MULTILEVEL METHODS FOR ELLIPTIC PROBLEMS WITH HIGHLY VARYING COEFFICIENTS ON NONALIGNED COARSE GRIDS

ROBERT SCHEICHL†, PANAYOT S. VASSILEVSKI‡, AND LUDMIL T. ZIKATANOV§

Abstract. In this paper we generalize the analysis of classical multigrid and two-level overlapping Schwarz methods for 2nd order elliptic boundary value problems to problems with large discontinuities in the coefficients that are not resolved by the coarse grids or the subdomain partition. The theoretical results provide a recipe for designing hierarchies of standard piecewise linear coarse spaces such that the multigrid convergence rate and the condition number of the Schwarz preconditioned system do not depend on the coefficient variation or on any mesh parameters. An assumption we have to make is that the coarse grids are sufficiently fine in the vicinity of cross points or where regions with large diffusion coefficients are separated by a narrow region where the coefficient is small. We do not need to align them with possible discontinuities in the coefficients. The proofs make use of novel stable splittings based on weighted quasi-interpolants and weighted Poincaré-type inequalities. Numerical experiments are included that illustrate the sharpness of the theoretical bounds and the necessity of the technical assumptions.

Key words. coarse spaces, multigrid, overlapping Schwarz method, large coefficient jumps

AMS subject classifications. 65F10, 65N20, 65N30

DOI. 10.1137/100805248

1. Introduction. We are interested in 2nd order elliptic boundary value problems posed in variational form as

\[
\int_{\Omega} \alpha(x) \nabla u^* \cdot \nabla v \, dx = \int_{\Omega} f(x)v(x) \, dx \quad \text{for all } v \in H^1_0(\Omega)
\]

and to be solved for \( u^* \in H^1_0(\Omega) \) on a given polygonal (polyhedral) domain \( \Omega \subset \mathbb{R}^d \) for \( d = 2 \) or \( 3 \), where \( H^1_0(\Omega) \) is the usual Sobolev space of functions defined on \( \Omega \) with vanishing trace on \( \partial \Omega \). We are interested in the case where the diffusion coefficient \( \alpha = \alpha(x) \) may have large variations within \( \Omega \). To be more specific and to simplify the presentation below, we assume that \( \alpha \) is piecewise constant such that \( \alpha|_Y_m \equiv \alpha_m \) on a finite but possibly large number of regions \( Y_m \).

We consider standard finite element (FE) discretizations of this problem on a conforming mesh \( T_h \) on \( \Omega \), which we assume to resolve any discontinuities in the coefficients. To be specific, let \( V_h \) be the \( H^1_0 \)-conforming FE space of piecewise linear
functions associated with $\mathcal{T}_h$. We are interested in multilevel approaches to construct preconditioners for this problem within the subspace correction framework. Our study includes the classical two-level overlapping Schwarz and geometric multigrid (MG) methods.

For both types of subspace correction methods we need a coarse space $V_0 := \text{span}\{\Phi_j\}$. In the MG setting this space is the coarsest in a hierarchy of $(L + 1)$ spaces $V_0 \subset V_1 \subset \cdots \subset V_L = V_h$. For simplicity, we consider the case when these spaces are standard piecewise linear FE spaces defined on a sequence of successively refined meshes $\mathcal{T}_0 = \mathcal{T}_H, \mathcal{T}_1, \ldots, \mathcal{T}_L = \mathcal{T}_h$ with decreasing mesh size. For the two-level Schwarz method, on the other hand, we assume that there is a finite (overlapping) covering $\{\Omega_i\}$ of $\Omega$. In this case the subspaces (in addition to $V_0$) are $\{V_i\}_{i=1}^s$, where $V_i := V_h \cap H^1_0(\Omega_i)$. Since this is a two-level method, we have more flexibility in the choice of the coarse space $V_0$. In particular, $V_0$ can be obtained via some form of agglomeration of fine-grid elements from $\mathcal{T}_h$. Moreover, the analysis that we present also goes through even when the two FE spaces $V_h$ and $V_0$ are not nested.

The case of elliptic problems with highly varying coefficients has been of interest for many years. Under the assumption that the discontinuities are resolved by the coarsest grid, early works on the hierarchical basis (HB) method (see, e.g., [23] and the references therein) provide bounds that are independent of the coefficient variation. A well-known issue with the HB method is that the condition number of the preconditioned system in three dimensions grows as $1/h$, rendering these methods impractical in many cases. However, the robustness with respect to the coefficient variation naturally extends to a stabilized version of HB, the algebraic multilevel iteration (AMLI) method, and in [22] it was shown that as a multilevel preconditioner AMLI exhibits uniform condition number bounds in three dimensions, with respect to both the coefficient variation and the mesh size. The same optimal convergence results hold if AMLI cycles are not used to stabilize the HB method but in the traditional MG setting (for details, see [23, section 5.6]). Note, however, that AMLI cycles are slightly more expensive than $V$-cycles but nevertheless of optimal cost. For overlapping Schwarz-type methods an overview of early theoretical results for the resolved coefficient case can be found in [5]. The three-dimensional case was treated in [9], where for certain (so-called quasi-monotone) coefficient distributions the near-optimality of Schwarz-type methods with standard (piecewise linear) coarse spaces was shown. These results are based on stability results for weighted $L_2$-projections in [3] which require that the coefficients be resolved by the coarse mesh.

If the coefficients are not quasi-monotone, it is necessary to resort to other (“exotic”) coarse spaces (see, e.g., [9, 19]). The role of such coarse spaces is to handle the singularities due to coefficient discontinuities across element boundaries, typically resulting in the violation of Poincaré-type inequalities, which are crucial for the analysis. For a detailed discussion on the topic of constructing exotic coarse spaces for the two-level Schwarz method, we refer the reader to the monograph [21]. As recently shown in [24] and [27], for MG and two-level Schwarz with standard coarse spaces, the stability results for weighted $L_2$-projections in [3] can also be used to establish a near-optimal bound on the effective condition number of the preconditioned system (discarding a small cluster of “bad” eigenvalues). It is well known that Krylov methods still perform well in this case. We refer the reader also to [12] for earlier work.

The literature on the case when the coarser grids are not aligned with the discontinuities of the coefficient is fairly recent. To the best of our knowledge the only paper on standard piecewise linear coarse spaces is [11]. This work is in the context of the two-level Schwarz method, and the results are under certain restrictions on the
shape of the regions $Y_m$ and the behavior of the coefficient. In particular, it is not possible to treat non–quasi-monotone coefficients as defined in [9]. All other works, in particular in the algebraic MG literature, resort to operator-dependent bases and coarse spaces (see, e.g., [23] and the references therein). The theoretical analysis of the operator-dependent bases in the case of highly varying coefficients is fairly limited (for two-level results see [10]). More recent theoretical works in the context of the Schwarz method with coarse spaces constructed via energy minimization can be found in [13, 20]. See also related work on FOSLS-type methods for (1.1) (with analysis) in [1, 2]. However, the focus there is again on coarse grids that are aligned with the coefficients and on problem specific bases.

All the references mentioned above either deal with the case when the coarse grid is aligned with the discontinuities of the coefficient or use coefficient- (operator-) dependent bases for the coarse spaces. In this paper, we prove convergence results for the case where (a) the coarse grids and the subdomain partition do not have to be aligned with the coefficient discontinuities and (b) the multilevel hierarchy consists of standard piecewise linear coarse spaces. We are able to achieve such a generality under the mild assumption that the coarse grids are suitably refined in certain areas of the domain, such as near cross points. The key tools to prove robustness of the preconditioners with respect to the coefficient variation and mesh size are novel weighted Poincaré-type inequalities established in [17, 18, 16]. The uniform bound on the Poincaré constants relies on our assumption on the coarse grids. (For some special cases of these inequalities see also [1, 11].)

The implementation of the multilevel method that we analyze can be done by locally rearranging a given sequence of meshes. Starting from the finest mesh that resolves the coefficient (by definition), the coarsening is performed gradually, so that the coarser meshes are locally refined in certain problematic areas known in advance. Although the main focus of this paper is theoretical, an example of such a strategy is given in the numerical experiments section. If the resulting coarse space $V_0$ is still too large, it is possible to continue coarsening with operator-dependent techniques.

The rest of the paper is structured as follows. In section 2 we formulate a set of assumptions on the coarse spaces. In section 3 we discuss the validity of the key assumption and give coefficient-independent bounds of the constants in weighted Poincaré-type inequalities. We prove a new stability result for quasi-interpolation in section 4. We then show uniform bounds on the condition number of the preconditioned systems in section 5 (two-level Schwarz preconditioner) and in section 6 (MG preconditioner). The numerical tests in section 7 show the sharpness of the theoretical bounds and the necessity of the technical assumptions. We finish in section 8 with some conclusions.

Throughout the paper, the notation $C \lesssim D$ (for two quantities $C, D$) means that $C/D$ is bounded above independently, not only of the mesh size $h$ and the method specific parameters (such as $H_K$ and $\delta_K$, defined below for $K \in T_0$, or the number of levels $L$) but also of the coefficient values $\alpha_m$. Moreover, $C \approx D$ means that $C \lesssim D$ and $D \lesssim C$.

2. Abstract theoretical assumptions on the coarse spaces. To simplify the presentation of our theoretical results let us assume that $\Omega \subset \mathbb{R}^3$. The two-dimensional case follows immediately. The choice of appropriate coarse spaces $V_H := \text{span}\{\Phi_j \mid j = 1, \ldots, N\}$ is at the heart of multilevel subspace correction methods. In particular, we will consider standard piecewise linear coarse spaces associated with coarse triangulations $T_H := \{K\}$ of $\Omega$, such that each $K$ is a shape regular tetrahedron, where each of the functions $\Phi_j$ is associated with a vertex of $T_0$. However,
our framework allows also for more general coarse spaces associated, e.g., with a set \( T_H := \{ K \} \) of aggregates of fine grid elements (not necessarily simplicial), where each of the functions \( \Phi_j \) is associated with one of the aggregates \( K \) and has support on \( K \) and all the adjacent aggregates \( K' \). We do not assume that the elements/aggregates \( K \) or the functions \( \Phi_j \) are chosen in any way related to the coefficient function \( \alpha \). However, the assumptions on \( \Phi_j \) below will implicitly restrict how coarse we may choose \( T_H \) and require a certain “adaptivity” near areas where two regions with high coefficients are separated by a narrow strip with a relatively low coefficient or where one such region comes close to the Dirichlet boundary. This also extends to the situation where high coefficient regions touch each other or the Dirichlet boundary in a single point. For simplicity we assume that \( \Phi_j \in V_h \), i.e., the coarse space is conforming, but we will come back to the nonconforming case in section 5.1 below.

Let

\[
\omega_j := \text{supp}(\Phi_j) \quad \text{and} \quad \omega_K := \bigcup_{\{ j : \omega_j \cap K \neq \emptyset \}} \omega_j,
\]

and set \( H_j := \text{diam}(\omega_j) \) and \( H_K := \text{diam}(\omega_K) \). In addition, we will also require the local fine grid mesh width \( h_K := \max_{\tau \in T_h} h_\tau \), where \( h_\tau \) is the diameter of \( \tau \in T_h \).

First, we make the following standard assumptions on our coarse space:

A1. \( \| \Phi_j \|_{L^\infty(\Omega)} \lesssim 1 \).

A2. \( \| \nabla \Phi_j \|_{L^\infty(\Omega)} \lesssim H_j^{-1} \).

A3. For all \( K \in T_H \), either \( \sum_{j=1}^N \Phi_j|_{\omega_K} \equiv 1 \) or \( \partial \omega_K \cap \partial \Omega \neq \emptyset \).

A4. If \( \omega_j \cap \omega_j' \neq \emptyset \), then \( H_j \approx H_j' \).

For a standard piecewise linear coarse space \( V_H \) associated with a coarse simplicial triangulation \( T_H \), assumptions A1–A4 are always satisfied provided \( T_H \) is locally quasi-uniform. In the more general case, i.e., when the underlying partitioning does not consist of tetrahedra but of more general aggregates of fine grid elements that still satisfy certain local quasi-uniformity properties, locally supported functions \( \Phi_j \) satisfying A1–A4 can still be constructed fairly simply (and locally), e.g., by harmonic extension of piecewise linear boundary data from the interfaces between aggregates to the interior of the aggregates.

The following assumption captures all the coefficient dependence of the coarse space, and as we shall see in the next section, it can always be satisfied by appropriate local refinement of \( T_H \).

A5. For each \( K \in T_H \), there exists a \( C_K^\alpha \) such that one of the following two conditions holds for all \( v \in V_h \):

\[
\inf_{c \in \mathbb{R}} \int_{\omega_K} \alpha(v - c)^2 \, dx \lesssim C_K^\alpha H_K^2 \int_{\omega_K} \alpha |\nabla v|^2 \, dx,
\]

\[
\partial \omega_K \cap \partial \Omega \neq \emptyset \quad \text{and} \quad \int_{\omega_K} \alpha v^2 \, dx \lesssim C_K^\alpha H_K^2 \int_{\omega_K} \alpha |\nabla v|^2 \, dx.
\]

This assumption postulates the existence of a discrete weighted Poincaré/Friedrichs-type inequality on each \( \omega_K \). From assumptions A1–A4 such an inequality clearly follows in the case of coefficients \( \alpha \approx 1 \) (i.e., mildly varying coefficients) with constants \( C_K \approx 1 \) independent of any mesh parameters. If \( \alpha \) is highly varying, then the constants \( C_K^\alpha \) may depend on \( \max_{x,y \in \omega_K} \alpha(x)/\alpha(y) \). However, it turns out that the simple requirement that \( T_H \) be sufficiently fine in a few “critical” areas of the domain, such as near cross points, is sufficient for assumption A5 to be satisfied with \( C_K \) independent of any mesh parameters and of any variation in \( \alpha \) on \( \omega_K \) for almost
all coefficients $\alpha$. Thus, before we present our new multilevel analysis, we turn our attention to assumption A5.

### 3. Weighted Poincaré inequalities

In this section we investigate in detail the ways in which the local coefficient variation may affect the size of the constant $C_K^*$ in the weighted Poincaré-type inequalities in assumption A5. In particular, we explain how to avoid deterioration of $C_K^*$ by a suitable refinement of the coarse grid near cross points and other “critical” areas.

To be more specific and to simplify the presentation, we assume that $\alpha$ is piecewise constant on a finite but possibly large number of regions. The results extend in a straightforward way to more general coefficients $\alpha$, and we will briefly discuss this in Remark 3.1 below. Following [17, 18, 16] we will define classes of quasi-monotone piecewise constant coefficients for which assumption A5 holds with $C_K^*$ independent of the variation of $\alpha$ in $\omega_K$. $C_K^*$ may depend on $H_K/h_K$ or on $\log(H_K/h_K)$ for some $K \in \mathcal{T}_H$, prompting a certain adaptivity of the coarse grid in those “critical” regions.

Let $\alpha$ be piecewise constant with respect to a set \{\(\gamma_m : m = 1, \ldots, M\)\} of connected (open) subdomains of $\Omega$, i.e., $\alpha|_{\gamma_m} \equiv \alpha_m$, where \(\bigcup_{m=1}^{M} \gamma_m = \Omega\) and $\gamma_m \cap \gamma_{m'} = \emptyset$ if $m \neq m'$. We need only very mild assumptions on the shape and the size of these regions $\gamma_m$. We do not require any form of shape regularity. Some of the regions may be long and thin (channels). The important parameter is the “width” of $\gamma_m$ at its narrowest point. For that purpose we make a mild technical assumption on the shape of these regions $\gamma_m$.

**Definition 3.1 ($\eta$-regular).** We say a polyhedral region $D \subset \mathbb{R}^3$ is $\eta$-regular if it can be triangulated into a quasi-uniform set of tetrahedra $T$ with $\text{diam}(T) \geq \eta$.

We assume that for every $m = 1, \ldots, M$, there exists an $\eta_m > 0$ such that $\gamma_m$ is $\eta_m$-regular. Note that our assumption that $\alpha$ is resolved by the fine grid $\mathcal{T}_h$ means that it is always possible to find such an $\eta_m > 0$. Let $\eta_m$ be the largest possible such value.

To study assumption A5 let us consider a generic coarse element $K \in \mathcal{T}_H$ and define the following subsets of $\omega_K$ where $\alpha$ is constant:

$$
\omega_K^m := \omega_K \cap \gamma_m, \quad \text{where} \quad m \in I_K := \{m : \omega_K \cap \gamma_m \neq \emptyset\}.
$$

Let us assume for simplicity that each of these subregions is connected, which does not add any further restrictions, since we can always subdivide $\gamma_m$ to satisfy this assumption.

Generalizing the notion of quasi-monotonicity coined in [9], we will now define three types of quasi-monotonicity: Type 0, Type 1, and Type 2. To do this let us consider the following three directed combinatorial graphs $\mathcal{G}^{(k)} = (\mathcal{N}, \mathcal{E}^{(k)})$, $k = 0, 1, 2$. The set of vertices $\mathcal{N}$ for all these graphs is the set of subregions $\omega_K^m$, $m \in I_K$. The edges are ordered pairs of vertices. To define the edges we now distinguish between three different types of connections.

**Definition 3.2.** Suppose that $\gamma_{K}^{m,m'} = \overline{\omega_{K}^{m}} \cap \overline{\omega_{K}^{m'}}$ is a nonempty manifold of dimension $k$ for $k = 0, 1, 2$. The ordered pair $(\omega_K^m, \omega_K^{m'})$ is an edge in $\mathcal{E}^{(k)}$ if and only if $\alpha_m \leq \alpha_{m'}$. The edges in $\mathcal{E}^{(k)}$ are said to be of type $k$.

In addition, for $k = 1, 2$, we assume that

* $\text{meas}(\gamma_{K}^{m,m'}) \approx \text{meas}(\omega_K^m \cup \omega_K^{m'})^{k/3}$, and
* $\gamma_{K}^{m,m'}$ is sufficiently regular; i.e., it is a finite union of shape-regular $k$-dimensional simplices of diameter $\approx \text{meas}(\gamma_{K}^{m,m'})^{1/k}$.
Fig. 1. Quasi-monotone coefficient distributions of Types 2, 1, and 0 in (a)–(c), respectively. A darker color indicates a larger coefficient. A typical non–quasi-monotone coefficient is shown in (d).

Quasi-monotonicity is related to the connectivity in these graphs. Let \( m^* \in I_K \) be the index of the region \( \omega_{m^*}^K \) with the largest coefficient, i.e., \( \alpha_{m^*} = \max_{m \in I_K} \alpha_m \).

Definition 3.3. The coefficient \( \alpha \) is type-\( k \) quasi-monotone on \( \omega_K \) if there is a path in \( G^{(k)} \) from any vertex \( \omega_m^K \) to \( \omega_{m^*}^K \).

Obviously \( E^{(2)} \subset E^{(1)} \subset E^{(0)} \), and type-\( k \) quasi-monotone implies type-\( (k - 1) \) quasi-monotone. The coefficients in Figure 1(a)–(c) are examples of quasi-monotone coefficients of Types 2, 1, and 0, respectively. The coefficient in Figure 1(d) is not quasi-monotone.

The following lemma summarizes the results in [17, 18, 16]. It relates the existence of a benign constant \( C_{K}^{*} \) in (2.1) that is independent of \( \alpha \) directly to quasi-monotonicity and the way in which \( C_{K}^{*} \) depends on the ratio \( H_K/h_K \) to the type of quasi-monotonicity.

Lemma 3.1. If \( \alpha \) is type-\( k \) quasi-monotone on \( \omega_K \), then (2.1) holds with

\[
C_{K}^{*} := \begin{cases} 
1 & \text{if } k = 2, \\
1 + \log \left( \frac{H_K}{h_K} \right) & \text{if } k = 1, \\
\frac{H_K}{h_K} & \text{if } k = 0.
\end{cases}
\]

Quasi-monotonicity is crucial. If the coefficient is not quasi-monotone, e.g., the situation in Figure 1(d), then (2.1) cannot hold with \( C_{K}^{*} \) independent of \( \alpha \).

Example 3.1 (counterexample). Let us assume \( \Omega = (0, 1)^3 \) in Figure 1(d) with \( \alpha(x) = \alpha_1 \gg 1 \) if \( x_1 < 1/4 \) or \( x_1 > 3/4 \), and \( \alpha = 1 \) otherwise. Take, for example, the function

\[
v := \begin{cases} 
1 & \text{for } x_1 < 1/4, \\
1 - 4x_1 & \text{for } x_1 \in [1/4, 3/4], \\
-1 & \text{for } x_1 > 3/4.
\end{cases}
\]

Then it is easy to verify that \( \inf_{c \in \mathbb{R}} \int_{\Omega} \alpha(v - c)^2 \, dx \geq \alpha_1/2 \) and \( \int_{\Omega} \alpha |\nabla v|^2 \, dx = 8 \), which means that \( C_{K}^{*} \geq \alpha_1/16 \), and so \( C_{K}^{*} \) grows linearly with the contrast in \( \alpha(x) \).

Let us now consider the case where \( \partial \omega_K \cap \partial \Omega \neq \emptyset \), i.e., the case of Friedrichs inequality (2.2). We assume without loss of generality that \( \text{meas}(\partial \omega_K \cap \partial \Omega) = H_K^2 \). If \( \text{meas}(\partial \omega_K \cap \partial \Omega) \ll H_K^2 \), we can simply extend \( \omega_K \) by a finite number of elements \( K \in T_H \) such that this assumption is satisfied. Of course (2.2) then needs to hold on the extended \( \omega_K \).
We proceed as above and define three graphs \( \mathcal{G}^{(k)} = (\mathcal{N}, \mathcal{E}^{(k)}) \), \( k = 0, 1, 2 \), all containing one extra node, namely, \( \omega_K^0 := \mathbb{R}^3 \setminus \overline{\Omega} \) (i.e., the outside of \( \Omega \)), such that \( \mathcal{N} = \mathcal{N} \cup \{\omega_K^0\} \). We set \( \alpha_0 = \infty \) and \( \mathcal{E}^{(k)} = \mathcal{E}^{(k)} \) and then add to the sets \( \mathcal{E}^{(k)} \) all connections from \( \omega_K^m \) to \( \omega_K^0 \) (if they exist). Since \( \alpha_0 > \alpha_m \) by definition, the ordered pair \( (\omega_K^m, \omega_K^0) \in \mathcal{E}^{(k)} \) for any region \( \omega_K^m \) that touches the Dirichlet boundary \( \partial \Omega \) in a \( k \)-dimensional manifold. Here, we require only that \( \text{meas}(\gamma_{K,m}^c) \approx \text{meas}(\omega_K^m)^{1/3} \) for \( k = 1, 2 \).

**Definition 3.4.** The coefficient \( \alpha \) is type-\( k \) \( \Gamma \)-quasi-monotone on \( \omega_K \) if there is a path in \( \mathcal{G}^{(k)} \) from any vertex \( \omega_K^m \) to \( \omega_K^0 \).

The following lemma can again be found in [17, 18, 16].

**Lemma 3.2.** If \( \alpha \) is type-\( k \) \( \Gamma \)-quasi-monotone on \( \omega_K \), then (2.2) holds with \( C_K^* \) as defined in Lemma 3.1.

Thus, combining the findings in Lemmas 3.1 and 3.2 and in Example 3.1, we can conclude that for assumption A5 to hold with benign constants \( C_K^* \), it suffices to make the coarse grid \( T_H \) sufficiently fine in certain “critical” areas of the domain:

1. The most important condition is that \( \alpha \) is quasi-monotone on all regions \( \omega_K \); otherwise \( C_K^* \approx \max_{x,y \in \omega_K} \alpha(x)/\alpha(y) \). In practice this means that we need to make sure that \( T_H \) is kept sufficiently fine in areas where two regions with large values of \( \alpha \) are separated by a narrow region \( \mathcal{Y}_m \) with a relatively small value \( \alpha_m \). A sufficient condition is that \( H_K \leq \eta_m \) for all \( K \) for which \( \omega_K^m \neq \emptyset \). Note that this also includes the case where a region with a large coefficient is separated from the Dirichlet boundary \( \partial \Omega \) by a narrow region \( \mathcal{Y}_m \) with a relatively small value \( \alpha_m \) to ensure \( \Gamma \)-quasi-monotonicity.

2. The second critical area is around so-called three-dimensional cross points, where the coefficient \( \alpha \) is only Type-0 quasi-monotone, e.g., the situation in Figure 1(c). Here \( C_K^* \approx H_K/h_K \), and so again it suffices to make sure the coarse mesh is sufficiently fine near the cross point, such that \( H_K \lesssim h_K \).

If both those conditions are satisfied, then all the constants \( C_K^*, K \in T_H \), depend at most logarithmically on \( H_K/h_K \), as is confirmed by the numerical tests in section 7.

**Remark 3.1.** Similar results can be proved in two dimensions. There, \( C_K^* = 1 \) if \( \alpha \) is Type-1 quasi-monotone on \( \omega_K \), and \( C_K^* = 1 + \log(H_K/h_K) \), if \( \alpha \) is Type-0 quasi-monotone on \( \omega_K \). Hence, in two dimensions cross points are a much less critical problem. The results can also be extended to more general coefficients (not piecewise constant). Obviously we can include mild local variation, i.e., \( \max_{x,y \in \mathcal{Y}_m} \alpha(x)/\alpha(y) \approx 1 \), but it is even possible to prove similar results to those in Lemmas 3.1 and 3.2 for arbitrary coefficients \( \alpha \), provided they satisfy certain monotonicity conditions on each patch \( \omega_K \) related to those discussed above. For details see [17, 18, 16].

4. **A new stability result for quasi-interpolation.** The crucial ingredient in the analysis of subspace correction methods is the existence of a stable splitting for any \( v \in V_h \) in appropriate subspaces of \( V_h \). To construct these stable splittings it is essential to have stable interpolation operators onto coarse spaces.

Let \( V_H \subset V_h \) be a generic coarse space as defined above. We define for any \( v \in V_h \) the following weighted quasi-interpolant onto \( V_H \), which is a straightforward generalization of usual quasi-interpolants, introduced first by Clement [8], to problems with highly varying coefficients (cf. also [11]):

\[
\Pi_H v := \sum_{j=1}^N \mathbf{\Phi}_j \mathbf{\Phi}_j, \quad \text{where} \quad \mathbf{\Phi}_j := \frac{\int_{\omega_j} \alpha v \, dx}{\int_{\omega_j} \alpha \, dx}.
\]

This quasi-interpolant has the following approximation and stability properties.
Lemma 4.1. Let assumptions A1–A5 hold. Then for \( v \in V_h \) and \( K \in \mathcal{T}_H \) we have

\[
\int_K \alpha (v - \Pi_H v)^2 \, dx \lesssim C_K^* H_K^2 \int_{\omega_K} \alpha |\nabla v|^2 \, dx,
\]

(4.2)

\[
\int_K \alpha |\nabla \Pi_H v|^2 \, dx \lesssim C_K^* \int_{\omega_K} \alpha |\nabla v|^2 \, dx.
\]

(4.3)

Proof. Note first that by Cauchy–Schwarz we have

\[
|\mathbf{v}_j|^2 \leq \int_{\omega_j} \alpha v^2 \, dx \int_{\omega_j} \alpha \, dx,
\]

(4.4)

and so, using assumption A1,

\[
\int_K \alpha (\Pi_H v)^2 \, dx \leq \sum_{j: \omega_j \cap K \neq \emptyset} \frac{\int_{\omega_j} \alpha v^2 \, dx}{\int_{\omega_j} \alpha \, dx} \int_K \alpha \Phi_j^2 \, dx \lesssim \int_{\omega_K} \alpha v^2 \, dx,
\]

(4.5)

which also implies

\[
\int_K \alpha (v - \Pi_H v)^2 \, dx \lesssim \int_{\omega_K} \alpha v^2 \, dx.
\]

(4.6)

Let \( c \in \mathbb{R} \) be an arbitrary constant. If \( \{ \Phi_j \} \) forms a partition of unity on all of \( \omega_K \), we can replace \( v \) on the right-hand side of (4.6) by \( \hat{v} := v - c \). Thus, by assumption A5 there exists a \( c \in \mathbb{R} \) such that

\[
\int_{\omega_K} \alpha \hat{v}^2 \, dx \lesssim C_K^* H_K^2 \int_{\omega_K} \alpha |\nabla v|^2 \, dx.
\]

(4.7)

Combining (4.6) and (4.7) completes the proof of (4.2).

If, on the other hand, \( \{ \Phi_j \} \) does not form a partition of unity on all of \( \omega_K \), then \( \partial \omega_K \cap \partial \Omega \neq \emptyset \), and so again by assumption A5 we have

\[
\int_{\omega_K} \alpha v^2 \, dx \lesssim C_K^* H_K^2 \int_{\omega_K} \alpha |\nabla v|^2 \, dx.
\]

(4.8)

To prove (4.3) we proceed similarly; i.e., using assumption A2 we have

\[
\int_K \alpha |\nabla \Pi_H v|^2 \, dx \leq \sum_{j: \omega_j \cap K \neq \emptyset} \frac{\int_{\omega_j} \alpha v^2 \, dx}{\int_{\omega_j} \alpha \, dx} \int_K \alpha |\nabla \Phi_j|^2 \, dx \lesssim H_j^{-2} \int_{\omega_K} \alpha v^2 \, dx,
\]

(4.9)

which can be bounded as for (4.2), using in addition assumption A4.

This lemma will be sufficient to find a stable splitting for the two-level overlapping Schwarz method. For multilevel methods we will need a further result that provides stability of interpolation between pairs of spaces. Let \( V_H \) and \( V_\eta \) be two subspaces of \( V_h \) such that \( V_H \subset V_\eta \), and let \( \Pi_H \) and \( \Pi_\eta \) be the corresponding quasi-interpolants as defined in (4.1). If \( V_\eta = V_h \), we set \( \Pi_\eta = I \). Furthermore, let

\[
\alpha^{\eta} |_{K'} := \frac{1}{|K'|} \int_{K'} \alpha \, dx \quad \text{for all } K' \in \mathcal{T}_\eta;
\]

(4.10)
i.e., \( \alpha \) is the piecewise constant coefficient function with respect to \( \mathcal{T}_n \) obtained by averaging the coefficient over each element \( K' \in \mathcal{T}_n \).

The following lemma can be proved in much the same way as Lemma 4.1.

**Lemma 4.2.** Let \( V_H \) be such that assumptions A1–A5 hold. Then for any \( v \in V_h \) and \( K \in \mathcal{T}_H \) we have

\[
(4.11) \quad \int_K \alpha^\tau (\Pi_H^\tau v - \Pi_H v)^2 \, dx \lesssim C_K^* H_K^2 \int_{\omega_K} \alpha |\nabla v|^2 \, dx.
\]

**Proof.** We proceed as in (4.5), using assumption A1 and (4.4) to get

\[
(4.12) \quad \int_K \alpha^\tau (\Pi_H^\tau v)^2 \, dx \leq \sum_{i: \omega_i = \omega_K \cap \omega_i \neq \emptyset} \int_{\omega_i} \alpha v^2 \, dx \lesssim \int_{\omega_K} \alpha v^2 \, dx,
\]

where in the last step we used the fact that \( \int_{\omega_i} \alpha \, dx = \int_{\omega_K} \alpha^\tau \, dx \).

Now let \( \{ \Phi_i \}_{i=1}^{N_n} \) denote the basis functions associated with \( V_n \), and set \( \omega_i^g := \text{supp} \Phi_i^g \). Then we can show similarly that

\[
(4.13) \quad \int_K \alpha^\tau (\Pi_H^\tau v)^2 \, dx \leq \sum_{i: \omega_i \cap \omega_i \neq \emptyset} \int_{\omega_i} \alpha v^2 \, dx \lesssim \int_{\omega_K} \alpha v^2 \, dx.
\]

This follows trivially if \( V_n = V_h \). Together, (4.12) and (4.13) imply that

\[
(4.14) \quad \int_K \alpha^\tau (\Pi_H^\tau v - \Pi_H v)^2 \, dx \lesssim \int_{\omega_K} \alpha v^2 \, dx.
\]

The result follows again by using assumption A5 to bound the right-hand side, where crucially we need that both \( \Pi_H^\tau v \) and \( \Pi_H v \) reproduce constants wherever \( \{ \Phi_j \} \) forms a partition of unity on all of \( \omega_K \). \( \square \)

**5. Analysis of two-level overlapping Schwarz.** Let us start by analyzing the two-level overlapping Schwarz method. To complete the setup for this method, in addition to a coarse space \( V_0 := V_h \), we also require a set of overlapping subdomains \( \{ \Omega_i \}_{i=1}^s \) that provide a finite covering of \( \Omega \). We assume that this set is chosen such that there exists a partition of unity \( \{ \chi_i \} \) subordinate to \( \{ \Omega_i \} \) with

OS1. \( \| \chi_i \|_{L^\infty(\Omega)} \lesssim 1 \) and
OS2. \( \| \nabla \chi_i \|_{L^\infty(\Omega)} \lesssim \delta_i^{-1} \) for some \( \delta_i > 0 \).

In other words, the overlap of \( \Omega_i \) with its neighbors has to be of order \( \delta_i \). To simplify the presentation below let \( \delta_K := \min\{\| \omega_i \cap \Omega_i \neq \emptyset \delta_i \}. \) Note again that the sets \( \Omega_i \) are chosen completely independently of the coefficient \( \alpha \). They may also be chosen completely independently of the coarse space, although to simplify the understanding of the theoretical results, it may help the reader to bear in mind the special case where \( s = N_H \) and \( \Omega_i = \omega_i = \text{supp} \Phi_i \) (or a union of such supports).

The above setting is a standard setting for two-level overlapping Schwarz preconditioners. For the convergence analysis, let us define the operator \( A : V_h \rightarrow V_h \):

\[
(A v, w) := a(v, w) \quad \text{for all} \quad v, w \in V_h.
\]

Then the following definitions of the additive Schwarz preconditioner are convenient (see [21, Chapter 2] and also [26, 14]):

\[
B_{AS}^{-1} := P_0 + \sum_{i=1}^s P_i, \quad \text{and} \quad (B_{AS} v, v) := \inf_{\sum_{k=0}^s v_k = v} \sum_{i=0}^s a(v_i, v_i).
\]
Here $P_i v$, $i = 1, \ldots, s$, are the elliptic (also called $a(.,.)$-orthogonal) projections of $v \in V_h$ on $V_i := V_h \cap H_0^1(\Omega_i)$ defined in a standard way, such that

$$a(P_i v, w) = a(v, w) \quad \text{for all } w \in V_i.$$ 

The elliptic projection on the coarse space $V_0$ is denoted with $P_0$ and is defined in the same way. To apply the classical Schwarz theory in this case (see, e.g., [21, Chapter 2]) it suffices to find, for any $v \in V_h$, a stable splitting $\{v_i\}_{i=0}^s$ such that $v_i \in V_i$,

$$v = \sum_{i=0}^s v_i \quad \text{and} \quad \sum_{i=0}^s a(v_i, v_i) \leq C_0 a(v, v).$$

Here, we choose

$$v_0 := \Pi_H v \quad \text{and} \quad v_i := I^h(\chi_i(v - v_0)),$$

where $\Pi_H$ is the quasi-interpolant on the coarse grid $T_H$, defined in (4.1), and $I^h$ is the nodal interpolant on the fine grid $T_h$. Since $\{\chi_i\}$ is a partition of unity on all of $\Omega$, $\{v_i\}_{i=0}^s$ obviously forms a splitting of $v$. The following lemma confirms that the splitting is stable.

**Lemma 5.1.** Under the assumptions A1–A5, we have for all $v \in V_h$ that

$$\sum_{i=0}^s a(v_i, v_i) \lesssim \max_{K \in T_H} C_K^* \left( 1 + \frac{H_K}{\delta_K} \right)^2 a(v, v).$$

**Proof.** The bound for the energy of $v_0$ follows immediately from Lemma 4.1.

It remains to bound the energy of $v_i$ for $i > 0$. It is a classical result (see [13, Lemma 3.3]) for the nonconstant coefficient case that

$$a(v_i, v_i) = \int_{\Omega_i} |\nabla I^h(\chi_i(v - v_0))|^2 \, dx \lesssim \left( \|\nabla \chi_i\|_{L^\infty(\Omega_i)} \int_{\Omega_i} |v - v_0|^2 \, dx + \|\chi_i\|_{L^\infty(\Omega_i)} \int_{\Omega_i} |\nabla(v - v_0)|^2 \, dx \right) \lesssim \left( \delta_i^{-2} \int_{\Omega_i} |v - v_0|^2 \, dx + \int_{\Omega_i} |\nabla v|^2 \, dx \right),$$

where in the last step we have used OS1 and OS2.

To bound the right-hand side of (5.1) we use Lemma 4.1; i.e.,

$$\int_{\Omega_i} |\nabla v|^2 \, dx \leq \sum_{K:K \cap \Omega_i \neq \emptyset} \int_K |\nabla \Pi_H v|^2 \, dx \lesssim \sum_{K:K \cap \Omega_i \neq \emptyset} C_K^* \int_K |\nabla v|^2 \, dx.$$ 

Substituting these two bounds into (5.1), summing up, and using the fact that the cover $\{\Omega_i\}_{i=1}^s$ is finite, we obtain the result.

Classical Schwarz theory then leads to the following bound on the condition number of $B_{AS} A$ (see [21, Chapters 2 and 3] for details).
Theorem 5.1. Under the assumptions A1–A5 and provided \( \{ \Omega_i \}_{i=1}^s \) is a finite cover of \( \Omega \) satisfying OS1–OS2, we have
\[
\kappa(B_{AS}^{-1}A) \lesssim \max_{K \in T_H} C_K^* \left( 1 + \frac{H_K}{\delta_K} \right)^2.
\]
(The hidden constant does not depend on \( \alpha \).)

Corollaries for the multiplicative and for the hybrid versions of two-level overlapping Schwarz follow in the usual way from Lemma 5.1.

Note that the quadratic dependence on \( H_K/\delta_K \) can be improved to a linear dependence, if we add a further (technical) assumption on the subdomain partition related to how the coefficient varies on the subdomain boundary layer \( \Omega_{i,\delta_i} := \Omega_i \setminus \bigcup_{i' \neq i} \overline{\Omega}_{i'} \), i.e., the part of \( \Omega_i \) that is overlapped by neighboring subdomains. Since \( \nabla \chi_i = 0 \) in the remainder of \( \Omega_i \), the first integral on the right-hand side of (5.1) needs only to be taken over \( \Omega_{i,\delta_i} \). If each coefficient region \( Y_m \) that overlaps \( \Omega_{i,\delta_i} \) has a sufficiently large intersection (\( \approx \delta_i \)) with the boundary of \( \Omega_i \), then we can apply [13, Lemma 3.4] to each of these coefficient subregions separately and reduce the condition number bound in Theorem 5.1 to
\[
\kappa(B_{AS}^{-1}A) \lesssim \max_{K \in T_H} C_K^* \left( 1 + \frac{H_K}{\delta_K} \right).
\]

Note that this is a sufficient, but by no means necessary, condition, and much more general partitions \( \{ \Omega_i \}_{i=1}^s \) are possible to obtain the linear dependence, but this would become too technical to describe here.

If we assume for simplicity generous overlap, that is, \( \delta_K \approx H_K \), e.g., in the case where \( s = N_H \) and \( \{ \Omega_i \}_{i=1}^s = \{ \omega_i \}_{i=1}^s \), then we get from Theorem 5.1 that
\[
\kappa(B_{AS}^{-1}A) \lesssim \max_{K \in T_H} C_K^*.
\]

Recalling our discussion in section 3 this means that it is not essential for the robustness of two-level overlapping Schwarz that discontinuities in the coefficient are resolved by the coarse grid and/or the subdomain partitioning. However, it also shows that a certain adaptivity of the coarse space is required near areas with high contrast in the coefficients, such that \( \max_{K \in T_H} C_K^* \lesssim 1 \) independent of any mesh parameters and independent of \( \alpha \). This provides a simple recipe for designing fully robust two-level Schwarz methods based on standard piecewise linear coarse spaces.

5.1. Nonconforming coarse spaces. We finish this section by making a comment about nonconforming coarse spaces \( V_H \nsubseteq V_h \). Robustness of two-level Schwarz methods for this case can still be proved adapting the proof techniques developed in [6] to the variable coefficient case (see also [21, Chapter 3]). The only assumption on the coarse space that has to be slightly modified is assumption A5. Essentially the proof is identical to the one above if we choose
\[
\nu_0 := \tilde{I}^h \left( \sum_{j=1}^N \overline{\nu}_j \tilde{\Phi}_j \right), \quad \text{with} \quad \overline{\nu}_j := \frac{\int_{\omega_j} \alpha v \, dx}{\int_{\omega_j} \alpha \, dx}
\]
and \( \omega_j := \text{supp}(\tilde{I}^h(\Phi_j)) \), where \( \tilde{I}^h \) is the following quasi-interpolant onto the fine grid:

For every function \( v \in L_1(\Omega) \) let
\[
\tilde{I}^h(v) := \sum_{\text{vertex } s_p \text{ in } T^h} \overline{\tau}_p \varphi_p, \quad \text{where} \quad \overline{\tau}_p := \frac{\int_{D_p} \alpha v \, dx}{\int_{D_p} \alpha \, dx}
\]
and $D_p := \bigcup_{\tau, \tau_\ell \in \tau} \tau$. This quasi-interpolant is stable in the weighted $L_2$-norm and in the weighted $H^1$-seminorm in the sense that

\[(5.2) \quad \int_\tau a \tilde{I}^h(v)^2 \, dx \lesssim \int_{D_\tau} \alpha v^2 \, dx \quad \text{and} \quad \int_\tau \alpha |\nabla \tilde{I}^h(v)|^2 \, dx \lesssim \int_{D_\tau} \alpha |\nabla v|^2 \, dx \]

for $D_\tau := \bigcup_{\tau, \tau_\ell \in \tau, \tau_\ell \neq \emptyset} \tau'$. The inequalities in (5.2) can be proved like (4.5) and (4.9) in the proof of Lemma 4.1, provided assumption A5 holds on a slightly extended region $\omega_K$ for every $K \in \mathcal{T}_H$. To be precise, setting $\tilde{\omega}_j := \bigcup_{\tau, v_\ell \in \tau, v_\ell \neq \emptyset} D_\tau$, we define $\omega_K := \bigcup_{j, \omega_j \cap K} \tilde{\omega}_j$, i.e., the original region $\omega_K$ extended by a layer of fine grid elements. If assumption A5 holds on every such $\omega_K$, then the proofs of Lemmas 4.1 and 5.1 can be adapted straightforwardly to the nonconforming case using (5.2), and Theorem 5.1 holds also for $V_h \not\subset V_{\ell}$. Note that the support $\Omega_H$ of the functions in $V_H$ does not even have to be equal to $\Omega$. It suffices that $\text{dist}(x, \partial \Omega) \approx H_h$ for all $x \in \omega_j$ (for details see [6, 21]).

This is particularly useful for unstructured fine grids $\mathcal{T}_h$ where it may be difficult to find a coarse space $V_{\ell} \subset V_h$ that satisfies assumptions A1–A4. See [12] for a practical coarse space $V_{\ell} \not\subset V_h$ for unstructured fine grids $\mathcal{T}_h$ that does satisfy assumptions A1–A4.

6. MG analysis. Lemmas 4.1 and 4.2 actually provide the basis for a complete multilevel theory, and in this section we will show how the analysis in the previous section can be extended to multilevel methods, such as standard geometric MG with piecewise linear coarse spaces. As for two-level Schwarz, we will see that the only requirement we eventually need from our coarse spaces is that the underlying meshes are sufficiently fine in certain “critical” areas of the domain. Provided this is the case, the convergence rate of standard geometric MG is independent of the coefficients, even when they are not resolved by any of the coarse meshes.

Let us assume we have a sequence of nested FE spaces $V_0 \subset V_1 \subset \cdots \subset V_L$, such that $V_L = V_h$ and $V_0 = V_\ell$ and such that $V_\ell$ satisfies assumptions A1–A5. For simplicity, in this section let us only consider spaces $\{V_\ell\}_{\ell=0}^{L-1}$ that consist of piecewise linear and continuous functions associated with some coarse triangulations $\mathcal{T}_\ell$ that are locally quasi-uniform, so that A1–A4 are naturally satisfied on all grids. To further fix the notation, we will consider here the multigrid V-cycle with weighted Jacobi smoother. Other types of smoothers (e.g., the Gauss–Seidel smoother) can be analyzed in a completely analogous fashion. For equivalence relations between the Jacobi, Gauss–Seidel, and other smoothers, see [23, 4, 15].

We now introduce some notation relevant to the MG analysis that we present below. We start by defining the popular Jacobi method using additive Schwarz notation. With a proper scaling it defines the smoother that we use in the MG analysis. Let $\{\Phi_j\}_{j=1}^{N_j}$ denote the basis functions associated with $V_\ell$, $\ell > 0$, and let $p^\ell_j$ denote the elliptic projection on the one-dimensional space span\{\Phi_j\}, that is,

$$p^\ell_j v = \frac{a(\Phi_j, v)}{a(\Phi_\ell, \Phi_\ell)} \Phi_j.$$

The scaled Jacobi operator $S_{\ell} = \sigma_S^\ell \sum_{j=1}^{N_j} p^\ell_j$ for any given $\sigma_S^\ell > 0$ is invertible and hence can be used to define the bilinear form

$$a_\ell(v_\ell, w_\ell) := a(S_{\ell}^{-1} v_\ell, w_\ell) \quad \text{for all } v_\ell, w_\ell \in V_\ell.$$
By expanding \( v_\ell = \sum_{j=1}^{N_\ell} \xi_j \phi_j^\ell \) and \( w_\ell = \sum_{j=1}^{N_\ell} \eta_j \phi_j^\ell \) we get
\[
a_\ell(v_\ell, w_\ell) = (\sigma_{S_\ell}^\ell)^{-1} \sum_{j=1}^{N_\ell} \xi_j a(\phi_j^\ell, \phi_j^\ell) \eta_j.
\]

Noticing that \( a(\phi_j^\ell, \phi_j^\ell) \) are the diagonal entries of the stiffness matrix, we see that the above form is simply an operator-function notation of the traditional Jacobi iteration matrix.

Here and in what follows, \( \alpha_{S_\ell}^\ell > 0 \) is chosen so that \( S_\ell \) is a contraction in the energy norm. For example, taking \( (\sigma_{S_\ell}^\ell)^{-1} \) equal to twice the number of nonzeros per row in the stiffness matrix on level \( \ell \) is sufficient to make both \( S_\ell \) and \( (I - S_\ell) \) contractive in the energy norm. We set
\[
\sigma_S := \min_{1 \leq \ell \leq L} \alpha_{S_\ell}^\ell > 0
\]
and observe that from the shape regularity of the meshes it follows that we have a bounded number of nonzeros per row in the stiffness matrices on every level. Hence, \( \sigma_S \) is independent of \( \alpha \) and of the mesh sizes.

We now introduce the norm associated with the bilinear form \( a_\ell(\cdot, \cdot) \):
\[
\| v_\ell \|_{*, \ell}^2 := a_\ell(v_\ell, v_\ell) = a(S_\ell^{-1} v_\ell, v_\ell) \quad \text{for all } v_\ell \in V_\ell.
\]

On the coarsest level \( V_0 \) we solve exactly, and so we choose \( a_0(\cdot, \cdot) := a(\cdot, \cdot) \), and \( \| \cdot \|_{*, 0} \) is the standard energy norm. The action of the V-cycle MG preconditioner \( B_{MG}^{-1} \) for a given \( f \in V_L \) can now be formulated as follows (see, for example, [4, 15, 23, 25, 7]).

ALGORITHM 6.1 (MG preconditioner). Given \( f \in V_L \), set \( u_{-L-1} = 0 \).

For \( \ell = -L : L \)
\[
\text{for } \ell = -L : L
\]
\[
\text{Let } e_\ell \in V_[\ell] \text{ be the solution of }
\]
\[
a_\ell(e_\ell, v_\ell) = (f, v_\ell) - a(u_{\ell-1}, v_\ell) \quad \text{for all } v_\ell \in V_[\ell].
\]
\[
\text{Define } u_\ell := u_{\ell-1} + e_\ell.
\]
\end{for}

Set \( B_{MG}^{-1} f = u_L \).

For \( \ell > 0 \), \( a_{-\ell}(\cdot, \cdot) \) is defined using the \( a(\cdot, \cdot) \)-adjoint of \( S_\ell \). In the case of weighted Jacobi, we have that \( a_{-\ell}(\cdot, \cdot) = a_{-\ell}(\cdot, \cdot) \). Note that, even though the steps in the algorithm above are on the fine grid, its implementation can be done efficiently using restrictions to coarse grid problems. We refer the reader to [15, 23] for implementation issues.

For any fixed \( 0 < \ell \leq L \), the bilinear form \( a_\ell(\cdot, \cdot) \) defines a linear operator \( T_\ell : V \mapsto V_\ell \) via the relation
\[
a_\ell(T_\ell v, v_\ell) = a(v, v_\ell), \quad \text{and hence } T_\ell = S_\ell P_\ell = \sum_{j=1}^{N_\ell} \xi_j \phi_j^\ell P_\ell,
\]
where \( P_\ell \) is the elliptic \( a(\cdot, \cdot) \)-orthogonal projection on \( V_\ell \). Indeed, by the definitions above we have
\[
a_\ell(T_\ell v, v_\ell) = a_\ell(S_\ell P_\ell v, v_\ell) = a(S_\ell^{-1} S_\ell P_\ell v, v_\ell) = a(v, v_\ell).
\]

One also easily verifies that \( T_\ell \) is self-adjoint in the \( a(\cdot, \cdot) \) inner product, i.e.,
\[
a(T_\ell v, w) = a(T_\ell v, P_\ell w) = a(S_\ell P_\ell v, P_\ell w) = a(P_\ell v, S_\ell P_\ell w) = a(v, T_\ell w).
\]
Finally, we will also need the symmetrization of $T_{\ell}$, namely,

$$
\mathbb{T}_{\ell} := S_{\ell} P_{\ell}, \quad \text{where} \quad S_{\ell} = (2S_{\ell} - S_{\ell}^2).
$$

First note that $\mathbb{T}_{\ell}$ acting on $V_{\ell}$ equals $S_{\ell}$. Then, by the construction of $S_{\ell}$, we notice that $S_{\ell} - S_{\ell} = S_{\ell}(I - S_{\ell})$ is symmetric positive definite in the $a$-inner product; hence $a(S_{\ell}^{-1} v_{\ell}, v_{\ell}) \leq a(S_{\ell}^{-1} v_{\ell}, v_{\ell})$, which implies that, for all $0 < \ell \leq L$,

$$
(6.3) \quad a(S_{\ell}^{-1} v_{\ell}, v_{\ell}) \leq \|v_{\ell}\|_{a,\ell}^2 \quad \text{and} \quad a(S_{\ell}^{-1} T_{\ell} v_{\ell}, T_{\ell} v_{\ell}) \leq a(v_{\ell}, v_{\ell}) \quad \text{for all} \quad v_{\ell} \in V_{\ell}.
$$

Since on the coarsest grid $V_0$ the subspace solver is exact, we use the elliptic projection $P_0$, satisfying

$$
a(P_0 v, w_0) = a(v, w_0) \quad \text{for all} \quad v \in V, \ w_0 \in V_0
$$

instead of $T_0$.

To show uniform convergence of the multilevel method we need the following result, referred to as the “XZ-identity”, in the form found in [23] or [7, Lemma 3.4].

**Lemma 6.1.** Assume that the preconditioner $B_{MG}$ is defined via Algorithm 6.1. Then, for $v_{\ell} \in V_{\ell}$, we have

$$
(6.4) \quad (B_{MG} v_{\ell}, v_{\ell}) = \inf_{\sum_{\ell} v_{\ell} = v_{\ell}} \sum_{\ell=0}^{L} a(T_{\ell}^{-1}(v_{\ell} + T_{\ell}^* w_{\ell}), v_{\ell} + T_{\ell}^* w_{\ell}),
$$

where $w_{\ell} = \sum_{i > \ell} v_i$.

In the above lemma, $T_{\ell}^*$ is the adjoint of $T_{\ell}$ with respect to the $a(., .)$-inner product, and, as we already observed in the case of the weighted Jacobi smoother, we have $T_{\ell}^* = T_{\ell}$. With the stability results established in section 4 it is then easy to see that the following convergence result follows directly from Lemma 6.1.

**Theorem 6.1.** Let us assume that assumption A5 holds for all $K \in \mathcal{T}_0$. Then we have the following estimate for all $v \in V_{L}$:

$$
(6.5) \quad a((I - B_{MG}^{-1} A)v, v) \leq 1 - \frac{1}{c},
$$

where $c \lesssim L (\max_{K \in \mathcal{T}_L} C_K^*)$ and the hidden constant in $\lesssim$ is independent of the PDE coefficient $\alpha$, of $L$, and of the mesh size $h$.

**Proof.** It follows from Lemma 6.1 that in order to prove (6.5) we need to show that

$$
(B_{MG} v, v) \leq c a(v, v).
$$

As in section 4, we define the following quasi-interpolants $\Pi_{\ell} : V_{L} \rightarrow V_{\ell}$, $\ell = 0, \ldots, L - 1$:

$$
\Pi_{\ell} v := \sum_{j=1}^{N_{\ell}} \lambda_j^\ell \Phi_j^\ell, \quad \text{where} \quad \lambda_j^\ell := \frac{f_{\omega_j}^\ell \alpha dx}{\int_{\omega_j}^\ell \alpha dx}
$$

and $\omega_j^\ell := \text{supp} \Phi_j^\ell$. We also set $\Pi_{L} := I$ and $\Pi_{-1} := 0$ and consider the decomposition

$$
(6.6) \quad V_h \ni v = \sum_{\ell=0}^{L} v_{\ell}, \quad \text{where} \quad v_{\ell} = (\Pi_{\ell} - \Pi_{\ell-1}) v.
$$

Note that this implies that $w_{\ell} = \sum_{i > \ell} v_i = (I - \Pi_{\ell}) v$ in Lemma 6.1.
Since \( T_\ell^* = T_\ell \) and the infimum in Lemma 6.1 is over all decompositions, it follows from (6.3) that with our specific choice of \( \{ v_\ell \} \) in (6.6)

\[
(B_{\text{MG}} v, v) \leq \sum_{\ell=0}^L a(S^{-1}_\ell (v_\ell + T_\ell w_\ell), v_\ell + T_\ell w_\ell)
\]

\[
(6.7)
\]

\[
\leq 2 \sum_{\ell=0}^L a(S^{-1}_\ell v_\ell, v_\ell) + 2 \sum_{\ell=0}^L a(S^{-1}_\ell T_\ell w_\ell, T_\ell w_\ell)
\]

\[
\leq 2 \sum_{\ell=0}^L \| (\Pi_\ell - \Pi_{\ell-1}) v \|_{\ell, \ell}^2 + 2 \sum_{\ell=0}^{L-1} a((I - \Pi_\ell) v, (I - \Pi_\ell) v).
\]

Now to bound the terms on the right side of (6.7) note first that it follows from the previous section via some simple numericalexperiments that are designed to verify the local quasi-uniformity of \( \Omega = (0,1)^3 \).

Applying (6.9) and (6.8) to each term on the right side of (6.7) completes the proof.

7. Numerical results. In this section we will confirm the theoretical results in the previous section via some simple numerical experiments that are designed to verify our assumptions and the statements made about the design of robust coarse spaces. We restrict ourselves for the most part to three dimensions and to problems on the unit cube \( \Omega = (0,1)^3 \). The multilevel preconditioner/method we use is a standard V-cycle geometric MG method with one pre- and one postsmoothing step, standard piecewise linear FE interpolation, and its adjoint as a restriction. The smoother is the symmetric Gauss–Seidel method.
The finest grid $\mathcal{T}_L$ is always a uniform grid obtained by $L$ refinements from the uniform simplicial grid $\widehat{\mathcal{T}}_0$, based on a uniform $6 \times 6 \times 6$ cubic grid as depicted in Figure 2 (left). Let (for simplicity) $h_L := 2^{-L/6}$ denote the mesh size of $\mathcal{T}_L$. In the majority of the examples we will choose $\mathcal{T}_0 = \widehat{\mathcal{T}}_0$ and use the sequence of grids obtained in the above refinement procedure as the intermediate coarse grids $\mathcal{T}_1, \ldots, \mathcal{T}_{L-1}$. However, in section 7.1 we will also introduce a different sequence of coarse grids that is locally refined near cross points (where the coefficient is only Type-0 quasi-monotone). The coarse grid matrices are always obtained via the Galerkin product.

In the tables below we will give estimates of condition numbers $\kappa$ and eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the preconditioned matrix $B_{MG}^{-1}A$ (numbered in ascending order). These are based on Ritz values obtained from applying the MG preconditioner within a conjugate gradient (CG) iteration with right-hand side zero and random initial guess. We will also give the number of preconditioned CG iterations ($\#\text{PCG}$) necessary to reduce the residual by a factor $10^{-8}$. An estimate of the MG V-cycle convergence rate can then be computed from the condition number estimate via $\rho = (\kappa - 1)/\kappa$. For certain examples we will also give the number of basic MG V-cycles ($\#\text{MG}$) necessary to reduce the residual by a factor $10^{-8}$ (without CG acceleration).

In all the examples the coefficient will be $\alpha = 1$ everywhere except in one or two islands where the coefficient will be $\alpha = \widehat{\alpha}$. These islands are in general resolved only on the finest grid. To set a familiar benchmark we first give results for $\alpha \equiv 1$ on all of $\Omega$, i.e., the three-dimensional Laplacian, in Table 1.

### 7.1. Suitable grid hierarchies for cross points.

In Table 2 we present the case of a (Type-0 quasi-monotone) three-dimensional cross point (cf. Figure 1(c)), where $\alpha = \widehat{\alpha}$ for $x \in \left(\frac{7}{24}, \frac{1}{2}\right) \times \left(\frac{1}{2}, \frac{23}{24}\right) \times \left(\frac{1}{2}, \frac{23}{24}\right) \cup \left(\frac{1}{2}, \frac{23}{24}\right) \times \left(\frac{1}{2}, \frac{23}{24}\right) \times \left(\frac{1}{2}, \frac{1}{2}\right)$ and 1 elsewhere. We see from the 4th column that with uniform coarse grids the condition number grows linearly with $h_0/h_L$, as predicted by our theory.

As suggested in section 3, a remedy for this lack of robustness is locally refined coarse grids near the three-dimensional cross point. Here, the locally refined coarse spaces (for the rightmost five columns in Table 2) were obtained by coarsening the

---

**Table 1**

<table>
<thead>
<tr>
<th>$L$</th>
<th>$N_L$</th>
<th>$N_0$</th>
<th>$\kappa$</th>
<th>$\rho$</th>
<th>$#\text{MG}$</th>
<th>$#\text{PCG}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$1.2 \times 10^4$</td>
<td>125</td>
<td>1.331</td>
<td>0.249</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>$1.0 \times 10^5$</td>
<td>125</td>
<td>1.365</td>
<td>0.267</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>$8.6 \times 10^5$</td>
<td>125</td>
<td>1.375</td>
<td>0.273</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>$7.0 \times 10^6$</td>
<td>125</td>
<td>1.379</td>
<td>0.275</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>
Three-dimensional cross point at \((\frac{1}{2}, \frac{1}{2}, \frac{1}{2})\). The coefficients are not resolved on \(T_0\) and \(T_1\). In the top table \(\hat{\alpha} = 10^4\). In the bottom table \(L = 4\).

<table>
<thead>
<tr>
<th>(L)</th>
<th>(N_L)</th>
<th>Uniform coarse grids</th>
<th>Locally refined coarse grids</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_0)</td>
<td>(\rho)</td>
<td>#MG</td>
<td>#PCG</td>
</tr>
<tr>
<td>2</td>
<td>(1.2 \times 10^4)</td>
<td>125</td>
<td>4.58</td>
</tr>
<tr>
<td>3</td>
<td>(1.0 \times 10^5)</td>
<td>125</td>
<td>9.62</td>
</tr>
<tr>
<td>4</td>
<td>(8.6 \times 10^5)</td>
<td>125</td>
<td>19.6</td>
</tr>
<tr>
<td>5</td>
<td>(7.0 \times 10^6)</td>
<td>125</td>
<td>38.2</td>
</tr>
</tbody>
</table>

The coefficients are not resolved on \(T_0\) and \(T_1\). In the top table \(\hat{\alpha} = 10^4\). In the bottom table \(L = 4\). The results in Table 2 confirm the theoretically predicted robustness of this coarsening procedure with no dependence on coefficient variation or mesh size ratio.

Proceeding like this all the way to level 0, we obtain a sequence of grids that are locally refined toward the center of the domain as depicted in Figure 2 (right) with coarse mesh size \(h_K \approx h_K\) for all \(K \in T_0\) locally near \((\frac{1}{2}, \frac{1}{2}, \frac{1}{2})\). The procedure also ensures that \(h_K\) grows only gradually away from \((\frac{1}{2}, \frac{1}{2}, \frac{1}{2})\), thus satisfying the local quasi-uniformity assumption A4. However, the mesh grading is geometric, and so the dimension \(N_0\) of the coarse problem grows only logarithmically with respect to the size \(N_L\) of the fine grid problem. Therefore, the grid complexity \(\sum_{\ell=0}^L N_\ell/N_L\) and the operator complexity \(\sum_{\ell=0}^L \#NNZ_\ell/\#NNZ_L\), where \#NNZ denotes the number of nonzero entries in the stiffness matrix on level \(\ell\), are virtually identical to those for uniform grids. For \(L = 4\) they change only from 1.1369 to 1.1372 and from 1.1353 to 1.1359, respectively. This remains true (asymptotically, as \(N_L \to \infty\)) also for more complicated problems with multiple cross points (at least in the case of a uniform fine mesh). The results in Table 2 confirm the theoretically predicted robustness of this coarsening procedure with no dependence on coefficient variation or mesh size ratio.

In Table 3 we see that a two-dimensional cross point is indeed much less troublesome. The example there is simply a projection of the problem in Table 2 to the \((x_2, x_3)\)-plane. We see that the growth of \(\kappa\) in two dimensions is indeed only logarithmic in \(h_0/h_f\) for uniform coarse grids, as predicted by our theory. Locally refined coarse grids, which can be obtained in the same way as in three dimensions, lead again to a fully robust method (although this may be unnecessary here). Similar behavior can be observed for Type-1 quasi-monotone coefficients in three dimensions.

### 7.2. Quasi-monotonicity and MG robustness

In Table 4 we confirm that quasi-monotonicity and \(\Gamma\)-quasi-monotonicity as defined in section 3 are necessary and...
Two-dimensional cross point at \((\frac{1}{5}, \frac{1}{5})\) with \(\tilde{\alpha} = 10^4\). The coefficients are not resolved on \(T_0\) and \(T_1\).

<table>
<thead>
<tr>
<th>(L)</th>
<th>(N_L)</th>
<th>Uniform coarse grids</th>
<th>Locally refined coarse grids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(N_0) (\lambda_1^{-1}) (\lambda_2^{-1}) (\rho) (# \text{PCG})</td>
<td>(N_0) (\lambda_1^{-1}) (\lambda_2^{-1}) (\rho) (# \text{PCG})</td>
</tr>
<tr>
<td>4</td>
<td>(9.0 \times 10^3)</td>
<td>25</td>
<td>2.73</td>
</tr>
<tr>
<td>5</td>
<td>(3.6 \times 10^4)</td>
<td>25</td>
<td>3.36</td>
</tr>
<tr>
<td>6</td>
<td>(1.5 \times 10^5)</td>
<td>25</td>
<td>4.08</td>
</tr>
<tr>
<td>7</td>
<td>(5.9 \times 10^5)</td>
<td>25</td>
<td>4.87</td>
</tr>
<tr>
<td>8</td>
<td>(2.4 \times 10^6)</td>
<td>25</td>
<td>5.74</td>
</tr>
</tbody>
</table>

Two islands with \(L = 4\). The coefficients are not resolved on \(T_0\) and \(T_1\). The coefficient \(\alpha(x)\) is quasi-monotone on \(\omega_K\), for all \(K \in T_0\), in the top left table. It fails to be quasi-monotone for some \(K \in T_0\) in the top right and bottom right tables. It fails to be \(\Gamma\)-quasi-monotone in both of the bottom tables.

<table>
<thead>
<tr>
<th>(\tilde{\alpha})</th>
<th>Quasi- and (\Gamma)-quasi-monotone</th>
<th>Only (\Gamma)-quasi-monotone</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10^3)</td>
<td>(\lambda_1^{-1}) (\lambda_2^{-1}) (\rho) (# \text{MG}) (# \text{PCG})</td>
<td>(\lambda_1^{-1}) (\lambda_2^{-1}) (\rho) (# \text{MG}) (# \text{PCG})</td>
</tr>
<tr>
<td>1.69</td>
<td>1.36</td>
<td>0.407</td>
</tr>
<tr>
<td>2.75</td>
<td>0.636</td>
<td>14</td>
</tr>
<tr>
<td>3.32</td>
<td>2.86</td>
<td>0.699</td>
</tr>
<tr>
<td>4.42</td>
<td>2.89</td>
<td>0.707</td>
</tr>
<tr>
<td>5.42</td>
<td>2.84</td>
<td>0.707</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\tilde{\alpha})</th>
<th>Only quasi-monotone</th>
<th>Neither quasi- nor (\Gamma)-quasi-monotone</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10^4)</td>
<td>(\lambda_1^{-1}) (\lambda_2^{-1}) (\rho) (# \text{MG}) (# \text{PCG})</td>
<td>(\lambda_1^{-1}) (\lambda_2^{-1}) (\rho) (# \text{MG}) (# \text{PCG})</td>
</tr>
<tr>
<td>33.6</td>
<td>1.81</td>
<td>0.970</td>
</tr>
<tr>
<td>319</td>
<td>1.82</td>
<td>0.997</td>
</tr>
<tr>
<td>3175</td>
<td>1.83</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Sufficient conditions for the robustness of classical geometric multigrid. We consider two isolated islands in \(\Omega\) where \(\alpha = \tilde{\alpha}\). The islands are \((\frac{5}{24}, \frac{11}{24}) \times (\frac{19}{24}, \frac{19}{24})\) and \((\frac{10}{24}, \frac{19}{24}) \times \frac{5}{24} \times \frac{11}{24}\) in the top left table, and \((\frac{5}{24}, \frac{11}{24}) \times (\frac{10}{24}, \frac{19}{24}) \times (\frac{5}{24}, \frac{11}{24})\) in the top right table. In the bottom two tables the only difference is that \(x_2 \notin (\frac{5}{24}, \frac{11}{24})\) instead of \((\frac{5}{24}, \frac{11}{24})\).

We see that standard geometric MG is robust only when the coefficient is quasi-monotone and \(\Gamma\)-quasi-monotone on \(\omega_K\) for all \(K \in T_0\). If either of these conditions is violated on any patch \(\omega_K\), then \(C^*_K\) and \(B_{MG}^{-1}\) grow linearly with the contrast \(\tilde{\alpha}\), and the MG convergence rate deteriorates rapidly.

Krylov methods such as CG still perform well in all cases, since there are at most two small eigenvalues of size \(\tilde{\alpha}^{-1}\) and the effective condition number is bounded. As mentioned in the introduction, this has already been pointed out in [24] for the case when the coarsest grid is aligned with the discontinuities in the coefficient. In our analysis we do not require any alignment of the coarser grids with the coefficient discontinuities. In addition, our numerical tests in Table 4 confirm the observation already made in [24] that the number of small eigenvalues is bounded by the number of disconnected regions \(\gamma_m\) where \(\alpha\) is large compared to the neighboring regions. Such observations are in turn again related to the local quasi-monotonicity and/or \(\Gamma\)-quasi-monotonicity of the coefficient. Note, however, that the method is
not fully robust and that the number of CG iterations grows linearly with the number of disconnected regions \( Y_m \) where \( \alpha \) is large. Repeating the experiment in the last row of the bottom right table of Table 4 with three and four islands, the number of PCG iterations grows from 17 to 23 and to 28, respectively. The behavior is identical for the fully resolved case.

### 7.3. Additional experiments.
Here we confirm first that (from an asymptotic point of view) it does not matter how many of the coarse grids are aligned with the coefficient and second that it is crucial to solve the problem on the coarsest grid exactly.

In the leftmost table in Table 5 we gradually change one island where \( \alpha = \tilde{\alpha} \) from being fully aligned on all coarse grids to not being aligned on any of the coarse grids. In the middle table we repeat the experiment with two islands that meet at \((\frac{1}{2}, \frac{1}{2}, \frac{1}{2})\) (three-dimensional cross point). We observe that aligning clearly has an effect on the constant, but asymptotically the method remains robust independent of the number \( \tilde{L} \) of grids on which the grid is not aligned with the coefficient.

In the rightmost table in Table 5 we see that it is crucial to solve the problem on the coarsest grid exactly in the case of highly varying coefficients. Otherwise the condition number and the MG convergence rate deteriorate with the size of the coefficient jump \( \tilde{\alpha} \). Note that this is not a consequence of the coarse grids not being aligned with the coefficient jumps. Such phenomena occur even in the fully resolved case, as demonstrated in Table 5.

### 8. Conclusions.
In this paper we have given a rigorous convergence analysis of multilevel iterative methods for elliptic problems with highly varying coefficients (in particular two-level additive Schwarz and geometric MG) that does not rely on aligning any coarse grids or subdomain partitions with large coefficient jumps or on building operator-dependent coarse spaces. The key result of the paper is a novel stability result for quasi-interpolation in coarse spaces which itself relies on novel uniform weighted Poincaré inequalities. In the context of standard piecewise linear coarse spaces, we are able to achieve such a generality under the mild assumption that the coarse grids are suitably refined in certain areas of the domain, such as near cross points. Some simple numerical experiments on academic examples confirm our theoretical claims. However, the theoretical results presented here provide ideas for various (rigorously justified) robust practical multilevel methods for elliptic problems with highly varying coefficients that can be used before resorting to potentially higher cost algebraic MG methods (such as the robust spectral algebraic MG methods) on coarser levels.
REFERENCES


