On the Asymptotic Optimality of Spectral Coarse Spaces

On the Asymptotic Optimality of Spectral Coarse Spaces

Gabriele Ciaramella and Tommaso Vanzan

1 Introduction

The goal of this work is to study the asymptotic optimality of spectral coarse spaces for two-level iterative methods. In particular, we consider a linear system $A\mathbf{u} = \mathbf{f}$, where $A \in \mathbb{R}^{n\times n}$ and $\mathbf{f} \in \mathbb{R}^n$, and a two-level method that, given an iterate $\mathbf{u}^k$, computes the new vector $\mathbf{u}^{k+1}$ as

$$\mathbf{u}^{k+1/2} = G\mathbf{u}^k + M^{-1}\mathbf{f}, \qquad \text{(smoothing step)} \qquad (1)$$

$$\mathbf{u}^{k+1} = \mathbf{u}^{k+1/2} + PA_c^{-1}R(\mathbf{f} - A\mathbf{u}^{k+1/2}). \qquad \text{(coarse correction)} \qquad (2)$$

The smoothing step (1) is based on the splitting $A = M - N$, where M is the preconditioner, and $G = M^{-1}N$ the iteration matrix. The correction step (2) is characterized by prolongation and restriction matrices $P \in \mathbb{R}^{n\times m}$ and $R = P^\top$, and a coarse matrix $A_c = RAP$. The columns of P are linearly independent vectors spanning the coarse space $V_c := \text{span}\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$. The convergence of the one-level iteration (1) is characterized by the eigenvalues of G, λ_j, $j = 1, \dots, n$ (sorted in descending order by magnitude). The convergence of the two-level iteration (1)-(2) depends on the spectrum of the iteration matrix T, obtained by substituting (1) into (2) and rearranging terms:

$$T = [I - P(RAP)^{-1}RA]G. \qquad (3)$$

The goal of this short paper is to answer, though partially, the fundamental question: **given an integer m, what is the coarse space of dimension m which minimizes the spectral radius $\rho(T)$?** Since step (2) aims at correcting the error components that the smoothing step (1) is not able to reduce (or eliminate), it is intuitive to think that an optimal coarse space V_c is obtained by defining $\mathbf{p}_j$ as the eigenvectors

Gabriele Ciaramella
Politecnico di Milano e-mail: gabriele.ciaramella@polimi.it

Tommaso Vanzan
CSQI Chair, EPFL Lausanne e-mail: tommaso.vanzan@epfl.ch

of G corresponding to the m largest (in modulus) eigenvalues. We call such a V_c *spectral coarse space*. Following the idea of correcting the 'badly converging' modes of G, several papers proposed new, and in some sense optimal, coarse spaces. In the context of domain decomposition methods, we refer, e.g., to [2, 3, 4], where efficient coarse spaces have been designed for parallel, restricted additive and additive Schwarz methods. In the context of multigrid methods, it is worth mentioning the work [6], where the interpolation weights are optimized using an approach based on deep-neural networks. Fundamental results are presented in [7]: for a symmetric A, it is proved that the coarse space of size m that minimizes the energy norm of T, namely $\|T\|_A$, is the span of the m eigenvectors of $\overline{M}A$ corresponding to the m lowest eigenvalues. Here, $\overline{M} := M^{-1}+M^{-\top}-M^{-\top}AM^{-1}$ is symmetric and assumed positive definite. If M is symmetric, a direct calculation gives $\overline{M}A = 2M^{-1}A - (M^{-1}A)^2$. Using that $M^{-1}A = I - G$, one can show that the m eigenvectors associated to the lowest m eigenvalues of $\overline{M}A$ correspond to the m largest modes of G. Hence, the optimal coarse space proposed in [7] is a spectral coarse space. The sharp result of [7] provides a concrete optimal choice of V_c minimizing $\|T\|_A$. This is generally an upper bound for the asymptotic convergence factor $\rho(T)$. As we will see in Section 2, choosing the spectral coarse space, one gets $\rho(T) = |\lambda_{m+1}|$. The goal of this work is to show that this is not necessarily the optimal asymptotic convergence factor. In Section 2, we perform a detailed optimality analysis for the case $m = 1$. The asymptotic optimality of coarse spaces for $m \geq 1$ is studied numerically in Section 3. Interestingly, we will see that by optimizing $\rho(T)$ one constructs coarse spaces that lead to preconditioned matrices with better condition numbers.

2 A perturbation approach

Let G be diagonalizable with eigenpairs $(\lambda_j, \mathbf{v}_j)$, $j = 1, \ldots, n$. Suppose that $\mathbf{v}_j$ are also eigenvectors of A: $A\mathbf{v}_j = \widetilde{\lambda}_j \mathbf{v}_j$. Concrete examples where these hypotheses are fulfilled are given in Section 3. Assume that $\operatorname{rank} P = m$ ($\dim V_c = m$). For any eigenvector $\mathbf{v}_j$, we can write the vector $T\mathbf{v}_j$ as

$$T\mathbf{v}_j = \sum_{\ell=1}^{n} \widetilde{t}_{j,\ell}\mathbf{v}_\ell, \ j = 1, \ldots, n. \tag{4}$$

If we denote by $\widetilde{T} \in \mathbb{R}^{n\times n}$ the matrix of entries $\widetilde{t}_{j,\ell}$, and define $V := [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, then (4) becomes $TV = V\widetilde{T}^\top$. Since G is diagonalizable, V is invertible, and thus T and $\widetilde{T}^\top$ are similar. Hence, T and $\widetilde{T}$ have the same spectrum. We can now prove the following lemma.

Lemma 1 (Characterization of $\widetilde{T}$)

Given an index $\widetilde{m} \geq m$ and assume that $V_c := \operatorname{span}\{\mathbf{p}_1, \ldots, \mathbf{p}_m\}$ satisfies

$$V_c \subseteq \operatorname{span}\{\mathbf{v}_j\}_{j=1}^{\widetilde{m}} \text{ and } V_c \cap \{\mathbf{v}_j\}_{j=\widetilde{m}+1}^{n} = \{0\}. \tag{5}$$

Then, it holds that

$$\widetilde{T} = \begin{bmatrix} \widetilde{T}_{\widetilde{m}} & 0 \\ X & \Lambda_{\widetilde{m}} \end{bmatrix}, \qquad \begin{array}{l} \Lambda_{\widetilde{m}} = \operatorname{diag}(\lambda_{\widetilde{m}+1}, \ldots, \lambda_n), \\ \widetilde{T}_{\widetilde{m}} \in \mathbb{R}^{\widetilde{m}\times\widetilde{m}}, X \in \mathbb{R}^{(n-\widetilde{m})\times\widetilde{m}}. \end{array} \tag{6}$$

Proof The hypothesis (5) guarantees that span $\{\mathbf{v}_j\}_{j=1}^{\widetilde{m}}$ is invariant under the action of T. Hence, $T\mathbf{v}_j \in \text{span}\,\{\mathbf{v}_j\}_{j=1}^{\widetilde{m}}$ for $j = 1, \ldots, \widetilde{m}$, and, using (4), one gets that $\widetilde{t}_{j,\ell} = 0$ for $j = 1, \ldots, \widetilde{m}$ and $\ell = \widetilde{m}+1, \ldots, n$. Now, consider any $j > \widetilde{m}$. A direct calculation using (4) reveals that $T\mathbf{v}_j = G\mathbf{v}_j - P(RAP)^{-1}RAG\mathbf{v}_j = \lambda_j\mathbf{v}_j - \sum_{\ell=1}^{\widetilde{m}} x_{j-\widetilde{m},\ell}\mathbf{v}_\ell$, where $x_{i,k}$ are the elements of $X \in \mathbb{R}^{(n-\widetilde{m})\times\widetilde{m}}$. Hence, the structure (6) follows. □

Notice that, if (5) holds, then Lemma 1 allows us to study the properties of T using the matrix $\widetilde{T}$ and its structure (6), and hence $\widetilde{T}_{\widetilde{m}}$.

Let us now turn to the questions posed in Section 1. Assume that $\mathbf{p}_j = \mathbf{v}_j$, $j = 1, \ldots, m$, namely $V_c = \text{span}\,\{\mathbf{v}_j\}_{j=1}^{m}$. In this case, (5) holds with $\widetilde{m} = m$, and a simple argument[1] leads to $\widetilde{T}_{\widetilde{m}} = 0$, $\widetilde{T} = \begin{bmatrix} 0 & 0 \\ X & \Lambda_{\widetilde{m}} \end{bmatrix}$. The spectrum of $\widetilde{T}$ is $\{0, \lambda_{m+1}, \ldots, \lambda_n\}$. This means that $V_c \subset \ker T$ and $\rho(T) = |\lambda_{m+1}|$. Let us now perturb the coarse space V_c using the eigenvector $\mathbf{v}_{m+1}$, that is $V_c(\varepsilon) := \text{span}\,\{\mathbf{v}_j + \varepsilon\,\mathbf{v}_{m+1}\}_{j=1}^{m}$. Clearly, $\dim V_c(\varepsilon) = m$ for any $\varepsilon \in \mathbb{R}$. In this case, (5) holds with $\widetilde{m} = m + 1$ and $\widetilde{T}$ becomes

$$\widetilde{T}(\varepsilon) = \begin{bmatrix} \widetilde{T}_{\widetilde{m}}(\varepsilon) & 0 \\ X(\varepsilon) & \Lambda_{\widetilde{m}} \end{bmatrix}, \tag{7}$$

where we make explicit the dependence on ε. Notice that $\varepsilon = 0$ clearly leads to $\widetilde{T}_{\widetilde{m}}(0) = \operatorname{diag}(0, \ldots, 0, \lambda_{m+1}) \in \mathbb{R}^{\widetilde{m}\times\widetilde{m}}$, and we are back to the unperturbed case with $\widetilde{T}(0) = \widetilde{T}$ having spectrum $\{0, \lambda_{m+1}, \ldots, \lambda_n\}$. Now, notice that $\min_{\varepsilon\in\mathbb{R}} \rho(\widetilde{T}(\varepsilon)) \le \rho(\widetilde{T}(0)) = |\lambda_{m+1}|$. Thus, it is natural to ask the question: is this inequality strict? Can one find an $\widetilde{\varepsilon} \neq 0$ such that $\rho(\widetilde{T}(\widetilde{\varepsilon})) = \min_{\varepsilon\in\mathbb{R}} \rho(\widetilde{T}(\varepsilon)) < \rho(\widetilde{T}(0))$ holds? If the answer is positive, then we can conclude that choosing the coarse vectors equal to the dominating eigenvectors of G is not an optimal choice. The next key result shows that, in the case $m = 1$, the answer is positive.

Theorem 1 (Perturbation of V_c)

Let $(\mathbf{v}_1, \lambda_1)$, $(\mathbf{v}_2, \lambda_2)$ and $(\mathbf{v}_3, \lambda_3)$ be three real eigenpairs of G, $G\mathbf{v}_j = \lambda_j\mathbf{v}_j$ such that with $0 < |\lambda_3| < |\lambda_2| \le |\lambda_1|$ and $\|\mathbf{v}_j\|_2 = 1$, $j = 1, 2$. Denote by $\widetilde{\lambda}_j \in \mathbb{R}$ the eigenvalues of A corresponding to $\mathbf{v}_j$, and assume that $\widetilde{\lambda}_1\widetilde{\lambda}_2 > 0$. Define $V_c :=$ span $\{\mathbf{v}_1 + \varepsilon\mathbf{v}_2\}$ with $\varepsilon \in \mathbb{R}$, and $\gamma := \mathbf{v}_1^\top\mathbf{v}_2 \in [-1, 1]$. Then

(A) *The spectral radius of $\widetilde{T}(\varepsilon)$ is $\rho(\widetilde{T}(\varepsilon)) = \max\{|\lambda(\varepsilon, \gamma)|, |\lambda_3|\}$, where*

$$\lambda(\varepsilon, \gamma) = \frac{\lambda_1\widetilde{\lambda}_2\varepsilon^2 + \gamma(\lambda_1\widetilde{\lambda}_2 + \lambda_2\widetilde{\lambda}_1)\varepsilon + \lambda_2\widetilde{\lambda}_1}{\widetilde{\lambda}_2\varepsilon^2 + \gamma(\widetilde{\lambda}_1 + \widetilde{\lambda}_2)\varepsilon + \widetilde{\lambda}_1}. \tag{8}$$

[1] Let $\mathbf{v}_j$ be an eigenvector of A with $j \in \{1, \ldots, m\}$. Denote by $\mathbf{e}_j \in \mathbb{R}^n$ the jth canonical vector. Since $P\mathbf{e}_j = \mathbf{v}_j$, $RAP\mathbf{e}_j = RA\mathbf{v}_j$. This is equivalent to $\mathbf{e}_j = (RAP)^{-1}RA\mathbf{v}_j$, which gives $T\mathbf{v}_j = \lambda_j(\mathbf{v}_j - P(RAP)^{-1}RA\mathbf{v}_j) = \lambda_j(\mathbf{v}_j - P\mathbf{e}_j) = 0$.

(B) *Let $\gamma = 0$. If $\lambda_1 > \lambda_2 > 0$ or $0 > \lambda_2 > \lambda_1$, then* $\min_{\varepsilon \in \mathbb{R}} \rho(\widetilde{T}(\varepsilon)) = \rho(\widetilde{T}(0))$.

(C) *Let $\gamma = 0$, If $\lambda_2 > 0 > \lambda_1$ or $\lambda_1 > 0 > \lambda_2$, then there exists an $\widetilde{\varepsilon} \neq 0$ such that* $\rho(\widetilde{T}(\widetilde{\varepsilon})) = |\lambda_3| = \min_{\varepsilon \in \mathbb{R}} \rho(\widetilde{T}(\varepsilon)) < \rho(\widetilde{T}(0))$.

(D) *Let $\gamma \neq 0$. If $\lambda_1 > \lambda_2 > 0$ or $0 > \lambda_2 > \lambda_1$, then there exists an $\widetilde{\varepsilon} \neq 0$ such that* $|\lambda(\widetilde{\varepsilon}, \gamma)| < |\lambda_2|$ *and hence* $\rho(\widetilde{T}(\widetilde{\varepsilon})) = \max\{|\lambda(\widetilde{\varepsilon}, \gamma)|, |\lambda_3|\} < \rho(\widetilde{T}(0))$.

(E) *Let $\gamma \neq 0$. If $\lambda_2 > 0 > \lambda_1$ or $\lambda_1 > 0 > \lambda_2$, then there exists an $\widetilde{\varepsilon} \neq 0$ such that* $\rho(\widetilde{T}(\widetilde{\varepsilon})) = |\lambda_3| = \min_{\varepsilon \in \mathbb{R}} \rho(\widetilde{T}(\varepsilon)) < \rho(\widetilde{T}(0))$.

Proof Since $m = 1$, a direct calculation allows us to compute the matrix

$$\widetilde{T}_{\widetilde{m}}(\varepsilon) = \begin{bmatrix} \lambda_1 - \frac{\lambda_1 \widetilde{\lambda}_1 (1+\varepsilon\gamma)}{g} & -\varepsilon \frac{\lambda_1 \widetilde{\lambda}_1 (1+\varepsilon\gamma)}{g} \\ -\frac{\lambda_2 \widetilde{\lambda}_2 (\varepsilon+\gamma)}{g} & \lambda_2 - \frac{(\varepsilon \lambda_2 \widetilde{\lambda}_2)(\varepsilon+\gamma)}{g} \end{bmatrix},$$

where $g = \widetilde{\lambda}_1 + \varepsilon\gamma[\widetilde{\lambda}_1 + \widetilde{\lambda}_2] + \varepsilon^2 \widetilde{\lambda}_2$. The spectrum of this matrix is $\{0, \lambda(\varepsilon, \gamma)\}$, with $\lambda(\varepsilon, \gamma)$ given in (8). Hence, point (A) follows recalling (7).

To prove points (B), (C), (D) and (E) we use some properties of the map $\varepsilon \mapsto \lambda(\varepsilon, \gamma)$. First, we notice that

$$\lambda(0, \gamma) = \lambda_2, \quad \lim_{\varepsilon \to \pm\infty} \lambda(\varepsilon, \gamma) = \lambda_1, \quad \lambda(\varepsilon, \gamma) = \lambda(-\varepsilon, -\gamma). \tag{9}$$

Second, the derivative of $\lambda(\varepsilon, \gamma)$ with respect to ε is

$$\frac{d\lambda(\varepsilon, \gamma)}{d\varepsilon} = \frac{(\lambda_1 - \lambda_2)\widetilde{\lambda}_1 \widetilde{\lambda}_2 (\varepsilon^2 + 2\varepsilon/\gamma + 1)\gamma}{(\widetilde{\lambda}_2 \varepsilon^2 + \gamma(\widetilde{\lambda}_1 + \widetilde{\lambda}_2)\varepsilon + \widetilde{\lambda}_1)^2}. \tag{10}$$

Because of $\lambda(\varepsilon, \gamma) = \lambda(-\varepsilon, -\gamma)$ in (9), we can assume without loss of generality that $\gamma \geq 0$.

Let us now consider the case $\gamma = 0$. In this case, the derivative (10) becomes $\frac{d\lambda(\varepsilon,0)}{d\varepsilon} = \frac{(\lambda_1 - \lambda_2)\widetilde{\lambda}_1 \widetilde{\lambda}_2 2\varepsilon}{(\widetilde{\lambda}_2 \varepsilon^2 + \widetilde{\lambda}_1^2)^2}$. Moreover, since $\lambda(\varepsilon, 0) = \lambda(-\varepsilon, 0)$ we can assume that $\varepsilon \geq 0$.

Case (B). If $\lambda_1 > \lambda_2 > 0$, then $\frac{d\lambda(\varepsilon,0)}{d\varepsilon} > 0$ for all $\varepsilon > 0$. Hence, $\varepsilon \mapsto \lambda(\varepsilon, 0)$ is monotonically increasing, $\lambda(\varepsilon, 0) \geq 0$ for all $\varepsilon > 0$ and, thus, the minimum of $\varepsilon \mapsto |\lambda(\varepsilon, 0)|$ is attained at $\varepsilon = 0$ with $|\lambda(0, 0)| = |\lambda_2| > |\lambda_3|$, and the result follows. Analogously, if $0 > \lambda_2 > \lambda_1$, then $\frac{d\lambda(\varepsilon,0)}{d\varepsilon} < 0$ for all $\varepsilon > 0$. Hence, $\varepsilon \mapsto \lambda(\varepsilon, 0)$ is monotonically decreasing, $\lambda(\varepsilon, 0) < 0$ for all $\varepsilon > 0$ and the minimum of $\varepsilon \mapsto |\lambda(\varepsilon, 0)|$ is attained at $\varepsilon = 0$.

Case (C). If $\lambda_1 > 0 > \lambda_2$, then $\frac{d\lambda(\varepsilon,0)}{d\varepsilon} > 0$ for all $\varepsilon > 0$. Hence, $\varepsilon \mapsto \lambda(\varepsilon, 0)$ is monotonically increasing and such that $\lambda(0, 0) = \lambda_2 < 0$ and $\lim_{\varepsilon \to \infty} \lambda(\varepsilon, 0) = \lambda_1 > 0$. Thus, the continuity of the map $\varepsilon \mapsto \lambda(\varepsilon, 0)$ guarantees the existence of an $\widetilde{\varepsilon} > 0$ such that $\lambda(\widetilde{\varepsilon}, 0) = 0$. Analogously, if $\lambda_2 > 0 > \lambda_1$, then $\frac{d\lambda(\varepsilon,0)}{d\varepsilon} < 0$ for all $\varepsilon > 0$ and the result follows by the continuity of $\varepsilon \mapsto \lambda(\varepsilon, 0)$.

Let us now consider the case $\gamma > 0$. The sign of $\frac{d\lambda(\varepsilon,\gamma)}{d\varepsilon}$ is affected by the term $f(\varepsilon) := \varepsilon^2 + 2\varepsilon/\gamma + 1$, which appears at the numerator of (10). The function $f(\varepsilon)$

is strictly convex, attains its minimum at $\varepsilon = -\frac{1}{\gamma}$, and is negative in $(\bar{\varepsilon}_1, \bar{\varepsilon}_2)$ and positive in $(-\infty, \bar{\varepsilon}_1) \cup (\bar{\varepsilon}_2, \infty)$, with $\bar{\varepsilon}_1, \bar{\varepsilon}_2 = -\frac{1 \mp \sqrt{1-\gamma^2}}{\gamma}$.

Case (D). If $\lambda_1 > \lambda_2 > 0$, then $\frac{d\lambda(\varepsilon,\gamma)}{d\varepsilon} > 0$ for all $\varepsilon > \bar{\varepsilon}_2$. Hence, $\frac{d\lambda(0,\gamma)}{d\varepsilon} > 0$, which means that there exists an $\widetilde{\varepsilon} < 0$ such that $|\lambda(\widetilde{\varepsilon}, \gamma)| < |\lambda(0, \gamma)| = |\lambda_2|$. The case $0 > \lambda_2 > \lambda_1$ follows analogously.

Case (E). If $\lambda_1 > 0 > \lambda_2$, then $\frac{d\lambda(\varepsilon,\gamma)}{d\varepsilon} > 0$ for all $\varepsilon > 0$. Hence, by the continuity of $\varepsilon \mapsto \lambda(\varepsilon, \gamma)$ (for $\varepsilon \geq 0$) there exists an $\widetilde{\varepsilon} > 0$ such that $\lambda(\widetilde{\varepsilon}, \gamma) = 0$. The case $\lambda_2 > 0 > \lambda_1$ follows analogously. □

Theorem 1 and its proof say that, if the two eigenvalues λ_1 and λ_2 have opposite signs (but they could be equal in modulus), then it is always possible to find an $\varepsilon \neq 0$ such that the coarse space $V_c := \text{span}\{\mathbf{v}_1 + \varepsilon \mathbf{v}_2\}$ leads to a faster method than $V_c := \text{span}\{\mathbf{v}_1\}$, even though both are one-dimensional subspaces. In addition, if $\lambda_3 \neq 0$ the former leads to a two-level operator T with a larger kernel than the one corresponding to the latter. The situation is completely different if λ_1 and λ_2 have the same sign. In this case, the orthogonality parameter γ is crucial. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal ($\gamma = 0$), then one cannot improve the effect of $V_c := \text{span}\{\mathbf{v}_1\}$ by a simple perturbation using $\mathbf{v}_2$. However, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are not orthogonal ($\gamma \neq 0$), then one can still find an $\varepsilon \neq 0$ such that $\rho(\widetilde{T}(\varepsilon)) < \rho(\widetilde{T}(0))$.

Notice that, if $|\lambda_3| = |\lambda_2|$, Theorem 1 shows that one cannot obtain a $\rho(T)$ smaller than $|\lambda_2|$ using a one-dimensional perturbation. However, if one optimizes the entire coarse space V_c (keeping m fixed), then one can find coarse spaces leading to better contraction factor of the two-level iteration, even though $|\lambda_3| = |\lambda_2|$. This is shown in the next section.

3 Optimizing the coarse-space functions

Consider the elliptic problem

$$-\Delta u + c\,(\partial_x u + \partial_y u) = f \ \text{ in } \Omega = (0,1)^2, \quad u = 0 \ \text{ on } \partial\Omega. \tag{11}$$

Using a uniform grid of size h, the standard second-order finite-difference scheme for the Laplace operator and the central difference approximation for the advection terms, problem (11) becomes $A\mathbf{u} = \mathbf{f}$, where A has constant and positive diagonal entries, $D = \text{diag}(A) = 4/h^2 I$. A simple calculation shows that, if $c \geq 0$ satisfies $c \leq 2/h$, then the eigenvalues of A are real. The eigenvectors of A are orthogonal if $c = 0$ and non-orthogonal if $c > 0$.

One of the most used smoothers for (11) is the damped Jacobi method: $\mathbf{u}^{k+1} = \mathbf{u}^k + \omega D^{-1}(\mathbf{f} - A\mathbf{u}^k)$, where $\omega \in (0, 1]$ is a damping parameter. The corresponding iteration matrix is $G = I - \omega D^{-1} A$. Since $D = 4/h^2 I$, the matrices A and G have the same eigenvectors. For $c = 0$, it is possible to show that, if $\omega = 1$ (classical Jacobi iteration), then the nonzero eigenvalues of G have positive and negative signs, while if $\omega = 1/2$, the eigenvalues of G are all positive. Hence, the chosen model problem allows us to work in the theoretical framework of Section 2.

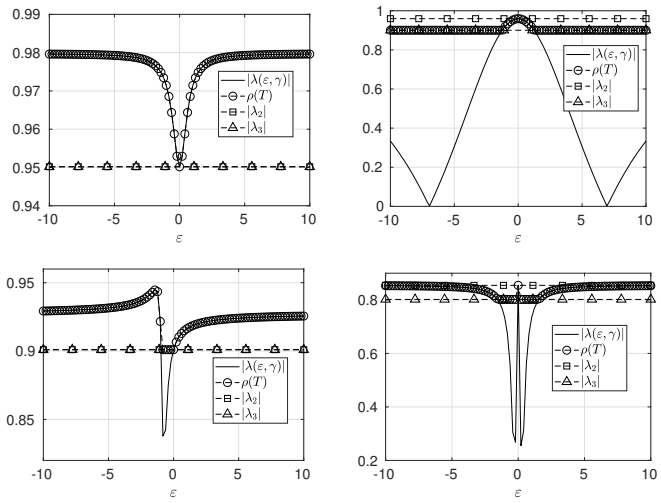

Fig. 1: Behavior of $|\lambda(\varepsilon,\gamma)|$ and $\rho(T(\varepsilon))$ as functions of ε for different c and γ. Top left panel: $c = 0$, $\omega = 1/2$; top right panel: $c = 0$, $\omega = 1$; bottom left panel: $c = 10$, $\omega = 1/2$; bottom right panel: $c = 10$, $\omega = 1$.

To validate numerically Theorem 1, we set $h = 1/10$ and consider $V_c := \{\mathbf{v}_1 + \varepsilon\mathbf{v}_2\}$. Figure 1 shows the dependence of $\rho(T(\varepsilon))$ and $|\lambda(\varepsilon,\gamma)|$ on ε and γ. On the top left panel, we set $c = 0$ and $\omega = 1/2$ so that the hypotheses of point (B) of Theorem 1 are satisfied, since $\gamma = 0$ and $\lambda_1 \geq \lambda_2 > 0$. As point (B) predicts, we observe that $\min_{\varepsilon\in\mathbb{R}} \rho(T(\varepsilon))$ is attained at $\varepsilon = 0$, i.e. $\min_{\varepsilon\in\mathbb{R}} \rho(T(\varepsilon)) = \rho(T(0)) = \lambda_2$. Hence, adding a perturbation does not improve the coarse space made only by $\mathbf{v}_1$. Next, we consider point (C), by setting $c = 0$ and $\omega = 1$. Through a direct computation we get $\lambda_1 = -0.95$, $\lambda_2 = -\lambda_1$ and $\lambda_3 = 0.90$. The top-right panel shows, on the one hand, that for several values of ε, $\rho(T(\varepsilon)) = \lambda_3 < \lambda_2$, that is with a one-dimensional perturbed coarse space, we obtain the same contraction factor we would have with the two-dimensional spectral coarse space $V_c = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. On the other hand, we observe that there are two values of ε such that $\lambda(\varepsilon,\gamma) = 0$, which (recalling (4) and (6)) implies that T is nilpotent over the span$\{\mathbf{v}_1, \mathbf{v}_2\}$. To study point (D), we set $c = 10$, $\omega = 1/2$, which lead to $\lambda_1 = 0.92$, $\lambda_2 = \lambda_3 = 0.90$. The left-bottom panel confirms there exists an $\varepsilon^* < 0$ such that $|\lambda(\varepsilon^*,\gamma)| \leq \lambda_2$, which implies $\rho(T(\varepsilon^*)) \leq \lambda_2$. Finally, we set $c = 10$ and $\omega = 1$. Point (E) is confirmed by the right-bottom panel, which shows that $|\lambda(\varepsilon,\gamma)| < |\lambda_2|$, and thus $\min_\varepsilon \rho(T(\varepsilon)) = |\lambda_3|$, for some values of ε.

We have shown both theoretically and numerically that the spectral coarse space is not necessarily the one-dimensional coarse space minimizing $\rho(T)$. Now, we wish to go beyond this one-dimensional analysis and optimize the entire coarse space V_c keeping its dimension m fixed. This is equivalent to optimizing the prolongation operator P whose columns span V_c. Thus, we consider the optimization problem

$$\min_{P \in \mathbb{R}^{n \times m}} \rho(T(P)). \tag{12}$$

To solve approximately (12), we follow the approach proposed by [6]. Due to the Gelfand formula $\rho(T) = \lim_{k \to \infty} \sqrt[k]{\|T^k\|_F}$, we replace (12) with the simpler optimization problem $\min_P \|T(P)^k\|_F^2$ for some positive k. Here, $\|\cdot\|_F$ is the Frobenius norm. We then consider the unbiased stochastic estimator [5]

$$\|T^k\|_F^2 = \text{trace}\left((T^k)^\top T^k\right) = \mathbb{E}_{\mathbf{z}}\left[\mathbf{z}^\top (T^k)^\top T^k \mathbf{z}\right] = \mathbb{E}_{\mathbf{z}}\left[\|T^k \mathbf{z}\|_2^2\right],$$

where $\mathbf{z} \in \mathbb{R}^n$ is a random vector with Rademacher distribution, i.e. $\mathbb{P}(\mathbf{z}_i = \pm 1) = 1/2$. Finally, we rely on a sample average approach, replacing the unbiased stochastic estimator with its empirical mean such that (12) is approximated by

$$\min_{P \in \mathbb{R}^{n \times m}} \frac{1}{N} \sum_{i=1}^{N} \|T(P)^k \mathbf{z}_i\|_F^2, \tag{13}$$

where $\mathbf{z}_i$ are a set of independent, Rademacher distributed, random vectors. The action of T onto the vectors $\mathbf{z}_i$ can be interpreted as the feed-forward process of a neural net, where each layer represents one specific step of the two-level method, that is the smoothing step, the residual computation, the coarse correction and the prolongation/restriction operations. In our setting, the weights of most layers are fixed and given, and the optimization is performed only on the weights of the layer representing the prolongation step. The restriction layer is constrained to have as weights the transpose of the weights of the prolongation layer. The cost of constructing coarse spaces using deep neural networks can be very high, and not practical if the problem needs to be solved only once. However, our interest here is on theoretical aspects, and deep neural networks are used only to show the existence of coarse spaces (asymptotically) better than the spectral ones.

We solve (13) for $k = 10$ and $N = n$ using Tensorflow [1] and its stochastic gradient descent algorithm with learning parameter 0.1. The weights of the prolongation layer are initialized with an uniform distribution. Table 1 reports both $\rho(T(P))$ and $\|T(P)\|_A$ using a spectral coarse space and the coarse space obtained solving (13). We can clearly see that there exist coarse spaces, hence matrices P, corresponding to values of the asymptotic convergence factor $\rho(T(P))$ much smaller than the ones obtained by spectral coarse spaces. Hence, Table 1 confirms that a spectral coarse space of dimension m is not necessarily a (global) minimizer for $\min_{P \in \mathbb{R}^{n \times m}} \rho(T(P))$. This can be observed not only in the case $c = 0$, for which the result of [7, Theorem 5.5] states that (recall that M is symmetric) the spectral coarse space minimizes $\|T(P)\|_A$, but also for $c > 0$, which corresponds to a nonsymmetric A. Interestingly, the coarse spaces obtained by our numerical optimizations lead to preconditioned matrices with better condition numbers, as shown in the last row of Table 1, where the condition number κ_2 of the matrix A preconditioned by the two-level method (and different coarse spaces) is reported.

	c	ω	$m = 1$	$m = 5$	$m = 10$	$m = 15$
$\rho(T)$	0	1/2	0.95 - 0.95	0.90 - 0.90	0.82 - 0.83	0.76 - 0.78
	0	1	0.95 - 0.90	0.90 - 0.80	0.80 - 0.65	0.74 - 0.53
	10	1/2	0.90 - 0.90	0.85 - 0.82	0.79 - 0.74	0.73 - 0.68
	10	1	0.85 - 0.80	0.80 - 0.67	0.71 - 0.55	0.66 - 0.37
$\|T\|_A$	0	1/2	0.95 - 0.95	0.90 - 0.90	0.82 - 0.84	0.76 - 0.77
	0	1	0.95 - 0.95	0.90 - 0.94	0.80 - 0.88	0.74 - 0.88
κ_2	0	1	46.91 - 29.45	18.48 - 14.40	9.37 - 8.22	6.69 - 8.53
	10	1	27.25 - 23.98	22.44 - 12.36	17.34 - 11.35	13.06 - 9.71

Table 1: Values of $\rho(T)$, $\|T\|_A$ and condition number κ_2 of the matrix A preconditioned by the two-level method for different c and ω and using either a spectral coarse space (left number), or the coarse space obtained solving (13) (right number).

Acknowledgements G. Ciaramella is a member of INdAM GNCS.

References

1. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
2. M. J. Gander, L. Halpern, and K. Repiquet. A new coarse grid correction for RAS/AS. In *Domain Decomposition Methods in Science and Engineering XXI*, pages 275–283. Springer, 2014.
3. M. J. Gander, L. Halpern, and K. Santugini-Repiquet. On optimal coarse spaces for domain decomposition and their approximation. In *Domain Decomposition Methods in Science and Engineering XXIV*, pages 271–280. Springer, 2018.
4. M. J. Gander and B. Song. Complete, optimal and optimized coarse spaces for additive Schwarz. In *Domain Decomposition Methods in Science and Engineering XXIV*, pages 301–309. Springer, 2019.
5. M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun. Stat.-Simul. C.*, 18(3):1059–1076, 1989.
6. A. Katrutsa, T. Daulbaev, and I. Oseledets. Deep multigrid: learning prolongation and restriction matrices. *arXiv preprint arXiv:1711.03825*, 2017.
7. J. Xu and L. Zikatanov. Algebraic multigrid methods. *Acta Numer.*, 26:591–721, 2017.