Nonoverlapping domain decomposition preconditioners for discontinuous Galerkin approximations of Hamilton–Jacobi–Bellman equations

Iain Smears

Nonoverlapping domain decomposition preconditioners for discontinuous Galerkin approximations of Hamilton–Jacobi–Bellman equations

Iain Smears

Received: date / Accepted: date

Abstract We analyse a class of nonoverlapping domain decomposition preconditioners for nonsymmetric linear systems arising from discontinuous Galerkin finite element approximation of fully nonlinear Hamilton–Jacobi–Bellman (HJB) partial differential equations. These nonsymmetric linear systems are uniformly bounded and coercive with respect to a related symmetric bilinear form, that is associated to a matrix $\mathbf{A}$. In this work, we construct a nonoverlapping domain decomposition preconditioner $\mathbf{P}$, that is based on $\mathbf{A}$, and we then show that the effectiveness of the preconditioner for solving the nonsymmetric problems can be studied in terms of the condition number $\kappa(\mathbf{P}^{-1}\mathbf{A})$. In particular, we establish the bound $\kappa(\mathbf{P}^{-1}\mathbf{A}) \lesssim 1 + p^6H^3/q^3h^3$, where H and h are respectively the coarse and fine mesh sizes, and q and p are respectively the coarse and fine mesh polynomial degrees. This represents the first such result for this class of methods that explicitly accounts for the dependence of the condition number on q; our analysis is founded upon an original optimal order approximation result between fine and coarse discontinuous finite element spaces. Numerical experiments demonstrate the sharpness of this bound. Although the preconditioners are not robust with respect to the polynomial degree, our bounds quantify the effect of the coarse and fine space polynomial degrees. Furthermore, we show computationally that these methods are effective in practical applications to nonsymmetric, fully nonlinear HJB equations under h-refinement for moderate polynomial degrees.

Keywords domain decomposition · preconditioners · GMRES · discontinuous Galerkin · finite element methods · approximation in discontinuous spaces · Hamilton–Jacobi–Bellman equations

Mathematics Subject Classification (2000) 65F10 · 65N22 · 65N55 · 65N30 · 35J66

1 Introduction

In [20–22], discontinuous Galerkin finite element methods (DGFEM) were introduced for the numerical solution of linear nondivergence form elliptic equations and fully nonlinear Hamilton–Jacobi–Bellman (HJB) equations with Cordes coefficients. In these applications,

Iain Smears
Inria Paris, 2 Rue Simone Iff, 75589, Paris, France. E-mail: iain.smears@inria.fr

the appropriate norm on the finite element space is a broken H^2-norm with penalization of the jumps in values and in first derivatives across the faces of the mesh. As a result, it is typical for the condition number of the discrete problems to be of order p^8/h^4, where h is the mesh size and p is the polynomial degree. The purpose of this work is to study the application of a commonly used class of nonoverlapping domain decomposition preconditioners to these problems.

Nonoverlapping domain decomposition methods, along with their overlapping counterparts, have been successfully developed for a range of applications of DGFEM by many authors [2–4, 6, 10, 11, 14]. In order to solve a problem on a fine mesh $\mathcal{T}_h$, these methods combine a coarse space solver, defined on a coarse mesh $\mathcal{T}_H$, with local fine mesh solvers, defined on a subdomain decomposition $\mathcal{T}_S$ of the domain Ω. The discontinuous nature of the finite element space leads to a significant flexibility in the choice of the decomposition $\mathcal{T}_S$, which can either be overlapping or nonoverlapping. As explained in the above references, these preconditioners possess many advantages in terms of simplicity and applicability, as they allow very general choices of basis functions, nonmatching meshes and varying element shapes, and are naturally suited for parallelization. It has been pointed out by various authors, such as Lasser and Toselli in [14, p. 1235], that nonoverlapping methods feature reduced inter-subdomain communication burdens, thus representing an advantage in parallel computations.

For problems involving H^1-type norms, such as divergence form second-order elliptic PDE, nonoverlapping additive Schwarz preconditioners for h-version methods [10] lead to condition numbers of order $1 + H/h$, where H is the coarse mesh size, while overlapping methods lead to a condition number of order $1 + H/\delta$, where δ is the subdomain overlap. For problems in H^2-type norms such as the biharmonic equation, the h-version analysis [11] leads to condition numbers of order $1 + H^3/h^3$. We remark that the analysis in these works leaves the polynomial degree implicit inside the generic constants. However an analysis that keeps track of all parameters is important in practice for determining their effect on the performance of the preconditioners, even if robustness of the condition number cannot be guaranteed. Antonietti and Houston [4] were the first to keep track of the dependence on the polynomial degrees for this class of preconditioners for problems in H^1-norms, where they showed a condition number bound of order $1 + p^2H/h$. However, their numerical experiments lead them to conjecture the improved bound of order $1 + p^2H/qh$, where q is the coarse space polynomial degree. This conjecture was recently proved in [5] using ideas first developed in this work.

As can be seen from the theoretical analysis in the above references, the effectiveness of the preconditioner depends in an essential way on the approximation properties between the coarse and fine spaces. In the analysis of h-version DGFEM, it is sufficient to consider low-order projection operators from the fine space to the coarse space; for example, coarse element mean-value projections are employed in [10] and local first-order elliptic projections are used in [11]. However, low-order projections lead to suboptimal bounds for the condition number with respect to polynomial degrees. This work resolves this suboptimality through an original optimal order approximation result between coarse and fine spaces.

There are further classes of preconditioners for p-version and hp-version methods for problems in H^1-norms that achieve condition numbers either independent or depending only polylogarithmically on the polynomial degree, such as Neumann–Neumann and FETI methods, see [17, 24] and the many references therein. We are aware of one work on generalising these methods to H^2-norm problems: Brenner and Wang [8] considered iterative substructuring methods for the h-version C^0 interior penalty discretizations of the biharmonic equation. They show that the usual choices of orthogonalised basis functions required

by these algorithms do not extend to the H^2-norm context, and that different basis functions must be used on different elements of the mesh. In comparison, the overlapping and nonoverlapping methods described above generalise straightforwardly to the H^2-norm context without additional difficulties. Moreover, a comparison of the computations in [7] and [8] suggests that the substructuring algorithms only yield a similar performance in practice to the two-level additive Schwarz methods for these problems.

1.1 Main results

The numerical scheme of [21] for fully nonlinear HJB equations leads to a discrete nonlinear problem that can be solved iteratively by a semismooth Newton method. The linear systems obtained from the Newton linearization are generally nonsymmetric but coercive with respect to a discrete H^2-type norm. In section 3, we apply existing GMRES convergence theory for SPD preconditioners [9, 15] to these nonsymmetric systems, leading to a guaranteed minimum convergence rate, with a contraction factor expressed in terms of the condition number $\kappa(\mathbf{P}^{-1}\mathbf{A})$, where $\mathbf{A}$ is the matrix of a related symmetric bilinear form that is spectrally equivalent to a discrete H^2-type norm, and where $\mathbf{P}$ is an arbitrary symmetric positive definite preconditioner. Thus, the construction and analysis of preconditioners for a symmetric problem can be used for the solution of the nonsymmetric systems appearing in applications to HJB equations [21]. A further benefit is that the preconditioner does not require re-assembly at each new semismooth Newton iteration.

Section 4 presents the specific construction of a nonoverlapping additive Schwarz preconditioner $\mathbf{P}$ based on $\mathbf{A}$, and sections 5 and 6 show the condition number bound

$$\kappa(\mathbf{P}^{-1}\mathbf{A}) \lesssim 1 + \frac{p^2 H}{q\, h} + \frac{p^6\, H^3}{q^3\, h^3}. \tag{1.1}$$

In comparison to the existing literature, this is the first bound for this class of preconditioners that explicity accounts for the coarse mesh polynomial degree. Unfortunately, (1.1) implies that this standard class of preconditioners cannot be expected to be robust with respect to the polynomial degree. Nevertheless, our result shows that the coarse space polynomial degree can contribute significantly to reducing the condition number.

The central original result underpinning our analysis is Theorem 4 of section 5, which shows that for any $v_h \in V_{h,\mathbf{p}}$, there is a function $v \in H^2(\Omega) \cap H^1_0(\Omega)$ such that

$$\|v_h - v\|_{L^2(\Omega)} + \frac{h}{p}\|v_h - v\|_{H^1(\Omega;\mathcal{T}_h)} \lesssim \frac{h^2}{p^2}|v_h|_{\mathrm{J},h}, \quad \|v\|_{H^2(\Omega)} \lesssim \|v_h\|_{2,h}, \tag{1.2}$$

where the piecewise Sobolev norms $\|\cdot\|_{H^s(\Omega;\mathcal{T}_h)}$, jump seminorm $|\cdot|_{\mathrm{J},h}$, and the discrete H^2-type norm $\|\cdot\|_{2,h}$ are defined in section 2. This result is a natural converse to classical direct approximation theory, since, here, the nonsmooth function from the discrete space $V_{h,\mathbf{p}}$ is approximated by a smoother function from an infinite dimensional space. It follows from (1.2) that there exists a function v_H in the coarse space $V_{H,\mathbf{q}}$, of polynomials of degree q on $\mathcal{T}_H$, such that

$$\|v_h - v_H\|_{H^k(\Omega;\mathcal{T}_h)} \lesssim \frac{H^{2-k}}{q^{2-k}}\|v_h\|_{2,h}, \quad k \in \{0,1,2\}, \tag{1.3}$$

thus yielding an approximation between coarse and fine meshes that is optimal in the orders of the mesh size and the polynomial degree. The approximation result is used to show the

stable decomposition property for the additive Schwarz preconditioner in section 6, thereby leading to the spectral bound (1.1).

The first numerical experiment, in section 7.1, confirms that (1.1) is sharp with respect to the orders in the polynomial degrees. The experiment of section 7.2 compares nonoverlapping methods with their overlapping counterparts, where it is found that they are competitive in both iteration counts and computational cost. Despite the polynomial degree suboptimality of these preconditioners, in section 7.3 we show computationally that for h-refinement, nonoverlapping methods can be efficient and competitive in challenging applications to fully nonlinear HJB equations.

2 Definitions

For real numbers a and b, we shall write $a \lesssim b$ to signify that there is a positive constant C such that $a \leq Cb$, where C is independent of the quantities of interest, such as the element sizes and polynomial degrees, but possibly dependent on other quantities, such as the mesh regularity parameters.

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a bounded convex polytopal domain. Note that convexity of Ω implies that the boundary $\partial\Omega$ of Ω is Lipschitz [13]. Let $\{\mathcal{T}_h\}_h$ be a sequence of shape-regular meshes on Ω, consisting of simplices or parallelepipeds. For each element $K \in \mathcal{T}_h$, let $h_K \coloneqq \operatorname{diam} K$. It is assumed that $h = \max_{K\in\mathcal{T}_h} h_K$ for each mesh $\mathcal{T}_h$. Let $\mathcal{F}_h^i$ denote the set of interior faces of the mesh $\mathcal{T}_h$ and let $\mathcal{F}_h^b$ denote the set of boundary faces. The set of all faces of $\mathcal{T}_h$ is denoted by $\mathcal{F}_h^{i,b} \coloneqq \mathcal{F}_h^i \cup \mathcal{F}_h^b$. Since each element has piecewise flat boundary, the faces may be chosen to be flat. For $K \in \mathcal{T}_h$ or $F \in \mathcal{F}_h^{i,b}$, we use $\langle\cdot,\cdot\rangle_K$, respectively $\langle\cdot,\cdot\rangle_F$, to denote the L^2-inner product over K, respectively F, of scalar functions, vector fields, and higher-order tensors.

Mesh conditions The meshes are allowed to be nonmatching, i.e. there may be hanging nodes. We assume that there is a uniform upper bound on the number of faces composing the boundary of any given element; in other words, there is a $c_{\mathcal{F}} > 0$, independent of h, such that

$$\max_{K\in\mathcal{T}_h} \operatorname{card}\{F \in \mathcal{F}_h^{i,b} \colon F \subset \partial K\} \leq c_{\mathcal{F}} \qquad \forall K \in \mathcal{T}_h. \tag{2.1}$$

It is also assumed that any two elements sharing a face have commensurate diameters, i.e. there is a $c_{\mathcal{T}} \geq 1$, independent of h, such that

$$\max(h_K, h_{K'}) \leq c_{\mathcal{T}} \min(h_K, h_{K'}), \tag{2.2}$$

for any K and K' in $\mathcal{T}_h$ that share a face. For each h, let $\mathbf{p} \coloneqq (p_K \colon K \in \mathcal{T}_h)$ be a vector of positive integers; note that this requires $p_K \geq 1$ for all $K \in \mathcal{T}_h$. We make the assumption that $\mathbf{p}$ has *local bounded variation*: there is a $c_{\mathcal{P}} \geq 1$, independent of h, such that

$$\max(p_K, p_{K'}) \leq c_{\mathcal{P}} \min(p_K, p_{K'}), \tag{2.3}$$

for any K and K' in $\mathcal{T}_h$ that share a face.

Function spaces For each $K \in \mathcal{T}_h$, let $\mathcal{P}_{p_K}(K)$ be the space of all real-valued polynomials in $\mathbb{R}^d$ with either total or partial degree at most p_K. In particular, we allow the combination of spaces of polynomials of fixed total degree on some parts of the mesh with spaces of polynomials of fixed partial degree on the remainder. We also allow the use of the space of polynomials of total degree at most p_K even when K is a parallelepiped. The discontinuous Galerkin finite element spaces $V_{h,\mathbf{p}}$ are defined by

$$V_{h,\mathbf{p}} \coloneqq \left\{ v \in L^2(\Omega) \colon \ v|_K \in \mathcal{P}_{p_K}(K), \ \forall K \in \mathcal{T}_h \right\}. \tag{2.4}$$

Let $\mathbf{s} \coloneqq (s_K \colon K \in \mathcal{T}_h)$ denote a vector of non-negative real numbers. The broken Sobolev space $H^{\mathbf{s}}(\Omega; \mathcal{T}_h)$ is defined by

$$H^{\mathbf{s}}(\Omega; \mathcal{T}_h) \coloneqq \left\{ v \in L^2(\Omega) \colon \ v|_K \in H^{s_K}(K), \ \forall K \in \mathcal{T}_h \right\}. \tag{2.5}$$

For $s \geq 0$, we set $H^s(\Omega; \mathcal{T}_h) \coloneqq H^{\mathbf{s}}(\Omega; \mathcal{T}_h)$, where $s_K = s$ for all $K \in \mathcal{T}_h$. The norm $\|\cdot\|_{H^{\mathbf{s}}(\Omega;\mathcal{T}_h)}$ and semi-norm $|\cdot|_{H^{\mathbf{s}}(\Omega;\mathcal{T}_h)}$ are defined on $H^{\mathbf{s}}(\Omega; \mathcal{T}_h)$ as

$$\|v\|_{H^{\mathbf{s}}(\Omega;\mathcal{T}_h)} \coloneqq \left(\sum_{K \in \mathcal{T}_h} \|v\|^2_{H^{s_K}(K)} \right)^{\frac{1}{2}}, \quad |v|_{H^{\mathbf{s}}(\Omega;\mathcal{T}_h)} \coloneqq \left(\sum_{K \in \mathcal{T}_h} |v|^2_{H^{s_K}(K)} \right)^{\frac{1}{2}}. \tag{2.6}$$

For a function $v_h \in V_{h,\mathbf{p}}$, the element-wise gradient $\nabla v_h|_K$ and the Hessian $D^2 v_h|_K$ are well-defined for all $K \in \mathcal{T}_h$ since v_h is smooth on K. Thus expressions such as $\langle D^2 u_h, D^2 v_h \rangle_K$ are well-defined for all $K \in \mathcal{T}_h$ and all $u_h, v_h \in V_{h,\mathbf{p}}$.

Jump and average operators For each face $F \in \mathcal{F}_h^{i,b}$, let $n_F \in \mathbb{R}^d$ denote a *fixed* choice of a unit normal vector to F. Since F is flat, n_F is constant over F. Let K be an element of $\mathcal{T}_h$ for which $F \subset \partial K$; then n_F is either inward or outward pointing with respect to K. Let $\tau_F \colon H^s(K) \to H^{s-1/2}(F)$, $s > 1/2$, denote the trace operator from K to F, and let τ_F be extended componentwise to vector-valued functions.

For each face F, define the jump operator $[\![\cdot]\!]$ and the average operator $\{\cdot\}$ by

$$\begin{aligned}
[\![\phi]\!] &\coloneqq \tau_F \left(\phi|_{K_{\mathrm{ext}}} - \phi|_{K_{\mathrm{int}}} \right), \quad & \{\phi\} &\coloneqq \frac{1}{2} \tau_F \left(\phi|_{K_{\mathrm{ext}}} + \phi|_{K_{\mathrm{int}}} \right), \quad & &\text{if } F \in \mathcal{F}_h^i, \\
[\![\phi]\!] &\coloneqq \tau_F \left(\phi|_{K_{\mathrm{ext}}} \right), & \{\phi\} &\coloneqq \tau_F \left(\phi|_{K_{\mathrm{ext}}} \right), & &\text{if } F \in \mathcal{F}_h^b,
\end{aligned}$$

where ϕ is a sufficiently regular scalar or vector-valued function, and K_{ext} and K_{int} are the elements to which F is a face, i.e. $F = \partial K_{\mathrm{ext}} \cap \partial K_{\mathrm{int}}$. Here, the labelling is chosen so that n_F is outward pointing with respect to K_{ext} and inward pointing with respect to K_{int}. Using this notation, the jump and average of scalar-valued functions, resp. vector-valued, are scalar-valued, resp. vector-valued.

Tangential differential operators For $F \in \mathcal{F}_h^{i,b}$, let $H^s_{\mathrm{T}}(F)$ denote the space of H^s-regular tangential vector fields on F, thus $H^s_{\mathrm{T}}(F) \coloneqq \{v \in H^s(F)^d \colon v \cdot n_F = 0 \text{ on } F\}$. We define the tangential gradient $\nabla_{\mathrm{T}} \colon H^s(F) \to H^{s-1}_{\mathrm{T}}(F)$ and the tangential divergence $\mathrm{div}_{\mathrm{T}} \colon H^s_{\mathrm{T}}(F) \to H^{s-1}(F)$, where $s \geq 1$, following [13]. Let $\{t_i\}_{i=1}^{d-1} \subset \mathbb{R}^d$ be an orthonormal coordinate system on F. Then, for $u \in H^s(F)$ and $v \in H^s_{\mathrm{T}}(F)$ such that $v = \sum_{i=1}^{d-1} v_i \, t_i$, with $v_i \in H^s(F)$ for $i = 1, \ldots, d-1$, we define

$$\nabla_{\mathrm{T}} u \coloneqq \sum_{i=1}^{d-1} t_i \frac{\partial u}{\partial t_i}, \qquad \mathrm{div}_{\mathrm{T}} v \coloneqq \sum_{i=1}^{d-1} \frac{\partial v_i}{\partial t_i}. \tag{2.7}$$

Mesh-dependent norms In the following, we let u_h and v_h denote functions in $V_{h,\mathbf{p}}$. For face-dependent positive real numbers μ_F and η_F, let the jump stabilization bilinear form $J_h\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}}$ be defined by

$$J_h(u_h, v_h) \coloneqq \sum_{F\in\mathcal{F}_h^i} \mu_F \langle [\![\nabla u_h \cdot n_F]\!], [\![\nabla v_h \cdot n_F]\!] \rangle_F + \sum_{F\in\mathcal{F}_h^{i,b}} \big[\mu_F \langle [\![\nabla_{\mathrm{T}} u_h]\!], [\![\nabla_{\mathrm{T}} v_h]\!] \rangle_F + \eta_F \langle [\![u_h]\!], [\![v_h]\!] \rangle_F\big]. \quad (2.8)$$

Define the jump seminorm $|\cdot|_{\mathrm{J},h}$ and the mesh-dependent norm $\|\cdot\|_{2,h}$ on $V_{h,\mathbf{p}}$ by

$$|v_h|_{\mathrm{J},h}^2 \coloneqq J_h(v_h, v_h), \qquad \|v_h\|_{2,h}^2 \coloneqq \sum_{K\in\mathcal{T}_h} \|v_h\|_{H^2(K)}^2 + |v_h|_{\mathrm{J},h}^2. \quad (2.9)$$

For each face $F \in \mathcal{F}_h^{i,b}$, define

$$\tilde{h}_F \coloneqq \begin{cases} \min(h_K, h_{K'}), & \text{if } F \in \mathcal{F}_h^i, \\ h_K, & \text{if } F \in \mathcal{F}_h^b, \end{cases} \qquad \tilde{p}_F \coloneqq \begin{cases} \max(p_K, p_{K'}), & \text{if } F \in \mathcal{F}_h^i, \\ p_K, & \text{if } F \in \mathcal{F}_h^b, \end{cases} \quad (2.10)$$

where K and K' are such that $F = \partial K \cap \partial K'$ if $F \in \mathcal{F}_h^i$ or $F \subset \partial K \cap \partial\Omega$ if $F \in \mathcal{F}_h^b$. The assumptions on the mesh and the polynomial degrees, in particular (2.2) and (2.3), show that if F is a face of an element K, then $h_K \le c_{\mathcal{T}}\, \tilde{h}_F$ and $\tilde{p}_F \le c_{\mathcal{P}}\, p_K$. Henceforth, it is assumed that the parameters μ_F and η_F in (2.8) are given by

$$\mu_F \coloneqq c_\mu \frac{\tilde{p}_F^2}{\tilde{h}_F}, \quad \eta_F \coloneqq c_\eta \frac{\tilde{p}_F^6}{\tilde{h}_F^3} \quad \forall\, F \in \mathcal{F}_h^{i,b}, \quad (2.11)$$

where c_μ and c_η are fixed positive constants independent of h and $\mathbf{p}$.

Approximation Under the hypothesis of shape-regularity of $\{\mathcal{T}_h\}$, for any function $u \in H^{\mathbf{s}}(\Omega; \mathcal{T}_h)$, there exists an approximation $\Pi_h u \in V_{h,\mathbf{p}}$, such that for each element $K \in \mathcal{T}_h$,

$$\|u - \Pi_h u\|_{H^r(K)} \lesssim \frac{h_K^{\min(s_K,\, p_K+1)-r}}{p_K^{s_K-r}} \|u\|_{H^{s_K}(K)} \quad \forall\, r,\ 0 \le r \le s_K, \quad (2.12a)$$

and, if $s_K > 1/2$,

$$\|D^\alpha (u - \Pi_h u)\|_{L^2(\partial K)} \lesssim \frac{h_K^{\min(s_K,\, p_K+1)-|\alpha|-1/2}}{p_K^{s_K-|\alpha|-1/2}} \|u\|_{H^{s_K}(K)} \quad \forall \alpha,\ |\alpha| \le k, \quad (2.12b)$$

where k is the greatest non-negative integer strictly less than $s_K - 1/2$. The constants in (2.12a) and (2.12b) do not depend on u, K, p_K, h_K or r, but depend possibly on $\max_{K\in\mathcal{T}_h} s_K$. Vector fields can be approximated componentwise.

3 HJB equations

We consider fully nonlinear HJB equations of the form

$$\begin{aligned} \sup_{\alpha\in\Lambda} [L^\alpha u - f^\alpha] = 0 \quad &\text{in } \Omega, \\ u = 0 \quad &\text{on } \partial\Omega, \end{aligned} \tag{3.1}$$

where Ω is a bounded convex domain, Λ is a compact metric space, and the operators L^α are given by

$$L^\alpha v \coloneqq a^\alpha \colon D^2 v + b^\alpha \cdot \nabla v - c^\alpha\, v, \quad v \in H^2(\Omega),\ \alpha \in \Lambda. \tag{3.2}$$

For simplicity of presentation here, we restrict our attention to the case $b^\alpha \equiv 0$, $c^\alpha \equiv 0$, and refer the reader to [21] for the general case. The matrix-valued function a and the scalar function f are assumed to be continuous on $\overline{\Omega}\times\Lambda$, and a is assumed to be uniformly elliptic, uniformly over $\overline{\Omega}\times\Lambda$. The PDE in (3.1) is fully nonlinear in the sense that the Hessian of the unknown solution appears inside the nonlinear term in (3.1). As a result of the nonlinearity, no weak form of the equation is available: this has constituted a long-standing difficulty in the development of high-order methods for this class of problems.

However, provided that the coefficients of L^α satisfy the Cordes condition, which, in the case of pure diffusion, requires that there exist $\varepsilon \in (0, 1]$ such that

$$\frac{|a^\alpha(x)|^2}{(\operatorname{Tr} a^\alpha(x))^2} \le \frac{1}{d-1+\varepsilon} \quad \forall\, \alpha \in \Lambda,\ \forall\, x \in \Omega. \tag{3.3}$$

then the boundary-value problem (3.1) has a unique solution in $H^2(\Omega) \cap H^1_0(\Omega)$, see [21, Theorem 3]. Observe that for problems in two spatial dimensions, condition (3.3) is equivalent to uniform ellipticity.

Defining the operator $F_\gamma[u] \coloneqq \sup_{\alpha\in\Lambda} [\gamma^\alpha (L^\alpha u - f^\alpha)]$, where $\gamma^\alpha = \operatorname{Tr} a^\alpha / |a^\alpha|^2$, the numerical scheme of [21] for solving (3.1) associated to a homogeneous Dirichlet boundary condition is to find $u_h \in V_{h,\mathbf{p}}$ such that

$$\mathcal{A}_h(u_h; v_h) = 0 \quad \forall\, v_h \in V_{h,\mathbf{p}}, \tag{3.4}$$

where the nonlinear form $\mathcal{A}_h$ is defined in [21, Eq. (5.3)], and can be equivalently given as

$$\begin{aligned} \mathcal{A}_h(u_h; v_h) \coloneqq &\sum_{K\in\mathcal{T}_h} \langle F_\gamma[u_h], \Delta v_h\rangle_K \\ &+ \frac{1}{2}\left(a_h(u_h, v_h) - \sum_{K\in\mathcal{T}_h} \langle \Delta u_h, \Delta v_h\rangle_K + J_h(u_h, v_h) \right), \end{aligned} \tag{3.5}$$

where the bilinear form $a_h \colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ is defined by

$$\begin{aligned} a_h(u_h, v_h) \coloneqq &\sum_{K\in\mathcal{T}_h} \langle D^2 u_h, D^2 v_h\rangle_K + J_h(u_h, v_h) \\ &+ \sum_{F\in\mathcal{F}_h^i} \left[\langle \operatorname{div}_{\mathrm{T}} \nabla_{\mathrm{T}} \{u_h\}, [\![\nabla v_h \cdot n_F]\!]\rangle_F + \langle \operatorname{div}_{\mathrm{T}} \nabla_{\mathrm{T}} \{v_h\}, [\![\nabla u_h \cdot n_F]\!]\rangle_F \right] \\ &- \sum_{F\in\mathcal{F}_h^{i,b}} \left[\langle \nabla_{\mathrm{T}} \{\nabla u_h \cdot n_F\}, [\![\nabla_{\mathrm{T}} v]\!]\rangle_F + \langle \nabla_{\mathrm{T}} \{\nabla v_h \cdot n_F\}, [\![\nabla_{\mathrm{T}} v_h]\!]\rangle_F \right]. \end{aligned} \tag{3.6}$$

3.1 Semismooth Newton method

In [21, Section 8], it is shown that the discretized nonlinear problem (3.4) can be solved by a semismooth Newton method, which leads to a sequence of nonsymmetric but positive definite linear systems to be solved at each iteration. We summarize here the essential ideas on the semismooth Newton, and refer the reader to [21] for the complete analysis.

For $x \in \Omega$ and $M \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$, define $F_\gamma(x, M) \coloneqq \sup_{\alpha\in\Lambda}[\gamma^\alpha(x)\,(a^\alpha(x){:}M - f^\alpha(x))]$, and let $\Lambda(x, M)$ denote the set of all $\alpha \in \Lambda$ that attain the supremum in $F_\gamma(x, M)$; note that $\Lambda(x, M)$ is always a non-empty subset of Λ due to the compactness of Λ and the continuity of the functions a, f and γ over $\overline{\Omega} \times \Lambda$. This defines a set-valued mapping $(x, M) \mapsto \Lambda(x, M)$. For a function $v \in H^2(\Omega; \mathcal{T}_h)$, let $\Lambda[v]$ denote the set of all Lebesgue measurable mappings $\alpha(\cdot)\colon \Omega \to \Lambda$ that satisfy $\alpha(x) \in \Lambda(x, D^2v(x))$ for almost every $x \in \Omega$; in [21, Theorem 10], it is shown that $\Lambda[v]$ is non-empty for any $v \in H^2(\Omega; \mathcal{T}_h)$.

The semismooth Newton method is now defined as follows. Start by choosing an initial iterate $u_h^0 \in V_{h,\mathbf{p}}$. Then, for each nonnegative integer j, given the previous iterate $u_h^j \in V_{h,\mathbf{p}}$, choose an $\alpha_j \in \Lambda[u_h^j]$. Next, the function $f^{\alpha_j}\colon \Omega \to \mathbb{R}$ is defined by $f^{\alpha_j}\colon x \mapsto f^{\alpha_j(x)}(x)$; the functions a^{α_j} and γ^{α_j} are defined in a similar way. Note that the measurability of the mappings α_j ensures the measurability of f^{α_j}, a^{α_j} and γ^{α_j}. Then, find the solution $u_h^{j+1} \in V_{h,\mathbf{p}}$ of the linearized system

$$B_h^j(u_h^{j+1}, v_h) = \sum_{K\in\mathcal{T}_h} \langle \gamma^{\alpha_j} f^{\alpha_j}, \Delta v_h\rangle_K \quad \forall\, v_h \in V_{h,\mathbf{p}}, \tag{3.7}$$

where the bilinear form $B_h^j\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ is defined by

$$\begin{aligned} B_h^j(w_h, v_h) \coloneqq &\sum_{K\in\mathcal{T}_h} \langle \gamma^{\alpha_j} a^{\alpha_j}{:}D^2 w_h, \Delta v_h\rangle_K \\ &+ \frac{1}{2}\left(a_h(u_h, v_h) - \sum_{K\in\mathcal{T}_h} \langle \Delta u_h, \Delta v_h\rangle_K + J_h(u_h, v_h)\right), \end{aligned} \tag{3.8}$$

In [21, Theorem 11], it was shown that $u_h^j \to u$ as $j \to \infty$ for a sufficiently close initial guess u_h^0, and moreover that the convergence is superlinear. It was also shown that the bilinear forms B_h^j are uniformly bounded and coercive in an H^2-type norm, with constants independent of the iterates. Since the preconditioners of this work take advantage of the coercivity of the B_h^j, we summarize the relevant results in the following lemma.

Lemma 1 *Let Ω be a bounded convex polytopal domain and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of meshes satisfying* (2.1)*. Let the bilinear forms B_h^j be defined by* (3.8)*. Then, there exist positive constants $\underline{c}_\mu$ and $\underline{c}_\eta$ such that if $c_\mu \geq \underline{c}_\mu$ and $c_\eta \geq \underline{c}_\eta$, then the bilinear forms a_h and B_h^j are uniformly coercive: for all v_h, $w_h \in V_{h,\mathbf{p}}$, we have*

$$\|v_h\|_{2,h}^2 \lesssim a_h(v_h, v_h), \qquad |a_h(v_h, w_h)| \lesssim \|v_h\|_{2,h}\|w_h\|_{h,2}, \tag{3.9}$$

$$\|v_h\|_{2,h}^2 \lesssim B_h^j(v_h, v_h), \qquad |B_h^j(v_h, w_h)| \lesssim \|v_h\|_{2,h}\|w_h\|_{2,h}, \tag{3.10}$$

where the constants are independent of the sequence $\{u_h^j\}_{j=0}^\infty$ and of the choice of the mappings $\alpha_j \in \Lambda[u_h^j]$ for each $j \geq 0$.

Proof First we prove (3.9). The continuity bound in (3.9) is a straightforward consequence of the trace and inverse inequalities. To show the coercivity bound, we first show that

$$\|v_h\|_{2,h}^2 \lesssim \sum_{K\in\mathcal{T}_h} \|D^2 v_h\|_{L^2(K)}^2 + |v_h|_{\mathrm{J},h}^2 =: |v_h|_{h,2}^2. \tag{3.11}$$

For any $v_h \in V_{h,\mathbf{p}}$, integration by parts gives

$$\sum_{K\in\mathcal{T}_h} \|\nabla v_h\|_{L^2(K)}^2 = \sum_{K\in\mathcal{T}_h} \langle v_h, -\Delta v_h\rangle_K + \sum_{F\in\mathcal{F}_h^{i,b}} \langle [\![v_h]\!], \{\nabla v_h\cdot n_F\}\rangle_F + \sum_{F\in\mathcal{F}_h^{i}} \langle \{v_h\}, [\![\nabla v_h\cdot n_F]\!]\rangle_F. \tag{3.12}$$

Hence, the trace and inverse inequalities imply that

$$\sum_{K\in\mathcal{T}_h} \|\nabla v_h\|_{L^2(K)}^2 \lesssim \|v_h\|_{L^2(\Omega)} |v_h|_{h,2}. \tag{3.13}$$

We recall the broken Poincaré inequality

$$\|v_h\|_{L^2(\Omega)}^2 \lesssim \sum_{K\in\mathcal{T}_h} \|\nabla v_h\|_{L^2(K)}^2 + \sum_{F\in\mathcal{F}_h^{i,b}} \frac{1}{\tilde{h}_F} \|[\![v_h]\!]\|_{L^2(F)}^2. \tag{3.14}$$

Therefore it follows from (3.13) that we have $\sum_{K\in\mathcal{T}_h} \|v_h\|_{H^1(K)}^2 \lesssim |v_h|_{h,2}^2$, from which we deduce (3.11). The proof of (3.9) is now completed by noting that [20, Lemma 7] implies that there exist $\underline{c}_\mu$ and $\underline{c}_\eta$ such that $|v_h|_{h,2}^2 \lesssim a_h(v_h, v_h)$, whenever $c_\mu \geq \underline{c}_\mu$ and $c_\eta \geq \underline{c}_\eta$, since a_h equals the bilinear form denoted by $B_{\mathrm{DG}(1)}$ in the notation of [20, Lemma 7]. The continuity bound for B_h^j in (3.10) can also be shown straightforwardly through the Cauchy–Schwarz inequality with the trace and inverse inequalities, where we note that the functions γ^{α_j}, respectively a^{α_j}, appearing in (3.8) are uniformly bounded in L^∞ by $\|\gamma\|_{C(\overline{\Omega}\times\Lambda)}$, respectively by $\|a\|_{C(\overline{\Omega}\times\Lambda;\mathbb{R}^{d\times d})}$; this implies that the continuity constants in (3.10) can be taken to be independent of the $\{u_h^j\}_{j=0}^\infty$. The coercivity bound in (3.10) was shown in [20, Theorem 8] and [21, Eq. (8.5)], where it is seen that the coercivity constant is independent of the iteration count j, but otherwise may depend on the constant ε from (3.3) and the choice of the penalty parameters c_μ and c_η. □

3.2 Iterative solution by the preconditioned GMRES method

Each step of the semismooth Newton method requires the solution of (3.7). These linear systems have a common general form, which consists of finding $\tilde{u}_h \in V_{h,\mathbf{p}}$ such that

$$B_h(\tilde{u}_h, v_h) = \ell_h(v_h) \quad \forall\, v_h \in V_{h,\mathbf{p}}, \tag{3.15}$$

where we shall henceforth omit to denote the dependence of the bilinear form B_h and of the right-hand side ℓ_h on the iteration number of the semismooth Newton method. It follows from Lemma 1 that there exist positive constants c_B and C_B such that, for any v_h and $w_h \in V_{h,\mathbf{p}}$,

$$a_h(v_h, v_h) \leq \frac{1}{c_B} B_h(v_h, v_h), \quad |B_h(v_h, w_h)| \leq C_B \sqrt{a_h(v_h, v_h)}\sqrt{a_h(w_h, w_h)}, \tag{3.16}$$

where c_B and C_B are independent of the iteration count of the semismooth Newton method and the discretization parameters. Therefore the sequence of linearisations of (3.4) are uniformly bounded and coercive with respect to the norm defined by the bilinear form a_h.

The coercivity and boundedness of B_h imply that an efficient preconditioner for a_h can also be used effectively as a preconditioner for the GMRES algorithm applied to (3.15). Indeed, assume that $\mathbf{P}$ is an SPD preconditioner for the matrix $\mathbf{A} \coloneqq (a_h(\phi_i, \phi_j))$ that satisfies

$$0 < c_{\mathbf{P}} \leq \frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\mathbf{v}^\top \mathbf{P}\mathbf{v}} \leq C_{\mathbf{P}} \quad \forall\, \mathbf{v} \in \mathbb{R}^{\dim V_{h,\mathbf{P}}} \setminus \{0\}, \tag{3.17}$$

where we assume that $c_{\mathbf{P}}$ and $C_{\mathbf{P}}$ are the best possible constants in (3.17). Thus the condition number $\kappa(\mathbf{P}^{-1}\mathbf{A}) = C_{\mathbf{P}}/c_{\mathbf{P}}$. Let the matrix $\mathbf{B} \coloneqq (B_h(\phi_j, \phi_i))$. Then, the preconditioner $\mathbf{P}$ can be used in either the right or left preconditioned GMRES method [18, 19] for solving (3.15) as follows. First, we define the norms $\|\cdot\|_{\mathbf{P}}$ and $\|\cdot\|_{\mathbf{P}^{-1}}$ on $\mathbb{R}^{\dim V_{h,\mathbf{P}}}$ by

$$\|\mathbf{v}\|_{\mathbf{P}}^2 \coloneqq \mathbf{v}^\top \mathbf{P}\, \mathbf{v}, \quad \|\mathbf{v}\|_{\mathbf{P}^{-1}}^2 \coloneqq \mathbf{v}^\top \mathbf{P}^{-1}\, \mathbf{v} \qquad \forall\, \mathbf{v} \in \mathbb{R}^{\dim V_{h,\mathbf{P}}}. \tag{3.18}$$

Applying k-steps of the right preconditioned GMRES method in the $\mathbf{P}^{-1}$-inner product computes $\mathbf{u}_k$ as the solution of

$$\mathbf{u}_k = \mathbf{P}^{-1}\mathbf{w}_k, \quad \mathbf{w}_k = \underset{\tilde{\mathbf{w}}_k \in \mathcal{K}_k(\mathbf{B}\mathbf{P}^{-1}, \mathbf{r}_0) + \mathbf{w}_0}{\operatorname{argmin}} \|\mathbf{B}\mathbf{P}^{-1}\,(\mathbf{w} - \tilde{\mathbf{w}}_k)\|_{\mathbf{P}^{-1}}, \tag{3.19}$$

where $\mathbf{r}_0$ denotes the initial residual, $\mathbf{w} \coloneqq \mathbf{P}\mathbf{u}$, $\mathbf{w}_0 \coloneqq \mathbf{P}\mathbf{u}_0$, and where $\mathcal{K}(\mathbf{B}\mathbf{P}^{-1}, \mathbf{r}_0)$ denotes the k-dimensional Krylov subspace generated by $\mathbf{B}\mathbf{P}^{-1}$ and $\mathbf{r}_0$. It is well-known that (3.19) is equivalent to

$$\mathbf{u}_k = \underset{\tilde{\mathbf{u}}_k \in \mathcal{K}_k(\mathbf{P}^{-1}\mathbf{B}, \mathbf{P}^{-1}\mathbf{r}_0) + \mathbf{u}_0}{\operatorname{argmin}} \|\mathbf{P}^{-1}\mathbf{B}\,(\mathbf{u} - \tilde{\mathbf{u}}_k)\|_{\mathbf{P}}, \tag{3.20}$$

which is obtained after k-steps of the left-preconditioned GMRES algorithm in the $\mathbf{P}$-inner product, see [18]. For a discussion of the implementation of the preconditioned GMRES method in the $\mathbf{P}$- and $\mathbf{P}^{-1}$-inner products, we refer the reader to [18, p. 269].

It follows from (3.16) and the hypothesis (3.17) that $\mathbf{B}$ is also coercive and bounded in the norm defined by $\mathbf{P}$: for any $\mathbf{v}$ and $\mathbf{w} \in \mathbb{R}^{\dim V_{h,\mathbf{P}}}$, we have

$$\|\mathbf{v}\|_{\mathbf{P}}^2 \leq \frac{1}{c_{\mathbf{P}} c_B} \mathbf{v}^\top \mathbf{B}\mathbf{v}, \quad |\mathbf{v}^\top \mathbf{B}\mathbf{w}| \leq C_{\mathbf{P}} C_B \|\mathbf{v}\|_{\mathbf{P}} \|\mathbf{w}\|_{\mathbf{P}}.$$

This enables us to appeal to the following well-known bound from GMRES convergence theory [9].

Theorem 1 *Let $\mathbf{u} \in \mathbb{R}^{\dim V_{h,\mathbf{P}}}$ be the vector representing the solution of* (3.15)*. For each $k \geq 1$, let $\mathbf{u}_k$ be defined by* (3.19) *or equivalently by* (3.20)*, with associated residual $\mathbf{r}_k$. Then*

$$\frac{\|\mathbf{r}_k\|_{\mathbf{P}^{-1}}}{\|\mathbf{r}_0\|_{\mathbf{P}^{-1}}} = \frac{\|\mathbf{P}^{-1}\mathbf{r}_k\|_{\mathbf{P}}}{\|\mathbf{P}^{-1}\mathbf{r}_0\|_{\mathbf{P}}} \leq \left(1 - \frac{c_{\mathbf{P}}^2 c_B^2}{C_{\mathbf{P}}^2 C_B^2}\right)^{k/2} = \left(1 - \frac{1}{\kappa(\mathbf{P}^{-1}\mathbf{A})^2} \frac{c_B^2}{C_B^2}\right)^{k/2}. \tag{3.21}$$

The bound (3.21) and the coercivity of $\mathbf{B}$ imply the following bound for the error:

$$\|\mathbf{u} - \mathbf{u}_k\|_{\mathbf{P}}^2 \leq \frac{1}{c_{\mathbf{P}} c_B} (\mathbf{u} - \mathbf{u}_k)^\top \mathbf{r}_k \leq \frac{1}{c_{\mathbf{P}} c_B} \|\mathbf{u} - \mathbf{u}_k\|_{\mathbf{P}} \|\mathbf{r}_k\|_{\mathbf{P}^{-1}},$$

thereby implying that

$$\|\mathbf{u}-\mathbf{u}_k\|_{\mathbf{P}} \leq \frac{\|\mathbf{r}_0\|_{\mathbf{P}^{-1}}}{c_{\mathbf{P}}c_B}\left(1-\frac{c_B^2}{C_B^2\,\kappa(\mathbf{P}^{-1}\mathbf{A})^2}\right)^{k/2}. \tag{3.22}$$

The bound (3.22) gives a guaranteed minimum convergence rate for GMRES in the $\mathbf{P}$-norm, which is equivalent to the a_h-norm up to the condition number $\kappa(\mathbf{P}^{-1}\mathbf{A})$. We recall that a_h defines a norm equivalent to $\|\cdot\|_{2,h}$, which is the norm of interest, as shown by Lemma 1. This strongly suggests that the $\mathbf{P}^{-1}$-norm, as opposed to the Euclidean norm, of the residual is a natural objective to be minimized by GMRES, as in (3.19) and (3.20).

The conclusion from (3.21) and (3.22) is that, if $\mathbf{P}$ is a robust preconditioner for $\mathbf{A}$ in the sense of yielding uniformly bounded condition numbers with respect to the parameters being varied, then $\mathbf{P}$ will also be a robust preconditioner for the nonsymmetric problems arising from linearizations of HJB equations. In section 4, we construct a specific symmetric positive definite preconditioner $\mathbf{P}$, based on a nonoverlapping domain decomposition method, that will be used to solve (3.7).

Remark 1 The general preconditioning strategy proposed here was largely motivated by the analysis in [15]. It is well-known [26] that convergence bounds for GMRES, such as (3.21), need not be descriptive of the convergence rate obtained in practice, i.e. GMRES may perform significantly better than what is predicted by (3.21) alone. In particular this is observed in some of the experiments of section 7.3 below. This implies that the efficiency of the preconditioners must generally be assessed from computations.

3.3 Condition number of the unpreconditioned problem

The condition number of the matrix $\mathbf{A} \coloneqq (a_h(\phi_i,\phi_j))$ depends on the choice of basis for $V_{h,\mathbf{p}}$. However, in practice, the basis is often chosen to be either a nodal basis or a mapped orthonormal basis. For example, let us assume that each basis function ϕ_i of $V_{h,\mathbf{p}}$ has support in only one element, and is mapped from a member of a set of functions that are L^2-orthonormal on a reference element. Then, arguments that are similar to those in [4] show that the ℓ^2-norm condition number $\kappa(\mathbf{A})$ of the matrix $\mathbf{A} \coloneqq (a_h(\phi_i,\phi_j))$ satisfies

$$\kappa(\mathbf{A}) \lesssim \max_{K\in\mathcal{T}_h}\frac{p_K^8}{h_K^4}\,\frac{\max_{K\in\mathcal{T}_h}h_K^d}{\min_{K\in\mathcal{T}_h}h_K^d}, \tag{3.23}$$

where it is recalled that d is the dimension of the domain Ω.

4 Domain decomposition preconditioners

Let Ω be partitioned into a set $\mathcal{T}_S \coloneqq \{\Omega_i\}_{i=1}^N$ of nonoverlapping Lipschitz polytopal subdomains Ω_i. The partition $\mathcal{T}_S$ is assumed to be conforming. A coarse simplicial or parallelepipedal mesh $\mathcal{T}_H$ is associated to each fine mesh $\mathcal{T}_h$. Let $H_D \coloneqq \operatorname{diam} D$ for each $D \in \mathcal{T}_H$ and suppose that $H \coloneqq \max_{D\in\mathcal{T}_H} H_D$. It is required that the sequence of meshes $\{\mathcal{T}_H\}_H$ satisfy the mesh conditions of section 2. Furthermore, the partitions $\mathcal{T}_S$, $\mathcal{T}_H$ and $\mathcal{T}_h$ are assumed to be *nested*, in the sense that no face of $\mathcal{T}_S$, respectively $\mathcal{T}_H$, cuts the interior of an element of $\mathcal{T}_H$, respectively $\mathcal{T}_h$. Hence, each element $D \in \mathcal{T}_H$ satisfies $\overline{D} = \bigcup \overline{K}$, where the union is over all elements $K \in \mathcal{T}_h$ such that $K \subset D$.

For each mesh $\mathcal{T}_H$, let $\mathbf{q} \coloneqq (q_D \colon D \in \mathcal{T}_H)$ be a vector of *positive* integers; so $q_D \geq 1$ for each element $D \in \mathcal{T}_H$. Assume that $\mathbf{q}$ satisfies the bounded variation property of (2.3), and that $q_D \leq \min_{K \subset D} p_K$ for all $D \in \mathcal{T}_H$. For each $D \in \mathcal{T}_H$, define the sets

$$\begin{aligned} \mathcal{T}_h(D) &\coloneqq \{K \in \mathcal{T}_h \colon K \subset D\}, \qquad & \mathcal{F}_h^i(D) &\coloneqq \left\{F \in \mathcal{F}_h^i \colon F \subset D\right\}, \\ \mathcal{F}_h^i(\partial D) &\coloneqq \left\{F \in \mathcal{F}_h^i \colon F \subset \partial D\right\}, & \mathcal{F}_h^{i,b}(\partial D) &\coloneqq \{F \in \mathcal{F}_h^{i,b} \colon F \subset \partial D\}. \end{aligned} \tag{4.1}$$

Although the sets $\mathcal{F}_h^i(D)$ and $\mathcal{F}_h^{i,b}(D)$ are not disjoint, the above assumptions on the meshes imply that $\mathcal{F}_h^{i,b} = \bigcup_D \mathcal{F}_h^i(D) \cup \mathcal{F}_h^{i,b}(\partial D)$ and that $\mathcal{F}_h^i = \bigcup_D \mathcal{F}_h^i(D) \cup \mathcal{F}_h^i(\partial D)$. Define the function spaces

$$V_{h,\mathbf{p}}^i \coloneqq \left\{v \in L^2(\Omega_i) \colon \ v|_K \in \mathcal{P}_{p_K}(K) \quad \forall\, K \in \mathcal{T}_h, K \subset \Omega_i\right\}, \quad 1 \leq i \leq N, \tag{4.2a}$$

$$V_{H,\mathbf{q}} \coloneqq \left\{v \in L^2(\Omega) \colon \ v|_D \in \mathcal{P}_{q_D}(D) \quad \forall\, D \in \mathcal{T}_H\right\}. \tag{4.2b}$$

For convenience of notation, let $V_{h,\mathbf{p}}^0 \coloneqq V_{H,\mathbf{q}}$. It follows from the above conditions on the meshes that every function $v_H \in V_{H,\mathbf{q}}$ also belongs to $V_{h,\mathbf{p}}$, so let $I_0 \colon V_{H,\mathbf{q}} \to V_{h,\mathbf{p}}$ denote the natural imbedding map. For $1 \leq i \leq N$, let $I_i \colon V_{h,\mathbf{p}}^i \to V_{h,\mathbf{p}}$ denote the natural injection operator defined by

$$I_i\, v_i \coloneqq \begin{cases} v_i & \text{on } \Omega_i, \\ 0 & \text{on } \Omega - \Omega_i, \end{cases} \qquad \forall\, v_i \in V_{h,\mathbf{p}}^i. \tag{4.3}$$

Then, any function $v_h \in V_{h,\mathbf{p}}$ can be decomposed as $v_h = \sum_{i=1}^N I_i \left(v_h|_{\Omega_i}\right)$. Let the bilinear forms $a_h^i \colon V_{h,\mathbf{p}}^i \times V_{h,\mathbf{p}}^i \to \mathbb{R}$, $0 \leq i \leq N$, be defined by

$$a_h^i(u_i, v_i) \coloneqq a_h(I_i\, u_i, I_i\, v_i) \qquad \forall\, u_i, v_i \in V_{h,\mathbf{p}}^i. \tag{4.4}$$

It is clear that the bilinear forms a_h^i are symmetric and coercive on $V_{h,\mathbf{p}}^i \times V_{h,\mathbf{p}}^i$. For each $0 \leq i \leq N$, let $\mathbf{A}_i$ denote the matrix that corresponds to the bilinear form a_h^i and let $\mathbf{I}_i$ denotes the matrix corresponding to the injection operator I_i. Therefore, for each $0 \leq i \leq N$, the matrix $\mathbf{A}_i$ has dimension $\dim V_{h,\mathbf{p}}^i \times \dim V_{h,\mathbf{p}}^i$, and the matrix $\mathbf{I}_i$ has dimension $\dim V_{h,\mathbf{p}} \times \dim V_{h,\mathbf{p}}^i$. Then, we define $\mathbf{P}_i^{-1} \coloneqq \mathbf{I}_i\, \mathbf{A}_i^{-1}\, \mathbf{I}_i^\top$, which therefore has dimension $\dim V_{h,\mathbf{p}} \times \dim V_{h,\mathbf{p}}$. The additive Schwarz preconditioner $\mathbf{P}$ is defined in terms of its inverse by

$$\mathbf{P}^{-1} \coloneqq \sum_{i=0}^N \mathbf{P}_i^{-1}. \tag{4.5}$$

Thus $\mathbf{P}^{-1}$ defines a symmetric positive definite preconditioner $\mathbf{P}$ that may be used as explained in section 3. Further preconditioners, such as multiplicative, symmetric multiplicative and hybrid methods, are presented in [23,25] and the references therein. The general theory of Schwarz methods [23,25] simplifies the analysis of these preconditioners to the verification of three key properties.

Property 1 Suppose that there exists a positive constant c_0 such that each $v_h \in V_{h,\mathbf{p}}$ admits a decomposition $v_h = \sum_{i=0}^N I_i\, v_i$, with $v_i \in V_{h,\mathbf{p}}^i$, for each $0 \leq i \leq N$, with

$$\sum_{i=0}^N a_h^i(v_i, v_i) \leq c_0\, a_h(v_h, v_h). \tag{4.6}$$

Property 2 Assume that there exist constants $\varepsilon_{ij} \in [0,1]$, such that

$$|a_h(I_i\, v_i, I_j v_j)| \leq \varepsilon_{ij} \sqrt{a_h(I_i\, v_i, I_i\, v_i)\; a_h(I_j\, v_j, I_j\, v_j)}, \tag{4.7}$$

for all $v_i \in V^i_{h,\mathbf{p}}$ and all $v_j \in V^j_{h,\mathbf{p}}$, $1 \leq i,j \leq N$. Let $\rho(\mathcal{E})$ denote the spectral radius of the matrix $\mathcal{E} \coloneqq (\varepsilon_{ij})$.

Property 3 Suppose that there exists a constant $\omega \in (0,2)$, such that

$$a_h(I_i\, v_i, I_i\, v_i) \leq \omega\, a^i_h(v_i, v_i) \quad \forall\, v_i \in V^i_{h,\mathbf{p}},\ 0 \leq i \leq N. \tag{4.8}$$

Properties 1–3 are sometimes referred to respectively as the stable decomposition property, the strengthened Cauchy–Schwarz inequality, and local stability.

The following theorem from the theory of Schwarz methods is quoted from [25].

Theorem 2 *If Properties 1–3 hold, then the condition number $\kappa(\mathbf{P}^{-1}\mathbf{A})$ obtained by the additive Schwarz preconditioner satisfies*

$$\kappa(\mathbf{P}^{-1}\mathbf{A}) \leq c_0\, \omega\, (\rho(\mathcal{E}) + 1)\,. \tag{4.9}$$

Remark 2 With the above choices of bilinear forms a^i_h and with the arguments presented in [4], it is seen that (4.8) holds in fact with equality for $\omega = 1$. Also, in (4.7), we can take $\varepsilon_{ij} = 1$ if $\partial\Omega_i \cap \partial\Omega_j \neq \emptyset$, and $\varepsilon_{ij} = 0$ otherwise. Therefore, as explained in [4], $\rho\,(\mathcal{E}) \leq N_c + 1$, where N_c is the maximum number of adjacent subdomains that a given subdomain might have. Therefore, Properties 2 and 3 hold, and it remains to verify Property 1.

The following theorem determines a bound on the constant appearing in (4.6), which can be used in conjunction with Theorem 2 to analyse the properties of the preconditioners. The proof of this result is given in the following sections.

Theorem 3 *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded convex polytopal domain, and let $\mathcal{T}_S$, $\{\mathcal{T}_H\}_H$ and $\{\mathcal{T}_h\}_h$ be successively nested shape-regular sequences of meshes, with $\mathcal{T}_S$ conforming, and $\{\mathcal{T}_H\}_H$ and $\{\mathcal{T}_h\}_h$ satisfying* (2.1)*,* (2.2) *and* (2.3)*. Let μ_F and η_F satisfy* (2.11) *for each face F, with c_μ and c_η chosen to satisfy the hypothesis of Lemma* 1*. Then, each $v_h \in V_{h,\mathbf{p}}$ admits a decomposition $v_h = \sum_{i=0}^N I_i\, v_i$, with $v_i \in V^i_{h,\mathbf{p}}$, $0 \leq i \leq N$, such that*

$$\sum_{i=0}^{N} a^i_h(v_i, v_i) \lesssim \tilde{c}_0\, a_h(v_h, v_h), \tag{4.10}$$

where the constant $\tilde{c}_0$ is given by

$$\tilde{c}_0 \coloneqq 1 + \max_{D\in\mathcal{T}_H} \left[\frac{q_D}{H_D} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^2}{h_K}\right] \max_{D\in\mathcal{T}_H} \frac{H_D^2}{q_D^2} + \max_{D\in\mathcal{T}_H} \left[\frac{q_D}{H_D} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^6}{h_K^3}\right] \max_{D\in\mathcal{T}_H} \frac{H_D^4}{q_D^4}. \tag{4.11}$$

It follows from Theorems 2 and 3 that the condition number satisfies

$$\kappa(\mathbf{P}^{-1}\mathbf{A}) \lesssim \tilde{c}_0\, (N_c + 2)\,, \tag{4.12}$$

where $\tilde{c}_0$ is given in (4.11) above, and N_c is the maximum number of adjacent subdomains that a given subdomain from $\mathcal{T}_S$ might have. Thus the condition number does not depend on the number N of subdomains, but may depend on the maximum number of neighbours any

subdomain possesses, denoted by N_c in (4.12). If the sequence of coarse spaces $\{V_{H,\mathbf{q}}\}_H$ satisfy the assumption that $H_D/q_D \lesssim \min_{D\in\mathcal{T}_H} H_D/q_D$ for all $D \in \mathcal{T}_H$, then the constant $\tilde{c}_0$ in the above proposition simplifies to

$$\tilde{c}_0 \simeq 1 + \max_{D\in\mathcal{T}_H} \left[\frac{H_D}{q_D} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^2}{h_K} + \frac{H_D^3}{q_D^3} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^6}{h_K^3} \right]. \tag{4.13}$$

Moreover, if the sequences of meshes $\{\mathcal{T}_H\}_H$ and $\{\mathcal{T}_h\}_h$ are quasiuniform, and if the polynomial degrees are also quasiuniform in the sense that $q \coloneqq \max_D q_D \lesssim q_D$ for all $D \in \mathcal{T}_H$ and $p \coloneqq \max_K p_K \lesssim p_K$ for all $K \in \mathcal{T}_h$, then the condition number of the preconditioned system satisfies the bound

$$\kappa(\mathbf{P}^{-1}\mathbf{A}) \lesssim (N_c + 2) \left(1 + \frac{p^2\, H}{q\, h} + \frac{p^6\, H^3}{q^3\, h^3} \right). \tag{4.14}$$

It is well-known that the above bound is optimal in terms of the powers of H and h, see [7, 11]. The numerical experiments of section 7 show that the bound (4.14) is also sharp in terms of the orders of p and q. Choosing the coarse space such that $H \simeq h$ and $q \simeq p$ implies that $\kappa(\mathbf{P}^{-1}\mathbf{A}) \lesssim p^3$, which shows that the preconditioner is robust with respect to h but not with respect to p. The explicit dependence of our bound on q shows nonetheless a significant improvement over the condition number of order p^8/h^4 for the unpreconditioned matrix. The preconditioner $\mathbf{P}$ can therefore be used to precondition the nonsymmetric systems (3.7) of the semismooth Newton method, where the convergence is guaranteed by Theorem 1 in combination with (4.14).

5 Approximation of discontinuous functions

As explained in the introduction, the optimal bound for the condition numbers, as given by Theorem 3, rests upon the optimality of approximation properties between coarse and fine spaces. Therefore, in this section, we first determine how closely a function in $V_{h,\mathbf{p}}$ can be approximated by functions in $H^2(\Omega) \cap H_0^1(\Omega)$. This leads to an approximation result for functions in $V_{h,\mathbf{p}}$ by functions in $V_{H,\mathbf{q}}$ that is of optimal order in both the coarse mesh size and polynomial degree.

5.1 Lifting operators

Let $V_{h,\mathbf{p}}^d$ denote the space of d-dimensional vector fields with components in $V_{h,\mathbf{p}}$. Let $\mathbf{r}_h \colon L^2(\mathcal{F}_h^{i,b}) \to V_{h,\mathbf{p}}^d$ and $r_h \colon L^2(\mathcal{F}_h^i) \to V_{h,\mathbf{p}}$ be defined by

$$\sum_{K\in\mathcal{T}_h} \langle \mathbf{r}_h(w), \mathbf{v}_h \rangle_K = \sum_{F\in\mathcal{F}_h^{i,b}} \langle w, \{\mathbf{v}_h \cdot n_F\} \rangle_F \qquad \forall\, \mathbf{v}_h \in V_{h,\mathbf{p}}^d, \tag{5.1}$$

$$\sum_{K\in\mathcal{T}_h} \langle r_h(w), v_h \rangle_K = \sum_{F\in\mathcal{F}_h^{i}} \langle w, \{v_h\} \rangle_F \qquad \forall\, v_h \in V_{h,\mathbf{p}}. \tag{5.2}$$

The following result is well-known; for instance, see [4] for a proof.

Lemma 2 *Let Ω be a bounded Lipschitz domain and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of meshes satisfying* (2.1)*,* (2.2) *and* (2.3)*. Then, the lifting operators satisfy the following bounds:*

$$\sum_{K\in\mathcal{T}_h}\|\mathbf{r}_h(w)\|_{L^2(K)}^2 \lesssim \sum_{F\in\mathcal{F}_h^{i,b}}\frac{\tilde{p}_F^2}{\tilde{h}_F}\|w\|_{L^2(F)}^2 \qquad \forall\, w\in L^2(\mathcal{F}_h^{i,b}), \tag{5.3a}$$

$$\sum_{K\in\mathcal{T}_h}\|r_h(w)\|_{L^2(K)}^2 \lesssim \sum_{F\in\mathcal{F}_h^{i}}\frac{\tilde{p}_F^2}{\tilde{h}_F}\|w\|_{L^2(F)}^2 \qquad \forall\, w\in L^2(\mathcal{F}_h^{i}). \tag{5.3b}$$

For $v_h\in V_{h,\mathbf{p}}$ and $\mathbf{v}_h\in V_{h,\mathbf{p}}^d$, define $G_h(v_h)\in V_{h,\mathbf{p}}^d$ and $D_h(\mathbf{v}_h)\in V_{h,\mathbf{p}}$ element-wise by

$$G_h(v_h)|_K \coloneqq \nabla v_h|_K - \mathbf{r}_h(\llbracket v_h\rrbracket)|_K, \tag{5.4a}$$

$$D_h(\mathbf{v}_h)|_K \coloneqq \operatorname{div}\mathbf{v}_h|_K - r_h(\llbracket \mathbf{v}_h\cdot n_F\rrbracket)|_K, \tag{5.4b}$$

for all $K\in\mathcal{T}_h$. Observe that $D_h(\mathbf{v}_h)$ belongs to $L^2(\Omega)$ for any $\mathbf{v}_h\in V_{h,\mathbf{p}}^d$.

Lemma 3 *Let Ω be a bounded Lipschitz polytopal domain, and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of meshes satisfying* (2.1)*,* (2.2) *and* (2.3)*. Let η_F and μ_F satisfy* (2.11) *for all $F\in\mathcal{F}_h^{i,b}$. Then, for any $v_h\in V_{h,\mathbf{p}}$, we have*

$$\sum_{K\in\mathcal{T}_h}\frac{p_K^4}{h_K^2}\|\mathbf{r}_h(\llbracket v_h\rrbracket)\|_{L^2(K)}^2 \lesssim |v_h|_{\mathrm{J},h}^2, \tag{5.5a}$$

$$\sum_{K\in\mathcal{T}_h}|\mathbf{r}_h(\llbracket v_h\rrbracket)|_{H^1(K)}^2 + \sum_{F\in\mathcal{F}_h^{i}}\mu_F\|\llbracket\mathbf{r}_h(\llbracket v_h\rrbracket)\cdot n_F\rrbracket\|_{L^2(F)}^2 \lesssim |v_h|_{\mathrm{J},h}^2. \tag{5.5b}$$

Proof The definition of the lifting operator gives

$$\begin{aligned}\sum_{K\in\mathcal{T}_h}\frac{p_K^4}{h_K^2}\|\mathbf{r}_h(\llbracket v_h\rrbracket)\|_{L^2(K)}^2 &= \sum_{F\in\mathcal{F}_h^{i,b}}\int_F\llbracket v_h\rrbracket\left\{\frac{p^4}{h^2}\mathbf{r}_h(\llbracket v_h\rrbracket)\cdot n_F\right\}\mathrm{d}s\\ &\lesssim \left(\sum_{F\in\mathcal{F}_h^{i,b}}\frac{\tilde{h}_F^3}{\tilde{p}_F^6}\frac{\tilde{p}_F^8}{\tilde{h}_F^4}\|\mathbf{r}_h(\llbracket v_h\rrbracket)\|_{L^2(F)}^2\right)^{\frac{1}{2}}|v_h|_{\mathrm{J},h}.\end{aligned}$$

The trace and inverse inequalities then yield

$$\sum_{K\in\mathcal{T}_h}\frac{p_K^4}{h_K^2}\|\mathbf{r}_h(\llbracket v_h\rrbracket)\|_{L^2(K)}^2 \lesssim \left(\sum_{K\in\mathcal{T}_h}\frac{p_K^4}{h_K^2}\|\mathbf{r}_h(\llbracket v_h\rrbracket)\|_{L^2(K)}^2\right)^{\frac{1}{2}}|v_h|_{\mathrm{J},h},$$

which implies (5.5a). The bound (5.5b) then follows from (5.5a) as a result of the trace and inverse inequalities. □

Corollary 1 *Under the hypotheses of Lemma 3, every $v_h\in V_{h,\mathbf{p}}$ satisfies*

$$\sum_{K\in\mathcal{T}_h}|G_h(v_h)|_{H^1(K)}^2 + \sum_{F\in\mathcal{F}_h^{i}}\mu_F\|\llbracket G_h(v_h)\cdot n_F\rrbracket\|_{L^2(F)}^2 \lesssim \|v_h\|_{2,h}^2. \tag{5.6}$$

We also have $\|D_h(G_h(v_h))\|_{L^2(\Omega)}\lesssim\|v_h\|_{2,h}$ for every $v_h\in V_{h,\mathbf{p}}$.

Proof Inequality (5.6) is an easy consequence of the definition of G_h in (5.4a) and of Lemma 3. For $v_h \in V_{h,\mathbf{p}}$ and $K \in \mathcal{T}_h$, we have

$$D_h(G_h(v_h))|_K = [\Delta v_h - \operatorname{div} \mathbf{r}_h([\![v_h]\!]) - r_h([\![\nabla v_h \cdot n_F]\!]) + r_h([\![\mathbf{r}_h([\![v_h]\!]) \cdot n_F]\!])]\,|_K. \tag{5.7}$$

In view of (5.3b), it is apparent that the global L^2-norms over Ω of the first and third terms on the right-hand side of (5.7) are bounded by $\|v_h\|_{2,h}$, whilst the bounds on the L^2-norms of the second and fourth terms follow from (5.5b). □

5.2 Approximation by H^2-regular functions

The first step towards the aforementioned approximation result is to consider the discrete analogue of the orthogonality of Helmholtz decompositions. In this section, we shall view the element-wise gradient of a function $v_h \in V_{h,\mathbf{p}}$ as an element of $L^2(\Omega)^d$, and thus we denote it by $\nabla_h v_h$.

Lemma 4 *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded Lipschitz polytopal domain, and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of meshes satisfying* (2.1), (2.2) *and* (2.3). *If μ_F and η_F satisfy* (2.11) *for every face $F \in \mathcal{F}_h^{i,b}$, then, for any $v_h \in V_{h,\mathbf{p}}$ and any $\psi \in H^1(\Omega)^{2d-3}$, we have*

$$\left|\int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi \,\mathrm{d}x\right| + \left|\int_\Omega \nabla_h v_h \cdot \operatorname{curl} \psi \,\mathrm{d}x\right| \lesssim \max_{K\in\mathcal{T}_h} \frac{h_K}{p_K^{3/2}} |v_h|_{\mathrm{J},h} \|\psi\|_{H^1(\Omega)}. \tag{5.8}$$

Proof It follows from (5.5a) that $\|\nabla_h v_h - G_h(v_h)\|_{L^2(\Omega)} \lesssim \max_K h_K/p_K^2 |v_h|_{\mathrm{J},h}$, so it is enough to show that (5.8) is satisfied by $G_h(v_h)$. Consider momentarily $\psi \in H^2(\Omega)^{2d-3}$; then, integration by parts yields

$$\int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi \,\mathrm{d}x = \sum_{F\in\mathcal{F}_h^{i,b}} \langle [\![v_h]\!], \{\operatorname{curl} \psi \cdot n_F\}\rangle_F - \sum_{K\in\mathcal{T}_h} \langle \mathbf{r}_h([\![v_h]\!]), \operatorname{curl} \psi\rangle_K.$$

Therefore, the definitions of the lifting operators $\mathbf{r}_h$ and r_h imply that

$$\begin{aligned}\int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi \,\mathrm{d}x = \sum_{F\in\mathcal{F}_h^{i,b}} \langle [\![v_h]\!], \{\operatorname{curl}(\psi - \Pi_h\psi) \cdot n_F\}\rangle_F \\ - \sum_{K\in\mathcal{T}_h} \langle \mathbf{r}_h([\![v_h]\!]), \operatorname{curl}(\psi - \Pi_h\psi)\rangle_K.\end{aligned}$$

Thus, if $\psi \in H^2(\Omega)^{2d-3}$, it is seen from the approximation bounds of (2.12) and from the lifting bound (5.5a) that

$$\left|\int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi \,\mathrm{d}x\right| \lesssim \max_{K\in\mathcal{T}_h} \frac{h_K^2}{p_K^3} |v_h|_{\mathrm{J},h} \|\psi\|_{H^2(\Omega)}. \tag{5.9}$$

Now, let $\psi \in H^1(\Omega)^{2d-3}$. We apply [1, Thm. 5.33] to the components of ψ: for each $\varepsilon > 0$, there exists a $\psi_\varepsilon \in C^\infty(\mathbb{R}^d)^{2d-3}$ such that

$$\|\psi - \psi_\varepsilon\|_{L^2(\Omega)} + \varepsilon \|\psi - \psi_\varepsilon\|_{H^1(\Omega)} \lesssim \varepsilon |\psi|_{H^1(\Omega)}, \tag{5.10a}$$

$$\|\psi_\varepsilon\|_{H^2(\Omega)} \lesssim \varepsilon^{-1} \|\psi\|_{H^1(\Omega)}, \tag{5.10b}$$

where, importantly, the constants in (5.10) do not depend on ε. Define $\phi_\varepsilon := \psi - \psi_\varepsilon$, so that

$$\int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi \, \mathrm{d}x = \int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi_\varepsilon \, \mathrm{d}x + \int_\Omega G_h(v_h) \cdot \operatorname{curl} \phi_\varepsilon \, \mathrm{d}x.$$

The bounds (5.9) and (5.10b) show that

$$\left| \int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi_\varepsilon \, \mathrm{d}x \right| \lesssim \varepsilon^{-1} \max_{K \in \mathcal{T}_h} \frac{h_K^2}{p_K^3} |v_h|_{\mathrm{J},h} \|\psi\|_{H^1(\Omega)}. \tag{5.11}$$

Integration by parts yields

$$\int_\Omega G_h(v_h) \cdot \operatorname{curl} \phi_\varepsilon \, \mathrm{d}x = \sum_{F \in \mathcal{F}_h^{i,b}} \langle [\![\nabla v_h \times n_F]\!], \phi_\varepsilon \rangle_F - \sum_{K \in \mathcal{T}_h} \langle \mathbf{r}_h([\![v_h]\!]), \operatorname{curl} \phi_\varepsilon \rangle_K.$$

Lemma 3 and (5.10a) imply that

$$\sum_{K \in \mathcal{T}_h} |\langle \mathbf{r}_h([\![v_h]\!]), \operatorname{curl} \phi_\varepsilon \rangle_K| \lesssim \max_{K \in \mathcal{T}_h} \frac{h_K}{p_K^2} |v_h|_{\mathrm{J},h} \|\psi\|_{H^1(\Omega)}. \tag{5.12}$$

Recall the continuous trace inequality [16]: for an element K and a face $F \subset \partial K$,

$$\|\phi_\varepsilon\|_{L^2(F)}^2 \lesssim |\phi_\varepsilon|_{H^1(K)} \|\phi_\varepsilon\|_{L^2(K)} + \frac{1}{h_K} \|\phi_\varepsilon\|_{L^2(K)}^2 \lesssim \frac{h_K}{p_K^2} |\phi_\varepsilon|_{H^1(K)}^2 + \frac{p_K^2}{h_K} \|\phi_\varepsilon\|_{L^2(K)}^2.$$

Therefore, the fact that $\mu_F = c_\mu \, \tilde{p}_F^2 / \tilde{h}_F$ leads to

$$\begin{aligned} \sum_{F \in \mathcal{F}_h^{i,b}} |\langle [\![\nabla v_h \times n_F]\!], \phi_\varepsilon \rangle_F| &\lesssim \left(\sum_{K \in \mathcal{T}_h} \left[\frac{h_K^2}{p_K^4} |\phi_\varepsilon|_{H^1(K)}^2 + \|\phi_\varepsilon\|_{L^2(K)}^2 \right] \right)^{\frac{1}{2}} |v_h|_{\mathrm{J},h} \\ &\lesssim \left(\max_{K \in \mathcal{T}_h} \frac{h_K}{p_K^2} |\phi_\varepsilon|_{H^1(\Omega)} + \|\phi_\varepsilon\|_{L^2(\Omega)} \right) |v_h|_{\mathrm{J},h}, \end{aligned}$$

where we have used the identity $|[\![\nabla v_h \times n_F]\!]| = |[\![\nabla_{\mathrm{T}} \, v_h]\!]|$ for each face F, because $\nabla_{\mathrm{T}} \, v_h$ is the component of ∇v_h that is orthogonal to n_F. Therefore, we deduce from (5.10a) and (5.12) that

$$\left| \int_\Omega G_h(v_h) \cdot \operatorname{curl} \phi_\varepsilon \, \mathrm{d}x \right| \lesssim \left(\max_{K \in \mathcal{T}_h} \frac{h_K}{p_K^2} + \varepsilon \right) |v_h|_{\mathrm{J},h} \|\psi\|_{H^1(\Omega)}. \tag{5.13}$$

Combining (5.11) and (5.13) yields

$$\left| \int_\Omega G_h(v_h) \cdot \operatorname{curl} \psi \, \mathrm{d}x \right| \lesssim \left(\varepsilon^{-1} \max_{K \in \mathcal{T}_h} \frac{h_K^2}{p_K^3} + \max_{K \in \mathcal{T}_h} \frac{h_K}{p_K^2} + \varepsilon \right) |v_h|_{\mathrm{J},h} \|\psi\|_{H^1(\Omega)}.$$

The bound (5.8) is then obtained by taking $\varepsilon := \max_{K \in \mathcal{T}_h} h_K / p_K^{3/2}$. □

Theorem 4 *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded convex polytopal domain, and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of meshes satisfying* (2.1)*,* (2.2) *and* (2.3)*. For a given $v_h \in V_{h,\mathbf{p}}$, let $v_{(h)} \in H^2(\Omega) \cap H_0^1(\Omega)$ be the unique solution of the boundary-value problem*

$$\Delta v_{(h)} = D_h(G_h(v_h)) \quad \text{in } \Omega, \tag{5.14a}$$

$$v_{(h)} = 0 \quad \text{on } \partial\Omega. \tag{5.14b}$$

Then, the approximation $v_{(h)}$ to v_h satisfies

$$\|v_h - v_{(h)}\|_{L^2(\Omega)} + \max_{K\in\mathcal{T}_h} \frac{h_K}{p_K}\|v_h - v_{(h)}\|_{H^1(\Omega;\mathcal{T}_h)} \lesssim \max_{K\in\mathcal{T}_h} \frac{h_K^2}{p_K^2}|v_h|_{\mathrm{J},h}, \tag{5.15a}$$

$$\|v_{(h)}\|_{H^2(\Omega)} \lesssim \|v_h\|_{2,h}. \tag{5.15b}$$

Remark 3 The above result is nearly optimal in the sense that only the jump seminorm $|v_h|_{\mathrm{J},h}$ appears on the right-hand side of the error bound (5.15a), and that the correct orders of convergence are established.

Proof Note that convexity of Ω implies that $v_{(h)}$ is well-defined, see [13], and that (5.15b) holds as a result of Corollary 1. First, we show that for any $p \in H^k(\Omega)\cap H_0^1(\Omega)$, $k \in \{1,2\}$, we have

$$\left|\int_\Omega \big(\nabla v_{(h)} - G_h(v_h)\big)\cdot\nabla p\,\mathrm{d}x\right| \lesssim \max_{K\in\mathcal{T}_h} \frac{h_K^k}{p_K^k}|v_h|_{\mathrm{J},h}\|p\|_{H^k(\Omega)}. \tag{5.16}$$

Indeed, since $v_{(h)}$ solves (5.14), integration by parts yields

$$\int_\Omega \big(\nabla v_{(h)} - G_h(v_h)\big)\cdot\nabla p\,\mathrm{d}x = \sum_{K\in\mathcal{T}_h} \langle r_h([\![G_h(v_h)\cdot n_F]\!]), p\rangle_K - \sum_{F\in\mathcal{F}_h^i} \langle [\![G_h(v_h)\cdot n_F]\!], \{p\}\rangle_F.$$

Then, the definition of the lifting operator gives

$$\int_\Omega \big(\nabla v_{(h)} - G_h(v_h)\big)\cdot\nabla p\,\mathrm{d}x = \sum_{K\in\mathcal{T}_h} \langle r_h([\![G_h(v_h)\cdot n_F]\!]), p - \Pi_h p\rangle_K - \sum_{F\in\mathcal{F}_h^i} \langle [\![G_h(v_h)\cdot n_F]\!], \{p - \Pi_h p\}\rangle_F. \tag{5.17}$$

Recalling that $p_K \geq 1$ for each element K, it is then seen that (5.16) follows from Corollary 1 and from the approximation bounds (2.12).

The remainder of the proof makes use of Helmholtz decompositions of vector fields [12]: for any $\mathbf{v} \in L^2(\Omega)^d$, there exists $p \in H_0^1(\Omega)$ and $\psi \in H^1(\Omega)^{2d-3}$, such that $\mathbf{v} = \nabla p + \operatorname{curl}\psi$ in Ω. Indeed, $p \in H_0^1(\Omega)$ is defined by

$$\int_\Omega \nabla p\cdot\nabla q\,\mathrm{d}x = \int_\Omega \mathbf{v}\cdot\nabla q\,\mathrm{d}x \qquad \forall\, q \in H_0^1(\Omega).$$

Then, $\mathbf{v} - \nabla p$ is divergence free, thus $\langle(\mathbf{v}-\nabla p)\cdot\mathbf{n}, 1\rangle_{\partial\Omega} = 0$, where $\mathbf{n}$ is the unit outward normal on $\partial\Omega$. Since the convex domain Ω has a connected boundary, it follows from [12,

Thms. 3.1 & 3.4 pp. 37–45] that there exists a $\psi \in H^1(\Omega)^{2d-3}$ such that $\mathbf{v} = \nabla p + \operatorname{curl}\psi$. Moreover, ψ may be chosen so that $\|p\|_{H^1(\Omega)} + \|\psi\|_{H^1(\Omega)} \lesssim \|\mathbf{v}\|_{L^2(\Omega)}$ for some constant independent of $\mathbf{v}$. This is a consequence of the Open Mapping Theorem and the facts that $\mathcal{V} \coloneqq \{\mathbf{v} \in L^2(\Omega)^d \colon \operatorname{div}\mathbf{v} = 0\}$ is a closed subspace of $L^2(\Omega)^d$, and that the mapping $\psi \mapsto \operatorname{curl}\psi$ is a surjective bounded linear mapping from $H^1(\Omega)^{2d-3}$ to $\mathcal{V}$.

Now, observe that $\|\nabla v_h - G_h(v_h)\|_{L^2(\Omega)} \lesssim \max_{K\in\mathcal{T}_h} h_K/p_K^2\,|v_h|_{\mathrm{J},h}$ by (5.5a), so it is enough to consider the error between $G_h(v_h)$ and $\nabla v_{(h)}$ to bound $|v_h - v_{(h)}|_{H^1(\Omega;\mathcal{T}_h)}$. Let $p \in H_0^1(\Omega)$ and $\psi \in H^1(\Omega)^{2d-3}$ satisfy $\nabla v_{(h)} - G_h(v_h) = \nabla p + \operatorname{curl}\psi$, with $\|p\|_{H^1(\Omega)} + \|\psi\|_{H^1(\Omega)} \lesssim \|\nabla v_{(h)} - G_h(v_h)\|_{L^2(\Omega)}$. Then, noting that $\nabla v_{(h)}$ and $\operatorname{curl}\psi$ are orthogonal, it is deduced that

$$\|\nabla v_{(h)} - G_h(v_h)\|_{L^2(\Omega)}^2 = \int_\Omega \big(\nabla v_{(h)} - G_h(v_h)\big)\cdot\nabla p\,\mathrm{d}x - \int_\Omega G_h(v_h)\cdot\operatorname{curl}\psi\,\mathrm{d}x. \tag{5.18}$$

Inequality (5.16) and the bound $\|p\|_{H^1(\Omega)} \lesssim \|\nabla v_{(h)} - G_h(v_h)\|_{L^2(\Omega)}$ give

$$\left|\int_\Omega \big(\nabla v_{(h)} - G_h(v_h)\big)\cdot\nabla p\,\mathrm{d}x\right| \lesssim \max_{K\in\mathcal{T}_h}\frac{h_K}{p_K}|v_h|_{\mathrm{J},h}\|\nabla v_{(h)} - G_h(v_h)\|_{L^2(\Omega)}.$$

The bounds of Lemma 4 show that

$$\left|\int_\Omega G_h(v_h)\cdot\operatorname{curl}\psi\,\mathrm{d}x\right| \lesssim \max_{K\in\mathcal{T}_h}\frac{h_K}{p_K^{3/2}}|v_h|_{\mathrm{J},h}\|\nabla v_{(h)} - G_h(v_h)\|_{L^2(\Omega)}.$$

Therefore, equation (5.18) and the above bounds yield

$$\|\nabla v_{(h)} - G_h(v_h)\|_{L^2(\Omega)} \lesssim \max_{K\in\mathcal{T}_h}\frac{h_K}{p_K}|v_h|_{\mathrm{J},h}. \tag{5.19}$$

We now consider the error $\|v_h - v_{(h)}\|_{L^2(\Omega)}$. Since Ω is convex, there is a unique $z \in H^2(\Omega)\cap H_0^1(\Omega)$ that solves $-\Delta z = v_h - v_{(h)}$ in Ω, with $\|z\|_{H^2(\Omega)} \lesssim \|v_h - v_{(h)}\|_{L^2(\Omega)}$. Then, it is found that

$$\begin{aligned}\|v_h - v_{(h)}\|_{L^2(\Omega)}^2 = \int_\Omega \big(G_h(v_h) - \nabla v_{(h)}\big)\cdot\nabla z\,\mathrm{d}x \\ + \sum_{K\in\mathcal{T}_h}\langle\mathbf{r}_h([\![v_h]\!]),\nabla z\rangle_K - \sum_{F\in\mathcal{F}_h^{i,b}}\langle[\![v_h]\!],\{\nabla z\cdot n_F\}\rangle_F.\end{aligned}$$

Applying the bound (5.16) to $z \in H^2(\Omega)\cap H_0^1(\Omega)$ gives

$$\left|\int_\Omega \big(G_h(v_h) - \nabla v_{(h)}\big)\cdot\nabla z\,\mathrm{d}x\right| \lesssim \max_{K\in\mathcal{T}_h}\frac{h_K^2}{p_K^2}|v_h|_{\mathrm{J},h}\|v_h - v_{(h)}\|_{L^2(\Omega)}.$$

Also, it is found that

$$\begin{aligned}&\sum_{K\in\mathcal{T}_h}\langle\mathbf{r}_h([\![v_h]\!]),\nabla z\rangle_K - \sum_{F\in\mathcal{F}_h^{i,b}}\langle[\![v_h]\!],\{\nabla z\cdot n_F\}\rangle_F \\ &\quad= \sum_{K\in\mathcal{T}_h}\langle\mathbf{r}_h([\![v_h]\!]),\nabla(z-\Pi_h z)\rangle_K - \sum_{F\in\mathcal{F}_h^{i,b}}\langle[\![v_h]\!],\{\nabla(z-\Pi_h z)\cdot n_F\}\rangle_F,\end{aligned}$$

which is bounded by $\max_K h_K^2/p_K^3|v_h|_{\mathrm{J},h}\|v_h - v_{(h)}\|_{L^2(\Omega)}$. Thus, we have shown that

$$\|v_h - v_{(h)}\|_{L^2(\Omega)} \lesssim \max_{K\in\mathcal{T}_h}\frac{h_K^2}{p_K^2}|v_h|_{\mathrm{J},h}. \tag{5.20}$$

The bounds (5.19) and (5.20) imply (5.15a). □

5.3 Approximation by coarse grid functions

Theorem 4 leads to the following approximation result between coarse and fine spaces.

Theorem 5 *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded convex polytopal domain, and let $\{\mathcal{T}_H\}_H$ and $\{\mathcal{T}_h\}_h$ be nested shape-regular sequences of meshes satisfying* (2.1)*,* (2.2) *and* (2.3)*. Then, for any $v_h \in V_{h,\mathbf{p}}$, there exists a $v_H \in V_{H,\mathbf{q}}$, such that*

$$\|v_h - v_H\|_{H^k(\Omega;\mathcal{T}_h)} \lesssim \left(\max_{D\in\mathcal{T}_H} \frac{H_D}{q_D}\right)^{2-k} \|v_h\|_{2,h}, \qquad k \in \{0,1,2\}. \tag{5.21a}$$

$$\|v_H\|_{2,h}^2 \lesssim \left(1 + \max_{D\in\mathcal{T}_H}\left[\frac{H_D}{q_D}\max_{K\in\mathcal{T}_h(D)}\frac{p_K^2}{h_K} + \frac{H_D^3}{q_D^3}\max_{K\in\mathcal{T}_h(D)}\frac{p_K^6}{h_K^3}\right]\right)\|v_h\|_{2,h}^2. \tag{5.21b}$$

Proof Let $v_{(h)} \in H^2(\Omega)\cap H_0^1(\Omega)$ be the approximation to v_h considered in Theorem 4. Let $v_H \in V_{H,\mathbf{q}}$ be the projection $\Pi_H v_{(h)}$. Since $\max_{K\in\mathcal{T}_h} h_K/p_K \leq \max_{D\in\mathcal{T}_H} H_D/q_D$, it is seen that (5.21a) follows easily from the triangle inequality in conjunction with (5.15b) and the approximation properties of v_H. In particular, it follows from $v_H = \Pi_H v_{(h)}$ that $\|v_H\|_{H^2(\Omega;\mathcal{T}_h)} \lesssim \|v_{(h)}\|_{H^2(\Omega)}$, and since Theorem 4 implies that $\|v_{(h)}\|_{H^2(\Omega)} \lesssim \|v_h\|_{2,h}$, we obtain $\|v_H\|_{H^2(\Omega;\mathcal{T}_h)} \lesssim \|v_h\|_{2,h}$.

It remains to show (5.21b) by bounding the jump seminorm of v_H as follows. If the face $F \in \mathcal{F}_h^i(D)$ for $D \in \mathcal{T}_H$, then the jumps of v_H and its first derivatives vanish because v_H is a polynomial over D. Since $v_{(h)} \in H^2(\Omega)\cap H_0^1(\Omega)$, $[\![v_H]\!] = [\![v_H - v_{(h)}]\!]$ and $[\![\nabla_{\mathrm{T}} v_H]\!] = [\![\nabla_{\mathrm{T}}(v_H - v_{(h)})]\!]$ for each face $F \in \mathcal{F}_h^{i,b}(\partial D)$, whilst $[\![\nabla v_H \cdot n_F]\!] = [\![\nabla(v_H - v_{(h)})\cdot n_F]\!]$ for each face $F \in \mathcal{F}_h^i(\partial D)$. Therefore, it is deduced from the mesh assumptions on $\mathcal{T}_h$ and $\mathcal{T}_H$ that

$$\begin{aligned}
\sum_{F\in\mathcal{F}_h^{i,b}} \eta_F \|[\![v_H]\!]\|_{L^2(F)}^2 &\leq \sum_{D\in\mathcal{T}_H}\sum_{F\in\mathcal{F}_h^{i,b}(\partial D)} \eta_F \|[\![v_H - v_{(h)}]\!]\|_{L^2(F)}^2 \\
&\lesssim \sum_{D\in\mathcal{T}_H} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^6}{h_K^3}\|v_H - v_{(h)}\|_{L^2(\partial D)}^2 \\
&\lesssim \max_{D\in\mathcal{T}_H}\left[\frac{H_D^3}{q_D^3}\max_{K\in\mathcal{T}_h(D)}\frac{p_K^6}{h_K^3}\right]\|v_{(h)}\|_{H^2(\Omega)}^2.
\end{aligned}$$

Similar bounds also yield

$$\begin{aligned}
&\sum_{F\in\mathcal{F}_h^{i,b}} \mu_F \|[\![\nabla_{\mathrm{T}} v_H]\!]\|_{L^2(F)}^2 + \sum_{F\in\mathcal{F}_h^i} \mu_F \|[\![\nabla v_H \cdot n_F]\!]\|_{L^2(F)}^2 \\
&\qquad\qquad \lesssim \max_{D\in\mathcal{T}_H}\left[\frac{H_D}{q_D}\max_{K\in\mathcal{T}_h(D)}\frac{p_K^2}{h_K}\right]\|v_{(h)}\|_{H^2(\Omega)}^2.
\end{aligned}$$

Since $\|v_{(h)}\|_{H^2(\Omega)} \lesssim \|v_h\|_{2,h}$, the proof of (5.21b) is complete.

Previous results on the approximation of fine mesh functions by coarse mesh functions typically involved lower-order projection operators, which were therefore suboptimal in terms of q in bounds such as (5.21a). The original result of an approximation with optimal orders in both H and q of Theorem 5 enables the sharp analysis of the nonoverlapping domain decomposition preconditioners in the next section.

6 Stable decomposition property

The following lemma, due to Feng and Karakashian in [10], provides a trace inequality for the boundaries ∂D of elements $D \in \mathcal{T}_H$. However, the inequality is not written there in the form that is required for our purposes. So, we present again the proof, with some variations from the arguments in [10].

Lemma 5 *Let $\{\mathcal{T}_H\}_H$ and $\{\mathcal{T}_h\}_h$ be shape-regular sequences of nested simplicial or parallelepipedal meshes satisfying the conditions* (2.1) *and* (2.2)*, and let* $\mathbf{p}$ *satisfy* (2.3)*. Let $v \in L^2(D)$ belong to $\mathcal{P}_{p_K}(K)$ for each $K \subset D$. Then, we have*

$$\|v\|^2_{L^2(\partial D)} \lesssim \sum_{K\in\mathcal{T}_h(D)} |v|_{H^1(K)}\|v\|_{L^2(K)} + \frac{1}{H_D}\|v\|^2_{L^2(D)} + \left(\sum_{F\in\mathcal{F}_h^i(D)} \frac{\tilde{p}_F^2}{\tilde{h}_F}\|[\![v]\!]\|^2_{L^2(F)}\right)^{\frac{1}{2}} \|v\|_{L^2(D)}. \quad (6.1)$$

Proof As shown in [10], since each element $D \in \mathcal{T}_H$ is an affine image of a convex reference element, it follows that there is a point $x_0 \in D$, such that $(x - x_0)\cdot n_{\partial D} \gtrsim H_D$ for each $x \in \partial D$, where $n_{\partial D}$ is the unit outward normal vector to ∂D. Therefore,

$$\|v\|^2_{L^2(\partial D)} \lesssim \frac{1}{H_D}\int_{\partial D} |v|^2\,(x - x_0)\cdot n_{\partial D}\,\mathrm{d}s. \quad (6.2)$$

Integration by parts shows that

$$\int_{\partial D} |v|^2\,(x - x_0)\cdot n_{\partial D}\,\mathrm{d}s = \sum_{K\in\mathcal{T}_h(D)}\int_K \Big[\mathrm{div}\,(x - x_0)\,|v|^2 + 2v\,\nabla v\cdot(x - x_0)\Big]\,\mathrm{d}x - \sum_{F\in\mathcal{F}_h^i(D)} \langle [\![v^2]\!], \{(x - x_0)\cdot n_F\}\rangle_F.$$

Since $[\![v^2]\!] = 2[\![v]\!]\,\{v\}$, it is found that

$$\int_{\partial D} |v|^2\,(x - x_0)\cdot n_{\partial D}\,\mathrm{d}s \lesssim H_D \sum_{K\in\mathcal{T}_h(D)} |v|_{H^1(K)}\|v\|_{L^2(K)} + \|v\|^2_{L^2(D)} + H_D\left(\sum_{F\in\mathcal{F}_h^i(D)} \frac{\tilde{p}_F^2}{\tilde{h}_F}\|[\![v]\!]\|^2_{L^2(F)}\right)^{\frac{1}{2}}\left(\sum_{F\in\mathcal{F}_h^i(D)} \frac{\tilde{h}_F}{\tilde{p}_F^2}\|\{v\}\|^2_{L^2(F)}\right)^{\frac{1}{2}}.$$

The inverse and trace inequalities imply that

$$\sum_{F\in\mathcal{F}_h^i(D)} \frac{\tilde{h}_F}{\tilde{p}_F^2}\|\{v\}\|^2_{L^2(F)} \lesssim \|v\|^2_{L^2(D)}.$$

Therefore, (6.1) follows from (6.2) and the above bounds. □

Equipped with the approximation result of Theorem 5, it is now possible to prove Theorem 3 using a similar approach to [4, 10, 11].

6.1 Proof of Theorem 3

Let v_H be given as in Theorem 5, set $v_0 \coloneqq v_H$, and denote by $v_i \in V^i_{h,\mathbf{p}}$ the restriction of $v_h - v_H$ to Ω_i, $1 \leq i \leq N$. Then, we have

$$\sum_{i=0}^{N} a_h^i(v_i, v_i) = a_h(v_H, v_H) + a_h(v_h - v_H, v_h - v_H) - \sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} a_h(I_i v_i, I_j v_j). \quad (6.3)$$

Observe that the constant appearing on the right-hand side of (5.21b) can be bounded in terms of $\tilde{c}_0$, which was defined in (4.11). So, Theorem 5 and Lemma 1 imply

$$a_h(v_H, v_H) \lesssim \|v_H\|^2_{2,h} \lesssim \tilde{c}_0\, a_h(v_h, v_h), \quad (6.4a)$$

$$a_h(v_h - v_H, v_h - v_H) \lesssim \|v_h\|^2_{2,h} + \|v_H\|^2_{2,h} \lesssim \tilde{c}_0\, a_h(v_h, v_h). \quad (6.4b)$$

It remains to bound the last term in (6.3) for the interface flux and jump terms at the boundaries of the subdomains of $\mathcal{T}_S$. Expanding this term leads to

$$\sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} |a_h(I_i v_i, I_j v_j)| \leq \sum_{k=1}^{5} E_k, \quad (6.5)$$

where the quantities E_k are defined by

$$E_1 \coloneqq \sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} \sum_{\substack{F \in \mathcal{F}_h^i \\ F \subset \partial\Omega_i \cap \partial\Omega_j}} \eta_F\, |\langle (v_h - v_H)|_{\Omega_i}\,,\, (v_h - v_H)|_{\Omega_j} \rangle_F|, \quad (6.6a)$$

$$E_2 \coloneqq \sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} \sum_{\substack{F \in \mathcal{F}_h^i \\ F \subset \partial\Omega_i \cap \partial\Omega_j}} \mu_F\, |\langle \nabla_{\mathrm{T}}(v_h - v_H)|_{\Omega_i}\,,\, \nabla_{\mathrm{T}}(v_h - v_H)|_{\Omega_j} \rangle_F|, \quad (6.6b)$$

$$E_3 \coloneqq \sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} \sum_{\substack{F \in \mathcal{F}_h^i \\ F \subset \partial\Omega_i \cap \partial\Omega_j}} \mu_F\, |\langle \nabla(v_h - v_H)|_{\Omega_i} \cdot n_F,\, \nabla(v_h - v_H)|_{\Omega_j} \cdot n_F \rangle_F|, \quad (6.6c)$$

$$E_4 \coloneqq \sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} \sum_{\substack{F \in \mathcal{F}_h^i \\ F \subset \partial\Omega_i \cap \partial\Omega_j}} |\langle \mathrm{div}_{\mathrm{T}}\, \nabla_{\mathrm{T}}\, (v_h - v_H)|_{\Omega_i}\,,\, \nabla(v_h - v_H)|_{\Omega_j} \cdot n_F \rangle_F|, \quad (6.6d)$$

$$E_5 \coloneqq \sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} \sum_{\substack{F \in \mathcal{F}_h^i \\ F \subset \partial\Omega_i \cap \partial\Omega_j}} |\langle \nabla_{\mathrm{T}}(\nabla(v_h - v_H)|_{\Omega_i} \cdot n_F),\, \nabla_{\mathrm{T}}(v_h - v_H)|_{\Omega_j} \rangle_F|. \quad (6.6e)$$

Note that in (6.6), we have made use of the symmetry of the sum over i, j, $i \neq j$, and the fact that any face $F \subset \partial\Omega_i \cap \partial\Omega_j$ must be an interior face.

Defining $\eta_D \coloneqq \max_{K \in \mathcal{T}_h(D)} p_K^6 / h_K^3$ for each $D \in \mathcal{T}_H$, the hypotheses (2.2) and (2.3) and the nestedness of the meshes imply that

$$E_1 \lesssim \sum_{D \in \mathcal{T}_H} \eta_D \|v_h - v_H\|^2_{L^2(\partial D)}.$$

Therefore, using the trace inequality of Lemma 5, we find that

$$E_1 \lesssim \sum_{D\in\mathcal{T}_H} \eta_D \left[\frac{H_D}{q_D} \sum_{K\in\mathcal{T}_h(D)} |v_h - v_H|^2_{H^1(K)} + \frac{H_D}{q_D} \sum_{F\in\mathcal{F}_h^i(D)} \frac{\tilde{p}_F^2}{\tilde{h}_F} \|[\![v_h]\!]\|^2_{L^2(F)} + \frac{q_D}{H_D} \sum_{K\in\mathcal{T}_h(D)} \|v_h - v_H\|^2_{L^2(K)} \right].$$

Notice that the jumps $[\![v_H]\!]$ vanish for faces $F \in \mathcal{F}_h^i(D)$. Therefore, the approximation bound of Theorem 5 gives

$$E_1 \lesssim \max_{D\in\mathcal{T}_H} \left[\eta_D \frac{H_D}{q_D} \right] \max_{D\in\mathcal{T}_H} \frac{H_D^2}{q_D^2} \|v_h\|^2_{2,h} + \max_{D\in\mathcal{T}_H} \left[\eta_D \frac{H_D}{q_D} \max_{F\in\mathcal{F}_h^i(D)} \frac{\tilde{h}_F^2}{\tilde{p}_F^4} \right] |v_h|^2_{\mathrm{J},h} + \max_{D\in\mathcal{T}_H} \left[\eta_D \frac{q_D}{H_D} \right] \max_{D\in\mathcal{T}_H} \frac{H_D^4}{q_D^4} \|v_h\|^2_{2,h}, \quad (6.7)$$

and thus it follows from (2.2) and (2.3) and coercivity of a_h that

$$E_1 \lesssim \max_{D\in\mathcal{T}_H} \left[\frac{q_D}{H_D} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^6}{h_K^3} \right] \max_{D\in\mathcal{T}_H} \frac{H_D^4}{q_D^4}\, a_h(v_h, v_h). \quad (6.8)$$

Remark that we have used the bounds $H_D/q_D \lesssim q_D/H_D \max_{D\in\mathcal{T}_H} H_D^2/q_D^2$ and also $H_D/q_D \max_{F\in\mathcal{F}_h^i(D)} \tilde{h}_F^2/\tilde{p}_F^4 \lesssim q_D/H_D \max_{D\in\mathcal{T}_H} H_D^4/q_D^4$ in going from (6.7) to (6.8). This is done because it is currently not possible to improve the last term in (6.7), as a consequence of the nonlocal form of the bounds in Theorems 4 and Theorem 5.

The Cauchy–Schwarz inequality with a parameter and the symmetry of the sum over i, j, $j \neq i$, imply that

$$\sum_{k=2}^{5} E_k \lesssim \sum_{\substack{1\le i\neq j\le N \\ F\in\mathcal{F}_h^i \\ F\subset\partial\Omega_i\cap\partial\Omega_j}} \mu_F^{-1} \| D^2(v_h - v_H)\big|_{\Omega_i} \|^2_{L^2(F)} + \mu_F \| \nabla(v_h - v_H)|_{\Omega_j} \|^2_{L^2(F)}. \quad (6.9)$$

Since $\mathcal{T}_S$ is conforming, each face F may appear at most twice in the above sum, and thus the trace and inverse inequalities imply that

$$\sum_{\substack{1\le i\neq j\le N \\ F\in\mathcal{F}_h^i \\ F\subset\partial\Omega_i\cap\partial\Omega_j}} \mu_F^{-1} \| D^2(v_h - v_H)\big|_{\Omega_i} \|^2_{L^2(F)} \lesssim \sum_{K\in\mathcal{T}_h} \|v_h - v_H\|^2_{H^2(K)} \lesssim \tilde{c}_0 a_h(v_h, v_h). \quad (6.10)$$

Defining $\mu_D := \max_{K\in\mathcal{T}_h(D)} p_K^2/h_K$, we apply Lemma 5 componentwise to the gradient of $v_h - v_H$ to find that

$$\sum_{\substack{i,\,j=1\\ i\neq j}}^{N} \sum_{\substack{F\in\mathcal{F}_h^i\\ F\subset\partial\Omega_i\cap\partial\Omega_j}} \mu_F \|\nabla(v_h - v_H)|_{\Omega_j}\|^2_{L^2(F)} \lesssim \sum_{D\in\mathcal{T}_H} \mu_D \|\nabla(v_h - v_H)\|^2_{L^2(\partial D)}$$
$$\lesssim \sum_{D\in\mathcal{T}_H} \mu_D \left[\frac{H_D}{q_D} \sum_{K\in\mathcal{T}_h(D)} |v_h - v_H|^2_{H^2(K)} + \frac{H_D}{q_D} \sum_{F\in\mathcal{F}_h^i(D)} \frac{\tilde{p}_F^2}{\tilde{h}_F} \|[\![\nabla v_h]\!]\|^2_{L^2(F)} \right.$$
$$\left. + \frac{q_D}{H_D} \sum_{K\in\mathcal{T}_h(D)} |v_h - v_H|^2_{H^1(K)} \right]. \quad (6.11)$$

It is important to observe that only terms involving interior faces of the mesh $\mathcal{T}_h$ appear on the right-hand side of the above inequality, so for each $F \in \mathcal{F}_h^i(D)$, we have $\|[\![\nabla v_h]\!]\|^2_{L^2(F)} = \|[\![\nabla_{\mathrm{T}} v_h]\!]\|^2_{L^2(F)} + \|[\![\nabla v_h \cdot n_F]\!]\|^2_{L^2(F)}$. So, we deduce that

$$\sum_{D\in\mathcal{T}_H} \mu_D \|\nabla(v_h - v_H)\|^2_{L^2(\partial D)} \lesssim \max_{D\in\mathcal{T}_H} \left[\mu_D \frac{H_D}{q_D} \right] \|v_h - v_H\|^2_{H^2(\Omega;\mathcal{T}_h)}$$
$$+ \max_{D\in\mathcal{T}_H} \left[\mu_D \frac{H_D}{q_D} \right] |v_h|^2_{\mathrm{J},h} + \max_{D\in\mathcal{T}_H} \left[\mu_D \frac{q_D}{H_D} \right] \|v_h - v_H\|^2_{H^1(\Omega;\mathcal{T}_h)},$$

and thus Theorem 5 and coercivity of a_h show that

$$\sum_{D\in\mathcal{T}_H} \mu_D \|\nabla(v_h - v_H)\|^2_{L^2(\partial D)} \lesssim \max_{D\in\mathcal{T}_H} \left[\frac{q_D}{H_D} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^2}{h_K} \right] \max_{D\in\mathcal{T}_H} \frac{H_D^2}{q_D^2}\, a_h(v_h, v_h). \quad (6.12)$$

Therefore, the inequalities (6.9), (6.10) and (6.12) show that

$$\sum_{k=2}^{5} E_k \lesssim \max_{D\in\mathcal{T}_H} \left[\frac{q_D}{H_D} \max_{K\in\mathcal{T}_h(D)} \frac{p_K^2}{h_K} \right] \max_{D\in\mathcal{T}_H} \frac{H_D^2}{q_D^2}\, a_h(v_h, v_h). \quad (6.13)$$

In summary, combining the inequalities (6.4), (6.8) and (6.13) implies that

$$\sum_{i=0}^{N} a_h^i(I_i v_i, I_i v_i) \lesssim \tilde{c}_0\, a_h(v_h, v_h) + \sum_{k=1}^{5} E_k \lesssim \tilde{c}_0\, a_h(v_h, v_h), \quad (6.14)$$

which completes the proof of the stable decomposition property of Theorem 3. □

The proof of Theorem 3 completes the verification of Properties 1–3, and thus gives the bound (4.12) for the condition number of the preconditioned system.

7 Numerical experiments

We test the theoretical results of section 4 and investigate the performance and competitiveness of the preconditioners in practical applications. Direct factorizations were used to form the coarse mesh and local solvers.

$\kappa(\mathbf{P}^{-1}\mathbf{A})$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$	q rate
$p=2$	2.16×10^{1}					
$p=3$	3.34×10^{2}	6.71×10^{1}				
$p=4$	1.94×10^{3}	3.16×10^{2}	1.35×10^{2}			
$p=5$	7.22×10^{3}	1.43×10^{3}	4.11×10^{2}	2.10×10^{2}		
$p=6$	2.12×10^{4}	4.40×10^{3}	1.31×10^{3}	6.44×10^{2}	3.03×10^{2}	3.60
$p=7$	5.31×10^{4}	1.10×10^{4}	3.50×10^{3}	1.70×10^{3}	8.97×10^{2}	3.35
$p=8$	1.18×10^{5}	2.46×10^{4}	7.91×10^{3}	4.27×10^{3}	2.10×10^{3}	3.25
$p=9$	2.38×10^{5}	4.88×10^{4}	1.61×10^{4}	8.68×10^{3}	4.55×10^{3}	3.10
$p=10$	4.48×10^{5}	9.17×10^{4}	3.00×10^{4}	1.64×10^{4}	8.86×10^{3}	3.00
$p=11$	7.92×10^{5}	1.61×10^{5}	5.29×10^{4}	2.90×10^{4}	1.58×10^{4}	2.97
$p=12$	1.33×10^{6}	2.71×10^{5}	8.89×10^{4}	4.87×10^{4}	2.66×10^{4}	2.97
p rate	5.97	5.94	5.96	5.97	6.03	

Table 1 Dependence of the condition number $\kappa(\mathbf{P}^{-1}\mathbf{A})$ on the coarse and fine mesh polynomial degrees for the experiment of section 7.1. The asymptotic rates are computed by regression on the last three entries of each column for p and each row for q. It is found that $\kappa(\mathbf{P}^{-1}\mathbf{A})$ is of order $1+p^6/q^3$, as predicted in section 4.

$\kappa(\mathbf{P}^{-1}\mathbf{A})$	$h=1/4$	$h=1/8$	$h=1/16$	$h=1/32$	rate
$H=1/2$	1.57×10^{2}	1.19×10^{3}	1.08×10^{4}	8.90×10^{4}	3.06

Table 2 Dependence of the condition number $\kappa(\mathbf{P}^{-1}\mathbf{A})$ on the ratio of mesh sizes H/h, for fixed polynomial degrees p and q. The asymptotic rates $\kappa(\mathbf{P}^{-1}\mathbf{A})$ is found to be of order H^3/h^3, in agreement with the bounds of section 4.

7.1 Sharpness of the bound

Since the bound (4.12) is the first to be explicit in both coarse and fine mesh polynomial degrees, it is important to ascertain its sharpness. Let $\Omega=(0,1)^2$, and let the fixed meshes $\mathcal{T}_H=\mathcal{T}_S$ be obtained by a uniform subdivision of Ω into 4 squares, and let $\mathcal{T}_h$ be obtained by uniform subdivision of Ω into 16 squares. We consider the sequence of spaces $V_{h,\mathbf{p}}$ of piecewise polynomials on $\mathcal{T}_h$ with total degree p, where $p=2,\ldots,12$, and the coarse spaces $V_{H,\mathbf{q}}$ of piecewise polynomials on $\mathcal{T}_H$ with total degree q, where $q=2,\ldots,6$. We apply the additive Schwarz preconditioner defined in section 4 to the bilinear form a_h defined in (3.6), where the penalty parameters are defined by $c_\mu=c_\eta=10$. These choices are made to ensure that the resulting number of degrees of freedom is small, being at most equal to 1456 in the case of $p=12$, thereby facilitating the accurate computation of the condition numbers $\kappa(\mathbf{P}^{-1}\mathbf{A})$ of the preconditioned matrix $\mathbf{P}$. The resulting condition numbers are given in Table 1, which shows that $\kappa(\mathbf{P}^{-1}\mathbf{A})$ is of order $1+p^6/q^3$, in agreement with the results of section 4 and in particular with the bound (4.14). This confirms that the predicted rates with respect to the polynomial degrees are optimal. We further verify the sharpness of the bounds with respect to the parameters H and h in Table 2, which presents the condition numbers for varying $h=2^{-m}$, $m=2,\ldots,5$, and fixed $H=1/2$, and fixed $p=q=2$. It is seen that the predicted rate $\kappa(\mathbf{P}^{-1}\mathbf{A})$ is of order H^3/h^3 in agreement with the theory.

7.2 Comparison with overlapping methods

In this section, we compare the efficiency of nonoverlapping methods with the closely related overlapping methods. It is found the methods achieve similar performances in terms

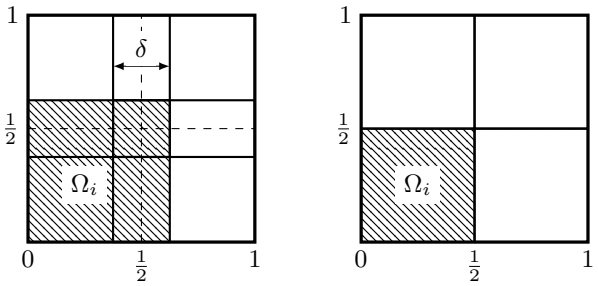

Fig. 1 Overlapping and nonoverlapping decompositions of $\Omega = (0,1)^2$ used in the experiment of section 7.2. Four subdomains are used for both the overlapping and nonoverlapping methods, with the overlap size δ defined as the length shown above.

of iteration counts, although nonoverlapping methods are often faster due to lower computational costs.

Let $\Omega := (0,1)^2$, and let $\mathcal{T}_h$ be obtained by uniform subdivision of Ω into squares of size $h = 2^{-k}$, $k = 3, \ldots, 8$. Let $V_{h,\mathbf{p}}$ consist of the space of polynomials of fixed partial degree $p = 2$ on each element $K \in \mathcal{T}_h$. Consider the model problem: find $u_h \in V_{h,\mathbf{p}}$ such that $a_h(u_h, v_h) = \ell(v_h)$ for all $v_h \in V_{h,\mathbf{p}}$, where the linear functional ℓ_h is chosen so that the solution u_h approximates the function $u(x,y) := \mathrm{e}^{xy}\sin(\pi x)\sin(\pi y)$; specifically, we define $\ell(v_h) = \sum_K \langle \Delta u, \Delta v_h \rangle_K$ for all $v_h \in V_{h,\mathbf{p}}$. It can then be shown that $\|u - u_h\|_{H^2(\Omega;\mathcal{T}_h)} \lesssim h^{p-1}$ [20]. The penalty parameters are chosen so that $\mu_F = 10/\tilde{h}_F$ and $\eta_F = 10/\tilde{h}_F^3$.

Overlapping domain decomposition Let $\delta \in (0,1)$ and let Ω be divided into overlapping subdomains $\mathcal{T}_S = \{\Omega_i\}_{i=1}^4$, as shown in the left-hand side diagram of Figure 1. This yields an overlapping decomposition of Ω with overlap δ; here, we use $\delta \in \{1/4, 1/8, 1/16\}$. Let $\mathcal{T}_H$ be a coarse mesh consisting of a uniform subdivision of Ω into 4 squares, thus yielding the ratios $H/\delta \in \{2, 4, 8\}$, and let $V_{H,\mathbf{q}}$ consist of the space of polynomials of fixed partial degree $q = 2$ on each element $D \in \mathcal{T}_H$. The local spaces $V_{h,\mathbf{p}}^i$ with associated solvers a_h^i, $1 \le i \le 4$, are defined analogously to the nonoverlapping case, described in section 4. The additive Schwarz preconditioner is also defined analogously to section 4.

Nonoverlapping domain decomposition The domain Ω is partitioned into four subdomains $\mathcal{T}_S = \{\Omega_i\}_{i=1}^4$, as shown in the right-hand side diagram of Figure 1. We consider three sequences of coarse meshes $\mathcal{T}_H$, also obtained by uniform subdivision of Ω into squares of size $H = 2^{-m}$, $m = 1, \ldots, k-1$, so that $H/h \in \{2, 4, 8\}$. The nonoverlapping additive Schwarz preconditioner is defined as in section 4.

Results The implementations of the overlapping and nonoverlapping methods were the same, except for the required difference in handling the subdomains. Since the parallelizations of overlapping and nonoverlapping methods differ, our implementation was in serial in order to permit a more straightforward comparison. Table 3 gives the number of iterations required to reduce the residual norms by a factor of 10^{-6}. The results for both methods are comparable to those in the literature: see for instance [2,4,7,14]. Table 3 also presents a representative sample of the CPU times required for the assembly of the preconditioner

		Iteration count					
		Overlapping			Nonoverlapping		
DoF	h	$H = 2\delta$	$H = 4\delta$	$H = 8\delta$	$H = 2h$	$H = 4h$	$H = 8h$
144	1/4				20		
576	1/8	18			22	29	
2304	1/16	18	24		22	30	43
9216	1/32	18	25	37	20	32	52
36864	1/64	18	25	41	18	30	50
147456	1/128	18	26	41	17	27	48
589824	1/256	18	26	42	17	25	40

	Timing					
	Overlapping			Nonoverlapping		
$h = 1/128$	$H = 2\delta$	$H = 4\delta$	$H = 8\delta$	$H = 2h$	$H = 4h$	$H = 8h$
Assembly time	18.6s	14.5s	13.0s	14.0s	11.9s	11.6s
Solver time	8.39s	9.56s	13.3s	6.51s	8.62s	14.4s

Table 3 Number of preconditioned CG iterations required to reduce the residual norm by a factor of 10^{-6} for overlapping and nonoverlapping methods, in the experiment of section 7.2, along with sample timings for assembly and timings of the preconditioned CG algorithm. The methods yield similar iteration counts for similar ratios of H/δ or H/h, but the nonoverlapping method is faster to assemble and apply, as a result of the smaller number of degrees of freedom in the local solvers.

and the application of the preconditioned CG method. The assembly timing strictly includes the time spent on assembling and factorizing the coarse and local mesh solvers, whereas the solver time strictly includes the time spent on applying the preconditioned CG method. These timings are meant to provide only a relative comparison of the methods, with better absolute timings achievable by parallelization.

For the same iteration count, the nonoverlapping methods are generally faster in both assembly and solution. This advantage in efficiency is essentially the result of the smaller dimension of the subdomain solvers. The nonoverlapping method is also generally cheaper in terms of memory costs. Our results show that both methods are efficient, with low iteration counts that remain bounded for fixed H/δ or H/h. In both cases, the results are comparable to computational results from the literature [2,10]. The extension of the analysis for nonoverlapping preconditioners from this work to the case of overlapping preconditioners is an interesting problem for future work.

7.3 Application to HJB equations

We will now consider applications of the preconditioning methods to problems of practical interest, namely fully nonlinear HJB equations. As explained above, this introduces several challenges, such as nonsymmetric linear systems that appear in the semismooth Newton method. Nevertheless, it is found that nonoverlapping methods in particular remain robust and lead to efficient solvers for these problems for h-version methods. The example presented here is closely related to the one from [21, Section 9.1]. Consider the boundary-value problem

$$\begin{aligned} \sup_{\alpha \in \Lambda} [L^\alpha u - f^\alpha] &= 0 \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{7.1}$$

Average GMRES iterations (Newton steps)							
		4 Subdomains			16 Subdomains		
DoF	h	$H = 2h$	$H = 4h$	$H = 8h$	$H = 2h$	$H = 4h$	$H = 8h$
144	1/4	14.3 (6)					
576	1/8	15.2 (5)	18.8 (5)		17.8 (5)		
2304	1/16	15.4 (5)	20.0 (5)	26.8 (5)	18.0 (5)	25.0 (5)	
9216	1/32	16.3 (6)	19.7 (6)	29.5 (6)	17.3 (6)	24.0 (6)	36.5 (6)
36864	1/64	16.0 (6)	18.3 (6)	26.3 (6)	17.2 (6)	22.0 (6)	32.8 (6)
147456	1/128	16.3 (6)	18.3 (6)	23.0 (6)	17.0 (6)	19.8 (6)	28.0 (6)

Table 4 Average number of GMRES iterations per Newton step, with total number of Newton steps in parentheses, for the problem of section 7.3 with both 4 and 16 subdomains.

Average GMRES iterations (Newton steps)					
DoF	p	$q = 3$	$q = 4$	$q = 5$	$q = 6$
16384	3	22.0 (6)			
25600	4	25.8 (6)	20.8 (6)		
36864	5	28.8 (6)	22.0 (6)	20.8 (6)	
50176	6	31.5 (6)	23.3 (6)	22.2 (6)	20.8 (6)
65536	7	35.2 (6)	24.0 (6)	23.5 (6)	21.2 (6)
82944	8	37.0 (6)	25.2 (6)	25.0 (6)	21.8 (6)

Table 5 Average number of GMRES iterations per Newton step, with total number of Newton steps in parentheses, for the problem of section 7.3 with varying polynomial degrees p and q, with fixed $h = 1/32$, $H = 2h$, 1024 elements and 256 subdomains.

Average GMRES iterations (Newton steps)						
Subdomains	$h = 1/4$	$h = 1/8$	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$
4	17.2 (5)	17.3 (6)	17.7 (6)	17.7 (6)	17.7 (6)	17.3 (6)
16		19.2 (6)	18.8 (6)	18.5 (6)	18.5 (6)	18.2 (6)
64			20 (6)	19.5 (6)	19.5 (6)	19.3 (6)
256				20.8 (6)	20.5 (6)	20.3 (6)

Table 6 Average number of GMRES iterations per Newton step required for a relative residual norm tolerance 10^{-4}, with total number of Newton steps in parentheses, for varying numbers of subdomains, using $H = 2h$ and $p = 4$.

where $\Omega = (0,1)^2$, $\Lambda \coloneqq [0, \pi/3] \times \mathrm{SO}(2)$, and where $L^\alpha v \coloneqq a^\alpha : D^2 v$, with

$$a^\alpha \coloneqq \frac{1}{2} R \begin{pmatrix} 1 + \sin^2\theta & \sin\theta\,\cos\theta \\ \sin\theta\,\cos\theta & \cos^2\theta \end{pmatrix} R^\top, \quad \alpha = (\theta, R) \in \Lambda. \tag{7.2}$$

The source terms f^α, $\alpha \in \Lambda$, are chosen so that the exact solution is given by $u(x,y) = \mathrm{e}^{xy}\sin(\pi x)\sin(\pi y)$, whilst yielding large variations in the values of α that attain the supremum in (7.1). As explained in [21], a key challenge in this example is that the diffusion coefficient a^α is highly anisotropic for θ near $\pi/3$, and the rotation matrices R may lead to large variations in the resulting diffusions across the domain and between Newton steps. As a result, significant anisotropic variations in the resulting linearizations are encountered in the application of the semismooth Newton method.

The numerical scheme (3.4) is applied on a sequence of fine meshes $\mathcal{T}_h$ obtained by uniform subdivision of Ω into squares of size $h = 2^{-k}$, $k = 3, \ldots, 7$, with polynomial degrees $2 \leq p \leq 5$. Each iteration of the semismooth Newton method for solving (3.4) leads to a nonsymmetric but positive definite linear system [21], which we solve using the GMRES

h	$p=2$	$p=3$	$p=4$	$p=5$
1/4	18	21	21	22
1/8	19	20	19	20
1/16	19	19	19	19
1/32	18	19	17	18
1/64	17	19	16	17

Table 7 Number of GMRES iterations on the first Newton step required at the first Newton step for a relative residual norm tolerance of 10^{-6}, for various polynomial degrees $2 \leq p \leq 5$, using $H = 2h$ and 4 subdomains. The results are better than theoretical predictions, see Remark 1.

	h					
Subdomains	1/4	1/8	1/16	1/32	1/64	1/128
4	18	19	19	18	17	15
16		22	20	19	18	17
64			21	19	18	17
256				21	19	18

Table 8 Number of GMRES iterations on the first Newton step required for a relative residual norm tolerance 10^{-6}, for varying numbers of subdomains, using $H = 2h$ and $p = 2$.

method (3.20) implemented as suggested in [18]. The nonoverlapping preconditioners are based on the bilinear form a_h, using between 4 and 256 regular subdomains and $q = p$.

To study the overall performance of the preconditioners, we computed the average number of GMRES iterations per Newton step required to reduce the residual norm $\|\mathbf{r}_k\|_{\mathbf{P}^{-1}}$ below a tolerance of 10^{-6} or a relative tolerance of 10^{-4}. Convergence of the Newton method was determined by requiring a step-increment L^2-norm below 10^{-6}. These tolerances were chosen to give a good balance between the different sources of error originating from discretization, linearization and algebraic solvers. The corresponding results are given in Table 4, showing the effectiveness of the preconditioners and their robustness with respect to the anisotropy of the diffusion term. Tables 5 and 7 shows the iteration counts for varying choices of the polynomial degrees. Tables 6 and 8 shows that the iteration counts are not affected by the number of subdomains. We point out that these iteration counts are comparable to those obtained by Lasser and Toselli in [14] for nonsymmetric H^1-type problems originating from advection-diffusion-reaction equations. In particular, for moderate polynomial degrees, the preconditioners are found to be efficient and robust under h-refinement.

Overall, these results show that nonoverlapping preconditioners are robust and efficient when confronted with the anisotropy, lack of symmetry and nonlinearity of this problem.

8 Conclusion

Original approximation results for discontinuous finite element spaces lead to optimal order spectral bounds for nonoverlapping domain decomposition preconditioners in H^2-norms. In the case of h-refinement, we have shown the robustness, efficiency and competitiveness of these preconditioning methods in applications to the nonsymmetric systems arising from fully nonlinear HJB equations.

References

1. Adams, R.A., Fournier, J.F.: Sobolev spaces, *Pure and Applied Mathematics*, vol. 140, second edition edn. Elsevier (2003).
2. Antonietti, P.F., Ayuso, B.: Schwarz domain decomposition preconditioners for discontinuous Galerkin approximations of elliptic problems: non-overlapping case. M2AN Math. Model. Numer. Anal. **41**(1), 21–54 (2007).
3. Antonietti, P.F., Ayuso, B.: Multiplicative Schwarz methods for discontinuous Galerkin approximations of elliptic problems. M2AN Math. Model. Numer. Anal. **42**(3), 443–469 (2008).
4. Antonietti, P.F., Houston, P.: A class of domain decomposition preconditioners for hp-discontinuous Galerkin finite element methods. Journal of Scientific Computing **46**(1), 124–149 (2011).
5. Antonietti, P.F., Smears, I., Houston, P.: A note on optimal spectral bounds for nonoverlapping domain decomposition preconditioners for hp-version discontinuous Galerkin methods. Int. J. Numer. Anal. Model. **13**(4), 513–524 (2016).
6. Antonietti, P.F., Süli, E.: Domain decomposition preconditioning for discontinuous Galerkin approximations of convection-diffusion problems. In: Domain decomposition methods in science and engineering XVIII, *Lect. Notes Comput. Sci. Eng.*, vol. 70, pp. 259–266. Springer, Berlin (2009).
7. Brenner, S.C., Wang, K.: Two-level additive Schwarz preconditioners for C^0 interior penalty methods. Numer. Math. **102**(2), 231–255 (2005).
8. Brenner, S.C., Wang, K.: An iterative substructuring algorithm for a C^0 interior penalty method. Electron. Trans. Numer. Anal. **39**, 313–332 (2012)
9. Eisenstat, S.C., Elman, H.C., Schultz, M.H.: Variational iterative methods for nonsymmetric systems of linear equations. SIAM J. Numer. Anal. **20**(2), 345–357 (1983).
10. Feng, X., Karakashian, O.A.: Two-level additive Schwarz methods for a discontinuous Galerkin approximation of second order elliptic problems. SIAM J. Numer. Anal. **39**(4), 1343–1365 (electronic) (2001).
11. Feng, X., Karakashian, O.A.: Two-level non-overlapping Schwarz preconditioners for a discontinuous Galerkin approximation of the biharmonic equation. J. Sci. Comput. **22/23**, 289–314 (2005).
12. Girault, V., Raviart, P.A.: Finite element methods for Navier-Stokes equations, *Springer Series in Computational Mathematics*, vol. 5. Springer-Verlag, Berlin (1986).
13. Grisvard, P.: Elliptic problems in nonsmooth domains, *Classics in Applied Mathematics*, vol. 69. SIAM, Philadelphia (2011).
14. Lasser, C., Toselli, A.: An overlapping domain decomposition preconditioner for a class of discontinuous Galerkin approximations of advection-diffusion problems. Math. Comp. **72**(243), 1215–1238 (electronic) (2003).
15. Loghin, D., Wathen, A.J.: Analysis of preconditioners for saddle-point problems. SIAM Journal on Scientific Computing **25**(6), 2029–2049 (electronic) (2004).
16. Monk, P., Süli, E.: The adaptive computation of far-field patterns by a posteriori error estimation of linear functionals. SIAM J. Numer. Anal. **36**(1), 251–274 (1999).
17. Pavarino, L.F.: Additive Schwarz methods for the p-version finite element method. Numer. Math. **66**(4), 493–515 (1994).
18. Saad, Y.: Iterative methods for sparse linear systems, second edn. Society for Industrial and Applied Mathematics, Philadelphia, PA (2003).
19. Saad, Y., Schultz, M.H.: GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Statist. Comput. **7**(3), 856–869 (1986).
20. Smears, I., Süli, E.: Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordes coefficients. SIAM J. Numer. Anal. **51**, 2088–2106 (2013)
21. Smears, I., Süli, E.: Discontinuous Galerkin finite element approximation of Hamilton–Jacobi–Bellman equations with Cordes coefficients. SIAM J. Numer. Anal. **52**(2), 993–1016 (2014).
22. Smears, I., Süli, E.: Discontinuous Galerkin finite element methods for time-dependent Hamilton–Jacobi–Bellman equations with Cordes coefficients. Numer. Math. **133**(1), 141–176 (2016).
23. Smith, B.F., Bjørstad, P.E., Gropp, W.D.: Domain decomposition. Cambridge University Press, Cambridge (1996).
24. Toselli, A., Vasseur, X.: Domain decomposition preconditioners of Neumann-Neumann type for hp-approximations on boundary layer meshes in three dimensions. IMA J. Numer. Anal. **24**(1), 123–156 (2004).
25. Toselli, A., Widlund, O.: Domain decomposition methods—algorithms and theory, *Springer Series in Computational Mathematics*, vol. 34. Springer-Verlag, Berlin (2005).
26. Wathen, A.J.: Preconditioning. Acta Numer. **24**, 329–376 (2015).