Discrete pseudo-differential operators and applications to numerical schemes

Erwan Faou [1] and Benoît Grébert [2]

[1] INRIA Rennes, Univ Rennes & Institut de Recherche Mathématiques de Rennes,
CNRS UMR 6625 Rennes & ENS Rennes, France.
Campus Beaulieu F-35042 Rennes Cedex, France
email: Erwan.Faou@inria.fr

[2] Laboratoire de Mathématiques Jean Leray, Université de Nantes,
2, rue de la Houssinière F-44322 Nantes cedex 3, France.
email: benoit.grebert@univ-nantes.fr

September 28, 2021

Abstract

We define a class of discrete operators acting on infinite, finite or periodic sequences mimicking the standard properties of pseudo-differential operators. In particular we can define the notion of order and regularity, and we recover the fundamental property that the commutator of two discrete operators gains one order of regularity. We show that standard differential operators acting on periodic functions, finite difference operators and fully discrete pseudo-spectral methods fall into this class of discrete pseudo-differential operators. As examples of practical applications, we revisit standard error estimates for the convergence of splitting methods, obtaining in some Hamiltonian cases no loss of derivative in the error estimates, in particular for discretizations of general waves and/or water-waves equations. Moreover, we give an example of preconditioner constructions inspired by normal form analysis to deal with the similar question for more general cases.

Contents

1 Introduction

The usefulness of pseudodifferential calculus is no longer in question [Hör87, Tay81] and if forms one of the corner stone of the analysis of Partial Differential Equations (PDEs). Indeed, even for simple differential operators, the notion of inverse, flow, approximations and any functional calculus in term of the operator can be expressed in term of pseudodifferential operators.

The undeniable power of this calculus resides, in particular, in a property of the commutators: if two pseudodifferential operators A and B are of order r_1 and r_2 respectively, then the order of the commutator $[A, B]$ is $r_1 + r_2 - 1$ and not $r_1 + r_2$ like the product AB. This property plays a central role in the consistency of the definition of pseudodifferential calculus, and is fundamental for obtaining estimates allowing to define the existence of solutions or to analyze their long time behavior. To cite a few examples, it has in particular been used in normal forms theory (see for instance [BGMR20]) but also in the definition solution with low regularity [DL89] or on the analysis of scattering effects [GV85] which often rely on commutator estimates.

Unfortunately, when we represent these operators in a Hilbert basis of the space where they act (typically the Fourier basis), we obtain matrices and we lose this property: the regularity of the operators leads to a certain decrease of the coefficients of the matrices which represent them, but this notion of decrease is usually not quantified in such a way that the "miracle of the commutators" is preserved. We thus lose an important property of the differential operators when we model them

by a numerical scheme. Despite this difficulty, commutator estimates plays a fundamental role in numerical analysis, in particular for splitting methods, see [JL00, Fao12, CCFM17], or Magnus expansions and Baker-Cambell-Hausdorf formula [HLW06]. But in general classical estimates can only be obtained in specific explicit cases and yield in general to loss of derivatives (although striking recent progresses has been done in this direction, see for instance [ORS21]).

Otis Chodosh overcame this difficulty in his thesis (see [Cho10, Cho11]) by characterizing the space of infinite matrices representing pseudo-differential operators. This space encodes a notion of order which leads to a property of commutators similar to the one enjoyed by the space of pseudo-differential operators. It is this space of infinite matrices that we take up here (see Definition 2.1) and that we call space of pseudo-differential matrices. In passing we re-demonstrate the "miracle of the commutators" in a direct way. We then show that the standard pseudo-differential operators used in the world of PDEs are represented by matrices of this type (see Section 2.3). One of the main result of this paper is then to extend Chodosh's class to a context of *periodic matrices* which allows us to show that standard discretized models of PDEs including finite difference methods, spectral or pseudo-spectral methods, are in fact represented by pseudo-differential periodic matrices (see section 3) and satisfy the wished commutator estimates. Of course to be useful, the important point is that these estimates are *uniform* with respect to the discretization parameters, to be consistent with the continuous limit. With these results in hand, we expect to revisit and extend many results of analysis and numerical analysis within this new framework. It include particular space discretized equations (or more generally lattice dynamics like FPUT or Discrete nonlinear Schrödinger equations), time discretization of PDEs (semi-discrete or fully discrete) or long time behavior of numerical schemes.

To highlight all the advantages we can get from this class of matrices, we give several examples where the discrete pseudo-differential algebra yields new results. In section 4.1, we revisit and amplify (see Theorem 4.1) a result of Jahnke and Lubich (see [JL00]) on error estimates in splitting schemes. Their result was based on an assumption of gain of regularity of the commutator between the two part of the splitting, which is immediately satisfied if the operators concerned are in our class. Thus while Jahnke and Lubich verify this assumption only in the case of the Schrödinger equation (or more generally PDEs where the commutators can be calculated explicitely, *i.e.* involve functions and classical differential operators) we can ensure with our technics that it is automatically satisfied for any reasonable PDE and their space discretizations. Moreover the result extends to operators of arbitrary orders without restrictions. As a consequence of this abstract analysis, we can characterize examples of situations where convergence of splitting schemes is guaranteed without any loss of regularity. This means that we can estimate the error committed by the scheme in the same regularity as that imposed on the solution we are estimating.

We notice that this latter situation occurs for splitting methods of high order as long as the modeled operator is of order strictly smaller than one (see Remark 4.3). We develop an example of application from the water waves models in section 4.3 where the linear water wave operator is a pseudo-differential operator of order $\frac{1}{2}$. Moreover, as a consequence of our analysis, the same result holds for fully discrete models obtained by spectral or finite difference approximations.

Nevertheless this constraint to start from an operator of order strictly smaller than one is very restrictive and clearly not satisfied in many interesting cases. As a second example of application of our discrete pseudo-differential calculs, in section 4.4, we show that by using technics inspired by normal forms theory (see in particular [BGMR20]) we can circumvent this constraint by changing the unknown. This change of variable acts as a *preconditioner* for the splitting scheme concerned and it is described by a discrete pseudo-differential operator that we can estimate. We develop this technique in the framework of the Schrödinger equation (see section 4.4) but we could describe the method in an abstract way in the formalism of the normal forms, as done for instance in [BGMR20]. In this paper, we focus on the emblematic example of the Schrödinger equation as proof of concept, and we propose in Proposition 4.9 a pre- and post- processed Lie-splitting scheme of order one to approximate the solutions of the Schrödinger equation (4.20) without loss of derivative.

Note that the use of our class of operators and commutator estimates could certainly be very useful for time dependent problems and the analysis of Magnus integrators, both from the convergence point of view or on the obtention of long time estimates. These analysis will be the subject of further studies.

In Section 4.2 we also consider as application of our analysis the problem of Sobolev norm growth for linear problem with time dependent potential. We give a fairly general result that applies both to continuous and space discretized models. This also shows that the assumptions made for splitting propagators are fulfilled in many situations.

Notations: In the following, $x \lesssim y$ means that $x \leq Cy$ for a constant independent of x and y, and $x \lesssim_\alpha y$ means that C depends on α.

Acknowledgement

During the preparation of this work the two authors benefited from the support of the Centre Henri Lebesgue ANR-11-LABX- 0020-01 and B.G. was supported by ANR -15-CE40-0001-02 "BEKAM" of the Agence Nationale de la Recherche. Furthermore B.G. thanks INRIA and particularly the MINGuS project for hosting him for a semester.

2 A class of pseudo differential matrices

We consider the space of summable squares of complex or real numbers $\ell^2(\mathbb{Z}^d)$ indexed by $\mathbb{Z}^d$, $d \in \mathbb{N}$ a positive integer. Typically, a sequence of this space represents (an approximation of) the Fourier coefficients of a function defined on a periodic torus $\mathbb{T}^d := (\mathbb{R}/2\pi\mathbb{Z})^d$, see Section 2.3 for examples.
For $s \in \mathbb{R}$ we define the discrete Sobolev space, h^s, by

$$h^s := \{(x_k)_{k\in\mathbb{Z}^d} \in \mathbb{C}^{\mathbb{Z}^d} \mid \sum_{k\in\mathbb{Z}^d} (1+|k|)^{2s}|x_k|^2 < +\infty\}$$

where we will use the norm $|k| = |k_1| + \cdots + |k_d|$ for elements $k = (k_1, \ldots, k_d)$ in $\mathbb{Z}^d$ and we equip this Hilbert space with its natural norm: $\|x\|_s = \left(\sum_{k\in\mathbb{Z}^d}(1+|k|)^{2s}|x_k|^2\right)^{\frac{1}{2}}$. We note that the dual of h^s is h^{-s}. An infinite matrix, $A : \mathbb{Z}^d \times \mathbb{Z}^d \mapsto \mathbb{C}$ is identified with the collection of its complex elements $A := \{A(m,n)\}_{(m,n)\in\mathbb{Z}^d\times\mathbb{Z}^d}$.

For an infinite matrices $A : \mathbb{Z}^d \times \mathbb{Z}^d \mapsto \mathbb{C}$ and $j \in \{1, \ldots, d\}$, we denote $A^{j,+}$ and $A^{j,-}$ the infinite matrices defined by

$$A^{j,+}(m,n) = A(m+e_j, n+e_j) \quad \text{and} \quad A^{j,-}(m,n) = A(m-e_j, n-e_j),$$

where e_j denotes the element of $\mathbb{Z}^d$ with components $(e_j)_n = \delta_{jn}$, the Kronecker symbol, for $n = 1, \ldots, d$. Then we define the following operators on matrices: $\Delta_j^+ A = A^{j,+} - A$ and $\Delta_j^- A = A^{j,-} - A$ and for $\alpha \in \mathbb{Z}$ we define

$$\Delta_j^\alpha = \begin{cases} (\Delta_j^+)^\alpha & \text{if} \quad \alpha \geq 0, \\ (\Delta_j^-)^{|\alpha|} & \text{if} \quad \alpha \leq 0. \end{cases}$$

By convention $\Delta_j^0 = \mathrm{Id}$. For $\alpha = (\alpha_j)_{j=1,\cdots,d} \in \mathbb{Z}^d$ we define the finite difference operator Δ^α acting on the set of infinite matrices $A : \mathbb{Z}^d \times \mathbb{Z}^d \mapsto \mathbb{C}$ by

$$\Delta^\alpha = \Delta_1^{\alpha_1} \cdots \Delta_d^{\alpha_d}.$$

Following the definition introduced by Chodosh (see [Cho10, Cho11]), we define:

Definition 2.1. *Let $r \in \mathbb{R}$ we define the class $\mathcal{A}_r$ of pseudo differential matrices of order r by: $A \in \mathcal{A}_r$ if for all $\alpha \in \mathbb{Z}^d$ and all $N \in \mathbb{N}$ there exists $C_{N,\alpha} \geq 0$ such that*

$$\left|(\Delta^\alpha A)(m,n)\right| \leq C_{N,\alpha}(1+|m|+|n|)^{r-|\alpha|}(1+|m-n|)^{-N}, \quad \forall m,n \in \mathbb{Z}^d$$

where $|m| = |m_1| + \cdots + |m_d|$ and $|\alpha| = |\alpha_1| + \cdots + |\alpha_d|$.

We denote

$$\mathcal{A} = \cup_{r\in\mathbb{R}} \mathcal{A}_r$$

which form a graded algebra, as the next lemma will show.
For $r \in \mathbb{R}$, $\mathcal{A}_r$ is a Frechet space when equipped with the family of semi-norms

$$\|A\|_{\alpha,N,r} := \sup_{m,n\in\mathbb{Z}^d} \frac{\left|(\Delta^\alpha A)(m,n)\right|(1+|m-n|)^N}{(1+|m|+|n|)^{r-|\alpha|}}.$$

Note that as $\Delta^\alpha A$ is linear in A, we immediately have the estimate

$$\|A+B\|_{\alpha,N,r} \leq \|A\|_{\alpha,N,r} + \|B\|_{\alpha,N,r}$$

and in particular, if A is of order r and B of order r', then $A+B$ is of order $\max(r,r')$.

2.1 Associated operators

Lemma 2.2. *Let $r \in \mathbb{R}$. An infinite matrix $A \in \mathcal{A}_r$ naturally defines a continuous operator, still denoted by A, from h^s to h^{s-r} for any $s \in \mathbb{R}$ via the formula*

$$(Ax)_m = \sum_{n\in\mathbb{Z}^d} A(m,n)x_n, \quad m \in \mathbb{Z}^d. \tag{2.1}$$

Furthermore we have

$$\|A\|_{\mathcal{L}(h^s,h^{s-r})} \leq C\|A\|_{0,|s|+|r|+d+1,r}$$

for some constant C depending only on d, r and s.

Proof. The proof is quite standard, for the sake of completeness, we include it.
Let $x \in h^s$ and $A \in \mathcal{A}_r$. We denote $\tilde{x}$ the element of $\ell^2(\mathbb{Z}^d) \equiv h^0$ defined by $\tilde{x}_k = (1+|k|)^s x_k$, $k \in \mathbb{Z}^d$. We have

$$\begin{aligned}\|Ax\|_{s-r}^2 &= \sum_n (1+|n|)^{2(s-r)} \Big| \sum_k A(n,k)x_k \Big|^2 \\ &\leq \|A\|_{0,N,r}^2 \sum_n \Big(\sum_k \frac{(1+|n|+|k|)^r (1+|n|)^{s-r}}{(1+|n-k|)^N (1+|k|)^s} |\tilde{x}_k| \Big)^2\end{aligned}$$

Now we use the fact that

$$(1+|n|+|k|)^r \lesssim_r (1+|n|)^r (1+|n-k|)^{|r|}$$

for all $r \in \mathbb{R}$. Indeed, for $r \geq 0$, it is a consequence of the triangle inequality

$$1+|n|+|k| \leq 1+2|n|+|n-k| \leq 2(1+|n|)(1+|n-k|)$$

and for $r < 0$, the inequality if equivalent to

$$(1+|n|+|k|)^{|r|} \gtrsim_r (1+|n|)^{|r|}(1+|n-k|)^{-|r|}$$

which is implied by

$$(1+|n|) \leq (1+|n|+|k|)(1+|n-k|)$$

which is true by the triangle inequality again. Similarly, we have $(1+|n|)^s \lesssim_r (1+|n-k|)^{|s|}(1+|k|)^s$, and from these inequalities, we deduce that

$$\|Ax\|_{s-r}^2 \lesssim_{r,s} \|A\|_{0,N,r}^2 \sum_n \Big(\sum_k \frac{|\tilde{x}_k|}{(1+|n-k|)^{N-|s|-|r|}} \Big)^2.$$

Now by using Young inequality for convolutions (see Lemma A.1 in appendix), we have for $N-|s|-|r| \geq d+1$

$$\sum_n \Big(\sum_k \frac{|\tilde{x}_k|}{(1+|n-k|)^{N-|s|-|r|}} \Big)^2 = \Big\| \Big(\frac{1}{(1+|k|)^{N-|s|-|r|}} \Big)_{k\in\mathbb{Z}^d} * \tilde{x} \Big\|_{\ell^2}^2 \lesssim \|\tilde{x}\|_{\ell^2}^2$$

as soon as $N-|s|-|r| \geq d+1$. For the smallest choice of N, we thus have

$$\|Ax\|_{s-r}^2 \leq C\|A\|_{0,N,r}^2 \|x\|_s^2$$

for some constant C depending only on d, r and s. □

2.2 Product and commutator

We generalize the matrix product as follows: let $A, B : \mathbb{Z}^d \times \mathbb{Z}^d \mapsto \mathbb{C}$ we define, when the series converge, the product AB by the formula

$$(AB)(m,n) = \sum_{k\in\mathbb{Z}^d} A(m,k)B(k,n) = \sum_{j=1}^{d}\sum_{k_j\in\mathbb{Z}} A(m, \sum_{i=1}^{d} k_i e_i)B(\sum_{i=1}^{d} k_i e_i, n).$$

We denote by $[A,B] := AB - BA$ the commutator of A and B.

We now introduce a convenient notion of interval of multi-indices: for $\alpha, \beta \in \mathbb{Z}^d$ we say that $\beta \in [0,\alpha]$ if for all $j = 1, \cdots, d$, $0 \le \beta_j \le \alpha_j$ or $\alpha_j \le \beta_j \le 0$ and we say that $\beta \in [0,\alpha)$ if for all $j = 1, \cdots, d$, $0 \le \beta_j \le \alpha_j - 1$ or $\alpha_j + 1 \le \beta_j \le 0$. The main result of this section is

Proposition 2.3. *Let $r_1, r_2 \in \mathbb{R}$.*

(i) The matrix product is a continuous map from $\mathcal{A}_{r_1} \times \mathcal{A}_{r_2} \to \mathcal{A}_{r_1+r_2}$. More precisely, for all $\alpha \in \mathbb{Z}^d$ and all $N \in \mathbb{N}$ there exists $C(\alpha, N, r_1, r_2) > 0$ such that

$$\|AB\|_{\alpha,N,r_1+r_2} \le C(\alpha,N,r_1,r_2)\Big(\sum_{\beta\in[0,\alpha]} \|A\|_{\beta,N+1+|r_2|,r_1}\Big)\Big(\sum_{\beta\in[0,\alpha]} \|B\|_{\beta,N+1+|r_1|,r_2}\Big). \tag{2.2}$$

(ii) The commutator gains one order : the map $\mathcal{A}_{r_1} \times \mathcal{A}_{r_2} \in (A,B) \mapsto [A,B] \in \mathcal{A}_{r_1+r_2-1}$ is continuous. More precisely, for all $\alpha \in \mathbb{Z}^d$ and all $N \in \mathbb{N}$ there exists $C(\alpha, N, r_1, r_2) > 0$ such that

$$\|[A,B]\|_{\alpha,N,r_1+r_2-1} \le C(\alpha,N,r_1,r_2)\Big(\sum_{\substack{\beta,\gamma\in[0,\alpha]\\ M_1,M_2\le 2N+|r_1|+|r_2|}} \|A\|_{\beta,M_1,r_1}\|B\|_{\gamma,M_2,r_2}\Big). \tag{2.3}$$

This result could be deduced from the original works of Chodosh [Cho10, Cho11] by using the correspondence between $\mathcal{A}_r$ and the class of pseudo-differential operators of order r on $\mathbb{T}^d$. Here we give a direct proof which gives the specified estimates (2.2) and (2.3) and will have the advantage to be directly carried over to numerical discretizations by finite difference, spectral or pseudo-spectral methods (see next Sections).

The direct proof of Proposition 2.3 requires some technical lemmas, the reader who is not interested in the technical aspects concerning the semi-norms introduced in Definition 2.1 can go directly to the section 2.3.

We begin with some algebraic calculus:

Lemma 2.4. *Let A, B two infinite matrices $A, B : \mathbb{Z}^d \times \mathbb{Z}^d \mapsto \mathbb{C}$ and $j = 1, \cdots, d$*

(i) $\Delta_j^+(AB) = \Delta_j^+ A\, B^{j,+} + A\, \Delta_j^+ B,$

(ii) $\Delta_j^-(AB) = \Delta_j^- A\, B^{j,-} + A\, \Delta_j^- B,$

(iii) $\Delta_j^+[A,B] = [\Delta_j^+ A, B] + [A, \Delta_j^+ B] + [\Delta_j^+ A, \Delta_j^+ B]$,

(iv) $\Delta_j^-[A,B] = [\Delta_j^- A, B] + [A, \Delta_j^- B] + [\Delta_j^- A, \Delta_j^- B]$.

Proof. It is just a calculus, for instance for (i):

$$\begin{aligned}
\Delta_j^+(AB)(m,n) &= \sum_k A(m+e_j,k)B(k,n+e_j) - \sum_k A(m,k)B(k,n) \\
&= \sum_k A(m+e_j,k+e_j)B(k+e_j,n+e_j) - \sum_k A(m,k)B(k,n) \\
&= \sum_k \Delta_j^+ A(m,k) B^{j,+}(k,n) + \sum_k A(m,k)\Big(B(k+e_j,n+e_j) - B(k,n)\Big) \\
&= (\Delta_j^+ A\ B^{j,+})(m,n) + (A\ \Delta_j^+ B)(m,n).
\end{aligned}$$

And for (iii):

$$\begin{aligned}
\Delta_j^+[A,B] &= \Delta_j^+(AB) - \Delta_j^+(BA) \\
&= \Delta_j^+ A\ B^{j,+} + A\ \Delta_j^+ B - \Delta_j^+ B\ A^{j,+} - B\ \Delta_j^+ A \\
&= \Delta_j^+ A\ \Delta_j^+ B + \Delta_j^+ A\ B + A\ \Delta_j^+ B - \Delta_j^+ B\ \Delta_j^+ A - \Delta_j^+ B\ A - B\ \Delta_j^+ A.
\end{aligned}$$

□

Lemma 2.5. *Let $A \in \mathcal{A}_{r_1}$ and $B \in \mathcal{A}_{r_2}$ for r_1 and r_2 in $\mathbb{R}$. Then*

$$\|AB\|_{0,N,r_1+r_2} \lesssim_{N,r_1,r_2} \|A\|_{0,N+d+|r_2|,r_1} \|B\|_{0,N+d+|r_1|,r_2}.$$

Proof. By definition we have for all $m, n \in \mathbb{Z}^d$

$$\begin{aligned}
&|(AB)(m,n)| \\
&\qquad \le \|A\|_{0,N+d+|r_2|,r_1} \|B\|_{0,N+d+|r_1|,r_2} \sum_k \frac{(1+|m|+|k|)^{r_1}(1+|n|+|k|)^{r_2}}{(1+|m-k|)^{N+|r_2|+d}(1+|n-k|)^{N+d+|r_1|}}.
\end{aligned}$$

So it suffices to prove that

$$\sum_k \frac{(1+|m|+|k|)^{r_1}(1+|n|+|k|)^{r_2}}{(1+|m-k|)^{N+|r_2|+d}(1+|n-k|)^{N+d+|r_1|}} \lesssim_{N,r_1,r_2} \frac{(1+|m|+|n|)^{r_1+r_2}}{(1+|m-n|)^N}. \tag{2.4}$$

To begin with, we deal with the denominators and the sum. Using that $(1+|m-k|)(1+|n-k|) \ge (1+|m-n|)$, we note that

$$\sum_{k\in\mathbb{Z}^d} \frac{1}{(1+|m-k|)^{N+d}(1+|n-k|)^{N+d}} \le (1+|m-n|)^{-N} \sum_{k\in\mathbb{Z}^d} \frac{1}{(1+|m-k|)^d(1+|n-k|)^d}.$$

But we have

$$\sum_{k\in\mathbb{Z}^d} \frac{1}{(1+|m-k|)^d(1+|n-k|)^d} = \sum_{k\in\mathbb{Z}^d} \frac{1}{(1+|k|)^d(1+|n-m-k|)^d}$$

We decompose the last sum into two parts according to $|n-m-k| \leq \frac{|k|}{2}$ or not. As $(1+|k|)(1+|m-n-k|) \geq (1+|m-n|)$ we obtain the bound

$$\sum_{|n-m-k|\leq\frac{|k|}{2}} \frac{1}{(1+|m-n|)^d} + \sum_{|n-m-k|>\frac{|k|}{2}} \frac{2^d}{(1+|k|)^{2d}} \lesssim_d 1$$

for all n and m, as in the first sum, we have $|k| \leq 2|n-m|$ and thus the number of term is $\mathcal{O}(|n-m|^d)$ up to constants depending on d. We thus see that to prove (2.4), it remains to prove that for all $k \in \mathbb{Z}^d$

$$(1+|m|+|k|)^{r_1} \lesssim_{r_1} (1+|m|+|n|)^{r_1}(1+|n-k|)^{|r_1|}$$

and similarly

$$(1+|n|+|k|)^{r_2} \lesssim_{r_2} (1+|m|+|n|)^{r_2}(1+|m-k|)^{|r_2|}.$$

Let us prove the first one. We consider two cases

- either $r_1 \geq 0$ then we use that, either $|k| \leq 2|n|$ or $|n-k| \geq |k|/2$ which leads to in both cases to $(1+|m|+|k|) \leq 2(1+|m|+|n|)(1+|n-k|)$;
- either $r_1 \leq 0$ then we use that $(1+|m|+|n|) \leq (1+|m|+|k|)(1+|n-k|)$ as in the proof of Lemma 2.2.

□

Lemma 2.6. *Let $A \in \mathcal{A}_{r_1}$ and $B \in \mathcal{A}_{r_2}$ then*

$$\|[A,B]\|_{0,K,r_1+r_2-1} \lesssim_{K,r_1,r_2} \sum_{\substack{|\alpha|+|\beta|=1 \\ N,M\leq 2(K+|r_1|+|r_2|+d+1)}} \|A\|_{\alpha,N,r_1} \|B\|_{\beta,M,r_2}. \tag{2.5}$$

Proof. Let $m, n \in \mathbb{Z}^d$, we have

$$\begin{aligned}[A,B](m,n) &= \sum_{k\in\mathbb{Z}^d} A(m,k)B(k,n) - B(m,k)A(k,n) \\ &= \sum_{\ell\in\mathbb{Z}^d} A(m,m+\ell)B(m+\ell,n) - B(m,n-\ell)A(n-\ell,n).\end{aligned}$$

Furthermore by using telescopic summation, we have

$$|B(m,n-\ell) - B(m+\ell,n)| \leq \sum_{\substack{j\in[0,\ell) \\ \alpha_i=\mathrm{sign}(\ell_i)}} (|\Delta_1^{\alpha_1}B| + \cdots + |\Delta_d^{\alpha_d}B|)(m+j,n+j-\ell)$$

and similarly

$$|A(n-\ell,n)-A(m,m+\ell)| \leq \sum_{\substack{j\in[0,m-n+\ell)\\ \alpha_i=\mathrm{sign}(m-n+\ell)}} (|\Delta_1^{\alpha_1}A|+\cdots+|\Delta_d^{\alpha_d}A|)(n-\ell+j,n+j).$$

Therefore we get

$$|[A,B](m,n)| \leq \sum_{\ell\in\mathbb{Z}^d} \Sigma_{1,\ell}+\Sigma_{2,\ell} \tag{2.6}$$

where

$$\Sigma_{1,\ell} = |A(n-\ell,n)| \sum_{\substack{j\in[0,\ell)\\ \alpha_i=\mathrm{sign}(\ell_i)}} (|\Delta_1^{\alpha_1}B|+\cdots+|\Delta_d^{\alpha_d}B|)(m+j,n+j-\ell)$$

and

$$\Sigma_{2,\ell} = |B(m+\ell,n)| \sum_{\substack{j\in[0,m-n+\ell)\\ \alpha_i=\mathrm{sign}(m-n+\ell)}} (|\Delta_1^{\alpha_1}A|+\cdots+|\Delta_d^{\alpha_d}A|)(n-\ell+j,n+j).$$

So it remains to estimate $\Sigma_{1,\ell}$ and $\Sigma_{2,\ell}$. We begin with $\Sigma_{1,\ell}$:

$$\Sigma_{1,\ell} \leq d\sum_{|\beta|=1} \|A\|_{0,2N,r_1}\|B\|_{\beta,2N,r_2} \sum_{j\in[0,\ell)} \frac{(1+|m+j|+|n+j-\ell|)^{r_2-1}(1+|n-\ell|+|n|)^{r_1}}{(1+|m-n+\ell|)^{2N}(1+|\ell|)^{2N}}.$$

We now want to choose N such that

$$\sum_{j\in[0,\ell)} \frac{(1+|m+j|+|n+j-\ell|)^{r_2-1}(1+|n-\ell|+|n|)^{r_1}}{(1+|m-n+\ell|)^{2N}(1+|\ell|)^{2N}} \lesssim_{K,r_1,r_2} \frac{(1+|m|+|n|)^{r_1+r_2-1}}{(1+|m-n|)^K(1+|\ell|)^{d+1}} \tag{2.7}$$

in such a way that

$$\sum_{\ell\in\mathbb{Z}^d} \Sigma_{1,\ell} \lesssim_{K,r_1,r_2,d} \sum_{|\beta|=1} \|A\|_{0,2N,r_1}\|B\|_{\beta,2N,r_2} \frac{(1+|m|+|n|)^{r_1+r_2-1}}{(1+|m-n|)^K}. \tag{2.8}$$

by summing in ℓ the series $\sum_\ell \frac{1}{(1+|\ell|)^{d+1}} < +\infty$. To prove (2.7), we first note that

$$(1+|m-n+\ell|)^2(1+|\ell|)^2 \geq (1+|m-n|)(1+|\ell|)$$

and thus

$$\begin{aligned}\sum_{j\in[0,\ell)} &\frac{(1+|m+j|+|n+j-\ell|)^{r_2-1}(1+|n-\ell|+|n|)^{r_1}}{(1+|m-n+\ell|)^{2N}(1+|\ell|)^{2N}}\\ &\leq \sum_{j\in[0,\ell)} \frac{(1+|m+j|+|n+j-\ell|)^{r_2-1}(1+|n-\ell|+|n|)^{r_1}}{(1+|m-n|)^{N}(1+|\ell|)^{N}}.\end{aligned}$$

On the other hand we have for $j \in [0, \ell)$

$$(1 + |m + j| + |n + j - \ell|) \leq (1 + |m| + |n| + |j| + |j - \ell|) \leq 2(1 + |m| + |n|)(1 + |\ell|)$$

and

$$\begin{aligned}(1 + |m| + |n|) &\leq 1 + |m + j| + |n + j - \ell| + |j| + |j - \ell| \\ &\leq 2(1 + |m + j| + |n + j - \ell|)(1 + |\ell|).\end{aligned}$$

This shows that

$$\frac{(1 + |m| + |n|)}{2(1 + |m - n|)(1 + |\ell|)} \leq (1 + |m + j| + |n + j - \ell|) \leq 2(1 + |m| + |n|)(1 + |\ell|)$$

and similarly

$$\frac{(1 + |m| + |n|)}{2(1 + |m - n|)(1 + |\ell|)} \leq (1 + |n - \ell| + |n|) \leq 2(1 + |m| + |n|)(1 + |\ell|).$$

By using these inequalities according to the sign of $r_2 - 1$ and r_1, this leads to (2.7) with $N = K + |r_1| + |r_2| + d + 1$. The estimate

$$\sum_{\ell \in \mathbb{Z}^d} \Sigma_{2,\ell} \lesssim_{K,r_1,r_2} \sum_{|\alpha|=1} \|A\|_{\alpha,2N,r_1} \|B\|_{0,2N,r_2} \frac{(1 + |m| + |n|)^{r_1+r_2-1}}{(1 + |m - n|)^K}. \tag{2.9}$$

is obtained in the same way for the same choice of N and thus, combining (2.6) with (2.8) and (2.9) we get (2.5). □

Proof of Proposition 2.3 Assertion *(i)* is a consequence of Lemma 2.5 combined with Lemma 2.4, assertions *(i)* and *(ii)* by noticing that $\|B^{j,+}\|_{\alpha,N,r} \lesssim_{\alpha,N,r} \|B\|_{\alpha,N,r}$ and a similar relation for $B^{j,-}$. Assertion *(ii)* is a consequence of Lemma 2.4 assertions *(iii)* and *(iv)*, and Lemma 2.6.

2.3 Representation of differential operators

Let $\mathbb{T}^d = (\mathbb{R}/(2\pi\mathbb{Z}))^d$ be the standard d-dimensional torus. A complex function $u : \mathbb{T}^d \to \mathbb{C}$ in $L^2(\mathbb{T}^d)$ is identified with its Fourier coefficients

$$\hat{u}(k) = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} e^{-ik\cdot x} u(x) \mathrm{d}x, \quad k \in \mathbb{Z}^d$$

and we have the correspondence, for $s \geq 0$,

$$\hat{u} \equiv (\hat{u}(k))_{k\in\mathbb{Z}^d} \in h^s \qquad \Longleftrightarrow \qquad \partial_x^\alpha u \in L^2(\mathbb{T}^d), \quad \alpha \in \mathbb{N}^d, \quad |\alpha| \leq s$$

where for a multiindex $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, $\partial_x^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}$.

For any function $\Phi : \mathbb{R}^d \to \mathbb{C}$, we define $\Phi(-i\partial_x)$ by the formula

$$\widehat{(\Phi(-i\partial_x)u)}(k) := \Phi(k)\hat{u}(k).$$

We notice that $\Phi(-i\partial_x)$ is a linear operator in ℓ^2 and we denote by A_Φ the corresponding diagonal matrix with components

$$A_\Phi(m,n) = \Phi(m)\delta_{mn}. \tag{2.10}$$

For a given function $V : \mathbb{T}^d \to \mathbb{C}$, we associate the operator

$$u(x) \mapsto V(x)u(x)$$

which, in Fourier, corresponds to the convolution operator

$$\widehat{(Vu)}(k) = \sum_{\ell \in \mathbb{Z}^d} \widehat{V}(k-\ell)\hat{u}(\ell).$$

We associate to the function V an infinite matrix B_V with components

$$B_V(m,n) = \widehat{V}(m-n), \quad m,n \in \mathbb{Z}^d \tag{2.11}$$

in such a way that we have

$$\widehat{(Vu)} = B_V\hat{u}.$$

Lemma 2.7. *Let $\Phi : \mathbb{R}^d \to \mathbb{C}$ and $V : \mathbb{T}^d \to \mathbb{C}$ two functions and A_Φ and B_V the matrices defined above.*

(i) If Φ is $\mathcal{C}^\infty$ and if there exists $r \in \mathbb{R}$ such that for all $\alpha \in \mathbb{N}^d$ and $x \in \mathbb{R}^d$, $|\partial_x^\alpha \Phi(x)| \lesssim_{r,\alpha} \langle x\rangle^{r-|\alpha|}$, then $A_\Phi \in \mathcal{A}_r$.

(ii) If V is $\mathcal{C}^\infty$ then $B_V \in \mathcal{A}_0$.

Proof. The operator A_Φ is diagonal and we have to prove that

$$|(\Delta^\alpha A_\Phi)(m,m)| \leq C_\alpha(1+|m|)^{r-|\alpha|}$$

for all $\alpha \in \mathbb{Z}^d$. In such an expression, Δ_j^+ and Δ_j^- are finite difference operators and acting only on Φ. By using classically Taylor estimates, we have for instance

$$(\Delta_j^+ A_\Phi)(m,m) = \Phi(m+e_j) - \Phi(m) = \partial_{x_j}\Phi(m) + \int_0^1 (1-t)\partial_{x_j}^2\Phi(m+te_j)\mathrm{d}t$$

which yields, by iterating this formula, estimates of the form

$$(\Delta^\alpha A_\Phi)(m,m) \lesssim_\alpha |\partial_x^{|\alpha|}\Phi(m)| + \sup_{y-m\in[-|\alpha|,|\alpha|]^d} |\partial_x^{|\alpha|+1}\Phi(y)|, \tag{2.12}$$

showing *(i)* under the assumption on Φ.

To prove *(ii)*, we first notice that $\Delta_j^+ B_V = \Delta_j^- B_V = 0$ as $B_V(m,n)$ depends only on $m-n$. Hence we only need to prove that for all N,

$$\left|B_V(m,n)\right| = \left|\hat{V}(m-n)\right| \lesssim_N (1+|m-n|)^{-N}, \quad \forall m,n \in \mathbb{Z}^d$$

which holds true for $V \in \mathcal{C}^\infty(\mathbb{T}^d)$. □

As a consequence of this result and Proposition (2.3), we obtain:

Corollary 2.8. *Let Φ_i, $i = 1,\dots,P$ functions satisfying condition (i) of the previous Lemma, for orders $r_i \in \mathbb{R}$, and V_i some smooth functions. Then*

$$A = \prod_{i=1}^{P} A_{\Phi_i} B_{V_i} \in \mathcal{A}_r, \quad \textit{with} \quad r = r_1 + \ldots + r_P. \tag{2.13}$$

With this result in hand, we see that all the standard pseudo-differential operators leads, in the Fourier side, to matrices belonging to a class $\mathcal{A}_r$ for a well chosen r[1]. For example:

- Fourier multipliers defined as polynomials of $D := -i\nabla_x$ which is the multiplication by $k \in \mathbb{Z}^d$ in Fourier. This includes the standard Laplace operator and linear KdV operator for instance.
- Transport operators of order one, of the form $u \mapsto \operatorname{div}(\rho(x)u)$ or $u \mapsto X(x)\cdot\nabla u$ for some smooth function ρ or smooth vector field X,
- Order two operators of the form $u \mapsto \operatorname{div}(a(x)\nabla u)$ for some smooth function a,
- All the pseudo-differential arising in fluid mechanics, for example the water wave operator

$$\Omega^2 := \frac{1}{\sqrt{\mu}}|D|\tanh(\sqrt{\mu}|D|) \tag{2.14}$$

  where μ is a small parameter and $|D|$ the Fourier multiplier $(|k_1|,|k_2|)$ in 2D. This operator encodes the pseudo-differential Dirichlet-to-Neumann operator arising in water wave theory. Note that the operator Ω is of order $r = \frac{1}{2}$.

We can also extend the definition of $\mathcal{A}_r$ to operators acting on vector fields with components in h^s:

Definition 2.9. *Let $p \geq 1$ be a given integer. Let $\boldsymbol{r} = (r_{ij})_{1\leq i,j\leq p}$ be a matrix of integers. We say that $\boldsymbol{A} \in \mathcal{A}_{\boldsymbol{b}}$ if $\boldsymbol{A} = (A_{ij})_{1\leq i,j\leq p}$ is p matrices of elements $A_{ij} \in \mathcal{A}_{r_{ij}}$ for all $i,j \in \{1,\dots,p\}$.*

[1] Actually this should also be recover from [Cho10].

We can then extend the notion of product and commutators and norms from the components A_{ij} to the system operator $\boldsymbol{A}$. In particular, the component of the product $(\boldsymbol{AB})_{i,j} = \sum_{k=1}^{p} A_{ik}B_{jk}$ is of order $\max_k(r_{ik} + r_{kj})$ and similar formula for the commutators.

For example for any vector field $\boldsymbol{u} = (u_1, u_2, u_3) : \mathbb{T}^d \to \mathbb{R}^3$ of zero average on $\mathbb{T}^d$, denoting by $\hat{u}_i(k)$, $i = 1, 2, 3$ the Fourier transform of its components, the Leray projection of this field onto the field of divergence free vector field is given by

$$P\boldsymbol{u} = \boldsymbol{u} - \nabla\Delta^{-1}(\nabla \cdot \boldsymbol{u})$$

and can be written

$$\widehat{(P\boldsymbol{u})}_i(k) = \hat{u}_i(k) - \sum_{j=1}^{3} \frac{k_i k_j}{|k|^2}\hat{u}_j(k), \quad k = (k_1, k_2, k_3) \in \mathbb{Z}^d.$$

It can also be written

$$\widehat{(P\boldsymbol{u})}_i = \sum_{j=1}^{3}(\delta_{ij} - A_{\Phi_{ij}})\widehat{u}_j,$$

with $\Phi_{ij}(0) = 0$, Φ_{ij} of class $\mathcal{C}^\infty$ and $\Phi_{ij}(x) = \frac{x_i x_j}{|x|^2}$, for $|x| > \frac{1}{2}$. Hence we see that P can be identified with a 3×3 matrix of infinite dimensional matrices belonging to $\mathcal{A}_{\mathbf{0}}$ as all the components of the matrix operator have order 0.

More generally, all the system of differential operators can be expressed as elements of $\mathcal{A}_{\boldsymbol{r}}$ for some matrix of orders, for instance elliptic system of Agmon, Douglis and Nirenberg type [ADN59] or hyperbolic systems such as system of conservation laws [Bre00].

2.4 Geometric conditions

In this section we consider subspaces of $\mathcal{A}_r$ that are stable by bracket and define Lie algebras. In these case, the flow operator is well defined and belong to an infinite dimensional Lie group. We give in sections 2.4.2 and 2.4.3 two examples but the reader can imagine many other situations. First we show in section 2.4.1 how we can encode different type of boundary condition in the matrix algebra $\mathcal{A}$.

2.4.1 Dirichlet boundary condition

On periodic functions, we can easily consider Dirichlet or Neumann boundary conditions by imposing some parity conditions. For example if $u : \mathbb{T}^d \to \mathbb{C}$ satisfies $u(-x) = -u(x)$, then $\hat{u}_k = \hat{u}_{-k}$ for $k \in \mathbb{Z}^d$, and u vanishes on the boundary of $\mathbb{T}^d$ represented as $[0, 2\pi]^d$, *i.e.* u satisfies Dirichlet boundary conditions on this domain.

In order to be able to consider operators preserving this boundary conditions, we define $\mathcal{A}^D$ the subclass of $\mathcal{A}$ defined by

$$A \in \mathcal{A}^D \iff A \in \mathcal{A} \quad \text{and} \quad A(-m, -n) = A(m, n),\ \forall m, n \in \mathbb{Z}^d.$$

We also define $h^s_{\rm odd}$ as the subspace of h^s formed by the odd sequences:

$$x \in h^s_{\rm odd} \iff x \in h^s \quad \text{and} \quad x_{-k} = -x_k,\ k \in \mathbb{Z}^d.$$

Matrices in $\mathcal{A}^D$ preserves oddness of sequences and is stable by multiplication and bracket. Thus as a consequence of Lemma 2.2 we get that every matrix $A \in \mathcal{A}^D_r$ naturally defines a continuous operator, still denoted by A, from $h^s_{\rm odd}$ to $h^{s-r}_{\rm odd}$ for any $s \in \mathbb{R}$ via the formula (2.1) and we have

$$\|A\|_{\mathcal{L}(h^s_{\rm odd}, h^{s-r}_{\rm odd})} \leq C \|A\|_{0,|s|+|r|+d+1,r}$$

for some constant C depending only on d.

We can of course also consider Neumann boundary conditions or mixed boundary conditions.

2.4.2 Hermitian operators

In order to consider equations of Schrödinger form, we define the set of Hermitian operators as the set of $H \in \mathcal{A}$ satisfying

$$H(m,n) = H^*(m,n) := \overline{H(n,m)}, \quad m,n \in \mathbb{Z}^d. \tag{2.15}$$

With obvious notations, we write this conditions $H = H^* := \overline{H}^T$ where the transpose matrix is defined by exchanging m and n in the coefficients. It is easy to check that for all Φ real-valued, the operator $H = A_\Phi$ is diagonal and Hermitian, and that $H = B_V$ is also Hermitian when V is a real function. More generally, representation of operator of the form

$$u \mapsto \operatorname{div}(\sigma(x)\nabla u) + X(x) \cdot \nabla u - \operatorname{div}(X(x)u) + V(x)u \tag{2.16}$$

for smooth vector field X with $\operatorname{div} X = 0$, and real functions σ and V, yields to Hermitian operators.

In section 4.4, we will consider the Hermitian operator $H = -\Delta + V$ for a smooth real function $V(x)$ on $\mathbb{T}^d$ and numerical schemes to approximate the solutions of the associate Schrödinger equation $i\partial_t u = (-\Delta + V)u$.

2.4.3 Symplectic matrices

We define symplectic systems as follows: for real matrices A, B and C (*i.e.* matrices with real coefficients) we set

$$\boldsymbol{S} = \begin{pmatrix} A & B \\ C & -A^T \end{pmatrix} \quad \text{with} \quad B^T = B \quad \text{and} \quad C^T = C, \tag{2.17}$$

which is an element of $\mathcal{A}_{\boldsymbol{r}}$ (see Definition (2.9)) with $\boldsymbol{r} = \begin{pmatrix} r(A) & r(B) \\ r(C) & r(A) \end{pmatrix}$ where $r(A)$, $r(B)$ and $r(C)$ denote the orders of A, B and C respectively. To $\boldsymbol{S}$ we associate the symplectic system on $\ell^2 \times \ell^2 \ni (p,q)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} p \\ q \end{pmatrix} = \boldsymbol{S}\begin{pmatrix} p \\ q \end{pmatrix}, \qquad \boldsymbol{S} = \begin{pmatrix} A & B \\ C & -A^T \end{pmatrix} \tag{2.18}$$

and, when it is well defined[2], its flow $e^{t\boldsymbol{S}}$ acting on $\ell^2 \times \ell^2$. This symplectic flow preserves the canonical symplectic form: defining $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ we have

$$(e^{t\boldsymbol{S}})^T J e^{t\boldsymbol{S}} = J.$$

A typical example is given by wave equations of the form $\partial_{tt} q - \Delta q = V(x)q$ that can be written

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 & \Delta + V \\ I & 0 \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix}$$

with $p = \partial_t$, where the right-hand sides defines a 2×2 system of order $\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$. Note however that this system can be also reformulated by using pseudo-differential transformations as a skew symmetric system of orders ≤ 1. We will give explicit examples in Section (4.3).

3 Periodic matrices and discretization

Let K be an even integer[3] and we define the grid points

$$x_a = \frac{2\pi a}{K}, \quad a \in \{-K/2, \dots, K/2 - 1\}^d := G_K \tag{3.1}$$

and we identify G_K with $(\mathbb{Z}/K\mathbb{Z})^d := \mathbb{Z}_K^d$ the set of equivalent class modulo K in each variable: to each $a \in \mathbb{Z}_K^d$ we associate $\hat{a}$ its unique representative within G_K. We set $h = \frac{2\pi}{K}$. The grid x_a can thus be written $x_a = ah \in \mathbb{T}^d$, $a \in \mathbb{Z}_K^d$. When this grid is used to discretize a function $u : \mathbb{T}^d \to \mathbb{C}$, we expect to approach $u(ah) \simeq u_a$, $a \in \mathbb{Z}_K^d$. Hence the function space is discretized by $u = (u_a)_{a \in \mathbb{Z}_K^d} \in \ell^2(\mathbb{Z}_K^d)$[4]. Very schematically a linear numerical scheme with mesh $h = \frac{2\pi}{K}$ is an application

$$\ell^2(\mathbb{Z}_K^d) \ni (u_a^K)_{a \in \mathbb{Z}_K^d} \mapsto (v_a^K)_{a \in \mathbb{Z}_K^d} \in \ell^2(\mathbb{Z}_K^d).$$

which can be represented by a periodic matrix M^K:

$$(v_a^K)_{a \in \mathbb{Z}_K^d} = M^K (u_a^K)_{a \in \mathbb{Z}_K^d}.$$

We thus naturally see the need of a concept of a *family* of periodic matrices M^K indexed by K, the number of points in our grid (or equivalently in the periodic case, by h the mesh size). Of course, as finite dimensional matrix, the norm of M^K is bounded but depends *a priori* on K. In order to evaluate the convergence of the scheme, or to study the global properties of numerical schemes at

[2] Using typically the fact that the unbounded part can be diagonalized explicitly in Fourier to define mild-solutions, an example is given below.

[3] Working with arbitrary integers is of course possible, by changing the structure of G_K according to the parity of K, see [Fao12]

[4] Notice that $\ell^2(\mathbb{Z}_K^d)$ is finite dimensional space equivalent to $\mathbb{C}^{K^d}$, we choose to equipped it with the ℓ_2-norm.

the continuous limit $K \to +\infty$, it will be essential to have norms on these families of matrices which are *uniform* in K. Moreover, the notion of pseudo-differential operator is well expressed in term of Fourier transform. Hence in the discrete case, we expect the Fourier transformation

$$A^K = \mathcal{F}_K^{-1} M^K \mathcal{F}_K$$

to inherit the commutator properties of continuous systems, uniformly in K. Here $\mathcal{F}_K$ stands for the discrete Fourier transform, see (3.4). This justifies the space of families of periodic matrices that we will define in the next section.

3.1 A class of families of periodic matrices

Notation 3.1. *Let $K \in 2\mathbb{N}^*$ an even integer. For $a \in \mathbb{Z}_K^d$, we denote by $\hat{a}$ its representative within $\{-K/2, \dots, K/2-1\}^d := G_K$ and we set $[a] = |\hat{a}_1| + \cdots + |\hat{a}_d|$.*

This quantity has some good properties:

Lemma 3.2. *For all $a, b, c \in \mathbb{Z}_K^d$ we have*

(i) $[a+b] \leq [a] + [b]$,

(ii) $(1 + [a] + [c]) \leq 2(1 + [a] + [b])(1 + [c-b])$.

Proof. (i) It suffices to consider the case $d = 1$. If $\hat{a} + \hat{b} \in G_K = \{-K/2, \dots, K/2-1\}$ then $\widehat{a+b} = \hat{a} + \hat{b}$ and thus $|\widehat{a+b}| \leq |\hat{a}| + |\hat{b}|$. If $\hat{a} + \hat{b} \notin G_K$ then $[a+b] = |\widehat{a+b}| \leq K/2 \leq |\hat{a} + \hat{b}| \leq [a] + [b]$.
(ii) is an easy consequence of *(i)*. □

We will consider familly of matrices $A^K : \mathbb{Z}_K^d \times \mathbb{Z}_K^d \mapsto \mathbb{C}$, indexed by K. For one give K, the matrix A^K is identified with the collection of its complex elements $A^K := \{A^K(m,n)\}_{(m,n) \in \mathbb{Z}_K^d \times \mathbb{Z}_K^d}$, where $A^K(m,n)$ is now K-periodic in m and n. For such a matrix, we extend easily the definition of Δ_j^+, Δ_j^- and Δ^α by periodicity.

Definition 3.3. *Let $r \in \mathbb{R}$. We define the class $\mathcal{A}_r^{\mathrm{per}}$ of families of periodic matrices of order r as follows: the family $A^\bullet = \{A^K\}_{K \in 2\mathbb{N}^*}$ belongs to $\mathcal{A}_r^{\mathrm{per}}$ if for all $\alpha \in \mathbb{Z}^d$ and all $N \in \mathbb{N}$ there exists $C_{N,\alpha} > 0$ such that*

$$\forall K \in \mathbb{N}, \quad \forall m, n \in \mathbb{Z}_K^d, \quad \left|(\Delta^\alpha A^K)(m,n)\right| \leq C_{N,\alpha} (1 + [m] + [n])^{r-|\alpha|} (1 + [m-n])^{-N}. \tag{3.2}$$

We denote $\mathcal{A}^{\mathrm{per}} = \cup_{r \in \mathbb{R}} \mathcal{A}_r^{\mathrm{per}}$ which form a graded algebra. We define the adapted family of semi-norms: for $A = \{A^K\}_{K \in 2\mathbb{N}^*}$,

$$[\![A^\bullet]\!]_{\alpha,N,r} := \sup_{K \in \mathbb{N}} \sup_{m,n \in \mathbb{Z}_K^d} \frac{\left|(\Delta^\alpha A^K)(m,n)\right| (1 + [m-n])^N}{(1 + [m] + [n])^{r-|\alpha|}}.$$

Remark 3.4. *In Equation (3.2) it is important to notice that the matrices K are of size $K \times K$, and the norm $[a]$ also depend on K. However, the key property is that N, α and the constant $C_{N,\alpha}$ are assumed to be uniform in K.*

We then define the product and commutator as follows: for $A^\bullet = \{A^K\}_{K\in 2\mathbb{N}*}$ and $B^\bullet = \{B^K\}_{K\in 2\mathbb{N}^*}$ in $\mathcal{A}^{\rm per}$, we set

$$A^\bullet B^\bullet = \{A^K B^K\}_{K\in 2\mathbb{N}^*} \quad \text{and} \quad [A^\bullet, B^\bullet] = \{\,[A^K, B^K]\,\}_{K\in 2\mathbb{N}^*}.$$

Thanks to Lemma 3.2, the proofs of Lemmas 2.5 and 2.6 are transposable, *mutatis mutandis*, to the periodic case. Thus we obtain

Proposition 3.5. *Let $r_1, r_2 \in \mathbb{R}$.*

(i) The matrix product is a continuous map from $\mathcal{A}^{\rm per}_{r_1} \times \mathcal{A}^{\rm per}_{r_2} \ni (A^\bullet, B^\bullet) \mapsto A^\bullet B^\bullet \in \mathcal{A}^{\rm per}_{r_1+r_2}$. More precisely, for all $\alpha \in \mathbb{Z}^d$ and all $N \in \mathbb{N}$ there exists $C(\alpha, N, r_1, r_2) > 0$, independent of K, such that

$$[\![A^\bullet B^\bullet]\!]_{\alpha,N,r_1+r_2} \leq C(\alpha, N, r_1, r_2)\Big(\sum_{\beta\in[0,\alpha]} [\![A^\bullet]\!]_{\beta,N+1+|r_2|,r_1}\Big)\Big(\sum_{\beta\in[0,\alpha]} [\![B^\bullet]\!]_{\beta,N+1+|r_1|,r_2}\Big).$$

(ii) The commutator gains one order : the map $\mathcal{A}^{\rm per}_{r_1} \times \mathcal{A}^{\rm per}_{r_2} \ni (A, B) \mapsto [A^\bullet, B^\bullet] \in \mathcal{A}^{\rm per}_{r_1+r_2-1}$ is uniformly continuous in K. More precisely, for all $\alpha \in \mathbb{Z}^d$ and all $N \in \mathbb{N}$ there exists $C(\alpha, N, r_1, r_2) > 0$, independent of K, such that

$$[\![\,[A^\bullet, B^\bullet]\,]\!]_{\alpha,N,r_1+r_2-1} \leq C(\alpha, N, r_1, r_2)\Big(\sum_{\substack{\beta,\gamma\in[0,\alpha]\\ M_1,M_2\leq 2N+|r_1|+|r_2|}} [\![A^\bullet]\!]_{\beta,M_1,r_1}\,[\![B^\bullet]\!]_{\gamma,M_2,r_2}\Big). \tag{3.3}$$

Families of matrices $A^\bullet = \{A^K\}_{K\in 2\mathbb{N}^*}$ in $\mathcal{A}^{\rm per}$ define naturally families of finite dimensional operators A^K on

$$\ell^2(\mathbb{Z}^d_K) := \{(x_k)_{k\in\mathbb{Z}^d} \in \mathbb{C}^{\mathbb{Z}^d} \mid x_k = x_j \text{ when } k_i \equiv j_i \bmod K \text{ for } i = 1, \cdots, d\} = \mathbb{C}^{\mathbb{Z}^d_K}.$$

for each $K \in 2\mathbb{N}^*$ via the formula

$$(A^K x)_a = \sum_{k\in\mathbb{Z}^d_K} A^K(a, k)x_k, \quad a \in \mathbb{Z}^d_K.$$

The space $\ell^2(\mathbb{Z}^d_K)$ has finite dimension nevertheless Lemma 2.2 has an interesting counterpart in the periodic case: for $x \in \ell^2(\mathbb{Z}^d_K)$ we denote

$$\|x\|^2_{s,K} = \sum_{k\in\mathbb{Z}^d_K} (1 + [k])^{2s}|x_k|^2,$$

and hence $\|x\|_{\ell^2(\mathbb{Z}_K^d)} = \|x\|_{0,K}$. Of course, since $\ell^2(\mathbb{Z}_K^d)$ has finite dimension, all the norms $\|\cdot\|_{s,K}$ are equivalent on $\mathbb{C}^{\mathbb{Z}_K^d}$, but with constant depending on K, for instance

$$\|x\|_{s,K} \leq (1+|K|)^s \|x\|_{0,K}$$

however, with the help of Definition 3.3, we have with the same proof as Lemma 2.2

Lemma 3.6. *Let $r, s \in \mathbb{R}$ and $A^\bullet = \{A^K\}_{K\in 2\mathbb{N}^*} \in \mathcal{A}_r^{\mathrm{per}}$ we have*

$$\forall\, K \in 2\mathbb{N}^* \quad \|A^K x\|_{s-r,K} \leq C[\![A^\bullet]\!]_{0,|s|+|r|+d+1,r} \|x\|_{s,K}$$

for some constant C depending only on d (and thus independent of K).

3.2 Representation of finite different schemes

For $j = 1, \ldots, d$, we define the finite difference operators from $\ell^2(\mathbb{Z}_K^d)$ into $\ell^2(\mathbb{Z}_K^d)$:

$$(\delta_{j,K}^+ u)_a = \frac{u_{a+e_j} - u_a}{h} \quad \text{and} \quad (\delta_{j,K}^- u)_a = \frac{u_a - u_{a-e_j}}{h}, \quad i = 1, \ldots, d, \quad h = \frac{2\pi}{K}$$

Another important tool is the discrete Fourier transform: $\mathcal{F}_K : \ell^2(\mathbb{Z}_K^d) \to \ell^2(\mathbb{Z}_K^d)$ such that for all $v = (v_a)_{a\in\mathbb{Z}_K^d} \in \ell^2(\mathbb{Z}_K^d)$,

$$(\mathcal{F}_K v)_a = \frac{1}{K^d} \sum_{b\in\mathbb{Z}_K^d} e^{-\frac{2i\pi a\cdot b}{K}} v_b = \frac{1}{K^d} \sum_{b\in\mathbb{Z}_K^d} e^{-ia\cdot x_b} v_b. \tag{3.4}$$

It inverse is given by

$$(\mathcal{F}_K^{-1} v)_a = \sum_{b\in\mathbb{Z}_K^d} e^{\frac{2i\pi a\cdot b}{K}} v_b = (K^d \mathcal{F}^* v)_a.$$

It is well known that the transformation $K^{d/2}\mathcal{F}_K$ is unitary, and that this transformation can be efficiently implemented by using Fast Fourier Transform algorithms.

3.2.1 Difference operators

In a finite dimensional setting, the operators $h\delta_{j,K}^\pm$ are represented by a matrices with ∓ 1 on the diagonal and ± 1 on one of the first diagonals, with complementary coefficient ± 1 in the corner to ensure the periodicity. For instance when $d = 1$ we have

$$\delta_{1,K}^+ = \frac{1}{h} \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & & & -1 & 1 \\ 1 & 0 & \cdots & 0 & -1 \end{pmatrix}.$$

Let us remark that this family of matrices, when $K \in 2\mathbb{N}^*$, does not define an element of any $\mathcal{A}_r^{\rm per}$, $r \in \mathbb{R}$. In particular because of the entry of index $(0,0)$ equals $\mp\frac{K}{2\pi}$ which cannot be estimated independently of K as it should be. Nevertheless we are going to prove that in the Fourier side these operators are in $\mathcal{A}_1^{\rm per}$ (and diagonal).

Let Q_K be the matrix associated with the discrete fourier transform $K^{d/2}\mathcal{F}_K$:

$$Q_K(a,b) = K^{-d/2} e^{-2i\pi a\cdot b/K} = \Big(\frac{h}{2\pi}\Big)^{d/2} e^{-iha\cdot b}.$$

By using the aliasing formula

$$\Big(\frac{h}{2\pi}\Big)^{d} \sum_{b\in\mathbb{Z}_K^d} e^{hij\cdot b} = \begin{cases} 1 & \text{if} \quad j = mK, \quad m\in\mathbb{Z}^d \\ 0 & \text{else,} \end{cases}$$

we see that Q_K is unitary, *i.e.* $Q_K^* Q_K = 1$, see (2.15). Moreover, we have the following Lemma:

Lemma 3.7. *Let $u = (u_a)_{a\in\mathbb{Z}_K^d}$ and $\hat{u} = \mathcal{F}_K u = (\hat{u}_b)_{b\in\mathbb{Z}_K^d}$. Then we have for $j = 1,\dots,d$,*

$$\left| \begin{array}{l} \delta_{j,K}^+ = Q_K^* D_{j,K}^+ Q_K = \mathcal{F}_K^{-1} D_{j,K}^+ \mathcal{F}_K \quad \textit{and} \\ \delta_{j,K}^- = Q_K^* D_{j,K}^- Q_K = \mathcal{F}_K^{-1} D_{j,K}^- \mathcal{F}_K \end{array} \right.$$

where $D_{j,K}^+$ and $D_{j,K}^-$ are the diagonal operators

$$\left| \begin{array}{lcll} (D_{j,K}^+ \hat{u})_a & = & \frac{1}{h}(e^{iha_j}-1)\hat{u}_a, & a\in\mathbb{Z}_K^d, \quad \textit{and} \\ (D_{j,K}^+ \hat{u})_a & = & \frac{1}{h}(1-e^{-iha_j})\hat{u}_a, & a\in Z_K^d. \end{array} \right.$$

Furthermore, the matrices $\{D_{j,K}^+\}_{K\in 2\mathbb{N}*}$ and $\{D_{j,K}^-\}_{K\in 2\mathbb{N}*}$ are in $\mathcal{A}_1^{\rm per}$.

Proof. We have $u = \mathcal{F}_K^{-1}\hat{u}$ which is written in coordinates

$$u_a = \sum_{b\in\mathbb{Z}_K} e^{\frac{2i\pi a\cdot b}{K}} \hat{u}_b.$$

Hence

$$u_{a+e_j} = \sum_{b\in\mathbb{Z}_K^d} e^{\frac{2i\pi a\cdot b}{K}} e^{\frac{2i\pi e_j\cdot b}{K}} \hat{u}_b = \sum_{b\in\mathbb{Z}_K^d} e^{\frac{2i\pi a\cdot b}{K}} e^{\frac{2i\pi b_j}{K}} \hat{u}_b.$$

This shows that

$$(u_{a+e_j} - u_a)_{a\in\mathbb{Z}_K^d} = \mathcal{F}_K^{-1}\Big\{((e^{\frac{2i\pi b_j}{K}}-1)\hat{u}_b)_{b\in\mathbb{Z}_K^d}\Big\}$$

and, as $h = \frac{2\pi}{K}$, we obtain the representation of $\delta_{j,K}^+$ given in the Lemma.
To prove that $\{D_{j,K}^+\}_{K\in 2\mathbb{N}*} \in \mathcal{A}_1^{\rm per}$, we note that the components of the matrix $D_{j,K}^+$ satisfy

$$D_{j,K}^+(a,b) = \frac{1}{h}(e^{iha_j}-1)\delta_{a,b}, \quad a,b\in\mathbb{Z}_K^d. \tag{3.5}$$

But we as $h = 2\pi/K$, we have $e^{iha_j} = e^{ih\hat{a}_j}$. Hence using the fact that $|e^{ix} - 1| \leq |x|$ for real numbers x, we have

$$|D^+_{j,K}(a,a)| \leq |\hat{a}_j| \leq [a] \quad \text{as} \quad a \in \mathbb{Z}^d_K.$$

Moreover, we have for $\alpha \geq 1$,

$$\partial^\alpha_{a_k} D^+_{j,K}(a,a) = i^\alpha h^{\alpha-1} \delta_{kj} e^{iha_j}.$$

By using Taylor expansion, we thus have, as in (2.12), that for $\alpha \in \mathbb{Z}^d \setminus \{0\}$,

$$|\Delta^\alpha D^+_{j,K}(a,a)| \lesssim_\alpha h^{1-|\alpha|} \lesssim \frac{1}{K^{|\alpha|-1}} \lesssim (1+[a])^{1-|\alpha|}$$

as we always have $[a] \leq K$. Therefore the familly $\{D^+_{j,K}\}_{K \in 2\mathbb{N}^*} \in \mathcal{A}^{\text{per}}_1$. □

3.2.2 Pointwise Multiplication

We consider now a periodic function $V(x)$, $x \in \mathbb{T}^d$. For $k \in \mathbb{Z}^d$, we denote by $(\mathcal{F}V)_k$ the Fourier transform of V. We thus have

$$V(x) = \sum_{k \in \mathbb{Z}^d} (\mathcal{F}V)_k e^{ik\cdot x}.$$

A natural discretization of the operator $u(x) \mapsto V(x)u(x)$ on a grid consists in the pointwise multiplication

$$(u_a)_{a \in \mathbb{Z}^d_K} \mapsto (V_a u_a)_{a \in \mathbb{Z}^d_K} =: \{(B_{V,K}u)_a\}_{a \in \mathbb{Z}^d_K}$$

where $V_a = V(ah)$ defines a element of $\ell^2(\mathbb{Z}^d_K)$. Denoting by $\hat{V}_a = \mathcal{F}_K\{(V_a)_{a \in \mathbb{Z}^d_K}\}$ the discrete Fourier transform of the sequence V_a, we thus have

$$V_a = V(ah) = \sum_{b \in \mathbb{Z}_K} \hat{V}_b e^{ihb\cdot a} = \sum_{k \in \mathbb{Z}^d} (\mathcal{F}V)_k e^{ihk\cdot a}. \tag{3.6}$$

In particular this shows that

$$\hat{V}_b = \sum_{\ell \in \mathbb{Z}^d} (\mathcal{F}V)_{b+\ell K}.$$

Using (3.6) for V_a and u_a we get

$$V_a u_a = \sum_{b,c \in \mathbb{Z}^d_K} \hat{V}_b e^{i(b+c)\cdot ah} \hat{u}_c = \sum_{f \in \mathbb{Z}^d_K} e^{ia\cdot fh} \Big(\sum_{b+c=f} \hat{V}_b \hat{u}_c \Big)$$

which leads to

$$\mathcal{F}_K\{(V_b u_b)_{b \in \mathbb{Z}^d_K}\}_a = \Big(\sum_{b \in \mathbb{Z}^d_K} \hat{V}_{a-b} \hat{u}_b \Big)_a.$$

As in the previous section, we obtain the following result:

Lemma 3.8. *For $K \in 2\mathbb{N}$, let $u = (u_a)_{a\in\mathbb{Z}^d_K}$ and $\hat{u} = \mathcal{F}_K u$ and we denote $\hat{u} = (\hat{u}_b)_{b\in\mathbb{Z}^d_K}$ its components. Let $V : \mathbb{T}^d \to \mathbb{C}$ be a smooth periodic function, and let $B_{V,K}$ be the diagonal operator acting on u defined by $(B_{V,K}u)_a = V_a u_a$ where $V_a = V(ah)$, $a \in \mathbb{Z}^d_K$. Then we have*

$$B_{V,K} = \mathcal{F}_K^{-1} M_{V,K} \mathcal{F}_K$$

where $M_{V,K}$ is the matrix with entries

$$M_{V,K}(a,b) = \hat{V}_{a-b} = \sum_{\ell\in\mathbb{Z}^d} (\mathcal{F}V)_{a-b+\ell K},\ a,b \in \mathbb{Z}^d_K \tag{3.7}$$

where $(\mathcal{F}V)_k$ denote the coefficients of the Fourier transform of the periodic function $V(x)$. Furthermore the family of matrices $M_{V,\bullet} := \{M_{V,K}\}_{K\in 2\mathbb{N}}$ belongs to $\mathcal{A}_0^{\mathrm{per}}$.

Proof. Only the last statement has not been proven. As a consequence of the smoothness of V we have that for all M there exits C_M such that

$$\forall\, j \in \mathbb{Z}^d, \quad |(\mathcal{F}V)_j| \leq C_M \frac{1}{(1+|j|)^M}.$$

Hence

$$|\hat{V}_{a-b}| \leq C_M \sum_{\ell\in\mathbb{Z}^d} \frac{1}{(a-b+\ell K)^M} = C_M \left(\frac{1}{(1+|a-b|)^M} + \sum_{\ell\neq 0} \frac{1}{(1+|a-b+\ell K|)^M} \right).$$

To bound the first term in the right-hand side, we note that when $a, b \in \mathbb{Z}^d_K$, then either $a-b \in G_K = \{-K/2, \dots, K/2-1\}$ and then $|a-b| = [a-b]$, or $a-b \notin G_K$, and then we have $|a-b| \geq K/2 \geq [a-b]$. In both case, we have $1+|a-b| \geq 1+[a-b]$. To deal with the second term, we note that when $|\ell| \geq 2$, we have $a-b+\ell K \geq K/2 \geq [a-b]$. Hence the sum of these terms is bounded by

$$C_M \frac{1}{(1+[a-b])^{M-d}} \sum_{|\ell|\geq 2} \frac{1}{(1+|a-b+\ell K|)^d}$$

and the last sum converges independently of K, a and b. It remains to consider the case $\ell = \pm 1$. In the case $a-b \in G_K$, we have as before $|a-b| \geq K/2 \geq [a-b]$ and hence we can control the term by $\frac{1}{(1+[a-b])^M}$. Finally, when $a-b \notin G_K$ we are in a situation where $a-b \pm K = [a-b]$ while $a-b \mp K \geq K/2 \geq [a-b]$ and we can control the term in both situations. This shows that for all N,

$$|\hat{V}_{a-b}| \leq C_N \frac{1}{(1+[a-b])^N}$$

We conclude by noticing that as in the infinite case, we have $\Delta_j^{\pm}(V_{a-b}) = 0$. □

3.2.3 General finite difference operators

By arguing as in Corollary 2.8, we prove the following result:

Corollary 3.9. *Let $K \in 2\mathbb{N}^*$, $u = (u_a)_{a\in\mathbb{Z}_K^d}$ a sequence, and $\hat{u} = \mathcal{F}_K u$ its discrete Fourier transform. Any composition of difference operators $\delta_{j,K}^{\pm}$ and multiplication operators of the form $B_{V,K}$ for some smooth periodic functions V acting on u, defines a discrete pseudo differential operator acting on $\hat{u}$. More precisely, for $P \in \mathbb{N}$ and for $p = 1, \dots, P$ let V_p be smooth functions, and $\epsilon_p \in \{0, \pm\}$. Then for all K we define*

$$A^K = \prod_{p=1}^{P} M_{V_p,K} D_{j,K}^{\epsilon_p}, \tag{3.8}$$

with the convention $D_{j,K}^0 = \mathrm{Id}_K$, and the family $A^\bullet = \{A^K\}_{K\in 2\mathbb{N}^*}$ defines an element of $\mathcal{A}_r^{\mathrm{per}}$ with $r = \sum_{p=1}^{P} |\epsilon_p|$.

Note that the decomposition (3.8) correspond after Fourier transform to matrix the decomposition

$$\mathcal{F}_K A^K \mathcal{F}_K^{-1} = \prod_{p=1}^{P} B_{V_p,K} \delta_{j,K}^{\epsilon_p},$$

acting on u. For example, a discretization of the order 2 operator (2.16) can be written (for example)

$$u_a \mapsto \sum_{j=1}^{d} \delta_j^+ B_{\sigma,V} \delta_j^- u + B_{X_j,K} \delta_j^+ - \sum_j \delta_j^+ (B_{X_j,K} u) + B_{V,K} u, \tag{3.9}$$

Which yields a familly of operators of order 2, belonging to the class $\mathcal{A}_2^{\mathrm{per}}$.

Similarly, all finite difference operators discretizing transport equation (like upwind, WENO schemes, ...) fall into the same category. Note that geometric considerations as in Section (2.4) can be made, according to the situation, the basic tool being given by the two previous Lemmas.

3.3 Pseudo-spectral methods

When discretizing general pseudo-differential equations, we face the problem of approximating operators of the form $\Phi(-i\partial_x)$ when Φ is not a polynomial but can be a rational or any function (see (2.14) in the case of water waves). In this case, one possibility is to consider spectral methods: A discretization of $u \mapsto \Phi(-i\partial_x)u$ is directly written in the discrete Fourier space

$$\hat{u}_a \mapsto \Phi(a)\hat{u}_a =: (A_{\Phi,K}\hat{u})_a, \qquad a \in \mathbb{Z}_K^d. \tag{3.10}$$

This yields to the evaluation of a diagonal operator in Fourier.

Now to approximate an operator of the for $u(x) \mapsto V(x)u(x)$, pseudo-spectral methods consist in calculating

$$\hat{u} \mapsto \mathcal{F}_K B_{V,K} \mathcal{F}_K^{-1} \hat{u}$$

where we recall that the operator $B_{V,K}$ is the pointwise multiplication by $V(ah)$ on the grid points ah, $a \in \mathbb{Z}^d_K$ (see section 3.2.2). Note that the evaluation of $\mathcal{F}_K B_{V,K} \mathcal{F}_K^{-1}$ doesn't cost too much since the operator $B_{V,K}$ is diagonal and the discrete Fourier transforms $\mathcal{F}_K$ and $\mathcal{F}_K^{-1}$ can be efficiently implemented by Fast Fourier transform algorithm.

This way, pseudo-spectral methods are efficient algorithm to discretize efficiently operators of the form (2.13). In echo with Lemma 2.7, we can state the following result

Lemma 3.10. *Let $\Phi : \mathbb{C}^d \to \mathbb{C}$ and $V : \mathbb{T}^d \to \mathbb{C}$. Then*

(i) If Φ is $\mathcal{C}^\infty$ and if there exists $r \in \mathbb{R}$ such that for all $\alpha \in \mathbb{N}^d$ and $x \in \mathbb{R}^d$, $|\partial_x^\alpha \Phi(x)| \lesssim_{r,\alpha} \langle x \rangle^{r-|\alpha|}$, then the family of matrices $A_{\Phi,\bullet} = \{A_{\Phi,K}\}_{K \in 2\mathbb{N}^*} \in \mathcal{A}_r^{\rm per}$.

(ii) If V is $\mathcal{C}^\infty$ then the family $\{\mathcal{F}_K B_{V,K} \mathcal{F}_K^{-1}\}_{K \in 2\mathbb{N}^*} \in \mathcal{A}_0^{\rm per}$.

As a corollary, pseudo-spectral discretization of compositions of the form (2.13) belong to $\mathcal{A}_r^K$, with r as in Corollary 2.8.

The proof is based on the same argument as the proof of Lemma (2.7) combined with the aliasing calculation of Lemma (3.8).

3.4 Approximation issues

As explained in the introduction of Section 3, one of the main motivation in the introduction of families of periodic matrices $A^\bullet = \{A^K\}_{K \in 2\mathbb{N}^*}$ in $\mathcal{A}_r^{\rm per}$ is to consider discretizations of PDEs, set on $\mathbb{T}^d$, on a grid of mesh $h = \frac{2\pi}{K}$. So naturally we are interested in the limit $K \to \infty$.
If we fix K, the matrix A^K can be embedded into $\mathcal{A}_r$ as an infinite dimensional operator that we denote by $\mathcal{I}A^K$ and simply defined as (see (3.1))

$$\mathcal{I}A^K(m,n) = \left| \begin{array}{ll} A^K(m,n) & \text{if} \quad (m,n) \in G_K \times G_K \\ 0 & \text{else.} \end{array} \right. \tag{3.11}$$

Then we can expect the convergence of A^K towards some infinite dimensional operator $A \in \mathcal{A}_r$. However, in general, $\|\mathcal{I}A^K\|_{\alpha,r,N}$ cannot be controlled uniformly by $[\![A^\bullet]\!]_{\alpha,r,N}$, as we can have $|m-n| > [m-n]$ even when $m, n \in \mathbb{Z}^d_K$. In fact the matrix was originally periodic and thus in the box $G_K \times G_K$ the top right corner is identified with the top left corner (consider the cas $d = 1$). In other words there are possible large coefficients $\mathcal{I}A^K(m,n)$ when typically $m = -K/2$ and $n = K/2 - 1$ where $|m-n| = K-1$ but $[m-n] = 1$. It is a typical *aliasing* phenomenon. For this reason, the convergence of A^K towards some infinite dimensional operator $A \in \mathcal{A}_r$ does not hold in the norm of $\mathcal{A}_r$ after using the embedding $\mathcal{I}$, but has to be understood in a weaker sense.

To illustrate this phenomenon, let us consider the approximation of the operator B_V given in (2.11) by the family of periodic matrices $M_{V,\bullet}$ defined in (3.7). We have

$$(\mathcal{I}M_{V,K})(m,n) - B_V(m,n) = \left| \begin{array}{ll} \sum_{\ell \neq 0} (\mathcal{F}V)_{m-n+\ell K} & \text{if} \quad (m,n) \in G_K \times G_K \\ (\mathcal{F}V)_{m-n} & \text{else} \end{array} \right.$$

We see again that as in the previous example, we do not have that $\|\mathcal{I}M_{V,K} - B_V\|_{\alpha,N,0}$ goes to zero, when $K \to \infty$. This can be seen by considering the entry (m,n) with $m = -K/2$ and $n = K/2-1$. This coefficient is equal to $\sum_{\ell\neq 0}(\mathcal{F}V)_{-K+1+\ell K}$ which is not small in general since it contains $(\mathcal{F}V)_1$ (for $\ell = 1$). However, we have the classical estimate with loss: for $0 \leq s' < s$ (see for instance [Lub08, Fao12])

$$\|(\mathcal{I}M_{V,K}) - B_V)x\|_{s'} \leq \frac{C}{K^{s-s'}}\|x\|_s .$$

Convergence issues can also appear because of the order of the operators we are trying to approximate. For instance let us consider the family of diagonal operators $D^+_{j,\bullet} = \{D^+_{j,K}\}_{K\in 2\mathbb{N}^*}$ defined in (3.5). In that case the previous aliasing phenomenon does not occur since the matrices are diagonal and for all $\alpha \in \mathbb{Z}^d$ and $N \in \mathbb{N}$ there exists C such that

$$\forall K \in 2\mathbb{N}^*, \quad \|\mathcal{I}D^+_{j,K}\|_{\alpha,N,1} \leq C[\![D^+_{j,\bullet}]\!]_{\alpha,N,1} .$$

On the other hand the limit operator of the $D^+_{j,K}$, when $K \to \infty$ is naturally the operator A_{Φ_j} with $\Phi(x) = ix_j$, for $j \in \{1,\dots,d\}$ and $x = (x_1,\dots,x_d)$. But the function $m \mapsto im - \frac{1}{h}(e^{ihm} - 1)$ does not go to 0 uniformly in h. Hence $\|A_{\Phi_j} - \mathcal{I}D^+_{j,K}\|_{\alpha,N,1}$ does not goes to 0 when $K \to \infty$. However, using Taylor expansion, we obtain easily the classical estimate with loss

$$\|(A_{\Phi_j} - \mathcal{I}D^+_{j,K})x\|_s \leq \frac{C}{K}\|x\|_{s+2} \tag{3.12}$$

for some constant independent of h. Note that in contrast with (3.12), we have the better estimates for spectral methods (see (3.10)): for $0 \leq s' < s$,

$$\|(A_\Phi - \mathcal{I}A_{\Phi,K})x\|_{s'} \leq \frac{C}{K^{s-s'-r}}\|x\|_s .$$

4 Applications

We now give several original applications of the previous results. The first one revisits the classical error bound for splitting schemes (see [JL00]).

4.1 Error bounds for splitting schemes

In this section we consider the error of a splitting scheme of order $k \geq 2$ for the abstract evolution equation

$$\dot{x} = iAx + iBx, \quad x \in h^s, \tag{4.1}$$

where $A \in \mathcal{A}_r$ and $B \in \mathcal{A}_\rho$ are Hermitian operators (see section 2.4.2), with $r, \rho \in \mathbb{R}$ and $\rho < r$. We also consider space discretization of this problem, of the form

$$\dot{y} = iA^K y + iB^K y \quad u \in \ell^2(\mathbb{Z}^d_K), \tag{4.2}$$

where $A^\bullet = \{A^K\}_{K\in 2\mathbb{N}^*}$ and $B^\bullet = \{B^K\}_{K\in 2\mathbb{N}^*}$ are in $\mathcal{A}_r^{\rm per}$ and $\mathcal{A}_\rho^{\rm per}$.

We assume that iA is the generator of a strongly continuous semigroup e^{itA} on $h^0 = \ell^2$, and that the flow of $i(A+B)$ and iB are well defined in h^s. We assume in particular that for all $s \geq 0$, we have the estimates

$$\exists\, M, C > 0, \quad \forall\, t \in \mathbb{R}, \quad \|e^{it(A+B)}\|_{\mathcal{L}(h^s,h^s)} + \|e^{itA}\|_{\mathcal{L}(h^s,h^s)} + \|e^{itB}\|_{\mathcal{L}(h^s,h^s)} \leq Me^{Ct} \tag{4.3}$$

where M and C depend on s but not on $t \in \mathbb{R}$.

In discrete case, and thus in finite dimension, the definition of the flow is not an issue by using matrix exponential. But we assume similarly that the following holds:

$$\exists\, M, C > 0, \text{ such that } \forall\, t \in \mathbb{R}, \quad \forall\, K \in 2\mathbb{N}^* \quad \forall\, y \in \ell^2(\mathbb{Z}_K^d)$$
$$\|e^{it(A^K+B^K)}y\|_{s,K} + \|e^{itA^K}y\|_{s,K} + \|e^{itB^K}y\|_{s,K} \leq Me^{Ct}\|y\|_{s,K}\,. \tag{4.4}$$

We will see below examples where such bounds (4.3) and (4.4) are satisfied.

Using our new class of discrete pseudo-differential operators we recover and amplify a result of Jahnke-Lubich proved in [JL00]:

Theorem 4.1 (local error bounds)**.** *Let $\rho, r \in \mathbb{R}$ with $\rho < r$.*

(i) Consider $A \in \mathcal{A}_r$ and $B \in \mathcal{A}_\rho$ and assume that the bound (4.3) holds. Then we have the following local error bounds for the Lie and Strang splitting: for all $s \geq 0$ there exists $C_s > 0$ and τ_0 such that for all $x \in \ell^2(\mathbb{Z}^d)$ and $|\tau| \leq \tau_0$,

$$\|e^{i\tau A}e^{i\tau B}x - e^{i\tau(A+B)}x\|_s \leq C_s\tau^2\|x\|_{s+r+\rho-1} \tag{4.5}$$
$$\|e^{\frac{i}{2}\tau B}e^{i\tau A}e^{\frac{i}{2}\tau B}x - e^{i\tau(A+B)}x\|_s \leq C_s\tau^3\|x\|_{s+2r+\rho-2}\,. \tag{4.6}$$

(ii) Consider $A^\bullet = \{A^K\}_{K\in 2\mathbb{N}^*}$ in $\mathcal{A}_r^{\rm per}$ and $B^\bullet = \{B^K\}_{K\in 2\mathbb{N}^*}$ in $\mathcal{A}_\rho^{\rm per}$ and assume that the bound (4.4) holds. Then for all $s \geq 0$ there exists $C_s > 0$ and τ_0 such that for all $|\tau| \leq \tau_0$, for all $K \in 2\mathbb{N}^*$ and all $y \in \ell^2(\mathbb{Z}_K^d)$,

$$\|e^{i\tau A^K}e^{i\tau B}y - e^{i\tau(A+B)}y\|_{s,K} \leq C_s\tau^2\|y\|_{s+r+\rho-1,K} \tag{4.7}$$
$$\|e^{\frac{i}{2}\tau B}e^{i\tau A}e^{\frac{i}{2}\tau B}y - e^{i\tau(A+B)}y\|_{s,K} \leq C_s\tau^3\|y\|_{s+2r+\rho-2,K}\,. \tag{4.8}$$

Comparing with Theorem 2.1 in [JL00], our result is more general in the following sense:

- We do not assume that B is bounded.
- The crucial assumption in [JL00] is $\|[A,B]v\|_s \leq c_1\|(-A)^\alpha v\|_s$ and $\|[A,[A,B]]v\|_s \leq c_1\|(-A)^\beta v\|_s$ for some $\alpha, \beta \geq 0$. Here these hypothesis are satisfied (with $\alpha = \frac{r+\rho-1}{r}$ and $\beta = \frac{2r+\rho-2}{r}$ as soon as A and B belong to $\mathcal{A}_r$ and $\mathcal{A}_\rho$ respectively. So α, β can be negative. Furthermore for the order 2 scheme, Jahnke-Lubich assume $\beta \geq 1 \geq \alpha$ which is not needed here. It will be important in section 4.3 to apply our result to water waves model.

- The local error bounds are readily carried over to discrete estimates by using the properties of $\mathcal{A}^{\mathrm{per}}$.

Proof. The proof is essentially the one given in [JL00] but with more refined commutator estimates. We consider only the continuous case, *i.e.* time splitting methods applied to (4.1). The discrete case (4.2) is exactly the same as we will only need commutator estimates (3.3) and the bound (4.4) which are both assumed to be independent of K.
To prove (4.5), owing to (4.3) and taking $|\tau| \leq \tau_0$, it is equivalent to prove the same bound for the operator

$$A(\tau) := e^{i\tau(A+B)}e^{-i\tau A}e^{-i\tau B} - \mathrm{Id}.$$

We have $A(0) = 0$ and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}A(\tau) &= e^{i\tau(A+B)}(i(A+B) - iA - ie^{-i\tau A}Be^{i\tau A})e^{-i\tau A}e^{-i\tau B}\\ &= e^{i\tau(A+B)}(iB - ie^{-i\tau A}Be^{i\tau A})e^{-i\tau A}e^{-i\tau B}.\end{aligned}$$

Hence using (4.3), we have if we assume $|\tau| \leq \tau_0$,

$$\|A(\tau)x\|_s \lesssim_{s,\tau_0} \int_0^\tau \|(B - e^{-i\sigma A}Be^{i\sigma A})x(\sigma)\|_s \,\mathrm{d}\sigma$$

where $x(\sigma) = e^{-i\tau A}e^{-i\tau B}x$ satisfies $\|x(\sigma)\|_s \lesssim_{s,\tau_0} \|x\|_s$ for all s. Let $B(\sigma) = B - e^{-i\sigma A}Be^{i\sigma A}$. We have $B(0) = 0$ and $\frac{\mathrm{d}}{\mathrm{d}\sigma}B(\sigma) = -ie^{-i\sigma A}[A,B]e^{i\sigma A}$. Hence for a given $\tilde{x}$ we have

$$\|(B - e^{-i\sigma A}Be^{i\sigma A})\tilde{x}\|_s \lesssim_{s,\tau_0} \int_0^\sigma \|[A,B]e^{i\alpha A}\tilde{x}\|_s \,\mathrm{d}\alpha.$$

Thus, using that $[A,B] \in \mathcal{A}_{r+\rho-1}$ (see Proposition 2.3) and Lemma 2.2, we deduce

$$\|A(\tau)x\|_s \lesssim_{s,\tau_0} \int_0^\tau \int_0^\sigma \|[A,B]e^{i\alpha A}x(\sigma)\|_s \,\mathrm{d}\sigma\mathrm{d}\alpha \lesssim \tau^2\|x\|_{s+r+\rho-1}.$$

which yields the result.
To prove (4.6), we proceed similarly by considering

$$A(\tau) := e^{i\tau(A+B)}e^{-i\frac{1}{2}\tau A}e^{-i\tau B}e^{-i\frac{1}{2}\tau A} - \mathrm{Id}.$$

We have

$$\frac{\mathrm{d}}{\mathrm{d}\tau}A(\tau) = e^{i\tau(A+B)}B(\tau)e^{-i\frac{1}{2}\tau A}e^{-i\tau B}e^{-i\frac{1}{2}\tau A}$$

with

$$\begin{aligned}B(\tau) &= i(A+B) - i\frac{1}{2}A - ie^{-i\frac{1}{2}\tau A}Be^{i\frac{1}{2}\tau A} - \frac{1}{2}ie^{-i\frac{1}{2}\tau A}e^{-i\tau B}Ae^{i\tau B}e^{i\frac{1}{2}\tau A}\\ &= +i(B - e^{-i\frac{1}{2}\tau A}Be^{i\frac{1}{2}\tau A}) + \frac{i}{2}e^{-i\frac{1}{2}\tau A}(A - e^{-i\tau B}Ae^{i\tau B})e^{i\frac{1}{2}\tau A}\\ &= ie^{-i\frac{1}{2}\tau A}C(\tau)e^{i\frac{1}{2}\tau A}\end{aligned}$$

where

$$C(\tau) = e^{i\frac{1}{2}\tau A} B e^{-i\frac{1}{2}\tau A} - B + \frac{1}{2}A - \frac{1}{2}e^{-i\tau B} A e^{i\tau B}.$$

We have $C(0) = B(0) = 0$. Moreover, we have

$$\frac{\mathrm{d}^k}{\mathrm{d}\tau^k} C(\tau) = \big(\frac{i}{2}\big)^k e^{i\frac{1}{2}\tau A} \mathrm{ad}_A^k B e^{-i\frac{1}{2}\tau A} - (-i)^k \frac{1}{2} e^{-i\tau B} (\mathrm{ad}_B^k A) e^{i\tau B}$$

where $\mathrm{ad}_B A = [B, A]$. Note that using Proposition 2.3 assertion *(ii)*, we have that $\mathrm{ad}_B^k A \in \mathcal{A}_{r+k\rho-k}$ and $\mathrm{ad}_A^k B \in \mathcal{A}_{\rho+kr-k}$. As $\rho < r$, we deduce using (4.3) that for $|\tau| \leq \tau_0$, we have

$$\left\| \partial_\tau^k C(\tau) x \right\|_s \lesssim_{s,k,\tau_0} \left\| x \right\|_{s+k\rho-k}.$$

This leads to the result by using $k = 2$ and the fact that $C(0) = C'(0) = 0$. □

Remark 4.2. *For a number $k \geq 1$, we denote by $SP_k(\tau, A, B)$ a splitting method of order k for (4.1) with time step τ. High order splitting methods can be easily constructed by using composition algorithms, see for instance [BCM08, HLW06] for a review. By using more elaborated algebraic formalism coming from the Baker-Campbell-Hausdorff formula, we infer the following: considering a splitting scheme of order $k \geq 2$ applied to (4.1) and (4.2), then for all $s \geq 0$ there exists $C_{s,k} > 0$ and τ_0 such that for all $x \in \ell^2(\mathbb{Z}^d)$ and $|\tau| \leq \tau_0$, $y \in \ell^2(\mathbb{Z}_K^d)$ and $K \in 2\mathbb{N}^*$,*

$$\left\| SP_k(\tau, A, B) x - e^{i\tau(A+B)} x \right\|_s \leq C_{s,k} \tau^{k+1} \left\| x \right\|_{s+kr+\rho-k} \tag{4.9}$$

$$\left\| SP_k(\tau, A^K, B^K) y - e^{i\tau(A+B)} y \right\|_{s,K} \leq C_{s,k} \tau^{k+1} \left\| y \right\|_{s+kr+\rho-k,K} \tag{4.10}$$

A complete proof in the general case is however out of the scope of this paper.

Remark 4.3. *It can be observed in the estimated (4.9) that the derivative loss, $\rho + k(r-1)$, decreases with the order of the scheme as soon as the order of A and B is strictly smaller than 1. Hence if furthermore $r < 1$, there exists $k \geq 2$ such that the error bound in the splitting scheme of order k does not require any loss of derivative, i.e. for any $s \geq 0$ there exists $C_s > 0$ such that:*

$$\| SP_k(\tau, A, B) x - e^{i\tau(A+B)} v \|_s \leq C_s \tau^{k+1} \| x \|_s. \tag{4.11}$$

We will give below an example of such situation for the Strang splitting ($k = 2$) in the case of the water wave system.

Remark 4.4. *We provide here only local error estimates, but global estimates can be easily obtained by following the classical argument of [JL00].*

Remark 4.5. *In the case of symplectic system of the form (2.18), the previous theorem readily applies, as the commutator of block matrix expresses in terms of commutators of the sub-matrices. Hence the previous theorem hold true for symplectic systems*

$$\dot{x} = (\boldsymbol{S}_1 + \boldsymbol{S}_2) x \tag{4.12}$$

where $x \in h^s \times h^s$, and $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ satisfy the decomposition (2.18), with blocks that are of order less than r and ρ respectively.

4.2 Growth of Sobolev norms

We give now an example of situations where the commutator estimates can be used to prove long time estimates of Sobolev norm infering in particular assumptions (4.3) and (4.4). We consider again the equations (4.1) and (4.2), but we assume that A and A^K are diagonal operators. In this situation, we can control the evolution of the Sobolev norm of the solutions as follows:

Theorem 4.6. *Let $\rho < 1$ and $r > \rho$. Let $A = A_\Phi \in \mathcal{A}_r$ where Φ satisfies the assumption* (i) *of Lemma 2.7, and for all $K \in 2\mathbb{N}^*$, let $A^K = A_{\Phi,K}$ the spectral approximation of A_Φ as defined in Lemma 3.10. Let $t \mapsto B(t)$ a continuous application from $\mathbb{R}$ to $\mathcal{A}_\rho$ with $B(t)^* = B(t)$, and let $t \mapsto B^\bullet(t) = \{B^K(t)\}_{K\in 2\mathbb{N}^*}$ a continuous application from $\mathbb{R}$ to $\mathcal{A}_\rho^{\mathrm{per}}$ be such that the $B^K(t)$ are hermitian for all K and all t. Assume that for all α, N, there exists $C_{\alpha,N,\rho}$ such that*

$$\forall\, t \in \mathbb{R} \qquad \|B(t)\|_{\alpha,N,\rho} + [\![B^\bullet(t)]\!]_{\alpha,N,\rho} \leq C_{\alpha,N,\rho}. \tag{4.13}$$

Let $x(t) \in h^s$ and $x^K(t) \in \ell^2(\mathbb{Z}^d_K)$ be the solution of the systems

$$\dot{x} = iAx + iB(t)x, \quad \text{and} \quad \dot{x}^K = iA^K x^K + iB^K(t)x^K.$$

Then we have for all $s \geq 0$, the existence of a constant $C_{s,\rho}$ such that with the notation $\langle t\rangle = (1+|t|^2)^{\frac{1}{2}}$,

$$\left|\begin{array}{lll} \forall\, t \in \mathbb{R} & & \|x(t)\|_s \leq C_{s,\rho}\langle t\rangle^{\frac{s}{1-\rho}}\|x(0)\|_s\,, \\ \forall\, K \in 2\mathbb{N}^* \quad \forall\, t \in \mathbb{R} & & \|x^K(t)\|_s \leq C_{s,\rho}\langle t\rangle^{\frac{s}{1-\rho}}\|x^K(0)\|_s\,. \end{array}\right.$$

Proof. We consider again only the continuous case, the discrete case being readily obtained by using the same calculations. Note that the proof follows arguments that can be found in [BFG20].

Let $(x,y) = \sum_{n\in\mathbb{Z}^d} \bar{x}_n y_n$ be the standard ℓ^2 scalar product. First we note that as A and $B(t)$ are hermitian, the ℓ^2 norm of x is preserved for all times: we have $\|x(t)\|_0 = \|x(0)\|_0 = (x(0),x(0))^{\frac{1}{2}}$. Let D be the diagonal operator defined by $D(m,n) = (1+|n|^2)^{\frac{1}{2}}\delta_{mn}$ for $(m,n) \in \mathbb{Z}^d \times \mathbb{Z}^d$. We have of course $\|x\|_s^2 = (D^s x, D^s x) \in \mathbb{R}$. We calculate then that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\|x\|_s^2 &= \mathrm{Re}(D^s x, iD^s(A+B(t))x) \\ &= -\mathrm{Im}(D^s x, (A+B(t))D^s x) - \mathrm{Im}(D^s x, [D^s, A+B(t)]x) \end{aligned}$$

As A and $B(t)$ are Hermitian, the first term vanishes, and as A is diagonal, it commutes with D^s. Hence, we have

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\|x\|_s^2\right| \leq \|x\|_s \|[D^s, B(t)]x\|_{\ell^2} \lesssim \|x\|_s \|x\|_{s+\rho-1}$$

where the last bound is obtained using commutator estimates and the assumption (4.13). By using a comparison Lemma with the ordinary differential equation $\dot{y} = C\sqrt{y}$ (see for instance Lemma 5.2 in [BFG20]) we obtain

$$\|x(t)\|_s \leq \|x(0)\|_s + C\int_0^t \|x(\sigma)\|_{s+\rho-1}\,.$$

for some constant C independent of t. By using the preservation of the ℓ^2 norm, we obtain (take $s = 1 - \rho$ in the previous estimate)

$$\|x(t)\|_{1-\rho} \lesssim \|x(0)\|_{1-\rho} + |t|\|x(0)\|_0$$

and by induction, for all $k \in \mathbb{N}$,

$$\|x(t)\|_{k(1-\rho)} \lesssim \langle t\rangle^k \|x(0)\|_{k(1-\rho)}$$

which shows the result by interpolation. □

Remark 4.7. *The previous result in the discrete case with a uniform bound in term of K is original. Note however that it is only interesting form times smaller than $t < K^{1-\rho}$ as we always know that $\|x(t)\|_{s,K} \leq K^s \|x(t)\|_{0,K} = K^s \|x(0)\|_{0,K}$. We refer to [FJ15] for the analysis of growth of Sobolev norm for fully discrete splitting schemes.*

4.3 Convergence without loss for water wave models

As an interesting and non trivial example of application, we consider a water-wave models with non constant topography. In this section $x \in \mathbb{T}$, $\zeta : \mathbb{T} \to \mathbb{R}$ models the free surface elevation and $\psi : \mathbb{T} \to \mathbb{R}$ models the trace of the velocity potential at the surface. The function $b : \mathbb{T} \to \mathbb{R}$ reflects the effect of the topography. The linearized water wave model around the flat surface and in presence of topography reads

$$\left| \begin{array}{l} \partial_t \zeta - G^N[b]\psi = 0 \\ \partial_t \psi + \zeta = 0 \end{array} \right.$$

where $G^N[b]$ is the Dirichlet to Neumann operator, see [Lan13] and the reference therein. We can expand it with respect to b. At first order we can expand the operator in powers of b as $G^N[b] = \Omega^2 + L_1(b) + O(b^2)$. If we retain only the first term in powers of b we obtain a system of the form

$$\left| \begin{array}{l} \partial_t \zeta - \Omega^2 \psi = G\nabla \cdot (b(x)\nabla G\psi), \\ \partial_t \psi + \zeta = 0, \end{array} \right. \tag{4.14}$$

where several choices for the operator G can be made. We can retain in a general approximation (see [Lan13, Section 3.7.2] and [CLS12])

$$\Omega^2 = \frac{1}{\sqrt{\mu}} |D| \tanh(\sqrt{\mu}|D|), \quad G = \operatorname{sech}(\sqrt{\mu}|D|). \tag{4.15}$$

where μ is a small parameter, and $|D| = |-i\nabla_x|$ is the Fourier multiplier by $|k|$ for $k \in \mathbb{Z}$. The limit $\mu \to 0$ (and thus $\Omega^2 = |D|^2$) yields the linearized St-Venant equations with $G = 1$. Several other models (when G is not trivial) with rational approximations of the water wave operators can be derived (for instance the Boussinesq approximations [Lan13, Section 5.1.3]).

Note moreover that splitting methods are particularly adapted to (4.14): the system can be divided at least into to systems

$$\left| \begin{array}{l} \partial_t \zeta - \Omega^2 \psi = 0 \\ \partial_t \psi + \zeta = 0 \end{array} \right. \quad \text{and} \quad \left| \begin{array}{l} \partial_t \zeta = G\nabla \cdot (b(x)\nabla G\psi) \\ \partial_t \psi = 0 \end{array} \right. \tag{4.16}$$

which can both be solved explicitly: the first one is diagonal in Fourier, and the second enjoys the conservation $\psi(t) = \psi(0)$ which allows an explicit solution $\zeta(t) = \zeta(0) + tG\nabla \cdot (b(x)\nabla G\psi(0))$ which can be computed easily using Fast Fourier transformations. Note that the first system could be also split into two pieces, yielding to Verlet-like algorithms, while the explicit solution corresponds to a Deuflhard algorithm, in the usual terminology of highly oscillatory systems (see [HLW06, Chapter XIII]).

To analyze the convergence of splitting schemes based on these decomposition, we make the usual change of variable for wave like systems: We define the new (symplectic) variables

$$\begin{pmatrix} \xi \\ v \end{pmatrix} = \begin{pmatrix} \Omega^{-\frac{1}{2}} & 0 \\ 0 & \Omega^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \zeta \\ \psi \end{pmatrix}. \tag{4.17}$$

After calculations, we find that the Hamiltonian system (4.14) writes in the new variables

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \xi \\ v \end{pmatrix} &= \begin{pmatrix} 0 & \Omega \\ -\Omega & 0 \end{pmatrix} \begin{pmatrix} \xi \\ v \end{pmatrix} + \begin{pmatrix} 0 & G\Omega^{-\frac{1}{2}}\nabla \cdot (b(x)G\Omega^{-\frac{1}{2}}\nabla \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \xi \\ v \end{pmatrix} \\ &:= \boldsymbol{S}_1 \begin{pmatrix} \xi \\ v \end{pmatrix} + \boldsymbol{S}_2 \begin{pmatrix} \xi \\ v \end{pmatrix} \end{aligned} \tag{4.18}$$

by using the notation (2.17) for symplectic operators. The energy associated with the system is

$$H(\xi, v) = \frac{1}{2} \int \Omega|\xi|^2 + \Omega|v|^2 + b(x)|G\Omega^{-\frac{1}{2}}\nabla v|^2.$$

Moreover, the flows $e^{t\boldsymbol{S}_1}$ and $e^{t\boldsymbol{S}_2}$ can be easily implemented: the first one decouples in Fourier modes, and the second one is triangular and can be calculated explicitly as explained above.

In the water wave case (4.15), we have by using Lemma 2.7 that $\Omega \in \mathcal{A}_{\frac{1}{2}}$. Moreover, we show that the operator

$$A = G\Omega^{-\frac{1}{2}}\nabla \cdot b(x)G\Omega^{-\frac{1}{2}}\nabla$$

is smoothing in the following sense: $A \in \mathcal{A}_\rho$ for all $\rho < 0$. Indeed, in dimension 1 this operator corresponds to the Fourier matrix with coefficients

$$A(n, m) = G_n \Omega_n^{-\frac{1}{2}} \hat{b}_{n-m} G_m \Omega_m^{-\frac{1}{2}} (in)(im). \tag{4.19}$$

Asymptotically, for large n and m, we have

$$|A(n, m)| \lesssim e^{-|n|-|m|} \hat{b}_{n-m}.$$

Thus we see that, even for rough bottom b (with bounded Fourier coefficients for instance), $A \in \mathcal{A}_\rho$ for all $\rho \leq 0$. Hence in this situation, applying Theorem 4.1, splitting methods based on the decomposition (4.16) converge in any Sobolev space without loss. We thus summarize the results of Theorem 4.1 applied to this situation[5]:

Theorem 4.8. *With the notation (4.18), if $b \in L^\infty(\mathbb{T})$, we have the following local error estimates for the water wave equation with (4.15) then for all $s \geq 0$ there exists $C_s > 0$ and τ_0 such that for all $x = (\xi, v)^T \in h^s \times h^s$ and $|\tau| \leq \tau_0$,*

$$\|e^{i\tau \boldsymbol{S}_1} e^{i\tau \boldsymbol{S}_2} x - e^{i\tau(\boldsymbol{S}_1+\boldsymbol{S}_2)} x\|_s \leq C_s \tau^2 \|x\|_s ,$$
$$\|e^{\frac{i}{2}\tau \boldsymbol{S}_1} e^{i\tau \boldsymbol{S}_2} e^{\frac{i}{2}\tau \boldsymbol{S}_1} x - e^{i\tau(\boldsymbol{S}_1+\boldsymbol{S}_2)} x\|_s \leq C_s \tau^3 \|x\|_s .$$

Note that these results translate automatically to the original variables by using the change of variables (4.17). Theorem 4.1 can also be applied in various situations where Ω and G are of some given orders.

4.4 Normal form as preconditioners

Considering splitting schemes for (4.1), we have seen in Remark 4.3 that when A is of order $r < 1$, it is possible to find a splitting methods based on the underlying decomposition that converge without loss of derivative. In this section, we show that in many situations, it is possible to make a change of variable putting the system into this form. It is based on a normal form transformation in the spirit of [BGMR20].

Rather than giving too general result, we will focus on the following system: We consider the Schrödinger equation in dimension 1,

$$\partial_t u = -i\Delta u + iV(x)u \tag{4.20}$$

where V is a smooth potential. Writing $x = \hat{u}$, the equation becomes

$$\partial_t x = Ax + Bx \quad x = (x_a)_{a\in\mathbb{Z}} = (\hat{u}_a)_{a\in\mathbb{Z}}, \tag{4.21}$$

where A and B are the matrix defined by

$$A(m,n) = |m|^2 \delta_{m,n} \quad \text{and} \quad B(m,n) = \hat{V}(m-n) = (\mathcal{F}V)(m-n).$$

So in this case $A \in \mathcal{A}_2$ and $B \in \mathcal{A}_0$ and if we apply directly Theorem 4.1, we will obtain a Lie spiting scheme that converges with a loss of 1 derivative and a Strang spliting scheme that converges with a loss of 2 derivatives.
Now let us define the operator X by the formula

$$X(m,n) = \begin{cases} -\dfrac{\hat{V}(m-n)}{i(|m|^2 - |n|^2)} & \text{when} \quad m \neq \pm n \\ 0 \quad \text{for} \quad m = \pm n \end{cases}$$

[5] It is also easy to prove the stability estimates (4.3)

By using the relation $m^2 - n^2 = (m+n)(m-n)$, we deduce that $|m^2 - n^2| \geq |m| + |n|$ for $m \neq \pm n$, and we deduce easily that $X \in \mathcal{A}_{-1}$. Furthermore X is hermitian.
In particular, X is a bounded operator from h^s to h^s, and we can define the transformation $y = e^{iX}x$, from h^s to h^s for all $s \geq 0$.

In the new variable $y = e^{iX}x$ the system reads

$$\dot{y} = ie^{iX}(A+B)e^{-iX}y$$

with $A \in \mathcal{A}_2$ and $X \in \mathcal{A}_{-1}$. Let $\mathrm{ad}_X(A) = i[X, A]$, we have

$$\mathrm{ad}_X^j(A) \in \mathcal{A}_{2-2j}.$$

In particular, for $j \geq 2$, we have $\mathrm{ad}_X^j(A) \in \mathcal{A}_{-2}$.
Now we taylor expand

$$e^{iX}\,(A+B)\,e^{-iX} = A + B + i[X, A] + i[X, B] + R$$

where the remainder

$$R := \int_0^1 (1-s)^2 \mathrm{e}^{-isX}\,\mathrm{ad}_X^2(A+B)\,\mathrm{e}^{isX}\mathrm{d}s \in \mathcal{L}(h^s, h^{s+2})$$

is smoothing and gain 2 derivatives. Moreover, as $B \in \mathcal{A}_0$ and $X \in \mathcal{A}_{-1}$, we have $[X, B] \in \mathcal{A}_{-2}$ which is also smoothing and gain 2 derivatives.

Eventually, we have by definition of X

$$A + B + i[X, A] = A + Z$$

where

$$Z(m,n) = B(m,n) \quad \text{for} \quad m = \pm n, \quad Z(m,n) = 0 \quad \text{for} \quad m \neq \pm n.$$

In particular Z is block diagonal and can be easily implemented.

So far, we have shown that the equation in y can be written

$$\dot{y} = i(A + Z + R)y$$

where $R \in \mathcal{L}(h^s, h^{s+2})$.

We can then split the system into

$$\dot{y} = i(A+Z)y \quad \text{and} \quad \dot{y} = iRy.$$

The first system is block diagonal and can be easily implemented, and the R part is smoothing and is thus nonstiff. We can therefore easily implement $e^{i(A+Z)}$ and e^{iR} and use splitting approximations.

In this case, the main error is dominated by the commutator

$$[A + Z, R] \in \mathcal{L}(h^s, h^s).$$

which can be seen by noticing that $R \in \mathcal{L}(h^s, h^{s+2})$ and $A + Z \in \mathcal{L}(h^s, h^{s-2})$ for all s. Therefore we obtain the following result:

Proposition 4.9. *The Lie-splitting method $e^{i\tau(A+Z)}e^{i\tau R}$ to approximate the solution of $e^{i\tau(A+Z+R)}$ converges without loss of derivative. As a consequence, the scheme*

$$e^{-iX}e^{i\tau(A+Z)}e^{i\tau R}e^{iX}$$

defines an order 1 *scheme for the equation* (4.21) *without loss of derivative: we have*

$$\|e^{i\tau(A+B)}x - e^{-iX}e^{i\tau(A+Z)}e^{i\tau R}e^{iX}x\|_s \leq C\tau^2\|x\|_s$$

for τ small enough, $s \geq 0$ and a constant C independent of x.

From the implementation point of view, e^{iX} can be seen as a preconditioner, that only need to be evaluated at the beginning and end of the simulation, owing to

$$\left(e^{-iX}e^{i\tau(A+Z)}e^{i\tau R}e^{iX}\right)^n = e^{-iX}\left(e^{i\tau(A+Z)}e^{i\tau R}\right)^n e^{iX}.$$

Furthermore we stress out that, since Z is block diagonal and R is smoothing , the splitting scheme $e^{i\tau(A+Z)}e^{i\tau R}$ can be easily implemented.
We conclude by emphasizing that this normal form strategy, viewed as a preconditioner construction, can be extended to any order of approximation and can be generalized to many situations, see [BGMR20]. Note also that, again, this result can be translated *mutatis mutandis* to pseudo-spectral and finite difference approximations using section 3. A general study is however out of the scope of this paper.

A Young inequality for convolution

Let $x = (x_n)_{n\in\mathbb{Z}^d}$ and $y = (y_n)_{n\in\mathbb{Z}^d}$ two sequences. We define $z = x * y = (z_n)_{n\in\mathbb{Z}^d}$ the sequence

$$z_n = \sum_{\substack{p,q\in\mathbb{Z}^d \\ n=p+q}} x_p y_q, \quad n \in \mathbb{Z}^d.$$

We also define for $p \geq 1$,

$$\|x\|_{\ell^p} = \left(\sum_{n\in\mathbb{Z}^d} |x_n|^p\right)^{\frac{1}{p}}.$$

And we recall the following Hölder inequality: for two sequences x, y

$$\left|\sum_k x_k y_k\right| \leq \|x\|_{\ell^p}\|y\|_{\ell^q} \quad \text{for} \quad 1 = \frac{1}{p} + \frac{1}{q},$$

which is itself a consequence of the Young inequality for product: $\forall a, b \geq 0$ we have $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$. This Hölder inequality is easily generalized by induction to

$$\left|\sum_k \prod_{i=1}^N x_k^{(i)}\right| \leq \prod_{i=1}^N \|x^{(i)}\|_{\ell^{p_i}} \quad \text{for} \quad \sum_{i=1}^N \frac{1}{p_i} = N,$$

for any sequences $x^{(i)}$, $i = 1, \ldots, N$ with $N \in \mathbb{N}$.

Lemma A.1. *For two sequences x and y indexed by $\mathbb{Z}^d$, we have*

$$\|x * y\|_{\ell^r} \leq \|x\|_{\ell^p} \|y\|_{\ell^q}, \quad \textit{for} \quad 1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}.$$

Proof. Let us denote $z = x * y$, we have

$$\begin{aligned}
|z_n| &\leq \sum_{k \in \mathbb{Z}^d} |x_{n-k}||y_k| \\
&= \sum_{k \in \mathbb{Z}^d} (|x_{n-k}|^p |y_k|^q)^{\frac{1}{r}} |x_{n-k}|^{\frac{r-p}{r}} |y_k|^{\frac{r-q}{r}} \\
&\leq \left(\sum_k |x_{n-k}|^p |y_k|^q \right)^{\frac{1}{r}} \left(\sum_k |x_{n-k}|^p \right)^{\frac{r-p}{rp}} \left(\sum_k |y_k|^q \right)^{\frac{r-q}{rq}} \\
&= \left(\sum_k |x_{n-k}|^p |y_k|^q \right)^{\frac{1}{r}} \|x\|_{\ell^p}^{\frac{r-p}{r}} \|y\|_{\ell^q}^{\frac{r-q}{r}},
\end{aligned}$$

where we used the generalized trilinear Hölder inequality for the decomposition

$$\frac{1}{r} + \frac{r-p}{rp} + \frac{r-q}{rq} = \frac{1}{p} + \frac{1}{q} - \frac{1}{r} = 1.$$

Then we obtain

$$\sum_n |z_n|^r \leq \|x\|_{\ell^p}^{r-p} \|y\|_{\ell^q}^{r-q} \left(\sum_{k,n} |x_{n-k}|^p |y_k|^q \right) = \|x\|_{\ell^p}^r \|y\|_{\ell^q}^r.$$

□

References

[ADN59] S. Agmon, A. Douglis and L. Nirenberg , *Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I*, Comm. Pure Appl. Math. 12, 623–727, (1959)

[Bam18] D. Bambusi. *Reducibility of 1-d Schrödinger equation with time quasiperiodic unbounded perturbations, I.* Trans. Amer. Math. Soc., 370(3), 1823–1865, (2018)

[BGMR20] D. Bambusi, B. Grébert, A. Maspero and D. Robert. Growth of Sobolev norms for abstract linear Schrödinger equations. Journal of the European Mathematical Society 23 (2), 557-583, (2020).

[BFG20] J. Bernier, E. Faou and B. Grébert, *Long time behavior of the solutions of NLW on the d-dimensional torus* Forum of Math. Sigma (2020) 26 pages.

[BCM08] S. Blanes, F. Casas, A. Murua, *Splitting and composition methods in the numerical integration of differential equations*, Bol. Soc. Esp. Mat. Apl. **45**, pp. 89–145 (2008).

[Bre00] A. Bressan, *Hyperbolic system of conservation laws. The one-dimensional Cauchy problem*, Oxford lecture series in mathematics and its applications 20, 2000.

[CCFM17] F. Casas, N. Crouseilles, E. Faou and M. Mehrenberger, *High-order Hamiltonian splitting for Vlasov-Poisson equations*. Numer. Math. 135, 769–801, (2017).

[CLS12] W. Craig, D. Lannes, C. Sulem. *Water waves over a rough bottom in the shallow water regime.* Ann. I. H. Poincaré - AN 29 (2012) 233–259.

[Cho10] O. Chodosh, Infinite matrix representations of classes of pseudodifferential operators. Undergraduate thesis (2010).

[Cho11] O. Chodosh, Infinite matrix representations of isotropic pseudodifferential operators, Methods Appl. Anal., vol. 18, no. 4, 351–372, (2011)

[DL89] R. Di Perna and P. L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. 98, 511–547, (1989).

[Fao12] E. Faou, *Geometric Numerical Integration and Schrödinger Equations*. European Math. Soc. 2012

[FJ15] E. Faou and T. Jézéquel, *Resonant time steps and instabilities in the numerical integration of Schrödinger equations*. Differential and Integral Equations 28, 221–238, (2015).

[GV85] J. Ginibre and G. Velo. *Scattering theory in the energy space for a class of nonlinear Schrödinger equations*. J. Math. Pures Appl. (9), 64(4):363–401, (1985)

[HLW06] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer, second edition, 2006.

[Hör87] L. Hörmander, *The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators.* Classics in Mathematics. Springer, 1987.

[JL00] T. Jahnke and Ch. Lubich. Error bounds for exponential operator splittings. BIT, vol. 40, No 4, pp 735-744 (2000)

[Lan13] D. Lannes, *The Water Waves Problem, Mathematical Analysis and Asymptotics*. Mathematical Surveys and Monographs, Vol. 188. American Mathematical Society, 2013.

[Lub08] C. Lubich, *From quantum to classical molecular dynamics: reduced models and numerical analysis*. European Math. Soc., 2008.

[MR17] A. Maspero and D. Robert, *On time dependent Schrödinger equations: Global well-posedness and growth of Sobolev norms.* Journal of Functional Analysis, 273(2):721 – 781, (2017). doi:10.1016/j.jfa.2017.02.029.

[ORS21] A. Ostermann, F. Rousset, K. Schratz. *Error estimates of a Fourier integrator for the cubic Schrödinger equation at low regularity*. Found. Comput. Math. 21:725–765 (2021)

[Tay81] M. E. Taylor, Pseudodifferential Operators, Princeton Univ. Press (1981).