Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Rapid error reduction for block Gauss-Seidel based on p-hierarchical bases

(revised version: June 2011)

by

Sabine Le Borne, and Jeffrey Ovall

Preprint no.: 33 2011

RAPID ERROR REDUCTION FOR BLOCK GAUSS-SEIDEL BASED ON P-HIERARCHICAL BASIS

SABINE LE BORNE * AND JEFFREY S. OVALL †

Abstract. We consider a two-level block Gauss-Seidel iteration for solving systems arising from finite element discretizations employing higher-order elements. A p-hierarchical basis is used to induce this block structure. Using superconvergence results normally employed in the analysis of gradient recovery schemes, we argue that a massive reduction of H^1-error occurs in the first iterate, so that the discrete solution is adequately resolved in very few iterates—sometimes a single iteration is sufficient. Numerical experiments support these claims.

Key words. higher-order finite elements, hierarchical bases, block Gauss-Seidel, hierarchical matrices

AMS subject classifications. 65N22, 65N55, 65N30, 65F08, 65F05

1. Introduction. We consider linear systems arising from the discretization of second-order, linear, elliptic problems via Lagrange finite elements of order > 1. We use a p-hierarchical basis of the finite element space V to split it as $V = V_1 \oplus V_2$, with V_1 containing the lower order elements (i.e., at least the affine elements). This splitting of V induces a natural 2×2 block structure and is designed so that the ill-conditioning of the system matrix is concentrated in its $(1,1)$-block and the $(2,2)$-block is well-conditioned (in the ideal case, it may even be spectrally equivalent to its own diagonal). We propose a block Gauss-Seidel iteration, either on its own as a classical linear iteration, or as a preconditioner for a Krylov iteration. The standard convergence analysis for such two-level schemes, at least in case of symmetric problems, is presented for example in [3].

Our analysis differs from others in our arguments concerning the *significant* error reduction in the first step of the iterative method, which in some cases already resolves the discrete solution up to the discretization error with respect to the H^1-norm. These arguments use superconvergence results commonly appearing in the analysis of gradient-recovery schemes. Our analysis requires that the block lower-triangular system (i.e., the system representing the preconditioner of the forward block Gauss-Seidel method) is solved accurately. In practice, this means that more effort must be spent on the $(1,1)$-block, in particular on that portion corresponding to the affine elements, and something cheap can be applied to the $(2,2)$-block. This approach shares a similar philosophical principle with cascadic multigrid [8] and other cascadic methods [12, 11, 28], in which more effort is taken to solve "coarser" problems accurately in order that less effort might be spent on "finer" problems, for a net reduction in solve-time. We also mention the so-called p-multigrid and hp-multigrid methods, which have the same basic structure as the classical h-multigrid methods, but with a p-hierarchical basis used to generate the hierarchy of levels. We refer interested readers to [26] and references therein for discussions of these multigrid variants.

The rest of this paper is organized as follows: In Section 2, we introduce our basic notation related to the elliptic problems under consideration and their finite element

*Tennessee Technological University, Department of Mathematics, Box 5054, Cookeville, TN 38505 — Phone: (931) 372-3690, E-Mail: sleborne@tntech.edu. This work has been partially supported by NSF Grant # DMS-0913017

†University of Kentucky, Department of Mathematics, 761 Patterson Office Tower, Lexington, KY 40506-0027 — Phone: (859) 257-6792, E-Mail: jovall@ms.uky.edu

discretization. Section 3 introduces the two-level block Gauss-Seidel iteration and contains our main results concerning the massive error reduction in the first iteration. Hierarchical matrices, and how we employ them in this context, are described in Section 4. Finally, in Section 5, we present numerical results to demonstrate our claims on both symmetric and non-symmetric problems for quadratic and quartic elements.

2. Preliminaries. We are interested in efficiently and reliably solving linear systems associated with finite element discretizations of problems of the form

$$\text{Find } u \in \mathcal{H} \text{ such that } B(u,v) = F(v) \text{ for all } v \in \mathcal{H} \ , \tag{2.1}$$

where

$$B(u,v) = \int_\Omega K\nabla u \cdot \nabla v + (\mathbf{b}\cdot\nabla u + cu)v \, dx \ , \tag{2.2}$$

$$F(v) = \int_\Omega fv \, dx + \int_{\Gamma_N} gv \, ds \ , \tag{2.3}$$

$$\mathcal{H} = H^1_{0,D}(\Omega) = \{v \in H^1(\Omega) : \ u = 0 \text{ on } \Gamma_D \text{ in the sense of trace}\} \ . \tag{2.4}$$

Here, $\Omega \subset \mathbb{R}^2$ is open, bounded and (for simplicity) polygonal, having boundary $\Gamma = \Gamma_D \cup \Gamma_N$. The *Dirichlet and Neumann* portions of the boundary, Γ_D and Γ_N respectively, are disjoint, with Γ_D closed in the relative topology. In the present work, for the sake of analysis, we make the assumption that the data functions, $K, \mathbf{b}, c, f, g$, are smooth on Ω. The matrix K is also assumed to be symmetric and uniformly positive definite throughout the domain. We assume that both F and B are bounded, and we assume, for convenience of exposition, that B is coercive. The key results, Lemma 3.1 and Theorem 3.3, carry over naturally to the more general inf-sup setting, assuming that similar inf-sup conditions are also uniformly satisfied by the discrete problems described below.

We discretize (2.1) by choosing a finite dimensional subspace $V \subset \mathcal{H}$ and restricting the problem:

$$\text{Find } \hat{u} \in V \text{ such that } B(\hat{u},v) = F(v) \text{ for all } v \in V \ . \tag{2.5}$$

Having chosen a basis $\{\psi_k : 1 \le k \le N\}$, we make the obvious identification between coefficient vectors $\mathbf{v} \in \mathbb{R}^N$ and functions $v \in V$, and obtain the following linear system corresponding to (2.5):

$$A\mathbf{u} = \mathbf{f} \text{ where } A_{ij} = B(\phi_j, \phi_i) \text{ and } \mathbf{f}_i = F(\phi_i) \ . \tag{2.6}$$

It is clear that $B(v,w) = \mathbf{w}^t A\mathbf{v}$, and when $B(\cdot,\cdot)$ is an inner-product ($\mathbf{b} = \mathbf{0}$), A is symmetric, positive-definite and we have

$$|\!|\!| v |\!|\!|^2 \doteq B(v,v) = \mathbf{v}^t A\mathbf{v} = \|\mathbf{v}\|_A^2 \ . \tag{2.7}$$

For finite element discretizations, the *stiffness matrix* A is large and sparse, and it is well-known that $\kappa(A)$ grows without bound as N increases.

Standard finite element discretizations begin with a conforming triangulation (mesh) $\mathcal{T}$ consisting of:

- Triangles $T \in \mathcal{T}$.
- Edges $e \in \overline{\mathcal{E}} = \mathcal{E}_D \cup \mathcal{E}$, with $\mathcal{E}_D$ denoting those edges contained in Γ_D, and $\mathcal{E}$ denoting those edges which are not.

- Vertices $z \in \overline{\mathcal{V}} = \mathcal{V}_D \cup \mathcal{V}$ with $\mathcal{V}_D$ denoting those vertices contained in Γ_D, and $\mathcal{V}$ denoting those vertices which are not.

Here, we have implicitly assumed that (the interior of) each boundary edge is contained entirely within either Γ_D or Γ_N. The conformity condition on $\mathcal{T}$ means that for any two triangles $T, T' \in \mathcal{T}$, $T \cap T'$ is either: empty, consists of a (complete) shared edge $e \in \mathcal{E}$, or consists of a single shared vertex $z \in \mathcal{V}$. We consider families of quasi-uniform, shape-regular meshes.

The p'th order, continuous Lagrange finite element spaces associated with the triangulation $\mathcal{T}$ are defined as

$$S_p = \{v \in C(\overline{\Omega}) \cap \mathcal{H} : \ v_{|_T} \in \mathbb{P}_p \text{ for each } T \in \mathcal{T}\} \ , \tag{2.8}$$

where $\mathbb{P}_p$ is the collection of all polynomials of (total) degree $\leq p$. For simple reference, we give the dimension of S_p, $|S_p|$, for $p = 1, 2, 3, 4$, in terms of the cardinalities of $\mathcal{V}$, $\mathcal{E}$ and $\mathcal{T}$:

$$|S_1| = |\mathcal{V}| \ , \ |S_2| = |\mathcal{V}| + |\mathcal{E}| \ , \ |S_3| = |\mathcal{V}| + 2|\mathcal{E}| + |\mathcal{T}| \ , \ |S_4| = |\mathcal{V}| + 3|\mathcal{E}| + 3|\mathcal{T}| \ .$$

For quasi-uniform, shape-regular meshes $\{\mathcal{T}_h\}$, and $\hat{u} \in S_p$ satisfying (2.5) the standard *a priori* error estimate is

$$|\!|\!| u - \hat{u} |\!|\!| \lesssim h^p \|u\|_{p+1} \ , \tag{2.9}$$

where $\|\cdot\|_j$ denotes the usual Sobolev norm on $H^j(\Omega)$. For quasi-uniform meshes in $\mathbb{R}^2$ it holds that $h \sim N^{-1/2}$, where N is the number of vertices, edges, or triangles in the mesh. In the case of adapted meshes, $N^{-p/2}$ is considered the optimal convergence rate.

3. Lagrange vs. Hierarchical Bases, Two-Level Block Gauss-Seidel. The most common basis for S_p is the so-called *Lagrange (nodal) basis*. Given an appropriate set of (vertex, edge and interior) nodes, all basis functions are globally continuous, piecewise of degree p, and the basis function associated with a given node has value 1 at that node and vanishes at all other nodes. These basis functions are most readily described on a given triangle in terms of its barycentric coordinates (ℓ_1, ℓ_2, ℓ_2). Although in principle the $\binom{p+2}{2}$ nodes on triangle T may be distributed in a number of ways—for example, at Fekete points [30]—we consider the usual case in which they are evenly distributed at the following barycentric coordinates

$$\left\{ \left(\frac{i}{p}, \frac{j}{p}, \frac{k}{p} \right) : \ i, j \geq 0 \ , \ i + j + k = p \right\} \ .$$

The corresponding basis functions, when restricted to T, may be expressed as

$$\psi_{(\frac{i}{p}, \frac{j}{p}, \frac{k}{p})} = \left(\prod_{i'=0}^{i-1} \frac{p\ell_1 - i'}{i - i'} \right) \left(\prod_{j'=0}^{j-1} \frac{p\ell_2 - j'}{j - j'} \right) \left(\prod_{k'=0}^{k-1} \frac{p\ell_3 - k'}{k - k'} \right) \ .$$

Such a function takes the value 1 at its associated node, and 0 at all other nodes. In contrast, a *p-hierarchical basis* for S_p is built up from functions of various degrees. In this work we consider hierarchical bases of S_2 and S_4 suggested by the hierarchical splittings

$$S_2 = S_1 \oplus (S_2 \setminus S_1) \quad , \quad S_4 = S_1 \oplus (S_2 \setminus S_1) \oplus (S_4 \setminus S_2) \ .$$

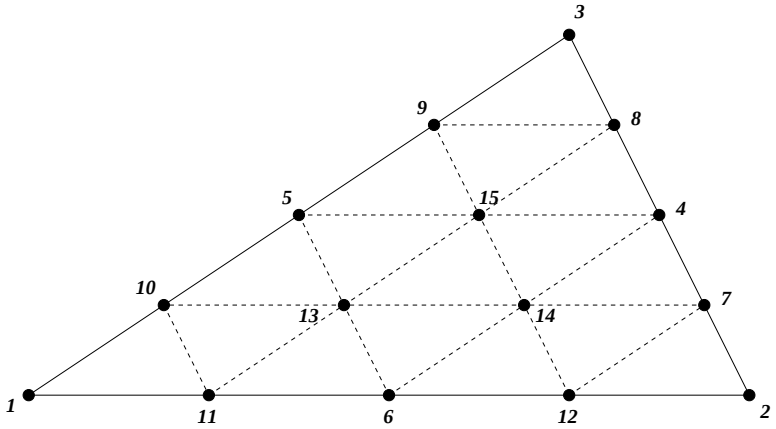

FIG. 3.1. *Node numbering for bases of degrees 1, 2 and 4.*

More precisely, on a triangle, the p-hierarchical basis of S_2 consists of linear functions $\psi_{\left(\frac{i}{1},\frac{j}{1},\frac{k}{1}\right)}$ associated with the nodes numbered $1,2,3$ in Figure 3.1, and the quadratic functions $\psi_{\left(\frac{i}{2},\frac{j}{2},\frac{k}{2}\right)}$ associated with the nodes numbered $4,5,6$. For S_4, we add to these the quartic functions $\psi_{\left(\frac{i}{4},\frac{j}{4},\frac{k}{4}\right)}$ associated with nodes 7 through 15. There are many systematic ways of generating hierarchical bases (cf. [1, 2, 29, 13]), and we have chosen $p = 2^m$ and the given splittings because of the convenient nesting of nodes and (sub-)spaces in these cases.

It is clear that the sparsity pattern of the stiffness matrix A and the error estimate (2.9) are independent of the choice of basis in the computation of $\hat{u} \in S_p$, but we argue that a hierarchical basis can provide *significant* gains in terms of solving the associated linear systems. In order to provide an intuitive sense of why hierarchical bases may be advantageous, we first consider the 6×6 element stiffness matrices for $V = S_2$ for both the standard Lagrange as well as the p-hierarchical basis. As before, the vertex unknowns are ordered first. Let $(\theta_1, \theta_2, \theta_3)$ denote the interior angles of the triangle T and define $s_j = \cot\theta_j$, $r_k = s_{k-1} + s_{k+1}$ and $s = s_1 + s_2 + s_3$. Then the element stiffness matrices are given by

$$A_T^{LB} = \frac{1}{6}\begin{pmatrix} 3r_1 & s_3 & s_2 & 0 & -4s_2 & -4s_3 \\ s_3 & 3r_2 & s_1 & -4s_1 & 0 & -4s_3 \\ s_2 & s_1 & 3r_3 & -4s_1 & -4s_2 & 0 \\ 0 & -4s_1 & -4s_1 & 8s & -8s_3 & -8s_2 \\ -4s_2 & 0 & -4s_2 & -8s_3 & 8s & -8s_1 \\ -4s_3 & -4s_3 & 0 & -8s_2 & -8s_1 & 8s \end{pmatrix}, \tag{3.1}$$

$$A_T^{HB} = \frac{1}{6}\begin{pmatrix} 3r_1 & -3s_3 & -3s_2 & -4r_1 & 4s_3 & 4s_2 \\ -3s_3 & 3r_2 & -3s_1 & 4s_3 & -4r_2 & 4s_1 \\ -3s_2 & -3s_1 & 3r_3 & 4s_2 & 4s_1 & -4r_3 \\ -4r_1 & 4s_3 & 4s_2 & 8s & -8s_3 & -8s_2 \\ 4s_3 & -4r_2 & 4s_1 & -8s_3 & 8s & -8s_1 \\ 4s_2 & 4s_1 & -4r_3 & -8s_2 & -8s_1 & 8s \end{pmatrix}. \tag{3.2}$$

These matrices are singular, with respective nullspaces $N(A_T^{LB}) = \text{span}\{(1,1,1,1,1,1)\}$ and $N(A_T^{LB}) = \text{span}\{(1,1,1,0,0,0)\}$, which leads to the global stiffness matrices being poorly conditioned. To put it another way, if the element stiffness matrices are

non-singular, it can be proved that the corresponding global stiffness matrix is spectrally equivalent to its diagonal (cf. [3, 19]), requiring no sophisticated machinery to solve the associated linear systems. In both cases, vectors in the nullspace represent (not surprisingly) functions which are locally constant. The key difference is *how* these constant functions are represented. In the first case, they are distributed evenly among the basis functions, and in the second they are concentrated in the linear basis functions. Suppose that the basei functions of $\mathcal{B}^{LB}$ and $\mathcal{B}^{HB}$ are ordered with those associated with vertices before those associated with edges. This induces the natural block structures

$$A^{LB} = \begin{pmatrix} A^{LB}_{11} & A^{LB}_{12} \\ A^{LB}_{21} & A^{LB}_{22} \end{pmatrix} \quad , \quad A^{HB} = \begin{pmatrix} A^{HB}_{11} & A^{HB}_{12} \\ A^{HB}_{21} & A^{HB}_{22} \end{pmatrix} \tag{3.3}$$

on the global stiffness matrices. It can be shown that both A^{LB}_{11} and A^{LB}_{22} are well-conditioned, so the ill-conditioning of A^{LB} is due to strong coupling by the off-diagonal blocks. In contrast, the ill-conditioning of A^{HB} is "concentrated" in A^{HB}_{11} while A^{HB}_{22} is well-conditioned and the off-diagonal coupling can be shown to be mild because of a *strong Cauchy inequality* in the H^1-inner-product between the spaces S_1 and $S_2 \setminus S_1$. We explain this further below.

Motivated by the discussion above, we consider a generic splitting of $V = S_p$, $V = V_1 \oplus V_2$, $V_1 \cap V_2 = \{0\}$. Choosing bases for V_1 and V_2, this splitting induces a natural 2×2 block structure for (2.6):

$$\begin{pmatrix} A_{11} & A_{21} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{pmatrix} \ . \tag{3.4}$$

Throughout, we we make the obvious identifications between coefficient vectors and functions:

$$\mathbf{v} \in \mathbb{R}^{|V|} \leftrightarrow v \in V \quad , \quad \mathbf{v}_1 \in \mathbb{R}^{|V_1|} \leftrightarrow v_1 \in V_1 \quad , \quad \mathbf{v}_2 \in \mathbb{R}^{|V_2|} \leftrightarrow v_2 \in V_2 \ ,$$

often without additional clarification. The block Gauss-Seidel iteration for solving $A\mathbf{u} = \mathbf{f}$ is described in linear-algebraic terms by the iteration

$$M\mathbf{u}^{k+1} = \mathbf{f} + N\mathbf{u}^k \quad , \quad M = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} \quad , \quad N = -\begin{pmatrix} 0 & A_{12} \\ 0 & 0 \end{pmatrix} \ , \tag{3.5}$$

and in terms of the bilinear form, by the iteration

$$B(u_1^{k+1}, v) = F(v) - B(u_2^k, v) \text{ for all } v \in V_1 \ , \tag{3.6}$$

$$B(u_2^{k+1}, v) = F(v) - B(u_1^{k+1}, v) \text{ for all } v \in V_2 \ . \tag{3.7}$$

When $B(\cdot,\cdot)$ is an inner-product, it can be shown (e.g. [14, 25, 3]) that there is a constant $\gamma = \gamma(B, V_1, V_2) \in [0, 1)$ for which

$$\max_{\substack{v_1 \in V_1, v_2 \in V_2 \\ v_1 \neq 0, v_2 \neq 0}} \frac{|B(v_1, v_2)|}{|\!|\!| v_1 |\!|\!| \, |\!|\!| v_2 |\!|\!|} \leq \gamma \ , \tag{3.8}$$

and the error in each block Gauss-Seidel iteration is guaranteed to contract by (at least) a factor of γ^2. This inequality is often called a *strong Cauchy inequality*, and it has played an important role in the analysis of multi-level linear solvers, as well

as hierarchical basis error estimation. The optimal constant γ is often estimated in terms of local quantities γ_T,

$$\gamma \leq \max_{T\in\mathcal{T}} \gamma_T \quad , \quad \max_{\substack{v_1\in V_1, v_2\in V_2 \\ v_1\neq 0, v_2\neq 0}} \frac{|B_T(v_1,v_2)|}{|\!|\!| v_1 |\!|\!|_T |\!|\!| v_2 |\!|\!|_T} \leq \gamma_T \ , \tag{3.9}$$

where $B_T(\cdot,\cdot)$ and $|\!|\!|\cdot|\!|\!|_T$ are defined by restricting to the element T the integrals defining $B(\cdot,\cdot)$ and $|\!|\!|\cdot|\!|\!|$. It is clear that, if v_1+v_2 is constant on T, with $v_j \in V_j$ and $v_j \neq 0$, then

$$\frac{\left|\int_T \nabla v_1 \cdot \nabla v_2 \, dx\right|}{|v_1|_{1,T}|v_2|_{1,T}} = 1 \ , \tag{3.10}$$

which indicates that one should require that locally constant functions should lie entirely in one these spaces, which we take to be V_1. In terms of the Laplacian example above, the local constant γ_T for A_T^{LB} is 1, and it can be shown that the global constant γ for A^{LB} approaches 1 as the mesh is refined. In contrast, it is well-known [25] that γ_T for A_T^{HB} is given by

$$\gamma_T^2 = \frac{1}{2} + \frac{1}{3}\sqrt{d-\frac{3}{4}} \quad , \quad d = \sum_{k=1}^{3} \cos^2\theta_k$$

where θ_k denote the interior angles of the triangle element. If the angles in T are non-obtuse, then $3/4 \leq d \leq 1$, so $1/2 \leq \gamma_T^2 \leq 2/3$. The optimal value of $1/2$ is achieved for equilateral triangles. For isoceles right triangles, $\gamma_T^2 = 2/3$.

At this stage we see that, at least in the case of symmetric problems, block Gauss-Seidel, either in the form presented or in its symmetric form, can provide an effective preconditioner for Krylov methods. A key benefit of hierarchical decompositions having $S_1 \subset V_1$ is that only the A_{11} block requires a "sophisticated" solver—such as a direct method, multi-grid method, or an $\mathcal{H}$-matrix method as used in this work—for each Gauss-Seidel iteration. The remaining block is spectrally-equivalent to its own diagonal, and therefore simple to handle.

We now argue that the initial step of the block Gauss-Seidel iteration provides error reduction which is *significantly* greater than what can be explained by these by now standard contraction arguments. These arguments follow the pattern given in [27], which were given in the context of hierarchical error estimation.

LEMMA 3.1. *Suppose that $u^0 = 0$, and $\phi = \phi_1 + \phi_2 \in V$. Then*

$$\|u-u^1\|_1 \lesssim \inf_{w\in V_2} \|u-u_1^1-w\|_1 \leq \inf_{\phi\in V} \left(\|u-\phi\|_1 + \|\phi_1 - u_1^1\|_1\right) \ .$$

The constant hidden in $\lesssim$ is determined by the constants of boundedness (continuity) and either coercivity or the inf-sup condition.

Proof. For the sake of clarity, let us assume that B is coercive, with constant m, and that M is its continuity constant. Using Galerkin orthogonality, we see that

$$m\|u-u^1\|_1 \leq B(u-u^1,u-u^1) = B(u-u^1,u-u^1-v) \leq M\|u-u^1\|_1\|u-u^1-v\|_1 \ ,$$

for all $v \in V_2$. Setting $w = u_2^1 + v$ establishes the first inequality. Now, for any $\phi = \phi_1 + \phi_2$, we see that

$$\inf_{w\in V_2} \|u-u_1^1-w\|_1 = \inf_{w\in V_2} \|u-\phi+\phi_1-u_1^1+\phi_2-w\|_1 \leq \|u-\phi\|_1 + |\!|\!|\phi_1-u_1^1|\!|\!| \ ,$$

which completes the proof. □

To obtain a practical estimate from Lemma 3.1, one generally chooses ϕ so that it and ϕ_1 are both suitable (quasi-)interpolants of u. Our arguments below employ super-convergence results of the sort found in [4, 31, 22], which in turn assume a mild approximate mesh symmetry condition. Although the results we present below allow for Neumann boundary conditions, for simplicity the mesh condition below is stated for homogeneous Dirichlet boundary conditions—the more technical condition for boundary edges/vertices can be found in [4] .

DEFINITION 3.2 ($O(h^{2\sigma})$-irregular Triangulation). *Let $\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2$ denote the set of interior edges in $\mathcal{T} = \mathcal{T}_h$. For each $e \in \mathcal{E}_1$, the adjacent triangles T and T' form an approximate parallelogram, in which the opposide sides differ in length by $\mathcal{O}(h^2)$. For $e \in \mathcal{E}_2$, we have the overall bound*

$$\sum_{e \in \mathcal{E}_2} (|T| + |T'|) = \mathcal{O}(h^{2\sigma}).$$

THEOREM 3.3. *Suppose that $\mathcal{T} = \mathcal{T}_h$ satisfies the approximate mesh symmetry condition above, and u is sufficiently regular. In the case $V_1 = S_1$ and $V = S_p$ for some $p > 1$, the standard nodal quadratic interpolant $u_q = u_\ell + u_b$, $u_\ell \in S_1$, $u_b \in S_2 \setminus S_1$, satisfies*

$$\|u - u^1\|_1 \lesssim \|u - u_q\|_1 + \|u_1^1 - u_\ell\|_1 \lesssim h^{1+\min(1,\sigma)} |\log h|^{1/2} \, \|u\|_{W^{3,\infty}(\Omega)} \, . \tag{3.11}$$

In the case $V_1 = S_2$ and $V = S_p$ for some $p > 2$, there is a cubic quasi-interpolant $u_c = u_q + u_w$, $u_q \in S_2$ (not the nodal interpolant), $u_w \in S_3 \setminus S_2$ such that

$$\|u - u^1\|_1 \lesssim \|u - u_c\|_1 + \|u_1^1 - u_q\|_1 \lesssim h^{2+\min(1/2,\sigma)} (\|u\|_4 + |u|_{W^{3,\infty}(\Omega)}) \, . \tag{3.12}$$

Proof. See [4] for (3.11) and [22] for (3.12). □

The results of primary interest from [4, 22] concern what we have denoted as $\|u_1^1 - \phi_1\|_1$ in Lemma 3.1. In both cases $u_1^1 = u_h \in V_1$ is the finite element solution. Those comparing [4] and [22] should note that what is called 2σ here and in [4] is called σ in [22]. Careful reading of these papers reveals when one can (slightly) relax the regularity assumptions in their proofs, and under what conditions optimal convergence rates can be expected. In brief, it is not unreasonable in practice to expect to see $\|u - u^1\|_1 = \mathcal{O}(h^2)$ or $\|u - u^1\|_1 = \mathcal{O}(h^2 |\log h|^{1/2})$ when $V_1 = S_1$, and $\|u - u^1\|_1 = \mathcal{O}(h^2)$ or $\|u - u^1\|_1 = \mathcal{O}(h^2 |\log h|^{1/2})$ when $V_1 = S_2$. Even in more realistic situations in which u possesses a few isolated singularities, and the triangulation is appropriately refined near those singularities, similar convergence is observed when one replaces h with $N^{-1/2}$, where N is the number of vertices, edges, or triangles. A partial explanation of this phenomenon, in the case of the Dirichlet Laplacian, is provided in [10]. The proofs of such superconvergence properties are generally quite technical and so have been limited to lower-order finite elements. One difficulty in extending these arguments to higher-order elements is finding a proper choice of quasi-interpolant—it is known that the standard Lagrange interpolant does not yield the desired superconvergence properties for higher-order elements [24]. Finally, it is clear that Lemma 3.1 is dimension-independent, so any superconvergence result of the sort above which holds in $\mathbb{R}^n$, would imply a similarly large first step in the corresponding block Gauss-Seidel iteration. One such superconvergence result for mildly structured tetrahedral meshes in $\mathbb{R}^3$ is provided in [9].

4. Concerning Hierarchical Matrices. The block Gauss-Seidel iteration (3.5) requires the solution of systems of equations involving the diagonal blocks A_{11} and A_{22}. In the numerical tests in the following section 5, both systems will be solved only approximately, i.e., they will be replaced by approximations $\widetilde{A_{11}}$ and $\widetilde{A_{22}}$. The 2×2 splitting of the matrix results in an ill-conditioned block A_{11} and a well-conditioned block A_{22}. We therefore propose to construct $\widetilde{A_{11}}$ as a (highly accurate) approximate Cholesky factorization using the technique of hierarchical ($\mathcal{H}$-)matrices to be described in this section. Since A_{22} is well-conditioned, it can be replaced by an approximation $\widetilde{A_{22}}$ representing a few steps of an inner iteration.

In the remainder of this section, we will specify the particular variant of $\mathcal{H}$-matrix construction which is proposed for the $\mathcal{H}$-Cholesky (or $\mathcal{H}$-LU factorization) of the A_{11}-block, and point to the literature for further details on $\mathcal{H}$-matrices.

$\mathcal{H}$-matrices were introduced in [20] and since then entered into a wide range of applications. The basic $\mathcal{H}$-matrix construction and corresponding arithmetic have reached a relatively mature state and are documented in the comprehensive lecture notes [7] and books [6, 21].

$\mathcal{H}$-matrices are based on a hierarchical subdivision of the matrix into subblocks and low-rank approximations of matrix data within (most of) these subblocks. Originally, $\mathcal{H}$-matrices were introduced in the context of fully populated matrices arising from solution operators of elliptic differential equations and in boundary element methods. In the finite element context, the stiffness matrix itself does not require an efficient approximation by an $\mathcal{H}$-matrix since it is sparse. Its Cholesky- or LU-factors, however, suffer from inacceptable fill-in if computed exactly and can be efficiently computed or approximated by $\mathcal{H}$-Cholesky or $\mathcal{H}$-LU factors, resp. [18, 5]. The $\mathcal{H}$-matrix construction and arithmetic as originally developed for fully populated matrices have a straightforward generalization to sparse matrices. However, there are two modifications for $\mathcal{H}$-matrices which have been designed for sparse matrices in particular and have been used here.

The first modification concerns the construction of the block structure of the $\mathcal{H}$-matrix. In the classical $\mathcal{H}$-matrix, the block structure is generated through a repeated bisection of the respective index sets, i.e., row and column index sets are divided into two subsets, respectively, which leads to four matrix subblocks. In the case of sparse matrices, the bisection has been replaced by a nested dissection approach in which row and column index sets are divided into three subsets each; two subsets S_1, S_2 of indices that are pairwise disconnected in the sense that $a_{ij} = 0 = a_{ji}$ if $i \in S_1$, $j \in S_2$ where $A = (a_{ij})$ denotes the stiffness matrix, and a third subset S_3 containing the remaining indices of the interior boundary. Such a subdivision results in a 3×3 matrix structure with zero blocks in the 1×2 and 2×1 positions. These zero blocks remain zero in a subsequent Cholesky or LU factorization which results in considerably faster ($\mathcal{H}$-)Cholesky factorizations compared to bisection-based $\mathcal{H}$-matrices [23, 17]. The blocks are subdivided recursively until a minimum blocksize $nmin$ is reached, i.e., a block is further subdivided if the minimum of its number of rows and columns is greater than $nmin$. In the subsequent numerical results in section 5, we set $nmin = 8$.

The second modification concerns the development of a "blackbox" clustering algorithm. The classical construction of $\mathcal{H}$-matrices requires geometric information associated with the underlying indices in order to determine a suitable block structure. For sparse matrices, the information contained in the associated matrix graph can replace the need for geometric information [16].

Whereas the classical $\mathcal{H}$-matrix uses a fixed rank for the low rank approximations

within matrix subblocks, it is possible to replace it by *adaptive ranks* in order to enforce a desired accuracy within the individual blocks. In particular, given a matrix block C and a desired $\mathcal{H}$-accuracy $0 < \delta_{\mathcal{H}} < 1$, we set the rank k_C of the approximation to C as

$$k_C := \min\{k' \mid \sigma_{k'} \leq \delta_{\mathcal{H}} \sigma_1\}$$

where σ_i denotes the i'th largest singular value of C. These modifications from the classical $\mathcal{H}$-matrix setting have led to highly efficient $\mathcal{H}$-Cholesky/$\mathcal{H}$-LU preconditioners for a wide range of sparse matrices [15].

5. Numerical Experiments. As a model problem, we consider the Laplace problem

$$-\Delta u = f \text{ in } \Omega = (-1,1) \times (-1,1) \text{ with } u = 0 \text{ on } \partial\Omega \; . \tag{5.1}$$

We choose the right hand side function f so that the exact solution is known to be

$$u = \cos\left(\frac{\pi}{2}x\right) \cos\left(\frac{\pi}{2}y\right) \tag{5.2}$$

with semi-norms

$$|u|_k^2 = \int_\Omega \sum_{j=0}^{k} \binom{k}{j} \left(\frac{\partial^k u}{\partial x^{k-j} \partial y^j}\right)^2 = \frac{\pi^{2k}}{2^k} \; .$$

We will use finite element discretizations of polynomial degrees 2 and 4, resp., with hierarchical bases, leading to the 2×2 block structure (3.4). Here, in the quadratic case, the first diagonal block contains the coupling of the linear basis functions whereas A_{22} contains the coupling of the quadratic basis functions. In the quartic case, there are different options for the splitting into two blocks; we will provide numerical tests where A_{11} contains the linear-linear coupling only as well as A_{11} including both the linear and quadratic basis functions.

In the following, we will provide numerical results which can roughly be grouped into two "themes", the first theme is a more theoretical perspective of convergence rates, whereas the second is a more practical view on iteration times:

- Our first set of tests (Fig. 5.1 through 5.5) confirms the theoretical result of Lemma 3.1, showing the predicted massive error reduction in the first iteration step, reaching discretization error in the case of quadratic elements. Here, we use the classical linear (forward) block Gauss-Seidel method (i.e., not accelerated by a Krylov method), and we vary the accuracies of our inner solvers for the diagonal blocks A_{11} and A_{22}.
- In our second set of experiments (Fig. 5.6 and 5.7), we will use a block Gauss-Seidel preconditioned GMRES method and show the iteration times necessary to reach a desired reduction in the residual for various accuracies of our inner solvers.

All numerical tests have been performed on a Dell 690n workstation (2.33GHz, 32GB memory) using the standard $\mathcal{H}$-matrix library HLIB (cf. `http://www.hlib.org`), and we always choose the zero vector as the initial iterate—which is not optional but mandatory to obtain the substantial decrease in the first iteration step.

In all tests, we will approximate the ill-conditioned block A_{11} by an $\mathcal{H}$-Cholesky factorization, resp.. Table 5.1 shows the set-up times and storage requirements for

such $\mathcal{H}$-Cholesky factorizations for various adaptive $\mathcal{H}$-accuracies $\delta_{\mathcal{H}}$. Both the setup times and storage requirements increase almost linearly in the problem size. The set-up times will be put into perspective and be compared with the iteration times in later experiments.

TABLE 5.1
Setup time (in sec.) and storage requirements (in MB) of $\mathcal{H}$-Cholesky factorizations of A_{11} for linear elements.

n_l	setup time (seconds)			storage (MB)		
$\delta_{\mathcal{H}}$	$1e-2$	$1e-4$	$1e-8$	$1e-2$	$1e-4$	$1e-8$
9,801	0.7	0.8	0.8	9	9	10
19,321	1.5	1.7	1.8	18	19	20
39,601	3.5	4.0	4.4	28	40	43
77,841	7.1	8.2	9.2	76	81	85
159,201	16	19	22	161	172	182
312,481	35	38	42	319	341	363
638,401	71	83	94	659	705	751

In Figures 5.1 and 5.3, we show the energy norm of the iteration error for quadratic and quartic finite elements, resp., on a log-log scale, i.e., given the exact solution u (5.2) and the iterative solution $\hat{u}^i \in V$ as determined by its coefficient vector $u^i \in \mathbb{R}^N$, we approximate the energy norm $\|e^i\|$ of the iteration error $e^i = u - u^i$ using a 25-point quadrature on each triangle in Ω_h. We show results for various accuracies for

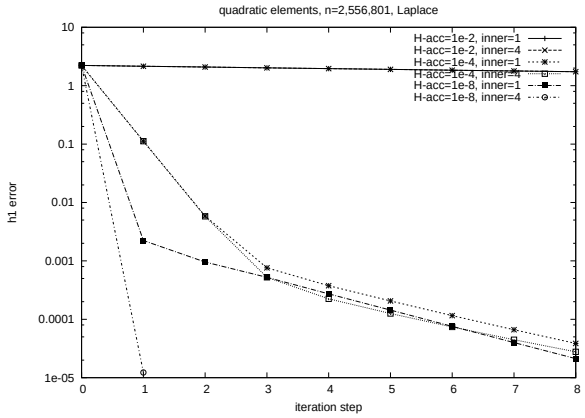

FIG. 5.1. *H_1-error per iteration step for quadratic elements*

the solvers for the diagonal blocks A_{11} and A_{22}. The block A_{11} is replaced by its $\mathcal{H}$-Cholesky factors using $\mathcal{H}$-accuracies 1e-2, 1e-4 and 1e-8, resp.. The block A_{22} is either replaced by its upper triangular part (backward Gauss-Seidel preconditioner) - which is denoted by "inner=1", or by (4 steps of) an inner (Jacobi preconditioned) GMRES iteration - which is denoted by "inner=4". The results in Fig. 5.1 show that an insufficient accuracy for solving the first diagonal block (i.e., H-acc=1e-2), does not lead to a successful method. When we solve the first diagonal block with an accuracy of H-acc=1e-4 (or better), we clearly see the big decrease in the first iteration step. If we solve both diagonal blocks highly accurately, i.e., H-acc=1e-8 for A_{11} and four steps of an inner GMRES method to solve for A_{22}, then the iteration error is

essentially of the same order as the discretization error within the first iteration step which is also supported by the following Figure 5.2. It illustrates the energy norm of the iteration error after the first step for various problem sizes. Here, both diagonal blocks have been solved very accurately (H-acc=1e-8, inner=4 and inner=8, resp.).

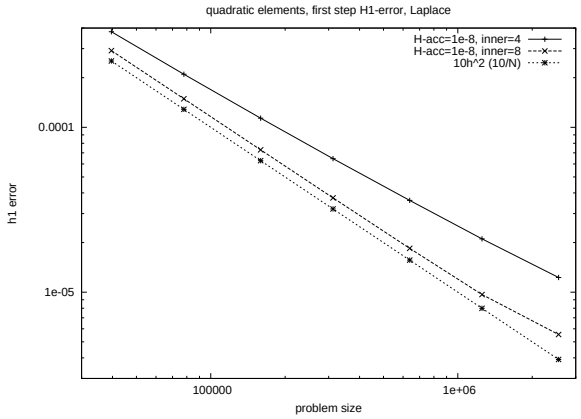

FIG. 5.2. *H_1-error after the first iteration step for quadratic elements*

In Fig. 5.3, we repeat this set of tests using quartic finite elements in the discretization (and up to 8 steps in the inner GMRES iteration for A_{22}). Initially, A_{11} contains the coupling of linear basis functions whereas A_{22} contains both quadratic and quartic basis functions. We see a big decrease of the error in the first iteration step as long as we choose an $\mathcal{H}$-accuracy H-acc of 1e-4 or smaller for $\widetilde{A_{11}}$. If the A_{22} block is not solved accurately (i.e., inner=1), then the better accuracy for A_{11} does not help to improve the convergence. In fact, the results for H-acc=1e-4, inner=1 and H-acc=1e-8, inner=1, are identical (the filled circle within the empty square almost appears to be a filled square).

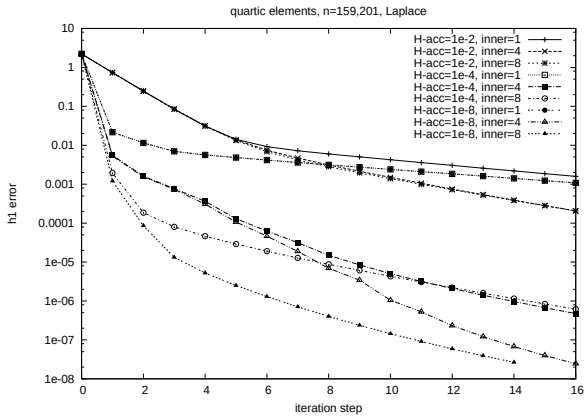

FIG. 5.3. *H_1-error per iteration step for quartic elements*

While the first step is quite large, Fig. 5.4 shows that this step is only of order $\mathcal{O}(h^2)$, i.e., the same as for quadratic elements (here, the problem sizes are N=150201,

312481 and 638401, resp.). This is in agreement with Lemma 3.1. In order to achieve a higher order of convergence in the first step, we simply regroup our basis functions to include the quadratic functions with the linear functions in block A_{11}. In this case, we see an error of $\mathcal{O}(h^3)$ after the first step which is also shown in Fig. 5.4 (curve with filled squares).

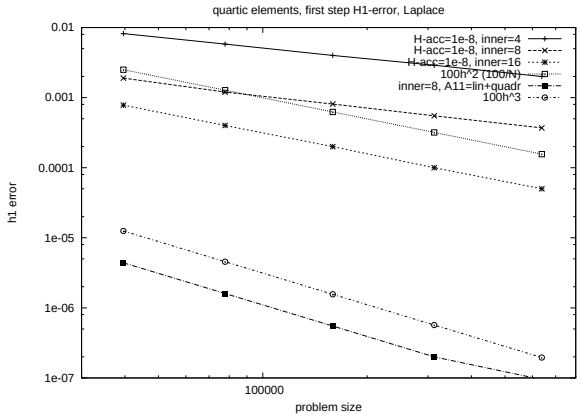

FIG. 5.4. *H_1-error after the first iteration step for quartic elements*

We have also performed experiments for a convection-diffusion problem

$$-\epsilon\Delta u + \mathbf{b}\cdot\nabla u = f \text{ in } \Omega = (-1,1)\times(-1,1) \text{ with } u = 0 \text{ on } \partial\Omega \tag{5.3}$$

with a recirculating convection

$$\mathbf{b}(x,y) = \begin{pmatrix} \frac{0.1e^{0.1y}}{2\pi(e^{0.1}-1)} \sin\left(\frac{2\pi(e^{0.1y}-1}{e^{0.1}-1}\right)\left(1-\cos\left(\frac{2\pi(e^{4x}-1}{e^4-1}\right)\right) \\ -\frac{4e^{4x}}{2\pi(e^4-1)} \sin\left(\frac{2\pi(e^{4x}-1}{e^4-1}\right)\left(1-\cos\left(\frac{2\pi(e^{0.1y}-1}{e^{0.1}-1}\right)\right) \end{pmatrix}.$$

Moderate convection $\epsilon \geq 10^{-3}$ in (5.3) hardly causes any difference in the iteration errors at all, and results very similar to those in Fig. 5.1 are shown in Figure 5.5 (strong convection $\epsilon \leq 10^{-4}$ led to meaningless results since we did not implement any upwinding in our discretization).

Finally, we performed tests using the block Gauss-Seidel preconditioned GMRES method. In the following tests, A_{11} is again approximated by its $\mathcal{H}$-Cholesky factors (of varying accuracy), and A_{22} is simply replaced by its lower triangular part. We did not perform inner iteration for A_{22} within the solver since the above approximation leads to quite good results already so that we do not expect any (timewise) improvements through inner iterations. In addition to the setup times for the $\mathcal{H}$-Cholesky factorization of A_{11}, we show the time necessary to reduce the initial residual by a factor of 1e-6. Figure 5.6 shows the results for quadratic elements while Fig. 5.7 shows results for quartic elements. In either case we observe a significant acceleration when we improve the $\mathcal{H}$-accuracy for the $\mathcal{H}$-Cholesky factorization of A_{11} from 1e-2 down to 1e-4. A further improvement down to 1e-8 yields only a slight improvement for quadratic elements and none for quartic elements. Whereas the setup times dominate the actual iteration times for quadratic elements, the setup and iteration times are quite similar for quartic elements.

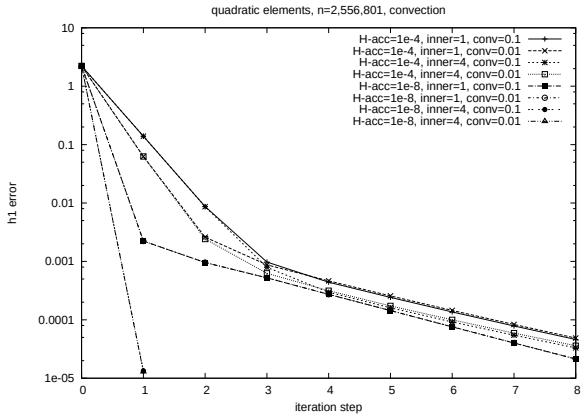

FIG. 5.5. *H_1-error convergence history for the convection diffusion problem*

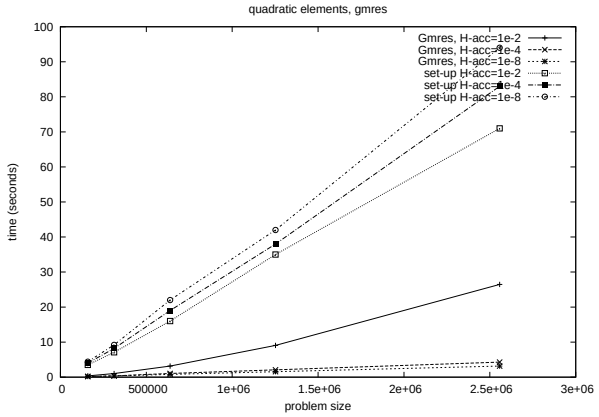

FIG. 5.6. *for quadratic elements*

REFERENCES

[1] I. Babuška, I. N. Katz, B. A. Szabo, and A. P. Greensfelder. Hierarchic families for the *p*-version of the finite element method. In *Advances in computer methods for partial differential equations, III (Proc. Third IMACS Internat. Sympos., Lehigh Univ., Bethlehem, Pa., 1979)*, pages 278–286. IMACS, New Brunswick, N.J., 1979.

[2] I. Babuška, B. A. Szabo, and I. N. Katz. The *p*-version of the finite element method. *SIAM J. Numer. Anal.*, 18(3):515–545, 1981.

[3] R. E. Bank. Hierarchical bases and the finite element method. In *Acta numerica, 1996*, volume 5 of *Acta Numer.*, pages 1–43. Cambridge Univ. Press, Cambridge, 1996.

[4] R. E. Bank and J. Xu. Asymptotically exact a posteriori error estimators. I. Grids with superconvergence. *SIAM J. Numer. Anal.*, 41(6):2294–2312 (electronic), 2003.

[5] M. Bebendorf. Why finite element discretizations can be factored by triangular hierarchical matrices. *SIAM Num. Anal.*, 45:1472–1494, 2007.

[6] M. Bebendorf. *Hierarchical Matrices. A Means to Efficiently Solve Elliptic Boundary Value Problems*, volume 63 of *Lecture Notes in Computational Science and Engineering*. Springer, 2008.

[7] S. Börm, L. Grasedyck, and W. Hackbusch. Hierarchical matrices, 2003. Lecture Notes No. 21, Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany, available online at www.mis.mpg.de/preprints/ln/, revised version June 2006.

[8] F. A. Bornemann and P. Deuflhard. The cascadic multigrid method for elliptic problems.

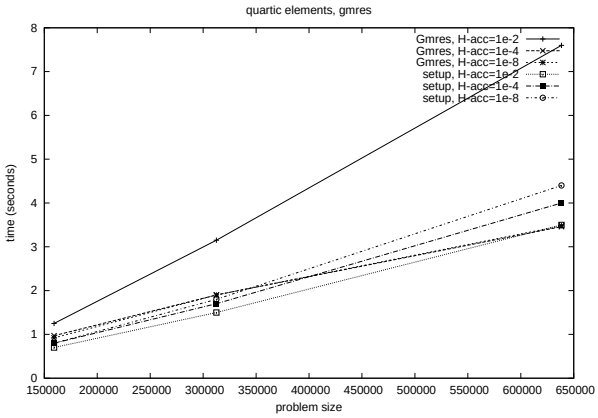

FIG. 5.7. *for quartic elements*

Numer. Math., 75(2):135–152, 1996.

[9] L. Chen. Superconvergence of tetrahedral linear finite elements. *Int. J. Numer. Anal. Model.*, 3(3):273–282, 2006.

[10] L. Chen and H. Li. Superconvergence of gradient recovery schemes on graded meshes for corner singularities. *J. Comput. Math.*, 28(1):11–31, 2010.

[11] P. Deuflhard. Cascadic conjugate gradient methods for elliptic partial differential equations: algorithm and numerical results. In *Domain decomposition methods in scientific and engineering computing (University Park, PA, 1993)*, volume 180 of *Contemp. Math.*, pages 29–42. Amer. Math. Soc., Providence, RI, 1994.

[12] P. Deuflhard, P. Leinen, and H. Yserentant. Concepts of an adaptive hierarchical finite element code. *IMPACT Comput. Sci. Eng.*, 1(1):3–35, 1989.

[13] C. Doucet, I. Charpentier, J.-L. Coulomb, and C. Guerin. Construction of hierarchic families of finite elements by localization of their degrees of freedom. *Magnetics, IEEE Transactions on*, 45(3):1304 –1307, march 2009.

[14] V. Eijkhout and P. Vassilevski. The role of the strengthened Cauchy-Buniakowskiĭ-Schwarz inequality in multilevel methods. *SIAM Rev.*, 33(3):405–419, 1991.

[15] L. Grasedyck, W. Hackbusch, and R. Kriemann. Performance of $\mathcal{H}$-LU preconditioning for sparse matrices. *Comput. Methods Appl. Math.*, 8(4):336–349, 2008.

[16] L. Grasedyck, R. Kriemann, and S. Le Borne. Parallel black box $\mathcal{H}$-LU preconditioning for elliptic boundary value problems. *Comput. Vis. Sci.*, 11(4-6):273–291, 2008.

[17] L. Grasedyck, R. Kriemann, and S. Le Borne. Domain decomposition based $\mathcal{H}$-LU preconditioning. *Numerische Mathematik*, 112:565–600, 2009.

[18] L. Grasedyck and S. Le Borne. $\mathcal{H}$-matrix preconditioners in convection-dominated problems. *SIAM J. Mat. Anal.*, 27:1172–1183, 2006.

[19] L. Grubišić and J. S. Ovall. On estimators for eigenvalue/eigenvector approximations. *Math. Comp.*, 78:739–770, 2009.

[20] W. Hackbusch. A sparse matrix arithmetic based on $\mathcal{H}$-matrices. Part I: Introduction to $\mathcal{H}$-matrices. *Computing*, 62:89–108, 1999.

[21] W. Hackbusch. *Hierarchische Matrizen*. Springer, 2009. (in German).

[22] Y. Huang and J. Xu. Superconvergence of quadratic finite elements on mildly structured grids. *Math. Comp.*, 77(263):1253–1268, 2008.

[23] I. Ibragimov, S. Rjasanow, and K. Straube. Hierarchical Cholesky decomposition of sparse matrices arising from curl-curl-equations. *J. Numer. Math.*, 15:31–58, 2007.

[24] B. Li. Lagrange interpolation and finite element superconvergence. *Numer. Methods Partial Differential Equations*, 20(1):33–59, 2004.

[25] J.-F. Maitre and F. Musy. The contraction number of a class of two-level methods; an exact evaluation for some finite element subspaces and model problems. In *Multigrid methods (Cologne, 1981)*, volume 960 of *Lecture Notes in Math.*, pages 535–544. Springer, Berlin, 1982.

[26] W. F. Mitchell. The *hp*-multigrid method applied to *hp*-adaptive refinement of triangular grids. *Numer. Linear Algebra Appl.*, 17(2-3):211–228, APR 2010.

[27] J. S. Ovall. Function, gradient, and Hessian recovery using quadratic edge-bump functions. *SIAM J. Numer. Anal.*, 45(3):1064–1080 (electronic), 2007.

[28] V. V. Shaidurov. Some estimates of the rate of convergence for the cascadic conjugate-gradient method. *Comput. Math. Appl.*, 31(4-5):161–171, 1996. Selected topics in numerical methods (Miskolc, 1994).

[29] B. Szabó and I. Babuška. *Finite element analysis*. A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1991.

[30] M. A. Taylor, B. A. Wingate, and R. E. Vincent. An algorithm for computing Fekete points in the triangle. *SIAM J. Numer. Anal.*, 38(5):1707–1720 (electronic), 2000.

[31] J. Xu and Z. Zhang. Analysis of recovery type a posteriori error estimators for mildly structured grids. *Math. Comp.*, 73(247):1139–1152 (electronic), 2004.