



HAL
open science

Toward an algebraic multigrid method for the indefinite Helmholtz equation

Robert D Falgout, Matthieu Lecouvez, Pierre Ramet, Clément Richefort

► **To cite this version:**

Robert D Falgout, Matthieu Lecouvez, Pierre Ramet, Clément Richefort. Toward an algebraic multigrid method for the indefinite Helmholtz equation. 2024. cea-04620991v2

HAL Id: cea-04620991

<https://cea.hal.science/cea-04620991v2>

Preprint submitted on 28 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **TOWARD AN ALGEBRAIC MULTIGRID METHOD FOR THE**
2 **INDEFINITE HELMHOLTZ EQUATION ***

3 ROBERT D. FALGOUT[†], MATTHIEU LECOUCVEZ[‡], PIERRE RAMET[§], AND CLÉMENT
4 RICHEFORT[‡]

5 **Abstract.** It is well known that multigrid methods are very competitive in solving a wide range
6 of SPD problems. However achieving such performance for non-SPD matrices remains an open prob-
7 lem. In particular, three main issues may arise when solving a Helmholtz problem : some eigenvalues
8 may be negative or even complex, requiring the choice of an adapted smoother for capturing them,
9 and because the near-kernel space is oscillatory, the geometric smoothness assumption cannot be
10 used to build efficient interpolation rules. Moreover, the coarse correction is not equivalent to a pro-
11 jection method since the indefinite matrix does not define a norm. We present some investigations
12 about designing a method that converges in a constant number of iterations with respect to the
13 wavenumber. The method builds on an ideal reduction-based framework and related theory for SPD
14 matrices to improve an initial least squares minimization coarse selection operator formed from a set
15 of smoothed random vectors. A new coarse correction is proposed to minimize the residual in an
16 appropriate norm for indefinite problems. We also present numerical results at the end of the paper.

17 **Key words.** Algebraic Multigrid, Helmholtz Equation, Linear Algebra, Indefinite matrix

18 **1. Introduction.** The numerical simulation of various physical phenomena leads
19 to potentially very large linear systems of equations written $A\mathbf{x} = \mathbf{b}$ in matrix form.
20 These systems can be solved directly by a convenient factorization of A , or iteratively
21 by computing and refining an approximation of the solution \mathbf{x} starting from an initial
22 guess \mathbf{x}_0 . Multigrid methods [7, 26] work iteratively and are known to be scalable
23 and quasi-optimal for solving sparse linear systems of equations for many classes of
24 problems. Each multigrid iteration combines a projection method on a coarser space
25 to capture the eigenvectors associated with the small eigenvalues, and a few iterations
26 of a smoothing method to capture the remaining eigenvectors generally associated
27 with the large eigenvalues.

28
29 In this paper, we investigate an algebraic multigrid method (AMG) for the indefi-
30 nite Helmholtz equation. We start by introducing an alternative smoother and new
31 interpolation rules. Thereafter, we demonstrate that the coarse correction process
32 can easily amplify the error in the indefinite case, and therefore needs to be updated.
33 Finally, numerical experiments are presented in the last section.

34
35 To simplify the discussion in what follows, we use the term "small/large eigenvec-
36 tor" to designate an eigenvector with small/large eigenvalue. We similarly say "pos-
37 itive/negative eigenvector" when referring to the eigenvalue sign. Additionally, capi-
38 tal italic Roman letters (A, E, P) denote matrices and bold lowercase letters denote
39 vectors ($\mathbf{u}, \mathbf{v}, \mathbf{r}, \boldsymbol{\alpha}$). Other lowercase letters denote scalars (σ, λ), while capital calli-
40 graphic letters denote sets and spaces ($\mathcal{C}, \mathcal{F}, \mathcal{K}$).

*This work was funded by CEA. This work was performed under the auspices of the U.S. Depart-
ment of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344
(LLNL-JRNL-865914).

[†]Lawrence Livermore National Laboratory, Livermore, CA (rvalgout@llnl.gov)

[‡]Commissariat à l’Energie Atomique et aux Energies Alternatives (CEA)
(matthieu.lecouvez@cea.fr, richefort.clement@protonmail.com)

[§]Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France
(pierre.ramet@inria.fr)

41 **1.1. Algebraic Multigrid methods.** While errors composed of small eigenvec-
 42 tors are known to be more difficult to eliminate for most iterative methods, multigrid
 43 methods accelerate the convergence to the solution by projecting them onto a coarser
 44 space. The coarse projection of those difficult eigenvectors is repeated recursively un-
 45 til reaching a small enough coarse matrix for which the factorization by a direct solver
 46 is fast. Assuming the matrix is symmetric positive definite (SPD), the best approxi-
 47 mation of the solution within the coarse projection space is computed by minimizing
 48 the approximation error in A -norm. The core idea in multigrid methods is to make
 49 this projection practical by recursively defining smaller subspaces by way of sparse
 50 operators P_l , called interpolation operators. The computation of \mathbf{x} is accelerated by
 51 way of a hierarchy of coarse problems $A_l \mathbf{x}_l = \mathbf{r}_l$, where \mathbf{r}_l is the residual of the level
 52 l in the grid hierarchy. P_l determines the coarse projection subspace of the level l ,
 53 and transfers the information from level $l + 1$ to l . In most symmetric applications,
 54 coarse matrices are constructed following the Galerkin formula $A_{l+1} = P_l^T A_l P_l$. The
 55 two-level coarse correction operator denoted by

$$56 \quad (1.1) \quad \Pi_A(P) := P(P^T A P)^{-1} P^T A$$

57 is an A -orthogonal projector onto $\text{range}(P)$ and coincides with a minimization problem
 58 in the SPD case such that

$$59 \quad (1.2) \quad \arg \min_{\tilde{\mathbf{x}} \in \text{span}\{P\}} \|\mathbf{x} - \tilde{\mathbf{x}}\|_A = \Pi_A(P) \mathbf{r}.$$

60 Two-level methods actually need both types of solvers. The coarse correction (1.1)
 61 requires a direct method for factorizing the coarsest matrix whereas the remaining
 62 error is eliminated on the fine level through a few iterations of an iterative method
 63 called a smoother. From Equation (1.1), the error propagation matrix for the coarse
 64 correction of a two-level method is

$$65 \quad (1.3) \quad E = I - \Pi_A(P).$$

66 Likewise, the error propagation matrix for the smoother is

$$67 \quad (1.4) \quad E_M = I - M^{-1} A$$

68 where M^{-1} is an approximation of A^{-1} . The smoother is applied before each restric-
 69 tion and after each interpolation, as illustrated in Algorithm 1.1.

Algorithm 1.1 Two-level cycle

```

1: Inputs :  $\mathbf{b}$  right-hand side,  $\tilde{\mathbf{x}}$  approximation of  $\mathbf{x}$  or initial guess,  $\mathbf{r} = \mathbf{b} - A\tilde{\mathbf{x}}$  residual
2:        $M$  smoother,  $P$  interpolation operator
3: for  $j = 1, \nu$  do
4:    $\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} + M^{-1} \mathbf{r}$ 
5:    $\mathbf{r} \leftarrow \mathbf{b} - A\tilde{\mathbf{x}}$ 
6: end for
7:  $\mathbf{r}_C \leftarrow P^T \mathbf{r}$ 
8:  $\tilde{\mathbf{e}}_C \leftarrow \text{Solve}(P^T A P, \mathbf{r}_C)$ 
9:  $\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} + P\tilde{\mathbf{e}}_C$ 
10:  $\mathbf{r} \leftarrow \mathbf{b} - A\tilde{\mathbf{x}}$ 
11: for  $j = 1, \nu$  do
12:    $\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} + M^{-1} \mathbf{r}$ 
13:    $\mathbf{r} \leftarrow \mathbf{b} - A\tilde{\mathbf{x}}$ 
14: end for
15: Output :  $\tilde{\mathbf{x}}$  approximation of  $\mathbf{x}$  at the end of the cycle

```

70 Finding a smoother and a coarse correction that are complementary is a major concern
 71 in the design of the method. Moreover, the context in which a multigrid method
 72 is applied determines what kind of operators should be used in the method. In particular,
 73 the near-kernel space of smallest eigenvectors is especially important in the design
 74 of interpolation.

75

76 In elliptic problems such as the Laplace equation whose spectrum is illustrated in Figure
 77 1.1, the convergence of multigrid methods is well known. The matrix A is SPD,
 78 so smoothers like w -Jacobi or Gauss-Seidel are known to be good smoothers since
 79 they damp the large eigenvectors without modifying the small ones. In this elliptical
 80 case, these small and large eigenvectors are characterized by low and high frequency
 81 oscillations respectively. Hence, while the smoother damps the oscillatory modes, the
 82 interpolation must target the slowly varying modes associated with small eigenvalues
 83 (see Figure 1.1b). For this reason, the geometric smoothness of the near-kernel space
 84 is generally a key assumption, and makes the construction of good interpolation rules
 85 more convenient in the initialization of the method.

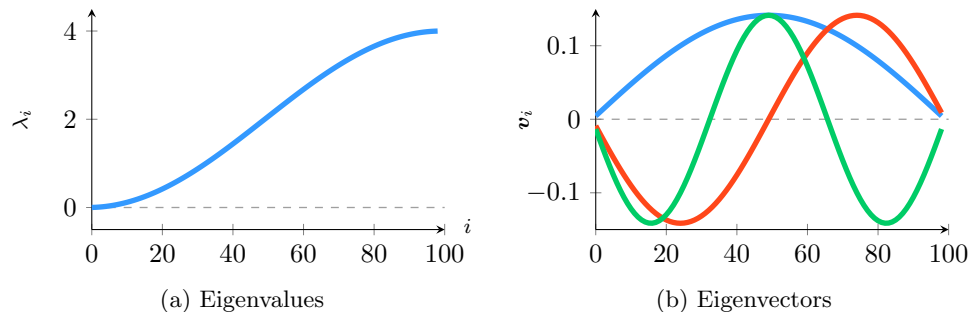


Fig. 1.1: Laplace eigenvalues and three smallest eigenvectors

86 Likewise in classic algebraic multigrid [4, 23, 27, 24, 12], the interpolation operators
 87 are designed to target what is called algebraically smooth components. The smoothed
 88 aggregation method [10] is particularly efficient for solving problems with an *a priori*
 89 known near-kernel space, for instance in diffusion [29] or elasticity [28] where the
 90 target small eigenvectors are the constant vector and rigid body modes respectively.
 91 Those vectors are split between disjoint aggregates over the entire domain to initiate a
 92 tentative block interpolation operator. A few smoothing iterations are applied to the
 93 tentative interpolation operator, extending its pattern and removing high frequency
 94 components from its range. Usually, a few iterations of Jacobi relaxation are enough,
 95 but this step of energy reduction has been generalized to Krylov methods such as the
 96 conjugate gradient [21] by enforcing sparsity constraints in the Krylov space to keep
 97 a practical interpolation operator. If near-kernel space information is lacking, test
 98 vectors can be computed algebraically, as in adaptive smoothed aggregation [6].

99

100 Furthermore, because the choice of the interpolation strategy is essential in the convergence
 101 of the method, an ideal framework maximizing the complementarity between the smoother
 102 and the coarse correction [13] has been established to guide algorithm development. While
 103 this idealistic scenario of convergence is mostly used as a theoretical tool, some reduction-based
 104 methods enable a good approximation of the ideal interpolation operator.
 105

106 **1.2. Why Helmholtz problems are difficult for multigrid.** The Helmholtz
 107 equation (1.5) involves indefinite matrices with potentially wide and oscillatory near-
 108 kernel spaces [11]. This equation is our target in this paper.

$$109 \quad (1.5) \quad (\text{Continuous Helmholtz problem}) \Leftrightarrow \begin{cases} -\Delta \mathbf{u} - k^2 \mathbf{u} = \mathbf{f} & \text{on } \Omega \\ + \text{b. c.} & \text{on } \partial\Omega \end{cases}$$

110 In fact, the Helmholtz equation can be seen as a shifted Poisson equation, where
 111 geometrically smooth eigenvectors (i.e., low Fourier modes, see Figure 1.1b) can be
 112 negative eigenvectors because of the shift. In the same way, the smallest eigenvectors
 113 of the shifted Laplacian are higher in frequency (see Figure 1.2b).

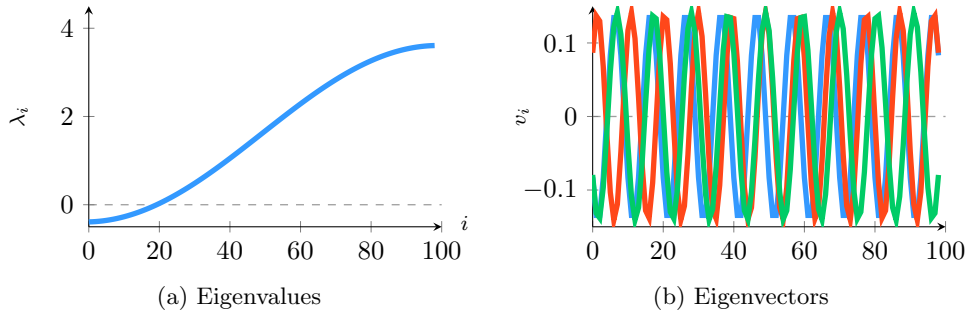


Fig. 1.2: Helmholtz eigenvalues and three smallest eigenvectors

114 This complication breaks the near-kernel space geometric smoothness assumption, a
 115 keystone of many multigrid methods. To design a coarse correction and smoothers
 116 that are complementary in this context, interpolation rules must reproduce the near-
 117 kernel oscillation, and contrary to usual relaxation methods, smoothers have to deal
 118 with both positive and negative eigenvalues. More importantly, the coarse correction
 119 is not equivalent to a minimization problem anymore since the indefinite matrix does
 120 not define a norm (i.e., the equality (1.2) is not valid for Helmholtz). Whereas the
 121 coarse correction is guaranteed to not amplify the error for SPD matrices, the ap-
 122 proximation error can be amplified in the indefinite case because the spectrum of the
 123 matrix has both signs.

124

125 For these reasons, finding a recurring process to build a scalable multilevel method
 126 is still an open question. Multiple correction [19], wave-ray [5, 18], and Complex-
 127 Shifted Laplacian [9] approaches have already been investigated to address this issue.
 128 In this paper, we present a fully algebraic approach built on ideal reduction-based
 129 ideas, and demonstrate its potential for solving the Helmholtz problem with constant
 130 iteration count independent of the wavenumber k . Certain discretization matrices re-
 131 sulting from the continuous problem (1.5) can be non-symmetric due to the boundary
 132 conditions. To center the discussion on the indefinite nature of Helmholtz, the next
 133 approaches address the symmetric indefinite shifted laplacian matrix arising from the
 134 following 5-pts stencil

$$135 \quad (1.6) \quad \hat{A} = \begin{bmatrix} & -1 & & & \\ -1 & 4 - (kh)^2 & -1 & & \\ & & & & \\ & & & & \\ & & & & -1 \end{bmatrix},$$

136 where kh is the shift and h the step size. In Section 2, we start by presenting a normal
 137 equation polynomial smoother specifically designed to damp the desired proportion

138 of largest eigenvalues independently of their signs, while interpolation rules for prop-
 139 agating oscillatory near-kernel information are established in Section 3. The Section
 140 4 gives more details on why the indefiniteness can corrupt the coarse correction by
 141 introducing a concept of pollution and Section 5 exposes an alternative coarse correc-
 142 tion to the classical one which avoids some divergence scenarios. Finally, Section 6
 143 presents benchmarks of this new multigrid method for different Helmholtz problems,
 144 with varying shift kh and wavenumber k .

145

146 In this paper, \mathbf{v}_i denotes the i^{th} eigenvector of A associated with the eigenvalue
 147 λ_i . Moreover, we always assume the eigenvalues are ordered in magnitude (i.e.,
 148 $\forall i < n, |\lambda_i| \leq |\lambda_{i+1}|$) such that $V_c := [\mathbf{v}_1, \dots, \mathbf{v}_{n_c}]$ and $V_f := [\mathbf{v}_{n_c+1}, \dots, \mathbf{v}_n]$
 149 contain the small and large eigenvector sets of size n_c and n_f respectively. Naturally,
 150 the full set of eigenvectors are given by $V = [V_c, V_f]$, and $n = n_c + n_f$.

151 **2. Polynomial Smoothers for Indefinite Problems.** Knowing the behavior
 152 of the smoother on the spectrum is interesting to guarantee the effectiveness of the
 153 cycle. Here, the smoother must damp large positive and negative eigenvalues, which
 154 is problematic for most standard methods. Generally, a polynomial method with
 155 degree greater than one can work. Krylov iterations are good polynomial smoothers
 156 in the indefinite case but they minimize the global residual norm regardless of the
 157 eigenvalues and are non-linear because of their right-hand side dependence. A linear
 158 polynomial is more convenient for generating the set of smoothed candidates vectors
 159 needed to construct the interpolation operator described in Section 3.

160 **2.1. General considerations on polynomial smoothers.** One way to en-
 161 sure that both positive and negative eigenvectors are damped is to consider a normal
 162 equation polynomial smoother i.e., a smoother working on A^2 . In general, the de-
 163 gree m of the polynomial must be greater than one to damp positive and negative
 164 eigenvectors, as the polynomial illustrated in Figure 2.1 does. Resorting to normal
 165 equations enables the polynomial to treat eigenvalues with respect to their magnitude
 166 rather than their sign, which is equivalent to working with even powers of A if the
 167 matrix is hermitian, which is what we assume in this section. In the future, it might
 168 be interesting to investigate more general polynomials with odd exponents. In this
 169 first approach, we use the convenient symmetry property enabled by normal equations
 170 in the Chebyshev framework.

171

172 Let $p_m(A^2)$ be a polynomial of degree m that approximates A^{-2} . From Equation
 173 (1.4), let $q_{m+1}(A^2)$ be the associated error propagation matrix of the polynomial
 174 smoother such that

$$175 \quad (2.1) \quad q_{m+1}(A^2) := I - p_m(A^2)A^2.$$

176 Additionally, let \mathbf{v} be an eigenvector of A associated with the eigenvalue λ . Hence,

$$177 \quad (2.2) \quad q_{m+1}(\lambda^2)\mathbf{v} = (1 - p_m(\lambda^2)\lambda^2)\mathbf{v}.$$

178 In multigrid methods, a good smoother eliminates the large eigenvalues that the coarse
 179 correction does not capture and vice-versa. Let a and b be real scalars such that
 180 $0 < a < b$. Assume these large squared eigenvalues are contained in the interval $[a, b]$.
 181 The construction of a relevant interval will be discussed in the next section. Since
 182 the polynomial smoother $p_m(A^2)$ is an inverse approximation of A^{-2} , the polynomial
 183 function $p_m(x)$ can be constructed to approximate the function x^{-1} [16] from $m +$

184 1 interpolation points x_i selected within the interval of large eigenvalues $[a, b]$. In
 185 particular, selecting the scaled first kind Chebyshev polynomial roots as interpolation
 186 points

$$187 \quad (2.3) \quad x_i := \frac{b+b}{2} + \frac{b-a}{2} \cos\left(\frac{(2i+1)\pi}{2(m+1)}\right), \quad i = 1, \dots, m+1.$$

188 gives the minimal error propagation function $q_{m+1}(x)$ on the interval $[a, b]$. The
 189 polynomial is constructed to satisfy the $m+1$ following constraints

$$190 \quad (2.4) \quad x_i \in [a, b], \quad p_m(x_i) = \frac{1}{x_i} \Leftrightarrow q_m(x_i) = 0, \quad i = 1, \dots, m+1.$$

191 Because the selected nodes x_i are the roots of q_{m+1} and $q_{m+1}(0) = 1$, then the
 192 Lagrange formula yields

$$193 \quad (2.5) \quad p_m(x) := \sum_{i=1}^{m+1} \frac{1}{x_i} \prod_{j=1, j \neq i}^{m+1} \frac{x-x_j}{x_i-x_j}, \quad q_{m+1}(x) = \prod_{i=1}^{m+1} \frac{x-x_i}{-x_i}.$$

194 First kind Chebyshev polynomials are defined by the three-terms recurrence relation

$$195 \quad (2.6) \quad \forall t \in [-1, 1], \quad C_0(t) = 1, \quad C_1(t) = t, \quad C_{m+1}(t) = 2tC_m(t) - C_{m-1}(t).$$

196 The roots of q_{m+1} are the roots of C_{m+1} but scaled on $[a, b]$, the error propagation
 197 function q_{m+1} can be derived as the following re-scaled Chebyshev polynomial

$$198 \quad (2.7) \quad q_{m+1}(x) = \frac{C_{m+1}\left(\frac{b+a-2x}{b-a}\right)}{C_{m+1}\left(\frac{b+a}{b-a}\right)}.$$

199 As explained in [2], the upper bound of $C_{m+1}(t)$ on $[-1, 1]$ equals one for $t = 1$
 200 and is strictly monotonically increasing for $t > 1$. Accordingly, the supremum of
 201 the numerator on $[a, b]$ equals one for $x = a$, and the denominator is strictly greater
 202 than one because $\frac{b+a}{b-a} > 1$. Last, $q_{m+1}(0) = 1$ and q_{m+1} is strictly monotonically
 203 decreasing for $x \in [0, a]$. As a consequence, $|q_{m+1}(x)| < 1$ on the interval $(0, b]$.
 204 Assuming $b \geq \lambda_{\max}^2$, then the spectral radius $\rho(q_{m+1}(A^2)) < 1$. In other words, the
 205 smoother is a convergent iterative method and does not amplify any region of the
 206 spectrum.

207 **2.2. Constructing an appropriate target interval.** One way to determine
 208 an interval $[a, b]$ without preliminary information [2, 1] is to compute a few power
 209 iterations to determine b by an overestimation of the largest eigenvalue, and choose
 210 the lower bound a according to b , for example $a = \frac{1}{2}b$. However, to respect the com-
 211 plementarity principle, the percentage of damped eigenvalues by the smoother must
 212 approximate the proportion of non-coarse variables (i.e. the n_f largest eigenvalues
 213 in our case). For instance, if a coarse level is one quarter the size of the finer level,
 214 then three-quarters of the largest amplitude eigenvectors should be damped by the
 215 smoother, while the coarse correction deals with the remaining small eigenvectors.
 216 Consequently, since eigenvalues are not necessarily uniformly separated, a should be
 217 determined so that a proportion of eigenvalues belongs to the interval $[a, b]$. More-
 218 over, the spectral distribution of coarse matrices are unknown in a multi-level setting.
 219 Therefore, a good interval should satisfy

$$220 \quad (2.8) \quad \lambda_i^2 \in [a, b] \Leftrightarrow \lambda_i \in [-\sqrt{b}, -\sqrt{a}] \cup [\sqrt{a}, \sqrt{b}], \quad i = n_c, \dots, n_f.$$

221 While this interval can be fixed using geometric information, we first compute a
 222 rough approximation of the matrix *spectral density* as detailed in [17]. This spectral
 223 density permits to determine which portion of the spectrum should be damped by
 224 the smoother, and is defined by the distribution function $\phi(t)$ that represents the
 225 probability of finding an eigenvalue at each point $t \in [-1, 1]$. We set the lower bound
 226 a of the Chebyshev node interval in a second step so that the probability within
 227 the interval equals the target proportion, for instance half of the total area in a
 228 scenario of exact balance between coarse and non-coarse variables. As defined in (2.6),
 229 the distribution function ϕ is approximated by a linear combination of orthogonal
 230 Chebyshev polynomial functions, such that

$$231 \quad (2.9) \quad \phi(t) = \sum_{k=1}^{\infty} \mu_k C_k(t) \approx \sum_{k=1}^{n_\mu} \mu_k C_k(t).$$

232 Because Chebyshev functions are naturally defined over $[-1, 1]$, the spectral density
 233 function must evaluate the spectral density of the scaled matrix $B = \frac{2}{b}A^2 - I$. Since b
 234 is assumed to bound the eigenvalues of A^2 , the spectrum of B belongs to $[-1, 1]$. The
 235 coefficients μ_k are then determined by a moments matching procedure, which gives

$$236 \quad (2.10) \quad \mu_k = \frac{2 - \delta_{k0}}{n\pi} \times \text{Trace}(C_k(B)).$$

237 Here, n corresponds to the matrix size and δ_{k0} the Kronecker symbol. The trace
 238 can be approximated by a stochastic trace estimation from a set of n_{vec} random
 239 and orthogonal vectors \mathbf{z}_l , where each element of these vectors is chosen following a
 240 normal distribution with zero mean and a unit standard deviation. Therefore, the
 241 trace approximations are given by

$$242 \quad (2.11) \quad \text{Trace}(C_k(B)) \approx \frac{1}{n_{\text{vec}}} \sum_{l=1}^{n_{\text{vec}}} \mathbf{z}_l^T C_k(B) \mathbf{z}_l, \quad k = 1, \dots, n_\mu.$$

243 According to (2.11), each trace can be estimated by a sample mean of n_{vec} products
 244 $\mathbf{z}_l^T C_k(B) \mathbf{z}_l$, and the n_μ vectors $C_k(B) \mathbf{z}_l$ can be computed from the three-term re-
 245 currence defined in (2.6). Once the distribution function ϕ is approximated following
 246 Equation (2.9), a rough area approximation by trapezoid rule yields a correct lower
 247 bound that satisfies a proportion around $\frac{n_f}{n}$. This lower bound only needs to be
 248 remapped on the initial interval to return the correct value for a . The interval $[a, b]$
 249 constitutes a purely algebraic interval in which the polynomial smoother is the most
 250 efficient. The bounds a and b are represented in Figure 2.1, where $x_{50\%}$ illustrates a
 251 theoretical lower bound target for the shifted laplacian matrix resulting from (1.6).
 252 Last, the total number of matrix vector products required by the spectral density
 253 approximation step for the construction of a relevant interval is $n_{\text{vec}} \times n_\mu$.

254 **3. Constructing good interpolation rules.** Interpolation operators are used
 255 both to construct the coarse level matrices and to transfer information across levels.
 256 SPD and geometric smoothness assumptions cannot be used to determine appropriate
 257 interpolation operators in our case. Some methods such as smoothed aggregation
 258 [10, 20] and bootstrap-AMG [3] use candidate vectors that are close to the near-
 259 kernel space to design the interpolation rules. These test vectors are either deduced
 260 from geometric information [5, 22] or algebraically as in adaptive multigrid methods
 261 [6]. Here, we prefer to use a fully algebraic and recurring process to create our

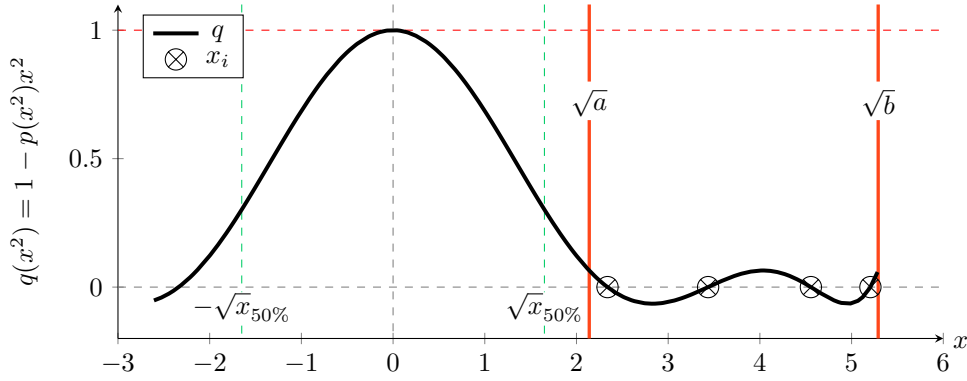


Fig. 2.1: Spectrum of the polynomial smoother error propagation matrix for $kh = 1.65$

262 interpolation operators. Candidate vectors will be generated from random vectors
 263 smoothed by the polynomial presented in Section 2, and used by the least squares
 264 minimization framework to determine good fine variable interpolation rules. This
 265 initial least squares interpolation operator is used as a coarse variable operator in the
 266 ideal reduction-based framework [13].

267 **3.1. Ideal framework.** Even though the ideal framework requires an SPD
 268 assumption and has not been generalized to indefinite problems, the idea of removing
 269 components that are handled by the smoother is of particular interest for capturing
 270 the near-kernel space of oscillatory problems, and will be our guiding principle. Ac-
 271 cordingly, we assume A is SPD in this section dedicated to the ideal framework.

272

273 Following [13], let \mathcal{C} and \mathcal{F} be complementary coarse and fine variable subsets of
 274 Ω respectively of size n_c and n_f . Let $R^T : \mathbb{R}^{n_c} \rightarrow \mathbb{R}^n$ and $S : \mathbb{R}^{n_f} \rightarrow \mathbb{R}^n$ be coarse
 275 and fine variable operators respectively, such that $RS = 0$. The space defined by the
 276 coarse variable operator R^T must be handled by the coarse correction, whereas the
 277 fine variable operator S defines a space where smoothing must operate in order to re-
 278 spect the complementarity principle. The *Ideal Interpolation* operator is a theoretical
 279 operator that is the best that satisfies $RP = I_c$, in the sense that it minimizes the
 280 difference between variables and interpolated coarse variables, within a space that is
 281 the most complementary to the range of the smoother M . The ideal interpolation
 282 operator is given by

$$283 \quad (3.1) \quad P_* = \arg \min_P \left(\max_{e \neq 0} \frac{\|(I - PR)e\|_M}{\|e\|_A} \right) = (I - S(S^T AS)^{-1} S^T A) R^T.$$

284 Let $P_{:,i}$ and $R_{:,i}^T$ be the i^{th} columns of P and R^T respectively. Each column of the
 285 ideal interpolation operator is therefore defined by

$$286 \quad (3.2) \quad P_{:,i} = R_{:,i}^T - s_i, \text{ with } s_i = \arg \min_{\tilde{s} \in \text{Range}(S)} \|R_{:,i}^T - \tilde{s}\|_A = S(S^T AS)^{-1} S^T A R_{:,i}^T$$

287 In fact, the matrix that multiplies each column of R^T in (3.2) and (3.1) is a pro-
 288 jection operator onto the A -orthogonal complement of the range of S . The ideal
 289 interpolation operator is constructed by removing components of $\text{range}(R^T)$ that are
 290 not A -orthogonal to $\text{range}(S)$. Under the assumption that the smoother captures the

291 space spanned by S , the best coarse matrix is therefore a matrix where S -related in-
 292 formation is subtracted. Even if applying $(S^T A S)^{-1}$ is too expensive, it gives insight
 293 for building a more practical method.

294 **3.2. Least Squares Minimization Interpolation Operator.** As mentioned
 295 at the beginning of Section 3.1, demonstrating that the interpolation operator (3.1)
 296 is ideal in the theoretical framework of [13] requires A to be symmetric positive-
 297 definite. However, the reduction viewpoint which consists in cleaning the range of
 298 interpolation by removing S -related components is of interest. In addition, numerical
 299 experiments reveal that the classical coarse variable operator $R^T = [0 \ I_c]^T$ does not
 300 have good approximation property for the oscillatory near-kernel space that char-
 301 acterizes Helmholtz. Therefore, a new coarse variable operator has to be designed
 302 algebraically. Using the smallest eigenvectors V_c from Section 3.1 to enforce the rep-
 303 resentation of the near-kernel space within the interpolation range is not practical.
 304 Instead, we construct a set of vectors approximating an oscillatory and potentially
 305 large near-kernel space by using the normal equations polynomial smoother developed
 306 in Section 2. In this section, we present a coarse variable operator \hat{R}^T of size $n \times n_c$
 307 constructed by a least squares minimization strategy [3]. Let the columns of T be a
 308 set of κ smoothed random vectors z_l approximating the near-kernel space such that

$$309 \quad (3.3) \quad T_{:,l} = q_{m+1}(A^2)z_l, \quad l = 1, \dots, \kappa.$$

310 where $T_{:,l}$ designates the l^{th} column of the set T . We assume a \mathcal{C}/\mathcal{F} splitting with n_c
 311 and n_f their respective size. \mathcal{C} -points are interpolated to the finer level with a sim-
 312 ple injection rule, while interpolation rules of \mathcal{F} -points are determined by the least
 313 squares minimization method presented in this section. We let \hat{R}^T have the form
 314 (3.8), where R_f^T designates the block of interpolation for the \mathcal{F} -points.

315 Let i be an \mathcal{F} -point and \hat{r}_i the vector containing the non-zero elements of the i^{th}
 317 row of \hat{R}^T . The idea consists in constructing each \mathcal{F} -point interpolation rule by min-
 318 imizing the squared difference between \mathcal{F} -values of the near-kernel candidate vectors
 319 and the interpolation from their connected \mathcal{C} -points in \mathcal{C}_i . Denote by $T_{i,:}$ a row vector
 320 containing the i^{th} values of each test vector, and $T_{\mathcal{C}_i,l}$ a vector containing the values
 321 in $T_{:,l}$ of the \mathcal{C} -points that are connected to variable i . Then

$$322 \quad (3.4) \quad \forall i \in \mathcal{F}, \quad \hat{r}_i = \arg \min_{\hat{r} \in \mathbb{C}^{\text{card}(\mathcal{C}_i)}} \sum_{l=1}^{\kappa} w_l (T_{i,l} - \hat{r} \cdot T_{\mathcal{C}_i,l})^2 =: \arg \min_{\hat{r} \in \mathbb{C}^{\text{card}(\mathcal{C}_i)}} \mathcal{L}_i(\hat{r})$$

323 where w_l are scaling weights (for instance $w_l = 1/|\lambda_l|$ if T contains near-kernel eigen-
 324 vectors). Finding the minimum of the convex loss function \mathcal{L}_i is equivalent to solving

$$325 \quad (3.5) \quad \nabla \mathcal{L}_i(\hat{r}_i) = 0.$$

326 Equation (3.5) can be rewritten element-wise

$$327 \quad (3.6) \quad \frac{\partial \mathcal{L}_i(\hat{r}_i)}{\partial \hat{r}_{ij}} = \sum_{l=1}^{\kappa} 2w_l (T_{i,l} - \hat{r}_i \cdot T_{\mathcal{C}_i,l}) T_{\mathcal{C}_i,l} = 0, \quad \forall j = 1, \dots, \text{card}(\mathcal{C}_i).$$

328 Finally, (3.6) leads to a system of linear equations to solve for each fine variable i

$$329 \quad (3.7) \quad \hat{r}_i T_{\mathcal{C}_i} W T_{\mathcal{C}_i}^T = T_i W T_{\mathcal{C}_i}^T$$

330 The matrix is full rank and the solution of Equation (3.7) is unique if we have at
 331 least $\kappa = \max_i \{\text{Card}(\mathcal{C}_i)\}$ locally linearly independent test vectors. Even if it
 332 is statistically always the case when starting from random candidate vectors, the
 333 matrix singularity can be detected during the factorization. In that special case, a
 334 pseudo-inverse can be computed to find an optimal solution in the least squares sense.

335 **3.3. Ideal approximation from least squares coarse operator.** In Section
 336 3.2, we presented a coarse variable operator for Helmholtz designed by a least squares
 337 minimization strategy. Using the framework presented in 3.1, define

$$338 \quad (3.8) \quad \hat{R}^T = \begin{bmatrix} R_f^T \\ I_c \end{bmatrix} \text{ and } \hat{S} = \begin{bmatrix} I_f \\ -R_f \end{bmatrix}$$

339 where \hat{R}^T is the least squares coarse variable operator presented in Section 3.2 and
 340 R_f^T is its \mathcal{F} -points interpolation block. Note that $\hat{R}\hat{S} = 0$ as required. Hence, since
 341 the least squares operator is designed to propagate the candidate vectors that are
 342 composed of small eigenvectors due to the Chebyshev polynomial smoother of Section
 343 2, the space spanned by \hat{S} is, by orthogonality, mostly composed of large eigenvectors.
 344 Accordingly, the aim of using the ideal framework in this oscillatory context is to im-
 345 prove the coarse variable operator by removing components that are already handled
 346 by the smoother.

347 However, two major issues arise in the use of the ideal interpolation operator (3.1).
 348 The first is a general concern related to the fine block $\hat{S}^T A \hat{S}$, which is usually not
 349 practical to invert, and would lead to a dense interpolation operator \hat{P} . To circumvent
 350 this problem, an approximation based on sparsity constraints must be applied. The
 351 second issue is related to the indefiniteness of the initial matrix. Indeed, as shown by
 352 the equation (3.2), applying the left operator is equivalent to solving a minimization
 353 problem in A -norm. However, such a norm does not exist in the indefinite case. Ignor-
 354 ing this problem may still give interesting results in practice, but we consider instead
 355 the $A^T A$ -norm to ensure the effectiveness of the interpolation operator. Since \hat{S} is
 356 sparse, we control the sparsity of \hat{P} by restricting the search space to a few columns
 357 of \hat{S} only. Define X_i to be the injection operator of ones and zeros of size $n_f \times n_i$
 358 with $n_i \leq n_f$ that selects n_i columns of \hat{S} , $\hat{S}X_i$. From (3.2), let \mathbf{s}_i be the solution of
 359 the ideal minimization problem such that

$$361 \quad (3.9) \quad \mathbf{s}_i := \arg \min_{\tilde{\mathbf{s}} \in \text{Range}(\hat{S}X_i)} \|\hat{R}_{:,i}^T - \tilde{\mathbf{s}}\|_{A^T A} = \hat{S}X_i \left(X_i^T \hat{S}^T A^T A \hat{S}X_i \right)^{-1} X_i^T \hat{S}^T A^T A \hat{R}_{:,i}^T.$$

362 Accordingly, columns of the reduction-based interpolation operator are computed by

$$363 \quad (3.10) \quad \hat{P}_{:,i} = \hat{R}_{:,i}^T - \mathbf{s}_i = \hat{R}_{:,i}^T - \hat{S}X_i \rho_{n_i},$$

364 where ρ_{n_i} is the solution of the $n_i \times n_i$ linear system

$$365 \quad (3.11) \quad X_i^T \hat{S}^T A^T A \hat{S}X_i \rho_{n_i} = X_i^T \hat{S}^T A^T A \hat{R}_{:,i}^T.$$

366 The choice of the non-zero pattern of \hat{P} must satisfy a good trade-off between ap-
 367 proximation properties of the near-kernel space and complexity. While improving the
 368 sparsity of this interpolation operator is a topic of future research, one strategy is
 369 to choose the columns of \hat{S} based on the entries of $\hat{S}^T A^T A \hat{R}_{:,i}^T$. In fact, each entry

370 corresponds to the scalar product between a column of \hat{S} and $\hat{R}_{:,i}^T$ in $A^T A$ -norm. A
 371 large entry designates a column of \hat{S} that contributes a lot to the solution of the
 372 minimization problem (3.9). The column selection phase iterates until the entries
 373 associated with the selected columns represent a percentage τ of the entire set of
 374 non-zero entries. At each iteration, the column associated with the largest entry of
 375 $\hat{S}^T A^T A \hat{R}_{:,i}^T$ is selected, which is equivalent to extending X_i with the euclidean basis
 376 vector with one at the index of the chosen column and zeros elsewhere. Because the
 377 columns with the largest entries in $\hat{S}^T A^T A \hat{R}_{:,i}^T$ are selected first, the set of selected
 378 columns is the smallest set that satisfies

379 (3.12)
$$\|X_i \hat{S}^T A^T A \hat{R}_{:,i}^T\|^2 \geq \tau \times \|\hat{S}^T A^T A \hat{R}_{:,i}^T\|^2, \text{ with } \tau \in [0, 1].$$

380 We note that even though setting $\tau = 1$ selects all the column associated with non-
 381 zero entries in the right-hand side, the remaining columns associated with zero entries
 382 are omitted, and therefore the matrix $X_i^T \hat{S}^T A^T A \hat{S} X_i$ still correspond to a principle
 383 sub-matrix of $\hat{S}^T A^T A \hat{S}$.

384
 385 The Figure 3.1 represents the error of interpolation of every eigenvector for two differ-
 386 ent shifted problems resulting from (1.6) with respect to τ . The red dots correspond
 387 to the error when no ideal approximation is used at all (i.e. $\tau = 0$ and therefore
 388 $\hat{P} = \hat{R}^T$), whereas blue and green dots represent the error of interpolation for $\tau = 0.5$
 389 and $\tau = 1$ respectively. The resulting operator complexity for a two-level method are
 390 given by Table 6.1 of the last section. Because the subspace $\hat{S} X_i$ grows with τ , larger
 391 values of τ lead to denser interpolation operators. For both shifts, the portion of the
 392 spectrum for which the least-squares minimization interpolation operator is the most
 393 accurate corresponds to the smallest eigenvalues in magnitude. This feature is an
 394 expected and desired effect of generating the set of test vectors from the polynomial
 395 smoother introduced in Section 2. However, the interpolation error increases with
 396 the shift. Therefore, the ideal approximation correction becomes necessary as the
 397 problem gets more indefinite. In particular, Figure 3.1 shows that the interpolation
 398 error decreases as more columns of \hat{S} are added to approximate the ideal interpolation
 operator. One drawback of this gain in accuracy is the fill-in of the matrix.

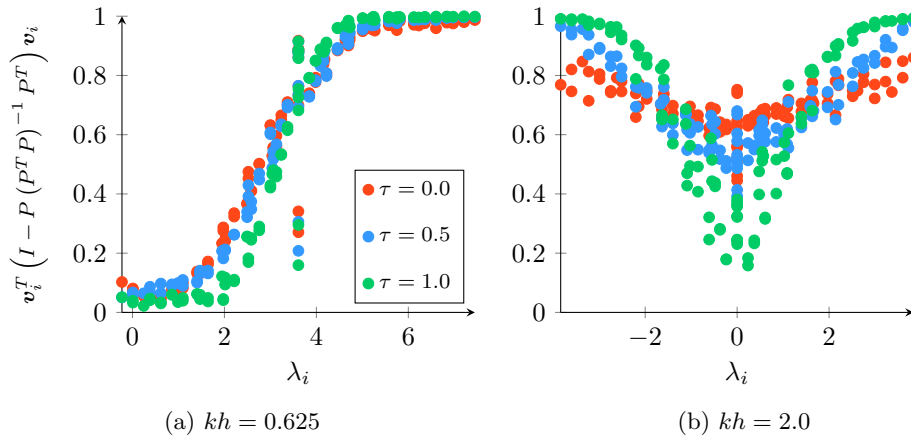


Fig. 3.1: Error of interpolation with respect to the shift and sparsity

399

400 **4. Alteration of the coarse correction in the indefinite case.** While both
 401 smoothers and interpolation operators are now designed to face two inconvenient prop-
 402 erties of the Helmholtz equation, signed eigenvalues and oscillatory near-kernel space,
 403 the effectiveness of the classical coarse correction is not guaranteed in an indefinite
 404 context. Worse still, the classical coarse correction can amplify the error associated
 405 with small eigenvectors although \hat{P} has good approximation properties, leading to
 406 a divergence of the method. Before discussing an alternative coarse correction, let
 407 us highlight how the matrix indefiniteness can corrupt the classical coarse correction
 with a simple illustration. The Figure 4.1 plots the smallest eigenvector of a two-

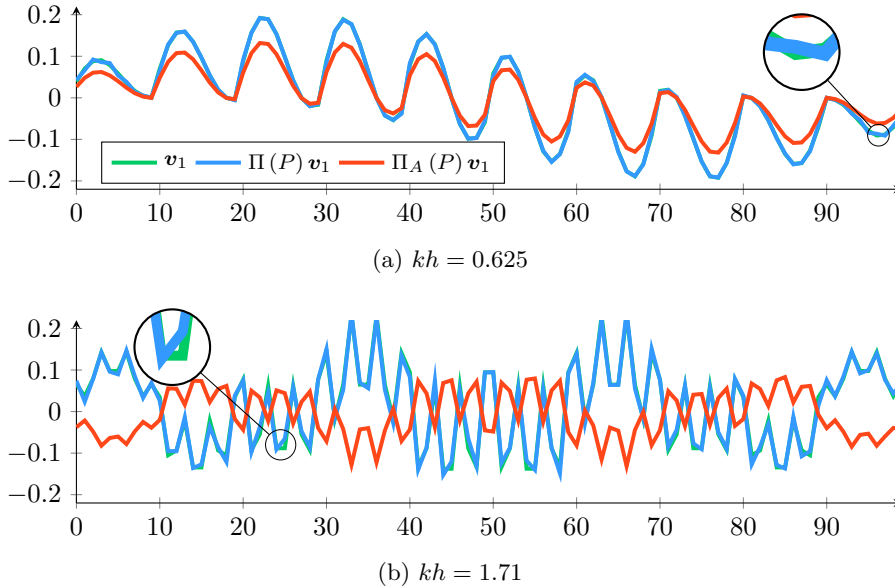


Fig. 4.1: Smallest eigenvector \mathbf{v}_1 (2D Shifted Laplacian) (blue) vs. its l_2 -projection $\Pi(P)\mathbf{v}_1$ (green) vs. its coarse correction $\Pi_A(P)\mathbf{v}_1$ (red), for two different shifts

408 dimensional shifted Laplacian matrix (1.6) in blue for two different shifts. The shift
 409 of 4.1b is greater than the shift of 4.1a. As expected, the higher the shift, the more
 410 oscillatory the problem. In red are plotted the results of the coarse correction $\Pi_A(P)$
 411 defined in Equation (1.1) when applied to the blue eigenvectors. In this example, the
 412 coarse correction is implemented with the reduction-based interpolation operator in-
 413 troduced in Section 3. Additionally, the green curves represent the best representation
 414 of both eigenvectors in the interpolation range by way of the l_2 -projection operator

$$416 \quad (4.1) \quad \Pi(P) := P(P^T P)^{-1} P^T$$

417 First, note that the blue and green curves align almost perfectly in both sub-figures,
 418 which means that the interpolation range introduced in Section 3 offers a good ap-
 419 proximation to the potentially oscillatory smallest eigenvector. In both cases, \hat{P} has
 420 good approximation properties. In Figure 4.1a, where the problem is discretized with
 421 10 points per wavelength, the red coarse correction vector is relatively close to the
 422 blue eigenvector. The slight difference between both is only a matter of amplitude.
 423 In contrast, while the oscillations of the coarse correction vector illustrated in Figure
 424 4.1b are synchronized with the oscillations of the smallest eigenvector, its direction

425 is reversed. In that case, while the interpolation range is almost perfect, the error of
 426 the smallest eigenvector is not reduced by the coarse correction, but amplified.
 427 At this stage, let us define a concept of pollution to better understand how the matrix
 428 indefiniteness can corrupt the coarse correction.

429 **THEOREM 4.1.** *Let A be an $n \times n$ matrix, and V its orthonormal set of eigen-*
 430 *vectors, each associated with the corresponding element of the diagonal eigenvalue*
 431 *matrix Λ . Also, let P be an $n \times n_c$ interpolation operator. Assuming $V_c^T P$ is non-*
 432 *singular, we write the linear decomposition of the post-scaled interpolation operator as*
 433 *$P(V_c^T P)^{-1} = VK$, where K is the following $n \times n_c$ matrix of coefficients*

$$434 \quad (4.2) \quad K := V^T P(V_c^T P)^{-1} = \begin{bmatrix} I_c \\ K_f \end{bmatrix}.$$

435 *The block I_c corresponds to the identity matrix of size $n_c \times n_c$, and the block K_f is*
 436 *a $n_f \times n_c$ matrix defined by $K_f := V_f^T P(V_c^T P)^{-1}$. The interpolation error of the*
 437 *eigenvector \mathbf{v}_i of V_c is given by*

$$438 \quad (4.3) \quad \mathbf{v}_i^T (I - \Pi(P)) \mathbf{v}_i = 1 - \left[(I_c + K_f^T K_f)^{-1} \right]_{i,i},$$

439 *where $[\cdot]_{j,k}$ denotes the entry (j, k) of the bracketed matrix.*

440 *Proof.* First, note that post-multiplying P by any non-singular matrix M_c of size
 441 $n_c \times n_c$ does not change the l_2 -projection (4.1)

$$442 \quad (PM_c)((PM_c)^T(PM_c))^{-1}(PM_c)^T = PM_c M_c^{-1} (P^T P)^{-1} M_c^{-T} M_c^T P^T \\ 443 \quad (4.4) \quad = P(P^T P)^{-1} P^T = \Pi(P).$$

445 In particular for $M_c = (V_c^T P)^{-1}$, then $PM_c = P(V_c^T P)^{-1} = VK$ implies that

$$446 \quad I - \Pi(P) = I - (VK)((VK)^T(VK))^{-1}(VK)^T \\ 447 \quad (4.5) \quad = I - VK(K^T K)^{-1} K^T V^T.$$

449 For any eigenvector \mathbf{v}_i of A , let $\mathbf{e}_i := V^T \mathbf{v}_i$ be the canonical unit vector with a one at
 450 the i^{th} position and zero elsewhere. Assuming $\mathbf{v}_i \in V_c$ ($i \leq n_c$), the vector $\mathbf{c}_i := K^T \mathbf{e}_i$
 451 of size n_c is also a unit vector with a one at the i^{th} position. Consequently,

$$452 \quad \mathbf{v}_i^T (I - \Pi(P)) \mathbf{v}_i = \mathbf{v}_i^T V (I - K(K^T K)^{-1} K^T) V^T \mathbf{v}_i \\ 453 \quad = \mathbf{e}_i^T (I - K(K^T K)^{-1} K^T) \mathbf{e}_i \\ 454 \quad (4.6) \quad = 1 - \mathbf{c}_i^T (K^T K)^{-1} \mathbf{c}_i = 1 - \left[(I_c + K_f^T K_f)^{-1} \right]_{i,i}. \quad \square$$

456 Since the l_2 -projection is unchanged by post-multiplication of P , we assume for what
 457 follows that K has the form (4.2). The block K_f designates what we call ‘‘pollution’’.
 458 This block of pollution causes the slight difference between an eigenvector \mathbf{v}_i of V_c
 459 and its best representation in the range of P . When a column of K_f is null, the
 460 interpolation error of the associated eigenvector equals zero, such that blue and green
 461 curves align perfectly. In practice however, this property is unlikely to be satisfied
 462 for Helmholtz, because P should be sparse for cost considerations and the smallest
 463 eigenvectors are usually unknown. Moreover, the near-kernel space of the Helmholtz
 464 equation is oscillatory. This makes the construction of good interpolation rules more
 465 difficult, and tends to pollute the interpolation range. While the pollution decreases

466 the convergence speed of multigrid methods for SPD problems, we demonstrate that
 467 it can corrupt the coarse correction and make the method diverge in the indefinite
 468 case, as illustrated by the reversed red vector of Figure 4.1(b).

469

470 To study the effectiveness of the coarse correction, consider the contraction of the
 471 n_c small eigenvectors V_c , assuming the n_f large eigenvectors V_f are damped by the
 472 smoother.

473 **THEOREM 4.2.** *Define A and P as in the setting of Theorem 4.1. Also, let the*
 474 *matrix K be defined as in (4.2). The contraction of an eigenvector \mathbf{v}_i of V_c after the*
 475 *coarse correction is given by*

$$476 \quad (4.7) \quad \mathbf{v}_i^T E \mathbf{v}_i = 1 - \lambda_i \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i}.$$

477

478 *Proof.* By the same reasoning of the proof for Theorem 4.1, we note that post-
 479 multiplying P by any non-singular matrix M_c of size $n_c \times n_c$ does not change the
 480 coarse correction

$$481 \quad (4.8) \quad (PM_c)((PM_c)^T A (PM_c))^{-1} (PM_c)^T = P(P^T A P)^{-1} P^T$$

482 For $PM_c = P(V_c^T P)^{-1} = VK$, we have

$$483 \quad (4.9) \quad E = I - (VK)((VK)^T A (VK))^{-1} (VK)^T A = V(I - K(K^T \Lambda K)^{-1} K^T \Lambda) V^T.$$

485 Define the euclidean basis vectors \mathbf{e}_i and \mathbf{c}_i as in the proof of Theorem 4.1. Subse-
 486 quently, the contraction of $\mathbf{v}_i \in V_c$ is

$$487 \quad \mathbf{v}_i^T E \mathbf{v}_i = \mathbf{v}_i^T V(I - K(K^T \Lambda K)^{-1} K^T \Lambda) V^T \mathbf{v}_i \\
 488 \quad = \mathbf{e}_i^T (I - K(K^T \Lambda K)^{-1} K^T \Lambda) \mathbf{e}_i \\
 489 \quad (4.10) \quad = 1 - \lambda_i \mathbf{c}_i^T (K^T \Lambda K)^{-1} \mathbf{c}_i = 1 - \lambda_i \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i}. \quad \square$$

491 Theorem 4.2 shows that the damping factors rely on a combination of the small ei-
 492 genvectors Λ_c plus the large eigenvectors Λ_f , such that the mix is given by the entries
 493 of the pollution K_f .

494

495 The effectiveness of the coarse correction is well-known in the SPD case. If all ei-
 496 genvectors are positives, one can remark that

$$497 \quad (4.11) \quad \forall i \leq n_c, \quad 0 \leq \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i} \leq [\Lambda_c^{-1}]_{i,i} = \lambda_i^{-1} \Rightarrow 0 \leq \mathbf{v}_i^T E \mathbf{v}_i \leq 1.$$

498 Hence, the coarse correction always operates a contraction on \mathbf{v}_i regardless the block
 499 of pollution K_f . In the indefinite case however, the property (4.11) does not hold. In
 500 fact, a necessary condition for the coarse correction to be a contraction is

$$501 \quad (4.12) \quad \forall i \leq n_c, \quad |\mathbf{v}_i^T E \mathbf{v}_i| \leq 1 \Rightarrow 0 \leq \lambda_i \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i} \leq 2.$$

502 From Equation (4.12), it follows that each diagonal entry must have the same sign as
 503 the associated eigenvalue, and be smaller than twice the inverse of the eigenvalue in
 504 magnitude. Nothing guarantee such conditions to be satisfied in the case where small

505 and large and either negative or positive eigenvalues are mixed. Especially for very
 506 small eigenvalues, the mix can easily lead to a diagonal entry of the opposite sign
 507 even though K_f is small, because its entries are weighted by the large eigenvalues
 508 Λ_f . Therefore, a good interpolation operator can still cause the coarse correction to
 509 amplify the error. For very near-zero eigenvalues, even a round-off error can eventually
 510 lead to divergence in the indefinite case. The following example better depicts how
 511 the pollution can cause divergence in the indefinite setting for a 2×2 matrix.

512 **EXAMPLE 4.3.** Let A be a 2×2 matrix, and \mathbf{v}_1 and \mathbf{v}_2 its eigenvectors respectively
 513 associated with eigenvalues $|\lambda_1| < |\lambda_2|$. Let P be an interpolation operator of size 2×1
 514 targeting the smallest eigenvector \mathbf{v}_1 , such that

515 (4.13)
$$P = \mathbf{v}_1 + \epsilon \mathbf{v}_2.$$

516 From definition (4.2), the K matrix can be derived by

517 (4.14)
$$K = V^T P (\mathbf{v}_1^T P)^{-1} = [\mathbf{v}_1, \mathbf{v}_2]^T \cdot [\mathbf{v}_1 + \epsilon \mathbf{v}_2] = \begin{bmatrix} 1 \\ \epsilon \end{bmatrix}.$$

518 From Theorem 4.2, the action of the coarse correction on \mathbf{v}_1 is given by

519 (4.15)
$$\mathbf{v}_1^T E \mathbf{v}_1 = 1 - \lambda_1 \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{1,1} = 1 - \frac{\lambda_1}{\lambda_1 + \epsilon^2 \lambda_2}$$

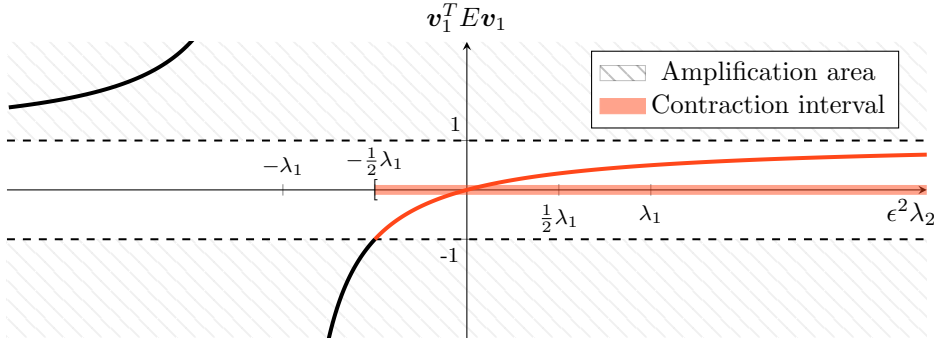


Fig. 4.2: Contraction of the coarse correction with respect to the pollution

520
 521 The figure 4.2 depicts the action of the coarse correction on \mathbf{v}_1 with respect to the
 522 pollution block $K_f^T \Lambda_f K_f = \epsilon^2 \lambda_2$. A first observation is that the coarse correction
 523 does not amplify the smallest eigenvector if eigenvalues have the same sign. If the
 524 eigenvalues are oppositely signed, then the coarse correction amplifies \mathbf{v}_1 for $\epsilon^2 \lambda_2 <$
 525 $-\lambda_1/2$. Therefore, the condition on the pollution $K_f = \epsilon$ that drives the error of
 526 interpolation is particularly difficult respective to small and large values of λ_1 and λ_2 .

527 The next theorem derives a more general condition on the spectral radius $\rho(K_f^T \Lambda_f K_f)$
 528 for the coarse correction to be a contraction of the smallest eigenvalues in the indefinite
 529 case based on the concept of pollution.

530 **THEOREM 4.4.** If A is indefinite, then

531 (4.16)
$$\rho(K_f^T \Lambda_f K_f) \leq \frac{1}{2} |\lambda_1| \Rightarrow \forall \mathbf{v}_i \in V_c, |\mathbf{v}_i^T E \mathbf{v}_i| \leq 1$$

532

533 *Proof.* Define $M_K = I_c + \Lambda_c^{-1} K_f^T \Lambda_f K_f$. From the shape of the matrix K defined
534 in Equation (4.9), we have

$$\begin{aligned} 535 \quad V_c^T E V_c &= V_c^T V (I - K (K^T \Lambda K)^{-1} K^T \Lambda) V^T V_c \\ 536 \quad &= I_c - (K^T \Lambda K)^{-1} \Lambda_c \\ 537 \quad &= I_c - (I_c + \Lambda_c^{-1} K_f^T \Lambda_f K_f)^{-1} \Lambda_c^{-1} \Lambda_c \\ 538 \quad (4.17) \quad &= I_c - M_K^{-1}. \end{aligned}$$

540 Hence, it follows that

$$541 \quad (4.18) \quad \forall \mathbf{v}_i \in V_c, \quad \mathbf{v}_i^T E \mathbf{v}_i = \mathbf{e}_i^T V_c^T E V_c \mathbf{e}_i = 1 - \mathbf{e}_i^T M_K^{-1} \mathbf{e}_i.$$

542 where \mathbf{e}_i is the i^{th} vector of the euclidean basis in \mathbb{R}^{n_c} . Therefore, $|\mathbf{v}_i^T E \mathbf{v}_i| \leq 1$ if

$$543 \quad (4.19) \quad \forall \mathbf{v}_i \in V_c, \quad -1 \leq \mathbf{v}_i^T E \mathbf{v}_i \leq 1 \Leftrightarrow 0 \leq \mathbf{e}_i^T M_K^{-1} \mathbf{e}_i \leq 2.$$

544 We begin by deriving a condition for the right bound of (4.19), and will show that
545 it also satisfies the left one. Let \mathbf{x} and \mathbf{y} be two vectors of \mathbb{R}^n linked by the relation
546 $\mathbf{x} = M_K \mathbf{y}$. The right bound is satisfied if

$$547 \quad (4.20) \quad \max_{\mathbf{x} \neq 0} \frac{\|M_K^{-1} \mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{y} \neq 0} \frac{\|\mathbf{y}\|}{\|M_K \mathbf{y}\|} = \left(\min_{\mathbf{y} \neq 0} \frac{\|M_K \mathbf{y}\|}{\|\mathbf{y}\|} \right)^{-1} \leq 2.$$

548 Therefore, the condition (4.20) is equivalent to

$$549 \quad (4.21) \quad \min_{\mathbf{y} \neq 0} \frac{\|M_K \mathbf{y}\|}{\|\mathbf{y}\|} \geq \frac{1}{2}.$$

550 Let $\sigma_i(M)$ be the i^{th} largest singular value of a given matrix M . In a same way,
551 $\lambda_i(M)$ designates the i^{th} largest eigenvalue in magnitude of M (we omit the matrix
552 between parenthesis when referring to the initial matrix A). In addition, let us recall
553 the following triangle inequality $\|\mathbf{y} + \mathbf{z}\| \geq \|\mathbf{y}\| - \|\mathbf{z}\|$, $\forall \mathbf{y}, \mathbf{z} \in \mathbb{R}^{n_c}$. Thus, we have
554 that

$$\begin{aligned} 555 \quad \min_{\mathbf{y} \neq 0} \frac{\|M_K \mathbf{y}\|}{\|\mathbf{y}\|} &= \min_{\mathbf{y} \neq 0} \frac{\|\mathbf{y} + \Lambda_c^{-1} K_f^T \Lambda_f K_f \mathbf{y}\|}{\|\mathbf{y}\|} \geq \min_{\mathbf{y} \neq 0} \left(1 - \frac{\|\Lambda_c^{-1} K_f^T \Lambda_f K_f \mathbf{y}\|}{\|\mathbf{y}\|} \right) \\ 556 \quad (4.22) \quad &= 1 - \max_{\mathbf{y} \neq 0} \frac{\|\Lambda_c^{-1} K_f^T \Lambda_f K_f \mathbf{y}\|}{\|\mathbf{y}\|} \\ 557 \quad &= 1 - \sigma_{n_c}(\Lambda_c^{-1} K_f^T \Lambda_f K_f). \end{aligned}$$

559 It follows that the condition (4.21) is satisfied if $\sigma_{n_c}(\Lambda_c^{-1} K_f^T \Lambda_f K_f) \leq \frac{1}{2}$. Finally,
560 since $\sigma_{n_c}(\Lambda_c^{-1} K_f^T \Lambda_f K_f) \leq \sigma_{n_c}(K_f^T \Lambda_f K_f) / \sigma_1$ and the singular values coincide with
561 eigenvalues in magnitude because both Λ_c and $K_f^T \Lambda_f K_f$ are hermitian, the right
562 bound of (4.19) is satisfied if

$$563 \quad (4.23) \quad |\lambda_{n_c}(K_f^T \Lambda_f K_f)| = \rho(K_f^T \Lambda_f K_f) \leq \frac{1}{2} |\lambda_1|.$$

564 We now address the left bound of (4.19) assuming the condition (4.23) holds. Our
565 goal is to prove that all diagonal entries of M_K^{-1} are positive. In that direction, let
566 $F(M)$ be the field of values of a given matrix M of size n_c such that

$$567 \quad (4.24) \quad F(M) := \{\mathbf{x}^* M \mathbf{x} \mid \forall \mathbf{x} \in \mathbb{C}^{n_c}, \mathbf{x}^* \mathbf{x} = 1\}.$$

568 If M is hermitian, one can show that (e.g. [15, chapter 4])

$$569 \quad (4.25) \quad \min_{\mathbf{x}^* \mathbf{x}=1} \mathbf{x}^* M \mathbf{x} = \lambda_{\min}(M) \text{ and } \max_{\mathbf{x}^* \mathbf{x}=1} \mathbf{x}^* M \mathbf{x} = \lambda_{\max}(M).$$

571 Accordingly, let $F(\Lambda_c)$ and $F(K_f^T \Lambda_f K_f)$ be the field of values of Λ_c and $K_f^T \Lambda_f K_f$
572 respectively. Since A is non-singular, then $0 \notin F(\Lambda_c)$. Therefore, the spectrum of
573 $\Lambda_c^{-1} K_f^T \Lambda_f K_f$ is included as follows (e.g. [14, chapter 1])

$$574 \quad (4.26) \quad \forall j \leq n_c, \lambda_j(\Lambda_c^{-1} K_f^T \Lambda_f K_f) \in F(K_f^T \Lambda_f K_f) / F(\Lambda_c).$$

575 The set ratio in (4.26) has the usual algebraic interpretation such that

$$576 \quad (4.27) \quad \forall \alpha \in \frac{F(K_f^T \Lambda_f K_f)}{F(\Lambda_c)}, -\frac{\max_{\mathbf{x}^* \mathbf{x}=1} |\mathbf{x}^* K_f^T \Lambda_f K_f \mathbf{x}|}{\min_{\mathbf{x}^* \mathbf{x}=1} |\mathbf{x}^* \Lambda_c \mathbf{x}|} \leq \alpha \leq \frac{\max_{\mathbf{x}^* \mathbf{x}=1} |\mathbf{x}^* K_f^T \Lambda_f K_f \mathbf{x}|}{\min_{\mathbf{x}^* \mathbf{x}=1} |\mathbf{x}^* \Lambda_c \mathbf{x}|}.$$

577 Furthermore, matrices Λ_c and $K_f^T \Lambda_f K_f$ are hermitian so the property (4.25) holds
578 for both of them. Because the spectrum belongs to the set ratio as in (4.26), we have

$$579 \quad (4.28) \quad -|\lambda_1|^{-1} \cdot |\lambda_{n_c}(K_f^T \Lambda_f K_f)| \leq \lambda_j(\Lambda_c^{-1} K_f^T \Lambda_f K_f) \leq |\lambda_{n_c}(K_f^T \Lambda_f K_f)| \cdot |\lambda_1|^{-1}.$$

580 Therefore, assuming the condition (4.23) is satisfied, it follows

$$581 \quad (4.29) \quad \lambda_j(\Lambda_c^{-1} K_f^T \Lambda_f K_f) \geq -|\lambda_{n_c}(K_f^T \Lambda_f K_f)| \times |\lambda_1|^{-1} \geq -\frac{1}{2}.$$

582 Adding one to each member of the inequality (4.29) finally gives

$$583 \quad (4.30) \quad \lambda_j(M_K) = \lambda_j(I + \Lambda_c^{-1} K_f^T \Lambda_f K_f) \geq \frac{1}{2}$$

584 Hence, the condition (4.23) implies that all eigenvalues of M_K are positive. Sub-
585 sequently, $\det(M_K) > 0$. The adjugate formula for the inverse of M_K shows that
586 diagonal entries are positive if the determinant of principal sub-matrices are also posi-
587 tive. Denote by $[\cdot]_{\Omega_{-i}}$ the principal sub-matrix obtained by deleting the i^{th} row and
588 column of a matrix. Since Λ_c is diagonal, one can show that

$$589 \quad (4.31) \quad [\Lambda_c^{-1} K_f^T \Lambda_f K_f]_{\Omega_{-i}} = [\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}}.$$

590 As in Equation (4.26), the spectrum is included such that

$$591 \quad \forall j \leq n_c - 1, \lambda_j([\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}}) \in F([K_f^T \Lambda_f K_f]_{\Omega_{-i}}) / F([\Lambda_c]_{\Omega_{-i}}),$$

592 and therefore the following bound holds

$$593 \quad (4.32) \quad \lambda_j([\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}}) \geq -|\lambda_{n_c-1}([K_f^T \Lambda_f K_f]_{\Omega_{-i}})| \times |\lambda_1|^{-1}.$$

594 The matrix $K_f^T \Lambda_f K_f$ being hermitian, Cauchy's interlace theorem states that

$$595 \quad (4.33) \quad \lambda_j(K_f^T \Lambda_f K_f) \leq \lambda_j([K_f^T \Lambda_f K_f]_{\Omega_{-i}}) \leq \lambda_{j+1}(K_f^T \Lambda_f K_f), \quad j = 1, \dots, n_c - 1.$$

596 As a consequence, and from the inequality (4.29), we have

$$597 \quad (4.34) \quad \lambda_j([\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}}) \geq -|\lambda_{n_c}(K_f^T \Lambda_f K_f)| \times |\lambda_1|^{-1} \geq -\frac{1}{2}.$$

598 Hence, eigenvalues of principal sub-matrices also satisfy

$$599 \quad (4.35) \quad \lambda_j \left([M_K]_{\Omega_{-i}} \right) = \lambda_j \left(I_{n_c-1} + [\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) \geq \frac{1}{2}.$$

600 Because eigenvalues of the principal sub-matrices are positive, so are the determinants.

601 From the adjugate formula of M_K^{-1} , it follows that

$$602 \quad (4.36) \quad \mathbf{e}_i^T M_K^{-1} \mathbf{e}_i = [M_K^{-1}]_{i,i} = \frac{\det \left([M_K]_{\Omega_{-i}} \right)}{\det \left(M_K \right)} \geq 0, \quad i = 1, \dots, n_c$$

603 As a consequence, both left and right bounds of (4.19) are satisfied. Finally,

$$604 \quad (4.37) \quad \rho \left(K_f^T \Lambda_f K_f \right) \leq \frac{1}{2} |\lambda_1| \Rightarrow \forall \mathbf{v}_i \in V_c, \quad |\mathbf{v}_i^T E \mathbf{v}_i| \leq 1 \quad \square$$

605 The condition provided by Theorem 4.4 is that the spectral radius of the block
 606 $K_f^T \Lambda_f K_f$ should not exceed half of the smallest eigenvalue in magnitude. No as-
 607 sumption can be made on the sign of eigenvalues in the indefinite case, so that the
 608 condition prevents the coarse correction from amplifying the error in the case where
 609 eigenvalues are oppositely signed. Applied to the previous example 4.2, Theorem 4.4
 610 states that $|\epsilon^2 \lambda_2| < |\lambda_1|/2$. That said, the condition is extremely strict and probably
 611 impossible to satisfy in practice for very small eigenvalues. In a practical method, the
 612 block K_f will never be sufficiently small for solving all types of indefinite problems
 613 because of a potentially near-zero eigenvalue. As illustrated by Figure 4.1, a good
 614 interpolation operator with small K_f can still cause divergence although it satisfies
 615 good approximation properties. The classical coarse correction appears hopeless for
 616 indefinite problems.

617 **5. Alternative coarse correction for indefinite problems.** As discussed in
 618 the previous section, the classical coarse correction is not equivalent to a minimization
 619 problem in the indefinite case, and improving P will never be enough to remedy
 620 this loss of equivalence. Moreover, because the interpolation operator developed in
 621 Section 3 targets the smallest eigenvectors of each level, every coarser matrix is more
 622 indefinite than its fine parent. Then, as the number of coarse levels increases, the
 623 balance between negative and positive eigenvalues reaches an equilibrium, and makes
 624 the effectiveness of the classical coarse correction difficult to predict. Nevertheless,
 625 Figure 4.1 shows that the interpolation operator has good approximation properties
 626 for the oscillatory near-kernel space. In particular, the Figure 4.1b suggests that only
 627 the direction of the coarse correction vector has to be changed; the shape is correct.
 628 Hence, a coarse correction that amplifies or flips the smallest eigenvectors can still
 629 provide pertinent information for solving the system. In this section, we propose to
 630 minimize the approximation error in a proper norm for indefinite problems and within
 631 a space composed of vectors returned by the classical coarse correction. Moreover, to
 632 decrease the eigenvector pollution, each coarse correction vector is smoothed by the
 633 polynomial smoother of Section 2.

634 **5.1. Notations and general considerations on GMRES.** The *Generalized*
 635 *Minimal RESidual* (GMRES) method [25] approximates the solution in a Krylov
 636 subspace by minimizing the residual in the Euclidean norm. The method can solve
 637 any class of matrix system since the norm is valid independent of the context, which
 638 is of particular interest for the indefinite case. Let us first define some notation before

639 introducing the alternative coarse correction. Let W_p be the $n \times p$ rectangular matrix
640 containing the p orthonormalized Krylov vectors such that

$$641 \quad (5.1) \quad \text{range}(W_p) = \text{span} \{ \mathbf{b}, A\mathbf{b}, A^2\mathbf{b}, \dots, A^{p-1}\mathbf{b} \}.$$

642 Each column of W_p is orthonormalized following a Gram-Schmidt process. The co-
643 efficients of the orthonormalization are stored in the rectangular Hessenberg matrix
644 \bar{H}_p of size $p+1 \times p$. The square matrix H_p is of size $p \times p$ and obtained from \bar{H}_p by
645 deleting its last row. Both matrices W_p and H_p are linked by

$$646 \quad (5.2) \quad AW_p = W_{p+1}\bar{H}_p \text{ and } W_p^T AW_p = H_p,$$

647 which leads to the following equality

$$648 \quad (5.3) \quad \min_{\tilde{\mathbf{x}} \in \text{range}(W_p)} \|\mathbf{b} - A\tilde{\mathbf{x}}\|_2 = \min_{\boldsymbol{\rho}_p \in \mathbb{C}^p} \|\mathbf{b} - AW_p\boldsymbol{\rho}_p\|_2 = \min_{\boldsymbol{\rho}_p \in \mathbb{C}^p} \|W_p^T \mathbf{b} - H_p\boldsymbol{\rho}_p\|_2$$

649 In practice, GMRES takes advantage of the convenient Hessenberg shape of \bar{H}_p to
650 construct an upper triangular matrix by applying Given's rotations. The minimization
651 of the residual then relies on a backward substitution. The relation (5.2) can be
652 generalized [8] to any arbitrary subspace $W_p = [\mathbf{w}_1, \dots, \mathbf{w}_p]$ such that

$$653 \quad (5.4) \quad \arg \min_{\tilde{\mathbf{x}} \in \text{range}(W_p)} \|\mathbf{b} - A\tilde{\mathbf{x}}\|_2 = W_p H_p^{-1} Z_p^T \mathbf{b} \quad \text{with} \quad AW_p = Z_p H_p$$

654 and where Z_p denotes the orthonormalized basis of AW_p . Note that the Arnoldi
655 relation (5.4) does not define any particular recurrence relation since W_p is arbitrary
656 and not necessarily designed by successive matrix vector products. In addition, the
657 only matrix that needs to be orthonormal in the generalized setting is Z_p .

658 **5.2. Minimization within a space of coarse correction vectors.** As men-
659 tioned in the introduction of this section, the interpolation operator has good ap-
660 proximation properties for the oscillatory near-kernel space. Even though the small
661 eigenvectors that constitute each coarse correction vector are likely to be oriented in
662 the wrong direction or amplified because of the pollution effect introduced in Section
663 4, they still provide useful information about the near-kernel space. For ease of dis-
664 cussion, we present this idea with a two-level method. The multi-level case will be
665 presented in the next section with numerical experiments.

666

667 Let W_i be the set of coarse correction vectors of the i^{th} iteration linked by the Arnoldi
668 relation (5.4) with its orthonormal counterpart Z_i . Accordingly, let $\mathbf{w}_j \in W_i$ and
669 $\mathbf{z}_j \in Z_i$ denote the j^{th} vectors of the set W_i and Z_i respectively. At each iteration i ,
670 the classical coarse correction returns a new coarse correction vector that is smoothed
671 by the Chebyshev polynomial smoother presented in Section 2. This new smoothed
672 coarse correction vector is therefore added to the previous set such that

$$673 \quad (5.5) \quad W_i = [W_{i-1}, \mathbf{w}_i] \text{ with } \mathbf{w}_i = q_{m+1}^\nu(A^2)\Pi_A(P)\mathbf{r}^{(i)},$$

674 where $\mathbf{r}^{(i)}$ designates the residual at the i^{th} iteration. From the Arnoldi relation (5.4),
675 we have

$$676 \quad (5.6) \quad H_i = Z_i^T AW_i = H_i^{-T} W_i^T A^T AW_i, \quad Z_i = AW_i H_i^{-1}.$$

677 Hence, solving the minimization problem (5.4) is equivalent to solving the normal
678 equations within the subspace spanned by W_i

$$\begin{aligned} 679 \quad W_i H_i^{-1} Z_i^T A &= W_i (W_i^T A^T A W_i)^{-1} H_i^T Z_i^T A \\ 680 \quad (5.7) \quad &= W_i (W_i^T A^T A W_i)^{-1} W_i^T A^T A = \Pi_{A^T A}(W_i). \end{aligned}$$

682 The concept of pollution also drives the convergence in the alternative setting. Sec-
683 tion 4 demonstrated that the block K_f pollutes the range of P and therefore impacts
684 the classical coarse correction. Because the minimization space W_i is generated with
685 the classical coarse correction by way of Equation (5.5), the block of pollution still
686 impacts the contraction of the small eigenvectors. Resorting to the Euclidean norm
687 in (5.4) prevents the divergence, but it also squares the eigenvalues of the initial prob-
688 lem because of the equivalence with an $A^T A$ -orthogonal projection. This naturally
689 increases the gap between small and large eigenvalues, and therefore decreases the
690 contraction of the smallest over the largest.

691

692 Smoothing the classical coarse correction vectors by way of the polynomial $q_{m+1}^\nu(A^2)$
693 compensates for this effect by reducing the prevalence of large eigenvectors in the
694 minimization space. This idea of damping the large eigenvalues to reveal the smaller
695 ones is also used to generate a relevant set of test vectors for the construction of
696 the least-squares minimization operator introduced in Section 3. Once the coarse
697 correction vector is smoothed and included in W_i , the set Z_i is extended as follows

$$698 \quad (5.8) \quad Z_i = [Z_{i-1}, \mathbf{z}_i] \text{ with } \mathbf{z}_i = \frac{1}{h_{i,i}} \left(A \mathbf{w}_i - \sum_{j=1}^{i-1} h_{j,i} \cdot \mathbf{z}_j \right),$$

699 where coefficients $h_{j,i}$ result from the orthogonalization process of the new vector
700 $A \mathbf{w}_i$. Those coefficients are stored in the squared upper triangular matrix

$$701 \quad (5.9) \quad H_i = \begin{bmatrix} & & & h_{1,p} \\ & H_{i-1} & & \vdots \\ & & & h_{i-1,i} \\ 0 & \dots & 0 & h_{i,i} \end{bmatrix} \text{ with } h_{j,i} = \begin{cases} \langle \mathbf{z}_j, \mathbf{z}_i \rangle & \text{if } j < i \\ \|\mathbf{z}_i\|_2 & \text{if } j = i \end{cases}.$$

702 The algorithm 5.1 presents the alternative two-level cycle, and can be compared with
703 the classic one in Algorithm 1.1.

704 **EXAMPLE 5.1.** Consider again Example 4.3, where A is a 2×2 matrix, with
705 \mathbf{v}_1 and \mathbf{v}_2 its eigenvectors respectively associated with eigenvalues $|\lambda_1| < |\lambda_2|$. The
706 interpolation operator P targets \mathbf{v}_1 as defined by (4.13). Let W_1 be the minimization
707 space of dimension 1 constructed following (5.5) such that

$$708 \quad (5.10) \quad W_1 = q_{m+1}(A^2) \Pi_A(P) \mathbf{v}_1 = \frac{\lambda_1^2}{\lambda_1 + \epsilon^2 \lambda_2} (q_{m+1}(\lambda_1^2) \mathbf{v}_1 + q_{m+1}(\lambda_2^2) \epsilon \mathbf{v}_2).$$

709 Furthermore, define E_{W_1} to be the error propagation matrix of the alternative coarse
710 correction. One can show that

$$711 \quad (5.11) \quad \mathbf{v}_1^T E_{W_1} \mathbf{v}_1 = \mathbf{v}_1^T (I - \Pi_{A^T A}(W_1)) \mathbf{v}_1 = 1 - \frac{q_{m+1}^2(\lambda_1^2) \lambda_1^2}{q_{m+1}^2(\lambda_1^2) \lambda_1^2 + q_{m+1}^2(\lambda_2^2) \epsilon^2 \lambda_2^2}.$$

Algorithm 5.1 Two-level cycle with the alternative coarse correction

Inputs : \mathbf{b} right-hand side, $\tilde{\mathbf{x}}$ approximation of \mathbf{x} , $\mathbf{r} = \mathbf{b} - A\tilde{\mathbf{u}}$ residual
 M smoother, P interpolation operator

for $j = 1, \nu$ **do**
 $\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} + p(A^2)\mathbf{r}$
 $\mathbf{r} \leftarrow \mathbf{b} - A\tilde{\mathbf{x}}$
end for

$\mathbf{r}_C \leftarrow P^T \mathbf{r}$
 $\mathbf{e}_C \leftarrow \text{Solve}(P^T A P, \mathbf{r}_C)$
 $\mathbf{w} \leftarrow q_{m+1}^\nu(A^2) P \mathbf{e}_C$
 $\tilde{\mathbf{w}}, H_i \leftarrow \text{Orthonormalize}(\mathbf{w}, Z_{i-1})$
 $W_i, Z_i \leftarrow [W_{i-1}, \mathbf{w}], [Z_{i-1}, \tilde{\mathbf{w}}]$

for $j = 1, \nu$ **do**
 $\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} + p(A^2)\mathbf{r}$
 $\mathbf{r} \leftarrow \mathbf{b} - A\tilde{\mathbf{x}}$
end for

$\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} + W_i H_i^{-1} Z_i^T \mathbf{r}$
 $\mathbf{r} \leftarrow \mathbf{b} - A\tilde{\mathbf{x}}$

Output : $\tilde{\mathbf{x}}$ approximation of \mathbf{x} at the end of the cycle

712 To simplify the discussion, let us assume that the smallest eigenvector is preserved
713 by the smoother, such that $q_{m+1}(\lambda_1^2) = 1$. Figure 5.1 illustrates the contraction of \mathbf{v}_1
714 after applying the alternative coarse correction with respect to the pollution and the
715 polynomial. As expected, the smoother increases the contraction and counter balances
716 the squared large eigenvalue λ_2 that weights the pollution $K_f = \epsilon$ when minimizing in
the Euclidean norm.

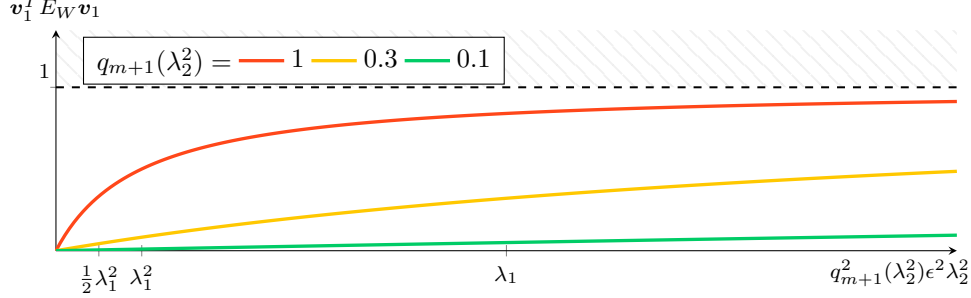


Fig. 5.1: Illustration of the contraction of a small eigenvector with respect to the pollution and the polynomial smoother

717

718 **6. Numerical Experiments.** In the following numerical experiments, the inter-
719 val of the Chebyshev polynomial smoother is determined following the spectral
720 density approximation method presented in Section 2.2. The number n_ν of coeffi-
721 cients μ_k in the moment matching procedure is fixed to 15, and n_{vec} fixed to 5. The
722 degree m of the polynomial is 3. Regarding the construction of the interpolation
723 operator, the number of smoothed test vectors is fixed to 15. Last, the number of
724 interpolation points in the least square minimization strategy used to construct the
725 coarse grid selection operator \hat{R}^T never exceeds 4 (i.e., $\max_{i \in \mathcal{F}} \{\text{Card}(C_i)\} = 4$).

726

6.1. Two-level experiment on the Two Dimensional Shifted Laplacian.

727

Let us first apply this new multigrid setting to the two-dimensional shifted laplacian
728 problem associated with the stencil matrix (1.6). The size of the shifted laplacian

729 matrix is fixed to $n = 100$. Figure 6.1 depicts the number of iterations with respect
 730 to the shift kh using either the classical or the alternative coarse correction. Recall
 731 that the matrix is the most indefinite (exact balance between negative and positive
 732 eigenvalues) when $kh = 2$, and that the near-kernel space becomes more oscillatory
 733 as kh increases. The number of iterations are also presented with respect to the
 734 percentage τ that governs the number of selected columns of \hat{S} in the approximation
 735 of ideal interpolation. The resulting operator complexity defined by $\phi := \frac{\sum_l \text{nnz}(A_l)}{\text{nnz}(A_0)}$
 736 for different values of τ is provided by Table 6.1. Last, the tolerance of the relative
 737 residual norm is set to 10^{-6} , and the maximal number of iterations is fixed to 100.
 738 Peak values of the standard multigrid setting on the left column denote divergence,
 739 whereas they stand for slow convergence in the alternative setting plotted on the right
 740 column.

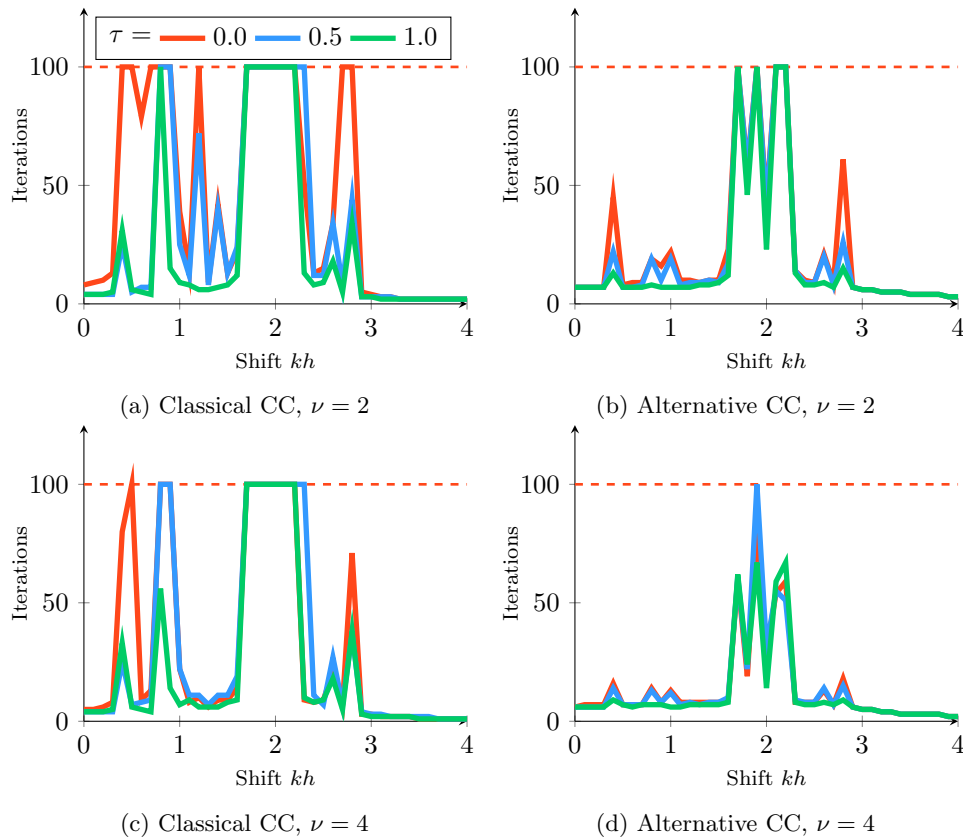


Fig. 6.1: Number of iterations of two-level methods with respect to kh and τ

τ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
ϕ	1.81	3.00	3.47	3.69	3.93	4.16	4.35	4.55	4.73	4.87	5.15

Table 6.1: Operators complexity of the two-level method with respect to τ

741 Both left figures 6.1c and 6.1a correspond to a two-level method built on the classical
 742 coarse correction respectively for $\nu = 2$ and $\nu = 4$. Whereas increasing the number
 743 of selected columns in \hat{S} for approximating the ideal interpolation operator by way

744 of the parameter τ generally helps the convergence, the method remains likely to
 745 diverge for the reasons explained in Section 4. Still, the best setting for the classical
 746 coarse correction is naturally $\tau = 1$ and $\nu = 4$. Certain divergence scenarios that
 747 happen for $\nu = 2$ (for instance around $kh = 0.8$) are fixed by doubling the number of
 748 smoothing iterations. Doing so improves the set of test vectors in approximating the
 749 near-kernel space, and therefore leads to a better least-squares minimization coarse
 750 variable operator that decreases the pollution K_f . It remains however impossible
 751 to derive a general setting that ensures the convergence of the standard method in
 752 all cases. Both right figures 6.1b and 6.1d represent the same experiment with the
 753 alternative coarse correction. The peaks around $kh = 2$ depict a slow convergence
 754 situation where the relative residual norm is stuck around 10^{-5} because of very near-
 755 zero eigenvalues. Except for these extremely indefinite cases, the method converges
 756 in all cases. We also remark that the divergence of the standard method correlates
 757 with more iterations in the alternative setting. At the cost of complexity, increasing
 758 τ or ν provides a better convergence factor.

759 6.2. Multi-level experiment on the Two Dimensional Helmholtz problem with absorbing boundary conditions.

760 The following numerical experiments depict the convergence for a two dimensional
 761 Helmholtz problem using absorbing boundary conditions and with a discretization
 762 coefficient set to $kh = 0.625$ (i.e. 10 points per wavelength, where k corresponds
 763 to the wavenumber). Therefore, the discretization matrix is indefinite, complex and
 764 non-hermitian, and grows with k . As a consequence, the restriction operation is made
 765 through the transpose conjugate \hat{P}^* . Moreover, the squared matrix in the polynomial
 766 setting is replaced by A^*A . Also note that these numerical experiments result from
 767 the alternative coarse correction only, and that Z^T is replaced by Z^* in (5.4). The
 768 first benchmark illustrated in Figure 6.2 explores the convergence of the method by
 769 fixing the number of selected column of \hat{S} to the maximum (i.e., $\tau = 1$). Each curve
 770 corresponds to a method following its number of levels. The y -axis corresponds to
 771 the number of iterations, while the wavenumber varies along the x -axis. The number
 772 of iterations is constant until the fourth level. The number of iterations of both the
 773 five-level and six-level methods increase with the wavenumber.

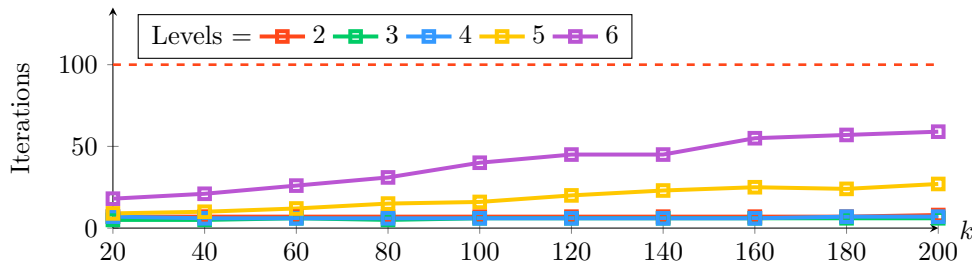


Fig. 6.2: Number of iterations following the wavenumber k , $\nu = 2$, $\tau = 1$

774 While setting $\tau = 1$ enables the method to converge almost constantly up to five
 775 levels, the operator complexity is too high for practical implementation. Therefore,
 776 the second benchmark explores the number of iterations of a two-level method with
 777 respect to the parameter τ . Figure 6.3 shows that the plain least-squares minimization
 778 operator (i.e. $\tau = 0$) is not a suitable choice as k increases. Even though larger
 779 sub-spaces $\hat{S}X_i$ in the approximation of the ideal interpolation operator yields denser
 780

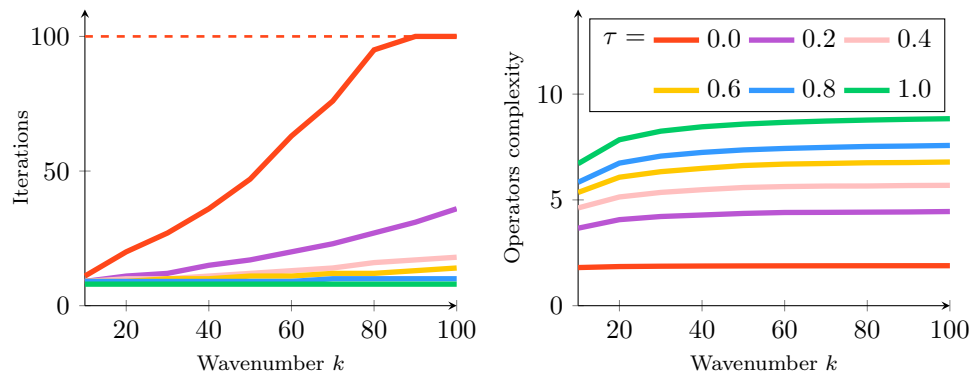


Fig. 6.3: Two-level method with alternative coarse correction - number of iterations and operators complexity with respect to k and τ , $\nu = 2$

781 matrices, the number of iterations tends toward size independence as τ grows. A trade-
 782 off between convergence and complexity may be possible depending on the problem
 783 size. More generally, Figure 6.3 reveals the important role of the ideal approximation
 784 step in the convergence. A better sparsification strategy is a topic of further research.
 785

786 **7. Conclusions.** Indefinite and oscillatory problems are difficult for multigrid
 787 methods. The negative eigenvalues require an adapted smoother, and the interpo-
 788 lation operator should capture the oscillatory near-kernel space. More importantly,
 789 the coarse correction should be adapted to the indefiniteness of the initial matrix,
 790 which does not define a norm. The normal equation polynomial smoother is designed
 791 to target a desired proportion of eigenvalues according to their amplitude, and the
 792 range of our interpolation operator offers a good approximation of the near-kernel
 793 space despite its oscillations. The alternative coarse correction space proposed in the
 794 paper minimizes the global residual in a proper norm for indefinite problems in a
 795 space approximating the set of smallest eigenvectors known to be difficult for most
 796 iterative methods. Finding a better trade-off between sparsity and accuracy of inter-
 797 polation and constructing a polynomial without resorting to normal equations will be
 798 important points in our future investigations.

799

REFERENCES

- 800 [1] M. F. ADAMS, M. BREZINA, J. J. HU, AND R. S. TUMINARO, *Parallel multigrid smoothing:*
 801 *polynomial versus gauss–seidel*, Journal of Computational Physics, 188 (2003), pp. 593–
 802 610.
 803 [2] A. H. BAKER, R. D. FALGOUT, T. V. KOLEV, AND U. M. YANG, *Multigrid smoothers for ultra-*
 804 *parallel computing*, SIAM Journal on Scientific Computing, 33 (2011), pp. 2864–2887, <https://doi.org/10.1137/100798806>, <https://doi.org/10.1137/100798806>, [https://arxiv.org/abs/](https://arxiv.org/abs/https://doi.org/10.1137/100798806)
 805 <https://doi.org/10.1137/100798806>.
 806 [3] A. BRANDT, J. BRANNICK, K. KAHL, AND I. LIVSHITS, *Bootstrap amg*, SIAM Journal of Scien-
 807 tific Computing, 33 (2011), pp. 612–632, <https://doi.org/10.1137/090752973>.
 808 [4] A. BRANDT, S. MCCORMICK, AND J. RUGE, *Algebraic multigrid (AMG) for sparse matrix*
 809 *equations*, in Sparsity and its Applications, D. J. Evans, ed., Cambridge University Press,
 810 Cambridge, 1985, pp. 257–284.
 811 [5] L. I. BRANDT A., *Wave-ray multigrid method for standing wave equations.*, ETNA. Electronic
 812 Transactions on Numerical Analysis [electronic only], 6 (1997), pp. 162–181, [http://eudml.](http://eudml.org/doc/119506)
 813 [org/doc/119506](http://eudml.org/doc/119506).
 814

- 815 [6] M. BREZINA, R. FALGOUT, S. MACLACHLAN, T. MANTEUFFEL, S. MCCORMICK, AND J. RUGE,
816 *Adaptive smoothed aggregation (α sa)*, SIAM Journal on Scientific Computing, 25 (2004),
817 pp. 1896–1920, <https://doi.org/10.1137/S1064827502418598>, <https://doi.org/10.1137/S1064827502418598>, <https://arxiv.org/abs/https://doi.org/10.1137/S1064827502418598>.
- 819 [7] W. BRIGGS, V. HENSON, AND S. MCCORMICK, *A Multigrid Tutorial, 2nd Edition*, 01 2000.
- 820 [8] O. COULAUD, L. GIRAUD, P. RAMET, AND X. VASSEUR, *Deflation and augmentation techniques*
821 *in krylov subspace methods for the solution of linear systems*, 2013, [https://arxiv.org/abs/](https://arxiv.org/abs/1303.5692)
822 [1303.5692](https://arxiv.org/abs/1303.5692).
- 823 [9] V. DWARKA AND C. VUIK, *Stand-alone multigrid for helmholtz revisited: Towards convergence*
824 *using standard components*, 2023, <https://arxiv.org/abs/2308.13476>.
- 825 [10] P. EK, M. BREZINA, AND J. MANDEL, *Convergence of algebraic multigrid based on smoothed*
826 *aggregation*, Computing, 56 (1998), <https://doi.org/10.1007/s002110000226>.
- 827 [11] O. G. ERNST AND M. J. GANDER, *Why it is difficult to solve helmholtz problems with classical*
828 *iterative methods*, (2010).
- 829 [12] R. D. FALGOUT, *An introduction to algebraic multigrid*, Computing in Science and Engineering,
830 vol. 8, no. 6, November 1, 2006, pp. 24–33, (2006), <https://www.osti.gov/biblio/897960>.
- 831 [13] R. D. FALGOUT AND P. S. VASSILEVSKI, *On generalizing the amg framework*, SIAM J. NUMER.
832 ANAL, 42 (2003), pp. 1669–1693.
- 833 [14] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press,
834 1 ed., Apr. 1991, <https://doi.org/10.1017/CBO9780511840371>, <https://www.cambridge.org/core/product/identifier/9780511840371/type/book> (accessed 2024-05-24).
- 835 [15] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 2 ed.,
836 Oct. 2012, <https://doi.org/10.1017/CBO9781139020411>, <https://www.cambridge.org/highereducation/product/9781139020411/book> (accessed 2024-05-24).
- 837 [16] J. K. KRAUS, P. S. VASSILEVSKI, AND L. T. ZIKATANOV, *Polynomial of best uniform approxima-*
838 *tion to x^{-1} and smoothing in two-level methods*, 2012, <https://arxiv.org/abs/1002.1859>.
- 840 [17] L. LIN, Y. SAAD, AND C. YANG, *Approximating spectral densities of large matrices*, SIAM Re-
841 view, 58 (2016), pp. 34–65, <https://doi.org/10.1137/130934283>, [https://doi.org/10.1137/](https://doi.org/10.1137/130934283)
842 [130934283](https://doi.org/10.1137/130934283), <https://arxiv.org/abs/https://doi.org/10.1137/130934283>.
- 843 [18] I. LIVSHITS, *A scalable multigrid method for solving indefinite helmholtz equations with constant*
844 *wave numbers*, Numerical Linear Algebra with Applications, 21 (2014), [https://doi.org/](https://doi.org/10.1002/nla.1926)
845 [10.1002/nla.1926](https://doi.org/10.1002/nla.1926).
- 846 [19] I. LIVSHITS, *Multiple galerkin adaptive algebraic multigrid algorithm for the helmholtz equa-*
847 *tions*, SIAM Journal on Scientific Computing, 37 (2015), pp. S195–S215, <https://doi.org/10.1137/140975310>, <https://doi.org/10.1137/140975310>, <https://arxiv.org/abs/https://doi.org/10.1137/140975310>.
- 848 [20] L. OLSON AND J. SCHRODER, *Smoothed aggregation for helmholtz problems*, Numerical Linear
849 Algebra with Applications, 17 (2010), pp. 361 – 386, <https://doi.org/10.1002/nla.686>.
- 850 [21] L. N. OLSON, J. B. SCHRODER, AND R. S. TUMINARO, *A general interpolation strategy*
851 *for algebraic multigrid using energy minimization*, SIAM Journal on Scientific Comput-
852 ing, 33 (2011), pp. 966–991, <https://doi.org/10.1137/100803031>, [https://doi.org/10.1137/](https://doi.org/10.1137/100803031)
853 [100803031](https://doi.org/10.1137/100803031), <https://arxiv.org/abs/https://doi.org/10.1137/100803031>.
- 854 [22] E. PAROLIN, D. HUYBRECHS, AND A. MOIOLA, *Stable approximation of helmholtz solutions*
855 *by evanescent plane waves*, 2022, <https://doi.org/10.48550/ARXIV.2202.05658>, [https://](https://arxiv.org/abs/2202.05658)
856 arxiv.org/abs/2202.05658.
- 857 [23] J. W. RUGE AND K. STÜBEN, *4. algebraic multigrid*, in Multigrid Methods, S. F. McCormick,
858 ed., Society for Industrial and Applied Mathematics, pp. 73–130, [https://doi.org/10.](https://doi.org/10.1137/1.9781611971057.ch4)
859 [1137/1.9781611971057.ch4](https://doi.org/10.1137/1.9781611971057.ch4), <http://epubs.siam.org/doi/10.1137/1.9781611971057.ch4> (ac-
860 cessed 2024-06-27).
- 861 [24] J. W. RUGE AND K. STÜBEN, *4. Algebraic Multigrid*, [https://doi.org/10.1137/1.9781611971057.](https://doi.org/10.1137/1.9781611971057.ch4)
862 [ch4](https://doi.org/10.1137/1.9781611971057.ch4), <https://epubs.siam.org/doi/abs/10.1137/1.9781611971057.ch4>, <https://arxiv.org/abs/https://epubs.siam.org/doi/pdf/10.1137/1.9781611971057.ch4>.
- 863 [25] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, [https://www-users.cs.umn.edu/~saad/](https://www-users.cs.umn.edu/~saad/IterMethBook.2ndEd.pdf)
864 [IterMethBook.2ndEd.pdf](https://www-users.cs.umn.edu/~saad/IterMethBook.2ndEd.pdf).
- 865 [26] G. STRANG, *Multigrid methods*, tech. report, MIT, 2006, [https://math.mit.edu/classes/18.086/](https://math.mit.edu/classes/18.086/2006/am63.pdf)
866 [2006/am63.pdf](https://math.mit.edu/classes/18.086/2006/am63.pdf).
- 867 [27] K. STÜBEN, *Algebraic multigrid (amg). an introduction with applications*, (1999).
- 868 [28] P. VANEK, J. MANDEL, AND M. BREZINA, *Algebraic multigrid by smoothed aggregation for*
869 *second and fourth order elliptic problems*, tech. report, USA, 1995.
- 870 [29] B. M. VANVEK PETR AND M. JAN, *Convergence of algebraic multigrid based on smoothed*
871 *aggregation*, (2001), <https://doi.org/10.1007/s211-001-8015-y>.