

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221660089>

# Adaptive Fourier–Galerkin Methods

Article in *Mathematics of Computation* · January 2012

DOI: 10.1090/S0025-5718-2013-02781-0 · Source: arXiv

CITATIONS

8

READS

28

3 authors, including:



**Claudio Canuto**

Politecnico di Torino

228 PUBLICATIONS 10,028 CITATIONS

[SEE PROFILE](#)



**Marco Verani**

Politecnico di Milano

98 PUBLICATIONS 899 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Multi-phase fluid flows [View project](#)



Numerical Methods for PDEs on Polygonal and Polyhedral Meshes [View project](#)

# Adaptive Fourier-Galerkin Methods

Claudio Canuto<sup>a</sup>, Ricardo H. Nochetto<sup>b</sup> and Marco Verani<sup>c</sup>

January 26, 2012

<sup>a</sup> Dipartimento di Scienze Matematiche, Politecnico di Torino  
Corso Duca degli Abruzzi 24, 10129 Torino, Italy  
E-mail: [claudio.canuto@polito.it](mailto:claudio.canuto@polito.it)

Department of Mathematics and Institute for Physical Science and Technology,  
University of Maryland, College Park, MD 20742, USAy  
E-mail: [rhn@math.umd.edu](mailto:rhn@math.umd.edu)

<sup>c</sup> MOX, Dipartimento di Matematica, Politecnico di Milano  
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy  
E-mail: [marco.verani@polimi.it](mailto:marco.verani@polimi.it)

## Abstract

We study the performance of adaptive Fourier-Galerkin methods in a periodic box in  $\mathbb{R}^d$  with dimension  $d \geq 1$ . These methods offer unlimited approximation power only restricted by solution and data regularity. They are of intrinsic interest but are also a first step towards understanding adaptivity for the *hp*-FEM. We examine two nonlinear approximation classes, one classical corresponding to algebraic decay of Fourier coefficients and another associated with exponential decay. We study the sparsity classes of the residual and show that they are the same as the solution for the algebraic class but not for the exponential one. This possible sparsity degradation for the exponential class can be compensated with coarsening, which we discuss in detail. We present several adaptive Fourier algorithms, and prove their contraction and optimal cardinality properties.

**Keywords:** Spectral methods, adaptivity, convergence, optimal cardinality.

## 1 Introduction

Adaptivity is now a fundamental tool in scientific and engineering computation. In contrast to the practice, which goes back to the 70's, the mathematical theory for multidimensional problems is rather recent. It started in 1996 with the convergence results by Dörfler [13] and Morin, Nochetto, and Siebert [18]. The first convergence rates were derived by Cohen, Dahmen, and DeVore [7] for wavelets in any dimensions  $d$ , and for finite element methods (AFEM) by Binev, Dahmen, and DeVore [2] for  $d = 2$  and Stevenson [21] for any  $d$ . The most comprehensive results for AFEM are those of Cascón, Kreuzer, Nochetto, and Siebert [6] for any  $d$  and  $L^2$  data, and Cohen, DeVore, and Nochetto [8] for  $d = 2$  and  $H^{-1}$  data; we refer to the survey [19] by Nochetto, Siebert and Veiser. This theory is quite satisfactory in that it shows that AFEM delivers a convergence rate compatible with that of the approximation classes where the solution and data belong. The recent results in [8] reveal that it is the approximation class of the solution that really matters. In all cases though the convergence rates are limited by the approximation

power of the method (both wavelets and FEM), which is finite and related to the polynomial degree of the basis functions, and the regularity of the solution and data. The latter is always measured in an *algebraic* approximation class.

In contrast very little is known for methods with infinite approximation power, such as those based on Fourier analysis. We mention here the results of DeVore and Temlyakov [12] for trigonometric sums and those of Binev et al [1] for the reduced basis method. A close relative to Fourier methods is the so-called *p*-version of the FEM (see e.g. [20] and [5]), which uses Legendre polynomials instead of exponentials as basis functions. The purpose of this paper is to present *adaptive Fourier-Galerkin methods (ADFOUR)*, and discuss their convergence and optimality properties. We do so in the context of both *algebraic* and *exponential* approximation classes, and take advantage of the orthogonality inherent to complex exponentials. We believe that this approach can be extended to the *p*-FEM. We view this theory as a first step towards understanding adaptivity for the *hp*-FEM, which combines mesh refinement (*h*-FEM) with polynomial enrichment (*p*-FEM) and is much harder to analyze.

Our investigation reveals some striking differences between ADFOUR and AFEM and wavelet methods. The basic assumption, underlying the success of adaptivity, is that the information read in the residual is quasi-optimal for either mesh design or choosing wavelet coefficients for the actual solution. This entails that the sparsity classes of the residual and the solution coincide. We briefly illustrate below, and fully discuss later in Sect. 5, that this basic premise is false for exponential classes even though it is true for algebraic classes. Confronted with this unexpected fact, we have no alternative but to implement and study ADFOUR with *coarsening* for the exponential case; see Sect. 6 and Sect. 8. This was the original idea of Cohen et al [7] and Binev et al [2] for the algebraic case, but it was subsequently removed by Stevenson [21].

We give now a brief description of the essential issues we are confronted with in designing and studying ADFOUR. To this end, we assume that we know the Fourier representation  $\mathbf{v} = \{v_k\}_{k \in \mathbb{Z}}$  of a periodic function  $v$ , and its non-increasing rearrangement  $\mathbf{v}^* = \{v_n^*\}_{n=1}^\infty$ , namely,  $|v_{n+1}^*| \leq |v_n^*|$  for all  $n \geq 1$ .

**Dörfler marking and best  $N$ -term approximation.** We recall the marking introduced by Dörfler [13], which is the only one for which there exist provable convergence rates. Given a parameter  $\theta \in (0, 1)$ , and a current set of Fourier frequencies or indices  $\Lambda$ , say the first  $N$  ones according to the labeling of  $\mathbf{v}$ , we choose the next set  $\partial\Lambda$  as the *minimal* set for which

$$\|P_{\partial\Lambda}\mathbf{r}\| \geq \theta\|\mathbf{r}\|, \quad (1.1)$$

where  $\mathbf{r} := \mathbf{v} - P_\Lambda\mathbf{v}$  is the *residual* and  $P_\Lambda$  is the orthogonal projection in the  $\ell^2$ -norm  $\|\cdot\|$  onto  $\Lambda$ . Note that, if  $\mathbf{r}_* := \mathbf{r} - P_{\partial\Lambda}\mathbf{r}$  and  $\Lambda_* := \Lambda \cup \partial\Lambda$ , then (1.1) can be equivalently written as

$$\|\mathbf{r}_*\| = \|\mathbf{r} - P_{\partial\Lambda}\mathbf{r}\| \leq \sqrt{1 - \theta^2}\|\mathbf{r}\|, \quad (1.2)$$

and that  $\mathbf{r} = \mathbf{v}|_{\Lambda^c}$  where  $\Lambda^c := \mathbb{N} \setminus \Lambda$  is the complement of  $\Lambda$  and likewise for  $\mathbf{r}_*$ . This is the simplest possible scenario because the information built in  $\mathbf{r}$  is exactly that of  $\mathbf{v}$ . Moreover,  $\mathbf{v} - \mathbf{r} = \{v_n^*\}_{n=1}^N$  is the best  $N$ -term approximation of  $\mathbf{v}$  in the  $\ell^2$ -norm and the corresponding error  $E_N(v)$  is given by

$$E_N(v) = \left( \sum_{n>N} |v_n^*|^2 \right)^{-\frac{1}{2}} = \|\mathbf{r}\|. \quad (1.3)$$

**Algebraic vs exponential decay.** Suppose now that  $\mathbf{v}$  has the precise *algebraic* decay<sup>1</sup>

$$|v_n^*| \simeq n^{-\frac{1}{\tau}} \quad \forall n \geq 1. \quad (1.4)$$

with

$$\frac{1}{\tau} = \frac{s}{d} + \frac{1}{2} \quad (1.5)$$

and  $s > 0$ . We denote by  $\|\mathbf{v}\|_{\ell_B^s}$  the smallest constant in the upper bound in (1.4). We thus have

$$E_N(v)^2 \simeq \|\mathbf{v}\|_{\ell_w^\tau}^2 \sum_{n>N} n^{-\frac{2}{\tau}} = \|\mathbf{v}\|_{\ell_B^s}^2 \sum_{n>N} n^{-\frac{2s}{d}-1} \simeq \|\mathbf{v}\|_{\ell_B^s}^2 N^{-\frac{2s}{d}}.$$

This decay is related to certain *Besov* regularity of  $v$  [12]. Note that the effect of Dörfler marking (1.2) is to reduce the residual from  $\mathbf{r}$  to  $\mathbf{r}_*$  by a factor  $\alpha = \sqrt{1 - \theta^2}$ , or equivalently

$$E_{N_*}(v) \leq \alpha E_N(v),$$

with  $N_* = |\Lambda_*|$ . Since the set  $\Lambda_*$  is minimal, we deduce that  $E_{N_*-1}(v) > \alpha E_N(v)$ , whence

$$\frac{N_*}{N} \simeq \alpha^{-\frac{d}{s}} \quad \Rightarrow \quad N_* - N \simeq \alpha^{-\frac{d}{s}} N \quad (1.6)$$

for  $\alpha$  small enough. This means that the number of degrees of freedom to be added is proportional to the current number. This simplifies considerably the complexity analysis since every step adds as many degrees of freedom as we have already accumulated.

The exponential case is quite different. Suppose that  $\mathbf{v}$  has a *genuinely exponential* decay

$$|v_n^*| \simeq e^{-\eta n} \quad \forall n \geq 1, \quad (1.7)$$

corresponding to analytic functions [14], and let  $\|\mathbf{v}\|_{\ell_G^\eta}$  be the smallest constant appearing in the upper bound in (1.7). These definitions are slight simplifications of the actual ones in Sect. 4.3 but enough to give insight on the main issues at stake. We thus have

$$E_N(v)^2 \simeq \|\mathbf{v}\|_{\ell_G^\eta}^2 \sum_{n>N} e^{-2\eta n} \simeq \|\mathbf{v}\|_{\ell_G^\eta}^2 e^{-2\eta N};$$

this and similar decays are related to *Gevrey* classes of  $C^\infty$  functions [14]. In contrast to (1.6), Dörfler marking now yields<sup>2</sup>

$$N_* - N \sim \frac{1}{\eta} \log \frac{1}{\alpha}. \quad (1.8)$$

This shows that the number of additional degrees of freedom per step is fixed and independent of  $N$ , which makes their counting as well as their implementation a very delicate operation.

**Plateaux.** We now consider a situation opposite to the ideal decay examined above. Suppose that the first  $K > 1$  Fourier coefficients of  $v$  are constant and either

$$|v_n^*| = \|\mathbf{v}\|_{\ell_B^s} n^{-\frac{1}{\tau}} \quad \text{or} \quad |v_n^*| = \|\mathbf{v}\|_{\ell_G^\eta} e^{-\eta n} \quad \forall n \geq K, \quad (1.9)$$

---

<sup>1</sup>Throughout the paper,  $A \lesssim B$  means  $A \leq cB$  for some constant  $c > 0$  independent of the relevant parameters in the inequality;  $A \simeq B$  means  $B \lesssim A \lesssim B$ .

<sup>2</sup>Throughout the paper,  $A \sim B$  means  $A = B + c$  for some quantity  $c \simeq 1$ .

for each approximation class. A simple calculation reveals that either

$$\|\mathbf{v}\| \simeq \|\mathbf{v}\|_{\ell_B^s} K^{-s/d} \quad \text{or} \quad \|\mathbf{v}\| \simeq \|\mathbf{v}\|_{\ell_G^\eta} e^{-\eta K}. \quad (1.10)$$

Repeating the argument leading to (1.6) and (1.8) with  $N = 1$ , we infer that either

$$N_* \simeq K \alpha^{-\frac{d}{s}} \quad \text{or} \quad N_* \sim K + \frac{1}{\eta} \log \frac{1}{\alpha}. \quad (1.11)$$

For  $K \gg 1$  this is a much larger number than the optimal values (1.6) and (1.8), and illustrates the fact that the Dörfler condition (1.1) adds many more frequencies in the presence of plateaux. We note that  $K$  is a multiplicative constant in the left of (1.11) and additive in the right of (1.11).

**Sparsity of the residual.** In practice we do not have access to the Fourier decomposition of  $v$  but rather of the residual  $r(v) = f - Lv$ , where  $f$  is the forcing function and  $L$  the differential operator. Only an operator  $L$  with constant coefficients leads to a spectral representation with diagonal matrix  $\mathbf{A}$ , in which case the components of the residual  $\mathbf{r} = \mathbf{f} - \mathbf{A}\mathbf{v}$  are directly those of  $\mathbf{f}$  and  $\mathbf{v}$ . In general  $\mathbf{A}$  decays away from the main diagonal with a law that depends on the regularity of the coefficients of  $L$ ; we will examine in Sect. 2.4 either algebraic or exponential decay. In this much more intricate and interesting endeavor, studied in this paper, the components of  $\mathbf{v}$  interact with entries of  $\mathbf{A}$  to give rise to  $\mathbf{r}$ . The question whether  $Lv$  belongs to the same approximation class of  $v$  thus becomes relevant because adaptivity decisions are made with  $r(v)$ , and thereby on the range of  $L$  rather than its domain.

We now provide insight on the key issues at stake via a couple of heuristic examples; we discuss this fully in Sect. 5.1 and Sect. 5.2. We start with the exponential case: let  $\mathbf{v} := \{v_k\}_{k \in \mathbb{Z}}$  be defined by

$$v_k = e^{-\eta n} \quad \text{if} \quad k = 2p(n-1), \quad v_k = 0 \quad \text{otherwise,}$$

for  $p \geq 2$  a given integer and  $n \geq 1$ . This sequence exhibits gaps of size  $2p$  between consecutive nonzero entries for  $k \geq 0$ . Its non-decreasing rearrangement  $\mathbf{v}^* = \{v_n^*\}_{n=1}^\infty$  is thus given by

$$v_n^* = e^{-\eta n} \quad n \geq 1,$$

whence  $\mathbf{v} \in \ell_G^\eta$  with  $\|\mathbf{v}\|_{\ell_G^\eta} = 1$ . Let  $\mathbf{A} := (a_{ij})_{i,j=1}^\infty$  be the Toeplitz bi-infinite matrix given by

$$a_{ij} = 1 \quad \text{if} \quad |i - j| \leq q, \quad a_{ij} = 0 \quad \text{otherwise,}$$

with  $1 \leq q < p$ . This matrix  $\mathbf{A}$  has  $2q + 1$  main nontrivial diagonals and is both of exponential and algebraic class according to the Definition 2.1 below. The product  $\mathbf{A}\mathbf{v}$  is much less sparse than  $\mathbf{v}$  but, because  $q < p$ , consecutive frequencies of  $\mathbf{v}$  do not interact with each other: the  $i$ -th component reads

$$(\mathbf{A}\mathbf{v})_i = e^{-\eta n} \quad \text{if} \quad |i - 2p(n-1)| \leq q \quad \text{for some} \quad n \geq 1,$$

or  $(\mathbf{A}\mathbf{v})_i = 0$  otherwise. The non-decreasing rearrangement  $(\mathbf{A}\mathbf{v})^*$  of  $\mathbf{A}\mathbf{v}$  becomes

$$(\mathbf{A}\mathbf{v})_m^* = e^{-\eta n} \quad \text{if} \quad (2q+1)(n-1) + 1 \leq m \leq (2q+1)n.$$

Consequently, writing  $(\mathbf{A}\mathbf{v})_m^* = e^{-\eta \frac{n}{m} m}$  and observing that

$$\frac{n}{m} \geq \frac{n}{(2q+1)n} = \frac{1}{2q+1}$$

and the equality is attained for  $m = (2q + 1)n$ , we deduce

$$\mathbf{A}\mathbf{v} \in \ell_G^{\bar{\eta}} \quad \text{with} \quad \|\mathbf{A}\mathbf{v}\|_{\ell_G^{\bar{\eta}}} = 1 \quad \bar{\eta} = \frac{\eta}{2q+1}.$$

We thus conclude that the action of  $\mathbf{A}$  may shift the exponential class, from the one characterized by the parameter  $\eta$  for  $\mathbf{v}$  to the one characterized by  $\bar{\eta} < \eta$  for  $\mathbf{A}\mathbf{v}$ . This uncovers the crucial feature that the image  $\mathbf{A}\mathbf{v}$  of  $\mathbf{v}$  may be substantially less sparse than  $\mathbf{v}$  itself. In Sect. 5.2 we present a rigorous construction with  $a_{ij}$  decreasing exponentially from the main diagonal and another, rather sophisticated, construction that illustrates the fact that the exponent  $\tau = 1$  in the bound  $|v_n^*| \lesssim e^{-\eta n} = e^{-\eta m^\tau}$  for  $\mathbf{v}$  may deteriorate to some  $\bar{\tau} < 1$  in the corresponding bound for  $\mathbf{A}\mathbf{v}$ .

It is remarkable that a similar construction for the algebraic decay would not lead to a change of algebraic class. In fact, let  $\mathbf{v} = \{v_k\}_{k \in \mathbb{Z}}$  be given by

$$v_k = \frac{1}{n} \quad \text{if} \quad k = 2p(n-1) \quad \text{for some} \quad n \geq 1,$$

and  $v_k = 0$  otherwise. The non-decreasing rearrangement  $\mathbf{v}^* = \{v_n^*\}_{n=1}^\infty$  of  $\mathbf{v}$  satisfies  $v_n^* = \frac{1}{n}$  whence

$$\mathbf{v} \in \ell_B^s \quad \text{with} \quad s = \frac{d}{2} \quad \|\mathbf{v}\|_{\ell_B^s} = 1.$$

On the other hand, the  $i$ -th component of  $\mathbf{A}\mathbf{v}$  reads

$$(\mathbf{A}\mathbf{v})_i = \frac{1}{n} \quad \text{if} \quad |i - 2p(n-1)| \leq q \quad \text{for some} \quad n \geq 1,$$

or  $(\mathbf{A}\mathbf{v})_i = 0$  otherwise. The non-decreasing rearrangement of  $(\mathbf{A}\mathbf{v})^*$  in turn satisfies

$$(\mathbf{A}\mathbf{v})_m^* = \frac{1}{n} \quad \text{if} \quad (2q+1)(n-1) + 1 \leq m \leq (2q+1)n,$$

whence writing  $(\mathbf{A}\mathbf{v})_m^* = \frac{m}{n} \frac{1}{m}$  and arguing as before we infer that

$$\mathbf{A}\mathbf{v} \in \ell_B^s \quad \text{with} \quad \|\mathbf{A}\mathbf{v}\|_{\ell_B^s} = 2q+1.$$

Since  $\|\mathbf{A}\mathbf{v}\|_{\ell_B^s} > \|\mathbf{v}\|_{\ell_B^s}$  we realize that  $\mathbf{A}\mathbf{v}$  is less sparse than  $\mathbf{v}$  but, in contrast to the exponential case, they belong to the same algebraic class  $\ell_B^s$ . Moreover, we will prove later in Sect. 5.1 that  $\mathbf{A}$  preserves the class  $\ell_B^s$  provided entries of  $\mathbf{A}$  possess a suitable algebraic decay away from the main diagonal.

Since Dörfler marking is applied to the residual  $\mathbf{r}$ , it is its sparsity class that determines the degrees of freedom  $|\partial\Lambda|$  to be added. The same argument leading to either (1.6) or (1.8) gives

$$|\partial\Lambda| \leq \left( \frac{\|\mathbf{r}\|_{\ell_B^s}}{\alpha\|\mathbf{r}\|} \right)^{\frac{d}{s}} + 1 \quad \text{or} \quad |\partial\Lambda| \leq \frac{1}{\eta} \log \frac{\|\mathbf{r}\|_{\ell_G^\eta}}{\alpha\|\mathbf{r}\|} + 1,$$

for each class. We thus see that the ratios  $\|\mathbf{r}\|_{\ell_B^s}/\|\mathbf{r}\|$  and  $\|\mathbf{r}\|_{\ell_G^\eta}/\|\mathbf{r}\|$  control the behavior of the adaptive procedure. This has already been observed and exploited by Cohen et al [7] in the context of wavelet methods for the class  $\ell_B^s$ . Our estimates, discussed in Sect. 5, are valid for both classes and use specific decay properties of the entries of  $\mathbf{A}$ .

**Coarsening.** Ever since its inception by Cohen et al [7] and Binev et al [2], this has been a controvertial issue for elliptic PDE. It was originally due to the lack of control on the ratio

$\|\mathbf{r}\|_{\ell_B^s}/\|\mathbf{r}\|$  for large  $s$  [7]. It was removed by Stevenson et al [16, 21] for the algebraic class  $\ell_B^s$  via a clever argument that exploits the minimality of Dörfler marking. This implicitly implies that the approximation classes for both  $v$  and  $Lv$  coincide, which we prove explicitly in Sect. 5.1 for the algebraic case. This is not true though for the exponential case and is discussed in Sect. 5.2. For the latter, we need to resort to *coarsening* to keep the cardinality of ADFOUR quasi-optimal. To this end, we construct an insightful example in Sect. 6 and prove a rather simple but sharp coarsening estimate which improves upon [7].

**Contraction constant.** It is well known that the contraction constant  $\rho(\theta) = \sqrt{1 - \frac{\alpha_*}{\alpha^*}\theta^2}$  cannot be arbitrarily close to 1 for estimators whose upper and lower constants,  $\alpha^* \geq \alpha_*$ , do not coincide. This is, however, at odds with the philosophy of spectral methods which are expected to converge superlinearly (typically exponentially). Assuming that the decay properties of  $\mathbf{A}$  are known, we can enrich Dörfler marking in such a way that the contraction factor becomes

$$\bar{\rho}(\theta) = \left(\frac{\alpha^*}{\alpha_*}\right)^{\frac{1}{2}} \sqrt{1 - \theta^2}.$$

This leads to  $\bar{\rho}(\theta)$  as close to 1 as desired and to *aggressive* versions of ADFOUR discussed in Sect. 3.

This paper can be viewed as a first step towards understanding adaptivity for the *hp*-FEM. However, the results we present are of intrinsic interest and of value for periodic problems with high degree of regularity and rather complex structure. One such problem is turbulence in a periodic box. Our techniques exploit periodicity and orthogonality of the complex exponentials, but many of our assertions and conclusions extend to the non-periodic case for which the natural basis functions are Legendre polynomials; this is the case of the *p*-FEM. In any event, the study of adaptive Fourier-Galerkin methods seems to be a new paradigm in adaptivity, with many intriguing questions and surprises, some discussed in this paper. In contrast to the *h*-FEM, they exhibit unlimited approximation power which is only restricted by solution and data regularity.

We organize the paper as follows. In Sect. 2 we introduce the Fourier-Galerkin method, present a posteriori error estimators, and discuss properties of the underlying matrix  $\mathbf{A}$  for both algebraic and exponential approximation classes. In Sect. 3 we deal with four algorithms, two for each class, and prove their contraction properties. We devote Sect. 4 to nonlinear approximation theory with an emphasis on the exponential class. In Sect. 5 we turn to the study of the sparsity classes for the residual  $\mathbf{r}$  along the lines outlined above. We examine the role of coarsening and prove a sharp coarsening estimate in Sect. 6. We conclude with optimality properties of ADFOUR for the algebraic class in Sect. 7 and for the exponential class in Sect. 8.

## 2 Fourier-Galerkin approximation

### 2.1 Fourier basis and norm representation

For  $d \geq 1$ , we consider  $\Omega = (0, 2\pi)^d$ , and the trigonometric basis

$$\phi_k(x) = \frac{1}{(2\pi)^{d/2}} e^{ik \cdot x}, \quad k \in \mathbb{Z}^d, \quad x \in \mathbb{R}^d,$$

which is orthonormal in  $L^2(\Omega)$ ; let

$$v = \sum_k \hat{v}_k \phi_k, \quad \hat{v}_k = (v, \phi_k), \quad \text{with } \|v\|_{L^2(\Omega)}^2 = \sum_k |\hat{v}_k|^2,$$

be the expansion of any  $v \in L^2(\Omega)$  and the representation of its norm via the Parseval identity. Let  $H_p^1(\Omega) = \{v \in H^1(\Omega) : v(x + 2\pi e_j) = v(x) \ 1 \leq j \leq d\}$ , and let  $H_p^{-1}(\Omega)$  be its dual. Since the trigonometric basis is orthogonal in  $H_p^1(\Omega)$  as well, one has for any  $v \in H_p^1(\Omega)$

$$\|v\|_{H_p^1(\Omega)}^2 = \sum_k (1 + |k|^2) |\hat{v}_k|^2 = \sum_k |\hat{V}_k|^2, \quad (\text{setting } \hat{V}_k := \sqrt{(1 + |k|^2)} \hat{v}_k); \quad (2.1)$$

here and in the sequel,  $|k|$  denotes the Euclidean norm of the multi-index  $k$ . On the other hand, if  $f \in H_p^{-1}(\Omega)$ , we set

$$\hat{f}_k = \langle f, \phi_k \rangle, \quad \text{so that } \langle f, v \rangle = \sum_k \hat{f}_k \hat{v}_k \quad \forall v \in H_p^1(\Omega);$$

the norm representation is

$$\|f\|_{H_p^{-1}(\Omega)}^2 = \sum_k \frac{1}{(1 + |k|^2)} |\hat{f}_k|^2 = \sum_k |\hat{F}_k|^2, \quad (\text{setting } \hat{F}_k := \frac{1}{\sqrt{(1 + |k|^2)}} \hat{f}_k). \quad (2.2)$$

Throughout the paper, we will use the notation  $\|\cdot\|$  to indicate both the  $H_p^1(\Omega)$ -norm of a function  $v$ , or the  $H_p^{-1}(\Omega)$ -norm of a linear form  $f$ ; the specific meaning will be clear from the context.

Given any finite index set  $\Lambda \subset \mathbb{Z}^d$ , we define the subspace of  $V := H_p^1(\Omega)$

$$V_\Lambda := \text{span} \{\phi_k \mid k \in \Lambda\};$$

we set  $|\Lambda| = \text{card } \Lambda$ , so that  $\dim V_\Lambda = |\Lambda|$ . If  $g$  admits an expansion  $g = \sum_k \hat{g}_k \phi_k$  (converging in an appropriate norm), then we define its projection  $P_\Lambda g$  upon  $V_\Lambda$  by setting

$$P_\Lambda g = \sum_{k \in \Lambda} \hat{g}_k \phi_k.$$

## 2.2 Galerkin discretization and residual

We now consider the elliptic problem

$$\begin{cases} Lu = -\nabla \cdot (\nu \nabla u) + \sigma u = f & \text{in } \Omega, \\ u \text{ } 2\pi\text{-periodic in each direction,} \end{cases} \quad (2.3)$$

where  $\nu$  and  $\sigma$  are sufficiently smooth real coefficients satisfying  $0 < \nu_* \leq \nu(x) \leq \nu^* < \infty$  and  $0 < \sigma_* \leq \sigma(x) \leq \sigma^* < \infty$  in  $\Omega$ ; let us set

$$\alpha_* = \min(\nu_*, \sigma_*) \quad \text{and} \quad \alpha^* = \max(\nu^*, \sigma^*).$$

We formulate this problem variationally as

$$u \in H_p^1(\Omega) \quad : \quad a(u, v) = \langle f, v \rangle \quad \forall v \in H_p^1(\Omega), \quad (2.4)$$

where  $a(u, v) = \int_\Omega \nu \nabla u \cdot \nabla \bar{v} + \int_\Omega \sigma u \bar{v}$  (bar indicating as usual complex conjugate). We denote by  $\|v\| = \sqrt{a(v, v)}$  the energy norm of any  $v \in H_p^1(\Omega)$ , which satisfies

$$\sqrt{\alpha_*} \|v\| \leq \|v\| \leq \sqrt{\alpha^*} \|v\|. \quad (2.5)$$



Given any finite set  $\Lambda \subset \mathbb{Z}^d$ , the Galerkin approximation is defined as

$$u_\Lambda \in V_\Lambda \quad : \quad a(u_\Lambda, v_\Lambda) = \langle f, v_\Lambda \rangle \quad \forall v_\Lambda \in V_\Lambda . \quad (2.6)$$

For any  $w \in V_\Lambda$ , we define the residual

$$r(w) = f - Lw = \sum_k \hat{r}_k(w) \phi_k , \quad \text{where} \quad \hat{r}_k(w) = \langle f - Lw, \phi_k \rangle = \langle f, \phi_k \rangle - a(w, \phi_k) .$$

Then, the previous definition of  $u_\Lambda$  is equivalent to the condition

$$P_\Lambda r(u_\Lambda) = 0 , \quad \text{i.e.,} \quad \hat{r}_k(u_\Lambda) = 0 \quad \forall k \in \Lambda . \quad (2.7)$$

On the other hand, by the continuity and coercivity of the bilinear form  $a$ , one has

$$\frac{1}{\alpha^*} \|r(u_\Lambda)\| \leq \|u - u_\Lambda\| \leq \frac{1}{\alpha_*} \|r(u_\Lambda)\| , \quad (2.8)$$

or, equivalently,

$$\frac{1}{\sqrt{\alpha^*}} \|r(u_\Lambda)\| \leq \|u - u_\Lambda\| \leq \frac{1}{\sqrt{\alpha_*}} \|r(u_\Lambda)\| . \quad (2.9)$$

### 2.3 Algebraic representations

Let us identify the solution  $u = \sum_k \hat{u}_k \phi_k$  of Problem (2.4) with the vector  $\mathbf{u} = (\hat{U}_k) = (c_k \hat{u}_k) \in \mathbb{C}^{\mathbb{Z}^d}$  of its  $H_p^1$ -normalized Fourier coefficients, where we set for convenience  $c_k = \sqrt{1 + |k|^2}$ . Similarly, let us identify the right-hand side  $f$  with the vector  $\mathbf{f} = (\hat{F}_\ell) = (c_\ell^{-1} \hat{f}_\ell) \in \mathbb{C}^{\mathbb{Z}^d}$  of its  $H_p^{-1}$ -normalized Fourier coefficients. Finally, let us introduce the bi-infinite, Hermitian and positive-definite matrix

$$\mathbf{A} = (a_{\ell,k}) \quad \text{with} \quad a_{\ell,k} = \frac{1}{c_\ell c_k} a(\phi_k, \phi_\ell) . \quad (2.10)$$

Then, Problem (2.4) can be equivalently written as

$$\mathbf{A} \mathbf{u} = \mathbf{f} . \quad (2.11)$$

We observe that the orthogonality properties of the trigonometric basis implies that the matrix  $\mathbf{A}$  is diagonal if and only if the coefficients  $\nu$  and  $\sigma$  are constant in  $\Omega$ .

Next, consider the Galerkin problem (2.6) and let  $\mathbf{u}_\Lambda \in \mathbb{C}^{|\Lambda|}$  be the vector collecting the coefficients of  $u_\Lambda$  indexed in  $\Lambda$ ; let  $\mathbf{f}_\Lambda \in \mathbb{C}^{|\Lambda|}$  be the analogous restriction for the vector of the coefficients of  $f$ . Finally, denote by  $\mathbf{R}_\Lambda$  the matrix that restricts a bi-infinite vector to the portion indexed in  $\Lambda$ , so that  $\mathbf{E}_\Lambda = \mathbf{R}_\Lambda^H$  is the corresponding extension matrix. Then, setting

$$\mathbf{A}_\Lambda = \mathbf{R}_\Lambda \mathbf{A} \mathbf{R}_\Lambda^H , \quad (2.12)$$

Problem (2.6) can be equivalently written as

$$\mathbf{A}_\Lambda \mathbf{u}_\Lambda = \mathbf{f}_\Lambda . \quad (2.13)$$

## 2.4 Properties of the stiffness matrix

It is useful to express the elements of  $\mathbf{A}$  in terms of the Fourier coefficients of the operator coefficients  $\nu$  and  $\sigma$ . Precisely, writing  $\nu = \sum_k \hat{\nu}_k \phi_k$  and  $\sigma = \sum_k \hat{\sigma}_k \phi_k$  and using the orthogonality of the Fourier basis, one easily gets

$$a_{\ell,k} = \frac{1}{(2\pi)^{d/2}} \left( \frac{\ell \cdot k}{c_\ell c_k} \hat{\nu}_{\ell-k} + \frac{1}{c_\ell c_k} \hat{\sigma}_{\ell-k} \right). \quad (2.14)$$

Note that the diagonal elements are uniformly bounded from below,

$$a_{\ell,\ell} \geq \frac{1}{(2\pi)^{d/2}} \min(\hat{\nu}_0, \hat{\sigma}_0) > 0, \quad \ell \in \mathbb{Z}^d, \quad (2.15)$$

whereas all elements are bounded in modulus by the elements of a *Toeplitz* matrix,

$$|a_{\ell,k}| \leq \frac{1}{(2\pi)^{d/2}} (|\hat{\nu}_{\ell-k}| + |\hat{\sigma}_{\ell-k}|), \quad \ell, k \in \mathbb{Z}^d, \quad (2.16)$$

which decay as  $|\ell - k| \rightarrow \infty$  at a rate dictated by the smoothness of the operator coefficients. Indeed, if  $\nu$  and  $\sigma$  are sufficiently smooth, their Fourier coefficients decay at a suitable rate and this property is inherited by the off-diagonal elements of the matrix  $\mathbf{A}$ , via (2.16). To be precise, if the coefficients  $\nu$  and  $\sigma$  have a finite order of regularity, then the rate of decay of their Fourier coefficients is algebraic, i.e.

$$|\hat{\nu}_k|, |\hat{\sigma}_k| \lesssim (1 + |k|)^{-\eta} \quad \forall k \in \mathbb{Z}^d, \quad (2.17)$$

for some  $\eta > 0$ . On the other hand, if the operator coefficients are real analytic in a neighborhood of  $\Omega$ , then the rate of decay of their Fourier coefficients is exponential, i.e.

$$|\hat{\nu}_k|, |\hat{\sigma}_k| \lesssim e^{-\eta|k|} \quad \forall k \in \mathbb{Z}^d. \quad (2.18)$$

Correspondingly, the matrix  $\mathbf{A}$  belongs to one of the following classes.

**Definition 2.1 (regularity classes for  $\mathbf{A}$ )** *A matrix  $\mathbf{A}$  is said to belong to*

- *the algebraic class  $\mathcal{D}_a(\eta_L)$  if there exists a constant  $c_L > 0$  such that its elements satisfy*

$$|a_{\ell,k}| \leq c_L (1 + |\ell - k|)^{-\eta_L} \quad \ell, k \in \mathbb{Z}^d; \quad (2.19)$$

- *the exponential class  $\mathcal{D}_e(\eta_L)$  if there exists a constant  $c_L > 0$  such that its elements satisfy*

$$|a_{\ell,k}| \leq c_L e^{-\eta_L |\ell - k|} \quad \ell, k \in \mathbb{Z}^d. \quad (2.20)$$

The following properties hold.

**Property 2.1 (continuity of  $\mathbf{A}$ )** *If either  $\mathbf{A} \in \mathcal{D}_a(\eta_L)$ , with  $\eta_L > d$ , or  $\mathbf{A} \in \mathcal{D}_e(\eta_L)$ , then  $\mathbf{A}$  defines a bounded operator on  $\ell^2(\mathbb{Z}^d)$ .*

*Proof.* See e.g. [17, 9]. □

**Property 2.2 (inverse of  $\mathbf{A}$ : algebraic case)** If  $\mathbf{A} \in \mathcal{D}_a(\eta_L)$ , with  $\eta_L > d$  and  $\mathbf{A}$  is invertible in  $\ell^2(\mathbb{Z}^d)$ , then  $\mathbf{A}^{-1} \in \mathcal{D}_a(\eta_L)$ .

*Proof.* See e.g. [17]. □

**Property 2.3 (inverse of  $\mathbf{A}$ : exponential case)** If  $\mathbf{A} \in \mathcal{D}_e(\eta_L)$  and there exists a constant  $c_L$  satisfying (2.20) such that

$$c_L < \frac{1}{2}(e^{\eta_L} - 1) \min_{\ell} a_{\ell, \ell}, \quad (2.21)$$

then  $\mathbf{A}$  is invertible in  $\ell^2(\mathbb{Z}^d)$  and  $\mathbf{A}^{-1} \in \mathcal{D}_e(\bar{\eta}_L)$  where  $\bar{\eta}_L \in (0, \eta_L]$  is such that  $\bar{z} = e^{-\bar{\eta}_L}$  is the unique zero in the interval  $(0, 1)$  of the polynomial

$$z^2 - \frac{e^{2\eta_L} + 2c_L + 1}{e^{\eta_L}(c_L + 1)}z + 1.$$

*Proof.* We follow the suggestion by Bini [3], and thus exploit the one-to-one correspondence between Toeplitz matrices and formal Laurent series (see e.g. [4]):

$$f(z) = \sum_{k=-\infty}^{\infty} a_k z^k \longleftrightarrow \mathbf{T}_f = (t_{i,j}), \quad t_{i,j} = a_{i-j}.$$

We refer to the function  $f(z)$  as to the symbol associated to the Toeplitz matrix  $\mathbf{T}_f$ . We recall now a few relations between  $f(z)$  and  $\mathbf{T}_f$ . If  $f(z)$  is analytic on  $\mathcal{A}_\alpha = \{z \in \mathbb{C} : e^{-\alpha} < |z| < e^\alpha\}$  with  $\alpha > 0$ , then there holds  $f(z) = \sum_{k=-\infty}^{+\infty} a_k z^k$ , where the coefficients  $a_k$  have exponential decay with rate  $e^{-\alpha}$  in the sense that for every  $0 < \rho < e^{-\alpha}$  there exists a constant  $\gamma > 0$  such that  $|a_k| \leq \gamma \rho^{|k|}$ . As a consequence, the symbol  $f(z)$  of the Toeplitz matrix  $\mathbf{T}_f$  is analytic on  $\mathcal{A}_\alpha$  for some  $\alpha > 0$  if and only if the elements of  $\mathbf{T}_f$  decay exponentially with rate  $e^{-\alpha}$ . Moreover, it is known that if  $f(z)$  is analytic on  $\mathcal{A}_\alpha$  and it is non-zero on  $\mathcal{A}_\beta \subset \mathcal{A}_\alpha$ , then the function  $g(z) = 1/f(z)$  is well defined and analytic on  $\mathcal{A}_\beta$ , the matrix  $\mathbf{T}_g$  is the inverse of  $\mathbf{T}_f$  and the elements of  $\mathbf{T}_g$  decay exponentially with rate  $e^{-\beta}$ .

We next introduce the analytic functions in  $\mathcal{A}_\alpha$

$$h(z) = \sum_{k=1}^{\infty} e^{-\alpha k} (z^k + z^{-k}) = \frac{z}{e^\alpha - z} + \frac{z^{-1}}{e^\alpha - z^{-1}}, \quad f_c(z) = 1 - ch(z),$$

with  $c > 0$ . For  $|z| = 1$  we deduce  $|h(z)| \leq 2 \sum_{k=1}^{\infty} e^{-\alpha k} = 2/(e^\alpha - 1)$ , whence  $c|h(z)| < 1$  provided that  $c < \frac{1}{2}(e^\alpha - 1)$ ; moreover  $\|\mathbf{T}_h\| \leq \|\mathbf{T}_h\|_\infty = 2/(e^\alpha - 1)$ , which is indeed a particular instance of Schur Lemma for symmetric matrices. For this range of  $c$ 's,  $f_c(z) \neq 0$  for  $|z| = 1$  and for continuity there exists  $\mathcal{A}_\beta \subset \mathcal{A}_\alpha$  on which  $f_c(z)$  is non-zero. This implies that  $g_c(z) := 1/f_c(z)$  is analytic on  $\mathcal{A}_\beta$  and the elements of the associated Toeplitz matrix  $\mathbf{T}_{g_c}$  decay exponentially with rate  $e^{-\beta}$ . The singularities of  $g_c$  correspond to zeros of  $f_c$ , which are in turn the roots  $\zeta_1, \zeta_2$  of the polynomial

$$z^2 - \frac{e^{2\alpha} + 2c + 1}{e^\alpha(c + 1)}z + 1.$$

These roots are real provided  $c < \frac{1}{2}(e^\alpha - 1)$ , in which case  $e^{-\beta} = \zeta_1 = \zeta_2^{-1} < 1$ .

Let  $\mathbf{A} \in \mathcal{D}_e(\alpha)$ , i.e. there exists a constant  $c$  such that  $|a_{\ell,k}| \leq ce^{-\alpha|\ell-k|}$  for  $\ell, k \in \mathbb{Z}^d$ . By rescaling of the rows of  $\mathbf{A}$ , it is not restrictive to assume that the diagonal elements  $\mathbf{A}$  are equal to 1. Then, it is possible to write  $\mathbf{A} = \mathbf{I} - \mathbf{S}$  with  $|\mathbf{S}| \leq c\mathbf{T}_h$ , the inequality being meant element by element, and  $\|\mathbf{S}\| < 1$ . Since  $g_c(z) = 1/(1 - ch(z)) = \sum_{k=0}^{\infty} c^k h(z)^k$  is well defined and analytic on  $\mathcal{A}_\beta \subset \mathcal{A}_\alpha$ , it follows that

$$\left| \sum_{k=0}^{\infty} \mathbf{S}^k \right| \leq \sum_{k=0}^{\infty} |\mathbf{S}|^k \leq \sum_{k=0}^{\infty} c^k \mathbf{T}_h^k = \mathbf{T}_{g_c}.$$

Hence, the elements of the matrix  $\mathbf{T}_{g_c}$  decay exponentially with rate  $e^{-\beta}$ . Property  $\|\mathbf{S}\| < 1$  yields  $\mathbf{A}^{-1} = (\mathbf{I} - \mathbf{S})^{-1} = \sum_{k=0}^{\infty} \mathbf{S}^k$  and  $|\mathbf{A}^{-1}| \leq \mathbf{T}_{g_c}$ , whence the coefficients of  $\mathbf{A}^{-1}$  being bounded by those of  $\mathbf{T}_{g_c}$  decay exponentially with rate  $e^{-\beta}$ , i.e.  $\mathbf{A}^{-1} \in \mathcal{D}_e(\beta)$  for some  $\beta < \alpha$ . This gives (2.21) once the row scaling of  $\mathbf{A}$  is taken into account.  $\square$

**Example 2.1 (sharpness of (2.21))** The following example illustrates that (2.21) is sharp. Let  $\mathbf{A}$  be

$$a_{ij} = -2^{-1-|i-j|} \quad i \neq j, \quad a_{ii} = 1,$$

which is singular because the sum of the coefficients in every row vanishes. This  $\mathbf{A}$  corresponds to  $e^{\eta_L} = 2$ ,  $c_L = \frac{1}{2}$  and  $\frac{1}{2}(e^{\eta_L} - 1) = \frac{1}{2}$ , which violates (2.21).

For any integer  $J \geq 0$ , let  $\mathbf{A}_J$  denote the following symmetric truncation of the matrix  $\mathbf{A}$

$$(\mathbf{A}_J)_{\ell,k} = \begin{cases} a_{\ell,k} & \text{if } |\ell - k| \leq J, \\ 0 & \text{elsewhere.} \end{cases} \quad (2.22)$$

Then, we have the following well-known results, whose proof is reported for completeness.

**Property 2.4 (truncation)** *The truncated matrix  $\mathbf{A}_J$  has a number of non-vanishing entries bounded by  $\omega_d J^d$ , where  $\omega_d$  is the measure of the Euclidean unit ball in  $\mathbb{R}^d$ . Moreover, under the assumption of Property 2.1, there exists a constant  $C_{\mathbf{A}}$  such that*

$$\|\mathbf{A} - \mathbf{A}_J\| \leq \psi_{\mathbf{A}}(J, \eta) := C_{\mathbf{A}} \begin{cases} (J+1)^{-(\eta_L-d)} & \text{if } \mathbf{A} \in \mathcal{D}_a(\eta_L) \text{ (algebraic case)}, \\ (J+1)^{d-1} e^{-\eta_L J} & \text{if } \mathbf{A} \in \mathcal{D}_e(\eta_L) \text{ (exponential case)}, \end{cases}$$

for all  $J \geq 0$ . Consequently, under the assumptions of Property 2.2 or 2.3, one has

$$\|\mathbf{A}^{-1} - (\mathbf{A}^{-1})_J\| \leq \psi_{\mathbf{A}^{-1}}(J, \bar{\eta}_L) \quad (2.23)$$

where we let  $\bar{\eta}_L = \eta_L$  in the algebraic case and  $\bar{\eta}_L$  be defined in Property 2.3 for the exponential case.

*Proof.* We use the Schur Lemma for symmetric matrices,  $\|\mathbf{B}\| \leq \|\mathbf{B}\|_{\infty} = \sup_{\ell} \sum_k |b_{\ell,k}|$  for  $\mathbf{B} = \mathbf{A} - \mathbf{A}_J$ . Thus, in the algebraic case

$$\begin{aligned} \sup_{\ell} \sum_{k:|\ell-k|>J} |a_{\ell,k}| &\leq C_L \sup_{\ell} \sum_{k:|\ell-k|>J} \frac{1}{(1+|\ell-k|)^{\eta_L}} \\ &\lesssim \sup_{\ell} \sum_{q=J+1}^{\infty} \sum_{k:|\ell-k|=q} \frac{1}{(1+q)^{\eta_L}} \lesssim \sup_{\ell} \sum_{q=J+1}^{\infty} \frac{q^{d-1}}{(1+q)^{\eta_L}} \lesssim (J+1)^{d-\eta_L}. \end{aligned}$$

A similar argument yields the result in the exponential case.  $\square$

## 2.5 An equivalent formulation of the Galerkin problem

For future reference, hereafter we rewrite the Galerkin problem (2.13) in an equivalent (infinite-dimensional) way. Let

$$\mathbf{P}_\Lambda : \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\mathbb{Z}^d)$$

be the projector operator defined as

$$(\mathbf{P}_\Lambda \mathbf{v})_\lambda = \begin{cases} v_\lambda & \text{if } \lambda \in \Lambda, \\ 0 & \text{if } \lambda \notin \Lambda. \end{cases}$$

Note that  $\mathbf{P}_\Lambda$  can be represented as a diagonal bi-infinite matrix whose diagonal elements are 1 for indexes belonging to  $\Lambda$ , zero otherwise. Let us set  $\mathbf{Q}_\Lambda = \mathbf{I} - \mathbf{P}_\Lambda$  and we introduce the bi-infinite matrix  $\hat{\mathbf{A}}_\Lambda := \mathbf{P}_\Lambda \mathbf{A} \mathbf{P}_\Lambda + \mathbf{Q}_\Lambda$  which is equal to  $\mathbf{A}_\Lambda$  for indexes in  $\Lambda$  and to the identity matrix, otherwise. The definitions of the projectors  $\mathbf{P}_\Lambda$  and  $\mathbf{Q}_\Lambda$  yield the following result.

**Property 2.5 (invertibility of  $\hat{\mathbf{A}}$ )** *If  $\mathbf{A}$  is invertible with either  $\mathbf{A} \in \mathcal{D}_a(\eta_L)$  or  $\mathbf{A} \in \mathcal{D}_e(\eta_L)$ , then the same holds for  $\hat{\mathbf{A}}_\Lambda$ .*

Now, let us consider the following extended Galerkin problem: find  $\hat{\mathbf{u}} \in \ell^2(\mathbb{Z}^d)$  such that

$$\hat{\mathbf{A}}_\Lambda \hat{\mathbf{u}} = \mathbf{P}_\Lambda \mathbf{f}. \quad (2.24)$$

Let  $\mathbf{E}_\Lambda : \mathbb{C}^{|\Lambda|} \rightarrow \ell^2(\mathbb{Z}^d)$  be the extension operator defined in Sect. 2.3 and let  $\mathbf{u}_\Lambda \in \mathbb{C}^{|\Lambda|}$  be the Galerkin solution to (2.13); then, it is easy to check that  $\hat{\mathbf{u}} = \mathbf{E}_\Lambda \mathbf{u}_\Lambda$ .

In the following, with an abuse of notation, the solution of (2.24) will be denoted by  $\mathbf{u}_\Lambda$ . We will refer to it as to the (extended) Galerkin solution, meaning the infinite-dimensional representant of the finite-dimensional Galerkin solution. In case of possible confusion, we will make clear which version (infinite-dimensional or finite-dimensional) has to be considered.

## 3 Adaptive algorithms with contraction properties

Our first algorithm will be an *ideal one*; it will serve as a reference to illustrate in the simplest situation the contraction property which guarantees the convergence of the algorithm, and it will be subsequently modified to get more efficient versions. The ideal algorithm uses as error estimator the ideal one, i.e., the norm of the residual in  $H_p^{-1}(\Omega)$ ; we thus set, for any  $v \in H_p^1(\Omega)$ ,

$$\eta^2(v) = \|r(v)\|^2 = \sum_{k \in \mathbb{Z}^d} |\hat{R}_k(v)|^2, \quad (3.1)$$

so that (2.8) can be rephrased as

$$\frac{1}{\alpha^*} \eta(u_\Lambda) \leq \|u - u_\Lambda\| \leq \frac{1}{\alpha_*} \eta(u_\Lambda); \quad (3.2)$$

recall that  $\hat{R}_k(v) = (1 + |k|^2)^{-1/2} r_k(v)$  according to (2.2). Obviously, this estimator is hardly computable in practice; in Sect. 3.2 we will introduce a feasible version, but for the moment we go through the ideal situation. Given any subset  $\Lambda \subseteq \mathbb{Z}^d$ , we also define the quantity

$$\eta^2(v; \Lambda) = \|P_\Lambda r(v)\|^2 = \sum_{k \in \Lambda} |\hat{R}_k(v)|^2,$$

so that  $\eta(v) = \eta(v; \mathbb{Z}^d)$ .

### 3.1 ADFOUR: an ideal algorithm

We now introduce the following procedures, which will enter the definition of all our adaptive algorithms.

- $u_\Lambda := \mathbf{GAL}(\Lambda)$   
Given a finite subset  $\Lambda \subset \mathbb{Z}^d$ , the output  $u_\Lambda \in V_\Lambda$  is the solution of the Galerkin problem (2.6) relative to  $\Lambda$ .
- $r := \mathbf{RES}(v_\Lambda)$   
Given a function  $v_\Lambda \in V_\Lambda$  for some finite index set  $\Lambda$ , the output  $r$  is the residual  $r(v_\Lambda) = f - Lv_\Lambda$ .
- $\Lambda^* := \mathbf{DÖRFLER}(r, \theta)$   
Given  $\theta \in (0, 1)$  and an element  $r \in H_p^{-1}(\Omega)$ , the output  $\Lambda^* \subset \mathbb{Z}^d$  is a finite set such that the inequality

$$\|P_{\Lambda^*} r\| \geq \theta \|r\| \quad (3.3)$$

is satisfied.

Note that the latter inequality is equivalent to

$$\|r - P_{\Lambda^*} r\| \leq \sqrt{1 - \theta^2} \|r\|. \quad (3.4)$$

If  $r = r(u_\Lambda)$  is the residual of a Galerkin solution  $u_\Lambda \in V_\Lambda$ , then by (2.7) we can trivially assume that  $\Lambda^*$  is contained in  $\Lambda^c := \mathbb{Z}^d \setminus \Lambda$ . For such a residual, inequality (3.3) can then be stated as

$$\eta(u_\Lambda; \Lambda^*) \geq \theta \eta(u_\Lambda), \quad (3.5)$$

a condition termed *Dörfler marking* in the finite element literature, or *bulk chasing* in the wavelet literature. Writing  $\hat{R}_k = \hat{R}_k(u_\Lambda)$ , the condition (3.5) can be equivalently stated as

$$\sum_{k \in \Lambda^*} |\hat{R}_k|^2 \geq \theta^2 \sum_{k \notin \Lambda} |\hat{R}_k|^2. \quad (3.6)$$

Also note that a set  $\Lambda^*$  of minimal cardinality can be immediately determined if the coefficients  $\hat{R}_k$  are rearranged in non-increasing order of modulus; however, the subsequent convergence result does not require the property of minimal cardinality for the sets of active coefficients.

In the sequel, we will invariably make the following assumption:

**Assumption 3.1 (Dörfler marking)** *The procedure **DÖRFLER** selects an index set  $\Lambda^*$  of minimal cardinality among all those satisfying condition (3.3).*

Given two parameters  $\theta \in (0, 1)$  and  $tol \in [0, 1]$ , we are ready to define our ideal adaptive algorithm.

**Algorithm ADFOUR**( $\theta, tol$ )

Set  $r_0 := f$ ,  $\Lambda_0 := \emptyset$ ,  $n = -1$

do

$n \leftarrow n + 1$

$$\partial\Lambda_n := \mathbf{DÖRFLER}(r_n, \theta)$$

$$\Lambda_{n+1} := \Lambda_n \cup \partial\Lambda_n$$

$$u_{n+1} := \mathbf{GAL}(\Lambda_{n+1})$$

$$r_{n+1} := \mathbf{RES}(u_{n+1})$$

while  $\|r_{n+1}\| > tol$

The following result states the convergence of this algorithm, with a guaranteed error reduction rate.

**Theorem 3.1 (convergence of ADFOUR)** *Let us set*

$$\rho = \rho(\theta) = \sqrt{1 - \frac{\alpha_*}{\alpha^*} \theta^2} \in (0, 1) . \quad (3.7)$$

*Let  $\{\Lambda_n, u_n\}_{n \geq 0}$  be the sequence generated by the adaptive algorithm **ADFOUR**. Then, the following bound holds for any  $n$ :*

$$\|u - u_{n+1}\| \leq \rho \|u - u_n\| .$$

*Thus, for any  $tol > 0$  the algorithm terminates in a finite number of iterations, whereas for  $tol = 0$  the sequence  $u_n$  converges to  $u$  in  $H_p^1(\Omega)$  as  $n \rightarrow \infty$ .*

*Proof.* For convenience, we use the notation  $e_n := \|u - u_n\|$  and  $d_n := \|u_{n+1} - u_n\|$ . As  $V_{\Lambda_n} \subset V_{\Lambda_{n+1}}$ , the following orthogonality property holds

$$e_{n+1}^2 = e_n^2 - d_n^2. \quad (3.8)$$

On the other hand, for any  $w \in H_p^1(\Omega)$ , one has in light of (2.5)

$$\|Lw\| = \sup_{v \in H_p^1(\Omega)} \frac{\langle Lw, v \rangle}{\|v\|} = \sup_{v \in H_p^1(\Omega)} \frac{a(w, v)}{\|v\|} \leq \|w\| \sup_{v \in H_p^1(\Omega)} \frac{\|v\|}{\|v\|} \leq \sqrt{\alpha^*} \|w\| .$$

Thus, using (3.3),

$$\begin{aligned} d_n^2 &\geq \frac{1}{\alpha^*} \|L(u_{n+1} - u_n)\|^2 = \frac{1}{\alpha^*} \|r_{n+1} - r_n\|^2 \\ &\geq \frac{1}{\alpha^*} \|P_{\Lambda_{n+1}}(r_{n+1} - r_n)\|^2 = \frac{1}{\alpha^*} \|P_{\Lambda_{n+1}} r_n\|^2 \geq \frac{\theta^2}{\alpha^*} \|r_n\|^2 . \end{aligned}$$

On the other hand, the rightmost inequality in (2.9) states that  $\|r_n\|^2 \geq \alpha_* e_n^2$ , whence the result.  $\square$

### 3.2 F-ADFOUR: A feasible version of ADFOUR

The error estimator  $\eta(u_\Lambda)$  based on (3.1) is not computable in practice, since the residual  $r(u_\Lambda)$  contains infinitely many coefficients. We thus introduce a new estimator, defined from an approximation of such residual with finite Fourier expansion (i.e., a trigonometric polynomial).

To this end, let  $\tilde{\nu}$ ,  $\tilde{\sigma}$  and  $\tilde{f}$  be suitable trigonometric polynomials, which approximate  $\nu$ ,  $\sigma$  and  $f$ , respectively, to a given accuracy. Then, the quantity

$$\tilde{r}(u_\Lambda) = \tilde{f} - \tilde{L}u_\Lambda = \tilde{f} + \nabla \cdot (\tilde{\nu} \nabla u_\Lambda) - \tilde{\sigma} u_\Lambda \quad (3.9)$$

belongs to  $V_{\tilde{\Lambda}}$  for some finite subset  $\tilde{\Lambda} \subset \mathbb{Z}^d$ , i.e., it has the finite (thus, computable) expansion

$$\tilde{r}(u_\Lambda) = \sum_{k \in \tilde{\Lambda}} \hat{r}_k(u_\Lambda) \phi_k .$$

The choice of the approximate coefficients has to be done in order to fulfil the following condition: for a fixed parameter  $\gamma \in (0, \theta)$ , we require that

$$\|r(u_\Lambda) - \tilde{r}(u_\Lambda)\| \leq \gamma \|\tilde{r}(u_\Lambda)\| . \quad (3.10)$$

Satisfying such a condition is possible, provided we have full access to the data. Indeed, on the one hand, the left-hand side tends to 0 as the approximation of the coefficients gets better and better, since (we keep here the full norm indication for a better clarity)

$$\begin{aligned} \|r(u_\Lambda) - \tilde{r}(u_\Lambda)\|_{H_p^{-1}(\Omega)} &\leq \|f - \tilde{f}\|_{H_p^{-1}(\Omega)} + \|\nu - \tilde{\nu}\|_{L^\infty(\Omega)} \|\nabla u_\Lambda\|_{L^2(\Omega)^d} + \|\sigma - \tilde{\sigma}\|_{L^\infty(\Omega)} \|u_\Lambda\|_{L^2(\Omega)} \\ &\leq \|f - \tilde{f}\|_{H_p^{-1}(\Omega)} + (\|\nu - \tilde{\nu}\|_{L^\infty(\Omega)} + \|\sigma - \tilde{\sigma}\|_{L^\infty(\Omega)}) \frac{1}{\alpha_*} \|f\|_{H_p^{-1}(\Omega)} , \end{aligned}$$

where we have used the bound on the solution of the Galerkin problem (2.6) in terms of the data. On the other hand, if  $u_\Lambda \neq u$ , then  $r(u_\Lambda) \neq 0$ , whence the right-hand side of (3.10) converges to a non-zero value as  $\tilde{\Lambda}$  increases.

With this remark in mind, we define a new error estimator by setting

$$\tilde{\eta}^2(u_\Lambda) = \|\tilde{r}(u_\Lambda)\|^2 = \sum_{k \in \tilde{\Lambda}} |\hat{R}_k(u_\Lambda)|^2 , \quad (3.11)$$

which, in view of (3.10), immediately yields

$$\frac{1-\gamma}{\alpha^*} \tilde{\eta}(u_\Lambda) \leq \|u - u_\Lambda\| \leq \frac{1+\gamma}{\alpha_*} \tilde{\eta}(u_\Lambda) . \quad (3.12)$$

**Lemma 3.1 (feasible Dörfler marking)** *Let  $\Lambda^*$  be any finite index set such that*

$$\tilde{\eta}(u_\Lambda; \Lambda^*) \geq \theta \tilde{\eta}(u_\Lambda) .$$

*Then,*

$$\eta(u_\Lambda; \Lambda^*) \geq \tilde{\theta} \eta(u_\Lambda) , \quad \text{with} \quad \tilde{\theta} = \frac{\theta - \gamma}{1 + \gamma} \in (0, \theta) . \quad (3.13)$$

*Proof.* One has

$$\begin{aligned} \|P_{\Lambda^*} r(u_\Lambda)\| &\geq \|P_{\Lambda^*} \tilde{r}(u_\Lambda)\| - \|P_{\Lambda^*} (r(u_\Lambda) - \tilde{r}(u_\Lambda))\| \\ &\geq \theta \|\tilde{r}(u_\Lambda)\| - \|r(u_\Lambda) - \tilde{r}(u_\Lambda)\| \\ &\geq (\theta - \gamma) \|\tilde{r}(u_\Lambda)\| \geq \frac{\theta - \gamma}{1 + \gamma} \|r(u_\Lambda)\| , \end{aligned}$$

which is the desired (3.13).  $\square$

The previous result suggests introducing the following feasible variant of the procedure **RES**:



- $r := \mathbf{F-RES}(v_\Lambda, \gamma)$

Given  $\gamma \in (0, \theta)$  and a function  $v_\Lambda \in V_\Lambda$  for some finite index set  $\Lambda$ , the output  $r$  is an approximate residual  $\tilde{r}(v_\Lambda) = \tilde{f} + \nabla \cdot (\tilde{\nu} \nabla v_\Lambda) - \tilde{\sigma} v_\Lambda$ , defined on a finite set  $\tilde{\Lambda}$  and satisfying

$$\|r(v_\Lambda) - \tilde{r}(v_\Lambda)\| \leq \gamma \|\tilde{r}(v_\Lambda)\| .$$

**Theorem 3.2 (contraction property of F-ADFOUR)** *Consider the feasible variant **F-ADFOUR** of the adaptive algorithm **ADFOUR**, where the step  $r_{n+1} := \mathbf{RES}(u_{n+1})$  is replaced by the step  $r_{n+1} := \mathbf{F-RES}(u_{n+1}, \gamma)$  for some  $\gamma \in (0, \theta)$ . Then, the same conclusions of Theorem 3.1 hold true for this variant, with the contraction factor  $\rho$  replaced by  $\rho = \rho(\tilde{\theta})$ , where  $\tilde{\theta}$  is defined in (3.13).  $\square$*

In the rest of the paper, we will develop our analysis considering Algorithm **ADFOUR** rather than **F-ADFOUR**; this is just for the sake of simplicity, since all the conclusions extend in a straightforward manner to the latter version as well.

### 3.3 A-ADFOUR: An aggressive version of ADFOUR

Theorem 3.1 indicates that even if one chooses  $\theta$  very close to 1, the predicted error reduction rate  $\rho = \rho(\theta)$  is always bounded from below by the quantity  $\sqrt{1 - \frac{\alpha_*}{\alpha^*}}$ . Such a result looks overly pessimistic, particularly in the case of smooth (analytic) solutions, since a Fourier method allows for an exponential decay of the error as the number of (properly selected) active degrees of freedom is increased. Fig 3.3 displays the influence of Dörfler parameter on the decay rate and number of solves: choosing  $\theta$  closer to 1 does not significantly affect the rate of decay of the error versus the number of activated degrees of freedom, but it significantly reduces the number of iterations. This in turn reduces the computational cost measured in terms of Galerkin solves.

Motivated by this observation, hereafter we consider a variant of Algorithm **ADFOUR**, which – under the assumptions of Property 2.2 or 2.3 – guarantees an arbitrarily large error reduction per iteration, provided the set of the new degrees of freedom detected by **DÖRFLER** is suitably enriched.

At the  $n$ -th iteration, let us define the set  $\Lambda_{n+1} := \Lambda_n \cup \partial\Lambda_n$  by setting

$$\begin{aligned} \widetilde{\partial\Lambda_n} &:= \mathbf{DÖRFLER}(r_n, \theta) \\ \partial\Lambda_n &:= \mathbf{ENRICH}(\widetilde{\partial\Lambda_n}, J) , \end{aligned} \tag{3.14}$$

where the latter procedure and the value of the integer  $J$  will be defined later on. We recall that the set  $\widetilde{\partial\Lambda_n}$  is such that  $g_n = P_{\widetilde{\partial\Lambda_n}} r_n$  satisfies

$$\|r_n - g_n\| \leq \sqrt{1 - \theta^2} \|r_n\|$$

(see (3.4)). Let  $w_n \in V$  be the solution of  $Lw_n = g_n$ , which in general will have infinitely many components, and let us split it as

$$w_n = P_{\Lambda_{n+1}} w_n + P_{\Lambda_{n+1}^c} w_n =: y_n + z_n \in V_{\Lambda_{n+1}} \oplus V_{\Lambda_{n+1}^c} .$$

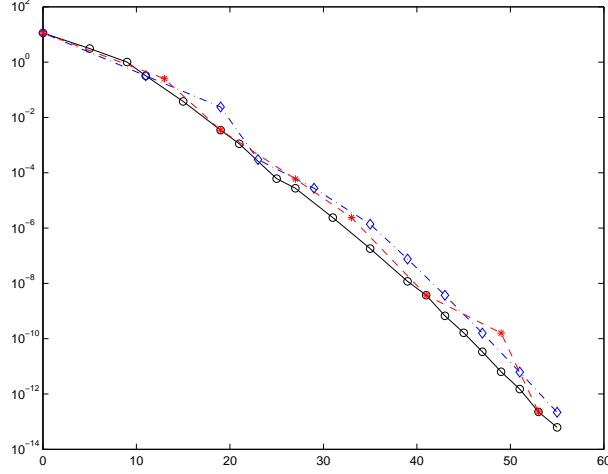


Figure 1: Residual norm vs number of degrees of freedom activated by **ADFOUR**, for different choices of Dörfler parameter  $\theta$ ; solid line:  $\theta = 1 - 10^{-1}$ ; dash-dotted line:  $\theta = 1 - 10^{-2}$ ; dashed line:  $\theta = 1 - 10^{-3}$ . The symbols (circles, diamonds, stars) identify the various **ADFOUR** iterations for the sample 1D problem (2.3) with analytic solution  $u(x) = \exp(\cos 2x + \sin x)$  and coefficients with  $\nu = 1 + \frac{1}{2} \sin 3x$  and  $\sigma = \exp(2 \cos 3x)$ .

Then, by the minimality property of the Galerkin solution in the energy norm and by (2.5) and (2.9), one has

$$\begin{aligned} \|u - u_{n+1}\| &\leq \|u - (u_n + y_n)\| \leq \|u - u_n - w_n + z_n\| \\ &\leq \frac{1}{\sqrt{\alpha_*}} \|L(u - u_n - w_n)\| + \sqrt{\alpha^*} \|z_n\| = \frac{1}{\sqrt{\alpha_*}} \|r_n - g_n\| + \sqrt{\alpha^*} \|z_n\|. \end{aligned}$$

Thus,

$$\|u - u_{n+1}\| \leq \frac{1}{\sqrt{\alpha_*}} \sqrt{(1 - \theta^2)} \|r_n\| + \sqrt{\alpha^*} \|z_n\|.$$

Now we can write  $z_n = (P_{\Lambda_{n+1}^c} L^{-1} P_{\widetilde{\partial\Lambda_n}}) r_n$ ; hence, if  $\Lambda_{n+1}$  is defined in such a way that

$$k \in \Lambda_{n+1}^c \quad \text{and} \quad \ell \in \widetilde{\partial\Lambda_n} \quad \Rightarrow \quad |k - \ell| > J,$$

then we have

$$\|P_{\Lambda_{n+1}^c} L^{-1} P_{\widetilde{\partial\Lambda_n}}\| \leq \|\mathbf{A}^{-1} - (\mathbf{A}^{-1})_J\| \leq \psi_{\mathbf{A}^{-1}}(J, \bar{\eta}_L),$$

where we have used (2.23). Now,  $J > 0$  can be chosen to satisfy

$$\psi_{\mathbf{A}^{-1}}(J, \bar{\eta}_L) \leq \sqrt{\frac{1 - \theta^2}{\alpha_* \alpha^*}}, \quad (3.15)$$

in such a way that

$$\|u - u_{n+1}\| \leq \frac{1}{\sqrt{\alpha_*}} \sqrt{1 - \theta^2} \|r_n\| \leq \left(\frac{\alpha^*}{\alpha_*}\right)^{1/2} \sqrt{1 - \theta^2} \|u - u_n\|. \quad (3.16)$$

Note that, as desired, the new error reduction rate

$$\bar{\rho} = \left( \frac{\alpha^*}{\alpha_*} \right)^{1/2} \sqrt{1 - \theta^2} \quad (3.17)$$

can be made arbitrarily small by choosing  $\theta$  arbitrarily close to 1. The procedure **ENRICH** is thus defined as follows:

- $\Lambda^* := \mathbf{ENRICH}(\Lambda, J)$

Given an integer  $J \geq 0$  and a finite set  $\Lambda \subset \mathbb{Z}^d$ , the output is the set

$$\Lambda^* := \{k \in \mathbb{Z}^d : \text{there exists } \ell \in \Lambda \text{ such that } |k - \ell| \leq J\} .$$

Note that since the procedure adds a  $d$ -dimensional ball of radius  $J$  around each point of  $\Lambda$ , the cardinality of the new set  $\Lambda^*$  can be estimated as

$$|\Lambda^*| \leq |\overline{B_d(0, J)} \cap \mathbb{Z}^d| |\Lambda| \sim \omega_d J^d |\Lambda| , \quad (3.18)$$

where  $\omega_d$  is the measure of the  $d$ -dimensional Euclidean unit ball  $B_d(0, 1)$  centered at the origin.

It is convenient for future reference to denote by  $\partial\Lambda_n := \mathbf{E-DÖRFLER}(r_n, \theta, J)$  the procedure described in (3.14). We summarize our results in the following theorem.

**Theorem 3.3 (contraction property of A-ADFOUR)** *Consider the aggressive variant **A-ADFOUR** of the adaptive algorithm **ADFOUR**, in which the step  $\partial\Lambda_n := \mathbf{DÖRFLER}(r_n, \theta)$  is replaced by*

$$\partial\Lambda_n := \mathbf{E-DÖRFLER}(r_n, \theta, J) ,$$

*where  $\theta$  is such that  $\bar{\rho}$  defined in (3.17) is smaller than 1, and  $J$  is the smallest integer for which (3.15) is fulfilled. Let the assumptions of Property 2.2 or 2.3 be satisfied. Then, the same conclusions of Theorem 3.1 hold true for this variant, with the contraction factor  $\rho$  replaced by  $\bar{\rho}$ .  $\square$*

### 3.4 C-ADFOUR and PC-ADFOUR: ADFOUR with coarsening

The adaptive algorithm **ADFOUR** and its variants introduced above are not guaranteed to be optimal in terms of complexity. Indeed, the discussion in the forthcoming Sect. 5 for the exponential case will indicate that the residual  $r(u_\Lambda)$  may be significantly less sparse than the corresponding Galerkin solution  $u_\Lambda$ ; in particular, we will see that many indices in  $\Lambda$ , activated in an early stage of the adaptive process, could be lately discarded since the corresponding components of  $u_\Lambda$  are zero. For these reasons, we propose here a new variant of algorithm **ADFOUR**, which incorporates a recursive coarsening step.

The algorithm is constructed through the procedures **GAL**, **RES**, **DÖRFLER** already introduced in Sect. 3.1, together with the new procedure **COARSE** defined as follows:

- $\Lambda := \mathbf{COARSE}(w, \epsilon)$

Given a function  $w \in V_{\Lambda^*}$  for some finite index set  $\Lambda^*$ , and an accuracy  $\epsilon$  which is known to satisfy  $\|u - w\| \leq \epsilon$ , the output  $\Lambda \subseteq \Lambda^*$  is a set of minimal cardinality such that

$$\|w - P_\Lambda w\| \leq 2\epsilon . \quad (3.19)$$

We will subsequently show (see Theorem 6.1) that the cardinality  $|\Lambda|$  is optimally related to the sparsity class of  $u$ . The following result will be used several times in the paper.

**Property 3.1 (coarsening)** *The procedure **COARSE** guarantees the bounds*

$$\|u - P_\Lambda w\| \leq 3\epsilon \quad (3.20)$$

and, for the Galerkin solution  $u_\Lambda \in V_\Lambda$ ,

$$\|u - u_\Lambda\| \leq 3\sqrt{\alpha^*}\epsilon. \quad (3.21)$$

*Proof.* The first bound is trivial, the second one follows from the minimality property of the Galerkin solution in the energy norm and from (2.5):

$$\|u - u_\Lambda\| \leq \|u - P_\Lambda w\| \leq \sqrt{\alpha^*}\|u - P_\Lambda w\| \leq 3\sqrt{\alpha^*}\epsilon. \quad \square$$

Given two parameters  $\theta \in (0, 1)$  and  $tol \in [0, 1)$ , we define the following adaptive algorithm with coarsening.

**Algorithm C-ADFOUR**( $\theta, tol$ )

Set  $r_0 := f$ ,  $\Lambda_0 := \emptyset$ ,  $n = -1$

do

$n \leftarrow n + 1$

set  $\Lambda_{n,0} = \Lambda_n$ ,  $r_{n,0} = r_n$

$k = -1$

do

$k \leftarrow k + 1$

$\partial\Lambda_{n,k} := \mathbf{DÖRFLER}(r_{n,k}, \theta)$

$\Lambda_{n,k+1} := \Lambda_{n,k} \cup \partial\Lambda_{n,k}$

$u_{n,k+1} := \mathbf{GAL}(\Lambda_{n,k+1})$

$r_{n,k+1} := \mathbf{RES}(u_{n,k+1})$

while  $\|r_{n,k+1}\| > \sqrt{1 - \theta^2}\|r_n\|$

$\Lambda_{n+1} := \mathbf{COARSE}\left(u_{n,k+1}, \frac{1}{\sqrt{\alpha_*}}\|r_{n,k+1}\|\right)$

$u_{n+1} := \mathbf{GAL}(\Lambda_{n+1})$

$r_{n+1} := \mathbf{RES}(u_{n+1})$

while  $\|r_{n+1}\| > tol$

We observe that the specific choice of accuracy  $\epsilon = \epsilon_n = \frac{1}{\sqrt{\alpha_*}}\|r_{n,k+1}\|$  in each call of **COARSE** in the algorithm above is motivated by the wish of guaranteeing a fixed reduction of the residual and error at each outer iteration. This is made precise in the following theorem.

**Theorem 3.4 (contraction property of C-ADFOUR)** *The algorithm **C-ADFOUR** satisfies*

- (i) The number of iterations of each inner loop is finite and bounded independently of  $n$ ;
- (ii) The sequence of residuals  $r_n$  and errors  $u - u_n$  generated for  $n \geq 0$  by the algorithm satisfies the inequalities

$$\|r_{n+1}\| \leq \rho \|r_n\| \quad (3.22)$$

and

$$\|u - u_{n+1}\| \leq \rho \|u - u_n\| \quad (3.23)$$

for

$$\rho = 3 \frac{\alpha^*}{\alpha_*} \sqrt{1 - \theta^2} . \quad (3.24)$$

In particular, if  $\theta$  is chosen in such a way that  $\rho < 1$ , for any  $\text{tol} > 0$  the algorithm terminates in a finite number of iterations, whereas for  $\text{tol} = 0$  the sequence  $u_n$  converges to  $u$  in  $H_p^1(\Omega)$  as  $n \rightarrow \infty$ .

*Proof.* (i) For any fixed  $n$ , each inner iteration behaves as the algorithm **ADFOUR** considered in Sect. 3.1. Hence, setting again  $\rho = \sqrt{1 - \frac{\alpha_*}{\alpha^*} \theta^2}$ , we have as in Theorem 3.1

$$\|u - u_{n,k+1}\| \leq \rho^{k+1} \|u - u_n\| ,$$

which implies, by (2.9),

$$\|r_{n,k+1}\| \leq \sqrt{\alpha^*} \|u - u_{n,k+1}\| \leq \sqrt{\alpha^*} \rho^{k+1} \|u - u_n\| \leq \sqrt{\frac{\alpha^*}{\alpha_*}} \rho^{k+1} \|r_n\| .$$

This shows that the termination criterion

$$\|r_{n,k+1}\| \leq \sqrt{1 - \theta^2} \|r_n\| \quad (3.25)$$

is certainly satisfied if

$$\sqrt{\frac{\alpha^*}{\alpha_*}} \rho^{k+1} \leq \sqrt{1 - \theta^2} ,$$

i.e., as soon as

$$k + 1 \geq \frac{\log\left(\frac{\alpha_*}{\alpha^*}(1 - \theta^2)\right)}{2 \log \rho} > k .$$

We conclude that the number  $K_n = k + 1$  of inner iterations is bounded by  $1 + \frac{\log\left(\frac{\alpha_*}{\alpha^*}(1 - \theta^2)\right)}{2 \log \rho}$ , which is independent of  $n$ .

(ii) By (2.8), we have

$$\|u - u_{n,k+1}\| \leq \frac{1}{\alpha_*} \|r_{n,k+1}\| .$$

At the exit of the inner loop, the quantity on the right-hand side is precisely the parameter  $\epsilon_n$  fed to the procedure **COARSE**; then, Property 3.1 yields

$$\|u - u_{n+1}\| \leq 3\sqrt{\alpha^*} \epsilon_n .$$

On the other hand, the termination criterion (3.25) yields

$$\epsilon_n \leq \frac{1}{\alpha_*} \sqrt{1 - \theta^2} \|r_n\| ,$$

so that

$$\|u - u_{n+1}\| \leq 3 \frac{\sqrt{\alpha^*}}{\alpha_*} \sqrt{1 - \theta^2} \|r_n\| .$$

This bound together with the left-hand inequality in (2.9) applied to  $r_{n+1}$  yields (3.22), whereas the same inequality applied to  $r_n$  yields (3.23).  $\square$

A coarsening step can also be inserted in the aggressive algorithm **A-ADFOUR** considered in Sect. 3.3; indeed, the enrichment step **ENRICH** could activate a larger number of degrees of freedom than really needed, endangering optimality. The algorithm we now propose can be viewed as a variant of **C-ADFOUR**, in which the use of **E-DÖRFLER** instead of **DÖRFLER** allows one to take a single inner iteration; in this respect, one can consider the enrichment step as a “prediction”, and the coarsening step as a “correction”, of the new set of active degrees of freedom. For this reason, we call this variant the **Predictor/Corrector-ADFOUR**, or simply **PC-ADFOUR**.

Given two parameters  $\theta \in (0, 1)$  and  $tol \in [0, 1)$ , we choose  $J \geq 1$  as the smallest integer for which (3.15) is fulfilled, and we define the following adaptive algorithm.

**Algorithm PC-ADFOUR**( $\theta, tol, J$ )

Set  $r_0 := f$ ,  $\Lambda_0 := \emptyset$ ,  $n = -1$

do

$n \leftarrow n + 1$

$\widehat{\partial}\Lambda_n := \mathbf{E-DÖRFLER}(r_n, \theta, J)$

$\widehat{\Lambda}_{n+1} := \Lambda_n \cup \widehat{\partial}\Lambda_n$

$\widehat{u}_{n+1} := \mathbf{GAL}(\widehat{\Lambda}_n)$

$\Lambda_{n+1} := \mathbf{COARSE}\left(\widehat{u}_{n+1}, \frac{1}{\alpha_*} \sqrt{1 - \theta^2} \|r_n\|\right)$

$u_{n+1} := \mathbf{GAL}(\Lambda_{n+1})$

$r_{n+1} := \mathbf{RES}(u_{n+1})$

while  $\|r_{n+1}\| > tol$

**Theorem 3.5 (contraction property of PC-ADFOUR)** *If the assumptions of Property 2.2 or Property 2.3 be satisfied, then the statement (ii) of Theorem 3.4 applies to Algorithm PC-ADFOUR as well.*

*Proof.* The first inequalities in both (3.16) and (2.5) yield

$$\|u - \widehat{u}_{n+1}\| \leq \frac{1}{\alpha_*} \sqrt{1 - \theta^2} \|r_n\| .$$

Since the right-hand side is precisely the parameter  $\epsilon_n$  fed to the procedure **COARSE**, one proceeds as in the proof of Theorem 3.4.  $\square$

## 4 Nonlinear approximation in Fourier spaces

### 4.1 Best $N$ -term approximation and rearrangement

Given any nonempty finite index set  $\Lambda \subset \mathbb{Z}^d$  and the corresponding subspace  $V_\Lambda \subset V = H_p^1(\Omega)$  of dimension  $|\Lambda| = \text{card } \Lambda$ , the best approximation of  $v$  in  $V_\Lambda$  is the orthogonal projection of  $v$  upon  $V_\Lambda$ , i.e. the function  $P_\Lambda v = \sum_{k \in \Lambda} \hat{v}_k \phi_k$ , which satisfies

$$\|v - P_\Lambda v\| = \left( \sum_{k \notin \Lambda} |\hat{V}_k|^2 \right)^{1/2}$$

(we set  $P_\Lambda v = 0$  if  $\Lambda = \emptyset$ ). For any integer  $N \geq 1$ , we minimize this error over all possible choices of  $\Lambda$  with cardinality  $N$ , thereby leading to the *best  $N$ -term approximation error*

$$E_N(v) = \inf_{\Lambda \subset \mathbb{Z}^d, |\Lambda|=N} \|v - P_\Lambda v\|.$$

A way to construct a *best  $N$ -term approximation*  $v_N$  of  $v$  consists of rearranging the coefficients of  $v$  in decreasing order of modulus

$$|\hat{V}_{k_1}| \geq \dots \geq |\hat{V}_{k_n}| \geq |\hat{V}_{k_{n+1}}| \geq \dots$$

and setting  $v_N = P_{\Lambda_N} v$  with  $\Lambda_N = \{k_n : 1 \leq n \leq N\}$ . As already mentioned in the Introduction, let us denote from now on  $v_n^* = \hat{V}_{k_n}$  the rearranged and rescaled Fourier coefficients of  $v$ . Then,

$$E_N(v) = \left( \sum_{n>N} |v_n^*|^2 \right)^{1/2}.$$

Next, given a strictly decreasing function  $\phi : \mathbb{N} \rightarrow \mathbb{R}_+$  such that  $\phi(0) = \phi_0$  for some  $\phi_0 > 0$  and  $\phi(N) \rightarrow 0$  when  $N \rightarrow \infty$ , we introduce the corresponding *sparsity class*  $\mathcal{A}_\phi$  by setting

$$\mathcal{A}_\phi = \left\{ v \in V : \|v\|_{\mathcal{A}_\phi} := \sup_{N \geq 0} \frac{E_N(v)}{\phi(N)} < +\infty \right\}. \quad (4.1)$$

We point out that in applications  $\|v\|_{\mathcal{A}_\phi}$  need not be a (quasi-)norm since  $\mathcal{A}_\phi$  need not be a linear space. Note however that  $\|v\|_{\mathcal{A}_\phi}$  always controls the  $V$ -norm of  $v$ , since  $\|v\| = E_0(v) \leq \phi_0 \|v\|_{\mathcal{A}_\phi}$ . Observe that  $v \in \mathcal{A}_\phi$  iff there exists a constant  $c > 0$  such that

$$E_N(v) \leq c\phi(N), \quad \forall N \geq 0. \quad (4.2)$$

The quantity  $\|v\|_{\mathcal{A}_\phi}$  dictates the minimal number  $N_\varepsilon$  of basis functions needed to approximate  $v$  with accuracy  $\varepsilon$ . In fact, from the relations

$$E_{N_\varepsilon}(v) \leq \varepsilon < E_{N_\varepsilon-1}(v) \leq \phi(N_\varepsilon-1) \|v\|_{\mathcal{A}_\phi},$$

and the monotonicity of  $\phi$ , we obtain

$$N_\varepsilon \leq \phi^{-1} \left( \frac{\varepsilon}{\|v\|_{\mathcal{A}_\phi}} \right) + 1. \quad (4.3)$$

The second addend on the right-hand side can be absorbed by a multiple of the first one, provided  $\varepsilon$  is sufficiently small; in other words, it is not restrictive to assume that there exists a constant  $\kappa$  slightly larger than 1 such that

$$N_\varepsilon \leq \kappa \phi^{-1} \left( \frac{\varepsilon}{\|v\|_{\mathcal{A}_\phi}} \right). \quad (4.4)$$

**Remark 4.1 (sparsity class for  $V'$ )** Replacing  $V$  by  $V'$  in (4.1) leads to the definition of a sparsity class, still denoted by  $\mathcal{A}_\phi$ , in the space of linear continuous forms  $f$  on  $H_p^1(\Omega)$ . This observation applies to the subsequent definitions as well (e.g., for the class  $\mathcal{A}_G^{\eta,t}$ ). In essence, we will treat in a unified way the nonlinear approximation of a function  $v \in H_p^1(\Omega)$  and of a form  $f \in H_p^{-1}(\Omega)$ .  $\square$

Throughout the paper, we shall consider two main families of sparsity classes, identified by specific choices of the function  $\phi$  depending upon one or more parameters. The first family is related to the best approximation in *Besov* spaces of periodic functions, thus accounting for a finite-order regularity in  $\Omega$ ; the corresponding functions  $\phi$  exhibit an algebraic decay as  $N \rightarrow \infty$ , which motivates our terminology of *algebraic classes*. The second family is related to the best approximation in *Gevrey* spaces of periodic functions, which are formed by infinitely-differentiable functions in  $\Omega$ ; the associated  $\phi$ 's exhibit an exponential decay, and for this reason such classes will be referred to as *exponential classes*. Properties of both families are collected hereafter.

## 4.2 Algebraic classes

The following is the counterpart for Fourier approximations of by now well-known nonlinear approximation settings [11], e.g. for wavelets or nested finite elements. For this reason, we just state definitions and properties without proofs.

For  $s > 0$ , let us introduce the function

$$\phi(N) = N^{-s/d} \quad \text{for } N \geq 1, \quad (4.5)$$

and  $\phi(0) = \phi_0 > 1$  arbitrary, with inverse

$$\phi^{-1}(\lambda) = \lambda^{-d/s} \quad \text{for } \lambda \leq 1, \quad (4.6)$$

and let us consider the corresponding class  $\mathcal{A}_\phi$  defined in (4.1).

**Definition 4.1 (algebraic class of functions)** We denote by  $\mathcal{A}_B^s$  the subset of  $V$  defined as

$$\mathcal{A}_B^s := \left\{ v \in V : \|v\|_{\mathcal{A}_B^s} := \|v\| + \sup_{N \geq 1} E_N(v) N^{s/d} < +\infty \right\}.$$

It is immediately seen that  $\mathcal{A}_B^s$  contains the Sobolev space of periodic functions  $H_p^{s+1}(\Omega)$ . On the other hand, it is proven in [12], as a part of a more general result, that for  $0 < \sigma, \tau \leq \infty$ , the Besov space  $B_{\tau,\sigma}^{s+1}(\Omega) = B_\sigma^{s+1}(L^\tau(\Omega))$  is contained in  $\mathcal{A}_B^{s*}$  provided  $s^* := s - d(1/\tau - 1/2)_+ > 0$ .

Let us associate the quantity  $\tau > 0$  to the parameter  $s$ , via the relation

$$\frac{1}{\tau} = \frac{s}{d} + \frac{1}{2}.$$



The condition for a function  $v$  to belong to some class  $\mathcal{A}_B^s$  can be equivalently stated as a condition on the vector  $\mathbf{v} = (\hat{V}_k)_{k \in \mathbb{Z}^d}$  of its Fourier coefficients, precisely, on the rate of decay of the non-increasing rearrangement  $\mathbf{v}^* = (v_n^*)_{n \geq 1}$  of  $\mathbf{v}$ .

**Definition 4.2 (algebraic class of sequences)** Let  $\ell_B^s(\mathbb{Z}^d)$  be the subset of sequences  $\mathbf{v} \in \ell^2(\mathbb{Z}^d)$  so that

$$\|\mathbf{v}\|_{\ell_B^s(\mathbb{Z}^d)} := \sup_{n \geq 1} n^{1/\tau} |v_n^*| < +\infty.$$

Note that this space is often denoted by  $\ell_w^r(\mathbb{Z}^d)$  in the literature, being an example of Lorentz space.

The relationship between  $\mathcal{A}_B^s$  and  $\ell_B^s(\mathbb{Z}^d)$  is stated in the following Proposition.

**Proposition 4.1 (equivalence of algebraic classes)** Given a function  $v \in V$  and the sequence  $\mathbf{v}$  of its Fourier coefficients, one has  $v \in \mathcal{A}_B^s$  if and only if  $\mathbf{v} \in \ell_B^s(\mathbb{Z}^d)$ , with

$$\|v\|_{\mathcal{A}_B^s} \lesssim \|\mathbf{v}\|_{\ell_B^s(\mathbb{Z}^d)} \lesssim \|v\|_{\mathcal{A}_B^s}.$$

At last, we note that the quasi-Minkowski inequality

$$\|\mathbf{u} + \mathbf{v}\|_{\ell_B^s(\mathbb{Z}^d)} \leq C_s \left( \|\mathbf{u}\|_{\ell_B^s(\mathbb{Z}^d)} + \|\mathbf{v}\|_{\ell_B^s(\mathbb{Z}^d)} \right)$$

holds in  $\ell_B^s(\mathbb{Z}^d)$ , yet the constant  $C_s$  blows up exponentially as  $s \rightarrow \infty$ .

### 4.3 Exponential classes

We first recall the definition of Gevrey spaces of periodic functions in  $\Omega = (0, 2\pi)^d$  (see [14]). Given reals  $\eta > 0$ ,  $0 < t \leq d$  and  $s \geq 0$ , we set

$$G_p^{\eta,t,s}(\Omega) := \left\{ v \in L^2(\Omega) : \|v\|_{G,\eta,t,s}^2 = \sum_{k \in \mathbb{Z}} e^{2\eta|k|^t} (1 + |k|^{2s}) |\hat{v}_k|^2 < +\infty \right\}.$$

Note that  $G_p^{\eta,t,s}(\Omega)$  is contained in all Sobolev spaces of periodic functions  $H_p^r(\Omega)$ ,  $r \geq 0$ . Furthermore, if  $t \geq 1$ ,  $G_p^{\eta,t,s}(\Omega)$  is made of analytic functions.

Gevrey spaces have been introduced to study the  $C^\infty$  and analytical regularity of the solutions of partial differential equations. For our elliptic problem (2.3), the following statement is an example of shift theorem in Gevrey spaces.

**Theorem 4.1 (shift theorem)** If the assumptions of Property 2.3 are satisfied, then for any  $\eta < \bar{\eta}_L$ ,  $0 < t \leq 1$  and  $s \geq -1$ ,  $L$  is an isomorphism between  $G_p^{\eta,t,s+2}(\Omega)$  and  $G_p^{\eta,t,s}(\Omega)$ .

*Proof.* Proceeding as in Sect. 2.3, it is immediate to see that the problem  $Lu = f$  can be equivalently formulated as  $\mathbf{A}\mathbf{u} = \mathbf{f}$ , where the vectors  $\mathbf{f}$  and  $\mathbf{u}$  contain the Fourier coefficients of functions  $f$  and  $u$  normalized in  $H_p^s(\Omega)$  and  $H_p^{s+2}(\Omega)$ , respectively. If  $\mathbf{W} = \text{diag}(e^{\eta|k|^t})$  is a bi-infinite diagonal exponential matrix, then we can write  $\mathbf{W}\mathbf{u} = \mathbf{W}\mathbf{A}^{-1}\mathbf{f} = (\mathbf{W}\mathbf{A}^{-1}\mathbf{W}^{-1})\mathbf{W}\mathbf{f}$ . We observe that property  $\|\mathbf{W}\mathbf{u}\|_{\ell^2} \lesssim \|\mathbf{W}\mathbf{f}\|_{\ell^2}$ , which implies the thesis, is a consequence of  $\|\mathbf{W}\mathbf{A}^{-1}\mathbf{W}^{-1}\|_{\ell^2} \lesssim 1$ .

To show the latter inequality, we let  $\mathbf{x}, \mathbf{y} \in \ell^2(\mathbb{Z}^d)$  and notice that

$$|\mathbf{y}^T \mathbf{W}\mathbf{A}^{-1}\mathbf{W}^{-1}\mathbf{x}| \leq c_L \sum_{m \in \mathbb{Z}^d} e^{-\bar{\eta}_L|m|} \sum_{k \in \mathbb{Z}^d} |y_{m+k}| e^{\eta|m+k|^t} e^{-\eta|k|^t} |x_k|.$$

Since  $0 < t \leq 1$ , we deduce  $|m+k|^t \leq |m|^t + |k|^t$  and  $e^{\eta(|m+k|^t - |k|^t)} \leq e^{\eta|m|^t}$ , whence

$$\|\mathbf{W}\mathbf{A}^{-1}\mathbf{W}^{-1}\mathbf{x}\| = \sup_{\mathbf{y} \in \mathbb{Z}^d} \frac{|\mathbf{y}^T \mathbf{W}\mathbf{A}^{-1}\mathbf{W}^{-1}\mathbf{x}|}{\|\mathbf{y}\|} \leq c_L \sum_{m \in \mathbb{Z}^d} e^{(-\bar{\eta}_L + \eta)|m|^t} \|\mathbf{x}\| \lesssim \|\mathbf{x}\|$$

because  $\bar{\eta}_L > \eta$  and the series converges. This implies the desired estimate.  $\square$

From now on, we fix  $s = 1$  and we normalize again the Fourier coefficients of a function  $v$  with respect to the  $H_p^1(\Omega)$ -norm. Thus, we set

$$G_p^{\eta,t}(\Omega) = G_p^{\eta,t,1}(\Omega) = \{v \in V : \|v\|_{G,\eta,t}^2 = \sum_k e^{2\eta|k|^t} |\hat{V}_k|^2 < +\infty\}. \quad (4.7)$$

Functions in  $G_p^{\eta,t}(\Omega)$  can be approximated by the linear orthogonal projection

$$P_M v = \sum_{|k| \leq M} \hat{V}_k \phi_k,$$

for which we have

$$\begin{aligned} \|v - P_M v\|^2 &= \sum_{|k| > M} |\hat{V}_k|^2 = \sum_{|k| > M} e^{-2\eta|k|^t} e^{2\eta|k|^t} |\hat{V}_k|^2 \\ &\leq e^{-2\eta M^t} \sum_{|k| > M} e^{2\eta|k|^t} |\hat{V}_k|^2 \leq e^{-2\eta M^t} \|v\|_{G,\eta,t}^2. \end{aligned}$$

As already observed in Property 2.4, setting  $N = \text{card}\{k : |k| \leq M\}$ , one has  $N \sim \omega_d M^d$ , so that

$$E_N(v) \leq \|v - P_M v\| \lesssim \exp\left(-\eta \omega_d^{-t/d} N^{t/d}\right) \|v\|_{G,\eta,t}. \quad (4.8)$$

Hence, we are led to introduce the function

$$\phi(N) = \exp\left(-\eta \omega_d^{-t/d} N^{t/d}\right) \quad (N \geq 0), \quad (4.9)$$

whose inverse is given by

$$\phi^{-1}(\lambda) = \frac{\omega_d}{\eta^{d/t}} \left(\log \frac{1}{\lambda}\right)^{d/t} \quad (\lambda \leq 1), \quad (4.10)$$

and to consider the corresponding class  $\mathcal{A}_\phi$  defined in (4.1), which therefore contains  $G_p^{\eta,t}(\Omega)$ .

**Definition 4.3 (exponential class of functions)** We denote by  $\mathcal{A}_G^{\eta,t}$  the subset of  $G_p^{\eta,t}(\Omega)$  defined as

$$\mathcal{A}_G^{\eta,t} := \left\{v \in V : \|v\|_{\mathcal{A}_G^{\eta,t}} := \sup_{N \geq 0} E_N(v) \exp\left(\eta \omega_d^{-t/d} N^{t/d}\right) < +\infty\right\}.$$

At this point, we make the subsequent notation easier by introducing the  $t$ -dependent function

$$\tau = \frac{t}{d} \leq 1.$$

As in the algebraic case, the class  $\mathcal{A}_G^{\eta,t}$  can be equivalently characterized in terms of behavior of rearranged sequences of Fourier coefficients.

**Definition 4.4 (exponential class of sequences)** Let  $\ell_G^{\eta,t}(\mathbb{Z}^d)$  be the subset of sequences  $\mathbf{v} \in \ell^2(\mathbb{Z}^d)$  so that

$$\|\mathbf{v}\|_{\ell_G^{\eta,t}(\mathbb{Z}^d)} := \sup_{n \geq 1} n^{(1-\tau)/2} \exp(\eta \omega_d^{-\tau} n^\tau) |v_n^*| < +\infty ,$$

where  $\mathbf{v}^* = (v_n^*)_{n=1}^\infty$  is the non-increasing rearrangement of  $\mathbf{v}$ .

The relationship between  $\mathcal{A}_G^{\eta,t}$  and  $\ell_G^{\eta,t}(\mathbb{Z}^d)$  is stated in the following Proposition.

**Proposition 4.2 (equivalence of exponential classes)** Given a function  $v \in V$  and the sequence  $\mathbf{v} = (\hat{V}_k)_{k \in \mathbb{Z}^d}$  of its Fourier coefficients, one has  $v \in \mathcal{A}_G^{\eta,t}$  if and only if  $\mathbf{v} \in \ell_G^{\eta,t}(\mathbb{Z}^d)$ , with

$$\|v\|_{\mathcal{A}_G^{\eta,t}} \lesssim \|\mathbf{v}\|_{\ell_G^{\eta,t}(\mathbb{Z}^d)} \lesssim \|v\|_{\mathcal{A}_G^{\eta,t}} .$$

*Proof.* Assume first that  $\mathbf{v} \in \ell_G^{\eta,t}(\mathbb{Z}^d)$ . Then,

$$E_N(v)^2 = \|v - P_N(v)\|^2 = \sum_{n > N} |v_n^*|^2 \lesssim \sum_{n > N} n^{\tau-1} \exp(-2\eta \omega_d^{-\tau} n^\tau) \|\mathbf{v}\|_{\ell_G^{\eta,t}(\mathbb{Z}^d)}^2 .$$

Now, setting for simplicity  $\alpha = 2\eta \omega_d^{-\tau}$ , one has

$$S := \sum_{n > N} n^{\tau-1} e^{-\alpha n^\tau} \sim \int_N^\infty x^{\tau-1} e^{-\alpha x^\tau} dx .$$

The substitution  $z = x^\tau$  yields

$$S \sim \frac{d}{t} \int_{N^\tau}^\infty e^{-\alpha z} dz = \frac{d}{\alpha t} e^{-\alpha N^\tau}$$

whence  $\|v\|_{\mathcal{A}_G^{\eta,t}} \lesssim \|\mathbf{v}\|_{\ell_G^{\eta,t}(\mathbb{Z}^d)}$ . Conversely, let  $v \in \mathcal{A}_G^{\eta,t}$ . We have to prove that for any  $n \geq 1$ , one has

$$n^{1-\tau} |v_n^*|^2 \lesssim e^{-\alpha n^\tau} \|v\|_{\mathcal{A}_G^{\eta,t}}^2 .$$

Let  $m < n$  be the largest integer such that  $n - m \geq n^{1-\tau}$  (note that  $0 \leq 1 - \tau < 1$ ), i.e.,  $m \sim n(1 - n^{-\tau})$ . Then,

$$n^{1-\tau} |v_n^*|^2 \leq (n - m) |v_n^*|^2 \leq \sum_{j=m+1}^n |v_j^*|^2 \leq \|v - P_m(v)\|^2 \leq e^{-\alpha m^\tau} \|v\|_{\mathcal{A}_G^{\eta,t}}^2 .$$

Now, by Taylor expansion,

$$m^\tau \sim n^\tau (1 - n^{-\tau})^\tau = n^\tau (1 - \tau n^{-\tau} + o(n^{-\tau})) = n^\tau - \tau + o(1) ,$$

so that  $e^{-\alpha m^\tau} \lesssim e^{-\alpha n^\tau}$ , and  $\|\mathbf{v}\|_{\ell_G^{\eta,t}(\mathbb{Z}^d)} \lesssim \|v\|_{\mathcal{A}_G^{\eta,t}}$  is proven.  $\square$

Next, we briefly comment on the structure of the set  $\ell_G^{\eta,t}(\mathbb{Z}^d)$ . This is not a vector space, since it may happen that  $\mathbf{u}, \mathbf{v}$  belong to this set, whereas  $\mathbf{u} + \mathbf{v}$  does not. Assume for simplicity that  $\tau = 1$  and consider for instance the sequences in  $\ell_G^{\eta,t}(\mathbb{Z}^d)$

$$\begin{aligned} \mathbf{u} &= (e^{-\eta}, 0, e^{-2\eta}, 0, e^{-3\eta}, 0, e^{-4\eta}, 0, \dots) , \\ \mathbf{v} &= (0, e^{-\eta}, 0, e^{-2\eta}, 0, e^{-3\eta}, 0, e^{-4\eta}, \dots) , \end{aligned}$$

Then,

$$\mathbf{u} + \mathbf{v} = (\mathbf{u} + \mathbf{v})^* = (e^{-\eta}, e^{-\eta}, e^{-2\eta}, e^{-2\eta}, e^{-3\eta}, e^{-3\eta}, e^{-4\eta}, e^{-4\eta}, \dots) ;$$

thus,  $(\mathbf{u} + \mathbf{v})_{2j}^* = e^{-\eta j}$ , so that  $e^{\eta 2j}(\mathbf{u} + \mathbf{v})_{2j}^* \rightarrow \infty$  as  $j \rightarrow +\infty$ , i.e.,  $\mathbf{u} + \mathbf{v} \notin \ell_G^{\eta,t}(\mathbb{Z}^d)$ . On the other hand, we have the following property.

**Lemma 4.1 (quasi-triangle inequality)** *If  $\mathbf{u}_i \in \ell_G^{\eta_i,t}(\mathbb{Z}^d)$  for  $i = 1, 2$ , then  $\mathbf{u}_1 + \mathbf{u}_2 \in \ell_G^{\eta,t}(\mathbb{Z}^d)$  with*

$$\|\mathbf{u}_1 + \mathbf{u}_2\|_{\ell_G^{\eta,t}} \leq \|\mathbf{u}_1\|_{\ell_G^{\eta_1,t}} + \|\mathbf{u}_2\|_{\ell_G^{\eta_2,t}}, \quad \eta^{-\frac{1}{\tau}} = \eta_1^{-\frac{1}{\tau}} + \eta_2^{-\frac{1}{\tau}}.$$

*Proof.* We use the characterization given by Proposition 4.2, so that

$$\|u_i - P_{N_i}(u_i)\| \leq \|u_i\|_{\mathcal{A}_G^{\eta_i,t} \exp(-\eta \omega_d^{-\tau} N_i^\tau)} \quad i = 1, 2.$$

Given  $N \geq 1$ , we seek  $N_1, N_2$  so that

$$N = N_1 + N_2, \quad \eta_1 N_1^\tau = \eta_2 N_2^\tau.$$

This implies

$$N = N_1 \eta_1^{\frac{1}{\tau}} \left( \eta_1^{-\frac{1}{\tau}} + \eta_2^{-\frac{1}{\tau}} \right) = N_1 \eta_1^{\frac{1}{\tau}} \eta^{-\frac{1}{\tau}},$$

and

$$\begin{aligned} \|(u_1 + u_2) - P_N(u_1 + u_2)\| &\leq \|u_1 - P_{N_1}(u_1)\| + \|u_2 - P_{N_2}(u_2)\| \\ &\leq \|u_1\|_{\mathcal{A}_G^{\eta_1,t} \exp(-\eta_1 \omega_d^{-\tau} N_1^\tau)} + \|u_2\|_{\mathcal{A}_G^{\eta_2,t} \exp(-\eta_2 \omega_d^{-\tau} N_2^\tau)} \\ &\leq (\|u_1\|_{\mathcal{A}_G^{\eta_1,t}} + \|u_2\|_{\mathcal{A}_G^{\eta_2,t}}) \exp(-\eta \omega_d^{-\tau} N^\tau). \end{aligned}$$

whence the assertion.  $\square$

Note that when  $\eta_1 = \eta_2$  we obtain  $\eta = 2^{-\tau} \eta_1 \leq 2^{-1} \eta_1$  thereby extending the previous counterexample.

## 5 Sparsity classes of the residual

For any finite index set  $\Lambda$ , let  $r = r(u_\Lambda)$  be the residual produced by the Galerkin solution  $u_\Lambda$ . Under Assumption 3.1, the step

$$\partial\Lambda := \mathbf{DÖRFLER}(r, \theta)$$

selects a set  $\partial\Lambda$  of minimal cardinality in  $\Lambda^c$  for which  $\|r - P_{\partial\Lambda} r\| \leq \sqrt{1 - \theta^2} \|r\|$ . Thus, if  $r$  belongs to a certain sparsity class  $\mathcal{A}_{\bar{\phi}}$ , identified by a function  $\bar{\phi}$ , then (4.3) yields

$$|\partial\Lambda| \leq \bar{\phi}^{-1} \left( \sqrt{1 - \theta^2} \frac{\|r\|}{\|r\|_{\mathcal{A}_{\bar{\phi}}}} \right) + 1. \quad (5.1)$$

Explicitly, if  $r \in \mathcal{A}_B^{\bar{s}}$  for some  $\bar{s} > 0$ , we have by (4.6)

$$|\partial\Lambda| \leq (1 - \theta^2)^{-d/2\bar{s}} \left( \frac{\|r\|_{\mathcal{A}_B^{\bar{s}}}}{\|r\|} \right)^{d/\bar{s}} + 1,$$

whereas if  $r \in \mathcal{A}_G^{\bar{\eta}, \bar{t}}$  for some  $\bar{\eta} > 0$  and  $\bar{t} > 0$ , we have by (4.10)

$$|\partial\Lambda| \leq \frac{\omega_d}{\eta^{d/\bar{t}}} \left( \log \frac{\|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}}}{\|r\|} + |\log \sqrt{1 - \theta^2}| \right)^{d/\bar{t}} + 1 .$$

We stress the fact that the cardinality of  $\partial\Lambda$  is related to the *sparsity class of the residual*. We will see in the rest of this section that such a class does coincide with the sparsity class of the solution in the algebraic case, whereas it is different (indeed, worse) in the exponential case. This is a crucial point to be kept in mind in the forthcoming optimality analysis of our algorithms.

The cardinality of  $\partial\Lambda$  depends indeed on how much the sparsity measure  $\|r\|_{\mathcal{A}_\phi}$  deviates from the Hilbert norm  $\|r\|$ . So, before embarking ourselves on the study of the relationship between the sparsity classes of the residual and of the solution, we make some brief comments on the ratio between these two quantities. For shortness, we only consider the exponential case, although similar considerations apply to the algebraic case as well. The size of the ratio

$$Q := \frac{\|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}}}{\|r\|}$$

depends on the relative behavior of the rearranged coefficients  $r_n^*$  of  $r$ , which by Definition 4.4 and Proposition 4.2 satisfy

$$|r_n^*| \leq \lambda^* n^{(\bar{\tau}-1)/2} e^{-\bar{\eta}\omega_d^{-\bar{\tau}} n^{\bar{\tau}}} \|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}} \quad (5.2)$$

for some constant  $\lambda^* > 0$ , with  $\bar{\tau} = \bar{t}/d$ . Let us consider two representative situations.

**Example 5.1 (*genuinely decaying functions*)** The most “favorable” situation is the one in which the sequence of rearranged coefficients decays precisely at the rate given by the right-hand side of (5.2); in other words, suppose that there exists a constant  $\lambda_* > 0$  such that for all  $n \geq 1$

$$\lambda_* n^{(\bar{\tau}-1)/2} e^{-\bar{\eta}\omega_d^{-\bar{\tau}} n^{\bar{\tau}}} \|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}} \leq |r_n^*| \leq \lambda^* n^{(\bar{\tau}-1)/2} e^{-\bar{\eta}\omega_d^{-\bar{\tau}} n^{\bar{\tau}}} \|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}} . \quad (5.3)$$

Then,

$$(\lambda_*)^2 \sum_{n \geq 1} n^{(\bar{\tau}-1)} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} n^{\bar{\tau}}} \|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}}^2 \leq \|r\|^2 \leq (\lambda^*)^2 \sum_{n \geq 1} n^{(\bar{\tau}-1)} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} n^{\bar{\tau}}} \|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}}^2 ,$$

and since

$$\sum_{n \geq 1} n^{(\bar{\tau}-1)} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} n^{\bar{\tau}}} \sim \int_1^{+\infty} x^{\bar{\tau}-1} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} x^{\bar{\tau}}} dx = \int_1^{+\infty} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} y} dy = C ,$$

we obtain

$$\frac{1}{C\lambda_*} \leq Q \leq \frac{1}{C\lambda^*} .$$

Thus, if (5.3) is a “tight” bound, the ratio  $Q$  is “small”, and the procedure **DÖRFLER** activates a moderate number of degrees of freedom at the current iteration.  $\square$

**Example 5.2 (plateaux)** The opposite situation, i.e., the worst scenario, occurs when the sequence of rearranged coefficients of  $r$  exhibits large “plateaux” consisting of equal (or nearly equal) elements in modulus. Fix an integer  $K$  arbitrarily large, and suppose that the  $K$  largest coefficients of  $r$  satisfy

$$|r_1^*| = |r_2^*| = \cdots = |r_{K-1}^*| = |r_K^*| = \lambda^* K^{(\bar{\tau}-1)/2} e^{-\bar{\eta}\omega_d^{-\bar{\tau}} K^{\bar{\tau}}} \|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{\tau}}}.$$

Since

$$\sum_{n>K} n^{(\bar{\tau}-1)} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} n^{\bar{\tau}}} \sim \int_{(K+1)^{\bar{\tau}}}^{+\infty} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} y} dy = e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} (K+1)^{\bar{\tau}}} < e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} K^{\bar{\tau}}},$$

there exists  $\delta \in (0, 1)$  such that

$$\|r\|^2 = (\lambda^*)^2 (K + \delta)^{\bar{\tau}} e^{-2\bar{\eta}\omega_d^{-\bar{\tau}} K^{\bar{\tau}}} \|r\|_{\mathcal{A}_G^{\bar{\eta}, \bar{\tau}}}^2.$$

We conclude that the ratio

$$Q = \frac{e^{\bar{\eta}\omega_d^{-\bar{\tau}} K^{\bar{\tau}}}}{\lambda^* (K + \delta)^{\bar{\tau}/2}}$$

turns out to be arbitrarily large, and indeed for such a residual it is easily seen that Dörfler’s condition  $\|P_{\partial\Lambda} r\| \geq \theta \|r\|$  requires  $|\partial\Lambda|$  to be of the order of  $\theta K$ .  $\square$

Let us now investigate the sparsity classes of the residual, treating the algebraic and exponential cases separately. Note that, in view of Propositions 4.1 or 4.2, for studying the sparsity classes of certain functions  $v$  and  $Lv$  we are entitled to study, equivalently, the sparsity classes of the related vectors  $\mathbf{v}$  and  $\mathbf{A}\mathbf{v}$ , where  $\mathbf{A}$  is the stiffness matrix (2.10).

## 5.1 Algebraic case

We first recall the notion of matrix compressibility (see [7] where the concept has been used in the wavelet context).

**Definition 5.1 (matrix compressibility)** For  $s^* > 0$ , a bounded matrix  $\mathbf{A} : \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\mathbb{Z}^d)$  is called  $s^*$ -compressible if for any  $j \in \mathbb{N}$  there exist constants  $\alpha_j$  and  $C_j$  and a matrix  $\mathbf{A}_j$  having at most  $\alpha_j 2^j$  non-zero entries per column, such that

$$\|\mathbf{A} - \mathbf{A}_j\| \leq C_j$$

where  $\{\alpha_j\}_{j \in \mathbb{N}}$  is summable, and for any  $s < s^*$ ,  $\{C_j 2^{sj/d}\}$  is summable.

Concerning the compressibility of the matrices belonging to the class  $\mathcal{D}_a(\eta_L)$  of Definition 2.1, the following result can be found in [9, Lemma 3.6]. We report here the proof for completeness.

**Lemma 5.1 (compressibility)** If  $s^* := \eta_L - d > 0$ , then any matrix  $\mathbf{A} \in \mathcal{D}_a(\eta_L)$  is  $s^*$ -compressible.

*Proof.* Let us take  $N_j = \lceil \frac{2^{j/d}}{(j+1)^2} \rceil$ , where  $\lceil \cdot \rceil$  denotes the integer part plus 1. Then by Property 2.4 (algebraic case) there holds  $\|\mathbf{A} - \mathbf{A}_{N_j}\| \lesssim 2^{-j(\eta_L-d)/d} (j+1)^{2(\eta_L-d)} =: C_j$  and  $\mathbf{A}_{N_j}$  has  $\alpha_j 2^j$  non-vanishing entries per column with  $\alpha_j \approx 2^d (j+1)^{-2d}$ . It is immediate to verify that  $\sum_j \alpha_j < \infty$ . Moreover, for  $s < s^*$  and setting  $\delta = s^* - s$ , we clearly have  $\sum_j C_j 2^{js/d} = \sum_j 2^{-j\delta/d} (j+1)^{2s^*} < \infty$ .  $\square$

We now consider the continuity properties of the operator  $L$  between sparsity spaces. The following result is well known (see e.g. [10]) and its proof is here reported for completeness.

**Proposition 5.1 (continuity of  $L$  in  $\mathcal{A}_B^s$ )** *Let  $\mathbf{A} \in \mathcal{D}_a(\eta_L)$ ,  $\eta_L > d$  and  $s^* = \eta_L - d$ . For any  $s < s^*$ , if  $v \in \mathcal{A}_B^s$  then  $Lv \in \mathcal{A}_B^s$ , with*

$$\|Lv\|_{\mathcal{A}_B^s} \lesssim \|v\|_{\mathcal{A}_B^s}.$$

*The constants appearing in the bounds go to infinity as  $s$  approaches  $s^*$ .*

*Proof.* Let us choose  $N_j = \lceil \frac{2^{j/d}}{(j+1)^2} \rceil$  as in the proof of Lemma 5.1. If we set  $\mathbf{A}_j := \mathbf{A}_{N_j}$ , then by Property 2.4 (algebraic case) we have

$$\|\mathbf{A} - \mathbf{A}_j\| \lesssim 2^{-j(\eta_L-d)/d} (j+1)^{2(\eta_L-d)} = 2^{-js^*/d} (j+1)^{2s^*}.$$

On the other hand, for any  $j \geq 0$ , let  $\mathbf{v}_j = P_j(\mathbf{v})$  be a best  $2^j$ -term approximation of  $\mathbf{v} \in \ell_B^s$ , which therefore satisfies  $\|\mathbf{v} - \mathbf{v}_j\| \leq 2^{-js/d} \|\mathbf{v}\|_{\ell_B^s}$ . Note that the difference  $\mathbf{v}_j - \mathbf{v}_{j-1}$  satisfies as well

$$\|\mathbf{v}_j - \mathbf{v}_{j-1}\| \lesssim 2^{-js/d} \|\mathbf{v}\|_{\ell_B^s}.$$

Let

$$\mathbf{w}_J = \sum_{j=0}^J \mathbf{A}_{J-j}(\mathbf{v}_j - \mathbf{v}_{j-1}),$$

where we set  $\mathbf{v}_{-1} = \mathbf{0}$ . Writing  $\mathbf{v} = \mathbf{v} - \mathbf{v}_J + \sum_{j=0}^J (\mathbf{v}_j - \mathbf{v}_{j-1})$ , we obtain

$$\mathbf{A}\mathbf{v} - \mathbf{w}_J = \mathbf{A}(\mathbf{v} - \mathbf{v}_J) + \sum_{j=0}^J (\mathbf{A} - \mathbf{A}_{J-j})(\mathbf{v}_j - \mathbf{v}_{j-1}).$$

The last equation yields

$$\begin{aligned} \|\mathbf{A}\mathbf{v} - \mathbf{w}_J\| &\leq \|\mathbf{A}\| \|\mathbf{v} - \mathbf{v}_J\| + \sum_{j=0}^J \|\mathbf{A} - \mathbf{A}_{J-j}\| \|\mathbf{v}_j - \mathbf{v}_{j-1}\| \\ &\lesssim \left( 2^{-Js/d} + \sum_{j=0}^J 2^{-(J-j)s^*/d} (J-j+1)^{2s^*} 2^{-js/d} \right) \|\mathbf{v}\|_{\ell_B^s} \\ &\lesssim 2^{-Js/d} \left( 1 + \sum_{j=0}^J 2^{-(J-j)(s^*-s)/d} (J-j+1)^{2s^*} \right) \|\mathbf{v}\|_{\ell_B^s} \\ &\lesssim 2^{-Js/d} \|\mathbf{v}\|_{\ell_B^s}, \end{aligned}$$

where the series  $\sum_k 2^{-k(s^*-s)/d} (k+1)^{2s^*}$  is convergent but degenerates as  $s$  approaches  $s^*$ . Finally, by construction  $\mathbf{w}_J$  belongs to a finite dimensional space  $V_{\Lambda_J}$ , where

$$|\Lambda_J| \lesssim \omega_d \sum_{j=0}^J N_{J-j}^d \lesssim 2^J \sum_{j=0}^J (J-j+1)^{-2d} \lesssim 2^J.$$

This implies  $\|\mathbf{A}\mathbf{v}\|_{\ell_B^s} \lesssim \|\mathbf{v}\|_{\ell_B^s}$  for any  $s < s^*$ .  $\square$

At last, we discuss the sparsity class of the residual  $r = r(u_\Lambda)$  for some Galerkin solution  $u_\Lambda$ .

**Proposition 5.2 (sparsity class of the residual)** *Let the assumptions of Property 2.2 be satisfied, and set  $s^* = \eta_L - d$ . For any  $s < s^*$ , if  $u \in \mathcal{A}_B^s$  then  $r(u_\Lambda) \in \mathcal{A}_B^s$  for any index set  $\Lambda$ , with*

$$\|r(u_\Lambda)\|_{\mathcal{A}_B^s} \lesssim \|u\|_{\mathcal{A}_B^s}.$$

*Proof.* Denoting by  $\mathbf{r}_\Lambda$  the vector representing  $r(u_\Lambda)$  and using Proposition 5.1, we get

$$\|\mathbf{r}_\Lambda\|_{\ell_B^s} = \|\mathbf{A}(\mathbf{u} - \mathbf{u}_\Lambda)\|_{\ell_B^s} \lesssim \|\mathbf{u} - \mathbf{u}_\Lambda\|_{\ell_B^s} \lesssim \|\mathbf{u}\|_{\ell_B^s} + \|\mathbf{u}_\Lambda\|_{\ell_B^s}. \quad (5.4)$$

At this point, we invoke the equivalent formulation of the Galerkin problem given by (2.24), which yields  $\hat{\mathbf{u}} = (\hat{\mathbf{A}}_\Lambda)^{-1}(\mathbf{P}_\Lambda \mathbf{f})$ . Using  $\mathbf{A} \in \mathcal{D}_a(\eta_L)$  and combining Property 2.5 together with Property 2.2, we obtain  $(\hat{\mathbf{A}}_\Lambda)^{-1} \in \mathcal{D}_a(\eta_L)$ . Hence, applying Proposition 5.1 to  $(\hat{\mathbf{A}}_\Lambda)^{-1}$  we get

$$\|\mathbf{u}_\Lambda\|_{\ell_B^s} = \|\hat{\mathbf{u}}\|_{\ell_B^s} = \|(\hat{\mathbf{A}}_\Lambda)^{-1}(\mathbf{P}_\Lambda \mathbf{f})\|_{\ell_B^s} \lesssim \|\mathbf{P}_\Lambda \mathbf{f}\|_{\ell_B^s} \leq \|\mathbf{f}\|_{\ell_B^s},$$

where the last step is an easy consequence of the definition of the projector  $\mathbf{P}_\Lambda$ . By substituting the above inequality into (5.4), we finally obtain

$$\|\mathbf{r}_\Lambda\|_{\ell_B^s} \lesssim \|\mathbf{u}\|_{\ell_B^s} + \|\mathbf{f}\|_{\ell_B^s} = \|\mathbf{u}\|_{\ell_B^s} + \|\mathbf{A}\mathbf{u}\|_{\ell_B^s} \lesssim \|\mathbf{u}\|_{\ell_B^s}, \quad (5.5)$$

where in the last inequality we used again Proposition 5.1.  $\square$

We observe that the previous bound is tailored to the “worst-scenario”: one expects indeed that for  $\Lambda$  large enough the residual becomes progressively smaller than the solution.

## 5.2 Exponential case

As already alluded to in the Introduction, and in striking contrast to the previous algebraic case, the implication  $v \in \mathcal{A}_G^{\eta,t} \Rightarrow Lv \in \mathcal{A}_G^{\eta,t}$  is false. The following counter-examples prove this fact, and shed light on which could be the correct implication.

**Example 5.3 (Banded matrices)** Fix  $d = 1$  and  $t = 1$  (hence,  $\tau = \frac{t}{d} = 1$ ). Recalling the expression (2.14) for the entries of  $\mathbf{A}$ , let us choose  $\hat{\nu}_0 = \hat{\sigma}_0 = \sqrt{2\pi}$ , which gives

$$a_{\ell,\ell} = 1 \quad \forall \ell \in \mathbb{Z}.$$

Next, let us choose  $\hat{\sigma}_h = 0$  for all  $h \neq 0$ , which implies (because  $d = 1$ )

$$|a_{\ell,k}| = \frac{1}{\sqrt{2\pi}} \frac{|\ell| |k|}{c_\ell c_k} |\hat{\nu}_{\ell-k}|, \quad \ell \neq k,$$



i.e.,

$$\frac{1}{2\sqrt{2\pi}} |\hat{\nu}_{\ell-k}| \leq |a_{\ell,k}| \leq \frac{1}{\sqrt{2\pi}} |\hat{\nu}_{\ell-k}|, \quad \ell \neq k, \quad |\ell|, |k| \geq 1.$$

At this point, let us fix a real  $\eta_L > 0$  and an integer  $p \geq 0$ , and let us choose the coefficients  $\hat{\nu}_h$  for  $h \neq 0$  to satisfy

$$|\hat{\nu}_h| = \begin{cases} \sqrt{2\pi} e^{-\eta_L |h|} & \text{if } 0 < |h| \leq p, \\ 0 & \text{if } |h| > p. \end{cases}$$

In summary, the coefficient  $\nu$  of the elliptic operator  $L$  is a trigonometric polynomial of degree  $p$ , whereas the coefficient  $\sigma$  is a constant. The corresponding stiffness matrix  $\mathbf{A}$  is banded with  $2p+1$  non-zero diagonals, and satisfies

$$\frac{1}{2} e^{-\eta_L |\ell-k|} \leq |a_{\ell,k}| \leq e^{-\eta_L |\ell-k|}, \quad 0 \leq |\ell-k| \leq p, \quad |\ell|, |k| \geq 1. \quad (5.6)$$

In order to define the vector  $\mathbf{v}$ , let us introduce the function  $\iota : \mathbb{N}_* \rightarrow \mathbb{N}_*$ ,  $\iota(n) = 2(p+1)n$ . Let us fix a real  $\eta > 0$  and let us define the components  $(\mathbf{v})_k = \hat{v}_k$  of the vector in such a way that

$$|(\mathbf{v})_k| = \begin{cases} e^{-\frac{\eta}{2}n} & \text{if } k = \iota(n) \text{ for some } n \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the rearranged components  $(\mathbf{v})_n^*$  satisfy  $|(\mathbf{v})_n^*| = e^{-\frac{\eta}{2}n}$ ,  $n \geq 1$ , whence  $\mathbf{v} \in \ell_G^{\eta,1}(\mathbb{Z})$  (or, equivalently,  $v \in \mathcal{A}_G^{\eta,1}$ ), with  $\|\mathbf{v}\|_{\ell_G^{\eta,1}(\mathbb{Z})} = 1$ , according to Definition 4.4.

The definition of the mapping  $\iota$  and the banded structure of  $\mathbf{A}$  imply that the only non-zero components of  $\mathbf{A}\mathbf{v}$  are those of indices  $\iota(n) + q$  for some  $n \geq 1$  and  $q \in [-p, p]$ . For these components one has

$$(\mathbf{A}\mathbf{v})_{\iota(n)+q} = a_{\iota(n)+q, \iota(n)} (\mathbf{v})_{\iota(n)},$$

thus, recalling (5.6), we easily obtain

$$\frac{1}{2} e^{-\eta_L p} e^{-\frac{\eta}{2}n} \leq |(\mathbf{A}\mathbf{v})_{\iota(n)+q}| \leq e^{-\frac{\eta}{2}n}, \quad q \in [-p, p]. \quad (5.7)$$

This shows that, for any integer  $N \geq 1$ ,

$$\#\{\ell : |(\mathbf{A}\mathbf{v})_\ell| \geq \frac{1}{2} e^{-\eta_L p} e^{-\frac{\eta}{2}N}\} \geq (2p+1)N,$$

hence

$$|(\mathbf{A}\mathbf{v})_{(2p+1)N}^*| e^{\frac{\eta}{2}(2p+1)N} \geq \frac{1}{2} e^{-\eta_L p} e^{\eta p N} \rightarrow +\infty \quad \text{as } N \rightarrow +\infty,$$

i.e.,  $\mathbf{A}\mathbf{v} \notin \ell_G^{\eta,1}(\mathbb{Z})$  (or, equivalently,  $Lv \notin \mathcal{A}_G^{\eta,1}$ ) regardless of the relative values of  $\eta_L$  and  $\eta$ .

On the other hand, let  $m_p$  be the smallest integer such that  $\frac{1}{2} e^{-\eta_L p} > e^{-\frac{\eta}{2}m_p}$ . Given any  $m \geq 1$ , let  $N \geq 1$  and  $Q \in [-p, p]$  be such that  $(\mathbf{A}\mathbf{v})_m^* = (\mathbf{A}\mathbf{v})_{\iota(N)+Q}$ , which combined with (5.7) yields

$$e^{-\frac{\eta}{2}(N+m_p)} < |(\mathbf{A}\mathbf{v})_m^*| \leq e^{-\frac{\eta}{2}N}.$$

The rightmost inequality in (5.7), namely  $|(\mathbf{A}\mathbf{v})_{\iota(N+m_p)+q}| \leq e^{-\frac{\eta}{2}(N+m_p)}$ , shows that there are at most  $(2p+1)(N+m_p)$  components of  $\mathbf{A}\mathbf{v}$  that are larger than  $e^{-\frac{\eta}{2}(N+m_p)}$  in modulus. This implies  $m \leq (2p+1)(N+m_p)$ , whence

$$e^{-\frac{\eta}{2}N} \leq e^{\frac{\eta}{2}m_p} e^{-\frac{\eta}{2(2p+1)}m}.$$

Setting  $\bar{\eta} = \frac{\eta}{2p+1}$ , we conclude that  $\mathbf{A}\mathbf{v} \in \ell_G^{\bar{\eta},1}(\mathbb{Z})$  (or, equivalently,  $Lv \in \mathcal{A}_G^{\bar{\eta},1}$ ), with

$$\|\mathbf{A}\mathbf{v}\|_{\ell_G^{\bar{\eta},1}(\mathbb{Z})} \leq e^{\frac{\eta}{2}m_p} \|\mathbf{v}\|_{\ell_G^{\eta,1}(\mathbb{Z})}.$$

Therefore, the sparsity class of  $\mathbf{A}\mathbf{v}$  deteriorates from  $\ell_G^{\eta,1}(\mathbb{Z})$  for  $\mathbf{v}$  to  $\ell_G^{\bar{\eta},1}(\mathbb{Z})$  with  $\bar{\eta} = \frac{\eta}{2p+1}$ .  $\square$

Next counter-example shows that, when the stiffness matrix  $\mathbf{A}$  is not banded, in order to have  $\mathbf{A}\mathbf{v} \in \ell_G^{\bar{\eta},\bar{t}}(\mathbb{Z})$  it is not enough to choose some  $\bar{\eta} < \eta$  as above, but a choice of  $\bar{t} < t$  is mandatory.

**Example 5.4 (Dense matrices)** Let us take again  $d = t = 1$  and modify the setting of the previous example, by assuming now that the coefficients  $\hat{\nu}_h$  satisfy

$$|\hat{\nu}_h| = \sqrt{2\pi}e^{-\eta_L|h|} \quad \text{for all } |h| > 0,$$

so that  $\mathbf{A}$  is no longer banded, and its elements satisfy

$$\frac{1}{2}e^{-\eta_L|\ell-k|} \leq |a_{\ell,k}| \leq e^{-\eta_L|\ell-k|} \quad \text{for all } |\ell|, |k| \geq 1. \quad (5.8)$$

If  $M > 0$  is an arbitrary integer, we now construct a vector  $\mathbf{v}^M = \sum_{n \geq 1} \mathbf{v}^{M,n}$  with gaps of size  $\lambda(M) \geq M$  between consecutive non-vanishing entries. To this end, we introduce the function  $\iota_M : \mathbb{N}_* \rightarrow \mathbb{N}_*$  defined as  $\iota_M(n) := \lambda(M)n$  and the vectors  $\mathbf{v}^{M,n}$  with components

$$|(\mathbf{v}^{M,n})_k| = e^{-\frac{\eta}{2}n} \delta_{k, \iota_M(n)}, \quad k \in \mathbb{Z}.$$

From (5.8) and the fact that only the  $\iota_M(n)$ -th entry of  $\mathbf{v}^{M,n}$  does not vanish, we obtain

$$\frac{1}{2}e^{-\eta_L|\ell-\iota_M(n)|}e^{-\frac{\eta}{2}n} \leq |(\mathbf{A}\mathbf{v}^{M,n})_\ell| \leq e^{-\eta_L|\ell-\iota_M(n)|}e^{-\frac{\eta}{2}n}. \quad (5.9)$$

As in Example 5.3, it is obvious that  $\mathbf{v}^M \in \ell_G^{\eta,1}(\mathbb{Z})$  with  $\|\mathbf{v}^M\|_{\ell_G^{\eta,1}(\mathbb{Z})} = 1$ . However, we will prove below that  $\|\mathbf{A}\mathbf{v}^M\|_{\ell_G^{\bar{\eta},\bar{t}}} \lesssim \|\mathbf{v}^M\|_{\ell_G^{\eta,1}}$  cannot hold uniformly in  $M$  for any  $\bar{\eta} > 0$  and  $\bar{t} > 1/2$ .

We start by examining the cardinality  $\#\mathcal{F}_n$  of the set

$$\mathcal{F}_n := \{\ell \in \mathbb{Z} : |(\mathbf{A}\mathbf{v}^{M,n})_\ell| > e^{-\frac{\eta}{2}M}\}$$

In view of (5.9), the condition  $|(\mathbf{A}\mathbf{v}^{M,n})_\ell| > e^{-\frac{\eta}{2}M}$  is satisfied by those  $\ell = \iota_M(n) + m$  such that

$$0 \leq |m| \leq \frac{\eta}{2\eta_L}(M - n),$$

whence  $n \leq M$  and  $\#\mathcal{F}_n \geq \frac{\eta}{\eta_L}(M - n) + 1$ . We now claim that

$$C_M := \#\{\ell : |(\mathbf{A}\mathbf{v}^M)_\ell| \geq e^{-\frac{\eta}{2}M}\} \geq \sum_{n=1}^M \#\mathcal{F}_n, \quad (5.10)$$

whose proof we postpone. Assuming (5.10) we see that

$$C_M \geq \sum_{n=1}^M \left( \frac{\eta}{\eta_L}(M - n) + 1 \right) \sim \frac{\eta}{2\eta_L}M^2,$$

or equivalently there are about  $N_M = \left\lceil \frac{\eta}{2\eta_L} M^2 \right\rceil$  coefficients of  $\mathbf{v}^M$  with values at least  $e^{-\frac{\eta}{2}M}$ . This implies that the  $N_M$ -th rearranged coefficient of  $\mathbf{A}\mathbf{v}^M$  satisfies

$$|(\mathbf{A}\mathbf{v}^M)_{N_M}^*| \geq e^{-\frac{\eta}{2}M} \geq e^{-\frac{1}{2}(2\eta_L\eta)^{1/2}N_M^{1/2}} \quad \text{for all } M \geq 1.$$

This proves that for any  $\bar{\eta} > 0$  and  $\bar{t} > \frac{1}{2}$ , one has

$$\|\mathbf{A}\mathbf{v}^M\|_{\ell_G^{\bar{\eta}, \bar{t}}(\mathbb{Z})} \geq |(\mathbf{A}\mathbf{v}^M)_{N_M}^*| e^{\frac{\bar{\eta}}{2}N_M^{\bar{t}}} \geq e^{\frac{\bar{\eta}}{2}N_M^{\bar{t}} - \frac{1}{2}(2\eta_L\eta)^{1/2}N_M^{1/2}} \rightarrow +\infty \quad \text{as } M \rightarrow \infty,$$

whence the following bound cannot be valid

$$\|\mathbf{A}\mathbf{v}\|_{\ell_G^{\bar{\eta}, \bar{t}}(\mathbb{Z})} \lesssim \|\mathbf{v}\|_{\ell_G^{\eta, 1}(\mathbb{Z})}, \quad \text{for all } \mathbf{v} \in \ell_G^{\eta, 1}(\mathbb{Z}).$$

It remains to prove (5.10). We first note that the sets  $\mathcal{F}_n$  are disjoint provided  $\iota_M(n+1) - \iota_M(n) = \lambda(M) \geq \frac{\eta}{\eta_L}M$ . We next set

$$\varepsilon_M := \min_{1 \leq n \leq M} \min_{\ell \in \mathcal{F}_n} |(\mathbf{A}\mathbf{v}^{M,n})_\ell| - e^{-\frac{\eta}{2}M} > 0$$

which is a constant only dependent on  $M$ . We observe that for every  $\ell \in \mathcal{F}_n$ , there holds

$$|(\mathbf{A}\mathbf{v}^M)_\ell| \geq |(\mathbf{A}\mathbf{v}^{M,n})_\ell| - \left| \sum_{p \neq n} (\mathbf{A}\mathbf{v}^{M,p})_\ell \right| \geq e^{-\frac{\eta}{2}M} + \varepsilon_M - \sum_{p \neq n} |(\mathbf{A}\mathbf{v}^{M,p})_\ell|. \quad (5.11)$$

We write  $\ell \in \mathcal{F}_n$  as  $\ell = \iota_M(n) + m$ , make use of (5.9) and the definition of  $\iota_M(n) = \lambda(M)n$  to deduce

$$\sum_{p \neq n} |(\mathbf{A}\mathbf{v}^{M,p})_\ell| \leq \sum_{p \neq n} e^{-\eta_L|\ell - \iota_M(p)|} e^{-\frac{\eta}{2}p} \leq \sum_{p \neq n} e^{-\eta_L|m + \lambda(M)(n-p)|} \leq \sum_{p \neq n} e^{-\eta_L(\lambda(M)|n-p| - |m|)}.$$

Since  $|m| \leq \frac{\eta}{2\eta_L}M$ , the above inequality gives

$$\sum_{p \neq n} |(\mathbf{A}\mathbf{v}^{M,p})_\ell| \leq 2e^{\eta_L|m|} \sum_{q \geq 1} e^{-\eta_L\lambda(M)q} \leq 2e^{\frac{\eta}{2}M} \sum_{q \geq 1} e^{-\eta_L\lambda(M)q}. \quad (5.12)$$

Combining (5.11) and (5.12) yields

$$|(\mathbf{A}\mathbf{v}^M)_\ell| \geq e^{-\frac{\eta}{2}M} + \varepsilon_M - 2e^{\frac{\eta}{2}M} \sum_{q \geq 1} e^{-\eta_L\lambda(M)q}.$$

By choosing  $\lambda(M)$  sufficiently large, the last term on the right-hand side of the above inequality can be made arbitrarily small, in particular  $\leq \varepsilon_M$ . We thus get  $|(\mathbf{A}\mathbf{v}^M)_\ell| \geq e^{-\frac{\eta}{2}M}$  and prove (5.10).  $\square$

Guided by Examples 5.3 and 5.4, we are ready to state the main result of this section. We define

$$\zeta(t) := \left( \frac{1+t}{\omega_d^{1+t}} \right)^{\frac{t}{d(1+t)}} \quad \forall 0 < t \leq d. \quad (5.13)$$

**Proposition 5.3 (continuity of  $L$  in  $\mathcal{A}_G^{\eta,t}$ )** *Let the differential operator  $L$  be such that the corresponding stiffness matrix satisfies  $\mathbf{A} \in \mathcal{D}_e(\eta_L)$  for some constant  $\eta_L > 0$ . Assume that  $v \in \mathcal{A}_G^{\eta,t}$  for some  $\eta > 0$  and  $t \in (0, d]$ . Let one of the two following set of conditions be satisfied.*

(a) *If the matrix  $\mathbf{A}$  is banded with  $2p + 1$  non-zero diagonals, let us set*

$$\bar{\eta} = \frac{\eta}{(2p+1)^\tau}, \quad \bar{t} = t.$$

(b) *If the matrix  $\mathbf{A}$  is dense, but the coefficients  $\eta_L$  and  $\eta$  satisfy the inequality  $\eta < \eta_L \omega_d^\tau$ , let us set*

$$\bar{\eta} = \zeta(t)\eta, \quad \bar{t} = \frac{t}{1+t}.$$

*Then, one has  $Lv \in \mathcal{A}_G^{\bar{\eta}, \bar{t}}$ , with*

$$\|Lv\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}} \lesssim \|v\|_{\mathcal{A}_G^{\eta, t}}. \quad (5.14)$$

*Proof.* We adapt to our situation the technique introduced in [7]. Let  $L_J$  ( $J \geq 0$ ) be the differential operator obtained by truncating the Fourier expansion of the coefficients of  $L$  to the modes  $k$  satisfying  $|k| \leq J$ . Equivalently,  $L_J$  is the operator whose stiffness matrix  $\mathbf{A}_J$  is defined in (2.22); thus, by Property 2.4 (exponential case) we have

$$\|L - L_J\| = \|\mathbf{A} - \mathbf{A}_J\| \leq C_{\mathbf{A}}(J+1)^{d-1} e^{-\eta_L J}.$$

On the other hand, for any  $j \geq 1$ , let  $v_j = P_j(v)$  be a best  $j$ -term approximation of  $v$  (with  $v_0 = 0$ ), which therefore satisfies  $\|v - v_j\| \leq e^{-\eta \omega_d^{-\tau} j^\tau} \|v\|_{\mathcal{A}_G^{\eta, t}}$ , with  $\tau = t/d$ . Note that the difference  $v_j - v_{j-1}$  consists of a single Fourier mode and satisfies as well

$$\|v_j - v_{j-1}\| \lesssim e^{-\eta \omega_d^{-\tau} j^\tau} \|v\|_{\mathcal{A}_G^{\eta, t}}.$$

Finally, let us introduce the function  $\chi : \mathbb{N} \rightarrow \mathbb{N}$  defined as  $\chi(j) = \lceil j^\tau \rceil$ , the smallest integer larger than or equal to  $j^\tau$ .

For any  $J \geq 1$ , let  $w_J$  be the approximation of  $Lv$  defined as

$$w_J = \sum_{j=1}^J L_{\chi(J-j)}(v_j - v_{j-1}).$$

Writing  $v = v - v_J + \sum_{j=1}^J (v_j - v_{j-1})$ , we obtain

$$Lv - w_J = L(v - v_J) + \sum_{j=1}^J (L - L_{\chi(J-j)})(v_j - v_{j-1}).$$

We now assume to be in Case (b). Since  $L : \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\mathbb{Z}^d)$  is continuous, the last equation yields

$$\|Lv - w_J\| \lesssim \left( e^{-\eta \omega_d^{-\tau} J^\tau} + \sum_{j=1}^J (\lceil (J-j)^\tau \rceil + 1)^{d-1} e^{-(\eta_L \lceil (J-j)^\tau \rceil + \eta \omega_d^{-\tau} j^\tau)} \right) \|v\|_{\mathcal{A}_G^{\eta, t}}. \quad (5.15)$$

The exponents of the addends can be bounded from below as follows because  $\tau \leq 1$

$$\begin{aligned}\eta_L \lceil (J-j)^\tau \rceil + \eta \omega_d^{-\tau} j^\tau &= \eta_L \lceil (J-j)^\tau \rceil - \eta \omega_d^{-\tau} (J-j)^\tau + \eta \omega_d^{-\tau} ((J-j)^\tau + j^\tau) \\ &\geq \eta_L (J-j)^\tau - \eta \omega_d^{-\tau} (J-j)^\tau + \eta \omega_d^{-\tau} ((J-j) + j)^\tau \\ &= \beta (J-j)^\tau + \eta \omega_d^{-\tau} J^\tau,\end{aligned}$$

with  $\beta = \eta_L - \eta \omega_d^{-\tau} > 0$  by assumption. Then, (5.15) yields

$$\|Lv - w_J\| \lesssim \left(1 + \sum_{j=0}^{J-1} (\lceil j^\tau \rceil + 1)^{d-1} e^{-\beta j^\tau}\right) e^{-\eta \omega_d^{-\tau} J^\tau} \|v\|_{\mathcal{A}_G^{\eta,t}} \lesssim e^{-\eta \omega_d^{-\tau} J^\tau} \|v\|_{\mathcal{A}_G^{\eta,t}}. \quad (5.16)$$

On the other hand, by construction  $w_J$  belongs to a finite dimensional space  $V_{\Lambda_J}$ , where

$$|\Lambda_J| \leq \omega_d \sum_{j=1}^J \chi(J-j)^d = \omega_d \sum_{j=0}^{J-1} \lceil j^\tau \rceil^d \sim \frac{\omega_d}{1+t} J^{1+t} \quad \text{as } J \rightarrow \infty. \quad (5.17)$$

This implies

$$\|Lv - w_J\| \lesssim e^{-\bar{\eta} \omega_d^{-\bar{\tau}} |\Lambda_J|^{\bar{\tau}}} \|v\|_{\mathcal{A}_G^{\eta,t}},$$

with  $\bar{\tau} = \frac{\tau}{1+d\tau} = \frac{t}{d(1+t)}$  and  $\bar{\eta} = \left(\frac{1+d\tau}{\omega_d^{1+d\tau}}\right)^{\bar{\tau}} \eta = \zeta(t)\eta$  as asserted.

We last consider Case (a). One has  $L_{\chi(J-j)} = L$  if  $\chi(J-j) \geq p$ , whence if  $j \leq J - p^{1/\tau}$ , then the summation in (5.15) can be limited to those  $j$  satisfying  $j_p \leq j \leq J$ , where  $j_p = \lceil J - p^{1/\tau} \rceil$ . Therefore

$$\|Lv - w_J\| \lesssim \left( e^{-\eta \omega_d^{-\tau} J^\tau} + \max_{j_p \leq j \leq J} \lceil (J-j)^\tau \rceil^{d-1} \sum_{j=j_p}^J e^{-\eta \omega_d^{-\tau} j^\tau} \right) \|v\|_{\mathcal{A}_G^{\eta,t}}.$$

Now,  $J - j \leq p^{1/\tau}$  if  $j_p \leq j \leq J$  and  $j^\tau \geq j_p^\tau \geq (J - p^{1/\tau})^\tau \geq J^\tau - p$ , whence

$$\|Lv - w_J\| \lesssim \left(1 + p^{d-1+1/\tau} e^{\eta \omega_d^{-\tau} p}\right) e^{-\eta \omega_d^{-\tau} J^\tau} \|v\|_{\mathcal{A}_G^{\eta,t}}.$$

We conclude by observing that  $|\Lambda_J| \leq (2p+1)J$ , since any matrix  $\mathbf{A}_J$  has at most  $2p+1$  diagonals.  $\square$

Finally, we discuss the sparsity class of the residual  $r = r(u_\Lambda)$  for any Galerkin solution  $u_\Lambda$ .

**Proposition 5.4 (sparsity class of the residual)** *Let  $\mathbf{A} \in \mathcal{D}_e(\eta_L)$  and  $\mathbf{A}^{-1} \in \mathcal{D}_e(\bar{\eta}_L)$ , for constants  $\eta_L > 0$  and  $\bar{\eta}_L \in (0, \eta_L]$  according to Property 2.3, and let  $1 \leq d \leq 10$ . If  $u \in \mathcal{A}_G^{\eta,t}$  for some  $\eta > 0$  and  $t \in (0, d]$ , such that  $\eta < \omega_d^{t/(d(1+2t))} \bar{\eta}_L$ , then there exist suitable positive constants  $\bar{\eta} \leq \eta$  and  $\bar{t} \leq t$  such that  $r(u_\Lambda) \in \mathcal{A}_G^{\bar{\eta}, \bar{t}}$  for any index set  $\Lambda$ , with*

$$\|r(u_\Lambda)\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}} \lesssim \|u\|_{\mathcal{A}_G^{\eta,t}}.$$

*Proof.* We first remark that the hypothesis  $1 \leq d \leq 10$  guarantees  $\omega_d \geq 2$  (see e.g. [15, Corollary 2.55]); this implies  $r < \omega_d^r$  for any  $r > 0$ , whence the function  $\zeta$  introduced in (5.13) satisfies  $\zeta(t) < 1$  for any  $t > 0$ . Assume for the moment we are given  $\bar{\eta}$  and  $\bar{t}$ . By using Proposition 5.3 and Lemma 4.1, we get

$$\|\mathbf{r}_\Lambda\|_{\ell_G^{\bar{\eta}, \bar{t}}} = \|\mathbf{A}(\mathbf{u} - \mathbf{u}_\Lambda)\|_{\ell_G^{\bar{\eta}, \bar{t}}} \lesssim \|\mathbf{u} - \mathbf{u}_\Lambda\|_{\ell_G^{\eta_1, t_1}} \lesssim \|\mathbf{u}\|_{\ell_G^{2\tau_1 \eta_1, t_1}} + \|\mathbf{u}_\Lambda\|_{\ell_G^{2\tau_1 \eta_1, t_1}}, \quad (5.18)$$

where,  $\bar{\tau} = \bar{t}/d$ ,  $\tau_1 = t_1/d$  and the following relations hold

$$\bar{\eta} = \zeta(t_1)\eta_1, \quad \bar{t} = \frac{t_1}{1+t_1} < t_1.$$

From (2.24) we have  $\mathbf{u}_\Lambda = (\hat{\mathbf{A}}_\Lambda)^{-1}(\mathbf{P}_\Lambda \mathbf{f})$ . Using Property 2.5 and applying Proposition 5.3 to  $(\hat{\mathbf{A}}_\Lambda)^{-1}$  we get

$$\|\mathbf{u}_\Lambda\|_{\ell_G^{2\tau_1 \eta_1, t_1}} = \|(\hat{\mathbf{A}}_\Lambda)^{-1}(\mathbf{P}_\Lambda \mathbf{f})\|_{\ell_G^{2\tau_1 \eta_1, t_1}} \lesssim \|\mathbf{P}_\Lambda \mathbf{f}\|_{\ell_G^{\eta_2, t_2}} \leq \|\mathbf{f}\|_{\ell_G^{\eta_2, t_2}},$$

with

$$2^{\tau_1} \eta_1 = \zeta(t_2)\eta_2 < \eta_2, \quad t_1 = \frac{t_2}{1+t_2} < t_2.$$

By substituting the above inequality into (5.18) and using again Proposition 5.3 we get

$$\|\mathbf{r}_\Lambda\|_{\ell_G^{\bar{\eta}, \bar{t}}} \lesssim \|\mathbf{u}\|_{\ell_G^{2\tau_1 \eta_1, t_1}} + \|\mathbf{f}\|_{\ell_G^{\eta_2, t_2}} = \|\mathbf{u}\|_{\ell_G^{2\tau_1 \eta_1, t_1}} + \|\mathbf{A}\mathbf{u}\|_{\ell_G^{\eta_2, t_2}} \lesssim \|\mathbf{u}\|_{\ell_G^{\eta, t}} \quad (5.19)$$

where

$$\eta_2 = \zeta(t)\eta < \eta, \quad t_2 = \frac{t}{1+t} < t.$$

This shows that the thesis holds true for the choice

$$\bar{\eta} = \left(\frac{1}{2}\right)^{\frac{t}{d(1+2t)}} \zeta\left(\frac{t}{1+2t}\right) \zeta\left(\frac{t}{1+t}\right) \zeta(t)\eta, \quad \bar{t} = \frac{t}{1+3t}.$$

It remains to verify the assumptions of Proposition 5.3 when  $\mathbf{A}$  is dense. Since  $\omega_d \geq 2$  and

$$t_1 = \frac{t}{1+2t} < t_2 = \frac{t}{1+t} < t,$$

we have  $\omega_d^{\tau_1} < \omega_d^{\tau_2} < \omega_d^\tau$ . Moreover, using  $\eta_1 < 2^{\tau_1} \eta_1 < \eta_2 < \eta$  and  $\eta_L \geq \bar{\eta}_L > \omega_d^{-\tau_1} \eta$  yields

$$\eta < \omega_d^\tau \eta_L, \quad \eta_1 < \omega_d^{\tau_1} \eta_L, \quad \eta_2 < \omega_d^{\tau_2} \bar{\eta}_L,$$

which are the required conditions to apply Proposition 5.3 when  $\mathbf{A}$  is dense. This concludes the proof.  $\square$

**Remark 5.1 (definition of  $\omega_d$ )** The limitation  $1 \leq d \leq 10$  stems from the fact that the measure of the unit Euclidean ball  $\omega_d$  in  $\mathbb{R}^d$  monotonically decreases to 0 as  $d \rightarrow \infty$ . To avoid such a restriction, one could modify the definition of the Gevrey classes  $G_p^{\eta, t}(\Omega)$  given in (4.7), by replacing the Euclidean norm  $|k| = \|k\|_2$  appearing in the exponential by the maximum norm  $\|k\|_\infty$ . Consequently, throughout the rest of the paper  $\omega_d$  would be replaced by the quantity  $2^d$ , strictly larger than 1 for any  $d$ .  $\square$

## 6 Coarsening

We start by considering an example that sheds light on the role of coarsening for the exponential case. We then state and prove a seemingly new coarsening result, which is valid for both classes.

### 6.1 Example of coarsening

Let  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$  for  $p \geq 1$  be the vectors

$$\mathbf{a} := (1, 0, \dots, 0), \quad \mathbf{b} := \frac{1}{p}(1, 1, \dots, 1).$$

Let  $\mathbf{v}, \mathbf{z}$  be the sequences defined by

$$\mathbf{v} := (e^{-\eta k} \mathbf{a})_{k=0}^\infty, \quad \mathbf{z} := (e^{-\eta k} \mathbf{b})_{k=0}^\infty.$$

We first observe that

$$\|\mathbf{v}\|^2 = p \|\mathbf{z}\|^2 = \frac{1}{1 - e^{2\eta}}, \quad \|\mathbf{v}\|_{\ell_G^{2\eta,1}(\mathbb{Z})} = p \|\mathbf{z}\|_{\ell_G^{2\eta/p,1}(\mathbb{Z})} = 1$$

(recall that  $\omega_d = 2$  for  $d = 1$ ). Given a parameter  $\varepsilon < 1$ , we now construct a perturbation  $\mathbf{w}$  of  $\mathbf{v}$  which is much less sparse than  $\mathbf{v}$  by simply scaling  $\mathbf{z}$  and adding it to  $\mathbf{v}$  (see Fig. 2 (a)):

$$\mathbf{w} := \mathbf{v} + \varepsilon \mathbf{z} = (e^{-\eta k} (\mathbf{a} + \varepsilon \mathbf{b}))_{k=1}^\infty.$$

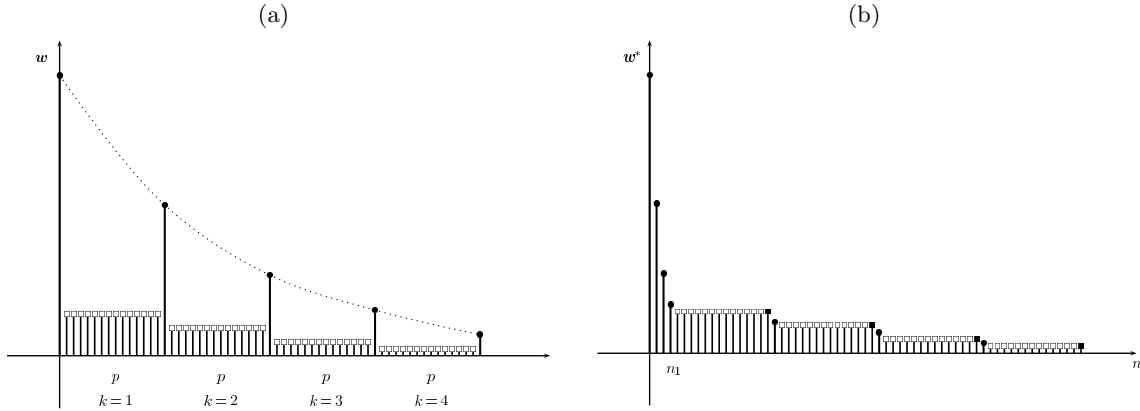


Figure 2: Pictorial representation of (a) the components of the vector  $\mathbf{w} = \mathbf{v} + \varepsilon \mathbf{z}$  and (b) its rearrangement  $\mathbf{w}^*$ . It turns out that  $\mathbf{w}^*$  exhibits the decay rate  $e^{-k\eta}$  of  $\mathbf{v}$  up to a level of accuracy  $\|\mathbf{w} - \mathbf{v}\|$  in  $\ell^2(\mathbb{Z})$  but a worse decay rate  $e^{-k\frac{\eta}{p}}$  of  $\mathbf{z}$  for smaller tolerances. Therefore, truncating  $\mathbf{w}^*$  with a threshold  $\delta \geq \|\mathbf{w} - \mathbf{v}\|$  captures the behavior of  $\mathbf{v}$ .

The first task is to compute the norms of  $\mathbf{w}$ . We obviously have  $\|\mathbf{w}\| \simeq \|\mathbf{v}\|$ . To determine the weak quasi-norm of  $\mathbf{w}$  we need to find the rearrangement  $\mathbf{w}^*$  (see Fig. 2 (b)). Let  $n_1$  be the smallest integer such that

$$\left(1 + \frac{\varepsilon}{p}\right) e^{-\eta n_1} \geq \frac{\varepsilon}{p} e^{-\eta} > \left(1 + \frac{\varepsilon}{p}\right) e^{-\eta(n_1+1)},$$

namely the index corresponding to the first crossing of the exponential curve  $e^{-\eta m}$  dictating the behavior of the first portion of the rearranged sequence  $\mathbf{w}^*$  (which coincides with the behavior of  $\mathbf{v}^*$ ), and the first plateau of  $\mathbf{z}$ . This implies

$$\frac{1}{\eta} \log \left( 1 + \frac{p}{\varepsilon} \right) < n_1 \leq 1 + \frac{1}{\eta} \log \left( 1 + \frac{p}{\varepsilon} \right) .$$

Next, let  $n_2$  be the smallest integer such that

$$\left( 1 + \frac{\varepsilon}{p} \right) e^{-\eta n_2} \geq \frac{\varepsilon}{p} e^{-2\eta} > \left( 1 + \frac{\varepsilon}{p} \right) e^{-\eta(n_2+1)} ,$$

which corresponds to the beginning of a number of decreasing exponentials preceeding the second plateau of  $\mathbf{w}^*$ . This implies

$$1 + \frac{1}{\eta} \log \left( 1 + \frac{p}{\varepsilon} \right) < n_2 \leq 2 + \frac{1}{\eta} \log \left( 1 + \frac{p}{\varepsilon} \right)$$

and shows that  $n_2 - n_1 = 1$ , and that there is exactly one exponential between the first and second plateaux. Iterating this argument, we see that the difference between two consecutive  $n_j$ 's is just 1, and that there is exactly one exponential between two consecutive plateaux (see Fig 2 (b)).

We are now ready to compute the weak quasi-norm of  $\mathbf{w}$ . Let  $\nu_k$  denote the index corresponding to the end of the  $k$ -th plateau of  $\mathbf{w}$ , which in turn corresponds to the value  $w_{\nu_k}^* = e^{-\eta k}$ . Then

$$\nu_k = pk + n_1 \sim pk + \frac{1}{\eta} \log \left( 1 + \frac{p}{\varepsilon} \right) .$$

To determine the class of  $\mathbf{w}$ , we seek  $\lambda$  so that  $\mathbf{w} \in \ell_G^{\lambda,1}(\mathbb{Z})$ , namely

$$\sup_{k \geq 0} \left( e^{\lambda \nu_k / 2} e^{-\eta k} \right) < \infty \quad \Leftrightarrow \quad \frac{1}{2} \lambda pk - \eta k \leq 0 \quad \Leftrightarrow \quad \lambda \leq \frac{2\eta}{p} .$$

We thus realize that  $\mathbf{w} \in \ell_G^{2\eta/p,1}(\mathbb{Z})$  belongs to a sparsity class much worse than that of  $\mathbf{v}$ , that deteriorates as the size  $p$  of the plateaux tends to  $\infty$ . On the other hand, we note that the restrictions  $\mathbf{w}_{[1,n_1]}^* = \mathbf{v}_{[1,n_1]}^*$  coincide, thereby showing that the decay rate of the first part of  $\mathbf{w}^*$  is the same as that of  $\mathbf{v}^*$  (see Fig 2(b)). This example explains the need to coarsen the vector  $\mathbf{w}$  starting at latest at  $n_1$ , to eliminate the tail of  $\mathbf{w}^*$  which decays with rate  $2\eta/p$  instead of the optimal rate  $2\eta$  of  $\mathbf{v}$ .

In addition, we observe that the best  $n_1$ -term approximation of  $\mathbf{w}$  satisfies

$$\|\mathbf{w} - \mathbf{w}_{n_1}\|^2 = \sum_{k=0}^{\infty} p \frac{\varepsilon^2}{p^2} e^{-2k\eta} = \frac{\varepsilon^2}{p} \frac{1}{1 - e^{-2\eta}} = \|\mathbf{v} - \mathbf{w}\|^2 = \varepsilon^2 \|\mathbf{z}\|^2 ,$$

which is precisely the size of the perturbation error of  $\mathbf{v}$ . Given an error tolerance  $\delta \geq \varepsilon \|\mathbf{z}\|$ , the best  $N$ -term approximation  $\mathbf{w}_N$  of  $\mathbf{w}$  satisfying  $\|\mathbf{w} - \mathbf{w}_N\| \leq \delta$  would require

$$N \sim \frac{1}{\eta} \log \frac{1}{\delta} = \frac{2}{2\eta} \log \frac{\|\mathbf{v}\|_{\ell_G^{2\eta,1}(\mathbb{Z})}}{\delta} .$$



## 6.2 New coarsening Result

We extract the following lesson from the example of Sect. 6.1: for as long as we deal with the first part of  $\mathbf{w}^*$ , which has a decay rate  $e^{-k\eta}$  dictated by that of  $\mathbf{v}^*$ , we could coarsen  $\mathbf{w}$  and obtain an approximation of both  $\mathbf{w}$  and  $\mathbf{v}$  with the decay rate  $e^{-k\eta}$  of  $\mathbf{v}$ . This requires limiting the accuracy to size  $\|\mathbf{v} - \mathbf{w}\|$  since a smaller accuracy utilizes the tail of  $\mathbf{w}$  which has a slower decay  $e^{-k\frac{\eta}{p}}$ .

We express this heuristics in the following theorem, which goes back to Cohen, Dahmen, and DeVore [7]. However, our proof is much more elementary and the statement much more precise. Although the result holds for the general setting of Sect. 4.1, we just present it for the exponential case, since it will be used only in this situation.

**Theorem 6.1 (coarsening)** *Let  $\varepsilon > 0$  and let  $v \in \mathcal{A}_G^{\eta,t}$  and  $w \in V$  be so that*

$$\|v - w\| \leq \varepsilon.$$

*Let  $N = N(\varepsilon)$  be the smallest integer such that the best  $N$ -term approximation  $w_N$  of  $w$  satisfies*

$$\|w - w_N\| \leq 2\varepsilon.$$

*Then,  $\|v - w_N\| \leq 3\varepsilon$  and*

$$N \leq \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|v\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon} \right)^{d/t} + 1.$$

*Proof.* Let  $\Lambda_\varepsilon$  be the set of indices corresponding to the best approximation of  $v$  with accuracy  $\varepsilon$ . So  $\Lambda_\varepsilon$  is a minimal set with properties

$$\|v - P_{\Lambda_\varepsilon} v\| \leq \varepsilon, \quad |\Lambda_\varepsilon| \leq \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|v\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon} \right)^{d/t} + 1.$$

If  $z = w - v$ , then

$$\begin{aligned} \|w - P_{\Lambda_\varepsilon} w\| &\leq \|(v + z) - P_{\Lambda_\varepsilon}(v + z)\| = \|(v - P_{\Lambda_\varepsilon} v) + (z - P_{\Lambda_\varepsilon} z)\| \\ &\leq \|v - P_{\Lambda_\varepsilon} v\| + \|z - P_{\Lambda_\varepsilon} z\| \leq \varepsilon + \|z\| \leq 2\varepsilon, \end{aligned}$$

because  $I - P_{\Lambda_\varepsilon}$  is the projector onto  $V_{\mathbb{Z}^d \setminus \Lambda_\varepsilon}$ . Since  $N$  is the cardinality of the smallest set satisfying the above relation, we deduce that  $N \leq |\Lambda_\varepsilon|$ . This concludes the proof.  $\square$

## 7 Optimality properties of adaptive algorithms: algebraic case

The rest of the paper will be devoted to investigating complexity issues for the sequence of approximations  $u_n = u_{\Lambda_n}$  generated by any of the adaptive algorithms presented in Sect. 3. In particular, we wish to estimate the cardinality of each  $\Lambda_n$  and check whether its growth is “optimal” with respect to the sparsity class  $\mathcal{A}_\phi$  of the exact solution, in the sense that  $|\Lambda_n|$  is comparable to the cardinality of the index set of the best approximation of  $u$  yielding the same error  $\|u - u_n\|$ .

The algebraic case will be dealt with in the present section, whereas the exponential case will be analyzed in the next one. The two cases differ in that no coarsening is needed for optimality

in the former case, whereas we will prove optimality in the latter case only for the algorithms that incorporate a coarsening step. The reason of such a difference can be attributed, on the one hand, to the slower growth of the activated degrees of freedom in the exponential case as opposed to the algebraic case and, on the other hand, to the discrepancy in the sparsity classes of the residuals and the solution in the exponential case, discussed in Sect. 5.2.

### 7.1 ADFOUR with moderate Dörfler marking

The approach followed in the sequel, which has been proposed in [16] in the wavelet framework and adopted in [21, 6] in the finite-element framework, allows us to prove the optimality of the algorithm in the algebraic case, provided Dörfler marking is not too aggressive.

The two following lemmas will be useful in the subsequent analysis.

**Lemma 7.1 (localized a posteriori upper bound)** *Let  $\Lambda \subset \Lambda_* \subset \mathbb{Z}^d$  be nonempty subsets of indices. Let  $u_\Lambda \in V_\Lambda$  and  $u_{\Lambda_*} \in V_{\Lambda_*}$  be the Galerkin approximations of Problem (2.4). Then*

$$\|u_{\Lambda_*} - u_\Lambda\|^2 \leq \frac{1}{\alpha_*} \sum_{k \in \Lambda_* \setminus \Lambda} |\hat{R}_k(u_\Lambda)|^2 = \frac{1}{\alpha_*} \eta^2(u_\Lambda, \Lambda_*).$$

*Proof.* One has

$$\|u_{\Lambda_*} - u_\Lambda\|^2 = a(u_{\Lambda_*} - u_\Lambda, u_{\Lambda_*} - u_\Lambda) = (f, u_{\Lambda_*} - u_\Lambda) - a(u_\Lambda, u_{\Lambda_*} - u_\Lambda) = \sum_{k \in \Lambda_*} \hat{r}_k(u_\Lambda)(\hat{u}_{\Lambda_*} - \hat{u}_\Lambda)_k$$

because  $\Lambda_*$  is the support of  $u_{\Lambda_*} - u_\Lambda$ . The asserted result follows immediately by the Cauchy-Schwarz inequality, upon recalling that  $\hat{r}_k(u_\Lambda) = 0$  for all  $k \in \Lambda$ .  $\square$

**Lemma 7.2 (Dörfler property)** *Let  $\Lambda \subset \Lambda_* \subset \mathbb{Z}^d$  be nonempty subsets of indices. Let  $u_\Lambda \in V_\Lambda$  and  $u_{\Lambda_*} \in V_{\Lambda_*}$  be the Galerkin approximations of Problem (2.4). Let the marking parameter  $\theta$  satisfies  $\theta \in (0, \theta_*)$ , where  $\theta_* = \sqrt{\frac{\alpha_*}{\alpha^*}}$ , and set  $\mu_\theta = 1 - \frac{\alpha^*}{\alpha_*} \theta^2 > 0$ . If*

$$\|u - u_{\Lambda_*}\|^2 \leq \mu \|u - u_\Lambda\|^2,$$

*for some  $\mu \in (0, \mu_\theta]$ , then  $\Lambda^*$  fulfils Dörfler's condition, i.e.,*

$$\eta(u_\Lambda, \Lambda^*) \geq \theta \eta(u_\Lambda).$$

*Proof.* Since  $u - u_{\Lambda_*} \perp u_\Lambda - u_{\Lambda_*}$  in the energy norm because of Pythagoras, the assumption yields

$$\|u - u_\Lambda\|^2 = \|u - u_{\Lambda_*}\|^2 + \|u_{\Lambda_*} - u_\Lambda\|^2 \leq \mu \|u - u_\Lambda\|^2 + \|u_{\Lambda_*} - u_\Lambda\|^2.$$

Invoking the lower bound in (3.2) gives

$$\|u_{\Lambda_*} - u_\Lambda\|^2 \geq (1 - \mu) \|u - u_\Lambda\|^2 \geq (1 - \mu) \frac{1}{\alpha^*} \eta^2(u_\Lambda),$$

whence applying Lemma 7.1 implies

$$\eta^2(u_\Lambda, \Lambda_*) \geq (1 - \mu) \frac{\alpha_*}{\alpha^*} \eta^2(u_\Lambda) \geq (1 - \mu_\theta) \frac{\alpha_*}{\alpha^*} \eta^2(u_\Lambda) = \theta^2 \eta^2(u_\Lambda).$$

This concludes the proof.  $\square$

We are ready to estimate the growth of degrees of freedom generated by the algorithm **ADFOUR** of Sect. 3.1. For the moment, we place ourselves in the abstract framework of Sect. 4.1, only the final result being specifically for the algebraic case.

**Proposition 7.1 (cardinality of  $\partial\Lambda_n$ )** *Let  $\theta$  satisfy the condition stated in Lemma 7.2, and let  $\mu \in (0, \mu_\theta]$  be fixed. Let  $\{\Lambda_n, u_n\}_{n \geq 0}$  be the sequence generated by the adaptive algorithm **ADFOUR**, and set  $\varepsilon_n^2 = \mu \|u - u_n\|^2$ . If the solution  $u$  belongs to the sparsity class  $\mathcal{A}_\phi$ , then*

$$|\partial\Lambda_n| = |\Lambda_{n+1}| - |\Lambda_n| \leq \kappa \phi^{-1} \left( \frac{\varepsilon_n}{\|u\|_{\mathcal{A}_\phi}} \right), \quad \forall n \geq 0, \quad (7.1)$$

where  $\kappa > 1$  is the constant in (4.4).

*Proof.* Let  $\varepsilon = \varepsilon_n$  and make use of (4.4) for  $u \in \mathcal{A}_\phi$ : there exists  $\Lambda_\varepsilon$  and  $w_\varepsilon \in V_{\Lambda_\varepsilon}$  such that

$$\|u - w_\varepsilon\|^2 \leq \varepsilon^2 \quad \text{and} \quad |\Lambda_\varepsilon| \leq \kappa \phi^{-1} \left( \frac{\varepsilon}{\|u\|_{\mathcal{A}_\phi}} \right).$$

Let  $\Lambda_* = \Lambda_n \cup \Lambda_\varepsilon$  be the overlay of the two index sets, and let  $u_* \in V_{\Lambda_*}$  be the Galerkin approximation of Problem (2.4). Then, since  $V_{\Lambda_\varepsilon} \subseteq V_{\Lambda_*}$ , we have

$$\|u - u_*\|^2 \leq \|u - w_\varepsilon\|^2 \leq \mu \|u - u_n\|^2.$$

Thus, we are entitled to apply Lemma 7.2 to  $\Lambda_n$  and  $\Lambda_*$ , yielding

$$\eta(u_n, \Lambda^*) \geq \theta \eta(u_n).$$

By the minimality property of the cardinality of  $\Lambda_{n+1}$  among all sets satisfying Dörfler property for  $u_n$  (Assumption 3.1), we deduce that  $|\Lambda_{n+1}| \leq |\Lambda_*| \leq |\Lambda_n| + |\Lambda_\varepsilon|$ , i.e.,

$$|\Lambda_{n+1}| - |\Lambda_n| \leq |\Lambda_\varepsilon|, \quad (7.2)$$

whence the result.  $\square$

**Corollary 7.1 (cardinality of  $\Lambda_n$ : general case)** *Let the assumptions of Proposition 7.1 be valid and  $\rho = \sqrt{1 - \frac{\alpha_*}{\alpha^*} \theta^2}$  be given by (3.7). Then*

$$|\Lambda_n| \leq \kappa \sum_{k=0}^{n-1} \phi^{-1} \left( \rho^{k-n} \frac{\varepsilon_n}{\|u\|_{\mathcal{A}_\phi}} \right), \quad \forall n \geq 0. \quad (7.3)$$

*Proof.* Recalling that  $|\Lambda_0| = 0$ , the previous proposition yields

$$|\Lambda_n| = \sum_{k=0}^{n-1} |\partial\Lambda_k| \leq \kappa \sum_{k=0}^{n-1} \phi^{-1} \left( \frac{\varepsilon_k}{\|u\|_{\mathcal{A}_\phi}} \right).$$

On the other hand, by Theorem 3.1 one has

$$\varepsilon_n = \sqrt{\mu} \|u - u_n\| \leq \sqrt{\mu} \rho^{n-k} \|u - u_k\| = \rho^{n-k} \varepsilon_k \quad \forall 0 \leq k \leq n-1, \quad (7.4)$$

and we conclude recalling the monotonicity of  $\phi$ .  $\square$

At this point, we assume to be in the algebraic case, i.e.  $u \in \mathcal{A}_B^s$  for some  $s > 0$ . Then, (7.3) reads

$$|\Lambda_n| \leq \kappa \mu^{-d/2s} \|u - u_n\|^{-d/s} \|u\|_{\mathcal{A}_B^s}^{d/s} \sum_{k=0}^{n-1} \left( \rho^{d/s} \right)^{n-k}, \quad \forall n \geq 0.$$

Summing-up the geometric series and using (2.5), we arrive at the following result.

**Theorem 7.1 (cardinality of  $\Lambda_n$ : algebraic case)** *Under the assumptions of Proposition 7.1, the growth of the active degrees of freedom produced by **ADFOUR** in the algebraic case is estimated as follows:*

$$|\Lambda_n| \leq C_* \|u - u_n\|^{-d/s} \|u\|_{\mathcal{A}_B^s}^{d/s}, \quad \forall n \geq 0,$$

where the constant  $C_*$  depends only on  $\alpha_*$ ,  $\mu$  and  $\rho$ . □

This result is “optimal” in that the number of active degrees of freedom is governed, up to a multiplicative constant, by the same law (4.4)-(4.5) as for the best approximation of  $u$ . The optimality of this result is related to the “sufficiently fast” growth of the active degrees of freedom: the increment of degrees of freedom at each iteration may be comparable to the total number of previously activated degrees of freedom (geometric growth).

## 7.2 A-ADFOUR: Aggressive ADFOUR

We now examine Algorithm **A-ADFOUR**, defined in Sect. 3.3, which allows the choice of the parameter  $\theta$  as close to 1 as desired. Such a feature is in the spirit of high regularity, or equivalently a large value of  $s$  for  $u \in \mathcal{A}_B^s$ . This is a novel approach which combines the contraction property in Theorem 3.3 and the key property of uniform boundedness of the residuals stated in Proposition 5.2.

**Theorem 7.2 (cardinality of  $\Lambda_n$  for A-ADFOUR)** *Let the assumptions of Property 2.2 and Theorem 3.3 be fulfilled, and let  $u \in \mathcal{A}_B^s$  for some  $s > 0$ . Then, the growth of the active degrees of freedom produced by **A-ADFOUR** is estimated as follows:*

$$|\Lambda_n| \leq C_* J^d \|u - u_n\|^{-d/s} \|u\|_{\mathcal{A}_B^s}^{d/s}, \quad \forall n \geq 0.$$

Here,  $J$  is the ( $\theta$ -dependent) input parameter of **ENRICH**, whereas the constant  $C_*$  is independent of  $\theta$ .

*Proof.* At each iteration  $n$ , the set  $\widetilde{\partial\Lambda}_n$  selected by **DÖRFLER** is minimal, hence by (3.4), (4.3) and (4.6), one has

$$|\widetilde{\partial\Lambda}_n| \leq \left( \sqrt{1 - \theta^2} \|r_n\| \right)^{-d/s} \|r_n\|_{\mathcal{A}_B^s}^{d/s} + 1.$$

Using (2.9) and Proposition 5.2, this bound becomes

$$|\widetilde{\partial\Lambda}_n| \lesssim \left( \sqrt{1 - \theta^2} \|u - u_n\| \right)^{-d/s} \|u\|_{\mathcal{A}_B^s}^{d/s}.$$

On the other hand, estimate (3.18) for the procedure **ENRICH** yields

$$|\partial\Lambda_n| \lesssim J^d \left( \sqrt{1 - \theta^2} \|u - u_n\| \right)^{-d/s} \|u\|_{\mathcal{A}_B^s}^{d/s}.$$

Now, as in the proof of Corollary 7.1,

$$|\Lambda_n| \lesssim J^d (1 - \theta^2)^{-d/s} \left( \sum_{k=0}^{n-1} \|u - u_k\|^{-d/s} \right) \|u\|_{\mathcal{A}_B^s}^{d/s}. \quad (7.5)$$

The contraction property of Theorem 3.3 yields for  $0 \leq k \leq n-1$

$$\|u - u_n\| \leq \bar{\rho}^{n-k} \|u - u_k\| ,$$

with  $\bar{\rho} = C_0 \sqrt{1 - \theta^2} < 1$  (see 3.17); thus,

$$\sum_{k=0}^{n-1} \|u - u_k\|^{-d/s} \leq \|u - u_n\|^{-d/s} \sum_{k=0}^{n-1} \bar{\rho}^{\frac{d}{s}(n-k)} \lesssim \bar{\rho}^{\frac{d}{s}} \|u - u_n\|^{-d/s} \lesssim (1 - \theta^2)^{d/s} \|u - u_n\|^{-d/s} .$$

Substituting into (7.5), the powers of  $1 - \theta^2$  cancel out, and the asserted estimate follows.  $\square$

## 8 Optimality properties of adaptive algorithms: exponential case

From now on, let us assume that  $u \in \mathcal{A}_G^{\eta,t}$  for some  $\eta > 0$  and  $t \in (0, d]$ . Let us first observe that none of the arguments that led to the complexity estimates of the previous section can be extended to the present situation.

For **ADFOUR** with moderate Dörfler marking, Corollary 7.1 in which  $\phi^{-1}$  is replaced by its logarithmic expression yields a bound for  $|\Lambda_n|$  which is at least  $n$  times larger than the optimal bound

$$|\Lambda_n^{\text{best}}| \leq \kappa \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon_n} \right)^{d/t}$$

for the given accuracy  $\varepsilon_n$  (see the proof of Proposition 8.1 for more details, in a similar situation). Manifestedly, the first cause of non-optimality is the crude bound (7.2), which in this case is no longer absorbed by the summation of a geometric series as in the algebraic case.

On the other hand, for **A-ADFOUR** a sharp estimate of the increment  $|\widetilde{\partial\Lambda}_n|$  is indeed used in the proof of Theorem 7.2, but this involves the sparsity class of the residual, which in the exponential case may be different from that of the solution, as discussed in Sect. 5.2.

Incorporating a coarsening step in the algorithms allows us to avoid, at least in part, these drawbacks. For these reasons, hereafter we investigate the optimality properties of the two algorithms with coarsening presented in Sect. 3

### 8.1 C-ADFOUR: ADFOUR with coarsening

Let us now discuss the complexity of Algorithm **C-ADFOUR**, defined in Sect. 3.4. The following optimal result holds.

**Theorem 8.1 (cardinality of  $\Lambda_n$  for C-ADFOUR)** *Assume that the solution  $u$  belongs to  $\mathcal{A}_G^{\eta,t}$ , for some  $\eta > 0$  and  $t \in (0, d]$ . Then, there exists a constant  $C > 1$  such that the cardinality of the set  $\Lambda_n$  of the active degrees of freedom produced by **C-ADFOUR** satisfies the bound*

$$|\Lambda_n| \leq \kappa \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|u - u_n\|} + \log C \right)^{d/t} , \quad \forall n \geq 0. \quad (8.1)$$

*Proof.* Since each Galerkin approximation  $u_{n+1}$  comes just after a call  $\Lambda_{n+1} := \mathbf{COARSE}(u_{n,k+1}, \varepsilon_n)$  with threshold  $\varepsilon_n = \alpha_*^{-1/2} \|r_{n,k+1}\| \geq \|u - u_{n,k+1}\|$ , Theorem 6.1 yields

$$|\Lambda_{n+1}| \leq \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon_n} \right)^{d/t} + 1.$$

On the other hand, (2.5) and Property 3.1 yield

$$\|u - u_{n+1}\| \leq \alpha_*^{-1/2} \|u - u_{n+1}\| \leq 3(\alpha^*/\alpha_*)^{1/2} \varepsilon_n. \quad (8.2)$$

Since  $n \geq -1$ , this gives the result, up to a shift in the index.  $\square$

Next, we investigate the optimality of each inner loop. We already know from Theorem 3.4 that the number  $K_n$  of inner iterations is bounded independently of  $n$ . So, we just estimate the growth of degrees of freedom when going from  $k$  to  $k+1$ . We only consider the case of a moderate Dörfler marking, i.e., we subject  $\theta$  to the condition stated in Lemma 7.2 (since the case of  $\theta$  close to 1 will be covered in the next subsection). The following result holds.

**Proposition 8.1 (cardinality of  $\Lambda_{n,k}$  for C-ADFOUR)** *Assume that  $u \in \mathcal{A}_G^{\eta,t}$  for some  $\eta > 0$  and  $t \in (0, d]$ , and that the marking parameter satisfies  $\theta \in (0, \theta_*)$ , where  $\theta_* = \sqrt{\frac{\alpha_*}{\alpha^*}}$ . Then, there exist constants  $C > 1$  and  $\bar{\eta} \in (0, \eta]$  such that, for all  $n \geq 0$  and all  $k = 1, \dots, K_n$ , one has*

$$|\Lambda_{n,k}| \leq \kappa \frac{\omega_d}{\bar{\eta}^{d/t}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|u - u_{n+1}\|} + \log C \right)^{d/t}.$$

*Proof.* Each inner loop of **C-ADFOUR** can be viewed as a truncated version of **ADFOUR**; hence, the analysis of this algorithm given in Sect. 7.1 can be adapted to the exponential case. In particular, for each increment  $\partial\Lambda_{n,j}$  of degrees of freedom, Proposition 7.1 gives

$$|\partial\Lambda_{n,j}| \leq \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon_{n,j}} \right)^{d/t} + 1, \quad \forall 0 \leq j \leq K_n.$$

Since,  $\varepsilon_{n,K_n} \leq \rho^{K_n-j} \varepsilon_{n,j}$  by (7.4), it follows that

$$|\partial\Lambda_{n,j}| \leq \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon_{n,K_n}} + (K_n - j) |\log \rho| \right)^{d/t} + 1.$$

Thus, recalling that  $t \leq d$  by assumption, we have

$$\begin{aligned} |\Lambda_{n,k}|^{t/d} &\leq |\Lambda_n|^{t/d} + \sum_{j=0}^{k-1} |\partial\Lambda_{n,j}|^{t/d} \\ &\leq |\Lambda_n|^{t/d} + \kappa \frac{\omega_d^{t/d}}{\eta} \left( k \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\varepsilon_{n,K_n}} + O(K_n^2) |\log \rho| \right). \end{aligned}$$

Combining (3.23), (8.1), and (8.2) with  $k \leq K_n \lesssim 1$ , we conclude the assertion with  $\bar{\eta} \leq \eta/(1 + K_n)$ .  $\square$

We remark that the previous result provides a complexity bound, relative to the sparsity class  $\mathcal{A}_G^{\eta,t}$  of the solution, which is optimal with respect to the index  $t$ , but suboptimal with respect to the index  $\bar{\eta} < \eta$ .

## 8.2 PC-ADFOUR: Predictor/Corrector ADFOUR

At last, we discuss the optimality of Algorithm **PC-ADFOUR**, presented in the second part of Sect. 3.4.

**Theorem 8.2 (cardinality of PC-ADFOUR)** *Suppose that  $u \in \mathcal{A}_G^{\eta,t}$ , for some  $\eta > 0$  and  $t \in (0, d]$ . Then, there exists a constant  $C > 1$  such that the cardinality of the set  $\Lambda_n$  of the active degrees of freedom produced by **PC-ADFOUR** satisfies the bound*

$$|\Lambda_n| \leq \kappa \frac{\omega_d}{\eta^{d/t}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|u - u_n\|} + \log C \right)^{d/t}, \quad \forall n \geq 0.$$

If, in addition, the assumptions of Proposition 5.4 are satisfied, then the cardinality of the intermediate sets  $\widehat{\Lambda}_{n+1}$  activated in the predictor step can be estimated as

$$|\widehat{\Lambda}_{n+1}| \leq |\Lambda_n| + \kappa J^d \frac{\omega_d^2}{\bar{\eta}^{d/\bar{t}}} \left( \log \frac{\|u\|_{\mathcal{A}_G^{\eta,t}}}{\|u - u_n\|} + |\log \sqrt{1 - \theta^2}| + \log C \right)^{d/\bar{t}}, \quad \forall n \geq 0,$$

where  $J$  is the input parameter of **ENRICH**, and  $\bar{\eta} \leq \eta$ ,  $\bar{t} \leq t$  are the parameters which occur in the thesis of Proposition 5.4.

*Proof.* The proof of the first bound is the same as that of Theorem 8.1. Concerning the second bound, we invoke Proposition 5.4 to write  $r_n \in \mathcal{A}_G^{\bar{\eta}, \bar{t}}$  and recall that  $\|r_n - P_{\widetilde{\partial\Lambda_n}} r_n\| \leq (1 - \theta^2)^{1/2} \|r_n\|$  for each iteration  $n$ . This, combined with the minimality of the set  $\widetilde{\partial\Lambda_n}$  selected by **DÖRFLER**, yields

$$|\widetilde{\partial\Lambda_n}| \leq \frac{\omega_d}{\bar{\eta}^{d/\bar{t}}} \left( \log \frac{\|r_n\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}}}{\sqrt{1 - \theta^2} \|r_n\|} \right)^{d/\bar{t}} + 1.$$

Estimate (3.18) for **ENRICH** yields

$$|\partial\Lambda_n| \leq \kappa J^d \frac{\omega_d^2}{\bar{\eta}^{d/\bar{t}}} \left( \log \frac{\|r_n\|_{\mathcal{A}_G^{\bar{\eta}, \bar{t}}}}{\sqrt{1 - \theta^2} \|r_n\|} \right)^{d/\bar{t}}.$$

Using (2.8) and Proposition 5.4, this time to replace  $r_n$  by  $u$  and  $u - u_n$ , we obtain the desired result.  $\square$

We observe that in the case  $\bar{\eta} < \eta$  and  $\bar{t} < t$ , the cardinalities  $|\widehat{\Lambda}_{n+1}|$  and  $|\Lambda_n|$  are not bounded by comparable quantities. This looks like a non-optimal result, yet it appears to be intimately related to the fact that in general the residuals belongs to a worse sparsity class than the solution.

## Acknowledgements

We wish to thank Dario Bini for providing us with Property 2.3, and Paolo Tilli for insightful discussions on the exponential classes.

The first and the third author have been partially supported by the Italian research fund PRIN 2008 “Analisi e sviluppo di metodi numerici avanzati per EDP”. The second author has been partially supported by NSF grants DMS-0807811 and DMS-1109325.

## References

- [1] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G.P. Petrova, and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis method*, SIAM J. Math. Anal. **43** (2011), no. 3, 1457–1472.
- [2] P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math. **97** (2004), no. 2, 219–268.
- [3] D. Bini, *Personal communication*.
- [4] A. Böttcher and B. Silbermann, *Introduction to large truncated Toeplitz matrices*, Universitext, Springer-Verlag, New York, 1999.
- [5] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral methods*, Scientific Computation, Springer-Verlag, Berlin, 2006, Fundamentals in single domains.
- [6] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal. **46** (2008), no. 5, 2524–2550.
- [7] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations – convergence rates*, Math. Comp **70** (1998), 27–75.
- [8] A. Cohen, R. DeVore, and R.H. Nochetto, *Convergence rates for afem with  $H^{-1}$  data*, (2011).
- [9] S. Dahlke, M. Fornasier, and K. Groechenig, *Optimal adaptive computations in the jaffard algebra and localized frames*, Journal of Approximation Theory **162** (201), no. 1, 153 – 185.
- [10] S. Dahlke, M. Fornasier, and T. Raasch, *Adaptive frame methods for elliptic operator equations*, Adv. Comput. Math. **27** (2007), no. 1, 27–63.
- [11] R. DeVore, *Nonlinear approximation*, Acta numerica, 1998, Acta Numer., vol. 7, Cambridge Univ. Press, Cambridge, 1998, pp. 51–150.
- [12] R. DeVore and V.N. Temlyakov, *Nonlinear approximation by trigonometric sums*, J. Fourier Anal. Appl. **2** (1995), no. 1, 29–48.
- [13] W. Dörfler, *A convergent adaptive algorithm for Poisson’s equation*, SIAM J. Numer. Anal. **33** (1996), no. 3, 1106–1124.
- [14] C. Foias and R. Temam, *Gevrey class regularity for the solutions of the Navier-Stokes equations*, J. Funct. Anal. **87** (1989), no. 2, 359–369.
- [15] G. B. Folland, *Real analysis*, second ed., John Wiley & Sons Inc., New York, 1999.
- [16] T. Gantumur, H. Harbrecht, and R. Stevenson, *An optimal adaptive wavelet method without coarsening of the iterands*, Math. Comp. **76** (2007), no. 258, 615–629.
- [17] S. Jaffard, *Propriétés des matrices ”bien localisées” près de leur diagonale et quelques applications*, Annales de l’I.H.P. **5** (1990), 461–476.



- [18] P. Morin, R. H. Nochetto, and K. G. Siebert, *Data oscillation and convergence of adaptive FEM*, SIAM J. Numer. Anal. **38** (2000), no. 2, 466–488 (electronic).
- [19] R. H. Nochetto, K. G. Siebert, and A. Veiser, *Theory of adaptive finite element methods: an introduction*, Multiscale, nonlinear and adaptive approximation, Springer, Berlin, 2009, pp. 409–542.
- [20] Ch. Schwab, *p- and hp-finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 1998.
- [21] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. **7** (2007), no. 2, 245–269.