

NORM PRECONDITIONERS FOR DISCONTINUOUS GALERKIN HP-FINITE ELEMENT METHODS

EMMANUIL H. GEORGOULIS * AND DANIEL LOGHIN †

Abstract. We consider a norm-preconditioning approach for the solution of discontinuous Galerkin finite element discretizations of second order PDE with non-negative characteristic form. In particular, we perform an analysis for the general case of discontinuous *hp*-finite element discretizations. Our solution method is a norm-preconditioned three-term GMRES routine. We find that for symmetric positive-definite diffusivity tensors the convergence of our solver is independent of discretization, while for the semidefinite case both theory and experiment indicate dependence on both h and p . Numerical results are included to illustrate performance on several test cases.

1. Introduction. Recent years have seen an increasing interest in a class of non-conforming finite element approximations of elliptic boundary-value problems with hyperbolic nature, usually referred to as *discontinuous Galerkin finite element methods*. Various families of discontinuous Galerkin finite element methods (DGFEMs) have been proposed, particularly for the numerical solution of convection-diffusion problems, due to the many attractive properties they exhibit. In particular, DGFEMs admit good stability properties, they offer flexibility in the mesh design (irregular meshes are admissible) and in the imposition of boundary conditions (Dirichlet boundary conditions are weakly imposed), and they are increasingly popular in the context of *hp*-adaptive algorithms. The increase in popularity for DGFEMs has created a corresponding demand for developing linear solvers.

Existing approaches to solving systems arising in DGFEMs include domain decomposition, either non-overlapping [12], [13] or overlapping [23] and multigrid [19], [5]. Another favoured approach consists in reformulating the problem as a system of PDE which is then solved using a mixed finite element method for which block-preconditioners can be devised [29], [24], [22].

Remarkably, preconditioned Krylov methods feature rarely and have been mostly employed for the case of time-dependent problems, whether using time-discontinuous discretizations [10], or space-discontinuous ones [3]. One of the reasons for this notable absence may be explained through the ill-conditioning the resulting linear systems suffer from [7] coupled with a need for preconditioner design.

In this work we extend results available for the case of standard finite element methods to the *hp*-discontinuous case. More precisely, it is known that finite element matrices belong to a framework that allows the design of useful preconditioners [11], [26], as well as meaningful monitoring of convergence (e.g., see [1]). In particular, it is now known that the norms associated with the finite element analysis can be employed as preconditioners. The analysis of their performance was first carried out using the concept of norm-equivalence.

The concept of norm-equivalence was formally introduced in [11] in the context of preconditioning for standard finite element discretisation of elliptic problems. The authors of [11] realized that finite element matrices usually are provided with useful preconditioners in the form of the matrix-norm associated with the finite element spaces involved in the solution process. In particular, they were concerned with (and showed mesh-independence for) the distribution of eigenvalues and singular values of the norm-preconditioned system with a view to employing an expensive method, such as Conjugate Gradients on the Normal Equations (CGNE). However, modern methods such as GMRES, cannot be analysed in terms of the distribution of the spectrum alone. Thus, norm-equivalence had to be replaced by a stronger property, *field-of-value equivalence*, which can be shown to be relevant in the convergence analysis of GMRES. This is the concept we employ in order to identify a useful preconditioner as well as to derive useful analytical results regarding convergence properties of our iterative scheme.

*Department of Mathematics, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom, e-mail: Emmanuil.Georgoulis@mcs.le.ac.uk

†School of Mathematics, University of Birmingham, Edgbaston, Birmingham BS15 2TT, United Kingdom, e-mail: d.loghin@bham.ac.uk

The paper is structured as follows. We first introduce the formulation of the problem, together with notation. In section 4 we derive some (continuous) norm-equivalences which we need in order to perform the analysis of our preconditioners. Section 5 describes our iterative approach and includes theoretical convergence results which are then verified in section 6.

2. Model Problem. Let Ω be a bounded open (curvilinear) polygonal domain in \mathbb{R}^d , and let Γ_∂ signify the union of its one-dimensional open edges, which are assumed to be sufficiently smooth (in a sense defined rigorously later). We consider the convection-diffusion-reaction equation

$$\mathcal{L}u \equiv -\nabla \cdot (\bar{\mathbf{a}}\nabla u) + \mathbf{b} \cdot \nabla u + cu = f \quad \text{in } \Omega, \quad (2.1)$$

where $f \in L^2(\Omega)$, $c \in L^\infty(\Omega)$, \mathbf{b} is a vector function whose entries are Lipschitz continuous real-valued functions on $\bar{\Omega}$, and $\bar{\mathbf{a}}$ is the *symmetric* diffusion tensor whose entries are bounded, piecewise continuous real-valued functions defined on $\bar{\Omega}$, with

$$\zeta^T \bar{\mathbf{a}}(x) \zeta \geq 0 \quad \forall \zeta \in \mathbb{R}^d, \quad \text{a.e. } x \in \bar{\Omega}. \quad (2.2)$$

Under this hypothesis, (2.1) is termed a *partial differential equation with nonnegative characteristic form*. By \mathbf{n} we denote the unit outward normal vector to Γ_∂ . We define

$$\Gamma_0 = \{x \in \Gamma_\partial : \mathbf{n}(x)^T \bar{\mathbf{a}}(x) \mathbf{n}(x) > 0\},$$

$$\Gamma_- = \{x \in \Gamma_\partial \setminus \Gamma_0 : \mathbf{b}(x) \cdot \mathbf{n}(x) < 0\}, \quad \Gamma_+ = \{x \in \Gamma_\partial \setminus \Gamma_0 : \mathbf{b}(x) \cdot \mathbf{n}(x) \geq 0\}.$$

The sets Γ_- and Γ_+ are referred to as *inflow* and *outflow* boundary, respectively. We can also see that $\Gamma_\partial = \Gamma_0 \cup \Gamma_- \cup \Gamma_+$. If Γ_0 has positive one-dimensional Hausdorff measure, we also decompose Γ_0 into two parts Γ_D, Γ_N and we impose Dirichlet and Neumann boundary conditions, respectively, via

$$\begin{aligned} u &= g_D \text{ on } \Gamma_D \cup \Gamma_-, \\ (\bar{\mathbf{a}}\nabla u) \cdot \mathbf{n} &= g_N \text{ on } \Gamma_N, \end{aligned} \quad (2.3)$$

where we adopt the (physically reasonable) hypothesis that $\mathbf{b} \cdot \mathbf{n} \geq 0$ on Γ_N , whenever the latter is nonempty.

Existence and uniqueness of solutions (in the weak sense) for the corresponding homogeneous problem were considered by Houston & Süli [21] (cf. also [27] and the references therein, for existence and uniqueness results for classical solutions), under the assumption that there exists a positive constant γ_0 such that

$$c - \frac{1}{2} \nabla \cdot \mathbf{b} \geq \gamma_0 \text{ almost everywhere in } \Omega. \quad (2.4)$$

3. Discontinuous Galerkin Finite Element Method. We denote by $L^p(\omega)$, $1 \leq p \leq \infty$, the standard Lebesgue spaces, $\omega \subset \mathbb{R}^d$ with corresponding norms $\|\cdot\|_{L^p(\omega)}$; the norm of $L^2(\omega)$ will be denoted by $\|\cdot\|_\omega$ for brevity. We also denote by $H^s(\omega)$, the standard Hilbertian Sobolev space of index $s \geq 0$ of real-valued functions defined on $\omega \subset \mathbb{R}^d$.

Let \mathcal{T} be a subdivision of Ω into disjoint open elements $\kappa \in \mathcal{T}$ such that each side of κ has at most one regular hanging node. We assume that the subdivision \mathcal{T} is constructed via mappings F_κ , where $F_\kappa : \hat{\kappa} := (-1, 1)^2 \rightarrow \kappa$ is a C^1 -diffeomorphism, with non-singular Jacobian. The above mappings are assumed to be constructed so as to ensure that the union of the closures of the elements $\kappa \in \mathcal{T}$ forms a covering of the closure of Ω , i.e., $\bar{\Omega} = \cup_{\kappa \in \mathcal{T}} \bar{\kappa}$.

We assign to the subdivision \mathcal{T} the broken Sobolev space of order s ,

$$H^s(\Omega, \mathcal{T}) := \{u \in L^2(\Omega) : u|_\kappa \in H^s(\kappa) \text{ for all } \kappa \in \mathcal{T}\},$$

equipped with the standard broken Sobolev norm.

For a nonnegative integer p , we denote by $\mathcal{Q}_p(\hat{\kappa})$, the set of all tensor-product polynomials on $\hat{\kappa}$ of degree p in each coordinate direction. To each $\kappa \in \mathcal{T}$ we assign nonnegative integers p_κ (the local polynomial degree), we collect the p_κ and F_κ into piecewise continuous functions $\mathbf{p} : \Omega \rightarrow \mathbb{R}$ and $\mathbf{F} : \Omega \rightarrow \mathbb{R}$, with $\mathbf{p}|_\kappa = p_\kappa$ and $\mathbf{F}|_\kappa = F_\kappa$, $\kappa \in \mathcal{T}$, respectively, and consider the finite element space

$$S \equiv S^{\mathbf{P}}(\Omega, \mathcal{T}, \mathbf{F}) := \{v \in L^2(\Omega) : v|_\kappa \circ F_\kappa \in \mathcal{Q}_{p_\kappa}(\hat{\kappa}), \kappa \in \mathcal{T}\}. \quad (3.1)$$

Similarly, we introduce the mesh quantities $\mathbf{h}_\perp : \cup_{\kappa \in \mathcal{T}} \partial\kappa \rightarrow \mathbb{R}$ and $\mathbf{r} : \Omega \rightarrow \mathbb{R}$ where $(\mathbf{h}_\perp)|_{e \subset \partial\kappa} = |\kappa|/|e|$, for all elemental faces $e \subset \Gamma$, and $(\mathbf{r}_\perp)|_\kappa = \rho_\kappa$, $\kappa \in \mathcal{T}$, respectively, where ρ_κ denotes the radius of the largest inscribed circle in the element κ .

We assume that the discretisation parameters satisfy the following conditions: on each elemental face $e = \partial\kappa \cap \partial\kappa'$, with e having positive $(d-1)$ -dimensional measure, and κ, κ' two generic neighbouring elements of the subdivision \mathcal{T} , we have

$$c_1^{-1} \leq p_\kappa/p_{\kappa'} \leq c_1, \quad c_2^{-1} \leq \mathbf{h}_\perp|_{e \subset \partial\kappa}/\mathbf{h}_\perp|_{e \subset \partial\kappa'} \leq c_2, \quad c_3^{-1} \leq \rho_\kappa/\rho_{\kappa'} \leq c_3,$$

for all $e \subset \Gamma$, with c_1, c_2, c_3 positive constants.

By Γ we denote the union of all one-dimensional element faces associated with the subdivision \mathcal{T} (including the boundary). Further we decompose Γ into three disjoint subsets $\Gamma = \Gamma_\partial \cup \Gamma_{\text{int}} = \Gamma_D \cup \Gamma_N \cup \Gamma_{\text{int}}$, where $\Gamma_{\text{int}} := \Gamma \setminus \Gamma_\partial$.

Next, we introduce some trace operators. Let κ, κ' be two (generic) elements sharing a face $e := \bar{\kappa} \cap \bar{\kappa}' \subset \Gamma_{\text{int}}$. Define the outward normal unit vectors \mathbf{n}^+ and \mathbf{n}^- on e corresponding to $\partial\kappa$ and $\partial\kappa'$, respectively. For functions $q : \Omega \rightarrow \mathbb{R}$ and $\phi : \Omega \rightarrow \mathbb{R}^2$ that may be discontinuous across Γ , we define the following quantities. For $q^+ := q|_{\partial\kappa}$, $q^- := q|_{\partial\kappa'}$ and $\phi^+ := \phi|_{\partial\kappa}$, $\phi^- := \phi|_{\partial\kappa'}$, we set

$$\{q\} := \frac{1}{2}(q^+ + q^-), \quad \{\phi\} := \frac{1}{2}(\phi^+ + \phi^-), \quad [q] := q^+ \mathbf{n}^+ + q^- \mathbf{n}^-, \quad [\phi] := \phi^+ \cdot \mathbf{n}^+ + \phi^- \cdot \mathbf{n}^-;$$

if $e \in \partial\kappa \cap \Gamma_\partial$, these definitions are modified to

$$\{q\} := q^+, \quad \{\phi\} := \phi^+, \quad [q] := q^+ \mathbf{n}, \quad [\phi] := \phi^+ \cdot \mathbf{n}.$$

Further, we divide the union of all open edges $\partial\kappa$ of every element κ into two subsets

$$\partial_- \kappa := \{x \in \partial\kappa : \mathbf{b}(x) \cdot \mathbf{n}(x) < 0\}, \quad \partial_+ \kappa := \{x \in \partial\kappa : \mathbf{b}(x) \cdot \mathbf{n}(x) > 0\},$$

where $\mathbf{n}(\cdot)$ denotes the unit outward normal vector function associated with the element κ ; we call these the *inflow* and *outflow* parts of $\partial\kappa$ respectively. Then, for every element $\kappa \in \mathcal{T}$, we denote by u_κ^+ the trace of a function u on $\partial\kappa$ taken from within the element κ (interior trace). We also define the exterior trace u_κ^- of $u \in H^1(\Omega, \mathcal{T})$ for almost all $x \in \partial_- \kappa \setminus \Gamma$ to be the interior trace $u_{\kappa'}^+$ of u on the element(s) κ' that share the edges contained in $\partial_- \kappa \setminus \Gamma$ of the boundary of element κ . Then, the *upwind jump* of u across $\partial_- \kappa \setminus \Gamma$ is defined by

$$[u]_\kappa := u_\kappa^+ - u_\kappa^-.$$

We note that this definition of jump is not the same as the one above; here the sign of the jump depends on the direction of the flow, whereas $[\cdot]$ depends only on the element-numbering. We note that the subscript in this definition will be suppressed when no confusion is likely to occur.

The broken weak formulation of the problem (2.1), (2.3), from which the interior penalty DGFEM will emerge, reads

$$\text{find } u \in A \text{ such that } B(u, v) = l(v) \quad \forall v \in H^{3/2+\epsilon}(\Omega, \mathcal{T}), \quad (3.2)$$

where

$$A := \{u \in H^{3/2+\epsilon}(\Omega, \mathcal{T}) : u, \bar{\mathbf{a}}\nabla u \cdot \mathbf{n} \text{ are continuous across } \Gamma_{\text{int}}\},$$

$$\begin{aligned} B(u, v) := & \sum_{\kappa \in \mathcal{T}} \int_{\kappa} (\bar{\mathbf{a}}\nabla u \cdot \nabla v + (\mathbf{b} \cdot \nabla u) v + c u v) \, dx \\ & - \sum_{\kappa \in \mathcal{T}} \int_{\partial_{-\kappa} \cap (\Gamma_- \cup \Gamma_D)} (\mathbf{b} \cdot \mathbf{n}) u^+ v^+ \, ds - \sum_{\kappa \in \mathcal{T}} \int_{\partial_{-\kappa} \setminus \Gamma_{\partial}} (\mathbf{b} \cdot \mathbf{n}) [u] v^+ \, ds \\ & + \int_{\Gamma_D \cup \Gamma_{\text{int}}} (\theta \{\bar{\mathbf{a}}\nabla v\} \cdot [u] - \{\bar{\mathbf{a}}\nabla u\} \cdot [v] + \sigma[u] \cdot [v]) \, ds, \end{aligned}$$

and

$$\begin{aligned} l(v) := & \sum_{\kappa \in \mathcal{T}} \int_{\kappa} f v \, dx - \sum_{\kappa \in \mathcal{T}} \int_{\partial_{-\kappa} \cap (\Gamma_- \cup \Gamma_D)} (\mathbf{b} \cdot \mathbf{n}) g_D v^+ \, ds \\ & + \int_{\Gamma_D} \theta (\bar{\mathbf{a}}\nabla v \cdot \mathbf{n} + \sigma v) g_D \, ds + \int_{\Gamma_N} g_N v \, ds, \end{aligned} \quad (3.3)$$

for $\theta \in \{-1, 1\}$, with the function σ defined by

$$\sigma|_e := C_{\sigma} \left\{ \frac{\mathbf{a} \mathbf{p}^2}{\mathbf{h}_{\perp}} \right\},$$

where \mathbf{p} and \mathbf{h}_{\perp} as above, and $\mathbf{a} : \Omega \rightarrow \mathbb{R}$ with $\mathbf{a}|_{\kappa} = \|\sqrt{\bar{\mathbf{a}}}\|_2 \| \cdot \|_{L^{\infty}(\kappa)}$, $\kappa \in \mathcal{T}$ where $\|\cdot\|_2$ denotes the matrix-2-norm, and C_{σ} positive constant. Here and in the sequel, we assume that there exists a constant $\alpha > 0$, such that

$$\alpha^{-1} \leq \mathbf{a}|_{\kappa} / \mathbf{a}|_{\kappa'} \leq \alpha,$$

for all pairs of neighbouring elements κ and κ' having a common face e , whenever \mathbf{a} is non-zero.

Then, the *interior penalty DGFEM* for the problem (2.1), (2.3) is defined as follows:

$$\text{find } u_{\text{DG}} \in S \text{ such that } B(u_{\text{DG}}, v) = l(v) \quad \forall v \in S. \quad (3.4)$$

We refer to the DGFEM with $\theta = -1$ as the *symmetric interior penalty DGFEM (SIPG)*, whereas for $\theta = 1$ the DGFEM will be referred to as the *non-symmetric interior penalty DGFEM (NIPG)*. This terminology stems from the fact that when $\mathbf{b} \equiv \vec{0}$, the bilinear form $B(\cdot, \cdot)$ is symmetric if and only if $\theta = -1$. Various types of error analysis for the variants of interior penalty DGFEMs can be found in [2, 28, 20, 14, 18, 16].

We make some assumptions on the regularity of the solution and on the functions in the finite element space S . We assume that $p_i^{\kappa} \geq 1$, $i = 1, 2$, $\kappa \in \mathcal{T}$, whenever diffusion is present, in order to ensure that the matrix of the system of linear algebraic equations that arises from (3.4) is nonsingular. When the analytical solution $u \in A$, the following *Galerkin orthogonality property* holds: $B(u - u_{\text{DG}}, v) = 0$ for every $v \in S$. If the continuity assumptions involved in the definition of A are violated, as is the case, for example, in an elliptic transmission problem, the DGFEM has to be modified accordingly.

We conclude this section by introducing the notion of inverse property for the diffusion tensor $\bar{\mathbf{a}}$.

DEFINITION 3.1 (inverse property). *We shall say that the tensor $\bar{\mathbf{a}}$ has the inverse property if an inverse inequality of the form*

$$\|\sqrt{\bar{\mathbf{a}}}\nabla v\|_e^2 \leq C_{\text{inv}} p_{\kappa}^2 |e| / |\kappa| \|\sqrt{\bar{\mathbf{a}}}\nabla v\|_{\kappa}^2, \quad (3.5)$$

holds, for all $v \in \mathcal{Q}_p(\kappa)$, where e is a face of the element κ , and C_{inv} is a positive constant independent of $\bar{\mathbf{a}}$ and the discretisation parameters.

We refer to [15] for a detailed discussion on diffusion tensors satisfying the inverse property. For a discussion on how to circumvent the necessity of the inverse property assumption, by modifying the DGFEM, we refer to [16].

4. Norm Equivalences. We define the *energy norm* $||| \cdot |||$ by

$$|||w||| := \left(\sum_{\kappa \in \mathcal{T}} \|\sqrt{\bar{\mathbf{a}}} \nabla w\|_{\kappa}^2 + \|c_0 w\|_{\Omega}^2 + \frac{1}{2} \|b_n[w]\|_{\Gamma}^2 + \|\sqrt{\sigma}[w]\|_{\Gamma_D \cup \Gamma_{\text{int}}}^2 \right)^{\frac{1}{2}},$$

where $c_0 := \sqrt{c - 1/2 \nabla \cdot \mathbf{b}}$, $b_n := \sqrt{|\mathbf{b} \cdot \mathbf{n}|}$, with \mathbf{n} on $\partial\kappa$ denoting the outward normal to $\partial\kappa$ and σ as above.

We are interested in studying the symmetric and the skew-symmetric parts of the bilinear form $B(\cdot, \cdot)$. For this reason, we rewrite the numerical fluxes for the convection part of the boundary as described in the following lemma.

LEMMA 4.1. *Using the notation above, the following identity holds:*

$$\begin{aligned} & - \sum_{\kappa \in \mathcal{T}} \left(\int_{\partial_{-\kappa} \cap (\Gamma_D \cup \Gamma_-)} (\mathbf{b} \cdot \mathbf{n}) u^+ v^+ ds + \int_{\partial_{-\kappa} \setminus \Gamma_{\partial}} (\mathbf{b} \cdot \mathbf{n}) [u] v^+ ds \right) \\ & = \int_{\Gamma} \left(\frac{1}{2} |\mathbf{b} \cdot \mathbf{n}| [u] \cdot [v] - [u] \cdot \{\mathbf{b}v\} \right) ds + \frac{1}{2} \int_{\Gamma_{\partial}} (\mathbf{b} \cdot \mathbf{n}) u^+ v^+ ds. \end{aligned} \quad (4.1)$$

Proof. On each elemental inflow boundary, we have $-(\mathbf{b} \cdot \mathbf{n}) = |\mathbf{b} \cdot \mathbf{n}|$. Thus, on each $\partial_{-\kappa} \setminus \Gamma_{\partial}$, we have

$$\begin{aligned} -(\mathbf{b} \cdot \mathbf{n}) [u] v^+ &= |\mathbf{b} \cdot \mathbf{n}| [u] v^+ = \frac{1}{2} |\mathbf{b} \cdot \mathbf{n}| [u] [v] + |\mathbf{b} \cdot \mathbf{n}| [u] \{v\} \\ &= \frac{1}{2} |\mathbf{b} \cdot \mathbf{n}| [u] \cdot [v] - (\mathbf{b} \cdot \mathbf{n}) [u] \{v\} = \frac{1}{2} |\mathbf{b} \cdot \mathbf{n}| [u] \cdot [v] - [u] \cdot \{\mathbf{b}v\} \end{aligned}$$

Hence,

$$- \sum_{\kappa \in \mathcal{T}} \int_{\partial_{-\kappa} \setminus \Gamma_{\partial}} (\mathbf{b} \cdot \mathbf{n}) [u] v^+ ds = \int_{\Gamma_{\text{int}}} \left(\frac{1}{2} |\mathbf{b} \cdot \mathbf{n}| [u] \cdot [v] - [u] \cdot \{\mathbf{b}v\} \right) ds. \quad (4.2)$$

Recalling the definitions of $[\cdot]$ and $\{\cdot\}$ on the boundary Γ_{∂} , along with $-(\mathbf{b} \cdot \mathbf{n}) = |\mathbf{b} \cdot \mathbf{n}|$ and $(\mathbf{b} \cdot \mathbf{n}) = |\mathbf{b} \cdot \mathbf{n}|$ on the inflow and outflow parts of the boundary, respectively, it is immediate that

$$\int_{\Gamma_{\partial}} \left(\frac{1}{2} |\mathbf{b} \cdot \mathbf{n}| [u] \cdot [v] - [u] \cdot \{\mathbf{b}v\} \right) ds + \frac{1}{2} \int_{\Gamma_{\partial}} (\mathbf{b} \cdot \mathbf{n}) u^+ v^+ ds = - \sum_{\kappa \in \mathcal{T}} \int_{\partial_{-\kappa} \cap (\Gamma_D \cup \Gamma_-)} (\mathbf{b} \cdot \mathbf{n}) u^+ v^+ ds. \quad (4.3)$$

Summing (4.2) and (4.3), the result follows. \square

Splitting ideas for the numerical fluxes of DGFEM for first order hyperbolic problems are presented in [6] when the convection is in divergence form; here we modified the argument presented in that work for the case where the convection is of the form $\mathbf{b} \cdot \nabla u$ (cf. also Section 2.7 in [14]).

Motivated by identity (4.1), we decompose $B(\cdot, \cdot)$ into symmetric and skew-symmetric components.

LEMMA 4.2. *The bilinear form can be decomposed into symmetric and skew-symmetric parts:*

$$B(u, v) = B_{\text{symm}}(u, v) + B_{\text{skew}}(u, v),$$

for all $u, v \in H^{3/2+\varepsilon}(\Omega, \mathcal{T})$, where

$$\begin{aligned} B_{\text{symm}}(u, v) &:= \sum_{\kappa \in \mathcal{T}} \int_{\kappa} (\bar{\mathbf{a}} \nabla u \cdot \nabla v + c_0^2 u v) \, dx + \frac{1}{2} \int_{\Gamma} |\mathbf{b} \cdot \mathbf{n}| [u] \cdot [v] \, ds \\ &\quad + \int_{\Gamma_D \cup \Gamma_{\text{int}}} \left(\frac{\theta - 1}{2} (\{\bar{\mathbf{a}} \nabla v\} \cdot [u] + \{\bar{\mathbf{a}} \nabla u\} \cdot [v]) + \sigma [u] \cdot [v] \right) \, ds, \end{aligned} \quad (4.4)$$

and

$$\begin{aligned} B_{\text{skew}}(u, v) &:= \frac{1}{2} \sum_{\kappa \in \mathcal{T}} \int_{\kappa} ((\mathbf{b} \cdot \nabla u) v - (\mathbf{b} \cdot \nabla v) u) \, dx + \frac{1}{2} \int_{\Gamma_{\text{int}}} ([v] \cdot \{\mathbf{b} u\} - [u] \cdot \{\mathbf{b} v\}) \, ds \\ &\quad + \frac{\theta + 1}{2} \int_{\Gamma_D \cup \Gamma_{\text{int}}} (\theta \{\bar{\mathbf{a}} \nabla v\} \cdot [u] - \{\bar{\mathbf{a}} \nabla u\} \cdot [v]) \, ds. \end{aligned}$$

Proof. Straightforward calculation yields that

$$\begin{aligned} B(u, v) &= B_{\text{symm}}(u, v) + \sum_{\kappa \in \mathcal{T}} \int_{\kappa} ((\mathbf{b} \cdot \nabla u) v + \frac{\nabla \cdot \mathbf{b}}{2} u v) \, dx - \int_{\Gamma} [u] \cdot \{\mathbf{b} v\} \, ds \\ &\quad + \frac{\theta + 1}{2} \int_{\Gamma_D \cup \Gamma_{\text{int}}} (\theta \{\bar{\mathbf{a}} \nabla v\} \cdot [u] - \{\bar{\mathbf{a}} \nabla u\} \cdot [v]) \, ds, \end{aligned}$$

with $B_{\text{symm}}(u, v)$ as defined in (4.4). Integration by parts of the second term in the first integral on the right-hand side of (4.5) yields

$$\sum_{\kappa \in \mathcal{T}} \int_{\kappa} \frac{\nabla \cdot \mathbf{b}}{2} u v \, dx = -\frac{1}{2} \sum_{\kappa \in \mathcal{T}} \int_{\kappa} ((\mathbf{b} \cdot \nabla u) v + (\mathbf{b} \cdot \nabla v) u) \, dx + \frac{1}{2} \sum_{\kappa \in \mathcal{T}} \int_{\partial \kappa} (\mathbf{b} \cdot \mathbf{n}) u^+ v^+ \, ds.$$

The result now follows by making use of the (standard) identity (see, e.g., [6])

$$\sum_{\kappa \in \mathcal{T}} \int_{\partial \kappa} (\mathbf{b} \cdot \mathbf{n}) u^+ v^+ \, ds = \int_{\Gamma} [u] \cdot \{\mathbf{b} v\} \, ds + \int_{\Gamma_{\text{int}}} \{u\} [\mathbf{b} v] \, ds,$$

and observing that $\{u\} [\mathbf{b} v] = \{\mathbf{b} u\} \cdot [\mathbf{v}]$. \square

Note that for the skew-symmetric part of the bilinear form we have $B_{\text{skew}}(u, u) = 0$, for all $u \in H^{3/2+\varepsilon}(\Omega, \mathcal{T})$.

LEMMA 4.3. *Let the domain Ω , its subdivision \mathcal{T} and the positive function σ be defined as above, with the constant C_{σ} sufficiently large when $\theta = -1$ (see [2, 14] for details). Let also $\bar{\mathbf{a}}$ satisfy the inverse property. Then, the following equivalence holds*

$$\kappa_1 \|u\|^2 \leq B(u, u) \leq \kappa_2 \|u\|^2, \quad (4.5)$$

for all $u \in S$, where κ_1 and κ_2 are positive constants independent of the mesh parameters.

Proof. We have

$$B(u, u) = B_{\text{symm}}(u, u) = \|u\|^2 + (\theta - 1) \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{\bar{\mathbf{a}} \nabla u\} \cdot [u] \, ds.$$

Thus, when $\theta = 1$, (4.5) holds with $\kappa_1 = 1 = \kappa_2$. When $\theta = -1$, we work as follows. Using the Cauchy-Schwarz inequality together with the inverse inequality (3.5), we obtain

$$\begin{aligned} 2 \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{\bar{\mathbf{a}} \nabla u\} \cdot [u] \, ds &= 2 \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{\sqrt{\bar{\mathbf{a}}} \nabla u\} \cdot [\sqrt{\bar{\mathbf{a}}} u] \, ds \\ &\leq \frac{1}{2} \sum_{\kappa \in \mathcal{T}} \|\sqrt{\bar{\mathbf{a}}} \nabla u\|_{\kappa}^2 + \frac{1}{2} \|\sqrt{\sigma} [u]\|_{\Gamma_D \cup \Gamma_{\text{int}}}^2, \end{aligned} \quad (4.6)$$

using Young's inequality and choosing C_σ sufficiently large. Thus, the left inequality in (4.5) holds with $\kappa_1 = \frac{1}{2}$ and the right inequality holds with $\kappa_2 = \frac{3}{2}$. \square

Along the same lines, we can show the continuity for the full bilinear form $B(\cdot, \cdot)$.

THEOREM 4.4. *Let the domain Ω , its subdivision \mathcal{T} and the positive function σ be defined as above, with the constant C_σ sufficiently large when $\theta = -1$ (see [2, 14] for details). Let also $\bar{\mathbf{a}}$ satisfy the inverse property. Further assume that $c_0(x) \geq \gamma_0^\kappa > 0$ for all $x \in \kappa$, for each $\kappa \in \mathcal{T}$. Then, for $u, v \in S$ and $C > 0$ constant, independent of the problem and the mesh parameters, we have*

$$|B(u, v)| \leq C \max_{\kappa \in \mathcal{T}} \{1, C_\kappa\} \|u\| \|v\|,$$

where $C_\kappa := \min\{\|(\gamma_0^\kappa \sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)}, \|\frac{\mathbf{p}^2 \mathbf{b}}{(\gamma_0^\kappa)^2 \mathbf{r}}\|_{L^\infty(\kappa)}\}$, assuming the convention

$$\|(\gamma_0^\kappa \sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)} = +\infty$$

on the elements where $\sqrt{\bar{\mathbf{a}}}$ is not positive definite.

Proof. We have

$$B_{\text{symm}}(u, v) \leq \|u\| \|v\| + \frac{\theta - 1}{2} \left(\left| \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{\bar{\mathbf{a}} \nabla u\} \cdot [v] \, ds \right| + \left| \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{\bar{\mathbf{a}} \nabla v\} \cdot [u] \, ds \right| \right). \quad (4.7)$$

Using Cauchy-Schwarz inequality together with the inverse inequality (3.5), and working as in (4.6), we conclude that

$$\int_{\Gamma_D \cup \Gamma_{\text{int}}} \{\bar{\mathbf{a}} \nabla u\} \cdot [v] \, ds \leq \frac{1}{2} \sum_{\kappa \in \mathcal{T}} \|\sqrt{\bar{\mathbf{a}}} \nabla u\|_\kappa^2 + \frac{1}{2} \|\sqrt{\sigma} [v]\|_{\Gamma_D \cup \Gamma_{\text{int}}}^2, \quad (4.8)$$

and similarly for the last term on the right-hand side of (4.7).

Next, we bound each of the terms of $B_{\text{skew}}(\cdot, \cdot)$ separately, for every $\kappa \in \mathcal{T}$. To this end, we consider 2 cases:

Case 1. $\bar{\mathbf{a}}$ is positive definite.

We have

$$\left| \int_\kappa b \cdot \nabla u \, v \, dx \right| = \left| \int_\kappa b^T \nabla u \, v \, dx \right| \leq \|(\gamma_0^\kappa \sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)} \|\sqrt{\bar{\mathbf{a}}} \nabla u\|_\kappa \|c_0 v\|_\kappa;$$

for $e \subset \Gamma_{\text{int}}$, where $e = \bar{\kappa}_+ \cap \bar{\kappa}_-$, for two (generic) elements $\kappa_+, \kappa_- \in \mathcal{T}$, we have

$$\begin{aligned} \left| \int_e [u] \cdot \{\mathbf{b}v\} \, ds \right| &\leq \int_e |\{\sqrt{\bar{\mathbf{a}}}\}_2| \frac{\mathbf{p}}{\sqrt{\mathbf{h}_\perp}} [u] \|(\sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\| \frac{\sqrt{\mathbf{h}_\perp}}{\mathbf{p}} |\{v\}| \, ds \\ &\leq \sum_{\kappa \in \{\kappa_+, \kappa_-\}} \|(\gamma_0^\kappa \sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)} \|\sqrt{\sigma} u\|_e \|c_0 v\|_\kappa, \end{aligned}$$

and similarly for the skew-symmetric counterparts. If $\theta = 1$, the remaining terms of the skew-symmetric part of the bilinear form are bounded using (4.8).

Case 2. $\bar{\mathbf{a}}$ is positive semi-definite, but *not* positive definite.

We have

$$\left| \int_\kappa b \cdot \nabla u \, v \, dx \right| \leq \|\mathbf{b}\|_{L^\infty(\kappa)} \|\nabla u\|_\kappa \|v\|_\kappa \leq C \|(\gamma_0^\kappa)^{-2} \mathbf{b}\|_{L^\infty(\kappa)} \frac{\mathbf{p}^2}{\mathbf{r}} \Big|_\kappa \|c_0 u\|_\kappa \|c_0 v\|_\kappa.$$

Also, for $e \in \Gamma_{\text{int}}$, where $e = \bar{\kappa}_+ \cap \bar{\kappa}_-$, for two (generic) elements $\kappa_+, \kappa_- \in \mathcal{T}$, we have

$$\begin{aligned} \left| \int_e [u] \cdot \{\mathbf{b}v\} ds \right| &\leq \|\mathbf{b}\|_{L^\infty(e)} \| [u] \|_e \| \{v\} \|_e \\ &\leq C \sum_{\kappa \in \{\kappa_+, \kappa_-\}} \left\| \frac{\mathbf{p}^2 \mathbf{b}}{(\gamma_0^\kappa)^2 \mathbf{h}_\perp} \right\|_{L^\infty(\kappa)} \|c_0 v\|_\kappa \|c_0 u\|_\kappa \\ &\leq C \sum_{\kappa \in \{\kappa_+, \kappa_-\}} \left\| \frac{\mathbf{p}^2 \mathbf{b}}{(\gamma_0^\kappa)^2 \mathbf{r}} \right\|_{L^\infty(\kappa)} \|c_0 v\|_\kappa \|c_0 u\|_\kappa, \end{aligned}$$

where the last inequality follows by observing that $\mathbf{r}|_\kappa = \mathbf{r}|_{e \subset \partial \kappa} \leq \mathbf{h}_\perp|_{e \subset \partial \kappa}$. Similarly we can bound the rest of the terms. If $\theta = 1$, the remaining terms of the skew-symmetric part of the bilinear form are bounded using (4.8). Summing up the resulting bounds, and using the discrete version of the Cauchy-Schwarz inequality, the result follows. \square

Next we consider the case when $c_0 \equiv 0$ on Ω . This includes the case of pure convection-diffusion where the wind \mathbf{b} is an incompressible vector field.

THEOREM 4.5. *Let the domain Ω , its subdivision \mathcal{T} and the positive function σ be defined as above, with the constant C_σ sufficiently large when $\theta = -1$ (see [2, 14] for details). We assume that the diffusion tensor $\bar{\mathbf{a}}$ is positive definite, and that $c_0 \equiv 0$ in Ω . Let also the mesh-regularity assumption*

$$|e|^{1/(1-d)} \leq \tilde{C} \frac{\mathbf{p}^2}{\mathbf{h}_\perp} \Big|_{e \subset \partial \kappa}$$

hold for all elemental faces $e \in \Gamma_{\text{int}}$, for some constant $\tilde{C} > 0$ independent of the problem and the mesh parameters. Then, for $u, v \in S$, we have

$$B(u, v) \leq C \max_{\kappa \in \mathcal{T}} \{1, \|(\sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)}\} \|(\sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_2 \|u\| \|v\|,$$

for $C > 0$ constant independent of the problem and the mesh parameters

Proof. The symmetric part of the bilinear form along with the face integrals that contain the diffusion tensor can be bounded as in the proof of Theorem 4.4. We bound each of the terms of $B_{\text{skew}}(\cdot, \cdot)$ separately, for every $\kappa \in \mathcal{T}$. We have

$$\left| \int_\kappa b \cdot \nabla u v \, dx \right| = \left| \int_\kappa b^T \nabla u v \, dx \right| \leq \|(\sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)} \|\sqrt{\bar{\mathbf{a}}} \nabla u\|_\kappa \|v\|_\kappa.$$

Also, for $e \in \Gamma_{\text{int}}$, where $e = \bar{\kappa}_+ \cap \bar{\kappa}_-$, for two (generic) elements $\kappa_+, \kappa_- \in \mathcal{T}$, we have

$$\begin{aligned} \left| \int_e [u] \cdot \{\mathbf{b}v\} ds \right| &\leq \int_e |\{\sqrt{\bar{\mathbf{a}}}\}_2| \frac{\mathbf{p}}{\sqrt{\mathbf{h}_\perp}} [u] \|(\sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\| \frac{\sqrt{\mathbf{h}_\perp}}{\mathbf{p}} \|\{v\}\| ds \\ &\leq \sum_{\kappa \in \{\kappa_+, \kappa_-\}} \|(\sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)} \|\sqrt{\sigma} u\|_e \|v\|_\kappa, \end{aligned}$$

and similarly for the skew-symmetric counterparts. Summing up and using the discrete version of the Cauchy-Schwarz inequality, we deduce

$$B(u, v) \leq C \max_{\kappa \in \mathcal{T}} \{1, \|(\sqrt{\bar{\mathbf{a}}})^{-1} \mathbf{b}\|_{L^\infty(\kappa)}\} (\|u\|^2 + \|u\|_\Omega^2)^{\frac{1}{2}} (\|v\|^2 + \|v\|_\Omega^2)^{\frac{1}{2}},$$

Using the Poincaré-Friedrichs inequality for broken H^1 -functions

$$\|w\|_\Omega^2 \leq C \left(\sum_{\kappa \in \mathcal{T}} \|\nabla w\|_\kappa^2 + \sum_{e \in \Gamma_{\text{int}}} |e|^{1/(1-d)} \| [w] \|_e^2 + \|w\|_{\Gamma_D}^2 \right)$$

presented in [4] (inequality (1.8) therein; following the notation of that work we have chosen Γ to be the Dirichlet boundary Γ_D), along with the mesh-regularity assumption, we deduce

$$\|w\|_\Omega^2 \leq C \|(\sqrt{\mathbf{a}})^{-1}\|_{L^\infty(\Omega)}^2 \left(\sum_{\kappa \in \mathcal{T}} \|\sqrt{\mathbf{a}} \nabla w\|_\kappa^2 + \|\sqrt{\sigma}[w]\|_{\Gamma_{\text{int}}}^2 + \|\sqrt{\sigma} w\|_{\Gamma_D}^2 \right),$$

i.e., $\|w\|_\Omega^2 \leq C \|(\sqrt{\mathbf{a}})^{-1}\|_{L^\infty(\Omega)}^2 \|w\|^2$, which yields the result. \square

5. Norm preconditioners. We now turn to the issue of preconditioner-design by first considering an analysis tool for GMRES. In the following, we use the standard definition $\|\mathbf{x}\|_H^2 = \langle \mathbf{x}, \mathbf{x} \rangle_H = \mathbf{x}^T H \mathbf{x}$ where $H \in \mathbb{R}^{n \times n}$ is symmetric and positive-definite and $\mathbf{x} \in \mathbb{R}^n$. We also introduce the following definition.

DEFINITION 5.1. Field-of-values (FOV) equivalence *Non-singular matrices $A, B \in \mathbb{R}^{n \times n}$ are said to be FOV-equivalent if there exist constants ξ_1, ξ_2 independent of n such that for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$*

$$\xi_1 \leq \frac{\langle \mathbf{x}, AB^{-1}\mathbf{x} \rangle_H}{\langle \mathbf{x}, \mathbf{x} \rangle_H}, \quad \frac{\|AB^{-1}\mathbf{x}\|_H}{\|\mathbf{x}\|_H} \leq \xi_2.$$

We write

$$A \approx_H B.$$

We remark here that FOV-equivalence implies that $\xi_1 < |\lambda(AB^{-1})| < \xi_2$, where $\lambda(AB^{-1})$ denote the eigenvalues of AB^{-1} . This is usually a desirable property in the context of preconditioning. The above definition was introduced with the following convergence result in mind.

LEMMA 5.2. *If $A \approx_H B$ the GMRES algorithm converges with respect to $\langle \cdot, \cdot \rangle_H$ in a number of iterations independent of n . Moreover, the residuals satisfy [9], [30, Thm 6.7]*

$$\frac{\|\mathbf{r}^k\|_H}{\|\mathbf{r}^0\|_H} \leq \left(1 - \frac{\xi_1^2}{\xi_2^2}\right)^{k/2}, \quad (5.1)$$

where ξ_1, ξ_2 are the constants in Definition 5.1.

The requirements of Definition 5.1 can be shown to be satisfied for the matrices resulting from our DG formulation. We first note that the coercivity result of Lemma 4.3 and the continuity result of Thms 4.4 have the following discrete counterparts

$$\min_{\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^t K \mathbf{w}}{\|\mathbf{w}\|_H^2} \geq \eta_1 \quad (5.2a)$$

$$\max_{\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^t K \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \leq \eta_2 \quad (5.2b)$$

where $\eta_1 = \kappa_1$ and η_2 is the continuity constant either in Thm 4.4 or in Thm 4.5. We denoted here the matrix representation of the discrete bilinear form $B(\cdot, \cdot)$ by K and by H the discrete representation of norm $\|\cdot\|$. Furthermore, the equivalence (4.5) has the following discrete version

$$\kappa_1 \|\mathbf{u}\|_H^2 \leq \|\mathbf{u}\|_{K_s}^2 \leq \kappa_2 \|\mathbf{u}\|_H^2, \quad (5.3)$$

where $K_s = (K + K^T)/2$. The following two results taken from [26] are sufficient to ensure that H is FOV-equivalent to K and therefore represents a good preconditioning candidate.

PROPOSITION 5.3. *Let (5.2) hold. Then $K \approx_{H^{-1}} H$.*

Given equivalence (5.3), one deduces that (5.2) holds also with H replaced by K_s so that we get the following

PROPOSITION 5.4. *Let (5.2), (5.3) hold. Then $K \approx_{K_s^{-1}} K_s$.*

The last two results suggest that both H and K_s are candidates for optimal preconditioning inside a GMRES algorithm (cf. Lemma 5.2). However, only the latter can be employed in an efficient fashion, as we describe below.

5.1. Three-term GMRES. It is well-known that while GMRES is one of the most robust methods available, it is not also the most efficient. In particular, the construction and storage of orthonormal Arnoldi vectors is the main hindrance in a practical context. A short (m -)term recurrence for GMRES (which entails the storage of only m vectors) is not guaranteed to exist for any given matrix. However, there is a certain class of matrices which affords a 3-term recurrence, as the following result shows. For a proof see [1]. See also [8] for related work on preconditioning with the symmetric part of a matrix.

LEMMA 5.5. *Let $A = I + S$, where $S = -S^T$. Then the GMRES algorithm applied to matrix A is a 3-term recurrence.* Given the above result, it is straightforward to choose a preconditioner for K since

$$K_s^{-1/2} K K_s^{-1/2} = I + S$$

where S is a skew-symmetric matrix. From an implementation point of view, employing GMRES with system matrix $K_s^{-1/2} K K_s^{-1/2}$ is equivalent to running GMRES in the K_s -inner product and using K_s as a left-preconditioner (see [26] for more details). This is our approach below. We end with the theoretical bound on GMRES convergence for the DG convection-diffusion problem.

THEOREM 5.6. *Let (5.2) hold. Then the residuals of the 3-term GMRES algorithm in the K_s -inner product satisfy*

$$\frac{\|\mathbf{r}^k\|_{K_s^{-1}}}{\|\mathbf{r}^0\|_{K_s^{-1}}} \leq \left(1 - \frac{\eta_1^2}{\eta_2^2}\right)^{k/2}. \quad (5.4)$$

The above bound is discretization-independent only if the diffusion tensor $\bar{\mathbf{a}}$ is positive definite. In this case, the constant η_2 is independent of both h and p , though it remains dependent on the PDE coefficients. For the case when the diffusion tensor is positive semi-definite we cannot expect optimal performance. While the constant η_1 is still innocuous, the continuity constant η_2 depends effectively on both discretization and coefficients. In particular, for uniform meshes of size h with constant degree p of the approximating polynomials, we essentially have at points where the diffusion vanishes

$$\eta_2 = \frac{p^2}{h} \max_{\kappa \in \mathcal{T}} \|\mathbf{b}/(\gamma_0^\kappa)^2\|_{L^\infty(\kappa)}.$$

This is a pessimistic estimate, as we shall see below.

6. Numerical Examples. We now demonstrate the validity of our theoretical bounds on two classes of problems. The first corresponds to a diffusion tensor which is positive-definite. For this class we chose to present standard convection-diffusion problems, given their ubiquitous nature in fluid modelling. We note that for non-diagonal diffusion tensors the results are expected to be similar. The second class of problems corresponds to a singular diffusion tensor. In this case, performance should deteriorate with the discretization.

6.1. The positive definite case. Our first example is that of a standard convection-diffusion operator. In this case $\bar{\mathbf{a}} = \epsilon I_2$ is positive definite. We chose the work with two values for \mathbf{b} : constant and circular. We solved

$$-\epsilon \Delta u + \mathbf{b} \cdot \nabla u = f \quad \text{for } (x, y) \in (0, 1)^2,$$

with $\mathbf{b} = (1, 1)$, subject to a Dirichlet boundary condition, which, along with the forcing function f , is chosen so that the analytical solution is

$$u(x, y) = x + y(1 - x) + \frac{e^{-\frac{1}{\epsilon}} - e^{-\frac{(1-x)(1-y)}{\epsilon}}}{1 - e^{-\frac{1}{\epsilon}}}.$$

This problem was considered in [20] (Example 3). The solution exhibits boundary layer behaviour along $x = 1$ and $y = 1$, and the layers become steeper as $\epsilon \rightarrow 0$. Note that the theory developed above includes this case, as here $\nabla \cdot \mathbf{b} = 0$ on Ω .

We solved the problem for a range of ϵ . Discretisations for a range of uniform h (meshsize) and p (degree of polynomial approximation) were employed. The results are presented Table 6.1.

p	n	$\epsilon = 0.5$	$\epsilon = 0.1$	$\epsilon = 0.05$	$\epsilon = 0.01$
1	2,500	7	15	22	77
	10,000	7	15	22	80
	40,000	7	14	22	80
2	5,625	7	14	22	80
	22,500	6	14	22	80
	90,000	6	14	21	78
3	10,000	6	14	22	79
	40,000	6	14	22	78
	160,000	6	13	21	78

TABLE 6.1

GMRES iterations for DGFEM-discretisation of the convection diffusion problem with constant wind $b = (1, 1)$ and with preconditioner K_s .

As predicted by theory, the number of iterations is independent of discretisation parameters. The dependence of ϵ of bound (5.4) gives also a description of convergence behaviour.

p	n	$\epsilon = 0.5$	$\epsilon = 0.1$	$\epsilon = 0.01$
1	2,500	12	13	7
	10,000	36	40	29
	40,000	124	117	69
2	5,625	18	17	12
	22,500	61	59	60
	90,000	235	231	137
3	10,000	39	29	23
	40,000	112	114	100
	160,000	> 300	> 300	> 300

TABLE 6.2

GMRES iterations for DGFEM-discretisation of the convection diffusion problem with $ILU(10^{-2})$ preconditioning.

For comparison purposes, we included the corresponding GMRES runs for the choice of a black-box preconditioner such as ILU. The results are presented in Table 6.2. We can see that while the number of iterations is low for some values of the parameters, the overall convergence behaviour is quite undesirable, with iteration counts growing with both discretisation parameters. Thus, while the number of iterations appears to be decreasing with ϵ , it is exactly for this range that the discretisation parameters have to be increased in order to resolve layers. The resulting convergence behaviour becomes

rapidly too costly to implement in practice. We note here that the ILU preconditioner is implemented with a standard *full* GMRES routine, which means that the storage increases with every iteration.

We now consider the “driven cavity lid” singularly perturbed convection-diffusion equation

$$-\epsilon \Delta u + \mathbf{b} \cdot \nabla u = 0 \quad \text{for } (x, y) \in (0, 1)^2,$$

with $0 < \epsilon \leq 1$ and

$$\mathbf{b}^T = (2(2y - 1)(1 - (2x - 1)^2), -2(2x - 1)(1 - (2y - 1)^2),$$

subject to the Dirichlet boundary conditions

$$u(x, y) = xy(1 - y), \quad \text{for } (x, y) \in \partial\Omega.$$

The solution exhibits boundary layer behaviour along $x = 1$ and the layers become steeper as $\epsilon \rightarrow 0$. We note that $c_0 = -1/2 \nabla \cdot \mathbf{b} \equiv 0$ on Ω . The iteration count for this case is presented in Table 6.3. We remark here that the same behaviour is expected and observed.

p	n	$\epsilon = 0.5$	$\epsilon = 0.1$	$\epsilon = 0.05$	$\epsilon = 0.01$
1	2,500	5	12	20	67
	10,000	5	11	19	67
	40,000	5	11	19	66
2	5,625	5	11	19	66
	22,500	5	11	19	65
	90,000	5	11	18	65
3	10,000	5	11	18	65
	40,000	5	11	18	64
	160,000	5	10	17	64

TABLE 6.3

GMRES iterations for DGFEM-discretisation of the convection diffusion problem with circular wind and with preconditioner K_s .

6.2. The positive semi-definite case. The second class of examples we consider is that for symmetric positive semidefinite diffusion tensors. We first consider the Grusin-type boundary-value problem

$$\begin{aligned} -u_{xx} - 16x^6 u_{yy} + (1 - y)u_y &= f \quad \text{in } \Omega \equiv (-1, 1)^2, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{6.1}$$

with f is chosen so that the analytical solution is

$$u(x, y) = (1 - x^2)(1 - y^2)(x^8 + y^2)^{1/4}. \tag{6.2}$$

The diffusion tensor is positive-definite everywhere except at $x = 0$. Note that the analytical solution does not belong globally to $H^1(\Omega)$ due to a singularity of the gradient at the origin; nevertheless, it is analytic in $\bar{\Omega} \setminus \{(0, 0)\}$. The iteration count for this case is presented in Table 6.4. As expected, the results depend on both discretization parameters. However, while the dependence of the number of iterations is roughly like h^{-1} , the p -dependence is better than the predicted p^2 .

p	n	its
1	2,304	22
	10,000	39
	40,000	67
2	5,184	42
	22,500	74
	90,000	121
3	9,216	62
	40,000	107
	160,000	183

TABLE 6.4

GMRES iterations for DGFEM-discretisation of the degenerate convection diffusion problem with preconditioner K_s .

Finally, we consider the following equation on $\Omega = (-1, 1)^2$

$$\begin{aligned} -x^2 u_{yy} + u_x + u &= 0, & \text{for } -1 \leq x \leq 1, y > 0, \\ u_x + u &= 0, & \text{for } -1 \leq x \leq 1, y \leq 0, \end{aligned}$$

with analytical solution

$$u(x, y) = \begin{cases} \sin\left(\frac{1}{2}\pi(1+y)\right) \exp\left(-\left(x + \frac{\pi^2}{4} \frac{x^3}{3}\right)\right), & \text{for } -1 \leq x \leq 1, y > 0; \\ \sin\left(\frac{1}{2}\pi(1+y)\right) \exp(-x), & \text{for } -1 \leq x \leq 1, y \leq 0, \end{cases}$$

along with an appropriate Dirichlet boundary condition. This problem is of changing-type, as there exists a second order term for $y > 0$, which is no longer present for $y \leq 0$. Moreover, we can easily verify that its analytical solution u exhibits a discontinuity along $y = 0$, although the derivative of u , in the direction normal to this line of discontinuity in u , is continuous across $y = 0$. We test the performance of DGFEM with $\theta = 1$ by employing various meshes. As mentioned at the end of Section 3, we have to modify the method by setting $\sigma|_e = 0$ for all element edges $e \subset (-1, 1) \times \{0\}$, where σ_e denotes the discontinuity-penalisation parameter; this is done in order to avoid penalising physical discontinuities. Note that the diffusive flux $(\sqrt{\mathbf{a}} \nabla u) \cdot \mathbf{n}$ is still continuous across $y = 0$, and thus the method still applies.

The results are shown in Table 6.5. While the dependence on parameters h and p has not changed radically, the actual iteration count is much greater, highlighting the fact that coercivity is weaker in that part of the domain for which $y \leq 0$.

6.3. Implementation issues. We used the 3-term GMRES algorithm in the K_s -inner product with zero initial guess and a tolerance of 10^{-6} for the relative residual measured in the norm $\|\cdot\|_{K_s^{-1}}$. We note here that this norm is relevant from the point of view of finite element convergence (see [1] for more details). Regarding the implementation of our preconditioner, we have to point out that this is not necessarily a cheap procedure. In fact, our preconditioning strategy has shifted the focus from the original non-symmetric linear system involving matrix K to solving a problem with a symmetric and positive-definite matrix K_s . Apart from allowing for a storage-free method, this is an easier task which can be approached in a variety of ways. Moreover, one can suitably relax this procedure, without deterioration in the GMRES convergence. We describe this procedure in greater detail in [17]. Our solver of choice, is a generalization of that described in [25] for standard finite element methods. Essentially, this is domain decomposition nested inside a GMRES routine and with adaptive preconditioning on the internal boundary which exhibits convergence independent of discretisation parameters. For more details see [17].

p	n	its
1	576	51
	2,304	120
	9,216	287
2	1,296	120
	5,184	301
	20,736	727
3	2,304	197
	9,216	535
	36,684	1,247

TABLE 6.5

GMRES iterations for DGFEM-discretisation of the degenerate convection diffusion problem with preconditioner K_s .

7. Summary. Discontinuous Galerkin methods raise new challenges with regard to the solution of the ensuing linear system. Due to the nature of the discretisation, the problems can become very quickly very large, particularly when the degree of the polynomial approximation is also increased. In the case of general elliptic problems, useful iterative methods can be designed by taking into account the finite element formulation. In this work we devised a preconditioner based on equivalence to the norm inherited from the finite element space. The preconditioner is employed together with a 3-term GMRES routine in order to maintain storage at a minimum. The resulting solver was applied to the case of DGFEM discretizations of problems with non-negative characteristic form. Theoretical results were derived to explain the convergence behaviour.

REFERENCES

- [1] M. ARIOLI, D. LOGHIN, AND A. J. WATHEN, *Stopping criteria for iterations in finite element methods.*, Numer. Math., 99 (2005), pp. 381–410.
- [2] D. N. ARNOLD, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.
- [3] F. BASSI AND S. REBAY, *GMRES discontinuous Galerkin solution of the compressible Navier-Stokes equations*, in Discontinuous Galerkin methods (Newport, RI, 1999), vol. 11 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, 2000, pp. 197–208.
- [4] S. C. BRENNER, *Poincaré-Friedrichs inequalities for piecewise H^1 functions*, SIAM J. Numer. Anal., 41 (2003), pp. 306–324 (electronic).
- [5] S. C. BRENNER AND J. ZHAO, *Convergence of multigrid algorithms for interior penalty methods*, Appl. Numer. Anal. Comput. Math., 2 (2005), pp. 3–18.
- [6] F. BREZZI, L. D. MARINI, AND E. SÜLI, *Discontinuous Galerkin methods for first-order hyperbolic problems*, Math. Models Methods Appl. Sci., 14 (2004), pp. 1893–1903.
- [7] P. CASTILLO, *Performance of discontinuous Galerkin methods for elliptic PDEs*, SIAM J. Sci. Comput., 24 (2002), pp. 524–547 (electronic).
- [8] P. CONCUS AND G. H. GOLUB, *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, in Proc. Second Internat. Symp. on Computing Methods in Applied Sciences and Engineering, R. Glowinski and J. L. Lions, eds., vol. 134 of Lecture Notes in Economics and Mathematical Systems, Berlin, 1976, Springer Verlag.
- [9] H. C. ELMAN, *Iterative methods for large sparse non-symmetric systems of linear equations*, PhD thesis, Yale University, New Haven, 1982.
- [10] D. J. ESTEP AND R. W. FREUND, *Using Krylov-subspace iterations in discontinuous Galerkin methods for nonlinear reaction-diffusion systems*, in Discontinuous Galerkin Methods, B. Cockburn, G. E. Karniadakis, and C.-W. Shu, eds., Springer-Verlag, Berlin, 2000, pp. 327–335.
- [11] V. FABER, T. A. MANTEUFFEL, AND S. V. PARTER, *On the theory of equivalent operators and application to the numerical solution of uniformly elliptic partial differential equations*, Adv. Appl. Math., 11 (1990), pp. 109–163.
- [12] X. FENG AND O. A. KARAKASHIAN, *Two-level additive Schwarz methods for a discontinuous Galerkin approximation*

- of second order elliptic problems, SIAM J. Numer. Anal., 39 (2001), pp. 1343–1365 (electronic).
- [13] ———, *Two-level non-overlapping Schwarz preconditioners for a discontinuous Galerkin approximation of the bi-harmonic equation*, J. Sci. Comput., 22/23 (2005), pp. 289–314.
 - [14] E. H. GEORGIOULIS, *Discontinuous Galerkin methods on shape-regular and anisotropic meshes*, D.Phil. Thesis, University of Oxford, (2003). Available also at <http://www.math.1e.ac.uk/PEOPLE/eg64/pubs.html>.
 - [15] ———, *hp-version interior penalty discontinuous galerkin finite element methods on anisotropic meshes.*, International Journal of Numerical Analysis and Modeling, 3 (2006), pp. 52–79.
 - [16] E. H. GEORGIOULIS AND A. LASIS, *A note on the design of hp-version interior penalty discontinuous galerkin finite element methods for degenerate problems.*, IMA J. Numer. Anal., 26 (2006), pp. 381–390.
 - [17] E. H. GEORGIOULIS AND D. LOGHIN, *A multilevel domain decomposition method for discontinuous Galerkin methods.* in preparation, 2006.
 - [18] E. H. GEORGIOULIS AND E. SÜLI, *Optimal error estimates for the hp-version interior penalty discontinuous Galerkin finite element method*, IMA J. Numer. Anal., 25 (2005), pp. 205–220.
 - [19] J. GOPALAKRISHNAN AND G. KANSCHAT, *A multilevel discontinuous Galerkin method*, Numer. Math., 95 (2003), pp. 527–550.
 - [20] P. HOUSTON, C. SCHWAB, AND E. SÜLI, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 39 (2002), pp. 2133–2163 (electronic).
 - [21] P. HOUSTON AND E. SÜLI, *Stabilised hp-finite element approximation of partial differential equations with nonnegative characteristic form*, Computing, 66 (2001), pp. 99–119.
 - [22] G. KANSCHAT, *Preconditioning methods for local discontinuous Galerkin discretizations*, SIAM J. Sci. Comput., 25 (2003), pp. 815–831.
 - [23] C. LASSER AND A. TOSELLI, *An overlapping domain decomposition preconditioner for a class of discontinuous Galerkin approximations of advection-diffusion problems*, Math. Comp., 72 (2003), pp. 1215–1238.
 - [24] R. D. LAZAROV, S. Z. TOMOV, AND P. S. VASSILEVSKI, *Interior penalty discontinuous approximations of elliptic problems*, Comput. Methods Appl. Math., 1 (2001), pp. 367–382.
 - [25] D. LOGHIN, D. RUIZ, AND A. TOUHAMI, *Adaptive preconditioners for nonlinear systems of equations*, J. Comp. Applied Math., (2006), pp. 362–374.
 - [26] D. LOGHIN AND A. J. WATHEN, *Analysis of block preconditioners for saddle-point problems*, SIAM J. Sci. Comput., 25 (2004), pp. 2029–2049.
 - [27] O. A. OLEINIK AND E. V. RADKEVIČ, *Second order equations with nonnegative characteristic form*, Plenum Press, New York, 1973. Translated from the Russian by Paul C. Fife.
 - [28] B. RIVIÈRE, M. F. WHEELER, AND V. GIRAULT, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. I*, Comput. Geosci., 3 (1999), pp. 337–360.
 - [29] T. RUSTEN, P. S. VASSILEVSKI, AND R. WINTHER, *Interior penalty preconditioners for mixed finite element approximations of elliptic problems*, Math. Comp., 65 (1996), pp. 447–466.
 - [30] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Co., Boston, 1996.