n}\in\Gamma} c_{\mathbf{n}}\right) \quad \text{as well as} \quad E_{wp}^{max} := \max_{\mathbf{n}\in\Gamma} E_{wp}(u_{\mathbf{n}}^\tau),$$

we therefore conclude that

$$\|U(\tau)\psi^{l,\tau} - \psi^{l+1,\tau}\| \leq c^{max} \left(\tau \left(\frac{h_\tau^s}{\varepsilon} + \sqrt{\varepsilon} \right) + E_{wp}^{max} \right) \|c_{\mathbf{n}}^{l,\tau}\|_1.$$

Consequently, the bound in (5.19) follows for the constant

$$C = c^{max} \max_{l=0,\dots,n-1} \|c_{\mathbf{n}}^{l,\tau}\|_1.$$

□

Theorem 78 proves that the error of the TSTG method increases linearly with the number n of concatenations, where the corresponding constant depends on the errors arising from (i) the discretisation of the wave packet transform, (ii) the thawed Gaussian propagation of the basis functions, and (iii) the integration of the equations of motion. In the following numerical experiments we employ a practical error bound based on a direct calculation of

$$\mathrm{err}_\Gamma^{l,\tau} := \sum_{\mathbf{n}\in\Gamma} |c_{\mathbf{n}}^{l,\tau}| \Big(E_{\mathbf{n}}^\tau + E_{wp}(u_{\mathbf{n}}^\tau) \Big) \tag{5.20}$$

for all $l = 1, 2, \dots, n-1$, using the split-step Fourier method as a reference solver for the propagation of the basis functions. We present a detailed description of the reference solver in Appendix 7.7.

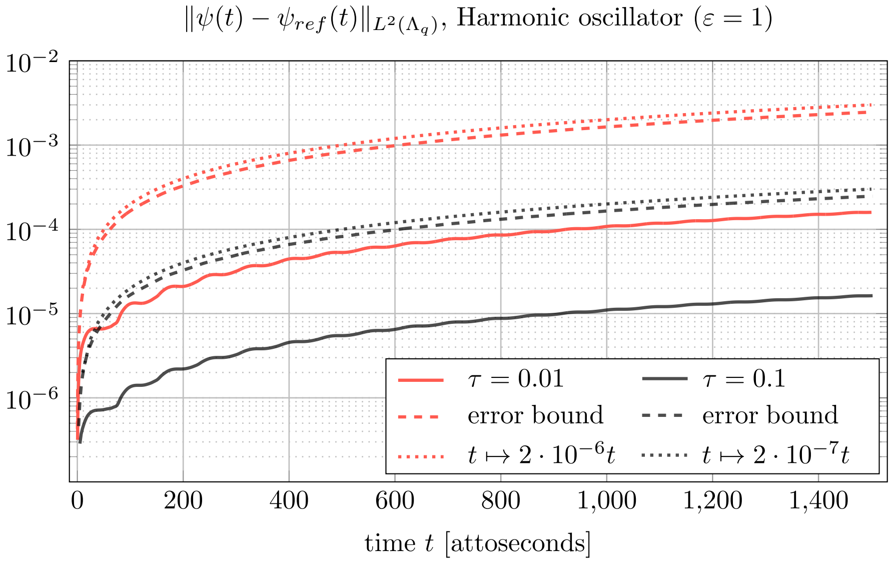


Figure 5.1: Evolution of the L^2-error between the TSTG method and the analytical solution ψ_{ref} for the harmonic oscillator ($\varepsilon = 1$). The errors increase linearly with the number of concatenated steps (logarithmic scaling of the y-axis). The time range covers about 15 oscillations of the harmonic oscillator.

5.4 Numerical experiments

We demonstrate the capabilities of the TSTG method with two numerical experiments. First, we test the method by calculating the full wave function of the one-dimensional harmonic oscillator for different propagation times τ and step sizes h_τ. We then reproduce the results of Kong *et al.* for a one-dimensional double-well potential.

Remark 79. *So far, only non-variationally evolving Gaussians have been used for the TSTG method, and the following experiments are the first to compare non-variational with variational Gaussians. Although we only present one-dimensional experiments to support our theoretical results for the error representation, we note that the capabilities of the TSTG method have already been demonstrated for multidimensional systems by Kong* et al.*, see [KMB16, Results].*

5.4.1 One-dimensional harmonic oscillator

We consider the quantum harmonic oscillator corresponding to the quadratic potential $V(x) = x^2/2$. As initial data we choose the Gaussian wave packet $\psi_0 = g_{z_0}^{\gamma_0,\varepsilon}$ with $z_0 = (1, 0), \gamma_0 = i$ and $\varepsilon \in \{0.1, 1\}$. In particular, the analytic solution is known to be, see [Hag98, Theorem 2.5],

$$\psi_{ref}(t) = (\pi\varepsilon)^{-1/4} \exp\left(-\frac{1}{2\varepsilon}\big(x - q(t)\big)^2 + \frac{i}{\varepsilon}p(t)\big(x - q(t)\big) + \frac{i}{\varepsilon}S(t) - \frac{i}{2}t\right),$$

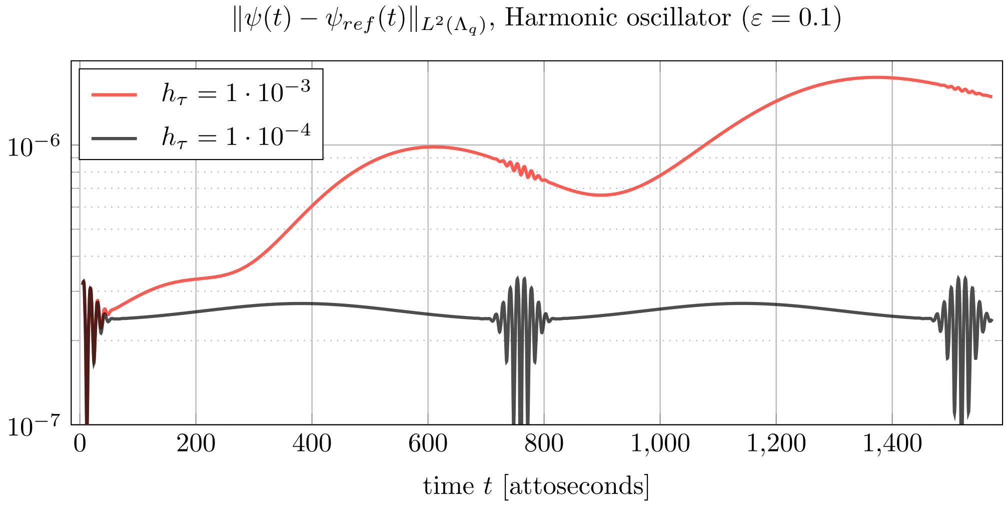


Figure 5.2: Evolution of the L^2-error between the TSTG method and the analytical solution for the harmonic oscillator ($\varepsilon = 0.1$) for different step sizes h_τ. The error increases faster for the coarser time grid (red curve).

where $q(t), p(t)$ and $S(t)$ are given by

$$q(t) = q_0 \cos(t) + p_0 \sin(t), \quad p(t) = p_0 \cos(t) - q_0 \sin(t),$$
$$S(t) = -\frac{1}{2} \sin(t)\Big(\big(q_0^2 - p_0^2\big)\cos(t) + 2q_0p_0 \sin(t)\Big).$$

The discretisation of the wave packet transform according to Section 3.2 was based on the phase space box $\Lambda = [-8, 8] \times [-8\pi, 8\pi]$, where we used 64 uniform grid points in position space, 32 uniform grid points in momentum space and the width parameter $\gamma = 4i$ for the basis functions. The propagation of the basis functions was implemented with the second-order variational splitting integrator in [LL20, Section 7.5].

Figure 5.1 shows the L^2-error between the TSTG method and the analytical solution on the spatial interval $\Lambda_q = [-8, 8]$ for $\varepsilon = 1$ and two choices of $\tau = 0.1$ (red) and $\tau = 0.01$ (black). The step size for the time integration was $h_\tau = 1 \cdot 10^{-3}$. The dashed lines indicate the error bound of Theorem 78 based on a direct evaluation of the error bounds $\mathrm{err}_\Gamma^{l,\tau}$ in (5.20), where we have again used the analytical solution to calculate the errors $E_{\mathbf{n}}^\tau$. Due to the logarithmic scaling of the y-axis, we have added the linear functions $t \mapsto 2 \cdot 10^{-6}t$ (dashed red) and $t \mapsto 2 \cdot 10^{-7}t$ (dashed black) to check whether the error actually increases linearly with the number of TSTG steps. Note that for $\tau = 0.01$ we need 10 times the number of concatenations compared to $\tau = 0.1$ and therefore the slopes of the red and black lines differ by a factor of 10. To keep the number of TSTG steps small, we note that τ should be chosen as large as possible. Furthermore, Figure 5.2 shows the L^2-error for $\varepsilon = 0.1$. The calculations were based on 128 uniform grid points in position and momentum space and $\tau = 0.01$ for two step sizes $h_\tau = 1 \cdot 10^{-3}$ and $h_\tau = 1 \cdot 10^{-4}$. For the smaller choice of h_τ (red curve), we see that the error increases faster, which is consistent with our theoretical result in Proposition 74.

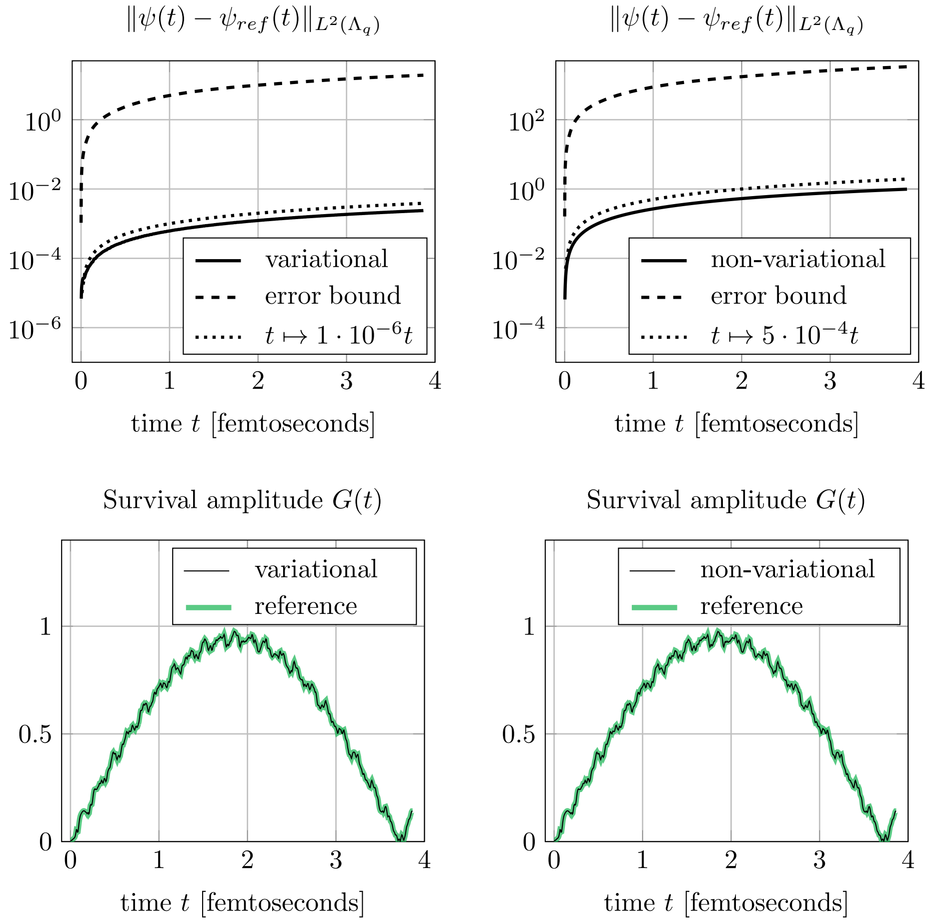


Figure 5.3: Comparison between variational (left) and non-variational Gaussians (right) for the double-well potential. Both variants of the TSTG method show good agreement as compared to benchmark results based on the Fourier method. Top: Error for the full wave function; Bottom: Survival amplitude.

5.4.2 One-dimensional double-well potential

In this experiment we follow the presentation of Kong *et al.* in [KMB16, Results] and use the one-dimensional double-well potential

$$V(x) = \frac{x^4}{16\eta} - \frac{x^2}{2}, \qquad \eta = 1.3544,$$

together with the initial datum ψ_0 defined in (3.41) for $\varepsilon = 1$, which is a model for quantum tunnelling. A short calculation shows that the total energy is given by

$$\langle \psi_0 \mid H\psi_0 \rangle = -\frac{64\eta^2 - 48\eta - 3}{64\eta} \approx -0.57. \tag{5.21}$$

As for the harmonic oscillator, we used the phase space box $\Lambda = [-8, 8] \times [-8\pi, 8\pi]$, but this time with 64 equally spaced grid points in both position and momentum space and

$\gamma = 4i$ for the width of the basis functions. In addition to the variational Gaussians, we implemented the non-variational Gaussians based on the Störmer–Verlet method, see e.g. [HLW03], which have also been used by Kong *et al.*. For the reference solution, the split-step Fourier method was implemented with 256 points in the range $\Lambda_q = [-8, 8]$ and the time increment $\tau = 0.01$. The step size $h_\tau = 0.001$ was used for both the variational and the non-variational Gaussian propagation. The top panels of Figure 5.3 show the L^2-error between the TSTG method and the reference solution for the variational (left) and the non-variational Gaussian (right) together with the error bounds of Theorem 78 (dashed lines). In the lower panels, the TSTG method is compared with the reference solution for the so-called "survival amplitude" (overlap between $\psi(x,t)$ and the mirror image of the initial state on the opposite side of the double well), which is defined by

$$G(t) := \int_{-\infty}^{\infty} \overline{\psi_0(-x)}\psi(x,t)\,\mathrm{d}x$$

and is a measure for the tunnelling amplitude. The results in Figure 5.3 show that the TSTG method accurately reproduced the full wave function and the survival amplitude. The experiments also show that the L^2-error increases linearly (approx. as $t \mapsto 10^{-6}t$ for the variable Gaussians), while for the non-variational Gaussians the rate is larger (approx. $t \mapsto 5 \cdot 10^{-4}t$). Furthermore, in Figure 5.4 we compare the TSTG method with the reference solution for the energy expected values (top) and their relative errors (bottom). For better illustration, we have only plotted the time range of the last 4,000 of a total of 16,000 propagation steps. It can be seen that the expected values of the reference solution are well approximated even after long running times. In particular, the slopes of the blue lines in the lower panel show that the error for the non-variational Gaussians (upper curves) increases faster.

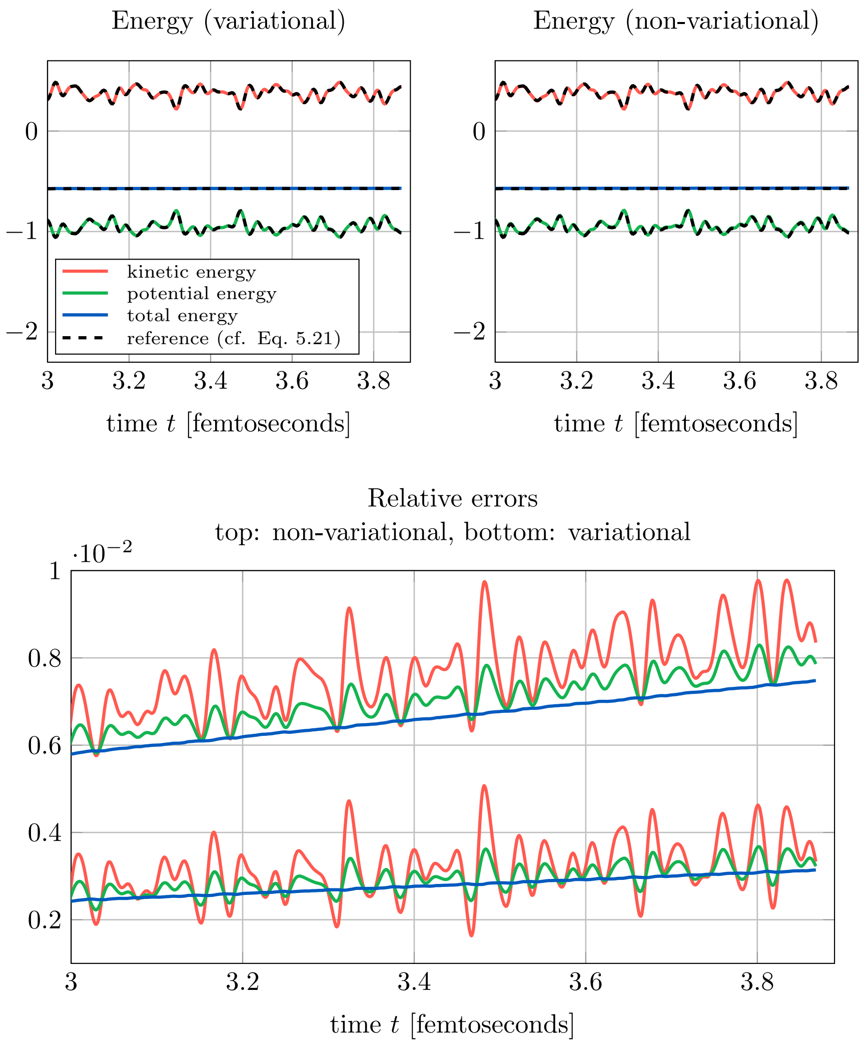


Figure 5.4: Evolution of energy expected values (top) and relative errors (bottom) for the variational Gaussians and the non-variational Gaussians between 12,000 and 16,000 TSTG propagation steps.

5.5 Summary of this chapter

In the previous sections we have derived an error representation for the time-sliced thawed Gaussian propagation method that combines representations of Gaussian wave packets based on the discretisation of the FBI formula with thawed Gaussian approximations for the propagation of the basis functions. To provide a mathematical formulation of the TSTG method, we combined the quadrature-based analysis and synthesis operators $\mathcal{A}_\Gamma$ and $\mathcal{S}_\Gamma$ with the reinitialisation operator $\mathcal{R}_\Gamma^\tau$, which allow to write the approximate solution at time $t_n = n\tau$ as

$$\psi(t_n) \approx \psi_\Gamma^{n,\tau} = \mathcal{S}_\Gamma \left(\mathcal{R}_\Gamma^\tau\right)^n \mathcal{A}_\Gamma \psi_0.$$

The algorithm was implemented in MATLAB to underline our theoretical results and to show that the global error of the method increases linearly with the number of time steps, independent of the thawed Gaussian method (variational or non-variational) and the order of the underlying time integrator.

5.6 Suggestions for further research

The TSTG method avoids multidimensional numerical quadrature by using an explicit formula for calculating the representation coefficients of the underlying Gaussian wave packet transform. So far, only Gaussian basis functions have been considered and the coefficients have been calculated using the formula for inner products in Lemma 7. However, other basis functions could also be used. For example, Hagedorn's wave packets provide a natural extension. With the additional polynomial prefactor, the accuracy of the approximation can be further improved, while the inner products can still be calculated analytically. Moreover, with regard to the numerical implementation of the method, the further use of the error estimator in (5.20) could be profitably employed. Future research could address the derivation of a practical a posteriori error bound that can be used to implement the TSTG method with adaptive step sizes or adaptive mesh refinements. To make the TSTG method applicable to high-dimensional systems, the calculation of the update coefficients given by the tensor-valued matrix-vector product according to (5.4) could be performed using low-rank approximations. We will describe this idea in more detail in the next chapter.

6 Multidimensional quantum dynamics with tensor trains

In the previous chapter we analysed the TSTG method and showed that the underlying decomposition of Gaussian wave packets rely on discretisations of the FBI formula based on fully tensorised grids in phase space, which limits the numerical implementation, because the number of basis functions grows exponentially with the dimension. In Section 3.2.3 we have already pointed out that sparse grids and Monte Carlo methods can overcome the curse of dimensionality to some extent, but for the TSTG method these approaches can only be used for the representation of the initial wave function, since the re-expansion of the time-evolved Gaussian basis functions is performed on a time-independent uniform grid (we worked with a time-independent approximation space). It is important to understand that in the TSTG method we do not face the problem of high-dimensional quadrature, but that the size of the coefficient tensors to represent the wave functions grows exponentially. In practical applications, we therefore have to work with high-dimensional tensors, where it is usually not even possible to store all elements explicitly. The following sections deal with tensor-train (TT) decompositions, which are a special variant of low-rank tensor decompositions that can overcome the curse and are currently used in many different fields, in particular for the computation of multidimensional quantum dynamics. In Section 6.1 we introduce the TT format and recall important results such as storage, existence and arithmetic operations. We then present results on the tensor-train Chebyshev (TTC) method recently published in the joint paper [SBGB22] with M. B. Soley, A. A. Gorodetsky and V. S. Batista, showing that the TT format can be used for high-dimensional quantum dynamics simulations.

6.1 The tensor-train format

For positive integers $n_1, \dots, n_d$, $d > 1$, we consider a multivariate function

$$C\colon [n_1] \times \cdots \times [n_d] \to \mathbb{R}, \ (i_1, \dots, i_d) \mapsto C(i_1, \dots, i_d),$$

where the sets $[n_k] \subset \mathbb{N}$ are defined for all $k = 1, \dots, d$ as

$$[n_k] := \{1, \dots, n_k\}.$$

Multivariate functions as defined above on a finite multi-index set $\Gamma = [n_1] \times \cdots \times [n_d]$ occur in many numerical applications. For instance, if $f\colon [0,1]^d \to \mathbb{R}$ is a given real-valued function and for $N \geq 1$ we consider the uniform grid points

$$x_i = ih, \quad i = 1, \dots, N, \quad h = \frac{1}{N},$$

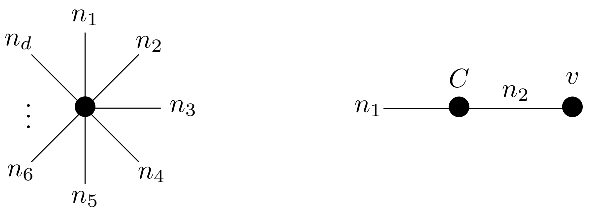


Figure 6.1: Graphical notation of tensors and tensor operations. Left: Tensor of order d. Each "arm" corresponds to an input variable; Right: Matrix-vector product. The summation over the auxiliary variable is visualised by the connecting arm between C and v.

one could think of the following natural connection:

$$C(i_1, \ldots, i_d) = f(x_{i_1}, \ldots, x_{i_d}) \quad \text{for all } i_k \in [N],\ k = 1, \ldots, d.$$

In particular, for the special case $d = 2$ we can identify C with an element of $\mathbb{R}^{n_1 \times n_2}$, where the image of (i_1, i_2) is given by the matrix element c_{i_1,i_2}. Accordingly, in the general case $d \geq 2$ we can identify C with an element of $\mathbb{R}^{n_1 \times \cdots \times n_d}$, usually called "tensor of order d and size $n_1 \times \cdots \times n_d$". Tensors of order d can be represented in various ways, e.g. as a single point with d "arms", see Figure 6.1 (left). Using the convention that connected compatible arms represent summation over the corresponding tensor indices, the graphical notation also allows visualisation of basic tensor operations. For example, the right-hand side of Figure 6.1 shows a matrix-vector product

$$(Cv)_{i_1} = \sum_{\alpha_1=1}^{n_2} c_{i_1,\alpha_1} v_{\alpha_1}, \quad i_1 \in [n_1].$$

As a natural generalisation of matrices, the analysis of tensors is not only of great interest from a theoretical point of view, but also for practical algorithms. However, it is known that tensors in high dimensions such as $d = 100$ or $d = 1000$ cannot be used explicitly on computers, which leads to the question of whether and how it is possible to represent multidimensional tensors by a much smaller number of parameters. Low rank tensor decompositions, which can be seen as a generalisation of the truncated singular value decomposition (SVD) for matrices, are an efficient tool for reducing the number of parameters. Many different techniques have been developed in the last decades and we refer to [GKT13] for an overview of the existing methods.

In the following, we focus on the approximation of tensors by tensors in tensor-train format, which are also called "matrix-product states" and represent a special form of the so-called "hierarchical tensor format". A detailed mathematical introduction to tensor methods, describing in particular the relationship between the various formats and their origins, can be found in [Hac14]. Let us now introduce the TT format:

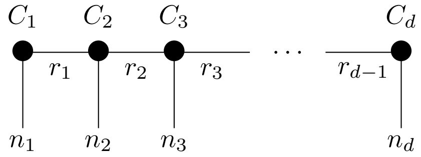


Figure 6.2: Graphical notation of a tensor in tensor-train format

Definition 80. *For positive integers $n_1, \dots, n_d$, let $C \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be a order-d-tensor. A decomposition of the form*

$$C(i_1, \dots, i_d) = C_1[i_1]C_2[i_2] \cdots C_{d-1}[i_{d-1}]C_d[i_d] \quad \text{for all } i_k \in [n_k],\ k = 1, \dots, d,$$

where $C_1[i_1] \in \mathbb{R}^{1 \times r_1}$ is a row vector, $C_k[i_k] \in \mathbb{R}^{r_{k-1} \times r_k}, k = 2, \dots, d-1$ are matrices and $C_d[i_d] \in \mathbb{R}^{r_{d-1}}$ is column vector, is called tensor-train (TT) decomposition. *In particular, the numbers $r_1, \dots, r_{d-1} \geq 1$ are called the* compression ranks.

By identifying vectors and matrices as order-2 and order-3 tensors, respectively, we see that tensor-train decompositions can also be written as follows, see [HRS12, Equation 4]:

$$C(i_1, \dots, i_d) = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{d-1}=1}^{r_{d-1}} C_1(i_1, \alpha_1) \left(\prod_{k=2}^{d-1} C_k(\alpha_{k-1}, i_k, \alpha_k) \right) C_d(\alpha_{d-1}, i_d).$$

Since the corresponding graph representation resembles a train structure, see Figure 6.2, Oseledets introduced the name "tensor train", see [Ose11], which seems to be the best known description for this type of decomposition today. However, this type of tensor decomposition was introduced earlier in the quantum-physics community, but under the name "matrix product format", see e.g. [Hac14, Section 9]. Indeed, the individual tensor elements $C(i_1, \dots, i_d)$ are obtained by products of matrices, and if each component is represented as a fibre of matrices of the same dimension, we obtain an alternative picture of the TT decomposition:

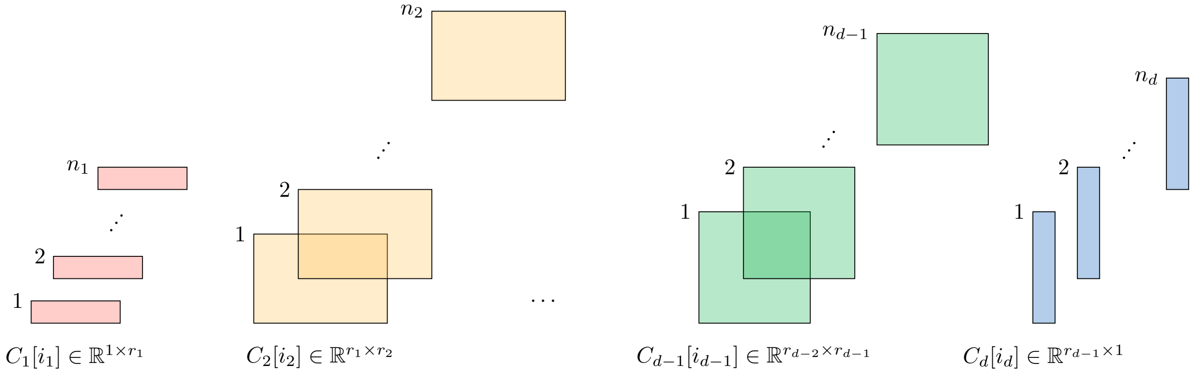

n_1
$\vdots$
2
1
$C_1[i_1] \in \mathbb{R}^{1 \times r_1}$
n_2
2
1
$C_2[i_2] \in \mathbb{R}^{r_1 \times r_2}$
$\cdots$
n_{d-1}
2
1
$C_{d-1}[i_{d-1}] \in \mathbb{R}^{r_{d-2} \times r_{d-1}}$
n_d
2
1
$C_d[i_d] \in \mathbb{R}^{r_{d-1} \times 1}$

Figure 6.3: Alternative visualisation of the TT format. Each entry of the original tensor corresponds to a product of matrices.

The advantages of TT decompositions can be easily illustrated with a short example, which is a special case of Example 4.2 in [KK18]. Let us consider the tensor $C \in \mathbb{R}^{3\times 4\times 5}$ of size $3 \cdot 4 \cdot 5 = 60$ defined by

$$C(i_1, i_2, i_3) = i_1 + i_2 + i_3, \quad i_1 \in [3], i_2 \in [4], i_3 \in [5].$$

Since we have

$$i_1 + i_2 + i_3 = \begin{pmatrix} i_1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ i_2 + i_3 \end{pmatrix} = \begin{pmatrix} i_1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ i_2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ i_3 \end{pmatrix},$$

we can define the matrices

$$C_1[i_1] := \begin{pmatrix} i_1 & 1 \end{pmatrix} \in \mathbb{R}^{1\times 2}, \quad C_2[i_2] := \begin{pmatrix} 1 & 0 \\ i_2 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \quad C_3[i_3] := \begin{pmatrix} 1 \\ i_3 \end{pmatrix} \in \mathbb{R}^{2\times 1},$$

to obtain the following TT decomposition:

$$C(i_1, i_2, i_3) = C_1[i_1]C_2[i_2]C_3[i_3].$$

Note that $n_1 = 3, n_2 = 4, n_3 = 5$ and $r_1 = r_2 = 2$. The total number of elements to store the matrices $C_1[i_1], C_2[i_2]$ and $C_3[i_3]$ is given by 32, while the size of C is almost twice as large! This simple example illustrates very clearly that the TT decomposition can be used to reduce memory requirements. If we set $r_0 = r_d = 1$, $R := \max\{r_1, \dots, r_{d-1}\}$ and $N := \max\{n_1, \dots, n_d\}$, the total number M of matrix entries for the TT decomposition in Definition 80 is bounded as follows,

$$M = \sum_{k=1}^{d} n_k \cdot (r_{k-1} \cdot r_k) \leq d \cdot N \cdot R^2, \tag{6.1}$$

which has to be compared with the size of the tensor given by $n_1 \times \cdots \times n_d$. Since the upper bound in (6.1) depends only linearly on the dimension d, TT decompositions are often said to break the curse of dimensionality. Indeed, various numerical experiments show that a large class of tensors encountered in practical applications can be handled efficiently in TT format, allowing high-dimensional computations that could not be performed on computers if tensors were implemented directly as high-dimensional arrays.

Remark 81. *We already mentioned that tensors are closely related to grid functions. In the one-dimensional case, for example, the function values on a grid can be stored in a row-vector. It has proven advantageous to convert such vectors into tensors with a special shape and then decompose the resulting tensors in TT format. One way to reshape vectors is "q-adic unfolding", where the indices of the resulting tensor correspond to the q-adic representation of the original indices. More precisely, a vector $v \in \mathbb{R}^N$ of length $N = q^L$ is transformed into an order-L tensor in $\mathbb{R}^{q\times\cdots\times q}$, where the jth element of v is represented by the index $(i_1, \dots, i_d)$ of the reshaped tensor via*

$$j - 1 = \sum_{k=1}^{L} (i_k - 1)q^{k-1}.$$

Such representations are called "Quantics Tensor Trains" (QTTs) and it has been shown that in many applications the ranks of the components are small, making QTTs a powerful tool for high-dimensional problems, see [KK18, Chapter 4.2] and references therein. In particular, we would like to mention that QTTs have been successfully applied to global optimisation problems in the joint paper [SBB21] with M. B. Soley and V. S. Batista, essentially by implementing the power iteration known from numerical linear algebra with QTTs for $q = 2$. The corresponding algorithm is known as the "Iterative Power Algorithm" (IPA), and it has been shown that IPA can be used to approximate solutions to high-dimensional optimisation problems such as those encountered in molecular and electronic structure calculations.

Before presenting numerical results related to quantum dynamics, let us recall some important results of tensor trains.

6.1.1 Existence and uniqueness of TT decompositions

The decomposition of a given tensor in TT format depends crucially on the compression ranks $r_1, \dots, r_{d-1}$. One of the first questions is therefore for which choice of ranks we can guarantee the existence of a TT decomposition. To answer this question, we need the following definition (see also [HRS12, Section 2.5]):

Definition 82. *Let $C \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be a tensor of order d and let*

$$\nu_k := \prod_{s=1}^{k} n_s \quad \textit{and} \quad \mu_k := \prod_{s=k+1}^{d} n_s, \quad k = 1, \dots, d-1. \tag{6.2}$$

The <u>kth canonical matrix</u> $C_k \in \mathbb{R}^{\nu_k \times \mu_k}$ of C is defined by

$$C_k\big((i_1, \dots, i_k); (i_{k+1}, \dots, i_d)\big) := C(i_1, \dots, i_d), \quad i_s \in [n_s],\ s = 1, \dots, d,$$

where the indices $(i_1, \dots, i_k)$ enumerate the rows and the indices $(i_{k+1}, \dots, i_d)$ the columns of the matrix C_k in colexicographical order, i.e. in column-major order. Moreover, the rank of the kth unfolding matrix, in the following denoted by s_k, is called the <u>kth separation rank</u> of C.

A short example will illustrate the definition of unfolding matrices. Therefore, let us consider the following order-3 tensor $C \in \mathbb{R}^{2 \times 2 \times 2}$, defined by

$$C(i_1, i_2, i_3) := i_1 \cdot 10^2 + i_2 \cdot 10 + i_3, \quad i_1, i_2, i_3 \in [2].$$

Then, the unfolding matrices $C_1 \in \mathbb{R}^{2 \times 4}$ and $C_2 \in \mathbb{R}^{4 \times 2}$ are given by

$$C_1 = \begin{pmatrix} 111 & 121 & 112 & 122 \\ 211 & 221 & 212 & 222 \end{pmatrix} \quad \text{and} \quad C_2 = \begin{pmatrix} 111 & 112 \\ 211 & 212 \\ 121 & 122 \\ 221 & 222 \end{pmatrix}.$$

In particular, in MATLAB the unfolding matrix C_k can be generated with the command

$$\texttt{reshape}(C, \nu_k, \mu_k),$$

where the positive integers ν_k and μ_k are defined according to (6.2).

We are now ready to prove the existence of TT decompositions. The following result was taken from [Ose11, Theorem 2.1].

Proposition 83. *Let C be an arbitrary tensor of order d with separation ranks $s_k \geq 1$. There exists a TT decomposition with compression ranks r_k not higher than s_k, that is, $r_k \leq s_k$ for all $k = 1, \dots, d-1$.*

For the proof we refer to [Ose11, Theorem 2.1]. □

Oseledets' proof not only shows that a TT decomposition exists, but also allows the construction of a practical algorithm for its computation. Since the resulting algorithm relies on successive SVDs of auxiliary matrices to compute the tensor-train components, this algorithm is called the "TT-SVD algorithm". It should be noted that the separation ranks of a tensor are not invariant under permutations of the indices, which has the consequence that the ranks (and thus the storage) increase if an unfavourable order is chosen. However, it is possible to approximate a given tensor by a tensor train with fixed ranks by replacing the SVDs in the TT-SVD algorithm with best-rank approximations; a result for the corresponding error can be found in [Ose11, Theorem 2.2].

Remark 84. *In many situations it happens that a tensor is in TT format but the ranks of the components are too large (see Section 6.1.2 for examples). Of course, the ranks can be reduced with the variant of the SVD algorithm described above, but as it turns out, this is not the best way. Instead, there is a method that directly uses the TT format. This procedure is called "rounding", which is important for numerical applications because without rounding the ranks would explode after repeated calculations. In [Ose11, Section 3], Oseledets presents an algorithm that requires $\mathcal{O}(dNR^3)$ operations.*

The next question that arises is whether the TT decompositions are unique. Since the TT format is based on matrix products, it is easy to see that the TT components cannot be uniquely determined, since, for example, for an invertible matrix $G \in \mathbb{R}^{r_k \times r_k}$ it follows that

$$C_k[i_k]C_{k+1}[i_{k+1}] = C_k[i_k]GG^{-1}C_{k+1}[i_{k+1}],$$

which implies that the components $C_k[i_k]$ and $C_{k+1}[i_{k+1}]$ can be replaced by

$$\tilde{C}_k[i_k] = C_k[i_k]G \quad \text{and} \quad \tilde{C}_{k+1}[i_{k+1}] = G^{-1}C_{k+1}[i_{k+1}].$$

Hence, it makes sense to use further conditions that uniquely determine the components. The next definition introduces special unfolding matrices that can be used to standardise TT decompositions (see also [HRS12, Section 2.4]):

Definition 85. *The* left unfolding *of a TT component* $C_k \in \mathbb{R}^{r_{k-1}\times n_k \times r_k}$, $k = 2, \dots, d-1$, *is denoted by* $C_k^L \in \mathbb{R}^{(r_{k-1}n_k)\times r_k}$ *and defined by*

$$C_k^L((\alpha_{k-1}, i_k); r_k) := C_k(\alpha_{k-1}, i_k, \alpha_k),$$

where the indices (α_{k-1}, i_k) *enumerate the rows and the index* α_k *the columns of the matrix* C_k^L *in colexicographical order. The rank* r_k^L *of the left unfolding is called the* left rank *of the component* C_k*. The* right unfolding $C_k^R \in \mathbb{R}^{r_{k-1}\times(n_k r_k)}$ *and the corresponding* right rank r_k^R *are defined analogously. Furthermore, a TT decomposition is called* minimal *if* $r_k^L = r_k$ *and* $r_k^R = r_{k-1}$.

We are now ready to prove the uniqueness of TT decompositions. The following result was taken from [HRS12, Theorem 1].

Proposition 86. *There is exactly one rank vector* $r = (r_1, \dots, r_{d-1})$ *such that a given order-*d *tensor* C *admits for a minimal TT decomposition and if* $s = (s_1, \dots, s_{d-1})$ *denotes the (unique) separation rank of* C*, there holds* $r = s$*. In particular, a minimal decomposition can be chosen such that the components are left-orthogonal, that is,*

$$\left(C_k^L\right)^T C_k^L = \mathrm{Id} \in \mathbb{R}^{r_k \times r_k} \quad \textit{for all } k = 1, \dots, d-1.$$

Under this condition, the decomposition is unique up to insertion of orthogonal matrices: For any two left-orthogonal minimal decompositions of C *for which*

$$C(i_1, \dots, i_d) = C_1[i_1]C_2[i_2]\cdots C_{d-1}[i_{d-1}]C_d[i_d] = D_1[i_1]D_2[i_2]\cdots D_{d-1}[i_{d-1}]D_d[i_d]$$

holds for all $i_k \in [n_k]$*, there exist orthogonal* $Q_1, \dots, Q_{d-1}, Q_k \in \mathbb{R}^{r_k\times r_k}$ *such that*

$$C_1[i_1]Q_1 = D_1[i_1], \quad Q_{d-1}^T C_d[i_d] = D_d[i_d], \quad Q_{k-1}^T C_k(i_k)Q_k = D_k(i_k).$$

For the proof we refer to [HRS12, Theorem 1]. □

The TT-SVD algorithm can be used to compute TT decompositions. Now that we have specified these decompositions with left-orthogonal unfolding matrices, we should add the fact that Oseledets' TT-SVD algorithm produces (in exact arithmetic) minimal TT decompositions with left-orthogonal components. As it can be advantageous in some applications, we also note that the algorithm can also be adapted to compute right-orthogonal or mixed (left- and right-orthogonal) components.

The TT format has many other interesting properties that are of particular interest for the development of numerical methods. For example, it can be shown that tensors in TT format of fixed rank locally form an embedded manifold in $\mathbb{R}^{n_1\times\cdots\times n_d}$, see [HRS12]. Moreover, we refer the interested reader to [GH21, Section 4], where the authors analyse TT approximation schemes for continuous functions.

6.1.2 Operations in TT format

In the previous section we saw that the TT format can be used to efficiently represent high-dimensional tensors, and once the tensors are in TT format, arithmetic operations usually have to be performed. In the following, we describe how these operations are implemented using the examples of addition, elementwise multiplication and inner products. In doing so, we summarise the description of Oseledets in [Ose11, Section 4].

Let us consider two tensors $A, B \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ of the same dimension, which are in tensor-train format, *i.e.*, $A(i_1, \ldots, i_d) = A_1[i_1] \cdots A_d[i_d], B(i_1, \ldots, i_d) = B_1[i_1] \cdots B_d[i_d]$. The sum $C = A + B$ is given by

$$\begin{aligned} C(i_1, \ldots, i_d) &= A(i_1, \ldots, i_d) + B(i_1, \ldots, i_d) \\ &= A_1[i_1] \cdots A_d[i_d] + B_1[i_1] \cdots B_d[i_d], \end{aligned}$$

and a simple calculation shows that the components of C in TT format are given by

$$C_k[i_k] = \begin{pmatrix} A_k[i_k] & 0 \\ 0 & B_k[i_k] \end{pmatrix}$$

for all $i_k \in [n_k]$, $k = 1, \ldots, d-1$, and

$$C_1[i_1] = \begin{pmatrix} A_1[i_1] & B_1[i_1] \end{pmatrix}, \quad C_d[i_d] = \begin{pmatrix} A_d[i_d] \\ B_d[i_d] \end{pmatrix}, \quad i_1 \in [n_1],\ i_d \in [n_d].$$

On the one hand, this shows that no arithmetic operations are necessary to build the TT format of the sum, since the components of C simply result from "connecting" the components of A and B in a common matrix. On the other hand, it can be seen that the size of the resulting components increases (the ranks are summed). Rounding in TT format (cf. Remark 84) can be used to avoid the rank increasing too much due to successive additions. In particular, if Oseledets' rounding algorithm is used after each addition, the total number of operations grows as $\mathcal{O}(dNR^3)$, where now R denotes the maximum rank of the components of A and B.

Another important operation is the elementwise product of tensors, also known as the "Hadamard product", denoted by $C = A \circ B$. It is given by

$$\begin{aligned} C(i_1, \ldots, i_d) &= A(i_1, \ldots, i_d) B(i_1, \ldots, i_d) \\ &= A_1[i_1] \cdots A_d[i_d] B_1[i_1] \cdots B_d[i_d], \end{aligned}$$

and a simple calculation (see e.g. [Ose11, Section 4.2]) shows that the components of C are given by

$$C_k[i_k] = A_k[i_k] \otimes B_k[i_k] \in \mathbb{R}^{r_{k-1}^{(A)} r_{k-1}^{(B)} \times r_k^{(A)} r_k^{(B)}}, \quad i_k \in [n_k],\ k = 1, \ldots, d,$$

where $\otimes$ denotes the Kronecker product of matrices and the superscripts (A) and (B) indicate that the ranks correspond to either the tensor A or B. In particular, it follows that the ranks of the components of C are given by the products of the ranks. We also note that in the special case $B(i_1, \dots, i_d) = b$ only one (freely chosen) component of A has to be multiplied by the scalar b and the others remain unchanged.

The Hadamard product $C = A \circ B$ is needed, for example, to calculate the inner product of A and B, defined by

$$\langle A \mid B \rangle := \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} A(i_1, \dots, i_d) B(i_1, \dots, i_d) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} C(i_1, \dots, i_d). \tag{6.3}$$

The computation of the sum in (6.3) is called a "multidimensional contraction". Due to the special structure of the matrices $A_k[i_k] \otimes B_k[i_k]$, the inner product can be implemented with a total of $\mathcal{O}(dNR^3)$ operations. Furthermore, the inner product algorithm can be used to calculate the Frobenius norm of a tensor, which is defined as follows:

$$\|A\|_F = \sqrt{\langle A \mid A \rangle}$$

In addition to the basic arithmetic operations mentioned above, advanced operations can also be performed in the TT format. For instance, with regard to algorithms for solving the Schrödinger equation, the Fourier transform of tensor trains is of particular interest, which makes it possible to transfer the Laplace operator into the frequency domain, where it can then be executed as a multiplication operator. Since this operation is also used in the TT Chebyshev method below, we refer interested readers to [GB17], where the authors introduce an extension of the split-step Fourier method in TT format, the so-called "tensor-train split-operator Fourier transform (TT-SOFT)" method.

Remark 87. *Like many other authors, we have treated only real-valued tensors for the sake of simplicity. However, we point out that the above results can also be extended to complex-valued tensors.*

6.2 Tensor-train Chebyshev method

As described in detail in Section 1.1, parts of the present section (Sec. 6.2) overlap to a large extent with the joint publication "Functional Tensor-Train Chebyshev Method for Multidimensional Quantum Dynamics Simulations" with M. B. Soley, A. A. Gorodetsky and V. S. Batista appeared in *Journal of Chemical Theory and Computation*, 18(1):25–36, 01 2022.

We now present the tensor-train Chebyshev (TTC) method, which is essentially a tensor-train implementation of the celebrated Chebyshev propagation scheme introduced by Tal-Ezer and Kosloff in [TK84]. This method approximates the unitary propagator $U(t) = e^{-iHt/\varepsilon}$ of the time-dependent Schrödinger equation (1.1) for a fixed time t by a linear combination of Chebyshev polynomials of the Hamiltonian. In contrast to methods based on the concatenation of short-time propagators, the Chebyshev method has

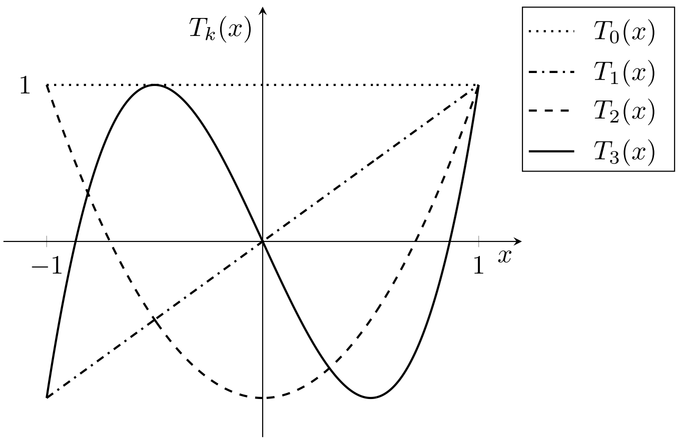


Figure 6.4: Plot of the first four Chebyshev polynomials.

the advantage that it can be implemented without error accumulation, since it allows the computation of the time-evolved state $U(t)\psi_0$ directly at the final time without having to compute intermediate states at earlier times. The original method has been successfully applied to molecular systems with low dimensionality, see e.g. [GG02, CA13], but applications to high-dimensional systems have been hindered by the exponential scaling of memory and computational costs, as the method relies on full grid representations. In the following, we present a viable solution to the exponential scaling by applying the tensor-train format.

The section is structured as follows: Section 6.2.1 introduces Chebyshev polynomials, Section 6.2.2 describes how to generate Chebyshev expansions of complex-valued functions, and Section 6.2.3 describes Chebyshev propagation based on discrete space representations. Finally, after discussing the tensor-train implementation in Section 6.2.4, numerical experiments of the TTC method are presented in Section 6.2.5.

6.2.1 Chebyshev polynomials

For all integers $k \geq 0$, the *kth Chebyshev polynomial* is defined as follows:

$$T_k \colon [-1,1] \to [-1,1],\, T_k(x) := \cos\big(k \arccos(x)\big)$$

We note that the Chebyshev polynomials satisfy the following recurrence relation

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x), \tag{6.4}$$

and the first four polynomials are given by (see also Figure 6.4)

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x.$$

Chebyshev polynomials have a number of remarkable properties and are therefore an important tool in approximation theory, see e.g. [Tre19, Chapter 10] and [FP68, Chapter 1]. For instance, they satisfy the following orthogonality relation for all $j, k \geq 1$, $j \neq k$,

$$\int_{-1}^{1} T_j(x) T_k(x) \frac{\mathrm{d}x}{\sqrt{1 - x^2}} = \frac{\pi}{2} \delta_{j,k}, \tag{6.5}$$

showing that the Chebyshev polynomials are orthogonal with respect to the weighted inner product defined by the left-hand side of (6.5).

6.2.2 Chebyshev expansion of complex-valued functions

Chebyshev polynomials can be used to approximate a given complex-valued function $f\colon [-1, 1] \to \mathbb{C}$ via its Fourier series, see e.g. [FP68, Chapter 2.6]. To show how, we introduce the 2π-periodic function

$$g(x) = f(\cos(x)),$$

which can be represented in $(-\pi, \pi)$ in terms of its Fourier series as follows:

$$g(x) = \sum_{k=0}^{\infty} (2 - \delta_{k,0}) a_k \cos(kx), \quad a_k = \frac{1}{\pi} \int_0^{\pi} g(x) \cos(kx)\, \mathrm{d}x$$

Therefore, the original function $f(y) = g(\arccos(y))$ can be represented for all $y \in (-1, 1)$ in terms of the Chebyshev polynomials as

$$f(y) = \sum_{k=0}^{\infty} (2 - \delta_{k,0})\, c_k T_k(y), \quad c_k = \frac{1}{\pi} \int_{-1}^{1} f(y) T_k(y) \frac{\mathrm{d}y}{\sqrt{1 - y^2}}. \tag{6.6}$$

Equation (6.6) is called the “Chebyshev expansion” of f. It can be used to approximate f as the linear combination of the first $N \geq 1$ Chebyshev polynomials as follows:

$$f(y) \approx S_N f(y) = \sum_{k=0}^{N-1} (2 - \delta_{k,0})\, c_k T_k(y) \tag{6.7}$$

The coefficients c_k defined by (6.6) are essentially the Fourier coefficients of g, which for analytic functions decay exponentially with k, see e.g. [Tad07, Section 2], and thus provide fast convergence of the partial sums $S_N f$. In particular, the resulting Chebyshev approximation is a polynomial of degree N, which is known to be close to the polynomial of the same degree with minimal error in the interval $[-1, 1]$, see [TE89, Section 2].

6.2.3 Chebyshev propagation in discrete representations

We obtain an approximation of the operator $U(t) = e^{-iHt/\varepsilon}$ at a given time $t > 0$ by considering the function $f(y) = e^{-iyt/\varepsilon}$ for which the coefficients c_k defined according to (6.6) can be expressed in terms of the Bessel functions J_k (of the first kind) as follows, see e.g. [AS64, Chapter 9],

$$c_k = (-i)^k J_k(t/\varepsilon),$$

yielding the following approximation for all $y \in (-1, 1)$:

$$e^{-iyt/\varepsilon} \approx \sum_{k=0}^{N-1} (2 - \delta_{k,0})\,(-i)^k J_k(t/\varepsilon) T_k(y) \tag{6.8}$$

Using a linear transformation of the argument y, we can restate (6.8) for an arbitrary Hermitian matrix $H \in \mathbb{C}^{d\times d}$ with eigenvalues contained in a finite interval $[a, b]$ as

$$e^{-iHt/\varepsilon} \approx e^{-it^+} \sum_{k=0}^{N-1} (2 - \delta_{k,0})\,(-i)^k J_k(t^-) T_k(H_0), \tag{6.9}$$

where we have introduced the rescaled variables $t^-, t^+ \in \mathbb{R}$ and the matrix $H_0 \in \mathbb{C}^{d\times d}$ with eigenvalues in $[-1, 1]$ defined by

$$t^\pm := \frac{t}{2\varepsilon}(b \pm a) \quad \text{and} \quad H_0 := \frac{2}{b-a}\left(H - \frac{b+a}{2}\,\mathrm{Id}\right), \tag{6.10}$$

where Id is the $d \times d$ identity matrix. In particular, since $e^{-iyt/\varepsilon}$ is an analytic function, we obtain fast convergence of the approximation in (6.9). However, the number of required polynomials increases with t since $e^{-iyt/\varepsilon}$ is oscillatory and thus a sufficiently large number N of Chebyshev polynomials is needed to resolve the oscillations. In fact, it has been shown that the error falls like the Nth order in $|t^-|/(2N)$ for sufficiently large N, see [Lub08, Chapter III.2.1, Theorem 2.4].

Remark 88. *It is important to note that (6.9) can be used more generally than in the current implementation to approximate the solution to any linear system of the form $i\dot{u} = Hu$. Such linear systems typically arise in space discretisation methods, including the Fourier collocation method, the Fourier–Galerkin method, or the Hermite–Galerkin method, see [Lub08, Chapter III]. Hence, we anticipate that the TTC method should also be valuable for solving high-dimensional linear systems in a wide range of applications beyond the solution of the time-dependent Schrödinger equation.*

6.2.4 Tensor-train implementation

Discrete tensor-train approximations of the propagator $U(t) = e^{-iHt/\varepsilon}$ are obtained by discretising the d-dimensional space in each coordinate direction in the range $x_{j,\min}$ to $x_{j,\max}$ with $n_j > 1$ points and the grid size $\Delta x_j > 0$. Point evaluations $\psi(x_k)$ of the wave function ψ, which are represented as a tensor $W \in \mathbb{C}^{n_1 \times \cdots \times n_d}$, are approximated for fixed compression ranks $r_1, \ldots, r_{d-1}$ by complex-valued low-rank tensor trains with components $W_j[k_j] \in \mathbb{C}^{r_{j-1} \times r_j}$ as follows,

$$W(k_1, \ldots, k_d) \approx W_1[k_1] \cdots W_d[k_d],$$

where $k = (k_1, \ldots, k_d)$ is the index corresponding to the gird point $x_k = (x_{1,k_1}, \ldots, x_{d,k_d})$. The action of the Hamiltonian

$$H = T + V = -\frac{\varepsilon^2}{2}\Delta_x + V$$

is represented by a Hermitian operator $\mathcal{H} = \mathcal{T} + \mathcal{V}$ on the tensor space. By this we mean that if tensors in $\mathbb{C}^{n_1 \times \cdots \times n_d}$ are identified with row vectors of length $D = n_1 \cdots n_d$, the discretised Hamiltonian can be represented as a Hermitian matrix in $\mathbb{C}^{D \times D}$. As usual, the potential energy operator $\mathcal{V}$ is given by the elementwise multiplication operator (Hadamard product), *i.e.*,

$$(\mathcal{V}W)(k_1, \ldots, k_d) := V_1[k_1] \cdots V_d[k_d] W_1[k_1] \cdots W_d[k_d],$$

where the matrices $V_j[k_j]$ are the cores of the potential in TT format. Furthermore, the kinetic energy operator

$$(\mathcal{T}W)(k_1, \ldots, k_d) \approx -\frac{\varepsilon^2}{2}\Delta_x \psi(x_k),$$

is defined by the Laplacian that acts as a multiplication operator in momentum space. Therefore, we apply the kinetic energy operator in momentum space by exploiting the implementation of multidimensional discrete Fourier transforms of tensor trains to switch between position and momentum space. With the help of the FFT we obtain a very efficient and accurate implementation of the discretised kinetic energy operator.

Recall that for the approximation of the propagator the discrete Hamiltonian must be rescaled according to (6.10) as follows,

$$\mathcal{H}_0 = \frac{2}{E_{\max} - E_{\min}} \left(\mathcal{H} - \frac{E_{\max} + E_{\min}}{2} \, \mathrm{Id} \right),$$

where now Id denotes the identity on the tensor space. The bounds for the eigenvalues $E_{\min}$ and $E_{\max}$ depend on the extension of the position grid and are given by

$$E_{\min} = \min_k V(x_k), \quad E_{\max} = \frac{\varepsilon^2 \pi^2}{2} \left(\frac{1}{\Delta x_1^2} + \cdots + \frac{1}{\Delta x_d^2} \right) + \max_k V(x_k),$$

where we used $\Delta p_j = 2\pi/(x_{j,\max} - x_{j,\min})$ for the grid spacing in momentum space of the jth coordinate, giving the maximum kinetic energy

$$T_{j,\max} = \frac{\varepsilon^2}{2} p_{j,\max}^2 = \frac{\varepsilon^2}{2} \frac{\pi^2}{\Delta x_j^2}.$$

Consequently, the solution $\psi(t)$ of the time-dependent Schrödinger equation (1.1) is approximated with $N \geq 1$ Chebyshev polynomials as follows,

$$\psi(t) = e^{-iHt/\varepsilon}\psi_0 \approx e^{-it^+} \sum_{k=0}^{N-1} (2 - \delta_{k,0})\,(-i)^k J_k(t^-) T_k(\mathcal{H}_0) W_0, \tag{6.11}$$

where $t^\pm = tE^\pm/2\varepsilon$, $E^\pm = E_{\max} \pm E_{\min}$, and W_0 samples the initial wave function ψ_0. We implement (6.11) as a one-step propagator to compute $\psi(t)$ directly from the initial data by using the Clenshaw algorithm, see Appendix 7.8. Alternatively, one could obtain the time-dependent states $T_k(\mathcal{H}_0)W_0$ according to the recurrence relation (6.4) as

$$\begin{aligned} T_0(\mathcal{H}_0)W_0 &= W_0,\ T_1(\mathcal{H}_0)W_0 = \mathcal{H}_0 W_0, \\ T_{k+1}(\mathcal{H}_0)W_0 &= 2\mathcal{H}_0 T_k(\mathcal{H}_0)W_0 - T_{k-1}(\mathcal{H}_0)W_0, \quad \text{for } k \geq 1. \end{aligned}$$

We note that the Chebyshev propagation scheme can be alternatively implemented by using the continuous analogue functional tensor-train decomposition of wave functions as described in [SBGB22, Section 3].

6.2.5 Numerical experiments

We present numerical experiments for two systems. To demonstrate the dependence of the error on the number of Chebyshev polynomials and the propagation time, we first test the TTC method for a Gaussian wave packet in a two-dimensional harmonic oscillator potential. We then present results for simulating the dynamics of protons in a 50-dimensional model of hydrogen-bonded DNA.

Remark 89. *In the following experiments, the construction of the tensor trains and the operations in TT format were performed using Oseledets' TT-Toolbox, see [Ose20]. Moreover, all operations were followed by rounding (cf. Remark 84) to avoid an artificial growth of the ranks.*

Two-dimensional harmonic oscillator

We consider the two-dimensional quantum harmonic oscillator potential

$$V(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2).$$

The discretisation in position space was based on a 32×32 grid on $\Lambda_q = [-6, 6]^2$ and the maximum rank for rounding as described in Remark 89 was chosen as $r_{\max} = 3$.

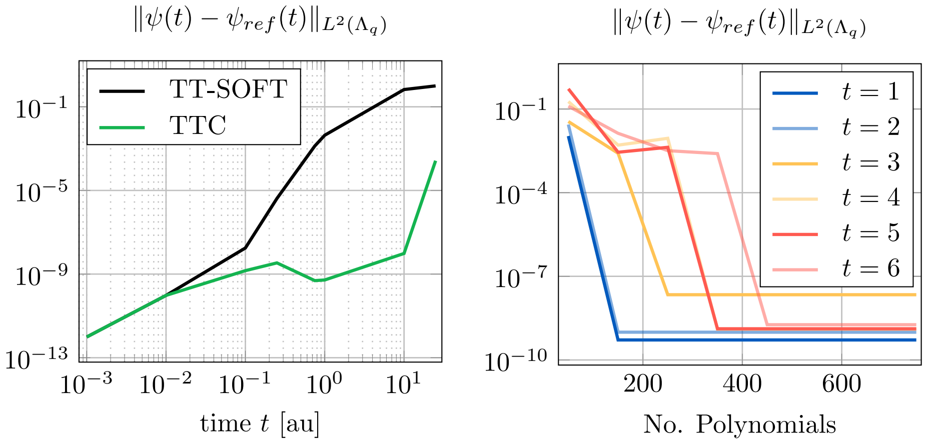


Figure 6.5: Left: Evolution of the L^2-error between the TTC method and the analytical solution for the two-dimensional harmonic oscillator ($\varepsilon = 1$). TTC provides more accurate approximations than the TT-SOFT method for longer times. Right: L^2-error for different numbers of polynomials. The number of polynomials required for accurate simulation increases with larger times.

The left-hand side of Figure 6.5 shows the L^2-error between the TTC method and the analytical solution (green line). For comparison, the L^2-error produced by the TT-SOFT method is plotted (black line), which is essentially the split-step Fourier method in tensor-train format, see [GB17]. The number of Chebyshev polynomials was $N = 750$ and the initial wave function was chosen as the Gaussian wave packet $\psi_0 = g_{z_0}^{C_0,\varepsilon}$ with $\varepsilon = 1$, $q_0 = (1, 0)$, $p_0 = (0, 0)$ and $C_0 = i\,\mathrm{Id}$. While the errors for short time steps cannot be distinguished, TTC produces the complete wave function with an error several orders of magnitude smaller for longer time steps given the sufficient large number of Chebyshev polynomials. In addition, the right-hand side of the Figure 6.5 shows the L^2-error for different numbers of polynomials, but this time for the initial Gaussian wave packet centred at $q_0 = (0, 0)$. We observe that the TTC method requires less than 200 polynomials to accurately approximate the full wave function for the final times $t = 1$ and $t = 2$. As expected, the number of polynomials required for accurate simulation increases with larger times. In particular, the errors converge for all final times up to $t = 6$ for fewer than 500 polynomials, demonstrating the robustness of the TTC method for the simulation of long-time dynamics.

50-dimensional model of hydrogen-bonded DNA

To demonstrate the capabilities of the TTC method for high-dimensional systems, we simulated the dynamics of protons in a 50-dimensional model of hydrogen-bonded DNA adenine-thymine base pairs, which is described by the potential

$$V(x_1, \dots, x_{50}) = \sum_{k=1}^{50} \alpha \left(0.429x_k - 1.126x_k^2 - 0.143x_k^3 + 0.563x_k^4\right) - \sum_{k=2}^{50} \alpha\beta x_k x_{k-1},$$

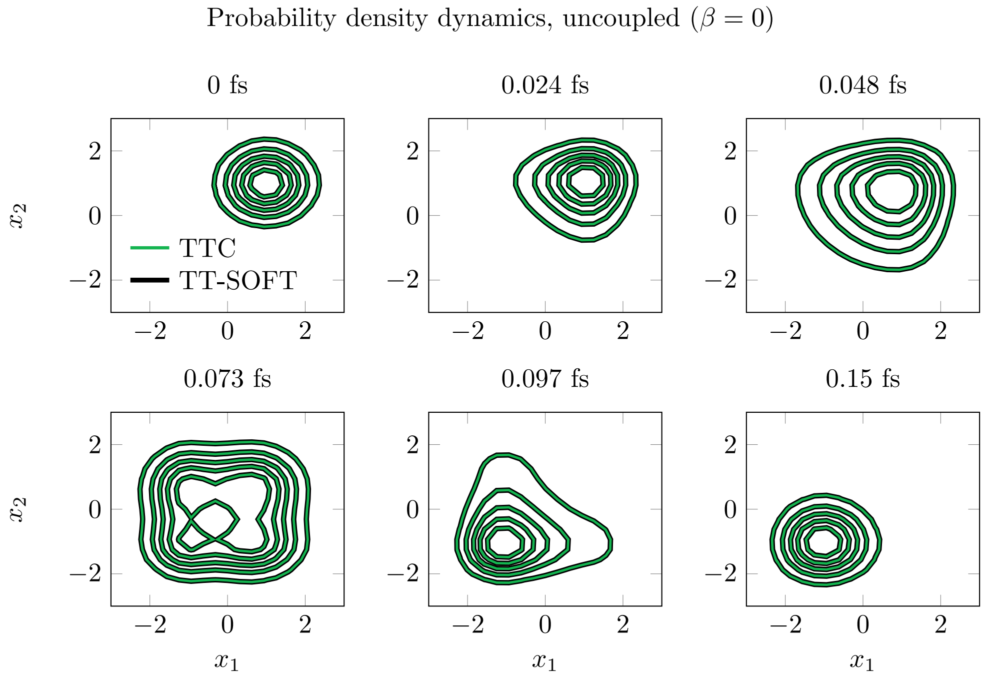


Figure 6.6: Comparison of two-dimensional slices of the 50-dimensional time-dependent wave packet obtained from TTC (green) as compared to TT-SOFT (black) for uncoupled DNA base pairs.

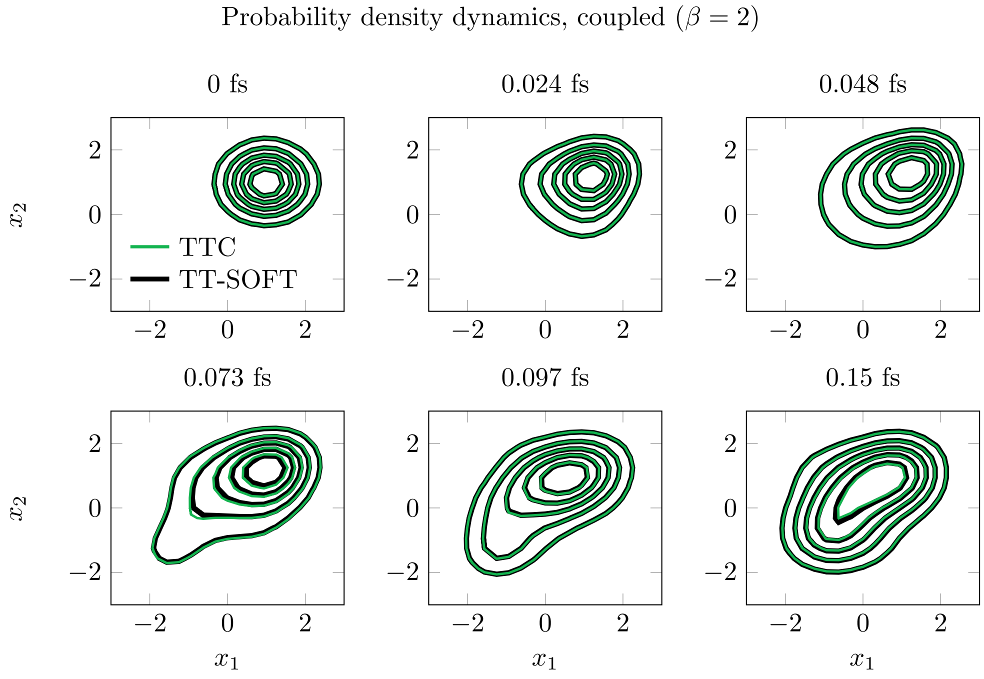


Figure 6.7: Comparison of two-dimensional slices of the 50-dimensional time-dependent wave packet obtained from TTC (green) as compared to TT-SOFT (black) for coupled DNA base pairs.

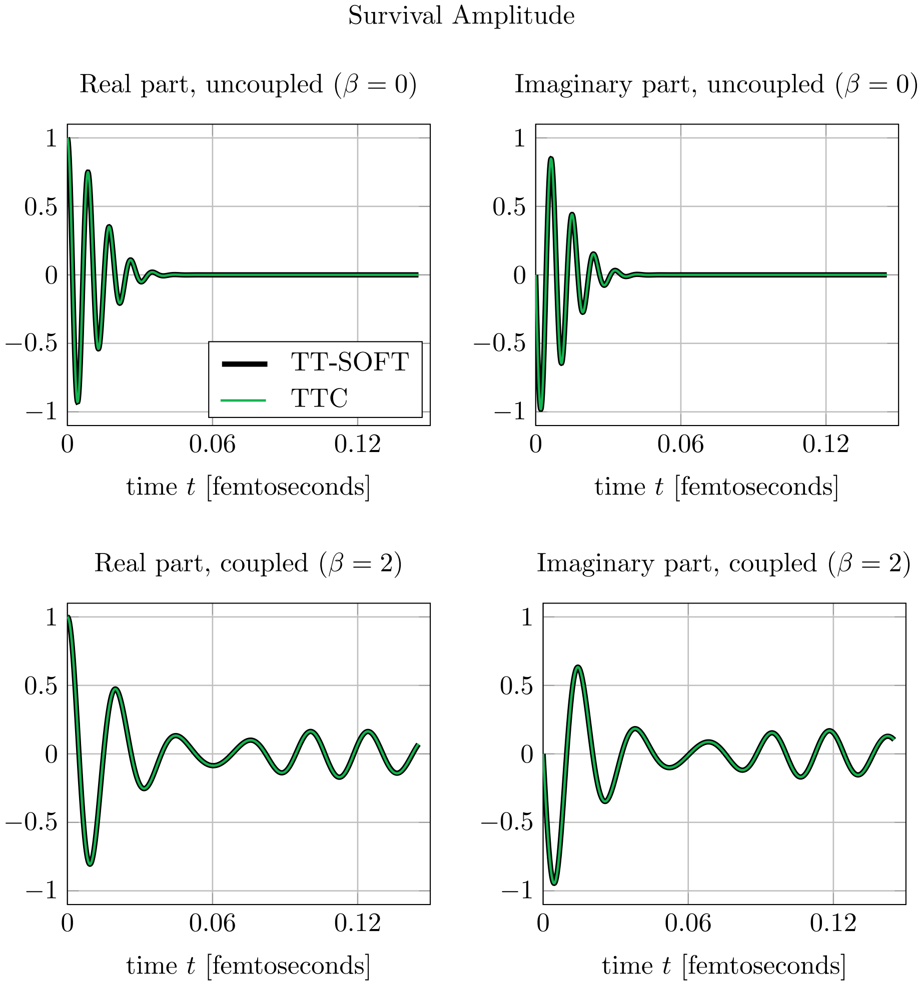


Figure 6.8: Comparison of survival amplitudes for the dynamics of uncoupled (top) and coupled (bottom) DNA base pairs, including the real (left) and imaginary (right) parts, obtained with TTC (green) and benchmark TT-SOFT (black).

where $\alpha = 0.1$ determines the energy scaling of the potential and $\beta \geq 0$ is the coupling parameter of the hydrogen bonds, see [GAKS15]. Due to the strongly anharmonic modes in the potential, this model is a challenging test case for tensor-train-based methods, especially for coupled DNA pairs ($\beta \neq 0$), which is beyond the reach of the original grid-based Chebyshev approach or other methods based on full grid representations.

The discretisation in position space was based on a uniform grid with 32 grid points in the range $x_k \in [-5, 5]$ for each coordinate direction. The number of Chebyshev polynomials was $N = 50$ and the initial wave function was chosen as $\psi_0 = g_{z_0}^{C_0,\varepsilon}$ with $\varepsilon = 1$, $q_{0,k} = 1$, $p_{0,k} = 0$ and $C_0 = i\,\mathrm{Id}$. For both TTC and TT-SOFT, the solution was calculated at intermediate times with a time step of $\tau = 0.01$ by defining each intermediate time as an end point. Figure 6.6 shows the comparison of two-dimensional slices in the x_1x_2-plane of the 50-dimensional wave packet obtained with the TTC method (green line) and the benchmark method TT-SOFT (black) for the uncoupled system ($\beta = 0$) at six different times. The maximum rank for rounding was chosen as $r_{\max} = 3$. Figure 6.7 shows the corresponding simulations for the coupled system ($\beta = 2$) with the maximum rank $r_{\max} = 10$. In addition to computing the full wave function, the method was analysed by computing survival amplitudes, which can be seen in Figure 6.8 for the uncoupled system (top) and the coupled system (bottom). The results for the full wave function and the survival amplitudes show excellent agreement between the methods and efficient performance, even without relying on high-performance computing facilities.

6.3 Summary of this chapter

Numerical methods based on naive space discretisations are not applicable to high-dimensional systems, since the size of the tensors resulting e.g. from function evaluations increases exponentially with increasing dimension. Both the storage and processing of operations is therefore only possible, if at all, with the use of high-performance computers. In this chapter we introduced the tensor-train decomposition, which allows the storage of tensors that usually require $\mathcal{O}(N^d)$ data points for a d-dimensional grid with N points to be reduced to $O(dNR^2)$ data points, where R denotes the maximum rank. It was also discussed how basic operations such as addition, multiplication and rounding can be implemented in TT format. The presented numerical results on the tensor-train implementation of the Chebyshev propagation method have shown that the TT format can potentially be used for other grid-based methods, e.g. the TSTG method.

7 Appendix

7.1 Inner products of Gaussian wave packets

Proof (of Lemma 7). The product of the functions $\overline{g_{z_1}^{C_1,\varepsilon}}$ and $g_{z_2}^{C_2,\varepsilon}$ is again a Gaussian, and to obtain an explicit representation, we rewrite the sum of the exponents

$$-\frac{i}{\varepsilon}\left(\frac{1}{2}(x-q_1)^T\bar{C}_1(x-q_1)+p_1^T(x-q_1)\right)+\frac{i}{\varepsilon}\left(\frac{1}{2}(x-q_2)^T C_2(x-q_2)+p_2^T(x-q_2)\right)$$

as a quadratic function

$$\frac{i}{\varepsilon}\left(\frac{1}{2}(x-q_2)^T B(x-q_2)+(x-q_2)^T b+c\right),$$

where a short calculation shows that $B\in\mathbb{C}^{d\times d}$, $b\in\mathbb{C}^d$ and $c\in\mathbb{C}$ are given by

$$\begin{aligned} B &:= C_2-\bar{C}_1, \quad b:=(p_2-p_1)-\bar{C}_1(q_2-q_1) \quad \text{and} \\ c &:= -\frac{1}{2}(q_2-q_1)^T\bar{C}_1(q_2-q_1)-p_1^T(q_2-q_1). \end{aligned} \tag{7.1}$$

In particular, since $\operatorname{Im}(-\bar{C}_1)=\operatorname{Im}(C_1)$ is positive definite and the sum of two real positive definite matrices is again positive definite, we conclude that B is an element of the Siegel space $\mathfrak{S}^+(d)$. This yields the following representation for all $x\in\mathbb{R}^d$:

$$\overline{g_{z_1}^{C_1,\varepsilon}(x)}g_{z_2}^{C_2,\varepsilon}(x)=\alpha\exp\left[\frac{i}{\varepsilon}\left(\frac{1}{2}(x-q_2)^T B(x-q_2)+(x-q_2)^T b+c\right)\right],$$

where the positive constant $\alpha>0$ is given by

$$\alpha:=(\pi\varepsilon)^{-d/2}\det(\operatorname{Im}C_1\operatorname{Im}C_2)^{1/4}.$$

Therefore, we conclude that

$$\begin{aligned}\langle g_{z_1}^{C_1,\varepsilon}\mid g_{z_2}^{C_2,\varepsilon}\rangle &= \int_{\mathbb{R}^d}\alpha\exp\left[\frac{i}{\varepsilon}\left(\frac{1}{2}(x-q_2)^T B(x-q_2)+(x-q_2)^T b+c\right)\right]\mathrm{d}x \\ &= \alpha\int_{\mathbb{R}^d}\exp\left[\frac{i}{\varepsilon}\left(\frac{1}{2}y^T By+y^T b+c\right)\right]\mathrm{d}y.\end{aligned}$$

In particular, since the last integral can be solved analytically by using a formula for multivariate Gaussian integrals, see e.g. [Fol89, Appendix A (Theorem 1)], we get

$$\int_{\mathbb{R}^d}\exp\left[\frac{i}{\varepsilon}\left(\frac{1}{2}y^T By+y^T b+c\right)\right]\mathrm{d}y=\frac{(2\pi\varepsilon)^{d/2}}{\sqrt{\det(-iB)}}\exp\left(-\frac{i}{2\varepsilon}b^T B^{-1}b+\frac{i}{\varepsilon}c\right),$$

where the branch of the square root is determined by the requirement

$$\det(-iB)^{-1/2} > 0$$

if $-iB$ is real and positive definite. Moreover, using the formulas in equation (7.1), we obtain the following representation:

$$\begin{aligned}
&\alpha\frac{(2\pi\varepsilon)^{d/2}}{\sqrt{\det(-iB)}}\exp\left(-\frac{i}{2\varepsilon}b^TB^{-1}b+\frac{i}{\varepsilon}c\right)\\
&=\frac{2^{d/2}\det(\operatorname{Im}C_1\operatorname{Im}C_2)^{1/4}}{\sqrt{\det(-iB)}}\exp\left(\frac{i}{2\varepsilon}(p_1+p_2)^T(q_1-q_2)\right)\cdots\\
&\quad\exp\left(\frac{i}{2\varepsilon}(p_2-p_1)^T(C_2-\bar{C}_1)^{-1}(C_2+\bar{C}_1)(q_2-q_1)\right)\cdots\\
&\quad\exp\left(\frac{i}{2\varepsilon}(p_2-p_1)^T(-B^{-1})(p_2-p_1)\right)\exp\left(\frac{i}{2\varepsilon}(q_2-q_1)^T(-\bar{C}_1-\bar{C}_1B^{-1}\bar{C}_1)(q_2-q_1)\right)
\end{aligned}$$

In the last line we have two Gaussians: One with respect to the difference p_2-p_1 with width matrix $-B^{-1}$ and one for q_2-q_1 with width matrix $-\bar{C}_1-\bar{C}_1B^{-1}\bar{C}_1$. In particular, the Woodbury matrix identity, see e.g. [Hig02, Page 258], yields

$$-\bar{C}_1-\bar{C}_1B^{-1}\bar{C}_1=\left(C_2^{-1}-\bar{C}_1^{-1}\right)^{-1}.$$

Hence, since $Z\in\mathfrak{S}^+(d)$ implies $-Z^{-1}\in\mathfrak{S}^+(d)$, see e.g. [Fol89, Theorem 4.64], we conclude that both width matrices

$$-B^{-1}\quad\text{and}\quad\left(C_2^{-1}-\bar{C}_1^{-1}\right)^{-1}$$

are in $\mathfrak{S}^+(d)$ and therefore we conclude that the block diagonal matrix M in (2.9) is an element of $\mathfrak{S}^+(2d)$. Putting together the above calculations we arrive at (2.8).

To prove the bound in (2.10), we follow the idea of [Swa08, 11.4 Lemma] and assume that the eigenvalues of $\operatorname{Im}(C_k)$ and $\operatorname{Im}(-C_k^{-1})$ are bounded from below by $\theta>0$ and from above by $\Theta>0$. Furthermore, let us introduce the real-valued Gaussian function

$$g_k^\theta(x)=(\pi\varepsilon)^{-d/4}\theta^{d/4}\exp\left(-\frac{\theta}{2\varepsilon}\|x-q_k\|_2^2\right),\quad k=1,2,\ x\in\mathbb{R}^d.$$

Then, for all $x\in\mathbb{R}^d$, the spectral bounds imply that

$$|g_{z_k}^{C_k,\varepsilon}(x)|\le\det(\operatorname{Im}C_k)^{1/4}\theta^{-d/4}g_k^\theta(x)\le\Theta^{d/4}\theta^{-d/4}g_k^\theta(x),$$

and therefore we obtain the following bound:

$$\begin{aligned}\left|\langle g_{z_1}^{C_1,\varepsilon}\mid g_{z_2}^{C_2,\varepsilon}\rangle\right|&\le\theta^{-d/2}\Theta^{d/2}\langle g_1^\theta\mid g_2^\theta\rangle\\&=\theta^{-d/2}\Theta^{d/2}\exp\left(-\frac{\theta}{4\varepsilon}\|q_2-q_1\|_2^2\right),\end{aligned}\tag{7.2}$$

where the last equality follows by (2.8). Furthermore, combining Plancherel's theorem with a formula for the Fourier transform $\mathcal{F}_\varepsilon g_{z_k}^{C_k,\varepsilon}$, implies

$$\begin{aligned} \left|\langle g_{z_1}^{C_1,\varepsilon} \mid g_{z_2}^{C_2,\varepsilon}\rangle\right| &= \left|\langle \mathcal{F}_\varepsilon g_{z_1}^{C_1,\varepsilon} \mid \mathcal{F}_\varepsilon g_{z_2}^{C_2,\varepsilon}\rangle\right| \\ &\leq \theta^{-d/2}\Theta^{d/2}\exp\left(-\frac{\theta}{4\varepsilon}\|p_2 - p_1\|_2^2\right). \end{aligned} \tag{7.3}$$

Consequently, combining the bounds in (7.2) and (7.3) proves (2.10) for

$$\zeta = \left(\frac{\Theta}{\theta}\right)^d. \tag{7.4}$$

□

7.2 Summation curve

To prove the existence of a constant $C_{\Gamma_q} > 0$ with

$$\sup_{x\in\Lambda_q} \frac{1}{S(x)} \sum_{k_1=1}^{K} \cdots \sum_{k_d=1}^{K} |g_0(x - q_k)| < C_{\Gamma_q}^d \tag{7.5}$$

as used in the proof of Theorem 17, we use two one-dimensional bounds: The first one is an upper bound for the infinite series

$$\sum_{k\in\mathbb{Z}} |g_0(x - q_k)|$$

and the second one a lower bound for the summation curve $S(x)$.

Lemma 90 (Upper bound)**.** *For $d = 1$ consider the Gaussian g defined in (2.1) with width parameter $\gamma = \gamma_r + i\gamma_i \in \mathbb{C}$, $\gamma_i > 0$ and the uniform grid points $q_k = k\Delta q$ with distance $\Delta q > 0$. Then, for all $x \in \mathbb{R}$, we have*

$$\sum_{k\in\mathbb{Z}} |g_0(x - q_k)| < c_{\Delta q,\varepsilon,\gamma} \tag{7.6}$$

with upper bound

$$c_{\Delta q,\varepsilon,\gamma} = \sqrt{2}(\pi\varepsilon)^{1/4}\gamma_i^{-1/4}\frac{1}{\Delta q}\left(1 + \Delta q\sqrt{\frac{\gamma_i}{2\pi\varepsilon}}\right).$$

Proof. Let $x \in \mathbb{R}$. Using formula (3.5), we get

$$\sum_{k\in\mathbb{Z}} |g_0(x - q_k)| \leq \sqrt{2}(\pi\varepsilon)^{1/4}\gamma_i^{-1/4}\frac{1}{\Delta q}\left(1 + 2\sum_{n=1}^{\infty}\exp\left(-\frac{2\varepsilon\pi^2 n^2}{\gamma_i\Delta q^2}\right)\right).$$

Hence, (7.6) follows by the estimate

$$\sum_{n=1}^{\infty} \exp\left(-\frac{2\varepsilon\pi^2 n^2}{\gamma_i \Delta q^2}\right) \leq \int_0^\infty \exp\left(-\frac{2\varepsilon\pi^2 z^2}{\gamma_i \Delta q^2}\right) \mathrm{d}z = \frac{\Delta q}{2}\sqrt{\frac{\gamma_i}{2\pi\varepsilon}}.$$

□

Lemma 91 (Lower bound)**.** *For $d = 1$ consider the Gaussian g defined in (2.1) with width parameter $\gamma = \gamma_r + i\gamma_i \in \mathbb{C}$, $\gamma_i > 0$. Moreover, consider the uniform grid*

$$q_k = q_0 - L_q + \frac{2k-1}{2}\Delta q, \quad k = 1, \ldots, K,$$

where $\Delta q = 2L_q/K$. Then, for all $x \in [q_0 - L_q, q_0 + L_q]$ we have

$$S(x) = \sum_{k=1}^{K} |g_0(x - q_k)|^2 > C_{\Delta q,\varepsilon,\gamma,L_q} \tag{7.7}$$

with lower bound

$$C_{\Delta q,\varepsilon,\gamma,L_q} = \frac{1}{2\Delta q}\left(\operatorname{erf}\left(2L_q\sqrt{\frac{\gamma_i}{\varepsilon}}\right) - \operatorname{erf}\left(\Delta q\sqrt{\frac{\gamma_i}{\varepsilon}}\right)\right). \tag{7.8}$$

Proof. Let $\bar{x} \in [q_0 - L_q, q_0 + L_q]$ and denote by $q_{\bar{K}}$ the nearest grid point, *i.e.*,

$$\bar{K} = \operatorname*{arg\,min}_{k=1,\ldots,K} |\bar{x} - q_k|.$$

There exists $t \in (-\Delta q/2, \Delta q/2]$ such that $\bar{x} = q_{\bar{K}} + t$, and without any loss of generality we assume that $t \geq 0$. Now, we decompose the one-dimensional summation curve as

$$S(x) = S_{\bar{K}}(x) + \big(S(x) - S_{\bar{K}}(x)\big),$$

where $S_{\bar{K}}$ is defined by

$$S_{\bar{K}}(x) := \sum_{k=1}^{\bar{K}} |g_0(x - q_k)|^2.$$

In particular, since $S_{\bar{K}}$ is symmetric to the axis

$$x = \frac{q_1 + q_{\bar{K}}}{2} = \begin{cases} q_l, & \text{if } \bar{K} = 2l - 1, \\ q_l + \Delta q/2, & \text{if } \bar{K} = 2l, \end{cases}$$

we have $S_{\bar{K}}(\bar{x}) = S_{\bar{K}}(q_{\bar{K}} + t) = S_{\bar{K}}(q_1 - t)$. Hence, since $S_{\bar{K}}$ is increasing on $[q_0 - L_q, q_1]$ and $S - S_{\bar{K}}$ is increasing on $[q_0 - L_q, q_{\bar{K}+1}]$, we conclude that

$$S(q_0 - L_q) \leq S_{\bar{K}}(q_1 - t) + \big(S(\bar{x}) - S_{\bar{K}}(\bar{x})\big) = S(\bar{x}),$$

and since $\bar{x}$ was chosen arbitrarily, this shows that on the interval $[q_0 - L_q, q_0 + L_q]$ the summation curve $S(x)$ attains its minimum at $x = q_0 - L_q$. Consequently, using that

$$S(q_0 - L_q) > \sum_{k=1}^{K} (\pi\varepsilon)^{-1/2} \gamma_i^{1/2} \exp\left(-\frac{\gamma_i}{\varepsilon}(k\Delta q)^2\right)$$

and estimating the sum from below by an integral as follows,

$$\sum_{k=1}^{K} (\pi\varepsilon)^{-1/2} \gamma_i^{1/2} \exp\left(-\frac{\gamma_i}{\varepsilon}(k\Delta q)^2\right) > \frac{1}{\Delta q\sqrt{\pi}} \int_{\Delta q\sqrt{\gamma_i/\varepsilon}}^{K\Delta q\sqrt{\gamma_i/\varepsilon}} e^{-z^2}\,\mathrm{d}z,$$

the lower bound in (7.8) follows by the definition of the error function. □

Proof for the upper bound in (7.5)*.* We have

$$\frac{1}{S(x)} \sum_{k_1=1}^{K} \cdots \sum_{k_d=1}^{K} |g_0(x - q_k)| < \frac{1}{S(x)} \sum_{k\in\mathbb{Z}^d} |g_0(x - q_k)|,$$

and since the grid is aligned with the eigenvalues of the matrix $\operatorname{Im} C$, we obtain the following factorization:

$$\frac{1}{S(x)} \sum_{k\in\mathbb{Z}^d} |g_0(x - q_k)| = \prod_{n=1}^{d} \frac{1}{S_n(x)} \sum_{k_n\in\mathbb{Z}} |g_n(x^T u_n - k_n\Delta q)|,$$

where the one-dimensional summation curves $S_n(x)$ are defined according to (3.6) for $\Gamma_q^{(n)} = \{1, \ldots, K\}$ and the Gaussian functions g_n are given in (3.7). In particular, using the upper bound in (7.6) and the lower bound in (7.7), we conclude that

$$\sup_{x\in\Lambda_x} \frac{1}{S_n(x)} \sum_{k_n\in\mathbb{Z}} |g_n(x^T u_n - k_n\Delta q)| < \frac{c_{\Delta q,\varepsilon,\lambda_n}}{C_{\Delta q,\varepsilon,\gamma,L_q}}, \qquad n = 1, \ldots, d.$$

Since in the limit $K \to \infty$ we get

$$\frac{c_{\Delta q,\varepsilon,\lambda_n}}{C_{\Delta q,\varepsilon,\gamma,L_q}} \to \frac{2\sqrt{2}(\pi\varepsilon)^{1/4}\lambda_n^{-1/4}}{\operatorname{erf}\left(2L_q\sqrt{\lambda_n/\varepsilon}\right)},$$

(use that $\Delta q \to 0$ as $K \to \infty$ and $\operatorname{erf}(x) \to 0$ as $x \to 0$) the bound in (7.5) follows for

$$C_{\Gamma_q} = \frac{2\sqrt{2}(\pi\varepsilon)^{1/4}\sigma^{-1/4}}{\operatorname{erf}\left(2L_q\sqrt{\sigma/\varepsilon}\right)},$$

where $\sigma > 0$ denotes the smallest eigenvalue of $\operatorname{Im}(C)$. □

7.3 Computing Fourier integrals using the FFT

For the computation of Fourier-type integrals, such as

$$I^{(k)}(\psi) := \int_{t-\lambda}^{t+\lambda} \psi(x)w(x-t)e^{-i\xi\frac{\pi}{\lambda}x}\,\mathrm{d}x, \quad \xi = \frac{k}{M},\ k\in\mathbb{Z},\ M\in\mathbb{N},$$

we used the fast Fourier transform. Let $v = (v_1,\dots,v_d)\in\mathbb{C}^d$ and

$$\hat{v}_l := \sum_{j=1}^{d} v_j e^{-2\pi i\cdot(j-1)(l-1)/d}, \quad l\in\{1,\dots,d\}.$$

For the computation of the integral $I^{(k)}(\psi)$ let us consider the composite trapezoidal rule on a uniform grid. Let $N\in\mathbb{N}$ be a power of 2, as well as $m\in\mathbb{N}$, $m\le N-1$ and

$$x_j := t-\lambda+\Delta\cdot j, \quad \Delta := \frac{2\lambda}{m}, \quad j\in\{0,1,\dots,m\}.$$

The trapezoidal rule with grid $\{x_j\}_{j\in\{0,1,\dots,m\}}$ yields the following approximation:

$$I^{(k)}(\psi) \approx e^{-i\frac{k}{M}\cdot\frac{\pi}{\lambda}t}e^{i\frac{k}{M}\pi}\Delta\Bigg(\frac{\psi(t+\lambda)w(\lambda)}{2}e^{-2\pi i\frac{k}{M}} - \frac{\psi(t-\lambda)w(-\lambda)}{2}\dots$$
$$+\sum_{j=1}^{m}\psi(x_{j-1})w(x_{j-1}-t)e^{-2\pi i\frac{k}{Mm}(j-1)}\Bigg).$$

Consequently, if $k = mn$ for some $n\in\{0,1,\dots,N-1\}$, as well as $M = N$ and $v_j := \psi(x_{j-1})w(x_{j-1}-t)$ for $j = 1,\dots,m$, $v_j := 0$ for $j = m+1,\dots,N$, then,

$$\begin{aligned} I^{(k)}(\psi) &\approx e^{-i\frac{mn}{N}\cdot\frac{\pi}{\lambda}t}e^{i\frac{mn}{N}\pi}\Delta\left(r_1e^{-2\pi i\frac{mn}{N}} - r_2 + \sum_{j=1}^{N} v_j e^{-2\pi i(j-1)n/N}\right) \\ &= e^{-i\frac{mn}{N}\cdot\frac{\pi}{\lambda}t}e^{i\frac{mn}{N}\pi}\Delta\Big(r_1e^{-2\pi i\frac{mn}{N}} - r_2 + \hat{v}_{n+1}\Big), \end{aligned}$$

where the constants $r_1, r_2\in\mathbb{R}$ are given by

$$r_1 = \frac{\psi(t+\lambda)w(\lambda)}{2} \quad\text{and}\quad r_2 = \frac{\psi(t-\lambda)w(-\lambda)}{2}.$$

In particular, the vector $\hat{v}$ can be calculated with the FFT. For sufficiently large values of m and N we get

$$\frac{1}{2\lambda}\int_{t-\lambda}^{t+\lambda}\psi(x)w(x-t)e^{-i\xi\frac{\pi}{\lambda}x}\,\mathrm{d}x \approx \frac{e^{-i\xi\frac{\pi}{\lambda}t}e^{i\xi\pi}}{m}\left(r_1e^{-2\pi i\xi} - r_2 + \hat{v}_{n+1}\right).$$

Recall that the window w is compactly supported. Provided that both the functions ψ and w are smooth on $(t-\lambda, t+\lambda)$, the trapezoidal rule gives accurate results. The actual rates of convergence are based on the Euler–Maclaurin formula and can be found e.g. in [DR07, Chapter 2.9]. In particular, the difference between the exact Fourier coefficients and their discrete approximation using the trapezoidal rule is known to be spectrally small, see [GT85, Equation 1.5].

7.4 Upper bound for K_s

Recall the representation of the combinatorial constant K_s in (4.10). We want to find an estimate for the following sum (cf. Equation (4.11)):

$$\sum_{k=0}^{s-1} \frac{(s+k)! \cdot (2s-k-1)!}{(2k+2)! \cdot (s-k-1)! \cdot (2s-2k)! \cdot k!} = \frac{K_s}{2^{2s}(s+1)!(2s)^2}.$$

For the summand we calculate

$$\begin{aligned} &\frac{(s+k)! \cdot (2s-k-1)!}{(2k+2)! \cdot (s-k-1)! \cdot (2s-2k)! \cdot k!} \\ &\quad = \frac{(s+k)! \cdot (2s-k-1)!}{s! \cdot k! \cdot (s-k-1)! \cdot s!} \cdot \frac{s! \cdot s!}{(2s+2)!} \cdot \binom{2s+2}{2k+2}. \end{aligned} \tag{7.9}$$

Recall Vandermonde's theorem, see e.g. [Sea91, Equation 2.43]:

$$\sum_{k=0}^{m} \frac{(a)_k}{k!} \frac{(b)_{m-k}}{(m-k)!} = \frac{(a+b)_m}{m!}, \quad a, b \in \mathbb{C},\, m \geq 0.$$

In particular, for $m = s-1$ and $a = b = s+1$, $s \geq 1$, we obtain

$$\sum_{k=0}^{s-1} \frac{(s+k)! \cdot (2s-k-1)!}{s! \cdot k! \cdot (s-k-1)! \cdot s!} = \binom{3s}{s-1}. \tag{7.10}$$

Hence, since

$$\binom{2s+2}{2s-2k} \leq \binom{2s+2}{s} \quad \text{for} \quad 0 \leq k \leq s-1,\, s \geq 2,$$

by (7.9) and (7.10) we conclude that

$$\begin{aligned} \sum_{k=0}^{s-1} \frac{(s+k)! \cdot (2s-k-1)!}{(2k+2)! \cdot (s-k-1)! \cdot (2s-2k)! \cdot k!} &\leq \frac{s! \cdot s!}{(2s+2)!} \binom{2s+2}{s} \binom{3s}{s-1} \\ &= \frac{(3s)!}{s!(2s+2)!} \cdot \frac{2s}{s+2}. \end{aligned}$$

This proves that

$$K_s \leq \frac{2^{2s+1} s^2 (3s)!}{(2s+1)!} \cdot \frac{2s}{s+2}, \quad s \geq 2.$$

Consequently, the true value of K_s is overestimated by the factor $2s/(s+2)$.

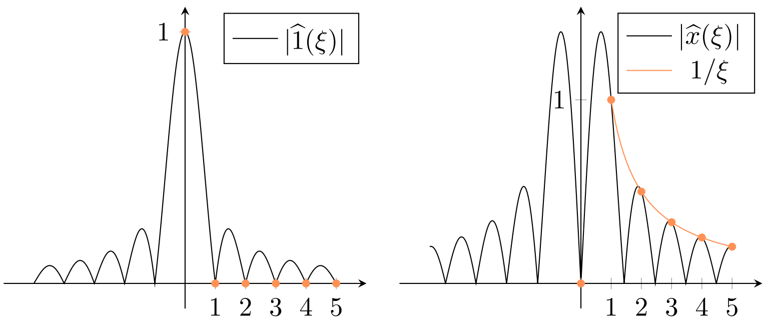


Figure 7.1: Absolute values of the Fourier coefficients (bullets) for $f(x) = 1$ (left) and $f(x) = x$ (right). The extension of the domain of the Fourier coefficients leads to a non-trivial function in ξ, but the restriction to integer values might result in a smooth decay (orange line).

7.5 Oscillations of the coefficients

We focus once more on the windowed coefficients c_ψ^w. In the plot at the upper left-hand side of Figure 4.2 the green line falls in a trembling way. To explain this phenomenon, we extend the domain of the Fourier coefficients. For a 2π-periodic function $f \in \mathrm{BV}_{\mathrm{loc}}$ and $\xi \in \mathbb{R}$ consider the number

$$\widehat{f}(\xi) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-i\xi x} \, \mathrm{d}x.$$

This means, that we calculate the Fourier coefficients not only for integer values, but for all real numbers ξ. For example, the extended Fourier coefficients of the function $f \equiv 1$ are given by

$$\widehat{1}(\xi) := \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-i\xi x} \, \mathrm{d}x = \frac{\sin(\pi\xi)}{\pi\xi} = \mathrm{sinc}(\xi) \, .$$

In particular, if k is an integer, we obtain the simple Fourier coefficients. Indeed,

$$|\widehat{1}(k)| = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{else.} \end{cases}$$

As we see in the left plot of Figure 7.1, for $k \neq 0$ the simple Fourier coefficients of $f \equiv 1$ correspond to the zeros of $\xi \mapsto |\,\mathrm{sinc}(\xi)\,|$. For the saw wave in Section 4.4.1 we can do the same calculation. Here we obtain

$$\widehat{x}(\xi) := \frac{1}{2\pi} \int_{-\pi}^{\pi} x e^{-i\xi x} \, \mathrm{d}x = i \cdot \frac{(\pi\xi\cos(\pi\xi) - \sin(\pi\xi))}{\pi\xi^2}.$$

Therefore, if $k \neq 0$ is an integer, we conclude that

$$|\widehat{x}(k)| = \frac{1}{|k|}.$$

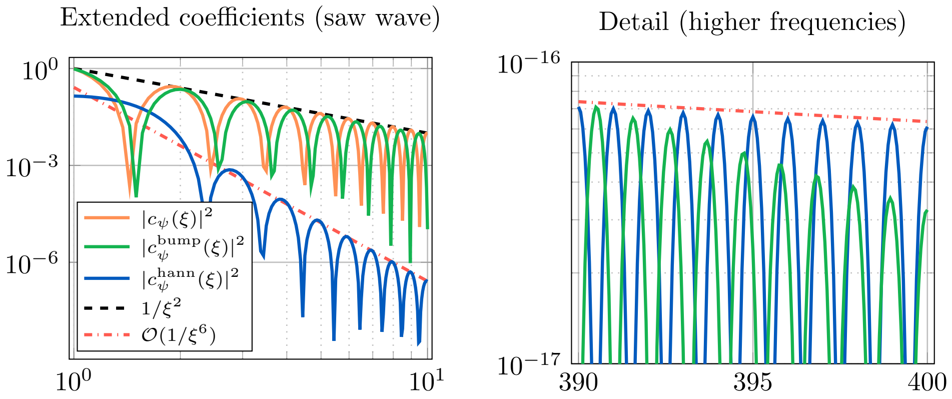


Figure 7.2: Due to the extension of the domain of the coefficients, we are able to resolve the trembling pattern in the upper left plot in Figure 4.2. The left plot shows low frequencies ($\xi \in [1, 10]$) and in the right plot ($\xi \in [390, 400]$) we observe that the bump coefficients fall below the Hann coefficients.

Thus, the coefficients of the saw wave function have a smooth decay, as we see at the right-hand side of Figure 7.1 (orange line). We computed the extended (windowed) coefficients for $\xi \in [1, 10]$ and $\xi \in [390, 400]$ for the saw wave. The result can be found in Figure 7.2. By extending the domain of the Fourier coefficients, we observe that the trembling also occurs for the other coefficients (plain and hann).

7.6 Discrete Gaussian convolution

Lemma 92. *For $\sigma > 0$ consider the one-dimensional Gaussian function*

$$f_\sigma(t) := \exp\left(-\frac{1}{2\sigma}t^2\right) \quad \textit{for all } t \in \mathbb{R}.$$

For arbitrary grid points $t_1 < t_2 < \ldots < t_N$ let

$$h_i := t_{i+1} - t_i,\ i = 1, \ldots, N-1 \quad \textit{and} \quad h := \min_{i=1,\ldots,N-1} h_i. \tag{7.11}$$

Then, for all $\sigma_1, \sigma_2 > 0$, there exists a constant $c > 0$ such that for all $s \in \mathbb{R}$ we have

$$\sum_{k=1}^{N} f_{\sigma_1}(t_k) f_{\sigma_2}(s - t_k) \leq c f_{\sigma_1+\sigma_2}(s), \tag{7.12}$$

where c depends on σ_1, σ_2 and h, but not on N.

Proof. Let $s \in \mathbb{R}$. A short calculation shows that

$$f_{\sigma_1}(t_k) f_{\sigma_2}(s - t_k) = f_{\sigma_1+\sigma_2}(s) f_{\sigma_3}\left(s' - t_k\right),$$

where we introduced the parameters

$$\sigma_3 = \frac{\sigma_1\sigma_2}{\sigma_1+\sigma_2} \quad \text{and} \quad s' = \frac{\sigma_1 s}{\sigma_1+\sigma_2}.$$

Consequently, the sum in (7.12) can be written as

$$\sum_{k=1}^{N} f_{\sigma_1}(t_k) f_{\sigma_2}(s-t_k) = f_{\sigma_1+\sigma_2}(s) \sum_{k=1}^{N} f_{\sigma_3}\left(s'-t_k\right).$$

In particular, the sum at the right-hand side can be bounded independently of s' as

$$\sum_{k=1}^{N} f_{\sigma_3}\left(s'-t_k\right) \leq \sum_{k\in\mathbb{Z}} f_{\sigma_3}\left(hk\right),$$

where the minimal distance $h > 0$ between consecutive grid points is defined in (7.11). Since the last sum can be viewed as a Riemann sum approximation to the integral

$$\frac{1}{h}\int_{\mathbb{R}} f_{\sigma_3}\left(t\right) \mathrm{d}t = \frac{\sqrt{2\pi\sigma_3}}{h},$$

we find a positive constant $c > 0$, depending on σ_3 and h, such that

$$\sum_{k\in\mathbb{Z}} f_{\sigma_3}\left(s'-t_k\right) \leq c,$$

which makes the proof complete. □

7.7 Reference solver: The split-step Fourier method

Let $\varepsilon, \mu > 0$ and $V\colon \mathbb{R} \to \mathbb{R}$ be a smooth potential. We are interested in the numerical solution to the one-dimensional time-dependent Schrödinger equation

$$\begin{cases} i\varepsilon\partial_t\psi(x,t) = -\dfrac{\varepsilon^2}{2\mu}\psi''(x,t) + V(x)\psi(x,t), \\ \psi(\bullet,0) = \psi_0, \end{cases} \tag{7.13}$$

for a given initial wave function $\psi_0 \in \mathcal{S}(\mathbb{R})$.

7.7.1 Transformation of the Schrödinger equation

Let ψ be the solution to (7.13). For a given time interval $[0, t_{max}]$, we assume that $|\psi(x,t)|$ is negligible outside the spatial interval $[a,b]$, *i.e.*,

$$|\psi(x,t)| \approx 0 \quad \text{for all } x \in \mathbb{R}\setminus[a,b] \text{ and } t \in [0,t_{max}]. \tag{7.14}$$

For the parameters

$$\alpha := \frac{b-a}{2\pi} \quad \text{and} \quad \beta := a$$

we introduce the complex-valued function

$$f\colon \mathbb{R}\times\mathbb{R} \to \mathbb{C},\ f(y,\tau) := \psi(\alpha y+\beta, \varepsilon\tau) \quad \text{for all } (y,\tau)\in\mathbb{R}\times\mathbb{R}.$$

Note that for $\tau\in\mathbb{R}$ we have

$$f(0,\tau) = \psi(a,\tau) \quad \text{and} \quad f(2\pi,\tau) = \psi(b,\tau).$$

In particular, f satisfies the following equation:

$$\begin{cases} i\partial_\tau f(y,\tau) = -\dfrac{\varepsilon^2}{2\mu}\dfrac{1}{\alpha^2} f''(y,\tau) + V(\alpha y+\beta) f(y,\tau), \\ f(y,0) = \psi_0(\alpha y+\beta). \end{cases} \tag{7.15}$$

7.7.2 Approximation by trigonometric polynomials

Let f be the solution to (7.15). Our assumption in (7.14) yields that $f(y,\tau)$ is negligible outside $[0,2\pi]$ for all times $\tau\in[0,t_{max}/\varepsilon]$. For those times we are interested in an approximation of f in $[0,2\pi]$ by trigonometric polynomials.

The discrete Fourier transform

We denote by $\mathcal{F}\colon \mathbb{C}^d\to\mathbb{C}^d$ the discrete Fourier transform:

$$\hat v = \mathcal{F}v \quad \text{with} \quad \hat v_k = \sum_{j=1}^{d} v_j e^{-2\pi i (j-1)(k-1)/d}, \quad k = 1,\dots,d.$$

In particular, the inverse transform $\mathcal{F}^{-1}\colon \mathbb{C}^d\to\mathbb{C}^d$ is given by

$$v = \mathcal{F}^{-1}\hat v \quad \text{with} \quad v_j = \frac{1}{d}\sum_{k=1}^{d} \hat v_k e^{2\pi i (j-1)(k-1)/d}, \quad j = 1,\dots,d.$$

Lemma 93. *Let K be a power of 2 and $c := \left(c_{-K/2},\dots,c_{K/2-1}\right)^T \in \mathbb{C}^K$ a given vector. Moreover, let*

$$v_j := \sum_{k=-K/2}^{K/2-1} c_k e^{2\pi i (j-1)k/K}, \quad j = 1,\dots,K,$$

and $\sigma\colon \mathbb{C}^K\to\mathbb{C}^K$ the linear map defined by

$$\sigma(c) := \left(c_0,\dots,c_{K/2-1},c_{-K/2},\dots,c_{-1}\right)^T.$$

Then, the discrete Fourier transform of v is given by

$$\mathcal{F}v = K\sigma(c).$$

Proof. For $k = 1, \ldots, K$ we obtain

$$\begin{aligned}
\hat{v}_k &= \sum_{j=1}^{K} v_j e^{-2\pi i(j-1)(k-1)/K} \\
&= \sum_{j=1}^{K} \left(\sum_{\nu=-K/2}^{K/2-1} c_\nu e^{2\pi i(j-1)\nu/K} \right) e^{-2\pi i(j-1)(k-1)/K} \\
&= \sum_{\nu=-K/2}^{K/2-1} c_\nu \sum_{j=1}^{K} \left(e^{2\pi i(\nu-k+1)/K} \right)^{j-1},
\end{aligned}$$

and by the formula of the geometric sum we find

$$\sum_{j=1}^{K} \left(e^{2\pi i(\nu-k+1)/K} \right)^{j-1} = \begin{cases} K, & \text{if } (\nu - k + 1) \in \{0, -K\}, \\ 0, & \text{else.} \end{cases}$$

Consequently, we conclude that

$$\hat{v}_k = \begin{cases} c_{k-1}, & \text{if } k \in \{1, \ldots, K/2\}, \\ c_{k-1-K}, & \text{if } k \in \{K/2+1, \ldots, K\}, \end{cases}$$

and therefore $\mathcal{F}v = \hat{v} = K\sigma(c)$. □

Collocation

Lemma 94. *Let K be a power of 2. For all times $\tau \in \mathbb{R}$, let $f_K(\tau)$ be a trigonometric polynomial of degree $K/2$, that is, there exists a time-dependent vector*

$$c(\tau) = \left(c_{-K/2}(\tau), \ldots, c_{K/2-1}(\tau) \right)^T \in \mathbb{C}^K$$

such that

$$f_K(y, \tau) = \sum_{k=-K/2}^{K/2-1} c_k(\tau) e^{iky} \quad \textit{for all } (y, \tau) \in \mathbb{R} \times \mathbb{R}.$$

Moreover, define the K equidistant grid points

$$y_j := (j-1)\frac{2\pi}{K}, \quad j = 1, \ldots, K.$$

Then, the trigonometric polynomial f_K satisfies the equation

$$i\partial_\tau f_K(y_j, \tau) = -\frac{\varepsilon^2}{2\mu} \frac{1}{\alpha^2} f_K''(y_j, \tau) + V(\alpha y_j + \beta) f_K(y_j, \tau), \tag{7.16}$$

for $\tau \in \mathbb{R}$ and all $j = 1, \dots, K$, if and only if

$$i\sigma(\dot{c}) = \frac{\varepsilon^2}{2\mu}\frac{1}{\alpha^2}D^2\sigma(c) + \mathcal{F}V\mathcal{F}^{-1}\sigma(c),$$

where $\dot{c}(\tau) = \partial_\tau c(\tau)$ and the matrices $D, V \in \mathbb{R}^{K\times K}$ are given by

$$D = \operatorname{diag}\big(0, \dots, K/2-1, -K/2, \dots, -1\big) \quad \text{and} \quad V = \operatorname{diag}\big(V(\alpha y_j + \beta)\big).$$

Proof. For $\tau \in \mathbb{R}$ and $j = 1, \dots, K$ we introduce

$$u_j(\tau) := f_K(y_j, \tau),$$
$$u_j^\Delta(\tau) := f_K''(y_j, \tau) = -\sum_{k=-K/2}^{K/2-1} k^2 c_k(\tau) e^{iky_j}.$$

Using the vectors u and u^Δ, we can rewrite the system in (7.16) equivalently as

$$i\dot{u} = -\frac{\varepsilon^2}{2\mu}\frac{1}{\alpha^2}u^\Delta + Vu.$$

In particular, by Lemma 93 we conclude that

$$\mathcal{F}u = K\sigma(c), \;\; \mathcal{F}\dot{u} = K\sigma(\dot{c}), \quad \text{as well as} \quad \mathcal{F}u^\Delta = -D^2\mathcal{F}u.$$

Therefore we obtain

$$i\dot{u} = \frac{\varepsilon^2}{2\mu}\frac{1}{\alpha^2}\mathcal{F}^{-1}D^2\mathcal{F}u + Vu. \tag{7.17}$$

Hence, applying the Fourier transform on both sides, we finally conclude that

$$i\dot{\sigma}(c) = \frac{\varepsilon^2}{2\mu}\frac{1}{\alpha}D^2\sigma(c) + \mathcal{F}V\mathcal{F}^{-1}\sigma(c).$$

□

7.7.3 The split-step Fourier method

For an initial vector $u_0 \in \mathbb{C}^K$, we consider the following system of ordinary differential equations (cf. (7.17)):

$$i\dot{u} = \frac{\varepsilon^2}{2\mu}\frac{1}{\alpha^2}\mathcal{F}^{-1}D^2\mathcal{F}u + Vu, \quad u(0) = u_0. \tag{7.18}$$

We denote by $F \in \mathbb{C}^{K\times K}$ the complex Fourier matrix defined by

$$F_{j,k} = \omega^{(j-1)(k-1)}, \quad \text{where} \quad \omega := e^{-2\pi i/K}.$$

In particular, we have $\mathcal{F}v = Fv$ for all $v \in \mathbb{C}^K$. Moreover, let us define the matrices

$$D' := \frac{\varepsilon^2}{2\mu}\frac{1}{\alpha^2}D^2 \in \mathbb{R}^{K\times K} \quad \text{and} \quad T := F^{-1}D'F \in \mathbb{C}^{K\times K}.$$

Then, the unique solution to the IVP (7.18) is given by

$$u(t) = e^{-it(T+V)}u_0 \quad \text{for all } t \in \mathbb{R}.$$

Strang splitting

Define the grid points

$$x_j := \alpha y_j + \beta = a + (j-1)\frac{b-a}{K}, \quad j = 1, \dots, K,$$

and recall that for all $t \in [0, t_{max}]$ the following approximation holds:

$$u_j(t/\varepsilon) = f_K(y_j, t/\varepsilon) \approx f(y_j, t/\varepsilon) = \psi(x_j, t).$$

Moreover, for $m \in \mathbb{N}$ let

$$\Delta t := \frac{t_{max}}{m} \quad \text{and} \quad t^n := n\Delta t, \quad n = 0, 1, \dots, m.$$

An approximation of $u(t^1/\varepsilon) = u(\Delta t/\varepsilon)$ is given by

$$u^1 := \exp\left(-i\frac{\Delta t}{2\varepsilon}V\right)\exp\left(-i\frac{\Delta t}{\varepsilon}T\right)\exp\left(-i\frac{\Delta t}{2\varepsilon}V\right)u_0. \tag{7.19}$$

In particular, using that $T = F^{-1}D'F$, we get

$$\exp\left(-i\frac{\Delta t}{\varepsilon}T\right) = F^{-1}\exp\left(-i\frac{\Delta t}{\varepsilon}D'\right)F.$$

For a vector $v \in \mathbb{C}^K$, let us introduce

$$v^+ := \exp\left(-i\frac{\Delta t}{2\varepsilon}V\right)v \quad \text{and} \quad v^{++} := \exp\left(-i\frac{\Delta t}{\varepsilon}D'\right)v,$$

and let $d \in \mathbb{R}^K$ be the real-valued vector defined by

$$d_j := \frac{\varepsilon}{2\mu}\frac{(2\pi)^2}{(b-a)^2} \cdot \begin{cases} (j-1)^2, & \text{if } j \in \{1, \dots, K/2\}, \\ (j-1-K)^2, & \text{if } j \in \{K/2+1, \dots, K\}. \end{cases}$$

Note that the components $j = 1, \dots, K$ of v^+ and v^{++} are given by

$$v_j^+ = \exp\left(-i\frac{\Delta t}{2\varepsilon}V(x_j)\right)v_j \quad \text{and} \quad v_j^{++} = \exp\Big(-i\Delta t \cdot d_j\Big)v_j.$$

Therefore, iteration of (7.19) yields the following algorithm:

Reference solver (Split-step Fourier method)

Input: $u^n \in \mathbb{R}^K$ (approximation to the solution u of (7.18) at time t^n/ε)
Output: $u^{n+1} = v^+$ (approximation at time t^{n+1}/ε).

```
1. set $v = u^n$;
2. calculate $v^+$;
3. calculate $v = Fv^+$; (via FFT)
4. calculate $v^{++}$;
5. calculate $v = F^{-1}v^{++}$; (via IFFT)
6. calculate $v^+$;
```

7.7.4 Computation of expected values

We are interested in the numerical computation of the expected values

$$\langle \hat{q} \rangle_t := \int_{\mathbb{R}} x|\psi(x,t)|^2 \,\mathrm{d}x \quad \text{and} \quad \langle V \rangle_t := \int_{\mathbb{R}} V(x)|\psi(x,t)|^2 \,\mathrm{d}x,$$

$$\langle \hat{p} \rangle_t := \int_{\mathbb{R}} p|\mathcal{F}_\varepsilon\psi(p,t)|^2 \,\mathrm{d}p \quad \text{and} \quad \langle T \rangle_t := \frac{1}{2\mu}\int_{\mathbb{R}} p^2|\mathcal{F}_\varepsilon\psi(p,t)|^2 \,\mathrm{d}p,$$

$$\langle H \rangle_t := \langle T + V \rangle_t = \langle T \rangle_t + \langle V \rangle_t.$$

For the numerical computation of $\langle \hat{q} \rangle_{t^n}$ and $\langle V \rangle_{t^n}$ at time t^n we use that

$$u_j^n \approx \psi(x_j, t^n)$$

and approximate the integrals over position space via the composite trapezoidal rule based on the grid points x_j. Furthermore, to compute the integrals in momentum space, we use that the Fourier transform $\mathcal{F}_\varepsilon\psi(p,t^n)$ can be approximated for all $p \in \mathbb{R}$ as

$$\begin{aligned}\mathcal{F}_\varepsilon\psi(p,t^n) &\approx \frac{1}{\sqrt{2\pi\varepsilon}}\frac{b-a}{K}\sum_{j=1}^{K}\psi(x_j,t^n)e^{-\frac{i}{\varepsilon}px_j} \\ &= e^{-\frac{i}{\varepsilon}pa}\cdot\frac{1}{\sqrt{2\pi\varepsilon}}\frac{b-a}{K}\sum_{j=1}^{K}u_j^n e^{-\frac{i}{\varepsilon}p(j-1)\frac{b-a}{K}}.\end{aligned}$$

Consequently, using the following grid points in momentum space

$$p_k := \frac{2\pi\varepsilon}{b-a}\cdot\begin{cases} k-1, & \text{if } k \in \{1,\dots,K/2\}, \\ k-1-K, & \text{if } k \in \{K/2+1,\dots,K\}, \end{cases}$$

we conclude that

$$\begin{aligned}\mathcal{F}_\varepsilon\psi(p_k, t^n) &\approx e^{-\frac{i}{\varepsilon}p_k a} \cdot \frac{1}{\sqrt{2\pi\varepsilon}}\frac{b-a}{K}\sum_{j=1}^{K} u_j^n e^{-2\pi i(k-1)(j-1)/K} \\ &= e^{-\frac{i}{\varepsilon}p_k a} \cdot \frac{1}{\sqrt{2\pi\varepsilon}}\frac{b-a}{K}(\mathcal{F}u^n)_k,\end{aligned}$$

where $\mathcal{F}u^n$ can be computed via the FFT. For the computation of $\langle\hat{p}\rangle_{t^n}$ and $\langle T\rangle_{t^n}$ we then approximate the integrals over momentum space via the composite trapezoidal rule based on the grid points p_k.

7.8 Clenshaw Algorithm

The direct computation of the Chebyshev expansion in (6.11) based on the usual summation algorithm has two disadvantages: (1) all summands have to be kept in the memory of the computer, which can be very expensive in practical applications since the tensors $T_k(\mathcal{H}_0)W_0$ (and also their low-rank approximations) are typically large objects, and (2) it is known that the worst-case error generated by the floating point operations grows proportionally to the number N of summands, see e.g. [Hig02, Chapter 4]. We therefore use the Clenshaw algorithm, see [Cle55], which offers a stable alternative to evaluate linear combinations of polynomials that satisfy a linear recurrence relation such as the Chebyshev polynomials, see e.g. [FP68, Chapter 3.11].

Assuming that for given coefficients $c_0, c_1, \dots, c_{N-1} \in \mathbb{C}$ we are interested in the value of the partial Chebyshev sum in (6.7), the Clenshaw algorithm replaces the summation by the evaluation of the following backward recurrence system

$$\begin{cases} B_r(y) = 2yB_{r+1}(y) - B_{r+2}(y) + c_r, & r = N-1, \dots, 0, \\ B_N(y) = 0, \quad B_{N+1}(y) = 0, \end{cases}$$

and then expresses the partial Chebyshev sum as

$$\sum_{k=0}^{N-1}(2-\delta_{k,0})c_k T_k(y) = B_0(y) - B_2(y).$$

To obtain the approximation of the solution $\psi(t)$, we adapted the Clenshaw algorithm by first solving the backward recurrence system

$$\begin{cases} B_r = 2\mathcal{H}_0 B_{r+1} - B_{r+2} + (-i)^r J_r(t^-)W_0, & r = N-1, \dots, 0, \\ B_N = 0, \quad B_{N-1} = 0, \end{cases}$$

and then computing the approximation

$$\psi(t) \approx e^{-it^+}\left(B_0 - B_2\right).$$

Note that this numerically stable procedure needs to keep only three tensors in memory.

Bibliography

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards Applied Mathematics Series. U.S. Government Printing Office, Tenth edition, 1964.

[BBCN17] M. Berra, I. M. Bulai, E. Cordero, and F. Nicola. Gabor frames of Gaussian beams for the Schrödinger equation. *Applied and Computational Harmonic Analysis*, 43(1):94–121, 2017.

[BBGK71] V. Bargmann, P. Butera, L. Girardello, and J. R. Klauder. On the completeness of the coherent states. *Reports on Mathematical Physics*, 2(4):221–228, 1971.

[BD93] G. Baszenki and F.-J. Delvos. Multivariate Boolean midpoint rules. In *Numerical Integration IV: Proceedings of the Conference at the Mathematical Research Institute*, International Series of Numerical Mathematics. Springer, 1993.

[BG20] S. Blanes and V. Gradinaru. High order efficient splittings for the semiclassical time-dependent Schrödinger equation. *Journal of Computational Physics*, 405:109157, 2020.

[BGZ75] H. Bacry, A. Grossmann, and J. Zak. Proof of completeness of lattice states in the kq representation. *Physical Review B*, 12(4):1118–1120, 1975.

[BIKS14] P. Bader, A. Iserles, K. Kropielnicka, and P. Singh. Effective Approximation for the Semiclassical Schrödinger Equation. *Foundations of Computational Mathematics*, 14(4):689–720, 2014.

[BIKS16] P. Bader, A. Iserles, K. Kropielnicka, and P. Singh. Efficient methods for linear Schrödinger equation in the semiclassical regime with time-dependent potential. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2193):20150733, 2016.

[BL20a] P. Bergold and C. Lasser. Fourier Series Windowed by a Bump Function. *Journal of Fourier Analysis and Applications*, 26(4):65, 2020.

[BL20b] P. Bergold and C. Lasser. The Gaussian Wave Packet Transform via Quadrature Rules. Preprint on arXiv, https://arxiv.org/abs/2010.03478, 2020.

[BL21] P. Bergold and C. Lasser. An Error Representation for the Time-Sliced Thawed Gaussian Propagation Method. Preprint on arXiv, https://arxiv.org/abs/2108.12182, 2021.

[BN71] P. L. Butzer and R. J. Nessel. *Fourier Analysis and Approximation: One Dimensional Theory*, volume 1 of *Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften*. Birkhäuser Basel, 1971.

[Boy96] J. P. Boyd. The Erfc-Log Filter and the Asymptotics of the Euler and Vandeven Sequence Accelerations. In *Proceedings of the Third International Conference on Spectral and High Order Methods*, pages 267–276, 1996.

[Boy06] J. P. Boyd. Asymptotic Fourier Coefficients for a C^∞ Bell (Smoothed-"Top-Hat") & the Fourier Extension Problem. *Journal of Scientific Computing*, 29:1–24, 2006.

[CA13] M. T. Cvitaš and S. C. Althorpe. A Chebyshev method for state-to-state reactive scattering using reactant-product decoupling: $\mathrm{OH}+\mathrm{H}_2 \to \mathrm{H}_2\mathrm{O}+\mathrm{H}$. *The Journal of Chemical Physics*, 139(6):064307, 2013.

[CDS03] M. Chiani, D. Dardari, and M. K. Simon. New exponential bounds and approximations for the computation of error probability in fading channels. *IEEE Transactions on Wireless Communications*, 2(4):840–845, 2003.

[CK90] R. D. Coalson and M. Karplus. Multidimensional variational Gaussian wave packet dynamics with application to photodissociation spectroscopy. *The Journal of Chemical Physics*, 93(6):3919–3930, 1990.

[Cle55] C. W. Clenshaw. A Note on the Summation of Chebyshev Series. *Mathematics of Computation*, 9(51):118–120, 1955.

[CNR09] E. Cordero, F. Nicola, and L. Rodino. Sparsity of Gabor representation of Schrödinger propagators. *Applied and Computational Harmonic Analysis*, 26(3):357–370, 2009.

[CR12] M. Combescure and D. Robert. *Coherent States and Applications in Mathematical Physics*. Theoretical and Mathematical Physics. Springer Cham, Second edition, 2012.

[Dau92] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), 1992.

[DIS00] T. Damour, B. R. Iyer, and B. S. Sathyaprakash. Frequency-domain P-approximant filters for time-truncated inspiral gravitational wave signals from compact binaries. *Physical Review D*, 62(8):084036, 2000.

[DLCS00] G. Dattoli, S. Lorenzutta, C. Cesarano, and D. Sacchetti. Higher order derivatives of exponential functions and generalized forms of Kampé de Fériet-Bell polynomials. Technical report, ENEA, Dipartimento Innovazione, 2000.

[DR07] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, Second edition, 2007.

[DS52] R. J. Duffin and A. C. Schaeffer. A Class of Nonharmonic Fourier Series. *Transactions of the American Mathematical Society*, 72(2):341–366, 1952.

[DT10] S. Descombes and M. Thalhammer. An exact local error representation of exponential operator splitting methods for evolutionary problems and applications to linear Schrödinger equations in the semi-classical regime. *BIT Numerical Mathematics*, 50:729–749, 2010.

[Edw82] R. E. Edwards. *Fourier Series: A Modern Introduction*, volume 1 of *Graduate Texts in Mathematics*. Springer-Verlag, Second edition, 1982.

[FF15] B. Fornberg and N. Flyer. Solving PDEs with radial basis functions. *Acta Numerica*, 24:215–258, 2015.

[FGL09] E. Faou, V. Gradinaru, and C. Lubich. Computing Semiclassical Quantum Dynamics with Hagedorn Wavepackets. *SIAM Journal on Scientific Computing*, 31(4):3027–3041, 2009.

[FL06] E. Faou and C. Lubich. A Poisson Integrator for Gaussian Wavepacket Dynamics. *Computing and Visualization in Science*, 9(2):45–55, 2006.

[FLF11] B. Fornberg, E. Larsson, and N. Flyer. Stable Computations with Gaussian Radial Basis Functions. *SIAM Journal on Scientific Computing*, 33(2):869–892, 2011.

[Fol89] G. B. Folland. *Harmonic Analysis in Phase Space*. Annals of Mathematics Studies. Princeton University Press, 1989.

[Fol99] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons, Second edition, 1999.

[FP68] L. Fox and I. B. Parker. *Chebyshev Polynomials in Numerical Analysis*. Oxford University Press, 1968.

[FR05] M. Fornasier and H. Rauhut. Continuous Frames, Function Spaces, and the Discretization Problem. *Journal of Fourier Analysis and Applications*, 11:245–287, 2005.

[FS98] H. G. Feichtinger and T. Strohmer. *Gabor Analysis and Algorithms: Theory and Applications*. Applied and Numerical Harmonic Analysis. Springer Science & Business Media, 1998.

[Gab46] D. Gabor. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[GAKS15] A. D. Godbeer, J. S. Al-Khalili, and P. D. Stevenson. Modelling proton tunneling in the adenine-thymine base pair. *Physical Chemistry Chemical Physics*, 17:13034, 2015.

[GB17] S. M. Greene and V. S. Batista. Tensor-Train Split-Operator Fourier Transform (TT-SOFT) Method: Multidimensional Nonadiabatic Quantum Dynamics. *Journal of Chemical Theory and Computation*, 13(9):4034–4042, 2017.

[GG98] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 18(3):209–232, 1998.

[GG02] E. M. Goldfield and S. K. Gray. A quantum dynamics study of $H_2 + OH \rightarrow H_2O + H$ employing the Wu-Schatz-Lendvay-Fang-Harding potential function and a four-atom implementation of the real wave packet method. *The Journal of Chemical Physics*, 117(4):1604–1613, 2002.

[GH14] V. Gradinaru and G. A. Hagedorn. Convergence of a semiclassical wavepacket based time-splitting for the Schrödinger equation. *Numerische Mathematik*, 126(1):53–73, 2014.

[GH21] M. Griebel and H. Harbrecht. Analysis of Tensor Approximation Schemes for Continuous Functions. *Foundations of Computational Mathematics*, 2021.

[GKT13] L. Grasedyck, D. Kressner, and C. Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78, 2013.

[Gla63] R. J. Glauber. Coherent and Incoherent States of the Radiation Field. *Physical Review*, 131(6):2766–2788, 1963.

[Gou56] H. W. Gould. Some Generalizations of Vandermonde's Convolution. *The American Mathematical Monthly*, 63(2):84–91, 1956.

[Grö93] K. Gröchenig. Acceleration of the frame algorithm. *IEEE Transactions on Signal Processing*, 41(12):3331–3340, 1993.

[Grö01] K. Gröchenig. *Foundations of Time-Frequency Analysis*. Applied and Numerical Harmonic Analysis. Springer Science & Business Media, 2001.

[GS00] M. Griebel and M. A. Schweitzer. A Particle-Partition of Unity Method for the Solution of Elliptic, Parabolic, and Hyperbolic PDEs. *SIAM Journal on Scientific Computing*, 22(3):853–890, 2000.

[GT85] D. Gottlieb and E. Tadmor. Recovering Pointwise Values of Discontinuous Data within Spectral Accuracy. In *Progress and Supercomputing in Computational Fluid Dynamics: Proceedings of U.S.-Israel Workshop, 1984*, pages 357–375, 1985.

[Hab70] S. Haber. Numerical Evaluation of Multiple Integrals. *SIAM Review*, 12(4):481–526, 1970.

[Hac14] W. Hackbusch. Numerical tensor calculus. *Acta Numerica*, 23:651–742, 2014.

[Hag80] G. A. Hagedorn. Semiclassical quantum mechanics. I. The $\hbar \to 0$ limit for coherent states. *Communications in Mathematical Physics*, 71(1):77–93, 1980.

[Hag98] G. A. Hagedorn. Raising and Lowering Operators for Semiclassical Wave Packets. *Annals of Physics*, 269(1):77–104, 1998.

[Hal13] B. C. Hall. *Quantum Theory for Mathematicians.* Graduate Texts in Mathematics. Springer, 2013.

[Har78] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[Hel75] E. J. Heller. Time-dependent approach to semiclassical dynamics. *The Journal of Chemical Physics*, 62(4):1544–1555, 1975.

[Hel76] E. J. Heller. Time dependent variational approach to semiclassical dynamics. *The Journal of Chemical Physics*, 64(1):63–73, 1976.

[Hel81] E. J. Heller. Frozen Gaussians: A very simple semiclassical approximation. *The Journal of Chemical Physics*, 75(6):2923–2931, 1981.

[Hig02] N. J. Higham. *Accuracy and Stability of Numerical Algorithms.* Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Second edition, 2002.

[HK84] M. F. Herman and E. Kluk. A semiclasical justification for the use of non-spreading wavepackets in dynamics calculations. *Chemical Physics*, 91(1):27–34, 1984.

[HLW03] E. Hairer, C. Lubich, and G. Wanner. Geometric numerical integration illustrated by the Störmer–Verlet method. *Acta Numerica*, 12:399–450, 2003.

[HLW06] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations.* Springer Series in Computational Mathematics. Springer Berlin, Heidelberg, Second edition, 2006.

[HRS12] S. Holtz, T. Rohwedder, and R. Schneider. On manifolds of tensors of fixed TT-rank. *Numerische Mathematik*, 120(4):701–731, 2012.

[IMKNZ00] A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna. Lie-group methods. *Acta Numerica*, 9:215–365, 2000.

[IN99] A. Iserles and S. P. Nørsett. On the solution of linear differential equations in Lie groups. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1754):983–1019, 1999.

[Jac94] D. Jackson. *The Theory of Approximation.* Colloquium Publications. American Mathematical Society, 1994. Reprint of the 1930 original.

[KK18] V. Khoromskaia and B. N. Khoromskij. *Tensor Numerical Methods in Quantum Chemistry.* De Gruyter, 2018.

[KKR15] E. Kieri, G. Kreiss, and O. Runborg. Coupling of Gaussian Beam and Finite Difference Solvers for Semiclassical Schrödinger Equations. *Advances in Applied Mathematics and Mechanics*, 7(6):687–714, 2015.

[KLY19] K. Kormann, C. Lasser, and A. Yurova. Stable Interpolation with Isotropic and Anisotropic Gaussians Using Hermite Generating Function. *SIAM Journal on Scientific Computing*, 41(6):A3839–A3859, 2019.

[KMB16] X. Kong, A. Markmann, and V. S. Batista. Time-Sliced Thawed Gaussian Propagation Method for Simulations of Quantum Dynamics. *The Journal of Physical Chemistry A*, 120(19):3260–3269, 2016.

[KU98] A. R. Krommer and C. W. Ueberhuber. *Computational Integration.* Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), 1998.

[LF03] E. Larsson and B. Fornberg. A numerical study of some radial basis function based solution methods for elliptic PDEs. *Computers and Mathematics with Applications*, 46(5):891–902, 2003.

[LL20] C. Lasser and C. Lubich. Computing quantum dynamics in the semiclassical regime. *Acta Numerica*, 29:229–401, 2020.

[LQ09] S. Leung and J. Qian. Eulerian Gaussian beams for Schrödinger equations in the semi-classical regime. *Communications in Computational Physics*, 228(8):2951–2977, 2009.

[LRT13] H. Liu, O. Runborg, and N. M. Tanushev. Error Estimates for Gaussian Beam Superpositions. *Mathematics of Computation*, 82(282):919–952, 2013.

[LRT16] H. Liu, O. Runborg, and N. M. Tanushev. Sobolev and Max Norm Error Estimates for Gaussian Beam Superpositions. *Communications in Mathematical Sciences*, 14(7):2037–2072, 2016.

[LS17] C. Lasser and D. Sattlegger. Discretising the Herman–Kluk Propagator. *Numerische Mathematik*, 137(1):119–157, 2017.

[LST18] C. Lasser, R. Schubert, and S. Troppmann. Non-Hermitian propagation of Hagedorn wavepackets. *Journal of Mathematical Physics*, 59(8):082102, 2018.

[Lub08] C. Lubich. *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis.* Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), 2008.

[MA01] L. Mauritz Andersson. Quantum dynamics using a discretized coherent state representation: An adaptive phase space method. *The Journal of Chemical Physics*, 115(3):1158–1165, 2001.

[Mal09] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way.* Elsevier, Third edition, 2009.

[Mar02] A. Martinez. *An Introduction to Semiclassical and Microlocal Analysis.* Universitext. Springer New York, 2002.

[MG16] W. Markoff and J. Grossmann. Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen. *Mathematische Annalen*, 77(2):213–258, 1916.

[ML01] G. F. Margrave and M. P. Lamoureux. Gabor deconvolution. *CREWES Research Report*, 13:241–276, 2001.

[MM94] G. Mastroianni and G. Monegato. Error Estimates for Gauss–Laguerre and Gauss–Hermite Quadrature Formulas. In *Approximation and Computation: A Festschrift in Honor of Walter Gautschi*, ISNM International Series of Numerical Mathematics, pages 421–434. Springer, 1994.

[MQ02] R. I. McLachlan and G. R. W. Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002.

[MRS10] D. J. A. McKechan, C. Robinson, and B. S. Sathyaprakash. A tapering window for time-domain templates and simulated signals in the detection of gravitational waves from coalescing compact binaries. *Classical and Quantum Gravity*, 27(8):084020, 2010.

[Ore38] O. Ore. On Functions with Bounded Derivatives. *Transactions of the American Mathematical Society*, 43(2):321–326, 1938.

[Ose11] I. V. Oseledets. Tensor-Train Decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[Ose20] I. V. Oseledets. https://www.github.com/oseledets/TT-Toolbox, 2020.

[Per71] A. M. Perelomov. On the completeness of a system of coherent states. *Theoretical and Mathematical Physics*, 6(2):156–164, 1971.

[PP02] A. Papoulis and S. U. Pillai. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill series in electrical engineering: Communications and signal processing. McGraw-Hill, Fourth edition, 2002.

[QY10] J. Qian and L. Ying. Fast Gaussian wavepacket transforms and Gaussian beams for the Schrödinger equation. *Journal of Computational Physics*, 229(20):7848–7873, 2010.

[RS75] M. Reed and B. Simon. *II: Fourier Analysis, Self-Adjointness*. Methods of Modern Mathematical Physics. Academic Press, 1975.

[SBB21] M. B. Soley, P. Bergold, and V. S. Batista. Iterative Power Algorithm for Global Optimization with Quantics Tensor Trains. *Journal of Chemical Theory and Computation*, 17(6):3280–3291, 2021.

[SBGB22] M. B. Soley, P. Bergold, A. A. Gorodetsky, and V. S. Batista. Functional Tensor-Train Chebyshev Method for Multidimensional Quantum Dynamics Simulations. *Journal of Chemical Theory and Computation*, 18(1):25–36, 2022.

[Sea91] J. B. Seaborn. *Hypergeometric Functions and Their Applications*. Texts in Applied Mathematics. Springer Science & Business Media, 1991.

[SI88] A. Sidi and M. Israeli. Quadrature methods for periodic singular and weakly singular Fredholm integral equations. *Journal of Scientific Computing*, 3(2):201–231, 1988.

[Sie39] C. L. Siegel. Einführung in die Theorie der Modulfunktionen n-ten Grades. *Mathematische Annalen*, 116(1):617–657, 1939.

[Sto32] M. H. Stone. On One-Parameter Unitary Groups in Hilbert Space. *Annals of Mathematics*, 33(3):643–648, 1932.

[Swa08] T. C. Swart. *Initial Value Representations*. Dissertation, Freie Universität Berlin, 2008.

[Tad86] E. Tadmor. The exponential accuracy of Fourier and Chebyshev differencing methods. *SIAM Journal on Numerical Analysis*, 23(1):1–10, 1986.

[Tad07] E. Tadmor. Filters, mollifiers and the computation of the Gibbs phenomenon. *Acta Numerica*, 16:305–378, 2007.

[Tan06] J. Tanner. Optimal Filter and Mollifier for Piecewise Smooth Spectral Data. *Mathematics of Computation*, 75(254):767–790, 2006.

[TE89] H. Tal-Ezer. Polynomial approximation of functions of matrices and applications. *Journal of Scientific Computing*, 4(1):25–60, 1989.

[TK84] H. Tal-Ezer and R. Kosloff. An accurate and efficient scheme for propagating the time dependent Schrödinger equation. *The Journal of Chemical Physics*, 81(9):3967–3971, 1984.

[Tre19] L. N. Trefethen. *Approximation Theory and Approximation Practice, Extended Edition.* Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), 2019.

[Tro17] S. Troppmann. *Non-Hermitian Schrödinger dynamics with Hagedorn's wave packets* . Dissertation, Technische Universität München, 2017.

[TSM85] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions.* Wiley Series in Probability and Statistics - Applied Probability and Statistics Section. John Wiley & Sons, 1985.

[Tu11] L. W. Tu. *An Introduction to Manifolds.* Universitext. Springer, New York, Second edition, 2011.

[Tuk67] J. W. Tukey. An introduction to the calculations of numerical spectrum analysis. *Spectra Analysis of Time Series*, pages 25–46, 1967.

[WRB04] G. A. Worth, M. A. Robb, and I. Burghardt. A novel algorithm for non-adiabatic direct dynamics using variational Gaussian wavepackets. *Faraday Discussions*, 127:307–323, 2004.

[WW96] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis.* Cambridge Mathematical Library. Cambridge University Press, Fourth edition, 1996.

[Zhe14] C. Zheng. Optimal Error Estimates for First-Order Gaussian Beam Approximations to the Schrödinger Equation. *SIAM Journal on Numerical Analysis*, 52(6):2905–2930, 2014.