



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-04620991v1

HAL Id: cea-04620991

<https://cea.hal.science/cea-04620991v1>

Preprint submitted on 22 Jun 2024 (v1), last revised 28 Jun 2024 (v2)

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, Polynomial Smoother,
18 Indefinite matrix

19 **1. Introduction.** The numerical simulation of various physical phenomena leads
20 to potentially very large linear systems of equations written $A\mathbf{x} = \mathbf{b}$ in matrix form.
21 These systems can be solved directly by a convenient factorization of A , or iteratively
22 by computing and refining an approximation of the solution \mathbf{x} starting from an initial
23 guess \mathbf{x}_0 . Multigrid methods [6, 25] work iteratively and are known to be scalable
24 and quasi-optimal for solving sparse linear systems of equations for many classes of
25 problems. Each multigrid iteration combines a projection method on a coarser space
26 to capture the eigenvectors associated with the small eigenvalues, and a few iterations
27 of a smoothing method to capture the remaining eigenvectors generally associated
28 with the large eigenvalues.

29
30 To simplify the discussion in what follows, we use the term "small/large eigenvec-
31 tor" to designate an eigenvector with small/large eigenvalue. We similarly say "pos-
32 itive/negative eigenvector" when referring to the eigenvalue sign. Additionally, capi-
33 tal italic Roman letters (A, E, P) denote matrices and bold lowercase letters denote
34 vectors ($\mathbf{u}, \mathbf{v}, \mathbf{r}, \boldsymbol{\alpha}$). Other lowercase letters denote scalars (σ, λ), while capital calli-
35 graphic letters denote sets and spaces ($\mathcal{C}, \mathcal{F}, \mathcal{K}$).

36 **1.1. Multigrid methods.** While errors composed of small eigenvectors are
37 known to be more difficult to eliminate for most iterative methods, multigrid methods
38 accelerate the convergence to the solution by projecting them onto a coarser space.
39 The coarse projection of those difficult eigenvectors is repeated recursively until reach-
40 ing a small enough coarse matrix for which the factorization by a direct solver is fast.

*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-xxxxxx).

[†]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 Assuming the matrix is symmetric positive definite (SPD), the best approximation
 42 of the solution within the coarse projection space is computed by minimizing the ap-
 43 proximation error in A -norm. The core idea in multigrid methods is to make this
 44 projection practical by recursively defining smaller subspaces by way of sparse op-
 45 erators P_l , called interpolation operators. The computation of \mathbf{x} is accelerated by
 46 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
 47 l in the grid hierarchy. P_l determines the coarse projection subspace of the level l ,
 48 and transfers the information from level $l + 1$ to l . In most symmetric applications,
 49 coarse matrices are constructed following the Galerkin formula $A_{l+1} = P_l^T A_l P_l$. The
 50 two-level coarse correction operator denoted by

$$51 \quad (1.1) \quad \Pi_A(P) := P(P^T A P)^{-1} P^T A$$

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

$$54 \quad (1.2) \quad \arg \min_{\tilde{\mathbf{x}} \in \text{span}\{P\}} \|\mathbf{x} - \tilde{\mathbf{x}}\|_A = \Pi_A(P) \mathbf{r}.$$

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

$$60 \quad (1.3) \quad E = I - \Pi_A(P).$$

61 Likewise, the error propagation matrix for the smoother is

$$62 \quad (1.4) \quad E_M = I - M^{-1} A$$

63 where M^{-1} is an approximation of A^{-1} . The smoother is applied before each restric-
 64 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:        $A$  initial matrix,  $M$  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

```

65 Finding a smoother and a coarse correction that are complementary is a major concern
 66 in the design of the method. Moreover, the context in which a multigrid method
 67 is applied determines what kind of operators should be used in the method. In

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

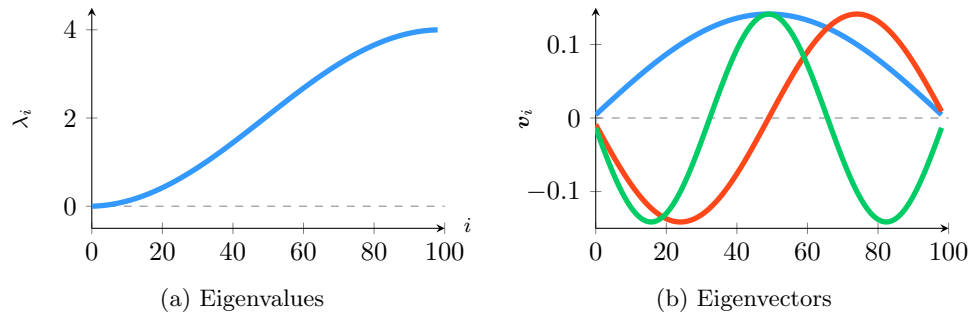


Fig. 1.1: Laplace eigenvalues and three smallest eigenvectors

79 Likewise in classic algebraic multigrid [26, 23, 11], the interpolation operators are
 80 designed to target what is called algebraically smooth components. The smoothed
 81 aggregation method [9] is particularly efficient for solving problems with an *a pri-*
 82 *ori* known near-kernel space, for instance in diffusion [28] or elasticity [27] where
 83 the target small eigenvectors are the constant vector and rigid body modes respec-
 84 tively. Those vectors are split between disjoint aggregates over the entire domain
 85 to initiate a tentative block interpolation operator. A few smoothing iterations are
 86 applied to the tentative interpolation operator to extend its pattern, but especially
 87 in order to clean the tentative interpolation range from high frequencies. Usually, a
 88 few iterations of the Jacobi relaxation method are enough, but this step of energy
 89 minimization has been generalized to Krylov methods such as the conjugate gradient
 90 [21] by enforcing sparsity constraints in the minimization space to keep a practical
 91 interpolation operator. If near-kernel space information is lacking, test vectors can
 92 still be computed algebraically, as in adaptive smoothed aggregation [5]. Furthermore,
 93 because the choice of the interpolation strategy is essential in the convergence of the
 94 method, an ideal framework maximizing the complementarity between the smoother
 95 and the coarse correction [12] has been established to guide the algorithm develop-
 96 ment. While this idealistic scenario of convergence is mostly used as a theoretical tool,
 97 some reduction-based methods enable a good approximation of the ideal interpolator
 98 given some initial coarse and fine variable splitting [19, 29].

99 **1.2. Why Helmholtz problems are difficult for multigrid.** The Helmholtz
 100 equation (1.5) involves indefinite matrices with potentially wide and oscillatory near-
 101 kernel spaces [10]. This equation is our target in this paper.

$$102 \quad (1.5) \quad (\text{Continuous Helmholtz problem}) \Leftrightarrow \begin{cases} -\Delta \mathbf{u} - k^2 \mathbf{u} = \mathbf{f} & \text{on } \Omega \\ + \text{b. c.} & \text{on } \partial\Omega \end{cases}$$

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

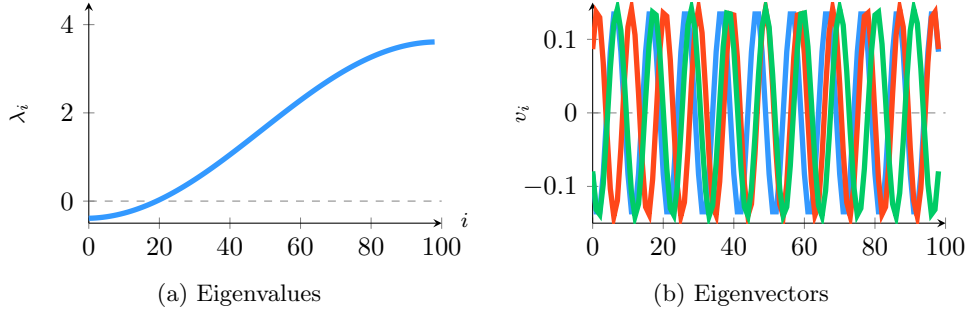


Fig. 1.2: Helmholtz eigenvalues and three smallest eigenvectors

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

117

118 For these reasons, finding a recurring process to build a scalable multilevel method
 119 is still an open question. Multiple correction [18], wave-ray [4, 17], and Complex-
 120 Shifted Laplacian [8] approaches have already been investigated to address this issue.
 121 In this paper, we present a fully algebraic approach built on ideal reduction-based
 122 ideas, and demonstrate its potential for solving the Helmholtz problem with constant
 123 iteration count independent of the wavenumber k . Certain discretization matrices re-
 124 sulting from the continuous problem (1.5) can be non-symmetric due to the boundary
 125 conditions. To center the discussion on the indefinite nature of Helmholtz, the next
 126 approaches address the symmetric indefinite shifted laplacian matrix arising from the
 127 following 5-pts stencil

$$128 \quad (1.6) \quad \hat{A} = \begin{bmatrix} & -1 & & & \\ -1 & 4 - (kh)^2 & & & \\ & & -1 & & \\ & & & -1 & \\ & & & & \end{bmatrix}.$$

129 In Section 2, we start by presenting a normal equation polynomial smoother specifi-
 130 cally designed to damp the desired proportion of largest eigenvalues independently
 131 of their signs, while interpolation rules for propagating oscillatory near-kernel infor-
 132 mation are established in Section 3. The Section 4 gives more details on why the
 133 indefiniteness can corrupt the coarse correction by introducing a concept of pollution
 134 and Section 5 exposes an alternative coarse correction to the classical one which avoids
 135 the divergence scenarios. Finally, Section 6 presents benchmarks of this new multigrid

136 method for different Helmholtz problems, with varying shift kh and wavenumber k .
137

138 Along the different approaches presented in this paper, \mathbf{v}_i denotes the i^{th} eigenvector
139 of A associated with the eigenvalue λ_i . Moreover, we always assume the eigenvalues
140 to be ordered in magnitude (i.e., $\forall i < n$, $|\lambda_i| \leq |\lambda_{i+1}|$) such that $V_c := [\mathbf{v}_1, \dots, \mathbf{v}_{n_c}]$
141 and $V_f := [\mathbf{v}_{n_c+1}, \dots, \mathbf{v}_n]$ contain the small and large eigenvector sets of size n_c and
142 n_f respectively. Naturally, the full set of eigenvectors are given by $V = [V_c, V_f]$, and
143 $n = n_c + n_f$.

144 **2. Polynomial Smoothers for Indefinite Problem.** Working with a smooth-
145 ing method whose behavior on the spectrum is *a priori* known is interesting to guar-
146 antee the effectiveness of the cycle. Here, the smoother must damp large positive
147 and negative eigenvalues, which is problematic for most standard methods. Gener-
148 ally, a polynomial method with degree greater than one can work. Krylov iterations
149 are good polynomial smoothers in the indefinite case but they minimize the global
150 residual norm regardless of the eigenvalues and are non-linear because of their right-
151 hand side dependence. A linear polynomial is more convenient for generating the
152 set of smoothed candidates vectors needed to construct the interpolation operator
153 described in Section 3.

154 **2.1. General considerations on polynomial smoothers.** One way to en-
155 sure that both positive and negative eigenvectors are damped is to consider a normal
156 equation polynomial smoother. In general, the degree m of the polynomial must be
157 greater than one to damp positive and negative eigenvectors, as the polynomial il-
158 lustrated in Figure 2.1 does. Resorting to normal equations enables the polynomial
159 to treat eigenvalues with respect to their magnitude rather than their sign, which is
160 equivalent to work with even powers of A if the matrix is hermitian, which is what
161 we assume in this section. In the future, it might be interesting to investigate more
162 general polynomials to avoid normal equations and consider odd exponents. In this
163 first approach, we use the convenient symmetry property enabled by normal equations
164 in the Chebyshev framework.

165
166 Let $p_m(A^2)$ be a polynomial of degree m that approximates A^{-2} . From Equation
167 (1.4), let $q_{m+1}(A^2)$ be the associated error propagation matrix of the polynomial
168 smoother such that

$$169 \quad (2.1) \quad q_{m+1}(A^2) := I - p_m(A^2)A^2.$$

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

$$171 \quad (2.2) \quad q_{m+1}(\lambda^2)\mathbf{v} = (1 - p_m(\lambda^2)\lambda^2)\mathbf{v}.$$

172 In multigrid methods, a good smoother eliminates the large eigenvalues that the coarse
173 correction does not capture and vice-versa. Let a and b be real scalars such that $0 <$
174 $a < b$. Assume these large squared eigenvalues are contained in the interval $[a, b]$. The
175 construction of a relevant interval will be discussed in the next. Since the polynomial
176 smoother $p_m(A^2)$ is an inverse approximate of A^{-2} , the polynomial function $p_m(x)$
177 can be constructed to approximate the function x^{-1} [15] from $m + 1$ interpolation
178 points x_i selected within the interval of large eigenvalues $[a, b]$. In particular, selecting
179 the scaled first kind Chebyshev polynomial roots as interpolation points

$$180 \quad (2.3) \quad x_i := \frac{b+b}{2} + \frac{b-a}{2} \cos\left(\frac{(2i+1)\pi}{2(m+1)}\right), \quad i = 1, \dots, m+1.$$

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

$$183 \quad (2.4) \quad x_i \in [a, b], p_m(x_i) = \frac{1}{x_i} \Leftrightarrow q_m(x_i) = 0, i = 1, \dots, m + 1.$$

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

$$186 \quad (2.5) \quad p_m(x) := \sum_{i=1}^{m+1} \frac{1}{x_i} \prod_{j=1, j \neq i}^{m+1} \frac{x - x_j}{x_i - x_j}, q_{m+1}(x) = \prod_{i=1}^{m+1} \frac{x - x_i}{-x_i}.$$

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

$$188 \quad (2.6) \quad \forall t \in [-1, 1], C_0(t) = 1, C_1(t) = t, C_{m+1}(t) = 2tC_m(t) - C_{m-1}(t).$$

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

$$191 \quad (2.7) \quad q_{m+1}(x) = \frac{C_{m+1}\left(\frac{b+a-2x}{b-a}\right)}{C_{m+1}\left(\frac{b+a}{b-a}\right)}.$$

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

200 **2.2. Constructing an appropriate target interval.** One way to determine
 201 an interval $[a, b]$ without preliminary information [2, 1] is to compute a few power
 202 iterations to determine b by an overestimation of the largest eigenvalue, and choose
 203 the lower bound a according to b , for example $a = \frac{1}{2}b$. However, to respect the com-
 204 plementarity principle, the percentage of damped eigenvalues by the smoother must
 205 approximate the proportion of non-coarse variables (i.e. the n_f largest eigenvalues
 206 in our case). For instance, if a coarse level is one quarter the size of the finer level,
 207 then three-quarters of the largest amplitude eigenvectors should be damped by the
 208 smoother, while the coarse correction deals with the remaining small eigenvectors.
 209 Consequently, since eigenvalues are not necessarily uniformly separated, a should be
 210 determined so that a proportion of eigenvalues belongs to the interval $[a, b]$. More-
 211 over, the spectral distribution of coarse matrices are unknown in a multi-level setting.
 212 Therefore, a good interval should satisfy

$$213 \quad (2.8) \quad \lambda_i^2 \in [a, b] \Leftrightarrow \lambda_i \in [-\sqrt{b}, -\sqrt{a}] \cup [\sqrt{a}, \sqrt{b}], i = n_c, \dots, n_f.$$

214 While this interval can be fixed using geometric information, we first compute a
 215 rough approximation of the matrix *spectral density* as detailed in [16]. This spectral
 216 density permits to determine which portion of the spectrum should be damped by

217 the smoother, and is defined by the distribution function $\phi(t)$ that represents the
 218 probability of finding an eigenvalue at each point $t \in [-1, 1]$. We set the lower bound
 219 a of the Chebyshev node interval in a second step so that the probability within
 220 the interval equals the target proportion, for instance half of the total area in a
 221 scenario of exact balance between coarse and non-coarse variables. As defined in (2.6),
 222 the distribution function ϕ is approximated by a linear combination of orthogonal
 223 Chebyshev polynomial functions, such that

$$224 \quad (2.9) \quad \phi(t) = \sum_{k=1}^{\infty} \mu_k C_k(t) \approx \sum_{k=1}^{n_\mu} \mu_k C_k(t).$$

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

$$229 \quad (2.10) \quad \mu_k = \frac{2 - \delta_{k0}}{n\pi} \times \text{Trace}(C_k(B)).$$

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

$$235 \quad (2.11) \quad \text{Trace}(C_k(B)) \approx \frac{1}{n_{\text{vec}}} \sum_{l=1}^{n_{\text{vec}}} \mathbf{z}_l^T C_k(B) \mathbf{z}_l, \quad k = 1, \dots, n_\mu.$$

236 According to (2.11), each trace can be estimated by a sample mean of n_{vec} products
 237 $\mathbf{z}_l^T C_k(B) \mathbf{z}_l$, and the n_μ vectors $C_k(B) \mathbf{z}_l$ can be computed from the three-term re-
 238 currence defined in (2.6). Once the distribution function ϕ is approximated following
 239 Equation (2.9), a rough area approximation by trapezoid rule yields a correct lower
 240 bound that satisfies a proportion around $\frac{n_f}{n}$. This lower bound only needs to be
 241 remapped on the initial interval to return the correct value for a . The interval $[a, b]$
 242 constitutes a purely algebraic interval in which the polynomial smoother is the most
 243 efficient. The bounds a and b are represented in Figure 2.1, where $x_{50\%}$ illustrates a
 244 theoretical lower bound target for the shifted laplacian matrix resulting from (1.6).
 245 Last, the total number of matrix vector products required by the spectral density
 246 approximation step for the construction of a relevant interval is $n_{\text{vec}} \times n_\mu$.

247 **3. Constructing good interpolation rules.** Interpolators are used both to
 248 construct the coarse level matrices and to transfer information across levels. SPD and
 249 geometric smoothness assumptions cannot be used to determine appropriate interpo-
 250 lation operators in our case. Some methods such as smoothed aggregation [9, 20] and
 251 bootstrap-AMG [3] use candidate vectors that are close to the near-kernel space to
 252 design the interpolation rules. These test vectors are either deduced from geometric
 253 information [4, 22] or algebraically as in adaptive multigrid methods [5]. Here, we
 254 prefer to stick to a fully algebraic and recurring process to create our interpolation
 255 operators. Candidate vectors will be generated from random vectors smoothed by the
 256 polynomial presented in Section 2, and used by the least squares minimization frame-
 257 work to determine good fine variable interpolation rules. This initial least squares
 258 interpolation operator is used as a coarse variable operator in the ideal reduction-
 259 based framework [12].

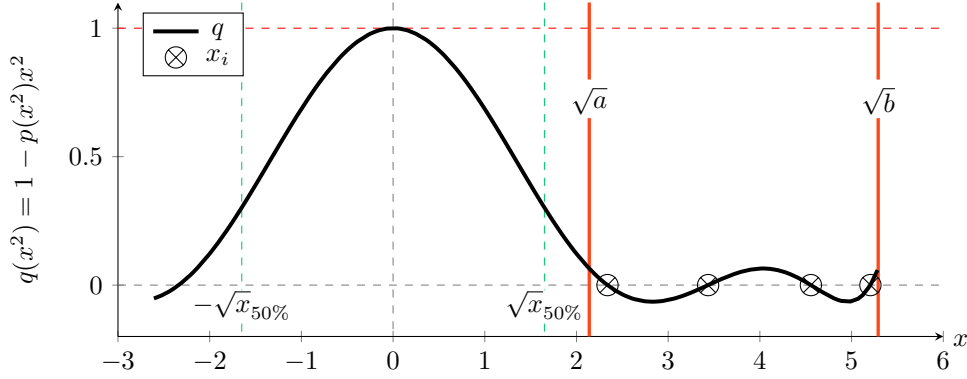


Fig. 2.1: Spectrum of the polynomial smoother error propagation matrix for $kh = 1.65$

260 **3.1. Ideal framework.** Even though the ideal framework requires an SPD as-
 261 sumption and has not been generalized to indefinite problems, the idea of removing
 262 irrelevant information from the interpolation range is of particular interest for cap-
 263 turing the near-kernel space of oscillatory problems, and will be our guiding principle
 264 in this section.

265

266 Accordingly, we assume A is SPD in this section dedicated to the ideal framework.
 267 Following [12], let \mathcal{C} and \mathcal{F} be complementary coarse and fine variable subsets of Ω
 268 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
 269 and fine variable operators respectively, such that $RS = 0$. The space defined by the
 270 coarse variable operator R^T must be handled by the coarse correction, whereas the
 271 fine variable operator S defines a space where smoothing must operate in order to re-
 272 spect the complementarity principle. The *Ideal Interpolation* operator is a theoretical
 273 operator that is the best that satisfies $RP = I_c$, in the sense that it minimizes the
 274 difference between variables and interpolated coarse variables, within a space that is
 275 the most complementary to the range of the smoother M . The ideal interpolation
 276 operator is given by

$$277 \quad (3.1) \quad P_* = \arg \min_P \left(\max_{e \neq 0} \frac{\|(I - PR)e\|_M}{\|e\|_A} \right) = (I - S(S^T AS)^{-1} S^T A) R^T.$$

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

$$280 \quad (3.2) \quad P_{:,i} = R_{:,i}^T - s_i, \text{ with } s_i = \arg \min_{\tilde{s} \in \text{Range}(S)} \|R_{:,i}^T - \tilde{s}\|_A = S(S^T AS)^{-1} S^T A R_{:,i}^T$$

281 In fact, the matrix that multiplies each column of R^T in (3.2) and (3.1) is a projection
 282 operator onto the A -orthogonal complement of the range of S . The ideal interpolation
 283 operator is constructed by extracting from R^T the information that can already be
 284 solved in the subspace S . Such information is irrelevant at a coarse level and should
 285 be handled by the smoother. Under the assumption that the smoother captures the
 286 space spanned by S , the best coarse matrix is therefore a matrix where S -related
 287 information is subtracted. Even if applying $(S^T AS)^{-1}$ is too expensive, it gives
 288 insight for building a more practical method.

289 **3.2. Least Squares Minimization Interpolation Operator.** As mentioned
 290 at the beginning of Section 3.1, demonstrating that the interpolation operator (3.1) is
 291 ideal in the theoretical framework of [12] requires A to be symmetric positive-definite.
 292 However, the reduction viewpoint which consists in cleaning the range of interpola-
 293 tion by extracting irrelevant information at a coarse level perspective is of interest.
 294 In addition, numerical experiments reveal that the classical coarse variable operator
 295 $R^T = [0 \ I_c]^T$ does not have good approximation property for the oscillatory near-
 296 kernel space that characterizes Helmholtz. Therefore, a new coarse variable operator
 297 has to be designed algebraically. Using the smallest eigenvectors V_c from Section 3.1
 298 to enforce the representation of the near-kernel space within the interpolation range
 299 is not practical. Instead, we construct a set of vectors approximating an oscillatory
 300 and potentially large near-kernel space by using the normal equations polynomial
 301 smoother developed in Section 2.

302
 303 In this section, we present a coarse variable operator \hat{R}^T of size $n \times n_c$ construc-
 304 ted by a least squares minimization strategy [3]. Let the columns of T be a set of κ
 305 smoothed random vectors z_l that approximates the near-kernel space such that

$$306 \quad (3.3) \quad T_{:,l} = q_{m+1}(A^2)z_l, \quad l = 1, \dots, \kappa.$$

307 where $T_{:,l}$ designate the l^{th} column of the set T . We assume a \mathcal{C}/\mathcal{F} splitting with n_c
 308 and n_f their respective size. \mathcal{C} -points are interpolated to the finer level with a simple
 309 injection rule, while interpolation rules of \mathcal{F} -points are determined by the least
 310 squares minimization method presented in this section. Due to this splitting, the
 311 coarse interpolation block in \hat{R}^T corresponds to a $n_c \times n_c$ identity matrix denoted by
 312 I_c , while R_f^T designate the block of interpolation for the \mathcal{F} -points.

313
 314 Let i be an \mathcal{F} -point and \hat{r}_i the vector containing the non-zero elements of the i^{th}
 315 row of \hat{R}^T . The idea consists of constructing each \mathcal{F} -point interpolation rule by min-
 316 imizing the squared difference between \mathcal{F} -values of the near-kernel candidate vectors
 317 and the interpolation from their connected \mathcal{C} -points in \mathcal{C}_i . Denote by $T_{i,:}$ a row vector
 318 containing the i^{th} values of each test vector, and $T_{\mathcal{C}_i,l}$ a vector containing the values
 319 in $T_{:,l}$ of the \mathcal{C} -points that are connected to variable i . Then

$$320 \quad (3.4) \quad \forall i \in \mathcal{F}, \quad \hat{r}_i = \arg \min_{\hat{r} \in \mathbb{C}^{\text{card}(\mathcal{C}_i)}} \sum_{l=1}^{\kappa} w_l (T_{i,l} - \hat{r} \cdot T_{\mathcal{C}_i,l})^2 =: \arg \min_{\hat{r} \in \mathbb{C}^{\text{card}(\mathcal{C}_i)}} \mathcal{L}_i(\hat{r})$$

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

$$323 \quad (3.5) \quad \nabla \mathcal{L}_i(\hat{r}_i) = 0.$$

324 Equation (3.5) can be rewritten element-wise

$$325 \quad (3.6) \quad \frac{\partial \mathcal{L}_i(\hat{r}_i)}{\partial \hat{r}_{ij}} = \sum_{l=1}^{\kappa} 2w_l (T_{i,l} - \hat{r}_i \cdot T_{\mathcal{C}_i,l}) T_{\mathcal{C}_i,l} = 0, \quad \forall j = 1, \dots, \text{card}(\mathcal{C}_i).$$

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

$$327 \quad (3.7) \quad \hat{r}_i T_{\mathcal{C}_i} W T_{\mathcal{C}_i}^T = T_i W T_{\mathcal{C}_i}^T$$

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

333 **3.3. Ideal approximation from least squares coarse operator.** In Section
 334 3.2, we presented a coarse variable operator for Helmholtz designed by a least squares
 335 minimization strategy. Using the framework presented in 3.1, define

$$336 \quad (3.8) \quad \hat{R}^T = \begin{bmatrix} R_f^T \\ I_c \end{bmatrix} \text{ and } \hat{S} = \begin{bmatrix} I_f \\ -R_f \end{bmatrix}$$

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

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

$$360 \quad (3.9) \quad \mathbf{s}_i := \arg \min_{\tilde{\mathbf{s}} \in \text{Range}(\hat{S}X_i)} \|\hat{R}_{:,i}^T - \tilde{\mathbf{s}}\|_{A^T A} = \hat{S}X_i \left(X_i^T \hat{S}^T A^T A \hat{S}X_i \right)^{-1} X_i^T \hat{S}^T A^T A \hat{R}_{:,i}^T.$$

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

$$362 \quad (3.10) \quad \hat{P}_{:,i} = \hat{R}_{:,i}^T - \mathbf{s}_i = \hat{R}_{:,i}^T - \hat{S}X_i \rho_{n_i},$$

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

$$364 \quad (3.11) \quad X_i^T \hat{S}^T A^T A \hat{S}X_i \rho_{n_i} = X_i^T \hat{S}^T A^T A \hat{R}_{:,i}^T.$$

365 The choice of the non-zero pattern of \hat{P} must satisfy a good trade-off between ap-
 366 proximation properties of the near-kernel space and complexity. While improving the
 367 sparsity of this interpolation operator is a topic of future research, one strategy is
 368 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

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

378 (3.12)
$$\|X_i \hat{S}^T A^T A \hat{R}_{:,i}^T\|^2 \geq \tau \times \|\hat{S}^T A^T A \hat{R}_{:,i}^T\|^2, \text{ with } \tau \in [0, 1].$$

379 We note that even though setting $\tau = 1$ selects all the column associated with non-
 380 zero entries in the right-hand side, the remaining columns associated with zero entries
 381 are omitted, and therefore the matrix $X_i^T \hat{S}^T A^T A \hat{S} X_i$ still correspond to a principle
 382 sub-matrix of $\hat{S}^T A^T A \hat{S}$. The Figure 3.3 represents the error of interpolation of every
 383 eigenvector for two different shifted problems resulting from (1.6) with respect to τ .
 384 The red dots correspond to the error when no ideal approximation is used at all (i.e.
 385 $\tau = 0$ and therefore $\hat{P} = \hat{R}^T$), whereas blue and green dots represent the error of
 386 interpolation for $\tau = 0.5$ and $\tau = 1$ respectively. The legend for each color associates
 387 the percentage τ with the average number of non-zero entries $\frac{\text{nnz}}{n}$ in the resulting
 388 interpolation operator. Because the subspace $\hat{S} X_i$ grows with τ , larger values of τ
 389 leads to denser interpolation operators. For both shifts, the portion of the spectrum
 390 for which the least-squares minimization interpolation operator is the most accurate
 391 corresponds to the smallest eigenvalues in magnitude. This feature is an expected
 392 and desired effect of generating the set of test vectors from the polynomial smoother
 393 introduced in Section 2. However, the interpolation error increases with the shift.
 394 Therefore, the ideal approximation correction becomes necessary as the problem gets
 395 more indefinite. In particular, Figure 3.3 shows that the interpolation error decreases
 396 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.

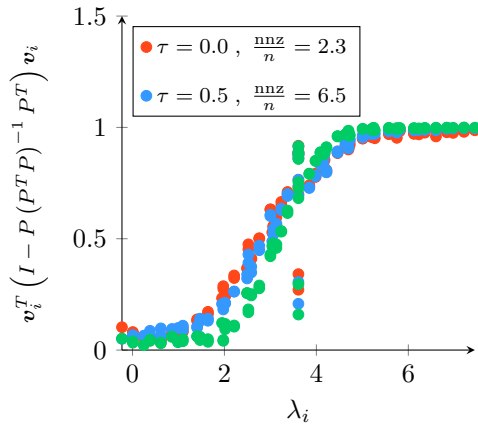


Fig. 3.1: $kh = 0.625$

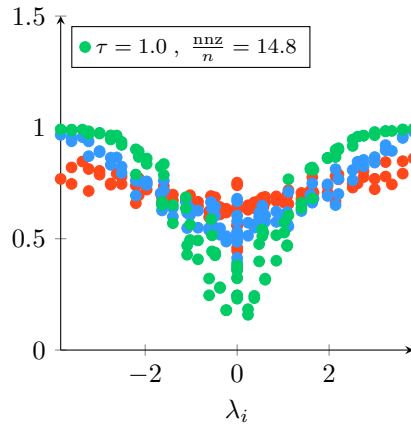


Fig. 3.2: $kh = 2.0$

Fig. 3.3: Error of interpolation with respect to the shift and sparsity

397

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

406
 407 The Figure 4.3 plots the smallest eigenvector of a two-dimensional shifted Lapla-
 408 cian matrix in blue for two different shifts. The shift of 4.2 is greater than the shift
 409 of 4.1. As expected, the higher the shift, the more oscillatory the problem. In red are
 410 plotted the results of the coarse correction when applied to the blue eigenvectors. In
 411 this example, the coarse correction is implemented with the reduction-based interpo-
 412 lation operator introduced in Section 3. Additionally, the green curves represent the
 413 best representation of both eigenvectors in the interpolation range. First, note that
 414 the blue and green curves align almost perfectly in both sub-figures, which means that
 415 the interpolation range introduced in Section 3 offers a good approximation to the
 416 potentially oscillatory smallest eigenvector. In both cases, \hat{P} has good approximation
 417 properties. In Figure 4.1, where the problem is discretized with 10 points per wave-
 418 length, the red coarse correction vector is relatively close to the blue eigenvector. The
 419 slight difference between both is only a matter of amplitude. In contrast, while the
 420 oscillations of the coarse correction vector illustrated in Figure 4.2 are synchronized
 421 with the oscillations of the smallest eigenvector, its direction is reversed. In that case,
 422 while the interpolation range is almost perfect, the error of the smallest eigenvector
 423 is not reduced by the coarse correction, but amplified.

424 At this stage, let us define a concept of pollution to better understand how the matrix
 425 indefiniteness can corrupt the coarse correction.

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

$$431 \quad (4.1) \quad K := V^T P(V_c^T P)^{-1} = \begin{bmatrix} I_c \\ K_f \end{bmatrix}.$$

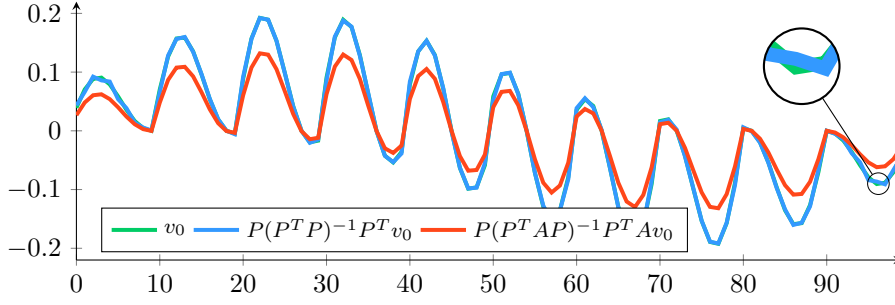
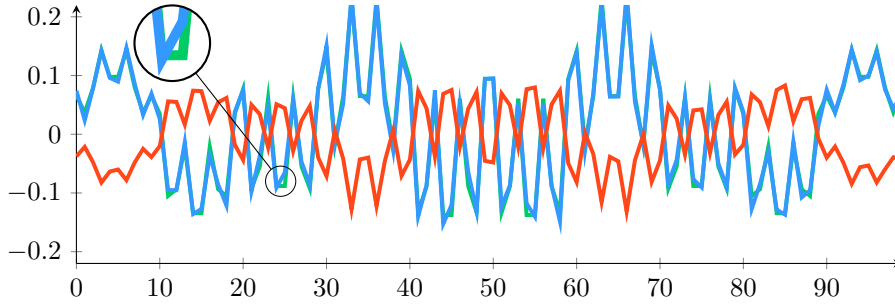
432 *The block I_c corresponds to the identity matrix of size $n_c \times n_c$, and the block K_f*
 433 *is a $n_f \times n_c$ matrix such that $K_f := V_f^T P(V_c^T P)^{-1}$. The interpolation error of the*
 434 *eigenvector \mathbf{v}_i of V_c is given by*

$$435 \quad (4.2) \quad \mathbf{v}_i^T (I - \Pi(P)) \mathbf{v}_i = 1 - \left[(I_c + K_f^T K_f)^{-1} \right]_{i,i},$$

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

437 *Proof.* First, note that post-multiplying P by any non-singular matrix M_c of size
 438 $n_c \times n_c$ does not change the l_2 -projection

$$439 \quad (PM_c)((PM_c)^T(PM_c))^{-1}(PM_c)^T = PM_c M_c^{-1} (P^T P)^{-1} M_c^{-T} M_c^T P^T \\ 440 \quad (4.3) \quad = P(P^T P)^{-1} P^T = \Pi(P).$$

Fig. 4.1: $kh = 0.625$ Fig. 4.2: $kh = 1.71$ Fig. 4.3: Layering of : \mathbf{v}_1 (blue) vs. $P(P^T P)^{-1} P^T \mathbf{v}_1$ (green) vs. $P(P^T A P)^{-1} P^T A \mathbf{v}_1$ (red), for two different shifts

442 In particular for $M_c = (V_c^T P)^{-1}$,

$$443 \quad I - \Pi(P) = I - P(P^T P)^{-1} P^T$$

$$444 \quad (4.4) \quad = I - P(V_c^T P)^{-1} (P(V_c^T P)^{-1})^T P(V_c^T P)^{-1} (P(V_c^T P)^{-1})^T.$$

446 Since $P(V_c^T P)^{-1} = VK$, it follows that

$$447 \quad I - \Pi(P) = I - (VK)((VK)^T(VK))^{-1}(VK)^T$$

$$448 \quad (4.5) \quad = I - VK(K^T K)^{-1} K^T V^T.$$

450 For any eigenvector \mathbf{v}_i of A , let $\mathbf{e}_i := V^T \mathbf{v}_i$ be the canonical unit vector with a
 451 one at the i^{th} position and zero elsewhere. Assuming $\mathbf{v}_i \in V_c$ ($i \leq n_c$), the vector
 452 $\mathbf{c}_i := K^T \mathbf{e}_i$ of size n_c is also a unit vector with a one at the i^{th} position. Consequently,
 453 the damping factor of $\mathbf{v}_i \in V_c$ is

$$454 \quad \mathbf{v}_i^T (I - \Pi(P)) \mathbf{v}_i = \mathbf{v}_i^T V (I - K(K^T K)^{-1} K^T) V^T \mathbf{v}_i$$

$$455 \quad = \mathbf{e}_i^T (I - K(K^T K)^{-1} K^T) \mathbf{e}_i$$

$$456 \quad (4.6) \quad = 1 - \mathbf{c}_i^T (K^T K)^{-1} \mathbf{c}_i = 1 - [(I_c + K_f^T K_f)^{-1}]_{i,i}. \quad \square$$

458 Since the l_2 -projection is unchanged by post-multiplication of P , we assume for what
 459 follows that K has the form (4.1). The block K_f designates what we call “pollu-
 460 tion”. This block of pollution causes the slight difference between an eigenvector

461 \mathbf{v}_i of V_c and its best representation in the range of P . When a column of K_f is
 462 null, the interpolation error of the associated eigenvector equals zero, such that blue
 463 and green curves align perfectly. In practice however, this property is unlikely to be
 464 satisfied for Helmholtz, because P should be sparse for cost considerations and the
 465 smallest eigenvectors are usually unknown. Moreover, the near-kernel space of the
 466 Helmholtz equation is oscillatory. This makes the construction of good interpolation
 467 rules more difficult, and tends to pollute the interpolation range. In fact, this pollu-
 468 tion is probably unavoidable and the columns of K_f are unlikely to be zero. While
 469 the pollution decreases the convergence speed of multigrid methods for SPD prob-
 470 lems, we demonstrate that it can corrupt the coarse correction and make the method
 471 diverge in the indefinite case, as illustrated by the reversed red vector of Figure 4.3(b).
 472

473 In that direction, let us discuss the effectiveness of the coarse correction by looking at
 474 the contraction of the n_c small eigenvectors V_c only, assuming the n_f large eigenvectors
 475 V_f are damped by the smoother.

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

$$479 \quad (4.7) \quad \mathbf{v}_i^T E \mathbf{v}_i = 1 - \lambda_i \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i}.$$

480

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

$$484 \quad (PM_c)((PM_c)^T A (PM_c))^{-1} (PM_c)^T = PM_c M_c^{-1} (P^T A P)^{-1} M_c^{-T} M_c^T P^T \\ 485 \quad (4.8) \quad = P(P^T A P)^{-1} P^T$$

487 In particular for $M_c = (V_c^T P)^{-1}$,

$$488 \quad E = I - P(P^T A P)^{-1} P^T A \\ 489 \quad (4.9) \quad = I - P(V_c^T P)^{-1} (P(V_c^T P)^{-1})^T A P(V_c^T P)^{-1} (P(V_c^T P)^{-1})^T A.$$

491 Similar to (4.5), the equality $P(V_c^T P)^{-1} = VK$ leads to

$$492 \quad (4.10) \quad E = I - (VK)((VK)^T A (VK))^{-1} (VK)^T A = V(I - K(K^T \Lambda K)^{-1} K^T \Lambda) V^T.$$

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

$$496 \quad \mathbf{v}_i^T E \mathbf{v}_i = \mathbf{v}_i^T V(I - K(K^T \Lambda K)^{-1} K^T \Lambda) V^T \mathbf{v}_i \\ 497 \quad = \mathbf{e}_i^T (I - K(K^T \Lambda K)^{-1} K^T \Lambda) \mathbf{e}_i \\ 498 \quad (4.11) \quad = 1 - \lambda_i \mathbf{c}_i^T (K^T \Lambda K)^{-1} \mathbf{c}_i = 1 - \lambda_i \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i}. \quad \square \\ 499$$

500 Theorem 4.2 shows that the damping factors rely on a combination of the small ei-
 501 genvalues Λ_c plus the large eigenvalues Λ_f , such that the mix is given by the entries
 502 of the pollution K_f .
 503

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

$$506 \quad (4.12) \quad \forall i \leq n_c, 0 \leq \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i} \leq [\Lambda_c^{-1}]_{i,i} = \lambda_i \Rightarrow 0 \leq \mathbf{v}_i^T E \mathbf{v}_i \leq 1.$$

507 Hence, the coarse correction always operates a contraction on \mathbf{v}_i regardless the block
508 of pollution K_f . In the indefinite case however, the property (4.12) does not hold. In
509 fact, a necessary condition for the coarse correction to be a contraction is

$$510 \quad (4.13) \quad \forall i \leq n_c, |\mathbf{v}_i^T E \mathbf{v}_i| \leq 1 \Rightarrow 0 \leq \lambda_i \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{i,i} \leq 2.$$

511 From Equation (4.13), it follows that each diagonal entry must have the same sign as
512 its associated eigenvalue, and be smaller than twice the inverse of the eigenvalue in
513 magnitude. Nothing guarantee such conditions to be satisfied in the case where small
514 and large and either negative or positive eigenvalues are mixed. Especially for very
515 small eigenvalues, the mix can easily lead to a diagonal entry of the opposite sign
516 even though K_f is small, because its entries are weighted by the large eigenvalues
517 Λ_f . Therefore, a good interpolation operator can still cause the coarse correction to
518 amplify the error. For very near-zero eigenvalues, even a round-off error can eventually
519 lead to divergence in the indefinite case. The following example better depicts how
520 the pollution can cause divergence in the indefinite setting for a 2×2 matrix.

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

$$524 \quad (4.14) \quad P = \mathbf{v}_1 + \epsilon \mathbf{v}_2.$$

525 *From definition (4.1), the K matrix can be derived by*

$$526 \quad (4.15) \quad K = V^T P (\mathbf{v}_1^T P)^{-1} = [\mathbf{v}_1, \mathbf{v}_2]^T \cdot [\mathbf{v}_1 + \epsilon \times \mathbf{v}_2] = \begin{bmatrix} 1 \\ \epsilon \end{bmatrix}.$$

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

$$528 \quad (4.16) \quad \mathbf{v}_1^T E \mathbf{v}_1 = 1 - \lambda_1 \left[(\Lambda_c + K_f^T \Lambda_f K_f)^{-1} \right]_{1,1} = 1 - \frac{\lambda_1}{\lambda_1 + \epsilon^2 \lambda_2}$$

529

530 *The figure 4.4 depicts the action of the coarse correction on \mathbf{v}_1 with respect to the*
531 *pollution block $K_f^T \Lambda_f K_f = \epsilon^2 \lambda_2$. A first observation is that the coarse correction*
532 *does not amplify the smallest eigenvector if eigenvalues have the same sign. If the*
533 *eigenvalues are oppositely signed, then the coarse correction amplifies \mathbf{v}_1 for $\epsilon^2 \lambda_2 <$
534 $-\lambda_1/2$. Therefore, the condition on the pollution $K_f = \epsilon$ that drives the error of*
535 *interpolation is particularly difficult respectively for small and large values of λ_1 and*
536 λ_2 .

537 The next theorem derives a more general condition for the coarse correction to be a
538 contraction of the smallest eigenvalues in the indefinite case based on the concept of
539 pollution.

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

$$541 \quad (4.17) \quad \left| \lambda_{n_c} (K_f^T \Lambda_f K_f) \right| \leq \frac{1}{2} |\lambda_1| \Rightarrow \forall \mathbf{v}_i \in V_c, |\mathbf{v}_i^T E \mathbf{v}_i| \leq 1$$

542

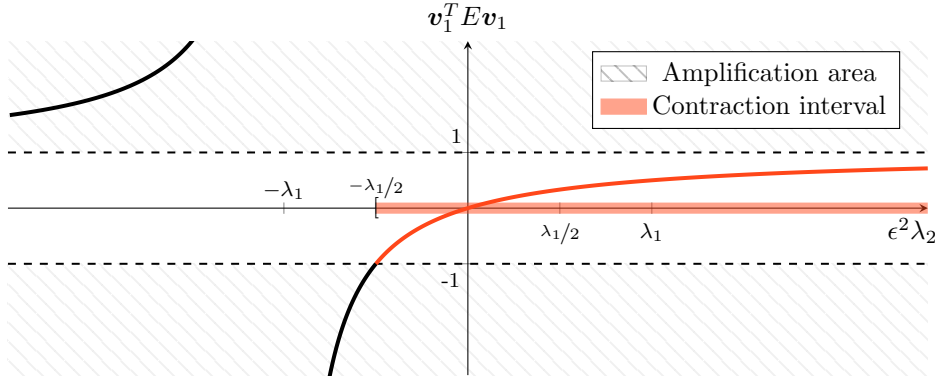


Fig. 4.4: Contraction of the coarse correction with respect to the pollution

543 *Proof.* Define $M_K = I_c + \Lambda_c^{-1} K_f^T \Lambda_f K_f$. From the shape of the matrix K defined
 544 in Equation (4.10), we have

$$\begin{aligned}
 545 \quad V_c^T E V_c &= V_c^T V (I - K(K^T \Lambda K)^{-1} K^T \Lambda) V^T V_c \\
 546 &= I_c - (K^T \Lambda K)^{-1} \Lambda_c \\
 547 &= I_c - (I_c + \Lambda_c^{-1} K_f^T \Lambda_f K_f)^{-1} \Lambda_c^{-1} \Lambda_c \\
 548 \quad (4.18) \quad &= I_c - M_K^{-1}.
 \end{aligned}$$

550 Hence, it follows that

$$551 \quad (4.19) \quad \forall \mathbf{v}_i \in V_c, \quad \mathbf{v}_i^T E \mathbf{v}_i = \mathbf{e}_i^T V_c^T E V_c \mathbf{e}_i = 1 - \mathbf{e}_i^T M_K^{-1} \mathbf{e}_i.$$

552 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

$$553 \quad (4.20) \quad \forall \mathbf{v}_i \in V_c, \quad -1 \leq \mathbf{v}_i^T E \mathbf{v}_i \leq 1 \Leftrightarrow 0 \leq \mathbf{e}_i^T M_K^{-1} \mathbf{e}_i \leq 2.$$

554 We begin by deriving a condition for the right bound of (4.20), and will show, in a
 555 second time, that it also satisfies the left one. Let \mathbf{x} and \mathbf{y} two vectors of \mathbb{R}^n linked
 556 by the relation $\mathbf{x} = M_K \mathbf{y}$. The right bound is satisfied if

$$557 \quad (4.21) \quad \max_{\mathbf{x} \neq 0} \frac{\|M_K^{-1} \mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{y} \neq 0} \frac{\|\mathbf{y}\|}{\|M_K \mathbf{y}\|} = \left(\min_{\mathbf{y} \neq 0} \frac{\|M_K \mathbf{y}\|}{\|\mathbf{y}\|} \right)^{-1} \leq 2.$$

558 Therefore, the condition (4.21) is equivalent to

$$559 \quad (4.22) \quad \min_{\mathbf{y} \neq 0} \frac{\|M_K \mathbf{y}\|}{\|\mathbf{y}\|} \geq \frac{1}{2}.$$

560 Let $\sigma_i(M)$ be the i^{th} largest singular value of a given matrix M (we omit the matrix
 561 between parenthesis when referring to the singular values of the initial matrix A). In
 562 addition, let us recall the following triangle inequality $\|\mathbf{y} + \mathbf{z}\| \geq \|\mathbf{y}\| - \|\mathbf{z}\|$, $\forall \mathbf{y}, \mathbf{z} \in$
 563 \mathbb{R}^{n_c} . Thus, we have that

$$\begin{aligned}
 564 \quad \min_{\mathbf{y} \neq 0} \frac{\|M_K \mathbf{y}\|}{\|\mathbf{y}\|} &= \min_{\mathbf{y} \neq 0} \frac{\|\mathbf{y} + \Lambda_c^{-1} K_f^T \Lambda_f K_f \mathbf{y}\|}{\|\mathbf{y}\|} \geq \min_{\mathbf{y} \neq 0} \left(1 - \frac{\|\Lambda_c^{-1} K_f^T \Lambda_f K_f \mathbf{y}\|}{\|\mathbf{y}\|} \right) \\
 565 \quad (4.23) \quad &= 1 - \max_{\mathbf{y} \neq 0} \frac{\|\Lambda_c^{-1} K_f^T \Lambda_f K_f \mathbf{y}\|}{\|\mathbf{y}\|} \\
 566 \quad &= 1 - \sigma_{n_c}(\Lambda_c^{-1} K_f^T \Lambda_f K_f).
 \end{aligned}$$

568 It follows that the condition (4.22) is satisfied if $\sigma_{n_c}(\Lambda_c^{-1}K_f^T\Lambda_fK_f) \leq \frac{1}{2}$. Finally,
 569 since $\sigma_{n_c}(\Lambda_c^{-1}K_f^T\Lambda_fK_f) \leq \sigma_1^{-1}\sigma_{n_c}(K_f^T\Lambda_fK_f)$ and the singular values coincide with
 570 eigenvalues in magnitude because both Λ_c and $K_f^T\Lambda_fK_f$ are hermitian, the right
 571 bound of (4.20) is satisfied if

$$572 \quad (4.24) \quad |\lambda_{n_c}(K_f^T\Lambda_fK_f)| \leq \frac{1}{2}|\lambda_1|.$$

573 We now address the left bound of (4.20) assuming the condition (4.24) holds. Our
 574 goal is to prove that all diagonal entries of M_K^{-1} are positives. In that direction, let
 575 $F(M)$ be the field of values of a given matrix M of size n_c such that

$$576 \quad (4.25) \quad F(M) := \{\mathbf{x}^*M\mathbf{x} \mid \forall \mathbf{x} \in \mathbb{C}^{n_c}, \mathbf{x}^*\mathbf{x} = 1\}.$$

577 If M is hermitian, one can show that (e.g. [14, chapter 4])

$$578 \quad (4.26) \quad \min_{\mathbf{x}^*\mathbf{x}=1} \mathbf{x}^*M\mathbf{x} = \lambda_{\min}(M) \text{ and } \max_{\mathbf{x}^*\mathbf{x}=1} \mathbf{x}^*M\mathbf{x} = \lambda_{\max}(M).$$

580 Accordingly, let $F(\Lambda_c)$ and $F(K_f^T\Lambda_fK_f)$ be the field of values of Λ_c and $K_f^T\Lambda_fK_f$
 581 respectively. Since A is non-singular, then $0 \notin F(\Lambda_c)$. Therefore, the spectrum of
 582 $\Lambda_c^{-1}K_f^T\Lambda_fK_f$ is included as follows (e.g. [13, chapter 1])

$$583 \quad (4.27) \quad \forall j \leq n_c, \lambda_j(\Lambda_c^{-1}K_f^T\Lambda_fK_f) \in F(K_f^T\Lambda_fK_f)/F(\Lambda_c).$$

584 The set ratio in (4.27) has the usual algebraic interpretation such that

$$585 \quad (4.28) \quad \forall \alpha \in \frac{F(K_f^T\Lambda_fK_f)}{F(\Lambda_c)}, -\frac{\max_{\mathbf{x}^*\mathbf{x}=1} |\mathbf{x}^*K_f^T\Lambda_fK_f\mathbf{x}|}{\min_{\mathbf{x}^*\mathbf{x}=1} |\mathbf{x}^*\Lambda_c\mathbf{x}|} \leq \alpha \leq \frac{\max_{\mathbf{x}^*\mathbf{x}=1} |\mathbf{x}^*K_f^T\Lambda_fK_f\mathbf{x}|}{\min_{\mathbf{x}^*\mathbf{x}=1} |\mathbf{x}^*\Lambda_c\mathbf{x}|}.$$

586 Furthermore, matrices Λ_c and $K_f^T\Lambda_fK_f$ are hermitian so the property (4.26) holds
 587 for both of them. Because the spectrum belongs to the set ratio as in (4.27), we have

$$588 \quad (4.29) \quad -|\lambda_1|^{-1} \cdot |\lambda_{n_c}(K_f^T\Lambda_fK_f)| \leq \lambda_j(\Lambda_c^{-1}K_f^T\Lambda_fK_f) \leq |\lambda_{n_c}(K_f^T\Lambda_fK_f)| \cdot |\lambda_1|^{-1}.$$

589 Therefore, assuming the condition (4.24) is satisfied, it follows

$$590 \quad (4.30) \quad \lambda_j(\Lambda_c^{-1}K_f^T\Lambda_fK_f) \geq -|\lambda_{n_c}(K_f^T\Lambda_fK_f)| \times |\lambda_1|^{-1} \geq -\frac{1}{2}.$$

591 Adding one to each member of the inequality (4.30) finally gives

$$592 \quad (4.31) \quad \lambda_j(M_K) = \lambda_j(I + \Lambda_c^{-1}K_f^T\Lambda_fK_f) \geq \frac{1}{2}$$

593 The condition (4.24) implies that all eigenvalues of M_K are positives. Subsequently,
 594 $\det(M_K) > 0$. The adjugate formula for the inverse of M_K shows that diagonal
 595 entries are positives if the determinant of principal sub-matrices are also positives. In
 596 that direction, denote by $[\cdot]_{\Omega_{-i}}$ the principal sub-matrix obtained by deleting the i^{th}
 597 row and column of a matrix. Since Λ_c is diagonal, one can show that

$$598 \quad (4.32) \quad [\Lambda_c^{-1}K_f^T\Lambda_fK_f]_{\Omega_{-i}} = [\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T\Lambda_fK_f]_{\Omega_{-i}}.$$

599 As in Equation (4.27), the spectrum is included such that

$$600 \quad \forall j \leq n_c - 1, \lambda_j \left([\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) \in F \left([K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) / F \left([\Lambda_c]_{\Omega_{-i}} \right),$$

601 and therefore the following bound holds

$$602 \quad (4.33) \quad \lambda_j \left([\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) \geq - \left| \lambda_{n_c-1} \left([K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) \right| \times |\lambda_1|^{-1}.$$

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

$$604 \quad (4.34) \quad \lambda_j (K_f^T \Lambda_f K_f) \leq \lambda_j \left([K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) \leq \lambda_{j+1} (K_f^T \Lambda_f K_f), \quad j = 1, \dots, n_c - 1.$$

605 As a consequence, and from the inequality (4.30), we have

$$606 \quad (4.35) \quad \lambda_j \left([\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) \geq - |\lambda_{n_c} (K_f^T \Lambda_f K_f)| \times |\lambda_1|^{-1} \geq -\frac{1}{2}.$$

607 Hence, eigenvalues of principal sub-matrices also satisfy

$$608 \quad (4.36) \quad \lambda_j \left([M_K]_{\Omega_{-i}} \right) = \lambda_j \left(I_{n_c-1} + [\Lambda_c]_{\Omega_{-i}}^{-1} [K_f^T \Lambda_f K_f]_{\Omega_{-i}} \right) \geq \frac{1}{2}.$$

609 Because eigenvalues of the principal sub-matrices are positives, so are the determi-
610 nants. From the adjugate formula of M_K^{-1} , it follows that

$$611 \quad (4.37) \quad \mathbf{e}_i^T M_K^{-1} \mathbf{e}_i = [M_K^{-1}]_{i,i} = \frac{\det \left([M_K]_{\Omega_{-i}} \right)}{\det \left(M_K \right)} \geq 0, \quad i = 1, \dots, n_c$$

612 As a consequence, both left and right bounds of (4.20) are satisfied. Finally,

$$613 \quad (4.38) \quad \left| \lambda_{n_c} (K_f^T \Lambda_f K_f) \right| \leq \frac{1}{2} |\lambda_1| \Rightarrow \forall \mathbf{v}_i \in V_c, \quad |\mathbf{v}_i^T E \mathbf{v}_i| \leq 1 \quad \square$$

614 The condition provided by Theorem 4.4 is that the amplitude of the block $K_f^T \Lambda_f K_f$
615 never exceeds half of the smallest eigenvalue in magnitude. No assumption can be
616 made on the sign of eigenvalues in the indefinite case, so that the condition pre-
617 vents the coarse correction from amplifying the error in the case where eigenvalues
618 are oppositely signed. Applied to the previous example 4.4, Theorem 4.4 states that
619 $|\epsilon^2 \lambda_2| < |\lambda_1|/2$. That said, the condition is extremely strict and probably impossible
620 to satisfy in practice for very small eigenvalues. In a practical method, the block K_f
621 will never be sufficiently small for solving all type of indefinite problems because of
622 a potentially very near-zero eigenvalue. As illustrated by Figure 4.3, a good inter-
623 polation operator with small K_f can still cause divergence although it satisfies good
624 approximation properties. The classical coarse correction is hopeless for indefinite
625 problems.

626 **5. Alternative coarse correction for indefinite problems.** As discussed in
627 the previous section, the classical coarse correction is not equivalent to a minimization
628 problem in the indefinite case, and improving P will never be enough to remedy
629 this loss of equivalence. Moreover, because the interpolation operator developed in
630 Section 3 targets the smallest eigenvectors of each level, every coarser matrix is more
631 indefinite than its fine parent. Then, as the number of coarse levels increases, the

632 balance between negative and positive eigenvalues reaches an equilibrium, and makes
 633 the effectiveness of the classical coarse correction difficult to predict. Nevertheless,
 634 Figure 4.3 shows that the interpolation operator has good approximation properties
 635 for the oscillatory near-kernel space. In particular, the Figure 4.3(b) suggests that only
 636 the direction of the coarse correction vector has to be changed; the shape is correct.
 637 Hence, a coarse correction that amplifies or flips the smallest eigenvectors can still
 638 provide pertinent information for solving the system. In this section, we propose to
 639 minimize the approximation error in a proper norm for indefinite problems and within
 640 a space composed of vectors returned by the classical coarse correction. Moreover, to
 641 decrease the eigenvector pollution, each coarse correction vector is smoothed by the
 642 polynomial smoother of Section 2.

643 **5.1. Notations and general considerations on GMRES.** The *Generalized*
 644 *Minimal RESidual* (GMRES) method [24] approximates the solution in a Krylov
 645 subspace by minimizing the residual in the Euclidean norm. The method can solve
 646 any class of matrix system since the norm is valid independent of the context, which
 647 is of particular interest for the indefinite case. Let us first define some notation before
 648 introducing the alternative coarse correction. Let W_p be the $n \times p$ rectangular matrix
 649 containing the p orthonormalized Krylov vectors such that

$$650 \quad (5.1) \quad \text{range}(W_p) = \text{span} \{ \mathbf{b}, A\mathbf{b}, A^2\mathbf{b}, \dots, A^{p-1}\mathbf{b} \}.$$

651 Each column of W_p is orthonormalized following a Gram-Schmidt process. The co-
 652 efficients of the orthonormalization are stored in the rectangular Hessenberg matrix
 653 \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
 654 deleting its last row. Both matrices W_p and H_p are linked by

$$655 \quad (5.2) \quad AW_p = W_{p+1}\bar{H}_p \text{ and } W_p^T AW_p = H_p,$$

656 which leads to the following equality

$$657 \quad \arg \min_{\tilde{\mathbf{x}} \in \text{range}(W_p)} \|\mathbf{b} - A\tilde{\mathbf{x}}\|_2 = \arg \min_{\boldsymbol{\rho}_m \in \mathbb{C}^p} \|\mathbf{b} - AW_p\boldsymbol{\rho}_m\|_2 = \arg \min_{\boldsymbol{\rho}_p \in \mathbb{C}^p} \|W_p^T \mathbf{b} - H_p\boldsymbol{\rho}_p\|_2$$

$$658 \quad (5.3) \quad = W_p H_p^{-1} W_p^T \mathbf{b}.$$

660 In practice, GMRES takes advantage of the convenient Hessenberg shape of \bar{H}_p to
 661 construct an upper triangular matrix by applying Given's rotations. The minimization
 662 of the residual then relies on a backward substitution. The relation (5.2) can be
 663 generalized [7] to any arbitrary subspace $W_p = [\mathbf{w}_1, \dots, \mathbf{w}_p]$ such that

$$664 \quad (5.4) \quad \arg \min_{\tilde{\mathbf{x}} \in \text{range}(W_p)} \|\mathbf{b} - A\tilde{\mathbf{x}}\|_2 = W_p H_p^{-1} Z_p^T \mathbf{b} \text{ with } AW_p = Z_p H_p$$

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

669 **5.2. Minimization within a space of coarse correction vectors.** As men-
 670 tioned in the introduction of this section, the interpolation operator has good ap-
 671 proximation properties for the oscillatory near-kernel space. Even though the small
 672 eigenvectors that constitute each coarse correction vector are likely to be oriented in

673 the wrong direction or amplified because of the pollution effect introduced in Sec-
 674 tion 4, they still provide useful information about the near-kernel space. For ease of
 675 discussion, we present this idea on a two-level method. The multi-level case will be
 676 depicted in the next section dedicated to numerical experiments.

677

678 In that direction, let W_i be the set of coarse correction vectors of the i^{th} iteration
 679 linked by the Arnoldi relation (5.4) with its orthonormal counterpart Z_i . Accordingly,
 680 let $\mathbf{w}_j \in W_i$ and $\mathbf{z}_j \in Z_i$ denote the j^{th} vectors of the set W_i and Z_i respectively.
 681 At each iteration i , the classical coarse correction returns a new coarse correction
 682 vector that is smoothed by the Chebyshev polynomial smoother presented in Section
 683 2. This new smoothed coarse correction vector is therefore added to the previous set
 684 such that

$$685 \quad (5.5) \quad W_i = [W_{i-1}, \mathbf{w}_i] \text{ with } \mathbf{w}_i = q_{m+1}^\nu(A^2)\Pi_A(P)\mathbf{r}^{(i)},$$

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

$$688 \quad (5.6) \quad H_i = Z_i^T A W_i = H_i^{-T} W_i^T A^T A W_i, \quad Z_i = A W_i H_i^{-1}.$$

689 Hence, solving the minimization problem (5.4) is equivalent to solve the normal equa-
 690 tions within the subspace spanned by W_i

$$691 \quad W_i H_i^{-1} Z_i^T A = W_i (W_i^T A^T A W_i)^{-1} H_i^T Z_i^T A$$

$$692 \quad (5.7) \quad = W_i (W_i^T A^T A W_i)^{-1} W_i^T A^T A = \Pi_{A^T A}(W_i).$$

694 The concept of pollution also drives convergence in the alternative setting. Section
 695 4 demonstrated that the block K_f pollutes the range of P and therefore impacts the
 696 classical coarse correction. Because the minimization W_i resorts to the classical coarse
 697 correction by way of Equation (5.5), the block of pollution still impact the capture of
 698 the small eigenvectors. Resorting to euclidean norm in (5.4) prevents from divergence,
 699 but it also squares the eigenvalues of the initial problem because of the equivalence
 700 with an $A^T A$ -orthogonal projection. This naturally increases the gap between small
 701 and large eigenvalues, and therefore decreases the contraction of the smallest over the
 702 largest.

703

704 Smoothing the classical coarse correction vectors by way of the polynomial $q_{m+1}^\nu(A^2)$
 705 compensates this effect by decreasing the distribution of large eigenvectors in the min-
 706 imization space. This idea of damping the large eigenvalues to reveal the smaller ones
 707 is also used to generate a relevant set of test vectors for the construction of the least-
 708 squares minimization operator introduced in Section 3. Once the coarse correction
 709 vector is smoothed and included in W_i , the set Z_i is extended as follows

$$710 \quad (5.8) \quad Z_i = [Z_{i-1}, \mathbf{z}_i] \text{ with } \mathbf{z}_i = \frac{1}{h_{i,i}} \left(A \mathbf{w}_i - \sum_{j=1}^{i-1} h_{j,i} \cdot \mathbf{z}_j \right),$$

711 where coefficients $h_{j,i}$ result from the orthogonalization process of the new vector
 712 $A \mathbf{w}_i$. Those coefficients are stored in the squared upper triangular matrix

$$713 \quad (5.9) \quad H_i = \begin{bmatrix} & & & h_{1,p} \\ & & & \vdots \\ & H_{i-1} & & h_{i-1,i} \\ \hline 0 & \cdots & 0 & h_{i,i} \end{bmatrix} \text{ with } h_{j,i} = \begin{cases} \langle \mathbf{z}_j, \mathbf{z}_i \rangle & \text{if } j < i \\ \|\mathbf{z}_i\|_2 & \text{if } j = i \end{cases}.$$

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

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
 A initial matrix, 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

715

716 **EXAMPLE 5.1.** Let us pursue Example 4.3, where A is a 2×2 matrix, with \mathbf{v}_1
 717 and \mathbf{v}_2 its eigenvectors respectively associated with eigenvalues $|\lambda_1| < |\lambda_2|$. The in-
 718 terpolation operator P targets \mathbf{v}_1 as defined by (4.14). Let W_1 be the minimization
 719 space of dimension 1 constructed following (5.5) such that

$$720 \quad (5.10) \quad W_1 = q_{m+1}(A^2) \Pi_A(P) \mathbf{v}_1 = \frac{\lambda_1^2}{\lambda_1 + \epsilon^2 \lambda_2} (q_{m+1}(\lambda_1^2) \mathbf{v}_1 + q_{m+1}(\lambda_2^2) \epsilon \mathbf{v}_2).$$

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

$$723 \quad (5.11) \quad \mathbf{v}_1^T E_{W_1} \mathbf{v}_1 = \mathbf{v}_1^T (I - \Pi_{A^T A}(W_1)) \mathbf{v}_1 = 1 - \frac{q_{m+1}^2(\lambda_1^2) \lambda_1^2}{q_{m+1}^2(\lambda_1^2) \lambda_1^2 + q_{m+1}^2(\lambda_2^2) \epsilon^2 \lambda_2^2}.$$

724 To simplify the discussion, let us assume that the smallest eigenvector is preserved by
 725 the smoother, such that $q_{m+1}(\lambda_1^2) = 1$. The following figure illustrates the contrac-
 726 tion of \mathbf{v}_1 after applying the alternative coarse correction with respect to the pollution
 727 and the polynomial. As expected, the smoother increases the contraction and counter-
 728 balance the squared large eigenvalue λ_2 that weights the pollution $K_f = \epsilon$ when mini-
 729 mizing in euclidean norm.

730

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

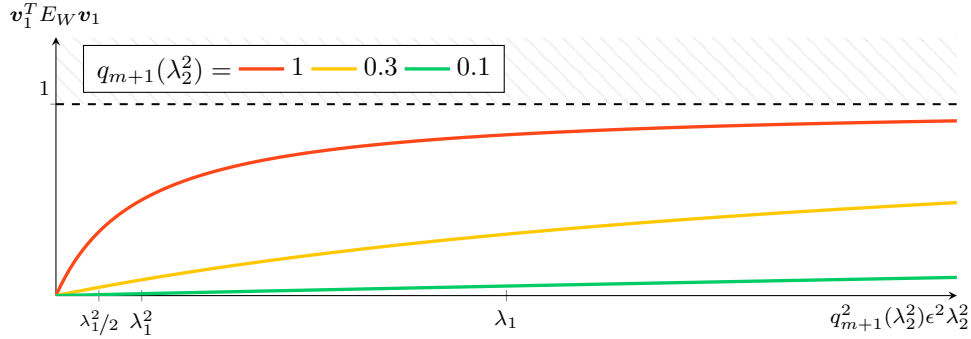


Fig. 5.1: Illustration of the contraction of a small eigenvector with respect to the pollution and the polynomial smoother

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

739

740

741

742

743

744

745

746

747

748

749

Let us first apply this new multigrid setting to the two-dimensional shifted laplacian problem associated with the stencil matrix (1.6). The size of the shifted laplacian matrix is fixed to $n = 100$. The following figures depict the number of iterations with respect to the shift kh and using either the classical or the alternative coarse correction. Recall that the matrix is the most indefinite (exact balance between negative and positive eigenvalues) when $kh = 2$, and that the near-kernel space becomes more oscillatory as kh increases. Those number of iterations are also presented with respect to the percentage τ that governs the number of selected columns of \hat{S} in the approximation of ideal interpolation. The resulting operator complexity defined by $\phi := \frac{\sum_l \text{nnz}(A_l)}{\text{nnz}(A_0)}$ for different values of τ is provided by Table 6.1.

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

750

751

752

753

754

755

756

757

758

759

760

761

762

763

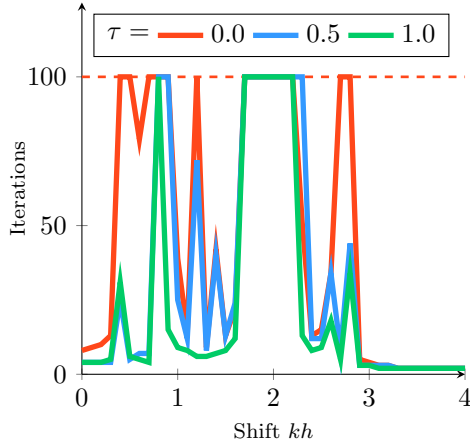
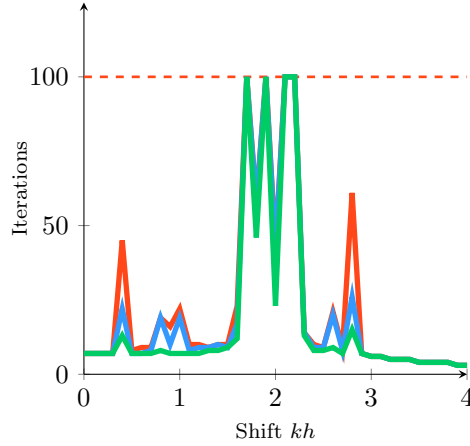
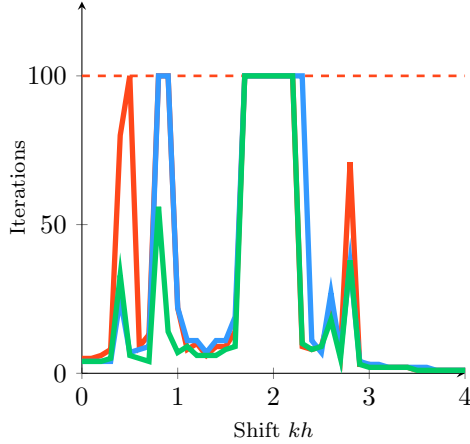
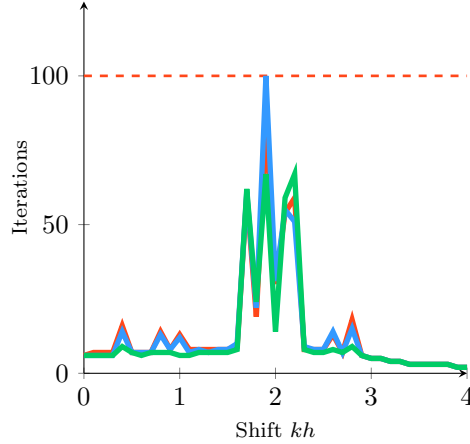
764

765

766

767

Last, the tolerance of the relative residual norm is set to 10^{-6} , and the maximal number of iterations is fixed to 100. Peak values of the standard multigrid setting on the left column denote divergence, whereas they stand for slow convergence in the alternative setting plotted on the right column. Both left figures 6.3 and 6.1 correspond to a two-level method built on the classical coarse correction respectively for $\nu = 2$ and $\nu = 4$. Whereas increasing the number of selected columns in \hat{S} for approximating the ideal interpolation operator by way of the parameter τ generally helps the convergence, the method remains likely to diverge for the reasons explained in Section 4. Still, the best setting for the classical coarse correction is naturally $\tau = 1$ and $\nu = 4$. Certain divergence scenarios that happens for $\nu = 2$ (for instance around $kh = 0.8$) are fixed by doubling the number of smoothing iterations. Doing so improves the set of test vectors in approximating the near-kernel space, and therefore leads to a better least-squares minimization coarse variable operator that decreases the pollution K_f . It remains however impossible to derive a general setting that ensures the convergence of the standard method in all cases. Both right figures 6.2 and 6.4 represent the same experiment with the alternative coarse correction. The peaks around $kh = 2$ depict slow convergence situation where the relative residual norm is stuck around 10^{-5} because of very near-zero eigenvalues. Beside those extremely

Fig. 6.1: Classical CC, $\nu = 2$ Fig. 6.2: Alternative CC, $\nu = 2$ Fig. 6.3: Classical CC, $\nu = 4$ Fig. 6.4: Alternative CC, $\nu = 4$ Fig. 6.5: Number of iterations of two-level methods with respect to kh and τ

768 indefinite cases, the method converges in all cases. We also remark that the divergence
 769 of the standard method correlate with more iterations in the alternative setting. At
 770 the cost of complexity, increasing τ or ν provides a better convergence factor.

771 6.2. Multi-level experiment on the Two Dimensional Helmholtz problem 772 with absorbing boundary conditions.

773 The following numerical experiments depict the convergence for a two dimensional
 774 Helmholtz problem using absorbing boundary conditions and with a discretization
 775 coefficient set to $kh = 0.625$ (i.e. 10 points per wavelength, where k corresponds
 776 to the wavenumber). Therefore, the discretization matrix is indefinite, complex and
 777 non-hermitian, and grows with k . As a consequence, the restriction operation is made
 778 through the transpose conjugate \hat{P}^* . Moreover, the squared matrix in the polynomial
 779 setting is replaced by A^*A . Also note that those numerical experiments result from
 780 the alternative coarse correction only, and that Z^T is replaced by Z^* in (5.4). The
 781 first benchmark illustrated by Figure 6.6 exposes the convergence of the method by
 782 fixing the number of selected column of \hat{S} to the maximum (i.e., $\tau = 1$). Each curve
 783 corresponds to a method following its number of levels. The y -axis corresponds to

784 the number of iterations, while the wavenumber varies along the x -axis. The number of
 785 iterations is constant until the fourth level. The number of iterations of both the
 five-level and six-level methods increase with the wavenumber.

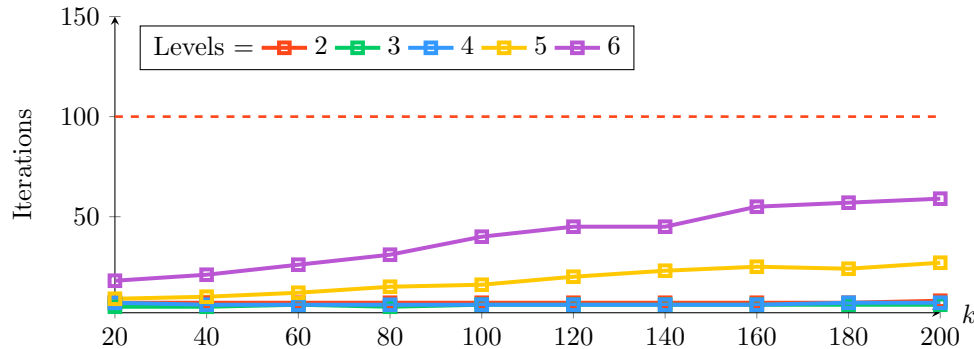


Fig. 6.6: Number of iterations following the wavenumber k , $\nu = 2$, $\tau = 1$

786

787 While setting $\tau = 1$ enables the method to converge almost constantly up to five
 788 levels, the operator complexity is too high for practical implementation. Therefore,
 789 the second benchmark exposes the number of iterations of a two-level method with
 respect to the parameter τ . Figure 6.7 shows that the plain least-squares minimization

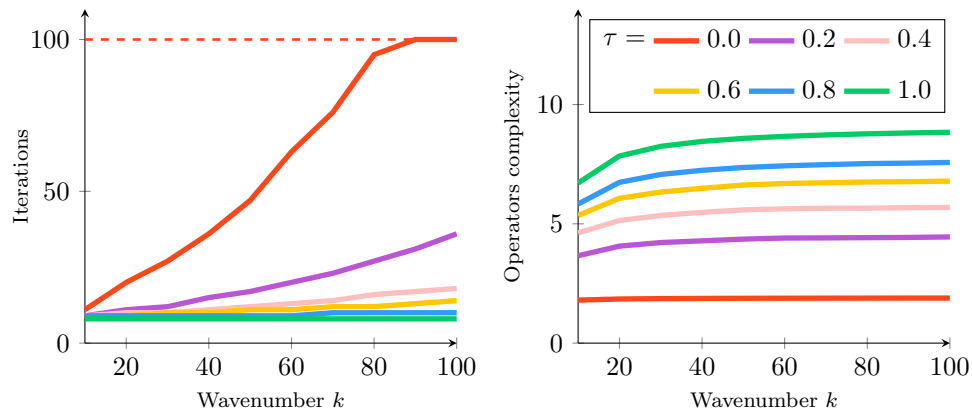


Fig. 6.7: Two-level method with alternative coarse correction - number of iterations
 and operators complexity with respect to k and τ , $\nu = 2$

790

791 operator (i.e. $\tau = 0$) is not a suitable choice as k is growing. Even though larger sub-
 792 spaces $\hat{S}X_i$ in the approximation of the ideal interpolation operator yields denser
 793 matrices, the number of iterations tends to size independence as τ grows. A trade-off
 794 between convergence and complexity may be possible depending on the problem-size.
 795 More generally, Figure 6.7 reveals how important is the role that plays the ideal
 796 approximation step in the convergence. A better sparsification strategy is a topic of
 797 further research.

798

799 **7. Conclusions.** Indefinite and oscillatory problems are difficult for multigrid
 methods. The negative eigenvalues require an adapted smoother, and the interpolator

800 should capture the oscillatory near-kernel space. More importantly, the coarse correc-
 801 tion should be adapted to the indefiniteness of the initial matrix, which does not define
 802 a norm. The normal equation polynomial smoother is designed to target a desired
 803 proportion of eigenvalues according to their amplitude, and the range of our inter-
 804 polator offers a good approximation of the near-kernel space despite its oscillations.
 805 The alternative coarse correction space proposed in the paper permits to minimize the
 806 global residual in a proper norm for indefinite problems, in a space approximating the
 807 set of smallest eigenvectors known to be difficult for most iterative methods. Finding
 808 a better trade-off between sparsity and accuracy of interpolation, and constructing
 809 a polynomial without resorting to normal equations will be important points in our
 810 future investigations.

811

REFERENCES

- 812 [1] M. F. ADAMS, M. BREZINA, J. J. HU, AND R. S. TUMINARO, *Parallel multigrid smoothing:*
 813 *polynomial versus gauss–seidel*, Journal of Computational Physics, 188 (2003), pp. 593–
 814 610.
- 815 [2] A. H. BAKER, R. D. FALGOUT, T. V. KOLEV, AND U. M. YANG, *Multigrid smoothers for ultra-*
 816 *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)
 817 <https://doi.org/10.1137/100798806>.
- 818 [3] A. BRANDT, J. BRANNICK, K. KAHL, AND I. LIVSHITS, *Bootstrap amg*, SIAM Journal of Scien-
 819 tific Computing, 33 (2011), pp. 612–632, <https://doi.org/10.1137/090752973>.
- 820 [4] L. I. BRANDT A., *Wave-ray multigrid method for standing wave equations.*, ETNA. Electronic
 821 Transactions on Numerical Analysis [electronic only], 6 (1997), pp. 162–181, [http://eudml](http://eudml.org/doc/119506).
 822 [org/doc/119506](http://eudml.org/doc/119506).
- 823 [5] M. BREZINA, R. FALGOUT, S. MACLACHLAN, T. MANTEUFFEL, S. MCCORMICK, AND J. RUGE,
 824 *Adaptive smoothed aggregation (asa)*, SIAM Journal on Scientific Computing, 25 (2004),
 825 pp. 1896–1920, <https://doi.org/10.1137/S1064827502418598>, [https://doi.org/10.1137/](https://doi.org/10.1137/S1064827502418598)
 826 [S1064827502418598](https://doi.org/10.1137/S1064827502418598), <https://arxiv.org/abs/https://doi.org/10.1137/S1064827502418598>.
- 827 [6] W. BRIGGS, V. HENSON, AND S. MCCORMICK, *A Multigrid Tutorial, 2nd Edition*, 01 2000.
- 828 [7] O. COULAUD, L. GIRAUD, P. RAMET, AND X. VASSEUR, *Deflation and augmentation techniques*
 829 *in krylov subspace methods for the solution of linear systems*, 2013, [https://arxiv.org/abs/](https://arxiv.org/abs/1303.5692)
 830 [1303.5692](https://arxiv.org/abs/1303.5692).
- 831 [8] V. DWARKA AND C. VUIK, *Stand-alone multigrid for helmholtz revisited: Towards convergence*
 832 *using standard components*, 2023, <https://arxiv.org/abs/2308.13476>.
- 833 [9] P. EK, M. BREZINA, AND J. MANDEL, *Convergence of algebraic multigrid based on smoothed*
 834 *aggregation*, Computing, 56 (1998), <https://doi.org/10.1007/s002110000226>.
- 835 [10] O. G. ERNST AND M. J. GANDER, *Why it is difficult to solve helmholtz problems with classical*
 836 *iterative methods*, (2010).
- 837 [11] R. D. FALGOUT, *An introduction to algebraic multigrid*, Computing in Science and Engineering,
 838 vol. 8, no. 6, November 1, 2006, pp. 24–33, (2006), <https://www.osti.gov/biblio/897960>.
- 839 [12] R. D. FALGOUT AND P. S. VASSILEVSKI, *On generalizing the amg framework*, SIAM J. NUMER.
 840 ANAL., 42 (2003), pp. 1669–1693.
- 841 [13] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press,
 842 1 ed., Apr. 1991, <https://doi.org/10.1017/CBO9780511840371>, [https://www.cambridge](https://www.cambridge.org/core/product/identifier/9780511840371/type/book).
 843 [org/core/product/identifier/9780511840371/type/book](https://www.cambridge.org/core/product/identifier/9780511840371/type/book) (accessed 2024-05-24).
- 844 [14] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 2 ed.,
 845 Oct. 2012, <https://doi.org/10.1017/CBO9781139020411>, [https://www.cambridge.org/](https://www.cambridge.org/highereducation/product/9781139020411/book)
 846 [highereducation/product/9781139020411/book](https://www.cambridge.org/highereducation/product/9781139020411/book) (accessed 2024-05-24).
- 847 [15] J. K. KRAUS, P. S. VASSILEVSKI, AND L. T. ZIKATANOV, *Polynomial of best uniform approximation*
 848 *to x^{-1} and smoothing in two-level methods*, 2012, <https://arxiv.org/abs/1002.1859>.
- 849 [16] L. LIN, Y. SAAD, AND C. YANG, *Approximating spectral densities of large matrices*, SIAM Re-
 850 view, 58 (2016), pp. 34–65, <https://doi.org/10.1137/130934283>, [https://doi.org/10.1137/](https://doi.org/10.1137/130934283)
 851 [130934283](https://doi.org/10.1137/130934283), <https://arxiv.org/abs/https://doi.org/10.1137/130934283>.
- 852 [17] I. LIVSHITS, *A scalable multigrid method for solving indefinite helmholtz equations with constant*
 853 *wave numbers*, Numerical Linear Algebra with Applications, 21 (2014), [https://doi.org/](https://doi.org/10.1002/nla.1926)
 854 [10.1002/nla.1926](https://doi.org/10.1002/nla.1926).
- 855 [18] I. LIVSHITS, *Multiple galerkin adaptive algebraic multigrid algorithm for the helmholtz equa-*
 856

- 857 *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>.
- 858
- 859
- 860 [19] S. MACLACHLAN AND Y. SAAD, *A greedy strategy for coarse-grid selection*, SIAM Journal on
861 Scientific Computing, 29 (2007), pp. 1825–1853, <https://doi.org/10.1137/060654062>, <https://doi.org/10.1137/060654062>, <https://arxiv.org/abs/https://doi.org/10.1137/060654062>.
- 862
- 863 [20] L. OLSON AND J. SCHRODER, *Smoothed aggregation for helmholtz problems*, Numerical Linear
864 Algebra with Applications, 17 (2010), pp. 361 – 386, <https://doi.org/10.1002/nla.686>.
- 865 [21] L. N. OLSON, J. B. SCHRODER, AND R. S. TUMINARO, *A general interpolation strategy
866 for algebraic multigrid using energy minimization*, SIAM Journal on Scientific Comput-
867 ing, 33 (2011), pp. 966–991, <https://doi.org/10.1137/100803031>, [https://doi.org/10.1137/](https://doi.org/10.1137/100803031)
868 [100803031](https://doi.org/10.1137/100803031), <https://arxiv.org/abs/https://doi.org/10.1137/100803031>.
- 869 [22] E. PAROLIN, D. HUYBRECHS, AND A. MOIOLA, *Stable approximation of helmholtz solutions
870 by evanescent plane waves*, 2022, <https://doi.org/10.48550/ARXIV.2202.05658>, [https://](https://arxiv.org/abs/2202.05658)
871 arxiv.org/abs/2202.05658.
- 872 [23] J. W. RUGE AND K. STÜBEN, *4. Algebraic Multigrid*, [https://doi.org/10.1137/1.9781611971057.](https://doi.org/10.1137/1.9781611971057.ch4)
873 [ch4](https://doi.org/10.1137/1.9781611971057.ch4), <https://epubs.siam.org/doi/abs/10.1137/1.9781611971057.ch4>, [https://arxiv.org/](https://arxiv.org/abs/https://epubs.siam.org/doi/pdf/10.1137/1.9781611971057.ch4)
874 [abs/https://epubs.siam.org/doi/pdf/10.1137/1.9781611971057.ch4](https://epubs.siam.org/doi/pdf/10.1137/1.9781611971057.ch4).
- 875 [24] 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)
876 [IterMethBook_2ndEd.pdf](https://www-users.cs.umn.edu/~saad/IterMethBook_2ndEd.pdf).
- 877 [25] 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)
878 [2006/am63.pdf](https://math.mit.edu/classes/18.086/2006/am63.pdf).
- 879 [26] K. STÜBEN, *Algebraic multigrid (amg). an introduction with applications*, (1999).
- 880 [27] P. VANEK, J. MANDEL, AND M. BREZINA, *Algebraic multigrid by smoothed aggregation for
881 second and fourth order elliptic problems*, tech. report, USA, 1995.
- 882 [28] B. M. VANVEK PETR AND M. JAN, *Convergence of algebraic multigrid based on smoothed
883 aggregation*, (2001), <https://doi.org/10.1007/s211-001-8015-y>.
- 884 [29] T. U. ZAMAN, S. P. MACLACHLAN, L. N. OLSON, AND M. WEST, *Coarse-grid selection
885 using simulated annealing*, Journal of Computational and Applied Mathematics, 431
886 (2023), p. 115263, <https://doi.org/https://doi.org/10.1016/j.cam.2023.115263>, [https://](https://doi.org/https://doi.org/10.1016/j.cam.2023.115263)
887 [www.sciencedirect.com/science/article/pii/S0377042723002078](https://doi.org/https://doi.org/10.1016/j.cam.2023.115263).