Krylov-Subspace Preconditioners
for
Discontinuous Galerkin Finite Element Methods

Emmanuil H. Georgoulis∗ and Daniel Loghin †
May 20, 2006

Abstract

Standard (conforming) finite element approximations of convection-dominated convection-diffusion problems often exhibit poor stability properties that manifest themselves as non-physical oscillations polluting the numerical solution. Various techniques have been proposed for the stabilisation of finite element methods (FEMs) for convection-diffusion problems, such as the popular streamline upwind Petrov-Galerkin (SUPG) method, and its variants. During the last decade, families of discontinuous Galerkin finite element methods (DGFEMs) have been proposed 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 corresponding linear solvers.

This work aims to provide an overview of the current state of affairs in the solution of DGFEM-linear problems and present some recent results on the preconditioning of stiffness matrices arising from DGFEM discretisations of steady-state convection-diffusion boundary-value problems. More specifically, preconditioners are derived for which theoretical results indicate GMRES performance independent of discretisation parameters.

1 Introduction

In many practical examples of convection-diffusion boundary value problems, diffusion can be small compared to convection. Such cases usually give rise to stability concerns of classical finite element discretisations. Various techniques have been proposed for the stabilisation of

∗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
finite element methods (FEMs) for convection-diffusion problems, such as the popular streamline upwind Petrov-Galerkin (SUPG) method, the residual-free bubble method, and other up-winding techniques.

In recent years, there has been 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 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.

Much work has been put into developing optimal iterative solvers for standard finite element discretisations of convection-diffusion problems, the driving force being the need for fast solvers for CFD. Among the successful approaches are Krylov methods, multigrid methods, domain decomposition or combinations thereof. In the case of DG methods, the main approaches to date include domain decomposition methods [14] and block-preconditioned Krylov methods of the problem reformulated as a system of PDE [13].

Our approach is based on a number of existing results regarding preconditioning for elliptic problems [6], [15], together with a suitable Krylov subspace solver. In particular, we use the concept of field-of-value equivalence to identify a useful preconditioner as well as to derive useful analytical results regarding convergence properties of our iterative scheme.

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 $\Omega$ be a bounded open (curvilinear) polygonal domain in $\mathbb{R}^2$, and let $\Gamma \partial$ 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 steady state convection-diffusion equation

$$Lu \equiv -\epsilon \Delta u + b \cdot \nabla u = f \quad \text{in } \Omega,$$

(1)

where $f \in L^2(\Omega)$, and $b = (b_1, b_2)^T$ is an incompressible vector field (i.e., $\nabla \cdot b = 0$) whose entries $b_i, i = 1, 2$, are Lipschitz continuous real-valued functions on $\Omega$.

We decompose $\Gamma \partial$ into two parts $\Gamma_D, \Gamma_N$ and we impose Dirichlet and Neumann boundary conditions, respectively, via

$$u = g_D \text{ on } \Gamma_D,$$

$$\epsilon \nabla u \cdot n = g_N \text{ on } \Gamma_N,$$

(2)

where we adopt the (physically reasonable) hypothesis that $b \cdot n \geq 0$ on $\Gamma_N$, whenever the latter is nonempty, where $n$ denotes the outward normal to $\Gamma \partial$. 

2
3 Discontinuous Galerkin Finite Element Method

We shall denote by $H^s(\Omega)$ the standard Hilbertian Sobolev space of index $s \geq 0$ of real-valued functions defined on $\Omega$.

Let $T$ be a subdivision of $\Omega$ into disjoint open elements $\kappa \in T$ such that each side of $\kappa$ has at most one regular hanging node. We let $h_\kappa := \text{diam}(\bar{\kappa})$. We assume that the subdivision $T$ is constructed via mappings $F_\kappa$, where $F_\kappa : \bar{\kappa} := (-1,1)^2 \to \kappa$ 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 $\kappa \in T$ forms a covering of the closure of $\Omega$, i.e., $\bar{\Omega} = \bigcup_{\kappa \in T} \bar{\kappa}$.

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

We assign to the subdivision $T$ an appropriate sub-index on $\partial \kappa$ that may be discontinuous across $\Gamma$, we define the following quantities. For $q : \Omega \to \mathbb{R}$ and $\phi : \Omega \to \mathbb{R}^2$ that may be discontinuous across $\Gamma$, 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 $\kappa$ 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 $\kappa$; we call these the inflow and outflow parts of $\partial \kappa$ respectively. Then, for every element $\kappa \in T$, we denote by $u_\kappa^+$ the trace of a function $u$ on $\partial \kappa$ taken from within the element $\kappa$ (interior trace). We also define the exterior trace $u_\kappa^-$ of $u \in H^1(\Omega, T)$ for almost all $x \in \partial_- \kappa \setminus \Gamma$ to be the interior trace $u_\kappa^+$ of $u$ on the element(s) $\kappa'$ that share the edges contained in $\partial_- \kappa \setminus \Gamma$ of the boundary of element $\kappa$. 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 (1), (2), from which the interior penalty DGFEM will emerge, reads

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

where

$$A := \{ u \in H^{3/2+\epsilon}(\Omega, T) : u, \epsilon \nabla u \cdot \nu \text{ are continuous across } \Gamma \} ,$$

$$B(u, v) := \sum_{\kappa \in T} \int_{\kappa} (\epsilon \nabla u \cdot \nabla v + (b \cdot \nabla u) v) \, dx$$

$$- \sum_{\kappa \in T} \int_{\partial_- \kappa \cap \Gamma_D} (b \cdot n) u^+ v^+ \, ds - \sum_{\kappa \in T} \int_{\partial_- \kappa \cap \Gamma_D} (b \cdot n) |u|^+ v^+ \, ds$$

$$+ \int_{\Gamma_D \cup \Gamma_{\text{int}}} \left( \theta \left\{ \epsilon \nabla v \right\} \cdot [u] - \left\{ \epsilon \nabla u \right\} \cdot [v] + \sigma [u] \cdot [v] \right) ds,$$

and

$$l(v) := \sum_{\kappa \in T} \int_{\kappa} fv \, dx - \sum_{\kappa \in T} \int_{\partial_- \kappa \cap \Gamma_D} (b \cdot n) g_D v^+ \, ds$$

$$+ \int_{\Gamma_D} \theta (\epsilon \nabla v \cdot n + \sigma v) g_D \, ds + \int_{\Gamma_N} g_N v \, ds,$$

for $\theta \in \{-1, 0, 1\}$, with the function $\sigma$ defined by

$$\sigma|_e := C_\sigma \frac{\epsilon \{p^2\}}{h^*_{\perp}},$$

where $h^*_{\perp} = \frac{|\kappa| + |\kappa'|}{2|e|}$, for $e \subset \bar{\kappa} \cap \bar{\kappa}'$, and $C_\sigma$ sufficiently large positive constant.

Then, the interior penalty DGFEM for the problem (1), (2) is defined as follows:

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

We refer to the DGFEM with $\theta = -1$ as the symmetric interior penalty DGFEM (SIPG), for $\theta = 0$ the DGFEM is referred as the incomplete interior penalty DGFEM (HIPG) 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 $b \equiv 0$, 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, 12, 8, 10] (cf. also the references therein).

We make some assumptions on the regularity of the solution and on the functions in the finite element space $S$. We assume that $p^i_\kappa \geq 1$, $i = 1, 2$, $\kappa \in T$, whenever diffusion is present, in order to ensure that the matrix of the system of linear algebraic equations that arises from (6) 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.
4 Norm Equivalences

We define the energy norm $|||\cdot|||$ by

$$|||w|||^2 := \sum_{\kappa \in T} \left( \|\sqrt{\kappa} w\|_{\kappa}^2 + \frac{1}{2} \|b_n w^+\|_{\partial_{\kappa} \cap \Gamma_D}^2 + \frac{1}{2} \|b_n w^+\|_{\partial_{\kappa} \setminus \Gamma_D}^2 + \frac{1}{2} \|b_n [w]\|_{\partial_{\kappa} \setminus \Gamma_D}^2 \right) + \|\sqrt{\kappa} w\|_{\Gamma_D \cup \Gamma_{int}}^2,$$

where $b_n := \sqrt{b \cdot n}$, with $n$ on $\partial \kappa$ denoting the outward normal to $\partial \kappa$ and $\sigma$ as above.

The coercivity of the bilinear form $B(\cdot, \cdot)$ with respect to this norm is given below, rendering $|||\cdot|||$ as a natural choice for energy norm for the DGFEM.

**Lemma 4.1** Let the domain $\Omega$, its subdivision $T$ and the positive function $\sigma$ be defined as above, with the constant $C_{\sigma}$ sufficiently large. Then, assuming that $w \in S$, we have

$$C_\theta |||w|||^2 \leq B(w, w),$$

with the constant $C_\theta > 0$ depending on the choice of $\theta = \{-1, 0, 1\}$.

The proof can be found in [2, 16, 12].

In the sequel, we shall be concerned with the case of the symmetric interior penalty DGFEM (i.e., the case $\theta = 1$) since this is the most relevant and popular method in the literature.

We are interested in studying the symmetric part of the bilinear form $B(\cdot, \cdot)$. For this reason, we split $B(\cdot, \cdot)$ as follows:

$$B(u, v) = B_{\text{diff}}(u, v) + B_{\text{conv}}(u, v),$$

where

$$B_{\text{diff}}(u, v) := \sum_{\kappa \in T} \left( \int_{\kappa} \epsilon \nabla u \cdot \nabla v \, dx + \int_{\Gamma_D \cup \Gamma_{int}} \left( - \{\epsilon \nabla v\} \cdot [u] - \{\epsilon \nabla u\} \cdot [v] + \sigma [u] \cdot [v] \right) \, ds \right), \quad (7)$$

and

$$B_{\text{conv}}(u, v) := \sum_{\kappa \in T} \left( \int_{\kappa} (b \cdot \nabla u) v \, dx - \int_{\partial_{\kappa} \cap \Gamma_D} (b \cdot n) u^+ v^+ \, ds - \int_{\partial_{\kappa} \setminus \Gamma_D} (b \cdot n) [u] v^+ \, ds \right). \quad (8)$$

It is clear that $B_{\text{diff}}(v, u) = B_{\text{diff}}(u, v)$, i.e., $B_{\text{diff}}(\cdot, \cdot)$ is symmetric. Further, we study the symmetric part of $B_{\text{conv}}(\cdot, \cdot)$.

**Lemma 4.2** Using the notation above, the following identity holds:

$$-\sum_{\kappa \in T} \left( \int_{\partial_{\kappa} \cap \Gamma_D} (b \cdot n) u^+ v^+ \, ds + \int_{\partial_{\kappa} \setminus \Gamma_D} (b \cdot n) [u] v^+ \, ds \right) = \int_{\Gamma} \left( \frac{1}{2} (b \cdot n) [u] \cdot [v] - [u] \cdot \{bv\} \right) \, ds. \quad (9)$$

For a proof we refer to [9]. Similar splitting ideas for the numerical fluxes of DGFEM for first order hyperbolic problems are presented in [3] when the convection is in divergence form; here, the argument presented in that work has been modified for the case where the convection is of the form $b \cdot \nabla u$.

Motivated by the identity (9), we decompose $B_{\text{conv}}(\cdot, \cdot)$ into symmetric and skew-symmetric components. To this end, we consider the splitting

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

5
where

\[ B_{\text{conv}}(u, v) := \int_{\Gamma} \frac{1}{2} \mathbf{b} \cdot \mathbf{n}[u] \cdot [v] \, ds \]

and

\[ B_{\text{conv}}^{\text{skew}}(u, v) := \sum_{\kappa \in T} \int_{\kappa} (\mathbf{b} \cdot \nabla u) v \, dx - \int_{\Gamma} [u] \cdot \{ bv \} \, ds. \]

In other words, the symmetric part of the bilinear form \( B(\cdot, \cdot) \), denoted by \( B_{\text{symm}}(\cdot, \cdot) \) is given by the sum of the symmetric part of the convection part of the bilinear form and the diffusion part of the bilinear form, i.e.,

\[ B_{\text{symm}}(u, v) = B_{\text{diff}}(u, v) + B_{\text{conv}}^{\text{symm}}(u, v). \]

**Lemma 4.3** Using the notation above, the following equivalence holds:

\[ \kappa_1 \| u \|_2^2 \leq B_{\text{symm}}(u, u) \leq \kappa_2 \| u \|_2^2, \quad (10) \]

for all \( u \in S \).

**Proof.** We have

\[ B_{\text{symm}}(u, u) = \| u \|_2^2 - 2 \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{ \epsilon \nabla u \} \cdot [u] \, ds. \]

Using the Cauchy-Schwarz inequality together with the standard \( hp \)-version inverse inequality

\[ \| v \|_2^2 \leq C_{\text{inv}} p^2 |\epsilon| / |e| \| v \|_r^2, \quad (11) \]

(see, e.g., [18]) we obtain

\[ 2 \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{ \epsilon \nabla u \} \cdot [u] \, ds \leq \tau^{-1} \| \{ \epsilon \nabla u \} \|_{L^2(\Gamma_D \cup \Gamma_{\text{int}})}^2 + \tau \| [u] \|_{L^2(\Gamma_D \cup \Gamma_{\text{int}})}^2 \]

\[ \leq \sum_{\kappa \in T} \tau^{-1} C_{\text{inv}} \frac{p^2 |\epsilon|}{|e|} \| \sqrt{\epsilon} \nabla u \|_{L^2(\Gamma_D \cup \Gamma_{\text{int}})}^2 + \tau \| [u] \|_{L^2(\Gamma_D \cup \Gamma_{\text{int}})}^2. \quad (12) \]

Choosing \( \tau = \frac{1}{2} C_{\text{inv}} p^2 |\epsilon| / |e| \) then, for sufficiently large \( C_{\sigma} \), we conclude that

\[ 2 \int_{\Gamma_D \cup \Gamma_{\text{int}}} \{ \epsilon \nabla u \} \cdot [u] \, ds \leq \frac{1}{2} \sum_{\kappa \in T} \| \sqrt{\epsilon} \nabla u \|_{L^2(\Gamma_D \cup \Gamma_{\text{int}})}^2 + \frac{1}{2} \| [\sigma] \|_{L^2(\Gamma_D \cup \Gamma_{\text{int}})}^2. \]

Thus, the left inequality in (10) holds with \( \kappa_1 = \frac{1}{2} \) and the right inequality holds with \( \kappa_2 = \frac{3}{2} \).

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

**Lemma 4.4** Let the domain \( \Omega \), its subdivision \( T \) and the positive function \( \sigma \) be defined as above, with the constant \( C_{\sigma} \) sufficiently large. Then, assuming that \( w, v \in S \), we have

\[ |B(w, v)| \leq C(\| \mathbf{b} \|_{L^\infty(\Omega)}/\epsilon) \| w \| \| v \|. \]

A proof can be found in [9].
5 Norm-preconditioners

Finite element methods represent one of the most significant sources of large sparse problems for which preconditioning inside a Krylov method becomes mandatory. These matrices, however, belong to a framework that allows the design of useful preconditioners, as well as meaningful monitoring of convergence. 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 [6] in the context of preconditioning for standard finite element discretisation of elliptic problems. The authors of [6] 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. We summarise the relevant results below.

We use the standard definition
\[ \| x \|_H = \langle x, x \rangle_H = x^T H x \]
where \( H \in \mathbb{R}^{n \times n} \) is symmetric and positive-definite and \( x \in \mathbb{R}^n \). We also introduce the following

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 \( \gamma, \Gamma \) independent of \( n \) such that for all \( x \in \mathbb{R}^n \setminus \{0\} \)
\[ \gamma \leq \frac{\langle x, A^{-1} B x \rangle_H}{\langle x, x \rangle_H}, \quad \| A^{-1} B x \|_H \leq \Gamma \| x \|_H. \]
We write
\[ A \approx_H B. \]

We remark here that FOV-equivalence implies that \( \gamma < |\lambda(A^{-1})| < \Gamma \), where \( \lambda(A^{-1}) \) 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 [5]:

Lemma 5.2 : Generalized Minimum Residual (GMRES). 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 [5], [17, Thm 6.7]
\[ \frac{\| r_k \|_H}{\| r_0 \|_H} \leq \left( 1 - \frac{\gamma^2}{\Gamma^2} \right)^{k/2}, \quad (13) \]
where \( \gamma, \Gamma \) 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 4.1 and the continuity result of Lemma 4.4 have the following discrete counterparts
\[ \max_{w \in \mathbb{R}^n \setminus \{0\}} \max_{v \in \mathbb{R}^n \setminus \{0\}} \frac{w^T K v}{\| w \|_H \| v \|_H} \leq C \left( \| b \|_{L^\infty(\Omega)} / \epsilon \right) \]
\[ \min_{w \in \mathbb{R}^n \setminus \{0\}} \frac{w^T K w}{\| w \|_H^2} \geq C \theta \]
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 (10) has the following discrete version

$$\kappa_1 \|u\|_H^2 \leq \|u\|_{K_s}^2 \leq \kappa_2 \|u\|_H^2,$$  \hspace{1cm} (15)

where $K_s = (K + K^T)/2$. The following result taken from [15] is sufficient to ensure that $H$ is $FOV$-equivalent to $K$ and therefore represents a good preconditioning candidate.

**Proposition 5.3** Let (14) hold. Then $K \approx H^{-1} H$.

Given equivalence 15, one deduces that 14 holds also with $H$ replaced by $K_s$ so that we get the following

**Proposition 5.4** Let (14), (15) 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 [4] 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 [15] 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** The residuals of the 3-term GMRES algorithm in the $K_s$-inner product satisfy

$$\frac{\|r^k\|_{K_s^{-1}}}{\|r^0\|_{K_s^{-1}}} \leq \left(1 - \frac{\gamma^2}{\Gamma^2}\right)^{k/2},$$  \hspace{1cm} (16)

where $\gamma = C_\theta/\kappa_1$ and $\Gamma = C(\|b\|_{L^\infty(\Omega)}/\epsilon)/\kappa_2$.

We note here that the bounds are both $h$- and $p$-independent, a result which we expect to see reflected in the convergence behaviour of GMRES. On the other hand, the presence of $\epsilon$ in $\Gamma$ corresponds to deterioration of the bound with decreasing $\epsilon$. As shown below, this leads to deterioration in convergence also.
6 Numerics

For \( b = (1, 1)^T \) and \( 0 < \epsilon \leq 1 \), we consider the singularly perturbed convection-diffusion equation

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

subject to a Dirichlet boundary condition, which, along with the forcing function \( f \), is chosen so that the analytical solution is

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

This problem was considered in [12] (Example 3). The solution exhibits boundary layer behaviour along \( x = 1 \) and \( y = 1 \), and the layers become steeper as \( \epsilon \to 0 \). Note that the theory developed above includes this case, as here \( \nabla \cdot b = 0 \) on \( \Omega \).

We solved the problem for a range of \( \epsilon \). Discretisations for a range of \( h \) (meshsize) and \( p \) (degree of polynomial approximation) were employed. 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). The results are presented Table 1. As predicted by theory, the number of iterations is independent of discretisation parameters. The dependence of \( \epsilon \) of bound 16 gives also a description of convergence behaviour.

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 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 \( \epsilon \), 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.

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

<table>
<thead>
<tr>
<th>( p )</th>
<th>( n )</th>
<th>( \epsilon = 0.5 )</th>
<th>( \epsilon = 0.1 )</th>
<th>( \epsilon = 0.01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,500</td>
<td>7</td>
<td>15</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>7</td>
<td>15</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>40,000</td>
<td>7</td>
<td>14</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>5,625</td>
<td>7</td>
<td>14</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>22,500</td>
<td>6</td>
<td>14</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>90,000</td>
<td>6</td>
<td>14</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>6</td>
<td>14</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>40,000</td>
<td>6</td>
<td>14</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>160,000</td>
<td>6</td>
<td>13</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 1: GMRES iterations for DGhp-discretisation of the convection diffusion problem with preconditioner \( K_s \).
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. Our solver of choice is domain decomposition nested inside a GMRES routine and which exhibits convergence independent of discretisation parameters. For more details see [7].

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 keep storage down to a minimum. The resulting solver was applied to the case of the singularly-perturbed convection-diffusion problem. The performance showed no discretisation dependence and only a dependence on the viscosity parameter. Theoretical results were derived to explain this behaviour.

References


