Technische Universität München
Zentrum Mathematik
Lehrstuhl für Wissenschaftliches Rechnen

Approximation of PDEs with Underlying Continuity Equations

Ilja Klebanov

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.

Vorsitzender:		Univ.-Prof. Dr. Tim N. Hoffmann
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Caroline Lasser
	2.	Univ.-Prof. Dr. Christof Schütte, Freie Universität Berlin (nur schriftliche Beurteilung)
	3.	Univ.-Prof. Dr. Stefan Teufel, Eberhard Karls Universität Thübingen

Die Dissertation wurde am 13.04.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 02.12.2015 angenommen.

Contents

Chapter 1

Introduction and Motivation

The continuity equation, also known as *transport equation*, *advection equation* or *conservation of mass formula*, is one of the vital equations in mathematical physics. It is the basic equation of fluid mechanics and appears in its direct form e.g. in the Vlasov equation and as part of the Navier-Stokes equations. It also describes the flow of the position density of the Schrödinger equation.

In this thesis we want to present a possibility to use this equation for the numerical computation of the solution of the partial differential equation (PDE) it "originates from". The class of equations we are going to treat is given in the following definition and our presentation will be using the Schrödinger equation as the showcase PDE.

Definition 1.1 (evolutionary PDE with underlying continuity equation)**.**
Let $1 \leq p < \infty$. We will call a partial differential equation (PDE) of the form

$$\partial_t \psi_t = F_t(\psi_t), \qquad \psi_0 = \psi_{\text{in}}, \tag{1.0.1}$$

with right hand side $F_t \colon L^p(\mathbb{R}^d, \mathbb{K}) \to L^p(\mathbb{R}^d, \mathbb{K})$ an *evolutionary PDE with underlying continuity equation*, if

- it has a unique solution $(\psi_t)_{t\in\mathbb{R}} \in C^1\left(\mathbb{R}, L^p \cap C^\infty(\mathbb{R}^d, \mathbb{K}) \setminus \{0\}\right)$ for each $\psi_{\text{in}} \in L^p \cap C^\infty(\mathbb{R}^d, \mathbb{K}) \setminus \{0\}$,
- the corresponding probability density $\rho_t = \dfrac{|\psi_t|^p}{\|\psi_t\|_{L^p}^p}$ fulfills the continuity equation

$$\partial_t \rho_t + \operatorname{div} j_t = 0 \tag{1.0.2}$$

for some current $(j_t)_{t\in\mathbb{R}} \in C^1\left(\mathbb{R}, L^1 \cap C^\infty(\mathbb{R}^d, \mathbb{R}^d)\right)$.

1. INTRODUCTION AND MOTIVATION

Numerical methods for many of the mentioned PDEs face at least some of the following challenges:

(1) high dimensions,

(2) unbounded domains (e.g. $\Omega = \mathbb{R}^d$),

(3) structure preservation,

(4) high oscillations.

We will mainly address the first two difficulties, while (3) and (4) will be treated only in the case of the Schrödinger equation.

(1) and (2) make it impossible to discretize the whole domain and much effort has been spent in order to filter out the "significant part", which we will call *region of interest*, for a proper discretization. In the case of the Schrödinger equation, this was often realized by the use of so-called semiclassical methods, which, in the most simple example, transport the region of interest via classical trajectories. In [Hel81], Heller presents such a method, which chooses the approximant from a linear span of Gaussian functions, the centers of which are initially chosen in the region of interest and propagated classically. For more sophisticated methods, see e.g. [Lub08] for an overview or [Fao09] for a special example.

Semiclassical methods reach their limits, if the behavior of the quantum mechanical system differs significantly from the classical one. In these situations, classical trajectories can leave the region of interest after a short period of time.

It is surprising that only few approaches make use of the density $\rho_t = |\psi_t|^2$ to determine the region of interest. One of these methods was presented by Kormann and Larsson in [Kor12], who modified Heller's approach by adapting the region of interest and setting the centers of the Gaussian basis functions *manually*. However, this appears to be a difficult choice to make. Therefore, they decide on equidistant centers and equal widths for all Gaussians, though the method itself allows for more general settings.

Our work can be seen as an automation and generalization of their algorithm. We explain why equidistant centers, equal widths and radial symmetry are poor choices. Instead, the presented method selects suitable centers and covariance matrices and adapts them in time and space *automatically*. The trouble of choosing a proper region of interest is avoided.

We also choose the approximant from a linear span of Gaussian functions. In order to place the centers of the basis functions in the region of interest, we exploit the density function ρ_t in the following way: the higher it gets (in some region), the finer we discretize (the more basis functions we put in this region). This way, only the relevant part of $\mathbb{R}^d$ is discretized.

This formulation of the idea might not sound much different to the previous one, however, we hope that the reader will find its realization elegant:

Roughly speaking, instead of manually checking the values of ρ_t and adapting the basis functions at each time step, we use of a transport map from $\mathbb{P}_{\text{uni}}$ to $\mathbb{P}_{\rho_t}$ to determine the centers (see Sections 3.1 or 3.2 for details). The Jacobian of said transport map yields proper covariance matrices for the Gaussian basis functions, see the discussion in Section 3.1.5.

The adaptation in time happens automatically by solving properly chosen ordinary differential equations (ODEs) for the centers and the covariance matrices, which originate from the evolution of the transport map. For this purpose, it is crucial to understand the connection between the continuity equation and its characteristic ODEs, see Section 2.2. In the case of the Schrödinger equation, the resulting propagation of the centers happens along Bohmian trajectories.

The propagation of the coefficients (the approximant is a linear combination of the basis functions) does not lie in the focus of our discussion. We will use the Galerkin approximation associated with the constructed manifold, which in the case of the Schrödinger equation is referred to as Dirac-Frenkel variational principle and provides strong error estimates. It is explained in detail in Section 2.5.

The thesis is structured as follows:

In Chapter 2, we will introduce the main ideas of and some theory on the continuity equation, its probabilistic viewpoint and its connection to transport maps (Section 2.2). We also present a new method of applying the continuity equation to the construction of transport maps and sampling points (Section 2.3). This is followed by an introduction to the Schrödinger equation and its numerical treatment (Sections 2.4 and 2.5), with special focus on the Bohmian interpretation of quantum mechanics. In this context, we also illustrate the concept of phase space by explaining the ideas behind the Fourier, FBI and Wigner transforms and their connection (Section 2.6). In Section 2.7 we

study a generalization to the concept of convolutions required for the choice of a suited approximation space, while Section 2.8.2 provides us with the proper tool to analyze its approximation properties.

Chapter 3 starts with a motivational outline of how to construct a reasonable approximation space in arbitrary dimensions (Section 3.1). Once the approximation space is chosen, the resulting algorithm to solve a PDE with underlying continuity equation is determined and written down in detail in Section 3.2. Section 3.3 presents a rigorous error analysis of the constructed approximation space.

In Chapter 4 the algorithm is finally applied to the Schrödinger equation with two different potentials, followed by a short conclusion of our results in Chapter 5.

Appendix A demonstrates some connections between quantum mechanics in phase space and Bohmian mechanics, while Appendix B aims to analyze a new type of phase space transforms, which generalize the FBI transform and emerged from a collaboration with my colleagues Johannes Keller and Stephanie Troppmann.

Chapter 2

Basics and Notation

2.1 Probability Theory and Transport Maps

We will start by introducing some standard notation from probability theory.

Definition 2.1 (probability space, uniform distribution)**.** A set Ω with a σ-algebra $\mathcal{B} \subseteq 2^\Omega$ is called *measurable space* $(\Omega, \mathcal{B})$. If it is further equipped with a probability measure $\mathbb{P}\colon \mathcal{B} \to [0,1]$, it is called a *probability space* $(\Omega, \mathcal{B}, \mathbb{P})$.

If $\Omega \subseteq \mathbb{R}^d$ ($d \in \mathbb{N}_{>0}$ is the dimension), we will always assume that the σ-algebra is the Borel σ-algebra $\mathcal{B}(\Omega)$ on Ω. In this case and if the probability measure $\mathbb{P}$ is given by a probability density function $\rho \in L^1(\Omega, \mathbb{R}_{\geq 0})$, $\|\rho\|_{L^1(\Omega)} = 1$, we will denote it by $\mathbb{P}_\rho$:

$$\mathbb{P}_\rho(B) = \int_B \rho(x)\,\mathrm{d}x \qquad \text{for all} \quad B \in \mathcal{B}(\Omega).$$

$\mathbb{P}_\rho$ for $\rho \equiv 1$ on $\Omega = (0,1)^d$ will be called the *uniform distribution* and denoted by $\mathbb{P}_{\text{uni}}$.

Definition 2.2 (random variable, pushforward measure, transport map, expectation value, variance, standard deviation)**.**
Let $(\Omega_1, \mathcal{B}_1, \mathbb{P})$ be a probability space, $(\Omega_2, \mathcal{B}_2)$ a measurable space and $X\colon \Omega_1 \to \Omega_2$ a measurable map (i.e. $X^{-1}(B) \in \mathcal{B}_1$ for all $B \in \mathcal{B}_2$). Then X induces a probability measure on $(\Omega_2, \mathcal{B}_2)$ via

$$\mathbb{P}_X(B) := \mathbb{P}\left(X^{-1}(B)\right) \qquad \text{for all} \quad B \in \mathcal{B}_2$$

and we call

- X an Ω_2-valued *random variable* on Ω_1,
- $\mathbb{P}_X$ the *probability distribution* of X and X $\mathbb{P}_X$-distributed,

- $\mathbb{P}_X$ the *pushforward measure* of $\mathbb{P}$ (by X): $\quad \mathbb{P}_X = X_\# \mathbb{P}$,
- X a *transport map* from $\mathbb{P}$ to $\mathbb{P}_X$,
- $\mathbb{E}_\mathbb{P}(X) := \int_{\Omega_1} X \, \mathrm{d}\mathbb{P}$ the *expectation value* of X, if X is real-valued,
- $\mathbb{V}_\mathbb{P}(X) := \mathbb{E}_\mathbb{P}\left(\left[X - \mathbb{E}_\mathbb{P}(X) \right]^2 \right)$ the *variance* of X,
- $\sigma_\mathbb{P}(X) := \sqrt{\mathbb{V}_\mathbb{P}(X)}$ the *standard deviation* of X.

We will use the notations $\mathbb{E}_\mathbb{P}$, $\mathbb{V}_\mathbb{P}$ and $\sigma_\mathbb{P}$, if $X = \mathrm{Id}$, and $\mathbb{E}_\rho(X)$, $\mathbb{V}_\rho(X)$ and $\sigma_\rho(X)$, if $\mathbb{P}$ is given by the probability density ρ. When talking about the distribution $\mathbb{P}_X$ of a random variable $X \colon \Omega_1 \to \Omega_2$, the domain $(\Omega_1, \mathcal{B}_1, \mathbb{P})$ is often omitted and only the codomain $(\Omega_2, \mathcal{B}_2, \mathbb{P}_X)$ is mentioned.

Let us state two simple properties of transport maps:

Lemma 2.3. Let $(\Omega_1, \mathcal{B}_1, \mathbb{P}_1)$, $(\Omega_2, \mathcal{B}_2, \mathbb{P}_2)$, $(\Omega_3, \mathcal{B}_3, \mathbb{P}_3)$ be probability spaces and $X \colon \Omega_1 \to \Omega_2$, $Y \colon \Omega_2 \to \Omega_3$ be transport maps from $\mathbb{P}_1$ to $\mathbb{P}_2$, from $\mathbb{P}_2$ to $\mathbb{P}_3$ respectively, i.e. $X_\# \mathbb{P}_1 = \mathbb{P}_2$, $Y_\# \mathbb{P}_2 = \mathbb{P}_3$. Then:

(i) $Y \circ X$ is a transport map from $\mathbb{P}_1$ to $\mathbb{P}_3$.

(ii) If X is a bijection, then X^{-1} is a transport map from $\mathbb{P}_2$ to $\mathbb{P}_1$.

Proof. (i) For all $B \in \mathcal{B}_3$ we have:

$$\begin{aligned} \mathbb{P}_{Y \circ X}(B) &= \mathbb{P}_1\left(X^{-1}(Y^{-1}(B)) \right) = X_\# \mathbb{P}_1(Y^{-1}(B)) \\ &= \mathbb{P}_2(Y^{-1}(B)) = Y_\# \mathbb{P}_2(B) = \mathbb{P}_3(B). \end{aligned}$$

(ii) For all $B \in \mathcal{B}_1$ we have:

$$X^{-1}_\# \mathbb{P}_2(B) = \mathbb{P}_2(X(B)) = X_\# \mathbb{P}_1(X(B)) = \mathbb{P}_1(B).$$

□

We are going to use transport maps to generate $\mathbb{P}$-distributed points for a given probability measure $\mathbb{P}$. Before elaborating this issue, let us first specify the term "$\mathbb{P}$-distributed points". Following our definition (see below), this formulation can refer to Monte Carlo points (i.e. independent realizations of a $\mathbb{P}$-distributed random variable), quasi-Monte Carlo points or just equidistant points in $(0,1)^d$, if $\mathbb{P} = \mathbb{P}_{\mathrm{uni}}$. What all these sets of points have in common is characterized by the following definition.

Definition 2.4 (distribution of point sequences)**.** Let $\Omega \subseteq \mathbb{R}^d$ be a domain and $\mathbb{P}$ be a probability measure on Ω. A sequence of finite point sequences $\left(\left(x_1^{(N)}, \dots, x_N^{(N)}\right)\right)_{N\in\mathbb{N}}$ in Ω, often laxly denoted by the sequence $(x_1, \dots, x_N)$ or even by the points $x_1, \dots, x_N$, will be called $\mathbb{P}$-*distributed*, if for each $B \in \mathcal{B}(\Omega)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \chi_B\big(x_j^{(N)}\big) = \mathbb{P}(B)\,.$$

Remark 2.5. In the case of the points $x_j \in \Omega_2 \subseteq \mathbb{R}^d$ being independent realizations of a $\mathbb{P}_X$-distributed random variable $X\colon (\Omega_1, \mathcal{B}_1, \mathbb{P}) \to (\Omega_2, \mathcal{B}_2)$, the convergence has to be understood in an "almost surely" sense and can be proven by the law of large numbers:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \chi_B\big(x_j^{(N)}\big) = \mathbb{E}(\chi_B \circ X) = 1 \cdot \mathbb{P}(X \in B) + 0 \cdot \mathbb{P}(X \notin B) = \mathbb{P}_X(B)\,.$$

In this case (Monte Carlo points) and the one of quasi-Monte Carlo points, one can give convergence rates, which are, roughly speaking, $O(N^{-1/2})$ in the Monte Carlo case (again in an "almost surely" sense) and $O(\log(N)^d N^{-1})$ in the quasi-Monte Carlo case, see e.g. [Nie92, Theorems 1.1, 2.11 and 3.8]. For point sequences with even higher order of convergence, see [Dic10]. Equidistant points have a convergence rate of $O(N^{-1/d})$, which is known as the "curse of dimensionality", since the number of points needed to reach a certain precision grows exponentially with the dimension.

Example 2.6. *Equidistant points in* $(0,1)^d$ *given by*

$$\{y_1, \dots, y_N\} = \left\{ \left(\frac{2k_1 - 1}{2n}, \dots, \frac{2k_d - 1}{2n} \right)^T \;\middle|\; k_j \in \{1, \dots, n\}\ \forall j \right\}, \tag{2.1.1}$$

where $n \in \mathbb{N}$ *and* $N = n^d$*, are* $\mathbb{P}_{uni}$*-distributed.*

Proof. We only have to consider sets of the form $B = (0, a_1] \times \cdots \times (0, a_d]$, since they form a generator of the Borel σ-algebra on $(0,1)^d$, which is closed with respect to intersections.

$$\begin{aligned}
\frac{1}{N}\sum_{j=1}^{N} \chi_B(y_j) &= \frac{1}{n^d} \sum_{k_1,\dots,k_d=1}^{n} \prod_{\ell=1}^{d} \chi_{(0,a_l]}\left(\frac{2k_\ell - 1}{2n}\right) &&= \prod_{\ell=1}^{d} \frac{1}{n} \sum_{k_\ell=1}^{n} \chi_{(0,a_l]}\left(\frac{2k_\ell - 1}{2n}\right) \\
&= \prod_{\ell=1}^{d} \frac{1}{n} \left| \left\{ k_\ell \,:\, \frac{2k_\ell - 1}{2n} \le a_\ell \right\} \right| &&= \prod_{\ell=1}^{d} \frac{1}{n} \left| \left\{ k_\ell \,:\, k_\ell \le \frac{2na_\ell + 1}{2} \right\} \right| \\
&= \prod_{\ell=1}^{d} \frac{1}{n} \left\lfloor \frac{2na_\ell + 1}{2} \right\rfloor &&\xrightarrow{n\to\infty} \prod_{\ell=1}^{d} a_\ell = \mathbb{P}_{\mathrm{uni}}(B).
\end{aligned}$$

□

Let us return to the issue of generating $\mathbb{P}$-distributed points for a given probability measure $\mathbb{P}$. Naturally, transport maps X from $\mathbb{P}_1$ to $\mathbb{P}_2$ can be used to transform $\mathbb{P}_1$-distributed points into $\mathbb{P}_2$-distributed, which is the statement of the following proposition.

Proposition 2.7. Let $\Omega_1, \Omega_2 \subseteq \mathbb{R}^d$, $\mathbb{P}$ be a probability measure on Ω_1, $X\colon \Omega_1 \to \Omega_2$ a (Borel-) measurable map and $x_1, \dots, x_N$ $\mathbb{P}$-distributed points in Ω_1.
Then $X(x_1), \dots, X(x_N)$ are $\mathbb{P}_X$-distributed points in Ω_2.

Proof. For each $B \in \mathcal{B}(\Omega_2)$ we have:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \chi_B\left(X(x_j)\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \chi_{[X^{-1}(B)]}\left(x_j\right) = \mathbb{P}\left(X^{-1}(B)\right) = \mathbb{P}_X(B)\,.$$

□

The natural question arising here is, of course, how to construct a transport map from $\mathbb{P}_1$ to $\mathbb{P}_2$, once the probability measures are given[1].

We will always assume that we are able to produce $\mathbb{P}_{\text{uni}}$-distributed points (e.g. by choosing equidistant points as in (2.1.1)), therefore and in view of Proposition 2.7, we are mostly interested in a transport map from $\mathbb{P}_{\text{uni}}$ to $\mathbb{P}$ and, if it is bijective, in its inverse. Let us give a characterization of this kind of transport maps:

Lemma 2.8. Let $\Omega \subseteq \mathbb{R}^d$ be a domain, $\rho \in C(\Omega)$ be a probability density function and $R \in C^1\left(\Omega, (0,1)^d\right)$ be a C^1-diffeomorphism. Then R is a transport map from $\mathbb{P}_\rho$ to $\mathbb{P}_{\text{uni}}$ if and only if $|\det DR| = \rho$.

Proof. By the transformation formula we obtain for each $B \in \mathcal{B}(\Omega)$:

$$\mathbb{P}_{\text{uni}}(B) = \int_B 1 \,\mathrm{d}y = \int_{R^{-1}(B)} |\det DR(x))| \,\mathrm{d}x \qquad \text{and}$$
$$R_\# \mathbb{P}_\rho(B) = \mathbb{P}_\rho\left(R^{-1}(B)\right) = \int_{R^{-1}(B)} \rho(x) \,\mathrm{d}x.$$

Therefore the continuity of ρ and DR implies that the equality $\mathbb{P}_{\text{uni}} = R_\#\mathbb{P}_\rho(B)$ holds if and only if $\rho = |\det DR|$. □

[1] One could also ask for an "optimal" transport map from $\mathbb{P}_1$ to $\mathbb{P}_2$, where optimality is usually defined by the minimization of a cost functional, e.g. $c(X) = \int_{\Omega_1} |X(x) - x|^2 \,\mathrm{d}\mathbb{P}_1(x)$. This is the standard question of Optimal Transport, see e.g. [Vil08], which will not concern us here.

This lemma already answers the question as to how to produce $\mathbb{P}_\rho$-distributed points in one dimension:

Corollary 2.9. Let $\rho \in C((a,b), \mathbb{R}_{>0})$ be a probability density function ($-\infty \leq a < b \leq \infty$) and $R\colon (a,b) \to (0,1)$ its cumulative distribution function. Then R is a C^1-diffeomorphism and R^{-1} is a transport map from $\mathbb{P}_{\text{uni}}$ to $\mathbb{P}_\rho$.

Proof. Since ρ is positive and continuous, R is a C^1-diffeomorphism. Lemma 2.8 and Lemma 2.3 yield the transport map property of R^{-1}. □

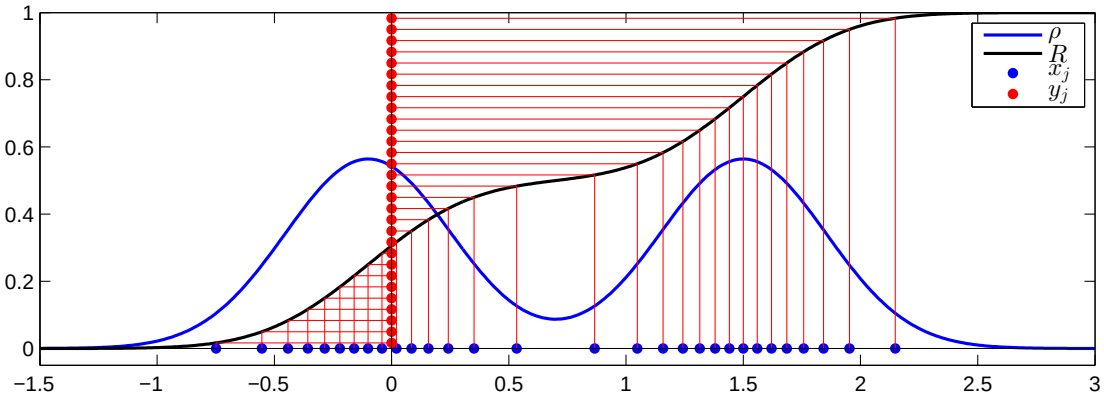


Figure 2.1: The inverse of the cumulative distribution function R of ρ is a transport map from $\mathbb{P}_{\text{uni}}$ to $\mathbb{P}_\rho$. Therefore the points $x_j = R^{-1}(y_j)$ are $\mathbb{P}_\rho$-distributed, if the points y_j are $\mathbb{P}_{\text{uni}}$-distributed (here equidistant in $(0,1)$).

Constructing transport maps from $\mathbb{P}_1$ to $\mathbb{P}_2$ in higher dimensions is a much harder task, since finding an "anti-derivative" R of ρ with $|\det DR| = \rho$ is complicated, and will be treated in a special case in Section 2.3. We will use the continuity equation as a technique for the construction. The continuity equation is crucial for this thesis and will be dealt with in the following chapter.

2.2 The Continuity Equation

The continuity equation is probably the most essential tool in this thesis, since a lot of theoretical and practical results are based on it. We will use it to

- explain and formalize Bohmian Mechanics (see Section 2.4),

- derive the theory of quasi-Bohmian trajectories (see Section 2.7.4 and Definition 2.56),
- build up the initial set of $\mathbb{P}_0$-distributed points $x_{j,0},\ j = 1, \dots, N$, by constructing a transport map R_0^{-1} from $\mathbb{P}_{\text{uni}}$ to $\mathbb{P}_0$ (see Section 2.3),
- propagate $\mathbb{P}_t$-distributed points $x_{j,t}$, the transport map R_t and its Jacobian $D_x R_t$ in time (see Proposition 2.11, Proposition 2.12 and Corollary 2.15 below).

The continuity equation describes the continuous flow of a conserved quantity along a velocity field v_t in terms of its density function ρ_t, such as a liquid flowing in a stream of water, electrical charge moving in an electrical field or the probability distribution describing the position of one or more objects. The third example is the one which will be crucial for us; therefore, we will choose it in order to present the ideas of the continuity equation, for which we will discuss two possible points of view:

The first starts with an initial random variable $X_{\text{in}} \in \mathbb{R}^d$ describing the (collective) position of objects (most often $d = 3n$, where n is the number of objects), which is $\mathbb{P}_{\rho_{\text{in}}}$-distributed, $\rho_{\text{in}} : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ being a probability density function. Assume that the movement of these objects is governed by a well-known (possibly time-dependent) velocity field $v_t : \mathbb{R}^d \to \mathbb{R}^d$, i.e. we consider the initial value problem

$$\dot{X}_t = v_t(X_t) \quad , \quad X_0 = X_{\text{in}} \tag{2.2.1}$$

(since X_{in} is a random variable, all X_t also become random variables).
In this case, the probability density function ρ_t of X_t will be governed by the continuity equation

$$\partial_t \rho_t = -\operatorname{div}(\rho_t v_t) \quad , \quad \rho_0 = \rho_{\text{in}}, \tag{2.2.2}$$

i.e. the continuity equation tells us how the probability density will evolve in time. The vector field

$$j_t := \rho_t v_t \colon \mathbb{R}^d \to \mathbb{R}^d$$

is referred to as *current, flux, current density* or *flux density* and describes the direction and amount of the quantity which flows through a unit area perpendicular to the direction of the current per time unit.

The second point of view starts with a time-dependent probability density function $\rho_t : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, which fulfills an equation of the form (2.2.2) for some velocity field

$v_t : \mathbb{R}^d \to \mathbb{R}^d$. Let us consider a $\mathbb{P}_{\rho_{\text{in}}}$-distributed random variable X_{in}, which we want to evolve in such a way that it stays $\mathbb{P}_{\rho_t}$-distributed for all times t. Then choosing it as the solution of (2.2.1) is one correct possibility, since it will guarantee that X_t stays $\mathbb{P}_{\rho_t}$-distributed.

Remark 2.10. While in the first case the resulting continuity equation is given uniquely, the choice for v_t in (2.2.1) usually is not. Consider for example a radially symmetric and time-independent probability density $\rho_t = \rho_0 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, e.g. $\rho_t(x) = \pi^{-d/2} \exp\left(-|x|^2\right)$. Then both choices $v_t \equiv 0$ and $\tilde{v}_t(x) = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} x$ yield a $\mathbb{P}_{\rho_t}$-distributed X_t.

The connection between the ordinary differential equation (2.2.1) and the partial differential equation (2.2.2) described above has been treated in high generality by, among others, Ambrosio, DiPerna and Lions (see e.g. [Amb08a], [Amb04], [DiP89]), who treat the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_t = -\operatorname{div}(\mathbb{P}_t v_t), \qquad \mathbb{P}_{\text{in}} = \mathbb{P}_0, \qquad (t,x) \in [0,T] \times \mathbb{R}^d, \tag{2.2.3}$$

in the distributional sense, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\mathbb{P}_t = \int_{\mathbb{R}^d} v_t^T \nabla \varphi \,\mathrm{d}\mathbb{P}_t \qquad \text{for any test function} \quad \varphi \in C_c^\infty(\mathbb{R}^d).$$

We choose to quote the following result:

Proposition 2.11. Let $T > 0$ and $(\Phi_t)_{t\in[0,T]} : \mathbb{R}^d \to \mathbb{R}^d$ be the flow of the ODE (2.2.1), where

$$v = (v_t)_{t\in[0,T]} \in L^1\left([0,T], W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)\right).$$

Then for any initial probability distribution $\mathbb{P}_{\text{in}}$ the continuity equation (2.2.3) has the unique solution

$$\mathbb{P}_t = \Phi_{t\#}\mathbb{P}_{\text{in}}, \qquad \text{i.e.} \quad \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\mathbb{P}_t = \int_{\mathbb{R}^d} \varphi(\Phi_t(x)) \,\mathrm{d}\mathbb{P}_{\text{in}}(x) \quad \text{for any} \quad \varphi \in C_c^\infty(\mathbb{R}^d).$$

In other words, Φ_t transports $\mathbb{P}_{\text{in}}$ to $\mathbb{P}_t$.

Proof. See e.g. [Amb08a, Proposition 2.1]. □

In some of our applications of the continuity equation we will not have the boundedness of $v_t \in L^\infty$ and even the weaker assumption (see also [Amb08a])

$$v \in L^1\left([0,T], W^{1,\infty}_{\mathrm{loc}}(\mathbb{R}^d, \mathbb{R}^d)\right), \qquad \frac{\|v\|}{1+\|x\|} \in L^1\left([0,T], L^\infty(\mathbb{R}^d)\right)$$

can not be guaranteed.

Therefore we state another result, which relies only on v_t being locally Lipschitz. However, it *assumes* that $\mathbb{P}_t$ solves the continuity equation (2.2.3) and concludes "only" the transport property of the ODE:

Proposition 2.12. Let $(\mathbb{P}_t)_{t\in[0,T]}$ be a continuous (in the weak sense) family of Borel probability measures solving the continuity equation (2.2.3) with respect to the family of Borel vector fields $(v_t)_{t\in[0,T]}$, which satisfies the following conditions:

(i) $\displaystyle\int_0^T \int_{\mathbb{R}^d} |v_t(x)| \,\mathrm{d}\mathbb{P}_t(x)\,\mathrm{d}t < \infty,$

(ii) $\displaystyle\int_0^T \left(\sup_B |v_t| + \mathrm{Lip}(v_t, B)\right)\mathrm{d}t < \infty$ for every compact $B \subset \mathbb{R}^d$, where $\mathrm{Lip}(v_t, B)$ denotes the (minimal) Lipschitz constant of v_t in B.

Then $\mathbb{P}_0$-almost every $X_{\mathrm{in}} \in \mathbb{R}^d$ the ordinary differential equation (2.2.1) has a unique solution $(X_t)_{t\in[0,T)}$ and its flow Φ_t transports $\mathbb{P}_0$ to $\mathbb{P}_t$ for each $t \in [0,T]$.

Proof. See [Amb08b, Proposition 8.1.8] □

Remark 2.13. If the probability measures $\mathbb{P}_t = \mathbb{P}_{\rho_t}$ are given by a family of probability density functions $(\rho_t)_{t\in[0,T]} \in C^1(\mathbb{R}^{1+d}, \mathbb{R}_{\geq 0})$ and the velocity field fulfills $(v_t)_{t\in[0,T]} \in C^1(\mathbb{R}^{1+d}, \mathbb{R}^d)$, the condition (ii) in the above proposition is always fulfilled as well as the continuity of $(\mathbb{P}_t)_{t\in[0,T]}$. Condition (i) can then be replaced by

$$\rho_t v_t \in L^1(\mathbb{R}^d, \mathbb{R}^d) \qquad \text{for each } t \in [0,T].$$

Further, if the assumptions are fulfilled for every $T \in \mathbb{R}$, the result can be extended from the interval $[0,T]$ to all of $\mathbb{R}$.

Another version of this result, which is tailored to the theory of Bohmian mechanics, was proven by Teufel and Tumulka [Teu05] and shows the existence of Bohmian trajectories (we present a slightly weaker version):

Theorem 2.14. Let $(\rho_t)_{t\in\mathbb{R}}$ be a time-dependent probability density which solves the continuity equation (2.2.2) for a velocity field v_t, such that

$$J := (\rho_t, \rho_t v_t)_{t\in\mathbb{R}} \in C^1\left(\mathbb{R}^{1+d}, \mathbb{R}^{1+d}\right).$$

For $X_{\text{in}} = x \in \mathbb{R}^d \setminus \mathcal{N}_0$, where $\mathcal{N}_t := \{x \in \mathbb{R}^d \mid \rho_t(x) = 0\}$, there exists a maximal solution

$$(\Phi_t(x))_{t\in(\tau_x^-,\tau_x^+)} := (X_t)_{t\in(\tau_x^-,\tau_x^+)}$$

of (2.2.1) and we set for $t \in \mathbb{R}$:

$$\Phi_t \colon \mathbb{R}^d \to \mathbb{R}^d \cup \{\diamondsuit\}, \qquad \Phi_t(x) := \begin{cases} \Phi_t(x) & \text{if } \ t \in (\tau_x^-, \tau_x^+) \\ \diamondsuit & \text{otherwise.} \end{cases}$$

If for every $T > 0$ and $r > 0$

$$\int_0^T \int_{\Phi_t(B_r(0))\setminus\{\diamondsuit\}} \left|\rho_t(x) \operatorname{div} v_t(x)\right| + \left|\rho_t(x) v_t(x)^\intercal \frac{x}{|x|}\right| \mathrm{d}x\,\mathrm{d}t \ < \ \infty,$$

then for $\mathbb{P}_{\rho_0}$-almost every $X_{\text{in}} \in \mathbb{R}^d$ the ordinary differential equation (2.2.1) has a unique solution $(X_t)_{t\in[0,\infty)}$ and the flow Φ_t transports $\mathbb{P}_{\rho_0}$ to $\mathbb{P}_{\rho_t}$ for each $t \geq 0$:

$$\mathbb{P}_{\rho_t} = \Phi_{t\#}\mathbb{P}_{\rho_0}.$$

Proof. See [Teu05, Theorem 1]. □

Corollary 2.15. Let $(\rho_t)_{t\in\mathbb{R}}$, $(v_t)_{t\in\mathbb{R}}$ and $(\Phi_t)_{t\in\mathbb{R}}$ fulfill the assumptions of Proposition 2.12 and Remark 2.13 (or alternatively of Theorem 2.14). Further, let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and $Y_0 \in C^1\left(\Omega, \mathbb{R}^d\right)$ be a transport map from $\mathbb{P}$ to $\mathbb{P}_{\rho_0}$. Then

(i) $Y_t := \Phi_t \circ Y_0$ is a transport map from $\mathbb{P}$ to $\mathbb{P}_{\rho_t}$,

(ii) its Jacobian $\Psi_t = D_x Y_t$ solves the variational equation

$$\partial_t \Psi_t(x) = D_x v_t(Y_t(x)) \cdot \Psi_t(x).$$

Proof. (i) is a straightforward application of Lemma 2.3 to the results of Proposition 2.12 (or alternatively Theorem 2.14).

(ii) The differentiability of Y_t is guaranteed by $Y_0 \in C^1$ and $\Phi_t \in C^1$ (the latter follows from $v_t \in C^1$). A simple computation shows:

$$\partial_t D_x Y_t = D_x \partial_t(\Phi_t \circ Y_0) = D_x\left(v_t \circ Y_t\right) = \left[(D_x v_t) \circ Y_t\right] \cdot D_x Y_t.$$

□

Corollary 2.15, though simple to prove, gives us the possibility to find transport maps Y_t to $\mathbb{P}_{\rho_t}$ for each $t \geq 0$, once we have found a transport map Y_0 to $\mathbb{P}_{\rho_0}$, by just solving the ordinary differential equation 2.2.1. In other words, we can produce $\mathbb{P}_{\rho_t}$-distributed points from $\mathbb{P}_{\rho_0}$-distributed points (based on Proposition 2.7)!

The last question left is the existence of the initial transport map Y_0 and how to construct such a transport map numerically, which leads us to the next section. Surprisingly, the continuity equation will play an important role once again.

2.3 Existence and Construction of Transport Maps

In this section we will address the existence of transport maps and construct a transport map in a special case. Both issues make use of the continuity equation in the following way:

(1) Assume we want to construct a transport map from $\mathbb{P}_{\rho_{\text{in}}}$ and $\mathbb{P}_{\rho_{\text{fin}}}$, where $\rho_{\text{in}},\ \rho_{\text{fin}} \in L^1(\mathbb{R}^d)$ are probability densities. Choose a time-dependent probability density $(\rho_t)_{t\in[0,1]}$, with $\rho_0 = \rho_{\text{in}}$ and $\rho_1 = \rho_{\text{fin}}$. The obvious choice, but by far not the only interesting one (especially for numerical computations) is

$$\rho_t = (1-t)\rho_{\text{in}} + t\rho_{\text{fin}}.$$

(2) The next step (which is the difficult one) is to find a (possibly time-dependent) velocity field $(v_t)_{t\in[0,1]}$, such that $(\rho_t)_{t\in[0,1]}$ solves the continuity equation

$$\partial_t \rho_t = -\operatorname{div}(\rho_t v_t),$$

in the upper case $-\operatorname{div}(\rho_t v_t) = \partial_t \rho_t = \rho_{\text{fin}} - \rho_{\text{in}}$.

(3) By Proposition 2.12, the flow $(\Phi_t)_{t\in[0,1]}$ of the ordinary differential equation $\dot{x}_t = v_t(x_t)$ is a transport map from $\mathbb{P}_{\rho_0}$ to $\mathbb{P}_{\rho_t}$. In particular, $\Phi_{t=1}$ transports $\mathbb{P}_{\rho_{\text{in}}}$ to $\mathbb{P}_{\rho_{\text{fin}}}$.

Let us start with a result on the existence, the idea of which was first presented by Moser in [Mos65] for compact manifolds. Roughly speaking, he follows the above steps and shows the existence of $(v_t)_{t\in[0,1]}$. For a summary see e.g. [Vil08, Appendix to Chapter 1]). It was later generalized to non-compact manifolds by Greene and Shiohama.

Lemma 2.16. Let $\rho_1, \rho_2 \in C^\infty\left((0,1)^d, \mathbb{R}_{>0}\right)$ be two positive probability density functions. Then there exists a transport map $X \in C^\infty\left((0,1)^d, (0,1)^d\right)$ from $\mathbb{P}_{\rho_1}$ to $\mathbb{P}_{\rho_2}$.

Proof. Since ρ_1 and ρ_2 are probability densities,

$$\rho_1(B) < \infty \qquad \text{and} \qquad \rho_2(B) < \infty$$

for all $B \in \mathcal{B}\left((0,1)^d\right)$. The claim follows from [Gre79, Theorem 1]. □

Proposition 2.17. Let $\rho \in C^\infty(\mathbb{R}^d, \mathbb{R}_{>0})$ be a positive probability density function. Then there exists a transport map $R \in C^\infty\left(\mathbb{R}^d, (0,1)^d\right)$ from $\mathbb{P}_\rho$ to $\mathbb{P}_{\text{uni}}$.

Proof. We will prove this result by first pushing forward $\mathbb{P}_\rho$ to a proper probability measure $\mathbb{P}_{\tilde{\rho}}$ on $(0,1)^d$ and then apply the upper lemma.

Consider a Gaussian density $g(z) = \pi^{-1/2}\, e^{-z^2}$ in one dimension and its cumulative distribution function $G(x) = \frac{1}{2}\,(1 + \operatorname{erf}(x))$. Then

$$X_1 \colon \mathbb{R}^d \to (0,1)^d, \qquad X_1(x) := (G(x_1), \dots, G(x_d))$$

is a C^∞-diffeomorphism and a transport map from $\mathbb{P}_\rho$ to $\mathbb{P}_{\tilde{\rho}}$, where

$$\tilde{\rho} = \frac{\rho}{|\det D_x X_1|} \circ X_1^{-1},$$

since the transformation formula implies for all $A \in \mathcal{B}\left((0,1)^d\right)$

$$\begin{aligned} \int_A \left(\frac{\rho}{|\det D_x X_1|} \circ X_1^{-1} \right)(y)\, \mathrm{d}y &= \int_{X_1^{-1}(A)} \frac{\rho(x)}{|\det D_x X_1(x)|}\, |\det D_x X_1(x)| \\ &= \int_{X_1^{-1}(A)} \rho(x)\, \mathrm{d}x. \end{aligned}$$

From $|\det D_x X_1(x)| = \prod_{j=1}^d g(x_j) = \pi^{-d/2} e^{-|x|^2}$ we deduce that $\tilde{\rho} \in C^\infty\left((0,1)^d, \mathbb{R}_{>0}\right)$ is a smooth and positive probability density, therefore there exists a transport map $X_2 \in C^\infty\left((0,1)^d, (0,1)^d\right)$ from $\mathbb{P}_{\tilde{\rho}}$ to $\mathbb{P}_{\text{uni}}$ by Lemma 2.16. As a consequence,

$$Y = X_2 \circ X_1 \in C^\infty\left(\mathbb{R}^d, (0,1)^d\right)$$

is a transport map from $\mathbb{P}_\rho$ to $\mathbb{P}_{\text{uni}}$ by Lemma 2.3(i). □

Now let us discuss how to produce $\mathbb{P}_\rho$-distributed points by constructing a transport map from $\mathbb{P}_{\text{uni}}$ to $\mathbb{P}_\rho$.

We will treat the special case of ρ being a weighted sum of normal distributions. In the general case of $\rho \in \mathcal{S}(\mathbb{R}^d)$, one would have to first approximate ρ by a weighted sum of normal distributions and then apply our construction, which in this case would only give an approximate transport map.

Let us start with an introductory example:

Example 2.18. *Assume we are able to produce $\mathbb{P}_{\rho_{in}}$-distributed points, where*

$$\rho_{in}(x) := \rho(x) := (2\pi)^{-d/2}\, e^{-|x|^2/2}$$

is the standard normal distribution, and our aim is to sample $\mathbb{P}_{\rho_{fin}}$-distributed points, where

$$\rho_{fin}(x) = \sum_{k=1}^{K} w_k\, \rho(x - a_k) \qquad \text{with} \quad a_k \in \mathbb{R}^d,\ w_k > 0,\ \sum_{k=1}^{K} w_k = 1, \tag{2.3.1}$$

is a weighted sum of shifted standard normal distributions. Defining

$$\rho_{k,t} := \rho(x - t a_k), \qquad \rho_t := \sum_{k=1}^{K} w_k \rho_{k,t} \qquad \text{and} \qquad v_{k,t} := a_k,$$

it is easy to see that each $\rho_{k,t},\ k = 1, \dots, K$, solves a continuity equation:

$$\operatorname{div}(\rho_{k,t} v_{k,t})(x) = v_{k,t}(x)^T \nabla \rho_{k,t}(x) + \rho_{k,t}(x) \underbrace{\operatorname{div} v_{k,t}(x)}_{=\operatorname{div} x_k = 0} = x_k \nabla \rho(x - t x_k) = -\partial_t \rho_{k,t}(x).$$

By linearity of the derivatives ∂_t and div*, the probability density ρ_t solves the continuity equation*

$$\partial_t \rho_t = \sum_{k=1}^{K} w_k\, \partial_t \rho_{k,t} = -\sum_{k=1}^{K} w_k \operatorname{div}(\rho_{k,t} v_{k,t}) = -\operatorname{div} j_t \qquad \text{for} \qquad j_t := \sum_{k=1}^{K} w_k \rho_{k,t} v_{k,t}.$$

By Proposition 2.12, the flow Φ_t corresponding to the velocity field

$$v_t := \frac{j_t}{\rho_t} = \frac{\sum_{k=1}^{K} w_k \rho_{k,t} a_k}{\sum_{k=1}^{K} w_k \rho_{k,t}}$$

is a transport map from $\mathbb{P}_{\rho_0}$ to $\mathbb{P}_{\rho_t}$ for each t. In particular, $\Phi_{t=1}$ is a transport map from $\mathbb{P}_{\rho_{in}}$ to $\mathbb{P}_{\rho_{fin}}$.

A straightforward generalization of the upper ideas is given in the following proposition:

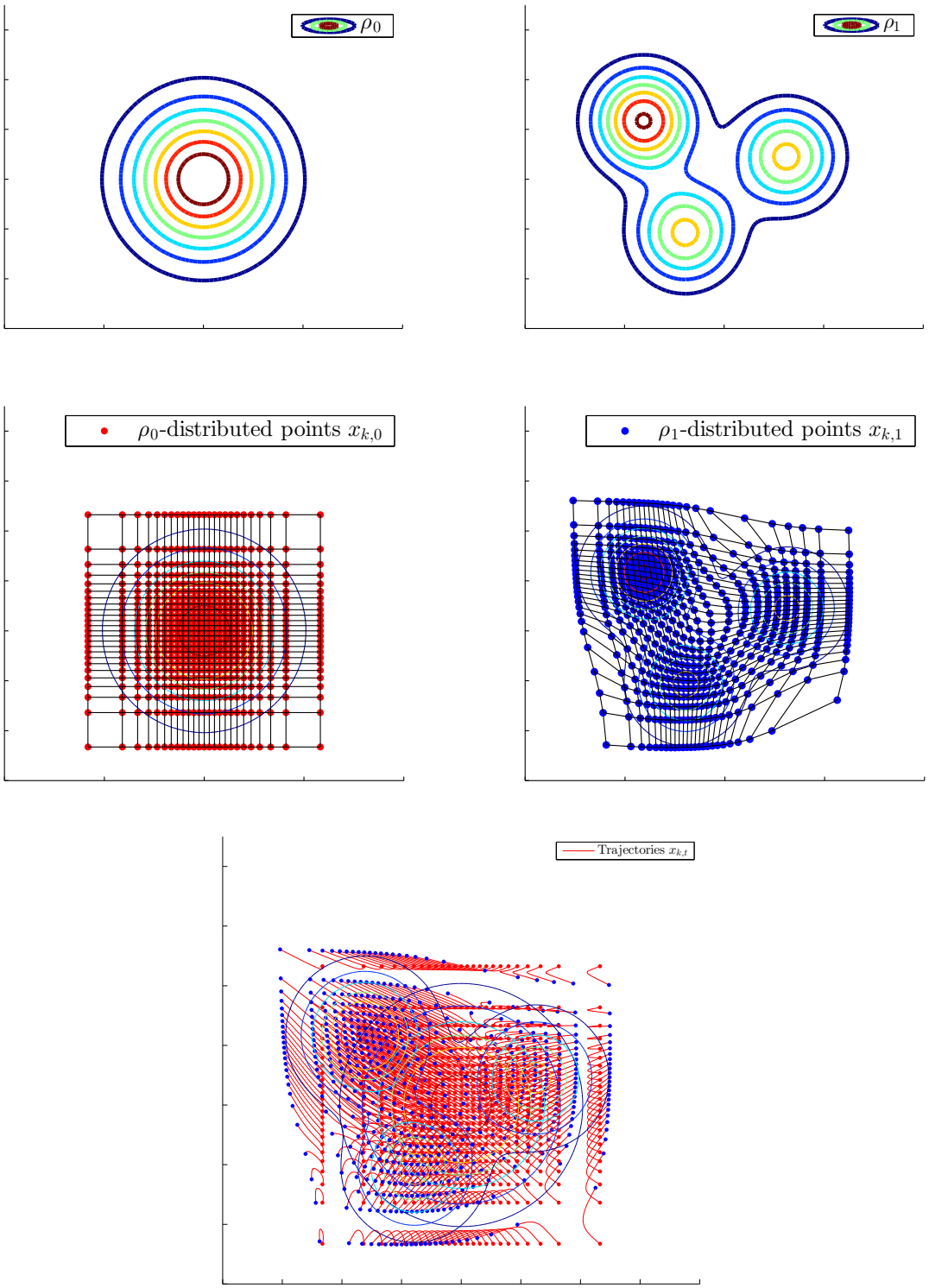


Figure 2.2: $\mathbb{P}_{\rho_0}$-distributed points are transported to $\mathbb{P}_{\rho_1}$-distributed points by solving proper initial value problems.

Proposition 2.19. Let $(\rho_{k,t})_{t\in[0,1]}$, $k = 1, \dots, K$, be time-dependent probability density functions, which fulfill the continuity equations

$$\partial_t \rho_{k,t} = -\operatorname{div}(\rho_{k,t} v_{k,t}) = -\operatorname{div} j_{k,t}, \qquad \rho_{k,0} = \rho_{\text{in}},$$

where the initial density ρ_{in} coincides for all $k = 1, \dots, K$. Then

$$\rho_t = \sum_{k=1}^{N} w_k \, \rho_{k,t}, \qquad \text{where the weights } \; w_k > 0 \; \text{ fulfill } \; \sum_{k=1}^{K} w_k = 1,$$

solves the continuity equation

$$\partial_t \rho_t = -\operatorname{div} j_t, \qquad \rho_0 = \rho_{\text{in}}, \qquad \text{for} \qquad j_t = \sum_{k=1}^{K} w_k j_{k,t}.$$

Proof. As in the upper example, this is a consequence of the linearity of differentiation:

$$\partial_t \rho_t = \sum_{k=1}^{K} w_k \, \partial_t \rho_{k,t} = -\sum_{k=1}^{K} w_k \operatorname{div}(j_{k,t}) = -\operatorname{div} j_t.$$

□

Corollary 2.20. In the situation of Proposition 2.19, if ρ_t and

$$v_t := \frac{j_t}{\rho_t} = \frac{\sum_{k=1}^{K} w_k j_{k,t}}{\sum_{k=1}^{K} w_k \rho_{k,t}}$$

fulfill the assumptions of Proposition 2.12, the flow $(\Phi_t)_{t\in\mathbb{R}}$ corresponding to the ordinary differential equation

$$\dot{x}_t = v_t(x_t) \tag{2.3.2}$$

is a transport map from $\mathbb{P}_{\rho_0}$ to $\mathbb{P}_{\rho_t}$ for each $t \in [0, 1]$.

We would like to give another application of the upper Proposition and Corollary, which is slightly more general than the previous example and is based on the following lemma:

Lemma 2.21. Let $X : (\Omega, \mathcal{B}, \mathbb{P}) \to \mathbb{R}^d$ be an $\mathcal{N}(0, \mathrm{Id})$-distributed random variable, $a \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d\times d}$ an invertible matrix. Then $Y = a + BX$ is an $\mathcal{N}(a, BB^\intercal)$-distributed random variable. In other words, $x \mapsto a + Bx$ is a transport map from $\mathcal{N}(0, \mathrm{Id})$ to $\mathcal{N}(a, BB^\intercal)$.

Proof. For all $A \in \mathcal{B}(\mathbb{R}^d)$ the transformation $y = a + Bx$ yields:

$$\begin{aligned}\mathbb{P}(Y \in A) &= \mathbb{P}\left(X \in B^{-1}(A-a)\right) \\ &= (2\pi)^{-d/2} \int_{B^{-1}(A-a)} e^{-x^\intercal x/2}\,\mathrm{d}x \\ &= (2\pi)^{-d/2} |\det(B)|^{-1} \int_A \exp\left[-\frac{1}{2}(y-a)^\intercal B^{-\intercal} B^{-1}(y-a)\right]\,\mathrm{d}y,\end{aligned}$$

i.e. $\mathbb{P}_Y$ is given by the probability density function

$$\rho(y) = (2\pi)^{-d/2} |\det(BB^\intercal)|^{-1/2} \exp\left[-\frac{1}{2}(y-a)^\intercal (BB^\intercal)^{-1}(y-a)\right].$$

□

Example 2.22. *Again, we assume that we are able to produce $\mathbb{P}_{\rho_{in}}$-distributed points, where*

$$\rho_{in}(x) := \rho(x) := (2\pi)^{-d/2}\, e^{-|x|^2/2}$$

is the standard normal distribution, and our aim is to sample $\mathbb{P}_{\rho_{fin}}$-distributed points, where ρ_{fin} is a weighted sum of shifted (not necessarily standard) normal distributions,

$$\rho_{fin}(x) = (2\pi)^{-d/2} \sum_{k=1}^{K} w_k\, |\det B_k|^{-1} \exp\left[-\frac{1}{2}(x-a_k)^\intercal (B_k B_k^\intercal)^{-1}(x-a_k)\right], \quad (2.3.3)$$

where $a_k \in \mathbb{R}^d$, $w_k > 0$, $\sum_{k=1}^K w_k = 1$ and $B_k B_k^\intercal$ is the Cholesky decomposition of the k-th covariance matrix. For $K = 1$ (or alternatively for each $k = 1, \dots, K$ separately) Lemma 2.21 (together with Proposition 2.12 and Corollary 2.15) tells us how to get $\mathbb{P}_{\rho_{fin}}$-distributed points . For each $k = 1, \dots, K$ we have:

- *The solution of the initial value problem*

$$\dot{x}_t = v_{k,t}(x_t), \qquad v_{k,t}(x) := a_k + (B_k - \mathrm{Id})x_0, \qquad x_0 = X_{in}$$

  is given by

$$x_t = x_0 + t\,[a_k + (B_k - \mathrm{Id})x_0] = ta_k + [tB_k + (1-t)\mathrm{Id}]x_0.$$

- *If X_{in} is $\mathbb{P}_{\rho_{in}} = \mathcal{N}(0, \mathrm{Id})$-distributed, then x_t is $\mathbb{P}_{\rho_{k,t}} = \mathcal{N}(ta_k, B_{k,t}B_{k,t}^\intercal)$-distributed by Lemma 2.21, where $B_{k,t} := tB_k + (1-t)\mathrm{Id}$ and*

$$\rho_{k,t}(x) = (2\pi)^{-d/2} |\det B_{k,t}|^{-1} \exp\left[-\frac{1}{2}(x-ta_k)^\intercal (B_{k,t}B_{k,t}^\intercal)^{-1}(x-ta_k)\right].$$

- *By Proposition 2.12, the probability density $(\rho_{k,t})_{t\in\mathbb{R}}$ has to be the solution of the continuity equation*

$$\partial_t \rho_{k,t} = -\operatorname{div}(j_{k,t}), \qquad j_{k,t} = \rho_{k,t} v_{k,t}, \qquad \rho_{k,0} = \rho_{in}$$

and x_1 is $\rho_{k,1}$-distributed. This can also be seen by direct computation instead of using Lemma 2.21 and Proposition 2.12.

The case $K > 1$ follows by the same argumentation as in Example 2.18 or from Proposition 2.19 and Corollary 2.20:
Both imply that the probability density function

$$\rho_t(x) = \sum_{k=1}^{K} w_k \rho_{k,t}$$

solves the continuity equation

$$\partial_t \rho_t = -\operatorname{div}(\rho_t v_t) \qquad \text{for} \qquad v_t = \frac{\sum_{k=1}^{K} w_k j_{k,t}}{\sum_{k=1}^{N} w_k \rho_{k,t}}.$$

By Proposition 2.12, the flow Φ_t corresponding to the initial value problem

$$\dot{x}_t = v_t(x_t), \qquad x_0 = x_{in},$$

is a transport map from $\mathbb{P}_{\rho_{in}} = \mathcal{N}(0, \mathrm{Id})$ to $\mathbb{P}_{\rho_t}$. In particular, $\Phi_{t=1}$ is a transport map from $\mathbb{P}_{\rho_{in}}$ to $\mathbb{P}_{\rho_{fin}}$.

Remark 2.23. So far, we have discussed how to transport $\mathcal{N}(0, \mathrm{Id})$ to a weighted sum of normal distributions. In order to (approximately) transport $\mathbb{P}_{\text{uni}}$-distributed points to $\mathbb{P}_\rho$-distributed points for a given probability density ρ, we have to

(i) approximate ρ by a weighted sum of normal distributions ρ_{fin},

(ii) transport the $\mathbb{P}_{\text{uni}}$-distributed points to $\mathcal{N}(0, \mathrm{Id})$-distributed points,

(iii) transport the $\mathcal{N}(0, \mathrm{Id})$-distributed points to $\mathbb{P}_{\rho_{\text{fin}}}$-distributed points by the upper algorithm.

The second step (ii) can be performed by the transport map

$$Y \colon (0,1)^d \to \mathbb{R}^d, \qquad Y_j(y) = \sqrt{2}\,\mathrm{erf}^{-1}\left(2y_j - 1\right), \quad j = 1, \dots, d.$$

This follows from Corollary 2.9 and the fact that the standard normal distribution factorizes. Alternatives are the Box-Muller method (see [Box58]) and comparable transformations.

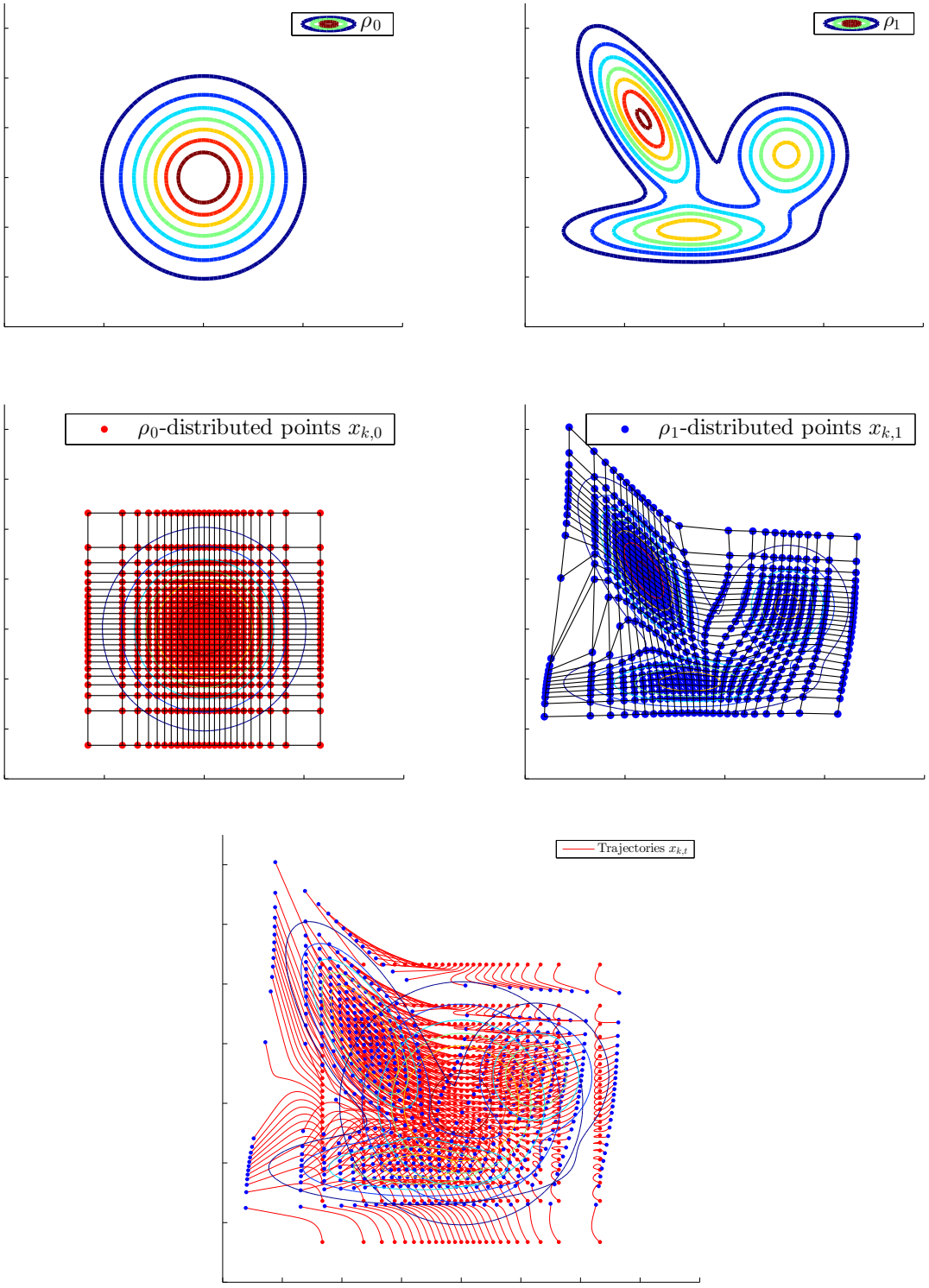


Figure 2.3: $\mathbb{P}_{\rho_0}$-distributed points are transported to $\mathbb{P}_{\rho_1}$-distributed points by solving proper initial value problems.

For our algorithm, we will not only need $\Phi_{t=1}$ for constructing $\mathbb{P}_\rho$-distributed points, but also its Jacobian $\Psi_{t=1} = D_x\Phi_{t=1}$. The latter can be computed by solving a variational equation along with the ordinary differential equation (2.3.2), which is the statement of the following proposition (this is basically a reformulation of Corollary 2.15).

Proposition 2.24. Let $(\Phi_t)_{t\in\mathbb{R}}$ be the flow corresponding to the ordinary differential equation

$$\dot{x}_t = v_t(x_t),$$

where $v_t \in C^1\left(\mathbb{R}^d, \mathbb{R}^d\right)$ is a velocity field. Then the propagation of the Jacobian $\Psi_t = D_x\Phi_t$ is given by the variational equation

$$\partial_t\Psi_t = [(D_x v_t) \circ \Phi_t] \cdot \Psi_t, \qquad \Psi_0 = E_d.$$

Proof. $\Phi_0 = \mathrm{Id}$ implies $\Psi_0 = E_d$. The rest of the proof is analogous to the one of Corollary 2.15 (with $Y_0 = \mathrm{Id}$). □

2.4 The Schrödinger Equation and Bohmian Mechanics

In quantum mechanics, the state of a d-dimensional physical system (often $d = 3n$, where n is the number of considered particles) with potential $V \in C^1(\mathbb{R}^d, \mathbb{R})$ at time t is described by the so-called *wave function* $\psi_t \in L^2(\mathbb{R}^d, \mathbb{C})$, $\|\psi_t\|_{L^2} = 1$. The time evolution of the wave function is given by the *Schrödinger equation*

$$i\partial_t\psi_t = H\psi_t \quad , \quad H = -\frac{1}{2}\Delta + V \quad , \quad \psi_0 = \psi_{in}\,. \tag{2.4.1}$$

The Schrödinger equation was introduced by Erwin Schrödinger in 1926 (see [Sch26]) in a time-independent version, i.e. in form of an eigenvalue problem.

The theory on the existence and uniqueness of solutions of (2.4.1) relies on the self-adjointness of the Hamiltonian H (see e.g. [Gus11, Theorem 2.16]). However, not all potentials V yield a self-adjoint Hamiltonian H. A discussion of this issue can be found in [Ree75]. We will always *assume* that H is a self-adjoint operator on $L^2(\mathbb{R}^d)$. Since a self-adjoint Hamiltonian implies a unitary propagation

$$\psi_t = e^{-itH}\,\psi_{\mathrm{in}}\,,$$

the L^2-norm of the wave function is preserved in time, so the above criterion $\|\psi_t\|_{L^2} = 1$ is meaningful. We will also assume that V grows at most polynomially:

$$\exists k \in \mathbb{N}\ \exists C > 0 : \qquad V(x) < C(1 + |x|)^k \quad \text{for all } x \in \mathbb{R}^d.$$

The probability density function $\rho_t(x) = |\psi_t(x)|^2$ is usually interpreted as the position density for the system to be found in state $x \in \mathbb{R}^d$ at time t (and that is what we mean by the above formulation "the state is described by the wave function"). This interpretation dates back to a paper of Max Born, in which he introduced this so-called *Born rule* in a footnote, see [Bor26], and can now be found in every textbook on quantum mechanics.

In his famous papers from 1952 (see [Boh52a] and [Boh52b]), David Bohm proposed an alternative to the standard interpretation, the so-called Copenhagen interpretation of quantum mechanics in terms of hidden variables, suggesting that the particles follow deterministic trajectories, which are governed by the wave function. The randomness described by the probability density ρ_t originates only from our lack of knowledge on the initial position of the particles.

In the resulting theory, which is usually referred to as Bohmian mechanics[1] or de Broglie-Bohm theory, the system state, previously described only by the wave function ψ_t, is extended by the (collective) position of the particle(s) $q_t \in \mathbb{R}^d$ (again $d = 3n$, where n is the number of particles, is the most common case), which evolves in time in the following way:

$$\dot{q}_t = v_t(q_t) \quad , \quad v_t = \Im\left[\frac{\nabla\psi_t}{\psi_t}\right] \quad , \quad q_0 = q_{\text{in}}. \tag{2.4.2}$$

Since the initial position q_0 is afflicted by some uncertainty, so is the position q_t at any time $t \in \mathbb{R}$. Following the Born rule, q_t may therefore be viewed as a $\mathbb{P}_{\rho_t}$-distributed random variable. In order to guarantee that q_t actually stays $\mathbb{P}_{\rho_t}$-distributed for all times $t \in \mathbb{R}$, the formula for the velocity field v_t in (2.4.2) had to be chosen in such a way, that ρ_t fulfills the corresponding continuity equation:

Proposition 2.25. The time-dependent probability density function $\rho_t = |\psi_t|^2$ solves the continuity equation

$$\partial_t \rho_t + \operatorname{div} j_t = 0, \qquad \text{where} \quad j_t = \rho_t v_t = \Im\left[\overline{\psi}_t \nabla \psi_t\right].$$

[1] For a beautiful presentation of Bohmian mechanics see e.g. [Dur09]

Proof. By the Schrödinger equation (2.4.1), we have

$$\begin{aligned}\partial_t \rho_t &= \partial_t \bar{\psi}_t\, \psi_t + \bar{\psi}_t\, \partial_t \psi_t = \frac{1}{2i}\Delta\bar{\psi}_t\, \psi_t - \frac{1}{i} V\bar{\psi}_t\, \psi_t - \frac{1}{2i}\bar{\psi}_t\, \Delta\psi_t + \frac{1}{i} V\bar{\psi}_t\, \psi_t \\ &= -\frac{1}{2i}\left(\bar{\psi}_t \Delta\psi_t - \psi_t \Delta\bar{\psi}_t\right) = -\frac{1}{2i}\operatorname{div}\left(\bar{\psi}_t \nabla\psi_t - \psi_t \nabla\bar{\psi}_t\right) = -\operatorname{div} j_t\,.\end{aligned}$$

□

Remark 2.26. In summary, if we choose q_0 as a $\mathbb{P}_{\rho_0}$-distributed random variable or N $\mathbb{P}_{\rho_0}$-distributed initial points $q_{j,0},\ j = 1, \ldots, N$, the resulting trajectories of (2.4.2) will stay $\mathbb{P}_{\rho_t}$-distributed for all $t > 0$, as long as the assumptions of Theorem 2.14 are fulfilled (for a detailed discussion, see [Teu05]). The trajectories $\left(q_{j,t}\right)_{t\in\mathbb{R}}$ are called *Bohmian trajectories*.

This property is of high interest for the numerical treatment of the Schrödinger equation, since the points $q_{j,t}$ stay in the "region of interest" (the region where ψ_t is far from zero, the higher ρ_t is in some region, the more points will be lying there) for all time, see Figure 2.4.

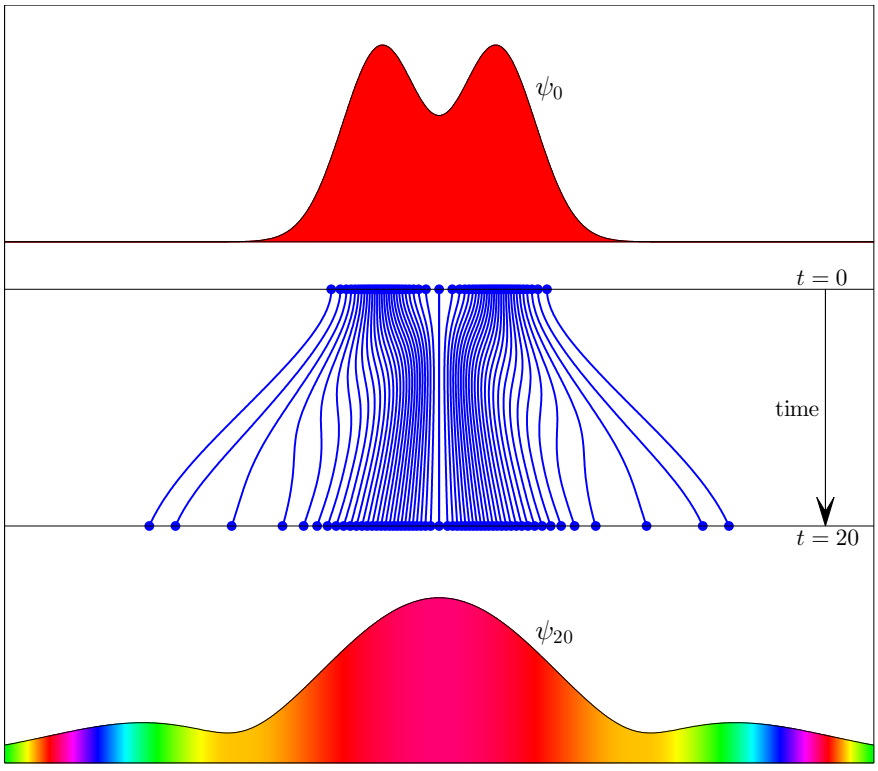


Figure 2.4: Bohmian trajectories stay $|\psi_t|^2$-distributed for all times t. The example shown here is described by formula (4.1.1).

Remark 2.27. We will use further notations for the velocity field v_t. Namely, under the additional assumption that $\psi_t(x) \neq 0$ for all $x \in \mathbb{R}^d$, we obtain from rewriting $\psi_t = \exp(T_t) = R_t \exp(S_t)$ (R_t and S_t being real-valued):

$$v_t = \Im\left[\frac{\nabla\psi_t}{\psi_t}\right] = \Im[\nabla T_t] = \nabla S_t.$$

The Schrödinger equation can be reformulated to:

$$\begin{cases} \partial_t R_t &= -\frac{1}{2} R_t \Delta S_t - \nabla R_t^\intercal \nabla S_t, \\ \partial_t S_t &= -\frac{1}{2}|\nabla S_t|^2 + \frac{\Delta R_t}{2R_t} - V, \end{cases} \tag{2.4.3}$$

$$\partial_t T_t = \frac{i}{2}\left(\Delta T_t + \nabla T_t^\intercal \nabla T_t\right) - iV. \tag{2.4.4}$$

Remark 2.28. Before Bohm's publications of the equations for v_t, they were independently discovered by Erwin Madelung in 1926 ([Mad27]) and Louis de Broglie in 1927. However, they were not well received by the scientific community, which made them fall into oblivion for over 20 years. Bohm revived, extended and stood up for this theory and by now, though still lacking acceptance, it has has taken a more prominent role.

Let us say a few words about the theory of Madelung, who gave a hydrodynamic formulation of quantum mechanics by considering PDEs for the pair ($\rho_t = R_t^2$, $v_t = \nabla S_t$), which provides a beautiful, self-consistent theory:

$$\begin{aligned} \partial_t \rho_t &= -\operatorname{div}(\rho_t v_t), \\ \partial_t v_t &= -\nabla V - \frac{1}{2}\nabla\left(|v_t|^2 + \frac{\Delta\sqrt{\rho_t}}{\sqrt{\rho_t}}\right) \end{aligned}$$

(the first equation is, of course, just the continuity equation). The flow along one trajectory is given by the ODEs

$$\begin{aligned} \dot{x}_t &= p_t := v_t(x_t), \\ \dot{p}_t &= \partial_t v_t(x_t) + D_x v_t(x_t) \cdot \underbrace{\dot{x}_t}_{v_t(x_t)} = -\nabla\left(V + \frac{\Delta\sqrt{\rho_t}}{2\sqrt{\rho_t}}\right)(x_t), \end{aligned}$$

which look a lot like classical equations of motion except for the additional term $V_{\text{quant}} := \frac{\Delta\sqrt{\rho_t}}{2\sqrt{\rho_t}}$, which is added to the potential V. This term was later referred to as *quantum potential*, meaning that it has to be added to the usual potential in order to pass from classical to quantum mechanics. Note that the quantum potential depends on ρ_t and thereby indirectly on the wave function, so the Schrödinger equation still has to be solved in some way for the computation of Bohmian trajectories (this observation also holds for the upper reformulations of v_t).

2.5 The Dirac-Frenkel Variational Principle

Following the presentation in [Lub08, Chapter II.1], we will now introduce the Dirac-Frenkel ansatz for approximating the solution of the Schrödinger equation on a smooth submanifold of the Hilbert space it is defined on.

In Subsection 2.5.1 we will explain how this ansatz has to be modified in the case of time-dependent parametrized submanifolds.

Assume that the wave function ψ_t lies in a Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ and is the solution to the abstract Schrödinger equation

$$\dot{\psi}_t = -iH\psi_t \tag{2.5.1}$$

with linear and self-adjoint Hamiltonian H. Assume further that it is approximated by $u_t \in M$, where M is a smooth submanifold of $\mathcal{H}$ and let $\mathcal{T}_u M$ denote the tangent space of M in the point u. The Galerkin ansatz in this context is to construct the approximation u_t by choosing its time derivative $\dot{u}_t \in \mathcal{T}_{u_t} M$ in such a way that the residual of the Schrödinger equation applied to u_t is orthogonal to the tangent space $T_{u_t} M$:

$$\text{Choose} \quad \dot{u}_t \in \mathcal{T}_{u_t} M \quad \text{such that} \quad \langle \dot{u}_t + iHu_t \ , \ v\rangle = 0 \qquad \text{for all } v \in \mathcal{T}_{u_t} M.$$

In other words, $\dot{u}_t$ is the best approximation of $-iHu_t$ in $\mathcal{T}_{u_t} M$ or the orthogonal projection of $-iHu_t$ onto $\mathcal{T}_{u_t} M$:

$$\dot{u}_t = P_{\mathcal{T}_{u_t} M}\left(-iHu_t\right). \tag{2.5.2}$$

This ansatz, called the *Dirac-Frenkel variational principle* or *variational approximation*, has quite a few remarkable properties listed in [Lub08, Chapter II.1], amongst others conservation of the norm $\|u_t\|$ (under weak additional assumptions) and the total energy $\langle u_t, Hu_t\rangle$ in time and the existence of the following a posteriori error bound for the approximant u_t:

Proposition 2.29. The error of the variational approximation at time t is bounded by

$$\|u_t - \psi_t\| \leq \|u_0 - \psi_0\| + \int_0^t \operatorname{dist}\left(-iHu_s\,, \mathcal{T}_{u_s} M\right) \mathrm{d}s.$$

Proof. See [Lub08, Chapter II.1, Theorem 1.5]. □

2.5.1 Time-dependent, Parametrized Manifolds

As mentioned in the introduction, our approximation manifold will be a complex linear space spanned by Gaussian basis functions, which adapt in time and space to the function ψ_t we want to approximate. This makes our manifold time-dependent: $M_t := \operatorname{span}_{\mathbb{C}} \{\eta_{1,t}, \dots, \eta_{N,t}\}$. More generally, let

$$M_t = \{\chi_t(c) \mid c \in \mathbb{R}^m\} \subseteq \mathcal{H}$$

be a time-dependent smooth manifold parametrized by a map $\chi_t : \mathbb{R}^m \to \mathcal{H}$ which is differentiable in time. Again, ψ_t will be approximated by

$$\psi_t \approx u_t = \chi_t(c_t) \in M_t,$$

but in contrast to the time-independent case, the time derivative of u_t lies in an *affine* linear space, namely the tangent space of M_t in u_t shifted by the time derivative of χ_t in c_t:

$$\dot{u}_t \;\in\; \dot{\chi}_t(c_t) + T_{u_t} M_t \;=:\; \tilde{T}_{u_t} M_t.$$

Projecting the right-hand side of the Schrödinger equation (2.5.1) orthogonally onto the affine space $\tilde{T}_{u_t} M_t$ corresponds to subtracting the shift $\sigma_t = \dot{\chi}_t(c_t)$ from the right-hand side, projecting the result $-iHu_t - \sigma_t$ onto $T_{u_t} M_t$ and then adding the shift again:

$$\dot{u}_t \;\overset{!}{=}\; P_{\tilde{T}_{u_t} M_t}(-iHu_t) \;=\; \sigma_t + P_{T_{u_t} M_t}(-iHu_t - \sigma_t) \tag{2.5.3}$$

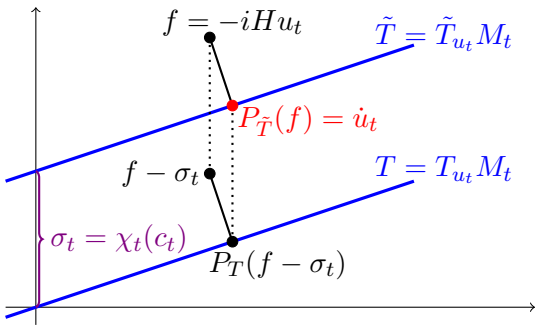


Figure 2.5: Projecting $f := -iHu_t$ onto an affine linear space $\tilde{T} = \sigma_t + T$ by subtracting the shift σ_t from f, projecting the result onto T and adding σ_t again.

Therefore, the time-dependent formulation of the Dirac-Frankel variational principle is:

$$\boxed{\text{Choose} \quad \dot{u}_t \in \tilde{\mathcal{T}}_{u_t} M_t \quad \text{such that} \quad \langle \dot{u}_t + iHu_t \ , \ v \rangle = 0 \qquad \forall v \in \mathcal{T}_{u_t} M_t.}$$

The error analysis looks very close to the unmodified version and the proof goes analogously:

Theorem 2.30. Let $M_t = \{\chi_t(y) \mid y \in \mathbb{R}^m\} \subseteq \mathcal{H}$ be a time-dependent smooth manifold and its parametrization $\chi_t : \mathbb{R}^m \to \mathcal{H}$ be differentiable in time. The error at time t of the variational approximation $u_t = \chi_t(c_t) \in M_t$ is bounded by

$$\|u_t - \psi_t\| \ \leq \ \|u_0 - \psi_0\| + \int_0^t \operatorname{dist}\left(-iHu_s - \dot{\chi}_t(c_t)\,, \mathcal{T}_{u_s} M_s\right) \mathrm{d}s.$$

Proof. Let $e_t := u_t - \psi_t$, $P^\perp := \mathrm{Id} - P_{T_{u_t} M_t}$ and $\sigma_t = \dot{\chi}_t(c_t)$. Subtracting (2.5.1) from (2.5.3) yields:

$$\dot{e}_t = -iHe_t - P^\perp\left(-iHu_t - \sigma_t\right).$$

Since H is self-adjoint, taking the inner product with e_t on both sides and considering its real part leads to:

$$\begin{aligned}
& \Re\,\langle \dot{e}_t, e_t\rangle && = && \underbrace{\Re\,\langle -iHe_t\,, e_t\rangle}_{=0} - \Re\left\langle P^\perp\left(-iHu_t - \sigma_t\right), e_t\right\rangle \\
\implies \quad & \|e_t\| \frac{d}{dt}\|e_t\| && \leq && \left\| P^\perp\left(-iHu_t - \sigma_t\right)\right\| \|e_t\| \\
\implies \quad & \frac{d}{dt}\|e_t\| && \leq && \operatorname{dist}\big(-iHu_t - \sigma_t\,, \mathcal{T}_{u_t} M\big).
\end{aligned}$$

□

The norm of the approximant u_t is again conserved, but this time the total energy is not:

Proposition 2.31. The norm of the variational approximation u_t is conserved, i.e. $\frac{d}{dt}\|u_t\| = 0$, while the total energy is perturbed in the following way:

$$\frac{d}{dt}\langle u_t, Hu_t\rangle = -2\,\Re\langle \sigma_t \mid P^\perp(Hu_t)\rangle\,,$$

where again $\sigma_t := \dot{\chi}_t(c_t)$ and $P^\perp := \mathrm{Id} - P_{T_{u_t} M_t}$.

Proof. Using the abbreviations $f_t = -iHu_t$, $\dot{u}_t = \sigma_t + w_t$ with $w_t \in T_{u_t}M_t$ we get:

$$\begin{aligned}
\frac{d}{dt}\langle u_t \mid Hu_t\rangle &= 2\,\Re\langle \dot{u}_t \mid Hu_t\rangle = 2\,\Re\langle \sigma_t + w_t \mid Hu_t\rangle = -2\,\Im\langle \sigma_t + w_t \mid -iHu_t\rangle \\
&= -2\,\Im\big[\langle \sigma_t \mid \dot{u}_t - \sigma_t + P^\perp(f_t - \sigma_t)\rangle + \langle w_t \mid \dot{u}_t\rangle\big] \\
&= -2\,\Im\big[\langle \underbrace{\sigma_t + w_t}_{\dot{u}_t} \mid \dot{u}_t\rangle - \langle \sigma_t \mid \sigma_t\rangle + \langle \sigma_t \mid P^\perp f_t\rangle - \underbrace{\langle \sigma_t \mid P^\perp \sigma_t\rangle}_{\langle P^\perp\sigma_t \mid P^\perp\sigma_t\rangle}\big] \\
&= -2\,\Im\langle \sigma_t \mid P^\perp f_t\rangle \\
&= 2\,\Re\langle \sigma_t \mid P^\perp(Hu_t)\rangle\,, \\
\frac{d}{dt}\langle u_t \mid u_t\rangle &= 2\,\Re\langle u_t \mid \dot{u}_t\rangle = 2\,\Re\langle u_t \mid -iHu_t\rangle = 2\,\Im\underbrace{\langle u_t \mid Hu_t\rangle}_{\in\mathbb{R}} = 0.
\end{aligned}$$

□

2.5.2 Application to a Time-dependent Vector Space

We will now derive the ordinary differential equations for the coefficients and the approximation error when applying the variational principle to a time-dependent vector space

$$M_t := \operatorname{span}\{\eta_{1,t}, \ldots, \eta_{N,t}\}$$

with basis functions $\eta_{1,t}, \ldots, \eta_{N,t} \in \mathcal{H}$, which are differentiable in time.

Proposition 2.32. Let $u_t = \sum_{j=1}^N c_{j,t}\eta_{j,t} \in M_t = \operatorname{span}\{\eta_{1,t}, \ldots, \eta_{N,t}\}$ be the variational approximation of the wave function ψ_t and $H = -\frac{1}{2}\Delta + V$ be the Schrödinger operator. Then the coefficients $c_{j,t}$ and the error bound E_t from Theorem 2.30 fulfill the ordinary differential equations:

$$A_t\dot{c}_t = B_t c_t\,, \tag{2.5.4}$$

$$\dot{E}_t = \sqrt{(c_t^* G_t - \dot{c}_t^* B_t)\, c_t}\,, \tag{2.5.5}$$

where

$$\begin{aligned}
A_t &= \left(\langle \eta_{j,t} \mid \eta_{k,t}\rangle\right)_{j,k=1,\ldots,N}\,, \\
B_t &= \left(\langle \eta_{j,t} \mid \theta_{k,t}\rangle\right)_{j,k=1,\ldots,N}\,, \\
G_t &= \left(\langle \theta_{j,t} \mid \theta_{k,t}\rangle\right)_{j,k=1,\ldots,N}\,, \\
c_t &= \left(c_{j,t}\right)_{j=1,\ldots,N}\,, \\
\theta_{j,t} &= -\dot{\eta}_{j,t} + \frac{i}{2}\Delta\eta_{j,t} - iV\eta_{j,t} \quad \forall j = 1, \ldots, N.
\end{aligned}$$

Proof. The application of the variational principle leads to (for all $j = 1, \dots, N$):

$$0 \stackrel{!}{=} \langle \eta_{j,t} \mid \dot{u}_t + iHu_t \rangle = \left\langle \eta_{j,t} \;\middle|\; \sum_{k=1}^{N} \dot{c}_{k,t}\eta_{k,t} + c_{k,t}\dot{\eta}_{k,t} - \frac{i}{2}c_{k,t}\Delta\eta_{k,t} + ic_{k,t}V\eta_{k,t} \right\rangle$$

$$\iff \quad \sum_{k=1}^{N} \dot{c}_{k,t} \underbrace{\langle \eta_{j,t} \mid \eta_{k,t} \rangle}_{=:\, a_{jk,t}} = \sum_{k=1}^{N} c_{k,t} \underbrace{\left\langle \eta_{j,t} \;\middle|\; \overbrace{-\dot{\eta}_{k,t} + \frac{i}{2}\Delta\eta_{k,t} - iV\eta_{k,t}}^{=:\,\theta_{k,t}} \right\rangle}_{=:\, b_{jk,t}}$$

$$\iff \quad A_t\dot{c}_t = B_t c_t \,.$$

By Theorem 2.30 the error bound E_t for $\|u_t - \psi_t\|$ is the solution of the initial value problem

$$\dot{E}_t = \operatorname{dist}\left(-iHu_t - \sigma_t \,, \mathcal{T}_{u_t}M_t\right) \qquad , \qquad E_0 = \|u_0 - \psi_0\|,$$

where as usual $\sigma_t := \dot{\chi}_t(c_t) = \sum_{j=1}^{N} c_{j,t}\,\dot{\eta}_{j,t}$ the shift of the tangent space. Since $T_{u_t}M_t = M_t$, we have to find the best approximation $w_t = \sum_{j=1}^{N} d_{j,t}\eta_{j,t}$ of $(-iHu_t - \sigma_t)$ in M_t and compute its distance to M_t. The best approximation step leads to the same formula for d_t as we got for $\dot{c}_t$:

For all $j = 1, \dots, N$ we have

$$0 \stackrel{!}{=} \langle \eta_{j,t} \mid w_t - (-iHu_t - \sigma_t) \rangle = \left\langle \eta_{j,t} \mid \sum_{k=1}^{N} d_{k,t}\eta_{k,t} - \frac{i}{2}c_{k,t}\Delta\eta_{k,t} + ic_{k,t}V\eta_{k,t} + c_{k,t}\dot{\eta}_{k,t} \right\rangle$$

$$\iff \quad \sum_{k=1}^{N} d_{k,t} \langle \eta_{j,t} \mid \eta_{k,t} \rangle = \sum_{k=1}^{N} c_{k,t} \left\langle \eta_{j,t} \;\middle|\; \frac{i}{2}\Delta\eta_{k,t} - iV\eta_{k,t} - \dot{\eta}_{k,t} \right\rangle$$

$$\iff \quad A_t d_t = B_t c_t \,.$$

The computation of the distance is an application of the theorem of Pythagoras:

$$\begin{aligned}
\operatorname{dist}(-iHu_t - \sigma_t \,,\, M_t)^2 &= \| - iHu_t - \sigma_t - w_t\|^2 \\
&= \| - iHu_t - \sigma_t\|^2 - \|w_t\|^2 \\
&= \sum_{j,k=1}^{N} \overline{c}_{j,t}\, c_{k,t} \underbrace{\langle \theta_{j,t} \mid \theta_{k,t} \rangle}_{=:\, g_{jk,t}} - \sum_{j,k=1}^{N} \overline{d}_{j,t}\, d_{k,t} \underbrace{\langle \eta_{j,t} \mid \eta_{k,t} \rangle}_{=:\, a_{jk,t}} \\
&= c_t^* G_t c_t - d_t^* A_t d_t \\
&= (c_t^* G_t - \dot{c}_t^* B_t)\, c_t.
\end{aligned}$$

□

Example 2.33. *Let $\mathcal{H} = L^2(\mathbb{R}, \mathbb{R})$ and $M_t \subseteq \mathcal{H}$ be spanned by the basis functions (see Chapter 3)*

$$\eta_{j,t}(x) = \left(\frac{2\epsilon_{j,t}}{\pi}\right)^{\frac{1}{4}} \exp\left(-\epsilon_{j,t}\,(x - x_{j,t})^2 + ip_{j,t}(x - x_{j,t})\right),$$

where $\epsilon_{j,t} > 0,\ p_{j,t},\ x_{j,t} \in \mathbb{R}$ are properly chosen parameters. We omit the index t and denote $x_{jk} := x_j - x_k$, $p_{jk} := p_j - p_k$. Using Wolfram Mathematica 9.0, we compute:

$$a_{00}^{jk} := \langle \eta_j \mid \eta_k \rangle = \sqrt{\frac{2}{\epsilon_j + \epsilon_k}}\,(\epsilon_j\epsilon_k)^{\frac{1}{4}} \exp\left[-\frac{4\epsilon_j\epsilon_k x_{jk}^2 + p_{jk}^2 - 4i(\epsilon_k p_j + \epsilon_j p_k)x_{jk}}{4(\epsilon_j + \epsilon_k)}\right],$$

$$a_{01}^{jk} := \langle \eta_j \mid (\bullet - x_k)\eta_k \rangle = \frac{a_{00}^{jk}}{2(\epsilon_j + \epsilon_k)}\left[2\epsilon_j x_{jk} - ip_{jk}\right],$$

$$a_{02}^{jk} := \langle \eta_j \mid (\bullet - x_k)^2\eta_k \rangle = \frac{a_{00}^{jk}}{4(\epsilon_j + \epsilon_k)^2}\left[2(\epsilon_j + \epsilon_k) - (p_{jk} + 2i\epsilon_j x_{jk})^2\right],$$

$$a_{11}^{jk} := \langle (\bullet - x_j)\eta_j \mid (\bullet - x_k)\eta_k \rangle = \frac{a_{00}^{jk}}{4(\epsilon_j + \epsilon_k)^2}\left[2(\epsilon_j + \epsilon_k) - (p_{jk} + 2i\epsilon_j x_{jk})(p_{jk} - 2i\epsilon_k x_{jk})\right],$$

$$\begin{aligned} a_{12}^{jk} :=& \langle (\bullet - x_j)\eta_j \mid (\bullet - x_k)^2\eta_k \rangle = \frac{a_{00}^{jk}}{8(\epsilon_j + \epsilon_k)^3}\Big[ip_{jk}^3 - 2p_{jk}^2(2\epsilon_j - \epsilon_k)x_{jk} - \\ & 2ip_{jk}\left[3(\epsilon_j + \epsilon_k) + 2\epsilon_j(\epsilon_j - 2\epsilon_k)x_{jk}^2\right] + 4x_{jk}\left[2\epsilon_j^2 + \epsilon_j\epsilon_k - \epsilon_k^2 - 2\epsilon_j^2\epsilon_k x_{jk}^2\right]\Big], \end{aligned}$$

$$\begin{aligned} a_{22}^{jk} :=& \langle (\bullet - x_j)^2\eta_j \mid (\bullet - x_k)^2\eta_k \rangle \\ =& \frac{a_{00}^{jk}}{16(\epsilon_j + \epsilon_k)^4}\Big[p_{jk}^4 + 4ip_{jk}^3 x_{jk}(\epsilon_j - \epsilon_k)\ - 4p_{jk}^2\left[3(\epsilon_j + \epsilon_k) + x_{jk}^2(\epsilon_j^2 + \epsilon_k^2 - 4\epsilon_j\epsilon_k)\right] - \\ & 8ip_{jk}x_{jk}(\epsilon_j - \epsilon_k)\left[3(\epsilon_j + \epsilon_k) - 2\epsilon_j\epsilon_k x_{jk}^2\right] + \\ & 12(\epsilon_j + \epsilon_k)^2 + 8x_{jk}^2(\epsilon_j^3 + \epsilon_k^3 - 3\epsilon_j^2\epsilon_k - 3\epsilon_j\epsilon_k^2) + 16\epsilon_j^2\epsilon_k^2 x_{jk}^4\Big], \end{aligned}$$

$$a_{10}^{jk} := \overline{a_{01}^{kj}} \qquad , \qquad a_{20}^{jk} := \overline{a_{02}^{kj}} \qquad , \qquad a_{21}^{jk} := \overline{a_{12}^{kj}}\,.$$

This implies the following results for A_t, B_t and G_t:

$$\begin{aligned} \eta_{j,t}(x) &= \left(\frac{2\epsilon_{j,t}}{\pi}\right)^{\frac{1}{4}} \exp\left(-\epsilon_{j,t}\,(x - x_{j,t})^2 + ip_{j,t}(x - x_{j,t})\right), \\ \nabla\eta_j(x) &= \eta_j(x)\left[-2\epsilon_j(x - x_j) + ip_j\right], \\ \Delta\eta_j(x) &= \eta_j(x)\left[(-2\epsilon_j(x - x_j) + ip_j)^2 - 2\epsilon_j\right] \\ &= \eta_j(x)\left[4(\epsilon_j)^2(x - x_j)^2 - 4i\epsilon_j p_j(x - x_j) - (p_j)^2 - 2\epsilon_j\right], \\ \dot{\eta}_j(x) &= \eta_j(x)\left[\left(\frac{1}{4\epsilon_j} - (x - x_j)^2\right)\dot{\epsilon}_j + (2\epsilon_j(x - x_j) - ip_j)\,\dot{x}_j + i(x - x_j)\dot{p}_j\right], \end{aligned}$$

$$\theta_j(x) = -\dot{\eta}_j(x) + \frac{i}{2}\Delta\eta_j(x) - iV\eta_j(x) = \eta_j(x)\Big[(x-x_j)^2 \underbrace{\left(\dot{\epsilon}_j + 2i\left(\epsilon_j\right)^2\right)}_{=:\Gamma_2^j} +$$

$$(x-x_j)\underbrace{\left(-2\epsilon_j\dot{x}_j - i\dot{p}_j + 2\epsilon_j p_j\right)}_{=:\Gamma_1^j} + \underbrace{\left(-\frac{\dot{\epsilon}_j}{4\epsilon_j} + ip_j\dot{x}_j - \frac{i}{2}(p_j)^2 - i\epsilon_j\right)}_{=:\Gamma_0^j} - iV(x)\Big],$$

$$b_{jk} = \langle \eta_j \mid \theta_k\rangle = a_{00}^{jk}\Gamma_0^k + a_{01}^{jk}\Gamma_1^k + a_{02}^{jk}\Gamma_2^k - i\langle\eta_j \mid V\eta_k\rangle,$$

$$\begin{aligned} g_{jk} = \langle \theta_j \mid \theta_k\rangle = {} & a_{00}^{jk}\overline{\Gamma_0^j}\Gamma_0^k + a_{01}^{jk}\overline{\Gamma_0^j}\Gamma_1^k + a_{02}^{jk}\overline{\Gamma_0^j}\Gamma_2^k + a_{10}^{jk}\overline{\Gamma_1^j}\Gamma_0^k + a_{11}^{jk}\overline{\Gamma_1^j}\Gamma_1^k + a_{12}^{jk}\overline{\Gamma_1^j}\Gamma_2^k + \\ & + a_{20}^{jk}\overline{\Gamma_2^j}\Gamma_0^k + a_{21}^{jk}\overline{\Gamma_2^j}\Gamma_1^k + a_{22}^{jk}\overline{\Gamma_2^j}\Gamma_2^k + \langle\eta_j \mid V\eta_k\rangle \\ & - i\left\langle\left(\Gamma_0^j + \Gamma_1^j(x-x_j) + \Gamma_2^j(x-x_j)^2\right)\eta_j \mid V\eta_k\right\rangle \\ & + i\left\langle V\eta_j \mid \left(\Gamma_0^k + \Gamma_1^k(x-x_k) + \Gamma_2^k(x-x_k)^2\right)\eta_k\right\rangle. \end{aligned}$$

Note that all integrals exist since we assumed that V grows at most polynomially (see Section 2.4).

2.6 Phase Space Transformations

In this section, we will discuss three transformations, which will be important for a deeper understanding of quantum mechanics and its Bohmian interpretation (see Section 2.4), adapted convolutions (see Section and 2.7.3), and which prepare us for Appendices A and B: the Fourier transform, the Fourier-Bros-Iagolnitzer (FBI) transform and the Wigner transform.

All three transformations can be viewed from various perspectives and we will try to present two of them. The first is the quantum mechanical point of view, where the phase space is the Cartesian product of position space and momentum space. The second viewpoint is time-frequency analysis, where the phase space is the Cartesian product of time and frequency (in this case the dimension is $d = 1$ and f is usually considered real-valued). In the following, $\mathcal{S}(\mathbb{R}^d, \mathbb{C})$ will denote the Schwartz space of rapidly decreasing functions.

Definition 2.34 (Fourier transform)**.** The *Fourier transform* $\mathcal{F} : \mathcal{S}(\mathbb{R}^d, \mathbb{C}) \to \mathcal{S}(\mathbb{R}^d, \mathbb{C})$ is defined by

$$\mathcal{F}f(\xi) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(y)\, e^{-iy^\intercal\xi}\,\mathrm{d}y\,.$$

The Fourier transform has many important properties, which can be found in e.g. [Fol89], from which we will only use the following:

Proposition 2.35 (Plancherel Theorem and Fourier Inversion Formula)**.**
The Fourier transform is an isometric isomorphism with inverse

$$\mathcal{F}^{-1}g(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} g(\xi)\, e^{-ix^\intercal \xi}\, \mathrm{d}\xi\,.$$

Proof. See [Fol89]. □

The isometry property is usually referred to as Plancherel theorem and implies that the squared modulus of the Fourier transform $\mathcal{F}\psi$ of the wave function ψ is also a probability density function: $\|\mathcal{F}\psi\|_{L^2} = 1$. While $|\psi|^2$ is usually interpreted as the probability density for the (joint) position of the considered particles, $|\mathcal{F}\psi|^2$ is viewed as the probability density for their (joint) momentum.

The Fourier inversion formula implies for $f \in \mathcal{S}(\mathbb{R}^d, \mathbb{C})$

$$\begin{aligned}(2\pi)^{-d} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)\, e^{-ix^\intercal \xi}\, \mathrm{d}\xi\, \mathrm{d}x &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x)\, e^{-ix^\intercal \xi}\, \mathrm{d}x\, \mathrm{d}\xi \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \mathcal{F}f(\xi)\, e^{i0^\intercal \xi}\, \mathrm{d}\xi \\ &= \mathcal{F}^{-1}\mathcal{F}f(0) = f(0).\end{aligned}$$

The technique of using "$(2\pi)^{-d}\int_{\mathbb{R}^d} e^{-ix^\intercal \xi}\, \mathrm{d}\xi$" as a δ-distribution will be used several times in this section, Section 2.7.3 and Appendices A and B.

From the point of view of time-frequency analysis, the Fourier transform yields a decomposition of the signal f into its frequencies: it indicates to which extent which frequency ξ occurs in f. However, one is often interested in the local frequencies of f, meaning, which frequencies of f occur at (or around) a specific time x. This can be analyzed using a windowed Fourier transform, also called Gabor transform, which does not "see" the values of f far from x. Applied to each time x, this yields a mapping defined on the phase space:

Definition 2.36 (windowed Fourier transform, FBI transform)**.** The *windowed Fourier transform* $\mathcal{F}_\sigma : \mathcal{S}(\mathbb{R}^d, \mathbb{C}) \to \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{C})$ is defined by

$$\mathcal{F}_\sigma f(x,\xi) = \pi^{-d/4} \int_{\mathbb{R}^d} f(y)\, g_\sigma(x-y)\, e^{-iy^\intercal \xi}\, \mathrm{d}y\,, \quad \text{where} \tag{2.6.1}$$

$$g_\sigma(x) = (2\pi\sigma^2)^{-d/2} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right), \tag{2.6.2}$$

In the case $\sigma = 1$, we will refer to it as the *Fourier-Bros-Iagolnitzer (FBI) transform* and denote it by $\mathcal{T} := \mathcal{F}_1$.

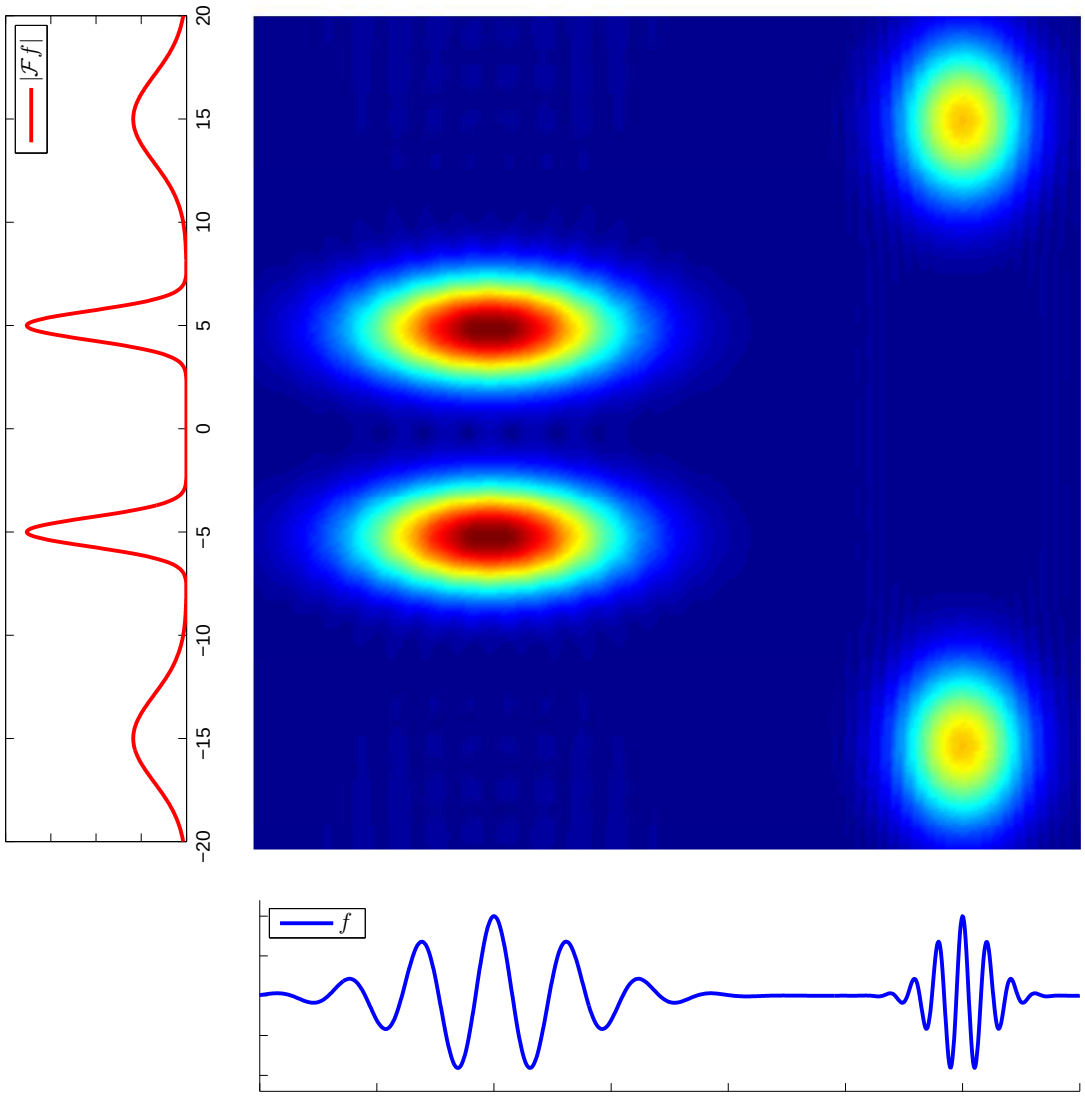


Figure 2.6: Visualisation of how different frequencies of a function f are represented by the Fourier transform $\mathcal{F}f$ and the FBI transform $\mathcal{F}_\sigma f$ (only the moduli of $\mathcal{F}f$ and $\mathcal{F}_\sigma f$ are plotted).

The width σ of the Gaussian function g_σ is a double-edged sword: The smaller it is chosen, the more accurate the time-frequency description becomes in x-direction, since only values very close to the considered time x are considered for $\mathcal{F}_\sigma f(x, \bullet)$. However,

the smaller the window, the more "difficult" it becomes to estimate the frequencies in this small region and the more "blurred" the frequency decomposition $\mathcal{F}_\sigma f(x, \bullet)$ becomes in ξ-direction (in particular, taking the limit $\sigma \to 0$ does not provide any useful information).

In quantum mechanics, this statement corresponds to the fact that there can be no joint probability distribution for position and momentum in phase space, which is a consequence of Heisenberg's uncertainty principle. Only a "blurred" version of such a probability density function can exist, as we will see in Proposition 2.39.

However, there is a replacement for such a probability density, namely the Wigner transform of f:

Definition 2.37 (Wigner transform)**.** The *Wigner transform* $W : \mathcal{S}(\mathbb{R}^d, \mathbb{C}) \to \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})$ is defined by

$$Wf(x, \xi) = (2\pi)^{-d} \int_{\mathbb{R}^d} \overline{f\left(x + \frac{y}{2}\right)} f\left(x - \frac{y}{2}\right) e^{iy^\intercal \xi} \, \mathrm{d}y.$$

Remark 2.38. Usually, the Wigner transform $W : \mathcal{S}(\mathbb{R}^d, \mathbb{C}) \times S(\mathbb{R}^d, \mathbb{C}) \to \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{C})$ is defined by

$$W(f, g)(x, \xi) = (2\pi)^{-d} \int_{\mathbb{R}^d} \overline{f\left(x + \frac{y}{2}\right)} g\left(x - \frac{y}{2}\right) e^{iy^\intercal \xi} \, \mathrm{d}y.$$

We will only use its definition on the diagonal, $Wf = W(f, f)$, where it is real-valued, which can be seen by taking the transformation $y \mapsto -y$ in the integral.

Just as expected from a joint probability density in phase space, the Wigner transform Wf of a function $f \in \mathcal{S}(\mathbb{R}^d, \mathbb{C})$ has the marginal densities $|f|^2$ and $|\mathcal{F}f|^2$:

$$\begin{aligned} |f(x)|^2 &= \int_{\mathbb{R}^d} Wf(x, \xi) \, \mathrm{d}\xi \,, \\ |\mathcal{F}f(\xi)|^2 &= \int_{\mathbb{R}^d} Wf(x, \xi) \, \mathrm{d}x \,. \end{aligned}$$

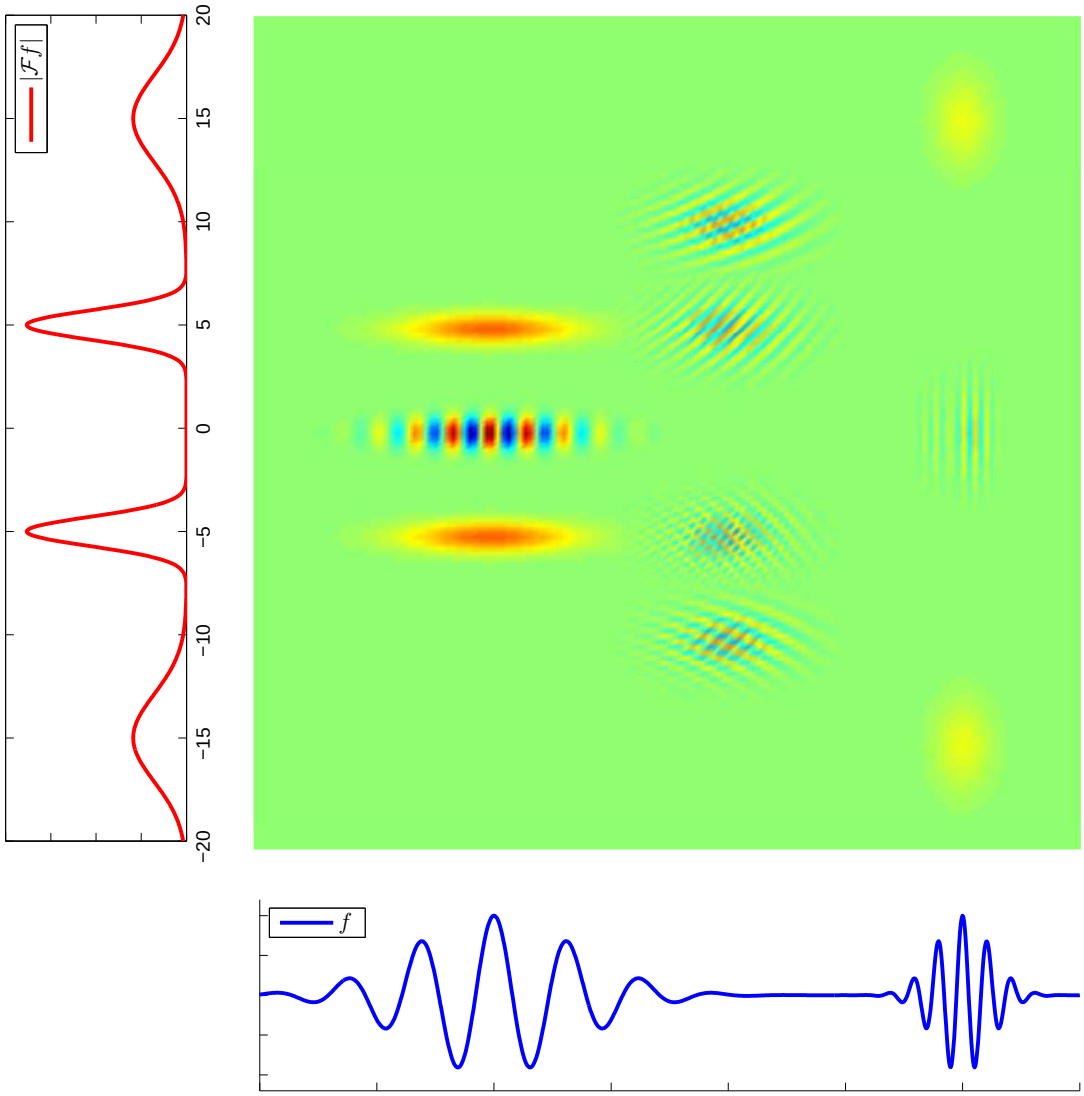


Figure 2.7: Visualisation of how different frequencies of a function f are represented by the Fourier transform $\mathcal{F}f$ and the Wigner transform Wf.

However, it may attain negative values and is therefore not a probability density function in the first place, therefore it is often referred to as *Wigner quasi-probability distribution*. Still, there are a few results, which can be deduced by treating it as though it was a probability density function, some of which are presented in Section 2.7 and Appendix A. Another important trick is to force the Wigner transform to become non-negative by smoothing it via convolution with a properly scaled Gaussian in phase

space. This results in a blurred joint probability distribution in phase space, which is strongly connected to the FBI-transform:

Proposition 2.39. Let

$$G_\sigma(x,\xi) = \left(2\pi\sigma^2\right)^{-d/2} \exp\left(-\frac{\|x\|^2 + \|\xi\|^2}{2\sigma^2}\right).$$

denote a Gaussian in phase space. Then we have

$$Wf * G_{1/\sqrt{2}} = |\mathcal{T}f|^2.$$

Proof. See [Hil97, (3.10)]. □

Remark 2.40. The resulting transform $Hf := Wf * G_{1/\sqrt{2}} = |\mathcal{T}f|^2$ is called the *Husimi transform* of f and plays an important role in microlocal analysis. It is the closest one can get to a joint non-negative probability distribution for position and momentum in phase space, since for $\sigma < 1/\sqrt{2}$ the non-negativity of $Wf * G_{1/\sqrt{2}}$ can no longer be guaranteed. More precisely, one can show that for each $\sigma < 1/\sqrt{2}$ there is a function $f \in \mathcal{S}(\mathbb{R}^d, \mathbb{C})$, such that $Wf * G_\sigma$ attains negative values.

2.7 Convolutions and Adapted Convolutions

2.7.1 Motivation

The convolution of two integrable functions $f, g\colon \mathbb{R}^d \to \mathbb{R}$

$$(f * g)(x) = \int_{\mathbb{R}^d} f(y)\, g(x-y)\, \mathrm{d}y$$

is a basic mathematical tool with applications in probability theory, image processing, optics, acoustics and many others. If g is a probability density, say a Gaussian density

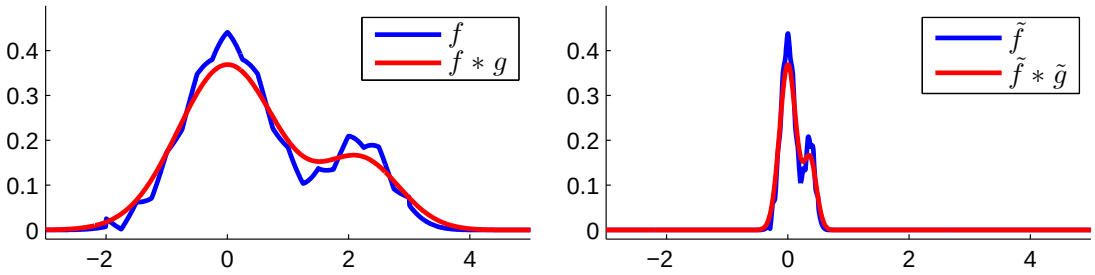
$$g(x) = (2\pi\sigma^2)^{-d/2} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right), \tag{2.7.1}$$

the convolution $(f * g)(x)$ at point x can be viewed as the mean over all $f(y),\ y \in \mathbb{R}^d$, weighted by the density $g(x-y)$. Convolutions are therefore often used to "flatten" or "smooth" a function f by some probability density g and g is called a *smoothing kernel* in this case.

A natural question arising here is how to choose the standard deviation σ of g, i.e. how strong we want to smooth the function f. Normally the aim is to flatten out the bumps and edges without losing the shape of the function completely.

Assume we have found a proper σ to smooth f, and $\tilde{f}(x) = f(\alpha x)$ is a scaled version of f by some factor $\alpha > 0$. In this case we also have to scale the density g by the same factor, $\tilde{g} = \alpha^d g(\alpha x)$ (the prefactor α^d is just a normalization factor), in order to get an analogous result (see Figure 2.8):

$$(\tilde{f} * \tilde{g})(x) = \alpha^d \int_{\mathbb{R}^d} f(\alpha y)\, g(\alpha(x-y))\, \mathrm{d}y = \int_{\mathbb{R}^d} f(y)\, g(\alpha x - y))\, \mathrm{d}y = (f * g)(\alpha x)$$



(a) The standard deviation of the Gaussian g from 2.7.1 chosen appropriately to smooth the function f, here $\sigma = 0.4$.

(b) In order to choose $\tilde{f}(x) = f(\alpha x)$ the function g has to be scaled in the same way to get an analogous result, here $\alpha = 6$.

Figure 2.8: Choosing proper standard deviations of the density g to smooth differently scaled versions of the function f.

One difficulty occurs if the function we want to smooth consists of two well-separated parts, one with low and the other with high variation, e.g. if we build up a function h from f and a scaled version of $\tilde{f}(x) = f(\alpha x)$ ($\alpha > 0$), separating them in space by a shift $a > 0$:

$$h(x) = f(x) + \tilde{f}(x - a).$$

Choosing g as a smoothing kernel will be unappropriate for the right part of the function (not enough smoothing), choosing $\tilde{g}$ for the left part (too much smoothing), see Figure 2.9 (a) and (b).

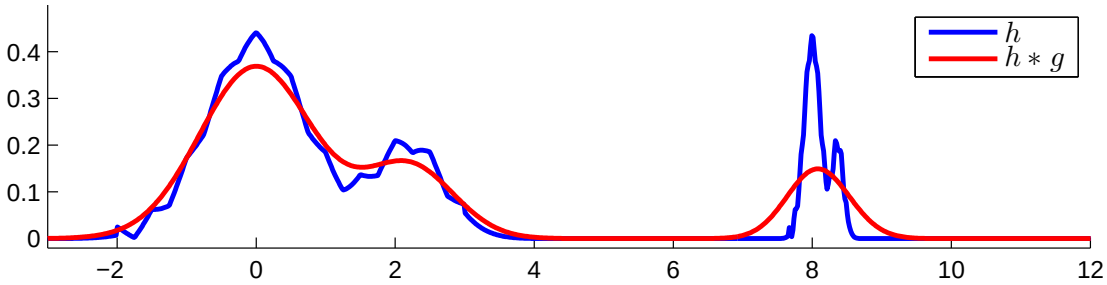
One possible way of finding the proper scaling of the smoothing kernel for both parts is to adapt it locally by replacing $(f * g)(x)$ by

$$(f *_\mu g)(x) := \int f(y)\ \mu(y)^d\, g\big(\mu(y)(x-y)\big)\, \mathrm{d}y\,,$$

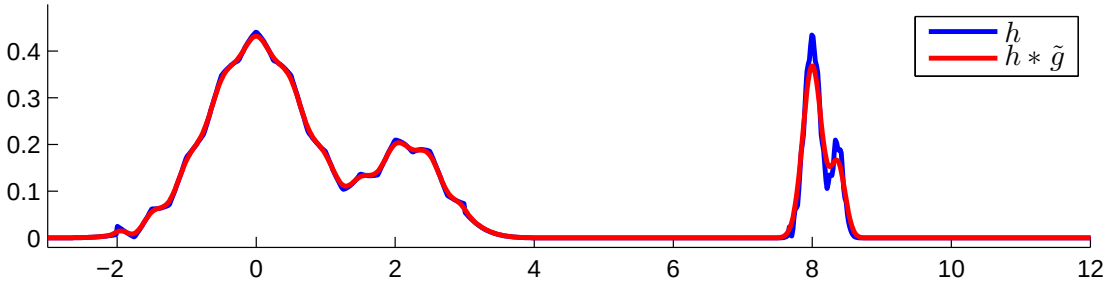
where $\mu : \mathbb{R}^d \to \mathbb{R}$ is a measurable function which scales the density g locally by different factors $\mu(y)$. In our example with shift $a = 8$ and scaling factor $\alpha = 6$, the choice

$$\mu(y) = \begin{cases} 1 & \text{if } x < 4, \\ 6 & \text{if } x \geq 7, \end{cases} \tag{2.7.2}$$

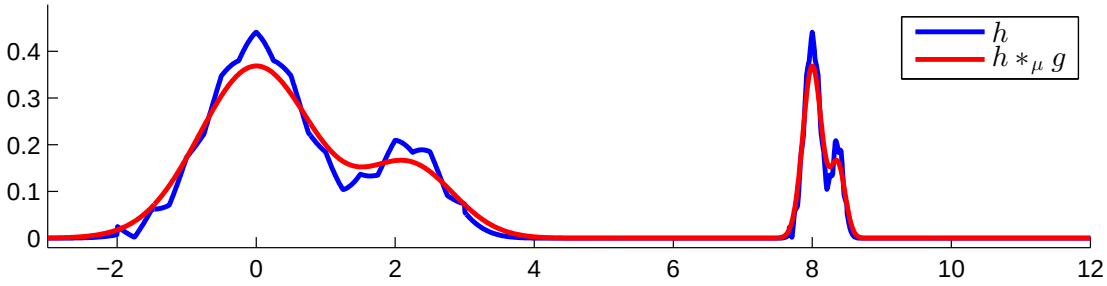
seems suitable (see Figure 2.9a).



(a) g is an appropriate smoothing kernel for the "left part", but not for the right one.



(b) $\tilde{g}$ is an appropriate smoothing kernel for the "right part", but not for the left one.



(c) Adapted convolutions guarantee an appropriate "width" of the smoothing kernel everywhere. μ is chosen as in equation (2.7.2).

Figure 2.9: Adapted convolutions guarantee an appropriate "width" of the smoothing kernel everywhere. Here, the shift is $a = 8$ and the scaling factor is $\alpha = 6$.

Remark 2.41. There are two possible points of view to describe the smoothing process of f with g in the convolution $f * g$:

- To compute $(f*g)(x)$ we "sum up" all values $f(y),\ y \in \mathbb{R}^d$, weighted with $g(x-y)$.
- Each value $f(y)$ contributes to the value $(f * g)(x)$ for each x, weighted with $g(x-y)$.

These viewpoints lead to two possible definitions for the μ-adapted convolution:

- $(f *_\mu g)(x) := \int_{\mathbb{R}^d} f(y)\,\mu(x)^d\, g\big(\mu(x)(x-y)\big)\mathrm{d}y,$
- $(f *_\mu g)(x) := \int_{\mathbb{R}^d} f(y)\,\mu(y)^d\, g\big(\mu(y)(x-y)\big)\mathrm{d}y.$

While the first definition does not lead to a reasonable theory (e.g. the adapted convolution would not be L^1-norm preserving), the second definition does. The theory is presented in the following section in a slightly more general setup.

2.7.2 Theory

Definition 2.42 (adapted convolutions, adaptation function)**.** Let $f \in L^1(\mathbb{R}^d),\ g \in L^p(\mathbb{R}^d),\ 1 \le p \le \infty$, $\mu\colon \mathbb{R}^d \to \mathbb{R}_{>0}$ be a measurable function and

$$g_\mu(x,y) := \mu(y)^{d/p}\, g\big(\mu(y)(x-y)\big).$$

We define the *μ-adapted convolution of f with g* by

$$(f *^p_\mu g)(x) := \int_{\mathbb{R}^d} f(y)\, g_\mu(x,y)\,\mathrm{d}y\,,$$

where $1/p := 0$ for $p = \infty$. μ will be called *adaptation function*. In the case $p = 1$, we will omit the upper index and just write $f *_\mu g$. We will allow the function μ to attain the values zero and infinity, if $\mathrm{supp}(f) \subseteq \mathrm{supp}(\mu)$, i.e. if $\mu(y) = 0$ implies $f(y) = 0$, since this does not affect the integral.

Remark 2.43.
1. This type of convolution is not symmetric and the notation $f *_\mu g$ indicates that g is scaled by $\mu(y)$, while $f\,{}_\mu{*}\, g$ can be used, if f is to be scaled (we will not need the second notation).
2. The μ-adapted convolution reduces to the common convolution $f * g$ for $\mu \equiv 1$.

Proposition 2.44 (Young's inequality)**.** Let $f \in L^1(\mathbb{R}^d),\ g \in L^p(\mathbb{R}^d),\ 1 \le p \le \infty$ and $\mu\colon \mathbb{R}^d \to \mathbb{R}_{>0}$ be a measurable function. Then $f *^p_\mu g \in L^p(\mathbb{R}^d)$ and

$$\|f *^p_\mu g\|_p \le \|f\|_1\, \|g\|_p\,.$$

Proof. First note that for $p < \infty$ and $y \in \mathbb{R}^d$ the transformation formula implies:

$$\|g_\mu(\bullet, y)\|_p^p = \int_{\mathbb{R}^d} \mu(y)^d \left|g\big(\mu(y)(x-y)\big)\right|^p \,\mathrm{d}x = \int_{\mathbb{R}^d} |g(x)|^p \,\mathrm{d}x = \|g\|_p^p . \tag{2.7.3}$$

The cases $p = 1$ and $p = \infty$ are straightforward:

$$\begin{aligned} \|f *_\mu g\|_1 &\le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)|\,|g_\mu(x,y)|\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} |f(y)| \int_{\mathbb{R}^d} |g_\mu(x,y)|\,\mathrm{d}x\,\mathrm{d}y \\ &= \|f\|_1 \,\|g\|_1 , \end{aligned}$$

$$\begin{aligned} \|f *_\mu^\infty g\|_\infty &\le \operatorname*{ess\,sup}_{x\in\mathbb{R}^d} \int_{\mathbb{R}^d} |f(y)| \left|g\left(\mu(y)^{\frac{1}{d}}(x-y)\right)\right| \mathrm{d}y \\ &\le \int_{\mathbb{R}^d} |f(y)|\,\mathrm{d}y \operatorname*{ess\,sup}_{x\in\mathbb{R}^d} |g(x)| \\ &\le \|f\|_1 \,\|g\|_\infty , \end{aligned}$$

$$\text{where} \quad \operatorname*{ess\,sup}_{x\in\mathbb{R}^d} |f(x)| := \inf_{\substack{N\in\mathcal{B}(\mathbb{R}^d)\\ \mu(N)=0}} \sup_{x\in\mathbb{R}^d\setminus N} |f(x)|$$

denotes the essential supremum. Now let $1 < p < \infty$ and p' be its conjugate exponent, i.e. $\frac{1}{p} + \frac{1}{p'} = 1$. Hölder's inequality yields

$$|f *_\mu^p g|(x) \le \int_{\mathbb{R}^d} |f(y)|^{\frac{1}{p'}} |f(y)|^{\frac{1}{p}} |g_\mu(x,y)|\,\mathrm{d}y \le \left\| |f|^{\frac{1}{p'}} \right\|_{p'} \left\| |f|^{\frac{1}{p}} |g_\mu(x,\bullet)| \right\|_p .$$

Together with equation (2.7.3) and the Fubini-Tonelli theorem this implies

$$\begin{aligned} \|f *_\mu^p g\|_p^p &= \int_{\mathbb{R}^d} |f *_\mu^p g|^p(x)\,\mathrm{d}x \\ &\le \left\| |f|^{\frac{1}{p'}} \right\|_{p'}^p \int_{\mathbb{R}^d} \left\| |f|^{\frac{1}{p}} |g_\mu(x,\bullet)| \right\|_p^p \mathrm{d}x \\ &= \left\|f\right\|_1^{p/p'} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)|\,|g_\mu(x,y)|^p \,\mathrm{d}y\,\mathrm{d}x \\ &= \left\|f\right\|_1^{p/p'} \int_{\mathbb{R}^d} |f(y)|\,\|g_\mu(\bullet,y)\|_p^p \,\mathrm{d}y \\ &= \left\|f\right\|_1^{1+p/p'} \|g\|_p^p \\ &\le \|f\|_1^p \,\|g\|_p^p . \end{aligned}$$

□

Remark 2.45. We will apply this proposition mostly for $p = 1$, in which case we also have:

$$\begin{aligned}\int_{\mathbb{R}^d} (f *_\mu g)(x)\,\mathrm{d}x &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(y)\, g_\mu(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} g_\mu(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \left(\int_{\mathbb{R}^d} f(y)\,\mathrm{d}y\right)\left(\int_{\mathbb{R}^d} g(x)\,\mathrm{d}x\right)\end{aligned}$$

and, if g is a probability density on $\mathbb{R}^d$,

$$\begin{aligned}\|f *_\mu g\|_1 &= \int_{\mathbb{R}^d} |f *_\mu g|(x)\,\mathrm{d}x &&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)|\, g_\mu(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} |f(y)| \int_{\mathbb{R}^d} g_\mu(x,y)\,\mathrm{d}x\,\mathrm{d}y &&= \left(\int_{\mathbb{R}^d} |f(y)|\,\mathrm{d}y\right)\left(\int_{\mathbb{R}^d} g(x)\,\mathrm{d}x\right) \\ &= \|f\|_1\,.\end{aligned}$$

Remark 2.46. The generalization of Young's inequality,

$$\|f * g\|_r \le \|f\|_p \, \|g\|_p$$

for $1 \le p, q \le \infty$, $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$ and $f \in L^p(\mathbb{R}^d)$, $g \in L^q(\mathbb{R}^d)$, does not hold for general adapted convolutions. Also, the adapted convolution is not associative. However, distributivity and associativity with scalar multiplication can be generalized to adapted convolutions (the proofs are straightforward).

Also, similar to common convolutions there are slightly modified (but non-symmetric!) rules for the differentiation of adapted convolutions. We will use the standard multiindex notation,

$$|\alpha| := \alpha_1 + \cdots + \alpha_d\,, \tag{2.7.4}$$

$$\partial^\alpha := \partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \cdots \partial_{x_d}^{\alpha_d}\,, \tag{2.7.5}$$

$$x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}, \qquad x \in \mathbb{R}^d\,, \tag{2.7.6}$$

where $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$.

Proposition 2.47. Let $f \in L^1(\mathbb{R}^d)$, $g \in C^\alpha(\mathbb{R}^d)$ for some $\alpha \in \mathbb{N}^d$, $\mu\colon \mathbb{R}^d \to \mathbb{R}_{>0}$ be a measurable function and $1 \le p \le \infty$, such that

$$\partial^\beta g \in L^p(\mathbb{R}^d) \qquad \text{and} \qquad f \cdot \mu^{|\beta|} \in L^1(\mathbb{R}^d) \qquad \text{for all} \qquad \beta \in \mathbb{N}^d,\ \beta \le \alpha.$$

Then $f *_\mu^p g \in C^\alpha(\mathbb{R}^d)$ and for all $\beta \in \mathbb{N}^d,\ \beta \leq \alpha$ the derivative $\partial^\beta (f *_\mu^p g) \in L^p(\mathbb{R}^d)$ is given by

$$\partial^\beta \left(f *_\mu^p g\right) = \left(f \cdot \mu^{|\beta|}\right) *_\mu^p \partial^\beta g\,.$$

Proof. For all $j = 1, \ldots, d$ we have

$$\begin{aligned}
\partial_{x_j} \left(f *_\mu^p g\right)(x) &= \partial_{x_j} \left(\int_{\mathbb{R}^d} f(y)\, \mu(y)^{\frac{d}{p}}\, g\big(\mu(y)(x-y)\big)\, \mathrm{d}y \right) \\
&= \int_{\mathbb{R}^d} f(y)\, \mu(y)^{\frac{d}{p}+1}\, (\partial_{x_j} g)\big(\mu(y)(x-y)\big)\, \mathrm{d}y \\
&= \left[(f \cdot \mu) *_\mu^p \partial_{x_j} g\right](x)
\end{aligned}$$

The claim follows by induction and $\partial^\beta (f *_\mu^p g) \in L^p(\mathbb{R}^d)$ follows from Proposition 2.44. □

2.7.3 Choosing a proper Adaptation Function μ

In the example from Section 2.7.1 the adaptation function μ was chosen *manually* to smooth the function $f : \mathbb{R}^d \to \mathbb{R}$ in a reasonable way, when taking the μ-adapted convolution with the density g. Let us now discuss how this choice can be performed *automatically* in dependence of the function f we want to smooth. We are mostly interested in the case where g is a probability density function used as a smoothing kernel, therefore we will restrict ourselves to the case $p = 1$.

Finding a good dependence for the adaptation function $\mu = \mu_f$ on the function f is a difficult task and will be answered here only partially. We will make the choice

$$\mu_f(x) = \sqrt{\left| \frac{\|\nabla f(x)\|^2 - f(x)\, \Delta f(x)}{2|f(x)|^2} \right|}$$

plausible, but neither prove uniqueness nor any kind of optimality.

Let us first gather the criteria which we would like our adaptation function μ_f to fulfill (see also the motivation section 2.7.1):

Condition 2.48 (Adaptation Conditions)**.** *From now on we will say that the mapping $f \mapsto \mu_f$ fulfills the* adaptation conditions*, if it has the following properties:*

(i) Invariance under shifting by $a \in \mathbb{R}^d$: $\mu_{f(\bullet - a)}(x) = \mu_f(x - a)$

(ii) Invariance under multiplication by a factor $\alpha \neq 0$: $\mu_{\alpha \cdot f} = \mu_f$

(iii) Proper scaling when f is scaled by some $\alpha > 0$: $\mu_{f(\alpha \cdot \bullet)}(x) = \alpha \cdot \mu_f(\alpha x)$

(iv) $\mu_f(x)$ should describe some kind of "variation" of f locally around $x \in \mathbb{R}^d$.

Obviously, the fourth property is not a rigorous condition, but just a rule of thumb, and we will discuss it now. We will start with a *global* version. A natural way to describe the variation of a function $f \in L^1(\mathbb{R}^d, \mathbb{R})$ globally is to consider its Fourier transform since functions with high oscillations tend to have high values of $\mathcal{F}f$ away from the origin and, if a function $f \in L^1(\mathbb{R}^d)$ is scaled by some factor $\alpha > 0$, $\tilde{f}(x) = f(\alpha x)$, the frequencies "appearing" in the Fourier transform are also scaled by α:

$$\mathcal{F}\tilde{f}(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(\alpha y)\, e^{-iy^\intercal \xi}\, \mathrm{d}y = \frac{(2\pi)^{-d/2}}{\alpha^d} \int_{\mathbb{R}^d} f(y)\, e^{-iy^\intercal \xi/\alpha}\, \mathrm{d}y = \frac{\mathcal{F}f(\xi/\alpha)}{\alpha^d}\,. \tag{2.7.7}$$

In order to assign a value for the variation to a function f we will therefore consider the expectation value and variance defined in the following proposition:

Proposition 2.49. Let $f \in W^{2,2}(\mathbb{R}^d, \mathbb{R}) \setminus \{0\}$. The expectation value and variance of the probability distribution $\mathbb{P}_\rho$ given by the density

$$\rho = \frac{|\mathcal{F}f|^2}{\|\mathcal{F}f\|_{L^2}^2} = \frac{|\mathcal{F}f|^2}{\|f\|_{L^2}^2}$$

(here we used the Plancherel theorem 2.35) are:

$$\mathbb{E}_\rho = 0 \qquad \text{and} \qquad \mathbb{V}_\rho = \frac{\int_{\mathbb{R}^d} \|\nabla f(z)\|^2 - f(z)\, \Delta f(z)\, \mathrm{d}z}{2\|f\|_{L^2}^2}\,.$$

Proof. Since for real-valued functions f

$$\mathcal{F}f(-\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(y)\, e^{iy^\intercal \xi}\, \mathrm{d}y = (2\pi)^{-d/2} \overline{\int_{\mathbb{R}^d} f(y)\, e^{-iy^\intercal \xi}\, \mathrm{d}y} = \overline{\mathcal{F}f(\xi)},$$

the expectation value of $\mathbb{P}_\rho$ vanishes. Using the transformation

$$z_1 = y_1 - y_2, \qquad z_2 = y_1 + y_2$$

we compute:

$$\begin{aligned}
\int_{\mathbb{R}^d} |\mathcal{F}f(\xi)|^2 \, \|\xi\|^2 \, \mathrm{d}\xi &= (2\pi)^{-d} \int_{\mathbb{R}^{3d}} f(y_1)\, f(y_2)\, e^{-i(y_1-y_2)^\intercal \xi} \, \|\xi\|^2 \, \mathrm{d}y_1 \, \mathrm{d}y_2 \, \mathrm{d}\xi \\
&= \frac{(2\pi)^{-d}}{2^d} \int_{\mathbb{R}^{3d}} \underbrace{f\left(\frac{z_2+z_1}{2}\right) f\left(\frac{z_2-z_1}{2}\right)}_{=: F(z_1,z_2)} e^{-iz_1^\intercal \xi} \, \xi^\intercal \xi \, \mathrm{d}z_1 \, \mathrm{d}z_2 \, \mathrm{d}\xi \\
&= -\frac{i(2\pi)^{-d}}{2^d} \int_{\mathbb{R}^{3d}} D_{z_1} F(z_1,z_2) \cdot \xi \; e^{-iz_1^\intercal \xi} \, \mathrm{d}z_1 \, \mathrm{d}z_2 \, \mathrm{d}\xi \\
&= -\frac{(2\pi)^{-d}}{2^d} \int_{\mathbb{R}^{3d}} \operatorname{tr}\left[D^2_{z_1} F(z_1,z_2)\right] \; e^{-iz_1^\intercal \xi} \, \mathrm{d}z_1 \, \mathrm{d}\xi \, \mathrm{d}z_2 \\
&= -\frac{1}{2^d} \int_{\mathbb{R}^d} \operatorname{tr}\left[D^2_{z_1} F(0,z_2)\right] \, \mathrm{d}z_2 \\
&= \frac{1}{2} \int_{\mathbb{R}^d} \|\nabla f(z)\|^2 - f(z)\, \Delta f(z) \, \mathrm{d}z
\end{aligned}$$

This proves the formula for the variance. □

The adaptation function μ_f, which in this global setting is just an adaptation *value* $\mu_f \in \mathbb{R}$, can be now assigned the standard deviation of $\mathbb{P}_\rho$,

$$\mu_f := \sigma_\rho = \sqrt{\mathbb{V}_\rho},$$

and the adaptation conditions 2.48 can easily be varified (in a global, x-independent sense).

However, we are not interested in a *global*, but in a *local* adaptation. Therefore we will study the "local frequency" of f by taking its windowed Fourier transform $\mathcal{F}_\sigma f(x,\xi)$ instead of its Fourier transform.

Again, let us consider the expectation value and variance of the corresponding probability density in ξ:

Proposition 2.50. Let $f \in W^{2,2}(\mathbb{R}^d, \mathbb{R}) \setminus \{0\}$. Then for each $x \in \mathbb{R}^d$ the expectation value and variance of the probability distribution $\mathbb{P}_{\rho_x}$ given by the density

$$\rho_x = \frac{|\mathcal{F}_\sigma f(x,\bullet)|^2}{\|\mathcal{F}_\sigma f(x,\bullet)\|^2_{L^2}}$$

are:

$$\mathbb{E}_{\rho_x} = 0 \qquad \text{and} \qquad \mathbb{V}_{\rho_x} = \frac{1}{2\sigma^2} + \frac{\Big(\left(\|\nabla f\|^2 - f\, \Delta f\right) * g^2_\sigma \Big)(x)}{2\left(f^2 * g^2_\sigma\right)(x)} \,.$$

Proof. Since for real-valued functions f

$$\begin{aligned}\mathcal{F}_\sigma f(x,-\xi) &= (\pi)^{-d/4}\int_{\mathbb{R}^d} f(y)\, g_\sigma(x-y)\, e^{iy^\intercal\xi}\,\mathrm{d}y \\ &= (\pi)^{-d/4}\overline{\int_{\mathbb{R}^d} f(y)\, g_\sigma(x-y)\, e^{-iy^\intercal\xi}\,\mathrm{d}y} = \overline{\mathcal{F}_\sigma f(x,\xi)},\end{aligned}$$

the expectation value of $\mathbb{P}_{\rho_x}$ vanishes. Using the transformation

$$z_1 = y_1 - y_2, \qquad z_2 = y_1 + y_2$$

and denoting

$$F_x(z_1,z_2,x) := f\left(\frac{z_2+z_1}{2}\right) f\left(\frac{z_2-z_1}{2}\right) g_\sigma\left(x-\frac{z_2+z_1}{2}\right) g_\sigma\left(x-\frac{z_2-z_1}{2}\right),$$

we compute

$$\begin{aligned}\int_{\mathbb{R}^d} |\mathcal{F}_\sigma f(x,\xi)|^2\,\mathrm{d}\xi &= \pi^{-d/2}\int_{\mathbb{R}^{3d}} f(y_1)\, f(y_2)\, g_\sigma(x-y_1)\, g_\sigma(x-y_2)\, e^{-i(y_1-y_2)^\intercal\xi}\,\mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}\xi \\ &= \frac{\pi^{-d/2}}{2^d}\int_{\mathbb{R}^{3d}} F(z_1,z_2,x)\, e^{-iz_1^\intercal\xi}\,\mathrm{d}z_1\,\mathrm{d}z_2\,\mathrm{d}\xi \\ &= \pi^{d/2}\int_{\mathbb{R}^d} F(0,z_2,x)\,\mathrm{d}z_2 \\ &= 2^d\pi^{d/2}\int_{\mathbb{R}^d} f(z)^2\, g_\sigma(x-z)^2\,\mathrm{d}z \\ &= 2^d\pi^{d/2}\left(f^2 * g_\sigma^2\right)(x)\end{aligned}$$

and, using Wolfram Mathematica 9.0,

$$\begin{aligned}&\int_{\mathbb{R}^d} |\mathcal{F}_\sigma f(x,\xi)|^2\,\|\xi\|^2\,\mathrm{d}\xi \\ &= \pi^{-d/2}\int_{\mathbb{R}^{3d}} f(y_1)\, f(y_2)\, g_\sigma(x-y_1)\, g_\sigma(x-y_2)\, e^{-i(y_1-y_2)^\intercal\xi}\,\|\xi\|^2\,\mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}\xi \\ &= \frac{\pi^{-d/2}}{2^d}\int_{\mathbb{R}^{3d}} F(z_1,z_2,x)\, e^{-iz_1^\intercal\xi}\,\xi^\intercal\xi\,\mathrm{d}z_1\,\mathrm{d}z_2\,\mathrm{d}\xi \\ &= -\frac{i\pi^{-d/2}}{2^d}\int_{\mathbb{R}^{3d}} D_{z_1}F(z_1,z_2,x)\cdot\xi\; e^{-iz_1^\intercal\xi}\,\mathrm{d}z_1\,\mathrm{d}z_2\,\mathrm{d}\xi \\ &= -\frac{\pi^{-d/2}}{2^d}\int_{\mathbb{R}^{3d}} \mathrm{tr}\left[D_{z_1}^2F(z_1,z_2,x)\right]\, e^{-iz_1^\intercal\xi}\,\mathrm{d}z_1\,\mathrm{d}\xi\,\mathrm{d}z_2 \\ &= -\pi^{d/2}\int_{\mathbb{R}^d} \mathrm{tr}\left[D_{z_1}^2F(0,z_2,x)\right]\,\mathrm{d}z_2 \\ &= 2^{d-1}\pi^{d/2}\int_{\mathbb{R}^d}\left(\sigma^{-2}f(z)^2 + \|\nabla f(z)\|^2 - f(z)\,\Delta f(z)\right) g_\sigma(x-z)^2\,\mathrm{d}z \\ &= 2^{d-1}\pi^{d/2}\Big[\left(d\,\sigma^{-2}f^2 + \|\nabla f\|^2 - f\,\Delta f\right) * g_\sigma^2\Big](x)\end{aligned}$$

Taking the quotient proves the formula for the variance. □

Again we can set

$$\mu_f(x) := \sigma_{\rho_x} = \sqrt{\mathbb{V}_{\rho_x}} = \sqrt{\frac{d}{2\sigma^2} + \frac{\Big((\|\nabla f\|^2 - f\,\Delta f) * g_\sigma^2 \Big)(x)}{2\,(f^2 * g_\sigma^2)\,(x)}}\,.$$

However, while the adaptation conditions 2.48 (i) and (ii) are fulfilled, the scale invariance (iii) is violated! The reason for this is that the window of the FBI transform $\mathcal{F}_\sigma f$ does not scale with f, producing the obstructive term $\frac{1}{2\sigma^2}$. Therefore a formula analogous to (2.7.7) does not hold for FBI-transforms. One might try to adapt the width σ of the window beforehand, but this would require a priori knowledge of the local variation μ_f of f, which we are trying to find by taking the FBI transform in the first place. Another disadvantage of this choice of μ_f is the difficulty to calculate the convolutions in the numerator and denominator in practice.

Since the obstructive term is caused by the width of the window, or in other words, by the blurry way we look at the function, we will "unblurr" it by replacing the term $|\mathcal{F}_\sigma f(x,\xi)|^2$ in the probability density ρ_x from Proposition 2.50 with the Wigner transform $Wf(x,\xi)$. This replacement is motivated by the discussion in Section 2.6 and by Proposition 2.39.

We would like to remind the reader, that the Wigner transform can take negative values and is therefore not a probability density function. However, since we are going to use the Wigner transform only to get a good guess on the choice of the adaptation function μ_f, we are going to ignore this detail and just compute the expectation value and variance as we did before:

Proposition 2.51. Let $f \in W^{2,2}(\mathbb{R}^d, \mathbb{R}) \setminus \{0\}$. Treating

$$\rho_x = \frac{Wf(x,\bullet)}{\int_{\mathbb{R}^d} Wf(x,\xi)\,\mathrm{d}\xi} = \frac{Wf(x,\bullet)}{|f(x)|^2}$$

as a probability density function for each $x \in \mathbb{R}^d$, the expectation value and variance of the probability distribution $\mathbb{P}_{\rho_x}$ are given by

$$\mathbb{E}_{\rho_x} = 0 \qquad \text{and} \qquad \mathbb{V}_{\rho_x} = \frac{\|\nabla f(x)\|^2 - f(x)\,\Delta f(x)}{2|f(x)|^2}\,.$$

Proof. Since the Wigner transform is real-valued, we get for real-valued functions f

$$\begin{aligned} Wf(x,-\xi) &= (2\pi)^{-d} \int_{\mathbb{R}^d} f\left(x+\frac{y}{2}\right)\, f\left(x-\frac{y}{2}\right)\, e^{-iy^\intercal \xi}\,\mathrm{d}y \\ &= (2\pi)^{-d} \overline{\int_{\mathbb{R}^d} f\left(x+\frac{y}{2}\right)\, f\left(x-\frac{y}{2}\right)\, e^{iy^\intercal \xi}\,\mathrm{d}y} = \overline{Wf(x,\xi)} = Wf(x,\xi), \end{aligned}$$

and therefore the expectation value of $\mathbb{P}_{\rho_x}$ vanishes. For the variance we compute:

$$\begin{aligned} \int_{\mathbb{R}^d} Wf(x,\xi)\,\|\xi\|^2\,\mathrm{d}\xi &= (2\pi)^{-d} \int_{\mathbb{R}^{2d}} \underbrace{f\left(x+\frac{y}{2}\right)\, f\left(x-\frac{y}{2}\right)}_{=:F(x,y)}\, e^{iy^\intercal\xi}\,\xi^\intercal\xi\;\mathrm{d}y\,\mathrm{d}\xi \\ &= -i(2\pi)^{-d} \int_{\mathbb{R}^{3d}} D_y F(x,y)\cdot\xi\; e^{iy^\intercal\xi}\,\mathrm{d}y\,\mathrm{d}\xi \\ &= -(2\pi)^{-d} \int_{\mathbb{R}^{3d}} \mathrm{tr}\left[D_y^2 F(x,y)\right]\; e^{iy^\intercal\xi}\,\mathrm{d}y\,\mathrm{d}\xi \\ &= -\mathrm{tr}\left[D_y^2 F(x,0)\right] \\ &= \frac{1}{2}\left(\|\nabla f(x)\|^2 - f(x)\,\Delta f(x)\right). \end{aligned}$$

Taking the quotient proves the formula for the variance. □

This time, if we choose

$$\boxed{\mu_f(x) := \sigma_{\rho_x} = \sqrt{\mathbb{V}_{\rho_x}} = \sqrt{\frac{|\|\nabla f(x)\|^2 - f(x)\,\Delta f(x)|}{2f(x)^2}}\,,} \tag{2.7.8}$$

the adaptation conditions 2.48 are fulfilled! Also, the cumbersome integrals resulting from convolutions no longer exist, making the application of μ_f very simple in practise, once the first and second derivatives of f are known.

Remark 2.52. μ_f is ill-defined in the nodes of f. But, since we are going to use μ_f as an adaptation function for the adapted convolution

$$(f *_{\mu_f} g)(x) = \int_{\mathbb{R}^d} f(y)\,\mu_f(y)^d\, g\big(\mu_f(y)(x-y)\big)\,\mathrm{d}y\,,$$

and since for nodes y of f the term $f(y)=0$ appears as a factor, this does not cause any problems.

μ_f might also have nodes where f does not, which conflicts with Definition 2.42. In order to avoid all of these issues (and also numerical issues for very small and very

large values of μ_f), we will usually add small positive numbers $0 < \epsilon_1, \epsilon_2 \ll 1$ to the numerator and the denominator:

$$\mu_f(x) = \sqrt{\frac{|\|\nabla f(x)\|^2 - f(x)\,\Delta f(x)| + \epsilon_1}{2f(x)^2 + \epsilon_2}}\,.$$

This also guarantees that $c < \mu_f < C$ for some positive constants $c, C > 0$, if $f, \nabla f$ and Δf are bounded.

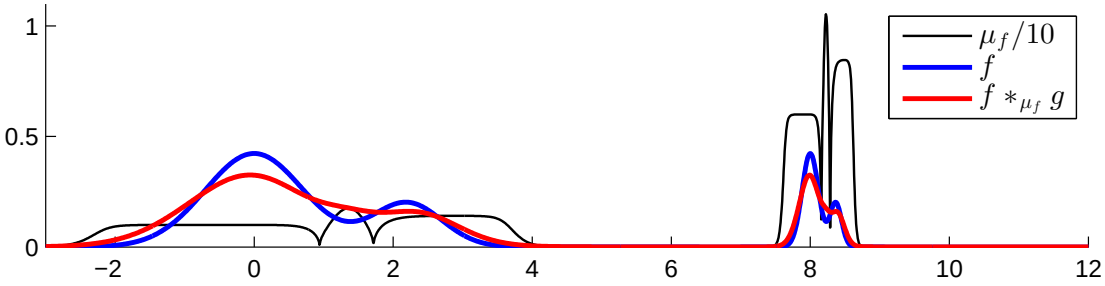


Figure 2.10: μ_f as given by formula (2.7.8) describes locally the variation of f. Choosing it as an adaptation function yields a proper scaling of g and thereby a proper smoothing of f everywhere.

2.7.4 Continuity Equation for Convolutions

Assume that $(\rho_t)_{t\geq 0}$ is a time-dependent probability density which fulfills the continuity equation

$$\partial_t \rho_t = -\operatorname{div}(\rho_t v_t) = -\operatorname{div}(j_t).$$

Assume further that we want to smooth ρ_t by considering a convolution $\rho_{g,t} = \rho_t * g_{\delta_t}$ with a smoothing kernel $g_{\delta_t}(x) = \delta_t^d g(\delta_t x)$, δ_t being a time-dependent parameter, or by taking an adapted convolution $\rho_{g,t} = \rho_t *_{\mu_t} g$ with time-dependent adaptation function $\mu_t = \mu_{\rho_t}$. How does the continuity equation have to be modified in order to describe $\rho_{g,t}$?

We were surprised that in both cases (see the following two propositions), we could find explicit formulas for the modified continuity equation:

Proposition 2.53. Let $\rho_t \in L^1(\mathbb{R}^d)$ be a time-dependent probability density function which fulfills the continuity equation

$$\partial_t \rho_t + \operatorname{div} j_t = 0 \qquad (t \in \mathbb{R})$$

for some current $j_t \in L^1(\mathbb{R}^d, \mathbb{R}^d)$, such that $(\rho_t,\, j_t)_{t\in\mathbb{R}} \in C^1\left(\mathbb{R}^{1+d}, \mathbb{R}^{1+d}\right)$.
Further, let $g \in L^1(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$ be another probability density function such that

$$g_\delta(x) := \delta^d g(\delta x), \qquad \gamma_\delta(x) := \frac{x}{\delta}\, g_\delta(x) = \delta^{d-1} x\, g(\delta x) \qquad (x \in \mathbb{R}^d,\ \delta > 0)$$

and all their first derivatives are (essentially) bounded: $g_\delta,\, \gamma_\delta \in W^{1,\infty}$ for all $\delta > 0$.
Finally, let $(\delta_t)_{t\in[0,\infty)} \in C^2(\mathbb{R}, \mathbb{R}_{>0})$. Then for each $t \in \mathbb{R}$

$$\rho_{g,t} := \rho_t * g_{\delta_t}$$

is a probability density function, which fulfills the continuity equation

$$\boxed{\partial_t \rho_{g,t} = -\operatorname{div} j_{g,t} \qquad \text{for} \qquad j_{g,t} = j_t * g_{\delta_t} - \dot\delta_t \rho_t * \gamma_{\delta_t}\,.}$$

Further, $(\rho_{g,t},\, j_{g,t})_{t\in[0,\infty)} \in C^1\left(\mathbb{R}^{1+d}, \mathbb{R}^{1+d}\right)$.

Proof. First we observe that for every $\delta > 0$

$$\partial_\delta g_\delta(x) \;=\; d\delta^{d-1} g(\delta x) + \delta^d x^\intercal \nabla g(\delta x) \;=\; \frac{d}{\delta} g_\delta(x) + \frac{x}{\delta} \nabla g_\delta(x) \;=\; \operatorname{div} \gamma_\delta(x).$$

As a consequence,

$$\begin{aligned}
\partial_t \rho_{g,t}(q) \;&=\; \int_{\mathbb{R}^d} \partial_t \rho_t(x)\, g_{\delta_t}(q-x)\,\mathrm{d}x \;+\; \int_{\mathbb{R}^d} \rho_t(x)\, \partial_\delta g_{\delta_t}(q-x)\, \dot\delta_t\,\mathrm{d}x \\
&=\; -\int_{\mathbb{R}^d} \operatorname{div} j_t(x)\, g_{\delta_t}(q-x)\,\mathrm{d}x \;+\; \dot\delta_t \int_{\mathbb{R}^d} \rho_t(x)\, \operatorname{div} \gamma_{\delta_t}(q-x)\,\mathrm{d}x \\
&=\; -\int_{\mathbb{R}^d} j_t(x)^\intercal\, \nabla g_{\delta_t}(q-x)\,\mathrm{d}x \;+\; \dot\delta_t \int_{\mathbb{R}^d} \rho_t(x)\, \operatorname{div} \gamma_{\delta_t}(q-x)\,\mathrm{d}x \\
&=\; -\operatorname{div}\left(\int_{\mathbb{R}^d} j_t(x)\, g_{\delta_t}(q-x)\,\mathrm{d}x \;-\; \dot\delta_t \int_{\mathbb{R}^d} \rho_t(x)\, \gamma_{\delta_t}(q-x)\,\mathrm{d}x \right) \\
&=\; -\operatorname{div} j_{g,t}(q).
\end{aligned}$$

The existence of all integrals follows directly from the assumptions and Proposition 2.44, while $(\rho_{g,t},\, j_{g,t})_{t\in[0,\infty)} \in C^1\left(\mathbb{R}^{1+d}, \mathbb{R}^{1+d}\right)$ follows from $(\delta_t)_{t\in[0,\infty)} \in C^2(\mathbb{R}, \mathbb{R}_{>0})$ and Proposition 2.47. □

Proposition 2.54. Let $\rho_t \in L^1(\mathbb{R}^d)$ be a time-dependent probability density function which fulfills the continuity equation

$$\partial_t \rho_t + \operatorname{div} j_t = 0 \qquad\qquad (t \in \mathbb{R})$$

for some current $j_t \in L^1(\mathbb{R}^d, \mathbb{R}^d)$, such that $(\rho_t,\, j_t)_{t\in\mathbb{R}} \in C^1\left(\mathbb{R}^{1+d}, \mathbb{R}^{1+d}\right)$.
Further, let $g \in L^1(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$ be another probability density function and

$$\gamma(x) := x\, g(x) \qquad\qquad (x \in \mathbb{R}^d),$$

such that g, γ and all their first derivatives are (essentially) bounded: $g,\, \gamma \in W^{1,\infty}$. Finally, let $(\mu_t)_{t\in\mathbb{R}} \in C^2\left(\mathbb{R}^{1+d}, \mathbb{R}_{>0}\right)$, such that $\mu_t, 1/\mu_t, \nabla\mu_t, \partial_t\mu_t \in L^\infty$ are (essentially) bounded for each $t \in \mathbb{R}$. Then for each $t \in \mathbb{R}$

$$\rho_{g,t} := \rho_t *_{\mu_t} g$$

is a probability density function, which fulfills the continuity equation

$$\boxed{\partial_t \rho_{g,t} = -\operatorname{div} j_{g,t} \qquad \text{for} \qquad j_{g,t} = j_t *_{\mu_t} g - \frac{j_t^\intercal \nabla\mu_t + \rho_t \partial_t \mu_t}{\mu_t^2} *_{\mu_t} \gamma\,.}$$

Further, $(\rho_{g,t},\, j_{g,t})_{t\in[0,\infty)} \in C^1\left(\mathbb{R}^{1+d}, \mathbb{R}^{1+d}\right)$.

Proof. We will use the abbreviation $f_\mu(x,y) := f\big(\mu(y)(x-y)\big)$ for functions $f,\, \mu\colon \mathbb{R}^d \to \mathbb{R}$ (note that this notation differs by a prefactor from the one used in Definition 2.42).

First, we make the following two observations:

$$\operatorname{div}_x\left[(x-y)\, g_{\mu_t}(x,y)\right] \;=\; d\, g_{\mu_t}(x,y) + \mu_t(y)\,(x-y)^\intercal\,(\nabla g)_{\mu_t}(x,y) \tag{2.7.9}$$

and

$$\begin{aligned}
-\int_{\mathbb{R}^d} &\operatorname{div} j_t(y)\, \mu_t(y)^d\, g_{\mu_t}(x,y)\, \mathrm{d}y \\
&= \int_{\mathbb{R}^d} j_t(y)^\intercal\, \nabla_y \left(\mu_t(y)^d\, g_{\mu_t}(x,y)\right) \mathrm{d}y \\
&= \int_{\mathbb{R}^d} j_t(y)^\intercal \Big[d\, \mu_t(y)^{d-1} \nabla\mu_t(y)\, g_{\mu_t}(x,y) \\
&\qquad\qquad + \left(\mu_t(y)^d\, \nabla\mu_t(y)(x-y)^\intercal - \mu_t(y)^{d+1}\, \mathrm{Id}\right) (\nabla g)_{\mu_t}(x,y) \Big]\, \mathrm{d}y \\
&= \int_{\mathbb{R}^d} j_t(y)^\intercal \Big[\mu_t(y)^{d-1} \nabla\mu_t(y) \Big(d\, g_{\mu_t}(x,y) + \mu_t(y)(x-y)^\intercal (\nabla g)_{\mu_t}(x,y) \Big) \\
&\qquad\qquad - \mu_t(y)^d\, \nabla_x \Big(g_{\mu_t}(x,y) \Big) \Big]\, \mathrm{d}y \\
&= \operatorname{div} \int_{\mathbb{R}^d} \mu_t(y)^{d-1}\, \left(j_t(y)^\intercal \nabla\mu_t(y)\right) g_{\mu_t}(x,y)\, (x-y) - j_t(y)\, \mu_t(y)^d\, g_{\mu_t}(x,y)\, \mathrm{d}y\,,
\end{aligned}$$

where we used equation (2.7.9) in the last step. Combining these two we get:

$$\begin{aligned}
\partial_t \rho_{g,t}(x) &= \int_{\mathbb{R}^d} \partial_t \rho_t(y)\, \mu_t(y)^d\, g_{\mu_t}(x,y)\, \mathrm{d}y \\
&\quad + \int_{\mathbb{R}^d} d\, \rho_t(y)\, \mu_t^{d-1}\, \partial_t \mu_t(y)\, g_{\mu_t}(x,y) + \rho_t(y)\, \mu_t(y)^d\, \partial_t \mu_t(y)\, (x-y)^\intercal\, (\nabla g)_{\mu_t}(x,y)\, \mathrm{d}y \\
&= \operatorname{div} \int_{\mathbb{R}^d} \mu_t(y)^{d-1}\, (j_t(y)^\intercal \nabla \mu_t(y))\, g_{\mu_t}(x,y)\, (x-y) - j_t(y)\, \mu_t(y)^d\, g_{\mu_t}(x,y)\, \mathrm{d}y \\
&\quad + \int_{\mathbb{R}^d} \rho_t(y)\, \mu_t(y)^{d-1} \partial_t \mu_t(y)\, \left[d\, g_{\mu_t}(x,y) + \mu_t(y)\, (x-y)^\intercal\, (\nabla g)_{\mu_t}(x,y) \right] \mathrm{d}y \\
&= \operatorname{div} \int_{\mathbb{R}^d} \mu_t(y)^{d-1}\, (j_t(y)^\intercal \nabla \mu_t(y))\, g_{\mu_t}(x,y)\, (x-y) - j_t(y)\, \mu_t(y)^d\, g_{\mu_t}(x,y)\, \mathrm{d}y \\
&\quad + \operatorname{div} \int_{\mathbb{R}^d} \rho_t(y)\, \mu_t(y)^{d-1} \partial_t \mu_t(y)\, g_{\mu_t}(x,y)\, \mathrm{d}y \\
&= -\operatorname{div} \int_{\mathbb{R}^d} \mu_t(y)^d\, g_{\mu_t}(x,y) \left[j_t(y) - \frac{x-y}{\mu_t(y)} \Big(j_t(y)^\intercal \nabla \mu_t(y) + \rho_t(y)\, \partial_t \mu_t(y) \Big) \right] \mathrm{d}y\,.
\end{aligned}$$

The existence of all integrals follows directly from the assumptions and Proposition 2.44, while $(j_{g,t})_{t\in[0,\infty)} \in C^1\left([0,\infty)\times\mathbb{R}^d, \mathbb{R}^d\right)$ follows from Proposition 2.47. □

Corollary 2.55. In the situation of Proposition 2.53, Proposition 2.54 respectively, the initial value problem

$$\dot{x}_t = v_{g,t}\,(x_t)\,, \qquad v_{g,t} = \frac{j_{g,t}}{\rho_{g,t}}\,(x_t)\,, \qquad x_0 = x_{\mathrm{in}} \tag{2.7.10}$$

has a unique solution $(x_t)_{t\geq 0}$ for $\rho_{g,t}$-almost every initial value x_{in}.

If x_{in} is a $\mathbb{P}_{\rho_{g,0}}$-distributed random variable (or $x_{k,\mathrm{in}}$ are $\mathbb{P}_{\rho_{g,0}}$-distributed points), then x_t (or $x_{k,t}$) will stay $\mathbb{P}_{\rho_{g,t}}$-distributed for all times $t \geq 0$.

Proof. This is a direct consequence of Propositions 2.53, 2.54 and Proposition 2.12. □

Definition 2.56 (Quasi-Bohmian Trajectories)**.** In the case of Bohmian mechanics, i.e. $v_t = \Im\left[\frac{\nabla\psi_t}{\psi_t}\right]$, where ψ_t is the wave function, the trajectories resulting from the initial value problem (2.7.10) will be referred to as *quasi-Bohmian trajectories*.

2.8 Radial Basis Functions

Approximations with radial basis functions (see. e.g. [Buh03], [Wen05]. [Fas07b]) is a growing field in numerical analysis. Its aim is to approximate a given function $f : \Omega \subseteq$

$\mathbb{R}^d \to \mathbb{R}$ by a linear combination of radially symmetric functions $\eta_j(x) = \phi(\|x - x_j\|)$, which are identical up to their centers:

$$f(x) \approx u(x) = \sum_{j=1}^{N} c_j \, \phi(\|x - x_j\|), \qquad c_j \in \mathbb{R},\ x_j \in \Omega,\ \phi \in C(\mathbb{R}^d, \mathbb{R}).$$

One important advantage of this class of approximations is the arbitrariness of the choice of the set of centers $\mathcal{X} = \{x_j \mid j = 1, \dots, N\}$. Instead of having to form a specific grid, the centers may be scattered in the domain of approximation Ω. The bounds for the approximation error are usually expressed in terms of the fill distance

$$h_{\mathcal{X},\Omega} = \max_{x \in \Omega} \min_{j=1,\dots,N} \|x - x_j\|.$$

However, this suggests, though by no means necessary, to position the points x_j on some kind of grid or close to one in order to minimize $h_{\mathcal{X},\Omega}$.

Once the approximation space $M = \operatorname{span}\{\eta_j \mid j = 1, \dots, N\}$ is fixed, one has to decide on the type of approximation, i.e. on how to choose the coefficients c_j. A very common approach is the bestapproximation $u = \sum_{j=1}^{N} c_j \eta_j$ with respect to some norm $\|\bullet\|$, which is normally induced by a scalar product $\langle \bullet, \bullet \rangle$, e.g. the L^2 norm:

$$b_k := \langle \eta_k, f \rangle \stackrel{!}{=} \langle \eta_k, u \rangle = \sum_{j=1}^{N} c_j \underbrace{\langle \eta_k, \eta_j \rangle}_{=: a_{kj}} \qquad \forall k = 1, \dots, N$$

$$\text{i.e.} \qquad b \stackrel{!}{=} Ac$$

in the matrix-vector notation $A = (a_{kj})_{k,j}$, $b = (b_k)_k$, $c = (c_k)_k$. The numerical solution of this system of linear equations requires for the so-called stiffness matrix A not to be ill-conditioned, which we will discuss below.

Remark 2.57. Note that the Gramian matrix A appears in the Dirac-Frenkel variational principle: When applying a time step method to the ODE (2.5.4), in our case it will be an explicit Runge-Kutta method, one has to solve such linear systems of equations in each time step.

Another common approximation approach is interpolation of f in $\mathcal{X}$:

$$\tilde{b}_k := f(x_k) \stackrel{!}{=} u(x_k) = \sum_{j=1}^{N} c_j \underbrace{\eta_j(x_k)}_{\tilde{a}_{kj}} \qquad \forall k = 1, \dots, N$$

$$\text{i.e.} \qquad \tilde{b} \stackrel{!}{=} \tilde{A}c.$$

Again the interpolation matrix $\tilde{A}$ must not be ill-conditioned.

Remark 2.58. Note that the least squares method, which is yet another common approach, can be expressed as a bestapproximation problem by choosing a proper (semi-)norm.

2.8.1 The Condition Number of Stiffness and Interpolation Matrices

Since it is difficult to give a general characterization of ill-conditioned matrices, we will settle for a special case and give an intuitive reasoning: Assume that $\eta_{j_1} \approx \eta_{j_2}$, say $\|\eta_{j_1} - \eta_{j_2}\| = \delta$. Then their scalar products with all basis functions η_k are also close to each other:

$$|\langle \eta_{j_1}, \eta_k \rangle - \langle \eta_{j_2}, \eta_k \rangle| = |\langle \eta_{j_1} - \eta_{j_2}, \eta_k \rangle| \leq \delta \|\eta_{k,t}\| = \delta \qquad \text{for all } k = 1, \ldots, N$$

by the Cauchy-Schwarz inequality (we used that in our case $\|\eta_{k,t}\| = 1$).

The same argumentation works for interpolation matrices: Assume $\|\eta_{j_1} - \eta_{j_2}\|_{L^\infty} = \delta$. Then

$$|\eta_{j_1}(x_k) - \eta_{j_2}(x_k)| \leq \delta \qquad \text{for all } k = 1, \ldots, N.$$

Therefore we would get two nearly identical rows in the stiffness matrix A and the interpolation matrix $\tilde{A}$, respectively, making them nearly rank deficient and thereby ill-conditioned.
In the case of Gaussian radial basis functions

$$\eta_j(x) = \left(\frac{2\epsilon^2}{\pi}\right)^{\frac{d}{4}} \exp\left(-\epsilon^2 \|x - x_j\|^2\right) \qquad \text{for } L^2\text{-bestapproximation,} \tag{2.8.1}$$

$$\eta_j(x) = \exp\left(-\epsilon^2 \|x - x_j\|^2\right) \qquad \text{for interpolation,} \tag{2.8.2}$$

this leads to a simple rule of thumb:

The closer the center $x_{j,t}$ is to the neighboring centers $x_{k,t}$, the more peaked the Gaussian $\eta_{j,t}$ must be chosen!

Remark 2.59. The prefactors are chosen in such a way, that the diagonals of the stiffness and interpolation matrices consist only of ones, which improves their condition numbers. Of course, the prefactors do not vary the approximation space.

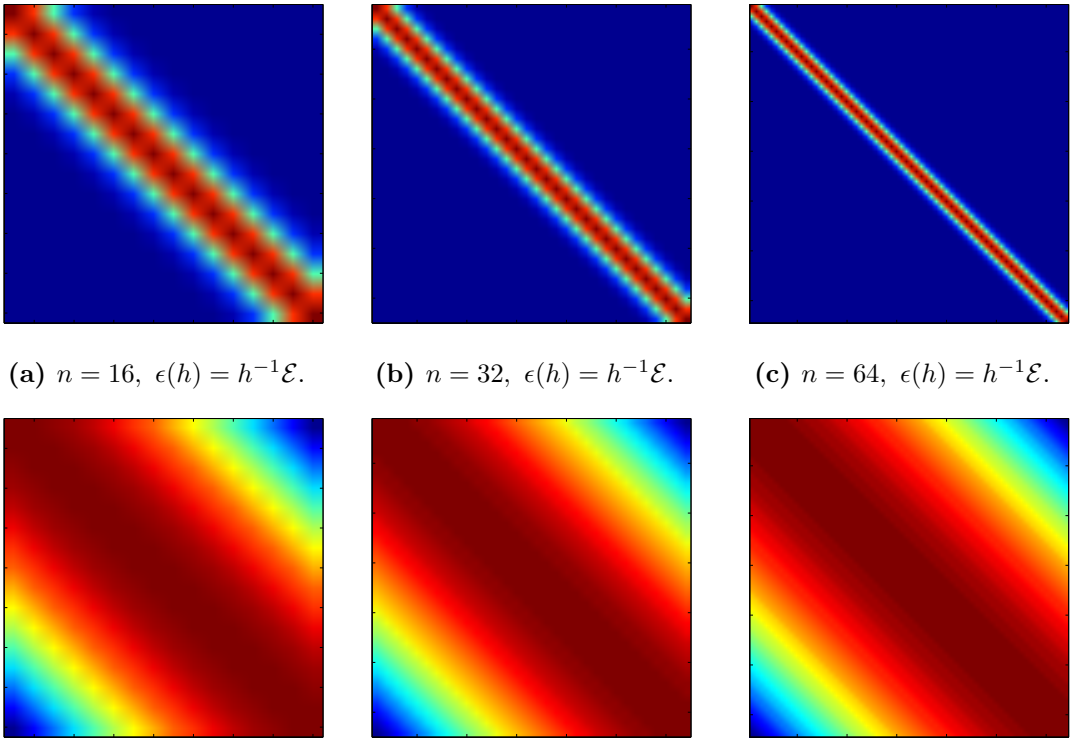
In fact, the so-called shape parameter ϵ satisfies a trade-off principle. Roughly speaking, the smaller it is, the better the approximation properties, but the worse the condition number of the stiffness and interpolation matrices. This dependence has

stirred up a wide discussion on how to choose the "optimal" shape parameter, see e.g. [Fas07a].

To make these connections between the distance of the centers x_j, the shape parameter ϵ of the basis functions and the condition number of the stiffness and interpolation matrices more concrete, let us consider the following example:

Given $f \in L^2(\mathbb{R}^d)$ with support $\text{supp}(f) \subseteq \Omega = (0,1)^d$. In order to approximate f we choose Gaussian radial basis functions of the form (2.8.1) centered on an equidistant grid $\mathcal{X} = \{x_1, \ldots, x_N\} \subseteq \Omega,\ N = n^d$ for some $n \in \mathbb{N}$. Assuming that we let the mesh size $h \sim 1/n$ go to zero, if we let the shape parameter ϵ constant in h, neighbored basis functions would "overlap" more and more (thereby becoming close to one another in the sense discussed above). Since matrices with at least two similar rows (or columns) are nearly rank deficient, the condition numbers of the the stiffness and the interpolation matrices would explode.

However, if we choose the shape parameter $\epsilon = \epsilon(h) = h^{-1}\mathcal{E}$ for some constant basic shape parameter $\mathcal{E}$, the "overlap" of neighbored basis functions stays constant in h and the condition number, though growing due to a larger matrix size, will not explode as in the upper example.



(a) $n = 16,\ \epsilon(h) = h^{-1}\mathcal{E}$. **(b)** $n = 32,\ \epsilon(h) = h^{-1}\mathcal{E}$. **(c)** $n = 64,\ \epsilon(h) = h^{-1}\mathcal{E}$.

(d) $n = 16,\ \epsilon(h) = \mathcal{E}$. **(e)** $n = 32,\ \epsilon(h) = \mathcal{E}$. **(f)** $n = 64,\ \epsilon(h) = \mathcal{E}$.

Figure 2.11: Visualization of the entries of the stiffness matrix (the results for the interpolation matrix are analogous) in dependence of n for various choices of $\epsilon(h)$.

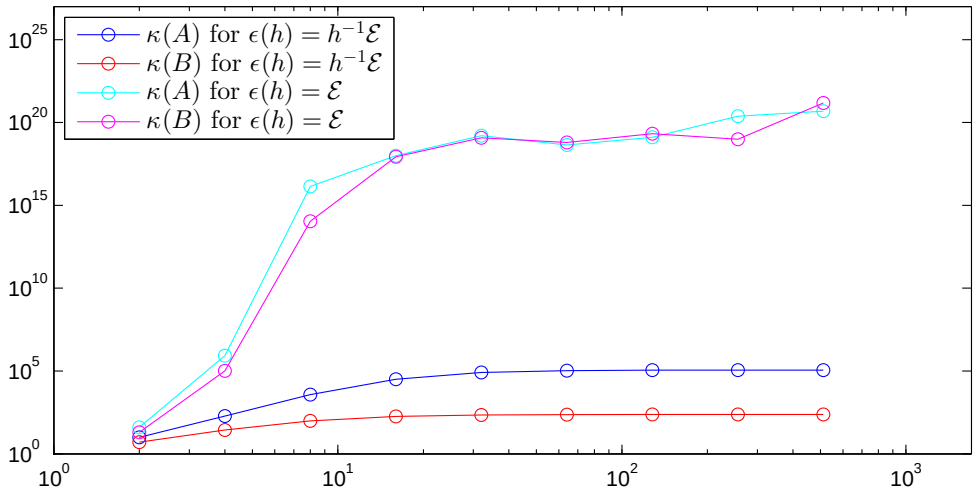


Figure 2.12: Condition numbers of the stiffness matrix A and interpolation matrix B plotted over n for two different choices for $\epsilon(h)$.

2.8.2 Approximate Approximations

The above discussion of the condition numbers of stiffness and interpolation matrices suggests the following choice of basis functions for an equidistant grid $\mathcal{X} = \{x_1, \ldots, x_N\} \subseteq \Omega = (0,1)^d,\ N = n^d,\ h = 1/n$:

$$\eta_j(x) = \kappa(h)\, \eta\left(h^{-1}\mathcal{E}(x - x_j)\right).$$

Here, $\mathcal{E} > 0$ is constant and $\kappa(h) > 0$ is a suitably chosen prefactor, which can be ignored, since it does not influence the approximation space. Further, the so-called generating function $\eta = \phi(\|\bullet\|) \in C(\mathbb{R}^d, \mathbb{R})$ denotes a radially symmetric function, which decays fast enough and thereby keeps the overlap small enough (this last property is specified below).

This dependence of the shape parameter on the point density, which is referred to as *stationary approximation* by Fasshauer (see [Fas07b]) or as *approximate approximation* by Maz'ya and Schmidt (see [Maz07]), results in approximation methods that do not converge in general. Instead, the convergence of the approximation error to zero can only be reached up to a certain saturation error. The latter can be tuned to be arbitary small by choosing a proper parameter $\mathcal{E} > 0$ small enough and therefore it can be neglected in many numerical computations (hence the term "approximate approximations"). Here we will follow the presentation of [Maz07]. We will start with the introduction of quasi-interpolation, which will be followed by its error analysis. We will use the standard multiindex notation (2.7.4) – (2.7.6) extended by:

$$\begin{aligned} &\nabla_k f := (\partial^\alpha f)_{|\alpha|=k}\,, && k \in \mathbb{N},\ f \in C^k\,, \\ &\|\nabla_k f\|_{L^p(\Omega)} := \left\|(\|\partial^\alpha f\|_{L^p(\Omega)})_{|\alpha|=k}\right\|_p\,, && k \in \mathbb{N},\ f \in W^{k,p}(\Omega), \end{aligned}$$

where $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ and $(\partial^\alpha f)_{|\alpha|=k}$ denotes the vector of partial derivatives in lexicographic order. Further, we will assume that the generating function η fulfills the following conditions:

Definition 2.60 ((extended) decay condition, moment condition)**.** Let $d \in \mathbb{N}$ and $\mu = \lfloor \frac{d}{2} \rfloor + 1$. We say that $\eta \in C(\mathbb{R}^d, \mathbb{R})$ fulfills

1. the *decay condition*, if

$$\exists K_1 > 0,\ K_2 > d: \quad |\eta(x)| \leq K_1(1 + |x|)^{-K_2}, \tag{2.8.3}$$

2. the *extended decay condition*, if

$$\eta \in C^\mu \text{ and } \partial^\alpha \eta \text{ fulfills (2.8.3) for all } \ 0 \le |\alpha| \le \mu, \tag{2.8.4}$$

3. the *moment condition* of order $K_3 \in \mathbb{N}$, if

$$\int_{\mathbb{R}^d} \eta(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \int_{\mathbb{R}^d} x^\alpha \eta(x)\,\mathrm{d}x = 0 \quad \text{for all } \ 1 \le |\alpha| \le K_3, \tag{2.8.5}$$

These conditions allow us to define quasi-interpolation and discuss its error analysis.

Definition 2.61 (quasi interpolation)**.** Let $\Omega \subseteq \mathbb{R}^d$ be a domain, $f \in L^p(\Omega)$, $1 \le p \le \infty$, and $\eta \in C(\mathbb{R}^d, \mathbb{R})$ fulfill the decay condition (2.8.3). We define the *quasi-interpolanion* with mesh size $h > 0$ and basic shape parameter $\mathcal{E} > 0$ of f as

$$(Q_h f)(x) := (Q_{h,\mathcal{E},\eta} f)(x) := \mathcal{E}^d \sum_{m \in \mathbb{Z}^d} f(hm)\ \eta\left(\frac{\mathcal{E}(x - hm)}{h}\right), \tag{2.8.6}$$

where $f(hm) := 0$ for $hm \notin \Omega$.

$Q_h f$ can be viewed as the *semi-discrete convolution* with mesh size h

$$(g_1 *_h g_2)(x) = \sum_{m \in \mathbb{Z}^d} g_1(hm) g_2(x/h - m)$$

of f with the generating function $\eta(\mathcal{E}\cdot\bullet)$. Using this observation, one can apply Young's and Hölder's inequalities to prove the well-definedness of and an upper bound for $Q_h f$:

Lemma 2.62. Let $h, \mathcal{E} > 0$, $\eta \in C(\mathbb{R}^d, \mathbb{R})$ fulfill the decay condition (2.8.3), $f \in L^p(\mathbb{R}^d, \mathbb{R})$, $1 \le p \le \infty$ and $|f|^p$ be Riemann integrable if $p < \infty$. Then

$$\|f\|_{p,h} := \begin{cases} \left(h^d \sum_{m \in \mathbb{Z}^d} |f(hm)|^p\right)^{1/p}, & 1 \le p < \infty, \\ \sup_{m \in \mathbb{Z}^d} |f(hm)|, & p = \infty \end{cases}$$

is uniformly bounded with respect to h and $Q_h f \in L^p(\mathbb{R}^d, \mathbb{R})$ with

$$\|Q_h f\|_{L^p} \le C \|f\|_{p,h}$$

for some constant $C = C(p, \mathcal{E}, \eta) > 0$, which is independent of f and h.

Proof. See [Maz07, Lemma 2.1, Corollary 2.2 and Remark 2.3]. □

One possible error estimate of quasi-interpolation is given by the following theorem, where $W^{k,p}$ denotes the Sobolev space of order k corresponding to the L^p-norm:

Theorem 2.63. Let $\Omega \subseteq \mathbb{R}^d$ be a domain, $1 \leq p \leq \infty$, $f \in W^{k,p}(\Omega, \mathbb{R})$ and $\eta \in C(\mathbb{R}^d, \mathbb{R})$ fulfill the extended decay condition (2.8.4) with constants $K_1 > 0$, $K_2 > d$ and the moment condition of order K_3, such that $d/p < k < K_2$.
Then there exists a constant $C_\eta > 0$ independent of $f, h, \mathcal{E}$ and for every $\varepsilon > 0$ an $\mathcal{E}_0 > 0$ such that for every $0 < \mathcal{E} \leq \mathcal{E}_0$ there exists $\kappa = \kappa(\varepsilon, \mathcal{E}) > 0$, such that for all $h > 0$ the quasi-interpolant $Q_h f$ of f satisfies
(i.e. $\exists C_\eta > 0\ \forall \varepsilon > 0\ \exists \mathcal{E}_0 > 0\ \forall 0 < \mathcal{E} \leq \mathcal{E}_0\ \exists \kappa > 0\ \forall h > 0$)

$$\|f - Q_h f\|_{L^p(\Omega_{\kappa h})} \leq C_\eta \left(\mathcal{E}^{-1} h\right)^M \|\nabla_M f\|_{L^p(\Omega)} + \varepsilon \sum_{j=0}^{M-1} \left(\mathcal{E}^{-1} h\right)^j \|\nabla_j f\|_{L^p(\Omega)},$$

where $M = \min(K_3, k)$ and $\Omega_\tau := \{x \in \Omega : B_\tau(x) \subseteq \Omega\}$.

Proof. See [Maz07, Theorem 2.28 and Lemma 2.30]. □

For a plot showing the convergence of the approximation up to a saturation error (and how to tune the latter one by decreasing $\mathcal{E}$) see Figures 4.3 and 4.6.

Chapter 3

Choice of the Approximation Manifold

The main challenge when applying a numerical method to a PDE, like the variational principle to the Schrödinger equation, is the choice of a "good" approximation manifold, meaning that it is (and stays) close to the true solution ψ_t of the PDE. In our case, the manifold will be a time-dependent N-dimensional complex vector space $M_t = \text{span}_{\mathbb{C}}\{\eta_1, \dots, \eta_N\}$ spanned by Gaussian basis functions η_j and the approximant will be denoted by

$$u_t = \sum_{j=1}^{N} c_{j,t}\eta_{j,t} \approx \psi_t.$$

We want to stress that the decision for the basis functions to be Gaussian in not essential for our algorithm, They should, however, fulfill the extended decay condition (2.8.4) and the moment condition (2.8.5).

The main advantage of our ansatz is the time-dependence of the approximation manifold - it adapts automatically to the wave function ψ_t and changes as ψ_t evolves in time.

For sake of notation, we will mostly omit the index t in our presentation, since the choice of the manifold will mainly be explained at a fixed time t. We ask the reader to keep in mind that ψ, ρ, ρ_g, R, R_g, v, M, η_j, ϵ_j, p_j, q_j, x_j, δ and μ are all time-dependent quantities and we will deal with their time-evolution in Section 3.2, where the index t will recur.

We will begin with an intuitive explanation of our choice of the approximation manifold in Section 3.1, which will be followed by an accurate presentation of the resulting algorithm in Section 3.2.

Section 3.3 will address the mathematically rigorous analysis of the manifold's approximation properties. The chapter will be completed by a proof of the linear independence of our so-called basis functions in a special case.

3.1 Outline in 1-D

The basis functions will be adapted to the wave function in five steps, which we will present now and four of which will be illustrated in Figure 3.1. This section is meant to give the reader an intuitive understanding of our ansatz and does not aim at mathematical rigour (which will follow in Section 3.3).

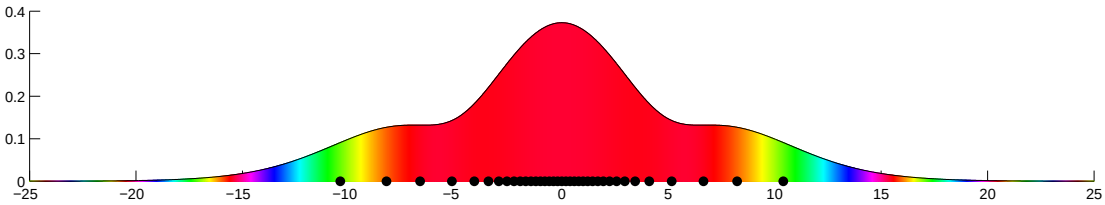
As in the last section, $N = n^d$ will denote the number of basis functions and $h = n^{-1}$ the "mesh size", even though we will not have an actual mesh here. The visualizations use the example (4.1.1) (for various values of t).

3.1.1 Step 1: Centers of the Basis Functions (see Figure 3.1b)

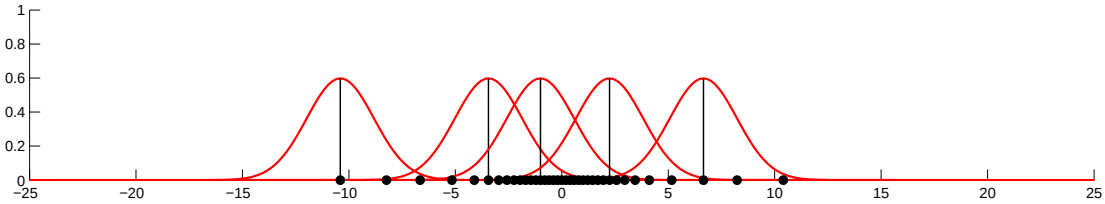
In order to get a good approximation, we want our basis functions to lie in the region of interest (i.e. where the wave function is far from zero) and to adapt to it, when it changes in time. As mentioned in Section 2.4 (see e.g. Remark 2.26), that is exactly what Bohmian trajectories do - they group up in regions of high values of $|\psi_t|$ and thin out with decaying $|\psi_t|$. Therefore we will propagate N Bohmian trajectories $q_j = q_{j,t}$ and choose the basis functions to be to be centered at these:

$$\eta_j(x) = \left(\frac{2\epsilon}{\pi}\right)^{\frac{d}{4}} \exp\left(-\epsilon\,(x - q_j)^2\right),$$

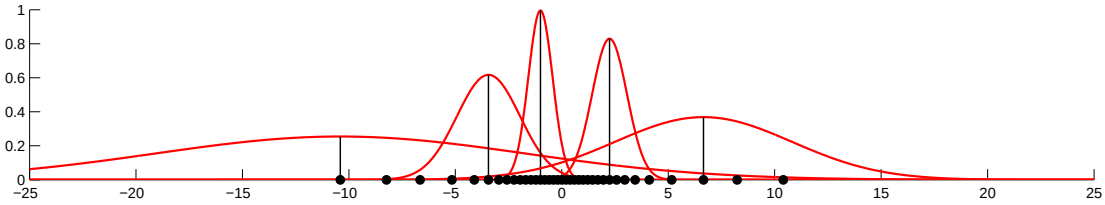
where the shape parameter is chosen as $\epsilon = h^{-2}\mathcal{E}^2$ with constant basic shape parameter $\mathcal{E} > 0$, as motivated in Section 2.8. The reason for the choice of the prefactor is given in Remark 2.59 (scaling the basis functions clearly does not modify the vector space).



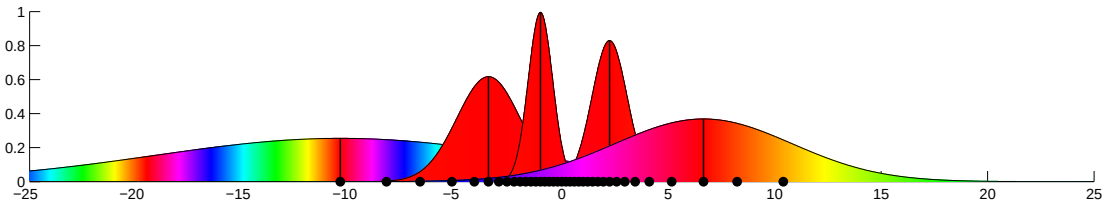
(a) The wavefunction we want to approximate, together with ρ_t-distributed points q_j.



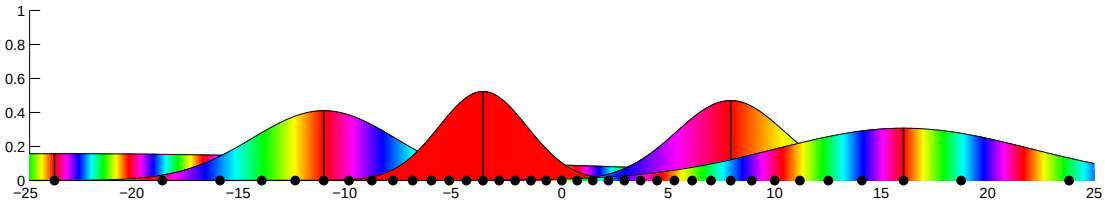
(b) Gaussian basis functions with constant width, centered in the points q_j.



(c) Gaussian basis functions with adapted widths.



(d) A momentum term was added.



(e) The basis functions after "pulling apart" their centers.

Figure 3.1: The construction of a proper approximation basis in four steps (for reasons of visualisation only a few basis functions are plotted).

3.1.2 Step 2: Widths of the Basis Functions (see Figure 3.1c)

There are three reasons why a constant (in time *and* space) shape parameter ϵ is a bad choice:

1. The wave function is changing in time. If, for example, it diffuses as in the case of the free Schrödinger equation, it would be appropriate to flatten out our Gaussians in order to keep a good approximation.

2. In regions, where the position density is low and we have few centers q_j, one would get a better approximation by choosing flat Gaussians, while a choice of peaked Gaussians is favorable in high-density-regions.

3. The resulting stiffness matrix A_t gets ill-conditioned if we choose ϵ too small in regions where we have many centers (see section 2.8.1). On the other hand, the approximation gets terrible if we choose our basis functions too peaked, especially in regions of low density (and thereby few basis functions).

These disadvantages motivate the following choice of basis functions:

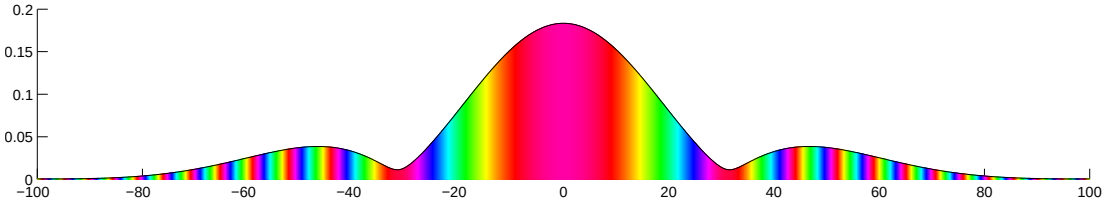
$$\eta_j(x) = \left(\frac{2\epsilon_j}{\pi}\right)^{\frac{d}{4}} \exp\left(-\epsilon_j \left(x - q_j\right)^2\right), \quad \text{where} \;\; \epsilon_j = \epsilon_{j,t} = h^{-2}\mathcal{E}^2 \rho_t(q_j)^\beta, \quad \mathcal{E}, \beta > 0.$$

The precise choice of β is discussed in Section 3.1.4.

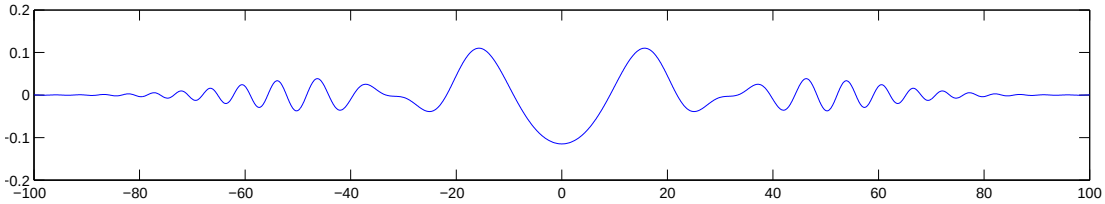
This way, the basis functions are flat in regions of low density (where there are few centers q_j) and peaked in regions of high density (with many centers q_j), leading to a good approximation while avoiding the ill-conditioning of the stiffness matrix. Also, the adaption in time happens automatically - if the wave function flattens out, so does $\rho_t = |\psi_t|^p / \|\psi_t\|_{L^p}^p$ and thereby also the basis functions.

3.1.3 Step 3: Adding a momentum (see Figure 3.1d)

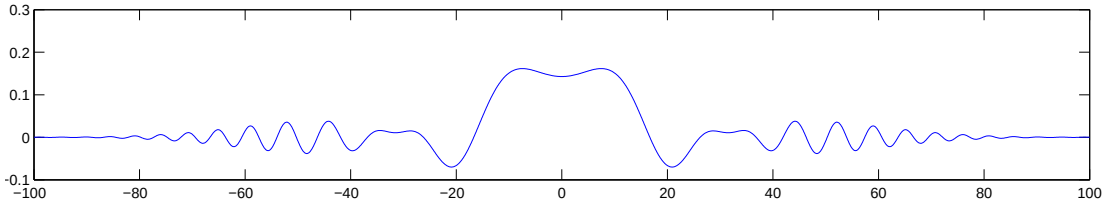
Consider the free Schrödinger equation with a real valued initial wave function. After some time it will dissolve and become highly oscillatory at the edges (see Figure 3.1a).



(a) The wavefunction at time $t = 60$.



(b) Real part of upper wavefunction.



(c) Imaginary part of upper wavefunction.

Figure 3.2: In the case of e.g. the free Schrödinger equation (meaning $V = 0$, see example (4.1.1)) the wavefunction tends to become highly oscillatory at the edges.

With our ansatz (which approximates the real and the imaginary parts of the wavefunction ψ separately, since the basis functions are real-valued and only the coefficients may be complex-valued), these oscillations cause serious problems, especially since we have very few points in these regions. Therefore we will add a "momentum" term to our basis functions and we already know how to choose it from Bohmian mechanics: Rewriting the wave function $\psi = Re^{iS}$ in the polar form, the best linear approximation to S at q_j is given

$$S(x) \approx S(q_j) + \nabla S(q_j)^\intercal (x - q_j) = S(q_j) + v^\intercal (x - q_j),$$

where $v = \Im\left[\frac{\nabla\psi}{\psi}\right]$ is the Bohmian velocity, see (2.4.2) and Remark 2.27. Since the constant term is handled by the coefficient c_j in the approximation $\psi \approx u = \sum_{j=1}^{N} c_j \eta_j$

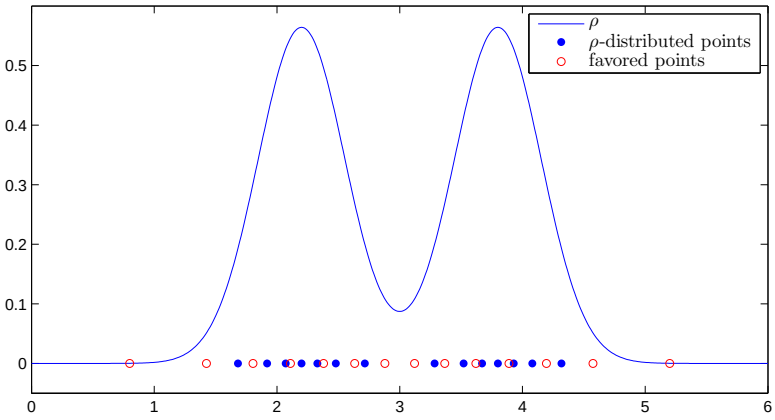
we arrive at following choice for the basis functions:

$$\eta_j(x) = \left(\frac{2\epsilon_j}{\pi}\right)^{\frac{d}{4}} \exp\left(-\epsilon_j (x-q_j)^2 + ip_j(x-q_j)\right), \quad \text{where} \quad p_j = v(q_j).$$

Remark 3.1. This modification is tailored to the specific case of the Schrödinger equation and will not be considered for the error analysis in Section 3.3. In the numerical experiments on the Schrödinger equation (Section 4) we will always compare both cases, with and without momentum term, where we will see that the latter one performs poorly when the oscillations get higher.

3.1.4 Step 4: Pulling the Centers Apart (see Figure 3.1e)

Our experiments have shown, that the ρ-distributed centers q_j tend to stick too close to each other:



This has two negative effects:

1. The points do not cover the whole region of interest, but only a small part of very high density (and there we get too many). This yields a bad approximation space.

2. If the points are too close to each other, the stiffness matrix A becomes ill-conditioned (see subsection 2.8.1).

The reason why we chose the centers to be ρ_t-distributed in the first place was their localization in the region of interest. In order to keep this property but avoid the

above mentioned disadvantages, we will choose new centers $x_j = x_{j,t}$, these being ρ_g-distributed instead of ρ-distributed, where we smoothed $\rho = \rho_t$ by taking its convolution with another Gaussian centered at zero:

$$\rho_g = \rho_{g,t} := \rho * g_\delta \quad , \quad g_\delta(x) = \delta^d g(\delta x) \quad , \quad g(x) = \pi^{-d/2} e^{-|x|^2}.$$

Again, the choice of g being a Gaussian is not essential.

The resulting basis functions are:

$$\eta_j(x) = \left(\frac{2\epsilon_j}{\pi}\right)^{\frac{d}{4}} \exp\left(-\epsilon_j (x - x_j)^2 + i p_j (x - x_j)\right), \tag{3.1.1}$$

where $p_j = v_t(x_j)$, $\epsilon_j = h^{-2}\mathcal{E}^2 \rho_g(x_j)^\beta$ for some $\kappa_\epsilon, \beta > 0$ and the points x_j are ρ_g-distributed.

The parameter $\delta = \delta_t$ is chosen in dependence of the density ρ: high for peaked ρ and low for wide ρ to get analogous results. We will now discuss the precise choice of the parameters δ and β using the concept of scale invariance.

Assume our whole approximation problem

(1) $\psi \in L^p(\mathbb{R}^d, \mathbb{C})$ and $\rho = \dfrac{|\psi|^p}{\|\psi\|_{L^p}^p} \in L^1(\mathbb{R}^d, \mathbb{R})$, $1 \le p < \infty$,

(2) $\rho_g = \rho * g_\delta$ for some $\delta > 0$,

(3) x_j, ϵ_j, p_j and η_j as in (3.1.1),

(4) $\psi \approx u = \sum_{j=1}^N c_j \eta_j$ for some $c_1, \dots, c_N \in \mathbb{C}$

is scaled by a factor $\alpha > 0$ to

(5) $\tilde{\psi}(x) = \alpha^{d/p} \psi(\alpha x)$ and $\tilde{\rho}(x) = \dfrac{|\tilde{\psi}(x)|^p}{\|\tilde{\psi}\|_{L^p}^p} = \alpha^d \rho(\alpha x)$,

(6) $\tilde{\rho}_g = \tilde{\rho} * g_{\tilde{\delta}}$ for some $\tilde{\delta} > 0$,

(7) $\tilde{x}_j = \frac{x_j}{\alpha}$, $\tilde{\epsilon}_j = h^{-2}\mathcal{E}^2 \tilde{\rho}_g(\tilde{x}_j)^\beta$, $\tilde{p}_j = \tilde{v}_t(\tilde{x}_j)$ (where $\tilde{v}_t = \Im\left[\frac{\nabla\tilde{\psi}}{\tilde{\psi}}\right]$) and

$$\tilde{\eta}_j(x) = \left(\frac{2\tilde{\epsilon}_j}{\pi}\right)^{\frac{d}{4}} \exp\left(-\tilde{\epsilon}_j (x - \tilde{x}_j)^2 + i\tilde{p}_j (x - \tilde{x}_j)\right),$$

(8) $\tilde{\psi} \approx \tilde{u} := \sum_{j=1}^N \tilde{c}_j \tilde{\eta}_j$ for $\tilde{c}_j = \alpha^{\frac{(2-p)d}{2p}} c_j$, $j = 1, \dots, N$,

then we demand, in order to keep the approximation properties of the original problem, that our basis functions and the density ρ_g are scaled by the same factor:

$$\tilde{\eta}_j(x) \stackrel{!}{=} \alpha^{\frac{d}{2}} \eta_j(\alpha x) \quad , \quad \tilde{\rho}_g(q) \stackrel{!}{=} \alpha^d \rho_g(\alpha q) \tag{3.1.2}$$

This demand is justified by the following

Proposition 3.2. Let $\psi, \tilde{\psi}, u, \tilde{u} \in L^p(\mathbb{R}^d, \mathbb{C}),\ 1 \leq p < \infty$, fulfill conditions (1) – (8) and (3.1.2). Then

$$\|\tilde{\psi} - \tilde{u}\|_{L^p} = \|\psi - u\|_{L^p}.$$

Proof. A simple applications of the transformation formula yields:

$$\begin{aligned}
\|\tilde{\psi} - \tilde{u}\|_{L^p}^p &= \int_{\mathbb{R}^d} \left| \tilde{\psi}(x) - \sum_{j=1}^N \tilde{c}_j \tilde{\eta}_j(x) \right|^p \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left| \alpha^{\frac{d}{p}} \psi(\alpha x) - \sum_{j=1}^N \alpha^{\frac{(2-p)d}{2p}} c_j \alpha^{\frac{d}{2}} \eta_j(\alpha x) \right|^p \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \alpha^d \left| \psi(\alpha x) - \sum_{j=1}^N c_j \eta_j(\alpha x) \right|^p \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left| \psi(x) - \sum_{j=1}^N c_j \eta_j(x) \right|^p \mathrm{d}x \\
&= \|\psi - u\|_{L^p}^p .
\end{aligned}$$

□

Remark 3.3. The prefactors $\alpha^{d/p}$ in $\tilde{\psi}(x) = \alpha^{d/p}\psi(\alpha x)$ and $\alpha^{d/2}$ in $\tilde{\eta}(x) = \alpha^{d/2}\eta(\alpha x)$ are optional and tailored to the case of the Schrödinger equation. Any other factors can be chosen, they just have to be compensated by corresponding prefactors in the coefficients $\tilde{c}_j$. The scaling of the probability density $\tilde{\rho}(x) = \alpha^d \rho(\alpha x)$ remains unaffected by these manipulations.

We will now show that the choices $\beta = \frac{2}{d}$ and $\delta(\rho) = \kappa_\delta \int_{\mathbb{R}^d} \rho^{\frac{d+1}{d}}(x)\, dx$ fulfill our demand (3.1.2):

Proposition 3.4. Under conditions (1) – (8), the choices

$$\beta = \frac{2}{d} \quad \text{and} \quad \delta(\rho) = \kappa_\delta \int_{\mathbb{R}^d} \rho^{\frac{d+1}{d}}(x)\, \mathrm{d}x,$$

where $\kappa_\delta > 0$ is a positive constant, fulfill the required relations (3.1.2). Furthermore,

$$\langle \tilde{\eta}_j, \tilde{\eta}_k \rangle_{L^2} = \langle \eta_j, \eta_k \rangle_{L^2} .$$

Proof. From

$$\begin{aligned}\tilde{\delta} &= \delta(\tilde{\rho}) = \kappa_\delta \int_{\mathbb{R}^d} \tilde{\rho}^{\frac{d+1}{d}}(x)\,\mathrm{d}x = \kappa_\delta \int_{\mathbb{R}^d} \alpha^{d+1}\rho^{\frac{d+1}{d}}(\alpha x)\,\mathrm{d}x \overset{y=\alpha x}{=} \\ &= \alpha\kappa_\delta \int_{\mathbb{R}^d} \rho^{\frac{d+1}{d}}(y)\,\mathrm{d}y = \alpha\delta\end{aligned}$$

we conclude

$$\begin{aligned}\tilde{\rho}^{g,\tilde{\delta}}(q) &= (\tilde{\rho} * g_{\tilde{\delta}})(q) = \int_{\mathbb{R}^d} \tilde{\rho}(x)\, g_{\tilde{\delta}}(q-x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \alpha^d \rho(\alpha x)\,\alpha^d \delta^d g\big(\alpha\delta(q-x)\big)\,\mathrm{d}x \overset{y=\alpha x}{=} \alpha^d \int_{\mathbb{R}^d} \rho(y)\,\delta^d g\big(\delta(\alpha q - y)\big)\,\mathrm{d}y \\ &= \alpha^d(\rho * g_\delta)(\alpha q) = \alpha^d \rho^{g,\delta}(\alpha q).\end{aligned}$$

Then

$$\begin{aligned}\tilde{\epsilon}_j &= h^{-2}\mathcal{E}^2\tilde{\rho}(\tilde{q}_j)^\beta = h^{-2}\mathcal{E}^2\left(\alpha^d\rho\left(\alpha\frac{q_j}{\alpha}\right)\right)^{\frac{2}{d}} = \alpha^2\epsilon_j, \\ \tilde{p}_j &= \Im\left[\frac{\nabla\tilde{\psi}\big(\tilde{x}_j\big)}{\tilde{\psi}\big(\tilde{x}_j\big)}\right] = \Im\left[\frac{\alpha^{\frac{d}{2}+1}\nabla\psi\big(\alpha\frac{x_j}{\alpha}\big)}{\alpha^{\frac{d}{2}}\psi\big(\alpha\frac{x_j}{\alpha}\big)}\right] = \alpha\, p_j,\end{aligned}$$

implies:

$$\begin{aligned}\tilde{\eta}_j(x) &= \left(\frac{2\tilde{\epsilon}_j}{\pi}\right)^{\frac{d}{4}} \exp\Big(-\tilde{\epsilon}_j\big(x-\tilde{x}_j\big)^2 + i\tilde{p}_j\big(x-\tilde{x}_j\big)\Big) \\ &= \left(\frac{2\alpha^2\epsilon_j}{\pi}\right)^{\frac{d}{4}} \exp\left(-\alpha^2\epsilon_j\left(x-\frac{x_j}{\alpha}\right)^2 + i\alpha p_j\left(x-\frac{x_j}{\alpha}\right)\right) = \alpha^{\frac{d}{2}}\eta_j(\alpha x), \\ \langle\tilde{\eta}_j,\tilde{\eta}_k\rangle_{L^2} &= \int_{\mathbb{R}^d} \overline{\tilde{\eta}}_j(x)\,\tilde{\eta}_k(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \alpha^{d/2}\overline{\eta}_j(\alpha x)\,\alpha^{d/2}\eta_k(\alpha x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \overline{\eta}_j(x)\,\eta_k(x)\,\mathrm{d}x \\ &= \langle\eta_j,\eta_k\rangle_{L^2}.\end{aligned}$$

□

Remark 3.5. We do not need to compute δ precisely - a rough estimate of how well-localized our ρ is suffices completely. Therefore, we can estimate δ by a Monte-Carlo quadrature using the Bohmian points q_j (the ones which are ρ-distributed):

$$\delta = \delta(\rho) = \kappa_\delta \int_{\mathbb{R}^d} \rho^{\frac{d+1}{d}}(x)\,\mathrm{d}x = \kappa_\delta \mathbb{E}_{\rho_t}\big(\rho^{\frac{1}{d}}\big) \approx \frac{\kappa_\delta}{N}\sum_{j=1}^{N} \rho\big(q_j\big)^{\frac{1}{d}}.$$

Thus, the basis functions and the density ρ_g flatten out whenever the wave function does and in such a way, that the approximation quality (in $L^2(\mathbb{R}^d, \mathbb{C})$-sense) is not reduced.

In many cases, it appears reasonable to adapt the extent of flattening out the density locally (see Section 2.7.1), which leads to the use of adapted convolutions. In this case, the new density ρ_g is defined by

$$\rho_g := \rho *_{\mu_\rho} g$$

with proper adaptation function $\mu_\rho \colon \mathbb{R}^d \to \mathbb{R}$ (for details, see Section 2.7). This is a generalization to the common notion of convolutions, since for $\mu_\rho(x) = \delta^{1/d}$ we arrive at the former definition of ρ_g.

Again, we will check the scale invariance of our new choice for ρ_g, following the strategy from Proposition 3.4:

Proposition 3.6. Under conditions (1) – (8), where condition (2) is replaced by

$$\rho_g = \rho *_{\mu_\rho} g \tag{2*}$$

the relations (3.1.2) are fulfilled, if $\beta = \frac{2}{d}$ and

$$\mu_{\alpha^d \rho(\alpha \bullet)} = \alpha \mu_\rho(\alpha x) \tag{3.1.3}$$

(compare with the adaptation conditions 2.48). In this case we also have

$$\langle \tilde{\eta}_j, \tilde{\eta}_k \rangle_{L^2} = \langle \eta_j, \eta_k \rangle_{L^2} .$$

Proof. The proof is identical to the one of Proposition 3.4 except for the scaling of ρ_g:

$$\begin{aligned}
\tilde{\rho}_g(q) &= (\tilde{\rho} *_{\mu_{\tilde{\rho}}} g)(q) = \int_{\mathbb{R}^d} \tilde{\rho}(x)\, \mu_{\tilde{\rho}}(x)^d\, g\big(\mu_{\tilde{\rho}}(x)(q-x)\big)\, \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \alpha^d \rho(\alpha x)\, \alpha^d \mu_\rho(\alpha x)^d\, g\big(\alpha \mu_\rho(\alpha x)(q-x)\big)\, \mathrm{d}x \\
&\overset{y=\alpha x}{=} \alpha^d \int_{\mathbb{R}^d} \rho(y)\, \mu_\rho(y)^d\, g\big(\mu_\rho(y)(\alpha q - y)\big)\, \mathrm{d}y \\
&= \alpha^d \rho_g(\alpha q).
\end{aligned}$$

□

Remark 3.7. The most difficult part after defining a proper density $\rho_{g,t}$ and choosing $\rho_{g,0}$-distributed points $x_j = x_{j,0}$ at time $t = 0$ is to find a proper way to propagate them, such that they stay $\rho_{g,t}$-distributed for all times $t > 0$. This issue is elaborated in Section 2.7.4. We were surprised to find that it could be performed explicitly in both cases, $\rho_{g,t} = \rho_t * g_{\delta_t}$ and $\rho_{g,t} = \rho_t *_{\mu_t} g$.

3.1.5 Step 5: Generalization to Higher Dimensions

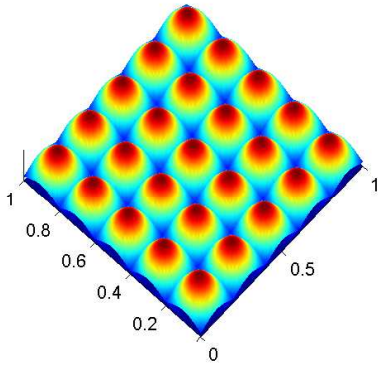
When choosing the approximation space in dimension $d > 1$, things get a little more subtle. We now have the freedom to break the radial symmetry of η_j by replacing the term $\epsilon_j \|x - x_j\|^2$ in the exponent by $(x-x_j)^T \Sigma_j^{-1} (x-x_j)$, where $\Sigma_j \in \mathbb{R}^{d\times d}$ are positive definite symmetric matrices (covariance matrices). Consider the following example:

Example 3.8. *For $\alpha > 0$, let $\Omega = (0,\alpha) \times (0,\alpha^{-1})$ be a rectangle equipped with the uniform distribution $\mathbb{P}_\rho$, $\rho \equiv 1$. The map*

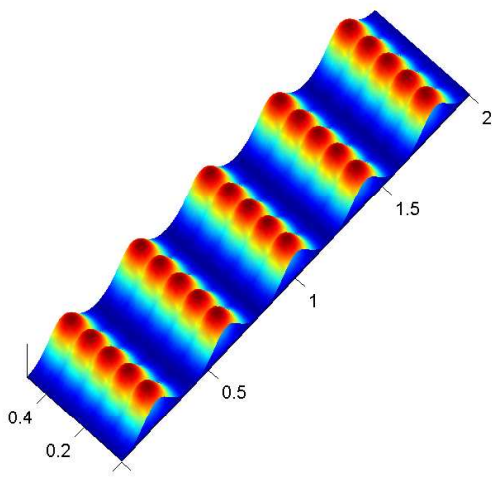
$$R_\alpha : \Omega \to (0,1)^2, \qquad x \mapsto \begin{pmatrix} \alpha^{-1} & 0 \\ 0 & \alpha \end{pmatrix} x$$

is a transport map from $\mathbb{P}_\rho$ to $\mathbb{P}_{uni}$, since $|D_x R_\alpha| \equiv 1$ (see Lemma 2.8). We choose $\mathbb{P}_\rho$-distributed centers of the basis functions by $x_j = {R_\alpha}^{-1}(y_j)$ with equidistant points $y_1, \dots, y_N \in (0,1)^d$ (see (2.1.1)). We will discuss two choices of basis functions: the first radially symmetric with parameter $\epsilon_j = h^{-2}\mathcal{E}^2 \rho(x_j)^{2/d}$ and the second with covariance matrices $\Sigma_j = \left(2h^{-2}\mathcal{E}^2 D_x R_\alpha(x_j)^\intercal D_x R_\alpha(x_j)\right)^{-1}$, i.e. (see Figure 3.3)

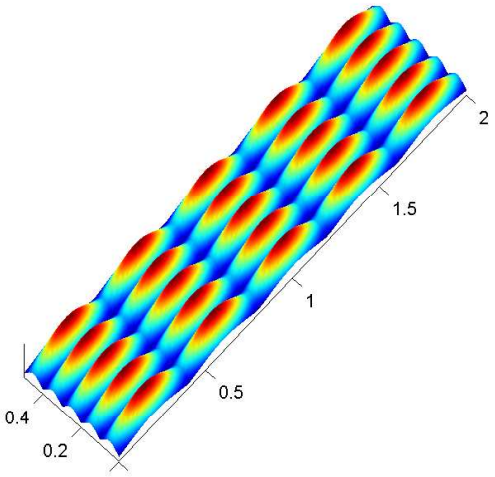
$$\eta_j^\alpha(x) = \left(\frac{2\epsilon_j}{\pi}\right)^{\frac{d}{4}} \exp\left(-\epsilon_j |x - x_j|^2\right) \quad \text{and}$$
$$\tilde{\eta}_j^\alpha(x) = \left(\frac{2}{\pi}\right)^{\frac{d}{4}} \left|h^{-1}\mathcal{E}\, D_x R_\alpha(x_j)\right|^{\frac{1}{2}} \exp\left(-\left|h^{-1}\mathcal{E}\, D_x R_\alpha(x_j)(x - x_j)\right|^2\right).$$



(a) For $\alpha = 1$ both cases coincide.



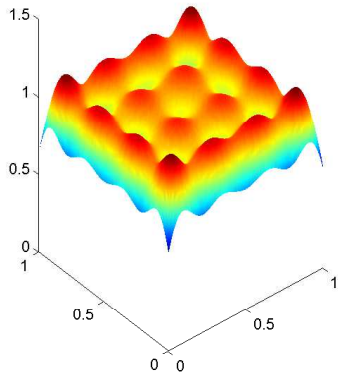
(b) For $\alpha > 1$ the radial basis functions η_j^α are too close too their neighbors in y-direction, raising the condition number of the stiffness matrix, see subsection 2.8.1, and too far in x-direction, worsening the approximation properties.



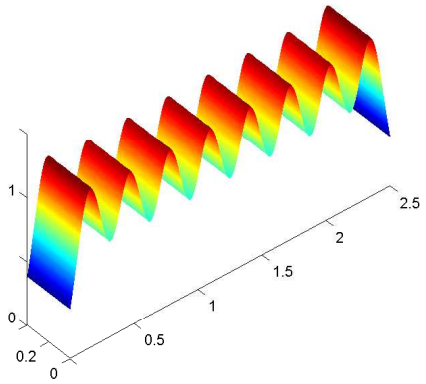
(c) These disadvantages can be compensated by scaling the basis functions – making them wider in x-direction and thinner in y-direction. In the case of the choice $\tilde{\eta}_j^\alpha$ both the stiffness matrix and the approximation properties are independent of α.

Figure 3.3: Comparison of the two choices of basis functions (η_j^α and $\tilde{\eta}_j^\alpha$) for $\alpha = 1, 2$ (here $N = 25,\ \mathcal{E} = 2$).

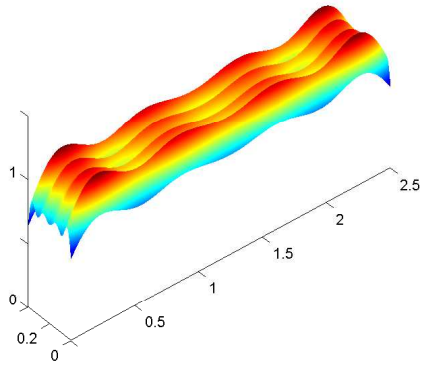
For the readers who still are not convinced, let us extend our example and compare the bestapproximation (in L^2-sense) of the function $f \equiv 1$ in $M = \operatorname{span}\{\eta_j^\alpha \mid j = 1, \ldots, N\}$ with the one in $\tilde{M} = \operatorname{span}\{\tilde{\eta}_j^\alpha \mid j = 1, \ldots, N\}$. The result for several α is illustrated in Figure 3.4.



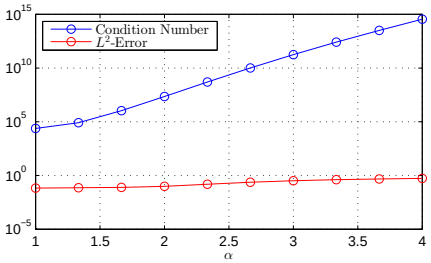
(a) For $\alpha = 1$ both cases coincide: $\kappa(A) = 2.4 \cdot 10^4,\ E = 0.067$.



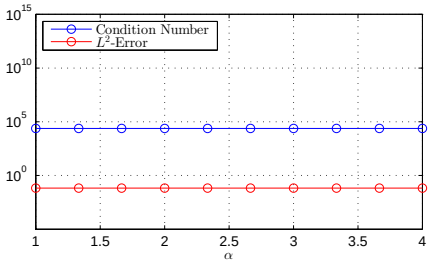
(b) For $\alpha = 2.5$ the radial basis functions η_j^α yield a bad approximation and an ill-conditioned stiffness matrix: $\kappa(A) = 2.3 \cdot 10^4,\ E = 0.1908$.



(c) For $\alpha = 2.5$ the condition number and the approximation properties of the basis functions $\tilde{\eta}_j^\alpha$ do not differ from the case $\alpha = 1$: $\kappa(A) = 2.4 \cdot 10^4,\ E = 0.067$.



(d) Condition number and L^2-Error of the bestappoximatoin for the basis functions η_j^α plotted over α.



(e) Condition Number and L^2-Error of the bestappoximatoin for the basis functions $\tilde{\eta}_j^\alpha$ plotted over α.

Figure 3.4: Comparison of the approximation errors E and the condition numbers of the stiffness matrices $\kappa(A)$ of the two choices of basis functions (η_j^α and $\tilde{\eta}_j^\alpha$) for $\alpha \in [1, 4]$. Here, $N = 8 \times 8 = 64$, $\mathcal{E} = 1.25$ and the function to be approximated is $f \equiv 1$ (note that the highest possible error of a bestapproximation is $\|f\|_{L^2} = 1$).

So, scaling $(x - x_j)$ by $D_xR(x_j)$ deforms the basis functions in a proper way. The theoretical reason why this is the "correct" way to scale, is given in all detail in Section 3.3 and will now be discussed roughly:

Let us denote the radial basis functions on the unit cube (equipped with the uniform distribution) and centered on a uniform grid $y_1, \dots, y_N$ by $\theta_1, \dots, \theta_N$ (in the upper example, $\theta_j = \eta_j^1$):

$$\theta_j(y) = \phi(\|y - y_j\|) \quad \text{for some function} \quad \phi : \mathbb{R}_{\geq 0} \to \mathbb{R}.$$

$R^{-1} : (0,1)^d \to \Omega$ is a deformation of the support of the density and we want not only our centers to be transformed by it ($x_j = R^{-1}(y_j)$), but also the basis functions θ_j:

$$\eta_j \approx \theta_j \circ R : \Omega \to \mathbb{R}.$$

Since the Jacobian $D_xR(x_j)$ is the best linear approximation of R near the point x_j, this results in:

$$(\theta_j \circ R)(x) \approx \theta_j\big(\underbrace{R(x_j)}_{=y_j} + D_xR(x_j)(x - x_j)\big) = \phi\big(\|D_xR(x_j)(x - x_j)\|\big) =: \eta_j(x).$$

Example 3.9. *As a final example consider R given by*

$$R^{-1}\Big((r,\varphi)\Big) = (r+1)\begin{pmatrix}\cos(\pi\varphi)\\ \sin(\pi\varphi)\end{pmatrix} \quad , \quad (r,\varphi) \in (0,1)^2.$$

In Figures 3.5 and 3.6 one can observe how R^{-1} not only transforms the centers, but also deforms the basis functions in a proper way.

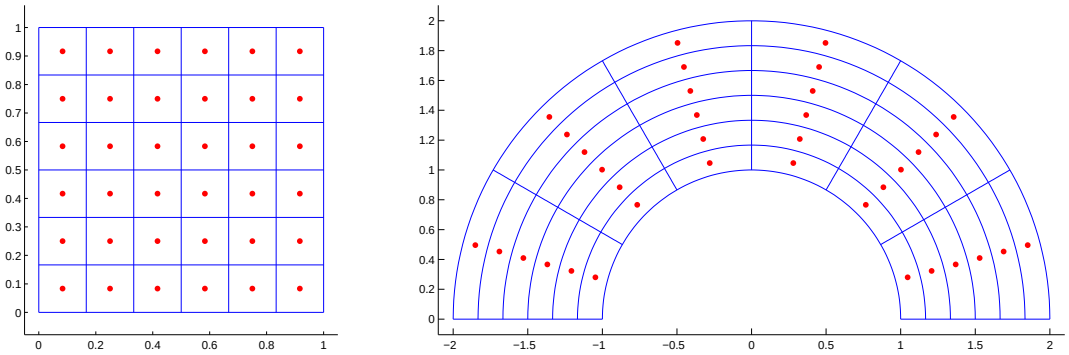


Figure 3.5: Visualisation of the transport map R^{-1} and the centers of the basis functions (in red).

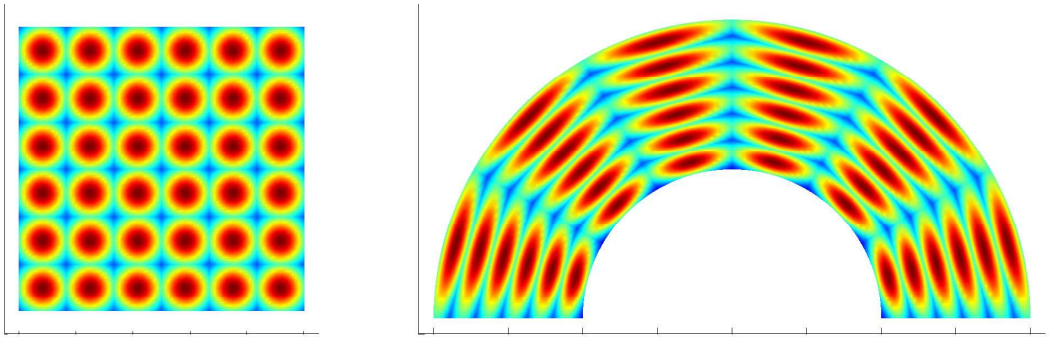


Figure 3.6: In order to keep good approximation properties without getting an ill-conditioned matrix the basis functions are deformed by $D_x R$.

3.2 Resulting Algorithm

In this section, we will sum up the the ideas from the previous section to get a better overview. Afterwards, we will write down the complete algorithm in bullet point form.

Let $1 \leq p < \infty$, $\psi = \psi_t \in L^p \cap C^\infty(\mathbb{R}^d, \mathbb{K}) \setminus \{0\}$ be the solution of (1.1) (and the function we want to approximate) and $\rho = \rho_t = \frac{|\psi_t|^p}{\|\psi_t\|_{L^p}^p}$ the corresponding probability density at a fixed time $t \in \mathbb{R}$. We define the Gaussian densities $g(x) = \pi^{-d/2} e^{-|x|^2}$, $g_\delta(x) = \delta^d g(\delta x)$ (for $\delta > 0$) and

$$\rho_g = \rho * g_\delta \qquad \text{or} \qquad \rho_g = \rho *_{\mu_\rho} g,$$

with proper scaling parameter $\delta > 0$ or adaptation function $\mu_\rho \in C^2\left(\mathbb{R}^d, \mathbb{R}_{>0}\right)$. Both choices are legitimate. The first is simpler to implement, while the second one is more general, see Section 2.7. By Proposition 2.17, there exists a transport map $R_g \in C^\infty\left(\mathbb{R}^d, (0,1)^d\right)$ from $\mathbb{P}_{\rho_g}$ to $\mathbb{P}_{\text{uni}}$ and by Proposition 2.7, the points $x_j = R_g^{-1}(y_j)$ are $\mathbb{P}_{\rho_g}$-distributed. Here, $y_1, \ldots, y_N$ are equidistant points in $(0,1)^d$ defined by (2.1.1) and $N = n^d$ for some $n \in \mathbb{N}$.

The approximation space $M = \operatorname{span}\{\eta_1, \ldots, \eta_N\} \subseteq L^2(\mathbb{R}^d)$ is now spanned by the basis functions

$$\boxed{\eta_j(x) = \left(\frac{2}{\pi}\right)^{\frac{d}{4}} |J_j|^{\frac{1}{2}} \exp\left(-|J_j(x - x_j)|^2\right),} \tag{3.2.1}$$

where $J_j := h^{-1} \mathcal{E} D_x R_g(x_j)$, $h = n^{-1}$ is the mesh size and $\mathcal{E} > 0$ is the basic shape

parameter. In dimension $d = 1$ this formula reduces to

$$\eta_j(x) = \left(\frac{2\epsilon_j}{\pi}\right)^{\frac{d}{4}} \exp\left(-\epsilon_j |x - x_j|^2\right) \tag{3.2.2}$$

with parameter $\epsilon_j = h^{-2}\mathcal{E}^2\rho_g(x_j)^2$. In the special case of the Schrödinger equation the approximation space can be "improved" by adding a momentum term:

$$\eta_j(x) = \left(\frac{2}{\pi}\right)^{\frac{d}{4}} |J_j|^{\frac{1}{2}} \exp\left(-|J_j(x - x_j)|^2 + ip_j(x - x_j)\right),$$

where $p_j = v_t(x_j) = \Im\left[\frac{\nabla\psi}{\psi}\right]$. In our numerical examples, we will always treat both cases: with and without momentum term (see Section 4).

For the time evolution of the approximation space we need to examine the propagation of $x_{j,t}$ and $J_{j,t}$ (the differential equations for the other parameters are straightforward to compute and are listed below):

By Proposition 2.12, the flow $\Phi_{g,t}$ of the dynamical system

$$\dot{x}_t = v_{g,t}(x_t), \qquad v_{g,t} := \frac{j_{g,t}}{\rho_{g,t}},$$

where $j_{g,t}$ is defined in Propositions 2.53, 2.54 respectively, defines a transport map from $\mathbb{P}_{\rho_{g,0}}$ to $\mathbb{P}_{\rho_{g,t}}$. Therefore, the points $x_{j,t} = \Phi_{g,t}(x_{j,0}) = \left(\Phi_{g,t} \circ R_{g,0}^{-1}\right)(y_j)$ propagated this way stay $\rho_{g,t}$-distributed for all $t \in \mathbb{R}$ by Proposition 2.7 (see also Corollary 2.55).

Let $Y_{g,t} := R_{g,t}^{-1} = \Phi_{g,t} \circ R_{g,0}^{-1}$, then the propagation of its Jacobian $D_y Y_{g,t}$ is described by Corollary 2.15:

$$\partial_t D_y Y_{g,t}(y) = D_x v_t\left(Y_{g,t}(y)\right) \cdot D_y Y_{g,t}.$$

Hence, its (scaled) inverse $J_{j,t} = h^{-1}\mathcal{E}\, D_x R_{g,t}(x_{j,t})$ evolves via

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} J_{j,t} &= -J_{j,t}\Big[\frac{\mathrm{d}}{\mathrm{d}t} \underbrace{J_{j,t}^{-1}}_{=h\mathcal{E}^{-1} D_y Y_{g,t}(y_j)}\Big] J_{j,t} \\
&= -h\mathcal{E}^{-1} J_{j,t} \cdot D_x v_t\left(Y_{g,t}(y_j)\right) \cdot D_y Y_{g,t}(y_j) \cdot J_{j,t} \\
&= -J_{j,t} \cdot D_x v_t\left(x_{j,t}\right)
\end{aligned} \tag{3.2.3}$$

Note that, since we will choose g to be a Gaussian, $\rho_{g,t}$, $j_{g,t}$, $v_{g,t}$, $\Phi_{g,t} \in C^\infty$ for all $t \in \mathbb{R}$ and therefore the differentiability properties of Y_0 are inherited by Y_t.

After discussing the main points for the propagation of the approximation space and how the propositions and theorems from Chapter 2 are applied, let us write down our algorithm for numerical approximation of an evolutionary PDE with underlying continuity equation (see Definition 1.1) in every detail:

Initialization:

- Define $n \in \mathbb{N}$, set $N = n^d,\ h = n^{-1}$ and define equidistant points $y_1, \dots, y_N$ in $(0,1)^d$ by (2.1.1).
- Choose suitable constants $\mathcal{E}, \kappa_\delta > 0$ and define for $x \in \mathbb{R}^d,\ \delta > 0$

$$g(x) = \pi^{-d/2}\, e^{-|x|^2}, \qquad g_\delta(x) := \delta^d g(\delta x), \qquad \gamma_\delta(x) := \frac{x}{\delta}\, g_\delta(x) = \delta^{d-1} x\, g(\delta x).$$

- Approximate ρ_0 by a weighted sum of normal distributions and construct $\mathbb{P}_{\rho_0}$-distributed points $q_{1,0}, \dots, q_{N,0}$ by the algorithm described in Section 2.3.
- Compute $\delta_0 = \frac{\kappa_\delta}{N} \sum_{j=1}^N \rho_0\big(q_{j,0}\big)^{\frac{1}{d}}$.
- Compute $\rho_{g,0} = \rho_0 * g_{\delta_0}$ numerically.
- Approximate $\rho_{g,0}$ by a weighted sum of normal distributions and construct a transport map $R_{g,0} \in C^\infty\big(\mathbb{R}^d, (0,1)^d\big)$ from $\mathbb{P}_{\rho_{g,0}}$ to $\mathbb{P}_{\text{uni}}$ and $\mathbb{P}_{\rho_{g,0}}$-distributed points $x_{1,0}, \dots, x_{N,0}$ by the algorithm described in Section 2.3. More precisely, it suffices to construct $x_{j,0} = R_{g,0}^{-1}(y_j)$ as in Section 2.3 and $D_y(R_{g,0}^{-1})(y_j)$ by using Proposition 2.24 instead of the whole transport map.
- Compute $p_{j,0} = v_0(x_{j,0})$ and $J_{j,0} := h^{-1}\mathcal{E}\, D_x R_{g,0}(x_{j,0}) = h^{-1}\mathcal{E}\, \Big[D_y(R_{g,0}^{-1})(y_j)\Big]^{-1}$. This defines the approximation space $M_0 = \operatorname{span}\{\eta_{1,0}, \dots, \eta_{N,0}\}$ by equation (3.2.1).
- Compute the bestapproximation $u_0 = \sum_{j=1}^N c_{j,0}\eta_{j,0}$ of ψ_0 in M_0.

Time Evolution:

Since the points $q_{j,t}$ are ρ_t-distributed, we may approximate

$$
\begin{aligned}
\rho_{g,t}(x) &= \int_{\mathbb{R}^d} \rho_t(q)\, g_{\delta_t}(x-q)\,\mathrm{d}q \approx \frac{1}{N}\sum_{j=1}^{N} g_{\delta_t}(x-q_{j,t}),\\
\nabla\rho_{g,t}(x) &= \int_{\mathbb{R}^d} \rho_t(q)\, \nabla g_{\delta_t}(x-q)\,\mathrm{d}q \approx \frac{1}{N}\sum_{j=1}^{N} \nabla g_{\delta_t}(x-q_{j,t}),\\
j_{g,t}(x) &= \int_{\mathbb{R}^d} j_t(q)\, g_{\delta_t}(x-q) - \dot\delta_t\, \rho_t(q)\, \gamma_{\delta_t}(x-q)\,\mathrm{d}q\\
&= \int_{\mathbb{R}^d} \Big[v_t(q) - \frac{\dot\delta_t}{\delta_t}(x-q)\Big]\rho_t(q)\, g_{\delta_t}(x-q)\,\mathrm{d}q\\
&\approx \frac{1}{N}\sum_{j=1}^{N} \Big[\underbrace{v_t(q_{j,t})}_{\dot q_{j,t}} - \frac{\dot\delta_t}{\delta_t}(x-q_{j,t})\Big] g_{\delta_t}(x-q_{j,t}),\\
D_x j_{g,t}(x) &= \int_{\mathbb{R}^d} \Big[v_t(q) - \frac{\dot\delta_t}{\delta_t}(x-q)\Big]\rho_t(q)\, \underbrace{\nabla g_{\delta_t}(x-q)^\intercal}_{-2\delta_t^2(x-q)^\intercal g_{\delta_t}(x-q)} - \frac{\dot\delta_t}{\delta_t}\rho_t(q)\, g_{\delta_t}(x-q)\,\mathrm{Id}\,\mathrm{d}q\\
&= \int_{\mathbb{R}^d} \Big[-2\delta_t^2\Big(v_t(q) - \frac{\dot\delta_t}{\delta_t}(x-q)\Big)(x-q)^\intercal - \frac{\dot\delta_t}{\delta_t}\,\mathrm{Id}\Big]\rho_t(q)\, g_{\delta_t}(x-q)\,\mathrm{d}q\\
&\approx \frac{1}{N}\sum_{j=1}^{N} \Big[-2\delta_t^2\Big(\dot q_{j,t} - \frac{\dot\delta_t}{\delta_t}(x-q_{j,t})\Big)(x-q_{j,t})^\intercal - \frac{\dot\delta_t}{\delta_t}\,\mathrm{Id}\Big] g_{\delta_t}(x-q_{j,t}).
\end{aligned}
$$

Using these formulas and

$$
v_{g,t}(x) = \frac{j_{g,t}}{\rho_{g,t}}(x)\,, \qquad D_x v_{g,t}(x) = \rho_{g,t}(x)^{-2}\left(D_x j_{g,t}\,\rho_{g,t} - j_{g,t}\nabla\rho_{g,t}^\intercal\right)(x),
$$

we can compute:

- $\dot q_{j,t} = v_t(q_{j,t})$,
- $\dot\delta_t = \dfrac{\kappa_\delta}{dN}\displaystyle\sum_{j=1}^{N} \rho_t\big(q_{j,t}\big)^{\frac{1}{d}-1}\big(\partial_t\rho_t\big(q_{j,t}\big) + \nabla\rho_t(q_{j,t})^\intercal\,\dot q_{j,t}\big)$,
- $\dot x_{j,t} = v_{g,t}(x_{j,t})$,
- $\dot p_{j,t} = \partial_t v_t(x_{j,t}) + D_x v_t(x_{j,t})\,\dot x_{j,t}$,
- $\dot J_{j,t} = -J_{j,t}\cdot D_x v_{g,t}(x_{j,t})$ by equation (3.2.3),
- $\dot c_{j,t}$ is computed via a Galerkin approximation, which results in the Dirac Frenkel variational principle in the case of the Schrödinger equation, see Section 2.5.2.

3.3 Rigorous Approximation Theory

We are finally ready to analyze the approximation error of the manifold constructed in Section 3.1 (as mentioned in Remark 3.1, the momentum term $ip_k(x - x_k)$ in the exponent of the basis functions is tailored particularly to the Schrödinger Equation and will not be considered here). Throughout this section we will use the following notation.

Notation 3.10. *Our ansatz is to approximate a continuous function $\psi \in W^{k,p}(\mathbb{R}^d, \mathbb{R})$, $1 \leq p < \infty$, with $k > d/p$ by the following subspace $M \subseteq L^p(\mathbb{R}^d, \mathbb{R})$:*

$$M = \operatorname{span}\left\{\eta_j \in L^p(\mathbb{R}^d, \mathbb{R}) \mid j = 1, \dots, N\right\} \quad , \quad \eta_j(x) = \eta\left(\frac{\mathcal{E}}{h} D_x R(x_j)(x - x_j)\right),$$

where $N = n^d$ for some odd $n \in 2\mathbb{N} + 1$, $h = 1/n$ is the mesh size, $\mathcal{E} > 0$ is the basic shape parameter from Definition 2.61 and

(1) $\eta \in C^1(\mathbb{R}^d, \mathbb{R})$ has bounded support $supp(\eta) \subseteq [-K, K]^d$ for some $K > 0$ (it therefore automatically fulfills the extended decay condition (2.8.4) with arbitrary large constant $K_2 > d$) and fulfills the moment condition (2.8.5) with constant $K_3 \geq 1$,

Remark 3.11. Strictly speaking, the Gaussian basis functions we consider do not have compact support and would have to be "cut off" outside a properly large region $[-K, K]^d$.

(2) $R\colon \mathbb{R}^d \to (0,1)^d$ is a C^k-diffeomorphism and a transport map from $\mathbb{P}_\rho$ to $\mathbb{P}_{uni}$,

(3) $\rho \in L^1(\mathbb{R}^d, \mathbb{R}_{>0})$ is a suitably chosen continuous and positive probability density function, usually

$$\rho = \frac{|\psi|^p}{\|\psi\|_{L^p}^p} * g_\delta,$$

$g_\delta(x) = \delta^d g(\delta x)$ being a properly scaled Gaussian probability density function (we omitted the index g to simplify the notation).

(4) the "centers" $x_j \in \mathbb{R}^d$ are given by $x_j = R^{-1}(y_j)$, where $y_1, \dots, y_N$ are equidistant points in $(0,1)^d$ defined by (2.1.1).

In order to analyze the approximation properties of ψ in M we will proceed in three steps (see figure 3.7):

Step 1: *We "push" the whole approximation problem from $L^p(\mathbb{R}^d)$ to $L^p_\omega((0,1)^d)$, $\omega(y) = 1/\rho(R^{-1}(y))$ for $y \in (0,1)^d$, via R, where L^p_ω denotes the weighted L^p space defined below. I.e. we consider*

$$\psi_\rho := \psi \circ R^{-1}, \quad \theta_j := \eta_j \circ R^{-1} \quad \text{and} \quad M_\rho := \operatorname{span}\{\theta_j \mid j = 1, \dots, N\},$$

Note that $\psi_\rho \in W^{k,p}\left((0,1)^d\right)$ and $\theta_j \in C^1\left((0,1)^d\right)$, since R is a C^k-diffeomorphism.

Step 2: *We understand the connection between the approximation errors of*

$$\psi \approx u = \sum_{j=1}^N c_j \eta_j \in M \qquad \text{and} \qquad \psi_\rho \approx u_\rho := u \circ R^{-1} = \sum_{j=1}^N c_j \theta_j \in M_\rho\,.$$

Step 3: *We analyze the approximation properties of ψ_ρ in M_ρ using the approximate approximation theory presented in Section 2.8.2.*

Definition 3.12 (weighted L^p norm and space)**.** Let $\Omega \subseteq \mathbb{R}^d$ be a domain and $\omega : \Omega \to \mathbb{R}_{>0}$ be a positive Lebesgue-measurable function. We define the *weighted L^p norm* of a function $g : \Omega \to \mathbb{R}$ and the *weighted L^p space* by

$$\begin{aligned} \|g\|_{L^p_\omega(\Omega)} &:= \left(\int_\Omega |g(x)|^p\, \omega(x)\, \mathrm{d}x\right)^{1/p} \\ \text{and} \qquad L^p_\omega(\Omega) := L^p_\omega(\Omega, \mathbb{R}) &:= \left\{g : \Omega \to \mathbb{R} \mid \|g\|_{L^p_\omega(\Omega)} < \infty\right\} \Big/ \mathcal{N}, \\ \text{where} \qquad \mathcal{N} &:= \left\{g : \Omega \to \mathbb{R} \mid \|g\|_{L^p_\omega(\Omega)} = 0\right\}. \end{aligned}$$

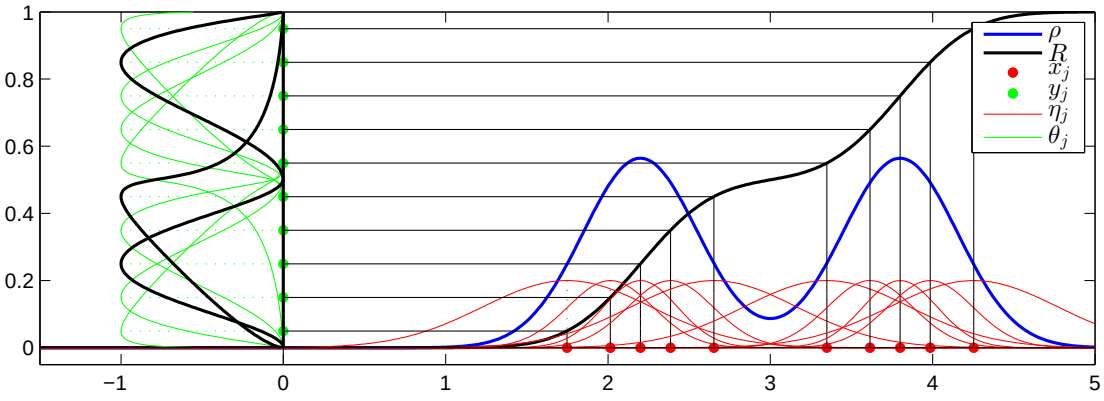
We will use the abbreviation L^p_ω, if it is clear, which domain Ω is considered.

Proposition 3.13. Let $\rho \in L^1(\mathbb{R}^d, \mathbb{R}_{>0})$, $R \in C^k\left(\mathbb{R}^d, (0,1)^d\right)$ and the weight function $\omega(y) = 1/\rho(R^{-1}(y))$ fulfill the assumptions of Notation 3.10, $A \subseteq \mathbb{R}^d$ be a domain and $\Omega := R(A)$. Then we have for $g \in L^p(A)$:

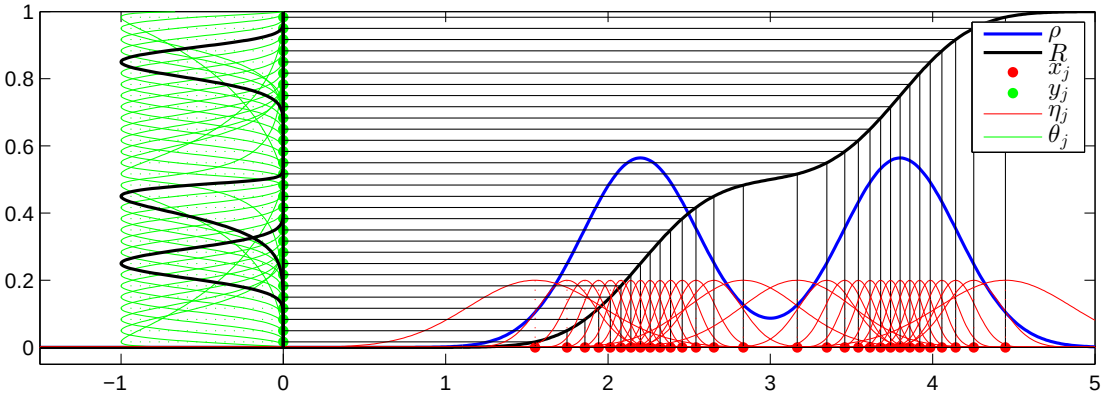
$$g_\rho := g \circ R^{-1} \in L^p_\omega(\Omega) \qquad \text{and} \qquad \|g\|_{L^p(A)} = \|g_\rho\|_{L^p_\omega(\Omega)}.$$

Further, if $g \in L^p_\rho(A)$ then

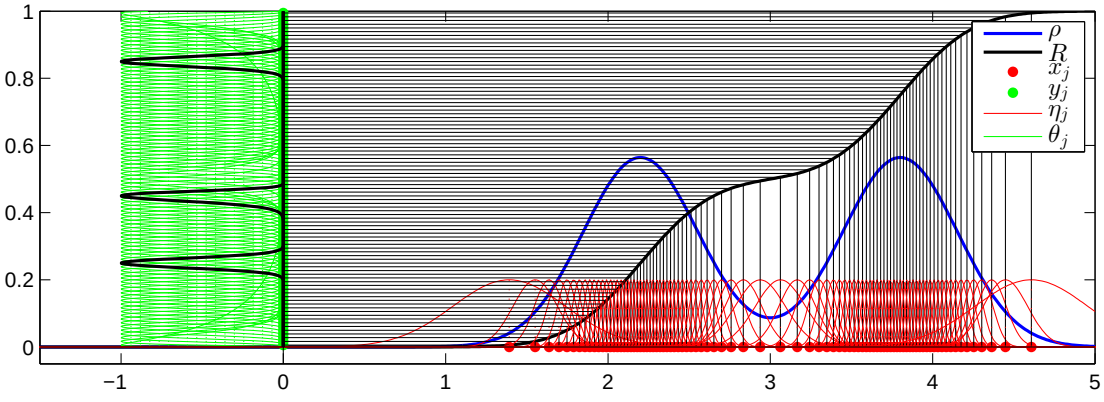
$$g_\rho \in L^p(\Omega) \qquad \text{and} \qquad \|g\|_{L^p_\rho(A)} = \|g_\rho\|_{L^p(\Omega)}\,.$$



(a) $N = 10$.



(b) $N = 30$.



(c) $N = 90$.

Figure 3.7: The basis functions η_j of M and θ_j of M_ρ for Gaussian η and $N = 10, 30, 90$. One can observe the assimilation of the three functions θ_j highlighted in black. Those θ_j closest to the boundary do not share this behavior, since their centers move closer and closer to the edges.

Proof. Using Lemma 2.8 and the transformation formula we get:

$$\begin{aligned}\|g\|_{L^p(A)}^p &= \int_A |g(x)|^p \,\mathrm{d}x = \int_A |g_\rho \circ R|^p(x)\,\mathrm{d}x = \int_A \left[\frac{|g_\rho|^p}{|D_x R| \circ R^{-1}} \circ R\right](x)\,|D_x R(x)|\,\mathrm{d}x \\ &= \int_\Omega \frac{|g_\rho|^p(y)}{\rho(R^{-1}(y))}\,\mathrm{d}y = \|g_\rho\|_{L^p_\omega(\Omega)}^p, \\ \|g\|_{L^p_\rho(A)}^p &= \int_A |g(x)|^p\,\rho(x)\,\mathrm{d}x = \int_A |g_\rho \circ R|^p(x)\,|D_x R(x)|\,\mathrm{d}x = \int_\Omega |g_\rho|^p(y)\,\mathrm{d}y \\ &= \|g_\rho\|_{L^p(\Omega)}^p.\end{aligned}$$

□

Corollary 3.14. Using Notation 3.10, Proposition 3.13 implies for $\psi, u \in L^p(\mathbb{R}^d, \mathbb{R})$ and $\psi_\rho = \psi \circ R^{-1}$, $u_\rho = u \circ R^{-1}$:

$$\|\psi - u\|_{L^p} = \|\psi_\rho - u_\rho\|_{L^p_\omega}.$$

To get the approximation properties of M_ρ under control, we will use the the theory of approximate approximations presented in Section 2.8.2. This relies on the fact that, while each basis function η_j is scaled in a different way, the functions θ_j become shifted copies of the same function for large N, namely

$$\theta_j(y) \approx \mathcal{E}^d\,\eta\left(\frac{\mathcal{E}}{h}(y - y_j)\right).$$

Proposition 3.15. Using Notation 3.10 (we allow the additional case $p = \infty$ here), let $x_* \in \mathbb{R}^d$, $y_* = R(x_*)$ and $\theta_* = \eta_* \circ R^{-1}$, where $\eta_*(x) := \mathcal{E}^d\,\eta\big(\frac{\mathcal{E}}{h} D_x R(x_*)(x - x_*)\big)$. Then there exists a constant

$$C = C\left(p, \mathcal{E}, K, \|D_x R(x_*)\|, \|D^2 R^{-1}(y_*)\|, \max_{y\in[-K,K]^d} \|\nabla\eta(y)\|_2, \right),$$

such that for all sufficiently small $h > 0$

$$\begin{cases} \left\|\theta_* - \mathcal{E}^d\,\eta\left(\frac{\mathcal{E}}{h}(\bullet - y_*)\right)\right\|_{L^p} \le C h^{1+\frac{d}{p}}, & \text{if } p < \infty, \\ \left\|\theta_* - \mathcal{E}^d\,\eta\left(\frac{\mathcal{E}}{h}(\bullet - y_*)\right)\right\|_{L^p} \le C h, & \text{if } p = \infty. \end{cases}$$

For $x_* \in A$, where $A \subseteq \mathbb{R}^d$ is a compact set, the constant

$$C = C\left(p, \mathcal{E}, K, \max_{y\in[-K,K]^d} \|\nabla\eta(y)\|_2, A\right)$$

can be chosen uniform with respect to x_*, y_* respectively.

Remark 3.16. The prefactor $h^{\frac{d}{p}}$ appears due to the fact that for $p < \infty$ and $g \in L^p(\mathbb{R}^d)$:

$$\left\|g\left(h^{-1}\cdot\bullet\right)\right\|_{L^p} = \left(\int_{\mathbb{R}^d}\left|g\left(h^{-1}x\right)\right|^p \,\mathrm{d}x\right)^{\frac{1}{p}} = \left(h^d\int_{\mathbb{R}^d}|g(x)|^p\,\mathrm{d}x\right)^{\frac{1}{p}} = h^{\frac{d}{p}}\|g\|_{L^p}$$

and therefore both $\|\theta_*\|_{L^p} \sim h^{\frac{d}{p}}$ and $\left\|\eta\left(\frac{\mathcal{E}}{h}(\bullet - y_*)\right)\right\|_{L^p} \sim h^{\frac{d}{p}}$.

Proof of Proposition 3.15. Let $B := (D_xR)^{-1}(x_*)\cdot[-K,K]^d$. Since $\operatorname{supp}(\eta) \subseteq [-K,K]^d$ and $D^2R(x) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a bilinear map for each $x \in \mathbb{R}^d$, we obtain the following supports for η_* and θ_* and for sufficiently small $h > 0$:

$$\begin{aligned}
\operatorname{supp}(\eta_*) &\subseteq x_* + \frac{h}{\mathcal{E}}B,\\
\operatorname{supp}(\theta_*) &\subseteq R\left(x_* + \frac{h}{\mathcal{E}}B\right)\\
&\subseteq R(x_*) + \frac{h}{\mathcal{E}}D_xR(x_*)B + \frac{h^2}{\mathcal{E}^2}\Bigg[\underbrace{D^2R\left(x_* + \frac{h}{\mathcal{E}}B\right)}_{\text{bounded on the region of evaluation}}\Bigg](B,B)\\
&\subseteq y_* + \frac{h}{\mathcal{E}}[-(K+1),K+1]^d.
\end{aligned}$$

From now on, let $y \in [-(K+1),K+1]^d$ and $h > 0$ sufficiently small. Viewing $D^2R^{-1}(w) : (0,1)^d \times (0,1)^d \to \mathbb{R}^d$ as a bilinear map for each $w \in (0,1)^d$ and Taylor expanding R^{-1} at y_*, we get for some $\zeta\colon \mathbb{R}^d \to (0,1)$

$$\begin{aligned}
&R^{-1}\left(y_* + h\mathcal{E}^{-1}y\right)\\
&= R^{-1}(y_*) + h\mathcal{E}^{-1}D_y(R^{-1})(y_*)y + h^2\mathcal{E}^{-2}\left(D^2R^{-1}(y_* + h\mathcal{E}^{-1}\zeta(h\mathcal{E}^{-1}y)y)\right)(y,y)\\
&= x_* + h\mathcal{E}^{-1}D_xR(x_*)^{-1}(y + hz_h(y)),
\end{aligned}$$

where we used that $D_y(R^{-1})(y_*) = D_xR(x_*)^{-1}$ and

$$z_h(y) := \mathcal{E}^{-1}J_R(x_*)\left[D^2R^{-1}(y_* + h\zeta(h\mathcal{E}^{-1}y)y)\right](y,y).$$

Since $y \in [-(K+1),K+1]^d$ and $R^{-1} \in C^2$, the function z_h is bounded uniformly with respect to y and h:

$$\exists L = L(\|D^2R^{-1}(y_*)\|) > 0 : \quad z_h(y) \in [-L,L]^d \ \text{ for all } \ y \in [-(K+1),K+1]^d.$$

By Taylor expanding once more, this time η at y, we get for some $\tilde{\zeta}\colon \mathbb{R}^d \to (0,1)$

$$\begin{aligned}
\theta_*\left(y_* + \frac{h}{\mathcal{E}}y\right) &= \mathcal{E}^d\,\eta\left(\frac{\mathcal{E}}{h}D_xR(x_*)\left[R^{-1}\left(y_* + \frac{h}{\mathcal{E}}y\right) - x_*\right]\right)\\
&= \mathcal{E}^d\,\eta\left(y + hz_h(y)\right)\\
&= \mathcal{E}^d\,\eta(y) + \mathcal{E}^d\,hz_h(y)^T\nabla\eta\left(y + h\,\tilde{\zeta}(hz_h(y))\,z_h(y)\right).
\end{aligned}$$

Since $z_h(y) \in [-L, L]^d$ and $\nabla\eta\left(\mathcal{E}y + \mathcal{E}h\tilde{\zeta}(y)z(y)\right)$ is uniformly bounded with respect to $y \in [-(K+1), K+1]^d$ for sufficiently small $h > 0$, this yields uniform convergence of the form

$$\left\| \theta_* - \mathcal{E}^d \, \eta\left(\frac{\mathcal{E}}{h}(\bullet - y_*)\right) \right\|_\infty \leq Ch,$$

as well as for $p < \infty$,

$$\begin{aligned}
\left(h^{-\frac{d}{p}} \left\| \theta_* - \mathcal{E}^d \, \eta\left(\frac{\mathcal{E}}{h}(\bullet - y_*)\right) \right\|_{L^p}\right)^p &= h^{-d} \int_{\mathbb{R}^d} \left| \theta_*(y) - \mathcal{E}^d \, \eta\left(\frac{\mathcal{E}}{h}(y - y_*)\right) \right|^p \, \mathrm{d}y \\
&= \mathcal{E}^{-d} \int_{[-(K+1),K+1]^d} \left| \theta_*\left(y_* + \frac{h}{\mathcal{E}}y\right) - \mathcal{E}^d \, \eta\,(y) \right|^p \, \mathrm{d}y \\
&= \int_{[-(K+1),K+1]^d} \left| h z_h(y)^T \nabla\eta\left(y + h\,\tilde{\zeta}(h z_h(y))\, z_h(y)\right) \right|^p \, \mathrm{d}y \\
&\leq (2K+2)^d \left[\max_{y \in [-(K+1),K+1]^d} \|z_h(y)\|_2 \|\nabla\eta(y)\|_2 \right]^p h^p \\
&\leq (2K+2)^d \left[(L\sqrt{d}) \max_{y \in [-K,K]^d} \|\nabla\eta(y)\|_2 \right]^p h^p.
\end{aligned}$$

□

Armed with the knowledge that all basis functions θ_j are, roughly speaking, shifted copies of the same continuous function, which fulfills the extended decay and moment conditions (2.8.4) and (2.8.5), we can now apply the theory of approximate approximations presented in Section 2.8.2 to analyze the error of ψ_ρ in M_ρ.

Theorem 3.17. Using Notation 3.10, let $\psi \in W^{k,p}(\mathbb{R}^d, \mathbb{R})$, $1 \leq p < \infty$, and $\varepsilon > 0$ be given. Then there exist constants $\mathcal{E}_0, h_0 > 0$ such that for all basic shape parameters $0 < \mathcal{E} < \mathcal{E}_0$ and all mesh sizes $0 < h < h_0$ (i.e. for sufficiently large $n \in 2\mathbb{N} + 1$), the approximant

$$\widehat{\psi} \;:=\; \sum_{j=1}^{N} \psi\big|_{A^{(2)}}(x_j)\, \eta_j \quad \in M$$

fulfills

$$\|\psi - \widehat{\psi}\|_{L^p(\mathbb{R}^d)} < \varepsilon.$$

Remark 3.18. Before proving this theorem, we have to deal with two issues:

(1) *Boundary problems:* The convergence of the basis functions θ_j is not uniform in x_j (or y_j) since the constant C from Proposition 3.15 depends on x_*. Since the outer centers y_j converge to the boundaries of $(0,1)^d$ as N goes to infinity, this problem

can not be solved by just choosing N "large enough", as can also be seen in Figure 3.7.
Also the approximate approximation theory only gives error bounds in L^p, not in L^p_ω. Since $\omega(y) = 1/\rho(R(y))$ goes to infinity for y converging to the boundary, the two norms $\|\bullet\|_{L^p\left((0,1)^d,\mathbb{R}\right)}$ and $\|\bullet\|_{L^p_\omega\left((0,1)^d,\mathbb{R}\right)}$ are not equivalent.

(2) In order to apply the theory from Section 2.8.2, we need to bring the points $y_1, \dots, y_N$ into the form $y_j = mh$ for some $m \in \mathbb{Z}^d$.

Solving these will require further notation.

Notation 3.19. *We will deal with the first problem by "cutting off" the function ψ outside a sufficiently large ball:*
For given $\epsilon > 0$ choose $r_1 > r_2 > r_3 > r_4 > 0$ such that

$$A^{(j)} := \overline{B_{r_j}(0)} \qquad , \qquad \Omega^{(j)} := R(A_j) \qquad \text{and} \qquad \widehat{\omega} := \max_{y \in \Omega^{(1)}} \omega(y)$$

satisfy

(i) $\|\psi\|_{L^p\left(\mathbb{R}^d \setminus A_3\right)} < \dfrac{\varepsilon}{4}$,

(ii) $\|\psi_\rho\big|_{\left(\Omega^{(2)} \setminus \Omega^{(4)}\right)}\|_{p,h} < \dfrac{\varepsilon}{4\widehat{\omega}C}$ *for all sufficiently small $h > 0$ with $C > 0$ and $\|\bullet\|_{p,h}$ as in Lemma 2.62,*

(iii) $r_2 = \dfrac{r_1 + r_3}{2}$.

In order to solve the second problem, we shift the unit square $\Omega = (0,1)^d$ and all its subsets and functions defined on it via $\tau(x) := x - \frac{1}{2}\mathbb{1}$ to

$$\tilde{\Omega} := \tau(\Omega) = \left(-\frac{1}{2}, \frac{1}{2}\right)^d, \qquad \tilde{\Omega}^{(j)} := \tau\left(\Omega^{(j)}\right) \ \forall j, \qquad \tilde{y_j} = \tau\left(y_j\right) \ \forall j,$$
$$\tilde{\psi}_\rho := \psi_\rho \circ \tau^{-1}, \qquad \tilde{\theta}_j := \theta_j \circ \tau^{-1}, \qquad \tilde{\omega} := \omega \circ \tau^{-1},$$
$$\tilde{M}_\rho := \operatorname{span}\left\{\tilde{\theta}_1, \dots, \tilde{\theta}_N\right\}.$$

This way the points $\tilde{y}_j$ coincide with the points required for quasiinterpolation as stated in the following lemma.

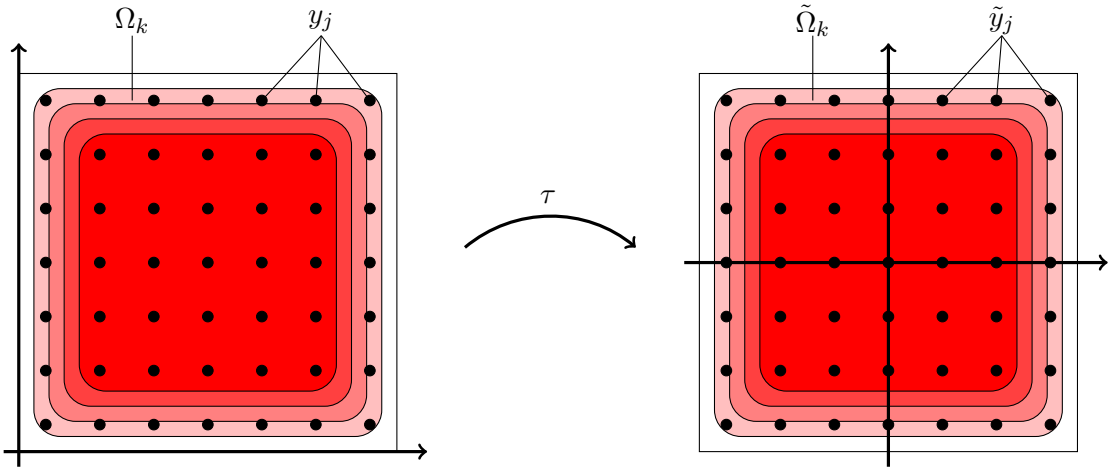


Figure 3.8: Bringing the points y_j into the form hm for some $m \in \mathbb{Z}^d$, by shifting Ω by $\tau(x) = x - \frac{1}{2}\mathbb{1}$. The sets $\Omega_1 \supset \Omega_2 \supset \Omega_3 \supset \Omega_4$ and $\tilde{\Omega}_1 \supset \tilde{\Omega}_2 \supset \tilde{\Omega}_3 \supset \tilde{\Omega}_4$ are visualized in red.

Lemma 3.20. Using Notations 3.10 and 3.19, we have for odd $n \in 2\mathbb{N}+1$, $N = n^d$, $h = 1/n$:

$$\{\tilde{y}_1, \dots, \tilde{y}_N\} = \left\{ hm \mid m \in \mathbb{Z}^d,\ hm \in \tilde{\Omega} \right\}.$$

Proof. Using the definition (2.1.1) of the points y_j, we get:

$$\begin{aligned}
\{\tilde{y}_1, \dots, \tilde{y}_N\} &= \left\{ \left(\frac{2k_1 - 1}{2n} - \frac{1}{2}, \dots, \frac{2k_d - 1}{2n} - \frac{1}{2} \right)^T \;\middle|\; k_j \in \{1, \dots, n\}\ \forall j \right\} \\
&= \left\{ \left(\frac{2k_1 - (n+1)}{2n}, \dots, \frac{2k_d - (n+1)}{2n} \right)^T \;\middle|\; k_j \in \{1, \dots, n\}\ \forall j \right\} \\
&= \left\{ \left(\left(k_1 - \frac{n+1}{2} \right) h, \dots, \left(k_d - \frac{n+1}{2} \right) h \right)^T \;\middle|\; k_j \in \{1, \dots, n\}\ \forall j \right\} \\
&= \left\{ (m_1 h, \dots, m_d h)^T \;\middle|\; m_j \in \left\{ -\frac{n-1}{2}, -\frac{n-3}{2}, \dots, \frac{n-1}{2} \right\}\ \forall j \right\} \\
&= \left\{ hm \mid m \in \mathbb{Z}^d,\ hm \in \tilde{\Omega} \right\}.
\end{aligned}$$

□

We are also going to use the following technical detail:

Lemma 3.21. Let $N \in \mathbb{N}$, $\alpha \in \mathbb{R}^N$ with exactly $\nu \leq N$ nonzero entries, $1 \leq p, q \leq \infty$ and, as usually, $1/q := 0$ for $q = \infty$. Then

$$\|\alpha\|_1 \leq \nu^{\frac{1}{q}} \|\alpha\|_p,$$

Proof. Let $v \in \mathbb{R}^N$ be given by $v_j = \operatorname{sgn}(\alpha_j),\ j = 1, \dots, N$. Hölder's inequality yields

$$\|\alpha\|_1 = |\langle v, \alpha \rangle| \leq \|v\|_q \|\alpha\|_p = \nu^{\frac{1}{q}} \|\alpha\|_p .$$

□

Proof of Theorem 3.17. Using Notations 3.10 and 3.19, we will consider two approximations of the function

$$\varphi = \tilde{\psi}_\rho\big|_{\tilde{\Omega}^{(2)}} \ \in W^{k,p}(\tilde{\Omega}^{(2)}, \mathbb{R}).$$

One is its quasi-interpolant

$$(Q_h\varphi)(y) = \mathcal{E}^d \sum_{\substack{m \in \mathbb{Z}^d \\ mh \in \Omega^{(2)}}} \psi(hm)\ \eta\left(\frac{\mathcal{E}(y - hm)}{h}\right) = \mathcal{E}^d \sum_{j=1}^{N} \varphi(\tilde{y}_j)\ \eta\left(\frac{\mathcal{E}(y - \tilde{y}_j)}{h}\right)$$

and the other is

$$\widehat{\varphi} = \sum_{j=1}^{N} \varphi(\tilde{y}_j)\ \tilde{\theta}_j \ \in \tilde{M}_\rho.$$

which are close to each other, since $\tilde{\theta}_j \approx \mathcal{E}^d\, \eta\left(\frac{\mathcal{E}(y-\tilde{y}_j)}{h}\right)$ by Proposition 3.15.

More precisely, since $\tilde{\theta}_j$ and $\mathcal{E}^d\eta\left(\frac{\mathcal{E}(\bullet-\tilde{y}_j)}{h}\right)$ are scaled by h and $\varphi(y) = 0$ for all $y \notin \tilde{\Omega}^{(2)}$ there exists a natural number $\nu \in \mathbb{N}$ independent of h, such that for each $y \in \mathbb{R}^d$

$$\left|\left\{ j = 1, \dots, N \;\middle|\; \varphi(\tilde{y}_j)\tilde{\theta}_j(y) \neq 0 \ \text{ or } \ \varphi(\tilde{y}_j)\mathcal{E}^d\eta\left(\frac{\mathcal{E}(y - \tilde{y}_j)}{h}\right) \neq 0 \right\}\right| \leq \nu$$

and therefore, Proposition 3.15 and Lemma 3.21 yield for $p < \infty$

$$\begin{aligned}
\|Q_h\varphi - \widehat{\varphi}\|^p_{L^p(\tilde{\Omega}^{(1)})} &\leq \int_{\mathbb{R}^d} \left(\sum_{j=1}^{N} |\varphi(\tilde{y}_j)| \left| \mathcal{E}^d\eta\left(\frac{\mathcal{E}(y - \tilde{y}_j)}{h}\right) - \tilde{\theta}_j(y) \right| \right)^p \mathrm{d}y \\
&\leq \int_{\mathbb{R}^d} \left\| \left(|\varphi(\tilde{y}_j)| \left| \mathcal{E}^d\eta\left(\frac{\mathcal{E}(y - \tilde{y}_j)}{h}\right) - \tilde{\theta}_j(y) \right| \right)_{j=1,\dots,N} \right\|^p_1 \mathrm{d}y \\
&\leq \int_{\mathbb{R}^d} \nu^{\frac{p}{q}} \left\| \left(|\varphi(\tilde{y}_j)| \left| \mathcal{E}^d\eta\left(\frac{\mathcal{E}(y - \tilde{y}_j)}{h}\right) - \tilde{\theta}_j(y) \right| \right)_{j=1,\dots,N} \right\|^p_p \mathrm{d}y \\
&\leq \nu^{\frac{p}{q}} \sum_{j=1}^{N} |\varphi(\tilde{y}_j)|^p \left\| \mathcal{E}^d\eta\left(\frac{\mathcal{E}(y - \tilde{y}_j)}{h}\right) - \tilde{\theta}_j(y) \right\|^p_{L^p(\mathbb{R}^d)} \\
&\leq \nu^{\frac{p}{q}} N \max_{y \in \tilde{\Omega}^{(2)}} |\varphi(y)|^p\ C^p \left(h^{1+\frac{d}{p}}\right)^p \\
&\leq \nu^{\frac{p}{q}} C^p h^p \max_{y \in \tilde{\Omega}^{(2)}} |\varphi(y)|^p,
\end{aligned}$$

where C is the constant from Proposition 3.15. Analogously, for $p = \infty$

$$\begin{aligned}\|Q_h\varphi - \widehat{\varphi}\|_{L^\infty(\tilde{\Omega}^{(1)})} &\leq \sup_{y\in\tilde{\Omega}^{(1)}} \sum_{j=1}^{N} |\varphi(\tilde{y}_j)| \left| \mathcal{E}^d \eta\left(\frac{\mathcal{E}}{h}(y-\tilde{y}_j)\right) - \tilde{\theta}_j(y)\right| \\ &\leq \nu C h \max_{y\in\tilde{\Omega}^{(2)}} |\varphi(y)|.\end{aligned}$$

So, for $1 \leq p \leq \infty$ and sufficiently small $h > 0$,

$$\|Q_h\varphi - \widehat{\varphi}\|_{L^p(\tilde{\Omega}^{(1)})} \leq \frac{\varepsilon}{8\hat{\omega}}. \tag{3.3.1}$$

Note that

$$\begin{aligned}\varphi \circ \tau \circ R &= \psi\big|_{A^{(2)}} \qquad \text{and} \\ \widehat{\varphi} \circ \tau \circ R &= \sum_{j=1}^{N} \varphi(\tilde{y}_j)\ \tilde{\theta}_j \circ \tau \circ R = \sum_{j=1}^{N} \psi\big|_{A^{(2)}}(x_j)\ \eta_j = \widehat{\psi} \in M.\end{aligned}$$

As shown below there are, roughly speaking, four errors we need to control:

(1) the quasi-interpolation error $\|Q_h\varphi - \varphi\|_{L^p\left(\tilde{\Omega}^{(2)}_{\kappa h}\right)}$, for which we will use Theorem 2.63 (κ is the constant from said theorem),

(2) The quasi-interpolation error "outside $\tilde{\Omega}^{(2)}_{\kappa h}$ ": $\|Q_h\varphi - \varphi\|_{L^p\left(\tilde{\Omega}\setminus\tilde{\Omega}^{(2)}_{\kappa h}\right)}$,

(3) the difference $\|Q_h\varphi - \hat{\varphi}\|_{L^p(\tilde{\Omega})}$, since our approximation has to lie in $\tilde{M}_\rho$ and the actual approximation is $\widehat{\varphi}$ instead of $Q_h\varphi$ – here we will apply Proposition 3.15,

(4) the error due to "cutting off" the function ψ, see Notation 3.19.

For sufficiently small h we can ensure

$$\operatorname{supp} Q_h\varphi \subseteq \tilde{\Omega}^{(1)} \qquad \text{and} \qquad \tilde{\Omega}^{(2)}_{\kappa h} := \{x \in \tilde{\Omega}^{(2)} : B_{\kappa h}(x) \subseteq \tilde{\Omega}^{(2)}\} \supseteq \tilde{\Omega}^{(3)}$$

and compute using Proposition 3.13:

$$\begin{aligned}&\|\psi - \widehat{\psi}\|_{L^p(\mathbb{R}^d)} \leq \|\psi - \widehat{\psi}\|_{L^p(A^{(3)})} + \|\widehat{\psi}\|_{L^p(\mathbb{R}^d\setminus A^{(3)})} + \|\psi\|_{L^p(\mathbb{R}^d\setminus A^{(3)})} \\ &\leq \|\varphi - \widehat{\varphi}\|_{L^p_{\tilde{\omega}}(\tilde{\Omega}^{(3)})} + \|\widehat{\varphi}\|_{L^p_{\tilde{\omega}}(\tilde{\Omega}^{(1)}\setminus\tilde{\Omega}^{(3)})} + \|\psi\|_{L^p(\mathbb{R}^d\setminus A^{(3)})} \\ &\leq \widehat{\omega}\left(\|\varphi - Q_h\varphi\|_{L^p(\tilde{\Omega}^{(3)})} + \|Q_h\varphi - \widehat{\varphi}\|_{L^p(\tilde{\Omega}^{(3)})}\right) + \\ &\quad + \widehat{\omega}\left(\|\widehat{\varphi} - Q_h\varphi\|_{L^p(\tilde{\Omega}^{(1)}\setminus\tilde{\Omega}^{(3)})} + \|Q_h\varphi\|_{L^p(\tilde{\Omega}^{(1)}\setminus\tilde{\Omega}^{(3)})}\right) + \|\psi\|_{L^p(\mathbb{R}^d\setminus A^{(3)})} \\ &\leq \widehat{\omega}\Big(\underbrace{\|\varphi - Q_h\varphi\|_{L^p(\tilde{\Omega}^{(3)})}}_{(1)} + 2\underbrace{\|Q_h\varphi - \widehat{\varphi}\|_{L^p(\tilde{\Omega}^{(1)})}}_{(3)} + \underbrace{\|Q_h\varphi\|_{L^p(\tilde{\Omega}^{(1)}\setminus\tilde{\Omega}^{(3)})}}_{(2)}\Big) + \underbrace{\|\psi\|_{L^p(\mathbb{R}^d\setminus A^{(3)})}}_{(4)} \\ &\leq \varepsilon,\end{aligned}$$

since

(1) $\leq \frac{\varepsilon}{4\hat{\omega}}$ for sufficiently small $\mathcal{E}, h > 0$ by Theorem 2.63,

(2) $\leq \|Q_h\big(\varphi|_{\tilde{\Omega}^{(2)}\setminus\tilde{\Omega}^{(4)}}\big)\|_{L^p(\mathbb{R}^d)} \leq \|Q_h\big(\psi_\rho|_{\Omega^{(2)}\setminus\Omega^{(4)}}\big)\|_{L^p(\mathbb{R}^d)} \leq \|\psi_\rho|_{\left(\Omega^{(2)}\setminus\Omega^{(4)}\right)}\|_{p,h} \leq \frac{\varepsilon}{4\hat{\omega}}$ by Lemma 2.62 and Notation 3.19,

(3) $\leq \frac{\varepsilon}{8\hat{\omega}}$ by (3.3.1),

(4) $\leq \frac{\varepsilon}{4}$ by Notation 3.19.

□

Remark 3.22. The proof shows further reasons to choose the (adapted) convolution $\rho_{g,t}$ instead of ρ_t for the density ρ:

(1) $\rho_{g,t}$ is strictly positive even if ρ_t is not.

(2) Due to the convolution, ψ_ρ is very small at the edges, therefore those basis functions θ_j, which are close to the edges and far from radial (see Figure 3.7 or Remark 3.18), mainly do not enter in the approximation of ψ_ρ.

3.4 Linear independence of the basis functions η_j

In the following, we will justify the term "basis function" by showing the linear independence of the functions η_j. Unfortunately, we were able to prove it in the radially symmetric case, for the general case we refer to the condition number plots of the Gramian matrix in all numerical experiments, see Chapter 4.

Proposition 3.23. Let $N \in \mathbb{N}$, $x_j, p_j \in \mathbb{R}^d$ and $\epsilon_j > 0$ for $j = 1, \ldots, N$. Then the functions

$$\eta_j(x) = \exp\left[-\epsilon_j(x - x_j)^2 + ip_j(x - x_j)\right]$$

are linearly independent in $L^2(\mathbb{R}^d, \mathbb{C})$, if the centers x_j are distinct.

Proof. Assume there exist $\lambda_1, \ldots, \lambda_N \neq 0$ (if some λ_j are zero, consider only those functions η_j with nonzero λ_j) , such that

$$\sum_{j=1}^{N} \lambda_j \eta_j = 0\,.$$

Without loss of generality, let $\epsilon_1 = \cdots = \epsilon_r < \epsilon_{r+1} \leq \cdots \leq \epsilon_N$ for some $r \geq 1$. Choose $j^* = \arg\max_{j=1,\ldots,r} \|x_k\|$, without loss of generality $j^* = 1$. Hence, $s := \max_{j=2,\ldots,r} x_1 x_j < x_1^2 \neq 0$ (it suffices to treat the case $N > 1$).
Using the notation $\mu_j := -\frac{\lambda_j}{\lambda_1}$, $\mu := \max\{|\mu_j| : j = 2, \ldots, N\}$, $\epsilon := \min\{\epsilon_j : j = r+1, \ldots, N\}$ we obtain from $\eta_1 = \sum_{j=2}^N \mu_j \eta_j$:

$$
\begin{aligned}
1 &= \lim_{t\to\infty} \frac{\left|\sum_{j=2}^N \mu_j \eta_j(tx_1)\right|}{|\eta_1(tx_1)|} \\
&\leq \lim_{t\to\infty} \left(\frac{\left|\sum_{j=2}^r \mu_j \eta_j(tx_1)\right|}{|\eta_1(tx_1)|} + \frac{\left|\sum_{j=r+1}^N \mu_j \eta_j(tx_1)\right|}{|\eta_1(tx_1)|} \right) \\
&\leq N\mu \lim_{t\to\infty} \left(\max_{j=2,\ldots,r} \frac{\exp[-\epsilon_1(tx_1 - x_j^2)]}{\exp[-\epsilon_1(tx_1 - x_1)^2]} \;+\; \max_{j=r+1,\ldots,N} \frac{\exp[-\epsilon(t\xi - x_j)^2]}{\exp[-\epsilon_1(tx_1 - x_1)^2]} \right) \\
&\leq N\mu \lim_{t\to\infty} \left(\max_{j=2,\ldots,r} \exp[-\epsilon_1(x_j^2 - x_1^2)] \exp[2\epsilon_1 t \underbrace{(s - x_1^2)}_{<0}] + \right. \\
&\qquad \left. + \;\exp[\underbrace{(\epsilon_1 - \epsilon)}_{>0}(tx_1)^2] \max_{j=r+1,\ldots,N} \exp[2(\epsilon x_j - \epsilon_1 x_1)x - \epsilon x_j^2 + \epsilon_1 x_1^2] \right) \\
&= 0. \qquad \lightning
\end{aligned}
$$

□

Chapter 4

Numerical Experiments

We are now ready to apply our method to an evolutionary PDE with underlying continuity equation

$$\begin{aligned} &\partial_t \psi_t = F_t(\psi_t), && \psi_0 = \psi_{\text{in}} \\ &\partial_t \rho_t = -\operatorname{div} j_t = -\operatorname{div}(\rho_t v_t), && \text{where } \rho_t = \frac{|\psi_t|^p}{\|\psi_t\|_{L^p}^p} \end{aligned}$$

(see Definition 1.1 for details) on the example of the Schrödinger equation (2.4.1). We will always assume that the velocity field v_t (or the current $j_t = \rho_t v_t$) is given analytically. This assumption seems absurd in the case of the Schrödinger equation, since $v_t = \Im\left[\frac{\nabla\psi_t}{\psi_t}\right]$ requires the knowledge of the function ψ_t, which we want to find in the first place. However, this procedure will suffice as a proof of concept.

The reason why taking the L^2-approximation u_t of ψ_t for the computation of v_t is a bad choice is given in Section 4.3. There, we will also give a possible solution to this problem, which so far has not been implemented numerically.

Unfortunately, this restriction forces us to treat examples for which the solution can be computed analytically. As a consequence, we only treat the free Schrödinger equation with potential $V = 0$ and the harmonic oscillator with potential $V(x) = \frac{x^2}{2}$. Please note that the L^2 errors are computed by first calculating the squared error and then taking the square root. Hence, the smallest possible error we can achieve is given by

$$\sqrt{\epsilon_{\text{machine}}} \approx 1.5 \cdot 10^{-8}.$$

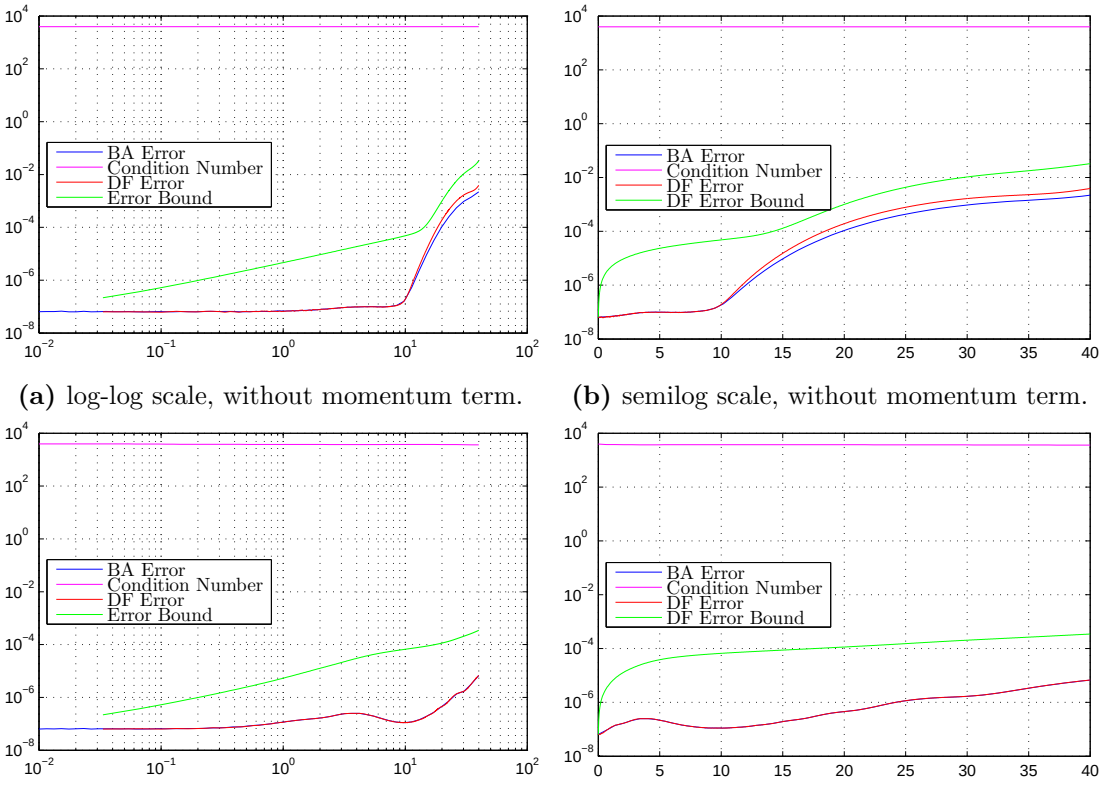
4.1 Free Dynamics

Let us approximate the following solution of the free Schrödinger equation (i.e. $V = 0$):

$$\psi_t = r(\psi_t^1 + \psi_t^2), \quad \text{where} \quad \psi_t^j(x) = \left(\frac{\sigma(t)}{\pi(1+i\sigma_0 t)}\right)^{\frac{1}{4}} \exp\left[-\frac{\sigma(t)}{2}(x-a_j)^2\right], \qquad (4.1.1)$$
$$\sigma(t) = \underbrace{\frac{\sigma_0}{1+\sigma_0^2 t^2}}_{\sigma_1(t)} + i\underbrace{\frac{-\sigma_0^2 t}{1+\sigma_0^2 t^2}}_{\sigma_2(t)}, \qquad a_{1,2} = \pm 3,$$

and $r > 0$ is just a normalization constant which guarantees $\|\psi_t\|_{L^2} = 1$. This example was taken from [Dec07].

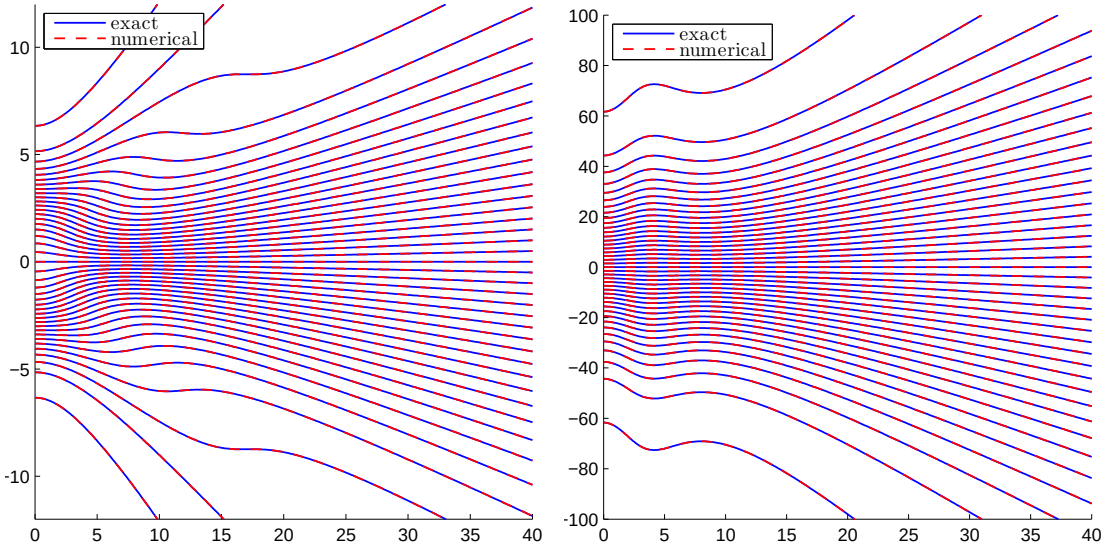
For our algorithm we choose the constants $N = 121$, $\mathcal{E} = 0.74$, $\kappa_\delta = 0.32$ and perform 30 time steps per time unit.



(a) log-log scale, without momentum term. **(b)** semilog scale, without momentum term.

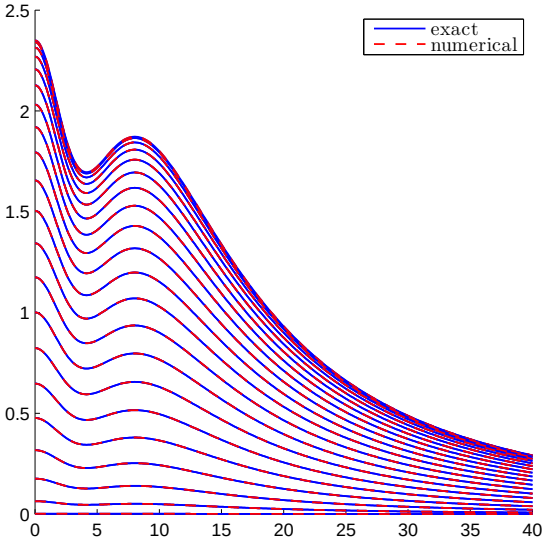
(c) log-log scale, with momentum term. **(d)** semilog scale, with momentum term.

Figure 4.1: Bestapproximation error, condition number of the stiffness matrix, error of the Dirac-Frenkel variational principle and its error bound plotted over time.

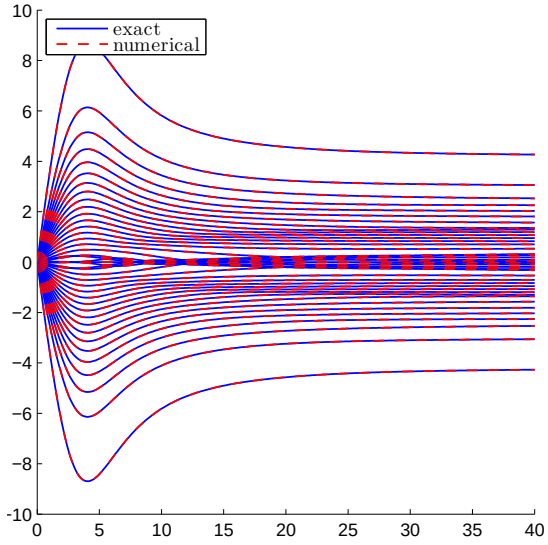


(a) Bohmian trajectories $q_{j,t}$.

(b) Quasi-Bohmian trajectories $x_{j,t}$.



(c) Parameters $\varepsilon_{j,t}$.



(d) Parameters $p_{j,t}$ (momentum terms).

Figure 4.2: Bohmian trajectories $q_{j,t}$, quasi-Bohmian trajectories $x_{j,t}$ and the parameters $\varepsilon_{j,t}$ and $p_{j,t}$ computed exactly and numerically over time (for reasons of visualisation every third trajectory/parameter is plotted).

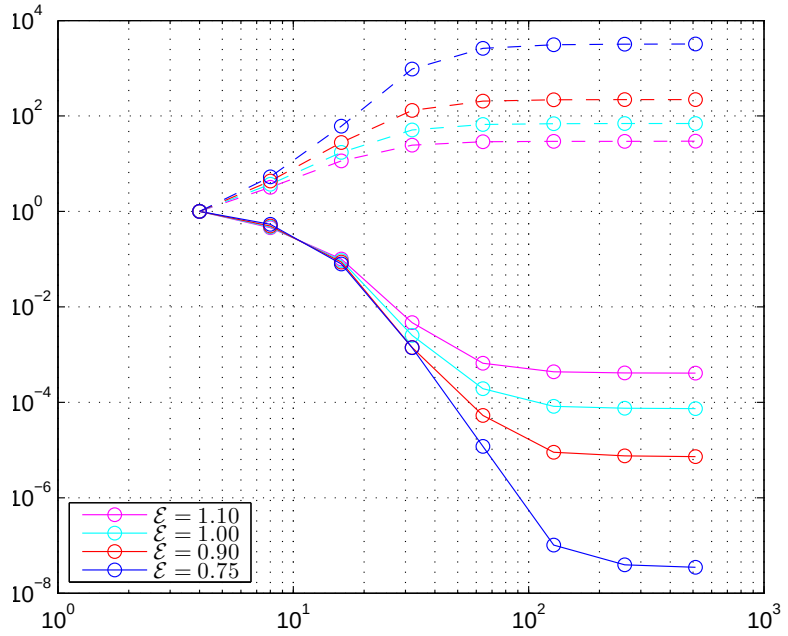


Figure 4.3: Bestapproximation error and condition number of the stiffness matrix plotted over the number N of basis functions for different values of the basic shape parameter $\mathcal{E}$. One can clearly see the saturation of the error for large values of N and also the trade-off-principle (the smaller $\mathcal{E}$, the better the approximation error, but the worse the condition number), both described in Section 2.8.2.

4.2 Harmonic Oscillator

Let us approximate the following solution of the harmonic oscillator, i.e. $V(x) = \frac{x^2}{2}$:

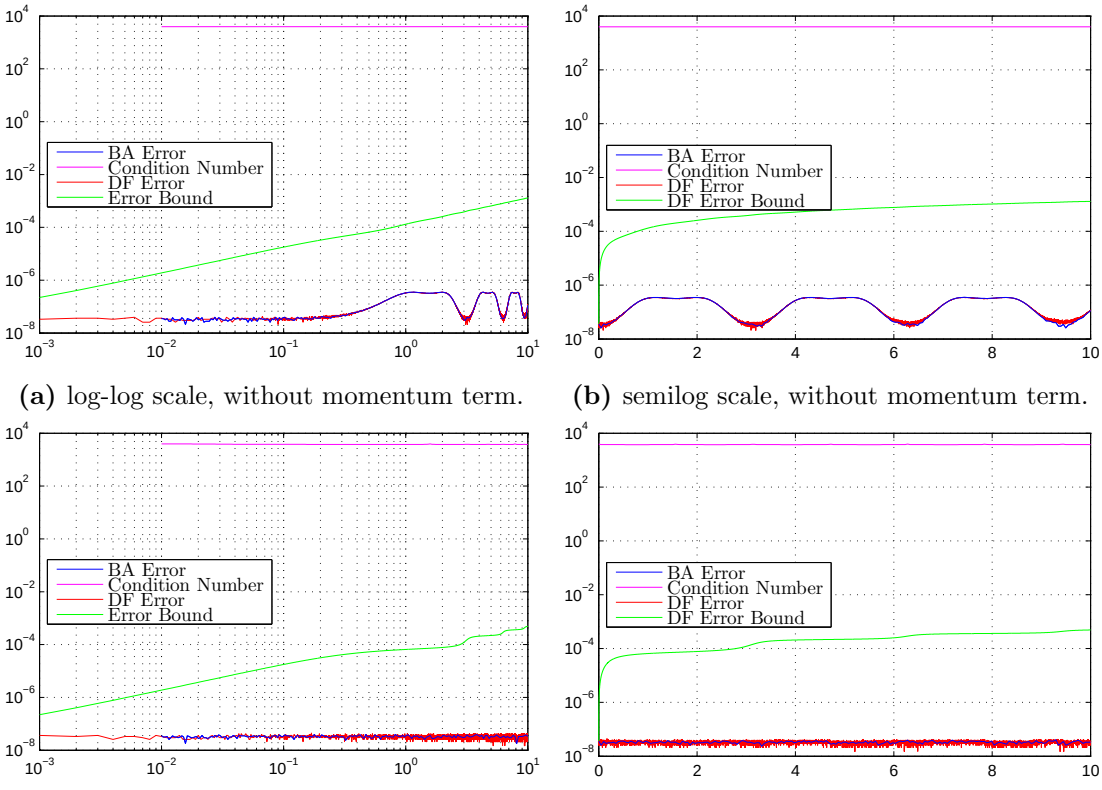
$$\psi_t(x) = \frac{1}{\pi^{1/4}\sqrt{Q_t}} \exp\left(-\frac{\sigma_t}{2}(x-q_t)^2 + ip_t(x-q_t) + iA_t\right)$$

where
$$\begin{aligned} q_t &= 2\cos(t), \\ p_t &= -2\sin(t), \\ A_t &= -\sin(2t), \\ Q_t &= \frac{\cos(t)}{\sqrt{2}} + i\sqrt{2}\sin(t), \\ \sigma_t &= \frac{2\cos(t) + i\sin(t)}{\cos(t) + 2i\sin(t)}. \end{aligned}$$

Note that the square root of Q_t has to be chosen in such a way that it stays continuous in time.

For our algorithm we choose the constants $N = 121$, $\mathcal{E} = 0.74$, $\kappa_\delta = 0.32$ and

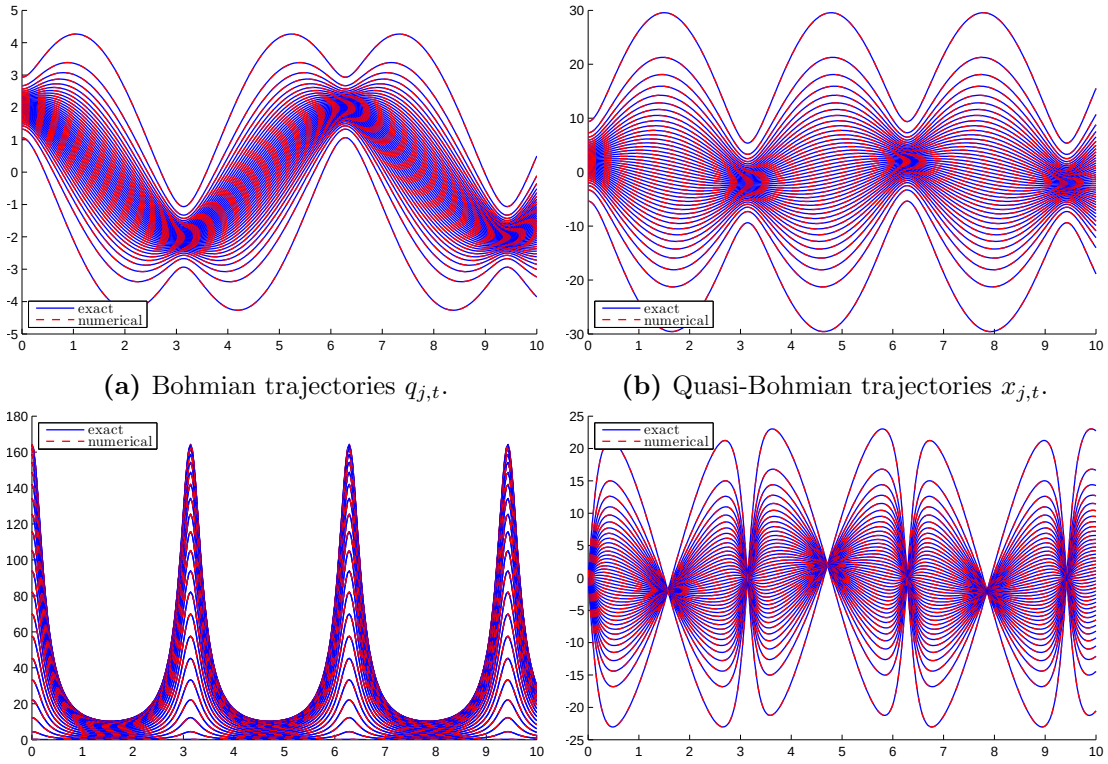
perform 1000 time steps per time unit.



(a) log-log scale, without momentum term. **(b)** semilog scale, without momentum term.

(c) log-log scale, with momentum term. **(d)** semilog scale, with momentum term.

Figure 4.4: Bestapproximation error, condition number of the stiffness matrix, error of the Dirac-Frenkel variational principle and its error bound plotted over time.



(a) Bohmian trajectories $q_{j,t}$.

(b) Quasi-Bohmian trajectories $x_{j,t}$.

(c) Parameters $\varepsilon_{j,t}$.

(d) Parameters $p_{j,t}$ (momentum terms).

Figure 4.5: Bohmian trajectories $q_{j,t}$, quasi-Bohmian trajectories $x_{j,t}$ and the parameters $\varepsilon_{j,t}$ and $p_{j,t}$ computed exactly and numerically over time (for reasons of visualisation every third trajectory/parameter is plotted).

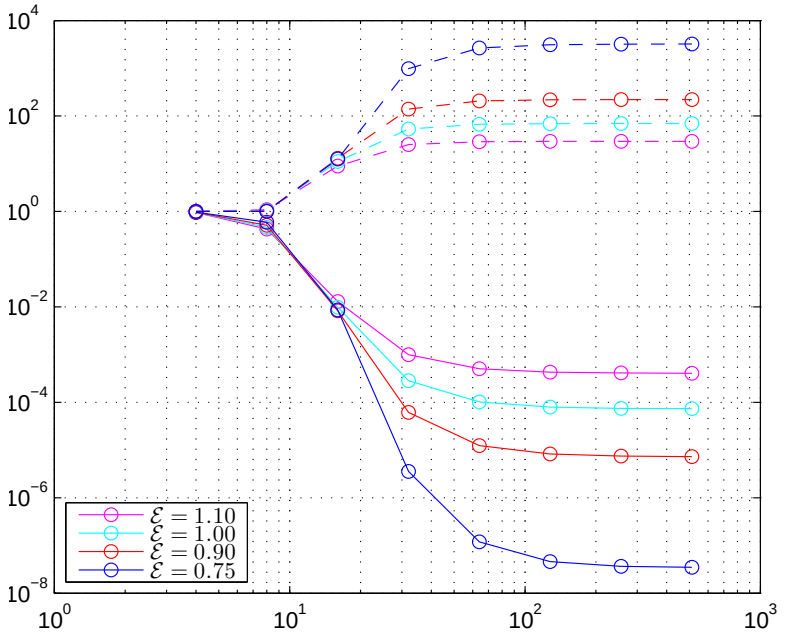


Figure 4.6: Bestapproximation error and condition number of the stiffness matrix plotted over the number N of basis functions for different values of the basic shape parameter $\mathcal{E}$. One can clearly see the saturation of the error for large values of N and also the trade-off-principle (the smaller $\mathcal{E}$, the better the approximation error, but the worse the condition number), both described in Section 2.8.2.

4.3 Numerical Computation of the Velocity Field v_t

As mentioned at the beginning of this chapter, we use an analytical formula for the computation of $v_t = \Im\left[\frac{\nabla\psi_t}{\psi_t}\right]$. The obvious choice of using our L^2-approximation $u_t = \sum_{j=1}^{N} c_{j,t}\eta_{j,t}$ of ψ_t in order to approximate $v_t \approx \Im\left[\frac{\nabla u_t}{u_t}\right]$ turns out to be unfit for the following reason:

No matter how well u_t approximates ψ_t (and ∇u_t approximates $\nabla\psi_t$) in the L^2 sense, the approximation can get oscillatory in regions of low values of ψ_t. Consider our standard example of ψ_0 being the weighted sum of two Gaussians and u_0 its bestapproximation in M_0. Figure 4.7 illustrates the oscillations of u_t and its result on the quotient $\frac{\nabla u_0}{u_0}$:

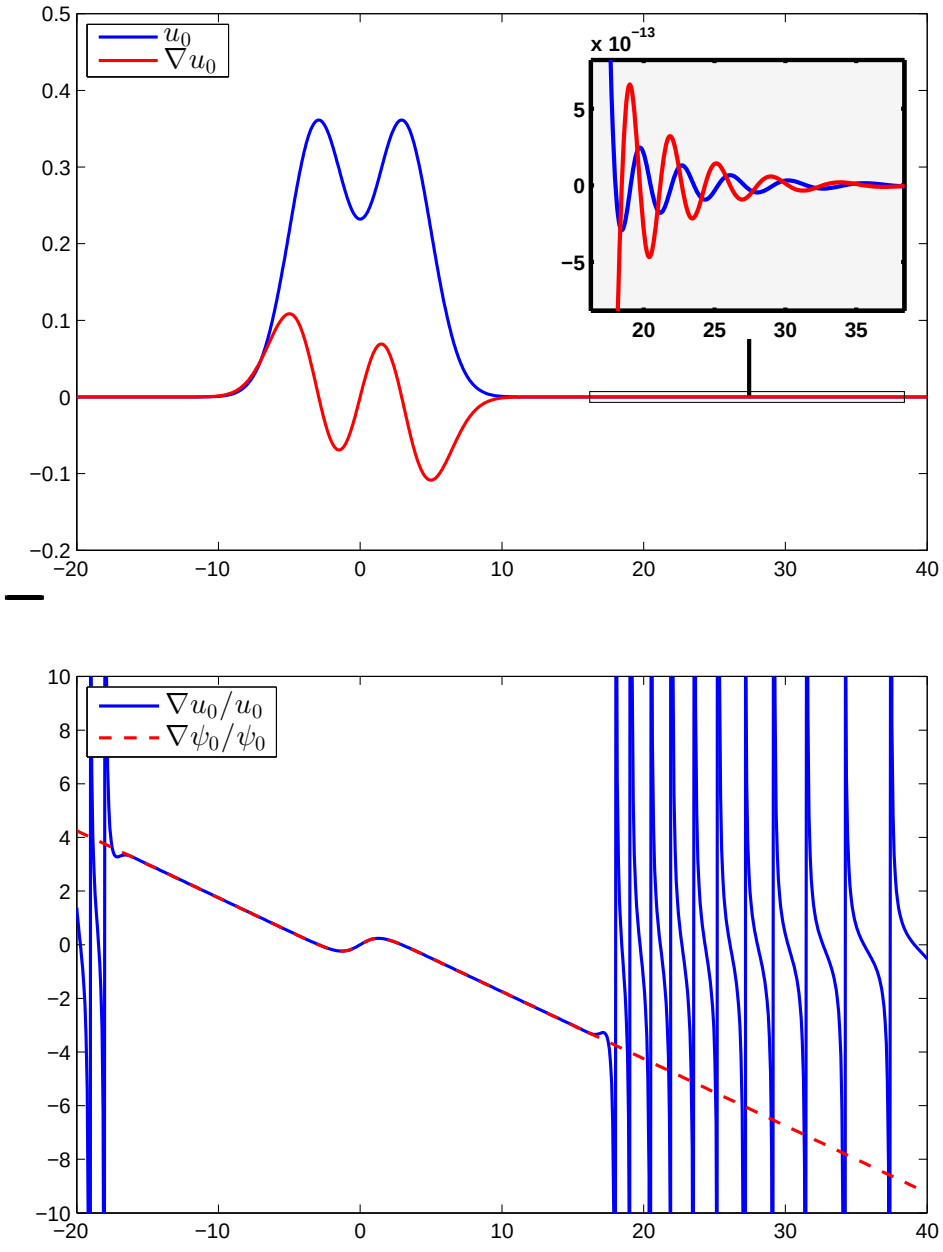


Figure 4.7: While ψ_0 is the sum of two Gaussians, its L^2-bestapproximation $u_0 \in M_0$ and its gradient get highly oscillatory in areas of low values of ψ_0. Though harmless for the L^2-approximation $u_0 \approx \psi_0$, the resulting approximation $\frac{\nabla u_0}{u_0} \approx \frac{\nabla \psi_0}{\psi_0}$ is catastrophic.

One possible solution to this problem is to rewrite $\psi_t = R_t e^{iS_t}$ or $\psi_t = e^{T_t}$ (see

Remark 2.27) for some functions $R_t, S_t : \mathbb{R}^d \to \mathbb{R}$ and $T_t : \mathbb{R}^d \to \mathbb{C}$ and use an approximation of S_t or T_t for the computation of v_t, since the formulas

$$v_t = \nabla S_t \qquad \text{and} \qquad v_t = \Im\left[\nabla T_t\right]$$

are much more pleasant from a numerical point of view. However, it is not clear which type of approximation to choose for R_t and S_t or for T_t, the time evolution of which is given by (2.4.3) and (2.4.4).

We suggest to treat the reformulation with T_t in order to avoid the unpleasant term $\frac{\Delta R_t}{2R_t}$ in (2.4.3) (it causes analogous problems to the one described above) and apply a collocation method for its time propagation. The algorithm will be explained only in dimension $d = 1$ and we will use basically the same basis function as before, except for three aspects:

- We do longer need the momentum term $p_{j,t}(x - y_{j,t})$ in the exponent, since in contrary to ψ_t, T_t does not show oscillatory behavior.
- $\psi_t \xrightarrow{|x|\to\infty} 0$ implies $\Re[T_t] \xrightarrow{|x|\to\infty} -\infty$, hence Gaussian basis functions alone are a bad choice for the approximation of T_t. Therefore, we will replace the outer for basis functions by the monomials $1, x, x^2, x^3$.
- Since we choose a collocation method instead of a Galerkin method, the prefactors are no longer needed.

In summary, we choose our basis functions to be

$$\zeta_{j,t}(x) = \begin{cases} 1 & \text{for } j = 1, \\ x & \text{for } j = 2, \\ \exp\left(-\epsilon_{j,t}\left|x - x_{j,t}\right|^2\right) & \text{for } j = 3, \dots, N-2, \\ x^2 & \text{for } j = N-1, \\ x^3 & \text{for } j = N, \end{cases}$$

where $\epsilon_{j,t} = h^{-2}\mathcal{E}^2\rho_{g,t}(x_{j,t})^2$ (see equation (3.2.2)). The time evolution of T_t,

$$\partial_t T_t = \frac{i}{2}\left(\Delta T_t + \nabla T_t^{\intercal}\nabla T_t\right) - iV,$$

makes the collocation method for the approximant $\Theta_t = \sum_{j=1}^N d_{j,t}\zeta_{j,t} \approx T_t$ gain the form

$$\begin{aligned}
&\partial_t\Theta_t(x_{k,t}) = \frac{i}{2}\left(\Theta_t'(x_{k,t}) + \Theta_t''(x_{k,t})^2\right) - iV(x_{k,t}) \qquad \forall k = 1,\dots,N \\
\iff \quad &\sum_{j=1}^N \dot{d}_{j,t} \underbrace{\zeta_{j,t}(x_{k,t})}_{=:\, a^{(1)}_{k,j,t}} = \sum_{j=1}^N d_{j,t}\Big(- \underbrace{\partial_t\zeta_{j,t}(x_{k,t})}_{=:\, a^{(2)}_{k,j,t}} + \frac{i}{2}\underbrace{\zeta_{j,t}''(x_{k,t})}_{=:\, a^{(3)}_{k,j,t}}\Big) \\
&\qquad\qquad + \frac{i}{2}\Big(\sum_{j=1}^N d_{j,t}\underbrace{\zeta_{j,t}'(x_{k,t})}_{=:\, a^{(4)}_{k,j,t}}\Big)^2 - iV(x_{k,t}) \qquad \forall k = 1,\dots,N \\
\iff \quad &A_t^{(1)}\dot{d}_t = \left(-A_t^{(2)} + \frac{i}{2}A_t^{(3)}\right) d_t + \frac{i}{2}\underbrace{\left(A_t^{(4)} d\right)^2}_{\text{componentwise}} - \underbrace{iV(x_t)}_{\text{componentwise}},
\end{aligned}$$

where $d_t := (d_{j,t})_j$, $x_t := (x_{j,t})_j$ and $A_t^{(\alpha)} := \left(a^{(\alpha)}_{k,j,t}\right)_{k,j}$ for $\alpha = 1,\dots,4$.

For $j = 3,\dots,N-2$ the derivatives of $\zeta_{j,t}$ are given by

$$\begin{aligned}
\zeta_{j,t}(x) &= \exp\left(-\epsilon_{j,t}\,(x - x_{j,t})^2\right), \\
\zeta_{j,t}'(x) &= -2\epsilon_{j,t}(x - x_{j,t})\,\zeta_{j,t}(x), \\
\zeta_{j,t}''(x) &= \left[4\epsilon_{j,t}^2(x - x_{j,t})^2 - 2\epsilon_{j,t}\right]\zeta_{j,t}(x), \\
\partial_t\zeta_{j,t}(x) &= \left[2\epsilon_{j,t}\dot{x}_{j,t}(x - x_{j,t}) - \dot{\epsilon}_{j,t}(x - x_{j,t})^2\right]\zeta_{j,t}(x).
\end{aligned}$$

Unfortunately, we were not able to produce a stable implementation of this method so far.

Instead, we follow a rather "clumsy" approach. We define

$$\begin{aligned}
&\tilde{\rho}_{g,t} = \rho_t * g_{\tilde{\delta}_t}, && \tilde{\delta}_t = \kappa_{\tilde{\delta}}\delta_t, \\
&\tilde{x}_{j,t} \;\; \mathbb{P}_{\tilde{\rho}_t}\text{-distributed}, && \Theta_{j,t} = T_t(\tilde{x}_{j,t}).
\end{aligned}$$

In each time step, we perform a cubic spline interpolation of the values $\Theta_{j,t}$ in the nodes $\tilde{x}_{j,t}$, which we use to compute the propagation of $\Theta_{j,t}$ (via a collocation method), $\tilde{x}_{j,t}$ and all the parameters of the manifold.

4.4 Morse Potential

We will now apply the above described method with the cubic spline interpolation to the Schrödinger equation with the Morse potential

$$V(x) = D\left(1 - e^{-\alpha(x-x_*)}\right)^2,$$

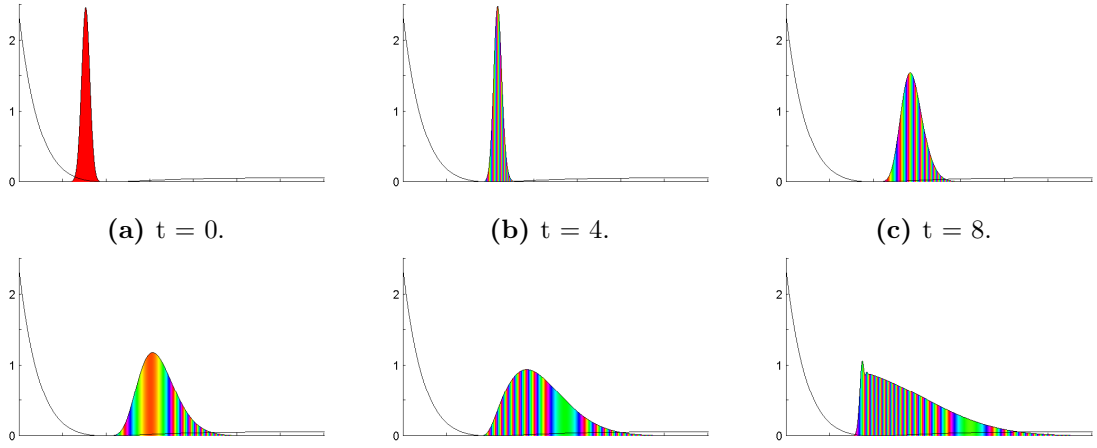
where $D = 0.0572,\ \alpha = 0.983,\ x_* = 5.03855$. The initial data will be a real Gaussian

$$\psi_{\text{in}}(x) = \left(\frac{\sigma}{\pi\varepsilon}\right)^{1/4} \exp\left(-\frac{\sigma}{2\varepsilon}(x - x_{\text{in}})\right),$$

where $\sigma = 0.3289,\ x_{\text{in}} = 4.53$. This example was taken from [Kel14], where the semiclassical formulation of the Schrödinger equation is used:

$$i\varepsilon\partial_t\psi_t = -\frac{\varepsilon^2}{2}\nabla\psi_t + V\psi_t,$$

where $0 < \varepsilon \ll 1$ is the so-called semiclassical parameter (here, $\varepsilon = 0.0029$). Thank you to Johannes Keller and Caroline Lasser for providing a reference solution for this example (computed via a Strang splitting):



(a) t = 0. **(b)** t = 4. **(c)** t = 8.

(d) t = 12. **(e)** t = 16. **(f)** t = 20.

Figure 4.8: Solution ψ_t of the above problem at different times t.

With only 20 nodes $\tilde{x}_{j,t}$ for the cubic spline interpolation, 100 time steps per time unit and $\kappa_{\tilde{\delta}} = 3.1$, we can propagate a suitable approximation of the manifold:

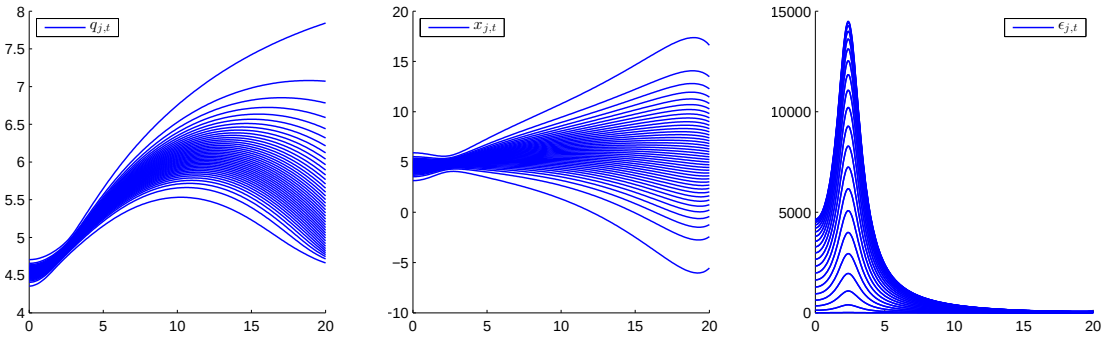


Figure 4.9: Bohmian trajectories $q_{j,t}$, quasi-Bohmian trajectories $x_{j,t}$ and the parameters $\varepsilon_{j,t}$ (for reasons of vizualisation every third trajectory/parameter is plotted). Note, how the Bohmian trajectories stay $|\psi_t|^2$-distributed.

The application of the Dirac Frenkel variational principle to the Morse potential causes problems. The potential does not fulfill the assumed conditions (at most polynomial growth, see Section 2.4). Therefore, the integrals $\langle \eta_{j,t} \mid V\eta_{k,t}\rangle_{L^2}$ appearing in the variational principle (see Example 2.33) might not exist. Nevertheless, we were able to evolve a reasonable approximation to the wavefunction up to time $T = 5$ (with $N = 121,\ \mathcal{E} = 0.74,\ \kappa_\delta = 0.32$ and 600 time steps per time unit.):

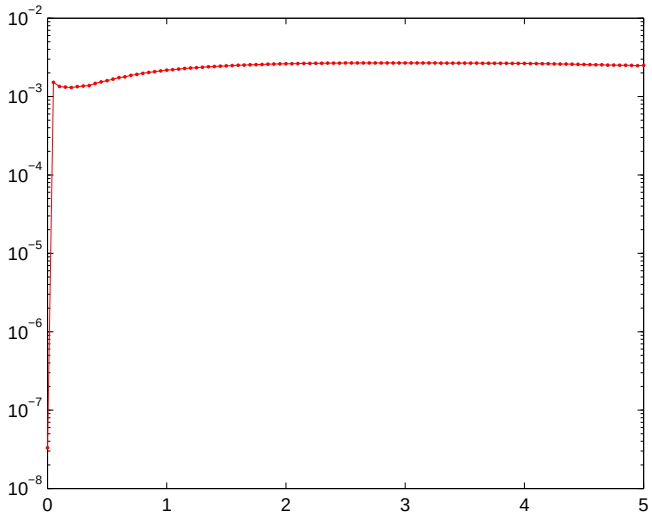


Figure 4.10: L^2-error of the Dirac Frenkel variational principle plotted over time.

Chapter 5

Conclusion and Future Directions

We developed a method using an approximation space, which automatically adapts to the specific function ψ_t we want to approximate. The basis functions $\eta_{j,t}$ we use for the approximation also adjust automatically, if the function changes in time. This was realized by considering the underlying probability density function ρ_t and its evolution in time, described by a continuity equation

$$\partial_t \rho_t = -\operatorname{div}(\rho_t v_t).$$

More precisely, the centers $x_{j,t}$ of the Gaussian basis functions $\eta_{j,t}$ were chosen $\mathbb{P}_{\rho_{g,t}}$-distributed, $\rho_{g,t}$ being a modified version of ρ_t, see Section 3.1.4. One way to put this into practice is by taking the images of equidistant points $y_1, \dots, y_N \in (0,1)^d$ under a transport map from $\mathbb{P}_{\text{uni}}$ to $\mathbb{P}_{\rho_{g,t}}$. The Jacobian of this transport map can be used for choosing suitable covariance matrices of the Gaussian basis functions, see Section 3.1.5.

This adaptation of the approximation space allows a good approximation with a rather small number of basis function, which is an important step towards high-dimensional problems. However, breaking the equidistance of the points y_j is an obstacle still to be cleared in order to reach this goal. Taking scattered $\mathbb{P}_{\text{uni}}$-distributed points y_j (e.g. Monte Carlo or quasi Monte Carlo points) has not yielded profitable results so far and would be an interesting direction for future research.

So far, our numerical experiments assumed that the velocity field v_t is given analytically (except for the rather negligible attempt in the last experiment). In many cases,

including the Schrödinger equation, the velocity field depends on the function ψ_t or the density ρ_t and needs to be computed numerically.

We explained the difficulties of such a computation in the case of the Schrödinger equation, which we could not solve until now. It therefore remains another challenging task for the future.

Appendix A

Bohmian Mechanics and the Wigner Transform

As mentioned in Section 2.6, some results can be deduced by treating the Wigner transform of the wave function as a probability density function, though in reality it may take negative values. Some of these results provide an insightful connection to Bohmian Mechanics, which (based on my knowledge) was first studied by Takabayasi (see [Tak54] or [Wya06, Chapter 3] for a summary) shortly after Bohm's famous publications on his statistical interpretation of quantum mechanics ([Boh52a] and [Boh52b]). Some of the ideas and formulas already appear in the appendix of Moyal's paper in 1949 [Moy49], who, of course, could not make the connection to Bohmian mechanics at that time. Further contributions were done by Hiley, see e.g. [Hil04].

The main idea is to treat the Wigner quasi-probability distribution $W\psi$ of the wave function $\psi \in L^2(\mathbb{R}^d, \mathbb{C})$ as a probability density function in phase space (ignoring the fact that it can take negative values) and "integrating out" the first or the second argument. Takabayasi refers to this as "projecting onto the coordinate space", while Hiley speaks of "shadow manifolds". The most simple result is to gain the marginal densities

$$\begin{aligned} \rho(x) \;=\; |\psi(x)|^2 \;&=\; \int_{\mathbb{R}^d} Wf(x,\xi)\,\mathrm{d}\xi\,, \\ |\mathcal{F}\psi(\xi)|^2 \;&=\; \int_{\mathbb{R}^d} W\psi(x,\xi)\,\mathrm{d}x\,. \end{aligned}$$

But there is more: Taking the expectation value with respect to the probability

density

$$\rho_x = \frac{W\psi(x,\bullet)}{\rho(x)}$$

(the denominator $\rho(x)$ is just a normalization term), just as we did in Section 2.7 (see Proposition 2.51), we arrive at the Bohmian velocity field $v = v_\psi = \Im\left[\frac{\nabla\psi}{\psi}\right]$. So, the velocity field suggested by Bohm is nothing other than the expected momentum (separately for each $x \in \mathbb{R}^d$) with respect to the Wigner distribution:

Proposition A.1. For every $x \in \mathbb{R}^d$, we have

$$\int_{\mathbb{R}^d} \xi\, \rho_x(\xi)\, \mathrm{d}\xi = \Im\left[\frac{\nabla\psi}{\psi}\right].$$

Proof. Using the Fourier inversion formula we get:

$$\begin{aligned}
\int_{\mathbb{R}^d} \xi\, \rho_x(\xi)\, d\xi \;&=\; \frac{(2\pi)^{-d}}{\rho(x)} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \overline{\psi}\left(x+\frac{y}{2}\right)\, \psi\left(x-\frac{y}{2}\right)\, \xi\, e^{i\xi^\intercal y}\, \mathrm{d}y\, \mathrm{d}\xi \\
&=\; \frac{(2\pi)^{-d}}{\rho(x)} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{2i}\left[\overline{\psi}\left(x+\frac{y}{2}\right)\, \nabla\psi\left(x-\frac{y}{2}\right) - \nabla\overline{\psi}\left(x+\frac{y}{2}\right)\, \psi\left(x-\frac{y}{2}\right)\right] e^{i\xi^\intercal y}\, \mathrm{d}y\, \mathrm{d}\xi \\
&=\; \frac{1}{\rho(x)}\frac{1}{2i}\left[\overline{\psi}(x)\,\nabla\psi(x) - \nabla\overline{\psi}(x)\,\psi(x)\right] \\
&=\; \frac{\Im\left[\overline{\psi}(x)\,\nabla\psi(x)\right]}{\rho(x)} \\
&=\; \Im\left[\frac{\nabla\psi}{\psi}\right].
\end{aligned}$$

□

Further connections, e.g. to the quantum potential introduced in Section 2.4, can be made by taking higher moments of $\mathbb{P}_{\rho_x}$ (see [Moy49], [Tak54], [Wya06]).

What we are going to study are the following two aspects:

1. Can an analogous statement be made with the Husimi transform (see Remark 2.40) in place of the Wigner distribution?

2. Using the terminology of Takabayasi, what happens if we "project onto other subspaces" instead of the coordinate space?

While the first question is answered by the following proposition, the second one will need a little more preparation.

Proposition A.2. Let $\psi \in L^2(\mathbb{R}^d)$ denote a wave function (i.e. $\rho = |\psi|^2$ is a probability density) and $H\psi = W\psi * G$, where $G(x,\xi) = G_1(x)G_2(\xi)$ and $G_1, G_2 \in \mathcal{S}(\mathbb{R}^d, \mathbb{R})$ are probability densities on $\mathbb{R}^d$, G_2 being an even function. Defining

$$\rho^H(q) = \int_{\mathbb{R}^d} H\psi(q,p)\,\mathrm{d}p \quad , \quad v_\psi^H(q) = \int_{\mathbb{R}^d} p\,\frac{H\psi(q,p)}{\rho^H(q)}\,\mathrm{d}p\,,$$

we get:

$$\rho^H = \rho * G_1 \qquad \text{and} \qquad \rho^H v_\psi^H = (\rho v_\psi) * G_1,$$

where as usually $v_\psi = \Im\left[\frac{\nabla\psi}{\psi}\right]$ denotes the Bohmian velocity field.

Proof. The first part is a simple computation:

$$\begin{aligned}\rho^H(q) &= \int_{\mathbb{R}^d} (W\psi * G)(q,p)\,\mathrm{d}p = \int_{\mathbb{R}^d}\int_{\mathbb{R}^{2d}} W\psi(q-x,p-\xi)\,G(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi\,\mathrm{d}p \\ &= \int_{\mathbb{R}^{2d}} \underbrace{\int_{\mathbb{R}^d} W\psi(q-x,p-\xi)\,\mathrm{d}p}_{\rho(q-x)}\; G_1(x)\,\mathrm{d}x\,G_2(\xi)\,\mathrm{d}\xi = (\rho * G_1)(q).\end{aligned}$$

For the second part first observe that (by the Fourier inversion formula)

$$\begin{aligned}&\int_{\mathbb{R}^{3d}} \xi\, W\psi(q-x,p-\xi)\,G(x,\xi)\,\mathrm{d}p\,\mathrm{d}x\,\mathrm{d}\xi \\ &= (2\pi)^{-d}\int_{\mathbb{R}^{4d}} \xi\,\overline{\psi}\left(q-x+\frac{y}{2}\right)\,\psi\left(q-x-\frac{y}{2}\right)\,e^{iy^\intercal(p-\xi)}\,G(x,\xi)\,\mathrm{d}y\,\mathrm{d}p\,\mathrm{d}x\,\mathrm{d}\xi \\ &= \int_{\mathbb{R}^{2d}} \xi\,\overline{\psi}\,(q-x)\;\psi\,(q-x)\;G(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi \\ &= \underbrace{\int_{\mathbb{R}^d} \xi\,G_2(\xi)\,\mathrm{d}\xi}_{=0,\text{ since } G_2 \text{ is even}}\;\int_{\mathbb{R}^d} \rho\,(q-x)\;G_1(x)\,\mathrm{d}x \\ &= 0.\end{aligned}$$

As a consequence,

$$\begin{aligned}\rho^H(q)\,v_\psi^H(q) &= \int_{\mathbb{R}^d} p\,(W\psi * G)(q,p)\,\mathrm{d}p \\ &= \int_{\mathbb{R}^{3d}} p\,W\psi(q-x,p-\xi)\,G(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi\,\mathrm{d}p \\ &= \int_{\mathbb{R}^{2d}} \underbrace{\int_{\mathbb{R}^d} (p-\xi)\,W\psi(q-x,p-\xi)\,\mathrm{d}p}_{\rho(q-x)\,v_\psi(q-x)}\;G(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi \\ &\qquad + \int_{\mathbb{R}^{3d}} \xi\,W\psi(q-x,p-\xi)\,\mathrm{d}p\,G(x,\xi)\,\mathrm{d}(x,\xi) \\ &= \big((\rho v_\psi) * G_1\big)(q).\end{aligned}$$

□

Let us now turn to the second question:

For simplicity, we will present the theory in dimension $d = 1$.

In the same way we did before, we can choose a 1-dimensional subspace U of the phase space other than the axes and "integrate out the orthogonal component" to get a probability densities "on U":

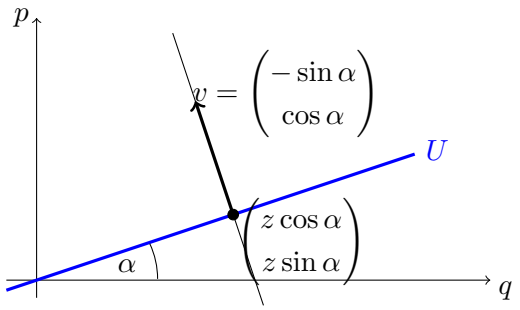


Figure A.1: 1-dimensional subspace U of the phase space and the direction vector of the corresponding orthogonal projection.

Definition A.3. We define the following densities on $\mathbb{R}$:

$$\rho^\alpha(z) = \int_{\mathbb{R}} W\psi\left(\begin{pmatrix} z\cos\alpha \\ z\sin\alpha \end{pmatrix} + h\begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}\right) \mathrm{d}h\,.$$

Since $\int_{\mathbb{R}^2} W_\psi = 1$, ρ^α is a probability density (the positivity will be seen in the proof of the following proposition). We further define the transformation

$$(\mathcal{F}^{\alpha,f_\alpha}\psi)(z) = e^{if_\alpha(z)} \int_{\mathbb{R}} \psi(y)\, K^\alpha(z,y)\, \mathrm{d}y,$$

where

$$K^\alpha(z,y) = \frac{1}{\sqrt{2\pi i \sin\alpha}} \exp\left[\frac{i}{2\sin\alpha}\left(\cos\alpha(y^2+z^2) - 2yz\right)\right]$$

is the Mehler kernel and $f_\alpha : \mathbb{R} \to \mathbb{R}$ is an arbitrary phase term.

Proposition A.4. For any $f_\alpha : \mathbb{R} \to \mathbb{R}$,

$$\rho^\alpha(z) = |(\mathcal{F}^{\alpha,f_\alpha}\psi)(z)|^2\,.$$

Proof. Using the linear transformation

$$\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} = \underbrace{\begin{pmatrix}-\sin\alpha & \frac{1}{2}\\ -\sin\alpha & -\frac{1}{2}\end{pmatrix}}_{A}\begin{pmatrix}h\\ y\end{pmatrix}, \qquad |\det A| = |\sin\alpha|, \quad \xi_1+\xi_2 = -2h\sin\alpha, \quad \xi_1-\xi_2 = y$$

and denoting $c := \cos(\alpha),\ s := \sin(\alpha)$, we get for any $f_\alpha : \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned}
\rho^\alpha(z) &= \int_{\mathbb{R}} W\psi\left(\begin{pmatrix}z\cos\alpha\\ z\sin\alpha\end{pmatrix} + h\begin{pmatrix}-\sin\alpha\\ \cos\alpha\end{pmatrix}\right)\mathrm{d}h\\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2}\bar{\psi}\left(zc - hs + \frac{y}{2}\right)\psi\left(zc - hs - \frac{y}{2}\right)e^{i(zs+hc)y}\,\mathrm{d}(y,h)\\
&= \frac{1}{2\pi|s|}\int_{\mathbb{R}^2}\bar{\psi}\left(zc+\xi_1\right)\psi\left(zc+\xi_2\right)e^{i\left(zs-\frac{c}{2s}(\xi_1+\xi_2)\right)(\xi_1-\xi_2)}\,\mathrm{d}(\xi_1,\xi_2)\\
&= \frac{1}{2\pi|s|}\int_{\mathbb{R}}\bar{\psi}\left(zc+\xi_1\right)e^{i\left(z\xi_1 s-\frac{c}{2s}\xi_1^2\right)}\,\mathrm{d}\xi_1\int_{\mathbb{R}}\psi\left(zc+\xi_2\right)e^{-i\left(z\xi_2 s-\frac{c}{2s}\xi_2^2\right)}\,\mathrm{d}\xi_2\\
&= \frac{1}{2\pi|s|}\left|\int_{\mathbb{R}}\psi\left(zc+\xi\right)e^{-i\left(z\xi s-\frac{c}{2s}\xi^2\right)}\,\mathrm{d}\xi\right|^2\\
&= \frac{1}{2\pi|s|}\left|\int_{\mathbb{R}}\psi\left(y\right)e^{\frac{i}{2s}\left(c(y^2+z^2)-2yz+z^2c\sin^2\alpha\right)}\,\mathrm{d}y\right|^2\\
&= \frac{1}{2\pi|s|}\left|e^{if_\alpha(z)}\int_{\mathbb{R}}\psi\left(y\right)e^{\frac{\pi i}{s}\left(c(y^2+z^2)-2yz\right)}\,\mathrm{d}y\right|^2.
\end{aligned}$$

□

Remark A.5.

1. For $f_\alpha = 0$ and $\alpha = \frac{\pi}{2}$, the transformation $\mathcal{F}^{\alpha,f_\alpha}\psi$ is just the Fourier transform of ψ.

2. The proposition also follows from the facts that the Mehler kernel is the evolution kernel of the harmonic oscillator and the evolution of the harmonic oscillator corresponds to a rotation of the Wigner function in phase space. Our proof is a simple alternative to this argumentation. Thank you to Johannes Keller for pointing that out to me.

3. The transformation $\mathcal{F}^{\alpha,f_\alpha}$ is not defined if α is a multiple of π, but it can be continuously extended to these cases. The following proposition treats the case $\alpha = 0$ exemplarily.

Proposition A.6. $(\mathcal{F}^{\alpha,f_\alpha}\psi)(z) \xrightarrow{\alpha\to 0} \psi(z)$ if $f_\alpha(z) \xrightarrow{\alpha\to 0} 0$ for all $z \in \mathbb{R}$.

Proof. It is sufficient to treat the case $\alpha \searrow 0$. Denoting once again $c := \cos(\alpha),\ s := \sin(\alpha)$, we obtain for $0 < \alpha < \frac{\pi}{2}$:

$$\begin{aligned}
(\mathcal{F}^{\alpha,f_\alpha}\psi)(z) &= e^{if_\alpha(z)} \int_{\mathbb{R}} \psi(y)\, \frac{1}{\sqrt{2\pi i s}} \exp\left[\frac{i}{2s}\left(c(y^2+z^2)-2yz\right)\right] \mathrm{d}y \\
&= \frac{e^{if_\alpha(z)}}{\sqrt{2\pi i s}} \int_{\mathbb{R}} \psi(y)\, \exp\left[\frac{ic}{2s}\left(\left(y-\frac{z}{c}\right)^2 - z^2\left(\frac{1}{c^2}-1\right)\right)\right] \mathrm{d}y \\
&= \frac{e^{if_\alpha(z)}}{\sqrt{2\pi i s}} \int_{\mathbb{R}} \psi(y)\, \exp\left[\frac{ic}{2s}\left(y-\frac{z}{c}\right)^2 - \frac{is}{2c}z^2\right] \mathrm{d}y \\
&= e^{if_\alpha(z)} \int_{\mathbb{R}} \psi\left(\sqrt{2\pi i s}\,\xi + \frac{z}{c}\right) \exp\left[-\pi c\,\xi^2 - \frac{is}{2c}z^2\right] \mathrm{d}\xi\,,
\end{aligned}$$

where we substituted $\xi = \frac{1}{\sqrt{2\pi i s}}\left(y - \frac{z}{c}\right)$ in the last step. Taking the limit $\alpha \searrow 0$, we get:

$$\begin{aligned}
\lim_{\alpha\searrow 0} \mathcal{F}^{\alpha,f_\alpha}\psi(z) &= \lim_{\alpha\searrow 0} \lim_{R\to\infty} e^{if_\alpha(z)} \int_{-R}^{R} \psi\left(\sqrt{2\pi i s}\,\xi + \frac{z}{c}\right) \exp\left[-\pi c\,\xi^2 - \frac{is}{2c}z^2\right] \mathrm{d}\xi \\
&= \lim_{R\to\infty} \int_{-R}^{R} \psi\left(z\right) \exp\left[-\pi\xi^2\right] \mathrm{d}\xi \\
&= \psi(z).
\end{aligned}$$

□

Proposition A.7. If $f_\alpha(x) = a_0 + a_1 x + a_2 x^2$ is a quadratic polynomial, the Wigner transform of ψ and $\mathcal{F}^{\alpha,f_\alpha}\psi$ have the following relation:

$$W\mathcal{F}^{\alpha,f_\alpha}\psi \begin{pmatrix} q \\ p \end{pmatrix} = W\psi \begin{pmatrix} q\cos\alpha - (p - a_1 - 2a_2 q)\sin\alpha \\ q\sin\alpha + (p - a_1 - 2a_2 q)\cos\alpha \end{pmatrix}.$$

In particular, if f_α is constant,

$$W\mathcal{F}^{\alpha,f_\alpha}\psi = W\psi \circ D_\alpha,$$

where $D_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ denotes the rotation matrix.

Proof. Denoting $c = \cos\alpha$ and $s = \sin\alpha$ we compute:

$$
\begin{aligned}
2\pi W\mathcal{F}^{\alpha,f_\alpha}\psi(q,p) &= \int_{\mathbb{R}} \overline{\mathcal{F}^{\alpha,f_\alpha}\psi\left(q+\frac{y}{2}\right)}\ \mathcal{F}^{\alpha,f_\alpha}\psi\left(q-\frac{y}{2}\right) e^{ipy}\,\mathrm{d}y \\
&= \frac{1}{2\pi|s|}\int_{\mathbb{R}^3} e^{-\frac{i}{2s}\left(cx_1^2+c\left(q+\frac{y}{2}\right)^2-2x_1\left(q+\frac{y}{2}\right)-cx_2^2-c\left(q-\frac{y}{2}\right)^2+2x_2\left(q-\frac{y}{2}\right)\right)} \\
&\qquad \exp\Big[i\left(py-a_1y-2a_2qy\right)\Big]\,\overline{\psi}(x_1)\,\psi(x_2)\,\mathrm{d}(x_1,x_2,y) \\
&= \frac{1}{2\pi|s|}\int_{\mathbb{R}^3} \exp\left[-\frac{i}{2s}\Big(c(x_1^2-x_2^2+2qy)-2(x_1-x_2)q-(x_1+x_2)y\Big)\right] \\
&\qquad \exp\Big[iy\left(p-a_1-2a_2q\right)\Big]\,\overline{\psi}(x_1)\,\psi(x_2)\,\mathrm{d}(x_1,x_2,y) \\
&\overset{(*)}{=} \frac{1}{2\pi}\int_{\mathbb{R}^3} \exp\left[-\frac{i}{2s}\Big(2cz_1z_2+2cqsz_3-2qz_2-2sz_3z_1\Big)+isz_3\left(p-a_1-2a_2q\right)\right] \\
&\qquad \overline{\psi}\left(z_1+\frac{z_2}{2}\right)\,\psi\left(z_1-\frac{z_2}{2}\right)\,\mathrm{d}(z_1,z_2,z_3) \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2} \exp\left[-iz_3\left(cq-z_1\right)+isz_3\left(p-a_1-2a_2q\right)\right] \\
&\qquad \int_{\mathbb{R}} \exp\left[iz_2\,\frac{q-cz_1}{s}\right]\overline{\psi}\left(z_1+\frac{z_2}{2}\right)\,\psi\left(z_1-\frac{z_2}{2}\right)\,\mathrm{d}z_2\,\mathrm{d}(z_1,z_3) \\
&\overset{(**)}{=} \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \exp\Big[iz_3\Big(s(p-a_1-2a_2q)-cq+z_1\Big)\Big]\ W\psi\left(z_1,\frac{q-cz_1}{s}\right)\mathrm{d}z_3\,\mathrm{d}z_1 \\
&= W\psi\left(cq-s(p-a_1-2a_2q),\frac{q-c^2q+cs(p-a_1-2a_2q)}{s}\right) \\
&= W\psi\Big(cq-s(p-a_1-2a_2q)\,,\,sq+c(p-a_1-2a_2q)\Big),
\end{aligned}
$$

where we used the Fourier inversion formula for $(**)$ and the following linear transformation for $(*)$:

$$
\begin{pmatrix} z_1\\ z_2\\ z_3\end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0\\ 1 & -1 & 0\\ 0 & 0 & \frac{1}{s}\end{pmatrix}}_{A}\begin{pmatrix} x_1\\ x_2\\ y\end{pmatrix}, \qquad |\det A| = \frac{1}{|s|}\ ,\quad x_1 = z_1+\frac{z_2}{2}\ ,\quad x_2 = z_1-\frac{z_2}{2}\,.
$$

□

Corollary A.8. Let f_α be constant and $\psi^\alpha = \mathcal{F}^{\alpha,f_\alpha}\psi$. Then

$$
\frac{1}{\rho^\alpha(z)}\int_{\mathbb{R}} W\psi\left(z\begin{pmatrix}\cos\alpha\\ \sin\alpha\end{pmatrix}+h\begin{pmatrix}-\sin\alpha\\ \cos\alpha\end{pmatrix}\right)\,h\,\mathrm{d}h = \Im\left[\frac{\nabla\psi^\alpha}{\psi^\alpha}\right](z).
$$

Proof. Propositions A.1 and A.7 yield

$$\begin{aligned}\frac{1}{\rho^\alpha(z)}\int_{\mathbb{R}} W\psi\left(z\begin{pmatrix}\cos\alpha\\ \sin\alpha\end{pmatrix}+h\begin{pmatrix}-\sin\alpha\\ \cos\alpha\end{pmatrix}\right)h\,\mathrm{d}h &= \int_{\mathbb{R}}\frac{(W\psi\circ D_\alpha)(z,h)}{|\psi^\alpha|^2(z)}\,h\,\mathrm{d}h\\ &= \int_{\mathbb{R}}\frac{W\psi^\alpha(z,h)}{|\psi^\alpha|^2(z)}\,h\,\mathrm{d}h\\ &= \Im\left[\frac{\nabla\psi^\alpha}{\psi^\alpha}\right](z).\end{aligned}$$

□

Corollary A.9. If $f_\alpha(x)=a_0+a_2x^2$ is a symmetric quadratic polynomial, the relation between the Wigner transform of ψ and $\mathcal{F}^{\alpha,f_\alpha}\psi$ becomes:

$$W\mathcal{F}^{\alpha,f_\alpha}\psi = W\psi\circ D_\alpha^{f_\alpha}\quad,\quad\text{where}\qquad D_\alpha^{f_\alpha}=\begin{pmatrix}\cos\alpha+2a_2\sin\alpha & -\sin\alpha\\ \sin\alpha-2a_2\cos\alpha & \cos\alpha\end{pmatrix}$$
$$\left(D_\alpha^{f_\alpha}\right)^{-1}=\begin{pmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha+2a_2\cos\alpha & \cos\alpha+2a_2\sin\alpha\end{pmatrix}.$$

In particular, $W\mathcal{F}^{\alpha,f_\alpha}\psi$ and $W\psi$ attain the same values, just the coordinates where the values are taken are "rotated". We want to visualize this pseudo-rotation $D_\alpha^{f_\alpha}$ on the example $f_\alpha(z)=\frac{z^2}{2}\cos\alpha\sin\alpha$ by plotting the trajectories:

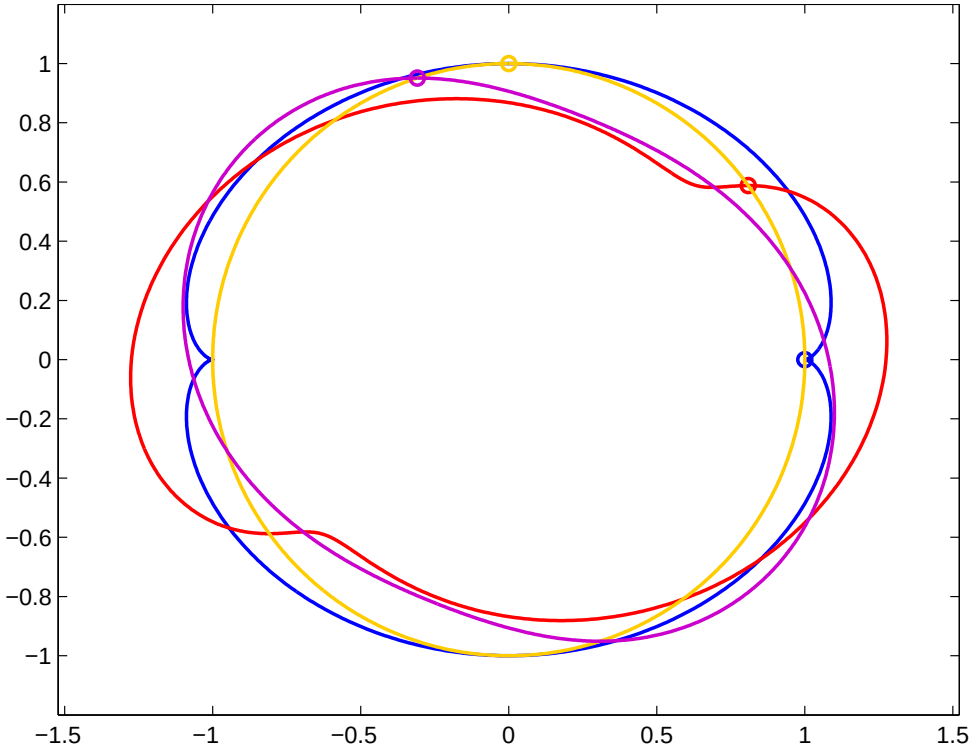


Figure A.2: Trajectories (in the variable α) under the pseudo-rotation $D_\alpha^{f_\alpha}$ for various starting points.

Appendix B

Generalized FBI Transform

The theory presented in this section originated from a collaboration with my colleagues Johannes Keller and Stephanie Troppmann.

In Section 2.6, we defined the FBI transform of a function $f \in S(\mathbb{R}^d, \mathbb{C})$ as its projection onto a complex Gaussian:

$$\mathcal{T}f(x,\xi) = (2\pi)^{-d/2}\big\langle g_{x,\xi},\, f\big\rangle_{L^2(\mathbb{R}^d)}, \qquad g_{x,\xi}(y) = \pi^{-d/4}\exp\left(-\frac{|x|^2}{2} + i\xi^\intercal(y-x)\right)$$

(the point of evaluation of $\mathcal{T}f$ coincides with the center of the Gaussian in phase space).

Our basic idea is to generalize this transform by plugging in various Hermite functions (i.e. polynomial $\times$ Gaussian) in place of $g_{x,\xi}$, or, even more general, so-called Hagedorn wave packets, which form an orthonormal basis of $L^2(\mathbb{R}^d)$. This approach will be presented in the following section.

In this chapter, we will use semiclassical scaling, which is reflected by adding a small parameter $\varepsilon > 0$ to the Fourier, the FBI and the Wigner transforms:

$$\begin{aligned}
\mathcal{F}^\varepsilon f(\xi) &= (2\pi\varepsilon)^{-d/2}\int_{\mathbb{R}^d} f(y)\, e^{-iy^\intercal\xi/\varepsilon}\,\mathrm{d}y\,,\\
\mathcal{T}^\varepsilon f(x,\xi) &= (2\pi\varepsilon)^{-d/2}(\pi\varepsilon)^{-d/4}\int_{\mathbb{R}^d} f(y)\,\exp\left(-\frac{|x|^2}{2\varepsilon} - \frac{i}{\varepsilon}\xi^\intercal(y-x)\right)\mathrm{d}y\,,\\
W^\varepsilon f(x,\xi) &= (2\pi\varepsilon)^{-d}\int_{\mathbb{R}^d} \overline{f\left(x+\frac{y}{2}\right)}\, f\left(x-\frac{y}{2}\right)\, e^{iy^\intercal\xi/\varepsilon}\,\mathrm{d}y.
\end{aligned}$$

We will further make use of the following Laguerre type polynomials:

Definition B.1. For $k \in \mathbb{N}$ and $\alpha \in \mathbb{R}$ let

$$L_k^\alpha(x) = \sum_{j=0}^{k}\binom{k+\alpha}{k-\alpha}\frac{(-x)^j}{j!}$$

be the (generalized) Laguerre polynomials associated with γ. For $m, n \in \mathbb{N}$, we define the polynomials

$$\mathcal{L}_{m,n}(x,y) = \begin{cases} 2^n m! y^{n-m} L_m^{n-m}(-2xy), & \text{if } m \leq n, \\ 2^m n! y^{m-n} L_n^{m-n}(-2xy), & \text{if } n \leq m. \end{cases}$$

B.1 Hagedorn Wave Packets

In his famous papers from 1980 and 1998 ([Hag80] and [Hag98]), George Hagedorn introduced an orthonormal basis of $L^2(\mathbb{R}^d)$, which generalizes the common Hermite basis (and its tensor-product version in dimensions $d > 1$). We will follow the notation in [Lub08, Chapter V] and [Las14]:

Definition B.2 (admissible pair)**.** We will call the pair (Q, P) with $Q, P \in \mathbb{C}^{d\times d}$ *admissible*, if

$$\begin{aligned} Q^\intercal P - P^\intercal Q &= 0, \\ Q^* P - P^* Q &= 2i\mathrm{Id}. \end{aligned} \tag{B.1.1}$$

For $q, p \in \mathbb{R}^d$ and an admissible pair (Q, P), we define the corresponding *complex Gaussian* by

$$\varphi_0^\varepsilon[q,p,Q,P](x) = (\pi\varepsilon)^{-d/4} (\det Q)^{-1/2} \exp\left(\frac{i}{2\varepsilon}(x-q)^\intercal P Q^{-1}(x-q) + \frac{i}{\varepsilon} p^\intercal (x-q)\right).$$

As in the case of Hermite functions, the orthonormal basis results from multiple applications of the ladder operators (often called raising and lowering operators) $A = (A_j)_{j=1}^d$ and $A^\dagger = (A_j^\dagger)_{j=1}^d$ to the Gaussian, in this case the proper generalization is

$$\begin{aligned} A = A[q,p,Q,P] &= -\frac{i}{\sqrt{2\varepsilon}} \left(P^\intercal \mathrm{op}_\varepsilon(x-q) - Q^\intercal(-i\varepsilon\nabla - p)\right), \\ A^\dagger = A^\dagger[q,p,Q,P] &= \frac{i}{\sqrt{2\varepsilon}} \left(P^* \mathrm{op}_\varepsilon(x-q) - Q^*(-i\varepsilon\nabla - p)\right). \end{aligned}$$

The kth Hagedorn wavepacket, $k \in \mathbb{N}^d$ is defined by

$$\varphi_k^\varepsilon = \varphi_k^\varepsilon[q,p,Q,P] = \frac{1}{\sqrt{k!}} (A^\dagger)^k \, \varphi_0^\varepsilon[q,p,Q,P],$$

where we use the multi-index notation

$$(A^\dagger)^k = (A_1^\dagger)^{k_1} \circ \cdots \circ (A_d^\dagger)^{k_d}, \qquad k! = \prod_{j=1}^d k_j!\,.$$

Remark B.3. For an admissible pair (Q, P), one can show that (see [Lub08, Chapter V])

- Q and P are invertible,
- $C = PQ^{-1}$ is a complex symmetric matrix with positive definite imaginary part,
- every complex symmetric matrix C with positive definite imaginary part has a decomposition of the form $C = PQ^{-1}$ with admissible pair (Q, P), but this decomposition is not unique,
- $QQ^* = \Im[C]^{-1}$ is symmetric and positive definite,
- $PP^* = -\Im[C^{-1}]^{-1}$ is symmetric,
- the ladder operators commute ($A_j^\dagger A_k^\dagger = A_k^\dagger A_j^\dagger$ for all $j, k = 1, \dots d$), therefore the upper definition is meaningful,
- the Hagedorn wave packets $(\varphi_k)_{k\in\mathbb{N}^d}$ form an orthonormal basis of $L^2(\mathbb{R}^d)$,
- In the case $q = p = 0,\ Q = \mathrm{Id},\ P = i\mathrm{Id}$ we regain the common Hermite basis for $d = 1$ and its tensor product form for $d > 1$.

Just as in the case of standard Hermite functions, the Hagedorn wavepackets are (by construction) a product of a polynomial and the (complex) Gaussian φ_0^ε. Let us introduce the notation (see [Las14])

$$\varphi_k^\varepsilon[q, p, Q, P](x) = \frac{1}{\sqrt{2^{|k|}k!}}\, p_k^\varepsilon[q, Q](x)\, \varphi_0^\varepsilon[q, p, Q, P](x), \qquad x \in \mathbb{R}^d,$$

where the multivariate polynomials $p_k^\varepsilon[q, Q]$, $k \in \mathbb{N}^d$ are recursively defined by

$$p_0^\varepsilon[q, Q] = 1, \qquad p_{k+e_j}^\varepsilon[q, Q] = B_j^\dagger p_k^\varepsilon[q, Q], \qquad \text{where}$$
$$B^\dagger = (B_j^\dagger)_{j=1}^d = \frac{2}{\sqrt{\varepsilon}} Q^{-1}\mathrm{op}_\varepsilon(x - q) - \frac{i}{\sqrt{\varepsilon}} Q^*(-i\varepsilon\nabla_x).$$

In the following proposition, we will see that varying $q \in \mathbb{R}^d$ only results in shifting the Hagedorn polynomials p_k^ε by q. Therefore, we will abbreviate $p_k^\varepsilon[0, Q]$ by p_k^ε and discuss most results only for p_k^ε.

Proposition B.4. The multivariate polynomials $p_k^\varepsilon[q, Q]$ have the following properties (here $x, y, z \in \mathbb{C}^d$ and $p_{k-e_j}^\varepsilon[q, Q] := 0$ for $k_j = 0$):

(1) $\left(p^\varepsilon_{k+e_j}(x)\right)_{j=1}^d = \frac{2}{\sqrt{\varepsilon}} Q^{-1}(x-q)p^\varepsilon_k[q,Q](x) - 2Q^{-1}\overline{Q}\left(k_j p^\varepsilon_{k-e_j}(x)\right)_{j=1}^d$,

(2) $p^\varepsilon_k(x+z) = \sum_{\nu\le k}\binom{k}{\nu}\left(\frac{2}{\sqrt{\varepsilon}}Q^{-1}z\right)^{k-\nu} p^\varepsilon_\nu(x)$,

(3) $\int_{\mathbb{R}^d} \overline{p^\varepsilon_l}(x+y)p^\varepsilon_k(x+z)|\varphi^\varepsilon_0(x)|^2dx = \prod_{j=1}^d \mathcal{L}_{l_j,k_j}\left(\frac{1}{\sqrt{\varepsilon}}(\overline{Q^{-1}}y)_j\,,\,\frac{1}{\sqrt{\varepsilon}}(Q^{-1}z)_j\right)$,

(4) $p^\varepsilon_k[q,Q](x) = p^\varepsilon_k[0,Q](x-q)$,

(5) $\nabla_x p^\epsilon_k = \frac{2}{\sqrt{\epsilon}}Q^{-T}(k_j p^\epsilon_{k-e_j})_{j=1}^d$, if $k_j > 0$,

(6) $p^\varepsilon_k(x+y) = 2^{-\frac{|k|}{2}}\sum_{\nu\le k}\binom{k}{\nu} p^\varepsilon_{k-\nu}\big(\sqrt{2}x\big)\,p^\varepsilon_\nu\big(\sqrt{2}y\big)$.

Proof. For the proofs of (1),(2) and (3) see [Las14].

(4) Roughly speaking, this identity results from the recursive definition of $p^\varepsilon_k[q,Q]$ and the fact that ∇_x and the shifting operator τ_q defined by $(\tau_q f)(x) = f(x-q)$ commute, while $\tau_q \circ \mathrm{op}_\varepsilon(x) = \mathrm{op}_\varepsilon(x-q)\circ\tau_q$. More precisely, we have by induction (we denote the entries of Q^{-1} by a_{ij} and the ones of Q^* by b_{ij}):

$$\begin{aligned}
p^\varepsilon_{k+e_j}[q,Q] &= \left[\frac{2}{\sqrt{\varepsilon}}Q^{-1}\mathrm{op}_\varepsilon(x-q) - \sqrt{\varepsilon}\,Q^*\,\nabla_x\right]_j p^\varepsilon_k[q,Q] \\
&= \frac{2}{\sqrt{\varepsilon}}\sum_{l=1}^d a_{jl}\,\mathrm{op}_\varepsilon(x_l-q_l)\,\tau_q\,p^\varepsilon_k[0,Q] - \sqrt{\varepsilon}\sum_{l=1}^d b_{jl}\,\nabla_{x_l}\,\tau_q\,p^\varepsilon_k[0,Q] \\
&= \frac{2}{\sqrt{\varepsilon}}\sum_{l=1}^d a_{jl}\,\tau_q\,\mathrm{op}_\varepsilon(x_l)\,p^\varepsilon_k[0,Q] - \sqrt{\varepsilon}\sum_{l=1}^d b_{jl}\,\tau_q\,\nabla_{x_l}\,p^\varepsilon_k[0,Q] \\
&= \tau_q\left[\frac{2}{\sqrt{\varepsilon}}Q^{-1}\,\mathrm{op}_\varepsilon(x_l) - \sqrt{\varepsilon}\,Q^*\,\tau_q\,\nabla_{x_l}\right]_j p^\varepsilon_k[0,Q] \\
&= \tau_q\;p^\varepsilon_{k+e_j}[0,Q].
\end{aligned}$$

(5) Combining the recursive definition of the Hagedorn polynomials $p^\varepsilon_k[q,Q]$ and the Rodriguez type formula (1), we obtain

$$-\frac{i}{\sqrt{\varepsilon}}Q^*\left(-i\varepsilon\nabla_x p^\varepsilon_k[q,Q]\right) = -2Q^{-1}\overline{Q}\left(k_j p^\varepsilon_{k-e_j}[q,Q]\right)_{j=1}^d\,.$$

Since $QQ^* = \Im[C]^{-1}$ is real-valued (see Remark B.3), we have $(Q^*)^{-1}Q^{-1}\overline{Q} = (QQ^*)^{-1}\overline{Q} = (\overline{(QQ^*)})^{-1}\overline{Q} = Q^{-T}$, which proves the claim.

(6) Applying the three-term recurrence formula thrice, we get by induction over $k \in \mathbb{N}^d$:

$$
\begin{aligned}
&2^{\frac{|k|+1}{2}} \left(p^\varepsilon_{k+e_j}(x+y)\right)_{j=1}^d = 2^{\frac{|k|+1}{2}} \left[\frac{2}{\sqrt{\varepsilon}} Q^{-1}(x+y) p^\varepsilon_k(x+y) - 2Q^{-1}\overline{Q}\left(k_j p^\varepsilon_{k-e_j}(x+y)\right)_{j=1}^d\right] \\
&= \frac{2}{\sqrt{\varepsilon}} Q^{-1}(\sqrt{2}x+\sqrt{2}y) \sum_{\nu\le k} \binom{k}{\nu} p^\varepsilon_{k-\nu}(\sqrt{2}x)\, p^\varepsilon_\nu(\sqrt{2}y) \\
&\quad - 2\cdot 2Q^{-1}\overline{Q}\Bigg(k_j \sum_{\nu\le k-e_j} \underbrace{\binom{k-e_j}{\nu}}_{\binom{k}{\nu}\frac{k_j-\nu_j}{k_j}} p^\varepsilon_{k-\nu-e_j}(\sqrt{2}x)\, p^\varepsilon_\nu(\sqrt{2}y)\Bigg)_{j=1}^d \\
&= \sum_{\nu\le k}\binom{k}{\nu}\left(\frac{2}{\sqrt{\varepsilon}} Q^{-1}(\sqrt{2}x) p^\varepsilon_{k-\nu}(\sqrt{2}x) - 2Q^{-1}\overline{Q}\left((k_j-\nu_j)\, p^\varepsilon_{k-\nu-e_j}(\sqrt{2}x)\right)_{j=1}^d\right) p^\varepsilon_\nu(\sqrt{2}y) \\
&\quad + \sum_{\nu\le k}\binom{k}{\nu}\left(\frac{2}{\sqrt{\varepsilon}} Q^{-1}(\sqrt{2}y) p^\varepsilon_{k-\nu}(\sqrt{2}y) - 2Q^{-1}\overline{Q}\left((k_j-\nu_j)\, p^\varepsilon_{k-\nu-e_j}(\sqrt{2}y)\right)_{j=1}^d\right) p^\varepsilon_\nu(\sqrt{2}x) \\
&= \Bigg(\sum_{\nu\le k}\binom{k}{\nu} p^\varepsilon_{k-\nu+e_j}(\sqrt{2}x)\, p^\varepsilon_\nu(\sqrt{2}y) + \underbrace{\sum_{\nu\le k}\binom{k}{\nu} p^\varepsilon_{k-\nu+e_j}(\sqrt{2}y)\, p^\varepsilon_\nu(\sqrt{2}x)}_{\substack{\nu\to k-\nu+e_j \\ = \sum_{\nu\le k+e_j}\binom{k}{k-\nu+e_j} p^\varepsilon_\nu(\sqrt{2}y) p^\varepsilon_{k-\nu+e_j}(\sqrt{2}x)}}\Bigg)_{j=1}^d \\
&= \left(\sum_{\nu\le k+e_j}\binom{k+e_j}{\nu} p^\varepsilon_{k+e_j-\nu}(\sqrt{2}x)\, p^\varepsilon_\nu(\sqrt{2}y)\right)_{j=1}^d.
\end{aligned}
$$

□

B.2 Generalized FBI transform

To simplify the presentation, we will use the following notation throughout this chapter:

Notation B.5. *We will abbreviate*

- $z := \begin{pmatrix} q \\ p \end{pmatrix} \in \mathbb{R}^{2d}, \qquad Z := \begin{pmatrix} Q \\ P \end{pmatrix} \in \mathbb{C}^{2d\times d}, \qquad \varphi^\varepsilon_k[z,Z] := \varphi^\varepsilon_k[q,p,Q,P],$
- $\Omega := \begin{pmatrix} 0 & -\mathrm{Id} \\ \mathrm{Id} & 0 \end{pmatrix} \in \mathbb{R}^{2d\times 2d},$
- $\Xi = \begin{pmatrix} 0 & PQ^* \\ -QP^* & 0 \end{pmatrix} \in \mathbb{C}^{2d\times 2d}.$

We are now ready to define the transforms we mentioned in the introduction of this chapter:

Definition B.6 (GFBI transform)**.** For an admissible matrix pair $Z = \begin{pmatrix} Q \\ P \end{pmatrix} \in \mathbb{C}^{2d\times d}$, $k \in \mathbb{N}^d$ and $\varepsilon > 0$, we define the kth *generalized FBI (GFBI) transform* via

$$(T^\varepsilon_{k,Z} f)(z) := (T^\varepsilon_{k,Q,P} f)(q,p) := (2\pi\varepsilon)^{-d/2} \langle \varphi^\varepsilon_k[z,Z], f\rangle_{L^2(\mathbb{R}^d)} . \tag{B.2.1}$$

Proposition B.7. The GFBI transforms $T^\varepsilon_{k,Z} f$ are well-defined isometries from $L^2(\mathbb{R}^d)$ to $L^2(R^{2d})$.

Proof. This is an application of [Com12, Proposition 4]. □

Remark B.8. If $Q = \mathrm{Id}$, $P = i\mathrm{Id}$, we regain the common FBI transform for $k = 0$.

Proposition B.9. The GFBI transform $T^\varepsilon_{k,Z}$ corresponding to the Hagedorn wavepacket $\varphi^\varepsilon_k[\bullet, Z]$ of another Hagedorn wavepacket $\varphi^\varepsilon_l[\zeta, Z]$ with phase space center $\zeta = (x,\xi)^T$ and the same admissible pair $Z = \begin{pmatrix} Q \\ P \end{pmatrix} \in \mathbb{C}^{2d\times d}$ is given by

$$\begin{aligned}\left(T^\varepsilon_{k,Z}\, \varphi^\varepsilon_l[\zeta,Z]\right)(z) &= (2\pi\varepsilon)^{-d/2} \langle \varphi^\varepsilon_k[z,Z], \varphi^\varepsilon_l[\zeta,Z]\rangle_{L^2(\mathbb{R}^d)} \\ &= \frac{e^{-\left(\|Z^*\Omega(z-\xi)\|^2 + (z-\xi)^T\Xi(z-\xi) - 4ip^T(q-x)\right)/4\varepsilon}}{\sqrt{2^{|k+l|+d} k!l!(\pi\varepsilon)^d}} \prod_{j=1}^d \mathcal{L}_{l_j,k_j}\left(\frac{i}{2\sqrt{\varepsilon}}[\overline{Z^*\Omega\tilde z}]_j\,,\, \frac{i}{2\sqrt{\varepsilon}}[Z^*\Omega\tilde z]_j\right).\end{aligned}$$

Proof. Using Proposition B.4(3) and abbreviating $p^\varepsilon_m := p^\varepsilon_m[0,Q]$, $\varphi^\varepsilon_0 := \varphi^\varepsilon_0[0,Z]$, $\tilde q := q - x$, $\tilde p := p - \xi$, $\tilde z := z - \zeta$, $C := PQ^{-1}$, $A := QQ^* = \Im[C]^{-1}$, we get

$$\begin{aligned}(\tilde p - \overline{C}\tilde q)^T A(\tilde p - \overline{C}\tilde q) &= \tilde p QQ^*\tilde p - 2\tilde p^T AC^*\tilde q + \tilde q^T C^* AC^*\tilde q \\ &= \tilde p QQ^*\tilde p - 2\tilde p^T QP^*\tilde q + \tilde q^T PP^*\tilde q - 2i\tilde q^T C^*\tilde q \\ &= (Q^*\tilde p - P^*\tilde q)^*(Q^*\tilde p - P^*\tilde q) + \tilde q^T PQ^*\tilde p - \tilde p^T QP^*\tilde q - 2i\tilde q^T C^*\tilde q \\ &= \|Z^*\Omega\tilde z)\|^2 + \tilde z^T\Xi\tilde z - 2i\tilde q^T C^*\tilde q\end{aligned}$$

and therefore

$$\begin{aligned}
&\frac{i}{2\varepsilon}\Big[y^TCy-(y-\tilde q)^T\overline{C}(y-\tilde q)+2\xi^Ty-2p^T(y-\tilde q)\Big]\\
&\qquad=-\frac{1}{\varepsilon}\left[y^TA^{-1}y-i\tilde q^T\overline{C}y+i\tilde p^Ty+\frac{i}{2}\tilde q^T\overline{C}\tilde q-ip^T\tilde q\right]\\
&\qquad=-\frac{1}{\varepsilon}\Bigg[\left(y+\frac{i}{2}A(\tilde p-\overline{C}\tilde q)\right)^TA^{-1}\left(y+\frac{i}{2}A(\tilde p-\overline{C}\tilde q)\right)\\
&\qquad\quad+\frac{1}{4}(\tilde p-\overline{C}\tilde q)^TA(\tilde p-\overline{C}\tilde q)+\frac{i}{2}\tilde q^T\overline{C}\tilde q-ip^T\tilde q\Bigg]\\
&\qquad=-\frac{1}{\varepsilon}\left[\left(y+\frac{i}{2}A(\tilde p-\overline{C}\tilde q)\right)^TA^{-1}\left(y+\frac{i}{2}A(\tilde p-\overline{C}\tilde q)\right)\right]\\
&\qquad\quad-\frac{1}{4\varepsilon}\Big[\|Z^*\Omega\tilde z)\|^2+\tilde z^T\Xi\tilde z-4ip^T\tilde q\Big].
\end{aligned}$$

Using these observations we get:

$$\begin{aligned}
&\langle\varphi_k^\varepsilon[q,p,Q,P],\varphi_l^\varepsilon[0,0,Q,P]\rangle\\
=&\frac{(\pi\varepsilon)^{-\frac{d}{2}}|\det Q|^{-1}}{\sqrt{2^{|k+l|}k!l!}}\int_{\mathbb{R}^d}\overline{p_k^\varepsilon}(y-q)\,p_l^\varepsilon(y-x)\;\times\\
&\qquad\qquad e^{\frac{i}{2\varepsilon}\left[(y-x)^TC(y-x)-(y-q)^T\overline{C}(y-q)+2\xi^T(y-x)-2p^T(y-q)\right]}\,\mathrm{d}y\\
=&\frac{(\pi\varepsilon)^{-\frac{d}{2}}|\det Q|^{-1}}{\sqrt{2^{|k+l|}k!l!}}\int_{\mathbb{R}^d}\overline{p_k^\varepsilon}(y-\tilde q)\,p_l^\varepsilon(y)\,e^{\frac{i}{2\varepsilon}\left[y^TCy-(y-\tilde q)^T\overline{C}(y-\tilde q)+2\xi^Ty-2p^T(y-\tilde q)\right]}\,\mathrm{d}y\\
=&\frac{1}{\sqrt{2^{|k+l|}k!l!}}\,\exp\left[-\frac{1}{4\varepsilon}\left(\|Z^*\Omega\tilde z\|^2+\tilde z^T\Xi\tilde z-4ip^T\tilde q\right)\right]\;\times\\
&\int_{\mathbb{R}^d}\overline{p_k^\varepsilon}\left(y-\frac{i}{2}A(\tilde p-\overline{C}\tilde q)-\tilde q\right)\,p_l^\varepsilon\left(y-\frac{i}{2}A(\tilde p-\overline{C}\tilde q)\right)\,|\varphi_0^\varepsilon(y)|^2\,\mathrm{d}y\\
=&\frac{1}{\sqrt{2^{|k+l|}k!l!}}\,\exp\left[-\frac{1}{4\varepsilon}\left(\|Z^*\Omega\tilde z\|^2+\tilde z^T\Xi\tilde z-4ip^T\tilde q\right)\right]\;\times\\
&\prod_{j=1}^d\mathcal{L}_{l_j,k_j}\left(-\frac{1}{\sqrt{\varepsilon}}\left[\overline{Q^{-1}}\left(\frac{i}{2}A(\tilde p-\overline{C}\tilde q)+\tilde q\right)\right]_j,\,-\frac{1}{\sqrt{\varepsilon}}\left[Q^{-1}\left(\frac{i}{2}A(\tilde p-\overline{C}\tilde q)\right)\right]_j\right)\\
=&\frac{e^{-\left[\|Z^*\Omega\tilde z\|^2+\tilde z^T\Xi\tilde z-4ip^T\tilde q\right]/4\varepsilon}}{\sqrt{2^{|k+l|}k!l!}}\prod_{j=1}^d\mathcal{L}_{l_j,k_j}\left(\frac{i}{2\sqrt{\varepsilon}}[\overline{Z^*\Omega\tilde z}]_j\,,\,\frac{i}{2\sqrt{\varepsilon}}[Z^*\Omega\tilde z]_j\right).
\end{aligned}$$

For the last step, we used the following observations:

$$\overline{Q^{-1}}QQ^*(\tilde p-\overline{C}\tilde q)=\overline{Q^{-1}QQ^*(\tilde p-C\tilde q)}=\overline{Q^*\tilde p-\underbrace{Q^*P}_{=P^*Q+2i\mathrm{Id}}Q^{-1}\tilde q}=-\overline{Z^*\Omega\tilde z}+2i\overline{Q^{-1}}\tilde q,$$

$$Q^*(\tilde p-\overline{C}\tilde q)=Q^*\tilde p-\underbrace{\overline{Q^TP}}_{=P^TQ}\,Q^{-1}\tilde q=-Z^*\Omega\tilde z.$$

□

Theorem B.10. For each $k \in \mathbb{N}^d$, the GFBI transform T_k^ε fulfills

$$|T_k^\varepsilon \psi|^2 = (W^\varepsilon \psi) * \sigma_k, \quad \text{where}$$
$$\sigma_k(x,\xi) = 2\kappa_2 e^{-\frac{1}{\varepsilon}\|Z^*\Omega\zeta\|^2} \sum_{\nu \le k} (-1)^{k-\nu} \left| \binom{k}{\nu} \left(\frac{2Z^*\Omega\zeta}{\sqrt{\varepsilon}} \right)^\nu \right|^2 .$$

Remark B.11. $\sigma_k \in \mathcal{S}(\mathbb{R}^{2d})$ decays exponentially in (x,ξ) because the kernel of the $\mathbb{R}$-linear map

$$(P^*, -Q^*) : \mathbb{R}^{2d} \to \mathbb{C}^d$$

is trivial (making it an isomorphism of $\mathbb{R}$-vector spaces!). In fact, for $(x,\xi) \in \mathbb{R}^{2d}$ we have

$$P^*x - Q^*\xi = 0 \quad \Longleftrightarrow \quad \xi = (Q^*)^{-1}P^*x = (PQ^{-1})^*x = \overline{C}x = \Re[C]x - i\Im[C]x\,.$$

Since $\Im[C] = (QQ^*)^{-1}$ is invertible, ξ can only be real-valued if $x = 0$ (which implies $\xi = 0$).

Definition B.12 (quadratic time-frequency representation, Cohen's class)**.** Let G be a sesquilinear form on $L^2(\mathbb{R}^d, \mathbb{C})$. Then we call

$$C\colon L^2(\mathbb{R}^d, \mathbb{C}) \to \mathbb{C}, \qquad Cf = G(f,f)$$

a *quadratic time-frequency representation*. If further for each $f \in L^2(\mathbb{R}^d, \mathbb{C})$

$$Cf = W^\varepsilon f * \sigma$$

for some $\sigma \in \mathcal{S}(\mathbb{R}^{2d})$, we say that C belongs to *Cohen's class*.

Remark B.13. For a discussion of quadratic time-frequency representations and Cohen's class see [Gro01, Chapter 4].

Corollary B.14. For each $k \in \mathbb{N}^d$, the modulus squared $|T_k^\varepsilon|^2$ of the GFBI transform T_k^ε belongs to Cohen's class.

Proof. $|T_k^\varepsilon|^2$ is a quadratic time-frequency representation, since

$$G_k(\psi_1, \psi_2) = (2\pi\varepsilon)^{-d} \overline{\langle \varphi_k, \psi_2 \rangle} \langle \varphi_k, \psi_1 \rangle$$

is sesquilinear and $|T_k^\varepsilon \psi|^2 = G_k(\psi, \psi)$. The claim follows from Theorem B.10. □

Proof of Theorem B.10. Denoting

$$A = \Im[C]^{-1} = QQ^* \qquad \text{and} \qquad R = \Re[C],$$

we get the following identities:

$$\begin{aligned}
PP^* &= CQQ^*C^* = (R + iA^{-1})A(R - iA^{-1}) = RAR + A^{-1}, && \text{(B.2.2)}\\
RA &= (C - iA^{-1})A = PQ^{-1}QQ^* - i\mathrm{Id} = PQ^* - i\mathrm{Id}, && \text{(B.2.3)}\\
AR &= A(C - iA^{-1}) = QQ^*PQ^{-1} - i\mathrm{Id} = Q(P^*Q + 2i\mathrm{Id})Q^{-1} - i\mathrm{Id} = QP^* + i\mathrm{Id}, && \text{(B.2.4)}\\
iQZ^*\Omega\zeta &= iQ(P^*x - Q^*\xi) = iA\xi - iQP^*x = iA\xi - i(AR - i\mathrm{Id})x = x + iA\xi - iARx, && \text{(B.2.5)}\\
\|Z^*\Omega\zeta\|^2 &= \|P^*x - Q^*\xi\|^2 = x^TPP^*x + \xi^TQQ^*\xi - x^TPQ^*\xi - \xi^TQP^*x\\
&= x^T(A^{-1} + RAR)x + \xi^TA^{-1}\xi - x^T(RA + i\mathrm{Id})\xi - \xi^T(AR - i\mathrm{Id})x && \text{(B.2.6)}\\
&= x^TA^{-1}x + x^TRARx + \xi^TA^{-1}\xi - 2x^TRA\xi.
\end{aligned}$$

Using the transformation $v_1 = x + \frac{y}{2},\ v_2 = x - \frac{y}{2}$ (i.e. $x = \frac{v_1+v_2}{2},\ y = v_1 - v_2$) and its application to the term

$$\begin{aligned}
&(v_1 - q)^T C\,(v_1 - q) - (v_2 - q)^T\,\overline{C}\,(v_2 - q)\\
&\quad= \left(x - q + \frac{y}{2}\right)^T C\left(x - q + \frac{y}{2}\right) - \left(x - q - \frac{y}{2}\right)^T\overline{C}\left(x - q - \frac{y}{2}\right)\\
&\quad= 2i\left((x - q)^T\,\frac{C - \overline{C}}{2i}\,(x - q) + \frac{1}{4}y^T\frac{C - \overline{C}}{2i}y - i\,(x - q)^T\,\frac{C + \overline{C}}{2}y\right)\\
&\quad= 2i\Bigg((x - q)^T\,\underbrace{\Im[C]}_{A^{-1}}\,(x - q) + \frac{1}{4}y^T\,\underbrace{\Im[C]}_{A^{-1}}\,y - i\,(x - q)^T\,\underbrace{\Re[C]}_{=R}\,y\Bigg),
\end{aligned}$$

we get

$$
\begin{aligned}
&2^k k!(2\pi\varepsilon)^d(\pi\varepsilon)^{d/2}\det Q\ \ ((W^\varepsilon\psi)*\sigma_k)\,(q,p)\\
&= 2^k k!(\pi\varepsilon)^{d/2}\det Q\ \int_{\mathbb{R}^{2d}}\sigma_k(q-x,p-\xi)\int_{\mathbb{R}^d}\overline{\psi}\left(x+\frac{y}{2}\right)\psi\left(x-\frac{y}{2}\right)e^{i\xi y/\varepsilon}\,\mathrm{d}y\,\mathrm{d}(x,\xi)\\
&\overset{(*)}{=}\int_{\mathbb{R}^{2d}}\exp\left[-\frac{1}{\varepsilon}\left((x-q)^TA^{-1}(x-q)+\frac{1}{4}y^TA^{-1}y-i(p+R(x-q))^Ty\right)\right]\\
&\qquad\qquad\times\ \overline{p}_k^\varepsilon\left(x-q-\frac{y}{2}\right)p_k^\varepsilon\left(x-q+\frac{y}{2}\right)\overline{\psi}\left(x+\frac{y}{2}\right)\psi\left(x-\frac{y}{2}\right)\,\mathrm{d}(x,y)\\
&= \int_{\mathbb{R}^{2d}}\exp\left[\frac{i}{2\varepsilon}\left((v_1-q)^TC(v_1-q)-(v_2-q)^T\overline{C}(v_2-q)+2p((v_1-q)-(v_2-q))\right)\right]\\
&\qquad\qquad\times\ \overline{p}_k^\varepsilon\,(v_2-q)\,p_k^\varepsilon\,(v_1-q)\,\overline{\psi}\,(v_1)\,\psi\,(v_2)\,\mathrm{d}(v_1,v_2)\\
&= \left|\int_{\mathbb{R}^d}\overline{p}_k^\varepsilon\,(v-q)\exp\left[-\frac{i}{2\varepsilon}\left((v-q)^T\overline{C}(v-q)+2p(v-q)\right)\right]\ \psi\,(v)\ \mathrm{d}v\right|^2\\
&= 2^k k!(\pi\varepsilon)^{d/2}\det Q\ \left|\langle\phi_k^\varepsilon[q,p,Q,P],\psi\rangle\right|^2\\
&= 2^k k!(2\pi\varepsilon)^d(\pi\varepsilon)^{d/2}\det Q\ \ \left|T_k^\varepsilon\psi\right|^2(q,p),
\end{aligned}
$$

where $(*)$ holds if and only if (here, $\kappa_1 := 2^k k!(\pi\varepsilon)^{d/2}\det Q$)

$$
\begin{aligned}
\kappa_1\int_{\mathbb{R}^d}\sigma_k(q-x,p-\xi)e^{i\xi y/\varepsilon}\,\mathrm{d}\xi \quad &= \quad e^{-\frac{1}{\varepsilon}\left((x-q)^TA^{-1}(x-q)+\frac{1}{4}y^TA^{-1}y-i(p+R(x-q))^Ty\right)}\\
&\qquad\times\ \overline{p}_k^\varepsilon\left(x-q-\frac{y}{2}\right)p_k^\varepsilon\left(x-q+\frac{y}{2}\right)\\
\Leftrightarrow\quad \kappa_1\int_{\mathbb{R}^d}\sigma_k(x,\xi)e^{i(p-\xi)y/\varepsilon}\,\mathrm{d}\xi \quad &= \quad e^{-\frac{1}{\varepsilon}\left(x^TA^{-1}x+\frac{1}{4}y^TA^{-1}y-i(p-Rx)^Ty\right)}\\
&\qquad\times\ \overline{p}_k^\varepsilon\left(x+\frac{y}{2}\right)p_k^\varepsilon\left(x-\frac{y}{2}\right)\\
\Leftrightarrow\quad \kappa_1\int_{\mathbb{R}^d}\sigma_k(x,\xi)e^{-i\xi y/\varepsilon}\,\mathrm{d}\xi \quad &= \quad e^{-\frac{1}{\varepsilon}\left(x^TA^{-1}x+\frac{1}{4}y^TA^{-1}y+ix^TRy\right)}\\
&\qquad\times\ \overline{p}_k^\varepsilon\left(x+\frac{y}{2}\right)p_k^\varepsilon\left(x-\frac{y}{2}\right)\\
\Leftrightarrow\quad \sigma_k(x,\xi) \quad &= \quad \frac{e^{-\frac{1}{\varepsilon}x^TA^{-1}x}}{\kappa_1}\,(\mathcal{F}_y^\varepsilon)^{-1}\Big[e^{-\frac{1}{\varepsilon}\left(\frac{1}{4}y^TA^{-1}y+ix^TRy\right)}\\
&\qquad\qquad\times\ \overline{p}_k^\varepsilon\left(x+\frac{y}{2}\right)p_k^\varepsilon\left(x-\frac{y}{2}\right)\Big](\xi).
\end{aligned}
$$

This means, the step $(*)$ is correct for

$$\begin{aligned}
\sigma_k(x,\xi) =& \underbrace{\frac{(2\pi\varepsilon)^{-d}}{\kappa_1}}_{=:\kappa_2} e^{-\frac{1}{\varepsilon}x^TA^{-1}x} \int_{\mathbb{R}^d} e^{-\frac{1}{\varepsilon}\left(\frac{1}{4}y^TA^{-1}y+ix^TRy-i\xi^Ty\right)} \\
&\qquad \times\ \overline{p}_k^\varepsilon\left(x+\frac{y}{2}\right) p_k^\varepsilon\left(x-\frac{y}{2}\right) \mathrm{d}y \\
=&\kappa_2 e^{-\frac{1}{\varepsilon}x^TA^{-1}x} \int_{\mathbb{R}^d} e^{-\frac{1}{\varepsilon}\left[\left(\frac{y}{2}+iARx-iA\xi\right)^TA^{-1}\left(\frac{y}{2}+iARx-iA\xi\right)+x^TRARx+\xi^TA\xi-2x^TRA\xi\right]} \\
&\qquad \times\ \overline{p}_k^\varepsilon\left(x+\frac{y}{2}\right) p_k^\varepsilon\left(x-\frac{y}{2}\right) \mathrm{d}y \\
=&2\kappa_2 e^{-\frac{1}{\varepsilon}\|Z^*\Omega\zeta\|^2} \int_{\mathbb{R}^d} e^{-\frac{1}{\varepsilon}z^TA^{-1}z} \\
&\qquad \times\ \overline{p}_k^\varepsilon\left(x-iARx+iA\xi+z\right) p_k^\varepsilon\left(x+iARx-iA\xi-z\right) \mathrm{d}z \\
=&2\kappa_2 e^{-\frac{1}{\varepsilon}\|Z^*\Omega\zeta\|^2} \sum_{\nu_1,\nu_2\le k} \binom{k}{\nu_1}\binom{k}{\nu_2} \left(\frac{2}{\sqrt{\varepsilon}}\overline{Q^{-1}(x+iARx-iA\xi)}\right)^{k-\nu_1} \\
&\times\ \left(\frac{2}{\sqrt{\varepsilon}}Q^{-1}(x+iARx-iA\xi)\right)^{k-\nu_2} \underbrace{\int_{\mathbb{R}^d} e^{-\frac{1}{\varepsilon}z^TA^{-1}z}\ \overline{p}_{\nu_1}^\varepsilon(z)\ p_{\nu_2}^\varepsilon(-z)\,\mathrm{d}z}_{(-1)^{|\nu_2|}\delta_{\nu_1,\nu_2}} \\
=&2\kappa_2 e^{-\frac{1}{\varepsilon}\|Z^*\Omega\zeta\|^2} \sum_{\nu\le k}(-1)^\nu \left|\binom{k}{\nu}\left(\frac{2iZ^*\Omega\zeta}{\sqrt{\varepsilon}}\right)^{k-\nu}\right|^2,
\end{aligned}$$

where we used the transformation $z=\frac{y}{2}+iARx-iA\xi$, equations (B.2.5) and (B.2.6) and the orthogonality relations of the polynomials p_k^ε. □

Table of Symbols

$\mathbb{K}$	field of real or complex numbers: $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$
$\|x\|_p$	ℓ^p-norm of a vector $x \in \mathbb{R}^d,\ 1 \leq d \leq \infty$
$\|x\| = \|x\|$	$:= \|x\|_2$ Euclidean norm of $x \in \mathbb{R}^d$
$\mathcal{B}(\Omega)$	Borel σ-algebra on a topological space Ω
$B_r(x)$	$:= \{y \in \mathbb{R}^d \mid \|y - x\| < r\}$ open ball of radius r centered at $x \in \mathbb{R}^d$
$\overline{B_r(x)}$	$:= \{y \in \mathbb{R}^d \mid \|y - x\| \leq r\}$ closed ball of radius r centered at $x \in \mathbb{R}^d$
$\operatorname{sgn}(x)$	$:= \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$ signum function
$\mathbb{1}$	vector in $\mathbb{R}^d$ with all entries equal to one
$C^k(\Omega, \tilde{\Omega})$	space of k times continuously differentiable functions with domain and codomain $\Omega, \tilde{\Omega} \subseteq \mathbb{R}^d$
$C^k(\Omega)$	$:= C^k(\Omega, \mathbb{R})$
$C_c^k(\Omega)$	$:= \left\{f \in C^k(\Omega, \mathbb{R}) \mid \operatorname{supp}(f) \text{ is compact}\right\}$ space of k times continuously differentiable functions with compact support
$L^p(\Omega, \mathbb{K})$	Lebesgue space with domain $\Omega \subseteq \mathbb{R}^d$ and codomain $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$
$L^p(\Omega)$	$:= L^p(\Omega, \mathbb{R})$
L^p	$:= L^p(\mathbb{R}^d, \mathbb{R})$
$\|f\|_p$	$= \|f\|_{L^p} = \|f\|_{L^p(\Omega)}$ L^p-norm of a function $f \in L^p(\Omega)$
$W^{k,p}(\Omega, \mathbb{K})$	Sobolev space of order k corresponding to the L^p-norm with domain $\Omega \subseteq \mathbb{R}^d$ and codomain $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$
$W^{k,p}(\Omega)$	$:= W^{k,p}(\Omega, \mathbb{R})$
$\mathcal{S}(\Omega, \mathbb{K})$	Schwartz space of rapidly decreasing functions on $\Omega \subseteq \mathbb{R}^d$ with codomain $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$

B. TABLE OF SYMBOLS

$\mathbb{P}_\rho$	probability distribution given by the density ρ: $\mathbb{P}_\rho(A) := \int_A \rho(x)\mathrm{d}x$ for $A \in \mathcal{B}(\mathbb{R}^d)$
$\mathbb{P}_{\text{uni}}$	uniform probability distribution on $(0,1)^d$
$\mathcal{N}(\mu, \Sigma)$	normal distribution with mean $\mu \in \mathbb{R}^d$ and positive definite covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ given by the probability density function $\rho(x) = (2\pi)^{-d/2}\lvert\det(\Sigma)\rvert^{-1/2}\exp\left[-\frac{1}{2}(x-\mu)^\intercal\Sigma^{-1}(x-\mu)\right]$
$\Phi_\#\mu$	pushforward measure of a probability distribution μ on $\mathcal{X}$ via a measurable map $\Phi\colon \mathcal{X} \to \mathcal{Y}$ ($(\mathcal{X}, \mathcal{B}_\mathcal{X})$ and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ being measurable spaces) defined by $(\Phi_\#\mu)(B) := \mu\left(\Phi^{-1}(B)\right)$ for all $B \in \mathcal{B}_\mathcal{Y}$
$Dv = D_x v$	Jacobian (in space) of a C^1 vector field $v : \mathbb{R}^d \to \mathbb{R}^d$
Id	Identity map of the considered space
E_d	d-dimensional identity matrix
e_j	$:= (\delta_{ij})_{i=1,\dots,d}$ unit vector in $\mathbb{R}^d$
$\mathrm{tr}(A)$	trace of a quadratic matrix A

Bibliography

[Amb04] L. Ambrosio. "Transport equation and Cauchy problem for BV vector fields". *Inventiones mathematicae*, Vol. 158, No. 2, pp. 227–260, 2004.

[Amb08a] L. Ambrosio. "Transport equation and Cauchy problem for non-smooth vector fields". In: *Calculus of variations and nonlinear partial differential equations*, pp. 1–41, Springer, 2008.

[Amb08b] L. Ambrosio, N. Gigli, G. Savaré, and M. Struwe. "Gradient Flows: in Metric Spaces and in the Space of Probability Measures". *Lectures in Mathematics ETH Zürich*, 2008.

[Boh52a] D. Bohm. "A suggested interpretation of the quantum theory in terms of" hidden" variables. I". *Physical Review*, Vol. 85, No. 2, p. 166, 1952.

[Boh52b] D. Bohm. "A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. II". *Phys. Rev.*, Vol. 85, pp. 180–193, 1952.

[Bor26] M. Born. "Zur Quantenmechanik der Stoßvorgänge". *Zeitschrift für Physik*, Vol. 38, No. 11-12, pp. 803–827, 1926.

[Box58] G. E. Box and M. E. Muller. "A note on the generation of random normal deviates". *The annals of mathematical statistics*, No. 29, pp. 610–611, 1958.

[Buh03] M. D. Buhmann, M. D. Buhmann, and M. D. Buhmann. *Radial basis functions: theory and implementations*. Vol. 5, Cambridge university press Cambridge, 2003.

[Com12] M. Combescure and D. Robert. *Coherent states and applications in mathematical physics*. Springer Science & Business Media, 2012.

BIBLIOGRAPHY

[Dec07] D.-A. Deckert, D. Dürr, and P. Pickl. "Quantum dynamics with Bohmian trajectories". *The Journal of Physical Chemistry A*, Vol. 111, No. 41, pp. 10325–10330, 2007.

[Dic10] J. Dick and F. Pillichshammer. "Discrepancy theory and quasi-Monte Carlo integration". *by WWL Chen, A. Srivastav, G. Travaglini (Springer, Berlin, 2012)*, 2010.

[DiP89] R. J. DiPerna and P.-L. Lions. "Ordinary differential equations, transport theory and Sobolev spaces". *Inventiones mathematicae*, Vol. 98, No. 3, pp. 511–547, 1989.

[Dur09] D. Dürr and S. Teufel. *Bohmian mechanics: the physics and mathematics of quantum theory*. Springer Science & Business Media, 2009.

[Fao09] E. Faou, V. Gradinaru, and C. Lubich. "Computing semiclassical quantum dynamics with Hagedorn wavepackets". *SIAM Journal on Scientific Computing*, Vol. 31, No. 4, pp. 3027–3041, 2009.

[Fas07a] G. E. Fasshauer and J. G. Zhang. "On choosing "optimal" shape parameters for RBF approximation". *Numerical Algorithms*, Vol. 45, No. 1-4, pp. 345–368, 2007.

[Fas07b] G. F. Fasshauer. *Meshfree approximation methods with MATLAB*. World Scientific Publishing Co., Inc., 2007.

[Fol89] G. B. Folland. *Harmonic analysis in phase space*. Princeton university press, 1989.

[Gre79] R. E. Greene and K. Shiohama. "Diffeomorphisms and volume-preserving embeddings of noncompact manifolds". *Transactions of the American Mathematical Society*, Vol. 255, pp. 403–414, 1979.

[Gro01] K. Gröchenig. *Foundations of time-frequency analysis*. Springer Science & Business Media, 2001.

[Gus11] S. J. Gustafson and I. M. Sigal. *Mathematical concepts of quantum mechanics*. Springer Science & Business Media, 2011.

[Hag80] G. A. Hagedorn. "Semiclassical quantum mechanics". *Communications in Mathematical Physics*, Vol. 71, No. 1, pp. 77–93, 1980.

[Hag98] G. A. Hagedorn. "Raising and lowering operators for semiclassical wave packets". *Annals of Physics*, Vol. 269, No. 1, pp. 77–104, 1998.

[Hel81] E. J. Heller. "Frozen Gaussians: a very simple semiclassical approximation". *The Journal of Chemical Physics*, Vol. 75, No. 6, pp. 2923–2931, 1981.

[Hil04] B. Hiley. "Phase space descriptions of quantum phenomena". In: *Proc. Int. Conf. Quantum Theory: Reconsideration of Foundations*, pp. 267–86, 2004.

[Hil97] M. Hillery, R. O'Connell, M. Scully, and E. P. Wigner. *Distribution functions in physics: fundamentals*. Springer, 1997.

[Kel14] J. Keller and C. Lasser. "Quasi-classical description of molecular dynamics based on Egorov's theorem". *The Journal of chemical physics*, Vol. 141, No. 5, p. 054104, 2014.

[Kor12] K. Kormann and E. Larsson. "An RBF-Galerkin approach to the time-dependent Schrödinger equation". Tech. Rep., Technical report 2012-024, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2012.

[Las14] C. Lasser and S. Troppmann. "Hagedorn wavepackets in time-frequency and phase space". *Journal of Fourier Analysis and Applications*, Vol. 20, No. 4, pp. 679–714, 2014.

[Lub08] C. Lubich. *From quantum to classical molecular dynamics: reduced models and numerical analysis*. European Mathematical Society, 2008.

[Mad27] E. Madelung. "Quantentheorie in hydrodynamischer Form". *Zeitschrift für Physik A Hadrons and Nuclei*, Vol. 40, No. 3, pp. 322–326, 1927.

[Maz07] V. Maz'ya and G. Schmidt. *Approximate approximations*. 2007.

[Mos65] J. Moser. "On the volume elements on a manifold". *Transactions of the American Mathematical Society*, pp. 286–294, 1965.

BIBLIOGRAPHY

[Moy49] J. E. Moyal. "Quantum mechanics as a statistical theory". In: *Mathematical Proceedings of the Cambridge Philosophical Society*, pp. 99–124, Cambridge Univ Press, 1949.

[Nie92] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods.* Vol. 63, SIAM, 1992.

[Ree75] M. Reed and B. Simon. *Methods of modern mathematical physics. 2. Fourier analysis, self-adjointness.* Vol. 2, Elsevier, 1975.

[Sch26] E. Schrödinger. "Quantisierung als Eigenwertproblem". *Annalen der Physik*, Vol. 385, No. 13, pp. 437–490, 1926.

[Tak54] T. Takabayasi. "The formulation of quantum mechanics in terms of ensemble in phase space". *Progress of Theoretical Physics*, Vol. 11, No. 4-5, pp. 341–373, 1954.

[Teu05] S. Teufel and R. Tumulka. "Simple proof for global existence of Bohmian trajectories". *Communications in mathematical physics*, Vol. 258, No. 2, pp. 349–365, 2005.

[Vil08] C. Villani. *Optimal transport: old and new.* Vol. 338, Springer, 2008.

[Wen05] H. Wendland. *Scattered data approximation.* Cambridge University Press, 2005.

[Wya06] R. E. Wyatt. *Quantum dynamics with trajectories: introduction to quantum hydrodynamics.* Vol. 28, Springer Science & Business Media, 2006.