FOURIER ANALYSIS OF PERIODIC STENCILS IN MULTIGRID METHODS[*]

M. BOLTEN[†] AND H. RITTICH[†]

Abstract. Many applications require the numerical solution of a partial differential equation (PDE), leading to large and sparse linear systems. Often a multigrid method can solve these systems efficiently. To adapt a multigrid method to a given problem, local Fourier analysis (LFA) can be used. It provides quantitative predictions about the behavior of the components of a multigrid method. In this paper we generalize LFA to handle what we call periodic stencils. An operator given by a periodic stencil has a block Fourier symbol representation. It gives a way to compute the spectral radius and norm of the operator. Furthermore block Fourier symbols can be used to find out how an operator acts on smooth/oscillatory input and whether its output will be smooth/oscillatory. This information can then be used to construct efficient smoothers and coarse grid corrections. We consider a particular PDE with jumping coefficients and show that it leads to a periodic stencil. LFA shows that the Jacobi method is a suitable smoother for this problem and an operator dependent interpolation is better than linear interpolation, as suggested by numerical experiments described in the literature. If an operator is given by an ordinary stencil, then block smoothers yield periodic stencils if the blocks correspond to rectangles in the domain. LFA shows that the block Jacobi and the red-black block Jacobi method efficiently reduce more frequencies than their pointwise versions. Further, it yields that a block smoother used in combination with aggressive coarsening can to some degree compensate for the reduced convergence rate caused by aggressive coarsening.

Key words. multigrid, local Fourier analysis, jumping coefficients, block smoothers, aggressive coarsening

AMS subject classifications. 65F10, 65N22, 65N55

DOI. 10.1137/16M1073959

1. Introduction. Local Fourier analysis (LFA) provides quantitative insight into the behavior of *multigrid methods*. In this paper we introduce the notion of *periodic stencils* and their Fourier analysis. We use this to analyze multigrid solvers for partial differential equations (PDEs) with jumping coefficients and *block smoothers* in combination with *aggressive coarsening*.

Multigrid methods [25, 7, 13, 5] often are an important ingredient in the efficient numerical solution of PDEs. In most cases the solution process involves the solution of a large sparse linear system. Multigrid methods solve these systems iteratively by combining a smoothing process that is accelerated by a hierarchy of coarse grids. However, multigrid methods need to be adapted to the application, which makes a quantitative analysis beneficial.

LFA was introduced in [3] to analyze smoothing processes. It was later extended in [24] and [4] to analyze all multigrid components. For an introduction to LFA see [25, 26].

Through the last decades LFA has been further generalized. A three-grid analysis was proposed in [27] which gives further insight into the behavior of multigrid

[†]Bergische Universität Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany (bolten@math.uni-wuppertal.de, rittich@math.uni-wuppertal.de).

methods. Two- and multicolor relaxations have been analyzed in [24, 29, 18, 19]. Aggressive coarsening has been considered with pointwise [26, 2, 11] and polynomial smoothers [6]. A Fourier analysis for triangular grids has been presented in [10]. Furthermore, overlapping smoothers [20] and multigrid methods for discontinuous Galerkin discretizations [14, 15, 16] have been investigated with the help of LFA. Recently, a framework has been introduced that couples LFA with algebraic computations [9] to enable the analysis of further problems, e.g., the parabolic diffusion equation. To do so, the dimensions that pose problems to the analysis, e.g., due to the nonnormality of the operator applied in this direction, are excluded from the LFA and treated separately.

In this paper we introduce the analysis of periodic stencil operators. This gives a general framework for carrying out LFAs for more complex problems. We provide two applications of this framework.

PDEs with jumping coefficients appear in many applications, for example, when analyzing diffusion processes in inhomogenous media. In these applications the coefficients are often not continuous and vary by several orders of magnitude. We analyze the multigrid method for these problems that was described in [1].

Aggressive coarsening gets interesting when multigrid methods run in parallel environments. If a grid is very coarse the work performed on this grid per iteration is relatively low. This leads to little work per core or even idle processors and the cost of communication in comparison to computation becomes dominant, in the end leading to a bad utilization of the hardware. This is especially a problem on modern super computers with millions of cores [23]. Aggressive coarsening creates fewer coarse grids such that more time is spent on finer grids. However, the convergence rate is increased. We show that this can partially be compensated by the use of block smoothers. In contrast to already existing analysis we choose our blocks to correspond to rectangles in the domain, instead of lines or planes [29, 25, 26].

This paper is structured as follows. We provide a short review of LFA in section 2. In section 3 a high-level view on LFA for periodic stencils is presented. Section 4 discusses the interpretation of block symbols. Then we apply these results to a multigrid solver for a PDE with jumping coefficients and a multigrid solver using aggressive coarsening and block smoothers in section 5. After that, in section 6 we fill in the details from section 3. The appendix contains a proof of the computation of the spectral radius and norm of a matrix multiplication operator by using the essential supremum of pointwise values. In the remainder of this section we introduce some notation that will be used throughout the paper.

In the following we denote the dimensionality by d. A d-dimensional vector in $\mathbb{R}^d$ will be denoted in bold letters. Most of the time they will be used as (multi-)indices. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$. We denote the *pointwise multiplication* of two vectors by $\cdot$ and the *pointwise division* by $/$, i.e.,

$$\mathbf{a} \cdot \mathbf{b} := (a_\ell b_\ell)_{\ell=1}^d \quad \text{and} \quad \mathbf{a}/\mathbf{b} := (a_\ell / b_\ell)_{\ell=1}^d .$$

Whenever we compare vectors it is to be understood componentwise as well, i.e.,

$$\mathbf{a} \le \mathbf{b} \iff (a_\ell \le b_\ell \quad \text{for all} \quad \ell = 1, \dots, d) .$$

We define

$$\sum_{\mathbf{j}=\mathbf{a}}^{\mathbf{b}} u_{\mathbf{j}} := \sum_{j_1=a_1}^{b_1} \sum_{j_2=a_2}^{b_2} \cdots \sum_{j_d=a_d}^{b_d} u_{(j_1, j_2, \dots, j_d)^T}$$

and

$$(u_j)_{\mathbf{j}=\mathbf{a}}^{\mathbf{b}} := \left(\cdots \left(\left(u_{(j_1, j_2, \ldots, j_d)^T} \right)_{j_1=a_1}^{b_1} \right)_{j_2=a_2}^{b_2} \cdots \right)_{j_d=a_d}^{b_d} .$$

Furthermore, let $\mathbf{1} := (1)_{\ell=1}^d$ be the vector whose entries are all one, and let $\mathrm{vol}_{\mathbf{a}} := \prod_{j=1}^d a_j$ be the volume of the hypercube with side lengths $a_1, \ldots, a_d$. The index set of $\mathbf{n} \in \mathbb{N}^d$ is given by

$$\mathcal{I}_{\mathbf{n}} := \{1, 2, \ldots, n_1\} \times \cdots \times \{1, 2, \ldots, n_d\} .$$

We denote the euclidean scalar product and norm by

$$\langle a, b \rangle := \sum_{k=1}^d \overline{b_k} a_k \quad \text{and} \quad \|a\| := \langle a, a \rangle^{1/2} ,$$

respectively. For $f, g \in L^2(K)$ with $K \subseteq \mathbb{R}^d$ we have

$$\langle f, g \rangle := \int_K \overline{g(x)} f(x) \, \mathrm{d}x \quad \text{and} \quad \|f\| := \langle f, f \rangle^{1/2} .$$

Furthermore we define

$$\delta_{\mathbf{k}\mathbf{j}}(\vartheta) := \delta_{\mathbf{k}\mathbf{j}} := \begin{cases} 1 & \text{if } \mathbf{k} = \mathbf{j}, \\ 0 & \text{otherwise.} \end{cases}$$

2. Elements of LFA. LFA is an idealized analysis. That means we analyze a problem that is similar to the original one but easier to analyze. For LFA we usually make two simplifying assumptions.

The *first simplification* is the neglect of the boundary conditions of the PDE in its discretization. This is done by assuming that we consider (bounded) functions on infinite grids and linear (bounded) operators that act on these functions. For example, in multigrid methods this can be a discrete differential operator or the error operator of a smoother that has been extended to an infinite grid. We have that every such operator can be represented by a *stencil*.

More precisely, the infinite *grid* with *step size* $\mathbf{h} \in \mathbb{R}^d$ is

$$\Omega_{\mathbf{h}} := \left\{ \mathbf{h} \cdot \mathbf{k} : \mathbf{k} \in \mathbb{Z}^d \right\} .$$

For functions $f, g : \Omega_{\mathbf{h}} \to \mathbb{C}$ we have with

$$\langle f, g \rangle := \sum_{\mathbf{x} \in \Omega_{\mathbf{h}}} \overline{g(\mathbf{x})} f(\mathbf{x}) \quad \text{and} \quad \|f\| := \langle f, f \rangle^{1/2}$$

a scalar product and a norm, respectively. The set of all (bounded) functions on the grid $\Omega_{\mathbf{h}}$ is

$$\mathcal{G}_{\mathbf{h}} := \{ f : \Omega_{\mathbf{h}} \to \mathbb{C} : \|f\| < \infty \} .$$

DEFINITION AND LEMMA 2.1. *For $A \in L(\mathcal{G}_{\mathbf{h}}; \mathcal{G}_{\mathbf{h}})$, i.e., $A : \mathcal{G}_{\mathbf{h}} \to \mathcal{G}_{\mathbf{h}}$ is a linear, bounded operator, there exists a family $\{s_{\mathbf{x}}\}_{\mathbf{x} \in \Omega_{\mathbf{h}}}$, $s_{\mathbf{x}} \in \mathcal{G}_{\mathbf{h}}$, such that*

$$\left(Au \right)(\mathbf{x}) = \sum_{\mathbf{y} \in \Omega_{\mathbf{h}}} s_{\mathbf{x}}(\mathbf{y}) \cdot u(\mathbf{x} + \mathbf{y}) .$$

We call $\{s_{\mathbf{x}}\}_{\mathbf{x} \in \Omega_{\mathbf{h}}}$ the stencil of A.

The *second simplification* is to assume that the operator is given by a *constant stencil*, i.e.,

$$s = s_{\mathbf{x}} \quad \text{for all} \quad x \in \Omega_{\mathbf{h}}$$

for some $s \in \mathcal{G}_{\mathbf{h}}$. This is helpful, because using the *discrete time Fourier transform* (DTFT) a constant stencil operator can be represented by a *symbol*. This symbol will give us some insight into the operator.

To explain this we first need the definition of the DTFT. Let

$$\Theta_{\mathbf{h}} := [0, 2\pi/h_1) \times \cdots \times [0, 2\pi/h_d)$$

and $\mathcal{H}_{\mathbf{h}} := L^2(\Theta_{\mathbf{h}}; \mathbb{C})$. The DTFT $\mathcal{F}_{\mathbf{h}} : \mathcal{G}_{\mathbf{h}} \to \mathcal{H}_{\mathbf{h}}$ is given by

$$(\mathcal{F}_{\mathbf{h}} f)(\vartheta) := \frac{\text{vol}_{\mathbf{h}}^{1/2}}{(2\pi)^{d/2}} \sum_{\mathbf{x} \in \Omega_{\mathbf{h}}} f(\mathbf{x}) e^{-i\langle \vartheta, \mathbf{x} \rangle} .$$

This is an isometry and its inverse is given by

$$(1) \qquad \left(\mathcal{F}_{\mathbf{h}}^{-1} \hat{f} \right)(\mathbf{x}) = \frac{\text{vol}_{\mathbf{h}}^{1/2}}{(2\pi)^{d/2}} \int_{\Theta_{\mathbf{h}}} \hat{f}(\vartheta) \, e^{i\langle \vartheta, \mathbf{x} \rangle} \, \mathrm{d}\vartheta .$$

In the case that A is a constant stencil operator, define its *Fourier representation* $\widehat{A} \in L(\mathcal{H}_{\mathbf{h}}; \mathcal{H}_{\mathbf{h}})$ by

$$(2) \qquad \widehat{A}\mathcal{F}_{\mathbf{h}} = \mathcal{F}_{\mathbf{h}} A .$$

It can be easily seen that

$$(3) \qquad \widehat{A}\hat{f} = \hat{a} \cdot \hat{f}, \qquad \text{where}$$

$$(4) \qquad \hat{a}(\vartheta) = \sum_{\mathbf{y} \in \Omega_{\mathbf{h}}} s(y) \cdot e^{i\langle \vartheta, \mathbf{y} \rangle} .$$

We say that $\widehat{A}$ is a *multiplication operator* with (scalar) *symbol* $\hat{a}$. If the Fourier representation $\widehat{A}$ of $A \in L(\mathcal{G}_{\mathbf{h}}; \mathcal{G}_{\mathbf{h}})$ has a symbol $\hat{a}$, we call $\hat{a}$ the *Fourier symbol* of A. Note that not every operator has a Fourier symbol.

To simplify the notation we will not distinguish notationally between the Fourier representation $\widehat{A}$ of A and its Fourier symbol in the following. It should be clear from the context when we refer to the Fourier representation and when we refer to the Fourier symbol. Note that every operator $A \in L(\mathcal{H}_{\mathbf{h}}; \mathcal{H}_{\mathbf{h}})$ has a Fourier representation but not necessarily a Fourier symbol.

In many applications simple iterative methods yield slowly varying error functions, and a slowly varying error can be represented on a coarser grid. Multigrid methods accelerate this iterative process by using the fact that solving on a coarse grid is cheaper then on a fine one. The representation $\widehat{A}\hat{f} = \hat{a} \cdot \hat{f}$ allows us to analyze an operator in terms of slowly varying and oscillatory functions, as follows. The formula for the inverse DTFT (1) shows that each function in $\mathcal{G}_{\mathbf{h}}$ can be represented by a combination of $e^{i\langle \vartheta, \mathbf{x} \rangle}$. We have that

$$(5) \qquad \left(e^{i\langle \vartheta, \mathbf{x} \rangle} = 1 \quad \text{for all} \quad \mathbf{x} \in \Omega_h \right) \quad \text{iff} \quad \vartheta \in 2\pi\mathbb{Z}^d/h .$$

In other words, the term $e^{i\langle \vartheta, x \rangle}$ is constant if all ϑ_i $(i = 1, \ldots, d)$ are a multiple of $2\pi/h_i$. Thus if all ϑ_i $(i = 1, \ldots, d)$ are *close* to a multiple of $2\pi/h_i$, the term $e^{i\langle \vartheta, x \rangle}$ is close to the constant function and hence slowly varying (on $\Omega_{\mathbf{h}}$). Those values for ϑ are called the *low frequencies*, the other ones the *high frequencies*. This distinction

FIG. 1. *Real part of the one-dimensional (1D) wave functions $e^{i\theta h}$. We choose $\theta = 0.95 \cdot \frac{2\pi}{h}$ (left), which is close to a multiple of $\frac{2\pi}{h}$. Therefore θ is a low frequency and we see a slowly varying function. Furthermore, we choose $\theta = 0.45 \cdot \frac{2\pi}{h}$ (right), which is far away from all multiples of $\frac{2\pi}{h}$. Therefore θ is a high frequency and we see an oscillatory function.*

is illustrated in Figure 1. The function $\hat{f}(\vartheta)$ characterizes the size of the frequencies. Hence if $\hat{f}$ is large on the low frequencies and small otherwise, we expect the function $f := \mathcal{F}_{\mathbf{h}}^{-1}\hat{f}$ to be slowly varying. This means that the Fourier symbol of an operator shows the difference in its action on slowly varying and on oscillatory functions.

The definition for low and high frequencies that we gave here is vague. To make a precise definition, however, it is necessary to take the coarse grid that should be used into account. For a discussion of the definition of low and high frequencies with respect to different coarse grids see, e.g., [25, 26]. Furthermore, we give a precise definition for one case in section 5.3.

Another very useful feature of the representation is that we can compute the spectral radius and norms, easily.

LEMMA 2.2. *If $a \in L^\infty(\Theta_{\mathbf{h}})$ is the symbol of the multiplication operator A, then*

$$r(A) = \|A\| = \operatorname*{ess\,sup}_{\vartheta} |a(\vartheta)|.$$

Proof. This is a direct consequence of the fact that the spectrum of A is the essential range of a [21, Chapter 12]. □

Due to Parseval's identity the DTFT is an isometric map and therefore we obtain the following.

COROLLARY 2.3. *If $\widehat{A}$ is the Fourier symbol of the operator A, then*

$$r(A) = \|A\| = \operatorname*{ess\,sup}_{\vartheta} |\widehat{A}(\vartheta)|.$$

Note that if E is the error operator of an iterative method, then $\|E\|$ is the worst case error reduction factor for one iteration. And since $\rho(E) = \lim_{n\to\infty} \|E^n\|^{1/n}$ the spectral radius gives us the asymptotic error reduction factor per iteration.

3. The analysis of periodic stencil operators I. In the previous section we saw that if we can approximate our operator of interest by a constant stencil we can analyze it by inspecting its Fourier symbol. However, this assumption can be too restrictive. In this section we present a similar analysis for the case that the operator is approximated by a *periodic stencil*. If we split the frequency function into slices corresponding to *harmonic* frequencies and put these slices into a vector, the periodic stencil operator can be represented by a *matrix symbol*. In this section we give a high-level description and delay the details to section 6.

We start by generalizing our notion of multiplication operators. We denote $L^\infty_{\mathbf{n} \times \mathbf{m}}(\Theta_{\mathbf{h}})$ as the set of matrices of functions in $L^\infty(\Theta_{\mathbf{h}})$, i.e., $a \in L^\infty_{\mathbf{n} \times \mathbf{m}}(\Theta_{\mathbf{h}})$ iff

$$a_{\mathbf{ij}} \in L^\infty(\Theta_{\mathbf{h}}) \quad \text{for} \quad \mathbf{i} \in \mathcal{I}_{\mathbf{n}}, \ \mathbf{j} \in \mathcal{I}_{\mathbf{m}}.$$

This matrix can act on a vector of $L^2(\Theta_{\mathbf{h}})$ functions by the following definition.

DEFINITION 3.1. *Let* $a \in L^\infty_{\mathbf{n} \times \mathbf{m}}(\Theta_\mathbf{h})$. *Then* a *is the* matrix symbol *of the* (ma-trix) multiplication operator $A : L^2(\Theta_\mathbf{h})^\mathbf{m} \to L^2(\Theta_\mathbf{h})^\mathbf{n}$ *given by*

$$(Af)_\mathbf{i} := \sum_{\mathbf{j}=\mathbf{1}}^\mathbf{m} a_{\mathbf{ij}} f_\mathbf{j} \quad \textit{for all } \mathbf{i} \in \mathcal{I}_\mathbf{n}.$$

Basically this is a matrix vector multiplication where the entries of the matrix and the vector are functions and the multiplication is replaced by pointwise function multiplication. Note that we also allow for multi-indices. The product of two matrices of functions is defined as the usual matrix-matrix product. In the following we will use the same letter for the operator and its symbol.

For the definition of the *harmonic frequencies* let $\vartheta \in \Theta_\mathbf{h}$. Then $\tilde\vartheta \in \Theta_\mathbf{h}$ is an $\mathbf{n}$-*harmonic* of ϑ if

$$e^{\mathrm{i}\langle\vartheta,\mathbf{x}\rangle} = e^{\mathrm{i}\langle\tilde\vartheta,\mathbf{x}\rangle} \quad \text{for all } \mathbf{x} \in \Omega_{\mathbf{h}\cdot\mathbf{n}},$$

in other words, if the wave functions of the frequencies coincide on the subset $\Omega_{\mathbf{h}\cdot\mathbf{n}}$ of $\Omega_\mathbf{h}$. This is the case iff

$$(6) \qquad\qquad \vartheta - \tilde\vartheta \in 2\pi\mathbb{Z}^d/(\mathbf{h}\cdot\mathbf{n})$$

due to (5) and the exponentiation identity. Note that (6) is an equivalence-relation. It is easy to see that for every equivalence class $[\vartheta]$ we find a unique value $\vartheta' \in \Theta_{\mathbf{h}\cdot\mathbf{n}}$ such that

$$(7) \qquad [\vartheta] = \left\{\vartheta' + \mathbf{s}_\mathbf{j}^{\mathbf{h},\mathbf{n}} : \mathbf{j} \in \mathcal{I}_\mathbf{n}\right\}, \quad \text{where} \quad \mathbf{s}_\mathbf{j}^{\mathbf{h},\mathbf{n}} := 2\pi(\mathbf{j} - \mathbf{1})/(\mathbf{h}\cdot\mathbf{n}).$$

We call ϑ' the *base frequency*. In unambiguous situations we will drop $\mathbf{h}$ and $\mathbf{n}$ from $\mathbf{s}$. All in all we have the disjoint union

$$(8) \qquad\qquad \Theta_\mathbf{h} = \dot{\bigcup_{\vartheta \in \Theta_{\mathbf{h}\cdot\mathbf{n}}}} [\vartheta].$$

The definition of the *block Fourier symbol* of an operator requires the *frequency splitting operator* $R_\mathbf{n} : \mathcal{H}_\mathbf{h} \to (\mathcal{H}_{\mathbf{h}\cdot\mathbf{n}})^\mathbf{n}$ given by

$$(R_{\mathbf{nj}}\hat{u})(\vartheta) = \hat{u}\left(\vartheta + \mathbf{s}_\mathbf{j}^{\mathbf{h},\mathbf{n}}\right) \qquad \text{for } \vartheta \in \Theta_{\mathbf{h}\cdot\mathbf{n}}.$$

This operator produces a vector of functions such that $(R_\mathbf{n}\hat{u})(\vartheta)$ is a vector whose entries are the values of $\hat{u}$ at all harmonics of ϑ. In other words the operator $R_\mathbf{n}$ groups the harmonic frequencies together. Figure 2 shows an example of the application of this operator. Note that due to the disjoint union (8) $R_\mathbf{n}$ is bijective. If we define the scalar product and norm on $\mathcal{H}_{\mathbf{h}\cdot\mathbf{n}}^\mathbf{n}$ by

$$\langle f,g\rangle := \sum_{\mathbf{k}=\mathbf{1}}^\mathbf{n} \langle f_\mathbf{k}, g_\mathbf{k}\rangle \quad \text{and} \quad \|f\| := \langle f,f\rangle^{1/2} \qquad \text{for} \qquad f,g \in \mathcal{H}_{\mathbf{h}\cdot\mathbf{n}}^\mathbf{n}$$

it is easy to check that the mapping $R_\mathbf{n}$ is an isometry.

Now, analogously to Fourier symbol of operators we have the following.

DEFINITION 3.2 (block Fourier symbol). *Let* $0 < \mathbf{h} \in \mathbb{R}^d$, $0 < \mathbf{m},\mathbf{n} \in \mathbb{N}^d$, *and* $A : \mathcal{G}_{\mathbf{h}/\mathbf{n}} \to \mathcal{G}_{\mathbf{h}/\mathbf{m}}$. *If the operator* $\widehat{A} : \mathcal{F}_\mathbf{h}^\mathbf{n} \to \mathcal{F}_\mathbf{h}^\mathbf{m}$ *given by*

$$\widehat{A}R_\mathbf{n}\mathcal{F}_{\mathbf{h}/\mathbf{n}} = R_\mathbf{m}\mathcal{F}_{\mathbf{h}/\mathbf{m}}A$$

has a matrix symbol $\hat{a} \in L^\infty_{\mathbf{m}\times\mathbf{n}}(\Theta_\mathbf{h})$ *we call* $\hat{a}$ *the* block Fourier symbol *of* A.
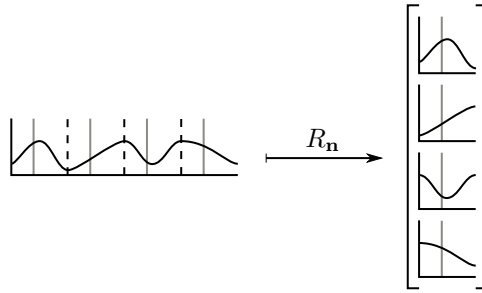
FIG. 2. *The action of frequency splitting operator $R_\mathbf{n}$ applied to a 1D function for $\mathbf{n} = 4$. The gray line marks a set of $\mathbf{n}$-harmonics. After application of the operator the harmonic frequencies all belong to the same value of the argument θ but to different components of the vector.*

In other words, if for $f = Au$ we have the DTFTs $\hat{f}$ and $\hat{u}$ of f and u, respectively, then

$$(9) \qquad \hat{f}\left(\vartheta + \mathbf{s_k}^{\mathbf{h}/\mathbf{m},\mathbf{m}}\right) = \sum_{\mathbf{j}} \widehat{A}_{\mathbf{k},\mathbf{j}}(\vartheta) \cdot \hat{u}\left(\vartheta + \mathbf{s_j}^{\mathbf{h}/\mathbf{n},\mathbf{n}}\right) \quad \text{for} \quad \vartheta \in \Theta_\mathbf{h}\,.$$

Thus in contrast to the constant stencil case where each output frequency just depends on one input frequency (3), one output frequency depends on the $\mathbf{n}$-harmonics of its frequency.

Note that we require that the operator A maps from a grid function with step size $\mathbf{h}/\mathbf{n}$ to one with step size $\mathbf{h}/\mathbf{m}$. This requirement means that the grids have to be refinements of one common grid.

As a first example for operators with block symbols consider the following lemma, which can be proven by a simple calculation.

LEMMA 3.3. *If A has a (nonblock) symbol $\widehat{A}$, then for any block size $\mathbf{n}$ the operator $(R_\mathbf{n}\widehat{A}R_\mathbf{n}^{-1})$ is a matrix multiplication operator with*

$$\left(R_\mathbf{n}\widehat{A}R_\mathbf{n}^{-1}\right)_{\mathbf{kj}}(\vartheta) = \delta_{\mathbf{kj}}\widehat{A}(\vartheta + \mathbf{s_k})\,.$$

Thus $(R_\mathbf{n}\widehat{A}R_\mathbf{n}^{-1})$ is the block symbol of A.

It states that every operator with (nonblock) symbol $\widehat{A}$ has a block symbol $\widehat{A}_\blacksquare$. This $\widehat{A}_\blacksquare(\vartheta)$ is a block diagonal matrix (a.e.); cf. the construction illustrated in Figure 3.

As already mentioned in the introduction, periodic stencil operators have block symbols. These operators are defined as follows.

DEFINITION 3.4 (periodic stencil operator). *Let $0 < \mathbf{n} \in \mathbb{N}^d$ be given. If A is given by a stencil $\{s_\mathbf{x}\}_{\mathbf{x}\in\Omega_\mathbf{h}}$ and*

$$s_\mathbf{x} = s_{\mathbf{x}+\mathbf{k}\cdot(\mathbf{h}\cdot\mathbf{n})} \quad \textit{for all} \quad \mathbf{x} \in \Omega_\mathbf{h},\ \mathbf{k} \in \mathbb{Z}^d\,,$$

then A is called a periodic stencil operator *and $\{s_\mathbf{x}\}_{\mathbf{x}\in\Omega_\mathbf{h}}$ a* periodic stencil *and $\mathbf{n}$ its* block size.

Thus a periodic stencil with period $\mathbf{n}$ is defined by $n_1 n_2 \cdots n_d$ distinct functions in $\mathcal{G}_\mathbf{h}$. After this preparation we can formulate the main result of this section which will be proven in section 6.
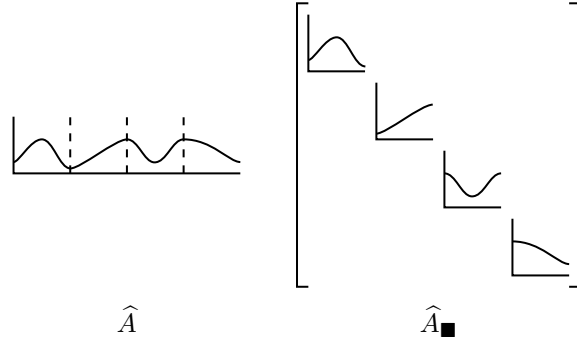
$$\widehat{A} \qquad\qquad \widehat{A}_{\blacksquare}$$

FIG. 3. *Comparison of a 1D symbol $\widehat{A}$ and the block symbol $\widehat{A}_{\blacksquare}$, which represent the same operator. The block symbol $\widehat{A}_{\blacksquare}$ acts on $R_{\mathbf{n}}\hat{u}$ in the same way as $\widehat{A}$ acts on $\hat{u}$.*

THEOREM 3.5. *Let $A \in L(\mathcal{G}_h; \mathcal{G}_h)$ be a block periodic stencil operator with stencil $a_{\mathbf{x}} \in \mathcal{G}_h$ and period $\mathbf{n}$. Let $A^{(\mathbf{k})}$ be the (constant) stencil operator corresponding to the stencil $(a^{(\mathbf{k})})_{\mathbf{x}} := a_{\mathbf{k}\cdot\mathbf{h}}$ (for all $\mathbf{x} \in \Omega_{\mathbf{h}}$). Then A has a block Fourier symbol $\widehat{A} \in L_{\mathbf{n}\times\mathbf{n}}^{\infty}(\Theta_{\mathbf{h}\cdot\mathbf{n}})$ with*

$$\widehat{A} = F^*G\,, \quad where \quad G_{\mathbf{kj}}(\vartheta) = F_{\mathbf{kj}}\widehat{A}^{(\mathbf{k})}(\vartheta + \mathbf{s_j})\,,$$

and $F \in \mathbb{C}^{\mathbf{n}\times\mathbf{n}}$ is the Fourier matrix, *i.e.,*

$$F_{\mathbf{jk}} := \tfrac{1}{\mathrm{vol}_{\mathbf{n}}^{1/2}} e^{\mathrm{i}2\pi\langle \mathbf{k}/\mathbf{n}, \mathbf{j}\rangle} \quad for \quad \mathbf{k}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}}\,.$$

The representation as a matrix symbol gives us the benefit of easy computation of the spectral radius and operator norm. Thus we get the worst case and asymptotic error reduction factors, analogously to the previous section.

THEOREM 3.6. *Let $A \in L(\mathcal{G}_{\mathbf{h}}; \mathcal{G}_{\mathbf{h}})$ with Fourier symbol $\widehat{A}$. Then*

$$r(A) = \operatorname*{ess\,sup}_{\vartheta\in\Theta_{\mathbf{h}}} r(\widehat{A}(\vartheta)) \qquad and \qquad \|A\| = \operatorname*{ess\,sup}_{\vartheta\in\Theta_{\mathbf{h}}} \|\widehat{A}(\vartheta)\|\,.$$

The proof of Theorem 3.6 can be found in Appendix A.

4. Interpretation and visualization of block symbols. For now, we can compute the norm and spectral radius of a matrix multiplication operator and therefore for a block stencil operator. However, in the case of scalar symbols we were able to extract further information from the symbol. If $\widehat{A}$ is a *scalar symbol*, then $\widehat{A}(\vartheta)$ is a scalar. This scalar provides a *direct relation* between *one* input frequency and *one* output frequency. The input frequency is multiplied by the scalar to obtain the corresponding output frequency, as can be seen in (3). This becomes handy if, e.g., a smoothing error operator E is analyzed. If $\widehat{E}(\vartheta)$ is small for high frequencies ϑ, then the norm of highly varying inputs will be substantially reduced *and* the output will consist mainly of low frequencies. Thus this answers two questions at once: First, how does a dominating frequency of the input influence the norm of the output? Second, which frequencies dominate the output? In the case where $\widehat{A}$ is a block symbol the relation is not as obvious. We only have that $\widehat{A}(\vartheta)$ is a matrix that relates the $\mathbf{m}$-harmonics of ϑ of the input the $\mathbf{n}$-harmonics of the output.

For $d \leq 3$ the scalar symbols $\widehat{A} \in L^\infty(\Theta_{\mathbf{h}})$ can be visualized, e.g., by a contour or isosurface plots. If we have a block symbol in $L^\infty_{\mathbf{n} \times \mathbf{m}}(\Theta_{\mathbf{h}})$ we have $\mathrm{vol}_{\mathbf{n}} \cdot \mathrm{vol}_{\mathbf{m}}$ functions from $L^\infty(\Theta_{\mathbf{h}})$. As this grows quickly, plotting every single function is not feasible. In this section we discuss how to reduce a block symbol to a scalar function to obtain the same pieces of information as in the scalar case. Unlike in the scalar case, two different transformations are needed. The choice of the transformation depends on whether the influence of certain frequencies on the norm are to be obtained or the dominating frequencies in the output.

4.1. Frequency damping. For a moment, consider a matrix $A \in \mathbb{C}^{m \times n}$ and a vector $\mathbf{x} \in \mathbb{C}^n$. In order to quantify the influence of a dominating component i in the vector $\mathbf{x}$ on the norm of the output we write $\mathbf{x} = \sum_i \xi_i e_i$, where e_i is the ith unit vector. Then $A\mathbf{x} = \sum_i \xi_i(Ae_i)$ and thus

$$\|A\mathbf{x}\| \leq \sum_i |\xi_i| \cdot \|Ae_i\|.$$

Consequently if $|\xi_j| \ll |\xi_i|$ for all $i \neq j$, then

$$\|A\mathbf{x}\| \approx \|Ae_i\|.$$

Therefore the term $\|Ae_i\|$ gives us an estimate on how strongly the ith component influences the result. Note that Ae_i is the ith column of A. Thus we compute the vector containing the norms of the columns of A, i.e.,

$$\mathrm{cn}(A)_j = \left(\sum_k |a_{kj}|^2 \right)^{1/2}.$$

By this we have a vector which contains in the ith component a measure for the influence of the ith component in $\mathbf{x}$ on the result of $A\mathbf{x}$.

Analogously we can define for a block symbol $\widehat{A} \in L^\infty_{\mathbf{m} \times \mathbf{n}}(\Theta_{\mathbf{h}})$ the function

$$\mathrm{cn}(\widehat{A})_{\mathbf{j}} := \left(\sum_{\mathbf{k}} |\widehat{A}_{\mathbf{kj}}|^2 \right)^{1/2}.$$

Recall that the matrix $\widehat{A}(\vartheta)$ relates the $\mathbf{n}$-harmonics of the input to the $\mathbf{m}$-harmonics (cf. (9)). If we consider the vector $\hat{u} \in \mathcal{H}^{\mathbf{n}}_{\mathbf{h}}$, then by definition of $R_{\mathbf{n}}$ the entry $\hat{u}_{\mathbf{j}}(\vartheta)$ corresponds to the frequency $\vartheta + \mathbf{s_j}$ (recall the definition $\mathbf{s_j}$ from (7)). Thus function $\mathrm{cn}(\widehat{A})_{\mathbf{j}}(\vartheta)$ is the influence of the frequency $\vartheta + s_{\mathbf{j}}$ on the result. And therefore

$$R_{\mathbf{m}}^{-1}\mathrm{cn}(\widehat{A})$$

gives the frequency function that describes how every frequency influences the result. We refer to this function as the *frequency damping* of the operator. It can be visualized easily.

4.2. Frequency emission. Again, let us consider a matrix $A \in \mathbb{C}^{m \times n}$. Assume we have a normalized input vector $\mathbf{x} \in \mathbb{C}^n$. Consider the question of how large the ith component of $|Ax|$ can be for all normalized vectors $\mathbf{x}$. That is, we want to determine

$$\max_{\|\mathbf{x}\|=1} |\langle e_i, A\mathbf{x}\rangle| = \max_{\|\mathbf{x}\|=1} |\langle A^* e_i, \mathbf{x}\rangle| \leq \max_{\|\mathbf{x}\|=1} \|A^* e_i\| \cdot \|\mathbf{x}\| = \|A^* e_i\|.$$

We define $\mathrm{rn}(A)_i := \|A^* e_i\|$, which is the norm of the ith row of A, i.e.,

$$\mathrm{rn}(A)_i := \left(\sum_j |a_{ij}|^2 \right)^{1/2} .$$

Analogously for a block symbol $\widehat{A} \in L_{\mathbf{m} \times \mathbf{n}}$ we define

$$\mathrm{rn}(\widehat{A})_{\mathbf{i}} := \left(\sum_{\mathbf{i}} |\widehat{A}_{\mathbf{ij}}|^2 \right)^{1/2} .$$

Like in the previous section the entry $\mathrm{rn}(\widehat{A})_{\mathbf{k}}$ belongs to the output frequency $\hat{f}(\vartheta + \mathbf{s_k})$. Hence to get a function which assigns every ϑ its worst case value for a normalized input, we need to compute

$$R_{\mathbf{n}}^{-1} \mathrm{rn}(\widehat{A}) .$$

We call this function the *frequency emission* of the operator.

5. Applications. In this section we use Theorem 3.5 to analyze a multigrid method for a PDE with jumping coefficients, block smoothers, and block smoothers in combination with aggressive coarsening.

5.1. Jumping coefficients. When analyzing a multigrid method for solving a PDE it is usually assumed that the coefficients of the differential operator are almost locally constant. Hence in the discretization the stencils will be almost constant between neighboring grid points. Thus the local behavior of a multigrid method can be analyzed by considering constant stencils. However, if the operator has a jump in the coefficients this assumption is not feasible. In this section we illustrate how discretizing jumping coefficient operators can be formulated using periodic stencils and thus can be analyzed in terms of Theorem 3.5.

We consider the PDE

$$-\nabla \cdot (b \nabla u) = f \quad \text{in } \Omega ,$$

where $0 < b : \Omega \to \mathbb{R}$. Usually Ω will be a suitable domain, i.e., $\Omega \subseteq \mathbb{R}^2$ would be Lipschitz, and u has to fulfill some boundary conditions. However, since we want to perform an LFA we neglect the domain and the boundary conditions. Furthermore, in this section we use the definitions

$$nw := \mathbf{h} \cdot \begin{pmatrix} -1 \\ -1 \end{pmatrix} , \qquad n := \mathbf{h} \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} , \qquad ne := \mathbf{h} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} ,$$

$$w := \mathbf{h} \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} , \qquad c := \mathbf{h} \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} , \qquad e := \mathbf{h} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} ,$$

$$sw := \mathbf{h} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} , \qquad n := \mathbf{h} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} , \qquad se := \mathbf{h} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} .$$

We discretize the operator $\nabla \cdot (b \nabla)$ by cell-centered finite volumes, using regular, rectangular control volumes. We assume that b is discontinuous, but only at the boundary of the control volumes, and that $h_1 = h_2$. In this case

$$a_{\mathbf{x}}(\mathbf{y}) = \begin{cases} -\frac{2}{\mathrm{vol}_{\mathbf{h}}} \cdot \frac{b(\mathbf{x}+\mathbf{y})b(\mathbf{x})}{b(\mathbf{x}+\mathbf{y})+b(\mathbf{x})} & \text{if } \mathbf{y} \in \{w, e, n, s\} , \\ -(a_{\mathbf{x}}(w) + a_{\mathbf{x}}(e) + a_{\mathbf{x}}(n) + a_{\mathbf{x}}(s)) & \text{if } \mathbf{y} = c , \\ 0 & \text{otherwise} \end{cases}$$
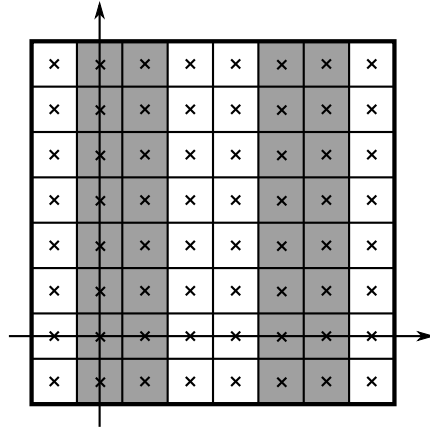
FIG. 4. *The coefficient pattern of the jumping coefficient problem. White: $b = 1$. Gray: $b = 10^6$. Cross: Control volume center.*

is the stencil of a reasonable discretization $A_h : \mathcal{G}_{\mathbf{h}} \to \mathcal{G}_{\mathbf{h}}$ [25, Chapter 7.7]. We are interested in the case when there is one jump in the domain. However, as LFA analyzes local behavior, this should have similar results as in the case where the jumps are far apart. Thus b is chosen to be periodic with period $\mathbf{p} = \mathbf{n} \cdot \mathbf{h}$ and

$$b(\mathbf{x}) = 1 \qquad \text{for all } 0 \le x_1 + (h_1/2) < p_1/2\,,$$
$$b(\mathbf{x}) = 10^6 \quad \text{for all } p_1/2 \le x_1 + (h_1/2) < p_1\,,$$

which is sketched in Figure 4. Then s is a periodic stencil with period $\mathbf{p}$.

Our first investigation concerns the question of whether the damped Jacobi method [22] is a suitable smoother for this problem. Its error propagation operator is

$$E_J := I - \omega D^{-1} A\,,$$

where $d_x(y) := s(y)\delta_{xy}$ is the stencil of D. Thus D is also given by a periodic stencil. By using the fact that $\mathcal{F}_h$ is an isomorphism the symbol of E_J is given by

$$\widehat{E_J} = \widehat{I} - \omega \widehat{D}^{-1} \widehat{A}\,.$$

Figure 5 is a plot of $\operatorname{rn}(\widehat{E_J})$ for the case $\omega = 0.8$. The values of this function which are close to one are located in the corners of the plot. Thus the low frequencies are dominating in the output of $\widehat{E_J}$, and therefore the Jacobi method is a suitable smoother for this problem.

Now, to construct an efficient *two-grid* method we have to construct a coarse grid correction. And for that we have to pick a coarse grid step size H, an interpolation $P : \mathcal{G}_H \to \mathcal{G}_h$, a restriction $R : \mathcal{G}_h \to \mathcal{G}_H$, and a coarse grid approximation $A_H : \mathcal{G}_H \to \mathcal{G}_H$. The two-grid method has an error operator that is given by [7, 25, 26, 13]

$$(10) \qquad\qquad E := E_J \left(I - P A_H^{-1} R A_h \right) E_J\,.$$

As we have seen in Figure 5 the Jacobi method produces a smooth error. A smooth error can be well approximated by linear interpolation from a coarse set of degrees of freedom. Thus we choose P as linear interpolation with full coarsening
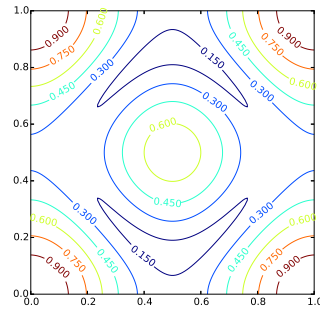
FIG. 5. *The frequency emission* $\mathrm{rn}(\widehat{E}_J)$ *with* $\omega = 0.8$ *for the jumping coefficient problem with period* $\mathbf{p} = (4,4)^T$.
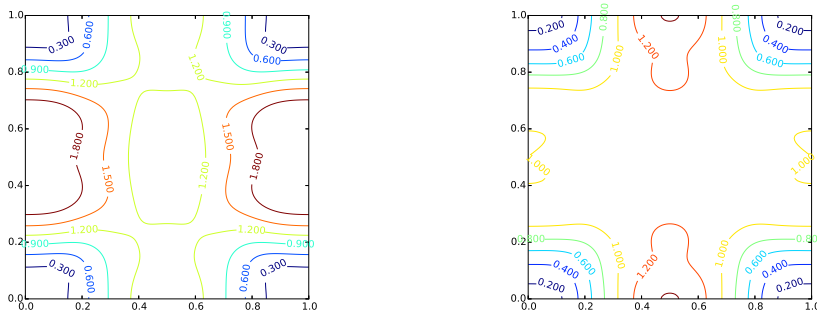

FIG. 6. *Frequency damping of the linear (left) and adaptive interpolation (right) coarse grid correction for the jumping coefficient problem.*

TABLE 1
Comparison of the linear interpolation and adaptive two-grid method.

	Linear interpolation	Adaptive
$r(E_{\mathrm{TG}})$	0.46	0.36
$\|E_{\mathrm{TG}}\|$	0.91	0.53

($H := 2h$) and $R := P^*$ [5]. For our discretization we required that the jump in the coefficients should be on the boundary of two control volumes. As this can be violated on coarser grids a rediscretization coarse grid correction is not feasible. Therefore, the Galerkin coarse grid approximation $A_H := RA_hP$ is used.

The block symbol $\widehat{E}_{TG}$ of the whole two-grid method (10) can be computed [26, 24]. In Figure 6 we plot $\mathrm{rn}(\widehat{E})$. We see that the coarse grid correction with linear interpolation amplifies some high frequencies substantially. Thus, as motivated, the smooth error is interpolated well. However, nonsmooth error results in amplification of the error. And since during the first few V-cycles the error is nonsmooth, this has a bad effect on the convergence rate.

This behavior can also be seen in Table 1. For the linear interpolation the norm of the error operator is close to one, while the spectral radius is much smaller. Hence the method works well in the asymptotic regime but can be bad for the first few iterations.

Thus we take a look at the *operator dependent interpolation* as described in [1, 8, 25]. Let $u \in \mathcal{G}_\mathbf{h}$ be the function on the coarse grid and $v \in \mathcal{G}_\mathbf{h}$ the result of the interpolation. For coarse grid points we have

$$v(\mathbf{x}) = u(\mathbf{x}).$$

For grid points on a horizontal line between two coarse grid points, we compute

$$v(\mathbf{x}) = \alpha_w \cdot u(\mathbf{x} + w) + \alpha_e \cdot u(\mathbf{x} + e), \quad \text{where}$$

$$\alpha_w = \frac{a_\mathbf{x}(nw) + a_\mathbf{x}(w) + a_\mathbf{x}(sw)}{a_\mathbf{x}(n) + a_\mathbf{x}(c) + a_\mathbf{x}(s)} \quad \text{and} \quad \alpha_e = \frac{a_\mathbf{x}(ne) + a_\mathbf{x}(e) + a_\mathbf{x}(se)}{a_\mathbf{x}(n) + a_\mathbf{x}(c) + a_\mathbf{x}(s)}.$$

Analogously for grid points on a vertical line

$$v(\mathbf{x}) = \alpha_n \cdot u(\mathbf{x} + n) + \alpha_s \cdot u(\mathbf{x} + s), \quad \text{where}$$

$$\alpha_n = \frac{a_\mathbf{x}(nw) + a_\mathbf{x}(n) + a_\mathbf{x}(ne)}{a_\mathbf{x}(w) + a_\mathbf{x}(c) + a_\mathbf{x}(e)} \quad \text{and} \quad \alpha_s = \frac{a_\mathbf{x}(sw) + a_\mathbf{x}(s) + a_\mathbf{x}(se)}{a_\mathbf{x}(w) + a_\mathbf{x}(c) + a_\mathbf{x}(e)}.$$

For the interpolation of the points in the center of four coarse grid points, we use the already interpolated points, by

$$\begin{aligned} v(\mathbf{x}) = -\tfrac{1}{a_\mathbf{x}(c)} \cdot \Big(& a_\mathbf{x}(e) \cdot v(\mathbf{x} + e) + a_\mathbf{x}(ne) \cdot v(\mathbf{x} + ne) + a_\mathbf{x}(n) \cdot v(\mathbf{x} + n) \\ & + a_\mathbf{x}(nw) \cdot v(\mathbf{x} + nw) + a_\mathbf{x}(w) \cdot v(\mathbf{x} + w) + a_\mathbf{x}(sw) \cdot v(\mathbf{x} + s) \\ & + a_\mathbf{x}(s) \cdot v(\mathbf{x} + s) + a_\mathbf{x}(se) \cdot v(\mathbf{x} + se) \Big). \end{aligned}$$

Figure 6 illustrates that the adaptive coarse grid correction performs much better than the one with linear interpolation. The results of the two-grid methods are shown in Table 1. The adaptive method has a smaller spectral radius. However, the norm of the two-grid operator is substantially reduced for the adaptive method.

As LFA considers a simplified problem we want to compare our predictions from Table 1 to a real run of the two-grid method. We choose $\Omega = (0,1) \times (0,1)$, the unit square, and apply Dirichlet boundary conditions on $\partial\Omega$. The domain is discretized by a regular 33×33 grid. Furthermore we choose $b(\mathbf{x})$ in the following way. In the center of the domain we let $b(\mathbf{x})$ alternate in the x_1 direction between 1 and 10^6 11 times. The value of $b(\mathbf{x})$ in this area changes every two control volumes. Thus, in the center of the domain the coefficient b consists of 11 strips of width $2h$. For this case we compute the spectral radius of the adaptive two-grid method with Jacobi ($\omega = 0.8$) and obtain $\rho(E_{\text{TG}}^{\text{real}}) = 0.36$, which is in perfect agreement with our prediction.

As mentioned earlier we observed that the number of jumps seems not to influence the convergence rate. Thus run the same experiment again with

$$b(\mathbf{x}) = \begin{cases} 10^6 & \text{if } x_1 < 1/2, \\ 1 & \text{if } x_1 \geq 1/2. \end{cases}$$

We compute the spectral radius again and obtain $\rho(E_{\text{TG}}^{\text{real}}) = 0.36$, which is still in perfect agreement with the prediction, even though we have just one jump in the domain and our analysis assumes that we have infinitely many jumps.

Remark 5.1. The problem of jumping coefficients has already been analyzed by the semialgebraic mode analysis (SAMA) [9]. This analysis exploits the tensor structure of a given problem and applies the Fourier analysis only to one part of the tensor product. More precisely, SAMA considers an idealized problem where the first part of the tensor product is considered on a *finite* grid and the second part of the product on an *infinite* grid.

In the case of the jumping coefficient problem, the spatial direction in which the coefficient jumps is represented on a finite grid and the direction in which the coefficient stays constant is represented on an infinite grid. This approach, however, has some drawbacks for this application.

The first drawback is that the SAMA symbols are larger than the symbols obtained by our periodic stencil approach. This is because SAMA uses a finite grid of high resolution. This high resolution is needed to obtain representative results.

The second drawback is that SAMA requires a tensor structure. In our approach we could easily analyze the case where the parts of the domain where the coefficient is large form rectangles, instead of lines. This problem has no tensor structure, thus the analysis by SAMA is not possible.

5.2. Block Jacobi. This section describes an LFA for the block Jacobi method [12, 22] (cf. [28]). To define the method, first *block Jacobi sweep*s are introduced. Assume that we are given a set S of (nonempty) subsets of $\mathcal{I}$ and a *weight* $\omega \in (0, 2)$. A *block Jacobi sweep* is the mapping from $\mathcal{G}_\mathbf{h} \to \mathcal{G}_\mathbf{h}$ given by

$$(11a) \qquad x \mapsto x + \omega \sum_{s \in S} y_s\,,$$

where the grid functions $y_s \in \mathrm{span}\{e_\ell : \ell \in s\}$ are choosen such that the *correction residuals* $r_s := b - A(x + y_s)$ fulfill

$$(11b) \qquad r_{s\mathbf{j}} = 0 \quad \text{for} \quad \mathbf{j} \in s\,.$$

In other words, each y_s manipulates the components x_i with $i \in s$ such that the (correction) residual of y_s vanishes for all those indices.

The block Jacobi method requires that the set S is a partition of the index set, i.e., $\mathcal{I} = \bigcup_{s \in S} s$, and for $s, s' \in S$ the statement $s \cap s' \neq \emptyset$ implies $s = s'$. Then one iteration of the block Jacobi method is given by a block Jacobi sweep.

In this case a few algebraic transformations yield

$$(12) \qquad \tilde{x} = x + \omega D^{-1}(b - Ax)\,,$$

where D is the *restriction of A to the blocks in S*. This restriction is defined with the help of the mapping

$$bl(i) := s \quad \text{such that} \quad i \in s, \, s \in S\,,$$

that assigns every element in $\Omega_\mathbf{h}$ the part of the partition that contains this element. If A is given by a stencil a, the *restriction of A to the blocks of S* is given by the stencil d with

$$(13) \qquad d_\mathbf{x}(\mathbf{y}) = \begin{cases} a_\mathbf{x}(\mathbf{y}) & \text{if } (\mathbf{x} + \mathbf{y}) \in b(\mathbf{x})\,, \\ 0 & \text{otherwise.} \end{cases}$$
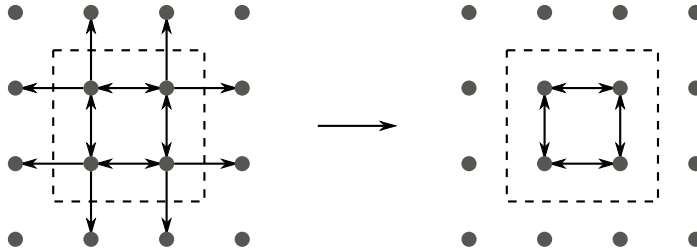
FIG. 7. *Computation of the stencil d from a in two dimensions.*

For $d = 2$, $\mathbf{n} = \binom{2}{2}$ and the five-point stencil the computation of d is illustrated in Figure 7. Note that if we restrict D to a finite subset of $\Omega_{\mathbf{h}}$ it can be (after a suitable permutation of the indices) represented by a block diagonal matrix.

Until now the partition was arbitrary. In this paper we focus on the case where the partition is given by regular squares of points in $\Omega_{\mathbf{h}}$, in contrast to [29, 25, 26]. More precisely, we choose the partition of $\Omega_{\mathbf{h}}$ by

$$(14) \qquad S_{\mathbf{i}} := (\mathbf{i} \cdot \mathbf{n}) + \mathcal{T}_{\mathbf{n}}, \quad \text{where} \quad \mathcal{T}_{\mathbf{n}} := \{\mathbf{h} \cdot \mathbf{k} : 0 \leq \mathbf{k} < \mathbf{n}\} .$$

We call $\mathbf{n}$ the *block size*. In this case we see that d is a periodic stencil and thus Theorem 3.5 can be applied.

From the iteration update (12) it follows that the error operator is

$$E = \left(I - \omega D^{-1} A\right) .$$

As the DTFT is an isomorphism we have that the matrix symbol of the error operator is

$$\widehat{E} = \left(\widehat{I} - \widehat{D}^{-1} \widehat{A}\right) .$$

Recall Lemma 3.3, which stated that every operator with a symbol has a corresponding block symbol of arbitrary block size. By that we get a block symbol $\widehat{A}$ of A with block size $\mathbf{n} \times \mathbf{n}$ and thus the matrix-matrix multiplication with D^{-1} is possible.

We consider the second order finite difference discretization of Poisson's equation $\Delta u = f$. In the following we choose $\omega = 0.8$. In Figure 8 we compare the regular Jacobi method with the 4×4 (block size) Jacobi method. The function has its largest values in the corners of the plot. Note that in the 4×4 method the high values are much closer to the corners and take less space. Thus the block smoother is effective for a larger subset of frequencies than the regular one.

We can quantify this behavior by computing *smoothing factors*. For this analysis, however, we need a precise definition of low and high frequencies.

The precise definition of low and high frequencies depends on the choice of the coarse grid. Assume we consider the fine grid $\Omega_{\mathbf{h}}$ and the coarse grid $\Omega_{\mathbf{c} \cdot \mathbf{n}}$. In this case we call $\mathbf{c}$ the *coarsening range* and say that a frequency $\theta \in \Theta_{\mathbf{h}}$ is *low* with respect to the coarsening range $\mathbf{c}$ if

$$\vartheta_i < \frac{\pi}{h c_i} \quad \text{or} \quad \frac{(2c_i - 1)\pi}{h c_i} \leq \vartheta_i \quad \text{for} \quad i = 1, \ldots, d .$$

This definition can be motivated as follows. When we restrict the wave functions corresponding to the $\mathbf{c}$-harmonics $[\theta]$ of some θ to the coarse grid $\Omega_{\mathbf{c} \cdot \mathbf{h}}$, they all give the same function. It turns out that every set of $\mathbf{c}$-harmonics contains exactly one
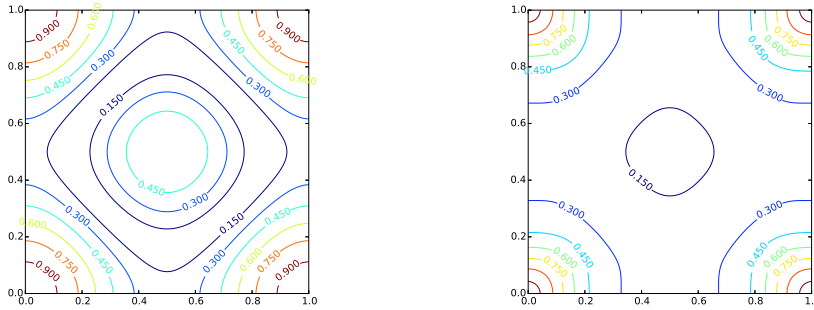
FIG. 8. *Frequency emission of the Jacobi method (left) and the 4×4 block Jacobi method (right) with $\omega = 0.8$.*

TABLE 2
Smoothing factor for the block Jacobi method with different block sizes and coarsening ranges.

Coarsening	Block size				
	1	2	4	6	8
2	0.60	0.40	0.42	0.42	0.42
4	0.88	0.80	0.66	0.63	0.61
6	0.95	0.90	0.84	0.76	0.73
8	0.97	0.94	0.89	0.86	0.80

low frequency. This low frequency has the minimal distance to the set $2\pi \mathbb{Z}^d/h$ from the frequencies in $[\theta]$. Thus, the corresponding wave function is the smoothest. In other words, the low frequencies correspond to the smoothest wave functions from all the wave functions that are not distinguishable on the coarse grid from this particular wave function. Using this definition we can introduce the smoothing factor.

The smoothing factor of a smoothing method indicates the effectiveness of the smoother when using the coarse grid $\Omega_{\mathbf{c} \cdot \mathbf{h}}$. The factor is defined using the filtering operator $\widehat{Q}$ that is given by the symbol

$$\widehat{Q}(\vartheta) := \begin{cases} 0 & \text{if } \vartheta \text{ is low w.r.t. } \mathbf{c}, \\ 1 & \text{otherwise.} \end{cases}$$

The smoothing factor of a smoother E is given by

$$\mathrm{sm}(E) := r(\widehat{Q}\widehat{E}) \,.$$

As only high frequencies pass the filtering operator $\widehat{Q}$ and an effective smoother reduces high frequencies, an effective smoother should have a small smoothing factor.

Table 2 lists the smoothing factor of the block Jacobi method for different block sizes and coarsening ranges. We observe that the smoothing factor increases when we increase the coarsening range. On the other hand the smoothing factor decreases when we increase the block size. Thus, we assume that the block Jacobi method with large block sizes is a suitable smoother for large coarsening ranges.

5.3. Red-black block Jacobi. The *red-black block Jacobi method* (RBBJ) is a modification of the block Jacobi method. Analogously to the block Jacobi method we start with a partition S of the grid $\Omega_{\mathbf{h}}$. Then we partition the blocks in S into the *red*
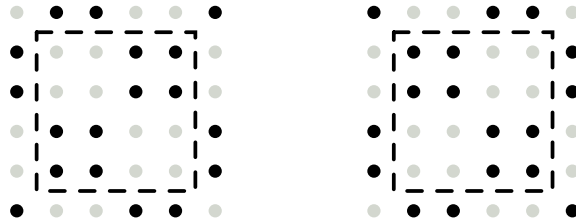
FIG. 9. *We consider the block size* $\mathbf{n} = \left(\begin{smallmatrix}2\\2\end{smallmatrix}\right)$ *The operator* Z_c *maps the value at black dots to itself and the others to zero. This pattern is repeated over the whole grid* $\Omega_{\mathbf{h}}$.

blocks S_R and the *black blocks* S_B (i.e., $\{S_R, S_B\}$ is a partition of S). Furthermore define the *red points* in $\Omega_{\mathbf{h}}$ as $R := \bigcup_{s \in S_R} s$ and the *black points* in $\Omega_{\mathbf{h}}$ as $S := \bigcup_{s \in S_B} s$.

Then one iteration of the RBBJ method is given by applying a block Jacobi sweep (11) w.r.t. S_R followed by a block Jacobi sweep w.r.t. S_B Thus each sweep updates only half of the blocks. The difference to the block Jacobi method is that the correction residuals for the second sweep are computed w.r.t. an intermediate iterate $\bar{x}$.

A few algebraic manipulations yield that one iteration of the RBBJ method applied to the iterate x yields the new iterate $\tilde{x}$ by

$$\bar{x} = x + \omega Z_{\mathrm{R}} D^{-1} (b - Ax),$$

$$\tilde{x} = \bar{x} + \omega Z_{\mathrm{B}} D^{-1} (b - A\bar{x}),$$

where D is the restriction of A to S (see (13)) and

$$\left(Z_c u\right)(\mathbf{x}) = \begin{cases} u(\mathbf{x}) & \text{if } \mathbf{x} \in c \\ 0 & \text{otherwise} \end{cases} \quad \text{for} \quad c \in \{R, B\}.$$

Note that the intermediate variable $\bar{x}$ is the result of the first Jacobi sweep and $\tilde{x}$ the result of the second. By combining the two steps the method is given by

$$E = E_B E_R, \quad \text{where} \quad E_c = \left(I - Z_c D^{-1} A\right) \quad \text{for } c \in \{R, B\}.$$

This can be formulated in the frequency domain

$$\widehat{E}_R = \left(I - \widehat{Z}_R \widehat{D}^{-1} A\right) \quad \text{and} \quad \widehat{E}_B = \left(I - \widehat{Z}_B \widehat{D}^{-1} A\right).$$

We are interested in the case where the partitions are given by

$$S_R := \{(\mathbf{k} \cdot \mathbf{n}) + \mathcal{T}_{\mathbf{n}} : \mathbf{k}_1 + \cdots + \mathbf{k}_d \text{ even}\},$$

$$S_B := \{(\mathbf{j} \cdot \mathbf{n}) + \mathcal{T}_{\mathbf{n}} : \mathbf{j}_1 + \cdots + \mathbf{j}_d \text{ odd}\},$$

where $\mathcal{T}_{\mathbf{n}}$ was defined in (14). In that case Z_c is given by a periodic stencil with period $2\mathbf{n}$ (see Figure 9). The stencil of D has period $\mathbf{n}$ and therefore also the period $2\mathbf{n}$; hence the operators in (5.3) are easily combined. The results for the RBBJ method applied to the second order finite difference discretization of Poisson's equation can be found in Figure 10. Similarly to the block Jacobi method the area of the undamped components becomes more concentrated in the corners of the plot when the block size increases.
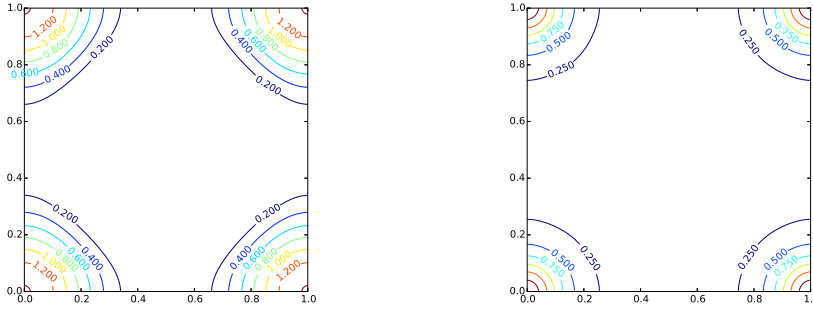
FIG. 10. *Frequency emission of the red-black Jacobi method (left) and the 4×4 RBBJ method (right).*

Let us consider a specific example. Set $d = 2$ and $\mathbf{n} = (1,1)^T$. If we order the multi-indices in $\mathcal{I}_{2\mathbf{n}}$ as $(0,0)^T$, $(1,1)^T$, $(1,0)$, and $(0,1)$ we get

$$(15a) \quad \widehat{E}_R(\vartheta) = \frac{1}{2} \begin{pmatrix} 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{00})}{\widehat{D}(\vartheta + \mathbf{s}_{00})} & -\frac{\widehat{A}(\vartheta + \mathbf{s}_{11})}{\widehat{D}(\vartheta + \mathbf{s}_{11})} & 0 & 0 \\ -\frac{\widehat{A}(\vartheta + \mathbf{s}_{00})}{\widehat{D}(\vartheta + \mathbf{s}_{00})} & 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{11})}{\widehat{D}(\vartheta + \mathbf{s}_{11})} & 0 & 0 \\ 0 & 0 & 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{10})}{\widehat{D}(\vartheta + \mathbf{s}_{10})} & -\frac{\widehat{A}(\vartheta + \mathbf{s}_{01})}{\widehat{D}(\vartheta + \mathbf{s}_{01})} \\ 0 & 0 & -\frac{\widehat{A}(\vartheta + \mathbf{s}_{10})}{\widehat{D}(\vartheta + \mathbf{s}_{10})} & 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{01})}{\widehat{D}(\vartheta + \mathbf{s}_{01})} \end{pmatrix}$$

and

$$(15b) \quad \widehat{E}_B(\vartheta) = \frac{1}{2} \begin{pmatrix} 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{00})}{\widehat{D}(\vartheta + \mathbf{s}_{00})} & \frac{\widehat{A}(\vartheta + \mathbf{s}_{11})}{\widehat{D}(\vartheta + \mathbf{s}_{11})} & 0 & 0 \\ \frac{\widehat{A}(\vartheta + \mathbf{s}_{00})}{\widehat{D}(\vartheta + \mathbf{s}_{00})} & 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{11})}{\widehat{D}(\vartheta + \mathbf{s}_{11})} & 0 & 0 \\ 0 & 0 & 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{10})}{\widehat{D}(\vartheta + \mathbf{s}_{10})} & \frac{\widehat{A}(\vartheta + \mathbf{s}_{01})}{\widehat{D}(\vartheta + \mathbf{s}_{01})} \\ 0 & 0 & \frac{\widehat{A}(\vartheta + \mathbf{s}_{10})}{\widehat{D}(\vartheta + \mathbf{s}_{10})} & 2 - \frac{\widehat{A}(\vartheta + \mathbf{s}_{01})}{\widehat{D}(\vartheta + \mathbf{s}_{01})} \end{pmatrix}.$$

This is the well-known LFA for red-black Jacobi [24, 25] illustrating that Theorem 3.5 is a generalization.

Remark 5.2. The block symbols of the red-black Jacobi method (15) are block diagonal. This fact is a consequence of the ordering of the multi-indices that we chose. For a general problem, however, we expect the block symbol to be a full matrix. Thus, in the general case the ordering $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$ would be more natural.

Remark 5.3. The analysis in this section can be generalized to multicoloring (more than two colors) block relaxation schemes. Assume we are given the colors $c_1, \ldots, c_m$; then we have to define Z_c for all $c \in C$ appropriately and

$$\widehat{E} = \widehat{E}_{c_m} \cdots \widehat{E}_{c_2} \widehat{E}_{c_1} .$$

5.4. Aggressive coarsening. The usual motivation for multigrid (see, e.g., [25, 26, 7]) is that a smooth error can be represented by fewer degrees of freedom than an oscillatory one. Block smoothers, as could be seen in the previous section, are effective
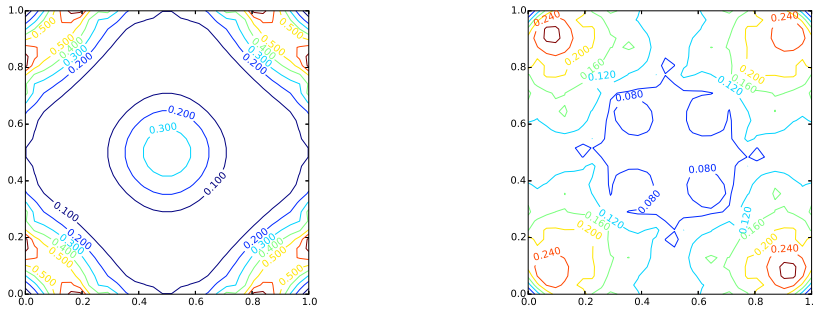
FIG. 11. *Frequency damping plot of a two-grid method with aggressive* 4 × 4 *coarsening and a Jacobi smoother (left) and a* 4 × 4 *block Jacobi smoother (right).*

TABLE 3
Spectral radius of the two-grid method with aggressive coarsening and block Jacobi smoothing as predicted by LFA.

Coarsening	Block size				
	1	2	4	6	8
2	0.36	0.32	0.27	0.26	0.25
4	0.76	0.62	0.38	0.45	0.35
6	0.88	0.79	0.68	0.47	0.58
8	0.92	0.85	0.76	0.74	0.55

in a larger subset of the frequency spectrum than point smoothers, so the smooth error should be representable by even fewer degrees of freedom. In the following, two-grid methods are analyzed where the distance between the points on the coarse grid is increased. To do so, instead of choosing the coarse grid step size H as $2\mathbf{h}$ the step size $H = \mathbf{c} \cdot \mathbf{h}$ is chosen. This is called *aggressive coarsening*.

The error operator of the two-grid method (see, e.g., [25, 26, 7]) is

$$E := M \left(I - PA_H^{-1}RA_h \right) M \, .$$

The operator M is the smoothing error operator, while P and R are the interpolation and restriction.

We choose P as the bilinear interpolation (with respect to the coarsening) and $R = \frac{1}{C}P^T$, where $C \in \mathbb{R}$ is chosen such that the constant function is transferred exactly to the coarse grid.

Figure 11 shows the frequency plot of a two-grid method and aggressive coarsening applied to the second order finite difference discretization of Poisson's equation. It demonstrates that the combination of (pointwise) Jacobi and aggressive coarsening fails to reduce some of the high frequency errors. However, this problem disappears if we use the block Jacobi method with a sufficiently large block size.

In Tables 3 and 4 the spectral radii of the two-grid error operator (10) for different coarsenings $\mathbf{c}$ and different block sizes of the smoothers are presented. It turns out that choosing the block size equal to the coarsening ratio can compensate for the growth in the spectral radius caused by aggressive coarsening to some extend. Furthermore, we observe that not choosing the block size as a multiple of the coarsening has a negative effect on the convergence rate.

TABLE 4
Spectral radius of the two-grid method with aggressive coarsening and RBBJ smoothing as predicted by LFA.

Coarsening	Block size				
	1	2	4	6	8
2	0.072	0.033	0.028	0.024	0.015
4	0.51	0.3	0.11	0.18	0.1
6	0.73	0.55	0.39	0.18	0.3
8	0.81	0.67	0.51	0.47	0.25

TABLE 5
Spectral radius of the two-grid method with aggressive coarsening and block Jacobi smoother with optimal weight (parentheses).

Coarsening	Block size				
	1	2	4	6	8
2	0.36 (0.80)	0.25 (0.74)	0.22 (0.75)	0.22 (0.75)	0.22 (0.75)
4	0.75 (0.93)	0.61 (0.88)	0.38 (0.79)	0.43 (0.84)	0.35 (0.80)
6	0.88 (0.96)	0.78 (0.94)	0.67 (0.90)	0.47 (0.83)	0.56 (0.87)
8	0.93 (0.98)	0.86 (0.96)	0.77 (0.93)	0.72 (0.92)	0.55 (0.86)

TABLE 6
Asymptotic convergence rate of the two grid method with aggressive coarsening and block Jacobi smoothing for a finite grid.

Coarsening	Block size				
	1	2	4	6	8
2	0.34	0.29	0.24	0.24	0.23
4	0.72	0.60	0.34	0.41	0.32
6	0.82	0.75	0.65	0.44	0.54
8	0.84	0.80	0.73	0.71	0.53

We can also use LFA to improve the two-grid method, by computing optimal weights. Until now we used $\omega = 0.8$; however, this is optimal only for a block size of one and a coarsening range of two. Table 5 shows the spectral radius of the two-grid method with block Jacobi smoother and an optimal weight ω, i.e., ω is chosen such that the spectral radius is minimized. We see that the optimal ω varies between 0.74 and 0.98. The effect is, however, small; i.e., in comparison to Table 3 the spectral radii change only little.

We now want to evaluate the quality of the LFA prediction. For this purpose we compute the asymptotic convergence rate of the two-grid method for a finite 96×96 grid after 10 iterations with periodic boundary conditions. Table 6 shows the results for the Jacobi smoother with weight $\omega = 0.8$, Table 7 shows the results for the RBBJ smoother, and Table 8 shows the results for the block Jacobi smoother with optimal weight. Comparing these tables to the LFA predictions in Tables 3, 4, and 5, we see that the predictions are quite accurate.

TABLE 7

Asymptotic convergence rate of the two-grid method with aggressive coarsening and RBBJ smoothing for a finite grid.

Coarsening	Block size				
	1	2	4	6	8
2	0.06	0.02	0.01	0.01	0.01
4	0.50	0.30	0.06	0.14	0.05
6	0.71	0.53	0.38	0.14	0.26
8	0.78	0.68	0.54	0.46	0.22

TABLE 8

Asymptotic convergence rate of the two-grid method with aggressive coarsening and block Jacobi smoother with optimal weight (parentheses) for a finite grid.

Coarsening	Block size				
	1	2	4	6	8
2	0.34 (0.80)	0.23 (0.74)	0.21 (0.75)	0.20 (0.75)	0.22 (0.75)
4	0.70 (0.93)	0.61 (0.88)	0.34 (0.79)	0.38 (0.84)	0.32 (0.80)
6	0.81 (0.96)	0.74 (0.94)	0.60 (0.90)	0.42 (0.83)	0.49 (0.87)
8	0.87 (0.98)	0.80 (0.96)	0.71 (0.93)	0.67 (0.92)	0.50 (0.86)

6. The analysis of periodic stencil operators II. Continuing section 3 we will now prove Theorem 3.5. The proof is based on the following observation: Define the operator $T_{\mathbf{k}} : \mathcal{G}_{\mathbf{h}} \to \mathcal{G}_{\mathbf{h} \cdot \mathbf{n}}$ by

$$(T_{\mathbf{k}} u)(\mathbf{x}) = u(\mathbf{x} + \mathbf{t_k}) \quad \text{for} \quad \mathbf{x} \in \Omega_{\mathbf{h} \cdot \mathbf{n}} \,,$$

where $\mathbf{t_k} := (\mathbf{k} - \mathbf{1}) \cdot \mathbf{h}$. If A is a periodic stencil operator we then have

$$(16) \qquad\qquad T_{\mathbf{k}} A = T_{\mathbf{k}} A^{(\mathbf{k})} \,,$$

where $A^{(\mathbf{k})}$ is the constant stencil operator given by $s_{\mathbf{t_k}}$. That means at the points $(\mathbf{t_k} + \Omega_{\mathbf{h} \cdot \mathbf{n}})$ the result $f = Au$ will behave like it has been computed by a constant stencil operator. Now $T_{\mathbf{k}} u$ is a grid function on $\Omega_{\mathbf{h} \cdot \mathbf{n}}$. Thus we can apply the DTFT to $T_{\mathbf{k}} u$ to get $\hat{g}_{\mathbf{k}} := \mathcal{F}_{\mathbf{h} \cdot \mathbf{n}} T_{\mathbf{k}} Au$. We will show that due to observation (16) the frequency function $\hat{g}_{\mathbf{k}}$ can be computed from $\hat{u}$ (instead of u) by the knowledge of the symbol $\widehat{A}^{(\mathbf{k})}$ which is easily obtained by (4) as $A^{(\mathbf{k})}$ is a constant stencil operator. In the next step we show that the functions $\hat{g}_{\mathbf{k}}$ can be combined to get $\hat{f} = R_{\mathbf{n}} \mathcal{F}_h Au$. Thus in the end we can compute $\hat{f}$ from $\hat{u}$.

In the following it will be helpful to combine the operators $T_{\mathbf{k}}$ into a single operator $T : \mathcal{G}_{\mathbf{h}} \to (\mathcal{G}_{\mathbf{h} \cdot \mathbf{n}})^{\mathbf{n}}$ by $(Tu)_{\mathbf{k}} = T_{\mathbf{k}} u$. The combined operator T will be called the *space splitting* operator. On the space $(\mathcal{G}_{\mathbf{h} \cdot \mathbf{n}})^{\mathbf{n}}$ a scalar product can be defined by

$$\langle f, g \rangle := \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{n}} \langle f_{\mathbf{k}}, g_{\mathbf{k}} \rangle \,.$$

Then T is an isometry between $\mathcal{G}_{\mathbf{h}}$ and $(\mathcal{G}_{\mathbf{h} \cdot \mathbf{n}})^{\mathbf{n}}$.

LEMMA 6.1. *The operator $T_{\mathbf{k}}$ has a block Fourier symbol, i.e.,*

$$\mathcal{F}_{\mathbf{h} \cdot \mathbf{n}} T_{\mathbf{k}} = \widehat{T}_{\mathbf{k}} R_{\mathbf{n}} \mathcal{F}_{\mathbf{h}} \,,$$

where

(17) $$\widehat{T}_{\mathbf{kj}} = D_{\mathbf{kk}} F_{\mathbf{kj}}, \qquad D_{\mathbf{kj}}(\vartheta) = \delta_{\mathbf{kj}} \cdot e^{\mathrm{i}\langle \vartheta, \mathbf{t_k} \rangle}.$$

Furthermore

(18) $$\mathcal{F}_{\mathbf{h \cdot n}} T = (DF) R_{\mathbf{n}} \mathcal{F}_{\mathbf{h}},$$

where we define for $u \in \mathcal{G}^{\mathbf{n}}_{\mathbf{h \cdot n}}$ *the elementwise DTFT* $\mathcal{F}_{\mathbf{h \cdot n}} : \mathcal{G}^{\mathbf{n}}_{\mathbf{h \cdot n}} \to \mathcal{H}^{\mathbf{n}}_{\mathbf{h \cdot n}}$ *by the equation* $(\mathcal{F}_{\mathbf{h \cdot n}} u)_{\mathbf{j}} := \mathcal{F}_{\mathbf{h \cdot n}} u_{\mathbf{j}}$.

Proof. The assertion will be proven in two steps. The first step is to show that the Fourier representation

$$\mathcal{F}_{\mathbf{h \cdot n}} T_{\mathbf{1}} = \widehat{T}_{\mathbf{1}} R_{\mathbf{n}} \mathcal{F}_{\mathbf{h}},$$

of $T_{\mathbf{1}}$ has a symbol $\widehat{T}_{\mathbf{1}}$, where $\widehat{T}_{\mathbf{1}}$ is a row vector with

$$\widehat{T}_{\mathbf{1j}}(\vartheta) = \tfrac{1}{\mathrm{vol}_{\mathbf{n}}^{1/2}}.$$

In the second step we use this result to prove the general case.

For $x \in \Omega_{\mathbf{h \cdot n}}$ we have

$$(T_{\mathbf{1}} u)(x) = u(x) = \tfrac{\mathrm{vol}_h^{1/2}}{(2\pi)^{d/2}} \int_{\Theta_h} \hat{u}(\vartheta) \, e^{\mathrm{i}\langle \vartheta, x \rangle} \, \mathrm{d}\vartheta.$$

By using the measurable partition $\Phi = (\mathbf{s_j} + \Theta_{\mathbf{h \cdot n}})_{\mathbf{j}=\mathbf{1}}^{\mathbf{n}}$ of $\Theta_{\mathbf{h}}$, i.e., $\Theta_{\mathbf{h}} = \dot{\bigcup}_{s \in \mathcal{S}} \Phi_s$ and Φ_s is measurable, it follows that

$$(T_{\mathbf{1}} u)(x) = \tfrac{\mathrm{vol}_h^{1/2}}{(2\pi)^{d/2}} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{n}} \int_{\Phi_j} \hat{u}(\vartheta) \, e^{\mathrm{i}\langle \vartheta, x \rangle} \, \mathrm{d}\vartheta$$

and a variable substitution leads to

$$(T_{\mathbf{1}} u)(x) = \tfrac{\mathrm{vol}_h^{1/2}}{(2\pi)^{d/2}} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{n}} \int_{\Theta_{\mathbf{h \cdot n}}} \hat{u}(\vartheta + \mathbf{s_j}) \, e^{\mathrm{i}\langle \vartheta + \mathbf{s_j}, x \rangle} \, \mathrm{d}\vartheta$$

$$= \tfrac{1}{\mathrm{vol}_{\mathbf{n}}^{1/2}} \cdot \tfrac{\mathrm{vol}_{\mathbf{h \cdot n}}^{1/2}}{(2\pi)^{d/2}} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{n}} \int_{\Theta_{\mathbf{h \cdot n}}} \hat{u}(\vartheta + \mathbf{s_j}) \, e^{\mathrm{i}\langle \vartheta, x \rangle} e^{\mathrm{i}\langle \mathbf{s_j}, x \rangle} \, \mathrm{d}\vartheta.$$

Now $e^{\mathrm{i}\langle \mathbf{s_j}, x \rangle} = 1$ for $x \in \Omega_{\mathbf{h \cdot n}}$ (cf. (5)), thus

$$(T_{\mathbf{1}} u)(x) = \tfrac{\mathrm{vol}_{\mathbf{h \cdot n}}^{1/2}}{(2\pi)^{d/2}} \int_{\Theta_{\mathbf{h \cdot n}}} \left(\tfrac{1}{\mathrm{vol}_{\mathbf{n}}^{1/2}} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{n}} R_{\mathbf{nj}} \hat{u}(\vartheta) \right) \cdot e^{\mathrm{i}\langle \vartheta, x \rangle} \, \mathrm{d}\vartheta = \left(\mathcal{F}_{\mathbf{h \cdot n}}^{-1} \widehat{T}_{\mathbf{1}} R_{\mathbf{n}} \mathcal{F}_h u \right)(x).$$

This completes the first step. Now we can turn to the second step. Consider the shift operator $(Z_{\mathbf{k}} u)(x) := u(x + \mathbf{t_k})$. We know that $\widehat{Z}_{\mathbf{k}}(\vartheta) = e^{\mathrm{i}\langle \vartheta, \mathbf{t_k} \rangle}$. Furthermore it is easy to see that $T_{\mathbf{k}} = T_{\mathbf{1}} Z_{\mathbf{k}}$. Hence

$$\mathcal{F}_{\mathbf{h \cdot n}} T_{\mathbf{k}} = \mathcal{F}_{\mathbf{h \cdot n}} T_{\mathbf{1}} Z_{\mathbf{k}} = \widehat{T}_{\mathbf{1}} R_{\mathbf{n}} \mathcal{F}_h Z_{\mathbf{k}} = \widehat{T}_{\mathbf{1}} \left(R_{\mathbf{n}} \widehat{Z}_{\mathbf{k}} R_{\mathbf{n}}^{-1} \right) R_{\mathbf{n}} \mathcal{F}_h.$$
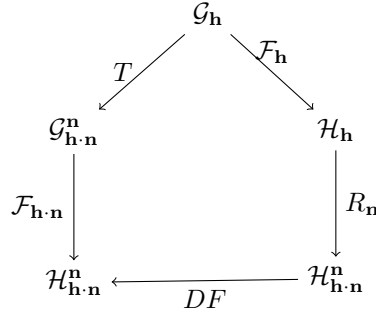
$$\mathcal{G}_{\mathbf{h}}$$

$$T \qquad \mathcal{F}_{\mathbf{h}}$$

$$\mathcal{G}_{\mathbf{h}\cdot\mathbf{n}}^{\mathbf{n}} \qquad\qquad \mathcal{H}_{\mathbf{h}}$$

$$\mathcal{F}_{\mathbf{h}\cdot\mathbf{n}} \qquad\qquad R_{\mathbf{n}}$$

$$\mathcal{H}_{\mathbf{h}\cdot\mathbf{n}}^{\mathbf{n}} \longleftarrow \qquad \mathcal{H}_{\mathbf{h}\cdot\mathbf{n}}^{\mathbf{n}}$$
$$DF$$

FIG. 12. *Commuting diagram of the block discrete-"time" Fourier transform.*

And this leads to

$$\widehat{T}_{\mathbf{kj}} = \left(\widehat{T}_{\mathbf{1}} \left(R_{\mathbf{n}} \widehat{Z}_{\mathbf{k}} R_{\mathbf{n}}^{-1} \right) \right)_{\mathbf{j}} = \sum_{\mathbf{r}=1}^{\mathbf{n}} \widehat{T}_{\mathbf{1r}} \cdot \left(R_{\mathbf{n}} \widehat{Z}_{\mathbf{k}} R_{\mathbf{n}}^{-1} \right)_{\mathbf{rj}} .$$

We have $(R_{\mathbf{n}} \widehat{Z}_{\mathbf{k}} R_{\mathbf{n}}^{-1})_{\mathbf{rj}}(\vartheta) = \delta_{\mathbf{rj}} \widehat{Z}_{\mathbf{k}}(\vartheta + \mathbf{s_r}) = \delta_{\mathbf{rj}} e^{i\langle \vartheta + \mathbf{s_r}, \mathbf{t_k}\rangle}$. Thus

$$\widehat{T}_{\mathbf{kj}} = \sum_{\mathbf{r}=1}^{\mathbf{n}} \frac{1}{\mathrm{vol}_{\mathbf{n}}^{1/2}} \cdot \delta_{\mathbf{rj}} e^{i\langle \vartheta + \mathbf{s_r}, \mathbf{t_k}\rangle} = \frac{1}{\mathrm{vol}_{\mathbf{n}}^{1/2}} e^{i\langle \vartheta + \mathbf{s_j}, \mathbf{t_k}\rangle} = e^{i\langle \vartheta, \mathbf{t_k}\rangle} \cdot \frac{1}{\mathrm{vol}_{\mathbf{n}}^{1/2}} e^{i\langle \mathbf{s_j}, \mathbf{t_k}\rangle}$$
$$= D_{\mathbf{kk}}(\vartheta) F_{\mathbf{kj}} . \qquad\qquad \square$$

Equation (18) basically states that the diagram in Figure 12 commutes. And thus as T, $\mathcal{F}_{\mathbf{h}}$, $\mathcal{F}_{\mathbf{h}\cdot\mathbf{n}}$, and $R_{\mathbf{n}}$ are isometries, so is (DF). We can now prove the main result.

Proof of Theorem 3.5. Let $u \in \mathcal{G}_{\mathbf{h}}$. We begin by computing the vector $\hat{g} := \mathcal{F}_{\mathbf{h}\cdot\mathbf{n}} T A u$.

We use the definition of the elementwise DTFT and observation (16), i.e., the fact that a periodic stencil operator looks like a constant stencil operator when multiplied from the left by $T_{\mathbf{k}}$, to obtain

$$\hat{g}_{\mathbf{k}} := \mathcal{F}_{\mathbf{h}\cdot\mathbf{n}} T_{\mathbf{k}} A u = \mathcal{F}_{\mathbf{h}\cdot\mathbf{n}} T_{\mathbf{k}} A^{(\mathbf{k})} u .$$

We can now use Lemma 6.1 and then Lemma 3.3 to express $\hat{g}$ in terms of $\hat{u} := R_{\mathbf{n}} \mathcal{F}_{\mathbf{h}} u$ by

$$\hat{g}_{\mathbf{k}} = \widehat{T}_{\mathbf{k}} \left(R_{\mathbf{n}} \widehat{A}^{(\mathbf{k})} R_{\mathbf{n}}^{-1} \right) R_{\mathbf{n}} \mathcal{F}_{\mathbf{h}} u = \widehat{T}_{\mathbf{k}} \left(R_{\mathbf{n}} \widehat{A}^{(\mathbf{k})} R_{\mathbf{n}}^{-1} \right) \hat{u} .$$

Recall that $\widehat{T}_{\mathbf{k}}$ is a row vector and $(R_{\mathbf{n}} \widehat{A}^{(\mathbf{k})} R_{\mathbf{n}}^{-1})$ is a diagonal matrix. Hence, $\widehat{T}_{\mathbf{k}}(R_{\mathbf{n}} \widehat{A}_{\mathbf{n}}^{(\mathbf{k})} R_{\mathbf{n}}^{-1})$ is a row vector with

$$\left(\widehat{T}_{\mathbf{k}} \left(R_{\mathbf{n}} \widehat{A}^{(\mathbf{k})} R_{\mathbf{n}}^{-1} \right) \right)_{\mathbf{j}} (\vartheta) = D_{\mathbf{kk}}(\vartheta) \cdot F_{\mathbf{kj}} \cdot \widehat{A}^{(\mathbf{k})}(\vartheta + \mathbf{s_j}) = D_{\mathbf{kk}}(\vartheta) \cdot G_{\mathbf{kj}}(\vartheta) ,$$

where we used (17). Thus

$$\hat{g}_{\mathbf{k}} = \sum_{\mathbf{j}} D_{\mathbf{kk}} \cdot G_{\mathbf{kj}} \cdot \hat{u}_{\mathbf{j}}$$

and the whole vector $\hat{g}$ can be computed by $\hat{g} = DG\hat{u}$.

As we now have a relation between $\hat{u}$ and $\hat{g}$ it remains to combine the $\hat{g}_\mathbf{k}$ to obtain $\hat{f} := R_\mathbf{n}\mathcal{F}_\mathbf{h} Au$. We do that by using Lemma 6.1 again (cf. Figure 12). From (18) we know that

$$\hat{f} = R_\mathbf{n}\mathcal{F}_\mathbf{h} Au = (DF)^{-1}\mathcal{F}_{\mathbf{h}\cdot\mathbf{n}} TAu = F^{-1}D^{-1}\hat{g} = F^{-1}D^{-1}DG\hat{u} = F^{-1}G\hat{u}\,.$$

Thus

$$R_\mathbf{n}\mathcal{F}_\mathbf{h} A = (F^*G)R_\mathbf{n}\mathcal{F}_\mathbf{h} = \widehat{A}R_\mathbf{n}\mathcal{F}_\mathbf{h}\,,$$

as F is the DFT matrix and therefore unitary. □

Appendix A. Matrix multiplication operators. In this appendix we want to discuss the proof of Theorem 3.6. To our knowledge the theorem and its proof cannot be found in the literature. This theorem can be seen as a generalization of Corollary 2.3. For its proof we used the essential range of a function in $L^\infty(X;\mathbb{C})$. Analogously, the essential union of the pointwise spectra $\sigma(a(x))$ for $a \in L^\infty_{\mathbf{n}\times\mathbf{m}}(X;\mathbb{C})$ can be defined [17].

In this section let μ be the Lebesgue measure and $X := \Theta_\mathbf{h}$. (However, (X, Σ, μ) can be any σ-finite measure space.) The essential union of the *pointwise computed* spectra $\sigma(a(x))$ is now given by

$$\text{ess-}\bigcup_{x\in X}\sigma(a(x)) := \bigcap_{b\in[a]}\overline{\bigcup_{x\in X}\sigma(b(x))}\,,$$

where $[a]$ is the equivalence class of functions that are almost everywhere equal to a. It can be shown that this is equivalent to

(19)
$$\text{ess-}\bigcup_{x\in X}\sigma(a(x)) = \{z \in \mathbb{C} : \text{ for all } \varepsilon > 0 : \mu(\{x \in X : \sigma(a(x)) \cap U_\varepsilon(z) \neq \emptyset\}) > 0\}\,.$$

PROPOSITION A.1. *Let* $A : L^2(X)^n \to L^2(X)^n$ *be a matrix multiplication operator with nonempty resolvent set* $\rho(A)$ *and matrix symbol* a. *Then*

$$\sigma(A) = \text{ess-}\bigcup_{x\in X}\sigma(a(x))\,.$$

Proof. See [17] for the proof. □

So the spectrum of the matrix multiplication operator is the essential union of the pointwise spectra. This result can be used to show the following result for the spectral radius.

LEMMA A.2. *Let* A *be as defined in Proposition* A.1. *The spectral radius* $r(A)$ *of* A *is the essential supremum of the pointwise computed spectral radii of* $a(\vartheta)$, *i.e.,*

$$r(A) = \operatorname*{ess\,sup}_{\vartheta\in X} r(a(\vartheta))\,.$$

Proof. We set

(20)
$$C := \operatorname*{ess\,sup}_{\vartheta\in X} r(a(\vartheta)) = \inf\{\alpha > 0 : \mu(\{x \in X : r(a(x)) > \alpha\}) = 0\}\,.$$

It is known that

$$r(A) = \max_{\lambda\in\sigma(A)} |\lambda|\,.$$

Thus we can prove the assertion by showing that

   (i) for every $\varepsilon' > 0$ and all $|\lambda| \geq C + \varepsilon'$ we have $\lambda \in \rho(A)$,
   (ii) for every $\varepsilon' > 0$ there exists $\lambda \in \sigma(A)$ with $|\lambda| \geq C - \varepsilon'$.
To show (i) we choose $\lambda \in \mathbb{C}$ such that $|\lambda| \geq C + \varepsilon'$. Then if we can find $\varepsilon > 0$ such that

$$\mu(\{x \in X : \sigma(a(x)) \cap U_\varepsilon(\lambda) \neq \emptyset\}) = 0 \tag{21}$$

the assertion (i) is proven due to Proposition A.1 and (19). If the set $\{x \in X : \sigma(a(x)) \cap U_\varepsilon(\lambda) \neq \emptyset\}$ is empty, then its measure is clearly zero ((21) holds). Thus we consider the case where this set is not empty. In this case we can choose $x_0 \in X$ such that $\sigma(a(x_0)) \cap U_\varepsilon(\lambda) \neq \emptyset$. And thus we can pick $\tilde{\lambda} \in \sigma(a(x_0)) \cap U_\varepsilon(\lambda)$. Then

$$|\tilde{\lambda}| = |\lambda - (\lambda - \tilde{\lambda})| \geq |\lambda| - |\lambda - \tilde{\lambda}| > C + \varepsilon' - \varepsilon \,,$$

as $\tilde{\lambda} \in U_\varepsilon(\lambda)$. If we choose $\varepsilon = \varepsilon'$ we have $|\tilde{\lambda}| > C$. Recall that we choose $\tilde{\lambda}$ to be an eigenvalue of the matrix $a(x_0)$. Thus we know that the spectral radius $r(a(x_0))$ fulfills

$$r(a(x_0)) \geq |\tilde{\lambda}| > C \,.$$

So as x_0 was an arbitrary element from $\sigma(a(x_0)) \cap U_\varepsilon(\lambda)$ we have shown that

$$\{x \in X : \sigma(a(x)) \cap U_\varepsilon(\lambda) \neq \emptyset\} \subseteq \{x \in X : r(a(x)) > C\} \,.$$

By definition of the essential supremum and the constant C in (20) the measure of the right-hand side is zero. Thus

$$\mu(\{x \in X : \sigma(a(x)) \cap U_\varepsilon(\lambda) \neq \emptyset\}) \leq \mu(\{x \in X : r(a(x)) > C\}) = 0$$

for all $|\lambda| > C + \varepsilon'$. Thus we arrive at (21) and the assertion (i) is established.

   We want to prove (ii) by contradiction. Assume that there exists $\varepsilon' > 0$ such that $|\lambda| < C - \varepsilon'$ implies $\lambda \in \rho(A)$. Consider the set $\{x \in X : r(a(x)) > C - \varepsilon\}$. We want to show that this set has measure zero, which would be a contradiction to the definition of C in (20). To this extent define the set

$$M := \{\lambda \in \mathbb{C} : C - \varepsilon' \leq |\lambda| \leq C\} \,.$$

Then

$$\{x \in X : r(a(x)) > C - \varepsilon\} \subseteq \{x \in X : r(a(x)) > C\} \cup \{x \in X : \sigma(a(x)) \cap M \neq \emptyset\} \,.$$

The first set on the right-hand size has zero measure due to the definition of C in (20). Thus it remains to show that the second set also has zero measure. By the definition of the essential union of the pointwise spectra (19) we have that for every $\lambda \in \rho(A)$ there exists $\varepsilon(\lambda) > 0$ such that

$$\mu(\{x \in X : \sigma(a(x)) \cap U_{\varepsilon(\lambda)}(\lambda) \neq \emptyset\}) = 0 \,. \tag{22}$$

We have that $M \subseteq (\bigcup_{\lambda \in M} U_{\varepsilon(\lambda)}(\lambda))$. Thus we have covered the compact set M by a family of open sets. By that we have finitely many $\lambda_1, \ldots, \lambda_m \in M$ such that

$$M \subseteq \bigcup_{i=1}^m U_{\varepsilon(\lambda_i)}(\lambda_i) \,.$$

This yields

$$\{x \in X : \sigma(a(x)) \cap M \neq 0\} \subseteq \bigcup_{i=1}^{m} \{x \in X : \sigma(a(x)) \cap U_{\varepsilon(\lambda_i)}(\lambda_i) \neq \emptyset\} \,,$$

and as the sets on the right-hand side all have zero measure due to (22) the assertion (ii) is proven. $\qquad \Box$

With that we can complete the final proof of this section.

Proof of Theorem 3.6. We have

$$r(L) = \operatorname*{ess\,sup}_{\vartheta \in \Theta_{\mathbf{h} \cdot \mathbf{n}}} r(\widehat{L}(\vartheta))$$

by Lemma A.2. Thus

$$\|L\|^2 = \|L^* L\| = r(L^* L) = \operatorname*{ess\,sup}_{\vartheta \in \Theta_{\mathbf{h} \cdot \mathbf{n}}} r(\widehat{L}(\vartheta)^* \widehat{L}(\vartheta)) \,. \qquad \Box$$

REFERENCES

[1] R. Alcouffe, A. Brandt, J. Dendy, Jr., and J. Painter, *The multi-grid method for the diffusion equation with strongly discontinuous coefficients*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 430–454.

[2] H. Bin Zubair, C. W. Oosterlee, and R. Wienands, *Multigrid for high-dimensional elliptic partial differential equations on non-equidistant grids*, SIAM J. Sci. Comput., 29 (2007), pp. 1613–1636.

[3] A. Brandt, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390.

[4] A. Brandt, *Rigorous quantitative analysis of multigrid,* I*: Constant coefficients two-level cycle with L_2-norm*, SIAM J. Numer. Anal., 31 (1994), pp. 1695–1730.

[5] A. Brandt and O. E. Livne, *Multigrid Techniques.* 1984 *Guide with Applications to Fluid Dynamics*, Classics in Appl. Math. 67, SIAM, Philadelphia, 2011.

[6] J. Brannick, X. Hu, C. Rodrigo, and L. Zikatanov, *Local Fourier analysis of multigrid methods with polynomial smoothers and aggressive coarsening*, Numer. Math. Theory Methods Appl., 8 (2015), pp. 1–21.

[7] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, 2nd ed., SIAM, Philadelphia, 2000.

[8] J. E. Dendy, Jr., *Black box multigrid*, J. Comput. Phys., 48 (1982), pp. 366–386.

[9] S. Friedhoff and S. MacLachlan, *A generalized predictive analysis tool for multigrid methods*, Numer. Linear Algebra Appl., 22 (2015), pp. 618–647.

[10] F. J. Gaspar, J. L. Gracia, and F. J. Lisbona, *Fourier analysis for multigrid methods on triangular grids*, SIAM J. Sci. Comput., 31 (2009), pp. 2081–2102.

[11] F. J. Gaspar, J. L. Gracia, F. J. Lisbona, and C. Rodrigo, *On geometric multigrid methods for triangular grids using three-coarsening strategy*, Appl. Numer. Math., 59 (2009), pp. 1693–1708.

[12] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, MD, 1989.

[13] W. Hackbusch, *Multi-Grid Methods and Applications*, Springer-Verlag, Berlin, 1985.

[14] P. W. Hemker, W. Hoffmann, and M. H. van Raalte, *Two-level Fourier analysis of a multigrid approach for discontinuous Galerkin discretization*, SIAM J. Sci. Comput., 25 (2003), pp. 1018–1041.

[15] P. W. Hemker, W. Hoffmann, and M. H. van Raalte, *Fourier two-level analysis for discontinuous Galerkin discretization with linear elements*, Numer. Linear Algebra Appl., 11 (2004), pp. 473–491.

[16] P. W. Hemker and M. H. van Raalte, *Fourier two-level analysis for higher dimensional discontinuous Galerkin discretisation*, Comput. Vis. Sci., 7 (2004), pp. 159–172.

[17] A. Holderrieth, *Matrix multiplication operators generating one parameter semigroups*, Semigroup Forum, 42 (1991), pp. 155–166.

[18] C.-C. J. KUO AND B. C. LEVY, *Two-color Fourier analysis of the multigrid method with red-black Gauss-Seidel smoothing*, Appl. Math. Comput., 29 (1989), pp. 69–87.

[19] O. E. LIVNE AND A. BRANDT, *Local mode analysis of multicolor and composite relaxation schemes*, Comput. Math. Appl., 47 (2004), pp. 301–317.

[20] S. P. MACLACHLAN AND C. W. OOSTERLEE, *Local Fourier analysis for multigrid with overlapping smoothers applied to systems of PDEs*, Numer. Linear Algebra Appl., 18 (2011), pp. 751–774.

[21] W. RUDIN, *Functional Analysis*, McGraw-Hill Series in Higher Mathematics, McGraw-Hill, New York, 1973.

[22] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.

[23] E. STROHMAIER, J. DONGARRA, H. SIMON, AND M. MEUER, *TOP500 November* 2015, http://top500.org/lists/2015/11/.

[24] K. STÜBEN AND U. TROTTENBERG, *Multigrid methods: Fundamental algorithms, model problem analysis and applications*, in Multigrid Methods (Cologne, 1981), Lecture Notes in Math. 960, Springer-Verlag, Berlin, 1982, pp. 1–176.

[25] U. TROTTENBERG, C. W. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, New York, 2001.

[26] R. WIENANDS AND W. JOPPICH, *Practical Fourier Analysis for Multigrid Methods*, Chapman & Hall/CRC Press, Boca Raton, FL, 2005.

[27] R. WIENANDS AND C. W. OOSTERLEE, *On three-grid Fourier analysis for multigrid*, SIAM J. Sci. Comput., 23 (2001), pp. 651–671.

[28] J. XU, *The method of subspace corrections*, J. Comput. Appl. Math., 128 (2001), pp. 335–362.

[29] I. YAVNEH, *Multigrid smoothing factors for red-black Gauss–Seidel relaxation applied to a class of elliptic operators*, SIAM J. Numer. Anal., 32 (1995), pp. 1126–1138.