Computational and Applied Mathematics (2024) 43:88
https://doi.org/10.1007/s40314-023-02586-2

The symbol approximation method: a numerical approach to the approximation of the symbol of self-adjoint operators

Jean-Paul Chehab[1]

Received: 28 July 2023 / Revised: 28 July 2023 / Accepted: 11 December 2023

Abstract
We propose a simple numerical procedure to approach the symbol of a self-adjoint linear operator $\mathcal{A}$ by using trace estimates of a corresponding discretization matrix A, with numerical data. The Symbol Approximation Method (SAM) is based on an adaptation of the matrix trace estimator to successive distinct numerical spectral bands in order to build a piecewise constant function as an approximation of the symbol σ of $\mathcal{A}$. The decomposition of the spectral interval into band of frequencies is proposed with several approaches, from the formal spectral to the multi-grid one. We apply the new method to different operators when discretized in finite differences or in finite elements. The SAM is also proposed as a tool for the modeling of waves equations, and is presented a means to capture an additional linear damping term (hidden operator) in hydrodynamics models such as Korteweig–de Vries or Benjamin Ono equations.

Keywords Operator symbol · Trace of a matrix estimate · Inverse problems · Preconditioning · Hydrodynamics damping model

Mathematics Subject Classification 65C05 · 65F15 · 65M55 · 65N06 · 65N30 · 65N55

1 Introduction

The knowledge of the the symbol of an operator is an important and practical question which arises in different applications, e.g., when finding a missing (linear) term (or hidden operator) in a PDE to take into account a phenomena and then to enrich the model; it can be also related to preconditioning techniques for differential operators (find a simple operator to approach a given one in order to build a simple preconditioner).

Communicated by Baisheng Yan.

✉ Jean-Paul Chehab
Jean-Paul.Chehab@u-picardie.fr

[1] Laboratoire LAMFA (UMR CNRS 7352), Université de Picardie Jules Verne, 33 rue Saint Leu, 80039 Amiens Cédex, France

Published online: 14 February 2024

Indeed, in a number of situations, a physical model has to be corrected (or enriched) by taking into account phenomena that have been ignored at first. For instance, in hydrodynamics, the damping of waves is observed experimentally, and a still challenging question is its mathematical representation, see, e.g. Chehab and Sadaka (2013a, b); Dumont and Duval (2013); Dumont and Manoubi (2018); Dutykh (2009); Ott and Sudan (1969, 1970). The most common approach is to try to express the damping force with a linear monotone self-adjoint operator $\mathcal{A}$; in particular cases it can be obtained by a formal derivation, but it can give rise to tricky expressions difficult to be handled and implemented numerically (Dutykh and Le Meur 2021; Le Meur 2015). Another approach on which we concentrate here consists in approaching $\mathcal{A}$ by using experimental or numerical samples.

Let A be a $n \times n$ discretization matrix of $\mathcal{A}$. We note that approximating the eigenvalues of A is not sufficient to approach the symbol σ of $\mathcal{A}$ since σ relies the frequency numbers to the eigenvalues; the symbol of an operator is defined as its Fourier multiplier, and is intrinsic to the operator; for instance when considering $\mathcal{A} = -\Delta$ on $\mathbb{R}$, we have $\sigma : \xi \in \mathbb{R} \mapsto \sigma(\xi) = \xi^2$, when $\mathcal{A} = -\Delta$ on $]0, 1[$ with homogeneous Dirichlet boundary conditions $\sigma : k \in \mathbb{N}^* \mapsto \sigma(\xi) = \pi^2 k^2$. Practically, we propose to extract information on the symbol of the operator through sequences of linear systems

$$AW^{(k)} = R^{(k)}, \quad k = 1, \ldots, N,$$

where $W^{(k)}$ and $R^{(k)}$ are given; $(\lambda_A^k)_{k=1}^n$ and $(\sigma_A(k)_{k=1}^k)$ denote respectively the eigenvalues and the symbol of A. Computing the coefficients of A is of course too costly, it is also to strongly related to the discretization scheme, and notably its dimension. We propose to approach σ by a piece-wise constant function on successive band of frequencies. This necessitates to produce an approximation of a mean value of the symbol in a given band of frequencies.To this end, we apply trace estimates techniques based on sample vectors and scalar product (see also Ubaru and Saad 2018):

$$Trace(A) = \sum_{k=1}^{N} \lambda_A^k \approx n \frac{\sum_{k=1}^N \langle AW^{(k)}, W^{(k)} \rangle}{\sum_{k=1}^N \langle W^{(k)}, W^{(k)} \rangle},$$

giving directly the mean value μ of the eigenvalues as

$$\mu = \frac{\sum_{k=1}^N \lambda_A^k}{n} = \frac{\sum_{k=1}^N \sigma_A(k)}{n} \approx \frac{\sum_{k=1}^N \langle AW^{(k)}, W^{(k)} \rangle}{\sum_{k=1}^N \langle W^{(k)}, W^{(k)} \rangle}.$$

The approximation of the symbol on band of frequencies is proposed considering the decomposition of the numerical number of frequencies interval $I = [1, n]$ of A into m sub-intervals as

$$I = \bigcup_{j=0}^{m-1} I_j = \bigcup_{j=0}^{m-1} [k_j + 1, k_{j+1}].$$

If we have at our disposal linear filters Π_j which allow to extract the eigen-components of a vector associated to eigenvalues belonging in I_j, then we can consider the local mean value estimates

$$\mu_j \approx \frac{\sum_{k=1}^N \langle A\Pi_j W^{(k)}, \Pi_j W^{(k)} \rangle}{\sum_{k=1}^N \langle \Pi_j W^{(k)}, \Pi_j W^{(k)} \rangle}, \quad j = 1, \ldots, m.$$

The number μ_j is then proposed as an estimate of the mean value of eigenvalues of A in I_j, and we can build the piecewise constant function

$$\sigma_m(x) = \sum_{j=1}^{m} \mu_j \chi_{I_j}(x),$$

as and approximation of $k \mapsto \sigma(k)$, where $\chi_{I_j}(x)$ denotes the characteristic function of the interval I_j, $j = 1, \ldots m$. The separation of band of frequencies (exact or approached) can be done with different techniques: spectral methods, multilevel methods (including multi-grid and hierarchical methods).

The article is organized as follows. In Sect. 2, we first recall basic and useful results on the estimate of the trace of a matrix using samples that we relay to a mean value estimate of the eigenvalues. Then, we can the define Symbol Approximation Method by first considering the spectral decomposition, applying the mean value estimates of to eigenvalues to successive band of frequencies. In Sect. 3, we adapt SAM to finite differences or finite elements discretization by using sequences of discretization (such as multi-grid techniques) to produce numerical filter band of frequency. We give illustrations of the multi-grid SAM (MSAM) when applied to different self-adjoint operators, their symbol is captured in a satisfactory way. In Sect. 4, we consider applications of multilevel SAM to time dependent problems. We first present the way the multilevel SAM can be applied to approach the symbol of a damping differential operator. We apply the method on damped dispersive models as Korteweig–de Vries (KdV) and Benjamin Ono (BO) equations. The numerical computations have been done using Matlab® and Freefem++ (http://www.freefem.org).

2 Numerical approximation of the symbol: the spectral case

2.1 Approximation of a matrix and of its trace using samples

2.1.1 Approximation of matrix coefficients

Let A be a matrix in $\mathcal{M}_n(\mathbb{R})$. We would like to compute the coefficients of A using a sequence of linear systems

$$AW^{(k)} = R^{(k)}, \quad k = 1, \ldots, N, \tag{1}$$

where $(W^{(k)})_{k=1}^N$ and $(R^{(k)})_{k=1}^N$ are two given sequences of vectors of $\mathbb{R}^n$, related by (1). We consider the associated least square problem

$$\min_{A \in \mathcal{M}_n(\mathbb{R})} \sum_{k=1}^{N} \|AW^{(k)} - R^{(k)}\|^2. \tag{2}$$

The functional $\Phi(A) = \sum_{k=1}^N \|AW^{(k)} - R^{(k)}\|^2$ is convex and coercive, the minimum values are reached at the critical points satisfying

$$\sum_{k=1}^{N} (AW^{(k)} - R^{(k)}) W_k^T = 0,$$

or

$$A\sum_{k=1}^{N}(W^{(k)}(W^{(k)})^T) = \sum_{k=1}^{N}(R^{(k)}(W^{(k)})^T),$$

which is equivalent to (2) iff $W = [W^{(1)}W^{(2)}\cdots W^{(N)}]$ is full rank.

We begin with the following simple but useful result:

Proposition 2.1 *Let $(W^{(k)})_{k=1}^N$ and $(R^{(k)})_{k=1}^N$ be two given sequence of vectors of $\mathbb{R}^n$ There exists a unique matrix $A \in \mathcal{M}_n(\mathbb{R})$ such that*

$$AW^{(k)} = R^{(k)}, \quad k = 1, \ldots, N$$

if and only if the matrix $[W^{(1)}W^{(2)}\cdots W^{(N)}]$. is full rank. In such a case, A is given by $A = TS^{-1}$ where $S = \sum_{k=1}^N W^{(k)}(W^{(k)})^T$ and $T = \sum_{k=1}^N R^{(k)}(W^{(k)})^T$

Proof The proof is classical. □

Of course, as an example of sequence (W_k) we can consider the case $S = WW^T = \gamma Id$, where $\gamma \in \mathbb{R}_+^*$. This occurs trivially when $(W_k = e_k)$, the k-th element of the canonical basis of $\mathbb{R}^n$, it is obviously too costly in practice to consider the whole collection of vector e_k. To reduce the computational cost of the problem, a first strategy consists in considering the approximations $(A_j)_{j\geq 1}$ of A using the truncated sum

$$A_j S_j = T_j, \quad j = 1, \ldots, N,$$

where $S_j = [W^{(1)}W^{(2)}\cdots W^{(j)}][W^{(1)}W^{(2)}\cdots W^{(j)}]^T$ and $T_j = [R^{(1)}R^{(2)}\cdots R^{(j)}]$ $[W^{(1)}W^{(2)}\cdots W^{(j)}]^T$. Note that $W^{(k)}$ can be a rectangular matrix $n \times s$ matrix.

Another strategy consists in choosing $(W^{(k)})_{k=1}^N$ as a Bernoulli test vector (or matrix) or a Hadamard matrix (for which $W^{(k)}(W^{(k)})^T = nId$). We recall the following two choices of samples of interest:

- Bernouilli test vector or matrices (also as i.i.d Rademacher random variables). They are defined by $(Pr((W_i^{(k)} = \pm 1) = 1/2)$.
- Hadamard Matrices. Following the Sylvester construction, we start from the partitioned matrix

$$\begin{bmatrix} H & H \\ H & -H \end{bmatrix}.$$

Starting from $H_1 = [1]$, it generates the sequences of $2n \times 2n$ matrices

$$H_{2^k} = \begin{bmatrix} H_{2^{k-1}} & H_{2^{k-1}} \\ H_{2^{k-1}} & -H_{2^{k-1}} \end{bmatrix} = H_2 \otimes H_{2^{k-1}}, \quad k \geq 1.$$

In Fig. 1, we give an illustration of the approximation of the coefficient of the laplacian matrix when using both Hadamard and Bernouilli samples; we observe that the convergence is slow in both the cases.

2.1.2 Approximation of the trace of a matrix using samples

Let $A \in \mathcal{M}_n(\mathbb{R})$. Consider the problem

$$\min_{D\in\mathcal{M}_n(\mathbb{R}),\text{D Diagonal}} \|A - D\|_F^2, \tag{3}$$

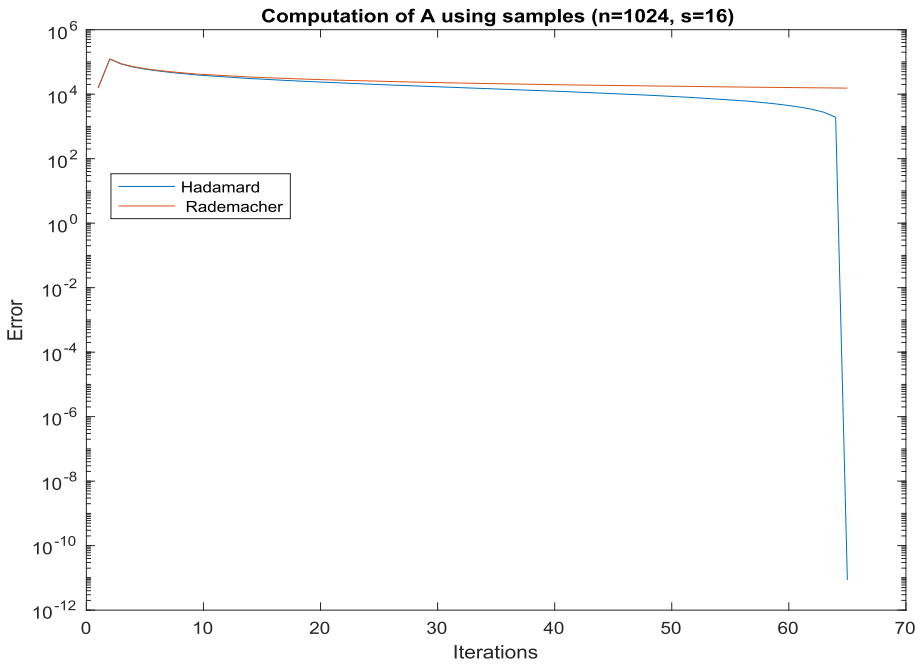

Fig. 1 Matrix estimation with Hadamard and Bernoulli samples $n \times s$, random matrix $n \times n$, $n = 4096$, $s = 16$

which solution is $D = diag(A)$. Here, $\|.\|_F$ denotes the Frobenius norm. We consider the simplest case when $D = \alpha Id$ so we look to

$$\min_{\alpha \in \mathbb{R}} \|A - \alpha Id\|_F^2. \tag{4}$$

This last problem solves easily: we set $\phi(\alpha) = \|A - \alpha Id\|_F^2$, we find that the minimum is reached for

$$\alpha = \frac{\langle A, Id \rangle_F}{\langle Id, Id \rangle_F} = \frac{trace(A)}{n} = \frac{\sum_{k=1}^{n} \lambda_k}{n}.$$

If A is only known trough a sequence of matrix–vector products (1), α will be approached numerically such as minimizing

$$\sum_{k=1}^{N} \|AW^{(k)} - \alpha W^{(k)}\|^2.$$

We find $\alpha_{opt} = \frac{\sum_{k=1}^{N} \langle AW^{(k)}, W^{(k)} \rangle}{\sum_{k=1}^{N} \langle W^{(k)}, W^{(k)} \rangle}$. We then obtain the following approximation of the trace

$$Trace(A) \simeq n \frac{\sum_{k=1}^{N} \langle AW^{(k)}, W^{(k)} \rangle}{\sum_{k=1}^{N} \langle W^{(k)}, W^{(k)} \rangle}.$$

From the previous relation, taking $(W^{(k)})_i = \pm 1$ with equiprobability, $\frac{1}{2}$, we find

$$Trace(A) \simeq \frac{1}{N}\sum_{k=1}^{N}\langle AW^{(k)}, W^{(k)}\rangle,$$

and we recover the method given proposed in Hutchinson (1989), then studied and applied, e.g. in Avron and Toledo (2011); Bai et al. (1996); Ubaru and Saad (2018). More generally, let $((W^{(k)})_{k=1}^{N})$ can be a sequence matrices of size $n \times s$ and we have

$$Trace(A) \approx \frac{1}{N\,s}\sum_{k=1}^{N}\langle (AW^{(k)}), W^{(k)}\rangle_F. \tag{5}$$

We now recall the result from Avron and Toledo (2011); Hutchinson (1989) which gives a justification of the above formula:

Lemma 2.2 *Let A be an $n \times n$ symmetric matrix with $trace(A) \neq 0$. Let z be a random vector whose entries are i.i.d Rademacher random variables ($Pr(zi = \pm 1) = 1/2$). $z^T Az$ is an unbiased estimator of trace(A) i.e.,*

$$E(zTAz) = trace(A),$$

and

$$Var(z^T Az) = 2\left(\|A\|_F^2 - \sum_{i=1}^{n} A_{ii}^2\right).$$

As an illustration, we display in Fig. 2, the approximation of the trace of the laplacian matrix when using both Hadamard and Bernouilli samples; we observe that the the convergence is fast in both the cases.

It is important to point out that the convergence curve shows that a reduced number of sample vectors is sufficient to obtain not a very precise but a good approximation of the trace value: this is characteristic to Monte Carlo's like-methods, in practice a moderate number of sample will be used.

We make now a simple but important remark regarding the applications we look to: when the matrix A is known only through matrix–vector products $AW_k = R_k$ the trace estimation (5) writes as

$$Trace(A) \simeq N\frac{\sum_{k=1}^{N}\langle R^{(k)}, W^{(k)}\rangle}{\sum_{k=1}^{m}\langle W^{(k)}, W^{(k)}\rangle}. \tag{6}$$

and particularly when $(W^{(k)})_i = \pm 1$ with equiprobability, $\frac{1}{2}$, we obtain

$$Trace(A) \simeq \frac{1}{N}\sum_{k=1}^{N}\langle R^{(k)}, W^{(k)}\rangle.$$

2.2 Estimates of spectral mean values on band of frequencies

Let A be a symmetric matrix of $\mathcal{M}_n(\mathbb{R})$. We introduce the following notation:

- $\sigma_A(k)$ is the value of the numerical symbol of the matrix A for the frequency number A
- σ is the symbol of the operator $\mathcal{A}$.

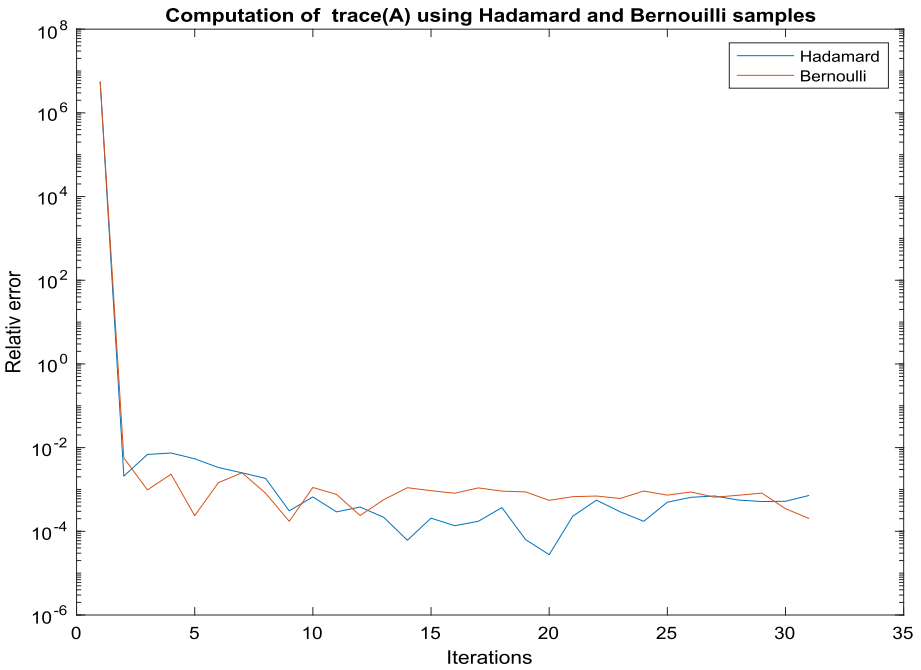

Fig. 2 Trace estimation with Hadamard and Bernoulli samples $n \times s$, random matrix $n \times n$, $n = 4096$, $s = 16$

Before presenting the method, we make the following remark: the eigenvalues λ_A^k of A coincide with the values of σ_A, however, the numbers $\sigma_A(k)$, $k = 1, \ldots, n$ are numbering in the increasing order of the frequency number k. Hence,

$$\sum_{k=1}^{n} \lambda_A^k = \sum_{k=1}^{n} \sigma_A(k).$$

We first consider the so-called spectral approach. Let A be a symmetric matrix. A being diagonalizable. We decompose the numerical frequency interval $I = [1, n]$ as

$$I = \cup_{j=0}^{m-1} I_j,$$

where $I_j = [k_j + 1, k_{j+1}]$, $k_0 = 0$, $k_m = n$, $k_j < k_{j+1}$. Let $(w_j)_{j=1}^n$ be the eigenvectors of A defined by the relations $Aw_j = \sigma_A(j)w_j$, $j = 1, \ldots, n$. We insist on the fact that $\{\lambda_A^k, k = 1, \ldots, n\} = \{\sigma_A(k)\ k = 1, \ldots, n\}$, however, the eigenvalues of A are not necessarily numbered in the increasing order of the numerical frequencies. We denote by

$$L_j = span(w_{k_j+1} \ldots w_{k_{j+1}})$$

and by P_j the $(k_{j+1} - k_j) \times n$-matrix $P_j = [w_{k_j+1}..w_{k_{j+1}}]^T$. The orthogonal projector onto L_j is then $\Pi_j = P_j^T P_j$. For any $u \in \mathbb{R}^n$, we can write $u = \sum_{i=1}^n \hat{u}_i w_i$, so $\Pi_j(u) = \sum_{i=k_j+1}^{k_{j+1}} \hat{u}_i w_i$ and, since A and Π_j commute, and $\Pi_j^2 = \Pi_j$, $j = 1, \ldots, m$, we have

$$\frac{\langle A\Pi_j u, \Pi_j u\rangle}{\langle \Pi_j u, \Pi_j u\rangle} = \frac{\langle AP_j u, P_j u\rangle}{\langle P_j u, P_j u\rangle} = \frac{\sum_{i=k_j+1}^{k_{j+1}} \hat{u}_i^2 \sigma_A(i)}{\sum_{i=k_j+1}^{k_{j+1}} \hat{u}_i^2}.$$

Now, for a sequence of vectors $u^{(\ell)} \in \mathbb{R}^n$, $\ell = 1, \ldots, N$, we define

$$\mu_j = \frac{\sum_{\ell=1}^{N} \langle AP_j u^{(\ell)}, P_j u^{(\ell)} \rangle}{\sum_{\ell=1}^{N} \langle P_j u^{(\ell)}, P_j u^{(\ell)} \rangle}.$$

We have then

$$\mu_j = \frac{\sum_{i=k_j+1}^{k_{j+1}} \sum_{\ell=1}^{N} (\hat{u}_i^{(\ell)})^2 \sigma_A(i)}{\sum_{i=k_j+1}^{k_{j+1}} \sum_{\ell=1}^{N} (\hat{u}_i^{(\ell)})^2},$$

so

$$\mu_j = \sum_{i=k_j+1}^{k_{j+1}} \gamma_i \sigma_A(i),$$

with $\gamma_i = \frac{\sum_{\ell=1}^{N} (\hat{u}_i^{(\ell)})^2}{\sum_{i=k_j+1}^{k_{j+1}} \sum_{\ell=1}^{N} (\hat{u}_i^{(\ell)})^2} \in [0, 1]$, $\sum_{i=k_j+1}^{k_{j+1}} \gamma_i = 1$. R_j is then a convex combination of λ_i and is proposed as an estimator of $\frac{\sum_{\ell=k_j+1}^{k_{j+1}} \sigma_A^{\ell}}{k_{j+1} - k_j}$, the mean value of the partial symbol in the numerical frequency number interval $[k_j + 1, k_{j+1}]$. Indeed, we have

$$\mu_j = \frac{\sum_{\ell=1}^{N} \langle (P_j^T A P_j) u^{(\ell)}, u^{(\ell)} \rangle}{\sum_{\ell=1}^{N} \langle P_j^T P_j u^{(\ell)}, u^{(\ell)} \rangle}.$$

When u is a random vector whose entries are i.i.d Rademacher random variables, we have

$$\mu_j = \frac{\sum_{\ell=1}^{N} \langle (P_j^T A P_j) u^{(\ell)}, u^{(\ell)} \rangle}{N(k_{j+1} - k_j)}.$$

At this point, we observe that, when u is a random vector whose entries are i.i.d Rademacher random variables, we obtain, applying Lemma 2.2,

$$E(u^T A_j u) = trace(P_j^T A P_j) = \sum_{\ell=k_j+1}^{k_{j+1}} \sigma_A(\ell),\ j = 1, \ldots, m,$$

where $A_j = P_j^T A P_j$, $j = 1, \ldots, m$. Therefore

$$\mu_j = \frac{\sum_{\ell=k_j+1}^{k_{j+1}} \sigma_A(\ell)}{k_{j+1} - k_j}.$$

Remark 2.3 This approach applies to the approximation of matrices expressed as $B^{-1}A$, which is related to generalized eigenvalues problems $Au = \lambda Bu$. This situation arises when considering finite elements discretization, as we will see later on. The ration μ_j is then replaced by

$$\mu_j = \frac{\sum_{\ell=1}^{N} \langle (P_j^T A P_j) u^{(\ell)}, u^{(\ell)} \rangle}{\sum_{\ell=1}^{N} \langle P_j B u^{(\ell)}, P_j^T u^{(\ell)} \rangle}$$

As pointed out above, when the matrix A is known only through matrix–vector products $A(P_j u^{(\ell)}) = v_j^{(\ell)}$ the trace estimation writes as

$$\mu_j = \frac{\sum_{\ell=1}^{m} \langle v^{(\ell)}, (P_j u^{(\ell)}) \rangle}{\sum_{\ell=1}^{m} \langle (P_j u^{(\ell)}), B(P_j u^{(\ell)}) \rangle}. \tag{7}$$

We resume as follows the spectral Symbol Approximation Method (SAM) to approach the spectrum of the generalized eigenvalue problem

Algorithm 1 : Spectral SAM

1: Given the projectors P_j, $j = 1, \ldots, m$ and a sequence of samples $(u^{(\ell)})_{\ell=1}^{N}$and $(v^{(\ell)})_{\ell=1}^{N} \in \mathbb{R}^n$ satisfying the relations $Bu^{(\ell)} = v^{(\ell)}$,
2: **for** $\ell = 1, \ldots, N$ **do**
3: **for** $j = 1, \ldots$ m, **do**
4: **Compute** $v_j^{(\ell)} = A(P_j u^{(\ell)})$
5: **Set** $N_j = N_j + \langle v^{(\ell)}, (P_j u^{(\ell)}) \rangle$
6: **Set** $D_j = D_j + \langle (P_j u^{(\ell)}), (P_j u^{(\ell)}) \rangle$
7: **end for**
8: **end for**
9: **for** $j = 1, \ldots$ m, **do**
10: **Set** $\mu_j = \frac{N_j}{D_j}$
11: **end for**
12: **Set** $\sigma_m(x) = \sum_{j=1}^{m} \mu_j \chi_{I_j}(x)$

At this point we give the following result:

Proposition 2.4 *Let $A \in \mathcal{M}_n(\mathbb{R})$ be a symmetric matrix with symbol σ_A. Decompose the numerical frequency interval $I = [1, n]$ as*

$$I = \cup_{j=0}^{m-1} [k_j + 1, k_{j+1}]$$

where $I_j = [k_j + 1, k_{j+1}]$, $k_0 = 0$, $k_m = n$, $k_j < k_{j+1}$.

Let (μ_j), $(j = 1, \ldots, m)$ and $(u^{(\ell)}$ defined as in Algorithm 1. Assume that $(|\Pi_j u^{(\ell)}\| \neq 0, \forall j = 0, \ldots, m, \forall \ell = 1, \ldots, N)$. Then, we have

$$\underline{\sigma_A}_j \leq \mu_j \leq \bar{\sigma_A} j, \quad j = 0, \ldots, m.$$

where $\underline{\sigma_A}_j = \min_{p \in I_j} \sigma_A(p)$ and $\bar{\sigma_A} j = \max_{p \in I_j} \sigma_A(p)$

Proof Let $(w_k)_{k=1}^{n}$ be the orthonormal eigenvectors of A, with the numbering $Aw_k = \sigma(k) w_k$, $k = 1, \ldots, n$. Let now $j \in 0, 1, \ldots, m$ be fixed. We have

$$\Pi_j u^{(\ell)} = \sum_{p=k_j+1}^{k_{j+1}} \hat{u}_p^{(\ell)} w_p \quad \text{and} \quad (A \Pi_j u^{(\ell)}) = \sum_{p=k_j+1}^{k_{j+1}} \lambda_p \hat{u}_p^{(\ell)} w_p.$$

Therefore,

$$\underline{\sigma_A}_j \sum_{\ell=1}^{N} \sum_{p=k_j+1}^{k_{j+1}} (\hat{u}_p^{(\ell)})^2 \leq \sum_{\ell=1}^{N} \langle A(\Pi_j u^{(\ell)}), \Pi_j u^{(\ell)} \rangle$$

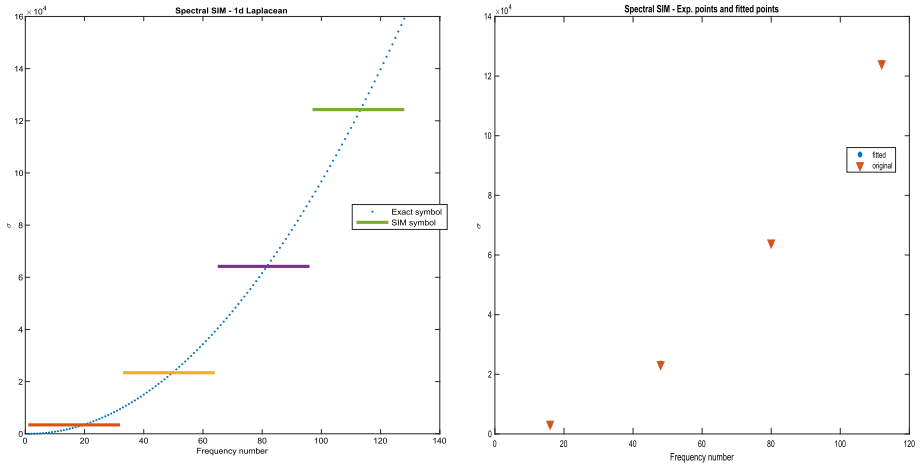

Fig. 3 Symbol Identikit and its polynomial fitting: the 1D case $n = 127$, A is $n \times n$

$$= \sum_{\ell=1}^{N} \sum_{p=k_j+1}^{k_{j+1}} \sigma(p)(\hat{u}_p^{(\ell)})^2 \leq \bar{\sigma_{A\,j}} \sum_{\ell=1}^{N} \sum_{p=k_j+1}^{k_{j+1}} (\hat{u}_p^{(\ell)})^2.$$

We have then

$$\underline{\sigma_{A\,j}} \leq \mu_j = \frac{\sum_{\ell=1}^{N} \langle A(P_j u^{(\ell)}),\, P_j u^{(\ell)} \rangle}{\sum_{\ell=1}^{N} \langle (P_j u^{(\ell)}),\, (P_j u^{(\ell)}) \rangle} = \sum_{p=k_j+1}^{k_{j+1}} \sigma(p)(\hat{u}_{j,p}^{(\ell)})^2 \leq \bar{\sigma_{A\,j}}$$

□

2.3 Illustrations

2.3.1 Finite differences discretization

We first consider A as the discretization matrix of the laplacian on $[0, 1]$ with periodic boundary conditions. We take $n = 256$ and a bandwidth $m = 4$ (the length of each bandwith is then 32).

In Fig. 3, we compare the piecewise constant symbol approximation obtained by the SAM Method and with the symbol of A. The symbol of the operator is $\sigma : k \in \mathbb{N}^* \mapsto \sigma(k) = \pi^2 k^2$. We find as a polynomial

$$\tilde{\sigma}_m(x) = 1.0e + 02 \left(1.911168150751541 + 0.367311968038462x + 0.094875519908414x^2\right).$$

The coefficient of x^2 is $9.48 \simeq \pi^2$, which is rather good and fits very well with the laplacian's symbol leading term. The large values of the other coefficients are attributable to the size of the bandwidth. The sample vectors are of Bernoulli type.

In space dimension 2, the symbol defines a surface that we fit with a polynomial of the two variables x and y.

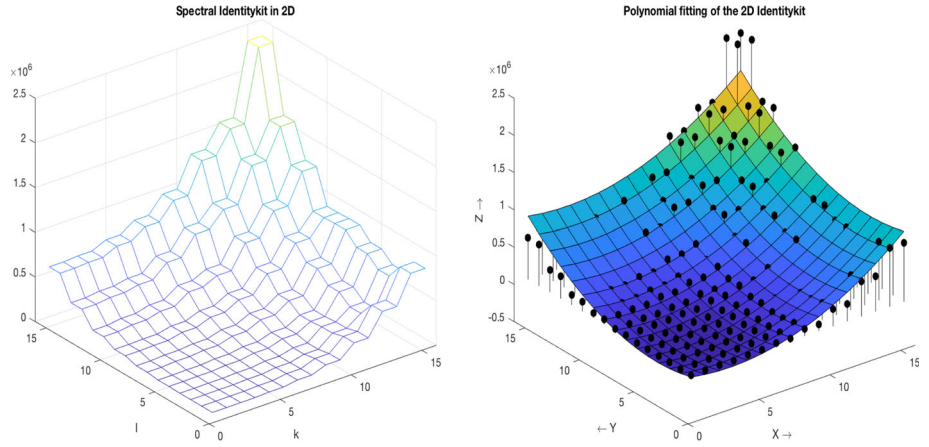

Fig. 4 Spectral Identikit and its polynomial fitting: the 2D case $n = 64$, A is $n^2 \times n^2$

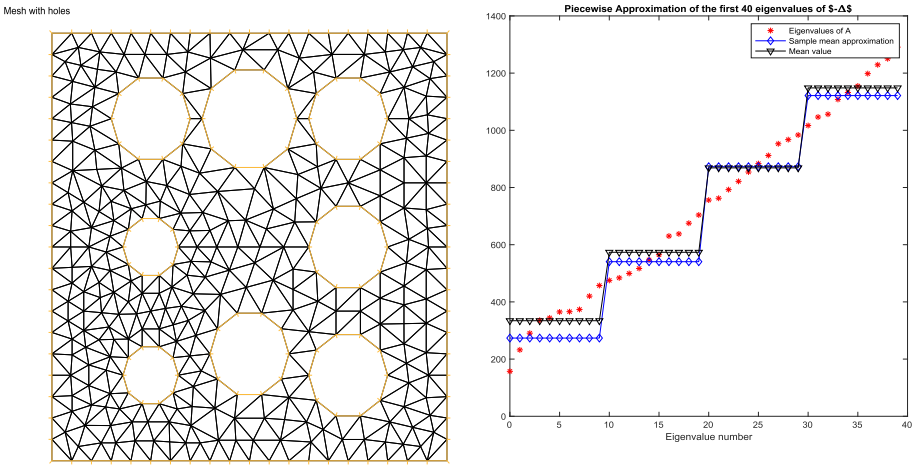

Fig. 5 Mesh of the domain Ω (left), first 40 eigenvalues of $-\Delta$ and their approximations by local means. $\mathbb{P}_1$ elements, 911 d.o.f

The computed symbol of the operator is

$$\tilde{\sigma}_m(x, y) = 1.0e + 04\,(6.702518213669119 - 4.960659057862875x - 5.612707457168954y) + 1.0e + 04\left(0.672528694583729(x^2 + y^2)\right),$$

which is a correct result allowing to propose the laplacian as the operator $\mathcal{A}$, see Fig. 4.

2.3.2 Finite elements discretization

We represented in Fig. 5 the approximation of the symbol of $\mathcal{A} = -\Delta$ in a square domain with holes. The SAM method allows to capture in a very satisfactory way σ.

3 The multigrid symbol approximation method (MSAM)

3.1 Derivation of the method

In the practical cases, the matrix A will be the discretization of a (local or nonlocal) differential operator which symbol can be assumed as a function defined on the spectral domain of $-\Delta$. Hence, the eigenvalues of A are related to the natural frequencies of the domain, e.g. to the eigenvalues λ_i of $-\Delta$. Computing these numbers can be costly, and a way to select a band of frequencies is to use numerical filters that will replace the projectors P_j. A classical procedure to realize numerically a separation of band of frequencies is to use grids or meshes of different characteristic mesh-size: this consists in handling different level of discretization (and associated spaces of approximation), from the coarsest to the finest, each one with a limited capability to capture a set of frequencies (by Nyquist's frequency sampling theorem). The coarse spaces allow to represent only signals supported by low frequencies while the fine ones cover a large band.

Consider the solution of the linear system

$$Ax = b,$$

Here A corresponds to the discretization matrix of an operator $\mathcal{A}$ but its coefficients are not known, A is only known through matrix–vector products; this e.g. is the case when A is a Schur complement. In such a situation, the use of Krylov methods is recommended and the numerical resolution is of course more efficient with the use of a pre-conditioner. The problem here is then to build a pre-conditioner of A without using the coefficient of A. A way to overcome this difficulty is then to approach the symbol σ of $\mathcal{A}$ by a simple polynomial (or rational function) $\tilde{\sigma}_m$. The pre-conditioner of A is then built as a discretization matrix of the operator of symbol σ_m.

To this end, we can of course apply the SAM method as presented above.

We first introduce here a simple variant of SAM when a sequence of matrices A is known, say when we have at our disposal discretization of $\mathcal{A}$ with different d.o.f. Let A_1 and A_2 be two discretization matrices of $\mathcal{A}$ of size $n_1 \times n_1$ and $n_2 \times n_2$, respectively, with $n_2 > n_1$. On one hand the trace approximation coming from the consistency of the discretization gives:

$$\begin{aligned} trace(A_1) &\simeq \sum_{i=1}^{n_1} \sigma(k), \\ trace(A_2) &\simeq \sum_{i=1}^{n_1} \sigma(k) + \sum_{i=n_1+1}^{n_2} \sigma(k), \end{aligned}$$

where $k \mapsto \sigma(k)$ is the symbol of $\mathcal{A}$.

On the other hand, the trace estimation of the SAM gives

$$\begin{aligned} \bar{\sigma}_1 = \frac{trace(A_1)}{n_1} &\simeq \frac{\sum_{\ell=1}^{N} \langle A_1 W_1^{(\ell)}, W_1^{(\ell)} \rangle}{\sum_{i=1}^{N} \langle W_1^{(\ell)}, W_1^{(\ell)} \rangle}, \\ \bar{\sigma}_2 = \frac{trace(A_2)}{n_2} &\simeq \frac{\sum_{\ell=1}^{N} \langle A_2 W_2^{(\ell)}, W_2^{(\ell)} \rangle}{\sum_{i=1}^{N} \langle W_2^{(\ell)}, W_2^{(\ell)} \rangle}. \end{aligned}$$

Here $W_1^{(\ell)} \in \mathbb{R}^{n_1}$ and $W_2^{(\ell)} \in \mathbb{R}^{n_2}$, $\ell = 1, \ldots, N$, are two sequences of sample vectors.

Now, applying the trace estimation, we deduce directly from above

$$\mu_1 = \bar{\sigma}_1, \tag{8}$$

$$\mu_2 = \frac{1}{n_2 - n_1}\left(n_2\bar{\sigma}_2 - n_1\bar{\sigma}_1\right) \tag{9}$$

More generally let n_j, $j = 1, \ldots, m$ be integers labelled in the increasing order. We consider the m matrices A_j of sizes $n_j \times n_j$. We denote by $\lambda_k, k \geq 1$ the eigenvalues of $\mathcal{A}$ and $\lambda_k^{(j)}, k = 1, \ldots, n_j$ the ones of A_j. The trace approximation results gives:

Algorithm 2 : MSAM method

1: **Compute** $\bar{\sigma}_j = \dfrac{\sum_{\ell=1}^{N} \langle A_j W_j^{(\ell)}, W_j^{(\ell)} \rangle}{\sum_{i=1}^{N} \langle W_j^{(\ell)}, W_j^{(\ell)} \rangle} \simeq \dfrac{1}{n_j} trace(A_{n_j})$, $j = 1, \ldots, m$
2: **Initialization** $\mu_1 = \bar{\sigma}_1$
3: **for** $j = 2, \ldots$ m, **do**
4: **Set** $\mu_j = \dfrac{1}{n_j - n_{j-1}}\left(n_j\bar{\sigma}_j - n_{j-1}\bar{\sigma}_{j-1}\right)$
5: **end for**
6: **Set** $\sigma_m(x) = \sum_{j=1}^{m} \mu_j \chi_{[n_{j-1}+1, n_j]}(x)$

Here, $W_j^{(\ell)} \in \mathbb{R}^{n_j}$, $\ell = 1, \ldots, N$, are the sequences of sample vectors.

Remark 3.1 The expression giving μ_j from $\bar{\sigma}_j$ and $\bar{\sigma}_{j-1}$ in Algorithm (4) is nothing else but a (simple) extrapolation formula; more general extrapolations techniques could be considered to improve the predictive formula, we refer to Brezinski (2000); Brezinski and Redivo-Zaglia (1991), e.g., for general techniques and surveys.

Proposition 3.2 *Let $\mathcal{A}$ be a self-adjoint positive definite operator and A_{n_j} a sequence of $n_j \times n_j$ discretization SPD matrices of $\mathcal{A}$. We note by $\sigma(k) > 0$ the symbol values of A and $\sigma_{A_j}(k)$ those of A_{n_j}. We introduce the error*

$$e^{(k,j)} = \sigma(k) - \sigma_{A_j}(k).$$

We have the relations:

(i)

$$\mu_j - \frac{1}{n_j - n_{j-1}} \sum_{k=n_j+1}^{n_j} \sigma(k) = \frac{1}{n_j - n_{j-1}} \left(\sum_{k=1}^{n_j} e^{(k,j)} - \sum_{k=1}^{n_{j-1}} e^{(k,j-1)} \right)$$

(ii)

$$\left(1 - \frac{1}{n_p}\right)\bar{\sigma}_{A_p} \leq \sum_{j=1}^{n_p} \left(\frac{n_j - n_{j-1}}{n_p}\right) \mu_j \leq \bar{\sigma}_{A_p},\ p = 2, \ldots, m.$$

Proof The proof is established by direct computations. □

3.2 Applications in spectral Fourier discretization

As a first illustration, we look to the performance of the MSAM method for approaching the symbol of a band-limited filter which symbol is $\theta(k) = \begin{cases} 1 & \text{if } 50 < k < 100, \\ 0 & \text{else,} \end{cases}$ say, a characteristic function in Fourier space. We display the comparison between the initial

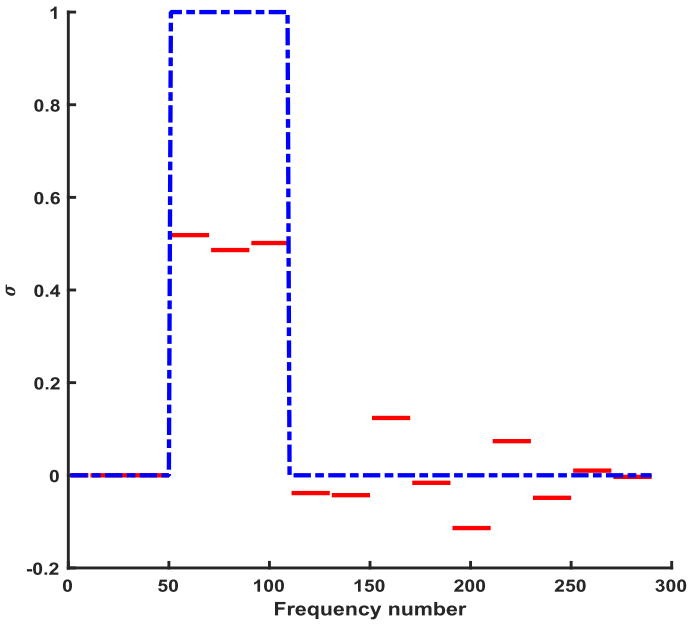

Fig. 6 Approximation of the symbol of band-limited filter Nmax = 290, $N = 30$, $m = 15$

symbol and its approximation in Fig. 6. The result is quite satisfactory, the symbol being difficult to approach; FFT are used to build the matrix vector product with θ in the physical space.

3.3 Applications in finite differences

As a first illustration, we compare in Fig. 7 the capability of the MSAM method to approximate the symbol of the laplacian matrix, when using spatial discretization schemes of order 2, 4 and 6, the two last ones being built with compact schemes.

In all the cases, we remark that the approached symbol is very closed to the exact symbol for the low frequencies: this is due to the very good approximation of the first eigenvalues by the discretization scheme, the accordance is even better when considering the high accurate compact schemes (order and 6), as expected by formula i. of Proposition 3.2.

We now consider the operator $\mathcal{A} = 100(-\partial_x^2)(1000 - \partial_x^2)^{-1}$ which symbol is $\sigma(k) = \frac{100\pi^2k^2}{1000 + \pi^2k^2}$, and a sequence of finite differences discretization matrices. We plot in Fig. 8 the approximated spectrum, the fitted curve, and the exact symbol. The nonlinear least-square fitting gave

$$\sigma_m(k) = \frac{8.2614 + 2.0832.\pi^2k^2}{2.5260 + 0.0208\pi^2k^2}$$

The first part of the symbol, which corresponds to the low frequencies, is not as well represented as in the laplacian case (see preceding remarks on Proposition 3.2). However, the result can be considered as relatively satisfactory since only matrix vector product have been used.

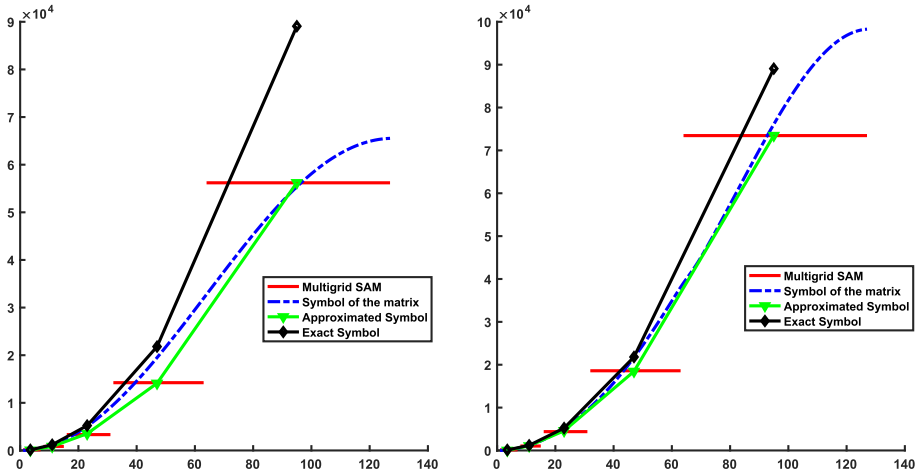

Fig. 7 Approximation of the spectrum of the 1D finite differences laplacian matrix, with 2d and 4th order discretization schemes. Nmax = 256, $N = 20$, $m = 6$

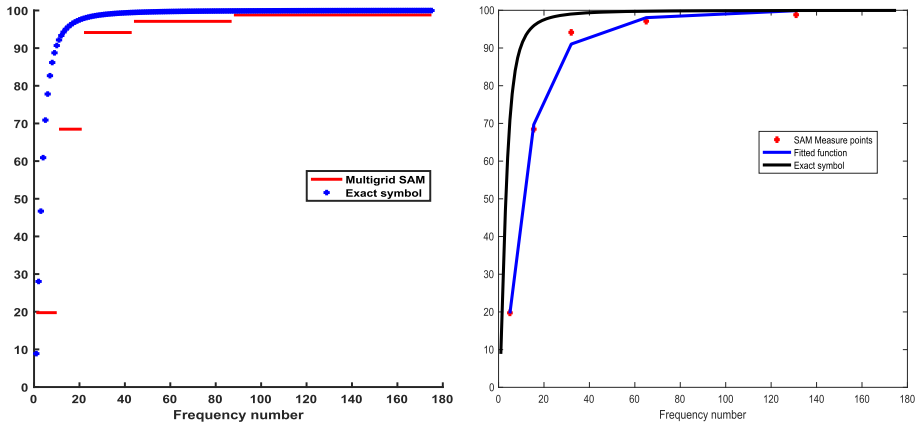

Fig. 8 Approximation of the spectrum of the 1D finite differences matrix of $\mathcal{A} = 100(-\partial_x^2)(1000 - \partial_x^2)^{-1}$, with 4th order discretization schemes. Nmax = 275, $N = 50$, $m = 6$

As a second type of illustration, we give application of the MSAM method to approximation of the spectrum of the Schur complement (or Uzawa's discrete operator) of the generalized Stokes problem:

$$\alpha \mathbf{u} - \frac{1}{Re}\Delta \mathbf{u} + \nabla p = f, \quad \text{in } \Omega =]0,1[^2 \tag{10}$$

$$\text{div}\mathbf{u} = 0 \qquad \text{in } \Omega, \tag{11}$$

$$\mathbf{u} = 0 \qquad \partial\Omega, \tag{12}$$

here $\mathbf{u} = (u, v)$ is the velocity field and the scaler p is the pression. The Uzawa operator is formally $\mathcal{U} = \text{div}\left(\alpha \mathbf{u} - \frac{1}{Re}\Delta\right)^{-1} \nabla$. When $\alpha = 0, \mathcal{U}$, it is closed to the $Re\, Id$ but for large values of α, $\mathcal{U}$ can be approached by a laplacian with Neumann Boundary conditions (see Peyret and Taylor (1983)), and an the Schur complement of the Uzawa-type method needs

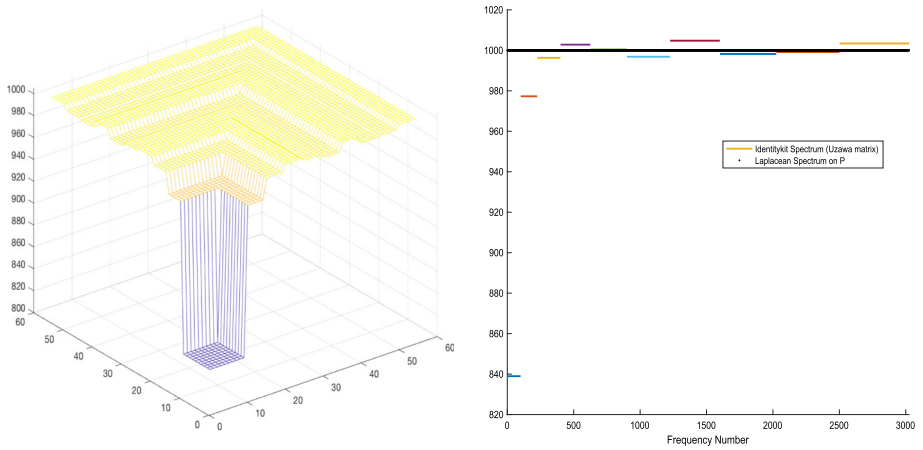

Fig. 9 Stokes problem with $\alpha = 0$. Spectrum Identikit of the Schur Complement. $Re = 1000$, $m = 12$, dof = 3025, $N = 50$. The approached symbol is $\sigma_m(x, y) = 932.79382.1243\, x + 2.1243\, y - 0.0278(x^2 + y^2)$

to be pre-conditioned in this case. The (second-order) discretization of the Stokes system on a MAC mesh with finite differences reads as

$$\alpha U + \frac{1}{Re} A_u.U + B_x P = 0, \tag{13}$$

$$\alpha V + \frac{1}{Re} A_v.V + B_y P = 0, \tag{14}$$

$$B_x^T U + B_y^T V = 0, \tag{15}$$

with natural notations. The Schur Complement is

$$\mathbf{S} = B_x^T \left(\alpha Id + \frac{1}{Re} A_U\right)^{-1} B_x + B_y^T \left(\alpha Id + \frac{1}{Re} A_V\right)^{-1} B_y.$$

The results of the MSAM method are displayed in Fig. 9 for $\alpha = 0$. We observe that expect the lowest eigenvalues close to $0,83\, Re$ most of the other ones are concentrated around $Re = 1000$; this agrees with Crouzeix (1974, 1997).

Now we take $\alpha = 1000$, so $\mathcal{U} \simeq -\Delta$. In Fig. 10 we display the computed MSAM approximation of the symbol of $\mathbf{S}$, and we compare it with the one of the discrete Laplacian, the eigenvalues being ranged in the increasing order. Then, we use the well-known pre-conditioner given by Cahouet and Chabard (1988) as a symmetric pre-conditioner of $\mathbf{S}$ build of (incomplete) Cholesky factors of the negative Laplacian matrix (associated to Homogeneous Diirichlet Boundary conditions), and we observe that it is concentrated around the point $(1, 0)$ in the complex plane.

3.4 Applications in finite elements

First of all, we must adapt the trace estimation formula. We consider for the sake of simplicity the homogeneous Dirichlet problem on the regular open and bounded domain $\Omega \subset \mathbb{R}^{2,3}$; the problem is discretized in finite elements on a (regular) triangulation $\mathcal{T}_h$ on which the finite

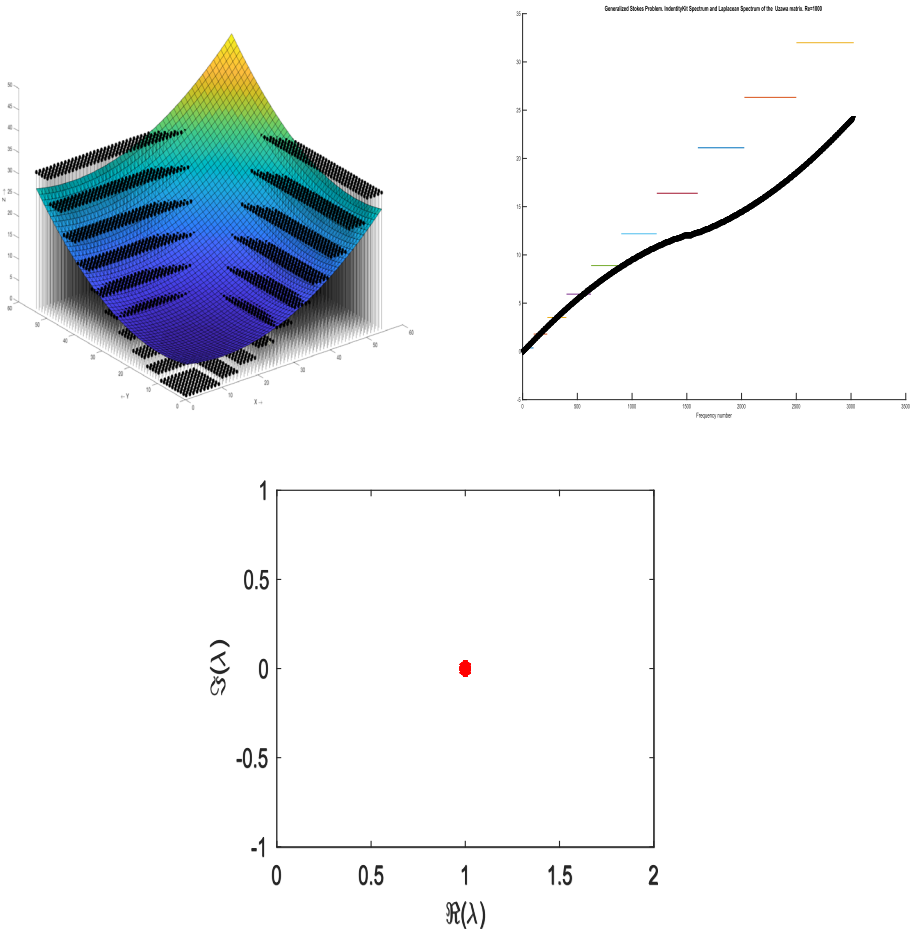

Fig. 10 Stokes problem with $\alpha = 1000$. Line 1: spectrum Identikit of the Schur complement in 2D (left) and 1D (right); Line 2 SAM estimate of the symbol of the preconditioned Uzawa Matrix in the complex plane. $Re = 1000$, $m = 12$, n d.o.f. = 3025, $N = 50$. The approached symbol is $\sigma_m(x, y) = 8.1173 - 0.2429\,x - 0.2429\,y + 0.0111(x^2 + y^2)$

elements space V_h is built. We obtain the linear system

$$A_h u_h = M_h f_h,$$

where $A_h$is the stiffness matrix and M_h the mass matrix; h is the characteristic mesh-size of the triangulation. The trace estimation formula writes here as

$$\frac{Trace(A_h)}{n} \simeq \frac{\sum_{k=1}^{N} \langle A_h W_k, W_k \rangle}{\sum_{k=1}^{N} \langle M_h W_k, W_k \rangle}.$$

where $W_k \in \mathcal{R}^n, k = 1, \ldots, N$ are sample vectors representing sample functions of $\mathcal{V}_h$. We have set $n = \dim(\mathcal{V}_h)$.

Now, to display the SAM method in a multi-grid like framework, we can consider an embedded sequence of finite element spaces, from the coarsest $\mathcal{V}_{h_1}$ to the finest $\mathcal{V}_{h_m}$, with

$$\mathcal{V}_{h_1} \subset \mathcal{V}_{h_2} \subset \cdots \subset \mathcal{V}_{h_m},$$

but also, in a more general way, increasing dimensional finite elements spaces,

$$\dim(\mathcal{V}_{h_1}) < \dim(\mathcal{V}_{h_2}) < \cdots < \dim(\mathcal{V}_{h_m}),$$

not necessarily embedded, and it will be the case for the numerical illustrations we present hereafter. We denote by A_{h_j} the associated stiffness and by M_{h_j}, the mass matrices $j = 1, \ldots, m$. We set $n_j = \dim(\mathcal{V}_{h_j})$.

Remark 3.3 It is not necessary in practice to have embedded finite elements spaces to apply the MSAM method, say $\mathcal{T}_{h_j} \subset \mathcal{T}_{h_{j+1}}$, $j = 1, \ldots, m-1$, as illustrated below (Figs. 11, 12).

The reader will find in Kuttler and Sigillito (1984) a extensive description of the eigenvalues and the eigenfunctions of the laplacian in space dimension two, for a number of domain shapes.

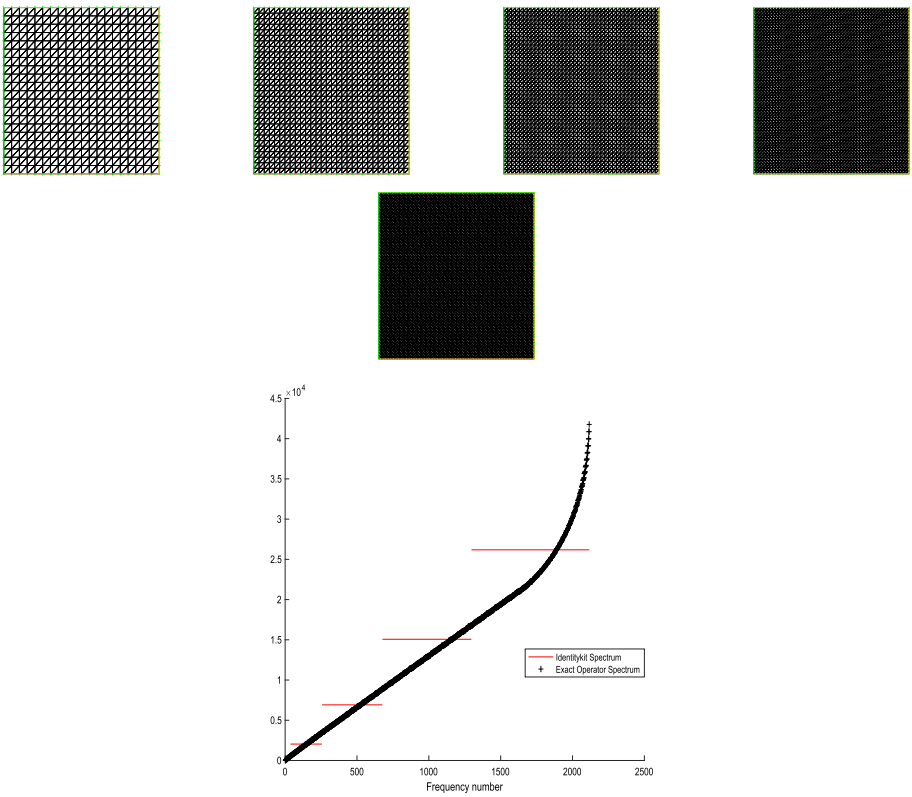

Fig. 11 Top: the successive meshes. Bottom: homogeneous Dirichlet problem on $\Omega =]0, 1[^2$. Comparison of Identikit spectrum of the stiffness matrix and of the exact symbol of the negative laplacian. $\mathbb{P}_1$ elements, n.d.o.f = 3721

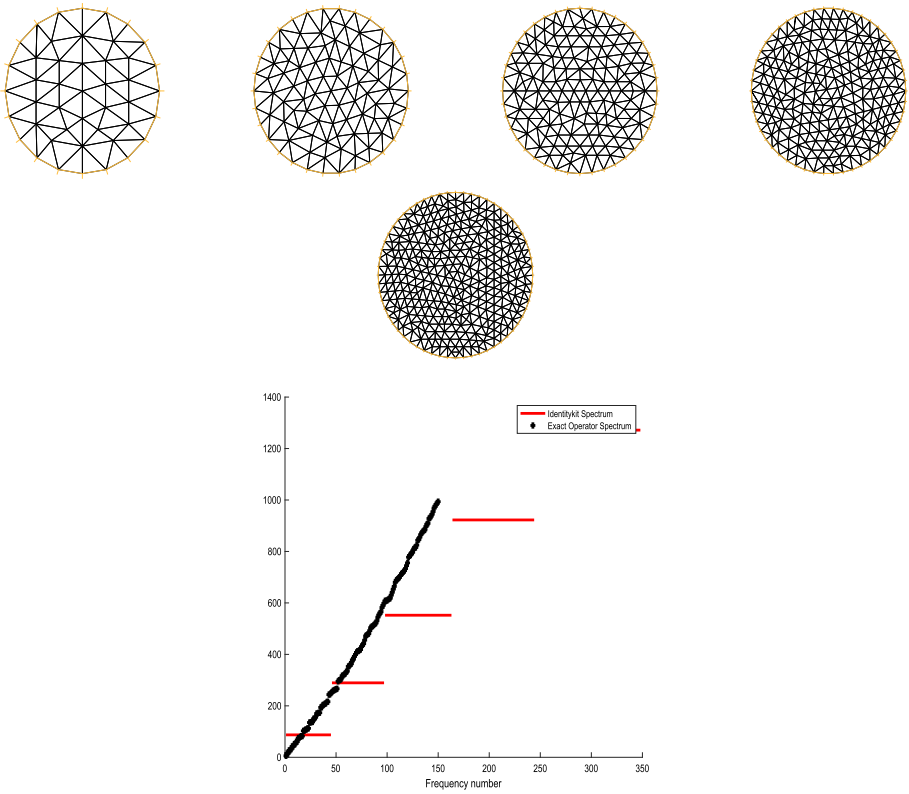

Fig. 12 Top: the successive meshes. Bottom: homogeneous Dirichlet problem on the unit Disk. Comparison of SAM of the stiffness matrix and of the exact symbol of the negative laplacian. $\mathbb{P}_1$ elements, n.d.o.f = 348

Algorithm 3 : MSAM method in Finite Elements

1: **Compute** $\bar{\lambda}_j \simeq \dfrac{\sum_{k=1}^{N} \langle A_{h_j} W_k^{(j)}, W_k^{(j)} \rangle}{\sum_{k=1}^{N} \langle M_{h_j} W_k^{(j)}, W_k^{(j)} \rangle}$.
2: **Initialisation** $\mu_1 = \bar{\lambda}_1$
3: **for** $j = 2, \ldots$ m, **do**
4: **Set** $\mu_j = \dfrac{1}{n_j - n_{j-1}} \left(n_j \bar{\lambda}_j - n_{j-1} \bar{\lambda}_{j-1} \right)$
5: **end for**
6: **Set** $\sigma_m(x) = \sum_{j=1}^{m} \mu_j \chi_{[n_{j-1}+1, n_j]}(x)$

3.5 Hierarchical methods

Another way to decompose a signal into a set of components, each one attached to a band of frequencies is the hierarchical approach. The separation of the components (or filtering) is realized considering several levels of discretization, as in the multi-grid methods in finite differences or finite elements. The leading idea is based on a numerical filtering of the solution; this principle can however be applied to other discretization such as wavelets (Mallat 2009), or spectral methods (Costa et al. 2001). We recall here briefly hereafter the hierarchy process in finite differences. We first consider for simplicity 2 levels of grid in the 1D case.

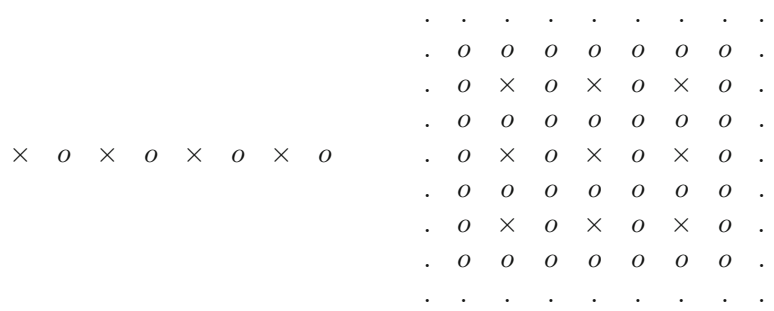

Fig. 13 Two levels hierarchical decompositions on the interval with periodic B.C. (left) and on the unit square with Dirichlet BC (right)

The use of different levels of discretization allows to decompose a signal (or values of a function at given grid points) into mean part and oscillatory (fluctuent) part. As in multigrid frameworks we consider two levels of discretization: given $h = \frac{1}{2n}$, the coarse grid $G_{2h} = \{x_i = 2hi, i = 0, \ldots n-1\}$ and the fine grid $G_h = \{x_i = ih, i = 0, \ldots 2n-1\}$. The filtering consists in computing proper local average of the function u at the grid points of $G_h \subset G_{2h}$ and to leave unchanged the values of u at the grid points of G_{2h}; the resulting vector $\bar{u}$ carries low frequencies components of u the high ones are smoothed by the average procedure; the resulting signal $z = u - \bar{u}$ is a high modes correction of $\bar{u}$ to u. The local average is based on a interpolation scheme and that the separation is more efficient when a high-order compact scheme is applied, see Chehab (2021, 1998) and the references therein.

We can resume the process with the two following steps:

- Step1 (Hierarchization): rewrite the vector of unknown as $u = (y, u_f)^t$; $y \in G_{2h}$, $u_f \in G_h \backslash G_{2h}$
- Step 2 (linear change of variable) replace u_f by $z = u_f - R(Y)$, where $R : G_{2h} \to G_h \setminus G_{2h}$ is a p-th order linear interpolation scheme; we have then $z = \mathcal{O}(h^p)$ for regular signals, the simplest one being the mid-point interpolation: $z_{2i+1} = u_{2i+1} - \frac{1}{2}(u_{2i} + u_{2i+2}) = \mathcal{O}(h^2)$. Higher interpolation schemes can be implemented using compact schemes.

As an illustration, we represent hereafter by a crux $'\times'$ the grid points belonging to G_{2h} and those of $G_h \backslash G_{2h}$ by a circle $'o'$, and by $'.'$ the points at the boundaries, See Fig. 13 hereafter:

When $m = d + 1$ levels of grids are used (with d dyadic refinements of the coarsest grid $G_{2^d h}$), $G_{2^d h} \subset \cdots G_{2h} \subset G_h$, we can rely the original (nodal) components to the new ones as

$$\begin{pmatrix} y \\ u_{f_1} \\ \vdots \\ u_{f_d} \end{pmatrix} = S \begin{pmatrix} y \\ z_1 \\ \vdots \\ z_d \end{pmatrix},$$

with obvious notations. Matrix S is called the transfer matrix. A a general construction in space dimension 2 and 3 is presented in Chehab (1998).

We give hereafter as an illustration in Figs. 14 and 15 the decomposition of approximation of $u(x) = \sin(100x(1-x))e^{-3x}$ at the fine grid points of G_h into several levels of oscillatory

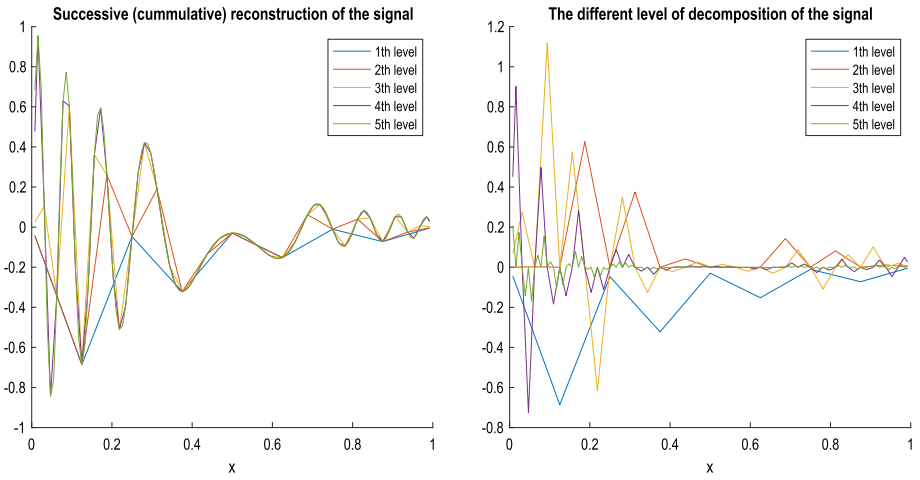

Fig. 14 Multilevel decomposition of the signal $u(x) = \sin(100x(1-x))$, $N = 127$. Second-order interpolation

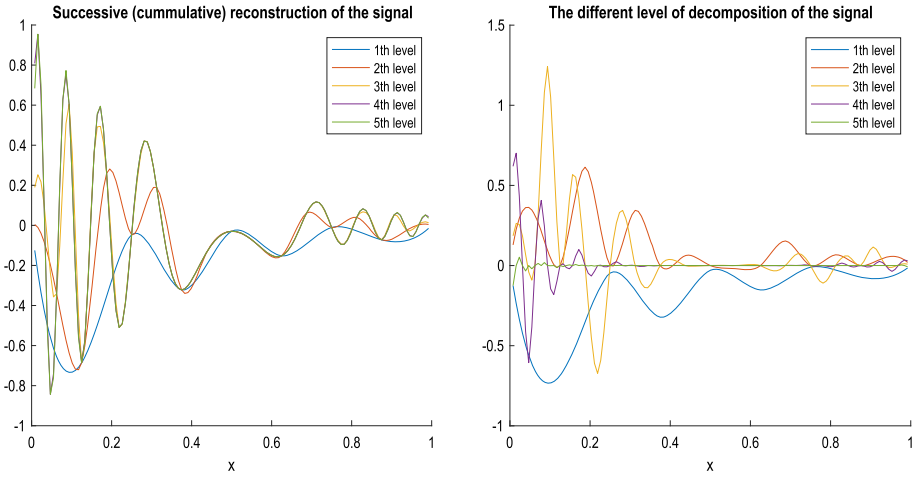

Fig. 15 Multilevel decomposition of the signal $u(x) = \sin(100x(1-x))$, $N = 127$. Fourth order interpolation

components, when second and fourth order interpolation formula are used, respectively, for the change of variable S. We observe that the fine grid corrections are smaller when using a higher order interpolation scheme.

To apply the SAM method, we must handle vectors of the same dimensions. We write

$$u = \begin{pmatrix} y \\ u_{f_1} \\ \vdots \\ u_{f_d} \end{pmatrix} = S\left(\begin{pmatrix} y \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ z_{f_1} \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ z_{f_d} \end{pmatrix}\right) = \tilde{u}_1 + \cdots \tilde{u}_m \tag{16}$$

We can now present the method as follows:

We give hereafter in Figs. 16 and 17, the approximation of the spectrum of the negative laplacian matrix in space dimensions 1 and 2, respectively.

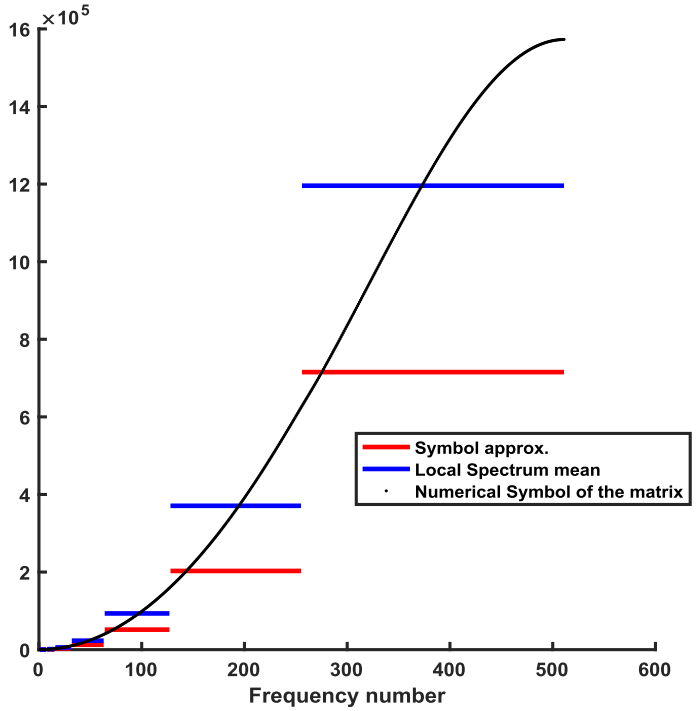

Fig. 16 Finite differences discretization. HSAM for the 1D Laplacian, $n = 511$, A is $n \times n$- 4th order discretization scheme

Algorithm 4 : Hierarchical SAM (HSAM)

Initialization: $S_{1,j} = S_{2,j} = 0$, $j = 1, \ldots, m$
for $k = 1, \ldots$ N **do**
  Decompose W_k as $W_k = \tilde{W}_k^{(1)} + \cdots + \tilde{W}_k^{(m)}$ following formula (16)
  for $j = 1, \ldots$ m **do**
    Compute $S_{1,j} = S_{1,j} + \langle A\tilde{W}_k^{(j)}, \tilde{W}_k^{(j)} \rangle$
    Set $S_{2,j} = S_{2,j} + \langle \tilde{W}_k^{(j)}, \tilde{W}_k^{(j)} \rangle$
  end for
end for
for $j = 1, \ldots$ m **do**
  Compute $\mu_j = \frac{S1_j}{S2_j}$
end for
Set $\sigma_m(x) = \sum_{j=1}^{m} \mu_j \chi_{[n_{j-1}+1, n_j]}(x)$

3.6 On the influence of the accuracy of the discretization schemes

We here give a description of the error of the SAM method. Let $A_n \in \mathcal{M}_n(\mathbb{R})$ be a symmetric matrix, obtained by a consistent discretization of the operator $\mathcal{A}$. We note by σ_{A_n} the numerical symbol of A_n and by $\sigma_{\mathcal{A}}$ the one of $\mathcal{A}$. We consider the decomposition of the numerical frequency interval $[1, n]$ as $[1, n] = \cup_{j=0}^{m-1}[k_j + 1, k_{j+1}]$. For a given sequence of sample

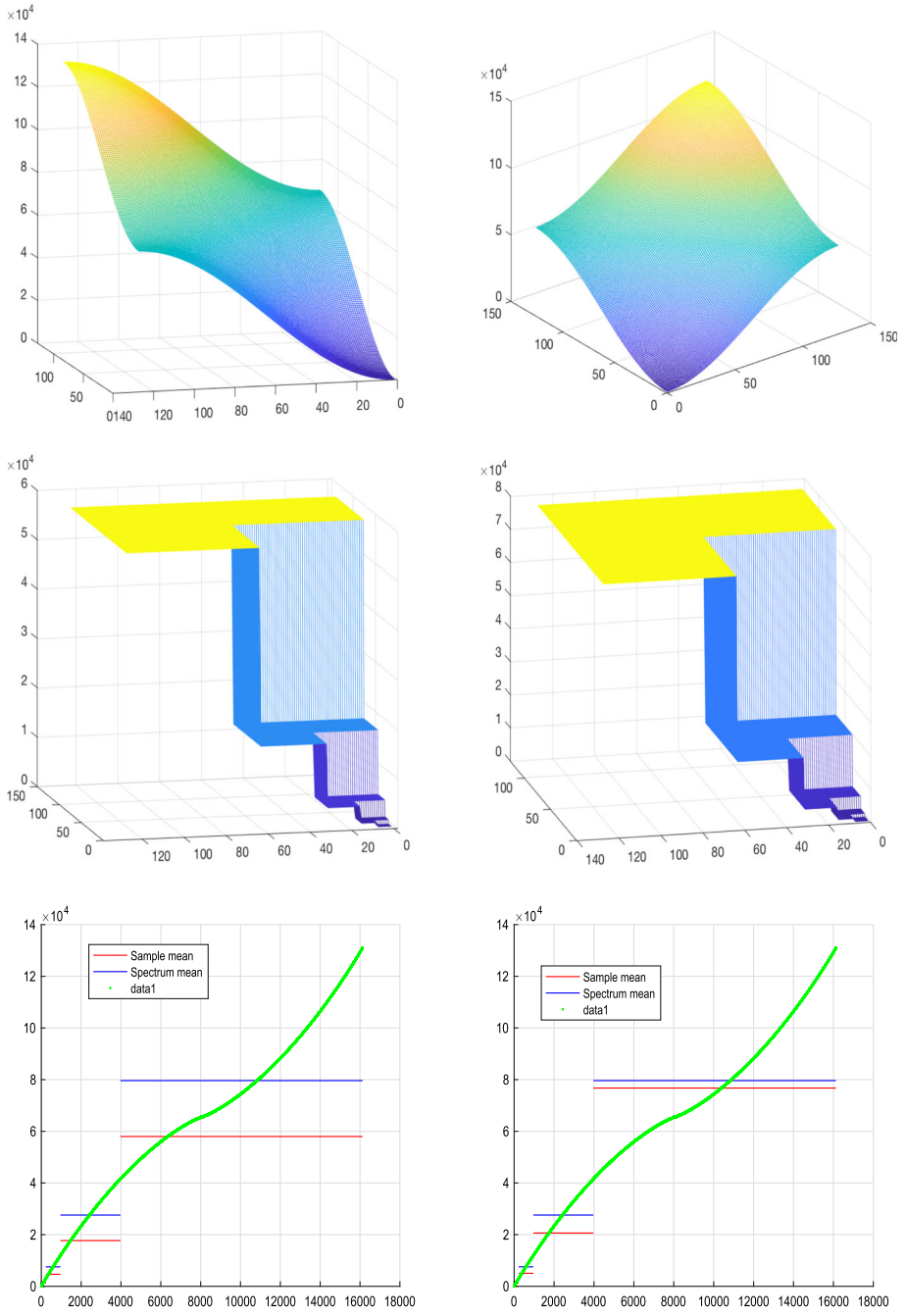

Fig. 17 Finite differences discretization. HSAM for the 2D Laplacian $n = 127^2$, A is $n \times n$-2d order discretization scheme (column 1), 4th order discretization scheme (column 2). Second order filter (2d order interpolation scheme)

vectors $(W^{(k)})_{k=1}^N \in \mathbb{R}^n$, the error of the SAM method satisfies the estimate:

$$|\sigma_{\mathcal{A}}(k) - \frac{\sum_{\ell=1}^N \langle AW_j^{(\ell)}, W_j^{(\ell)}\rangle}{\sum_{i=1}^N \langle W_j^{(\ell)}, W_j^{(\ell)}\rangle}| \leq |\sigma_{A_n}(k) - \sigma_{\mathcal{A}}(k)| + |\sigma_{A_n}(k) - \frac{\sum_{\ell=1}^N \langle AW_j^{(\ell)}, W_j^{(\ell)}\rangle}{\sum_{i=1}^N \langle W_j^{(\ell)}, W_j^{(\ell)}\rangle}|,$$
$$k_j + 1 \leq k \leq k_{j+1}, \forall j = 1, \ldots, m.$$

This is a simple consequence of the triangular inequality. It allows to decompose the error $E = E_1 + E_2$ into $E_1 = \sigma_{A_n}(k) - \sigma_{\mathcal{A}}(k)$ which depends only on the discretization scheme (and its accuracy), and $E_2 = \frac{\sum_{\ell=1}^N \langle AW_j^{(\ell)}, W_j^{(\ell)}\rangle}{\sum_{i=1}^N \langle W_j^{(\ell)}, W_j^{(\ell)}\rangle} - \sigma_{A_n}(k)$, which depends on the "quality" of the sample vectors and of the capability of the numerical filtering to separate group of frequencies. We now can give the following result:

Proposition 3.4 *Let $A_n \in \mathcal{M}_n(\mathbb{R})$ be a symmetric matrix, obtain by a consistent discretization of the operator $\mathcal{A}$. We note by σ_{A_n} the numerical symbol of A_n and by $\sigma_{\mathcal{A}}$ the one of $\mathcal{A}$. We consider the decomposition of the numerical frequency interval $[1, n]$ as $[1, n] = \cup_{j=0}^{m-1}[k_j + 1, k_{j+1}]$. For a given sequence of sample vectors $(W^{(k)})_{k=1}^N \in \mathbb{R}^n$, the error of the SAM method satisfies the estimate: If we assume in addition that for any fixed $M \in \mathbb{N}^*$, we have:*

$$\lim_{n\to+\infty} \max_{1\leq k\leq M} |\sigma_{A_n}(k) - \sigma_{\mathcal{A}}(k)| = 0,$$

then, if n is large enough,

$$|\sigma_{\mathcal{A}}(k) - \frac{\sum_{\ell=1}^N \langle AW_j^{(\ell)}, W_j^{(\ell)}\rangle}{\sum_{i=1}^N \langle W_j^{(\ell)}, W_j^{(\ell)}\rangle}| \leq 2|\sigma_{A_n}(k) - \frac{\sum_{\ell=1}^N \langle AW_j^{(\ell)}, W_j^{(\ell)}\rangle}{\sum_{i=1}^N \langle W_j^{(\ell)}, W_j^{(\ell)}\rangle}|, \forall k \leq M.$$

Proof The inequality shows that the more accurate is the discretization scheme, the smaller is the error term E_1 so, for n large enough and for a fixed M, we obtain the result. □

We now give an illustration of this result by considering the approximation of the symbol of $\mathcal{A} = -\Delta$ on $]0, 1[$ associated to Homogeneous Dirichlet Boundary conditions. We consider the discretization of $\mathcal{A}$ by 2d, 4th and 6th order discretization schemes, the two last ones being obtain by using compact scheme, Lele (1992), we used $N = 50$ sample Rademacher vectors. Figure 18 gives an illustration of Proposition 3.4: we observe that passing to the 4th and 6th order discretization schemes of $-\Delta$, MSAM gives a very satisfactory approximation σ. Also, in this context, the multi-grid filtering is effective.

4 SAM for time-dependent problems

In a number of situations, a physical model has to be corrected (or enriched) by taking into account phenomena that have been ignored at first. For instance, in hydrodynamics, the damping of waves is observed experimentally, and a still challenging question is its mathematical representation, see, e.g. Chehab and Sadaka (2013a, b); Dumont and Duval (2013); Dumont and Manoubi (2018); Dutykh (2009); Ott and Sudan (1969, 1970). The most common approach is to try to express the damping force with a linear operator $\mathcal{B}$; in particular cases, it can be obtained by a formal derivation, but it can give rise to tricky expressions difficult to be handled and implemented numerically (Dutykh and Le Meur 2021; Le Meur 2015). Another approach on which we concentrate here consists in approaching $\mathcal{B}$ by using experimental or numerical samples.

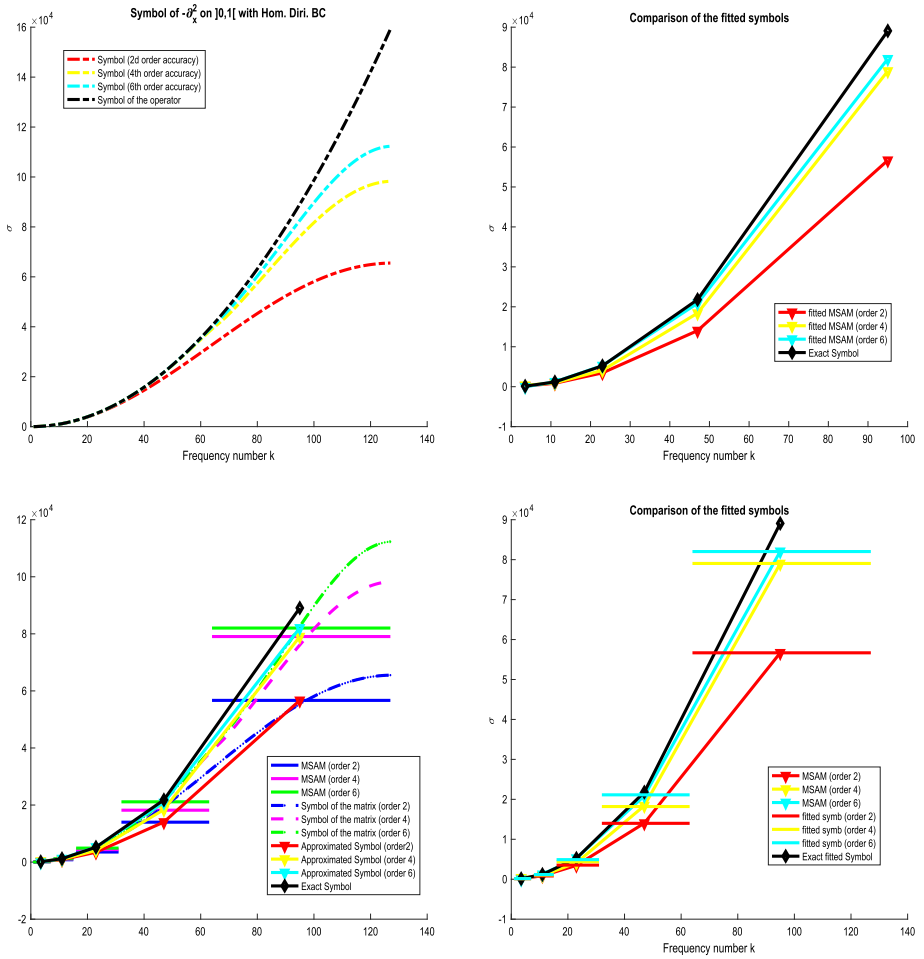

Fig. 18 Comparisons of the SMA method when using 2d, 4th and 6th order discretization schemes. $N = 50$, $n = 127$

4.1 Derivation of the method

We adapt here the numerical approximation to the symbol developed above. Consider the evolution system

$$\frac{\partial u}{\partial t} + \mathcal{A}u + F(u) = 0, t \in]0, T), \tag{17}$$

which is valid in absence of damping phenomena. We would like to identify a damping operator by fitting a set of experimental data with the enhanced model

$$\frac{\partial u}{\partial t} + \mathcal{A}u + F(u) + \mathcal{B}u = 0, \tag{18}$$

where $\mathcal{B}$ represents the (unknown) damping operator; $\mathcal{B}$ is assumed to be time independent, auto-adjoint and positive definite. When discretizing in space the last system, we get

$$\frac{du}{dt} + Au + F(u) + Bu = 0, t \in [0, T]. \tag{19}$$

Assume that we have at our disposal measured physical data $W(t)$ at discrete times $t_k \in [0, T]$; It is important to precise that the dimension of the differential system is supposed to be fixed, e.g. limited in practice by the size of exprimental measures $W(t)$.

We would like to compute B in such a way to fit with W. To this end, we consider a stable time marching scheme, e.g.

$$\frac{W^{(k+1)} - W^{(k)}}{\Delta t} + AW^{(k+1)} + F(W^{(k+1)}) + BW^{(k+1)} = 0,$$

or, equivalently,

$$BW^{(k+1)} = -\frac{W^{(k+1)} - W^{(k)}}{\Delta t} - AW^{(k+1)} - F(W^{(k+1)}) = R^{(k+1)}, k = 0, \ldots, N-1, \tag{20}$$

with $N\Delta t = T$. We propose to apply the multilevel SAM as follows:

- The sequence $W^{(k)}$ is supposed to be known, in practice, it can be given by experimental measures.
- Compute $R^{(k+1)}, k = 0, \ldots, N-1$ by formula (20),
- Decompose each $W^{(k+1)}$ and $R^{(k+1)}$ using frequency filters $\Pi_j,\ j = 1, \ldots, m$ (spectral, multi-grid or so) into $W_j^{(k)} = \Pi_j W^{(k)}$ and $R_j^{(k)} = \Pi_j R^{(k)}$, $j = 1, \ldots, m$,
- Compute

$$S_j^{(k)} = \langle R_j^{(k+1)}, W_j^{(k)} \rangle \quad \text{and} \quad T_j^{(k)} = \langle W_j^{(k+1)}, W_j^{(k)} \rangle, \quad j = 1, \ldots, m,$$

- Set

$$\mu_j = \frac{\sum_{k=0}^{N-1} S_j^{(k)}}{\sum_{k=0}^{N-1} T_j^{(k)}}, \quad j = 1, \ldots, m.$$

4.2 Application to the Korteweg–de-Vries equation

We consider the Korteweg–de-Vries model (KdV), which is obtained from Euler's equations by selecting a particular physical regime: small amplitude elevation, large wavelength, unidirectional propagation, see Miranville and Temam (2005); it addresses then to low frequency regimes. The long time behavior of dissipative asymptotic models is still an important issue: the capture of damping rates in several norms, the measure of regularization effects, the evidence of complex asymptotic dynamics, just to name but a few, are important questions to be considered when trying understand natural phenomena. Mostly, several of these questions are still open and the numerical simulation is a way to capture some properties, to select pertinent models and to develop theoretical as well as practical strategies.

We here focus on KdV equations on the torus $\mathbb{T} = \mathbb{T}(0, L)$; other dispersive models such as BBM (Bona-Benjamin-Mahony) or BO (Benjamin–Ono) equations could also studied following a similar approach.

Damped Korteweg–de Vries equations appear in different physical situations, they can be expressed in a large generality as

$$u_t + \mathcal{B}(u) + u_{xxx} + uu_x = 0, x \in \mathbb{T},\ t > 0, \tag{21}$$

where $\mathcal{B}$ is a linear operator, defined on a Hilbert space V, subspace of L^2 with values in L^2, and satisfying

$$\int_0^L \mathcal{B}(v)vdx \geq 0, \tag{22}$$

for all function $v \in V$, regular enough, in such a way the L^2-norm of the solution is decreasing in time as

$$\frac{1}{2}\frac{d|u|_{L^2}^2}{dt} + \int_0^L \mathcal{B}(u)udx = 0. \tag{23}$$

We find in the literature different choices for $\mathcal{L}$, depending on the physical situations, see Cabral and Rosa (2004); Chehab and Sadaka (2013a, b); Chehab et al. (2015); Chen et al. (2010); Dias and Dutykh (2007); Dumont and Duval (2013); Dumont and Manoubi (2018); Dutykh (2009); Ott and Sudan (1969, 1970) and Jäger (1982) for physical experiments.

A sampling we consider the solution generated by the damped KdV model, starting from a soliton initial datum. The initial datum is

$$u_0(x, 0) = \alpha \sec^2(\beta(x - x0) - C.t),$$

with $\epsilon = 1$, $L = 100$, $x_0 = L/2$, $\alpha = 0.8$, $C = 2\alpha/6$, $\beta = \frac{1}{2}\sqrt{C\epsilon}$. In Fig. 19, we represent the energy spectrum of u_0; we observe that the bandwidth is relatively important.

We first display the results when approaching the symbol of the damping operator by local spectral means.

The results show a very good agreement and the symbol of the damping operator is very correctly captured. This is also observed for operator like Id, $\sqrt{-\Delta}$.

The results of SAM when using spectral Fourier space discretization are reported in Fig. 20: the symbols of $\mathcal{B} = \nu Id$, $\mathcal{B} = -\nu\frac{\partial^2}{\partial x^2}$, $\mathcal{B} = \nu\sqrt{-\frac{\partial^2}{\partial x^2}}$ and $\mathcal{B} = \nu(Id - \frac{\partial^2}{\partial x^2})^{-1}$ are very correctly approached by the method.

We now consider the multi-grid case in 1D, base on a 4th order interpolation. The discretization is realized with 4th order compact finite differences schemes, we proceed as in see Calgaro et al. (2008).

In Fig. 21, we observe that the symbols of $\mathcal{B} = \nu Id$, $\mathcal{B} = -\nu\frac{\partial^2}{\partial x^2}$ and $\mathcal{B} = \nu\sqrt{-\frac{\partial^2}{\partial x^2}}$ are correctly recovered numerically by the method; in the case $\mathcal{B} = \nu(Id - \frac{\partial^2}{\partial x^2})^{-1}$, the symbol is only well approach for the lower and the higher frequencies, however it appears to be a low-pass filter.

4.3 Application to the Benjamin–Ono equation

The Benjamin–Ono equation (BO) describes one-dimensional internal waves in deep water, see Brooke Benjamin (1967); Ono (1975). It reads as

$$\begin{cases} u_t - \mathcal{H}(u_{xx}) + u^p u_x = 0 & x \in \mathbb{R},\ t > 0, \\ u(x, 0) = u_0(x) & x \in \mathbb{R}, \end{cases} \tag{24}$$

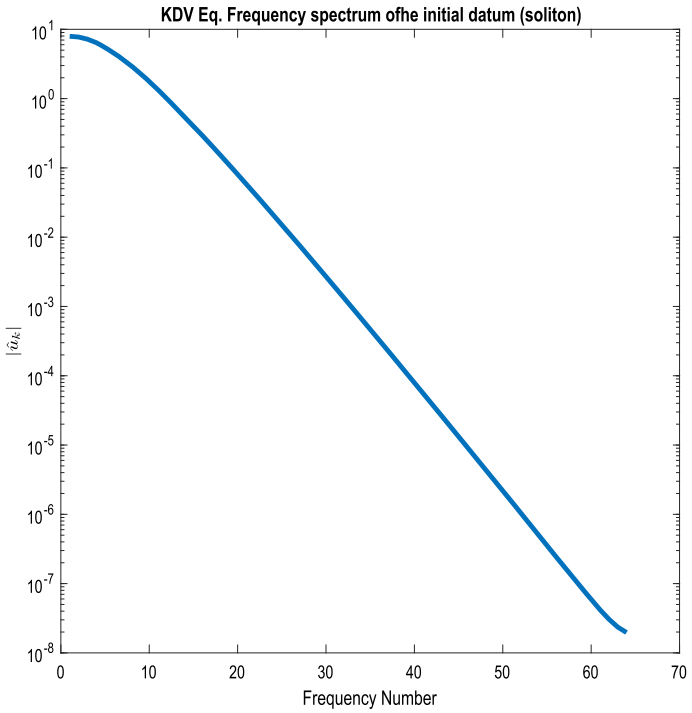

Fig. 19 KdV equation—spectral Fourier discretization in space. Energy spectrum of the initial soliton datum. $\nu = 1$, $\Delta t = 0.001$, $L = 100$, $n = 128$; Line 1 γId (left) and Δ (right), Line 2 $\sqrt{-\Delta}$ (left), $(1-\Delta)^{-1}$ (right)

with $p \geq 1$; here $\mathcal{H}$ is the Hilbert transform:

$$\mathcal{H}(u) = PV\frac{1}{\pi}\int_{\mathbb{R}} \frac{u(x-y)}{y}dy.$$

BO equation is rewritten into its conservative form:

$$\begin{cases} u_t - \partial_x\left(\mathcal{H}(u_x) + \frac{1}{p+1}u^{p+1}\right) = 0 & x \in \mathbb{R},\ t > 0, \\ u(x,0) = u_0(x) & x \in \mathbb{R}. \end{cases} \tag{25}$$

When $p = 1$, BO possess an infinity of invariants, see Abdelouhab et al. (1989), from which

$$\begin{aligned} I_{-1}(u) &= \int u dx, \\ I_0(u) &= \int u^2 dx, \\ I_1(u) &= \frac{1}{3}\int u^3 dx - \int u\mathcal{H}u_x dx, \\ I_2(u) &= \frac{1}{4}\int u^4 dx - \frac{3}{2}\int u^2\mathcal{H}u_x, dx + 2\int u_x^2 dx. \end{aligned}$$

When $p \geq 2$, they are only three invariants:

$$\begin{aligned} I_{-1}(u) &= \int u dx, \\ I_0(u) &= \int u^2 dx, \\ I_1(u) &= \frac{2}{(p+1)(p+2)}\int u^{p+2} dx - \int u\mathcal{H}u_x dx, \end{aligned}$$

see Bona and Kalisch (2004).

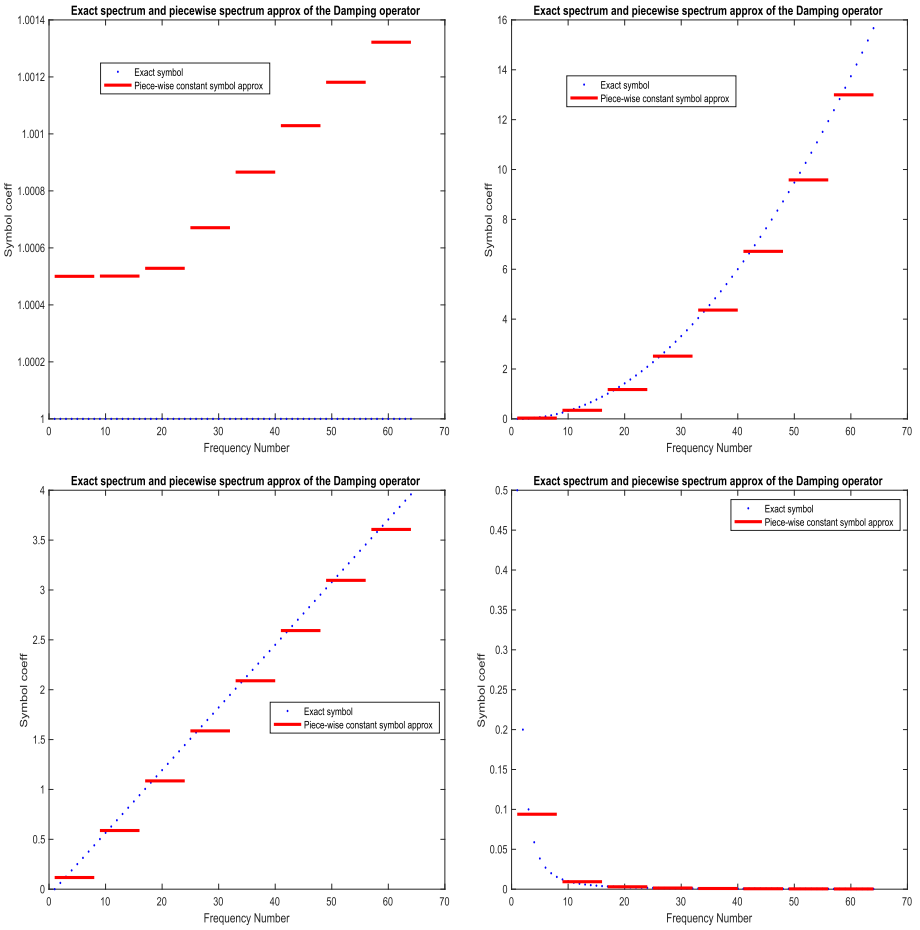

Fig. 20 KdV equation. Approximation of the symbol of the damping operators by spectral local means. $\nu = 1$, $\Delta t = 0.001$, $L = 100$, $N = 128$; Line 1 γId (left) and Δ (right), Line 2 $\sqrt{-\Delta}$ (left), $(1-\Delta)^{-1}$ (right)

We displayed a Sanz–Serna Scheme to simulate the equation and computed approximation to the symbol of the damping operator. We use both the spectral approach and the multi-grid one in finite differences. In Fig. 22, we represent the estimate of the damping operator $\mathcal{B} = \sqrt{-\Delta}$ when spectral Fourier discretization and spectral SAM methods are applied. We observe that the symbol is correctly approached only for low modes: this is due to the spectral profile of the soliton (build only low frequencies).

In Fig. 23, we have represented HSAM applied to finite differences discretization and hierarchical filtering. We observe as above, that the symbol is only well approached for very low frequencies. This is not surprising since the soliton is to poor in frequencies, the approximation of the symbol in the Fourier spectral case is also very limited to the low frequencies, as showed in Fig. 22.

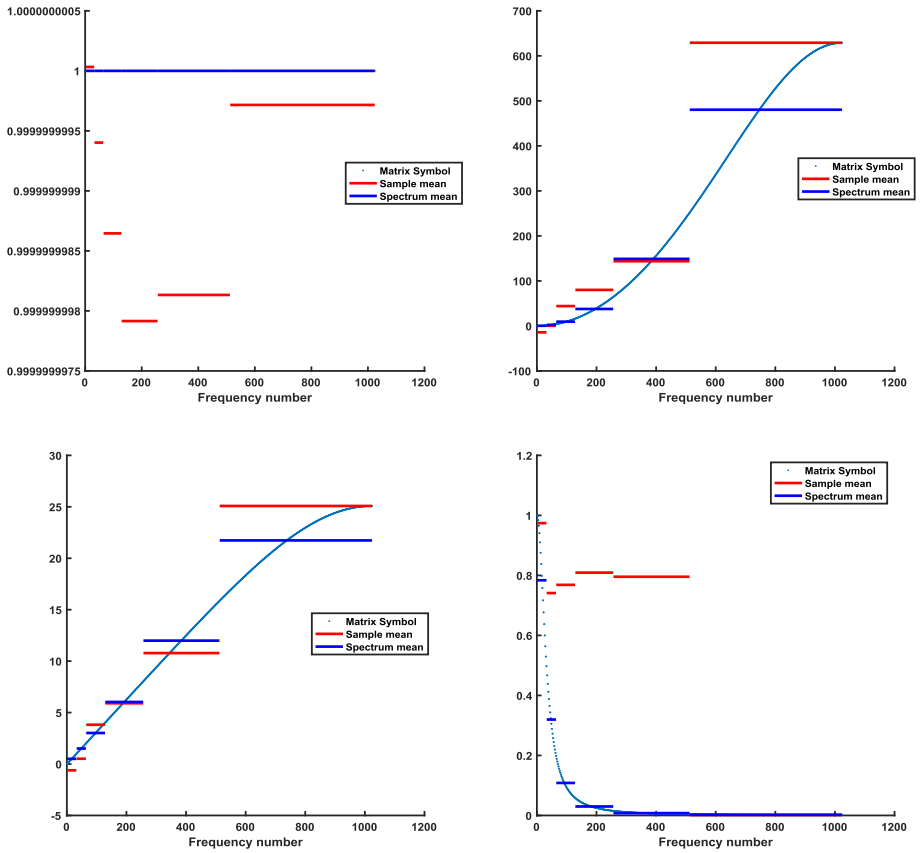

Fig. 21 KdV equation. Approximation of the symbol of the damping operator by hierarchical method in finite differences. $\nu = 1$, $n = 1024$, $\Delta t = 0.01$, $N = 5000$, $L = 100$. Initial datum: soliton

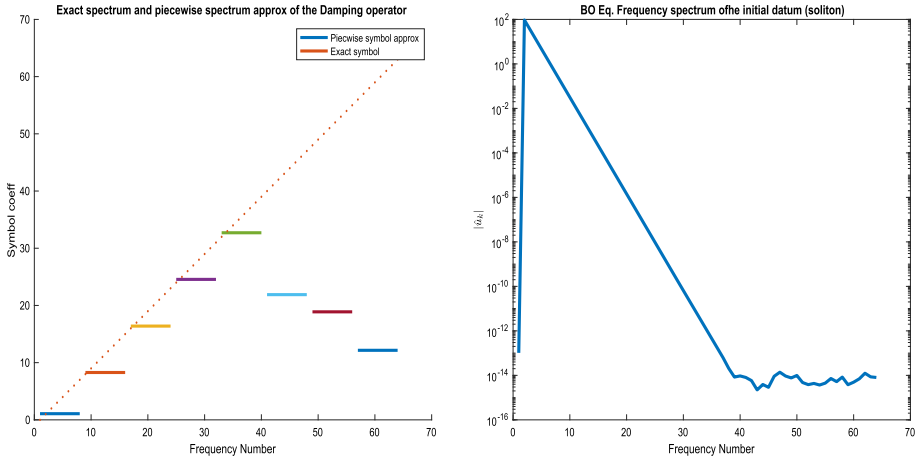

Fig. 22 Benjamin Ono equation. Approximation of the symbol of the damping operator $\sqrt{-\Delta}$ by spectral local means (left), spectrum energy of the initial datum (right) $n = 64$

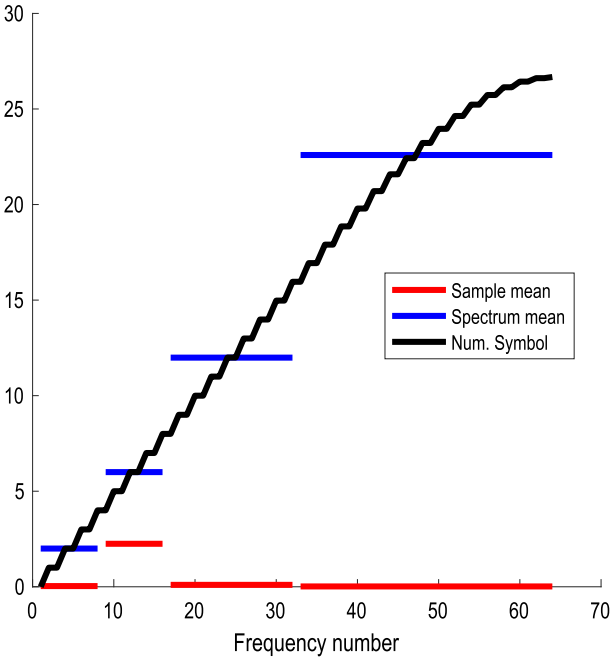

Fig. 23 Benjamin Ono equation. Approximation of the symbol of the damping operator $\sqrt{-\Delta}$ by spectral local means. HSMA and 4th order finite differences discretization. $n = 64$

5 Concluding remarks and further developments

The numerical procedures proposed in the present work (SAM, MSAM and HSAM) allow to produce a simple, fast and satisfactory approximation of the symbol of dissipative and damping operators, using different technique of discretization (Spectral Fourier, Finite Differences and Finite Elements). Of course natural extensions to orthogonal polynomials of wavelet discretization are possible since they allow to produce a separation of set of frequencies for a given data.

It appears (Proposition 3.4 and illustrations) that the three important ingredients from which depends the efficiency of the approach are:

- The discretization scheme
- The choice of the filtering procedure (Multi-grid or so)
- The choice of the sample (Rademacher, experimental data)

Working with samples data that cover an important band of frequencies seems essential to capture the part of the symbol associated to high frequency numbers. This was pointed out when considering KdV then BO equations.

As further developments, we can consider more models in hydrodynamics (such as from Boussinesq's family), and also the application of extrapolation formula to improve the SAM method, as pointed in Remark 3.1.

We can also consider the case in which the damping operator $\mathcal{B}$ is time dependent and derive corresponding SAM approximation:

$$\frac{du}{dt} + Au + F(u) + B(t)u = 0, t \in [0, T]. \tag{26}$$

The object here to be approached is the "spectrogram" of B. We decompose the time interval $[0, T]$ as $[0, T] = \cup_{p=0}^{q-1}[t_p, t_{p+1}]$ with

$$t_0 < t_1 < \cdots < t_P = T,$$

Let $\Delta t > 0$ be given, we consider for simplicity the case $t_{p+1} - t_p = cst = r\Delta t$ for a given integer r. The sequence $w^{(k)}$ satisfies

$$B_k W^{(k+1)} = -\frac{W^{(k+1)} - W^{(k)}}{\Delta t} - AW^{(k+1)} - F(W^{(k+1)}) = R^{(k+1)}, k = 0, \ldots, N = qr.$$

We propose to apply the SAM method as follows: let $p \in 1, \ldots, q$ be fixed

- Compute the sequence $R^{(k+1)} = -\frac{W^{k+1} - W^{(k)}}{\Delta t} - AW^{(k+1)} - F(W^{(k+1)}), k = (p-1)r + 1 \ldots pr$
- We decompose $W^{(k+1)}$) and $R^{(k+1)}$ using frequency filters $\Pi_j,\ j = 1, \ldots, m$ (spectral, multigrid or so)
- We compute

  $S_j^{(k,p)} = \langle \Pi_j R^{(k+1)}, \Pi_j W^k \rangle$ and $T_j^k = \langle \Pi_j W^{(k+1)}, \Pi_j W^{(k)} \rangle,\ k = (p-1)r + 1 \ldots pr$

- We set

$$\lambda_j^{(p)} = \frac{\sum_{k=(p-1)r+1}^{pr} S_j^{(k,p)}}{\sum_{k=0}^{N-1} T_j^{(k,p)}}, j = 1, \ldots, m.$$

The symbol of B in the interval $[t_p, t_{p+1}]$ will be approached by the piece-wise constant function based in the sequence $\lambda_j^{(p)}$, $j = 1, \ldots, m$.

Declarations

Conflict of interest The author declares no conflict of interest in this paper.

References

Abdelouhab L, Bona JL, Fellant M, Saut J-C (1989) Nonlocal models for nonlinear dispersive waves. Phys D 40:360–392

Avron H, Toledo S (2011) Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. JACM 58:1–34

Bai Z, Fahey G, Golub G (1996) Some large-scale matrix computation problems. J Comput Appl Math 74(1):71–89

Bona JL, Kalisch H (2004) Singularity formation in the generalized Benjamin-Ono equations. Discrete Contin Dyn Syst 11(1):27–45

Brezinski C (2000) Convergence acceleration during the 20th century. J Comput Appl Math 122:1–21. Numerical Analysis in the 20th Century Vol. II: Interpolation and Extrapolation

Brezinski C, Redivo-Zaglia M (1991) Extrapolation methods. Theory and practice. North-Holland, Amsterdam

Brooke Benjamin T (1967) Internal waves of permanent form in fluids of great depth. J Fluid Mech 29(3):559

Cabral M, Rosa R (2004) Chaos for a damped and forced KdV equation. Phys D 192:265–278

Cahouet J, Chabard J-P (1988) Some fast 3d finite element solvers for the generalized Stokes problem. Int J Numer Methods Fluids 8:869–895

Calgaro C, Chehab JP, Laminie J, Zahrouni E (2008) Séparation des échelles et schémas multiniveaux pour les équations d'ondes non-linéaires (French) [Scale separation and multilevel schemes for nonlinear wave equations]. CANUM ESAIM Proc 27:180–208

Chehab J-P (1998) Incremental unknowns method and compact schemes. ESAIM Math Model Numer Anal 32(1):51–83

Chehab J-P (2021) Stabilization and numerical filtering for the modeling and the simulation of time dependent PDEs. Discrete Contin Dyn Syst 14(8):2693–2728
Chehab J-P, Sadaka G (2013a) Numerical Study of a family of dissipative KdV equations. Commun Pure Appl Anal 12:519–546
Chehab J-P, Sadaka G (2013b) On Damping Rates of dissipative KdV equations. Discrete Contin Dyn Syst Ser S 6:1487–1506
Chehab JP, Garnier P, Mammeri Y (2015) Long-time behavior of solutions of a BBM equation with generalized damping. Discrete Contin Dyn Syst B, 1897–1915
Chen M, Dumont S, Dupaigne L, Goubet O (2010) Decay of solutions to a water wave model with nonlocal viscous dispersive term. Discrete Contin Dyn Syst 27:1473–1492
Costa B, Dettori L, Gottlieb D, Temam R (2001) Time marching techniques for the nonlinear Galerkin method. SIAM J Sci Comput 23:46–65
Crouzeix M (1974) Approximation et méthodes itératives de résolution d'inéquations variationelles et de problèmes non linéaires, I.R.I.A. Cahier 12:139–244
Crouzeix M (1997) On an operator related to the convergence of Uzawa's algorithm for the Stokes equation. Computational science for the 21st century, pp 242–249
Dias F, Dutykh D (2007) Viscous potential free-surface flows in a fluid layer of finite depth. C R Math Acad Sci Paris 345:113–118
Dumont S, Duval JB (2013) Numerical investigation of asymptotical properties of solutions to models for water waves with non local viscosity. Int J Numer Anal Model 10(2):333–349
Dumont S, Manoubi I (2018) Numerical analysis of a water wave model with a nonlocal viscous dispersive term using the diffusive approach. Math Methods Appl Sci 41(12):4810–4826
Dutykh D (2009) Viscous-potential free-surface flows and long wave modeling. Eur J Mech B Fluids 28:430
Dutykh D, Le Meur HVJ (2021) Derivation of a viscous Serre–Green–Naghdi equation: an impasse? Fluids 6(4):135
FreeFem++ page. http://www.freefem.org
Goubet O, Zahrouni E (2008) On time discretization of a weakly damped forced nonlinear Schrödinger equation. Commun Pure Appl Anal 7(6):1429–1442
Hedayat AS, Wallis WD (1978) Hadamard matrices and their applications. Ann Stat 6(6):1184–1238
Hutchinson MF (1989) A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. Commun Stat Simul Comput 18:1059–1076
Jäger D (1982) Experiments on KdV solitons. J Phys Soc Jpn 51:1686–1693
Kuttler JR, Sigillito VG (1984) Eigenvalues of the Laplacian in two dimensions. SIAM Rev 26(2):163–193
Le Meur HVJ (2015) Derivation of a viscous Boussinesq system for surface water waves. Asymptot Anal 94(3–4):30–345
Lele S (1992) Compact difference schemes with spectral like resolution. J Comput Phys 103:16–42
Mallat S (2009) A wavelet tour of signal processing. The sparse way book, 3rd edn. Elsevier, Amsterdam
Miranville A, Temam R (2005) Mathematical modeling in continuum mechanics, 2nd edn. Cambridge University Press, Cambridge
Ono H (1975) Algebraic solitary waves in stratified fluids. J Phys Soc Jpn 39(4):1082–1091
Ott E, Sudan RN (1969) Nonlinear theory of ion acoustic wave with Landau damping. Phys Fluids 12:2388–2394
Ott E, Sudan RN (1970) Damping of solitary waves. Phys Fluids 13:1432–1435
Peyret R, Taylor R (1983) Computational methods for fluid flow, Springer Series in Computational Physics. Springer, New York
Ubaru S, Saad Y (2018) Applications of trace estimation techniques. In: High performance computing in science and engineering, pp 19–33