A semi-algebraic approach that enables the design of inter-grid operators to optimize multigrid convergence

Pablo Navarrete Michelini[1, 2, *, †] and Edward J. Coyle[3]

[1]*Center for Wireless Systems and Applications, Purdue University, 465 Northwestern Ave., West Lafayette, IN 47907-2035, U.S.A.*
[2]*Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, Santiago, RM 8370451, Chile*
[3]*School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Dr. NW, Atlanta, GA 30332-0250, U.S.A.*

SUMMARY

We study the effect of inter-grid operators—the interpolation and restriction operators—on the convergence of two-grid algorithms for linear models. We show how a modal analysis of linear systems, along with some assumptions on the normal modes of the system, allows us to understand the role of inter-grid operators in the speed and accuracy of a full-multigrid step.

We state an assumption that generalizes local Fourier analysis (LFA) by means of a precise description of aliasing effects on the system. This assumption condenses, in a single algebraic property called the harmonic aliasing property, all the information needed from the geometry of the discretization and the structure of the system's eigenvectors. We first state a harmonic aliasing property based on the standard coarsening strategies of 1D problems. Then, we extend this property to a more aggressive coarsening typically used in 2D problems with the help of additional assumptions on the structure of the system matrix.

Under our general assumptions, we determine the exact rates at which groups of modal components of the error evolve and interact. With this knowledge, we are then able to design inter-grid operators that optimize the two-grid algorithm convergence. By different choices of operators, we verify the classic heuristics based on Fourier harmonic analysis, show a trade-off between the rate of convergence and the number of computations required per iteration, and show how our analysis differs from LFA.

KEY WORDS: multigrid algorithms; inter-grid operators; convergence analysis; modal analysis; aliasing

*Correspondence to: Pablo Navarrete Michelini, Departamento de Ingeniería Eléctrica, Universidad de Chile, Av. Tupper 2007, Santiago, RM 8370451, Chile.
†E-mail: pnavarre@purdue.edu

1. INTRODUCTION

We are interested in applications of the multigrid algorithm in the distributed sensing and processing tasks that arise in the design of wireless sensor networks. In such scenarios, the inexpensive, low-power, low-complexity sensor motes that are the nodes of the network must perform all computation and communication tasks. This is very different than the scenarios encountered in the implementation of multigrid algorithms on large parallel machines for the following reasons:

- Sensor motes are battery powered and must operate unattended for long periods of time. The design of algorithms that run on them must therefore attempt to minimize the number of computations each node must perform and the number of times it must communicate because both functions consume energy. Of the two functions, communication is the most energy intensive per bit of data.
- Communication between sensor motes is carried out in hop-by-hop fashion, since the energy required to send data over a distance d is proportional to d^α with $2 \leqslant \alpha \leqslant 4$. Thus, the sensor motes communicate directly only with their nearest neighbors in any direction.
- Re-executing an algorithm after adjusting parameters or models is very difficult or might not even be possible because of the remote deployment of the network. It is thus critical that the algorithms used to perform various tasks be as robust and well understood as possible before they are deployed.

In implementations of multigrid algorithms on networks like these, as in many other applications of multigrid algorithms, it is thus essential that the convergence rate of the algorithm be optimized. This minimizes the number of communication and computation steps of the algorithm. It also leads to interesting insights in the design of each step, highlighting both trade-offs between the different costs of computations within each node and communications between nodes, and the need for low complexity in each step of the algorithm.

Finally, in such applications the multigrid methods must be very robust in order to ensure the continuous operation of the whole system. This task is difficult because it is likely that the system model varies throughout the field. The current theory of algebraic multigrid (AMG) offers one possible solution to this problem [1–4]. Unfortunately, the convergence results obtained so far in the theory of AMG are not as strong as the theory for linear operators with constant stencil coefficients [5]. As optimal convergence behavior is critical under our particular distributed scenario, we seek a more flexible yet still rigorous convergence analysis.

The goal of this paper is thus to introduce a new convergence analysis based on a modal decomposition of the system and a precise description of aliasing phenomena on coarse systems. The purpose of this analysis is to provide tools that enable the design of coarsening strategies as well as inter-grid and smoothing operators. We try to stay close to the technique of local Fourier analysis (LFA)[‡] introduced by Achi Brandt [5, 6] as it is a powerful technique for quantitative convergence analysis. The essential difference between LFA and our approach is that we drop the requirement of constant stencil coefficients. By doing so, the eigenvectors of a linear operator will no longer be the so-called *harmonic grid functions* used in LFA [7], which in this paper we call

[‡]Originally called local mode analysis (LMA); we chose the nomenclature used in [7] as it emphasizes the essential difference with the approach introduced in this paper.

Fourier harmonic modes. The properties of the system must thus be constrained in some way in order to develop new tools for convergence analysis. The requirement we focus on is an explicit description of the aliasing effects produced by the coarsening strategy.

The aliasing of Fourier harmonic modes is present in LFA through the concept of *spaces of harmonics* [7]. We identify its simple form as one of the reasons why LFA is so powerful. Based on this fact, we assume a more general aliasing pattern that still allows us to characterize convergence behavior. This assumption condenses, in a single algebraic property called the harmonic aliasing property, all the information needed from the geometry of the discretization and the structure of the eigenvectors. If this property is satisfied, then no more information is needed from the system and the analysis is completely algebraic. Therefore, our analysis could be considered a semi-algebraic approach to the study of convergence issues and the design of efficient inter-grid operators.

One of the practical advantages of our approach is that we are able to separate the problem of coarsening from what we call *filtering*, i.e. interpolation/restriction weights and smoothing operations. The analysis of each problem makes no use of heuristics. The coarsening strategy is designed to ensure a convenient aliasing pattern whereas the design of the filters is meant to optimize multigrid convergence.

The main difficulty of our approach is the dependence of the assumptions on the eigenvectors of the system. In practical applications, it is very unlikely that this information is available. Therefore the verification of the assumptions remains unsolved. Nevertheless, this problem is also shared in many fields in which transient or local phenomena do not allow a proper use of Fourier analysis [8]. There have been many efforts to identify suitable bases for specific problems and the goal of this work is to open this problem in multigrid analysis. For these reasons, the results of this paper are not entirely conclusive about optimization strategies for coarsening and filtering. They are, however, an important first step toward this goal.

In Section 2 we provide the notation and the essential properties of the multigrid algorithm for further analysis. In Section 3 we list the assumptions needed on the algorithm and system in order to apply our analysis. In Section 4 we list the additional assumptions needed on 2D systems in order to extend our analysis. In Section 5 we derive the main results about the influence of inter-grid operators on multigrid convergence and verify the classic heuristics of Fourier harmonic analysis. In Section 6 we provide examples that show how to use our analysis and also on how our analysis differs from the classical LFA.

2. THE ELEMENTS OF MULTIGRID ALGORITHMS

We wish to solve discrete linear systems of the form $Au = f$, defined on a grid Ω^h with step size $h \in \mathbb{R}^+$ defined as the largest distance between neighboring grid nodes. A *coarse grid* Ω^s is defined as a set of nodes such that $\Omega^s \subset \Omega^h$ and $s > h$.

We define the so-called *inter-grid operators*, regardless of their use in the multigrid algorithm, as any linear transformation between scalar fields on Ω^h and Ω^s. That is,

$$I_s^h \in \mathbb{R}^{|\Omega^h| \times |\Omega^s|} \quad \text{and} \quad I_h^s \in \mathbb{R}^{|\Omega^s| \times |\Omega^h|} \tag{1}$$

where I_s^h is the *interpolation operator* and I_h^s is the *restriction operator*. We introduce a notation with markers ‘ ˇ ’ or ‘ ˆ ’ to indicate transfers from a finer or coarser grid, respectively. We are

then interested in the following operations:

$$\check{x} = I_h^s x, \quad x \in \mathbb{R}^{|\Omega^h|} \tag{2}$$

$$\hat{y} = I_s^h y, \quad y \in \mathbb{R}^{|\Omega^s|} \tag{3}$$

and

$$\check{A} = I_h^s A I_s^h, \quad A \in \mathbb{R}^{|\Omega^h| \times |\Omega^h|} \tag{4}$$

The definition of the coarsening operator in (4) follows the Galerkin condition and is standard in most multigrid applications [9].

We consider a full two-grid approach consisting of a *nested iteration* step, as shown in Figure 1, and μ_1 iterations of the *Correction Scheme*, including ν_1 pre-smoothing and ν_2 post-smoothing iterations, as shown in Figure 2. Here, the vector v_k is the kth approximation of the exact solution of the linear system, $u \in \mathbb{R}^{|\Omega^h|}$. Similarly, the vector $e_k = u - v_k$ is the approximation error after the kth step of the algorithm. One smoothing iteration is characterized by the *smoothing operator* S; after each iteration the approximation error evolves as $e_{k+1} = Se_k$. Because of this property we also call S the *smoothing filter*.

From these diagrams, it follows that the approximation error between smoothing iterations in the correction scheme is given by

$$e_{\nu_1+1} = K e_{\nu_1} \tag{5}$$

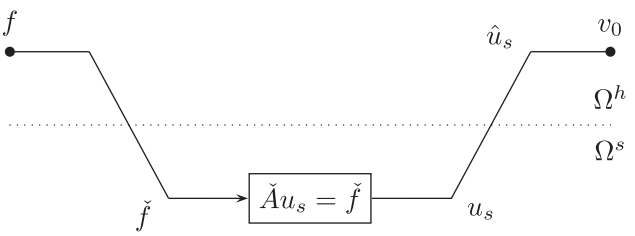

Figure 1. Diagram of a nested iterations step. The dotted line separates problems from the fine and coarse grid domains. The interpolation (restriction) operation is applied to vectors crossing the dotted line from below (above).

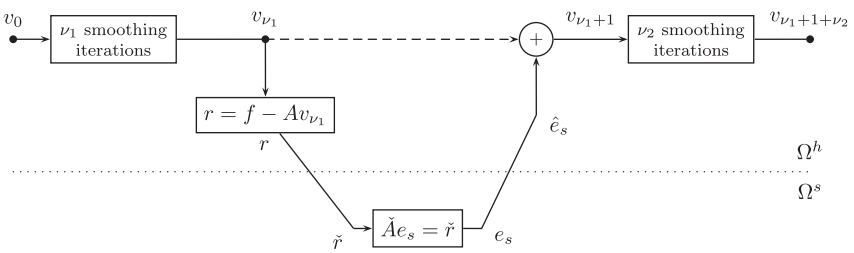

Figure 2. Diagram of a correction scheme step using ν_1 pre-smoothing iterations and ν_2 post-smoothing iterations (e.g. Gauss Seidel, Jacobi, Richardson, etc.). The dotted line separates problems from the fine and coarse grid domains. The interpolation (restriction) operation is applied to vectors crossing the dotted line from below (above).

and similarly, the initial approximation error, e_0, using nested iteration is given by

$$e_0 = Ku \tag{6}$$

where u is the exact solution of the linear system and K is the so-called *coarse grid correction* matrix [10] defined as

$$K = I - I_s^h \check{A}^{-1} I_h^s A \tag{7}$$

This matrix is the target of our analysis in Section 5 as it controls all of the convergence features of the two-grid scheme. Considering the effect of smoothing iterations, the error in the whole correction scheme evolves as

$$e_{\nu_1+1+\nu_2} = S^{\nu_2} K S^{\nu_1} e_0 \tag{8}$$

In the multiple-grid case, a recursive application of nested iterations and the correction scheme is used to solve coarse system equations, as shown in Figure 3. Since coarse systems are not solved with exact accuracy, the approximation error evolves differently. Here, the error depends on the accuracy of the solutions from the coarse grids. Thus, matrix K used above is replaced by a different matrix, denoted by K_1, which is obtained from the following recursions:

$$\begin{aligned} &K_L = 0, \quad A_1 = A \\ &A_j = \check{A}_{j-1}, \quad \text{with } j = 2, \ldots, L-1 \quad \text{and} \\ &K_{j-1} = I - I_j^{j-1}[I - (S_j^{\nu_2} K_j S_j^{\nu_1})^{\mu_j} K_j](\check{A}_{j-1})^{-1} I_{j-1}^j A_{j-1}, \quad \text{with } j = L, \ldots, 2 \end{aligned} \tag{9}$$

where S_j, I_j^{j-1}, and I_{j-1}^j are the smoothing, interpolation, and restriction operators chosen at level j, and μ_j is the number of iterations of the correction scheme used at level j. Then, the approximation error evolves as $e_0 = K_1 u$ in nested iterations and it evolves as $e_{\nu_1+1} = K_1 e_{\nu_1}$ between smoothing iterations of the correction scheme.

Although our analysis is technically applicable to the full multiple-grid case, the coupling between different levels makes the algebra tedious. Therefore, we concentrate on the two-level case and for the multiple-grid case we assume that the problem in coarse levels has been solved with enough accuracy so that matrices $(S_j^{\nu_2} K_j S_j^{\nu_1})^{\mu_j} K_j$ can be neglected and we can work under the two-grid assumptions.

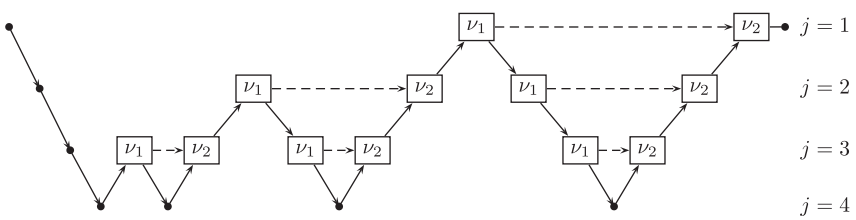

Figure 3. Diagram of the recursive full multigrid approach using one iteration of the correction scheme per level. Each box represents a number of pre- or post-smoothing iterations. The particular choice of using the same combination of pre-/post-smoothing iterations on different correction scheme steps is considered.

3. ASSUMPTIONS ABOUT THE ALGORITHM AND THE SYSTEM

Two assumptions are needed in order to derive our convergence results. First, we introduce a decomposition of the inter-grid interpolation/restriction operators into up-/down-sampling and filtering operations, a standard approach in digital signal processing [8, 11]. Second, we assume that the operators and the system possess the same basis of eigenvectors and we establish a condition on these eigenvectors under (up-/down)-sampling operations. These conditions are motivated by standard Fourier harmonic analysis but they are not restricted to systems with Fourier harmonic modes as eigenvectors.

3.1. System modes

Assuming that A is a diagonalizable square matrix, we define its eigen-decomposition as

$$A = W \Lambda V^{\mathrm{T}} \tag{10}$$

Here, the diagonal matrix Λ contains the eigenvalues of A on its diagonal. The columns of the matrix W are the right-eigenvectors of A, i.e. $AW = W\Lambda$. The columns of the matrix V contain the left-eigenvectors of A, i.e. $V^{\mathrm{T}}A = \Lambda V^{\mathrm{T}}$.

The column vectors of W and V form a *biorthogonal basis* since it follows from the above definitions that

$$V^{\mathrm{T}} W = I \tag{11}$$

If A is a symmetric matrix, then $V = W$ and the column vectors of W form an orthogonal basis.

It is important to note that from this point on our analysis differs from LFA. In LFA it is assumed that the stencil of A, denoted as the row vector s, is not dependent on the position of the grid nodes to which it is applied. When this is true, the operation Ax can be expressed as the convolution:

$$(Ax)_n = \sum_k (s)_k (x)_{n+k} \tag{12}$$

where $(Ax)_n$ denotes the nth component of the vector Ax. This implies that the eigenvectors of A are Fourier harmonic modes. In other words, if $(w)_k = \mathrm{e}^{i\omega k}$ then $Aw = s(\omega)w$ where $s(\omega)$ is the Fourier transform of the stencil sequence. In our analysis, the stencil can depend on the position of the grid nodes to which it is applied. In this case, the operation Ax can be expressed as

$$(Ax)_n = \sum_k (s_n)_k (x)_{n+k} \tag{13}$$

and then the eigenvectors of A need not be Fourier harmonic modes.

Later on we will make assumptions about the eigenvectors of A that are related to the coarsening strategy of the multigrid approach. This does, of course, limit the scope of our analytical approach, but it can still be applied to a broader family of operators than LFA. The examples in Sections 6.2 and 6.3 will make this point very clear.

3.2. Smoothing filters

We assume that the smoothing operator S used in the two-grid algorithm, as defined in Section 2, has the same eigenvectors as A. That is,

$$S = W \Sigma V^{\mathrm{T}} \tag{14}$$

where Σ is a diagonal matrix with the eigenvalues of matrix S. The diagonal values in Σ represent the factor by which each modal component of the approximation error is multiplied after one smoothing iteration.

As in LFA, our analysis is also applicable to smoothers of the form $A_+ e_{k+1} = A_- e_k$ with $A = A_+ - A_-$ [7], e.g. Gauss–Seidel with lexicographical ordering for constant stencil operators, assuming that both A_+ and A_- have the same eigenvectors as A. The smoothing operator is then given by

$$S = W(\Sigma_+)^{-1} \Sigma_- V^{\mathrm{T}} \tag{15}$$

where Σ_+ and Σ_- are diagonal matrices with the eigenvalues of A_+ and A_-, respectively.

3.3. Inter-grid filters

In our analysis of multigrid convergence, it is useful to decompose the inter-grid operators defined in Section 2 into two consecutive operations. For two grid levels, with the fine grid Ω^h and the coarse grid Ω^s, we first identify the operation of selecting nodes from the fine grid for the coarse grid. This leads to the following definitions:

Definition 1 (*Down-/up-sampling matrices*)
The down-sampling matrix $D \in \mathbb{R}^{|\Omega_s| \times |\Omega_h|}$ is defined as

$$(D)_{i,j} = \begin{cases} 1 & \text{if node } j \in \Omega^h \text{ is the } i\text{th selected node} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

The up-sampling matrix $U \in \mathbb{R}^{|\Omega_h| \times |\Omega_s|}$ is defined as

$$U = D^{\mathrm{T}} \tag{17}$$

A similar definition for an unselecting operation which will be useful in Section 6 is

Definition 2 (*Down-/up-unselecting matrices*)
The down-unselecting matrix $\bar{D}$ is defined as

$$(\bar{D})_{i,j} = \begin{cases} 1 & \text{if node } j \in \Omega^h \text{ is the } i\text{th unselected node} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

The up-unselecting matrix $\bar{U}$ is defined as $\bar{U} = \bar{D}^{\mathrm{T}}$.

An important property that follows from these definitions is

$$DU = \tilde{I} \tag{19}$$

where $\tilde{I} \in \mathbb{R}^{|\Omega_s| \times |\Omega_s|}$ is the identity matrix in the coarse grid. On the other hand, the matrix $UD \in \mathbb{R}^{|\Omega_h| \times |\Omega_h|}$ is a diagonal matrix with 1 in the diagonal whenever $i = j$ is a selected node and 0 otherwise.

Now, we can decompose the inter-grid operators I_s^h and I_h^s, as defined in Section 2, into the following matrix products:

$$\begin{aligned} I_s^h &= F_I U, \quad \text{with } F_I \in \mathbb{R}^{|\Omega^h| \times |\Omega^h|} \quad \text{and} \\ I_h^s &= D F_R, \quad \text{with } F_R \in \mathbb{R}^{|\Omega^h| \times |\Omega^h|} \end{aligned} \tag{20}$$

where the square matrices F_I and F_R are called the *interpolation* and *restriction filters*, respectively. Although this kind of decomposition is widely used in digital signal processing [8, 11], it has not been used for convergence analysis of multigrid algorithms. In the case that the variational property $I_h^s = c(I_s^h)^{\mathrm{T}}$ is assumed, the inter-grid filters reduce to a single filter F given by

$$F = F_R = c(F_I)^{\mathrm{T}} \tag{21}$$

The inter-grid operator decomposition applies to any kind of inter-grid operators. Now, we restrict our analysis to the set of inter-grid filters that have the same eigenvectors as the system matrix A. That is, we assume inter-grid filters of the form

$$\begin{aligned} F_I &= W \Pi_I V^{\mathrm{T}} \quad \text{and} \\ F_R &= W \Pi_R V^{\mathrm{T}} \end{aligned} \tag{22}$$

where Π_I and Π_R are diagonal matrices and their diagonal coefficients represent the damping effect of the filters on the corresponding eigenvector.

3.4. The harmonic aliasing property

From its earliest formulation, multigrid heuristics have always been based on Fourier harmonic analysis. The idea of reducing high- and low-frequency components of the approximation error can be found in almost any book or tutorial on the subject. In this paper, we generalize this to a modal analysis where the eigenvectors (or modes) are not necessarily Fourier harmonic modes. We keep the notion of *harmonic* analysis in a more general way. By *harmonic* modes now we mean a set of vectors with a certain property that, generally speaking, will preserve the notion of self-similarity through the aliasing of different modes after down-sampling. As an example, in Section 6.2 we will mention 'square-wave' like functions that do not fit within the scope of LFA. We introduce this property because the aliasing effects of Fourier harmonic modes are essential to revealing the role of the smoothing and inter-grid filters in multigrid convergence. Therefore, we need to define this property for our more general modal analysis.

Since the application of the following property will be constrained to 1D systems, we will start using a subindex x as a label that indicates the dimension where the operations apply. Then, we state the harmonic aliasing property as follows:

Definition 3 (*Harmonic aliasing property*)
A set of biorthogonal eigenvectors, W_x and V_x, and a down-sampling matrix D_x have the harmonic aliasing property if there exists an ordering of eigenvectors for which

$$V_x^{\mathrm{T}} U_x D_x W_x = N_x \tag{23}$$

where $U_x = D_x^{\mathrm{T}}$ is the up-sampling matrix and N_x is the harmonic aliasing pattern that we define to be

$$N_x = \frac{1}{2}\begin{bmatrix} \tilde{I}_x & \tilde{I}_x \\ \tilde{I}_x & \tilde{I}_x \end{bmatrix} \tag{24}$$

We must note that the harmonic aliasing property only involves the eigenvectors of the system and the down-/up-sampling operator. Although this is a strong assumption on the system, it only involves the down-sampling operator from the multigrid algorithm. It does not depend on the smoothing and inter-grid filters. This is an important consequence of the inter-grid operator decomposition.

The definition above implicitly assumes a down-sampling by a factor of 2 and naturally induces a partition of the eigenvectors into two sets, say $W_x = [W_{Lx} W_{Hx}]$ for the right-eigenvectors and $V_x = [V_{Lx} V_{Hx}]$ for the left-eigenvectors. The subscripts Lx and Hx resemble the standard Fourier harmonic analysis used to distinguish between low- and high-frequency modes (see for instance [10]). Using these partitions, we can restate the harmonic aliasing property. For that purpose we state the following definition:

Definition 4 (*Surjective property*)
A set of biorthogonal eigenvectors, W_x and V_x, and a down-sampling matrix D_x have the surjective property if there exists an ordering of the eigenvectors for which the partitions $W_x = [W_{Lx} W_{Hx}]$ and $V_x = [V_{Lx} V_{Hx}]$ fulfill the following conditions:

$$D_x W_{Lx} = D_x W_{Hx} \tag{25}$$

and

$$D_x V_{Lx} = D_x V_{Hx} \tag{26}$$

Theorem 1
The surjective property is equivalent to the harmonic aliasing property.

Proof
First, we have to note that, given the partitions $W_x = [W_{Lx} W_{Hx}]$ and $V_x = [V_{Lx} V_{Hx}]$, we can rewrite the harmonic aliasing property as the following set of biorthogonal relationships:

$$(D_x V_{Lx})^{\mathrm{T}} (D_x W_{Lx}) = \tfrac{1}{2} \tilde{I}_x \tag{27}$$

$$(D_x V_{Lx})^{\mathrm{T}} (D_x W_{Hx}) = \tfrac{1}{2} \tilde{I}_x \tag{28}$$

$$(D_x V_{Hx})^{\mathrm{T}} (D_x W_{Lx}) = \tfrac{1}{2} \tilde{I}_x \tag{29}$$

and

$$(D_x V_{Hx})^{\mathrm{T}} (D_x W_{Hx}) = \tfrac{1}{2} \tilde{I}_x \tag{30}$$

Then, since W_x and V_x form a biorthogonal basis, we have

$$W_x V_x^{\mathrm{T}} = W_{Lx} V_{Lx}^{\mathrm{T}} + W_{Hx} V_{Hx}^{\mathrm{T}} = I_x \tag{31}$$

By pre-multiplication by D_x and post-multiplication by U_x, we obtain

$$(D_x W_x)(D_x V_x)^{\mathrm{T}} = (D_x W_{Lx})(D_x V_{Lx})^{\mathrm{T}} + (D_x W_{Hx})(D_x V_{Hx})^{\mathrm{T}} = \tilde{I}_x \tag{32}$$

From here, if we assume the surjective property, then Equation (32) immediately implies the set of biorthogonal relationships above, and the harmonic aliasing property is fulfilled.

Now, we assume the harmonic aliasing property holds and we pre-multiply Equation (32) by $(D_x V_{Lx})^{\mathrm{T}}$. Using Equations (27) and (28) we obtain

$$\begin{aligned} (D_x V_{Lx})^{\mathrm{T}}(D_x W_{Lx})(D_x V_{Lx})^{\mathrm{T}} + (D_x V_{Lx})^{\mathrm{T}}(D_x W_{Hx})(D_x V_{Hx})^{\mathrm{T}} &= (D_x V_{Lx})^{\mathrm{T}} \\ \tfrac{1}{2}(D_x V_{Lx})^{\mathrm{T}} + \tfrac{1}{2}(D_x V_{Hx})^{\mathrm{T}} &= (D_x V_{Lx})^{\mathrm{T}} \\ (D_x V_{Lx})^{\mathrm{T}} &= (D_x V_{Hx})^{\mathrm{T}} \end{aligned} \tag{33}$$

Similarly, we post-multiply Equation (32) by $D_x W_{Hx}$. Using Equations (28) and (30), we obtain

$$\begin{aligned} (D_x W_{Lx})(D_x V_{Lx})^{\mathrm{T}}(D_x W_{Hx}) + (D_x W_{Hx})(D_x V_{Hx})^{\mathrm{T}}(D_x W_{Hx}) &= D_x W_{Hx} \\ (D_x W_{Lx})\tfrac{1}{2} + (D_x W_{Hx})\tfrac{1}{2} &= D_x W_{Hx} \\ D_x W_{Lx} &= D_x W_{Hx} \end{aligned} \tag{34}$$

Therefore, the harmonic aliasing property implies the surjective property. □

4. ASSUMPTIONS FOR SEPARABLE BASIS SYSTEMS

In Section 3 we stated assumptions that will allow us to understand the role of the smoothing and inter-grid filters in multigrid convergence. The assumptions stated in Section 3 do not allow the study of many multigrid applications. Specifically, when using the multigrid algorithm in d-dimensional problems, the down-sampling is often designed to reduce the number of grid nodes by a factor of 2^d. On the other hand, the harmonic aliasing property, as stated in Section 3.4, is essentially applicable only for cases where the grids are down-sampled by a factor of 2. The down-sampling by a factor of 2^d is important to reduce the computational and space costs of the algorithm. In this section, we assume further properties in the algorithm and system so that our analysis can be extended to these cases.

For these extensions we use the tensor product defined as:

Definition 5 (*Kronecker product*)
If A is an $m \times n$ matrix and B is a $p \times q$ matrix, then the Kronecker product $A \otimes B$ is the $mp \times nq$ block matrix:

$$A \otimes B = \begin{bmatrix} (A)_{1,1}B & \cdots & (A)_{1,n}B \\ \vdots & \ddots & \vdots \\ (A)_{m,1}B & \cdots & (A)_{m,n}B \end{bmatrix} \tag{35}$$

The most useful properties of Kronecker products for the purpose of our analysis are

$$(A \otimes B)(C \otimes D) = AC \otimes BD \tag{36}$$

and

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \tag{37}$$

For further properties, we refer the reader to [12, 13].

4.1. Separability assumptions

We now assume that we have a system matrix representing a 2D system with coordinates x and y. We denote the system matrix as $A_{xy} \in \mathbb{R}^{mn \times mn}$, where the integers m and n represent the discretization size of the dimensions corresponding to x and y, respectively. We assume that the system matrix can be expressed as the sum of Kronecker products:

$$A_{xy} = A_{x,1} \otimes A_{y,1} + \cdots + A_{x,r} \otimes A_{y,r} \tag{38}$$

$$= \sum_{i=1}^{r} A_{x,i} \otimes A_{y,i} \tag{39}$$

where $A_{x,i} \in \mathbb{R}^{m \times m}$ and $A_{y,i} \in \mathbb{R}^{n \times n}$, with $i = 1, \ldots, r$, representing r possible operators acting on the dimensions x and y, respectively.

We assume that the matrices $A_{x,i}$, $i = 1, \ldots, r$, have the same set of eigenvectors W_x and V_x, the matrices $A_{y,i}$, $i = 1, \ldots, r$, have the same set of eigenvectors W_y and V_y, but each matrix can have a different set of eigenvalues. We denote the matrix of eigenvalues as $\Lambda_{x,i}$ for each matrix $A_{x,i}$, and $\Lambda_{y,i}$ for each matrix $A_{y,i}$. Thus, we have the following eigen-decompositions:

$$A_{x,i} = W_x \Lambda_{x,i} V_x^{\mathrm{T}}, \quad i = 1, \ldots, r \tag{40}$$

and

$$A_{y,i} = W_y \Lambda_{y,i} V_y^{\mathrm{T}}, \quad i = 1, \ldots, r \tag{41}$$

for which the sets of eigenvectors satisfy the biorthogonal relationships $V_x^{\mathrm{T}} W_x = I_x$ and $V_y^{\mathrm{T}} W_y = I_y$, where I_x is an $m \times m$ identity matrix and I_y is an $n \times n$ identity matrix.

It follows from these assumptions that the right-eigenvectors of the system matrix A_{xy}, denoted as W_{xy}, and its eigenvalues, denoted as Λ_{xy}, are given by

$$W_{xy} = W_x \otimes W_y \quad \text{and} \quad \Lambda_{xy} = \sum_{i=1}^{r} \Lambda_{x,i} \otimes \Lambda_{y,i} \tag{42}$$

The left-eigenvectors, denoted as V_{xy}, are given by

$$V_{xy}^{\mathrm{T}} = W_{xy}^{-1} = (W_x \otimes W_y)^{-1} = W_x^{-1} \otimes W_y^{-1} = V_x^{\mathrm{T}} \otimes V_y^{\mathrm{T}} = (V_x \otimes V_y)^{\mathrm{T}} \tag{43}$$

We refer to the assumptions above as the *separability assumptions* because they allow us to apply the assumptions from Section 3 for separate sets of eigenvectors. This kind of factorization for the system matrix often appears in the discretization of partial differential equations (PDEs) (e.g. in finite difference discretization of the Laplacian, divergence and other operators). Thus, the analysis under these extended assumptions will be more suitable for applications.

4.2. Separable filters

The purpose of the assumptions in this section is to apply more aggressive coarsening in the multi-dimensional case. We start from two down-sampling matrices D_x and D_y independently designed to down-sample the nodes of the x- and y-dimensions by a factor of 2. Then, we define the down-sampling matrix for the 2D system, denoted as D_{xy}, as

$$D_{xy} = D_x \otimes D_y \tag{44}$$

In this way the down-sampling matrix D_{xy} is designed to reduce the total number of nodes by a factor of 4.

We use inter-grid filters, denoted by $F_{I,xy}$ and $F_{R,xy}$, and expressed as

$$\begin{aligned} F_{I,xy} &= F_{I,x} \otimes F_{I,y} \quad \text{and} \\ F_{R,xy} &= F_{R,x} \otimes F_{R,y} \end{aligned} \tag{45}$$

where $F_{I,x}$, $F_{R,x}$ and $F_{I,y}$, $F_{R,y}$ are restriction and interpolation filters with eigenvectors W_x and W_y, respectively, and with eigenvalues $\Pi_{I,x}$, $\Pi_{R,x}$ and $\Pi_{I,y}$, $\Pi_{R,y}$, respectively. Therefore, $F_{I,xy}$ and $F_{R,xy}$ have right-eigenvectors W_{xy}, left-eigenvectors V_{xy} and eigenvalues given by

$$\begin{aligned} \Pi_{I,xy} &= \Pi_{I,x} \otimes \Pi_{I,y} \quad \text{and} \\ \Pi_{R,xy} &= \Pi_{R,x} \otimes \Pi_{R,y} \end{aligned} \tag{46}$$

We note that due to the properties of Kronecker products, the decomposition in (20) is valid for both 1D and 2D operators.

Similarly, the smoothing operator S_{xy} is designed such that

$$S_{xy} = S_x \otimes S_y \tag{47}$$

where S_x and S_y are smoothing operators with eigenvectors W_x and W_y, respectively, with eigenvalues Σ_x and Σ_y, respectively. The eigenvalues of S_{xy} are given by

$$\Sigma_{xy} = \Sigma_x \otimes \Sigma_y \tag{48}$$

4.3. The separable harmonic aliasing property

Under the separability assumptions stated in the sections above, we assume the harmonic aliasing property on each set W_x, D_x and W_y, D_y. Then, a generalization of the harmonic aliasing property that we call the *separable harmonic aliasing property* follows for the set W_{xy}, D_{xy}. That is,

$$\begin{aligned} V_{xy}^{\mathrm{T}} U_{xy} D_{xy} W_{xy} &= (V_x \otimes V_y)^{\mathrm{T}} (D_x \otimes D_y)^{\mathrm{T}} (D_x \otimes D_y)(W_x \otimes W_y) \\ &= (V_x^{\mathrm{T}} U_x D_x W_x) \otimes (V_y^{\mathrm{T}} U_y D_y W_y) \\ &= N_x \otimes N_y \end{aligned} \tag{49}$$

where N_x and N_y are *harmonic aliasing patterns* as defined in (24).

5. ERROR ANALYSIS

In Section 2 the *coarse grid correction* matrix K was defined as

$$K = I - I_s^h \breve{A}^{-1} I_h^s A \tag{50}$$

This is the main object of study in this section as it shows the evolution of the approximation error in both nested iteration and the correction scheme. Namely, the approximation error after a full two-grid step with μ_1 correction scheme iterations, each of them with ν_1 pre-smoothing and ν_2 post-smoothing iterations, is given by

$$e_{(\nu_1+1+\nu_2)\mu_1} = (S^{\nu_2} K S^{\nu_1})^{\mu_1} K u \tag{51}$$

In the following sub-sections, we use the assumptions stated in Sections 3 and 4 to see how the eigenvectors of the system are affected by these iterations. Based on the partition of eigenvectors introduced in Section 3, we apply the same principle to create the following partition of eigenvalues:

$$\Lambda_x = \begin{bmatrix} \Lambda_{Lx} & 0 \\ 0 & \Lambda_{Hx} \end{bmatrix}, \quad \Sigma_x = \begin{bmatrix} \Sigma_{Lx} & 0 \\ 0 & \Sigma_{Hx} \end{bmatrix} \quad \text{and} \quad \Pi_x = \begin{bmatrix} \Pi_{Lx} & 0 \\ 0 & \Pi_{Hx} \end{bmatrix} \tag{52}$$

Within this section, we will use the convention to omit any subscript x, y or xy whenever the analysis leads to the same formulas. For example, the eigen-decomposition $A = W\Lambda V$ is valid in both 1D and 2D because the eigen-decomposition $A_x = W_x \Lambda_x V_x^{\mathrm{T}}$ is assumed in the 1D, and the properties of Kronecker products imply $A_{xy} = W_{xy}\Lambda_{xy} V_{xy}^{\mathrm{T}}$ in the 2D case.

5.1. Galerkin coarsening

From the assumptions in both Sections 3 and 4, the Galerkin condition stated in (4) can be expressed as

$$\begin{aligned} \breve{A}^{-1} &= \{I_h^s A I_s^h\}^{-1} \\ &= \{D F_R A F_I U\}^{-1} \\ &= \{(DW)\Pi_R \Lambda \Pi_I (DV)^{\mathrm{T}}\}^{-1} \end{aligned} \tag{53}$$

From here, we first consider the assumptions in Section 3. Using the partition of eigenvectors induced by the harmonic aliasing property, we define the matrix

$$\Delta_x = \Pi_{R,Lx}\Lambda_{Lx}\Pi_{I,Lx} + \Pi_{R,Hx}\Lambda_{Hx}\Pi_{I,Hx} \tag{54}$$

Then, we follow the last step in (53) and obtain

$$\begin{aligned} (\breve{A}_x)^{-1} &= \{(D_x W_x)\Pi_{R,x}\Lambda_x \Pi_{I,x}(D_x V_x)^{\mathrm{T}}\}^{-1} \\ &= \{(D_x W_{Lx})\Delta_x (D_x V_{Lx})^{\mathrm{T}}\}^{-1} \\ &= 4(D_x W_{Lx})\Delta_x^{-1}(D_x V_{Lx})^{\mathrm{T}} \end{aligned} \tag{55}$$

where we use, first, the surjective property and, second, the biorthogonal relationships (27) to (30).

Now we consider the assumptions in Section 4. Similarly, for this case we define the matrices

$$\Delta_{x,i} = \Pi_{R,Lx}\Lambda_{Lx,i}\Pi_{I,Lx} + \Pi_{R,Hx}\Lambda_{Hx,i}\Pi_{I,Hx} \tag{56}$$

$$\Delta_{y,i} = \Pi_{R,Ly}\Lambda_{Ly,i}\Pi_{I,Ly} + \Pi_{R,Hy}\Lambda_{Hy,i}\Pi_{I,Hy} \tag{57}$$

and, based on these definitions,

$$\Delta_{xy} = \sum_{i=1}^{r} \Delta_{x,i} \otimes \Delta_{y,i} \tag{58}$$

Then, we follow the last step in (53) to obtain

$$\begin{aligned}
(\breve{A}_{xy})^{-1} &= \{(D_{xy}W_{xy})\Pi_{R,xy}\Lambda_{xy}\Pi_{I,xy}(D_{xy}V_{xy})^{\mathrm{T}}\}^{-1} \\
&= \{(D_x W_{Lx} \otimes D_y W_{Ly})\Delta_{xy}(D_x V_{Lx} \otimes D_y V_{Ly})^{\mathrm{T}}\}^{-1} \\
&= 16(D_x W_{Lx} \otimes D_y W_{Ly})\Delta_{xy}^{-1}(D_x V_{Lx} \otimes D_y V_{Ly})^{\mathrm{T}} \\
&= 16(D_{xy} W_{Lxy})\Delta_{xy}^{-1}(D_{xy} V_{Lxy})^{\mathrm{T}}
\end{aligned} \tag{59}$$

where we use, first, the surjective property and, second, the biorthogonal relationships (27)–(30), and finally, we simply define $W_{Lxy} = W_{Lx} \otimes W_{Ly}$ and $V_{Lxy} = V_{Lx} \otimes V_{Ly}$.

We note that in both (55) and (59) the Galerkin coarse matrix $\breve{A}$ has an eigen-decomposition with eigenvectors given by the down-sampled eigenvectors of A. This is a nice property as it assures that the assumptions stated for the system on the fine grid are satisfied in coarser grids as well.

5.2. Convergence rates

Using the assumptions in Sections 3 and 4 and the results from Section 5.1, we can express the coarse grid correction matrix as follows:

$$\begin{aligned}
K &= I - I_s^h \breve{A}^{-1} I_h^s A \\
&= I - F_I U \breve{A}^{-1} D F_R W \Lambda V^{\mathrm{T}} \\
&= I - F_I W V^{\mathrm{T}} U \breve{A}^{-1} D W \Pi_R \Lambda V^{\mathrm{T}} \\
&= I - (2^{2d}) W \Pi_I (V^{\mathrm{T}} U D W_L) \Delta^{-1} (V_L^{\mathrm{T}} U D W) \Pi_R \Lambda V^{\mathrm{T}}
\end{aligned} \tag{60}$$

where d represents the dimension of the problem. In parentheses we see how the harmonic aliasing property appears naturally in this matrix.

For the assumptions from Section 3, we follow the algebra to obtain

$$\begin{aligned}
K_x &= I_x - 4W_x \Pi_{I,x} (V_x^{\mathrm{T}} U_x D_x W_{Lx}) \Delta_x^{-1} (V_{Lx}^{\mathrm{T}} U_x D_x W_x) \Pi_{R,x} \Lambda_x V_x^{\mathrm{T}} \\
&= W_x V_x^{\mathrm{T}} - 4W_x \Pi_{I,x} \left(\frac{1}{2} \begin{bmatrix} \tilde{I}_x \\ \tilde{I}_x \end{bmatrix} \right) \Delta_x^{-1} \left(\frac{1}{2} \begin{bmatrix} \tilde{I}_x & \tilde{I}_x \end{bmatrix} \right) \Pi_{R,x} \Lambda_x V_x^{\mathrm{T}}
\end{aligned}$$

$$= W_x \begin{bmatrix} \tilde{I}_x - \Pi_{I,Lx}\Delta_x^{-1}\Pi_{R,Lx}\Lambda_{Lx} & -\Pi_{I,Lx}\Delta_x^{-1}\Pi_{R,Hx}\Lambda_{Hx} \\ -\Pi_{I,Hx}\Delta_x^{-1}\Pi_{R,Lx}\Lambda_{Lx} & \tilde{I}_x - \Pi_{I,Hx}\Delta_x^{-1}\Pi_{R,Hx}\Lambda_{Hx} \end{bmatrix} V_x^{\mathrm{T}} \tag{61}$$

Note that matrix K is not diagonalized by the eigenvectors of the system. Instead, we obtain a block-tridiagonal matrix that shows how each group of modes from W_{Lx} and W_{Hx} are damped and mixed. In order to simplify this result, we define the *convergence operator*, Γ_x, such that

$$\begin{aligned} K_x &= W_x \Gamma_x V_x^{\mathrm{T}} \\ &= W_x \begin{bmatrix} \Gamma_{Lx\to Lx} & \Gamma_{Hx\to Lx} \\ \Gamma_{Lx\to Hx} & \Gamma_{Hx\to Hx} \end{bmatrix} V_x^{\mathrm{T}} \end{aligned} \tag{62}$$

Each one of the four submatrices in Γ_x is diagonal and we call them the *modal convergence operators*. Their diagonal values represent the factor by which each modal component of the error is multiplied and transferred between Lx and Hx modes according to the subscripts. Their diagonal values can be simplified as follows:

$$\begin{aligned} (\Gamma_{Lx\to Lx})_{i,i} &= \frac{1}{1+a_i b_i}, \quad (\Gamma_{Hx\to Lx})_{i,i} = \frac{-b_i}{1+a_i b_i} \\ (\Gamma_{Lx\to Hx})_{i,i} &= \frac{-a_i}{1+a_i b_i} \quad \text{and} \quad (\Gamma_{Hx\to Hx})_{i,i} = \frac{a_i b_i}{1+a_i b_i} \end{aligned} \tag{63}$$

where

$$a_i = \frac{(\Pi_{R,Lx})_{i,i}}{(\Pi_{R,Hx})_{i,i}} \frac{(\Lambda_{Lx})_{i,i}}{(\Lambda_{Hx})_{i,i}} \quad \text{and} \quad b_i = \frac{(\Pi_{I,Lx})_{i,i}}{(\Pi_{I,Hx})_{i,i}} \tag{64}$$

The convergence of a two-grid algorithm depends on the smoother S_x and the coarse grid correction matrix K_x, which in the domain of the system's eigenvectors is contained in the matrices Σ_x and Γ_x, respectively. Now, matrix Γ_x and its four modal convergence operators allow us to focus on the performance of the inter-grid operators; therefore, this is the main object of study for the design of inter-grid filters. In Section 6 we will show examples on how to apply this analysis.

From the assumptions in Section 4, we follow a different algebra. This is

$$\begin{aligned} K_{xy} &= I_{xy} - 16 W_{xy}\Pi_{I,xy}(V_{xy}^{\mathrm{T}}U_{xy}D_{xy}W_{Lxy})\Delta_{xy}^{-1}(V_{Lxy}^{\mathrm{T}}U_{xy}D_{xy}W_{xy})\Pi_{R,xy}\Lambda_{xy}V_{xy}^{\mathrm{T}} \\ &= I_{xy} - 16 W_{xy}\Pi_{I,xy}\left(\frac{1}{2}\begin{bmatrix}\tilde{I}_x \\ \tilde{I}_x\end{bmatrix} \otimes \frac{1}{2}\begin{bmatrix}\tilde{I}_y \\ \tilde{I}_y\end{bmatrix}\right)\Delta_{xy}^{-1}\left(\frac{1}{2}[\tilde{I}_x \;\; \tilde{I}_x] \otimes \frac{1}{2}[\tilde{I}_y \;\; \tilde{I}_y]\right)\Pi_{R,xy}\Lambda_{xy}V_{xy}^{\mathrm{T}} \\ &= I_{xy} - W_{xy}\left(\begin{bmatrix}\Pi_{I,Lx} \\ \Pi_{I,Hx}\end{bmatrix} \otimes \begin{bmatrix}\Pi_{I,Ly} \\ \Pi_{I,Hy}\end{bmatrix}\right)\Delta_{xy}^{-1}\sum_{i=1}^{r}\left(\begin{bmatrix}\Pi_{R,Lx}\Lambda_{Lx,i} \\ \Pi_{R,Hx}\Lambda_{Hx,i}\end{bmatrix}^{\mathrm{T}} \otimes \begin{bmatrix}\Pi_{R,Ly}\Lambda_{Ly,i} \\ \Pi_{R,Hy}\Lambda_{Hy,i}\end{bmatrix}^{\mathrm{T}}\right)V_{xy}^{\mathrm{T}} \\ &= W_{xy}\Gamma_{xy}V_{xy}^{\mathrm{T}} \end{aligned} \tag{65}$$

Here, a simple structure for the *convergence operator*, Γ_{xy}, does not appear clear because of the Kronecker products involved. Since the matrix Δ_{xy}^{-1} cannot in general be factored as a Kronecker

product, we cannot analyze the convergence of the algorithm for each dimension independent of the other. We then need to consider the four possible combinations of x, y-dimensions and L, H groups. The products considering these combinations are mixed in Γ_{xy} and we need to reorder them to identify the *modal convergence operators*. Thus, we introduce a permutation matrix $P \in \{0,1\}^{mn \times mn}$ such that for arbitrary matrices $X_L, X_H \in \mathbb{R}^{m/2 \times m/2}$ and $Y_L, Y_H \in \mathbb{R}^{n/2 \times n/2}$ one has

$$P\left(\begin{bmatrix} X_L \\ X_H \end{bmatrix} \otimes \begin{bmatrix} Y_L \\ Y_H \end{bmatrix}\right) = \begin{bmatrix} X_L \otimes Y_L \\ X_H \otimes Y_L \\ X_L \otimes Y_H \\ X_H \otimes Y_H \end{bmatrix} \tag{66}$$

Then, applying this permutation to reorder the rows and columns of Γ_{xy}, we obtain the following structure:

$$P\Gamma_{xy}P^{\mathrm{T}} = \begin{bmatrix} \Gamma_{LxLy\to LxLy} & \Gamma_{HxLy\to LxLy} & \Gamma_{LxHy\to LxLy} & \Gamma_{HxHy\to LxLy} \\ \Gamma_{LxLy\to HxLy} & \Gamma_{HxLy\to HxLy} & \Gamma_{LxHy\to HxLy} & \Gamma_{HxHy\to HxLy} \\ \Gamma_{LxLy\to LxHy} & \Gamma_{HxLy\to LxHy} & \Gamma_{LxHy\to LxHy} & \Gamma_{HxHy\to LxHy} \\ \Gamma_{LxLy\to HxHy} & \Gamma_{HxLy\to HxHy} & \Gamma_{LxHy\to HxHy} & \Gamma_{HxHy\to HxHy} \end{bmatrix} \tag{67}$$

where we identify the *modal convergence operators* representing the 16 possible ways to transfer modal components of the error between the four combinations of x, y-dimensions and L, H groups according to the subscripts. The values of each one of these groups can be expressed in a generic form as

$$\Gamma_{AxBy\to CxDy} = \delta_{AC}\delta_{BD} - (\Pi_{I,Cx} \otimes \Pi_{I,Dy})\Delta_{xy}^{-1} \sum_{i=1}^{r} (\Pi_{R,Ax}\Lambda_{Ax,i}) \otimes (\Pi_{R,By}\Lambda_{By,i}) \tag{68}$$

where $A \in \{H, L\}$, $B \in \{H, L\}$, $C \in \{H, L\}$, $D \in \{H, L\}$ and $\delta_{AC}\delta_{BD}$ is an identity matrix only if $A = C$ and $B = D$.

The convergence operator, Γ_{xy}, and its 16 modal convergence operators allow us to focus on the performance of the inter-grid operators and it is always the main object of study for the design of inter-grid filters. Compared with the 1D case, the analysis is now more complicated as the modal components of the error are transferred not only between two groups of modes but also between different dimensions. In Section 6.3 we will show an example on how to design inter-grid filters under this scenario.

5.3. The heuristics in error analysis

We consider an ideal scenario for a 1D problem in order to check the heuristic behavior of the multigrid algorithm. By using the variational property, we define the single inter-grid filter $F_{\mathrm{sharp},x}$ such that

$$\begin{aligned} \Pi_{Lx} &= I \quad \text{and} \\ \Pi_{Hx} &= 0 \end{aligned} \tag{69}$$

We call this filter the *sharp inter-grid filter*. In Fourier harmonic analysis, this would correspond to what is called a 'perfect low-pass filter' [11]. This definition is more general as we can now apply it to a more general kind of basis, that is, to any basis with the harmonic aliasing property.

By using the eigen-decomposition of A and the sharp inter-grid filter in (63), we obtain

$$K_{\text{sharp},x} = W_{Hx} W_{Hx}^{\mathrm{T}} \tag{70}$$

Therefore, for this choice of inter-grid operators, we can see that several applications of the coarse grid correction matrix do not help to reduce the error. It just cancels the W_{Lx} components of the error. We then need to apply smoothing iterations in order to reduce the W_{Hx} components of the error. We also verify that the error reduction achieved by multigrid iterations does not depend on the step size h as the iteration matrix does not depend on the eigenvalues of A. The simplicity of this result shows the general principles of multigrid algorithm design. In Section 6 we will see how this idealistic scenario does not always lead to an optimal algorithm for solving linear systems.

6. EXAMPLES OF INTER-GRID FILTER DESIGN

In Section 5 we obtained theoretical results for the convergence rates based on the assumptions stated in previous sections. In this section, we introduce examples to show how these results can be applied to different kinds of systems. We consider systems based on different sets of eigenvectors: Fourier harmonic modes, Hadamard harmonic modes, and a mixture of Fourier and Hadamard harmonic modes.

6.1. Fourier harmonic analysis: *trade-off between computational complexity and convergence rate*

We consider a 1D system in which A is a standard finite-difference discretization of a second-order derivative with step size $h=1$; i.e. the stencil of A is $s=[-1\,\underline{2}\,-1]$ (the underline denotes the diagonal element). We apply Dirichlet boundary conditions, i.e. stencil $[\underline{2}\,-1]$ at the left corner and $[-1\,\underline{2}]$ at the right corner, which lead to an invertible system. The number of nodes in the discretization is set to $N=16$ and we consider a two-grid algorithm with a coarse-grid step size of $2h=2$. In addition we assume the variational property that leads to a single inter-grid filter F.

The eigenvectors of A are given by $(W)_{i,j}=\sqrt{2/17}\sin(ij\pi/17)$, with $i, j=1,\ldots,16$. The eigenvector matrix W is orthonormal and, after reversing the order of the columns $j=9,\ldots,16$, it also fulfills the harmonic aliasing property. Therefore, our modal analysis can be directly applied to this system. On the other hand, the extension of Fourier analysis from complex- to real-valued harmonic functions is well known and LFA can therefore be applied to this system. Thus, the purpose of this example is to (i) show how our method is applied to a standard system in which the eigenvectors can be labeled by frequencies, thus giving an intuitive picture of what is happening and (ii) show how to design inter-grid filters within our new framework and thus demonstrate the issue we discover in this process.

For the inter-grid filter, we start with the common choice of linear-interpolation and full-weighting (LI/FW), and we consider their application on an increasing number of neighbors per node. The standard choice for this system considers two neighbors per node, which leads to an inter-grid filter F with stencil $s=[0.5\ \underline{1}\ 0.5]$ and Dirichlet boundary conditions. Considering more neighbors per node is equivalent to applying the inter-grid filter F several times in interpolation

or restriction operations. Thus, an inter-grid filter $F, F^2, F^3, F^4, \ldots$ represents LI/FW operations over $2, 4, 6, 8, \ldots$ neighbors per node, respectively.

In Table I we show the spectral radii of $\Gamma_{Lx\to Lx}$, $\Gamma_{Hx\to Lx}$, $\Gamma_{Lx\to Hx}$, and $\Gamma_{Hx\to Hx}$ for a two-grid approach using different numbers of LI/FW passes. Here, the most important factor is the spectral radius of $\Gamma_{Lx\to Lx}$. It shows the worst case reduction of modal components of the error for low-frequency modes that are mapped to themselves. In LFA the spectral radius $\|\Gamma_{Lx\to Lx}\|$ is called the *asymptotic convergence factor*, ρ_{loc} [7]. The reduction of these components of the error is the main task of the two-grid approach. We do not see much reduction of the high-frequency components of the error that are mapped to themselves, as the spectral radius of $\Gamma_{Hx\to Hx}$ is always close to 1, leaving this task to the smoothing iterations. The cross-frequency rates $\Gamma_{Hx\to Lx}$ and $\Gamma_{Lx\to Hx}$ represent the aliasing effect in which high- and low-frequency components of the error are reduced and mapped to low- and high-frequency components of the error, respectively.

The spectral radius of $\Gamma_{Hx\to Lx}$ in Table I appears to be close to 1, which means an almost complete transfer of high-to-low frequency components of the error at each iteration. A careful look at the convergence rates shows that this large number comes from the transfer of the highest-frequency error to the lowest-frequency error. Although this transfer is not ideal, it is not critical because the pre-smoothing iterations will reduce the highest-frequency error very effectively. As expected, all the convergence rates in Table I are further reduced as we increase the number of LI/FW passes. The disadvantage of increasing the number of passes is that the inter-grid filter, as well as the coarse system matrix, becomes less and less sparse (see Figure 4(a)–(d)), thus increasing the computational complexity of the algorithm.

To complete the convergence analysis, we need to consider a smoothing filter and select the number of smoothing iterations. A simple choice is to use the Richardson iteration scheme, which leads to a smoothing filter $S = I - (1/\tau)A$, with $\tau = 4$ obtained by the Gershgorin bound of A. This filter satisfies our assumptions because it has the same eigenvectors as A. Since the task of the smoothing filter is to reduce the high-frequency components of the error, we suggest choosing the number of smoothing iterations such that the reduction of the high-frequency components of the error, given by $\|\Sigma_{Hx}\|$, is equal to or less than the reduction of low-frequency components of the error achieved by the coarse grid correction matrix, given by $\|\Gamma_{Lx\to Lx}\|$. For this example, using a 1-pass LI/FW inter-grid filter we achieve the same reduction of low-frequency error as the reduction of high-frequency error achieved by one Richardson iteration. For instance, using one pre-smoothing ($\nu_1 = 1$) and one post-smoothing ($\nu_2 = 1$) Richardson iteration in the correction scheme, the approximation error after one full two-grid step ($\mu_1 = 1$) will be given by $e_3 = (SK)^2 u$ with a convergence rate of $\|(SK)^2\| = 0.2458$.

Table I. Spectral radii of modal convergence operators for the system in Section 6.1.

Filter	$\|\Gamma_{Lx\to Lx}\|$	$\|\Gamma_{Hx\to Lx}\|$	$\|\Gamma_{Lx\to Hx}\|$	$\|\Gamma_{Hx\to Hx}\|$
LI/FW 1-pass	0.4539	0.9915	0.4539	0.9915
LI/FW 2-passes	0.3647	0.5280	0.4388	1.0000
LI/FW 3-passes	0.2839	0.4946	0.4110	1.0000
LI/FW 4-passes	0.2149	0.4506	0.3745	1.0000
LI/FW 5-passes	0.1590	0.4011	0.3334	1.0000
LI/FW 6-passes	0.1155	0.3506	0.2914	1.0000

The results consider a two-grid approach using several passes of LI/FW as inter-grid operators.

As a different choice of inter-grid operators, we try to approach the sharp inter-grid filter with a common procedure used in signal processing. We select the eigenvalues of F in analogy with a Butterworth filter of order n [11]. We start at order $n=1$ with a cut-off frequency of $\pi/16$ that tries to reduce all frequencies except for the lowest frequency mode, and as we increase the order n the cut-off frequency approaches $\pi/2$ geometrically, at which point the filter becomes perfectly sharp. That is,

$$B_n(i) = \frac{1}{1+\left(\dfrac{2}{1-(7/8)^n}\dfrac{i-1}{N-1}\right)^{2n}}, \quad i=1,\ldots,16 \tag{71}$$

from which we construct the inter-grid filter as $F = W\Pi W^{\mathrm{T}}$ with $\Pi = \mathrm{diag}(B_n)$. The main reason to move the cut-off frequency with the order of the filter is to prevent the eigenvalues in Π_{Hx} from producing large cross-frequency convergence rates.

In Table II we show the spectral radii of $\Gamma_{Lx\to Lx}$, $\Gamma_{Hx\to Lx}$, $\Gamma_{Lx\to Hx}$, and $\Gamma_{Hx\to Hx}$ for a two-grid approach using Butterworth filters of different orders. The Butterworth filter is better than LI/FW, especially in terms of the cross-frequency convergence rate $\|\Gamma_{Hx\to Lx}\|$. The main disadvantage of the Butterworth filter is that it is always non-sparse, as shown in Figure 4(e)–(h). Even if increasing the order n makes the filter appear more and more sparse, the overall contribution of small terms is comparable to the largest entries. Now, increasing the order n also concentrates the largest entries close to the diagonal and the tridiagonal elements become similar to the LI/FW entries. This hints at the optimality of LI/FW as a tridiagonal inter-grid filter for this specific problem.

An important conclusion of these tests is that in the design of inter-grid filters for systems with Fourier harmonic modes as eigenvectors, we face a trade-off between the number of multigrid steps that can be saved by moving toward a sharp inter-grid filter and the number of communications between neighboring nodes required for interpolation/restriction tasks. This is a consequence of the Gibbs phenomenon, which is well known in Fourier analysis [11].

6.2. *Hadamard harmonic analysis*: *optimality of the sharp inter-grid filter*

Now, we consider a system based on an application of Markov chains. The system will have a variable size with 2^{l-1}, $l\in\mathbb{N}^+$, transient states and at least one recurrent state. We ignore the precise number of recurrent states and their interconnections as they will not play any role in the

Table II. Spectral radii of modal convergence operators for the system in Section 6.1.

Filter	$\|\Gamma_{Lx\to Lx}\|$	$\|\Gamma_{Hx\to Lx}\|$	$\|\Gamma_{Lx\to Hx}\|$	$\|\Gamma_{Hx\to Hx}\|$
B_1	0.4156	0.5826	0.4493	0.9982
B_2	0.2932	0.4994	0.4150	1.0000
B_3	0.1954	0.4350	0.3615	1.0000
B_4	0.1246	0.3623	0.3011	1.0000
B_5	0.0770	0.2925	0.2431	1.0000
B_6	0.0467	0.2314	0.1923	1.0000
B_7	0.0279	0.1807	0.1502	1.0000

The results consider a two-grid approach using Butterworth filters of different orders as the inter-grid filter.

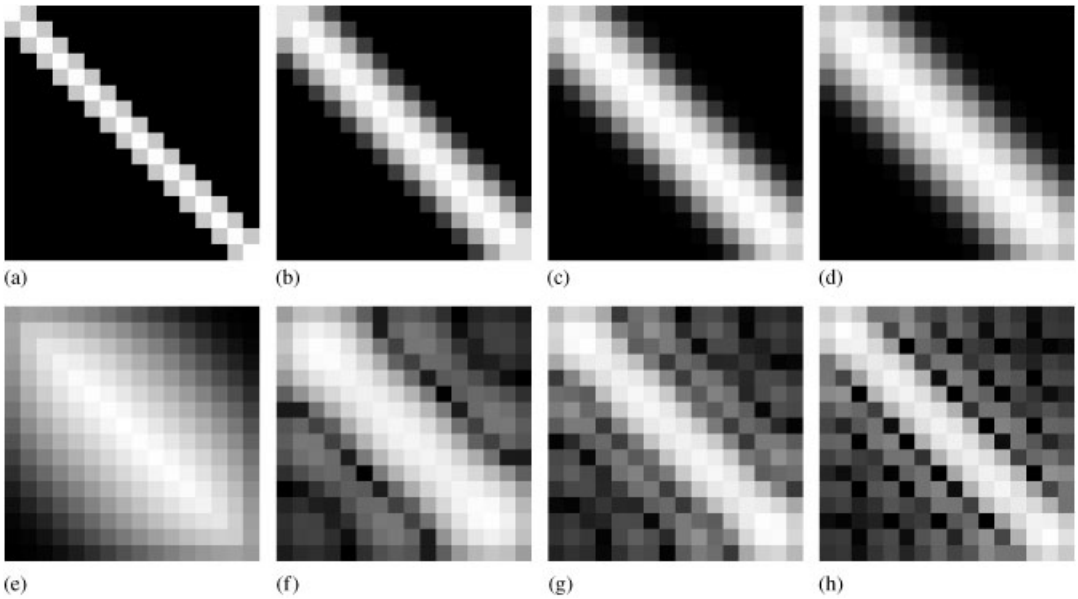

Figure 4. Images of the magnitude of entries for different inter-grid filter matrices. The intensity of gray color is white for the largest magnitude and black for the smallest magnitude. The scale between black and white is set in logarithmic scale in order to increase the visual difference between small and zero entries: (a) LI/FW 1-pass; (b) LI/FW 3-passes; (c) LI/FW 5-passes; (d) LI/FW 7-passes; (e) B_1; (f) B_3; (g) B_5; and (h) B_7.

solution of the problem. Thus, the structure of the system is given by the transition probability matrix within the transient states, which is obtained by the following recursion:

$$T_1 = \tfrac{1}{2} \tag{72}$$

$$T_l = \begin{bmatrix} T_{l-1} & 2^{-l} \cdot \tilde{I}_c \\ 2^{-l} \cdot \tilde{I}_c & T_{l-1} \end{bmatrix} \quad \text{for } l > 1 \tag{73}$$

where $\tilde{I}_c$ is a counter-diagonal matrix of the same size as T_{l-1}. The recursion (73) creates a matrix $T_l \in (\mathbb{R}^+)^{2^{l-1} \times 2^{l-1}}$ that is sub-stochastic since the sum of all of its entries in a row is always less than or equal to 1. In fact, the sum of all of the entries in a row is equal to $1 - 1/2^l$ for all the rows in T_l. Thus, in this Markov chain, each transient state has a probability of $1/2^l$ of jumping to one or more recurrent states in one step. An example of this structure is shown in Figure 5 where we can see the state transition diagram of the transient states for $l = 4$.

Since, by definition, no recurrent state is connected to any transient state, once the process jumps from a transient to a recurrent state it will never return to any transient state and it is said to have been *absorbed*. Starting from a given transient state i, $1 \leqslant i \leqslant 2^{l-1}$, the number of jumps within the transient states before jumping to a recurrent state is called the *absorbing time*, t_i. There are many applications associated with these so-called *absorbing chains* [14]; for instance, in the study of *discrete phase-type distributions* in queueing theory [15].

Here, we will consider the problem of computing the expected value of the absorbing time when we start at node i; denoted by $(x_l)_i = E[t_i]$. The vector $x_l \in \mathbb{R}^{2^{l-1}}$ is given by the solution of the linear

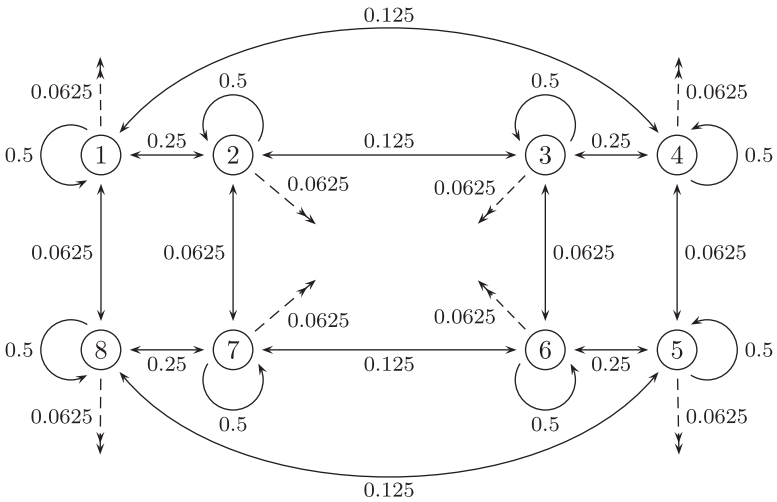

Figure 5. State transition diagram of the transient states for the Markov chain used in Section 6.2 with $l=4$ ($N=8$ nodes). Each connection with solid line shows the probability of state transitions. The dashed lines with double arrows show the probability of transition to one or more recurrent states that do not appear in this figure.

system

$$(I-T_l)x_l=\mathbf{1} \tag{74}$$

where $(\mathbf{1})_i=1$, for $i=1,\ldots,2^{l-1}$. Here, our system matrix is given by $A_l=I-T_l$, which is a non-singular, symmetric, positive-definite M matrix. Furthermore, the matrix A_l becomes ill-conditioned as we increase l, creating a problem similar to that found in the numerical solution of linear PDEs. In the general context of absorbing chains, the matrix $A_l=I-T_l$ is called the *fundamental matrix* [14]. The inversion of this matrix is important as it also appears in the computation of moments of discrete phase-type distributions and the probability of absorption by recurrent classes, among other problems.

In the transition graph of this Markov chain, each node representing a transient state is connected to l neighboring nodes. However, the structure of connections changes from node to node such that the stencil of A_l is not constant throughout the rows. For instance, in the Markov chain of Figure 5, the fundamental matrix is

$$A_4=\begin{bmatrix} 0.5 & -0.25 & 0 & -0.125 & 0 & 0 & 0 & -0.0625 \\ -0.25 & 0.5 & -0.125 & 0 & 0 & 0 & -0.0625 & 0 \\ 0 & -0.125 & 0.5 & -0.25 & 0 & -0.0625 & 0 & 0 \\ -0.125 & 0 & -0.25 & 0.5 & -0.0625 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.0625 & 0.5 & -0.25 & 0 & -0.125 \\ 0 & 0 & -0.0625 & 0 & -0.25 & 0.5 & -0.125 & 0 \\ 0 & -0.0625 & 0 & 0 & 0 & -0.125 & 0.5 & -0.25 \\ -0.0625 & 0 & 0 & 0 & -0.125 & 0 & -0.25 & 0.5 \end{bmatrix} \tag{75}$$

Here, the stencil at the 3rd row is $s_3 = [-0.125, \underline{0.5}, -0.25, 0, -0.0625]$ (the underline denotes the diagonal element), whereas the stencil at the 4th row is $s_4 = [-0.125, 0, -0.25, \underline{0.5}, -0.0625]$. Therefore, the assumptions of LFA are not fulfilled and its analysis does not apply for this system. Nevertheless, in the tests that follow we will ignore this fact as we wish to see what convergence rates LFA predicts for a system where its assumptions do not apply.

In fact, the eigenvectors of the fundamental matrix A_l do not correspond to the Fourier harmonic modes of LFA but instead form a Hadamard matrix of order $N = 2^{l-1}$. One of the standard ways to construct this matrix is Sylvester's construction [16], but the basis obtained by this procedure does not fulfill the harmonic aliasing property. As in the previous example, we need to reorder the columns of the eigenvector matrix in order to obtain the right structure. Therefore, we introduce a column-reordered variation of Sylvester's construction as follows:

$$W_1 = 1 \tag{76}$$

$$W_{l+1} = \frac{1}{\sqrt{2}} [U \;\; \bar{U}] \begin{bmatrix} W_l & W_l \\ W_l & -W_l \end{bmatrix} \tag{77}$$

where U and $\bar{U}$ correspond to uniform up-sampling and up-unselecting matrices of sizes $2^l \times 2^{l-1}$. The matrix $[U \;\; \bar{U}]$ acts as a permutation matrix that reorders the columns of the new basis. From this construction, it can be easily checked through induction arguments that the matrix W_l is orthonormal and that it fulfills the harmonic aliasing property. The same arguments could be used to check the fact that W_l diagonalizes the system matrix A_l. Furthermore, the orthogonality of W_l and Equation (77) allow us to obtain a closed-form expression for the sharp inter-grid filter, as defined in (69). That is,

$$F_{\text{sharp},l+1} = W_{l+1}^{\mathrm{T}} \begin{bmatrix} \tilde{I} & 0 \\ 0 & 0 \end{bmatrix} W_{l+1} = \frac{1}{2}(I + U\bar{D} + \bar{U}D) = \frac{1}{2} \begin{bmatrix} 1 & 1 & & & & & & \\ 1 & 1 & & & & & & \\ & & 1 & 1 & & & & \\ & & 1 & 1 & & & & \\ & & & & \ddots & & & \\ & & & & & & 1 & 1 \\ & & & & & & 1 & 1 \end{bmatrix} \tag{78}$$

The structure of the filter turns out to be very sparse, unlike the sharp filter for the previous example. This filter alternately averages the values at each node with its left neighbor and then its right neighbor.

In our analysis, the inter-grid filter F_l and the smoothing operator S_l should be designed to match the structure of the system. For this reason, our analysis would not work if we use standard inter-grid operators such us LI/FW, because the eigenvectors of the LI/FW filter are Fourier harmonic modes that are different than the Hadamard harmonic modes. As the sharp inter-grid filter in (78) has a sparse structure, we choose it as the inter-grid filter. As in the previous example, for the smoothing filter we use the Richardson iteration scheme, which leads to a smoothing filter $S_l = I - (1/\tau)A$, with $\tau = 1 - 2^{-l}$ obtained by the Gershgorin bound of A_l. Since the sharp inter-grid filter is removing all the Lx components of the error, the only parameters to configure are the

number of smoothing iterations. This means that we need only one iteration of the full two-grid algorithm with $\mathcal{O}(1)$ smoothing iterations to make the algorithm converge. On the other hand, a standard choice of LI/FW inter-grid operators does not work better than the sharp inter-grid configuration as shown in Table III.

As this scenario is rather unusual in the context of PDEs, where the eigenvectors are typically similar to Fourier harmonic modes (that come with Gibbs phenomenon, as shown in Section 6.1), we would like to understand how the sparse inter-grid filter arranges the information to reach convergence in one step. To understand this, we need to consider three facts. First, the fact that the sharp inter-grid filter is alternately averaging the values at each node with its left and then right neighbor. Second, we need to note that the coarse grid matrix $\breve{A}_l$ constructed from A_l and $F_{\text{sharp},l}$, using the Galerkin condition, is equal to our definition of A_{l-1} constructed by recursion (this can be checked by induction). This would not have been the case if we used a different inter-grid filter such as LI/FW. Then, we can say that the sharp inter-grid filter has been able to unveil the recursive structure by which we defined the system. It is also a nice property in the sense that the coarse grid problem also represents an absorbing Markov chain; thus the sharp inter-grid filter makes the two-grid algorithm an aggregation method similar to what is sought in [17] using a different multi-level approach.

The third fact is that the structure of our system induces a hierarchical classification of nodes. Namely, we can define classes of nodes by the strength of their connections, as is usually done in AMG methods [2]. Two nodes i and j belong to the same class if they have a transition probability $(P)_{i,j} \geqslant 1/2^c$, with $1 \leqslant c \leqslant l$. For instance, in the system of Figure 5 for $c=1$ we have eight singleton classes with the individual transition states in each one; for $c=2$ we have four classes: $\{1,2\}$, $\{3,4\}$, $\{5,6\}$, and $\{7,8\}$; for $c=3$ we have two classes: $\{1,2,7,8\}$ and $\{3,4,5,6\}$; and finally for $c=4$ we have one class with the whole set of nodes. This classification of nodes is shown in Figure 6.

Finally we can see how these three facts combine. The sharp inter-grid filter averages the strongest connected nodes, which correspond alternately to nodes at the left and right of each

Table III. Convergence rates of the full two-grid algorithm for different inter-grid operators and different sizes of the system in Section 6.2.

	$\|(SK)^2\|$	
N	Sharp filter	LI/FW
2	0.0000	0.2500
4	0.0816	0.2030
8	0.1600	0.2700
16	0.2040	0.3447
32	0.2268	0.3955
64	0.2383	0.4428
128	0.2442	0.4817
256	0.2471	0.5156

The configuration considers one step of the full two-grid algorithm with one pre-smoothing and one post-smoothing Richardson iteration. The results compare the convergence rates by using a sharp inter-grid filter or LI/FW for inter-grid operators.

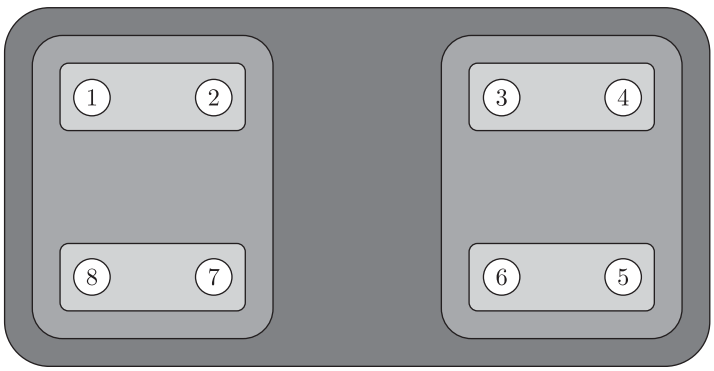

Figure 6. Classification of nodes by the strength of their connection for the Markov chain in Figure 5. By considering only the strongest connections, we start in the white color with eight singleton classes. As we consider weaker connections, we obtain four classes, two classes and finally one class with the whole set of nodes, represented in light to dark gray colors, respectively. The classification leads to a nested structure of classes.

Table IV. Spectral radii of modal convergence operators for the system in Section 6.2.

Analysis	$\|\Gamma_{Lx\to Lx}\|$	$\|\Gamma_{Hx\to Lx}\|$	$\|\Gamma_{Lx\to Hx}\|$	$\|\Gamma_{Hx\to Hx}\|$
MA $\forall N$	0	0	0	1
LFA $N=2$	0	0	0	1
LFA $N=4$	0.0528	0.2236	0.2236	0.9472
LFA $N=8$	0.1702	0.3758	0.3758	0.9803
LFA $N=16$	0.2877	0.4527	0.4527	0.9936
LFA $N=32$	0.3739	0.4838	0.4838	0.9981
LFA $N=64$	0.4283	0.4948	0.4948	0.9995
LFA $N=128$	0.4602	0.4984	0.4984	0.9999
LFA $N=256$	0.4783	0.4995	0.4995	1.0000

The results consider a two-grid approach using the sharp inter-grid filter from (78). The first row shows the results for our modal analysis (MA), which do not change with the problem size. The following rows show the estimation of LFA (working under incorrect assumptions) for systems with increasing size.

node. These nodes belong to the same class defined above for $c=2$ and, since the different classes for $1\leqslant c\leqslant l$ are nested (see Figure 6), the sharp inter-grid filter guarantees a similar structure in the coarse grid. This did not happen in the example of Section 6.1 because in that case we could not separate classes with a nested structure. This fact seems to be crucial in order to obtain an optimal inter-grid filter for the Markov chain problem.

In terms of convergence factors for this example, our analysis gives different results if we used LFA while ignoring the fact that the assumptions for LFA are not fulfilled. This is shown in Table IV, where we can see that the convergence estimated by our method compared with LFA is the same only for grid size $N=2$. This is because $N=2$ is the only size for which the Hadamard basis is the same as the Fourier basis. For $N>2$ we see how LFA gives increasingly pessimistic estimates of the convergence factors.

We can also check how different the convergence analysis would be if we chose LI/FW for the inter-grid operators. The multigrid algorithm lets us use these inter-grid operators but then neither LFA nor our analysis can be applied to get information about modal convergence. This is because

the Fourier harmonic modes of the LI/FW inter-gird filter do not match the Hadamard harmonic modes of the system. If we ignore this limitation and we use the Hadamard harmonic basis to estimate the convergence of a two-grid step, we obtain the results of Table V. On the other hand, if we use a Fourier harmonic basis to estimate convergence rates (which corresponds to LFA), we obtain the results in Table VI. The Hadamard analysis leads to a more pessimistic estimation but it is not possible to determine which result is more accurate because the definitions of the L and H groups of modes technically does not apply under both analyses.

The conclusion of this approach is that an arbitrary choice of inter-grid operators does not let us apply the heuristics of the multigrid methodology if we cannot define groups of L and H modes. The choice of LI/FW inter-grid operators still seems to make the algorithm stable because the estimated convergence factors are always less than 1, but its performance is obviously inferior to that of the optimal sharp inter-grid filter for this system.

Thus, in this case our analysis has been shown to be better than LFA in terms of its usefulness for studying convergence rates. Its main advantage appears in the design of inter-grid filters and smoothing operators.

6.3. Fourier–Hadamard harmonic analysis: the mixture of two different bases

We now consider a 2D system that corresponds to a mixture of the system from Section 6.1 and the system from Section 6.2. Let $A_x \in \mathbb{R}^{16\times 16}$ be the system matrix from Section 6.1 and let $A_y \in \mathbb{R}^{16\times 16}$ be the system matrix from Section 6.2 for $l=5$, $N=16$. Then, we define a 2D system by taking the Kronecker sum of these two operators. That is,

$$A_{xy} = A_x \oplus A_y \tag{79}$$

$$= A_x \otimes I_y + I_x \otimes A_y \tag{80}$$

Table V. Spectral radii of modal convergence operators for different sizes of the system in Section 6.2.

N	$\|\Gamma_{Lx\to Lx}\|$	$\|\Gamma_{Hx\to Lx}\|$	$\|\Gamma_{Lx\to Hx}\|$	$\|\Gamma_{Hx\to Hx}\|$
4	0.4375	0.7844	0.2296	0.8438
8	0.5179	0.8122	0.2641	0.9183
16	0.5843	0.8466	0.3737	0.9586
32	0.6279	0.8893	0.4322	0.9791
64	0.6624	0.9708	0.4645	0.9895

The results consider a two-grid approach, using LI/FW as the inter-grid operators, and assuming the Hadamard basis as eigenvectors of the system matrix (valid assumption) and inter-grid filter (wrong assumption).

Table VI. Spectral radii of modal convergence operators for different sizes of the system in Section 6.2.

N	$\|\Gamma_{Lx\to Lx}\|$	$\|\Gamma_{Hx\to Lx}\|$	$\|\Gamma_{Lx\to Hx}\|$	$\|\Gamma_{Hx\to Hx}\|$
4	0.2205	0.6765	0.3841	0.8843
8	0.2782	0.7038	0.4527	0.9630
16	0.3597	0.6907	0.4660	0.9915
32	0.4150	0.7805	0.4770	0.9978
64	0.4514	0.8945	0.4879	0.9995

The results consider a two-grid approach, using LI/FW as the inter-grid operators, and assuming Fourier harmonic modes as eigenvectors of the system matrix (wrong assumption) and inter-grid filter (valid assumption).

Thus, the system matrix $A_{xy} \in \mathbb{R}^{256\times256}$ is a mixture of matrices with different eigenvectors. Although the problem does not represent any well-known system in applications, we choose it in order to show how our analysis applies to mixtures of very different systems. A more realistic scenario of this kind would be, for example, a 2D diffusion equation with a diffusion coefficient that varies along one of the dimensions. The difficulty in that case is to check the harmonic aliasing property, which thus remains a problem for future research.

Since A_y does not have constant stencil coefficients, neither does A_{xy}. Therefore the assumptions of LFA are not fulfilled. However, since the system fulfills the assumptions introduced in Section 4, we are able to apply our modal analysis.

Here, the eigenvectors of the system matrix A_{xy} are given by $W_x \otimes W_y$, where W_x are Fourier harmonic modes and W_y are Hadamard harmonic modes. From the results of Section 5.2, we know that although the eigenvectors of a system represented by sums of Kronecker products are separable, the convergence rates are not. Thus, the problem of design of inter-grid operators cannot, in general, be considered with any one dimension independent of any other. Now, since in the y-dimension we can actually implement optimal inter-grid operators using the sharp inter-grid filter in (78), this allows us to decouple the two problems. Then, if we choose the inter-grid filter $F_{xy} = F_x \otimes F_y$ with the 1-pass LI/FW inter-grid filter as F_x (suitable for Fourier harmonic eigenvectors) and the sharp inter-grid filter in (78) as F_y (optimal for Hadamard harmonic modes) we obtain the convergence rates shown in Table VII for the two-grid algorithm. This combination of inter-grid filters completely removes the cross-modal convergence factors with modal transfers $Hy \to Ly$ and $Ly \to Hy$. For the modal transfers $Hy \to Hy$, we observe complete removal of cross-modal error components ($HxHy \to LxHy$ and $LxHy \to HxHy$) and complete transfer of self-mode error components ($LxHy \to LxHy$ and $HxHy \to HxHy$). For the modal transfers $Ly \to Ly$, we observe results similar to those obtained for the 1-pass LI/FW inter-grid filter in Section 6.1.

As we did in the previous example, we can ignore the fact that the assumptions for LFA are not fulfilled in this problem and we can compute its estimates for the convergence rates. These results are shown in Table VIII, where we see that the estimates are not too far from the estimates of our modal analysis. The disadvantage of LFA, other than working as an approximation, is in the interpretation of these results as it shows that there is no decoupling between the two dimensions of the problem.

Finally, we consider the use of different inter-grid operators for which we make a common choice of using a 2D LI/FW operator. This operator leads to an inter-grid filter $F_{xy} = F_x \otimes F_y$

Table VII. Spectral radii of modal convergence operators for the system in Section 6.3 using our modal analysis.

$\|\Gamma_{xy}\|$	*LxLy*	*HxLy*	*LxHy*	*HxHy*
$\hookrightarrow$ *LxLy*	0.4532	0.8503	0	0
$\hookrightarrow$ *HxLy*	0.4611	0.9994	0	0
$\hookrightarrow$ *LxHy*	0	0	1	0
$\hookrightarrow$ *HxHy*	0	0	0	1

The 16 convergence factors are organized according to the subscripts of modal convergence operators indicating transfer from the four combinations of modes in the columns to the four combinations of modes in the rows. The results consider a two-grid approach, using a 1-pass LI/FW inter-grid filter for the x-dimension and the sharp inter-grid filter in (78) for the y-dimension.

Table VIII. Spectral radii of modal convergence operators for the system in Section 6.3 using LFA (under wrong assumptions).

$\|\Gamma_{xy}\|$	*LxLy*	*HxLy*	*LxHy*	*HxHy*
$\hookrightarrow$ *LxLy*	0.6063	0.8420	0.4523	0.2935
$\hookrightarrow$ *HxLy*	0.4547	0.9995	0.2080	0.2024
$\hookrightarrow$ *LxHy*	0.4523	0.2935	0.9965	0.1878
$\hookrightarrow$ *HxHy*	0.2080	0.2024	0.1322	1.0000

The 16 convergence factors are organized according to the subscripts of modal convergence operators indicating transfer from the four combinations of modes in the columns to the four combinations of modes in the rows. The results consider a two-grid approach, using a 1-pass LI/FW inter-grid filter for the x-dimension and the sharp inter-grid filter in (78) for the y-dimension.

Table IX. Spectral radii of modal convergence operators for the system in Section 6.3 using our modal analysis (under incorrect assumptions).

$\|\Gamma_{xy}\|$	*LxLy*	*HxLy*	*LxHy*	*HxHy*
$\hookrightarrow$ *LxLy*	0.7126	0.8287	0.7548	0.2509
$\hookrightarrow$ *HxLy*	0.4533	0.9997	0.1892	0.1798
$\hookrightarrow$ *LxHy*	0.3730	0.2177	0.9982	0.2957
$\hookrightarrow$ *HxHy*	0.1432	0.1433	0.2226	1.0000

The 16 convergence factors are organized according to the subscripts of modal convergence operators indicating transfer from the four combinations of modes in the columns to the four combinations of modes in the rows. The results consider a two-grid approach, using a 1-pass LI/FW inter-grid filter in both x- and y-dimensions. It is assumed that Fourier harmonic modes are eigenvectors of the operators in the x-dimension (valid assumption) and Hadamard basis are eigenvectors of the operators in the y-dimension (valid for the system matrix and false for the inter-grid filter).

Table X. Spectral radii of modal convergence operators for the system in Section 6.3 using LFA (under incorrect assumptions).

$\|\Gamma_{xy}\|$	*LxLy*	*HxLy*	*LxHy*	*HxHy*
$\hookrightarrow$ *LxLy*	0.6722	0.8313	0.6119	0.3030
$\hookrightarrow$ *HxLy*	0.4553	0.9996	0.2253	0.2177
$\hookrightarrow$ *LxHy*	0.4714	0.3026	0.9999	0.2528
$\hookrightarrow$ *HxHy*	0.2257	0.2177	0.1890	1.0000

The 16 convergence factors are organized according to the subscripts of modal convergence operators indicating transfer from the four combinations of modes in the columns to the four combinations of modes in the rows. The results consider a two-grid approach, using a 1-pass LI/FW inter-grid filter in both x- and y-dimensions. It is assumed that Fourier harmonic modes are eigenvectors of the operators in both x- and y-dimensions (false only for the system matrix in the y-dimension).

where both F_x and F_y are 1D, 1-pass LI/FW filters. As in the example of Section 6.2, this choice of inter-grid operators makes both our modal analysis and LFA not applicable for this problem. In Tables IX and X, we can see the estimates of our analysis, based on a Fourier–Hadamard basis and LFA, respectively. The results are very similar and our analysis shows slightly pessimistic results compared with LFA.

There are many disadvantages for this choice of inter-grid operators. First and most important, it does not allow us to define groups of L and H modes. Second, by an arbitrary definition of these groups of modes using either our analysis or LFA, we see a high coupling in the cross-modal convergence rates. Finally, the convergence rate for the modal transfer $LxLy \rightarrow LxLy$ frequencies, which is the most important task for the two-grid algorithm, is far from the convergence rate achieved by the Fourier–Hadamard inter-grid operators in Table VII. This last fact has a consequence in the final algorithm which can be observed by using a smoothing filter $S_{xy} = S_x \otimes S_y$, where S_x and S_y correspond to the Richardson iteration scheme as configured in Sections 6.1 and 6.2, respectively. Then, a single full two-grid step ($\mu_1 = 1$) with $\nu_1 = \nu_2 = 1$ shows a convergence rate of $\|(SK)^2\| = 0.2301$ for our inter-grid configuration compared with $\|(SK)^2\| = 0.3037$ obtained by using a 2D LI/FW inter-grid operator.

Here, our analysis has been found to be better than LFA for the design of a 2D inter-grid filter, as the combination of LI/FW with a sharp inter-grid filter shows good performance and perfect decoupling between the convergence rates of different dimensions.

7. CONCLUSIONS

In this paper we introduced new tools for the analysis of the linear multigrid algorithm. These tools allowed us to reveal and study the roles of the smoothing and inter-grid operators in multigrid convergence. In most applications of multigrid methods, these operators are designed based on the geometry and heuristics of the problem. We see this as a big problem for distributed applications because in such scenarios it is essential to minimize the number of iterations the algorithm requires to converge.

The main contribution of this paper is the establishment of a new approach to convergence analysis and new design techniques for inter-grid and smoothing operators. We have shown how this analysis is different than LFA, which is considered to be the standard tool for the analysis and design of multigrid methods [7]. Our study shows the clear advantages of our approach when facing systems with non-uniform stencils. By considering different systems, we showed that there is no general approach to optimizing the multigrid operators for a given system. For systems with Fourier harmonic modes as eigenvectors, we face a trade-off between the computational complexity and the convergence rate of each multigrid step. For systems with a Hadamard basis as eigenvectors, we are able to obtain optimal multigrid operators that make the algorithm converge in one step, with $\mathcal{O}(1)$ smoothing iterations, which is possible due to the particular structure of the system. The same multigrid operators show a perfect decoupling in a mixture of two different systems where one of the operators has a Hadamard basis as eigenvectors. Our modal analysis has been shown to be crucial to unveil these properties and to show the exact influence of each operator on the convergence behavior of the algorithm.

We note that, given the assumptions imposed on the system, we were able to analyze multigrid convergence with no heuristics based on the geometry of the problem. This opens the possibility of designing a fully AMG method if the correct assumptions are satisfied. Nevertheless, this is not a straightforward step because the harmonic aliasing property is strongly connected with the geometry of the problem. The main difficulty in our approach is to check our assumptions on the eigenvectors of the system. For future research, we are studying practical methods to check these assumptions and modifications that can make them more flexible to check and manage.

REFERENCES

1. Brandt A. Algebraic multigrid theory: the symmetric case. *Applied Mathematics and Computations* 1986; **19**: 23–56.
2. Ruge JW, Stüben K. Algebraic multigrid (AMG). In *Multigrid Methods*, *Frontiers in Applied Mathematics*, vol. 3, McCormick SF (ed.). SIAM: Philadelphia, PA, 1987; 73–130.
3. Brandt A, McCormick SF, Ruge JW. Algebraic multigrid (AMG) for sparse matrix equations. In *Sparsity and its Applications*, Evans DJ (ed.). Cambridge University Press: Cambridge, 1984.
4. Yang UM. Parallel algebraic multigrid methods high performance preconditioners. In *Numerical Solutions of PDEs on Parallel Computers*, Bruaset AM, Bjrstad P, Tveito A (eds), Lecture Notes in Computational Science and Engineering: Springer: Berlin, 2005.
5. Brandt A. Rigorous quantitative analysis of multigrid, I: constant coefficients two-level cycle with l2-norm. *SIAM Journal on Numerical Analysis* 1994; **31**(6):1695–1730.
6. Brandt A. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation* 1977; **31**: 333–390.
7. Trottenberg U, Oosterlee CW, Schüller A. *Multigrid*. Academic Press: London, 2000.
8. Mallat S. *A Wavelet Tour of Signal Processing* (2nd edn), Wavelet Analysis and its Applications. Academic Press: New York, 1999.
9. Briggs WL, Henson VE, McCormick SF. *A Multigrid Tutorial* (2nd edn). SIAM: Philadelphia, PA, 2000.
10. Wesseling P. *An Introduction to Multigrid Methods*. Wiley: Chichester, 1992.
11. Proakis JG, Manolakis DG. *Digital Signal Processing* (2nd edn), Principles, Algorithms, and Applications. Macmillan: Indianapolis, IN, 1992.
12. Laub AJ. *Matrix Analysis for Scientists and Engineers*. SIAM: Philadelphia, PA, 2005.
13. Davis PJ. *Circulant Matrices*. A Wiley-Interscience Publication, Pure and Applied Mathematics. Wiley: New York, Chichester, Brisbane, 1979.
14. Bremaud P. *Markov Chains*: *Gibbs Fields*, *Monte Carlo Simulation*, *and Queues*. Springer: New York, 1999.
15. Neuts MF. *Matrix-Geometric Solutions in Stochastic Models*: *An Algorithmic Approach*. Johns Hopkins University Press: Baltimore, MD, 1981.
16. Sylvester JJ. Thoughts on inverse orthogonal matrices, simultaneous sign-successions, and tesselated pavements in two or more colours, with applications to newton's rule, ornamental tile-work, and the theory of numbers. *Philosophical Magazine* 1867; **34**:461–475.
17. De Sterck H, Manteuffel T, McCormick SF, Nguyen Q, Ruge JW. Markov chains and web ranking: a multilevel adaptive aggregation method. *Thirteenth Copper Mountain Conference on Multigrid Methods*, Copper Mountain, CO, U.S.A., 2007.