Journal of Computational Mathematics
Vol.27, No.5, 2009, 624–641.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2009.27.5.014

PARALLEL HIERARCHICAL MATRIX PRECONDITIONERS FOR THE CURL-CURL OPERATOR*

Mario Bebendorf
Institut für Numerische Simulation, Universität Bonn, Wegelerstrasse 6, 53115 Bonn, Germany
Email:bebendorf@ins.uni-bonn.de
Joerg Ostrowski
ABB Switzerland Ltd, Corporate Research, Segelhofstrasse 1 K / P.O. Box, CH-5405 Baden 5 Dättwil
Email:joerg.ostrowski@ch.abb.com

Abstract

This paper deals with the preconditioning of the curl-curl operator. We use $\mathbf{H}(\mathbf{curl})$-conforming finite elements for the discretization of our corresponding magnetostatic model problem. Jumps in the material parameters influence the condition of the problem. We will demonstrate by theoretical estimates and numerical experiments that hierarchical matrices are well suited to construct efficient parallel preconditioners for the fast and robust iterative solution of such problems.

Mathematics subject classification: 65N22, 65F10, 65N30
Key words: Computational electromagnetism, Preconditioners, Hierarchical matrices, Parallelization.

1. Introduction

A major field of occurrence of the curl-curl operator is computational electromagnetism. An example is the ungauged vector potential based magnetostatic problem

$$\mathbf{curl}\,\frac{1}{\mu}\mathbf{curl}\,\mathbf{u} = \mathbf{j}_0 \quad \text{in } \Omega, \tag{1.1a}$$

$$\mathbf{u_t} = 0 \quad \text{on } \partial\Omega, \tag{1.1b}$$

which we choose as our model problem with given source current $\mathbf{j}_0$. For theoretical purposes we assume that $\Omega \subset \mathbb{R}^3$ is a convex (curved) polyhedron, while in pratice this property does not seem to be required. We denote $\mathbf{n}$ as the exterior normal at the boundary $\partial\Omega$ and $\mathbf{u_t} := \mathbf{u} \times \mathbf{n}$ as the tangential surface trace of the vector potential $\mathbf{u}$. A typical setting to be simulated in magnetostatics is shown in Fig. 7.1. The computational domain $\Omega = \Omega_C \cup \Omega_I$ consists of different materials that are characterized by their material parameters, i.e. their *conductivity* σ satisfying $0 \le \sigma(x) \le \sigma_1$ and their *magnetic permeability* $\mu := \mu_r \cdot \mu_0 \in L^\infty(\Omega)$ with $1 \le \mu_r(x) \le \mu_1/\mu_0$ for some constants $\mu_1, \sigma_1 \in \mathbb{R}$ and $\mu_0 := 4\pi \cdot 10^{-7}$ (Vs)/(Am). $\mathbf{j}_0$ vanishes in the non-conducting domain Ω_I.

The ungauged magnetostatic problem is singular, because any gradient field $\mathbf{grad}\,\phi$ can be added to the solution. The magnetic field $\mathbf{B} := \mathbf{curl}\,\mathbf{u}$, which is the measurable physical quantity of interest, is not affected by this alternative solution $\mathbf{u}_{\text{new}} := \mathbf{u} + \mathbf{grad}\,\phi$. The vector potential $\mathbf{u}$ itself is not a measurable physical quantity.

* Received March 1, 2008 / Revised version received October 14, 2008 / Accepted February 5, 2009 /

The ungauged vector potential based magnetostatic problem is a special case of the vector potential based full Maxwell problem in frequency domain and temporal gauge

$$\mathbf{curl}\frac{1}{\mu}\mathbf{curl}\,\mathbf{u} + \mathrm{i}\omega(\sigma + \mathrm{i}\omega\epsilon)\mathbf{u} = -(\sigma + \mathrm{i}\omega\epsilon)\mathbf{grad}\,\varphi_0. \tag{1.2}$$

Herein, the electric permittivity is assigned by $\epsilon := \epsilon_r \cdot \epsilon_0$ with $\epsilon_r \geq 1$, $\epsilon_0 := 8.85 \cdot 10^{-12}$ (As)/(Vm) and φ_0 denotes the scalar electric potential. The magnetostatic equation (1.1a) follows from (1.2) in the case of vanishing angular frequency $\omega = 0$. The operator that arises from the full Maxwell problem is regular for all $\omega > 0$, whereas the curl-curl operator has a large kernel. Since the electromagnetic fields that emerge at low frequencies in the full Maxwell model are a good approximation of magnetostatic fields, it can be expected that the operator (1.2) at small frequencies is a good approximation for the operator (1.1). It is therefore an obvious idea to regularize the "magnetostatic operator" by adding a multiple of the identity. Hence, we consider the operator

$$\mathbf{L}_\alpha := \mathbf{curl}\frac{1}{\mu}\mathbf{curl} + \alpha\mathbf{I} \tag{1.3}$$

with constant $1/\mu_1 \leq \alpha \in \mathbb{R}$ as a preconditioner for the magnetostatic operator

$$\mathbf{L}_0 = \mathbf{curl}\frac{1}{\mu}\mathbf{curl}.$$

One of the most established methods for the iterative solution of electromagnetic problems are multigrid methods; see [1, 16]. Algebraic multigrid methods (AMG) can be applied if no finite element (FE) grid hierarchy is available; see [26] and [8] for an improved version. However, they lack a comprehensive theoretical analysis. A major difference of the method in [26] and the method presented in this article is that we do not regularize the problem itself. We rather use the regularized operator for generating preconditioners for the original problem (1.1), while in [26] an approximate solution which depends on the regularization parameter α is computed. See [18] for a preconditioning technique that relies on solvers for the discrete Poisson problem. In this article we propose the usage of hierarchical matrices ($\mathcal{H}$-matrices) [14, 15] due to their robustness with respect to non-smooth coefficients in the differential operator.

Hierarchical matrices provide a setting in which approximations of fully populated matrices (such as the inverse or the factors of the LU decomposition of sparse matrices) can be computed with logarithmic-linear complexity. The existence of such approximations in the case of FE discretizations was proved in [2, 4, 7] for general scalar elliptic boundary value problems. A strategy that is also based on $\mathcal{H}$-matrices was proposed in [27]. There, the discretization A of $\mathbf{L}_0$ is regularized by adding UU^T to A, where U is the matrix consisting of the kernel vectors of A (the so-called *discrete grad-div regularization*).

After the introduction of appropriate spaces and the variational formulation of our problem in Sect. 2, we will review hierarchical matrices in Sect. 3 in the context of nested dissection reorderings; see [10]. In Sect. 4 we will lay theoretical ground to the $\mathcal{H}$-matrix approximation of the factors of the LU decomposition in the case of discretizations of the operator (1.3) with Nédélec's edge elements [24]. The regularization (1.3) guarantees that low-precision LU factorizations can be computed in the methodology of hierarchical matrices, which can be used for preconditioning. In Sect. 5 we will investigate the influence of the regularization parameter α and the accuracy $\varepsilon_{\mathcal{H}}$ of the hierarchical matrix approximation on the condition number of the preconditioned problem. In Sect. 6 it will be shown how the nested dissection structure of the hierarchical LU decomposition can be exploited for parallelization. Finally, Sect. 7 will

contain numerical results which document the applicability of the presented methods. It will be seen that the preconditioner can be computed with almost linear complexity and that the number of iterations is bounded by a constant.

2. Variational Formulation

Let $\mathbf{L}^2(\Omega) = [L^2(\Omega)]^3$, $\mathbf{H}^1(\Omega) := [H^1(\Omega)]^3$, and $\mathbf{H}^1_{\rm loc}(\Omega) := [H^1_{\rm loc}(\Omega)]^3$. We define the spaces

$$\mathbf{H}(\mathbf{curl};\Omega) := \{\mathbf{v} \in \mathbf{L}^2(\Omega) : \mathbf{curl}\,\mathbf{v} \in \mathbf{L}^2(\Omega)\}, \quad \mathbf{H}_0(\mathbf{curl};\Omega) := \{\mathbf{v} \in \mathbf{H}(\mathbf{curl};\Omega) : \mathbf{v_t} = 0 \text{ on } \partial\Omega\}$$

equipped with the norm $\|\mathbf{u}\|^2_{\mathbf{H}(\mathbf{curl};\Omega)} := \|\mathbf{u}\|^2_{\mathbf{L}^2(\Omega)} + \|\mathbf{curl}\,\mathbf{u}\|^2_{\mathbf{L}^2(\Omega)}$. Testing equation (1.1a) and using the integration by parts formula

$$\int_\Omega \mathbf{v}\cdot\mathbf{curl}\,\mathbf{w} - \mathbf{w}\cdot\mathbf{curl}\,\mathbf{v}\,\mathrm{d}x = \int_{\partial\Omega} \mathbf{v_t}\cdot\mathbf{w}\,\mathrm{d}s$$

leads to the variational formulation of our magnetostatic model problem: search $\mathbf{u} \in \mathbf{H}_0(\mathbf{curl};\Omega)$ such that

$$(\frac{1}{\mu}\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} = (\mathbf{j}_0, \mathbf{v})_{\mathbf{L}^2(\Omega)} \quad \text{for all } \mathbf{v} \in \mathbf{H}_0(\mathbf{curl};\Omega). \tag{2.1}$$

The variational problem (2.1) is not uniquely solvable. The infinite dimensional kernel of the operator consists of all functions in

$$\mathbf{H}_0(\mathbf{curl}\,0;\Omega) := \{\mathbf{v} \in \mathbf{H}_0(\mathbf{curl};\Omega) : \mathbf{curl}\,\mathbf{v} = 0\}.$$

For domains of simple topology the orthogonal complement of $\mathbf{H}_0(\mathbf{curl}\,0;\Omega)$ is the set

$$\mathbf{Z}_0(\Omega) := \{\mathbf{v} \in \mathbf{H}_0(\mathbf{curl};\Omega) : (\mathbf{v}, \mathbf{grad}\,\varphi)_{\mathbf{L}^2(\Omega)} = 0 \text{ for all } \varphi \in H^1_0(\Omega)\}.$$

There it holds a Poincaré-Friedrichs type of inequality

$$\|\mathbf{u}\|_{\mathbf{L}^2(\Omega)} \le c\|\mathbf{curl}\,\mathbf{u}\|_{\mathbf{L}^2(\Omega)} \quad \text{for all } \mathbf{u} \in \mathbf{Z}_0(\Omega) \tag{2.2}$$

with a constant $c > 0$; see [17].

The regularized version of (2.1) is: search $\mathbf{u} \in \mathbf{H}_0(\mathbf{curl};\Omega)$ such that

$$(\frac{1}{\mu}\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + \alpha(\mathbf{u},\mathbf{v})_{\mathbf{L}^2(\Omega)} = (\mathbf{j}_0, \mathbf{v})_{\mathbf{L}^2(\Omega)} \quad \text{for all } \mathbf{v} \in \mathbf{H}_0(\mathbf{curl};\Omega). \tag{2.3}$$

The symmetric bilinear form

$$a(\mathbf{u},\mathbf{v}) := (\frac{1}{\mu}\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + \alpha(\mathbf{u},\mathbf{v})_{\mathbf{L}^2(\Omega)}$$

induced by equation (2.3) is continuous and $\mathbf{H}(\mathbf{curl};\Omega)$-coercive for all $\alpha > 0$. Continuity can be obtained by making use of the Cauchy-Schwarz inequality. Coercivity follows from

$$|a(\mathbf{u},\mathbf{u})| = \int_\Omega \frac{1}{\mu}|\mathbf{curl}\,\mathbf{u}|^2\,\mathrm{d}\mathbf{x} + \int_\Omega \alpha|\mathbf{u}|^2\,\mathrm{d}\mathbf{x} \ge \frac{1}{\mu_1}\|\mathbf{curl}\,\mathbf{u}\|^2_{\mathbf{L}^2} + \alpha\|\mathbf{u}\|^2_{\mathbf{L}^2} \tag{2.4a}$$

$$\ge \min\{\frac{1}{\mu_1}, \alpha\}\|\mathbf{u}\|^2_{\mathbf{H}(\mathbf{curl};\Omega)}. \tag{2.4b}$$

As a consequence, the $\mathbf{H}(\mathbf{curl})$-conforming edge element (see [17]) discretization results in a matrix $A_\alpha \in \mathbb{R}^{n\times n}$ with entries

$$(A_\alpha)_{ij} = a(\varphi_i, \varphi_j), \quad i,j = 1,\dots,n,$$

where φ_i, $i = 1,\dots,n$, denote the corresponding basis functions, is a symmetric positive definite matrix.

3. Hierarchical Matrices

The setting in which the preconditioner will be computed are hierarchical matrices. This methodology introduced by Hackbusch et al. [14, 15] is designed to handle fully populated matrices such as approximations of the inverse or the factors of the LU decomposition with logarithmic-linear complexity; see [5] and [21] for parallel $\mathcal{H}$-matrix algorithms. In this section we review $\mathcal{H}$-matrices in the context of nested dissection reorderings, which are well-suited for efficient LU factorizations.

The efficiency of hierarchical matrices is based on low-rank approximations of each submatrix of a suitable partition P of the full set of matrix indices $I \times I$, $I := \{1, \ldots, n\}$. The construction of P has to account for two aims. On one hand it has to guarantee that the rank k of the approximation

$$B_{ts} \approx XY^T, \quad X \in \mathbb{C}^{|t|\times k}, Y \in \mathbb{C}^{|s|\times k}, \tag{3.1}$$

to each block B_{ts}, $t \times s \in P$, of a given matrix $B \in \mathbb{R}^{n\times n}$ depends logarithmically on its approximation accuracy. Here, B denotes a fully populated matrix, e.g., the inverse or the factors of the LU decomposition of the matrix A resulting from a finite element discretization. On the other hand P must be computable with logarithmic-linear complexity. The former issue will be addressed by the so-called *admissibility condition* in Sect. 3.2, while the latter problem can be solved by so-called *cluster trees*.

3.1. Cluster tree

Searching the set of possible partitions of $I \times I$ for a partition P which guarantees (3.1) seems practically impossible since this set is considerably large. By restricting ourselves to blocks $t \times s$ made up from rows t and columns s which are generated by recursive subdivision, P can be found with almost linear complexity. The structure which describes the way I is subdivided into smaller parts is the cluster tree. Since we are aiming at nested dissection LU factorizations, we adapt the usual definition to this special situation. To this end, a cluster t_{sep} is referred to as a *separator* for t_1 and t_2 in t if $t = t_1 \cup t_{\text{sep}} \cup t_2$ and $A_{t_1 t_2} = A_{t_2 t_1} = 0$. A tree T_I is called a *cluster tree* for an index set I if it satisfies the following conditions:

(i) I is the root of T_I;

(ii) if $t \in T_I$ is not a leaf and if it is not a separator cluster, then t is a disjoint union of its sons $S_I(t) = \{t_1, t_{\text{sep}}, t_2\} \subset T_I$, where t_{sep} is a separator for t_1 and t_2 in t;

(iii) if $t \in T_I$ is not a leaf and if it is a separator cluster, then t is a disjoint union of its sons $S_I(t) = \{t_1, t_2\} \subset T_I$.

Due to (ii), the matrix A_{tt} for a cluster $t \in T_I \setminus \mathcal{L}(T_I)$ that is not a separator has the structure

$$A_{tt} = \begin{bmatrix} A_{t_1 t_1} & & A_{t_1 t_{\text{sep}}} \\ & A_{t_2 t_2} & A_{t_2 t_{\text{sep}}} \\ A_{t_{\text{sep}} t_1} & A_{t_{\text{sep}} t_2} & A_{t_{\text{sep}} t_{\text{sep}}} \end{bmatrix}.$$

We denote the set of leaves of the tree T_I by $\mathcal{L}(T_I)$. Each level of T_I contains a partition of the index set I.

A cluster tree for I can be computed for instance by the *bounding box method* [11] or the *principal component analysis* [5]. The latter methods take into account the geometric information associated with the matrix indices. A nested dissection approach based on the matrix graph $G_I = (I, E_I)$ with

$$E_I := \{(i,j) \in I \times I : a_{ij} \neq 0 \text{ or } a_{ji} \neq 0\}$$

of A has recently [6] turned out to give significantly better results. Each nested dissection step can be separated into two phases. In the partitioning phase the bipartition of t into t_1' and t_2' is computed using *spectral bisection* [25] combined with the multi-level ideas from [20]. Subdividing t in this manner in some sense minimizes the size of the *edge cut*, i.e., a set of edges $C \subset E_t$ such that $G_t' = (t, E_t \setminus C)$ is no longer connected. The partition phase of the cluster algorithm ensures that the cardinality of t_1' equals t_2'.

In the second step the vertex set t_{sep} which separates t_1 and t_2 is computed. To this end one considers the boundaries

$$\partial t_1' := \{i \in t_1' : \exists j \in t_2' \text{ such that } (i,j) \in E_t\},$$
$$\partial t_2' := \{j \in t_2' : \exists i \in t_1' \text{ such that } (j,i) \in E_t\}$$

of t_1' and t_2' and the edge set $E_{12} = \{(i,j) \in E_t : i \in \partial t_1', j \in \partial t_2'\}$ between $\partial t_1'$ and $\partial t_2'$. To get a small separator t_{sep}, the *minimal vertex cover algorithm* [19] is applied to the bipartite graph $(\partial t_1' \cup \partial t_2', E_{12})$. Finally, the vertices belonging to the minimal vertex cover are moved out of t_1' and t_2' to t_{sep}. This generates the desired partition $\{t_1, t_{\text{sep}}, t_2\}$ of t. The previous construction is recursively applied to t_1 and t_2, while t_{sep} is recursively subdivided using spectral bisection.

3.2. Block cluster tree

The approximation results from [2,7] show that in order to be able to guarantee a sufficient approximation of each submatrix B_{ts}, $t \times s \in P$, of B by a matrix of low rank, the subblock $t \times s$ has to satisfy the so-called *admissibility condition*

$$\min\{\operatorname{diam} X_t, \operatorname{diam} X_s\} \leq \eta \operatorname{dist}(X_t, X_s) \tag{3.2}$$

for a given parameter $\eta > 0$ or $\min\{|t|, |s|\} \leq n_{\min}$ holds for a given block size parameter $n_{\min} \in \mathbb{N}$. Here,

$$\operatorname{diam} X_t := \max_{x,y \in X_t} |x - y| \quad \text{and} \quad \operatorname{dist}(X_t, X_s) := \min_{x \in X_t,\, y \in X_s} |x - y|$$

and the support $X_t := \bigcup_{i \in t} X_i$ of a cluster $t \in T_I$ is the union of the supports of the basis functions corresponding to its indices. Notice that in order to satisfy (3.2) the supports of t and s have to be far enough away from each other. This condition is caused by the fact that the fundamental solution of elliptic differential operators possesses a singularity for $x = y$ only.

The results in [6] indicate that instead of the geometric condition (3.2) one is better off using the algebraic condition

$$\min\{\operatorname{diam}(t), \operatorname{diam}(s)\} \leq \eta \operatorname{dist}(t,s), \tag{3.3}$$

where for $t, s \subset I$ we set

$$\operatorname{diam}(t) := \max_{i,j \in t} d_{ij} \quad \text{and} \quad \operatorname{dist}(t,s) := \min_{i \in t,\, j \in s} d_{ij}.$$

Here, the expression d_{ij} is the length of a shortest path connecting two indices $i, j \in I$ within the matrix graph G. In [6] a multi-level Dijkstra algorithm is presented which computes approximations to the expressions appearing in (3.3) in such a way that the overall logarithmic-linear complexity is preserved.

The construction of the desired partition P can be done no matter what the actual admissibility condition is. The partition is usually generated by recursive subdivision of $I \times I$. The recursion stops in blocks which satisfy (3.2) or (3.3), respectively, or which are small enough. The numerical method from this article uses a *block cluster tree* $T_{I\times I}$, which is a cluster tree for the set of matrix indices $I \times I$ associated with the descendant mapping $S_{I\times I}$ defined by

$$S_{I\times I}(t,s) := \begin{cases} \emptyset, & \text{if } S_I(t) = \emptyset \text{ or } S_I(s) = \emptyset, \\ \emptyset, & \text{if } t \neq s \text{ and neither } t \text{ nor } s \text{ are separators}, \\ \emptyset, & \text{if } t \text{ or } s \text{ are separators and satisfy (3.3)}, \\ S_I(t) \times S_I(s), & \text{else.} \end{cases}$$

The set of leaves $\mathcal{L}(T_{I\times I})$ of the block cluster tree $T_{I\times I}$ forms a partition P of $I \times I$ such that (3.3) is satisfied for large enough blocks.

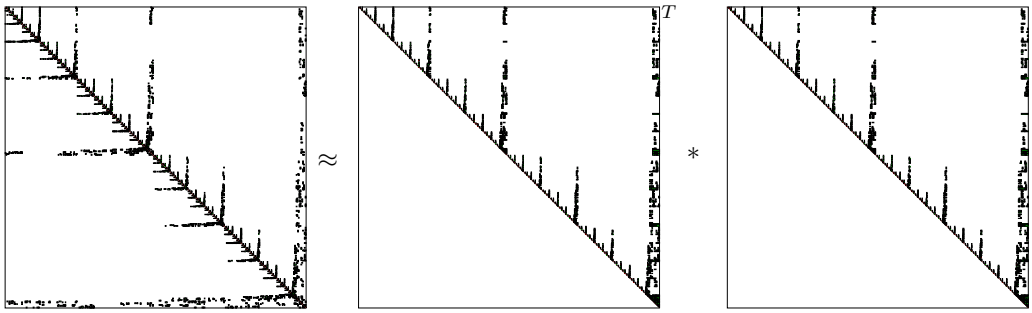

Fig. 3.1. Block structure of a nested dissection hierarchical Cholesky decomposition.

Given a suitable partition P, the set of $\mathcal{H}$-matrices with blockwise rank k is defined as

$$\mathcal{H}(P,k) := \{M \in \mathbb{R}^{I\times I} : \operatorname{rank} M_b \leq k \text{ for all } b \in P\}.$$

The storage requirement for $M \in \mathcal{H}(P,k)$ is of the order $kn \log n$. Multiplying M by a vector can be done with $\mathcal{O}(kn \log n)$ arithmetic operations. Since the sum of two $\mathcal{H}$-matrices $M_1, M_2 \in \mathcal{H}(P,k)$ exceeds blockwise rank k, the sum has to be truncated to $\mathcal{H}(P,k)$. This can be done with complexity $\mathcal{O}(k^2 n \log n)$ if an approximation error of controllable size can be tolerated. The complexity of computing an approximation to the product of two $\mathcal{H}$-matrices is $\mathcal{O}(k^2 n (\log n)^2)$; see [13–15].

4. Degenerate Approximation of Solution Operators

In [2, 7] it is proved that inverse finite element stiffness matrices of scalar elliptic boundary value problems can be approximated by hierarchical matrices with logarithmic-linear complexity. The core of the proof is the construction of a degenerate approximation to the Green function. It is remarkable that this result holds for arbitrary coefficients of the operator. Based

on the approximation of the inverse it was proved in [4] that also the factors of the LU decomposition can be handled by hierarchical matrices with logarithmic-linear complexity. This paves the way to approximate direct methods that do not suffer from the well-known effect of fill-in.

Our aim in this section is to show that discrete solution operators of homogeneous Dirichlet boundary value problems of the operator

$$\mathbf{L}_\alpha = \mathbf{curl}\frac{1}{\mu}\mathbf{curl} + \alpha\mathbf{I}, \quad \alpha \geq \frac{1}{\mu_1},$$

can be approximated by hierarchical matrices with logarithmic-linear complexity. While it is easy to extend the scalar situation to vector-valued problems in the usual Sobolev space setting, the $\mathbf{H}(\mathbf{curl})$ spaces require some non-trivial adaptation of the results from [2,7] to this situation.

Assume that there is a Green function $\mathbf{G}(x,y) \in \mathbb{R}^{3\times 3}$ for $\mathbf{L}_\alpha$ and Ω such that for each row $\mathbf{G}^{(i)}$, $i = 1,2,3$, and all $x \in \Omega$ it holds that

(i) $\mathbf{L}_\alpha \mathbf{G}^{(i)}(x,\cdot) = \delta_x e_i$ in Ω,

(ii) $[\mathbf{G}^{(i)}(x,\cdot)]_{\mathbf{t}} = 0$ on $\partial\Omega$.

Since we do not want to repeat the proofs from [2, 7], we concentrate on the central problem of constructing degenerate kernel expansions to $\mathbf{G}$, i.e., we will prove that for any $k \gtrsim |\log\varepsilon|^4$ there is

$$\mathbf{G}_k(x,y) := \sum_{\ell=1}^{k} \mathbf{u}_\ell(x)\mathbf{v}_\ell(y)^T$$

such that

$$\|\mathbf{G}(x,\cdot) - \mathbf{G}_k(x,\cdot)\|_{\mathbf{L}^2(D_2)} \leq \varepsilon \|\mathbf{G}(x,\cdot)\|_{\mathbf{L}^2(\hat{D}_2)} \quad \text{for all } x \in D_1,$$

where the pair of domains $D_1, D_2 \subset \Omega$ satisfies the admissibility condition (see (3.2))

$$\eta \operatorname{dist}(D_1, D_2) \geq \operatorname{diam} D_2 \tag{4.1}$$

and $\hat{D}_2$ is some domain enclosing D_2. The rest of the proof in [2] is based on this existence result and can be applied without changes.

Remark 4.1. Since the technique in [2] uses a relation between the FE inverse and the discretization of $\mathbf{L}_\alpha^{-1}$, it allows to prove any approximation accuracy up to the finite element error. Numerical experiments (see also Sect. 7), however, indicate that any precision can be reached.

Remark 4.2. In this section it will be assumed for theoretical purposes that Ω is convex (curved) polyhedron. Furthermore, in practice the admissibility condition (4.1) is checked for convex polyhedral supersets $D_1', D_2' \subset \mathbb{R}^3$ of D_1 and D_2, respectively. Hence, we may assume that D_2 is the intersection of two convex polyhedrons D_2' and Ω, which is a convex polyhedron, too.

Let $\emptyset \neq D \subset \Omega$ be a convex polyhedron and let K be a convex subset of D having positive distance to the boundary of D in Ω, i.e.

$$\sigma(K,D) := \operatorname{dist}(K, \partial D \cap \Omega) > 0.$$

By $X(D)$ we denote a given family of closed subspaces of $\mathbf{L}^2(D)$ which have the property that $\mathbf{u}|_K \in X(K)$ for $\mathbf{u} \in X(D)$. The following proof will rely on the existence of a subspace $V \subset X(K)$ of finite dimension $\dim V \le k$ such that

$$\inf_{\mathbf{v}\in V} \|\mathbf{u}-\mathbf{v}\|_{\mathbf{L}^2(K)} \le \frac{c_A}{\sqrt[3]{k}} \frac{\operatorname{diam} K}{\sigma(K,D)} \|\mathbf{u}\|_{\mathbf{L}^2(D)} \quad \text{for all } \mathbf{u} \in X(D). \tag{4.2}$$

Estimate (4.2) will be proved in the second part of this section for an appropriate choice of $X(D)$.

Lemma 4.1. *Assume that* (4.2) *is valid. Let $D_2 \subset D$ be a convex polyhedron satisfying*

$$0 < \operatorname{diam} D_2 \le \eta\, \sigma(D_2, D)$$

with given $\eta > 0$. Then for any $\varepsilon > 0$ there is a subspace $V \subset X(D_2)$ with $\dim V \lesssim c_\eta^3 |\log \varepsilon|^4$ so that

$$\inf_{\mathbf{v}\in V} \|\mathbf{u}-\mathbf{v}\|_{\mathbf{L}^2(D_2)} \le \varepsilon \|\mathbf{u}\|_{\mathbf{L}^2(D)} \quad \textit{for all } \mathbf{u} \in X(D), \tag{4.3}$$

where $c_\eta := c_A \mathrm{e}(2+\eta)$.

Proof. Let $\ell = \lceil |\log \varepsilon| \rceil$ and $r_0 = \sigma(D_2, D)$. We consider a nested sequence of domains

$$K_j = \{x \in \Omega : \operatorname{dist}(x, D_2) \le r_0(\ell - j)/\ell\}, \quad j = 0, \dots, \ell.$$

Notice that $D_2 = K_\ell \subset K_{\ell-1} \subset \cdots \subset K_0 \subset D$ with $\sigma(K_j, K_{j-1}) = r_0/\ell$, $j = 1, \dots, \ell$. Furthermore, it is easy to see that $\operatorname{diam} K_j \le (2+\eta) r_0$ and that each K_j is a convex (curved) polyhedron.

Let $\mathbf{u} \in X(D)$ and $\mathbf{e}_0 := \mathbf{u}|_{K_0} \in X(K_0)$. Applying estimate (4.2) to the pair (K_0, K_1) with the choice

$$k := \lceil (c_A(2+\eta)\ell \varepsilon^{-1/\ell})^3 \rceil,$$

there is $V_1 \subset X(K_1)$ such that $\mathbf{e}_0|_{K_1} = \mathbf{v}_1 + \mathbf{e}_1$ with $\mathbf{v}_1 \in V_1$ and

$$\|\mathbf{e}_1\|_{\mathbf{L}^2(K_1)} \le \frac{c_A}{\sqrt[3]{k}} \frac{\operatorname{diam} K_1}{\sigma(K_1, K_0)} \|\mathbf{e}_0\|_{\mathbf{L}^2(K_0)} \le \varepsilon^{1/\ell} \|\mathbf{e}_0\|_{\mathbf{L}^2(K_0)}.$$

Since $\mathbf{e}_0|_{K_1} \in X(K_1)$, the new remainder $\mathbf{e}_1$ also belongs to $X(K_1)$. Similarly, for all $j = 1, \dots, \ell$ we are able to find subspaces $V_j \subset X(K_j)$ with $\dim V_j \le k$ and approximants $\mathbf{v}_j \in V_j$ so that $\mathbf{e}_{j-1}|_{K_j} = \mathbf{v}_j + \mathbf{e}_j$ and

$$\|\mathbf{e}_j\|_{\mathbf{L}^2(K_j)} \le \varepsilon^{1/\ell} \|\mathbf{e}_{j-1}\|_{\mathbf{L}^2(K_{j-1})}.$$

Using the restrictions of V_j to the smallest domain $D_2 = K_\ell$, let

$$V := \operatorname{span}\{V_j|_{D_2},\, j = 1, \dots, \ell\}.$$

Then V is a subspace of $X(D_2)$ and, since $\mathbf{e}_0|_{D_2} = \mathbf{e}_\ell + \sum_{j=1}^{\ell} \mathbf{v}_j|_{D_2}$, we are led to

$$\inf_{\mathbf{v}\in V} \|\mathbf{e}_0 - \mathbf{v}\|_{\mathbf{L}^2(D_2)} \le \|\mathbf{e}_\ell\|_{\mathbf{L}^2(D_2)} \le \varepsilon \|\mathbf{e}_0\|_{\mathbf{L}^2(K_0)} \le \varepsilon \|\mathbf{u}\|_{\mathbf{L}^2(D)},$$

where the last inequality is due to $K_0 \subset D$.

The dimension of V is bounded by $\sum_{j=1}^{\ell} \dim V_j \le \ell k$. Since $\varepsilon^{-1/\ell} \le \mathrm{e}$ we obtain that $\dim V \le (c_A \mathrm{e}(2+\eta))^3 \ell^4 + \ell$. □

We define the closed space $X(D)$ by the following conditions: $\mathbf{u} \in X(D)$ if

(i) $\mathbf{u} \in \mathbf{H}(\mathbf{curl}; D)$,

(ii) $a(\mathbf{u}, \mathbf{v}) = 0$ for all $\mathbf{v} \in \mathbf{H}_0(\mathbf{curl}; D)$,

(iii) $\mathbf{u_t} = 0$ on $\partial D \cap \partial\Omega$.

Notice that (ii) implies that

$$(\mathbf{u}, \mathbf{grad}\, \varphi)_{\mathbf{L}^2(D)} = 0 \quad \text{for all } \varphi \in H_0^1(D), \tag{4.4}$$

which can be seen from using $\mathbf{v} := \mathbf{grad}\, \varphi$ as a test function and exploiting that $\mathbf{curl}\, \mathbf{grad}\, \varphi = 0$.

The previous lemma will now be applied to the i-th row $\mathbf{G}^{(i)}(x, \cdot)$ of $\mathbf{G}(x, \cdot)$. It is obvious that that $\mathbf{g}_x^{(i)} := \mathbf{G}^{(i)}(x, \cdot)|_{\Omega \setminus \overline{D}_1}$ is in $X(\Omega \setminus \overline{D}_1)$ for $x \in D_1$.

Theorem 4.1. *Let $D_1 \subset \Omega$ and let $D_2 \subset \Omega$ be a convex polyhedron satisfying* (4.1). *Then for any $\varepsilon > 0$ there is a separable approximation*

$$\mathbf{G}_k(x, y) = \sum_{\ell=1}^{k} \mathbf{u}_\ell(x) \mathbf{v}_\ell(y)^T \quad \textit{with} \quad k \lesssim c_\eta^3 |\log \varepsilon|^4,$$

so that for all $x \in D_1$

$$\|\mathbf{G}(x, \cdot) - \mathbf{G}_k(x, \cdot)\|_{\mathbf{L}^2(D_2)} \le \varepsilon \|\mathbf{G}(x, \cdot)\|_{\mathbf{L}^2(\hat{D}_2)}, \tag{4.5}$$

where $\hat{D}_2 := \{y \in \Omega : 2\eta \operatorname{dist}(y, D_2) < \operatorname{diam} D_2\}$ and $c_\eta = 2c_A \mathrm{e}(1 + \eta)$.

Proof. Notice that because of $\operatorname{dist}(D_1, \hat{D}_2) > 0$, we have $\mathbf{g}_x^{(i)} \in X(\hat{D}_2)$ for all $x \in D_1$. Since in addition $\operatorname{diam} D_2 \le 2\eta\, \sigma(D_2, \hat{D}_2)$, Lemma 4.1 can be applied with η replaced by 2η.

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a basis of the subspace $V \subset X(D_2)$ with $k = \dim V \lesssim c_{2\eta}^3 |\log \varepsilon|^4$. By means of (4.3) $\mathbf{g}_x^{(i)}$ can be decomposed into $\mathbf{g}_x^{(i)} = \hat{\mathbf{g}}_x^{(i)} + \mathbf{e}_x^{(i)}$ with $\hat{\mathbf{g}}_x^{(i)} \in V$ and

$$\|\mathbf{e}_x^{(i)}\|_{\mathbf{L}^2(D_2)} \le \varepsilon \|\mathbf{g}_x^{(i)}\|_{\mathbf{L}^2(\hat{D}_2)}, \quad i = 1, 2, 3.$$

Expressing $\hat{\mathbf{g}}_x^{(i)}$ in the basis of V, we obtain

$$\hat{\mathbf{g}}_x^{(i)} = \sum_{\ell=1}^{k} u_\ell^{(i)}(x) \mathbf{v}_\ell$$

with coefficients $u_\ell^{(i)}$ depending on $x \in D_1$. The function $\mathbf{G}_k(x, y) := \sum_{\ell=1}^{k} \mathbf{u}_\ell(x) \mathbf{v}_\ell(y)^T$ with $\mathbf{u}_\ell := (u_\ell^{(1)}, u_\ell^{(2)}, u_\ell^{(3)})^T$ satisfies estimate (4.5). □

It remains to show (4.2) for our choice of $X(D)$. The first step is the interior regularity for elements from $X(D)$, which is proved in the following lemma.

Lemma 4.2. *Let $D \subset \Omega$ be a convex polyhedron and let $\mathbf{u} \in X(D)$. Then for any set $K \subset D$ satisfying $\sigma(K, D) > 0$ it holds that*

$$\|\mathbf{u}\|_{\mathbf{H}^1(K)} \le \frac{c_{\mathbf{L}}}{\sigma(K, D)} \|\mathbf{u}\|_{\mathbf{L}^2(D)}, \qquad c_{\mathbf{L}} = 2\sqrt{\frac{\mu_1}{\mu_0} + 1},$$

where μ_0, μ_1 are defined in (1.1).

Proof. Let $\eta \in C^1(D)$ satisfy $0 \le \eta \le 1$, $\eta = 1$ in K, $\eta = 0$ in a neighborhood of $\partial D \cap \Omega$ and

$$\|\nabla \eta\|_{\infty,D} \le \frac{2}{\sigma(K,D)}.$$

We have that $\eta^2 \mathbf{u} \in \mathbf{H}(\mathbf{curl}; D)$ and $(\eta^2 \mathbf{u})_{\mathbf{t}} = 0$ on $\partial D = (\partial D \cap \Omega) \cup (\partial D \cap \partial\Omega)$. Hence, $\eta^2 \mathbf{u}$ can be used as a test function in the definition of $X(D)$, which leads to

$$(\frac{1}{\mu}\mathbf{curl}\,\mathbf{u}, \mathbf{curl}(\eta^2 \mathbf{u}))_{\mathbf{L}^2(D)} + \alpha \|\eta \mathbf{u}\|^2_{\mathbf{L}^2(D)} = 0.$$

Using $\mathbf{curl}(\varphi \mathbf{u}) = \varphi\,\mathbf{curl}\,\mathbf{u} + \nabla\varphi \times \mathbf{u}$ for $\varphi \in C^1(D)$, from

$$\begin{aligned}
(\frac{1}{\mu}\mathbf{curl}\,\mathbf{u}, \mathbf{curl}(\eta^2 \mathbf{u}))_{\mathbf{L}^2(D)} &= (\frac{1}{\mu}\mathbf{curl}\,\mathbf{u}, \eta\,\mathbf{curl}(\eta \mathbf{u}) + \eta \nabla\eta \times \mathbf{u})_{\mathbf{L}^2(D)} \\
&= (\frac{\eta}{\mu}\mathbf{curl}\,\mathbf{u}, \mathbf{curl}(\eta \mathbf{u}) + \nabla\eta \times \mathbf{u})_{\mathbf{L}^2(D)} \\
&= (\frac{1}{\mu}\mathbf{curl}(\eta \mathbf{u}) - \frac{1}{\mu}\nabla\eta \times \mathbf{u}, \mathbf{curl}(\eta \mathbf{u}) + \nabla\eta \times \mathbf{u})_{\mathbf{L}^2(D)} \\
&= \|\mu^{-1/2}\mathbf{curl}(\eta \mathbf{u})\|^2_{\mathbf{L}^2(D)} - \|\mu^{-1/2}\nabla\eta \times \mathbf{u}\|^2_{\mathbf{L}^2(D)}
\end{aligned}$$

we obtain that

$$\begin{aligned}
\|\eta \mathbf{u}\|^2_{\mathbf{H}(\mathbf{curl};D)} &\le \|\eta \mathbf{u}\|^2_{\mathbf{L}^2(D)} + \mu_1 \|\mu^{-1/2}\mathbf{curl}(\eta \mathbf{u})\|^2_{\mathbf{L}^2(D)} \\
&= \mu_1 \|\mu^{-1/2}\nabla\eta \times \mathbf{u}\|^2_{\mathbf{L}^2(D)} + (1 - \alpha\mu_1)\|\eta \mathbf{u}\|^2_{\mathbf{L}^2(D)} \\
&\le \frac{\mu_1}{\mu_0}\|\nabla\eta\|^2_{\infty,D}\|\mathbf{u}\|^2_{\mathbf{L}^2(D)}
\end{aligned}$$

due to $\alpha \ge 1/\mu_1$ and $\|\mu^{-1/2}\nabla\eta \times \mathbf{u}\|_{\mathbf{L}^2(D)} \le \mu_0^{-1/2}\|\nabla\eta\|_{\infty,D}\|\mathbf{u}\|_{\mathbf{L}^2(D)}$.

Furthermore, (4.4) shows that $\eta \mathbf{u}$ also belongs to $\mathbf{H}_0(\mathbf{curl}; D) \cap \mathbf{H}(\mathrm{div}; D)$ with $\mathrm{div}\,\eta \mathbf{u} = \mathbf{u} \cdot \nabla\eta$. The latter space can be imbedded into $\mathbf{H}^1(D)$; see [12, Theorem 3.7]. From [9, Theorem 2.3 and Remark 2.4] we obtain that

$$\|\mathbf{grad}\,\eta \mathbf{u}\|^2_{\mathbf{L}^2(D)} \le \|\mathbf{curl}\,\eta \mathbf{u}\|^2_{\mathbf{L}^2(D)} + \|\mathrm{div}\,\eta \mathbf{u}\|^2_{\mathbf{L}^2(D)},$$

because the second fundamental form is positive definite due to the convexity of Ω. The assertion follows from

$$\begin{aligned}
\|\mathbf{u}\|^2_{\mathbf{H}^1(K)} \le \|\eta \mathbf{u}\|^2_{\mathbf{H}^1(D)} &\le \|\eta \mathbf{u}\|^2_{\mathbf{H}(\mathbf{curl};D)} + \|\mathrm{div}\,\eta \mathbf{u}\|^2_{\mathbf{L}^2(D)} \\
&\le \|\nabla\eta\|^2_{\infty,D}\left(\frac{\mu_1}{\mu_0} + 1\right)\|\mathbf{u}\|^2_{\mathbf{L}^2(D)} \\
&\le \frac{4}{\sigma^2}\left(\frac{\mu_1}{\mu_0} + 1\right)\|\mathbf{u}\|^2_{\mathbf{L}^2(D)}.
\end{aligned}$$

□

Remark 4.3. Notice that the constant $c_{\mathbf{L}}$ depends on the ratio of μ_1 and μ_0 but not on the smoothness of μ. In the numerical experiments it will even be hard to observe a dependence on μ_1/μ_0.

Before we prove (4.2) in Lemma 4.4, let us state the following lemma from [7]. There we actually proved a scalar version; the generalization to vector-valued functions $\mathbf{u}$ is, however, obvious.

Lemma 4.3. *Let $D \subset \Omega$ be convex. If X is a closed subspace of $\mathbf{L}^2(D)$, then for any $k \in \mathbb{N}$ there is a subspace $V \subset X$ satisfying* $\dim V \le k$ *so that*

$$\inf_{\mathbf{v}\in V} \|\mathbf{u}-\mathbf{v}\|_{\mathbf{L}^2(D)} \le c\frac{\operatorname{diam} D}{\sqrt[3]{k}}\|\mathbf{grad}\,\mathbf{u}\|_{\mathbf{L}^2(D)} \quad \textit{for all } \mathbf{u} \in X \cap \mathbf{H}^1(D).$$

Lemma 4.4. *Let $D \subset \Omega$ be a convex polyhedron and let $K \subset D$, $\sigma(K,D) > 0$, be convex. Then for any $k \in \mathbb{N}$ there is a subspace $V \subset X(K)$ satisfying* $\dim V \le k$ *so that*

$$\inf_{\mathbf{v}\in V} \|\mathbf{u}-\mathbf{v}\|_{\mathbf{L}^2(K)} \le \frac{c_A}{\sqrt[3]{k}}\frac{\operatorname{diam} K}{\sigma(K,D)}\|\mathbf{u}\|_{\mathbf{L}^2(D)} \quad \textit{for all } \mathbf{u} \in X(D).$$

Proof. By Lemma 4.3 there is $V \subset X(K)$ such that

$$\inf_{\mathbf{v}\in V} \|\mathbf{u}-\mathbf{v}\|_{\mathbf{L}^2(K)} \le c\frac{\operatorname{diam} K}{\sqrt[3]{k}}\|\mathbf{grad}\,\mathbf{u}\|_{\mathbf{L}^2(K)} \quad \text{for all } \mathbf{u} \in X(K) \cap \mathbf{H}^1(K).$$

Hence, Lemma 4.2 gives the estimate

$$\inf_{\mathbf{v}\in V} \|\mathbf{u}-\mathbf{v}\|_{\mathbf{L}^2(K)} \le \frac{c_A}{\sqrt[3]{k}}\frac{\operatorname{diam} K}{\sigma(K,D)}\|\mathbf{u}\|_{\mathbf{L}^2(D)} \quad \text{for all } \mathbf{u} \in X(D),$$

where we set $c_A := cc_{\mathbf{L}}$. □

5. Preconditioning the Curl-Curl Operator

Let $A \in \mathbb{R}^{n\times n}$ be the coefficient matrix resulting from the edge element discretization of the variational formulation (2.1) and let

$$0 = \lambda_1 = \cdots = \lambda_{m-1} < \lambda_m \le \cdots \le \lambda_n$$

denote the eigenvalues of the symmetric positive semi-definite matrix A. Notice that the smallest non-vanishing eigenvalue λ_m is bounded by a constant from below due to (2.2).

In this section it will be shown that the regularization leads to a positive definite matrix which provides a spectrally equivalent preconditioner on the orthogonal complement $(\ker A)^\perp$ of the kernel of A. Note that the spectral equivalence on $(\ker A)^\perp$ is sufficient for a bounded number of iterations since the conjugate gradient method (CG) suppresses kernel components of the initial vector. Given $x_0 \in \mathbb{R}^n$, a scalar product $(\cdot,\cdot)$ and a positive-semidefinite matrix A which is self-adjoint with respect to $(\cdot,\cdot)$, CG generates a sequence $\{r_k\}$ defined by

$$\begin{aligned} &r_0 := b - Ax_0 \in \mathbb{R}^n, \quad r_0 \perp \ker A, \quad \omega_0 = 1,\\ &r_{k+1} = \omega_k r_k + (1-\omega_k)r_{k-1} - \alpha_k\omega_k A r_k, \quad k = 0,1,2,\dots. \end{aligned}$$

The choice of the parameters

$$\alpha_k = \frac{(r_k,r_k)}{(Ar_k,r_k)} \quad \text{and} \quad \omega_k^{-1} = 1 - \omega_{k-1}^{-1}\frac{\alpha_k}{\alpha_{k-1}}\frac{(r_k,r_k)}{(r_{k-1},r_{k-1})}$$

guarantee that $r_{k+1} \perp \ker A$. Let $A^+ \in \mathbb{R}^{n\times n}$ be the pseudo-inverse of A. Then the following generalization (see [23]) of the well-known CG error estimate holds

$$(r_k, A^+ r_k) \le 4\left(\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}}\right)^{2k}(r_0, A^+ r_0), \quad k = 0,1,2,\dots, \tag{5.1}$$

where $\xi = \lambda_m/\lambda_n$.

Consider a positive definite preconditioner $C \in \mathbb{R}^{n\times n}$. Then AC is positive semi-definite with respect to the scalar product $(x,y)_C := y^T Cx$. Hence, the mentioned properties of CG also hold for the preconditioned matrix AC if the scalar product $(\cdot,\cdot)_C$ is used. In particular, the residuals r_k are still in $(\ker A)^\perp$ provided $r_0 \perp \ker A$. This follows from $C^{-1}x \in \ker AC$ for $x \in \ker A$ and

$$r_k^T x = (r_k, C^{-1}x)_C = 0.$$

Furthermore, instead of (5.1) we have that

$$r_k^T A^+ r_k \le 4\left(\frac{1-\sqrt{\xi'}}{1+\sqrt{\xi'}}\right)^{2k} r_0^T A^+ r_0, \quad k = 0,1,2,\ldots,$$

where ξ' denotes the ratio of the m-th and the n-th eigenvalue of AC.

Due to (2.4) the regularization (1.3) leads to positive definite coefficient matrices $A_\alpha := A+\alpha M$, where M denotes the mass matrix $M_{ij} = (\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i)_{\mathbf{L}^2(\Omega)}$ and $\boldsymbol{\varphi}_i$, $i = 1,\ldots,n$, correspond to edge elements. The stiffness matrix $A \in \mathbb{R}^{n\times n}$ has entries $A_{ij} = (\mathbf{L}_0\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i)_{\mathbf{L}^2(\Omega)}$.

Let $C := L_{\mathcal{H}} L_{\mathcal{H}}^T$, $L_{\mathcal{H}} \in \mathcal{H}(P,k)$, be an approximate Cholesky decomposition of A_α satisfying

$$\|A_\alpha - C\|_2 \le \varepsilon \|A_\alpha\|_2. \tag{5.2}$$

Then an appropriate choice of the hierarchical matrix rounding precision $\varepsilon_{\mathcal{H}}$ guarantees that A and C are spectrally equivalent on the orthogonal complement $(\ker A)^\perp$ of the kernel of A. This can be seen from the following arguments. Let $\lambda_1', \lambda_n' > 0$ denote the smallest and the largest eigenvalue of M. From (5.2) it follows that

$$\begin{aligned}
&\|I - A_\alpha^{-1/2} C A_\alpha^{-1/2}\|_2 \\
=&\|A_\alpha^{-1/2}(A_\alpha - C)A_\alpha^{-1/2}\|_2 = \rho(A_\alpha^{-1/2}(A_\alpha - C)A_\alpha^{-1/2}) \\
=&\rho(A_\alpha^{-1}(A_\alpha - C)) \le \|A_\alpha^{-1}(A_\alpha - C)\|_2 \\
\le&\|A_\alpha^{-1}\|_2 \|A_\alpha - C\|_2 \le \varepsilon\kappa,
\end{aligned}$$

where $\rho(A)$ denotes the spectral radius of A and

$$\kappa := \|A_\alpha\|_2 \|A_\alpha^{-1}\|_2 \le \frac{\lambda_n + \alpha\lambda_n'}{\alpha\lambda_1'} = \frac{\lambda_n'}{\lambda_1'} + \frac{\lambda_n}{\alpha\lambda_1'}$$

denotes the spectral condition number of A_α. Setting $y := A_\alpha^{-1/2}x$ for all $x \in \mathbb{R}^n$, it holds that

$$\begin{aligned}
&\|I - A_\alpha^{-1/2} C A_\alpha^{-1/2}\|_2 \\
\ge&\frac{|x^T A_\alpha^{-1/2}(A_\alpha - C)A_\alpha^{-1/2}x|}{x^T x} = \frac{|y^T(A_\alpha - C)y|}{y^T A_\alpha y}.
\end{aligned}$$

We obtain that

$$(1-\varepsilon\kappa)x^T A_\alpha x \le x^T C x \le (1+\varepsilon\kappa)x^T A_\alpha x \quad \text{for all } x \in \mathbb{R}^n.$$

From $x^T Ax \ge \lambda_m \|x\|_2^2$ for all $x \in (\ker A)^\perp$ it follows that

$$x^T A_\alpha x \le x^T Ax + \alpha\lambda_n' \|x\|_2^2 \le (1 + \alpha\frac{\lambda_n'}{\lambda_m})x^T Ax.$$

Hence, we obtain for all $x \in (\ker A)^\perp$ and $\varepsilon\kappa < 1$ that

$$(1-\varepsilon\kappa)(1+\alpha\frac{\lambda_1'}{\lambda_n})x^TAx \le x^TCx \le (1+\varepsilon\kappa)(1+\alpha\frac{\lambda_n'}{\lambda_m})x^TAx$$

due to

$$x^TA_\alpha x \ge (1+\alpha\frac{\lambda_1'}{\lambda_n})x^TAx.$$

The choice

$$\varepsilon := \frac{1}{2\kappa} \ge \frac{\alpha\lambda_1'}{2(\alpha\lambda_n'+\lambda_n)}$$

leads to spectral equivalence

$$\frac{1}{2}(1+\alpha\frac{\lambda_1'}{\lambda_n})x^TAx \le x^TCx \le \frac{3}{2}(1+\alpha\frac{\lambda_n'}{\lambda_m})x^TAx, \quad x \in (\ker A)^\perp,$$

of A and C on $(\ker A)^\perp$ provided $\alpha \ge 1/\mu_1$ is chosen such that $\alpha\lambda_n' \sim \lambda_m$. Since the ratio λ_n'/λ_m is bounded by a constant, choosing α as a constant leads to spectral equivalence.

Remark 5.1. From the point of preconditioning it would be more efficient to use a multiple of the identity matrix I instead of the mass matrix because then $\lambda_n'/\lambda_1' = 1$.

6. Parallel Approximate LU Factorization

Assume that $p = 2^L$, $L \in \mathbb{N}$, processors are at our disposal. It is a well-known fact that in addition to reducing the fill-in, nested dissection reorderings have the advantage that the LU factorization allows for an efficient parallelization. This follows from the special structure of the matrix.

We present an LU factorization as a recursion over ℓ. Consider a diagonal block A from the ℓ-th level, $\ell < L$, of the block cluster tree. The LU factorization of A can be computed from

$$\begin{bmatrix} A_{11} & & A_{13} \\ & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} L_{11} & & \\ & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} U_{11} & & U_{13} \\ & U_{22} & U_{23} \\ & & U_{33} \end{bmatrix}.$$

Indices 1 and 2 correspond to clusters separated by an interface. The latter corresponds to the index 3. The previous factorization is equivalent to

$$\text{compute } L_{11}, U_{11} \text{ from } A_{11} = L_{11}U_{11}, \qquad \text{compute } L_{22}, U_{22} \text{ from } A_{22} = L_{22}U_{22}, \tag{6.1a}$$

$$\text{compute } U_{13} \text{ from } A_{13} = L_{11}U_{13}, \qquad \text{compute } U_{23} \text{ from } A_{23} = L_{22}U_{23}, \tag{6.1b}$$

$$\text{compute } L_{31} \text{ from } A_{31} = L_{31}U_{11}, \qquad \text{compute } L_{32} \text{ from } A_{32} = L_{32}U_{22}, \tag{6.1c}$$

$$\text{compute } X_1 := L_{31}U_{13}, \qquad \text{compute } X_2 := L_{32}U_{23}, \tag{6.1d}$$

and

$$\text{compute } L_{33}, U_{33} \text{ from } L_{33}U_{33} = A_{33} - X_1 - X_2. \tag{6.2}$$

Hence, each of the two tasks in (6.1a), (6.1b), (6.1c), and (6.1d) can be done in parallel.

Since in the ℓ-th level $2^{L-\ell}$ processors can be used to compute the factorization, we use $2^{L-\ell-1}$ processors to solve the left part of (6.1) (the part corresponding to the first cluster) and the other $2^{L-\ell-1}$ processors to solve the right part (the second cluster). If $\ell = L$, then the

sequential $\mathcal{H}$-matrix factorization algorithm introduced in [3] is used; see also [22] for the first algorithm for the computation of hierarchical LU decompositions.

Problem (6.1b) is a block forward substitution, i.e., a problem of type

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} = \begin{bmatrix} L_{11} & & \\ & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \end{bmatrix} \tag{6.3}$$

has to be solved for X_{ij}, $i = 1, 2, 3$, $j = 1, 2$. Assume this problem is to be solved with $2^{L-\ell}$ processors. Then $2^{L-\ell-1}$ processors will be used to solve the left and the other $2^{L-\ell-1}$ processors will solve the right column of the following block forward substitutions

$$\begin{aligned} A_{11} &= L_{11}X_{11}, \qquad & A_{21} &= L_{22}X_{21}, \\ A_{12} &= L_{11}X_{12}, \qquad & A_{22} &= L_{22}X_{22} \end{aligned}$$

and multiplications

$$\begin{aligned} Y_1 &:= L_{31}X_{11}, \qquad & Y_2 &:= L_{32}X_{21}, \\ Z_1 &:= L_{31}X_{12}, \qquad & Z_2 &:= L_{32}X_{22}. \end{aligned}$$

To parallelize these operations on p processors one can exploit the nested dissection substructure

$$C := [A_1, A_2, A_3]^T [B_1, B_2, B_3] = A_1B_1 + A_2B_2 + A_3B_3,$$

which can be accomplished by computing each of the products

$$C_1 := A_1B_1, \quad C_2 := A_2B_2,$$

on $p/2$ processors and computing $C = A_3B_3 + C_1 + C_2$ on one of the p processors.

The last step to solve (6.3) is the computation of X_{31} and X_{32} from

$$L_{33}X_{31} = A_{31} - Y_1 - Y_2 \quad \text{and} \quad L_{33}X_{32} = A_{32} - Z_1 - Z_2.$$

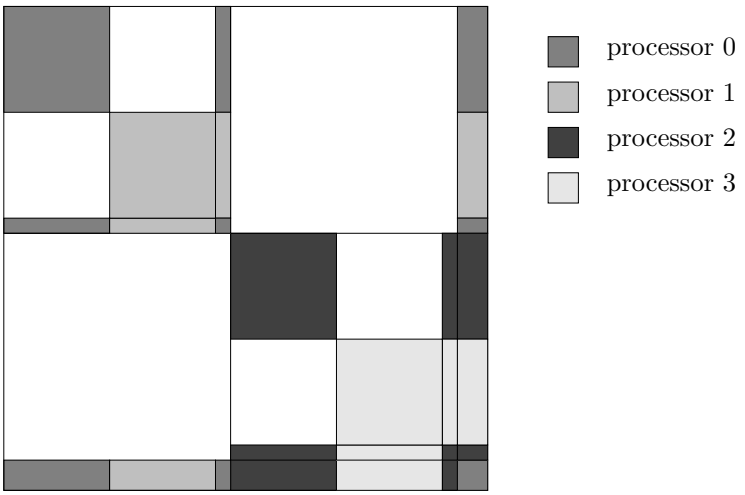

Fig. 6.1. Distribution of blocks among processors.

Since L_{33} does not possess a nested dissection substructure, we solve each of the two problems sequentially on the first processor of each half of the p processors. To this end, Z_1 and Y_2 have to be sent to the respective processor, while Z_2 and Y_1 are already in the right place. Again, the case $\ell = L$ is treated by applying the sequential $\mathcal{H}$-matrix block forward substitution presented in [3]. The backward substitution (6.1c) can be done analogously.

Only the last step (6.2), the computation of L_{33} and U_{33}, requires the completion of all previous steps (6.1a) through (6.1d). We solve (6.2) sequentially, which requires to communicate either X_1 or X_2. The parallelization of the Cholesky decomposition used in this article is now straightforward. Since the dimension of (6.2) is significantly smaller than the previous problems, one can expect a reasonable speedup. The resulting distribution of blocks (and hence the memory distribution) among the processors is shown in Fig. 6.1 in the case $p = 4$.

7. Numerical Experiments

Numerical experiments were made on the test geometry of Fig. 7.1. This geometry consists of a coil with material parameters $\sigma = 10^6$ $(\Omega\ \mathrm{m})^{-1}$, $\mu_r = 1$ and a highly permeable core with $\sigma = 0$ $(\Omega\ \mathrm{m})^{-1}$, $\mu_r = 500$ which are surrounded by air with $\sigma = 0$ $(\Omega\,\mathrm{m})^{-1}$, $\mu_r = 1$. Therefore the magnetic permeability jumps between a value of $1.3 \cdot 10^{-6}$ (Vs)/(Am) in the air and the coil to a value of $6.3 \cdot 10^{-4}$ (Vs)/(Am) in the core. The diameter of the coil is 2.45 cm.

Table 7.1 contains the time required to setup the hierarchical matrix preconditioner for various problem sizes n and different numbers of processors p, while Table 7.2 shows the time

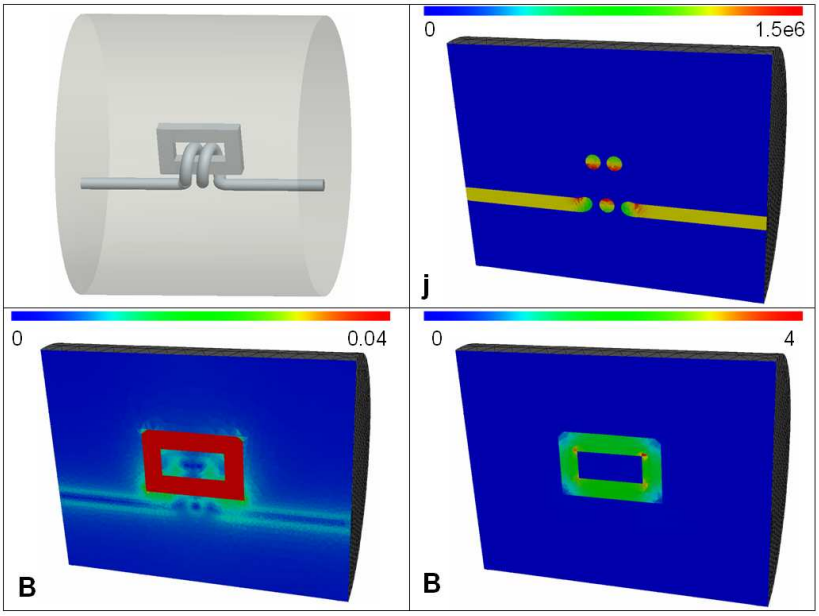

Fig. 7.1. Results of the magnetostatic field computations. The upper left picture shows the setting, the upper right picture shows the exciting current density $\mathbf{j}_0 = -\sigma\mathbf{grad}\,\varphi_0$ in A/m^2, and the lower pictures show the resulting magnetic field $\mathbf{B} = \mathbf{curl}\,\mathbf{u}$ in Tesla on two different scales. The magnetic flux is trapped in the highly permeable core of the coil.

required for the iterative solution. The columns labeled "E_p" contain the parallel efficiency

$$E_p := \frac{t(1)}{p \cdot t(p)},$$

i.e., the ratio of the execution time required on one processor and p times the time used on p processors.

Table 7.1: Cholesky factorization time on $p = 1, 2, 4$ processors.

n	non-zeros	partition	$p = 1$	$p = 2$	E_2	$p = 4$	E_4
163 693	2 679 725	3.4s	31.8s	16.1s	98.8%	9.7s	82.0%
297 302	4 884 262	7.2s	65.5s	33.5s	97.8%	17.8s	92.0%
420 881	6 909 745	11.3s	112.3s	59.3s	94.7%	33.9s	82.8%
523 989	8 626 747	14.2s	131.3s	66.4s	98.9%	40.5s	81.0%
664 539	10 921 019	20.0s	181.6s	91.1s	99.7%	50.0s	90.8%
742 470	12 192 476	22.3s	212.4s	115.6s	91.9%	60.3s	88.1%
810 412	13 284 530	25.2s	234.7s	131.2s	89.4%	70.6s	83.1%
955 968	15 715 398	29.5s	273.4s	159.6s	85.7%	82.7s	82.6%

Table 7.2: Preconditioned CG solution on $p = 1, 2, 4$ processors.

	$\varepsilon_{\mathrm{CG}} = 10^{-4}$						$\varepsilon_{\mathrm{CG}} = 10^{-6}$					
n	#It	$p = 1$	$p = 2$	E_2	$p = 4$	E_4	#It	$p = 1$	$p = 2$	E_2	$p = 4$	E_4
163 693	60	17.7s	9.3s	95%	6.2s	71%	82	24.2s	12.7s	95%	8.5s	71%
297 302	75	42.0s	24.2s	87%	14.6s	72%	103	57.6s	33.2s	87%	20.0s	72%
420 881	96	81.4s	43.5s	94%	28.8s	71%	131	110.8s	64.6s	86%	39.0s	71%
523 989	92	93.9s	50.3s	93%	33.6s	70%	124	126.6s	73.8s	85%	45.1s	70%
664 539	88	118.9s	63.6s	93%	42.2s	70%	120	162.3s	86.8s	93%	57.4s	70%
742 470	77	117.9s	69.9s	84%	42.1s	70%	106	162.3s	89.1s	91%	57.9s	70%
810 412	81	134.0s	74.0s	91%	47.3s	71%	109	180.3s	99.8s	90%	64.1s	70%
955 968	85	163.3s	92.5s	88%	58.4s	70%	115	221.0s	124.7s	89%	79.0s	70%

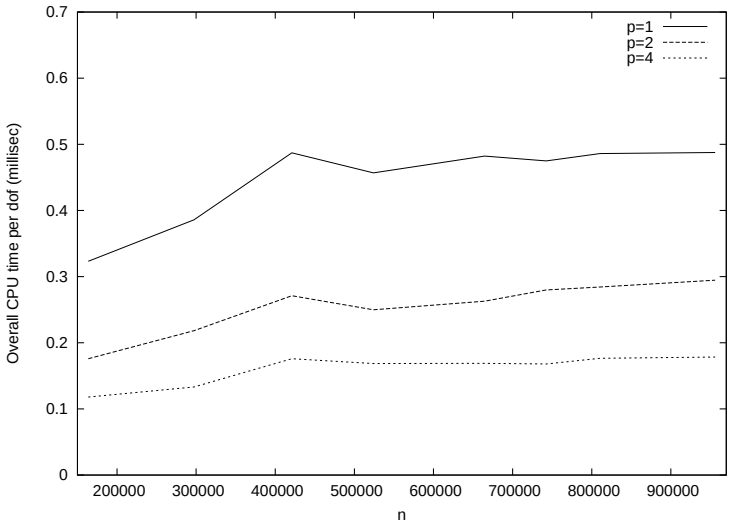

Fig. 7.2. Total solution time per dof for $p = 1, 2, 4$ processors.

The results were obtained on a system consisting of two Intel Xeon 5160 processors (dual core, 3 GHz). In all tests we have used a multiple of the identity matrix for regularization, where the regularization parameter α was set to $2\pi/\mu_0$. Notice that the size of the matrix entries in A is of the order $1/\mu_0$; see also the proof of Lemma 4.2 where $\alpha \geq 1/\mu_1$ is exploited. The rounding accuracy $\varepsilon_{\mathcal{H}}$ of the hierarchical matrix Cholesky factorization was chosen 10^{-2}.

Apparently, the complexity scales almost linearly (see Fig. 7.2 which depicts the overall solution time per degree of freedom for $\varepsilon_{\mathrm{CG}} = 10^{-4}$) and the parallelization of the hierarchical LU factorization algorithm shows a competitive speedup. The number of iterations is bounded independently of n. The major part of the total solution time is used to construct the preconditioner. This ensures a quick computation of the magnetic field in case of varying exciting currents $\mathbf{j}_0$, i.e. for multiple right-hand sides.

Acknowledgments. The authors wish to thank Ralf Hiptmair and the referees for helpful suggestions.

References

[1] D. N. Arnold, R. S. Falk, and R. Winther, Multigrid in $H(\mathrm{div})$ and $H(\mathrm{curl})$, *Numer. Math.*, **85**:2 (2000), 197-217.

[2] M. Bebendorf, Efficient inversion of Galerkin matrices of general second-order elliptic differential operators with nonsmooth coefficients, *Math. Comput.*, **74** (2005), 1179-1199.

[3] M. Bebendorf, Hierarchical *LU* decomposition based preconditioners for BEM, *Computing*, **74** (2005), 225-247.

[4] M. Bebendorf, Why finite element discretizations can be factored by triangular hierarchical matrices, *SIAM J. Numer. Anal.*, **45**:4 (2007), 1472-1494.

[5] M. Bebendorf, Hierarchical matrices: A means to efficiently solve elliptic boundary value problems, volume 63 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, 2008.

[6] M. Bebendorf and Th. Fischer, On the purely algebraic data-sparse approximation of the inverse and the triangular factors of sparse matrices, 2008, work in progress.

[7] M. Bebendorf and W. Hackbusch, Existence of $\mathcal{H}$-matrix approximants to the inverse FE-matrix of elliptic operators with L^∞-coefficients, *Numer. Math.*, **95**:1 (2003), 1-28.

[8] P. B. Bochev, C. J. Garasi, J. J. Hu, A. C. Robinson, and R. S. Tuminaro, An improved algebraic multigrid method for solving Maxwell's equations, *SIAM J. Sci. Comput.*, **25**:2 (2003), 623-642 (electronic).

[9] M. Costabel and M. Dauge, Maxwell and Lamé eigenvalues on polyhedra, *Math. Method. Appl. Sci.*, **22**:3 (1999), 243-258.

[10] A. George, Nested dissection of a regular finite element mesh, *SIAM J. Numer. Anal.*, **10** (1973), 345-363.

[11] K. Giebermann, Multilevel approximation of boundary integral operators, *Computing*, **67** (2001), 183-207.

[12] V. Girault and P.-A. Raviart, Finite Element Methods for Navier-Stokes Equations, volume 5 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1986.

[13] L. Grasedyck and W. Hackbusch, Construction and arithmetics of $\mathcal{H}$-matrices, *Computing*, **70** (2003), 295-334.

[14] W. Hackbusch, A sparse matrix arithmetic based on $\mathcal{H}$-matrices. Part I: Introduction to $\mathcal{H}$-matrices, *Computing*, **62**:2 (1999), 89-108.

[15] W. Hackbusch and B. N. Khoromskij, A sparse $\mathcal{H}$-matrix arithmetic. Part II: Application to multi-dimensional problems, *Computing*, **64**:1 (2000), 21-47.

[16] R. Hiptmair, Multigrid method for Maxwell's equations, *SIAM J. Numer. Anal.*, **36**:1 (1999), 204-225 (electronic).
[17] R. Hiptmair, Finite elements in computational electromagnetism, *Acta Numerica*, (2002), 237-339.
[18] R. Hiptmair and J. Xu, Nodal auxiliary space preconditioning in H(curl) and H(div) spaces, *SIAM J. Numer. Anal.*, **45**:6 (2007), 2483-2509 (electronic).
[19] J. E. Hopcroft and R. M. Karp, An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs, *SIAM J. Comput.*, **2**:4 (1973), 225-231.
[20] G. Karypis and V. Kumar, A fast and high quality multilevel scheme for partitioning irregular graphs, *SIAM J. Sci. Comput.*, **20**:1 (1999), 359-392.
[21] R. Kriemann, Parallel $\mathcal{H}$-matrix arithmetics on shared memory systems, *Computing*, **74**:3 (2005), 273-297.
[22] M. Lintner, Lösung der 2D Wellengleichung mittels Hierarchischer Matrizen, PhD thesis, Technische Universität München, Germany, 2002.
[23] D. G. Luenberger, Introduction to Linear and Nonlinear Programming. Addison-Wesly, Reading, 1973.
[24] J.-C. Nédélec, A new family of mixed finite elements in $\mathbb{R}^3$, *Numer. Math.*, **50**:1 (1986), 57-81.
[25] A. Pothen, H. D. Simon, and L. Wang, Spectral Nested Dissection, Technical Report RNR-92-003, Nasa Ames Research Center, Moffett Field, CA 94035, 1992.
[26] S. Reitzinger and J. Schöberl, An algebraic multigrid method for finite element discretizations with edge elements, *Numer. Linear Algebr.*, **9**:3 (2002), 223-2382.
[27] S. Rjasanow, I. Ibraghimov, and K. Straube, Hierarchical Cholesky decomposition of sparse matrices arising from curl-curl-equations, *J. Numer. Math.*, **15**:1 (2007), 31-58.