SIAM J. Numer. Anal.
Vol. 45, No. 4, pp. 1495–1516

MESH INDEPENDENT SUPERLINEAR PCG RATES VIA COMPACT-EQUIVALENT OPERATORS[*]

OWE AXELSSON[†] AND JÁNOS KARÁTSON[‡]

Abstract. The subject of the paper is the mesh independent convergence of the preconditioned conjugate gradient (PCG) method for nonsymmetric elliptic problems. The approach of equivalent operators is involved, in which one uses the discretization of another suitable elliptic operator as preconditioning matrix. By introducing the notion of compact-equivalent operators, it is proved that for a wide class of elliptic problems the superlinear convergence of the obtained PCG method is mesh independent under finite element discretizations; that is, the rate of superlinear convergence is given in the form of a sequence which is mesh independent and is determined only by the elliptic operators.

Key words. nonsymmetric elliptic problems, conjugate gradient method, preconditioning, equivalent operators, superlinear convergence, mesh independence

AMS subject classifications. 65J10, 65F10, 65N15

DOI. 10.1137/06066391X

1. Introduction.

The conjugate gradient (CG) method is a widespread way of solving large linear algebraic systems, such as those arising from discretized elliptic problems, in particular when combined with a suitable preconditioning. For nonsymmetric systems several CG algorithms exist [2, 4], including the common CGN method based on normal equations. Since its first presentation in [21] the convergence of the CG method has been well established, as summarized in [4]. The convergence theory of the CG method often involves linear operators in Hilbert space; see both classical and recent results [14, 15, 20, 27, 31, 32] and the authors' papers [6, 7, 8, 22, 24]. A basic reason to use Hilbert space theory is to derive mesh independence of the convergence estimates, by which it can be shown that the preconditioned CG (PCG) method can be competitive with multigrid methods [14].

The theory of equivalent operators in Hilbert space has proved to provide an efficient clear framework for the convergence study of the PCG method for elliptic problems [14, 18, 26]. Thereby one uses the discretization of a suitable linear elliptic operator as preconditioning matrix; see also [10, 12, 32]. As a main result, mesh independence of linear convergence rates is rigorously characterized in [14, 26]. We note that in [18], for proper boundary conditions, when the preconditioned operator is a compact perturbation of the identity, then convergence is expected to be faster than any linear rate.

Our goal is to complete the above results on the preconditioned CGN (PCGN) method by showing that for a class of elliptic problems, the superlinear convergence of the iteration is mesh independent under finite element method (FEM) discretizations. This means that a bound on the rate of superlinear convergence is given in the form of a sequence which is mesh independent and is determined only by the

[*] Received by the editors June 28, 2006; accepted for publication (in revised form) February 21, 2007; published electronically August 1, 2007.
http://www.siam.org/journals/sinum/45-4/66391.html

[†] Department of Information Technology, Uppsala University, SE-751 05 Uppsala, Sweden, and Institute of Geonics AS CR, 708 00 Ostrava, Czech Republic (owea@it.uu.se).

[‡] Department of Applied Analysis, ELTE University, H-1117 Budapest, Hungary (karatson@cs.elte.hu). This author was supported by NKTH Öveges Program and Hungarian Research grant OTKA T 043765.

elliptic operators. To describe the suitable class of problems, we introduce the notion of compact-equivalent operators, which expresses that preconditioning one with the other yields a compact perturbation of the identity. This notion and the convergence result give a refinement of the case of equivalent operators: roughly speaking, if the two operators (the original and preconditioner) are equivalent, then the corresponding PCG method provides mesh independent linear convergence, whereas if the two operators are compact-equivalent, then the PCG method provides mesh independent superlinear convergence.

Our present results are extensions of the earlier ones [8, 24], where such mesh independence was proved for the generalized conjugate gradient-least squares (GCG-LS) method for elliptic Dirichlet problems, but with severe restrictions: except for some special cases, both the original and preconditioning operators had to contain constant coefficients. Now we show that two elliptic operators, with homogeneous Dirichlet conditions on the same portion of the boundary, are compact-equivalent if and only if their principal parts coincide up to a constant factor. Within this class, the proof of the mesh independence result then contains no restrictions except standard smoothness and coercivity assumptions on the operators.

Our characterization of compact-equivalence provides, in fact, a limitation on the scope of the mesh independent superlinear convergence property. Since the principal parts of compact-equivalent operators must coincide (up to a constant), preconditioning methods like replacing rough diffusion coefficients by simpler, e.g., constant, ones are not covered by this setting except, of course, the case when the variable coefficient problem can be easily rewritten by suitable scaling to a constant coefficient problem, as for scalar coefficients. In fact, one cannot expect superlinear convergence for such non–compact-equivalent operators since, as shown in [15], convergence of the CG method may be only linear if an operator is not a compact perturbation of a constant times the identity.

The paper is organized as follows: the required background is given in section 2, compact-equivalent operators are introduced and characterized in section 3, and the mesh independence result is proved in section 4. Some closing remarks are found in section 5.

2. Background.

2.1. Conjugate gradient algorithms.

Let us consider a linear system

$$Bu = f \tag{1}$$

with a given nonsingular matrix $B \in \mathbf{R}^{n\times n}$, $f \in \mathbf{R}^n$ and solution u. Let $\langle .,. \rangle$ be a given inner product on $\mathbf{R}^n$ and, denoting by B^* the adjoint of B w.r.t. this inner product, assume that $B + B^* > 0$, i.e., is positive definite.

If B is self-adjoint, then the standard CG method reads as follows [4, 31]: let $u_0 \in \mathbf{R}^n$ be arbitrary, $d_0 := Bu_0 - f$; for given u_k and d_k, with $\hat{r}_k := Bu_k - f$, we let

$$u_{k+1} = u_k - \alpha_k d_k, \text{ where } \alpha_k = \frac{\langle \hat{r}_k, d_k\rangle}{\langle Bd_k, d_k\rangle}\,; \quad d_{k+1} = \hat{r}_{k+1} + \beta_k d_k, \text{ where } \beta_k = \frac{\|\hat{r}_{k+1}\|^2}{\|\hat{r}_k\|^2}. \tag{2}$$

Then, using the error vector $e_k = u_k - u$ and its energy norm $\|e_k\|_B = \langle Be_k, e_k\rangle^{1/2}$, respectively, and with the decomposition $B = I + C$ (where I is the identity matrix), the following celebrated estimate holds [4, 31]:

$$\left(\frac{\|e_k\|_B}{\|e_0\|_B}\right)^{1/k} \le \frac{2}{k}\,\|B^{-1}\|\sum_{j=1}^{k}\bigl|\lambda_j(C)\bigr| \qquad (k=1,2,\dots,n), \tag{3}$$

which shows superlinear convergence if the eigenvalues $|\lambda_1(C)| \ge |\lambda_2(C)| \ge \cdots$ approach zero.

Since this result is basic for the whole paper, and for completeness, we present a derivation of (3) following [4]. The optimality of the CG method implies

$$\frac{\|e_k\|_B}{\|e_0\|_B} \le \min_{P_k\in\pi_k^1}\ \max_{\lambda\in\sigma(B)\}} |P_k(\lambda)|,$$

where π_k^1 denotes the set of polynomials P_k of degree k with $P_k(0)=1$. Let $\lambda_j := \lambda_j(B)$ and $\mu_j := \lambda_j(C)\,(=\lambda_j-1)$. Then the polynomials $P_k(\lambda) := \prod_{j=1}^{k}(1-\frac{\lambda}{\lambda_j})$ satisfy $P_k(\lambda_i)=0$ $(i=1,\dots,k)$ and

$$\max_{\lambda\in\sigma(B)\}} |P_k(\lambda)| = \max_{i\ge k+1}\prod_{j=1}^{k}\left|1-\frac{\lambda_i}{\lambda_j}\right| = \max_{i\ge k+1}\prod_{j=1}^{k}\frac{|\mu_j-\mu_i|}{|1+\mu_j|} \le \max_{i\ge k+1}\prod_{j=1}^{k}\frac{2|\mu_j|}{|1+\mu_j|}\ ;$$

hence, using the arithmetic-geometric inequality,

$$\max_{\lambda\in\sigma(B)\}} |P_k(\lambda)|^{1/k} \le \frac{2}{k}\sum_{j=1}^{k}\frac{|\mu_j|}{|1+\mu_j|} \le \frac{2}{k}\left(\sup\frac{1}{|\lambda_j|}\right)\sum_{j=1}^{k}|\mu_j|,$$

which yields (3).

For nonsymmetric B, several CG algorithms exist (see, e.g., [2, 4, 13]). The GCG-LS method [3, 4] is defined directly for (1) and produces an estimate similar to that of (3) if B is normal. Mesh independent bounds in [8, 24] for (3) for some elliptic problems have been given using the GCG-LS method. Alternatively, one can consider the normal equation and apply a symmetric CG algorithm, which we will do in this paper. For clearness, let us hereby consider a nonsymmetric linear system

$$Au = b \tag{4}$$

with a given nonsingular matrix $A\in\mathbf{R}^{n\times n}$ and vector $b\in\mathbf{R}^n$. Let us apply the iteration (2) for the equation $A^*Au=A^*b$, i.e., with $B=A^*A$ and $f=A^*b$. Then, with notations $s_k=\hat r_k$ and $r_k=A^{-*}\hat r_k$, we obtain the following algorithmic form, often called the CGN method: let $u_0\in\mathbf{R}^n$ be arbitrary, $r_0 := Au_0-b$, $s_0 := d_0 := A^*r_0$; for given d_k, u_k, r_k, and s_k, we let

$$\left\{\begin{aligned}
&z_k = Ad_k,\\
&\alpha_k = \frac{\langle r_k, z_k\rangle}{\|z_k\|^2}\ ,\quad u_{k+1}=u_k-\alpha_k d_k\ ,\quad r_{k+1}=r_k-\alpha_k z_k\ ;\\
&s_{k+1}=A^*r_{k+1},\\
&\beta_k=\frac{\|s_{k+1}\|^2}{\|s_k\|^2}\ ,\quad d_{k+1}=s_{k+1}+\beta_k d_k.
\end{aligned}\right. \tag{5}$$

Let us consider the decomposition

$$A = I + K.$$

Then, using the relations $B = I + (K^* + K + K^*K)$, $\|e_k\|_B = \|Ae_k\| = \|r_k\|$, and $\|B^{-1}\| \le \nu^{-1}$, where $\nu := \min_{x \in \mathbf{R}^n} \frac{\|Ax\|^2}{\|x\|^2}$, estimate (3) can be reformulated as

$$(6) \qquad \left(\frac{\|r_k\|}{\|r_0\|}\right)^{1/k} \le \frac{2}{k\nu} \sum_{i=1}^{k} \Big(\big|\lambda_i(K^* + K)\big| + \lambda_i(K^*K) \Big) \qquad (k = 1, 2, \ldots, n).$$

The goal of this paper is to derive a mesh independent bound for (6) when (4) comes from a preconditioned discretized elliptic PDE using suitable equivalent operators.

2.2. Singular values of compact operators. Let H be a real Hilbert space. We shall consider compact operators, i.e., operators C such that the image (Cv_i) of any bounded sequence (v_i) contains a convergent subsequence.

Definition 2.1. (i) We call $\lambda_i(F)$ $(i = 1, 2, \ldots)$ the *ordered eigenvalues* of a compact self-adjoint linear operator F in H if each of them is repeated as many times as its multiplicity and $|\lambda_1(F)| \ge |\lambda_2(F)| \ge \cdots$.

(ii) The *singular values* of a compact operator C in H are

$$s_i(C) := \lambda_i(C^*C)^{1/2} \qquad (i = 1, 2, \ldots),$$

where $\lambda_i(C^*C)$ are the ordered eigenvalues of C^*C. In particular, if C is self-adjoint, then $s_i(C) = |\lambda_i(C)|$.

Some useful properties of compact operators are listed below.

Proposition 2.2. *Let C be a compact operator in H. Then the following properties hold.*

(a) *For any $k \in \mathbf{N}^+$ and any orthonormal vectors $u_1, \ldots, u_k \in H$,*

$$\sum_{i=1}^{k} \big|\langle Cu_i, u_i\rangle\big| \le \sum_{i=1}^{k} s_i(C)\,.$$

(b) *If B is a bounded linear operator in H, then*

$$s_i(BC) \le \|B\|\, s_i(C) \qquad (i = 1, 2, \ldots).$$

(c) *(Variational characterization of the eigenvalues.) If C is also self-adjoint, then*

$$\big|\lambda_i(C)\big| = \min_{H_{i-1} \subset H} \; \max_{\substack{u \perp H_{i-1} \\ u \ne 0}} \frac{\big|\langle Cu, u\rangle\big|}{\|u\|^2}\,,$$

where H_{i-1} stands for an arbitrary $(i-1)$-dimensional subspace.

(d) *If a sequence $(u_i) \subset H$ satisfies $\langle u_i, u_j\rangle = \langle Cu_i, u_j\rangle = 0$ $(i \ne j)$, then*

$$\inf_i \, |\langle Cu_i, u_i\rangle| / \|u_i\|^2 = 0.$$

Proof. Statements (a) and (b) are the consequences of [16, Chap. VI, Corollary 3.3 and Proposition 1.3, resp.]; for statement (c), see [17, Theorem III.9.1]. To prove (d), assume the contrary that the infimum equals $\delta > 0$. We may assume that $\langle Cu_i, u_i\rangle$ has constant sign (otherwise we can consider a subsequence that has constant sign). Then the orthonormal sequence $v_i := u_i/\|u_i\|$ satisfies for all $i \neq j$

$$\begin{aligned} 2\delta &\le |\langle Cv_i, v_i\rangle + \langle Cv_j, v_j\rangle| = |\langle C(v_i - v_j), v_i - v_j\rangle| \\ &\le \|C(v_i - v_j)\|\,\|v_i - v_j\| = \sqrt{2}\|C(v_i - v_j)\|; \end{aligned}$$

hence the image (Cv_i) of the bounded sequence (v_i) contains no convergent subsequence (i.e., C is not compact). □

3. Compact-equivalent operators in Hilbert space. In this section we introduce and characterize compact-equivalent operators. Roughly speaking, the compact-equivalence of the unbounded operators N and L expresses that $N^{-1}L$ is a compact perturbation of a constant times the identity. To avoid difficulties with domains and ranges, our definition will use a weak form of the operators in a suitable energy space H_S. In particular, no regularity is required in the case of elliptic operators.

The fact that a compact perturbation of a constant times identity is a bounded operator implies that compact-equivalent operators are equivalent in the sense of [14]. Hence, when we characterize compact-equivalent elliptic operators (under standard smoothness and coercivity assumptions), we can a priori assume that they have homogeneous Dirichlet conditions on the same portion of the boundary [26]. Within this class, compact-equivalence will hold if and only if the principal parts of the operators coincide up to some constant.

3.1. Basic definitions. In what follows, let H be a real Hilbert space. Let S be a (generally unbounded) linear symmetric operator in H which is coercive; i.e., there exists $p > 0$ such that $\langle Su, u\rangle \ge p\|u\|^2$ $(u \in D(S))$. Then the energy space H_S is the completion of $D(S)$ under the inner product $\langle u, v\rangle_S = \langle Su, v\rangle$, and the coercivity implies $H_S \subset H$. The corresponding S-norm is denoted by $\|u\|_S$, and the space of bounded linear operators on H_S by $B(H_S)$.

DEFINITION 3.1. *Let S be a linear symmetric coercive operator in H. We say that a linear operator L in H is S-bounded and S-coercive, and write $L \in BC_S(H)$ if the following properties hold:*

(i) *$D(L) \subset H_S$ and $D(L)$ is dense in H_S in the S-norm;*

(ii) *there exists $M > 0$ such that*

$$|\langle Lu, v\rangle| \le M\|u\|_S\|v\|_S \qquad (u, v \in D(L));$$

(iii) *there exists $m > 0$ such that*

$$\langle Lu, u\rangle \ge m\|u\|_S^2 \qquad (u \in D(L)).$$

DEFINITION 3.2. *For any $L \in BC_S(H)$, let $L_S \in B(H_S)$ be defined by*

$$\langle L_S u, v\rangle_S = \langle Lu, v\rangle \qquad (u, v \in D(L)).$$

Remark 1.

(a) The above definition makes sense since L_S is the bounded linear operator on H_S that represents the unique extension to H_S of the densely defined S-bounded bilinear form $u, v \mapsto \langle Lu, v\rangle$.

(b) L_S is coercive on H_S.

(c) If in particular $R(L) \subset R(S)$ (where $R(.)$ denotes the range), then $L_S\big|_{D(L)} = S^{-1}L$.

Remark 2. Definition 3.2 uses the idea of weak form of operators from [26]. Namely, if H_S is a subspace of $H^1(\Omega)$ consisting of functions vanishing on a fixed portion of the boundary, then L_S coincides with the weak operator L_w using (2.15) in [26].

Now let us consider an operator equation

$$Lu = g, \tag{7}$$

where $L \in BC_S(H)$ and $g \in H$.

DEFINITION 3.3. *We call $u \in H_S$ the weak solution of equation* (7) *if*

$$\langle L_S u, v\rangle_S = \langle g, v\rangle \qquad (v \in H_S). \tag{8}$$

Remark 3.

(a) For all $g \in H$ the weak solution of (7) exists and is unique. This follows in the usual way from the Lax–Milgram theorem, since $v \mapsto \langle g, v\rangle$ is a bounded linear functional on H_S by the coercivity of S.

(b) If $u \in D(L)$, then u satisfies (7) (i.e., it is a strong solution) if and only if u is a weak solution.

3.2. Compact-equivalent operators.

We can introduce the notion of compact-equivalence within the previously described setting as follows.

DEFINITION 3.4. *Let L and N be S-bounded and S-coercive operators in H. We call L and N compact-equivalent in H_S if*

$$L_S = \mu N_S + Q_S \tag{9}$$

for some constant $\mu > 0$ and compact operator $Q_S \in B(H_S)$.

Remark 4. (i) It follows in a straightforward way that the property compact-equivalence is an equivalence relation.

(ii) In the special case $R(L) \subset R(N)$, compact-equivalence of L and N means that $N^{-1}L$ is a compact perturbation of a constant times the identity in the space H_S. Indeed, it is easy to see that here $N^{-1}L = N_S^{-1}L_S\big|_{D(L)}$, and by definition the latter is the perturbation of μI with the operator $N_S^{-1}Q_S\big|_{D(L)}$, which is compact since N_S^{-1} is bounded. (In the general case the "weakly preconditioned" form $N_S^{-1}L_S$ is also a compact perturbation.)

Now we characterize compact-equivalence for elliptic operators. Let $H = L^2(\Omega)$ and let us define the operators

$$N_1 u \equiv -\mathrm{div}\,(A_1\,\nabla u) + \mathbf{b}_1\cdot\nabla u + c_1 u \qquad \text{for} \quad u_{|\Gamma_D} = 0,\ \tfrac{\partial u}{\partial \nu_{A_1}} + \alpha_1 u_{|\Gamma_N} = 0,$$
$$N_2 u \equiv -\mathrm{div}\,(A_2\,\nabla u) + \mathbf{b}_2\cdot\nabla u + c_2 u \qquad \text{for} \quad u_{|\Gamma_D} = 0,\ \tfrac{\partial u}{\partial \nu_{A_2}} + \alpha_2 u_{|\Gamma_N} = 0,$$

where $\frac{\partial u}{\partial \nu_{A_i}} = A_i\,\nu\cdot\nabla u$ denotes the weighted normal derivative. (The formal domain of N_i to be used in Definition 3.2 consists of those $u \in H^2(\Omega)$ that satisfy the above boundary conditions; however, this is used nowhere else.) The following properties hold, where $i = 1, 2$.

ASSUMPTIONS 3.2.

(i) $\Omega \subset \mathbf{R}^d$ is a bounded piecewise C^1 domain; Γ_D, Γ_N are disjoint open measurable subparts of $\partial\Omega$ such that $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$.

(ii) $A_i \in C^1(\overline{\Omega}, \mathbf{R}^{d\times d})$ and for all $x \in \overline{\Omega}$ the matrix $A_i(x)$ is symmetric; $\mathbf{b}_i \in C^1(\overline{\Omega})^d$, $c_i \in L^\infty(\Omega)$, $\alpha_i \in L^\infty(\Gamma_N)$.

(iii) We have the coercivity properties $\min_{\lambda\in\sigma(A_i(x))} \lambda \ge p > 0$ with p independent of x, $\hat{c}_i := c_i - \frac{1}{2}\operatorname{div}\mathbf{b}_i \ge 0$ in Ω and $\hat{\alpha}_i := \alpha_i + \frac{1}{2}(\mathbf{b}_i \cdot \nu) \ge 0$ on Γ_N.

(iv) Either $\Gamma_D \ne \emptyset$, or $\hat{c}_i$ or $\hat{\alpha}_i$ has a positive lower bound.

For the study of such operators we define the space

$$H^1_D(\Omega) := \{u \in H^1(\Omega) : \, u_{|\Gamma_D} = 0\} \quad \text{with } \langle u, v\rangle_S := \int_\Omega (G\,\nabla u \cdot \nabla v + huv) + \int_{\Gamma_N} \beta uv\, d\sigma, \tag{10}$$

where G has the same properties as A_i above in (ii)–(iii), and $h \in L^\infty(\Omega)$, $h \ge 0$, if $\Gamma_D \ne \emptyset$ and $h \ge \delta_0 > 0$ if $\Gamma_D = \emptyset$, and further, $\beta \in L^\infty(\Gamma_N)$ and $\beta \ge 0$. Then $H^1_D(\Omega)$ is the energy space H_S of the operator $Su := -\operatorname{div}(G\,\nabla u) + hu$ on $D(S) := \{u \in H^2(\Omega) : \, u_{|\Gamma_D} = 0, \, \frac{\partial u}{\partial \nu_{G+\beta u}}_{|\Gamma_N} = 0\}$. It is easy to check the properties in Definition 3.1 from the above assumptions, which means that $N_1, N_2 \in BC_S(L^2(\Omega))$.

PROPOSITION 3.5. *The elliptic operators N_1 and N_2 are compact-equivalent in $H^1_D(\Omega)$ if and only if their principal parts coincide up to some constant $\mu > 0$, i.e., $A_1 = \mu A_2$.*

Proof. We have for all $u, v \in H^1_D(\Omega)$

$$\langle (N_i)_S u, v\rangle_S = \int_\Omega \Big(A_i\, \nabla u \cdot \nabla v + (\mathbf{b}_i \cdot \nabla u)v + c_i uv \Big)\, dx + \int_{\Gamma_N} \alpha_i uv\, d\sigma\,.$$

Hence

$$(N_1)_S - \mu (N_2)_S = J_S + Q_S,$$

where, using notations $\mathbf{b} := \mathbf{b}_1 - \mu\mathbf{b}_2$, $c := c_1 - \mu c_2$, and $\alpha := \alpha_1 - \mu\alpha_2$, we have

$$\begin{aligned} \langle J_S u, v\rangle_S &= \int_\Omega (A_1 - \mu A_2)\; \nabla u \cdot \nabla v\, dx\,, \\ \langle Q_S u, v\rangle_S &= \int_\Omega \Big((\mathbf{b} \cdot \nabla u)v + cuv \Big)\, dx + \int_{\Gamma_N} \alpha uv\, d\sigma\,. \end{aligned} \tag{11}$$

Here Q_S is compact, which is known [18] when N_1 and N_2 have the same boundary conditions. Otherwise we use the equality

$$\begin{aligned} \int_\Omega (\mathbf{b} \cdot \nabla u) v\, dx &= -\int_\Omega u(\mathbf{b} \cdot \nabla v)\, dx - \int_\Omega (\operatorname{div}\mathbf{b}) uv\, dx \\ &\quad + \int_{\Gamma_N} (\mathbf{b} \cdot \nu)\, uv\, d\sigma \quad (u, v \in H^1_D(\Omega)) \end{aligned} \tag{12}$$

whence, using notations $\tilde{c} := c - \operatorname{div}\mathbf{b}$ and $\tilde{\alpha} := \alpha + \mathbf{b}\cdot\nu$,

$$\|Q_S u\|_S = \sup_{\substack{v\in H^1_D(\Omega)\\ \|v\|_S=1}} |\langle Q_S u, v\rangle_S| = \sup_{\substack{v\in H^1_D(\Omega)\\ \|v\|_S=1}} \left| -\int_\Omega u(\mathbf{b}\cdot\nabla v)\, dx + \int_\Omega \tilde{c} uv\, dx + \int_{\Gamma_N} \tilde{\alpha}\, uv\, d\sigma \right|.$$

Using the embedding estimates

$$\|v\|_{L^2(\Omega)} \le C_\Omega \|v\|_S\,, \qquad \|v\|_{L^2(\Gamma_N)} \le C_{\Gamma_N} \|v\|_S \qquad (v \in H^1_D(\Omega)) \tag{13}$$

(where $C_\Omega, C_{\Gamma_N} > 0$) and $\|\nabla v\|_{L^2(\Omega)} \le p^{-1/2}\|v\|_S$, and letting $K_1 := p^{-1/2}\|\mathbf{b}\|_{L^\infty(\Omega)} + C_\Omega \|\tilde{c}\|_{L^\infty(\Omega)}$, $K_2 := C_{\Gamma_N}\|\tilde{\alpha}\|_{L^\infty(\Gamma_N)}$, we obtain

$$\|Q_S u\|_S \le K_1 \|u\|_{L^2(\Omega)} + K_2 \|u\|_{L^2(\Gamma_N)} \tag{14}$$

whence Q_S is compact.

It remains to prove that if $A_1 \ne \mu A_2$, then J_S is not compact. Using Proposition 2.2(d), it suffices to find a sequence $(u_i) \subset H^1_0(\Omega) \subset H^1_D(\Omega)$ satisfying

$$\langle u_i, u_j \rangle_S = \langle J_S u_i, u_j \rangle_S = 0 \qquad (i \ne j), \tag{15}$$

$$\inf_i |\langle J_S u_i, u_i \rangle_S| / \|u_i\|^2_S = \delta > 0. \tag{16}$$

Let $A := A_1 - \mu A_2$. Since A is not identically zero, there is $x_0 \in \Omega$ such that $A_0 := A(x_0) \ne 0$. Here A_0 is symmetric; hence there is $u_0 \in H^1_0(\Omega)$ such that $\int_\Omega A_0\ \nabla u_0 \cdot \nabla u_0 \ne 0$. Let

$$\varepsilon := \left| \int_\Omega A_0\ \nabla u_0 \cdot \nabla u_0 \right| / \left(\int_\Omega |\nabla u_0|^2 \right), \qquad \Omega_{\varepsilon/2} := \{x \in \Omega :\ \|A(x) - A_0\| < \varepsilon/2\},$$

which is an open set since A is continuous. Fix $z' \in \Omega$, and for any $z \in \Omega$ and $R > 0$ let $\Omega_{z,R} := \{x \in \mathbf{R}^d :\ z' + R(x-z) \in \Omega\}$. Let $z_i \in \Omega$, $R_i > 0$ ($i \in \mathbf{N}^+$) such that $\Omega_i := \Omega_{z_i,R_i} \subset \Omega_{\varepsilon/2}$ and $\overline{\Omega}_i$ are pairwise disjoint sets. We define $u_i \in H^1_0(\Omega)$ by $u_i(x) := u_0\big(z' + R_i(x - z_i)\big)$ for $x \in \Omega_i$ and $u_i(x) := 0$ for $x \in \Omega \setminus \Omega_i$. Since $\operatorname{supp} u_i = \overline{\Omega}_i$ are disjoint, (15) is satisfied. Further, using the fact $\Omega_i \subset \Omega_{\varepsilon/2}$ and a linear transformation $\Omega_i \to \Omega$ in the integral, we obtain

$$\begin{aligned} \frac{|\langle J_S u_i, u_i \rangle_S|}{\int_{\Omega_i} |\nabla u_i|^2} &= \frac{\left|\int_{\Omega_i} A\ \nabla u_i \cdot \nabla u_i\right|}{\int_{\Omega_i} |\nabla u_i|^2} \ge \frac{\left|\int_{\Omega_i} A_0\ \nabla u_i \cdot \nabla u_i\right|}{\int_{\Omega_i} |\nabla u_i|^2} - \frac{\varepsilon}{2} \\ &= \frac{\left|\int_{\Omega} A_0\ \nabla u_0 \cdot \nabla u_0\right|}{\int_{\Omega} |\nabla u_0|^2} - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}. \end{aligned}$$

Since for $u \in H^1_0(\Omega)$ we have $\|u\|^2_S \le C \cdot \int_\Omega |\nabla u|^2$, the above estimate yields (16) with $\delta = \frac{\varepsilon}{2C} > 0$. □

4. Compact-equivalent preconditioning and mesh independent superlinear convergence rates.

We prove the mesh independent convergence results for the PCG method in four stages. First we consider symmetric preconditioning operators, which are more straightforward to handle. Then, by suitable modifications of the proof, we turn to arbitrary preconditioning operators (in the studied coercive framework) where the general result is obtained. In both the symmetric and nonsymmetric cases we first consider an abstract Hilbert space level and then derive the corresponding estimates for elliptic problems.

For simplicity we will consider compact-equivalence with $\mu = 1$ in (9), which is clearly no restriction, since if a preconditioner N_S satisfies $L_S = \mu N_S + Q_S$, then we can consider the preconditioner μN_S instead.

4.1. The abstract operator equation and its discretization. Let us consider the operator equation

$$Lu = g, \tag{17}$$

where $L \in BC_S(H)$ and $g \in H$, and let $u \in H_S$ be the weak solution as in Definition 3.3. Equation (17) will be solved numerically using a Galerkin discretization: let

$$V_h = span\{\varphi_1, \dots, \varphi_n\} \subset H_S\,,$$

where φ_i are linearly independent, be a given finite-dimensional subspace and let

$$\mathbf{L}_h := \Big\{ \langle L_S \varphi_i, \varphi_j \rangle_S \Big\}_{i,j=1}^n .$$

Finding the discrete solution $u_h \in V_h$ requires solving the $n \times n$ system

$$\mathbf{L}_h\, \mathbf{c} = \mathbf{b} \tag{18}$$

with $\mathbf{b} = \{\langle g, \varphi_j \rangle\}_{j=1}^n$. Since $L \in BC_S(H)$, the symmetric part of $\mathbf{L}_h$ is positive definite; hence system (18) has a unique solution. Moreover, if a sequence of such subspaces V_h satisfies $\inf_{v \in V_h} \|u - v\|_S \to 0$ for all $u \in H_S$, then the coercivity of L_S implies in the standard way [9] that u_h converges to the exact weak solution in the H_S-norm.

4.2. Symmetric preconditioning in Hilbert space. We introduce the stiffness matrix of S,

$$\mathbf{S}_h = \Big\{ \langle \varphi_i, \varphi_j \rangle_S \Big\}_{i,j=1}^n, \tag{19}$$

as preconditioner for system (18), and wish to solve

$$\mathbf{S}_h^{-1} \mathbf{L}_h\, \mathbf{c} = \tilde{\mathbf{b}} \tag{20}$$

(with $\tilde{\mathbf{b}} = \mathbf{S}_h^{-1}\mathbf{b}$) using the CG method . Let us endow $\mathbf{R}^n$ with the $\mathbf{S}_h$-inner product $\langle \mathbf{c}, \mathbf{d} \rangle_{\mathbf{S}_h} := \mathbf{S}_h\, \mathbf{c} \cdot \mathbf{d}$. Then the $\mathbf{S}_h$-adjoint of $\mathbf{S}_h^{-1}\mathbf{L}_h$ is $\mathbf{S}_h^{-1}\mathbf{L}_h^T$; hence we apply the CG algorithm (5) with $A = \mathbf{S}_h^{-1}\mathbf{L}_h$ and $A^* = \mathbf{S}_h^{-1}\mathbf{L}_h^T$.

Let us now assume that L and S are compact-equivalent with $\mu = 1$. In this special case (9) holds with $N_S = I$:

$$L_S = I + Q_S\,. \tag{21}$$

Hence, letting

$$\mathbf{Q}_h = \Big\{ \langle Q_S \varphi_i, \varphi_j \rangle_S \Big\}_{i,j=1}^n ,$$

system (20) takes the form

$$(\mathbf{I}_h + \mathbf{S}_h^{-1}\mathbf{Q}_h)\, \mathbf{c} = \tilde{\mathbf{b}}, \tag{22}$$

where $\mathbf{I}_h$ is the $n \times n$ identity matrix. Using (6), the CG algorithm (5) thus provides the estimate

$$\left(\frac{\|r_k\|_{\mathbf{S}_h}}{\|r_0\|_{\mathbf{S}_h}} \right)^{1/k} \le \frac{2}{k\nu_h} \sum_{i=1}^k \Big(\big|\lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T + \mathbf{S}_h^{-1}\mathbf{Q}_h)\big| + \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T\mathbf{S}_h^{-1}\mathbf{Q}_h) \Big) \tag{23}$$

$(k = 1, 2, \dots, n)$, where

$$\nu_h = \min_{\substack{\mathbf{c}\in\mathbf{R}^n\\ \mathbf{c}\neq\mathbf{0}}} \frac{\|\mathbf{S}_h^{-1}\mathbf{L}_h\mathbf{c}\|^2_{\mathbf{S}_h}}{\|\mathbf{c}\|^2_{\mathbf{S}_h}}\,. \tag{24}$$

Our goal is to give a bound on (23) that is independent of the subspace V_h.

PROPOSITION 4.1. *Let L be S-bounded and S-coercive. Let $\mathbf{S}_h$, $\mathbf{Q}_h$ be defined as above and let $s_i(Q_S)$ and $\lambda_i(Q_S^*+Q_S)$ $(i = 1, 2, \dots)$ denote the corresponding singular values, respectively, ordered eigenvalues where Q_S, defined in* (21), *is compact on H_S. Then the following relations hold:*

$$\text{(a)}\qquad \sum_{i=1}^{k} \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T\,\mathbf{S}_h^{-1}\mathbf{Q}_h) \le \sum_{i=1}^{k} s_i(Q_S)^2 \qquad (k = 1, \dots, n),$$

$$\text{(b)}\qquad \sum_{i=1}^{k}\big|\lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T + \mathbf{S}_h^{-1}\mathbf{Q}_h)\big| \le \sum_{i=1}^{k}\big|\lambda_i(Q_S^* + Q_S)\big| \qquad (k = 1, \dots, n),$$

$$\text{(c)}\qquad \nu_h \ge m^2 \quad \textit{for } \nu_h \textit{ in } (24), \quad \textit{where} \quad m := \inf_{\substack{u\in D(L)\\ u\neq 0}} \frac{\langle Lu, u\rangle}{\|u\|_S^2}\,.$$

Proof. (a) Let $\lambda_i := \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T\,\mathbf{S}_h^{-1}\mathbf{Q}_h)$ $(i = 1, \dots, n)$ and let $\mathbf{c}^i = (c_1^i, \dots, c_n^i) \in \mathbf{R}^n$ be corresponding eigenvectors such that

$$\mathbf{S}_h\,\mathbf{c}^i\cdot\mathbf{c}^l = \delta_{il} \qquad (i, l = 1, \dots, n), \tag{25}$$

where $\cdot$ denotes the ordinary inner product on $\mathbf{R}^n$. Then

$$\mathbf{S}_h^{-1}\mathbf{Q}_h\,\mathbf{c}^i\cdot\mathbf{Q}_h\,\mathbf{c}^i = \lambda_i \qquad (i = 1, \dots, n). \tag{26}$$

Let $\mathbf{d}^i := \mathbf{S}_h^{-1}\mathbf{Q}_h\,\mathbf{c}^i$ for all i; that is,

$$\mathbf{S}_h\,\mathbf{d}^i = \mathbf{Q}_h\,\mathbf{c}^i, \tag{27}$$

which turns (26) into

$$\mathbf{S}_h\mathbf{d}^i\cdot\mathbf{d}^i = \lambda_i\,. \tag{28}$$

Now let $u_i = \sum_{j=1}^n c_j^i\varphi_j \in V_h$ and $z_i = \sum_{j=1}^n d_j^i\varphi_j \in V_h$ $(i = 1, \dots, n)$. Then (28) yields

$$\|z_i\|_S^2 = \lambda_i\,. \tag{29}$$

Further, for all $v = \sum_{j=1}^n p_j\varphi_j \in V_h$, with notation $\mathbf{p} = (p_1, \dots, p_n) \in \mathbf{R}^n$, (27) yields $\mathbf{S}_h\,\mathbf{d}^i\cdot\mathbf{p} = \mathbf{Q}_h\,\mathbf{c}^i\cdot\mathbf{p}$, which implies

$$\langle z_i, v\rangle_S = \langle Q_S u_i, v\rangle_S \qquad (v \in V_h);$$

i.e., z_i is the orthogonal projection of $Q_S u_i \in H_S$ into V_h. Therefore $\|z_i\|_S \le \|Q_S u_i\|_S$, and (29) provides

$$\sum_{i=1}^{k}\lambda_i \le \sum_{i=1}^{k}\|Q_S u_i\|_S^2 = \sum_{i=1}^{k}\langle Q_S^* Q_S u_i, u_i\rangle_S. \tag{30}$$

Here $\langle u_i, u_l\rangle_S = \mathbf{S}_h\,\mathbf{c}^i\cdot\mathbf{c}^l$ for all $i,l = 1,\dots,n$; hence by (25) the vectors u_i are orthonormal in H_S. Therefore Proposition 2.2(a) for the operator $C = Q_S^* Q_S$ in the space H_S yields the desired estimate.

(b) The proof is similar to that of (a). Now let $\lambda_i := \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T + \mathbf{S}_h^{-1}\mathbf{Q}_h)$ and let $\mathbf{c}^i = (c_1^i,\dots,c_n^i)\in\mathbf{R}^n$ be corresponding eigenvectors with property (25). Then

$$(\mathbf{Q}_h^T + \mathbf{Q}_h)\,\mathbf{c}^i = \lambda_i\,\mathbf{S}_h\,\mathbf{c}^i \qquad (i = 1,\dots,n),$$

and (25) yields

$$\lambda_i \;=\; (\mathbf{Q}_h^T + \mathbf{Q}_h)\,\mathbf{c}^i\cdot\mathbf{c}^i = 2\,\mathbf{Q}_h\,\mathbf{c}^i\cdot\mathbf{c}^i.$$

For $u_i = \sum_{j=1}^n c_j^i\varphi_j \in V_h$ we thus obtain

$$\sum_{i=1}^k \big|\lambda_i\big| \;= 2\sum_{i=1}^k \big|\langle Q_S u_i, u_i\rangle_S\big| = \sum_{i=1}^k \big|\langle (Q_S^* + Q_S)u_i, u_i\rangle_S\big|\,, \tag{31}$$

and Proposition 2.2(a) for the operator $C = Q_S^* + Q_S$ in the space H_S yields the desired estimate.

(c) We have

$$\min_{\substack{\mathbf{c}\in\mathbf{R}^n\\ \mathbf{c}\neq\mathbf{0}}} \frac{\|\mathbf{S}_h^{-1}\mathbf{L}_h\mathbf{c}\|_{\mathbf{S}_h}}{\|\mathbf{c}\|_{\mathbf{S}_h}} = \min_{\substack{\mathbf{c}\in\mathbf{R}^n\\ \mathbf{c}\neq\mathbf{0}}} \frac{\|\mathbf{S}_h^{-1}\mathbf{L}_h\mathbf{c}\|_{\mathbf{S}_h}\|\mathbf{c}\|_{\mathbf{S}_h}}{\|\mathbf{c}\|^2_{\mathbf{S}_h}} \ge \min_{\substack{\mathbf{c}\in\mathbf{R}^n\\ \mathbf{c}\neq\mathbf{0}}} \frac{\langle\mathbf{S}_h^{-1}\mathbf{L}_h\mathbf{c},\mathbf{c}\rangle_{\mathbf{S}_h}}{\|\mathbf{c}\|^2_{\mathbf{S}_h}}$$

$$= \min_{\substack{\mathbf{c}\in\mathbf{R}^n\\ \mathbf{c}\neq\mathbf{0}}} \frac{\mathbf{L}_h\,\mathbf{c}\cdot\mathbf{c}}{\mathbf{S}_h\,\mathbf{c}\cdot\mathbf{c}} = \min_{\substack{u\in V_h\\ u\neq 0}} \frac{\langle L_S u, u\rangle_S}{\|u\|_S^2} \ge \inf_{\substack{u\in H_S\\ u\neq 0}} \frac{\langle L_S u, u\rangle_S}{\|u\|_S^2}$$

$$= \inf_{\substack{u\in D(L)\\ u\neq 0}} \frac{\langle L_S u, u\rangle_S}{\|u\|_S^2} = \inf_{\substack{u\in D(L)\\ u\neq 0}} \frac{\langle L u, u\rangle}{\|u\|_S^2} = m,$$

where the density of $D(L)$ in H_S has been used. □

In virtue of (23) and Proposition 4.1, we have proved the following theorem.

THEOREM 4.2. *Let L be S-bounded and S-coercive, and let L and S be compact-equivalent with $\mu = 1$. Let the compact operator Q_S be as in* (21). *Then for any subspace $V_h = span\{\varphi_1,\dots,\varphi_n\}\subset H_S$, the CG algorithm* (5) *with $\mathbf{S}_h$-inner product, applied for the $n\times n$ preconditioned system* (20), *yields*

$$\left(\frac{\|r_k\|_{\mathbf{S}_h}}{\|r_0\|_{\mathbf{S}_h}}\right)^{1/k} \le\; \varepsilon_k \qquad (k = 1,2,\dots,n), \tag{32}$$

$$\textit{where}\quad \varepsilon_k = \frac{2}{km^2}\sum_{i=1}^k\Big(\big|\lambda_i(Q_S^* + Q_S)\big| + s_i(Q_S)^2\Big) \;\to 0 \qquad (\textit{as } k\to\infty) \tag{33}$$

and $(\varepsilon_k)_{k\in\mathbf{N}^+}$ is a sequence independent of n and V_h.

4.3. Symmetric preconditioning for discretized elliptic problems.

4.3.1. General elliptic equations. Let us consider an elliptic problem

$$\begin{cases} Lu \equiv\ -\mathrm{div}\,(A\,\nabla u) +\ \mathbf{b}\cdot\nabla u + cu = g, \\ u_{|\Gamma_D} = 0, \qquad \frac{\partial u}{\partial \nu_A} + \alpha u_{|\Gamma_N} = 0, \end{cases} \tag{34}$$

where L satisfies Assumptions 3.2 and $g \in L^2(\Omega)$. We define $H^1_D(\Omega) = \{u \in H^1(\Omega) : u_{|\Gamma_D} = 0\}$; then Assumptions 3.2 ensure that problem (34) has a unique weak solution $u \in H^1_D(\Omega)$. Now let $V_h = span\{\varphi_1, \dots, \varphi_n\} \subset H^1_D(\Omega)$ be a given FEM subspace. We seek the FEM solution $u_h \in V_h$, which requires solving the $n \times n$ system

$$\mathbf{L}_h\,\mathbf{c} = \mathbf{b}, \tag{35}$$

where

$$\big(\mathbf{L}_h\big)_{i,j} = \int_\Omega \Big(A\,\nabla\varphi_i \cdot \nabla\varphi_j + (\mathbf{b}\cdot\nabla\varphi_i)\varphi_j + c\varphi_i\varphi_j\Big) \ + \int_{\Gamma_N} \alpha\varphi_i\varphi_j\,d\sigma$$

and $\mathbf{b}_j = \int_\Omega g\varphi_j$ $(j = 1, \dots, n)$. Following subsection 4.2, we define a preconditioner for system (35) as the discretization of a suitable symmetric elliptic operator. Let

$$Su :=\ -\mathrm{div}\,(A\,\nabla u) +\ hu \qquad \text{for}\ \ u \in H^2(\Omega) :\ u_{|\Gamma_D} = 0,\ \tfrac{\partial u}{\partial \nu_A} + \beta u_{|\Gamma_N} = 0, \tag{36}$$

where $h \in L^\infty(\Omega)$ and $h \geq 0$ if $\Gamma_D \neq \emptyset$ and $h \geq \delta_0 > 0$ if $\Gamma_D = \emptyset$, and, further, $\beta \in L^\infty(\Gamma_N)$ and $\beta \geq 0$. The corresponding inner product on $H^1_D(\Omega)$ is

$$\langle u, v\rangle_S := \int_\Omega (A\ \nabla u \cdot \nabla v + huv) + \int_{\Gamma_N} \beta uv\,d\sigma\,. \tag{37}$$

We introduce the matrix

$$\mathbf{S}_h = \Big\{\langle\varphi_i, \varphi_j\rangle_S\Big\}_{i,j=1}^n \tag{38}$$

as preconditioner for system (35), and then solve system (20) using the CG algorithm (5) with the $\mathbf{S}_h$-inner product and with $A = \mathbf{S}_h^{-1}\mathbf{L}_h$ and $A^* = \mathbf{S}_h^{-1}\mathbf{L}_h^T$.

THEOREM 4.3. *Let $V_h \subset H^1_D(\Omega)$ be an arbitrary FEM subspace and consider the FEM discretization* (35) *of problem* (34)*, using the stiffness matrix $\mathbf{S}_h$ as preconditioner. Then the superlinear convergence of the preconditioned CG method is mesh independent in the sense of Theorem* 4.2*; i.e., we have*

$$\left(\frac{\|r_k\|_{\mathbf{S}_h}}{\|r_0\|_{\mathbf{S}_h}}\right)^{1/k} \leq\ \varepsilon_k \qquad (k = 1, 2, \dots, n) \tag{39}$$

for the mesh independent sequence $\varepsilon_k \to 0$ from (33)*.*

Proof. The coercivity and boundedness assumptions on the coefficients of L and S imply in a standard way that L is S-bounded and S-coercive. Proposition 3.5 yields that L and S are compact-equivalent in $H^1_D(\Omega)$ if the latter is endowed with the inner product (37). Therefore Theorem 4.2 is valid with the compact operator Q_S defined via

$$\langle Q_S u, v\rangle_S = \int_\Omega \Big((\mathbf{b}\cdot\nabla u)v + (c-h)uv\Big) + \int_{\Gamma_N} (\alpha-\beta)uv\,d\sigma \qquad (u, v \in H^1_D(\Omega)), \tag{40}$$

which satisfies (21). □

We note that the above result is an extension of [8], where the mesh independence property has been proved for Dirichlet boundary conditions when either S is the symmetric part of L, or both L and S have constant coefficients.

Remark 5. Finding the correction terms in algorithm (5) with the present choice $A = \mathbf{S}_h^{-1}\mathbf{L}_h$ and $A^* = \mathbf{S}_h^{-1}\mathbf{L}_h^T$ is equivalent to the auxiliary problems

$$\begin{aligned} &\text{find } z_k \in V_h : \qquad \langle z_k, v\rangle_S = \langle L_S d_k, v\rangle_S \qquad (v \in V_h),\\ &\text{find } s_{k+1} \in V_h : \qquad \langle s_{k+1}, v\rangle_S = \langle L_S^* r_{k+1}, v\rangle_S \qquad (v \in V_h); \end{aligned}$$

i.e., z_k and s_{k+1} are the FEM solutions in V_h of the symmetric elliptic problems of the form $Sz_k = Ld_k$ and $Ss_{k+1} = L^* r_{k+1}$ with the boundary conditions of (36).

PROPOSITION 4.4. *Under the conditions of Theorem* 4.3*, the sequence* ε_k *in* (39) *satisfies*

$$\varepsilon_k \le \frac{4s}{k} \sum_{i=1}^{k} \frac{1}{\mu_i}, \tag{41}$$

where μ_i *(*$i \in \mathbf{N}^+$*) are the solutions of the eigenvalue problem*

$$Su = \mu u, \quad u_{|\Gamma_D} = 0, \quad r\left(\frac{\partial u}{\partial \nu_A} + \beta u\right)_{|\Gamma_N} = \mu u \tag{42}$$

and $s, r > 0$ *are constants defined below. When the asymptotics* $\mu_i = O(i^{2/d})$ *holds (in particular, for Dirichlet boundary conditions),*

$$\varepsilon_k \le O\left(\frac{\log k}{k}\right) \quad \textit{if} \;\; d = 2 \quad \textit{and} \quad \varepsilon_k \le O\left(\frac{1}{k^{2/d}}\right) \quad \textit{if} \;\; d \ge 3. \tag{43}$$

Proof. From (40) and (12) for $v = u$, letting $d = c - h$ and $\gamma = \alpha - \beta$, we obtain

$$\begin{aligned} \langle Q_S u, u\rangle_S &= \int_\Omega \left(d - \frac{1}{2}(\operatorname{div} \mathbf{b})\right) u^2 \; + \int_{\Gamma_N} \left(\gamma + \frac{1}{2}(\mathbf{b}\cdot\nu)\right) u^2 \, d\sigma \\ &\le C_1 \|u\|^2_{L^2(\Omega)} + C_2 \|u\|^2_{L^2(\Gamma_N)}. \end{aligned}$$

We have $\left|\langle (Q_S^* + Q_S)u, u\rangle_S\right| = 2\left|\langle Q_S u, u\rangle_S\right|$; hence the variational characterization of the eigenvalues yields

$$\begin{aligned} \left|\lambda_i(Q_S^* + Q_S)\right| &= \min_{H_{i-1}\subset H_S} \; \max_{\substack{u\perp H_{i-1}\\ u\ne 0}} \frac{\left|\langle (Q_S^* + Q_S)u, u\rangle_S\right|}{\|u\|_S^2} \\ &\le 2 \min_{H_{i-1}\subset H_S} \; \max_{\substack{u\perp H_{i-1}\\ u\ne 0}} \frac{C_1\|u\|^2_{L^2(\Omega)} + C_2\|u\|^2_{L^2(\Gamma_N)}}{\|u\|_S^2}, \end{aligned}$$

where H_{i-1} stands for an arbitrary $(i-1)$-dimensional subspace. On the other hand, here Q_S falls into the type (11), and hence (14) implies

$$\|Q_S u\|_S^2 \le 2K_1^2 \|u\|^2_{L^2(\Omega)} + 2K_2^2 \|u\|^2_{L^2(\Gamma_N)}.$$

Since $s_i(Q_S)^2 = \lambda_i(Q_S^* Q_S)$ and $\langle Q_S^* Q_S u, u\rangle_S = \|Q_S u\|_S^2$, we obtain as above that

$$s_i(Q_S)^2 = \min_{H_{i-1}\subset H_S} \max_{\substack{u\perp H_{i-1}\\ u\neq 0}} \frac{\langle Q_S^* Q_S u, u\rangle_S}{\|u\|_S^2}$$

$$\le \min_{H_{i-1}\subset H_S} \max_{\substack{u\perp H_{i-1}\\ u\neq 0}} \frac{2K_1^2\|u\|^2_{L^2(\Omega)} + 2K_2^2\|u\|^2_{L^2(\Gamma_N)}}{\|u\|_S^2}\,.$$

Altogether, letting $s := \frac{C_1+K_1^2}{m^2}$, $r := \frac{C_1+K_1^2}{C_2+K_2^2}$, formula (33) implies

$$\varepsilon_k \le \frac{4s}{k}\sum_{i=1}^{k}\hat\mu_i, \qquad \text{where } \hat\mu_i = \min_{H_{i-1}\subset H_S} \max_{\substack{u\perp H_{i-1}\\ u\neq 0}} \frac{\|u\|^2_{L^2(\Omega)} + \frac{1}{r}\|u\|^2_{L^2(\Gamma_N)}}{\|u\|_S^2},$$

in which the fraction equals $1/\mu$ for (42); hence the equality $\hat\mu_i = \frac{1}{\mu_i}$ follows from the variational characterization of the eigenvalues.

Estimate (43) follows from the asymptotics $\mu_i = O(i^{2/d})$ by an elementary calculation. For Dirichlet boundary conditions, this asymptotic behavior can be found in [11]. □

Remark 6. To the authors' knowledge the asymptotic behavior $\mu_i = O(i^{2/d})$ is not known for general (other than Dirichlet) boundary conditions. However, for the simple special case $-\Delta u = \mu u$, $\frac{\partial u}{\partial\nu}\big|_{\partial\Omega} = \mu u$, where Ω is a disc in $\mathbf{R}^2$, one can easily verify via the sign properties of the Bessel functions that μ_i are asymptotic to the Dirichlet eigenvalues and hence also satisfy $\mu_i = O(i^{2/d})$. This suggests a wider validity of this asymptotic rate.

Remark 7. It is of interest to compare the estimates (43), obtained in the context of the CGN method, to those valid for the GCG-LS method. In [8] we have proved $\varepsilon_k \le O\big(k^{-1/2}\big)$ in two dimensions on the unit square for the GCG-LS method under the same preconditioning (for Dirichlet boundary conditions, and using explicit formulae for the eigenvalues). Using the same technique, one can similarly derive $\varepsilon_k \le O\big(k^{-1/d}\big)$ in d dimensions (on the unit cube). That is, comparing with (43), we see that the decay rate of ε_k for the CGN method is almost or exactly (in two or more dimensions, respectively) the square of the decay rate for the GCG method, which compensates for the extra work of solving two auxiliary problems in the preconditioned CGN iteration.

4.3.2. An example: Convection-diffusion equations with Helmholtz preconditioners. As a special case of the preceding subsection, let us consider the case of a convection-diffusion operator L in (34) and a preconditioning operator S with constant coefficients. Namely, if $A \equiv I$ in (34), then we have the problem

$$(44)\qquad \begin{cases} Lu \equiv -\Delta u + \mathbf{b}(x)\cdot\nabla u + c(x)u = g(x), \\ u_{|\Gamma_D} = 0, \qquad \frac{\partial u}{\partial\nu} + \alpha(x)u_{|\Gamma_N} = 0, \end{cases}$$

where for clearness, the dependence of the coefficients on x has now been indicated unlike before. Let us define the preconditioning operator

$$(45)\qquad Su := -\Delta u + hu \qquad \text{for } u\in H^2(\Omega):\ u_{|\Gamma_D} = 0,\ \frac{\partial u}{\partial\nu} + \beta u_{|\Gamma_N} = 0,$$

where $h, \beta \in \mathbf{R}$ are constants such that $h \geq 0$ if $\Gamma_D \neq \emptyset$ and $h > 0$ if $\Gamma_D = \emptyset$, and further, $\beta \geq 0$. (For constant $\mathbf{b}$ and Dirichlet boundary conditions, the analysis of linear convergence in [25] proposes $h = O(|\mathbf{b}|^2)$ as an efficient choice.)

Then the auxiliary problems with this preconditioning are discrete Helmholtz problems with constant coefficients. For such problems various fast solvers are available (like fast Fourier transform, cyclic reduction, or multigrid; see, e.g., [19, 28, 30]), which, together with the mesh independence result of Theorem 4.3, turns $\mathbf{S}_h$ into an efficient preconditioner. We point out that this is an extension of [8], where the mesh independence property has been proved for Dirichlet boundary conditions under the strong restriction that the operator L itself has constant coefficients.

4.3.3. Elliptic systems. Analogously to subsection 4.3.1, we can consider elliptic systems

$$\left.\begin{aligned} L_i u \equiv -\mathrm{div}\,(A_i \nabla u_i) + \mathbf{b}_i \cdot \nabla u_i + \sum_{j=1}^{l} V_{ij} u_j = g_i, \\ u_{i\,|\Gamma_D} = 0, \qquad \tfrac{\partial u_i}{\partial \nu_A} + \alpha_i u_{i\,|\Gamma_N} = 0 \end{aligned}\right\} \qquad (i = 1, \dots, l), \tag{46}$$

where Ω, A_i, and α_i are as in Assumptions 3.2, $\mathbf{b}_i \in C^1(\overline{\Omega})^N$, $g_i \in L^2(\Omega)$, $V_{ij} \in L^\infty(\Omega)$. We assume that $\mathbf{b}_i$ and the matrix $V = \{V_{ij}\}_{i,j=1}^{l}$ satisfy the coercivity property

$$\lambda_{min}(V + V^T) - \max_i \mathrm{div}\,\mathbf{b}_i \geq 0$$

pointwise on Ω, where λ_{min} denotes the smallest eigenvalue; then system (46) has a unique weak solution $u \in H^1_D(\Omega)^l$. Such systems arise, e.g., from suitable time discretization and Newton linearization of transport systems, which often consist of a huge number of equations [33]. Now we choose an FEM subspace $V_h \subset H^1_D(\Omega)^l$ and look for the solution of the corresponding algebraic system $\mathbf{L}_h\, \mathbf{c} = \mathbf{b}$. We define the preconditioning operator $S = (S_1, \dots, S_l)$ as the l-tuple of independent operators

$$S_i u_i := -\mathrm{div}\,(A_i \nabla u) + h_i u \qquad \text{for} \;\; u_i \in H^2(\Omega): \; u_{i\,|\Gamma_D} = 0, \; \frac{\partial u_i}{\partial \nu_A} + \beta_i u_{i\,|\Gamma_N} = 0 \tag{47}$$

$(i = 1, \dots, l)$ with the conditions of (36), and let $\mathbf{S}_h$ be the stiffness matrix of S in $H^1_D(\Omega)^l$.

Then, similarly to subsection 4.3.1, one can verify that the superlinear convergence of the preconditioned CG method is mesh independent in the sense of Theorem 4.2; i.e., (32)–(33) hold.

This result is an extension of [24] where the above preconditioning has been introduced and its efficient parallelizability has been demonstrated; on the other hand, the mesh independence property was proved there for Dirichlet boundary conditions under strong restrictions on the matrix V (antisymmetric, or normal when the operator L itself has constant coefficients).

4.4. Nonsymmetric preconditioning in Hilbert space. Now let N be a general (possibly nonsymmetric) S-bounded and S-coercive operator which is compact-equivalent to L with $\mu = 1$; i.e., (9) becomes

$$L_S = N_S + Q_S\,. \tag{48}$$

We introduce the stiffness matrix of N_S,

$$\mathbf{N}_h = \Big\{ \langle N_S \varphi_i, \varphi_j \rangle_S \Big\}_{i,j=1}^n,$$

as preconditioner for system (18), and wish to solve

$$\mathbf{N}_h^{-1} \mathbf{L}_h \, \mathbf{c} = \hat{\mathbf{b}} \tag{49}$$

(with $\hat{\mathbf{b}} = \mathbf{N}_h^{-1}\mathbf{b}$) using the CG method . Since N is nonsymmetric, in order to define an inner product on $\mathbf{R}^n$ we preserve the stiffness matrix of S on V_h; i.e., using (19) we endow $\mathbf{R}^n$ with the $\mathbf{S}_h$-inner product $\langle \mathbf{c}, \mathbf{d} \rangle_{\mathbf{S}_h} := \mathbf{S}_h \, \mathbf{c} \cdot \mathbf{d}$ as earlier. Then the $\mathbf{S}_h$-adjoint of $\mathbf{N}_h^{-1}\mathbf{L}_h$ is $\mathbf{S}_h^{-1}\mathbf{L}_h^T \, \mathbf{N}_h^{-T}\mathbf{S}_h$; hence we apply the CG algorithm (5) with $A = \mathbf{N}_h^{-1}\mathbf{L}_h$ and $A^* = \mathbf{S}_h^{-1}\mathbf{L}_h^T\mathbf{N}_h^{-T}\mathbf{S}_h$.

Letting

$$\mathbf{Q}_h = \Big\{ \langle Q_S \varphi_i, \varphi_j \rangle_S \Big\}_{i,j=1}^n,$$

system (20) takes the form

$$(\mathbf{I}_h + \mathbf{N}_h^{-1}\mathbf{Q}_h) \, \mathbf{c} = \hat{\mathbf{b}}, \tag{50}$$

where $\mathbf{I}_h$ is the $n \times n$ identity matrix. Using (6), the CG algorithm (5) thus provides

$$\left(\frac{\|r_k\|_{\mathbf{S}_h}}{\|r_0\|_{\mathbf{S}_h}} \right)^{1/k} \le \frac{2}{k\nu_h} \sum_{i=1}^k \Big(\lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T \, \mathbf{N}_h^{-T}\mathbf{S}_h + \mathbf{N}_h^{-1}\mathbf{Q}_h) + \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T \, \mathbf{N}_h^{-T}\mathbf{S}_h\mathbf{N}_h^{-1}\mathbf{Q}_h) \Big) \tag{51}$$

($k = 1, 2, \dots, n$), where

$$\nu_h = \min_{\mathbf{c} \in \mathbf{R}^n} \frac{\|\mathbf{N}_h^{-1}\mathbf{L}_h\mathbf{c}\|_{\mathbf{S}_h}^2}{\|\mathbf{c}\|_{\mathbf{S}_h}^2} \,. \tag{52}$$

Again, our goal is to give a bound on (51) that is independent of V_h.

Proposition 4.5. *Let L and N be S-bounded and S-coercive operators, in particular,*

$$m := \inf_{\substack{u \in D(L) \\ u \neq 0}} \frac{\langle Lu, u \rangle}{\|u\|_S^2} > 0, \qquad \hat{m} := \inf_{\substack{u \in D(N) \\ u \neq 0}} \frac{\langle Nu, u \rangle}{\|u\|_S^2} > 0,$$
$$\hat{M} := \sup_{\substack{u \in D(N) \\ u \neq 0}} \frac{|\langle Nu, v \rangle|}{\|u\|_S \|v\|_S} > 0,$$

and let Q_S be a compact operator on H_S. Let $\mathbf{S}_h$, $\mathbf{N}_h$, and $\mathbf{Q}_h$ be defined as above, and let $s_i(Q_S)$ ($i = 1, 2, \dots$) denote the singular values of Q_S. Then the following relations hold:

(a) $$\sum_{i=1}^k \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T \, \mathbf{N}_h^{-T}\mathbf{S}_h\mathbf{N}_h^{-1}\mathbf{Q}_h) \le \frac{1}{\hat{m}^2} \sum_{i=1}^k s_i(Q_S)^2 \qquad (k = 1, \dots, n),$$

(b) $$\sum_{i=1}^k \big| \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T \, \mathbf{N}_h^{-T}\mathbf{S}_h + \mathbf{N}_h^{-1}\mathbf{Q}_h) \big| \le \frac{2}{\hat{m}} \sum_{i=1}^k s_i(Q_S) \qquad (k = 1, \dots, n),$$

(c) $$\nu_h \geq \frac{m^2}{\hat{M}^2}\,.$$

Proof. (a) We proceed in a manner similar to that of Proposition 4.1. Let $\lambda_i := \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T\,\mathbf{N}_h^{-T}\mathbf{S}_h\mathbf{N}_h^{-1}\mathbf{Q}_h)$ $(i = 1,\dots,n)$ and let $\mathbf{c}^i = (c_1^i,\dots,c_n^i) \in \mathbf{R}^n$ be corresponding eigenvectors with property (25). Then

$$\mathbf{S}_h\mathbf{N}_h^{-1}\mathbf{Q}_h\,\mathbf{c}^i \cdot \mathbf{N}_h^{-1}\mathbf{Q}_h\,\mathbf{c}^i = \lambda_i \qquad (i = 1,\dots,n). \tag{53}$$

Let $\mathbf{d}^i := \mathbf{N}_h^{-1}\mathbf{Q}_h\,\mathbf{c}^i$ for all i; that is,

$$\mathbf{N}_h\,\mathbf{d}^i = \mathbf{Q}_h\,\mathbf{c}^i\,. \tag{54}$$

For this $\mathbf{d}^i$ and λ_i, similarly to Proposition 4.1, we have (28) and, letting $u_i = \sum_{j=1}^n c_j^i\varphi_j \in V_h$ and $z_i = \sum_{j=1}^n d_j^i\varphi_j \in V_h$, we obtain (29). Further, for all $v = \sum_{j=1}^n p_j\varphi_j \in V_h$, with notation $\mathbf{p} = (p_1,\dots,p_n) \in \mathbf{R}^n$, (54) yields $\mathbf{N}_h\,\mathbf{d}^i \cdot \mathbf{p} = \mathbf{Q}_h\,\mathbf{c}^i \cdot \mathbf{p}$, which means

$$\langle N_S z_i, v\rangle_S = \langle Q_S u_i, v\rangle_S \qquad (v \in V_h).$$

From this we have

$$\|z_i\|_S^2 \leq \frac{1}{\hat{m}}\,\langle N_S z_i, z_i\rangle_S = \frac{1}{\hat{m}}\,\langle Q_S u_i, z_i\rangle_S \leq \frac{1}{\hat{m}}\,\|Q_S u_i\|_S\|z_i\|_S;$$

hence $\|z_i\|_S \leq \frac{1}{\hat{m}}\,\|Q_S u_i\|_S$. Then from (29)

$$\sum_{i=1}^k \lambda_i \leq \frac{1}{\hat{m}^2}\,\sum_{i=1}^k \|Q_S u_i\|_S^2 = \frac{1}{\hat{m}^2}\,\sum_{i=1}^k \langle Q_S^* Q_S u_i, u_i\rangle_S, \tag{55}$$

whence the desired estimate follows in the same way as from (30) in Proposition 4.1.

(b) Now let $\lambda_i := \lambda_i(\mathbf{S}_h^{-1}\mathbf{Q}_h^T\,\mathbf{N}_h^{-T}\mathbf{S}_h + \mathbf{N}_h^{-1}\mathbf{Q}_h)$ and let $\mathbf{c}^i = (c_1^i,\dots,c_n^i) \in \mathbf{R}^n$ be corresponding eigenvectors with property (25). Then

$$\lambda_i = \lambda_i\,\mathbf{S}_h\,\mathbf{c}^i\!\cdot\!\mathbf{c}^i = \mathbf{Q}_h^T\,\mathbf{N}_h^{-T}\mathbf{S}_h\,\mathbf{c}^i\!\cdot\!\mathbf{c}^i + \mathbf{S}_h\mathbf{N}_h^{-1}\mathbf{Q}_h\,\mathbf{c}^i\!\cdot\!\mathbf{c}^i = 2\,\mathbf{S}_h\mathbf{N}_h^{-1}\mathbf{Q}_h\,\mathbf{c}^i\!\cdot\!\mathbf{c}^i = 2\,\mathbf{Q}_h\,\mathbf{c}^i\!\cdot\!\mathbf{e}^i,$$

where $\mathbf{e}^i := \mathbf{N}_h^{-T}\mathbf{S}_h\,\mathbf{c}^i$ for all i. Here for all $v = \sum_{j=1}^n p_j\varphi_j \in V_h$, with notation $\mathbf{p} = (p_1,\dots,p_n) \in \mathbf{R}^n$, we obtain $\mathbf{e}^i \cdot \mathbf{N}_h\,\mathbf{p} = \mathbf{S}_h\,\mathbf{c}^i \cdot \mathbf{p}$, which means $\langle w_i, N_S v\rangle_S = \langle u_i, v\rangle_S$ for all $v \in V_h$, where $w_i = \sum_{j=1}^n e_j^i\varphi_j$ and $u_i = \sum_{j=1}^n c_j^i\varphi_j$, or

$$\langle N_S^* w_i, v\rangle_S = \langle u_i, v\rangle_S \qquad (v \in V_h). \tag{56}$$

Denote by P the orthogonal projection of H_S onto V_h. Then (56) yields $u_i = PN_S^* w_i$. Here the linear mapping $(PN_S^*)_{|V_h} : V_h \to V_h$ is one-to-one, since for all $v \in V_h$

$$\langle PN_S^* v, v\rangle_S = \langle N_S^* v, v\rangle_S = \langle N_S v, v\rangle_S \geq \hat{m}\|v\|_S^2\,. \tag{57}$$

Therefore

$$\mathbf{Q}_h\,\mathbf{c}^i \cdot \mathbf{e}^i = \langle Q_S u_i, w_i\rangle_S = \langle Q_S u_i, (PN_S^*)_{|V_h}^{-1} u_i\rangle_S = \langle u_i, Q_S^*(PN_S^*)_{|V_h}^{-1} u_i\rangle_S.$$

Here the operator $(PN_S^*)^{-1}_{|V_h}$ has a norm-preserving extension $\hat{N}$ from V_h onto H_S (namely, with $\hat{N}\big|_{(V_h)\perp} := 0$), and from (57) we have $\|\hat{N}\| \le \frac{1}{\hat{m}}$. Altogether, we obtain

$$\sum_{i=1}^{k}|\lambda_i| = 2\sum_{i=1}^{k}\left|\langle Q_S^*(PN_S^*)^{-1}_{|V_h}u_i,\, u_i\rangle_S\right| \,=\, 2\sum_{i=1}^{k}\left|\langle Q_S^*\hat{N}u_i,\, u_i\rangle_S\right| \,\le\, 2\sum_{i=1}^{k}s_i\big(Q_S^*\hat{N}\big)$$

$$\le \frac{2}{\hat{m}}\sum_{i=1}^{k}s_i\big(Q_S^*\big) = \frac{2}{\hat{m}}\sum_{i=1}^{k}s_i\big(Q_S\big)$$

(where, in the inequalities, statements (a) and (b) of Proposition 2.2 have been used, respectively).

(c) Let $\mathbf{c} \in \mathbf{R}^n$ be arbitrary, $\mathbf{d} := \mathbf{N}_h^{-1}\mathbf{L}_h\mathbf{c}$. Let $u = \sum_{j=1}^{n} c_j\varphi_j \in V_h$ and $z = \sum_{j=1}^{n} d_j\varphi_j \in V_h$. Then $m\|u\|_S^2 \le \langle L_S u, u\rangle_S = \mathbf{L}_h\,\mathbf{c}\cdot\mathbf{c} = \mathbf{N}_h\,\mathbf{d}\cdot\mathbf{c} = \langle N_S z, u\rangle_S \le \|N_S z\|_S\|u\|_S$; hence

$$m\|u\|_S \le \|N_S z\|_S$$

and

$$\frac{\|\mathbf{N}_h^{-1}\mathbf{L}_h\mathbf{c}\|^2_{\mathbf{S}_h}}{\|\mathbf{c}\|^2_{\mathbf{S}_h}} = \frac{\mathbf{S}_h\,\mathbf{d}\cdot\mathbf{d}}{\mathbf{S}_h\,\mathbf{c}\cdot\mathbf{c}} = \frac{\|z\|_S^2}{\|u\|_S^2} \ge m^2\,\frac{\|z\|_S^2}{\|N_S z\|_S^2} \,\ge\, \frac{m^2}{\hat{M}^2}. \qquad \square$$

By virtue of (51) and Proposition 4.5, we have proved the following theorem.

THEOREM 4.6. *Let L and N be S-bounded and S-coercive operators that are compact-equivalent in H_S with $\mu = 1$. Let the compact operator Q_S be as in* (48). *Then for any subspace $V_h = span\{\varphi_1, \dots, \varphi_n\} \subset H_S$, the CG algorithm* (5) *with $\mathbf{S}_h$-inner product, applied for the $n \times n$ preconditioned system* (49), *yields*

$$(58) \qquad \left(\frac{\|r_k\|_{\mathbf{S}_h}}{\|r_0\|_{\mathbf{S}_h}}\right)^{1/k} \le\; \varepsilon_k \qquad (k = 1, 2, \dots, n),$$

$$(59)\ \ \textit{where}\quad \varepsilon_k = \frac{2\hat{M}^2}{km^2}\sum_{i=1}^{k}\left(\frac{2}{\hat{m}}\,s_i(Q_S) + \frac{1}{\hat{m}^2}\,s_i(Q_S)^2\right) \;\to 0 \qquad (\textit{as } k \to \infty)$$

and $(\varepsilon_k)_{k\in\mathbf{N}^+}$ is a sequence independent of n and V_h.

Remark 8. When one preconditions L with N, a useful choice for the operator S is the symmetric part of N: i.e., if $D(N) = D(N^*)$, then $S = (N + N^*)/2$, and if $D(N) \ne D(N^*)$, then S is an operator that generates the inner product satisfying $\langle u, v\rangle_S := \frac{1}{2}(\langle Nu, v\rangle + \langle u, Nv\rangle)$ for $u, v \in D(N)$; see [23]. Then in Proposition 4.5 we have $\langle Nu, u\rangle = \|u\|_S^2$ $(u \in D(N))$, and hence $\hat{m} = 1$.

4.5. Nonsymmetric preconditioning for discretized elliptic problems.

This section contains our most general result for elliptic operators: in the studied coercive framework, preconditioning with an arbitrary operator N that is compact-equivalent with L provides mesh independent superlinear convergence. Besides its theoretical aspect, the importance of this property will be shown below by some practical examples as well. Let us first consider the elliptic problem (34)

$$(60) \qquad \begin{cases} Lu \equiv -\mathrm{div}\,(A\,\nabla u) + \mathbf{b}\cdot\nabla u + cu = g, \\ u_{|\Gamma_D} = 0, \qquad \frac{\partial u}{\partial \nu_A} + \alpha u_{|\Gamma_N} = 0, \end{cases}$$

and let us define the nonsymmetric preconditioning operator

$$Nu := -\operatorname{div}(A\,\nabla u) + \mathbf{w}\cdot\nabla u + zu \quad \text{for } \ u\in H^2(\Omega):\ u_{|\Gamma_D}=0,\ \frac{\partial u}{\partial \nu_A}+\eta u_{|\Gamma_N}=0 \tag{61}$$

for some properly chosen functions $\mathbf{w}, z, \eta$, where L and N satisfy Assumptions 3.2 in the obvious sense, and further, $g\in L^2(\Omega)$. Accordingly, the preconditioner for the discretized problem (35) is the nonsymmetric stiffness matrix

$$\left(\mathbf{N}_h\right)_{i,j} = \int_\Omega \Big(A\,\nabla\varphi_i\cdot\nabla\varphi_j + (\mathbf{w}\cdot\nabla\varphi_i)\,\varphi_j + z\varphi_i\varphi_j\Big) \ + \int_{\Gamma_N}\eta\varphi_i\varphi_j\,d\sigma\,.$$

We use the same energy space as in the symmetric case, i.e., $H_S = H^1_D(\Omega)$ with inner product (37). We then solve the preconditioned system using the CG algorithm (5) with the $\mathbf{S}_h$-inner product and with $A=\mathbf{N}_h^{-1}\mathbf{L}_h$ and $A^* = \mathbf{S}_h^{-1}\mathbf{L}_h^T\mathbf{N}_h^{-T}\mathbf{S}_h$.

THEOREM 4.7. *Let $V_h\subset H^1_D(\Omega)$ be an arbitrary FEM subspace and consider the FEM discretization* (35) *of problem* (34)*, using the stiffness matrix $\mathbf{N}_h$ as preconditioner. Then the superlinear convergence of the preconditioned CG method is mesh independent in the sense of Theorem* 4.6*; i.e.,* (58)*–*(59) *hold.*

Proof. The proof is similar to that of Theorem 4.3, but now Theorem 4.6 is applied in $H^1_D(\Omega)$. □

Examples. Let us consider problem (44); i.e., when in (60) we have

$$Lu = -\Delta u + \mathbf{b}(x)\cdot\nabla u + c(x)u,$$

where for clarity, the dependence of the coefficients on x has now been indicated. For convection-dominated problems (i.e., when $|\mathbf{b}|$ is large), the inclusion of nonsymmetric terms in N may turn it into a much better approximation of L than a symmetric preconditioner like (45). Although the preconditioner N thus becomes nonsymmetric as is L itself, the solution of the auxiliary problems can still remain considerably simpler than the original one. We illustrate this with two examples.

1. One can propose a preconditioning operator with constant coefficients:

$$Nu = -\Delta u + \ \mathbf{w}\cdot\nabla u + zu \qquad \text{for } \ u\in H^2(\Omega):\ u_{|\Gamma_D}=0,\ \frac{\partial u}{\partial\nu}+\eta u_{|\Gamma_N}=0, \tag{62}$$

where $\mathbf{w}\in\mathbf{R}^d$, $z,\eta\in\mathbf{R}$ are constants such that $z\ge 0$ if $\Gamma_D\ne\emptyset$ and $z>0$ if $\Gamma_D=\emptyset$, and further, $\eta\ge 0$. Owing to the fact that N has constant coefficients, one can rely on efficient solution methods for the auxiliary problems. Here one can use either multigrid or multilevel methods, or (if Ω is rectangular or the boundary conditions allow the problem to be easily embedded into a rectangular domain) fast direct solvers for separable equations are available; see, e.g., [29].

2. The preconditioning operator (62) can be further simplified if one convection coefficient is dominating. Assume that, say, $b_1(x)$ has considerably larger values than $b_j(x)$ $(j\ge 2)$. Then one can include only one nonsymmetric coefficient, i.e., propose the preconditioning operator

$$Nu = -\Delta u + w_1\frac{\partial u}{\partial x_1} + zu \qquad \text{for } \ u\in H^2(\Omega):\ u_{|\Gamma_D}=0,\ \frac{\partial u}{\partial\nu}+\eta u_{|\Gamma_N}=0, \tag{63}$$

where $w_1, z, \eta \in \mathbf{R}$ are constants with the same properties as those required for (62). In this case (above all, if $b_1(x)$ are large), the presence of the term $w_1 \frac{\partial u}{\partial x_1}$ itself may turn N into a much better approximation of L. Nevertheless, since this term is one-dimensional, the solution of the auxiliary problems remains considerably simpler than the original one, e.g., via local one-dimensional Green's functions [5]. (The above operator N has been proposed in [8], where the mesh independence result of the PCG method has been proved for Dirichlet boundary conditions under the strong restriction that the operator L itself has constant coefficients.)

Analogously to the symmetric case in subsection 4.3.3, the above results can be extended to systems in a straightforward way. Namely, let us consider system (46) and introduce the preconditioning operator N as an l-tuple of decoupled operators N_i, where each N_i is of the type (61). Then the superlinear convergence of the preconditioned CG method is mesh independent in the sense of Theorem 4.6; i.e., (58)–(59) hold. Since N_i are decoupled, the resulting algorithm is parallelizable. This turns it into an efficient method if, for instance, each N_i is like (62), or the problem itself is in one dimension, which may occur, e.g., after using some method of splitting in meteorological models with several components; see [33].

5. Some closing remarks.

5.1. Conclusions and notes on numerical realization. The main results of this paper can be summarized as follows. If two elliptic operators are compact-equivalent (which requires that their principal parts coincide up to a constant factor and they have homogeneous Dirichlet conditions on the same portion of the boundary), then the PCGN method provides mesh independent superlinear convergence; i.e., a bound on the rate of superlinear convergence is given in the form of a sequence which is mesh independent and is determined only by the elliptic operators. The analogous result holds for suitable elliptic systems where, as an additional advantage, the preconditioning operator can be chosen to be decoupled. Various further examples have been shown on the efficient choice of compact-equivalent preconditioners.

For the GCG-LS method we have obtained similar earlier results in [8, 24], but with severe restrictions: except for some special cases, both the original and preconditioning operators had to contain constant coefficients, and further, only Dirichlet boundary conditions have been considered. On the other hand, numerical experiments in [24] suggest that the restrictions are probably mostly technical, since a similar superlinear behavior has been observed for test problems with or without these conditions. Remark 7 suggests that the PCGN and GCG-LS methods require the same order of operations for prescribed accuracy; hence there is no a priori preference for one over the other. In any case, a favorable property for the PCGN iteration is the generality of the underlying theory, clarified in the present paper.

The PCGN algorithm has been applied in the setting of subsection 4.3.3 as an inner iterative solver for Newton's method for nonlinear nonsymmetric elliptic systems in [1]. As in the above-mentioned experiments in [24], an efficient performance of the compact-equivalent preconditioning has been observed. Further numerical experiments are beyond the scope and length of this paper.

When realizing the equivalent operator preconditioning for a problem with a second order operator with variable coefficients, one can use an inner-outer iteration method, i.e., precondition in the outer iterations with the given second order operator and use inner iterations to solve this equation. For the superlinear rate to remain, the inner iteration errors must not be of an order greater than that for the first order part of the operator. For optimal complexity of the overall computations to hold, one should then solve the arising inner systems with an optimal order of computational complexity, i.e., proportional to the degrees of freedom used in the discretization of the differential equation.

5.2. On singular perturbation problems. For singular perturbation problems such as

$$L_\varepsilon u \equiv -\varepsilon \Delta u + \mathbf{b} \cdot \nabla u + cu = f$$

(plus boundary conditions), where $\varepsilon > 0$ but $\varepsilon \ll \|\mathbf{b}\|$, one cannot neglect the first order term when forming an efficient preconditioner. Such problems are characterized by thin boundary and/or interior layers, and the diffusion term plays a noticeable role only in the layer. This property is not exploited in preconditioners like (62). A possible approach for handling such problems therefore is to use the following defect-correction method:

$$L_{\delta(x)}(u_{k+1} - u_k) = f - L_\varepsilon u \qquad (k \in \mathbf{N}^+),$$

where u_0 is given and in practice only one or two steps need to be performed. Here

$$L_{\delta(x)} u := -\delta(x) \Delta u + \mathbf{b} \cdot \nabla u + cu,$$

where $\delta(x) = 0$ outside the layers and increases continuously along each characteristic line (defined by the velocity vector $\mathbf{b}$) from zero to ε in the layers. The widths of the layers are typically chosen as $\varepsilon \log(1/\varepsilon)$. To solve the correction equation by iteration, one can form a preconditioner S by using the operator $\mathbf{b} \cdot \nabla u + hu$ outside the layers and $-\delta(x)\Delta u + \mathbf{b} \cdot \nabla u + hu$ in the layers for some properly chosen function $h \geq 0$. The analysis of the problem will not be considered further in the present paper.

REFERENCES

[1] I. Antal and J. Karátson, *A mesh independent superlinear algorithm for some nonlinear nonsymmetric elliptic systems*, Comput. Math. Appl., to appear.

[2] S. F. Ashby, T. A. Manteuffel, and P. E. Saylor, *A taxonomy for conjugate gradient methods*, SIAM J. Numer. Anal., 27 (1990), pp. 1542–1568.

[3] O. Axelsson, *A generalized conjugate gradient, least square method*, Numer. Math., 51 (1987), pp. 209–227.

[4] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, UK, 1994.

[5] O. Axelsson and S. V. Gololobov, *A combined method of local Green's functions and central difference method for singularly perturbed convection-diffusion problems*, J. Comput. Appl. Math., 161 (2003), pp. 245–257.

[6] O. Axelsson and J. Karátson, *On the rate of convergence of the conjugate gradient method for linear operators in Hilbert space*, Numer. Funct. Anal., 23 (2002), pp. 285–302.

[7] O. Axelsson and J. Karátson, *Symmetric part preconditioning for the conjugate gradient method in Hilbert space*, Numer. Funct. Anal. Optim., 24 (2003), pp. 455–474.

[8] O. Axelsson and J. Karátson, *Superlinearly convergent CG methods via equivalent preconditioning for nonsymmetric elliptic operators*, Numer. Math., 99 (2004), pp. 197–223.

[9] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North–Holland, Amsterdam, 1978.

[10] P. Concus and G. H. Golub, *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, in Computing Methods in Applied Sciences and Engineering, Lecture Notes in Econom. and Math. Systems 134, R. Glowinski and J.-L. Lions, eds., Springer, Berlin, 1976, pp. 56–65.

[11] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. II, Wiley Classics Library, John Wiley & Sons, New York, 1989.

[12] H. C. Elman and M. H. Schultz, *Preconditioning by fast direct methods for nonself-adjoint nonseparable elliptic equations*, SIAM J. Numer. Anal., 23 (1986), pp. 44–57.

[13] V. Faber and T. Manteuffel, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal., 21 (1984), pp. 352–362.

[14] V. Faber, T. Manteuffel, and S. V. Parter, *On the theory of equivalent operators and application to the numerical solution of uniformly elliptic partial differential equations*, Adv. in Appl. Math., 11 (1990), pp. 109–163.

[15] Z. Fortuna, *Some convergence properties of the conjugate gradient method in Hilbert space*, SIAM J. Numer. Anal., 16 (1979), pp. 380–384.

[16] I. Gohberg, S. Goldberg, and M. A. Kaashoek, *Classes of Linear Operators*, Vol. I, Oper. Theory Adv. Appl. 49, Birkhäuser Verlag, Basel, Switzerland, 1990.

[17] I. Gohberg and S., Goldberg, *Basic Operator Theory*, Birkhäuser, Boston, MA, 1981.

[18] C. I. Goldstein, T. A. Manteuffel, and S. V. Parter, *Preconditioning and boundary conditions without H_2 estimates: L_2 condition numbers and the distribution of the singular values*, SIAM J. Numer. Anal., 30 (1993), pp. 343–376.

[19] W. Hackbusch, *Multigrid Methods and Applications*, Springer Ser. Comput. Math. 4, Springer, Berlin, 1985.

[20] R. M. Hayes, *Iterative methods of solving linear problems in Hilbert space*, Nat. Bur. Standards Appl. Math. Ser, 39 (1954), pp. 71–104.

[21] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards, Sect. B, 49 (1952), pp. 409–436.

[22] J. Karátson, *Mesh independent superlinear convergence estimates of the conjugate gradient method for some equivalent self-adjoint operators*, Appl. Math., 50 (2005), pp. 277–290.

[23] J. Karátson, *Superlinear PCG Algorithms: Symmetric Part Preconditioning and Boundary Conditions*, Preprint 2006-10, Department of Applied Analysis, ELTE University, Budapest, Hungary; available online from http://www.cs.elte.hu/applanal/eng/preprint_eng.html.

[24] J. Karátson and T. Kurics, *Superlinearly convergent PCG algorithms for some nonsymmetric elliptic systems*, J. Comput. Appl. Math., to appear.

[25] T. Manteuffel and J. Otto, *Optimal equivalent preconditioners*, SIAM J. Numer. Anal., 30 (1993), pp. 790–812.

[26] T. A. Manteuffel and S. V. Parter, *Preconditioning and boundary conditions*, SIAM J. Numer. Anal., 27 (1990), pp. 656–694.

[27] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser Verlag, Basel, Switzerland, 1993.

[28] T. Rossi and J. Toivanen, *A parallel fast direct solver for block tridiagonal systems with separable matrices of arbitrary dimension*, SIAM J. Sci. Comput., 20 (1999), pp. 1778–1796.

[29] P. N. Swarztrauber, *A direct method for the discrete solution of separable elliptic equations*, SIAM J. Numer. Anal., 11 (1974), pp. 1136–1150.

[30] P. N. Swarztrauber, *The methods of cyclic reduction, Fourier analysis and the FACR algorithm for the discrete solution of Poisson's equation on a rectangle*, SIAM Rev., 19 (1977), pp. 490–501.

[31] R. Winther, *Some superlinear convergence results for the conjugate gradient method*, SIAM J. Numer. Anal., 17 (1980), pp. 14–17.

[32] O. Widlund, *A Lanczos method for a class of nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 15 (1978), pp. 801–812.

[33] Z. Zlatev, *Computer Treatment of Large Air Pollution Models*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1995.