

Adaptive semi-Lagrangian schemes for Vlasov equations

Martin Campos Pinto

Abstract. This lecture presents a new class of adaptive semi-Lagrangian schemes – based on performing a semi-Lagrangian method on adaptive interpolation grids – in the context of solving Vlasov equations with underlying "smooth" flows, such as the one-dimensional Vlasov-Poisson system. After recalling the main features of the semi-Lagrangian method and its error analysis in a uniform setting, we describe two frameworks for implementing adaptive interpolations, namely multilevel meshes and interpolatory wavelets. For both discretizations, we introduce a notion of good adaptivity to a given function and show that it is preserved by a low-cost prediction algorithm which transports multilevel grids along any "smooth" flow. As a consequence, error estimates are established for the resulting *predict and readapt* schemes under the essential assumption that the flow underlying the transport equation, as well as its approximation, is a stable diffeomorphism. Some complexity results are stated in addition, together with a conjecture of the convergence rate for the overall adaptive scheme. As for the wavelet case, these results are new and also apply to high-order interpolation.

Keywords. Fully adaptive scheme, semi-Lagrangian method, Vlasov equation, smooth characteristic flows, adaptive mesh prediction, error estimates, interpolatory wavelets.

AMS classification. 65M12, 65M50, 82D10.

1 Introduction

In this lecture we shall describe adaptive numerical methods for approximating Vlasov equations, i.e., kinetic equations which model in statistical terms the nonlinear evolution of a collisionless plasma. In order to give the reader a specific example, we shall somehow focus our presentation on the one-dimensional Vlasov-Poisson system

$$\partial_t f(t,x,v) + v \cdot \partial_x f(t,x,v) + E(t,x) \cdot \partial_v f(t,x,v) = 0, \quad t > 0, \quad x, v \in \mathbb{R},$$
$$\partial_x E(t,x) = \int_{\mathbb{R}} f(t,x,v)\,\mathrm{d}v - 1, \quad f(0,\cdot,\cdot) = f^0,$$

but we emphasize that our results actually apply to any nonlinear transport problem associated with a smooth characteristic flow, "smooth" meaning here that the flow is a Lipschitz diffeomorphism, see in particular Assumptions 3.2 and 3.3 below.

In order to save computational resources while approximating the complex and thin structures that may appear in the solutions as time evolves, several adaptive schemes have been proposed in the past few years, see e.g. [3, 6, 7, 20], all based on the semi-Lagrangian method originally introduced by Chio-Zong Cheng and Georg Knorr [9] and later revisited by Eric Sonnendrücker, Jean Rodolphe Roche, Pierre Bertrand and Alain Ghizzo [23]. A common feature of these new schemes lies in the multilevel, tree-structured discretization of the phase space, and a central issue appears to be the *prediction strategy* for generating adaptive meshes that are optimal, i.e., that only retain the "necessary" grid points.

Based on the regularity analysis of the numerical solution and how it gets transported by the numerical flow, it has recently been shown in [7] that a low-cost prediction strategy could achieve an accurate evolution of the adaptive multilevel meshes from one time step to the other, in the sense that the overall accuracy of the scheme was monitored by a prescribed tolerance parameter ε representing the local interpolation error at each time step. In this lecture we shall follow the same approach and propose new algorithms for high-order wavelet-based schemes, together with error estimates.

The lecture is organized as follows. We shall start by presenting the Vlasov-Poisson system in Section 2 and recall some important properties satisfied by its classical solutions. The backward semi-Lagrangian method is then described in Section 3, together with the main points of its error analysis, which yields a first convergence rate in the case where the discretization is uniform. We detail next two distinct frameworks for generating adaptive multilevel discretizations, namely adaptive meshes in Section 4 and interpolatory wavelets in Section 5. Finally, Section 6 is devoted to describing algorithms that build multilevel grids which either are well-adapted to given functions or *remain* well-adapted to transported functions. More precisely, for both discretizations we introduce a notion of ε-adaptivity to a given function, and show that it is preserved by a low-cost prediction algorithm which transports multilevel grids along any smooth flow. We end by proving new error estimates for the resulting semi-Lagrangian schemes (with arbitrary interpolation order, under a central assumption on the approximate flow) and state some partial complexity results.

In these notes we shall often write $A \lesssim B$ to express that $A \leq CB$ holds with a constant C independent of the parameters involved in the inequality, $A \sim B$ meaning that both $A \lesssim B$ and $B \lesssim A$ hold. As often as possible, we will nevertheless indicate in the text the parameters on which these "invisible" constants may depend.

2 The Vlasov-Poisson system

The Vlasov system, as introduced by Anatoliĭ Aleksandrovich Vlasov [24] in the late 1930s, describes in statistical terms the time evolution of rarefied plasmas, which are gases of charged particles such as ions and electrons. In this section we recall the form of the system as well as its interpretation, and describe some of its properties. Although we shall mention the general equations in $d = 3$ physical dimensions, this lecture will essentially focus on a simplified model in $d = 1$ physical dimension, leading to a two-dimensional transport problem in the phase space. Most of our algorithms and proofs, however, can be generalized to higher dimensions with no particular difficulty.

2.1 Description of the model

In the Vlasov model, the state of every species $\mathcal{E}$ of charged particles in the plasma is represented at time t by a *density function* $f_{\mathcal{E}}(t)$ defined in the phase space, i.e., the subset of $\mathbb{R}^d \times \mathbb{R}^d$ which contains every possible position x and speed v. In particular, the number of $\mathcal{E}$-type particles that are located at time t in a physical domain $\omega_x \subset \mathbb{R}^d$, with speed $v \in \omega_v \subset \mathbb{R}^d$, reads

$$Q_{\mathcal{E}}(t,\omega) = \iint_{\omega_x \times \omega_v} f_{\mathcal{E}}(t,x,v)\,\mathrm{d}x\,\mathrm{d}v. \tag{2.1}$$

For the sake of simplicity we shall assume that the effect of the magnetic field can be neglected – which corresponds to an electrostatic approximation – and consider only two species, namely:

- positive (and heavy) ions, assumed to be uniformly distributed in space and time. Their density function will be denoted by $f_p(t,x,v) = f_p(v)$, and we shall assume that it is normalized, i.e., $\int f_p(v)\,\mathrm{d}v = 1$;
- much lighter electrons. Their density function f is the main unknown of the model.

In dimension $d = 1$, the Vlasov-Poisson system reads as follows.

$$\partial_t f(t,x,v) + v \cdot \partial_x f(t,x,v) + E(t,x) \cdot \partial_v f(t,x,v) = 0, \tag{2.2}$$

$$\partial_x E(t,x) = \int_{\mathbb{R}} f(t,x,v)\,\mathrm{d}v - 1, \tag{2.3}$$

where E represents the normalized, self-consistent electric field. In order to consider the corresponding Cauchy problem we supplement (2.2)-(2.3) with an initial condition

$$f(0,\cdot,\cdot) = f^0 \tag{2.4}$$

assumed to be smooth (at least continuous) and compactly supported in $\mathbb{R}^2$.

2.2 Physical interpretation and characteristic flows

Since the end of the 18th century and the works of Charles-Augustin Coulomb, it is known that a closed system of charged particles $\{q_i \in \mathbb{R}, x_i(t) \in \mathbb{R}^3\}_{i\in\mathcal{I}}$ is subject to binary interactions – the so-called Coulomb interactions – yielding an N-body problem. A less expensive model follows by considering the electromagnetic field which, according to the theory that James Clerk Maxwell developed one century later, is created by the particles and simultaneously influences their motion through the Lorentz force. More precisely, if the associated charge and current densities are denoted by

$$\rho(t,x) := \sum_{i\in\mathcal{I}} q_i \delta_{\{x_i(t)\}}(x) \quad \text{and} \quad \jmath(t,x) := \sum_{i\in\mathcal{I}} v_i(t) q_i \delta_{\{x_i(t)\}}(x),$$

respectively, where $\delta_{x_i(t)}$ stands for the Dirac mass located at $x_i(t)$ and $v_i(t) := x_i'(t)$ is the particle's speed, then the electromagnetic field $(E,B)(t,x)$ created by the particles satisfies the following *Maxwell system*:

$$\nabla_x \cdot E = \rho/\varepsilon_0, \tag{2.5}$$

$$\nabla_x \times E = -\partial_t B, \tag{2.6}$$

$$\nabla_x \cdot B = 0, \tag{2.7}$$

$$\nabla_x \times B = \mu_0(\jmath + \varepsilon_0 \partial_t E). \tag{2.8}$$

Here ∇_x is the differential operator $(\partial_{x_1}, \partial_{x_2}, \partial_{x_3})$, $\times$ is the vector product in $\mathbb{R}^3$, and ε_0, μ_0 denote the permittivity and permeability constants. In turn, every particle is subject to the *Lorentz force*

$$F_i(t) = q_i[E(x_i(t)) + v_i(t) \times B(t, x_i(t))]. \tag{2.9}$$

According to Isaac Newton's fundamental law of dynamics $m_i v_i' = F_i$, it follows that every particle has a phase space trajectory that is a solution to the following ordinary differential equation:

$$x_i'(t) = v_i(t), \quad v_i'(t) = \frac{q_i}{m_i}[E(x_i(t)) + v_i(t) \times B(t, x_i(t))]. \tag{2.10}$$

The kinetic Vlasov model corresponds to a *continuous limit* of the above mean field model when the particles are so many that each species can be represented by a smooth, e.g. continuous, density function $f_{\mathcal{E}}$. In such a case, charge and current densities are defined as

$$\rho(t,x) := \sum_{\mathcal{E}} q_{\mathcal{E}} \int_{v\in\mathbb{R}^3} f_{\mathcal{E}}(t,x,v)\mathrm{d}v \quad \text{and} \quad \jmath(t,x) := \sum_{\mathcal{E}} q_{\mathcal{E}} \int_{v\in\mathbb{R}^3} v f_{\mathcal{E}}(t,x,v)\mathrm{d}v,$$

and the electromagnetic field is again given by Maxwell's system (2.5)-(2.8). As was written above, we shall only consider two species $\mathcal{E}$, namely positive ions and electrons. The so-called Vlasov-Poisson system – which corresponds to an electrostatic approximation, i.e., the effects of the magnetic field are neglected – can then be obtained by writing the pointwise conservation of the electron density f along the *characteristic curves*, which are defined as the solutions

$$t \mapsto (X(t), V(t)) = (X(t; s, x, v), V(t; s, x, v)) \tag{2.11}$$

to the ordinary differential system

$$\partial_t X(t) = V(t), \quad \partial_t V(t) = E(t, X(t)), \quad (X,V)(s) = (x,v). \tag{2.12}$$

Note that this is a natural extension of the trajectories (2.10) in the continuous limit. By writing that f satisfies

$$\partial_t f(t, X(t; 0, x, v), V(t; 0, x, v)) = 0 \tag{2.13}$$

for any (x, v) in the phase space, one indeed finds

$$\partial_t f(t,x,v) + v \cdot \nabla_x f(t,x,v) + E(t,x) \cdot \nabla_v f(t,x,v) = 0, \tag{2.14}$$

with an obvious definition for ∇_v. Because E has then a vanishing curl, it derives from an electric potential ϕ, i.e.,

$$E = -\nabla_x \phi, \tag{2.15}$$

and equation (2.5) is equivalent to the Poisson equation

$$\Delta_x \phi = -\frac{\rho}{\varepsilon_0}, \tag{2.16}$$

where $\Delta_x \equiv \nabla_x \cdot \nabla_x$ classically denotes the Laplace operator. Equations (2.14)-(2.16) form the Vlasov-Poisson system in three physical dimensions, from which the simplified model (2.2)-(2.3) easily derives (with normalized constants) in one dimension.

Now, as long as E is bounded and continuously differentiable with respect to x, it is known that the associated *characteristic flow*

$$\mathcal{F}_{t,s} : (x,v) \mapsto (X,V)(t; s, x, v) \tag{2.17}$$

is a measure preserving C^1-diffeomorphism with inverse $\mathcal{F}_{t,s}^{-1} = \mathcal{F}_{s,t}$, see e.g. [22]. Let us then notice that equation (2.13), which expresses the pointwise transport of f along the characteristic curves, also yields a local transport property for the charge density, in the sense that

$$\iint_\omega f(0,x,v)\,\mathrm{d}x\,\mathrm{d}v = \iint_{\mathcal{F}_{t,s}(\omega)} f(t,x,v)\,\mathrm{d}x\,\mathrm{d}v \quad \text{for any } \omega \subset \mathbb{R}^{2d}, \tag{2.18}$$

which is equivalent to saying that the flow is measure preserving. This property can be seen as a consequence of the vanishing divergence of the field $(x,v) \mapsto (v, E(t,x))$. Indeed, it first allows to write (2.2) in a conservative form, i.e.,

$$\partial_t f(t,x,v) + \nabla_{x,v} \cdot [(v, E(t,x)) f(t,x,v)] = 0. \tag{2.19}$$

Second, introducing the divergence-free field $\Phi(t,x,v) := (1, v, E(t,x))$ defined on $\mathcal{D}_{\tau,\omega} = \{(t,x,v) : t \in [0,\tau], \mathcal{F}_{t,s}^{-1}(x,v) \in \omega\} \subset [0,\tau] \times \mathbb{R}^{2d}$, a Stokes formula gives

$$\iint_{\mathcal{F}_{\tau,s}(\omega)} \mathrm{d}x\,\mathrm{d}v - \iint_\omega \mathrm{d}x\,\mathrm{d}v = \iint_{\partial \mathcal{D}_{\tau,\omega}} \Phi \cdot n\,\mathrm{d}\sigma = \iiint_{\mathcal{D}_{\tau,\omega}} \nabla_{t,x,v} \cdot \Phi\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}v = 0,$$

where n denotes the outward unit vector normal to $\partial\mathcal{D}_{\tau,\omega}$. Here the first equality comes from the fact that the boundary of $\mathcal{D}_{\tau,\omega}$ is parallel to the field lines of Φ outside the "faces" $\mathcal{F}_{\tau,s}(\omega)$ and ω. As the latter equality precisely means that $\mathcal{F}_{t,s}$ preserves the Lebesgue measure, we see that its Jacobian is equal to one. In particular, property (2.18) readily follows from equation (2.13).

2.3 Existence of smooth solutions

According to the previous section, we shall say that (f, E) is a classical solution of the Vlasov-Poisson system (2.2)-(2.4) if

(i) f is continuous

(ii) E is continuously differentiable

(iii) the Vlasov equation (2.2) is satisfied in the sense of distributions.

Now, because the characteristic curves are defined on any point (x, v) of the phase space, condition (iii) is equivalent to

(iii)' f is constant along the characteristic curves defined by (2.12).

Remark 2.1. If f is only continuous, the derivatives appearing in (2.2) must be understood in the (weak) sense of distributions. Nevertheless, the solution is said to be classical (or strong), because the characteristic curves are well-defined, and hence equation (2.13) is satisfied in a classical sense.

One of the first results is due to Sergei Iordanskii (see [21]), who has proven in the early 1960s global existence in time and uniqueness of classical solutions under certain conditions. First, the initial datum f^0 must be continuous, and it must be integrable in the following sense:

$$\rho^0(x) = \int_{\mathbb{R}} f^0(x, v)\,\mathrm{d}v - 1 < \infty \quad \text{and} \quad \int_{\mathbb{R}} v^2\theta(v)\,\mathrm{d}v < \infty, \tag{2.20}$$

where θ is a non-increasing function of $|v|$ that dominates f^0 and f_p (the density of the positive ions). Second, the electric field must satisfy the limit condition

$$\lim_{x\to-\infty} E(t, x) = 0, \quad t > 0, \tag{2.21}$$

which is a reasonable condition since E is only defined up to a constant. Note that the assumptions (2.20) are rather natural, since they allow to define the current and kinetic energy densities,

$$\jmath(t, x) := \int_{\mathbb{R}} v[f(t, x, v) - f_p(v)]\,\mathrm{d}v \quad \text{and} \quad \varepsilon_k(t, x) := \int_{\mathbb{R}} v^2[f(t, x, v) + f_p(v)]\,\mathrm{d}v,$$

respectively, which are two fundamental physical quantities. Finally, Iordanskii shows that f and E satisfy an additional equation, namely Ampère's equation

$$\partial_t E(t, x) = \int_{\mathbb{R}} v[f_p(v) - f(t, x, v)]\,\mathrm{d}v. \tag{2.22}$$

In 1980, Jeffery Cooper and Alexander Klimas [13] have extended these results to more general limit conditions than (2.21), in particular they have addressed the periodic case where $x \in \mathbb{R}/\mathbb{Z}$.

From now on, we shall only consider this case, moreover our initial data will always be assumed to have a compact support in the phase space

$$\Omega_\infty := (\mathbb{R}/\mathbb{Z}) \times \mathbb{R}. \tag{2.23}$$

Their result is the following.

Theorem 2.2. *If f^0 is continuous on Ω_∞ (hence 1-periodic with respect to x), if it satisfies*

$$\rho^0(x) := \int_{\mathbb{R}} f^0(x,v)\,\mathrm{d}v - 1 < \infty \quad \textit{and} \quad \int_{\mathbb{R}} |v|\theta(v)\,\mathrm{d}v < \infty, \tag{2.24}$$

where θ is defined as in (2.20)*, and if the plasma is globally neutral, i.e.,*

$$\int_0^1 \rho^0(x)\,\mathrm{d}x = \int_0^1 \int_{\mathbb{R}} f^0(x,v)\,\mathrm{d}v\,\mathrm{d}x - 1 = 0, \tag{2.25}$$

then there exists a unique classical solution to (2.2)-(2.4) *such that $\int_0^1 E(0,x)\,\mathrm{d}x = 0$, moreover this solution is 1-periodic with respect to x.*

Notice that the global neutrality, namely (2.25) at time $t = 0$, and

$$\iint_{\Omega_\infty} f(t,x,v)\,\mathrm{d}x\,\mathrm{d}v = \iint_{\Omega_\infty} f^0(x,v)\,\mathrm{d}x\,\mathrm{d}v = 1 \tag{2.26}$$

for positive times (which follows from the transport properties of f) is equivalent to the continuity of the 1-periodic field $E(t,\cdot)$. Now, it is possible to give an analytical expression for E: indeed if we denote by $-G(x,y)$ the Green function associated with the one-dimensional Poisson equation (2.3), defined in such a way that

$$\partial^2_{xx} G(\cdot,y) = \delta(\cdot - y) \quad \text{on} \quad (0,1) \tag{2.27}$$

holds for any $y \in (0,1)$ with periodic boundary conditions $G(0,y) = G(1,y)$, then E reads

$$E(t,x) = \int_0^1 K(x,y) \left(\int_{\mathbb{R}} f(t,y,v)\,\mathrm{d}v - 1 \right) \mathrm{d}y \tag{2.28}$$

with

$$K(x,y) = \partial_x G(x,y) = \begin{cases} y - 1 & \text{if } \ 0 < x < y, \\ y & \text{if } \ y \le x < 1. \end{cases} \tag{2.29}$$

In order to study later the accuracy of the numerical schemes, we now state some smoothness estimates for f and E. In general, it is known that any initial order of smoothness is preserved by the equation, see e.g. [14]. Since the analysis is simple in our case, we give a detailed proof for the following estimates, inspired by the techniques presented in the book of Robert Glassey [19]. Here and below we shall rely on usual notations for Sobolev spaces, see e.g. [1].

Lemma 2.3. *If f^0 belongs to $W^{1,\infty}(\Omega_\infty)$ and satisfies conditions* (2.24)-(2.25)*, then for any final time $T < \infty$, the solution f is compactly supported in the v-variable, i.e.,*

$$\Sigma_v(t) := \sup\{|v| : \exists x \in \mathbb{R}/\mathbb{Z}, f(t,x,v) > 0\} \le \Sigma_v(0) + 2T, \quad t \le T, \tag{2.30}$$

and satisfies the following smoothness estimates:

$$\|f(t,\cdot,\cdot)\|_{W^{1,\infty}(\Omega_\infty)} \le C \tag{2.31}$$

$$\|\partial_t f(t,\cdot,\cdot)\|_{L^\infty(\Omega_\infty)} \le C \tag{2.32}$$

$$\|E(t,\cdot)\|_{W^{2,\infty}(0,1)} \le C \tag{2.33}$$

$$\|\partial_t E(t,\cdot)\|_{W^{1,\infty}(0,1)} \le C \tag{2.34}$$

$$\|\partial^2_{tt} E(t,\cdot)\|_{L^\infty(0,1)} \le C \tag{2.35}$$

for all $t \in (0,T)$, with a constant $C > 0$ depending on f^0 and T only.

Proof. Let us first show the weaker assertion that

$$\sup_{t\in(0,T)} \|E(t,\cdot)\|_{W^{1,\infty}(0,1)} \le C \quad \text{and} \quad \sup_{t\in(0,T)} \|\partial_t E(t,\cdot)\|_{L^\infty(0,1)} \le C \tag{2.36}$$

hold as long as f^0 is continuous: Indeed the conservation of f along the characteristic curves (2.11) yields a maximum principle

$$0 \le f \le \|f^0\|_{L^\infty(\Omega_\infty)}, \tag{2.37}$$

and a bounded support in the v-direction, i.e., for all $t \in [0,T]$,

$$\Sigma_v(t) - \Sigma_v(0) \le \sup_{(x,v)\in\Omega_\infty} \int_0^T |\partial_t V(\tau;0,x,v)| \,\mathrm{d}\tau \le T\|E\|_{L^\infty((0,T)\times(0,1))}. \tag{2.38}$$

By using (2.28), (2.37) and (2.26), it follows that for all t, we have

$$\|E(t,\cdot)\|_{L^\infty(0,1)} \le \|K\|_{L^\infty} \left(\iint_{\Omega_\infty} |f(t,x,v)| \,\mathrm{d}x\,\mathrm{d}v + 1 \right) \le 2. \tag{2.39}$$

According to (2.38), the above inequality yields (2.30). We also have

$$\|\partial_x E(t,\cdot)\|_{L^\infty(0,1)} \le \Sigma_v(t)\|f^0\|_{L^\infty(\Omega_\infty)} + 1 \tag{2.40}$$

by using the Poisson equation (2.3), and

$$\|\partial_t E(t,\cdot)\|_{L^\infty(0,1)} \le \Sigma_v(t)^2 \|f^0\|_{L^\infty(\Omega_\infty)} + \int_{\mathbb{R}} v f_p(v) \,\mathrm{d}v$$

by using the Ampère equation (2.22), which establishes both inequalities in (2.36). If we now assume $f^0 \in W^{1,\infty}(\Omega_\infty)$, we can write

$$\begin{aligned} |f(t,x,v) - f(t,\tilde{x},\tilde{v})| &= |f^0(X_0(t),V_0(t)) - f^0(\tilde{X}_0(t),\tilde{V}_0(t))| \\ &\le |f^0|_{W^{1,\infty}(\Omega_\infty)} \big(|e_x(t)| + |e_v(t)|\big) \end{aligned}$$

where for any s such that $0 \le s \le t \le T$, we have denoted

$$\begin{cases} (X_0, V_0)(s) := (X, V)(t-s; t, x, v) \\ (\tilde{X}_0, \tilde{V}_0)(s) := (X, V)(t-s; t, \tilde{x}, \tilde{v}) \end{cases} \text{and} \quad \begin{cases} e_x(s) := X_0(s) - \tilde{X}_0(s) \\ e_v(s) := V_0(s) - \tilde{V}_0(s) \end{cases}. \tag{2.41}$$

By using the equations (2.12), we see that these quantities satisfy $e_x'(s) = e_v(s)$ and $e_v'(s) = E(t-s, X_0(s)) - E(t-s, \tilde{X}_0(s))$. The first bound in (2.36) thus yields

$$|e_x'(s)| + |e_v'(s)| \lesssim \big(|e_x(s)| + |e_v(s)|\big). \tag{2.42}$$

In particular, the function $\psi(s) := |e_x(s)| + |e_v(s)|$ satisfies

$$\psi(t) = \psi(0) + \Big|\int_0^t e_x'(s)\,\mathrm{d}s\Big| + \Big|\int_0^t e_v'(s)\,\mathrm{d}s\Big| \le \psi(0) + C(T)\int_0^t \psi(s)\,\mathrm{d}s, \tag{2.43}$$

and by applying the Gronwall lemma we find

$$\big(|e_x(t)| + |e_v(t)|\big) \lesssim \big(|e_x(0)| + |e_v(0)|\big) \lesssim \big(|x-\tilde{x}| + |v-\tilde{v}|\big) \tag{2.44}$$

with constants depending only on T, f^0. This shows that $\sup_{t\in(0,T)} \|f(t,\cdot,\cdot)\|_{W^{1,\infty}(\Omega_\infty)}$ is bounded by a constant that only depends on T and f^0, and we note that for all $t \in (0,T)$, the bound

$$\|\partial_t f(t,\cdot,\cdot)\|_{L^\infty(\Omega_\infty)} \le \big(Q(T) + \|E(t,\cdot)\|_{L^\infty(0,1)}\big)\|f(t,\cdot,\cdot)\|_{W^{1,\infty}(\Omega_\infty)}$$

where $Q(T) := \sup_{t\in(0,T)} \Sigma_v(t)$, follows from the Vlasov equation (2.2). Let us now turn to the electric field: By differentiating the Poisson equation (2.3) with respect to x and t, we respectively find

$$\|\partial_{xx}^2 E\|_{L^\infty((0,T)\times(0,1))} \le Q(T)\|\partial_x f\|_{L^\infty((0,T)\times\Omega_\infty)}, \tag{2.45}$$

$$\|\partial_{tx}^2 E\|_{L^\infty((0,T)\times(0,1))} \le Q(T)\|\partial_t f\|_{L^\infty((0,T)\times\Omega_\infty)}, \tag{2.46}$$

and the seminorm $\|\partial_{tt}^2 E\|_{L^\infty((0,T)\times(0,1))}$ is easily bounded by differentiating the Ampère equation (2.22) with respect to t. □

3 The backward semi-Lagrangian method

Based on the pointwise transport property (2.13), the semi-Lagrangian approach consists in combining a transport and a projection operator within every time step, as in

$$f_{n+1} := P\mathcal{T} f_n,$$

where $f_n \approx f(t_n)$ denotes the numerical solution. More precisely, the schemes that we will consider in this lecture decompose as follows (see e.g. [9] or [23]):

(i) given f_n, approach the exact backward flow $\mathcal{F}_{t_n,t_{n+1}}$, see (2.17), by some computable diffeomorphism $\mathcal{B}[f_n]$,

(ii) define an intermediate solution by transporting f_n along this approximate flow

$$\mathcal{T} f_n := f_n \circ \mathcal{B}[f_n], \tag{3.1}$$

(iii) obtain f_{n+1} by interpolating the intermediate $\mathcal{T} f_n$, for instance on the nodes of some triangulation $\mathcal{K}$.

Following the same principle, the *adaptive* semi-Lagrangian approach consists in interpolating $\mathcal{T} f_n$ on an adaptive grid of the phase space. A major issue resides then in transporting the grid along the flow in a such way that both its *sparsity* and its *accuracy* are guaranteed.

From now on, we shall assume that every solution is supported in $\Omega := (0,1)^2$. According to Lemma 2.3, we know that this holds true for sufficiently small supports of the initial datum f^0. Moreover we shall consider a uniform discretization involving N time steps and denote $\Delta t := T/N$ and $t_n := n\Delta t$ for $n = 0, \dots N$.

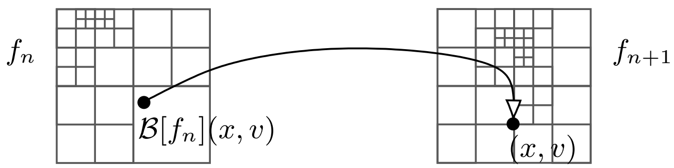

Figure 1. The backward semi-Lagrangian method (here with adaptive meshes).

Remark 3.1 (Computational cost). Like the numerical flow, the intermediate solution $\mathcal{T} f_n$ is computable everywhere, but only is *computed* on the interpolation grid corresponding to f_{n+1}. Hence if the latter is interpolated on a triangulation $\mathcal{K}$, the computational cost of one iteration is of the order $C\#(\mathcal{K})$, where C is the cost of applying the approximate flow $\mathcal{B}[f_n]$ to one point $(x,v) \in \Omega$.

3.1 Approximation of smooth flows

We mentioned before that our main results apply to any transport problem with an underlying smooth flow. Besides, we will need that the approximate flow is also smooth, and, moreover, stable and accurate in order to describe and further analyze our adaptive schemes. Let us formulate these properties as assumptions, for later reference.

Assumption 3.2. Let $s < t$ be arbitrary instants in $[0,T]$. The characteristic (backward) flow underlying the Vlasov equation, i.e., the mapping $\mathcal{F}_{s,t}$ for which we have

$$f(t,x,v) = f\big(s, \mathcal{F}_{s,t}(x,v)\big), \qquad x, v \in \Omega,$$

is a diffeomorphism from Ω into itself, i.e., it satisfies (with $\mathcal{B} = \mathcal{F}_{s,t}$)

$$\|\mathcal{B}(x+h) - \mathcal{B}(x)\| \le L_{\mathcal{B}}\|h\| \quad \text{and} \quad \|\mathcal{B}^{-1}(x+h) - \mathcal{B}^{-1}(x)\| \le L_{\mathcal{B}^{-1}}\|h\| \tag{3.2}$$

for all x, $x+h \in \Omega$, and with constants $L_{\mathcal{B}}$, $L_{\mathcal{B}^{-1}}$ independent of s, t (here and below, $\|\cdot\|$ denotes the Euclidean norm of $\mathbb{R}^2$).

Assumption 3.3. We are given a scheme $\mathcal{B}[\cdot] = \mathcal{B}_{\Delta t}[\cdot]$ which maps any Lipschitz function $g \in W^{1,\infty}(\Omega)$ to a mapping $\mathcal{B}[g] \colon \Omega \mapsto \Omega$ such that:

- for any numerical solution f_n, $n = 0, \cdots, N-1$, the approximate backward flow $\mathcal{B} = \mathcal{B}[f_n]$ is a diffeomorphism and (3.2) holds with constants independent of n,
- the mapping $\mathcal{B}[\cdot]$ is stable in the sense that there exists a constant independent of Δt such that
$$\big\|\mathcal{B}[g] - \mathcal{B}[\tilde{g}]\big\|_{L^\infty(\Omega)} \lesssim \Delta t \|g - \tilde{g}\|_{L^\infty(\Omega)} \tag{3.3}$$
holds for any pair g, $\tilde{g}$ of Lipschitz functions, and
- the approximation is locally r-th order accurate with $r > 1$, in the sense that
$$\big\|\mathcal{F}_{t_n,t_{n+1}} - \mathcal{B}[f(t_n)]\big\|_{L^\infty(\Omega)} \lesssim (\Delta t)^r \tag{3.4}$$
holds for all $n = 0, \dots N-1$ with a constant depending on f_0 and $T = N\Delta t$ only.

We observe that a smooth flow preserves the local regularity of the solutions, measured in terms of the first order modulus of smoothness ω_1,

$$\omega_1(g, \tau, A)_\infty := \sup_{\|h\| \le \tau} \|\Delta_h^1 g\|_{L^\infty(\{A\}_h)} \tag{3.5}$$

based on the finite difference $\Delta_h^1 g(x) := g(x+h) - g(x)$, and where we have set

$$\{A\}_h := \{x \in A : x + h \in A\} \tag{3.6}$$

for any open domain $A \subset \Omega$. Moduli of smoothness are classical functionals which enter for instance the definition of Besov spaces, see e.g. [11]. Clearly the quantity (3.5) is monotone with respect to t, and it is easily seen that if m is any positive integer, writing $\Delta_{mh}^1 g(x) = \Delta_h^1 g(x + (m-1)h) + \Delta_h^1 g(x + (m-2)h) + \cdots + \Delta_h^1 g(x)$ yields

$$\omega_1(g, m\tau, A)_\infty \le m\omega_1(g, \tau, A)_\infty. \tag{3.7}$$

We are now ready to prove the following result:

Lemma 3.4. *If the flow $\mathcal{B}$ satisfies* (3.2) *then for any $g \in L^\infty(\Omega)$, we have*

$$\omega_1(g \circ \mathcal{B}, \tau, A)_\infty \le \lceil L_{\mathcal{B}} \rceil \omega_1(g, \tau, A^{\mathcal{B},\tau})_\infty \tag{3.8}$$

where $A^{\mathcal{B},\tau} := \mathcal{B}(\tilde{A}^\tau)$, $\tilde{A}^\tau := A + B_{\ell^2}(0, L_{\mathcal{B}} L_{\mathcal{B}^{-1}} \tau)$ (here and below, we use the standard notation $B_{\ell^p}(x, r)$ for the open ℓ^p ball of $\mathbb{R}^2$ with center x and radius r) and where $\lceil L_{\mathcal{B}} \rceil$ denotes the smallest integer greater or equal to $L_{\mathcal{B}}$.

Proof. Because $\mathcal{B}$ is Lipschitz, for any h we have

$$\begin{aligned}\|f(B(\cdot+h))-f(B(\cdot))\|_{L^\infty(\{A\}_h)} &\le \sup_{x\in\{A\}_h}\ \sup_{\|h'\|\le L_{\mathcal{B}}\|h\|} |f(B(x)+h')-f(B(x))| \\ &\le \sup_{\|h'\|\le L_{\mathcal{B}}\|h\|} \|f(\cdot+h')-f\|_{L^\infty(\mathcal{B}(\{A\}_h))}.\end{aligned}$$

Now observe that $\mathcal{B}(\{A\}_h) \subset \mathcal{B}(A) \subset \{\mathcal{B}(A+B_{\ell^2}(0,L_{\mathcal{B}^{-1}}s))\}_{h'}$ holds for any h' with $\|h'\|\le s$, see (3.6). This yields

$$\|f(B(\cdot+h))-f(B(\cdot))\|_{L^\infty(\{A\}_h)} \le \sup_{\|h'\|\le L_{\mathcal{B}}\tau} \|f(\cdot+h')-f\|_{L^\infty(\{\mathcal{B}(A+B_{\ell^2}(0,L_{\mathcal{B}^{-1}}L_{\mathcal{B}}\tau))\}_{h'})}$$

for any h with $\|h\|\le\tau$, i.e., $\omega_1(g\circ\mathcal{B},\tau,A)_\infty \le \omega_1(g,L_{\mathcal{B}}\tau,A^{\mathcal{B},\tau})_\infty$, and (3.8) follows by applying (3.7). □

Remark 3.5. By observing that $|g|_{W^{1,\infty}(\Omega)} = \sup_{\tau>0}\tau^{-1}\omega_1(g,\tau,\Omega)_\infty$, Lemma 3.4 yields

$$|f(t,\cdot,\cdot)|_{W^{1,\infty}(\Omega)} \le C|f^0|_{W^{1,\infty}(\Omega)} \qquad \forall t\in[0,T]$$

under Assumption 3.2, with a constant $C>0$ independent of t.

Remark 3.6. If the flow $\mathcal{B}$ has more smoothness, it is possible to establish high-order estimates for the associated transport, involving moduli of (positive, integer) order ν

$$\omega_\nu(g,t,A)_\infty := \sup_{\|h\|\le t}\|\Delta_h^\nu g\|_{L^\infty(\{A\}_{h,\nu})} \tag{3.9}$$

based on the finite differences defined recursively by $\Delta_h^\nu g := \Delta_h^1(\Delta_h^{\nu-1}g)$, and now writing $\{A\}_{h,\nu} := \{x\in A : x+h\in A,\cdots,x+\nu h\in A\}$.

For the sake of completeness, we now describe one approximation scheme for the flow that is based on a Strang splitting in time, and which is standard in the context of semi-Lagrangian methods, see e.g. [9] or [23]. Denoting by

$$E[g] := \int_0^1 K(x,y)\left(\int_{\mathbb{R}} g(y,v)\,\mathrm{d}v - 1\right)\mathrm{d}y \tag{3.10}$$

the electric field associated with some arbitrary phase space density g, see (2.28), the scheme consists in defining one-directional flows

$$\mathcal{B}_x^{\frac12}(x,v) := (x-v\Delta t/2,v), \qquad \mathcal{B}_v[g](x,v) := (x,v-E[g]\Delta t), \tag{3.11}$$

and corresponding transport operators

$$\mathcal{T}_x^{\frac12}\colon g\mapsto g\circ\mathcal{B}_x^{\frac12}, \qquad \mathcal{T}_v\colon g\mapsto g\circ\mathcal{B}_v[g].$$

The full operator $\mathcal{T} := \mathcal{T}_x^{1/2}\mathcal{T}_v\mathcal{T}_x^{1/2} : g\mapsto g\circ\mathcal{B}[g]$ corresponds to the explicit flow

$$\mathcal{B}[g]\colon (x,v)\mapsto(\tilde{x},\tilde{v}), \qquad \begin{cases}\tilde{x} := x - v\Delta t + ((\Delta t)^2/2)E[\mathcal{T}_x^{\frac12}g](x-v\Delta t/2),\\ \tilde{v} := v - \Delta t E[\mathcal{T}_x^{\frac12}g](x-v\Delta t/2).\end{cases} \tag{3.12}$$

The following lemma states that this scheme is (locally) third order accurate in time, i.e., (3.4) is satisfied with $r = 3$. Readers who are mostly interested in the analysis of adaptive schemes might skip the technical details of the proof.

Lemma 3.7. *If the initial datum f^0 is in $W^{1,\infty}(\Omega)$, we have for all $n = 0, \dots, N-1$*

$$\sup_{(x,v)\in\Omega} \big\|\mathcal{F}_{t_n,t_{n+1}}(x,v) - \mathcal{B}[f(t_n)](x,v)\big\| \lesssim (\Delta t)^3 \tag{3.13}$$

with a constant that only depends on f^0 and on the final time $T = N\Delta t$.

Proof. For (x,v) fixed in $\mathbb{R}^2$, we let

$$(X,V)(s) := \mathcal{F}_{s,t_{n+1}}(x,v) \quad \text{and} \quad (X^n,V^n) := \mathcal{B}[f(t_n)](x,v),$$

thus we need to prove that $\max\big(|X^n - X(t_n)|, |V^n - V(t_n)|\big) \le C(\Delta t)^3$. Denoting by $E_X(t) := E(t, X(t))$ the exact field along the characteristic curve, we use Lemma 2.3 together with the characteristic equation (2.12) to bound the following time derivatives (for conciseness, here $\|\cdot\|_\infty$ stands for $\|\cdot\|_{L^\infty((0,T)\times(0,1))}$):

$$\|E_X\|_{L^\infty(0,T)} \le C \tag{3.14}$$

$$\|\dot E_X\|_{L^\infty(0,T)} \le \|\partial_t E\|_\infty + \|V\|_{L^\infty(0,T)}\|\partial_x E\|_\infty \lesssim \|\partial_t E\|_\infty + \|\partial_x E\|_\infty \le C \tag{3.15}$$

$$\begin{aligned}\|\ddot E_X\|_{L^\infty(0,T)} \le{}& \|\partial^2_{tt}E\|_\infty + 2\|V\|_{L^\infty(0,T)}\|\partial^2_{tx}E\|_\infty \\ &+ \|V^2\|_{L^\infty(0,T)}\|\partial^2_{xx}E\|_\infty + \|E\|_\infty\|\partial_x E\|_\infty \le C.\end{aligned} \tag{3.16}$$

We next decompose

$$\begin{aligned}X^n - X(t_n) &= X(t_{n+1}) - X(t_n) - v\Delta t + \frac{(\Delta t)^2}{2}\, E[\mathcal{T}_x^{\frac12} f(t_n)](x - v\frac{\Delta t}{2}) \\ &= \mathcal{E}_1 + \frac{(\Delta t)^2}{2}(\mathcal{E}_2 + \mathcal{E}_3)\end{aligned}$$

with auxiliary terms defined by

$$\begin{aligned}&\mathcal{E}_1 := X(t_{n+1}) - X(t_n) - v\Delta t + \frac{(\Delta t)^2}{2}\, E_X(t_{n+\frac12}) \quad \text{with} \quad t_{n+\frac12} = (n+\frac12)\Delta t, \\ &\mathcal{E}_2 := E(t_{n+\frac12}, x - v\frac{\Delta t}{2}) - E_X(t_{n+\frac12}), \\ &\mathcal{E}_3 := E[\mathcal{T}_x^{\frac12} f(t_n)](x - v\frac{\Delta t}{2}) - E(t_{n+\frac12}, x - v\frac{\Delta t}{2}).\end{aligned}$$

Similarly, we decompose

$$V^n - V(t_n) = V(t_{n+1}) - V(t_n) - \Delta t\, E[\mathcal{T}_x^{\frac12} f(t_n)](x - v\frac{\Delta t}{2}) = \mathcal{E}_4 + \Delta t(\mathcal{E}_2 + \mathcal{E}_3) \tag{3.17}$$

with $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$ defined as above and

$$\mathcal{E}_4 := V(t_{n+1}) - V(t_n) - \Delta t E_X(t_{n+\frac{1}{2}}).$$

It thus remains to establish that $|\mathcal{E}_1|$, $|\mathcal{E}_4| \lesssim (\Delta t)^3$ and $|\mathcal{E}_2|$, $|\mathcal{E}_3| \lesssim (\Delta t)^2$. For the first term, we calculate

$$\mathcal{E}_1 = \int_{t_n}^{t_{n+1}} (V(t) - v)\,\mathrm{d}t + \frac{(\Delta t)^2}{2} E_X(t_{n+\frac{1}{2}}) = \int_{t_n}^{t_{n+1}} \int_t^{t_{n+1}} (E_X(t_{n+\frac{1}{2}}) - E_X(s))\,\mathrm{d}s\,\mathrm{d}t$$

and from (3.15) we have $|E_X(t_{n+1/2}) - E_X(s)| \le |\dot{E}_X|_{L^\infty(0,T)} |t_{n+1/2} - s| \lesssim \Delta t$, hence $|\mathcal{E}_1| \lesssim (\Delta t)^3$. For the second term $\mathcal{E}_2$, we calculate

$$\begin{aligned}
|\mathcal{E}_2| &= \big|E(t_{n+\frac{1}{2}}, x - v\frac{\Delta t}{2}) - E(t_{n+\frac{1}{2}}, X(t_{n+\frac{1}{2}}))\big| \le \|\partial_x E\|_\infty \Big|X(t_{n+\frac{1}{2}}) - x + v\frac{\Delta t}{2}\Big| \\
&\lesssim |X(t_{n+\frac{1}{2}}) - X(t_{n+1}) + v\frac{\Delta t}{2}| \lesssim \int_{t_{n+\frac{1}{2}}}^{t_{n+1}} |v - V(t)|\,\mathrm{d}t \lesssim \|E_X\|_{L^\infty(0,T)} (\Delta t)^2 \lesssim (\Delta t)^2
\end{aligned}$$

where the last inequality comes from (3.14). According to the respective definitions of E and $E[\mathcal{T}_x^{1/2} f(t_n)]$, see (2.28) and (3.10), we next bound the third term $\mathcal{E}_3$ according to

$$\begin{aligned}
|\mathcal{E}_3| &= \left|\int_0^1 K(x - v\frac{\Delta t}{2}, y) \int_{\mathbb{R}} \big[f(t_n, y - \tilde{v}\frac{\Delta t}{2}, \tilde{v}) - f(t_{n+\frac{1}{2}}, y, \tilde{v})\big]\,\mathrm{d}\tilde{v}\,\mathrm{d}y\right| \\
&\le \int_0^1 \left|\int_{\mathbb{R}} \big[f(t_n, y - \tilde{v}\frac{\Delta t}{2}, \tilde{v}) - f(t_{n+\frac{1}{2}}, y, \tilde{v})\big]\,\mathrm{d}\tilde{v}\right|\,\mathrm{d}y = \int_0^1 \left|\int_{\mathbb{R}} \Phi(y, \tilde{v})\,\mathrm{d}\tilde{v}\right|\,\mathrm{d}y,
\end{aligned} \tag{3.18}$$

where we have denoted $\Phi(y, \tilde{v}) := f(t_n, y - \tilde{v}\Delta t/2, \tilde{v}) - f(t_{n+1/2}, y, \tilde{v})$. Denoting respectively $t_s := t_n + \Delta t/2 - s$ and $y_s(\tilde{v}) := y - \tilde{v}s$ for concise notations, we then observe that

$$\Phi(y, \tilde{v}) = \int_0^{\frac{\Delta t}{2}} \frac{\mathrm{d}}{\mathrm{d}s} f(t_s, y_s(\tilde{v}), \tilde{v})\,\mathrm{d}s = -\int_0^{\frac{\Delta t}{2}} (\partial_t f + \tilde{v}\partial_x f)(t_s, y_s(\tilde{v}), \tilde{v})\,\mathrm{d}s$$

hence it follows from the Vlasov equation that $\Phi(y, \tilde{v}) = -\int_0^{\Delta t/2} \Theta(s, y, \tilde{v})\,\mathrm{d}s$ with $\Theta(s, y, \tilde{v}) := E(t_s, y_s(\tilde{v}))\partial_v f(t_s, y_s(\tilde{v}), \tilde{v})$. Now, instead of writing a straightforward bound $|\mathcal{E}_3| \lesssim \Delta t$ (which is not sufficient), we integrate by parts by using $(y_s)' = -s$,

$$\begin{aligned}
\int_{\mathbb{R}} s\partial_x E(t_s, y_s(\tilde{v})) f(t_s, y_s(\tilde{v}), \tilde{v})\,\mathrm{d}\tilde{v} &= -\int_{\mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}\tilde{v}} \big[E(t_s, y_s(\tilde{v}))\big] f(t_s, y_s(\tilde{v}), \tilde{v})\,\mathrm{d}\tilde{v} \\
&= \int_{\mathbb{R}} E(t_s, y_s(\tilde{v})) \frac{\mathrm{d}}{\mathrm{d}\tilde{v}} \big[f(t_s, y_s(\tilde{v}), \tilde{v})\big]\,\mathrm{d}\tilde{v} \\
&= \int_{\mathbb{R}} E(t_s, y_s(\tilde{v})) \big[(-s\partial_x f + \partial_v f)(t_s, y_s(\tilde{v}), \tilde{v})\big]\,\mathrm{d}\tilde{v},
\end{aligned}$$

which yields

$$\int_{\mathbb{R}} \Theta(s,y,\tilde{v})\,\mathrm{d}\tilde{v} = s\int_{\mathbb{R}} \Big[\partial_x E(t_s,y_s(\tilde{v}))f(t_s,y_s(\tilde{v}),\tilde{v}) \\ + E(t_s,y_s(\tilde{v}))\partial_x f(t_s,y_s(\tilde{v}),\tilde{v})\Big]\,\mathrm{d}\tilde{v}.$$

It is then possible to write a satisfactory bound for Φ. Indeed by using Lemma 2.3 we know that the characteristic curves have a bounded support in the speed dimension, and that E, f have Lipschitz smoothness. It follows that

$$\begin{aligned}\left|\int_{\mathbb{R}} \Phi(y,\tilde{v})\,\mathrm{d}\tilde{v}\right| &= \left|\int_{\mathbb{R}}\int_0^{\frac{\Delta t}{2}} \Theta(s,y,\tilde{v})\,\mathrm{d}s\,\mathrm{d}\tilde{v}\right| \\ &\le \frac{\Delta t}{2}\sup_{|s|\le \Delta t/2}\left|\int_{-Q(T)}^{Q(T)} \Theta(s,y,\tilde{v})\,\mathrm{d}\tilde{v}\right| \lesssim (\Delta t)^2\end{aligned}$$

which together with (3.18) yields $|\mathcal{E}_3| \lesssim (\Delta t)^2$. For the fourth term $\mathcal{E}_4$ in (3.17), we finally write by using (2.12) and the fact that E_X is the field along exact curves,

$$\begin{aligned}\mathcal{E}_4 &= \big(V(t_{n+1}) - V(t_{n+\frac{1}{2}})\big) + \big(V(t_{n+\frac{1}{2}}) - V(t_n)\big) - \Delta t E_X(t_{n+\frac{1}{2}}) \\ &= \int_0^{\frac{\Delta t}{2}} \big[E_X(t_{n+1}-\tau) + E_X(t_n+\tau)\big]\,\mathrm{d}\tau - \Delta t E_X(t_{n+\frac{1}{2}}) \\ &= \int_0^{\frac{\Delta t}{2}} \big[E_X(t_{n+1}-\tau) - E_X(t_{n+\frac{1}{2}}) + E_X(t_n+\tau) - E_X(t_{n+\frac{1}{2}})\big]\,\mathrm{d}\tau \\ &= \int_0^{\frac{\Delta t}{2}}\int_\tau^{\frac{\Delta t}{2}} \big[\dot{E}_X(t_{n+1}-s) - \dot{E}_X(t_n+s)\big]\,\mathrm{d}s\,\mathrm{d}\tau,\end{aligned}$$

which yields $|\mathcal{E}_4| \le (\Delta t)^3 \|\ddot{E}_X\|_{L^\infty(0,T)} \lesssim (\Delta t)^3$ according to (3.16). □

3.2 Uniform interpolation and error analysis

In [2], Nicolas Besse established an a priori error estimate for the first order semi-Lagrangian method in the case where $\mathcal{K} = \mathcal{K}_h$ is a quasi-uniform (say, fixed) mesh of maximal diameter h and where the initial datum f^0 is in $W^{2,\infty}(\Omega)$. The scheme reads then:

$$f_0 := P_h f^0 \qquad \text{and} \qquad f_n := P_h \mathcal{T} f_{n-1} \qquad \text{for} \qquad n = 0,\dots,N-1,$$

with P_h denoting the associated $\mathcal{P}_1$ (continuous, piecewise affine) interpolation. Higher order schemes were also analyzed by Nicolas Besse and Michel Mehrenberger in [4], but we shall not describe their arguments here. The analysis consists in decomposing the error

$$e_{n+1} := \|f(t_{n+1}) - f_{n+1}\|_{L^\infty(\Omega)}$$

into three parts as follows: a first part

$$e_{n+1,1} := \|f(t_{n+1}) - \mathcal{T} f(t_n)\|_{L^\infty(\Omega)}$$

which corresponds to the approximation of the characteristics by the numerical transport operator $\mathcal{T}$, a second part

$$e_{n+1,2} := \|(I - P_h)\mathcal{T} f(t_n)\|_{L^\infty(\Omega)}$$

which corresponds to the interpolation error, and a third part

$$e_{n+1,3} := \|P_h\big(\mathcal{T} f(t_n) - \mathcal{T} f_n\big)\|_{L^\infty(\Omega)}$$

which can be seen as the propagation of the numerical errors from the previous time step. Estimating these three terms involves the properties (such as stability, smoothness and accuracy) of both the approximated characteristics and the interpolations. Let us detail the arguments.

In order to estimate the first term we can rely on the accuracy (3.4). Indeed, since the approximate and exact transport operators are characterized by the corresponding flows, according to $\mathcal{T} f(t_n) = f(t_n) \circ \mathcal{B}[f(t_n)]$ and $f(t_{n+1}) = f(t_n) \circ \mathcal{F}_{t_n,t_{n+1}}$, we have

$$e_{n+1,1} \le |f(t_n)|_{W^{1,\infty}(\Omega)} \sup_{(x,v)\in\Omega} \big\|\mathcal{F}_{t_n,t_{n+1}}(x,v) - \mathcal{B}[f(t_n)](x,v)\big\| \lesssim (\Delta t)^r$$

as long as $f^0 \in W^{1,\infty}(\Omega)$, by using (3.13) (with $r = 3$ if we use the Strang splitting scheme described above).

For the second term we need interpolation error estimates, and for that purpose we recall a classical result of Jacques Deny and Jacques-Louis Lions, involving the space $\mathbb{P}_m := \operatorname{span}\big\{x^a y^b : a, b \in \{0, 1, \ldots, m\}, a + b \le m\big\}$ of polynomials of total degree less or equal to m (see [10, Th. 14.1]).

Theorem 3.8. *For $1 \le p \le \infty$, $m \in \mathbb{N}$ and $A \subset \mathbb{R}^2$ being an open bounded connected domain with Lipschitz boundary,*

$$\inf_{q\in\mathbb{P}_m} \|g - q\|_{L^p(A)} \lesssim |g|_{W^{m+1,p}(A)} \tag{3.19}$$

holds with a constant independent of g.

In particular, it is possible to infer a local estimate by using a scaling argument: for any square, or triangle K_h of diameter h, we have

$$\inf_{q\in\mathbb{P}_m} \|g - q\|_{L^p(K_h)} \lesssim h^{m+1} |g|_{W^{m+1,p}(K_h)} \tag{3.20}$$

with a constant that depends on the angles of K_h, but not on its diameter. In order to derive an estimate for the $\mathcal{P}_m$ interpolation error in the supremum norm, we next observe that for any $K \in \mathcal{K}_h$, we have $\|P_h g\|_{L^\infty(K)} \lesssim \|g\|_{L^\infty(K)}$ (with constant one in the $\mathcal{P}_1$ case), hence for any $p \in \mathbb{P}_m$ we have

$$\|(I - P_h)g\|_{L^\infty(K)} \le \|g - q\|_{L^\infty(K)} + \|P_h(p - g)\|_{L^\infty(K)} \lesssim \|g - q\|_{L^\infty(K)}$$

which, according to (3.20), yields

$$\|(I - P_h)g\|_{L^\infty(K)} \lesssim h^{m+1}|g|_{W^{m+1,\infty}(K)}. \tag{3.21}$$

By using high-order smoothness estimates for the exact and approximate transport operators, which can be established by arguments similar to those in Lemmas 2.3 and 3.4, the second term is then estimated by

$$e_{n+1,2} \lesssim h^2|\mathcal{T}f(t_n)|_{W^{2,\infty}(\Omega)} \lesssim h^2|f(t_n)|_{W^{2,\infty}(\Omega)} \lesssim h^2,$$

as long as $f^0 \in W^{2,\infty}(\Omega)$.

Finally, we observe that the third part $e_{n+1,3}$ is bounded by $\|\mathcal{T}f(t_n) - \mathcal{T}f_n\|_{L^\infty(\Omega)}$, indeed piecewise affine interpolations never increase the L^∞ norm. Note that if $\mathcal{T}$ was linear, we would clearly have $\|\mathcal{T}f(t_n) - \mathcal{T}f_n\|_{L^\infty(\Omega)} = \|\mathcal{T}(f(t_n) - f_n)\|_{L^\infty(\Omega)} \le e_n$. Instead, the operator $\mathcal{T}$ is nonlinear but according to (3.3) it is stable, thus we have

$$\begin{aligned}
\|\mathcal{T}f(t_n) - \mathcal{T}f_n\|_{L^\infty(\Omega)} &\le \big\|f(t_n) \circ \mathcal{B}[f(t_n)] - f(t_n) \circ \mathcal{B}[f_n]\big\|_{L^\infty(\Omega)} \\
&\qquad + \big\|(f(t_n) - f_n) \circ \mathcal{B}[f_n]\big\|_{L^\infty(\Omega)} \\
&\le |f(t_n)|_{W^{1,\infty}(\Omega)}\big\|\mathcal{B}[f(t_n)] - \mathcal{B}[f(t_n)]\big\|_{L^\infty(\Omega)} + e_n \\
&\le (1 + C\Delta t)e_n
\end{aligned}$$

with a constant that only depends on the initial datum f^0 and the final time T. By gathering the above estimates, we find that

$$e_{n+1} \le e_{n+1,1} + e_{n+1,2} + e_{n+1,3} \le (1 + C\Delta t)e_n + C((\Delta t)^r + h^2)$$

holds with a constant depending on f^0 and T only, therefore it follows from a discrete Gronwall argument that

$$e_n \lesssim (\Delta t)^{r-1} + h^2/\Delta t \qquad \text{for} \quad n = 0, \dots, N.$$

By balancing $(\Delta t)^r \sim h^2$, we finally observe the following convergence rate in terms of the cardinality $\#(\mathcal{K}_h)$ of the triangulation:

$$\sup_{n=0,\dots,N} \|f(t_n) - f_n\|_{L^\infty(\Omega)} \lesssim h^{2(1-\frac{1}{r})} \lesssim \#(\mathcal{K}_h)^{-(1-\frac{1}{r})}. \tag{3.22}$$

In other terms, such a scheme is of global order $1 - 1/r$, at least. If need be, let us recall the reader that such an inequality somehow expresses a trade-off between the *accuracy* of the numerical approximations and their *complexity*, closely related to their computational cost. In particular, it allows

- to impose a maximal cardinality on the meshes, while guaranteeing the accuracy of the interpolations, or
- to prescribe a given accuracy on the interpolations, while giving a complexity bound for the associated meshes.

As we shall see, it is an essential purpose of adaptive strategies to improve the order of convergence. The remaining sections are then devoted to describe and further analyze such adaptive variants of the above scheme. In particular, we shall describe in Sections 4 and 5 two distinct frameworks, namely adaptive meshes and interpolatory wavelets, that both have a multilevel tree structure and are suitable for adaptive interpolation. In Section 6, we will give the details for algorithms automatically adapting the interpolation grids to a given function, and accurately transporting – that is, predicting – these grids along any given smooth flow.

4 Adaptive multilevel meshes

In this section, we describe a simple algorithmic setting for performing adaptive interpolations of finite element type. In order to motivate such constructions, we start by recalling how adaptive strategies can be proven to be more efficient than uniform ones when the unknown solution has a highly non-uniform smoothness (a precise meaning of this statement will be given in the text). Readers interested in more details about adaptive approximation and characterization of convergence rates in terms of smoothness spaces such as Sobolev or Besov spaces are strongly encouraged to read the excellent tutorial article of Ronald DeVore [17], and the book of Albert Cohen [11].

4.1 Why adaptive meshes?

We consider here the problem of interpolating some continuous function g known on the unit square $\Omega = (0,1)^2$. If we desire to use $\mathcal{P}_1$ finite elements, i.e., piecewise affine interpolations on conforming triangulations of Ω, we can think of two different approaches: The first one consists in using – as in the previous section – a sequence of uniform meshes $\mathcal{K}_h$ made of shape regular triangles of diameter $\mathcal{O}(h)$, i.e., triangles K that contain and that are contained in balls of respective diameter d_K and d'_K such that

$$c_* h \lesssim d_K \le d'_K \lesssim c^* h$$

with constants independent of h. If g belongs to the space $W^{2,\infty}(\Omega)$, i.e., if it is essentially bounded on Ω and if its second order derivatives are also essentially bounded on Ω, we have seen that estimate (3.20) allows to bound the global interpolation error on $\mathcal{K}_h$ by

$$\|(I - P_h)g\|_{L^\infty(\Omega)} \lesssim h^2 |g|_{W^{2,\infty}(\Omega)}. \tag{4.1}$$

Now, it is possible to write a convergence rate associated with this approximation, independently of the transport problem under consideration. Indeed by using that the cardinality $\#(\mathcal{K}_h)$ is of order h^{-2}, we have

$$\|(I - P_h)g\|_{L^\infty(\Omega)} \lesssim \#(\mathcal{K}_h)^{-1} |g|_{W^{2,\infty}(\Omega)}, \tag{4.2}$$

which yields a convergence rate for uniform $\mathcal{P}_1$ interpolation (see the discussion in previous section).

Now, to improve the trade-off between accuracy and complexity, a different approach consists in designing a mesh which is locally adapted to the target function g. A way of doing this could be, according to (3.20), to use bigger triangles (hence larger values of h) where g has a small $W^{2,\infty}$-seminorm, and smaller ones elsewhere. Intuitively, this should reduce the cardinality of the triangulation while not increasing much the global interpolation error. A more convenient setting, however, is given by the following local estimate – substantially stronger than (3.20) –

$$\|(I - P_K)g\|_{L^\infty(K)} \lesssim |g|_{W^{2,1}(K)} \tag{4.3}$$

that is valid with a constant that only depends on the angles of K. The foregoing estimate can be shown by using the continuous embedding of $W^{2,1}(K)$ into $L^\infty(K)$, see e.g. [1, Ch. 4], and a scaling argument. Note that the scale invariance, i.e., the fact that the constant does not depend on the diameter of K, corresponds to the fact that the Sobolev embedding is critical, indeed we have $\frac{1}{\infty} = \frac{1}{1} - \frac{2}{d}$ in dimension $d = 2$. According to this estimate, a natural desire is to find a triangulation $\mathcal{K}_\varepsilon$ that equilibrates the local seminorms $|g|_{W^{2,1}(K)}$, in the sense that it satisfies

$$\underline{c}\varepsilon \le |g|_{W^{2,1}(K)} \le \overline{c}\varepsilon \tag{4.4}$$

for constants $\underline{c}$, $\overline{c}$ independent of h, and any $K \in \mathcal{K}_\varepsilon$. Clearly, the associated interpolation P_ε would satisfy

$$\|(I - P_\varepsilon)g\|_{L^\infty(\Omega)} \lesssim \varepsilon,$$

and because summing over the left inequalities in (4.4) yields

$$\#(\mathcal{K}_\varepsilon) \le (\underline{c}\varepsilon)^{-1}|g|_{W^{2,1}(\Omega)}, \tag{4.5}$$

the resulting adaptive approximation $(\mathcal{K}_\varepsilon, P_\varepsilon g)$ would achieve a convergence rate of

$$\|(I - P_\varepsilon)g\|_{L^\infty(\Omega)} \lesssim \#(\mathcal{K}_\varepsilon)^{-1}|g|_{W^{2,1}(\Omega)}. \tag{4.6}$$

As this rate holds for functions which are only in $W^{2,1}(\Omega)$, we see that it indicates better performances in the case where g is not very smooth. And more generally, it reveals that such an adaptive approach should outperform a uniform one in the case where g is in $W^{2,\infty}(\Omega)$ but has a highly non-uniform smoothness, i.e., when $|g|_{W^{2,1}(\Omega)}$ is very small compared to $|g|_{W^{2,\infty}(\Omega)}$.

4.2 Multilevel FE-trees and associated quad-meshes

It thus appears that an adaptive strategy is likely to yield better results than a uniform one when interpolating functions of non-uniform smoothness. What we did not mention is an algorithm to design a triangulation $\mathcal{K}_\varepsilon$ that fulfills (4.4), and in practice this might be a quite difficult task. For the sake of simplicity, we shall therefore restrict ourselves to a particular class of triangulations that are obtained by recursive splittings of dyadic quadrangles. The resulting multilevel meshes should then be seen as a *compromise* between uniform and pure adaptive triangulations. As is usual in compromises, we need to choose between the two inequalities in (4.4), and because what we are first interested in is the accuracy of the approximations, we shall choose the right one. Nevertheless, we mention that such a choice still allows to derive complexity estimates, and we refer again to [17] for a survey on nonlinear (adaptive) and multilevel approximation.

Let us then introduce first multilevel *quad-meshes*, and later on derive conforming triangulations. To this end, we consider at any level $j \in \mathbb{N}$ the uniform partitions

$$\mathcal{Q}_j := \{\Omega_\gamma : \gamma \in \mathcal{I}_j\} \quad \text{with} \quad \mathcal{I}_j := \{(j,k,k') : 0 \le k, k' \le 2^j - 1\}$$

consisting of all *dyadic quadrangles*, i.e., quadrangles of the form

$$\Omega_{(j,k,k')} := (2^{-j}k, 2^{-j}(k+1)) \times (2^{-j}k', 2^{-j}(k'+1))$$

that are included in $\Omega = (0,1)^2$. In the sequel we shall denote by $|\gamma| = j$ the level of any index $\gamma \in \mathcal{I}_j$. Because the meshes $\mathcal{Q}_j$ are nested, we can equip the associated index sets with a natural tree structure: we define the *children* of γ as the set

$$\mathcal{C}^*(\gamma) := \{\mu \in \mathcal{I}_{|\alpha|+1} : \Omega_\mu \subset \Omega_\gamma\}$$

(here the superscript * is in order to distinguish this children set from another set that will be introduced in Section 5 when defining trees of dyadic points), and we say that λ is a parent of γ whenever $\gamma \in \mathcal{C}^*(\lambda)$. Obviously, every γ has four children and (as long as $|\gamma| \ge 1$) a unique parent $\mathcal{P}(\gamma)$. In graph theory, a *tree* is a connected acyclic graph, which here corresponds to considering only index sets $\Lambda \subset \mathcal{I}^\infty := \cup_{j\ge 0} \mathcal{I}_j$ that contain the parent of any of their elements, i.e., that satisfy

$$\mathcal{P}(\gamma) \in \Lambda \qquad \forall \gamma \in \Lambda. \tag{4.7}$$

We also recall that the (inner) *leaves* of Λ are the nodes with no children in Λ,

$$\mathcal{L}_{\text{in}}(\Lambda) := \{\gamma \in \Lambda : \mathcal{C}^*(\gamma) \cap \Lambda = \emptyset\}. \tag{4.8}$$

In the framework of multilevel meshes associated with finite element type interpolation, we consider the following definition.

Definition 4.1. The set $\Lambda \subset \mathcal{I}^\infty := \cup_{j\ge 0} \mathcal{I}_j$ is said to be a *FE-tree* if all its nodes (except the leaves) have exactly four children in Λ, i.e., if

$$\mathcal{C}^*(\gamma) \subset \Lambda \text{ or } \mathcal{C}^*(\gamma) \cap \Lambda = \emptyset \qquad \forall \gamma \in \Lambda. \tag{4.9}$$

Its associated quad-mesh is then defined as

$$M(\Lambda) := \{\Omega_\gamma : \gamma \in \mathcal{L}_{\text{in}}(\Lambda)\}, \tag{4.10}$$

see (4.8).

For later purposes, we will also need that the levels of two adjacent cells in a quad-mesh do not differ too much. This yields the following definition.

Definition 4.2. The FE-tree Λ is said to be *graded* if it satisfies

$$\big||\gamma| - |\mu|\big| \le 1 \qquad \forall\, \gamma, \mu \in \mathcal{L}_{\text{in}}(\Lambda) \quad \text{with} \quad \overline{\Omega}_\gamma \cap \overline{\Omega}_\mu \neq \emptyset. \tag{4.11}$$

Fundamentally, the tree structure should be seen as a convenient setting for algorithmic refinements: just as refining a cell in a mesh consists in replacing it by its four sub-cells, refining the corresponding node in Λ consists in adding its four children to Λ. To any quad-mesh (made of dyadic quadrangles), we can indeed associate the FE-tree

$$\Lambda(M) := \big\{\gamma \in \mathcal{I}^\infty : \exists \Omega_\lambda \in M,\ \Omega_\lambda \subset \Omega_\gamma\big\},$$

the leaves of which clearly coincide with M. Now, the simplest way to build a tree Λ is to recursively add new children to the root tree $\mathcal{I}_0 = \{(0,0,0)\}$, according to some growing criterion. Being interested in $\mathcal{P}_1$ interpolations, a natural criterion would be, according to (4.3), to check whether the local $W^{2,1}$-seminorm of the function g is larger than some prescribed tolerance ε. In the context of interpolating transported numerical solutions however, this is not well-posed since the second derivatives of a piecewise affine function g are not L^1-functions but only Radon measures supported on the edges of the underlying triangulation. Denoting by $\mathcal{M}(A) = (\mathcal{C}_c(A))'$ the set of Radon measure, that is the dual of the continuous functions with compact support on an open domain A, and by

$$\int_A \mu := \sup_{\substack{\varphi \in \mathcal{C}_c(A) \\ \|\varphi\|_{L^\infty(A)} \le 1}} |\mu(\varphi)|$$

the total mass of the measure $\mu \in \mathcal{M}(A)$, we then relax the $W^{2,1}$-seminorm of g into the total mass of its second derivatives,

$$|g|_{W^*(A)} := \int_A \Big(|\partial^2_{xx} g| + |\partial^2_{xv} g| + |\partial^2_{vv} g|\Big).$$

This defines a seminorm on any open domain A, and can be extended to any measurable (Borel) set such as a polygonal domain. Note that in this case it might include non-zero contributions from the edges. Accordingly, we denote by $W^*(\Omega)$ the space of any g such that $|g|_{W^*(\Omega)}$ is finite and we adopt the following definition.

Definition 4.3 (ε-adapted FE-trees and meshes). A mesh M consisting of quadrangles or triangles is said to be *ε-adapted* to g if it satisfies

$$\sup_{K \in M} |g|_{W^*(\overline{K})} \le \varepsilon,$$

and the FE-tree Λ is said to be ε-adapted to g if its associated mesh $M(\Lambda)$ does so.

In [7], it is shown that the W^*-seminorm satisfies two interesting properties: first it gives a generalization the error estimate (4.3), i.e., for any triangle K we have

$$\|(I - P_K)g\|_{L^\infty(K)} \lesssim |g|_{W^*(K)} \tag{4.12}$$

with a constant depending on the shape of K only, and (up to another relaxation involving a weighted Lipschitz seminorm) the numerical transport operator resulting from the splitting scheme (3.12) can be assumed stable with respect to the W^*-seminorm, i.e., we will consider that

$$|g \circ \mathcal{B}[f_n]|_{W^*(A)} \lesssim |g|_{W^*(\mathcal{B}(A))} \tag{4.13}$$

holds with a constant independent of g. For details about the exact stability satisfied by the splitting scheme, we refer to [7].

4.3 Adaptive interpolations based on quad-meshes

In order to perform continuous interpolation (and later we shall focus on first order, i.e., piecewise affine elements), we now associate to any graded quad-mesh M a conforming triangulation $\mathcal{K}(M)$, which is in some sense equivalent to M, see (4.14), as follows. First, we build a nonconforming triangulation $\tilde{\mathcal{K}}(M)$ by simply splitting each quadrangle K of M into two triangles. More precisely, if K is an upper left or a lower right child (of its parent cell), it is splitted into its lower left and upper right halves, and the splitting is symmetric in the other two cases. We observe on Figure 2 that unless M is uniform the resulting triangulation $\tilde{\mathcal{K}}(M)$ is indeed nonconforming: when a quadrangle K shares an edge with two finer quadrangles K_1 and K_2, the splitting produces one big triangle (say, K^-) that shares an edge with two smaller triangles (say K_1^- and K_2^+). Now, because M is graded, this is the only possible configuration where the triangles are nonconforming, and it is easily seen that a conforming triangulation $\mathcal{K}(M)$ is obtained by simply merging any such pair (K_1^-, K_2^+).

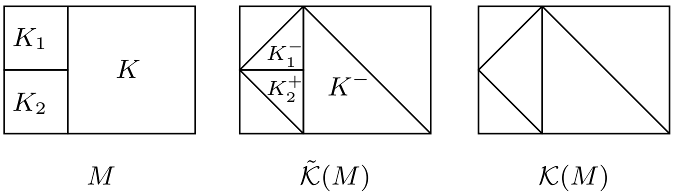

Figure 2. Deriving a conforming triangulation from a graded quad-mesh.

As every quadrangle in M (resp. every triangle in $\mathcal{K}(M)$) intersects at most two triangles in $\mathcal{K}(M)$ (resp. two quadrangles in M), we simultaneously have

$$\#(\mathcal{K}(M)) \sim \#(M) \qquad \text{and} \qquad \sup_{K \in \mathcal{K}(M)} |g|_{W^*(K)} \sim \sup_{K \in M} |g|_{W^*(K)} \tag{4.14}$$

for any g in $W^*(\Omega)$. It follows that the piecewise affine interpolation P_Λ associated with any graded FE-tree Λ via the conforming triangulation $\mathcal{K}(M(\Lambda))$ satisfies, in the case where Λ is ε-adapted to g,

$$\|(I - P_\Lambda)g\|_{L^\infty(\Omega)} \lesssim \sup_{K \in \mathcal{K}(M(\Lambda))} |g|_{W^*(K)} \lesssim \sup_{K \in M(\Lambda)} |g|_{W^*(K)} \lesssim \varepsilon. \tag{4.15}$$

5 Interpolatory wavelets

In this section we shall recall the construction of interpolatory wavelets of arbitrary (even) order $2R$, which rely on a discrete interpolation scheme first introduced by Gilles Deslauriers and Serge Dubuc, see [16]. We also review the main properties of the associated hierarchical basis that will be used later in the analysis of our adaptive semi-Lagrangian scheme. For more on wavelet constructions, we refer to the books of Ingrid Daubechies [15] and Albert Cohen [11].

For the sake of simplicity we describe the construction of interpolatory wavelets in the entire $\mathbb{R}^2$.

5.1 A discrete multilevel framework: the iterative interpolation scheme

We first denote the two-dimensional uniform dyadic grids at every level $j \in \mathbb{N}$ by

$$\Gamma_j := \left\{(2^{-j}k, 2^{-j}k') : k, k' \in \mathbb{Z}\right\} \;\subset\; \Gamma_{j+1} \;\subset\; \cdots \;\subset\; \mathbb{R}^2,$$

and let

$$\nabla_{j+1} := \Gamma_{j+1} \setminus \Gamma_j$$

be the set of nodes of *level* $j+1$, i.e., appearing in the refinement of Γ_j into Γ_{j+1}. In the sequel, the level of a dyadic cell γ will be denoted by the short notation $|\gamma|$, and the set of all dyadic nodes will be denoted by

$$\Gamma_\infty := \cup_{j\geq 0}\, \Gamma_j.$$

Note that if we let $\nabla_0 := \Gamma_0$, the sets ∇_j, $j \geq 0$, form a partition of Γ_∞.

The next ingredients are inter-grid operators P_j^{j+1}, P_{j+1}^j acting on sequences and standing for restriction and reconstruction, respectively: the general idea being that if the sequences $\mathbf{g}^{[j]} := \{g(\gamma) : \gamma \in \Gamma_j\} \in \ell^\infty(\Gamma_j)$, $j \in \mathbb{N}$, correspond to samples of a given $g \in \mathcal{C}(\mathbb{R}^2)$, the restricted sequences always satisfy

$$P_j^{j+1}\mathbf{g}^{[j+1]} = \mathbf{g}^{[j]},$$

whereas the predicted sequences $P_{j+1}^j\mathbf{g}^{[j]}$ generally differ from $\mathbf{g}^{[j+1]}$ on the finer nodes $\gamma \in \nabla_{j+1}$. On the other hand, the discrepancy should be small in the regions where g is smooth.

In order to be more specific we introduce stencils $S_\gamma \subset \Gamma_j$ associated with nodes of level $|\gamma| = j+1$, as follows (see Figure 3 for an illustration).

- If $\gamma = (2^{-(j+1)}(2k+1), 2^{-j}k')$ corresponds to a refinement of Γ_j in the first dimension, we set

$$S_\gamma := \{(2^{-j}(k+l), 2^{-j}k') : -R+1 \leq l \leq R\}. \tag{5.1}$$

- Similarly if $\gamma = (2^{-j}k, 2^{-(j+1)}(2k'+1))$ corresponds to a refinement in the second dimension, we set

$$S_\gamma := \{(2^{-j}k, 2^{-j}(k'+l)) : -R+1 \le l \le R\}. \tag{5.2}$$

- Finally if $\gamma = (2^{-(j+1)}(2k+1), 2^{-(j+1)}(2k'+1))$ corresponds to a refinement of Γ_j in both dimensions, we set

$$S_\gamma := \{(2^{-j}(2k+l), 2^{-j}(k'+l')) : -R+1 \le l, l' \le R\}. \tag{5.3}$$

Remark 5.1. Nodes of the latter type will play a particular role in the design of adaptive grids, and will be called **-nodes* in the sequel, see in particular Section 5.5.

The two inter-grid operators are then defined as follows:

- the restriction $P_j^{j+1} : \ell^\infty(\Gamma_{j+1}) \to \ell^\infty(\Gamma_j)$ is a simple decimation, i.e.,

$$\big(P_j^{j+1}\mathbf{g}^{[j+1]}\big)(\gamma) := \mathbf{g}^{[j+1]}(\gamma) \qquad \forall \gamma \in \Gamma_j,$$

- the prediction $P_{j+1}^j : \ell^\infty(\Gamma_j) \to \ell^\infty(\Gamma_{j+1})$ is a reconstruction given by

$$\big(P_{j+1}^j\mathbf{g}^{[j]}\big)(\gamma) := \begin{cases} \mathbf{g}^{[j]}(\gamma) & \text{if } \gamma \in \Gamma_j \\ \sum_{\mu \in S_\gamma} \pi(\gamma,\mu)\mathbf{g}^{[j]}(\mu) & \text{if } \gamma \in \nabla_{j+1} := \Gamma_{j+1} \setminus \Gamma_j \end{cases}$$

  for any $\gamma \in \Gamma_{j+1}$. Here the coefficients $\pi(\gamma,\mu)$, $\mu \in S_\gamma$, are defined in such a way that P_{j+1}^j corresponds to Lagrangian interpolation of maximal coordinate degree $2R-1$. More precisely, for any $\mu \in S_\gamma$ we let $L_{\gamma,\mu}$ denote the unique polynomial of $\mathbb{Q}_{2R-1} := \operatorname{span}\{x^a y^b : a, b \in \{0,1,\dots,2R-1\}\}$ that satisfies

$$L_{\gamma,\mu}(\lambda) = \delta_{\mu,\lambda} \qquad \forall \lambda \in S_\gamma$$

  (where δ stands for the Kronecker symbol), and is constant with respect to x or y in the case where S_γ is given by (5.1) or (5.2), respectively. Finally we set

$$\pi(\gamma,\mu) := L_{\gamma,\mu}(\gamma).$$

Remark 5.2. By using the shift invariance and self-similarity of the dyadic grids, we can check that the value of $\pi(\gamma,\mu)$ only depends on the relative positions of γ and μ. More precisely, for any R there exists a sequence $(h_{n,n'})_{n,n' \in \mathbb{Z}}$ such that

$$\pi(\gamma,\mu) = h_{k-2i,k'-2i'} \quad \text{for } \gamma = (2^{-(j+1)}k, 2^{-(j+1)}k'),\ \mu = (2^{-j}i, 2^{-j}i').$$

Moreover we note that $h_{n,n'} = 0$ for $\max(|n|,|n'|) \le 2R$, and it follows that the prediction coefficients $\pi(\gamma,\mu)$ are bounded by a constant that only depends on R.

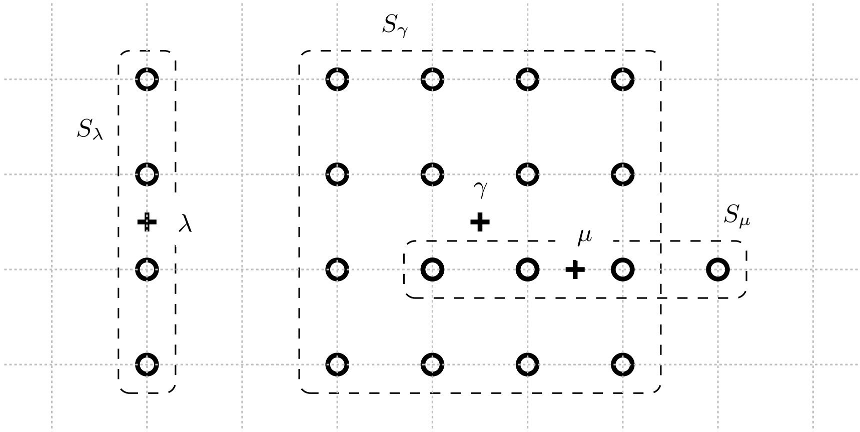

Figure 3. The three kinds of stencils corresponding to (5.1), (5.2) and (5.3), for $R = 2$.

Now, as previously announced we see that by restricting and further reconstructing samples at a given level one looses information, i.e. $P^j_{j+1}P^{j+1}_j \mathbf{g}^{[j+1]} = P^j_{j+1}\mathbf{g}^{[j]}$ generally differs from $\mathbf{g}^{[j+1]}$ on the finest nodes $\gamma \in \nabla_{j+1}$. The prediction error is then stored in a sequence

$$\mathbf{d}^{[j+1]} := \mathbf{g}^{[j+1]} - P^j_{j+1}\mathbf{g}^{[j]} \equiv \big\{d_\gamma(g) := \big(\mathbf{g}^{[j+1]} - P^j_{j+1}\mathbf{g}^{[j]}\big)(\gamma) : \gamma \in \nabla_{j+1}\big\},$$

and introducing the (Radon) measures $\tilde{\varphi}_\gamma := \delta_\gamma - \sum_{\mu\in S_\gamma} \pi(\gamma,\mu)\delta_\mu$, every $d_\gamma(g)$ rewrites as

$$d_\gamma(g) = g(\gamma) - \sum_{\mu\in S_\gamma} \pi(\gamma,\mu)g(\mu) = \langle\tilde{\varphi}_\gamma, g\rangle. \tag{5.4}$$

In the wavelet terminology these coefficients are called *details*, as they are seen as the additional information needed to recover the exact values of g from a coarse sampling. Intuitively, one would expect these coefficients to be small in the regions were g is smooth, and indeed it is easy to write a rigorous estimate: By observing that the prediction is exact for polynomials of order $2R$, i.e., we have $\langle\tilde{\varphi}_\gamma, p\rangle = 0$ for any $p \in \mathbb{Q}_{2R-1}$, we can exploit the bound on the coefficients $\pi(\gamma,\mu)$ (see Remark 5.2) together with the fact that every $\tilde{\varphi}_\gamma$ vanishes outside

$$\Sigma_\gamma := \overline{B}_{\ell^\infty}(\gamma, 2^{-|\gamma|}(2R-1)), \tag{5.5}$$

this yields (with constants that only depend on R)

$$|d_\gamma(g)| \le \inf_{p\in\mathbb{Q}_{2R-1}} |\langle\tilde{\varphi}_\gamma, g-p\rangle| \lesssim \inf_{p\in\mathbb{Q}_{2R-1}} \|g-p\|_{L^\infty(\Sigma_\gamma)} \lesssim 2^{-\nu|\gamma|}|g|_{W^{\nu,\infty}(\Sigma_\gamma)} \tag{5.6}$$

for any integer $\nu \le 2R$, where the third inequality follows from the Deny-Lions theorem, see (3.20). Note that it is also possible to estimate the details by using the modulus of smoothness of g already introduced in Section 3.1. According to a local variant of a theorem by Hassler Whitney (see e.g. [5] or [11]), we have indeed

$$|d_\gamma(g)| \lesssim \inf_{p\in\mathbb{Q}_{2R-1}} \|g-p\|_{L^\infty(\Sigma_\gamma)} \lesssim \omega_\nu(g, 2^{-|\gamma|}, \Sigma_\gamma)_\infty \tag{5.7}$$

with again $\nu \le 2R$, see (3.9).

From the above iterative interpolation scheme, it is possible to define a hierarchical wavelet basis for the full space $\mathcal{C}(\mathbb{R}^2)$ in which the details $d_\gamma(g)$ will play the role of the coefficients of g. In order to make this statement more precise, we introduce for any j and $\gamma \in \Gamma_j$ a sequence (of sequences) $\phi^{[j']}_{j,\gamma} \in \ell^\infty(\Gamma_{j'})$, $j' \ge j$, defined by

$$\phi^{[j]}_{j,\gamma}(\lambda) := \delta_{\gamma,\lambda} \quad \forall \lambda \in \Gamma_j \qquad \text{and} \qquad \phi^{[j'+1]}_{j,\gamma} := P^{j'}_{j'+1}\phi^{[j']}_{j,\gamma} \quad \forall j' \ge j. \tag{5.8}$$

Note that by definition of the prediction operator, we have

$$\phi^{[j'']}_{j,\gamma}(\lambda) = \phi^{[j']}_{j,\gamma}(\lambda) \qquad \forall \lambda \in \Gamma_{j'} \text{ and } j'' \ge j' \ge j,$$

therefore the above process essentially consists in refining growing sets of values. Now, as we will see in the next section (and as is well known), for any j and $\gamma \in \Gamma_j$ this process converges towards a continuous function $\varphi_{j,\gamma} : \mathbb{R}^2 \to \mathbb{R}$ in the sense that

$$\phi^{[j']}_{j,\gamma}(\lambda) = \varphi_{j,\gamma}(\lambda) \qquad \forall j' \ge j,\ \lambda \in \Gamma_{j'}. \tag{5.9}$$

Moreover the limit functions span nested spaces

$$V_j := \text{span}\{\varphi_{j,\gamma} : \gamma \in \Gamma_j\} \subset V_{j+1} \subset \cdots \tag{5.10}$$

which have a dense union in $\mathcal{C}(\mathbb{R}^2)$, and by keeping only the functions of the type

$$\varphi_\gamma := \varphi_{|\gamma|,\gamma}, \qquad \gamma \in \Gamma_\infty,$$

one obtains as announced a hierarchical basis of $\mathcal{C}(\mathbb{R}^2)$, i.e., every continuous function g reads as

$$g = \sum_{\gamma\in\Gamma_{j_0}} g(\gamma)\varphi_\gamma + \sum_{j\ge j_0+1}\sum_{\gamma\in\Gamma_j} d_\gamma(g)\varphi_\gamma, \tag{5.11}$$

for any $j_0 \in \mathbb{N}$, the limit holding in a pointwise sense.

5.2 Convergence of the iterative interpolation scheme

We shall now give a proof of the above claims (5.9), (5.10) (and defer (5.11) to the next section). As (5.8) is a particular instance of what is referred to as *stationary subdivision schemes*, its properties can be analyzed by using general tools such as those presented in the review articles [8] and [18]. In [15] and [11], the connections between wavelets and subdivision schemes are investigated in more details. Here we shall adapt the arguments given in [11] for our particular case of interest.

To begin with, we observe that it suffices to consider the case where $j = 0$ and $\gamma = 0$. Indeed by using the shift invariance and the self-similarity of the dyadic grids, we can check that for all j, j' such that $0 \le j \le j'$ and all γ, λ in $\Gamma_j, \Gamma_{j'}$, respectively, we have

$$\phi^{[j']}_{j,\gamma}(\lambda) = \phi^{[j']}_{j,0}(\lambda - \gamma) = \phi^{[j'-1]}_{j-1,0}\big(2(\lambda-\gamma)\big) = \cdots = \phi^{[j'-j]}_{0,0}\big(2^j(\lambda-\gamma)\big).$$

Hence the convergence of $\big(\phi^{[j']}_{j,\gamma}\big)_{j'\ge j}$ will follow from that of $(\phi^{[j']}_{0,0})_{j'\ge 0}$, moreover

$$\varphi_{j,\gamma}(x) = \varphi_{j,0}(x-\gamma) = \varphi_{j-1,0}\big(2(x-\gamma)\big) = \cdots = \varphi_{0,0}\big(2^j(x-\gamma)\big) \tag{5.12}$$

will be established for all $x \in \mathbb{R}^2$. In the sequel we shall drop the subscripts $_{0,0}$.

In a nutshell, it is possible to establish the convergence of (5.8) with two arguments: the first one consists in saying that it is equivalent to the existence of a continuous function satisfying an appropriate two-scale equation, and the second one in proving that such a *scaling function* indeed exists (and in this case it is the limit of the scheme). For the sake of simplicity, we shall give a detailed proof in the one-dimensional case (gathering arguments from [11]) and leave the two-dimensional case for the reader. Then the prediction operator is defined by univariate Lagrange interpolations, and the stencil corresponding to a node of level $j+1$, say $\gamma = 2^{-(j+1)}(2k+1)$, reads

$$S_\gamma = \{2^{-j}(k+l) : -R+1 \le l \le R\}.$$

In particular, for any integer k we have (writing $\phi^{[j]} = \phi^{[j]}_{0,0}$)

$$\phi^{[j+1]}(2^{-(j+1)}(2k+1)) = \sum_{i=k-R+1}^{k+R} \pi\big(2^{-(j+1)}(2k+1), 2^{-j}i\big)\phi^{[j]}(2^{-j}i) \tag{5.13}$$

whereas the values remain unchanged on nodes of level j, i.e.,

$$\phi^{[j+1]}(2^{-j}k) := \phi^{[j]}(2^{-j}k). \tag{5.14}$$

As in Remark 5.2, we then observe that there exist coefficients h_n, $n \in \mathbb{Z}$, such that $\pi\big(2^{-(j+1)}k, 2^{-j}i\big) = h_{k-2i}$ for any odd integer k, so that by setting the remaining (even) values to $h_{2n} := \delta_{n,0}$, our iterative scheme reads

$$\phi^{[j+1]}(2^{-(j+1)}k) = \sum_{i\in\mathbb{Z}} h_{k-2i}\phi^{[j]}(2^{-j}i) \qquad \forall k \in \mathbb{Z},\ j \in \mathbb{N}. \tag{5.15}$$

Remark 5.3. Before going further let us list a few important properties of the sequence h that will be of great importance in the sequel. From the fact that the prediction stencils are symmetric and locals, one easily infers that h shares the same properties: its only non-zero terms are $h_0 = 1$ and $h_{-1} = h_1, \ldots, h_{-(2R-1)} = h_{2R-1}$. Moreover it is possible to express the polynomial reproduction properties of the iterative scheme in terms of discrete moments of h. As already observed, defining $\mathbf{g}^{[0]}$ as samples of $p_l(x) := (2x+1)^l$ indeed leads to samples of the same polynomial for $P_1^0\mathbf{g}^{[0]}$, as long as $l \le 2R-1$. By linearity of P_1^0, this gives

$$\sum_{n\in\mathbb{Z}} h_{2n+1}(2n+1)^l = \sum_{n\in\mathbb{Z}} h_{-1-2n}\mathbf{g}^{[0]}(n) = P_1^0\mathbf{g}^{[0]}(-\frac{1}{2}) = p_l(-\frac{1}{2}) = \delta_{l,0} \quad (5.16)$$

for $l = 0,\dots,2R-1$. As $h_{2n} = \delta_{n,0}$, the left hand side is exactly $\sum_{n\in\mathbb{Z}} h_n n^l$.

The connection with the scaling function $\varphi = \varphi_{0,0}$ can then be stated as follows.

Lemma 5.4. *Let $(h_n)_{n\in\mathbb{Z}}$ be a sequence of real coefficients. If there exists a continuous function φ satisfying $\varphi(k) = \delta_{k,0}$, $k \in \mathbb{Z}$, and the two-scale equation*

$$\varphi(x) = \sum_{n\in\mathbb{Z}} h_n \varphi(2x-n) \qquad \forall x \in \mathbb{R}, \quad (5.17)$$

then the iterative scheme defined the refinement rule (5.15) *and the initial condition $\phi^{[0]}(k) = \delta_{k,0}$, $k \in \mathbb{Z}$, satisfies* (5.14) *and converges towards φ. Conversely, if the above iterative scheme satisfies* (5.14) *and if it converges towards a continuous function φ, then the latter satisfies $\varphi(k) = \delta_{k,0}$, $k \in \mathbb{Z}$, as well as* (5.17).

Proof. If φ satisfies the two-scale equation (5.17), then by induction we see that it belongs to $\mathrm{span}\{x \mapsto \varphi(2^j x - n) : n \in \mathbb{Z}\}$, for any $j \in \mathbb{N}$. In particular, there exist coefficients $c_{j,n}$ such that $\varphi(x) = \sum_{n\in\mathbb{Z}} c_{j,n}\varphi(2^j x - n)$ for all $x \in \mathbb{R}$, and by using that $\varphi(k) = \delta_{k,0}$ for $k \in \mathbb{Z}$, we find $c_{j,k} = \sum_{n\in\mathbb{Z}} c_{j,n}\varphi(k-n) = \varphi(2^{-j}k)$. Therefore

$$\varphi(x) = \sum_{n\in\mathbb{Z}} \varphi(2^{-j}n)\varphi(2^j x - n) \qquad \forall x \in \mathbb{R},\ j \in \mathbb{N}. \quad (5.18)$$

Now, assume by induction that $\varphi(2^{-j}k) = \phi^{[j]}(2^{-j}k)$, $k \in \mathbb{Z}$, holds for a given $j \in \mathbb{N}$, as it is the case for $j = 0$. By applying again the two-scale relation (5.17) to (5.18), we find

$$\begin{aligned}
\varphi(x) &= \sum_{n\in\mathbb{Z}} \phi^{[j]}(2^{-j}n)\varphi(2^j x - n) \\
&= \sum_{n\in\mathbb{Z}} \phi^{[j]}(2^{-j}n) \sum_{n'\in\mathbb{Z}} h_{n'}\varphi(2^{j+1}x - 2n - n') \\
&= \sum_{k\in\mathbb{Z}} \Big(\sum_{n\in\mathbb{Z}} h_{k-2n}\phi^{[j]}(2^{-j}n)\Big)\varphi(2^{j+1}x - k) \\
&= \sum_{k\in\mathbb{Z}} \phi^{[j+1]}(2^{-(j+1)}k)\varphi(2^{j+1}x - k).
\end{aligned}$$

According to (5.18), this yields $\varphi(2^{-(j+1)}k) = \phi^{[j+1]}(2^{-(j+1)}k)$ for all $k \in \mathbb{Z}$, hence the convergence. In particular, note that $\phi^{[j+1]}(2^{-j}k) = \varphi(2^{-j}k) = \phi^{[j]}(2^{-j}k)$ for all

$k \in \mathbb{Z}$, which is (5.14).

In the other direction, we first observe that the convergence of $\phi^{[j]}$ towards φ reads $\varphi(2^{-j}k) = \phi^{[j]}(2^{-j}k)$ for all $k \in \mathbb{Z}$ and $j \in \mathbb{N}$. In particular, we have $\varphi(k) = \delta_{k,0}$ for all $k \in \mathbb{Z}$, and

$$\varphi(2^{-1}k) = \phi^{[1]}(2^{-1}k) = \sum_{i \in \mathbb{Z}} h_{k-2i}\phi^{[0]}(i) = h_k = \sum_{n \in \mathbb{Z}} h_n \varphi(k-n),$$

which shows that the two-scale relation (5.17) holds for any $x \in \Gamma_1$. Now, by assuming that it holds for any $x \in \Gamma_j$, i.e.,

$$\phi^{[j]}(2^{-j}i) = \sum_{n \in \mathbb{Z}} h_n \phi^{[j-1]}(2^{-(j-1)}i - n) \qquad \forall i \in \mathbb{Z},$$

and by applying the latter to (5.15), we find

$$\begin{aligned}
\phi^{[j+1]}(2^{-(j+1)}k) &= \sum_{i \in \mathbb{Z}} h_{k-2i} \sum_{n \in \mathbb{Z}} h_n \phi^{[j-1]}(2^{-(j-1)}i - n) \\
&= \sum_{n \in \mathbb{Z}} h_n \sum_{m \in \mathbb{Z}} h_{(k-2^j n)-2m} \phi^{[j-1]}(2^{-(j-1)}m) \\
&= \sum_{n \in \mathbb{Z}} h_n \phi^{[j]}(2^{-j}(k - 2^j n)) = \sum_{n \in \mathbb{Z}} h_n \phi^{[j]}(2^{-j}k - n)
\end{aligned}$$

where we have used (5.15) again in the third equality. This shows that (5.17) also holds for all $x \in \Gamma_{j+1}$, and for all $x \in \mathbb{R}$ by density, since φ is continuous. □

Remark 5.5. By using $\phi^{[0]}(k) = \delta_{k,0}$, one easily checks that (5.14) is equivalent with

$$h_{2n} = \delta_{n,0} \qquad \forall n \in \mathbb{Z}, \tag{5.19}$$

which can also be inferred from the two-scale relation (5.17) and $\varphi(k) = \delta_{k,0}$.

It thus remains to prove that such a scaling function indeed exists. For that purpose we first observe that φ is a continuous solution of (5.17) if and only if its Fourier transform $\hat{\varphi} : \omega \mapsto \int_{\mathbb{R}} \varphi(x) e^{-ix\omega}\, \mathrm{d}x$ is an $L^1(\mathbb{R})$ function satisfying

$$\hat{\varphi}(\omega) = m\Big(\frac{\omega}{2}\Big)\hat{\varphi}\Big(\frac{\omega}{2}\Big), \tag{5.20}$$

where the trigonometric polynomial $m(\omega) := \frac{1}{2}\sum_{n \in \mathbb{Z}} h_n e^{-in\omega}$ is sometimes called the *symbol* of φ (remember that the sequence h is finite), and is non negative on $\mathbb{R}$, as we shall soon see.

Formally, (5.20) leads to consider $\hat{\varphi}(\omega) := \prod_{j=1}^{\infty} m(2^{-j}\omega)$. In order to make this rigorous, let $\hat{\varphi}_J(\omega) := \big[\prod_{j=1}^{J} m(2^{-j}\omega)\big]\chi_{[-\pi,\pi]}(2^{-J}\omega)$ and observe that (5.19) yields

$$m(\omega) + m(\omega+\pi) = \frac{1}{2}\sum_{n \in \mathbb{Z}} h_n e^{-in\omega}(1 + (-1)^n) = \sum_{n \in \mathbb{Z}} h_{2n} = 1. \tag{5.21}$$

By using this identity together with the periodicity of m, following [11] we calculate

$$\begin{aligned}\int_{\mathbb{R}} \hat\varphi_J &= \int_{-2^J\pi}^{2^J\pi} \Big[\prod_{j=1}^{J} m(2^{-j}\omega)\Big]\,\mathrm{d}\omega = 2^J \int_{-\pi}^{\pi} \Big[\prod_{j'=0}^{J-1} m(2^{j'}\xi)\Big]\,\mathrm{d}\xi \\ &= 2^J \int_0^{\pi} \big(m(\xi)+m(\xi+\pi)\big)\Big[\prod_{j'=1}^{J-1} m(2^{j'}\xi)\Big]\,\mathrm{d}\xi = 2^J \int_0^{\pi} \Big[\prod_{j'=1}^{J-1} m(2^{j'}\xi)\Big]\,\mathrm{d}\xi \\ &= 2^J \int_{-\pi/2}^{\pi/2} \Big[\prod_{j'=1}^{J-1} m(2^{j'}\xi)\Big]\,\mathrm{d}\xi = \int_{-2^{J-1}\pi}^{2^{J-1}\pi} \Big[\prod_{j=1}^{J-1} m(2^{-j}\omega)\Big]\,\mathrm{d}\omega = \int_{\mathbb{R}} \hat\varphi_{J-1} \\ &= \cdots = \int_{\mathbb{R}} \hat\varphi_1 = 2\int_{-\pi}^{-\pi} m(\omega)\,\mathrm{d}\omega = 2\pi.\end{aligned}$$

Now, if we have $m(\omega) \ge 0$ as claimed above, the latter equality shows that $\hat\varphi_J$ is uniformly bounded in $L^1(\mathbb{R})$, and it is an easy matter to check that $\hat\varphi_J$ converges towards $\hat\varphi$ in a pointwise sense. It thus follows by using Fatou's lemma that $\hat\varphi \in L^1(\mathbb{R})$, which establishes the continuity of its Fourier transform φ and the two-scale equation (5.17) follows. According to Lemma 5.4 this proves the convergence of the iterative interpolation scheme.

Therefore it only remains to verify the claim $m(\omega) \ge 0$, and for this purpose let us show that there exists a polynomial P_R of degree not larger than $R-1$ such that

$$m(\omega) = \big(\cos^2\frac{\omega}{2}\big)^R P_R\big(\sin^2\frac{\omega}{2}\big). \tag{5.22}$$

Indeed, according to Remark 5.3 we can write

$$m(\omega) = \frac{1}{2}\Big(h_0 + 2\sum_{n\ge 1}^{2R-1} h_n\cos(n\omega)\Big) = Q\Big(\cos^2\frac{\omega}{2}\Big)$$

with a polynomial Q of degree not larger than $2R-1$, as we know that

$$\begin{aligned}\cos(n\omega) = \Re\Big[\big(e^{i\frac{\omega}{2}}\big)^{2n}\Big] &= \Re\left[\sum_{k=0}^{2n}\tbinom{2n}{k}\big(i\sin\frac{\omega}{2}\big)^k\big(\cos\frac{\omega}{2}\big)^{2n-k}\right] \\ &= \sum_{k=0}^{n}\tbinom{2n}{2k}\big(\cos^2\frac{\omega}{2}-1\big)^k\big(\cos^2\frac{\omega}{2}\big)^{n-k}.\end{aligned}$$

Now, from Remark 5.3 we can also infer that

$$m^{(l)}(0) = \frac{1}{2}\sum_{n\in\mathbb{Z}}(-in)^l h_n = \frac{(-i)^l}{2}\Big(\delta_{l,0} + \sum_{n\in\mathbb{Z}}(2n+1)^l h_{2n+1}\Big) = \delta_{l,0}$$

for $l = 0, \cdots, 2R-1$. In particular it follows from (5.21) that $m(\omega)$ vanishes with order $2R-1$ at $\omega = \pi$, which in turn implies that $Q(X^2)$ can be factorized by X^{2R}, and establishes (5.22). In order to identify P_R, we next observe that equation (5.21) leads to see it as a polynomial solution to

$$X^R P(1-X) + (1-X)^R P(X) = 1. \tag{5.23}$$

For such an equation, Bezout's theorem ensures the existence of a minimal degree solution, but we shall give a direct argument: indeed we have

$$\begin{aligned} 1 = (X + (1-X))^{2R-1} &= \sum_{k=0}^{2R-1} \tbinom{2R-1}{k} X^k (1-X)^{R-1-k} \\ &= (1-X)^R \sum_{k=0}^{R-1} \Big[\tbinom{2R-1}{k} X^k (1-X)^{R-1-k} \Big] + X^R \sum_{k=0}^{R-1} \Big[\tbinom{2R-1}{k} X^{R-1-k} (1-X)^k \Big] \end{aligned}$$

so that $P(X) := \sum_{k=0}^{R-1} \tbinom{2R-1}{k} X^k (1-X)^{R-1-k}$ is a solution to (5.23) of degree $R-1$ or lower. Since such a solution is clearly unique, it coincides with P_R and the non negativity of $m(\omega)$ follows.

In conclusion we have proved (5.9) and (5.10) as well, since it follows from (5.12) and (5.17) that any $\varphi_{j,\gamma}$ can be written as a linear combination of $\varphi_{j+1,\lambda}$, $\lambda \in \Gamma_{j+1}$.

5.3 The hierarchical wavelet basis

In order to establish the validity of the hierarchical decomposition (5.11), we now study some properties of the scaling functions $\varphi_{j,\gamma}$. According to the previous section, it is easily seen that they are interpolatory in the sense that

$$\varphi_{j,\gamma}(\lambda) = \delta_{\gamma,\lambda}, \quad \forall \gamma, \lambda \in \Gamma_j. \tag{5.24}$$

In particular they form a nodal basis for the space $V_j := \mathrm{span}\{\varphi_{j,\gamma} : \gamma \in \Gamma_j\}$, in the sense that every $g \in V_j$ reads as $g = \sum_{\gamma \in \Gamma_j} g(\gamma) \varphi_{j,\gamma}$.

Remark 5.6 (Polynomial exactness). By using the polynomial reproduction properties of the linear prediction operator (as, for instance, in Remark 5.3) one easily sees that polynomials of coordinate degree less than $2R$ belong to the space V_0 (and hence to any V_j, $j \geq 0$).

Let us now estimate the support of $\varphi_{j,\gamma}$ from the locality of the prediction stencils. To do so, we set

$$\Sigma_{j,\gamma}^{[j]} := \{\gamma\} \quad \text{and} \quad \Sigma_{j,\gamma}^{[\ell+1]} := \big\{ \lambda \in \Gamma_{\ell+1} : (\{\lambda\} \cup S_\lambda) \cap \Sigma_{j,\gamma}^{[\ell]} \neq \emptyset \big\} \quad \text{for } \ell \geq j.$$

Clearly the sequence $(\Sigma_{j,\gamma}^{[\ell]})_{\ell \geq j}$ is increasing in the sense that $\Sigma_{j,\gamma}^{[\ell]} \subset \Sigma_{j,\gamma}^{[\ell+1]}$ for all $\ell \geq j$, and by using the size of the stencils we can check that $\Sigma_{j,\gamma}^{[j+i]} \subset B_{\ell^\infty}(\gamma, r_i)$ with

$r_i = 2^{-j}(R - 1/2)(1 + 2^{-1} + \cdots + 2^{-i})$ for $i \geq 0$, hence every $\Sigma^{[\ell]}_{j,\gamma}$ is a subset of $B_{\ell^\infty}(\gamma, 2^{-j}(2R-1))$. More precisely, we observe that

$$\Sigma^{[\ell]}_{j,\gamma} = \Big(\{\gamma\} \cup \bigcup_{i=j+1}^{\ell} \nabla_i\Big) \cap B_{\ell^\infty}(\gamma, 2^{-j}(2R-1)), \tag{5.25}$$

therefore the density of the dyadic points yields

$$\Sigma_{j,\gamma} := \overline{\cup_{\ell \geq j} \Sigma^{[\ell]}_{j,\gamma}} = \overline{B}_{\ell^\infty}(\gamma, 2^{-j}(2R-1)). \tag{5.26}$$

In particular, we have $\Sigma_{|\gamma|,\gamma} = \Sigma_\gamma$, see (5.5). Since $\mathrm{supp}(\phi^{[\ell]}_{j,\gamma}) \subset \Sigma^{[\ell]}_{j,\gamma}$ by construction of the sets $\Sigma^{[\ell]}_{j,\gamma}$, it follows that $\mathrm{supp}(\varphi_{j,\gamma}) \subset \Sigma_{j,\gamma}$. Hence the functions $\varphi_{j,\gamma}$, $\gamma \in \Gamma_j$, satisfy a bounded overlapping property, namely

$$\#\big(\{\lambda \in \Gamma_j : \mathrm{supp}(\varphi_{j,\gamma}) \cap \mathrm{supp}(\varphi_{j,\lambda}) \neq \emptyset\}\big) \leq C_s \quad \forall \gamma \in \Gamma_j \tag{5.27}$$

with a constant $C_s = C_s(R)$ independent of j. One major consequence of this fact is that the basis $\{\varphi_{j,\gamma} : \gamma \in \Gamma_j\}$ is *stable* in the sense that

$$c\|g\|_{L^\infty(\mathbb{R}^2)} \leq \sup_{\gamma \in \Gamma_j} |g(\gamma)| \leq \|g\|_{L^\infty(\mathbb{R}^2)} \qquad \forall g = \sum_{\gamma \in \Gamma_j} g(\gamma)\varphi_{j,\gamma} \in V_j \tag{5.28}$$

holds with $c = (C_s\|\varphi\|_{L^\infty(\mathbb{R}^2)})^{-1}$, see (5.12). Note that it is also possible to write a so-called Bernstein (i.e., inverse) estimate for the functions in V_j, according to

$$\Big|\sum_{\gamma \in \Gamma_j} c_\gamma \varphi_{j,\gamma}\Big|_{W^{\nu,\infty}(\mathbb{R}^2)} \leq C_s \sup_{\gamma \in \Gamma_j} |c_\gamma \varphi_{j,\gamma}|_{W^{\nu,\infty}(\mathbb{R}^2)} \leq 2^{\nu j} C \sup_{\gamma \in \Gamma_j} |c_\gamma| \tag{5.29}$$

with $C = C_s|\varphi|_{W^{\nu,\infty}(\mathbb{R}^2)}$. On every space V_j, we next define a linear projector P_j by $P_j g := \sum_{\gamma \in \Gamma_j} g(\gamma)\varphi_{j,\gamma}$ which, due to (5.24), is an interpolation. Moreover, P_j is uniformly stable in the sense that

$$\|P_j g\|_{L^\infty(\mathbb{R}^2)} \lesssim \sup_{\gamma \in \Gamma_j} |g(\gamma)| \leq \|g\|_{L^\infty(\mathbb{R}^2)} \quad \forall g \in \mathcal{C}(\mathbb{R}^2) \tag{5.30}$$

holds with a constant independent of j, by using (5.28). Note that the stability is local, i.e., for any cell $\Omega_{j,\gamma} := B_{\ell^\infty}(\gamma, 2^{-j})$ we have

$$\|P_j g\|_{L^\infty(\Omega_{j,\gamma})} \leq \sum_{\mu \in \tilde{\Omega}_{j,\gamma}} \|g(\mu)\varphi_{j,\mu}\|_{L^\infty(\mathbb{R}^2)} \lesssim \sup_{\mu \in \tilde{\Omega}_{j,\gamma}} |g(\mu)| \lesssim \|g\|_{L^\infty(\tilde{\Omega}_{j,\gamma})} \tag{5.31}$$

with $\tilde{\Omega}_{j,\gamma} := \{\mu :\in \Gamma_j, \Sigma_{j,\mu} \cap \Omega_{j,\gamma} \neq \emptyset\}$. By using arguments similar to those in Section 3.2, we can establish an error estimate for this uniform interpolation.

Lemma 5.7. *For any $g \in W^{1,\infty}(\mathbb{R}^2)$ and any integer $\nu \leq 2R$, we have*

$$\|g - P_j g\|_{L^\infty(\mathbb{R}^2)} \lesssim 2^{-\nu j}|g|_{W^{\nu,\infty}(\mathbb{R}^2)} \tag{5.32}$$

with a constant independent of j.

Proof. For any $\gamma \in \Gamma_j$, let $p_{j,\gamma}$ be a polynomial in $\mathbb{Q}_{2R-1}$ such that

$$\|g - p_{j,\gamma}\|_{L^\infty(\tilde\Omega_{j,\gamma})} \le 2 \inf_{p \in \mathbb{Q}_{2R-1}} \|g - p\|_{L^\infty(\tilde\Omega_{j,\gamma})} \lesssim 2^{-\nu j} |g|_{W^{\nu,\infty}(\tilde\Omega_{j,\gamma})}$$

where the second inequality follows from the Deny-Lions theorem, see (3.20). By using the fact that polynomials of degree less than $2R$ are contained in every V_j (see Remark 5.6), the local error is then estimated by

$$\begin{aligned}\|g - P_j g\|_{L^\infty(\Omega_{j,\gamma})} &\le \|g - p_{j,\gamma}\|_{L^\infty(\Omega_{j,\gamma})} + \|P_j(p_{j,\gamma} - g)\|_{L^\infty(\Omega_{j,\gamma})} \\ &\lesssim \|g - p_{j,\gamma}\|_{L^\infty(\Omega_{j,\gamma})} \lesssim 2^{-\nu j} |g|_{W^{\nu,\infty}(\tilde\Omega_{j,\gamma})}\end{aligned}$$

and the global estimate (5.32) follows by taking the supremum over $\gamma \in \Gamma_j$ and observing that the sets $\tilde\Omega_{j,\gamma}$ also satisfy a bounded overlapping property, see (5.27). □

Remark 5.8. Since $W^{1,\infty}(\mathbb{R}^2)$ is dense in $\mathcal{C}(\mathbb{R}^2)$, the previous estimate also shows that $P_j g \to g$ uniformly as $j \to \infty$. Indeed, if $g_\varepsilon \in W^{1,\infty}(\mathbb{R}^2)$ is such that $\|g - g_\varepsilon\|_{L^\infty} \le \varepsilon$, then for any $j \ge \ln(|g_\varepsilon|_{W^{1,\infty}}/\varepsilon)$, we have

$$\|g - P_j g\|_{L^\infty(\mathbb{R}^2)} \le \|g - g_\varepsilon\|_{L^\infty(\mathbb{R}^2)} + \|g_\varepsilon - P_j g_\varepsilon\|_{L^\infty(\mathbb{R}^2)} + \|P_j(g_\varepsilon - g)\|_{L^\infty(\mathbb{R}^2)} \lesssim \varepsilon$$

where we have used (5.30) and Lemma 5.7 in the second inequality.

Now, the detail coefficients defined in (5.4) also describe the fluctuations between successive continuous levels.

Lemma 5.9. *Remember that we have set $\varphi_\gamma := \varphi_{|\gamma|,\gamma}$. For any $J \ge j_0$ and any $g \in \mathcal{C}(\mathbb{R}^2)$, we have*

$$P_J g = P_{j_0} g + \sum_{j=j_0+1}^{J} \sum_{\gamma \in \nabla_j} d_\gamma(g) \varphi_\gamma.$$

Proof. Clearly, it suffices to show that for any $j \ge 1$, we have

$$P_j g = P_{j-1} g + \sum_{\gamma \in \nabla_j} d_\gamma(g) \varphi_\gamma. \tag{5.33}$$

To see this, let $\tilde{\mathbf{g}}^{[j]} := P_j^{j-1} \mathbf{g}^{[j-1]} + \sum_{\gamma \in \nabla_j} d_\gamma(g) \phi_{j,\gamma}^{[j]}$ and observe that for $\lambda \in \Gamma_{j-1}$ the definition (5.8) of $\phi_{j,\gamma}^{[j]}$ yields

$$\tilde{\mathbf{g}}^{[j]}(\lambda) = P_j^{j-1} \mathbf{g}^{[j-1]}(\lambda) = \mathbf{g}^{[j-1]}(\lambda) = g(\lambda) = \mathbf{g}^{[j]}(\lambda).$$

By using the definition of $d_\lambda(g)$ and again (5.8), for any $\lambda \in \nabla_j$, we next see that

$$\tilde{\mathbf{g}}^{[j]}(\lambda) = (P_j^{j-1} \mathbf{g}^{[j-1]})(\lambda) + d_\lambda(g) = \mathbf{g}^{[j]}(\lambda),$$

hence $\mathbf{g}^{[j]} = \tilde{\mathbf{g}}^{[j]}$. Applying then $P_J^j := P_J^{J-1} \cdots P_{j+1}^j$ to the latter equality gives

$$\sum_{\mu \in \Gamma_j} g(\mu) \phi_{j,\mu}^J = P_J^j \tilde{\mathbf{g}}^{[j]} = \sum_{\mu \in \Gamma_{j-1}} g(\mu) \phi_{j-1,\mu}^J + \sum_{\gamma \in \nabla_j} d_\gamma(g) \phi_{j,\gamma}^J.$$

So, letting $J \to \infty$ finally yields (5.33). □

Remark 5.10. From Remark 5.8 and Lemma 5.9, we easily infer the validity of the hierarchical decomposition (5.11).

5.4 Adaptive interpolations

Summing up, we now have a representation of any continuous function g in terms of the multilevel nodal functions φ_γ, $\gamma \in \Gamma_\infty$, with small coefficients in the regions where g is smooth, according to (5.6) or (5.7). Adaptivity will be achieved by discarding small coefficients in this expansion. In order to study such approximation schemes, we introduce the following definition.

Definition 5.11. Let $j_0 \in \mathbb{N}$. A grid $\Lambda \subset \Gamma_\infty$ is said to be *admissible* if it contains the coarse grid Γ_{j_0} and if it satisfies

$$\gamma \in \Lambda \quad \Longrightarrow \quad S_\gamma \subset \Lambda.$$

Next we define a mapping $P_\Lambda : \mathcal{C}(\mathbb{R}^2) \to V_\Lambda := \text{span}\{\varphi_\lambda : \lambda \in \Lambda\}$ by

$$P_\Lambda g := \sum_{\lambda \in \Lambda} d_\lambda(g) \varphi_\lambda. \tag{5.34}$$

Clearly this makes sense for any grid $\Lambda \subset \Gamma_\infty$, but if in addition Λ is admissible, then P_λ is an interpolation, which might be of interest for practical implementations.

Lemma 5.12. *If the grid Λ is admissible then $P_\Lambda g(\gamma) = g(\gamma)$ for all $\gamma \in \Lambda$.*

Proof. First, since (5.34) is a wavelet expansion of $P_\Lambda g$, we clearly have

$$d_\gamma(P_\Lambda g) = d_\gamma(g) \quad \forall \gamma \in \Lambda. \tag{5.35}$$

Observe next that any φ_μ, $|\mu| \geq j_0 + 1$, vanishes on any $\gamma \in \Gamma_{j_0}$, and calculate

$$P_\Lambda g(\gamma) = \sum_{|\lambda| \leq j_0} d_\lambda(g) \varphi_\lambda(\gamma) = \sum_{|\lambda| \geq 0} d_\lambda(g) \varphi_\lambda(\gamma) = g(\gamma), \quad \gamma \in \Gamma_{j_0}.$$

Now assume that $P_\Lambda g$ and g coincide on $\Lambda \cap \Gamma_{j-1}$, and consider $\gamma \in \Lambda \cap \nabla_j$: since Λ is admissible, we have by definition of the details d_γ

$$P_\Lambda g(\gamma) = \sum_{\mu \in S_\gamma} \pi(\gamma, \mu) P_\Lambda g(\mu) + d_\gamma(P_\Lambda g) = \sum_{\mu \in S_\gamma} \pi(\gamma, \mu) g(\mu) + d_\gamma(g) = g(\gamma).$$

□

As we said before, the error resulting from "discarding the small details" should be controlled by the amplitude of these details. Here is the precise statement that we shall use for an approximation in the supremum norm.

Lemma 5.13. *The approximation of g associated with the grid Λ is bounded by*

$$\|g - P_\Lambda g\|_{L^\infty(\mathbb{R}^2)} \leq C \sum_{j\geq 0} \sup_{\gamma\in\nabla_j\setminus\Lambda} |d_\gamma(g)|$$

with $C = C_s\|\varphi\|_{L^\infty(\mathbb{R}^2)}$, see (5.27).

Proof. In views of (5.11), g can be written as an infinite wavelet expansion. The approximation error thus satisfies

$$\begin{aligned}\|g - P_\Lambda g\|_{L^\infty(\mathbb{R}^2)} &= \Big\| \sum_{j\geq 0} \sum_{\gamma\in\nabla_j\setminus\Lambda} d_\gamma(g)\varphi_\gamma \Big\|_{L^\infty(\mathbb{R}^2)} \\ &\leq \sum_{j\geq 0} \Big\| \sum_{\gamma\in\nabla_j\setminus\Lambda} d_\gamma(g)\varphi_\gamma \Big\|_{L^\infty(\mathbb{R}^2)} \\ &\leq C_s\|\varphi\|_{L^\infty(\mathbb{R}^2)} \sum_{j\geq 0} \sup_{\gamma\in\nabla_j\setminus\Lambda} |d_\gamma(g)|,\end{aligned}$$

where we have employed (5.28) in the last inequality. □

Remark 5.14. The foregoing estimate is sharp. Indeed, consider the one-dimensional case (for the sake of simplicity) where for $R = 1$, the reference scaling function is given by $\varphi(x) = \max(1 - |x|, 0)$. Now let $\gamma_i := 2^{-2i}(1 + 4 + \cdots + 4^{(i-1)}) \in \nabla_{2i}$, and check that

$$\varphi_{\gamma_i} = \varphi(2^{2i}(x - \gamma_i)) \geq \frac{1}{2} \quad \text{on} \quad \Big[\frac{4^i - 1}{3\cdot 4^i}, \frac{4^i - 1}{3\cdot 4^i} + \frac{1}{2\cdot 4^i}\Big].$$

In particular, $\varphi_{\gamma_i}(1/3) \geq 1/2$ for all i, hence $\|\sum_{i\leq J}\varphi_{\gamma_i}\|_{L^\infty(\mathbb{R}^2)} \geq J/2$ for all J.

5.5 Connection with trees and meshes

In order to transport the numerical solutions along the flow, we will need to associate every (admissible) adaptive grid with a partition of the phase space. Moreover, our scheme will be based on tree algorithms. Hence we need to equip the dyadic grids with a tree structure. Here we describe how we shall do this.

First, we introduce the set of *-nodes of level $j \geq 1$ that correspond to a refinement of Γ_j in both directions,

$$\nabla_j^* := \big\{(2^{-j}(2k+1), 2^{-j}(2k'+1)) : k, k' \in \{0, 2^{j-1} - 1\}\big\} \subset \nabla_j,$$

and associate an (open) square cell to every *-node by setting

$$\Omega_\gamma := B_{\ell^\infty}(\gamma, 2^{-|\gamma|}) \qquad \forall\gamma \in \Gamma_\infty^* := \bigcup_{j\geq 1} \nabla_j^*.$$

Next, we equip the set of dyadic nodes with a tree structure by defining for every $\gamma \in \nabla_j$ one set of *children* in ∇_{j+1} as follows: If γ is a *-node of level j, we define its children as

$$\mathcal{C}(\gamma) := \{\gamma + 2^{-(j+1)}(l,l') : (l,l') \in \{-1,0,1\} \times \{-1,0,1\} \setminus (0,0)\} \subset \nabla_{j+1};$$

if $\gamma = (2^{-j}k, 2^{-(j-1)}k')$ (hence with k odd), we define its children as

$$\mathcal{C}(\gamma) := \{\gamma + 2^{-(j+1)}(l,0) : l \in \{-1,1\}\} \subset \nabla_{j+1};$$

and if $\gamma = (2^{-(j-1)}k, 2^{-j}k')$ (hence with k' odd), we define its children as

$$\mathcal{C}(\gamma) := \{\gamma + 2^{-(j+1)}(0,l') : l' \in \{-1,1\}\} \subset \nabla_{j+1}.$$

Finally, we say that λ is a *parent* of γ if $\gamma \in \mathcal{C}(\lambda)$. Note that this process partitions the levels in the sense that every dyadic node γ of positive level has one (and only one) parent $\mathcal{P}(\gamma)$, moreover $|\mathcal{P}(\gamma)| = |\gamma| - 1$. Now, as it can be checked, every parent of a *-node is also a *-node but the converse is not true, i.e., not every children of a *-node is a *-node itself. Hence we introduce the notion of **-children* and set

$$\mathcal{C}^*(\gamma) := \{\gamma + 2^{-(j+1)}(m,m') : (m,m') \in \{-1,1\}^2\} = \mathcal{C}(\gamma) \cap \nabla^*_{j+1}$$

for every $\gamma \in \nabla^*_j$. Let us adopt the following definition in the wavelet framework.

Definition 5.15. A grid $\Lambda \subset \Gamma_\infty$ is said to be a *W-tree* if it contains the coarse grid Γ_{j_0} and if it satisfies

$$\gamma \in \Lambda \quad \Longrightarrow \quad \mathcal{P}(\gamma) \subset \Lambda.$$

Clearly, the simplest way to build a tree Λ consists in starting from the coarsest grid Γ_{j_0} and adding recursively children, according to some criterion. Observe that by doing so, one also builds a non-uniform partition of $\mathbb{R}^2$, given by

$$M(\Lambda) := \{\Omega_\gamma : \gamma \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma^*_\infty\} \tag{5.36}$$

where

$$\mathcal{L}_{\text{out}}(\Lambda) := \{\gamma \notin \Lambda : \mathcal{P}(\lambda) \in \Lambda\}$$

denotes the set of *outer leaves* of the tree Λ. For later purposes, we will also need that the trees satisfy a stronger property.

Definition 5.16. A W-tree Λ is said to be *graded* if

$$\gamma \in \Lambda \cap \Gamma^*_\infty \quad \Longrightarrow \quad \{\mu \in \Gamma_{|\gamma|-1} : \Sigma_\mu \cap \Omega_\gamma \neq \emptyset\} \subset \Lambda,$$

see (5.5).

As we shall see, this definition is mostly motivated by the accuracy of the transported meshes. But it turns out that the use of graded trees is already imposed by the admissibility of the dyadic wavelet grids.

Lemma 5.17. *Every admissible wavelet grid is a graded W-tree.*

Proof. Clearly, every admissible grid is a W-tree (simply observe that the parent of γ is always contained in the stencil S_γ). Let us then show that for any $\gamma \in \Gamma^*_\infty$, $|\gamma| = j$, and $\mu \in \Gamma_{j-1}$, we have

$$\Omega_\gamma \cap \Sigma_\mu \neq \emptyset \iff \Omega_\gamma \subset \Sigma_\mu \iff \gamma \in (\Sigma_\mu)^\circ = B_{\ell^\infty}(\mu, 2^{-|\mu|}(2R-1)), \quad (5.37)$$

where $(\Sigma_\mu)^\circ$ denotes the interior of Σ_μ. Indeed, since

$$\Omega_\gamma := B_{\ell^\infty}(\gamma, 2^{-j}) = 2^{-(j-1)}(]k, k+1[\times]k', k'+1[)$$

with k, $k' \in \mathbb{N}$, and $\Sigma_\mu = \overline{B}_{\ell^\infty}(\mu, 2^{-|\mu|}(2R-1)) = 2^{-|\mu|}([m_1, m_2] \times [m_3, m_4])$ with $m_1, \dots, m_4 \in \mathbb{N}$, the equivalences in (5.37) easily follow from $|\mu| \leq j-1$.

Now assume in additon that γ belongs to an admissible grid Λ. Since $\lambda \in \nabla_j$ with $j \geq |\mu|+1$, according to (5.25) we see that (5.37) implies the existence of one minimal $\ell \geq |\mu|+1$ such that $\gamma \in \Sigma^{[\ell]}_{|\mu|,\mu}$, and this in turn implies that there exists $\gamma' \in S_\gamma \cap \Sigma^{[\ell-1]}_{|\mu|,\mu}$. By using that Λ is admissible, we see that γ' is also in Λ, and by repeating the argument we show that $\Lambda \cap \Sigma^{[|\mu|]}_{|\mu|,\mu} \neq \emptyset$, i.e., $\mu \in \Lambda$. Hence Λ is graded. □

In order to accurately transport the grids with a low-cost algorithm, we now introduce an important property which involve the *neighbors* of a *-node, i.e.,

$$\mathcal{N}(\gamma) := \{\mu \in \Gamma_\infty : \Sigma_\mu \cap \Omega_\gamma \neq \emptyset\}, \qquad \gamma \in \Gamma^*_\infty.$$

For the remainder of this lecture, we fix one positive constant $\kappa < 1$, and we remind that $\varepsilon > 0$ is an arbitrary tolerance.

Definition 5.18. We shall say that the W-tree Λ is *weakly ε-adapted* to $g \in \mathcal{C}(\mathbb{R}^2)$ if

$$|d_\mu(g)| \leq 2^{\kappa(|\gamma|-|\mu|)}\varepsilon \quad (5.38)$$

for all $\gamma \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma^*_\infty$ and all $\mu \in \mathcal{N}(\gamma) \setminus \Lambda$. If (5.38) holds for all $\mu \in \mathcal{N}(\gamma)$, we say that Λ is *strongly ε-adapted* to g.

We note that the foregoing criteria (5.38) is somehow related to the prediction strategy suggested by Albert Cohen, Sidi Mahmoud Kaber, Siegfried Müller and Marie Postel [12] in the context of wavelet-based finite volume schemes with guaranteed error estimates. A first property associated with these definitions is the following.

Lemma 5.19. *If Λ is admissible and weakly ε-adapted to g, then the associated interpolation error satisfies*

$$\|g - P_\Lambda g\|_{L^\infty(\mathbb{R}^2)} \lesssim \varepsilon$$

with a constant independent of g.

Proof. Let us begin with the following observation: If Λ is admissible (hence graded), then for any $\gamma \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma_\infty^*$ and $\mu \in \Gamma_{|\gamma|-2} = \Gamma_{|\mathcal{P}(\gamma)|-1}$ with $\mu \notin \Lambda$, we have $\Sigma_\mu \cap \Omega_\gamma \subset \Sigma_\mu \cap \Omega_{\mathcal{P}(\gamma)} = \emptyset$ since $\mathcal{P}(\gamma) \in \Lambda$. In other terms, we see that

$$|\mu| \geq |\gamma| - 1, \qquad \forall \mu \in \mathcal{N}(\gamma) \setminus \Lambda. \tag{5.39}$$

Now using that $\text{supp}(\varphi_\mu) \subset \Sigma_\mu$, we calculate

$$\begin{aligned}
\|g - P_\Lambda g\|_{L^\infty(\Omega_\gamma)} &= \Big\| \sum_{j \geq j_0+1} \sum_{\substack{\mu \in \nabla_j \setminus \Lambda \\ \Sigma_\mu \cap \Omega_\gamma \neq \emptyset}} d_\mu(g) \varphi_\mu \Big\|_{L^\infty(\Omega_\gamma)} \\
&\lesssim \sum_{j \geq j_0+1} \sup_{\substack{\mu \in \nabla_j \setminus \Lambda \\ \Sigma_\mu \cap \Omega_\gamma \neq \emptyset}} |d_\mu(g)| \lesssim \sum_{j \geq |\gamma|-1} \sup_{\substack{\mu \in \nabla_j \setminus \Lambda \\ \Sigma_\mu \cap \Omega_\gamma \neq \emptyset}} |d_\mu(g)| \\
&\lesssim \sum_{j \geq |\gamma|-1} 2^{\kappa(|\gamma|-j)} \varepsilon \lesssim \varepsilon
\end{aligned}$$

where the first inequality follows from Lemma 5.13, the second one from the above observation (which uses the gradedness of Λ), and the third one is the weak ε-adaptivity. The lemma follows from the fact that $\{\Omega_\gamma : \gamma \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma_\infty^*\}$ is a partition of $\mathbb{R}^2$. $\square$

6 Dynamic adaptivity

In Sections 4 and 5, we have defined tree-structured adaptive discretizations of $\mathcal{C}(\Omega)$, using multilevel meshes and wavelets, respectively. Note that wavelets have been constructed on the entire $\mathbb{R}^2$, therefore in order to represent a function $g \in \mathcal{C}(\Omega)$ in the wavelet basis we first need to extend it outside Ω. Due to the boundary conditions that we have considered in Theorem 2.2, we extend it by periodicity in the x dimension and by 0 in the v dimension. As the numerical flows is assumed to map Ω into itself, we note that this does not spoil the continuity of the numerical solutions inside Ω (but it might deteriorate the sparsity of smooth functions in the wavelet basis, at least in the vicinity of the boundary). Next, for both discretizations we have introduced a notion of ε-adaptivity to a given function and we have seen that interpolating on the corresponding adaptive mesh, respectively wavelet grid, is $C\varepsilon$ accurate in the supremum norm. In this section, we describe the algorithms that allow to build an ε-adapted mesh or grid to a given function and that transport these meshes along a smooth flow, while conserving the property of being adapted to the transported solution. We also describe algorithms that build graded refinements of given trees.

6.1 Adapting the trees

Remember that in the multilevel mesh case, the FE-trees Λ consist of indices corresponding to dyadic quadrangles and that the associated meshes $M(\Lambda)$ are defined as the inner leaves of Λ, see (4.10). In the wavelet case, the W-trees Λ consist of dyadic points and the associated quad-meshes $M(\Lambda)$, see (5.36), are defined as the outer leaves of the subtree consisting of the *-nodes of Λ, and has an FE-tree structure.

For constructing adapted multilevel meshes, we will use the following algorithm.

Algorithm 6.1 ($\mathbb{A}_\varepsilon^{\mathrm{FE}}(g)$: ε-adaption of FE-trees). Starting from $\Lambda_0 := \mathcal{I}_0$, set

$$\Lambda_{\ell+1} := \Lambda_\ell \cup \big\{\beta \in \mathcal{C}^*(\alpha) : \alpha \in M(\Lambda_\ell) \text{ such that } |g|_{W^{2,1}(\alpha)} > \varepsilon\big\}$$

for $\ell = 0, 1, \dots$ until $\Lambda_{L+1} = \Lambda_L$, and finally set $\mathbb{A}_\varepsilon^{\mathrm{FE}}(g) := \Lambda_L$.

The following algorithm builds a W-tree (only composed of *-nodes) that is strongly ε-adapted to g.

Algorithm 6.2 ($\mathbb{A}_\varepsilon^{\mathrm{W}}(g)$: strong ε-adaption of W-trees). Starting from $\Lambda_0^* := \Gamma_{j_0}^*$, set

$$\Lambda_{\ell+1}^* := \Lambda_\ell^* \cup \big\{\gamma \in M(\Lambda_\ell^*) : \max\{2^{\kappa|\mu|}|d_\mu(g)| : \mu \in \mathcal{N}(\gamma)\} > 2^{\kappa|\gamma|}\varepsilon\big\}$$

for $\ell = 0, 1, \dots$ until $\Lambda_{L+1}^* = \Lambda_L^*$, and finally set $\mathbb{A}_\varepsilon^{\mathrm{W}}(g) := \Lambda_L^*$.

As the resulting adapted trees have no reason to be graded, we also need algorithms that build graded refinements of any given tree.

Algorithm 6.3 ($\mathbb{G}^{\mathrm{FE}}(\Lambda)$: graded refinement of FE-trees). Starting from $\Lambda_0 = \Lambda$, build

$$\Lambda_{\ell+1} := \Lambda_\ell \cup \big\{\lambda \in \mathcal{C}(\gamma) : \gamma \in \Lambda_\ell \cap \mathcal{Q}_\ell,\ \exists \mu \in \Lambda_\ell \cap \mathcal{Q}_{\ell+2},\ \overline{\Omega}_\gamma \cap \overline{\Omega}_\mu \neq \emptyset\big\}$$

for $\ell = 0, 1, \dots$ until $\Lambda_{L-1} = \Lambda_L$, and set $\mathbb{G}^{\mathrm{FE}}(\Lambda) := \Lambda_{L-1}$.

As graded W-trees have a structure which involves stencils of possibly high order R, we use another approach for building them.

Algorithm 6.4 ($\mathbb{G}^{\mathrm{W}}(\Lambda)$: graded refinement of W-trees). Given any tree Λ, set

$$\mathbb{G}^{\mathrm{W}}(\Lambda) := \bigcup_{\gamma \in \Lambda} \big\{\gamma + 2^{-|\gamma|}(m_1, m_2) : m_1, m_2 \in \{-(2R-1), \dots, (2R-1)\}\big\}.$$

Remark 6.5. We could also give a variant of Algorithm 6.2 that builds *weakly* ε-adapted W-trees to some given g. According to Lemma 5.19, we know that interpolating g on the resulting grid (once graded) would be $C\varepsilon$ accurate, but this is not enough to ensure the accuracy of the adaptive semi-Lagrangian scheme (indeed, think of the case where g consists of one isolated basis function φ_γ).

Now, it is clear that the trees constructed by Algorithms 6.1 and 6.2 are ε-adapted in the sense of Definitions 4.3 and 5.18, respectively. It is also clear that Algorithm 6.3 yields a graded FE-tree in the sense of Definition 4.2, but the gradedness of W-trees resulting from Algorithm 6.4 is by no means obvious.

Lemma 6.6. *For any W-tree Λ, the resulting $\mathbb{G}^{\mathrm{W}}(\Lambda)$ is an admissible grid and hence a graded W-tree (according to Lemma 5.17).*

Proof. Let $\gamma \in \mathbb{G}^{\mathrm{W}}(\Lambda)$ and $\lambda \in \Lambda$ such that $\gamma_i = \lambda_i + 2^{-|\lambda|}m_i$ with $|m_i| \le 2R-1$, $i \in \{1,2\}$. In particular, we have $\gamma \in \Gamma_{|\lambda|}$, i.e., $j := |\gamma| \le |\lambda|$, and we can as well assume that $j = |\lambda|$. Indeed, if $j < |\lambda|$ then $\gamma \in \Gamma_{|\mathcal{P}(\lambda)|}$, i.e., there exist integers m_i', $i \in \{1,2\}$, such that $\gamma_i = \mathcal{P}(\lambda)_i + 2^{-|\mathcal{P}(\lambda)|}m_i'$. Moreover, we have

$$|m_i'| = 2^{j-1}|\gamma_i - \mathcal{P}(\lambda)_i| \le 2^{j-1}\big(|\gamma_i - \lambda_i| + |\lambda_i - \mathcal{P}(\lambda)_i|\big) \le \frac{|m_i|+1}{2} \le R \le 2R-1$$

therefore we can replace λ by its parent (which also belongs to Λ), hence assume $j = \lambda$. Now observe that any $\mu \in S_\gamma$ satisfies

$$\mu \in \Gamma_{j-1} \quad \text{and} \quad |\mu_i - \gamma_i| \le \begin{cases} 2^{-j}(2R-1) & \text{if } |\gamma_i| = j, \\ 0 & \text{if } |\gamma_i| < j. \end{cases}$$

Therefore, we can write $\mu = \mathcal{P}(\lambda) + 2^{-(j-1)}(k_1,k_2)$ and bound $|k_i|$ by

$$|k_i| = 2^{j-1}|\mu_i - \mathcal{P}(\lambda)_i| \le 2^{j-1}\big(|\mu_i - \gamma_i| + |\gamma_i - \lambda_i| + |\lambda_i - \mathcal{P}(\lambda)_i|\big).$$

We claim that $|k_i| \le 2R-1$, which, because of $\mathcal{P}(\lambda) \in \Lambda$, implies $\mu \in \mathbb{G}^{\mathrm{W}}(\Lambda)$. Indeed, we always have

$$|\mu_i - \gamma_i| \le 2^{-j}(2R-1), \quad |\gamma_i - \lambda_i| \le 2^{-j}|m_i| \le 2^{-j}(2R-1) \quad \text{and} \quad +|\lambda_i - \mathcal{P}(\lambda)_i| \le 2^{-j}$$

and observe that three cases may occur, according to $|\gamma| = |\lambda| = j$: either $|\gamma_i| < j$, in which case $|\mu_i - \gamma_i| = 0$, or $|\lambda_i| < j$, in which case $|\lambda_i - \mathcal{P}(\lambda)_i| = 0$, or $|\lambda_i| = |\gamma_i| = j$, in which case m_i must be even, hence $|\gamma_i - \lambda_i| \le 2^{-(j-1)}(R-1)$. By summing up, we find that

$$|\mu_i - \gamma_i| + |\gamma_i - \lambda_i| + |\lambda_i - \mathcal{P}(\lambda)_i| \le 2^{-(j-1)}(2R-1)$$

holds in every case, which ends the proof. □

6.2 Predicting the trees

For both discretizations, the algorithm that we shall use for predicting the meshes is based on partition trees (i.e., the leaves form a partition of Ω) and essentially consists in transporting the local space resolution along the (approximate) flow $\mathcal{B}$. We write it in terms of FE-trees, but already observe that it can also be applied to any W-tree Λ via its subtree $\Lambda \cap \Gamma_\infty^*$.

Algorithm 6.7 ($\mathbb{T}_{\mathcal{B}}^{\mathrm{FE}}(\Lambda)$: prediction of FE-trees). Starting from $\Lambda_0 := \mathcal{I}_{j_0}$, set

$$\Lambda_{\ell+1} := \Lambda_\ell \cup \big\{\mu \in \mathcal{C}^*(\gamma) : \gamma \in \mathcal{L}_{\mathrm{in}}(\Lambda_\ell), \min\{|\lambda| : \lambda \in \mathcal{L}_{\mathrm{in}}(\Lambda), \mathcal{B}(x_\gamma) \in \overline{\Omega}_\lambda\} > |\gamma|\big\}$$

(where x_γ denotes the center of Ω_γ) until $\Lambda_{L+1} = \Lambda_L$, and set $\mathbb{T}_{\mathcal{B}}^{\mathrm{FE}}(\Lambda) := \Lambda_L$.

As was previously said, the following variant for W-trees is almost the same algorithm. However, in order to establish the accuracy of the predicted grids, we now introduce a fixed parameter $\delta \in \mathbb{N}$ (the value of which will be chosen in the proof of Theorem 6.9), that correspond to a constant number of additional refinement levels. Remember that for any W-tree Λ, the set $\mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma^*_\infty$ consists of its outer *-leaves and that $\{\Omega_\lambda : \lambda \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma^*_\infty\}$ forms a partition of the phase space.

Algorithm 6.8 ($\mathbb{T}^{\text{W}}_{\mathcal{B}}(\Lambda)$: prediction of W-trees). Starting from $\Lambda^*_0 := \Gamma^*_{j_0}$, build

$$\Lambda^*_{\ell+1} := \Lambda^*_\ell \cup \big\{\gamma \in \mathcal{L}_{\text{out}}(\Lambda^*_\ell) \cap \Gamma^*_\infty : \min\{|\lambda| : \lambda \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma^*_\infty, \mathcal{B}(\gamma) \in \overline{\Omega}_\lambda\} > |\gamma| - \delta\big\}$$

until $\Lambda^*_{L+1} = \Lambda^*_L$, and set $\mathbb{T}^{\text{W}}_{\mathcal{B}}(\Lambda) := \Lambda^*_L$.

The main properties of these algorithms are summarized in the following theorem.

Theorem 6.9. *Let $\mathcal{B}$ be a diffeomorphism of Ω into itself, i.e., an invertible Lipschitz-continuous mapping with Lipschitz-continuous inverse. Then Algorithms 6.7 and 6.8 guarantee the accuracy of the predicted trees in the following sense:*

- *if Λ is a FE-tree ε-adapted to g, and if the flow $\mathcal{B}$ is stable in the sense of* (4.13)*, then the FE-tree $\mathbb{T}^{\text{FE}}_{\mathcal{B}}(\Lambda)$ is $C\varepsilon$-adapted to the $\mathcal{T}_{\mathcal{B}} g = g \circ \mathcal{B}$, with a constant C that only depends on $\mathcal{B}$;*
- *if Λ is a W-tree strongly ε-adapted to g, then the W-tree $\mathbb{T}^{\text{W}}_{\mathcal{B}}(\Lambda)$ is weakly $C\varepsilon$-adapted to $\mathcal{T}_{\mathcal{B}} g = g \circ \mathcal{B}$, with a constant C that only depends on $\mathcal{B}$.*

In addition, the cardinalities of the predicted trees are stable in the sense that

$$\#\big(\mathbb{T}^{\text{FE}}_{\mathcal{B}}(\Lambda)\big) \lesssim \#(\Lambda) \quad \textit{and} \quad \#\big(\mathbb{T}^{\text{W}}_{\mathcal{B}}(\Lambda)\big) \lesssim \#(\Lambda). \tag{6.1}$$

Proof. We shall only sketch the proof for the $C\varepsilon$-adaptivity of the predicted FE-trees (for details and for a proof of the complexity estimates (6.1), we refer to [7]). First, we introduce the set

$$\mathcal{I}(\Lambda, \mathcal{B}, \gamma) := \{\lambda \in \mathcal{L}_{\text{in}}(\Lambda) : \overline{\Omega}_\lambda \cap \mathcal{B}(\Omega_\gamma) \neq \emptyset\}$$

corresponding to the cells of the quad-mesh $M(\Lambda)$ that are even partly advected into Ω_γ. By using the gradedness of Λ and the smoothness of $\mathcal{B}$, one can show that there exists a constant C such that

$$\#\big(\mathcal{I}(\Lambda, \mathcal{B}, \gamma)\big) \leq C \qquad \forall \gamma \in \mathcal{L}_{\text{in}}\big(\mathbb{T}^{\text{FE}}_{\mathcal{B}}(\Lambda)\big). \tag{6.2}$$

According to the stability (4.13), we next estimate for any $\gamma \in \mathcal{L}_{\text{in}}(\mathbb{T}^{\text{FE}}_{\mathcal{B}}(\Lambda))$,

$$|g \circ \mathcal{B}|_{W^*(\Omega_\gamma)} \lesssim |g|_{W^*(\mathcal{B}(\Omega_\gamma))} \lesssim \sum_{\lambda \in \mathcal{I}(\Lambda, \mathcal{B}, \gamma)} |g|_{W^*(\overline{\Omega}_\lambda)} \lesssim \varepsilon,$$

where the second inequality follows from the fact that the cells $\overline{\Omega}_\lambda, \lambda \in \mathcal{I}(\Lambda, \mathcal{B}, \gamma)$, cover $\mathcal{B}(\Omega_\gamma)$, and the third inequality follows from the ε-adaptivity of Λ together with relation (6.2).

We now give a detailed proof of the weak $C\varepsilon$-adaptivity of the W-tree $\mathbb{T}^{\mathrm{W}}_{\mathcal{B}}(\Lambda)$. Let us recall that this amounts in proving that for any $\tilde\lambda \in \mathcal{L}_{\text{out}}\big(\mathbb{T}^{\mathrm{W}}_{\mathcal{B}}(\Lambda)\big) \cap \Gamma^*_\infty$ and any $\tilde\mu \in \mathcal{N}(\tilde\lambda) \setminus \mathbb{T}^{\mathrm{W}}_{\mathcal{B}}(\Lambda)$, we have

$$|d_{\tilde\mu}(g \circ \mathcal{B})| \le 2^{\kappa(\ell - j)} C\varepsilon \quad \text{where} \quad \ell := |\tilde\lambda|,\ \ j := |\tilde\mu|, \tag{6.3}$$

for a constant C that only depends on $\mathcal{B}$. By using (5.7) and Lemma 3.4, we obtain

$$|d_{\tilde\mu}(g \circ \mathcal{B})| \lesssim \omega_1(g \circ \mathcal{B}, 2^{-j}, \Sigma_{\tilde\mu})_\infty \lesssim \omega_1(g, 2^{-j}, \Sigma^{\mathcal{B}}_{\tilde\mu})_\infty \le \sum_{i \in \mathbb{N}} \omega_1(g_i, 2^{-j}, \Sigma^{\mathcal{B}}_{\tilde\mu})_\infty$$

where $\Sigma^{\mathcal{B}}_{\tilde\mu}$ denotes the set $(\Sigma_{\tilde\mu})^{\mathcal{B},\tau}$ with $\tau = 2^{-j}$, see Lemma 3.4, and where the "single layer" functions defined by

$$g_i := \sum_{\mu \in \mathcal{N}^{\mathcal{B}}_i(\tilde\mu)} d_\mu(g)\varphi_\mu \quad \text{with} \quad \mathcal{N}^{\mathcal{B}}_i(\tilde\mu) := \{\mu \in \nabla_i : \Sigma_\mu \cap \Sigma^{\mathcal{B}}_{\tilde\mu} \neq \emptyset\}$$

clearly satisfy $\sum_{i \in \mathbb{N}} g_i = g$ on $\Sigma^{\mathcal{B}}_{\tilde\mu}$. Note that since $\mathcal{B}$ is assumed to be Lipschitz, $\Sigma^{\mathcal{B}}_{\tilde\mu}$ is included in a ball of center $\tilde\mu$ and radius $C2^{-j}$ with a constant C depending on R only. Next we estimate the modulus of smoothness following rather classical techniques (see, e.g., [11, p. 183]): first, we observe that

$$\omega_1(g_i, 2^{-j}, \Sigma^{\mathcal{B}}_{\tilde\mu})_\infty \lesssim \|g_i\|_{L^\infty(\Omega)} \lesssim \sup_{\mu \in \mathcal{N}^{\mathcal{B}}_i(\tilde\mu)} |d_\mu(f)|$$

by using the definition of ω_1 and (5.28), and that

$$\begin{aligned} \omega_1(g_i, 2^{-j}, \Sigma^{\mathcal{B}}_{\tilde\mu})_\infty &\le \inf_{p \in \mathbb{Q}_0} \omega_1(g_i - p, 2^{-j}, \Sigma^{\mathcal{B}}_{\tilde\mu})_\infty \lesssim \inf_{p \in \mathbb{Q}_0} \|g_i - p\|_{L^\infty(\Sigma^{\mathcal{B}}_{\tilde\mu})} \\ &\lesssim 2^{-j}|g_i|_{W^{1,\infty}} \lesssim 2^{i-j} \sup_{\mu \in \mathcal{N}^{\mathcal{B}}_i(\tilde\mu)} |d_\mu(g)| \end{aligned}$$

by using the definition of ω_1, the Deny-Lions theorem and the Bernstein inequality (5.29). Note that the above estimates yield

$$|d_{\tilde\mu}(g \circ \mathcal{B})| \lesssim \sum_{i \in \mathbb{N}} \min\{2^{i-j}, 1\} \sup_{\mu \in \mathcal{N}^{\mathcal{B}}_i(\tilde\mu)} |d_\mu(g)|. \tag{6.4}$$

Next, we observe that there exists some $\lambda \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma^*_\infty$ such that $\mathcal{B}(\tilde\lambda) \in \overline{\Omega}_\lambda$ and

$$|\lambda| \le \ell - \delta, \tag{6.5}$$

(otherwise $\tilde\lambda$ would have been added to $\mathbb{T}^{\mathrm{W}}_{\mathcal{B}}(\Lambda)$), and we claim that any $\mu \in \mathcal{N}^{\mathcal{B}}_i(\tilde\mu)$ is a neighbor node of some leaf node adjacent to λ: in other terms, we claim that there exists $\gamma \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma^*_\infty$ that satisfies both

$$\overline{\Omega}_\gamma \cap \overline{\Omega}_\lambda \neq \emptyset \quad \text{and} \quad \mu \in \mathcal{N}(\gamma), \ \text{ i.e.,} \ \ \Sigma_\mu \cap \overline{\Omega}_\gamma \neq \emptyset. \tag{6.6}$$

Note that this would permit establishing the desired estimate (6.3): Indeed, if (6.6) holds then we have in particular $\lambda \in \mathcal{N}(\gamma) \setminus \Lambda$, and thus $|\gamma| \leq |\lambda| + 1$ according to (5.39). By using the strong ε-adaptivity of Λ to g, we write then

$$|d_\mu(g)| \leq 2^{\kappa(|\gamma|-|\mu|)}\varepsilon \lesssim 2^{\kappa(\ell-i)}\varepsilon \qquad \forall \mu \in \mathcal{N}_i^{\mathcal{B}}(\tilde{\mu}),$$

where we have used (6.5) in the second inequality. Inserting this estimate into (6.4), we finally find

$$|d_{\tilde{\mu}}(g \circ \mathcal{B})| \lesssim 2^{\kappa\ell - j}\varepsilon \sum_{i \leq j} 2^{(1-\kappa)i} + 2^{\kappa\ell}\varepsilon \sum_{i \geq j+1} 2^{-\kappa i} \lesssim 2^{\kappa(\ell-j)}\varepsilon$$

which is (6.3). It thus only remains to establish the claim (6.6), and for this purpose we let $\mathrm{dist}(A,B) := \inf_{a\in A, b \in B} \|a-b\|_{\ell^\infty}$ denote the minimal distance between two sets A, $B \subset \Omega$ (although this does not define a distance in the classical sense), for which we easily check that $\mathrm{dist}(A,B) = 0$ if and only if $\overline{A} \cap \overline{B} \neq \emptyset$, and $\mathrm{dist}(A,B) \leq \mathrm{dist}(A,C) + \mathrm{dist}(B,C) + \mathrm{diam}(C)$. By observing that none of the intersections

$$\Sigma_\mu \cap \Sigma_{\tilde{\mu}}^{\mathcal{B}}, \qquad \Sigma_{\tilde{\mu}}^{\mathcal{B}} \cap \mathcal{B}(\Sigma_{\tilde{\mu}}), \qquad \mathcal{B}(\Sigma_{\tilde{\mu}}) \cap \mathcal{B}(\Omega_{\tilde{\lambda}}) \quad \text{or} \quad \mathcal{B}(\Omega_{\tilde{\lambda}}) \cap \Omega_\lambda$$

is empty, we calculate

$$\mathrm{dist}(\Sigma_\mu, \Omega_\lambda) \leq \mathrm{diam}(\Sigma_{\tilde{\mu}}^{\mathcal{B}}) + \mathrm{diam}(\mathcal{B}(\Sigma_{\tilde{\mu}})) + \mathrm{diam}(\mathcal{B}(\Omega_{\tilde{\lambda}})) \lesssim 2^{-j} + 2^{-\ell} \leq 2^{-\ell} C_{\mathcal{B}}$$

with a constant $C_{\mathcal{B}}$ that only depends on the smoothness of $\mathcal{B}$. Therefore, choosing $\delta \geq \ln_2(C_{\mathcal{B}})$ yields $\mathrm{dist}(\Sigma_\mu, \Omega_\lambda) \leq 2^{-\ell+\delta} \leq 2^{-|\lambda|}$. In order to conclude, we infer from (5.39) that because $\lambda \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma_\infty^*$, every mesh cell $\overline{\Omega}_{\lambda'}$, $\lambda' \in \mathcal{L}_{\text{out}}(\Lambda) \cap \Gamma_\infty^*$, touching $\overline{\Omega}_\lambda$ is an ℓ^∞ ball of diameter $2^{1-|\lambda'|} \geq 2^{-|\lambda|}$. In particular, Σ_μ is in contact with one such cell $\overline{\Omega}_{\lambda'}$. Call it $\overline{\Omega}_\gamma$: we have just proved (6.6). □

6.3 The "predict and readapt" semi-Lagrangian scheme

Let us summarize what we have seen so far: In Section 3.1, we have described a time splitting scheme for computing an approximated backward flow $\mathcal{B}[f_n]$ from a given approximation f_n to the exact solution $f(t_n)$. This allows to compute every point value of $\mathcal{T} f_n := f_n \circ \mathcal{B}[f_n]$. Next, we have defined in Sections 4 and 5 two tree-structured discretizations of finite element and wavelet type, respectively, both suitable for adaptive interpolations. Finally, we have introduced in Section 6 algorithms for (i) building graded FE- and admissible W-trees which are ε-adapted to a given function g, and (ii) given a computable flow $\mathcal{B}$, predicting trees that stay well-adapted to the transported $g \circ \mathcal{B}$. Note that in views of Theorem 6.9, the quality of being well-adapted to the transported solution is only preserved by the predicted trees up to a multiplicative constant C that might be larger than 1. Hence it is necessary to readapt the trees once in a while in order to guarantee that *all* the interpolation errors stay within a bound of the order ε.

The resulting semi-Lagrangian scheme, which involves the algorithms $\mathbb{A}_\varepsilon = \mathbb{A}_\varepsilon^{\text{FE}}$ or

$\mathbb{A}_\varepsilon^{\mathrm{W}}$, $\mathbb{G} = \mathbb{G}^{\mathrm{FE}}$ or $\mathbb{G}^{\mathrm{W}}$ and $\mathbb{T}_\mathcal{B} = \mathbb{T}_\mathcal{B}^{\mathrm{FE}}$ or $\mathbb{T}_\mathcal{B}^{\mathrm{W}}$, depending whether one desires to implement an adaptive multilevel mesh scheme or a wavelet scheme, is as follows. Given an initial datum f^0, compute first

$$f_0 := P_{\Lambda^0} f^0, \tag{6.7}$$

where P_{Λ^0} denotes the adaptive (finite element or wavelet) interpolation associated with

$$\Lambda^0 := \mathbb{G}(\mathbb{A}_\varepsilon(f^0)), \tag{6.8}$$

then for $n = 1, \dots, N = T/\Delta t$,

$$f_n := P_{\Lambda^n} P_{\Lambda_p^n} \mathcal{T} f_{n-1}, \tag{6.9}$$

where the predicted and readapted trees are given by

$$\Lambda_p^n := \mathbb{G}(\mathbb{T}_{\mathcal{B}[f_{n-1}]}(\Lambda^{n-1})) \tag{6.10}$$

and

$$\Lambda^n := \mathbb{G}(\mathbb{A}_\varepsilon(f_{p,n}) \quad \text{with} \quad f_{p,n} := P_{\Lambda_p^n} \mathcal{T} f_{n-1}, \tag{6.11}$$

respectively.

Remark 6.10. If the flow $\mathcal{B}[f_n]$ is defined by the splitting method described in Section 3.1, we need to compute an auxiliary electric field from the intermediate solution $\mathcal{T}_x^{1/2} f_n$. The adaptive semi-Lagrangian scheme, therefore, should decompose into substeps corresponding to every one-dimensional transport operator. Such a decomposition, and the corresponding error analysis (for first order interpolations), is carried out in [7].

6.4 Error and complexity estimates (results and conjectures)

Our main result is that the above adaptive semi-Lagrangian schemes – of either multilevel mesh or wavelet type with arbitrary interpolation order – satisfy the following error estimate.

Theorem 6.11. *Under Assumptions 3.2 and 3.3, the numerical solution given by the adaptive semi-Lagrangian scheme* (6.7)-(6.11) *satisfies*

$$\|f(t_n) - f_n\|_{L^\infty(\Omega)} \lesssim (\Delta t)^{r-1} + \varepsilon/\Delta t \tag{6.12}$$

for $n = 0, \dots, N = T/\Delta t$, as long as the initial datum f^0 is in $W^{1,\infty}(\Omega)$.

Of course, it remains to precisely determine which schemes and which initial data yield approximate flows satisfying Assumption 3.3. We shall leave this issue to further studies.

Proof. The arguments are similar to those in Section 3.2 and as we shall see, they also apply to high-order interpolations. Let us describe them in detail: again we decompose the error $e_{n+1} := \|f(t_{n+1}) - f_{n+1}\|_{L^\infty(\Omega)}$ into three parts, which are now as follows. A first term $e_{n+1,1} := \|f(t_{n+1}) - \mathcal{T} f(t_n)\|_{L^\infty(\Omega)}$ as in Section 3.2, bounded for the same reasons by

$$e_{n+1,1} \lesssim |f(t_n)|_{W^{1,\infty}(\Omega)} (\Delta t)^r \lesssim (\Delta t)^r.$$

Then a second term $e_{n+1,2} := \|(I - P_{\Lambda^{n+1}} P_{\Lambda_p^{n+1}}) \mathcal{T} f_n\|_{L^\infty(\Omega)}$ which again corresponds to the interpolation error, and a third term $e_{n+1,3} := \|\mathcal{T} f(t_n) - \mathcal{T} f_n\|_{L^\infty(\Omega)}$ which again represents the (nonlinear) transport of the numerical error at time step n. Now, note that this decomposition slightly differs from that of Section 3.2 in that the interpolation error now involves the numerical solution instead of the exact one.This is in order to exploit the main properties of the predicted grids, i.e., the $C\varepsilon$-adaptivity to $\mathcal{T} f_n$. The good news is that it allows the third term, which propagates the error from the previous time steps, *not* to involve the interpolation operator. Hence for any interpolation order, we have

$$\begin{aligned} e_{n+1,3} &= \big\| f(t_n) \circ \mathcal{B}[f(t_n)] - f(t_n) \circ \mathcal{B}[f_n] \big\|_{L^\infty(\Omega)} + \big\| (f(t_n) - f_n) \circ \mathcal{B}[f_n] \big\|_{L^\infty(\Omega)} \\ &\le |f(t_n)|_{W^{1,\infty}(\Omega)} \big\| \mathcal{B}[f(t_n)] - \mathcal{B}[f(t_n)] \big\|_{L^\infty(\Omega)} + e_n \le (1 + C\Delta t) e_n \end{aligned}$$

by only using the stability (3.3) of the mapping $\mathcal{B}[\cdot]$. As for the second term, we write

$$e_{n+1,2} \lesssim \|(I - P_{\Lambda_p^{n+1}}) \mathcal{T} f_n\|_{L^\infty(\Omega)} + \|(I - P_{\Lambda^{n+1}}) P_{\Lambda_p^{n+1}} \mathcal{T} f_n\|_{L^\infty(\Omega)} \lesssim \varepsilon.$$

Here we have used the fact that according to Theorem 6.9 and by construction, respectively, the trees Λ_p^{n+1} and Λ^{n+1} are $C\varepsilon$ and ε-adapted to $\mathcal{T} f_n$ and $P_{\Lambda_p^{n+1}} \mathcal{T} f_n$, respectively. The above inequality follows then from (4.15) (in the multilevel mesh case) or Lemma 5.13 (in the wavelet case). Note that in the wavelet case, the above properties of the predicted and readapted trees are weak and strong, respectively, which suffices to our purposes. The error estimate follows then by gathering the above bounds, and by applying a discrete Gronwall lemma. □

Remark 6.12. In contrast to what happens with uniform schemes, an a priori L^∞ bound is available for high-order adaptive semi-Lagrangian schemes where the interpolations may, in general, increase the supremum norm. Indeed, by denoting $P_n := P_{\Lambda^n} P_{\Lambda_p^n}$ we can always bound $\tilde{e}_n := \|\mathcal{T}^n f_0 - f_n\|_{L^\infty(\Omega)}$ by

$$\tilde{e}_n \le \|\mathcal{T}^n f_0 - \mathcal{T} f_{n-1}\|_{L^\infty(\Omega)} + \|(I - P_n) \mathcal{T} f_{n-1}\|_{L^\infty(\Omega)} \le (1 + C\Delta t) \tilde{e}_{n-1} + C' \varepsilon$$

by using arguments from the above proof (and in particular, with $C = 0$ in the case of a linear transport). This yields, again by employing a discrete Gronwall lemma,

$$\|f_n\|_{L^\infty(\Omega)} \le \|\mathcal{T}^n f_0\|_{L^\infty(\Omega)} + \tilde{e}_n \le \|f_0\|_{L^\infty(\Omega)} + C\varepsilon/\Delta t$$

with a constant independent of ε and Δt.

In addition to the above error estimate, we can bound the cardinalities of the predicted and readapted trees as follows:

$$\#(\Lambda^{n+1}) \lesssim \#(\Lambda_p^{n+1}) \lesssim \#(\Lambda^n),$$

but at the present stage we do not know how to estimate the growth of these cardinalities on the overall time period. Our conjecture is that in the case of first order interpolations, they stay bounded by $C\varepsilon^{-1}$ as in (4.5), hence balancing $\varepsilon \sim (\Delta t)^r$ in estimate (6.12) would yield

$$\sup_{n\le N} \|f(t_n) - f_n\|_{L^\infty(\Omega)} \lesssim \varepsilon^{1-\frac{1}{r}} \lesssim \sup_{n\le N} \big(\#(\Lambda^n)\big)^{\frac{1}{r}-1}.$$

This "adaptive" convergence rate should require less regularity than its "uniform" version (3.22) – involving the $W^{2,\infty}(\Omega)$-seminorm of f^0 – and this would express an advantage of the adaptive method over its uniform counterpart. As for high-order interpolations, we expect them to achieve high convergence rates. In any case, a rigorous complexity analysis of our scheme is still to be done.

Acknowledgments. It is a pleasure to thank my colleagues from Strasbourg, as well as Albert Cohen, for very fruitful discussions. I would also like to thank Petra Wittbold and Etienne Emmrich for their kind invitation at the Technishe Universität Berlin, and for the careful proof-reading wich significantly improved the presentation of this work.

References

[1] R.A. Adams and J.J.F. Fournier, *Sobolev spaces*, second ed, Pure and Applied Mathematics 140, Elsevier/Academic Press, Amsterdam, 2003.

[2] N. Besse, Convergence of a semi-Lagrangian scheme for the one-dimensional Vlasov-Poisson system, *SIAM J. Numer. Anal.* **42** (2004), pp. 350–382.

[3] N. Besse, F. Filbet, M. Gutnic, I. Paun and E. Sonnendrücker, *An adaptive numerical method for the Vlasov equation based on a multiresolution analysis*, Numerical Mathematics and Advanced Applications ENUMATH 2001 (F. Brezzi, A. Buffa, S. Escorsaro and A. Murli, eds.), Springer, 2001, pp. 437–446.

[4] N. Besse and M. Mehrenberger, Convergence of classes of high-order semi-Lagrangian schemes for the Vlasov-Poisson system, *Math. Comput.* **77** (2008), pp. 93–123.

[5] Ju.A. Brudnyĭ, Approximation of functions of n variables by quasi-polynomials, *Izv. Akad. Nauk SSSR Ser. Mat.* **34** (1970), pp. 564–583.

[6] M. Campos Pinto and M. Mehrenberger, *Adaptive numerical resolution of the Vlasov equation*, Numerical methods for hyperbolic and kinetic problems, CEMRACS 2003/IRMA Lectures in Mathematics and Theoretical Physics (S. Cordier, T. Goudon, M. Gutnic and E. Sonnendrücker, eds.), European Mathematical Society, 2005.

[7] ______, Convergence of an adaptive semi-Lagrangian scheme for the Vlasov-Poisson system, *Numer. Math.* **108** (2008), pp. 407–444.

[8] A.S. Cavaretta, W. Dahmen and C.A. Micchelli, Stationary subdivision, *Mem. Amer. Math. Soc.* **93** (1991).

[9] C.Z. Cheng and G. Knorr, The integration of the Vlasov equation in configuration space, *J. Comput. Phys.* **22** (1976), pp. 330–351.

[10] P.G. Ciarlet, *Basic error estimates for elliptic problems*, Handbook of numerical analysis, Vol. II, North-Holland, Amsterdam, 1991, pp. 17–351.

[11] A. Cohen, *Numerical analysis of wavelet methods*, 32, North-Holland, Amsterdam, 2003.

[12] A. Cohen, S.M. Kaber, S. Müller and M. Postel, Fully adaptive multiresolution finite volume schemes for conservation laws, *Math. Comp.* **72** (2003), pp. 183–225.

[13] J. Cooper and A. Klimas, Boundary value problems for the Vlasov-Maxwell equation in one dimension, *J. Math. Anal. Appl.* **75** (1980), pp. 306–329.

[14] G.-H. Cottet and P.-A. Raviart, Particle methods for the one-dimensional Vlasov-Poisson equations, *SIAM J. Numer. Anal.* **21** (1984), pp. 52–76.

[15] I. Daubechies, *Ten lectures on wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics 61, SIAM, Philadelphia, 1992.

[16] G. Deslauriers and S. Dubuc, Symmetric iterative interpolation processes, *Constr. Approx.* **5** (1989), pp. 49–68, Fractal approximation.

[17] R. DeVore, Nonlinear approximation, *Acta Numerica* **7** (1998), pp. 51–150.

[18] N. Dyn and D. Levin, Subdivision schemes in geometric modelling, *Acta Numer.* **11** (2002), pp. 73–144.

[19] R.T. Glassey, *The Cauchy problem in kinetic theory*, SIAM, Philadelphia, 1996.

[20] M. Gutnic, M. Haefele, I. Paun and E. Sonnendrücker, Vlasov simulations on an adaptive phase-space grid, *Comput. Phys. Comm* **164** (2004), pp. 214–219.

[21] S.V. Iordanskii, The Cauchy problem for the Kinetic Equation of Plasma, *Amer. Math. Soc. Transl. Ser. 2* **35** (1964), pp. 351–363.

[22] G. Rein, *Collisionless kinetic equations from astrophysics – The Vlasov-Poisson System*, Handbook of Differential Equations, Evolutionary Equations, Vol. 3 (C.M. Dafermos and E. Feireisl, eds.), Elsevier, Oxford, 2005.

[23] E. Sonnendrücker, J. Roche, P. Bertrand and A. Ghizzo, The semi-Lagrangian method for the numerical resolution of the Vlasov equation, *J. Comput. Phys.* **149** (1999), pp. 201–220.

[24] A.A. Vlasov, A new formulation of the many particle problem (Russian), *Akad. Nauk SSSR. Zhurnal Eksper. Teoret. Fiz.* **18** (1948), pp. 840–856.

Author information

Martin Campos Pinto, IRMA institute (Université de Strasbourg & CNRS), 7 rue René Descartes, 67084 Strasbourg, France.
E-mail: campos@math.u-strasbg.fr