Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Existence of $\mathcal{H}$-Matrix Approximants to the Inverse FE-Matrix of Elliptic Operators with L^∞-Coefficients

by

Mario Bebendorf and Wolfgang Hackbusch

Preprint no.: 21 2002

Existence of $\mathcal{H}$-Matrix Approximants to the Inverse FE-Matrix of Elliptic Operators with L^∞-Coefficients

Mario Bebendorf and Wolfgang Hackbusch
Max-Planck-Institute for Mathematics in the Sciences
Inselstr. 22–26, D-04103 Leipzig, Germany
{bebendorf,wh}@mis.mpg.de

Abstract

This article deals with the existence of blockwise low-rank approximants — so-called $\mathcal{H}$-matrices — to inverses of FEM matrices in the case of uniformly elliptic operators with L^∞-coefficients. Unlike operators arising from boundary element methods for which the $\mathcal{H}$-matrix theory has been extensively developed, the inverses of these operators do not benefit from the smoothness of the kernel function. However, it will be shown that the corresponding Green functions can be approximated by degenerate functions giving rise to the existence of blockwise low-rank approximants of FEM inverses. Numerical examples confirm the correctness of our estimates. As a side-product we analyse the $\mathcal{H}$-matrix property of the inverse of the FE mass matrix.

1 Introduction

In a series of papers, the technique of hierarchical matrices ($\mathcal{H}$-matrices) has been introduced, which enable a cheap but sufficiently accurate representation of fully populated matrices (cf. [16], [17]). Since the method has its original from the panel clustering method (cf. [18]), it was first applied to dense matrices arising from the discretisation of boundary integral operators (see also [1]). Since hierarchical matrices allow the approximate computation of matrix-matrix multiplications and matrix inversions, also the inverse finite element (FE) stiffness matrix turns out to be computable with almost linear complexity.

A rigorous proof for the fact that the inverse A^{-1} of a FE stiffness matrix A can be approximated by means of hierarchical matrices was still missing. The heuristic argument is that A^{-1} is closely related to the Galerkin discretisation B of L^{-1}, where the inverse differential operator is written as the integral operator

$$(L^{-1}\varphi)(x) = \int_\Omega G(x,y)\varphi(y)\,dy \tag{1.1}$$

using Green's function $G(x,y)$ as Schwartz kernel. For differential operators L with constant (or analytic or at least sufficiently smooth) coefficients and for sufficiently smooth boundary $\partial\Omega$, one can show that $G(x,y)$ has the smoothness properties satisfying the following condition: Let ω_1 and ω_2 be two disjoint subsets of Ω. Then $G(x,y)$ restricted to $x \in \omega_1$ and $y \in \omega_2$ can be approximated sufficiently well by a (e.g., Taylor) polynomial $P(x,y)$. Since polynomials can be written in the separable form[1] $\sum_{i=1}^k u_i(x)v_i(y)$, we conclude that

$$G(x,y) \approx \sum_{i=1}^k u_i(x)v_i(y) \qquad \text{in } \omega_1 \times \omega_2. \tag{1.2}$$

Applying the Galerkin discretisation with FE basis functions φ_i, we obtain the matrix B, where $b_{\nu\mu} := \int_\Omega \int_\Omega \varphi_\nu(x)G(x,y)\varphi_\mu(y)dxdy$. Let I_{ω_1} and I_{ω_2} be the index sets with the property that $\operatorname{supp}(\varphi_\nu) \subset \overline{\omega}_1$ for $\nu \in I_{\omega_1}$ and $\operatorname{supp}(\varphi_\mu) \subset \overline{\omega}_2$ for $\mu \in I_{\omega_2}$. Approximating G by the right-hand side in (1.2), we get $b_{\nu\mu} \approx \tilde{b}_{\nu\mu} := \sum_{i=1}^k \int_\Omega \varphi_\nu(x)u_i(x)dx \int_\Omega \varphi_\mu(y)v_i(y)dy$. Hence, the block $(\tilde{b}_{\nu\mu})_{\nu\in I_{\omega_1},\mu\in I_{\omega_2}}$ is a rank-k matrix[2].

[1] In fact, we may write a polynomial $P(x,y)$ as $\sum_{i=1}^k p_i(x)y^{i-1}$, where $p_i(x)$ is a polynomial only in x. Therefore, set $u_i(x) := p_i(x)$ and $v_i(y) := y^{i-1}$.

[2] The rank-k matrix equals $\sum_{i=1}^k \mathbf{a}_i\mathbf{b}_i^T$ with the vectors $\mathbf{a}_i = \left(\int_\Omega \varphi_\nu(x)u_i(x)dx\right)_{\nu\in I_{\omega_1}}$ and $\mathbf{b}_i = \left(\int_\Omega \varphi_\mu(x)v_i(x)dx\right)_{\mu\in I_{\omega_2}}$.

This is the principle behind the representation of B by a hierarchical matrix $\widetilde{B}$: Blocks of appropriate size are replaced by low-rank matrices.

In short, the approximation of a dense matrix by an $\mathcal{H}$-matrix is a consequence of the exponential convergence of polynomials to certain parts $G(x,y)|_{\omega_1\times\omega_2}$ of the Green function. This argument fails if G is not smooth as it happens for non-smooth coefficients c_{ij} of the (uniformly elliptic) differential operator

$$Lu = -\sum_{i,j=1}^{d} \partial_j(c_{ij}\partial_i u) \tag{1.3}$$

or close to corners of the boundary $\partial\Omega$. In the case of $c_{ij} \in L^\infty(\Omega)$ the theorem of De Giorgi (1957; see [7, page 200]) guarantees only local Hölder continuity of G. In this article we consider the extreme case when $c_{ij} \in L^\infty(\Omega)$ and $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain, and prove that nevertheless B as well as the inverse FE stiffness matrix A^{-1} are well approximated by $\mathcal{H}$-matrices. On the other hand, we require no smoothness of the functions u_i, v_i in (1.2).

Usually, iterative methods are applied for the efficient numerical solution of elliptic partial differential equations, a prominent example are multigrid methods (cf. [12], [13, Chapter 10]). The first aim of iterative methods is a convergence rate independent[3] of the dimension of the problem ("optimality"). However, the influence of other problem parameters may still deteriorate the method and is not so easy to cure ("robustness"). Jumping coefficients and oscillatory coefficients (as it may happen for $c_{ij} \in L^\infty(\Omega)$) are two examples of this kind.

A weak point of traditional iterative methods is the treatment of arising Schur complements, since its explicit calculation is avoided but nevertheless a good preconditioning is required. This is hard to achieve for real life problems involving difficult problem parameters. The concept of $\mathcal{H}$-matrices allows to compute the Schur complement since the class of $\mathcal{H}$-matrices provides both efficient storage and efficient arithmetic of the matrix algebra.

Consequently, this article is designed to lay ground to future efficient and easy to implement algorithms for the solution of elliptic partial differential equations with extremely general coefficients. The efficient treatment of the inverse of the stiffness matrix might be used for (a) the direct solution of FEM systems, (b) for preconditioning another iterative method or (c) for the calculation of a Schur complement. It is interesting to remark that the easily available inverse enables also the calculation of matrix functions (e.g., $\exp(-tA)$; cf. [5]) or the solution of matrix equations (e.g., the Riccati equation; cf. [8]).

Since we emphasise the rather weak conditions $c_{ij} \in L^\infty(\Omega)$ on the coefficients and "Ω bounded Lipschitz" on the domain, we simplify other aspects in order not the distract the attention of the reader by other complications. These simplifications are listed below.

1. We consider L to be an differential operator (1.3) consisting only of the principal part. Lower order terms cause no problem as long as we can guarantee L^{-1} to exist. A *dominant* low order term changes the situation, since we obtain a singularly perturbed problem.

2. We consider a *second* order differential operator L as in (1.3).

3. L is assumed to be a scalar operator, systems are not considered.

4. L is assumed to be uniformly elliptic. Although the numerical examples presented in Section 6 do not show a dependence on the ratio $\sup\{\lambda_{\max}(x)/\lambda_{\min}(x) : x \in \Omega\}$ of the eigenvalues of the matrix $(c_{ij})_{i,j=1}^d$, the proof requires its boundedness.

5. The spatial dimension is assumed to be $d \geq 3$. This is not such restrictive since $d = 3$ is the interesting case. The true reason is that the result quoted from [11] is formulated only for $d \geq 3$, although there is no indication why it should not hold for $d = 1, 2$.

6. We consider Dirichlet boundary conditions, hence the Green functions satisfies zero boundary conditions.

7. As discretisation we require a finite element discretisation with a quasi-uniform triangulation. For other discretisations the proofs may be more involved, but there is no practical reason why they should behave worse. Adaptive meshes are no problem for hierarchical matrices (see [10]).

[3] A logarithmic dependence can be tolerated.

8. The estimates in Subsection 2.5 are proven for convex domains D_2. These domains D_2 will later correspond to cluster sets X in $\mathbb{R}^d$. Although the clusters X are in general not convex, they are usually constructed in such a way that $X \subset X_c$ and X_c is convex. Examples for X_c are Chebyshev spheres (cf. [18]) or bounding boxes (cf. [2]). Therefore, there is no need for a generalisation, although the estimates could be extended to non-convex domains D_2.

The structure of the rest of the article is as follows: Section 2 is devoted to the existence of degenerate approximations to the Green function G corresponding to the underlying boundary. The Green function G allows to define the solution operator by the integral operator (1.1). Its Galerkin discretisation with respect to the FE functions from above yields the matrix B. In Section 3 we show that B possesses the $\mathcal{H}$-matrix structure. The inverse stiffness matrix A^{-1} and B are connected via the mass matrix M which is considered in Section 4. Again M^{-1} can be approximated by an $\mathcal{H}$-matrix. Finally, in Section 5, we represent the inverse A^{-1} of the FE stiffness matrix by means of the foregoing quantities. Using results on the algebra of $\mathcal{H}$-matrices, we obtain the desired $\mathcal{H}$-matrix property of A^{-1}. Section 6 contains results of numerical tests that confirm the estimates made in this article.

2 Analysis of the Green Function

2.1 The Differential Operator

Let $L : V \to V'$ ($V = H_0^1(\Omega)$) be an (scalar) uniformly elliptic operator in divergence form

$$Lu = -\sum_{i,j=1}^{d} \partial_j(c_{ij}\partial_i u) \tag{2.1}$$

in a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d \geq 3$. The coefficient matrix $C = C(x) = (c_{ij})_{ij}$, $c_{ij} \in L^\infty(\Omega)$, shall be symmetric with

$$0 < \lambda_{\min} \leq \lambda \leq \lambda_{\max} \tag{2.2}$$

for all eigenvalues λ of $C(x)$ and almost all $x \in \Omega$. The ratio $\kappa_C = \lambda_{\max}/\lambda_{\min}$ is an upper bound on almost all spectral condition numbers $\mathrm{cond}_{\|\cdot\|_2} C(x)$. Under these assumptions it is shown in [11] that in the case $d \geq 3$, a Green function $G : \Omega \times \Omega \to \mathbb{R} \cup \{\infty\}$ with the following properties exists:

$$G(\cdot, y) \in H^1(\Omega \setminus B_r(y)) \cap W_0^{1,1}(\Omega) \qquad \text{for all } y \in \Omega \text{ and all } r > 0, \tag{2.3a}$$

$$a(G(\cdot, y), \varphi) = \varphi(y) \qquad \text{for all } \varphi \in C_0^\infty(\Omega) \text{ and } y \in \Omega, \tag{2.3b}$$

where $B_r(y)$ is the open ball centred at y with radius r and

$$a(u, v) = \int_\Omega \sum_{i,j=1}^{d} c_{ij}(\partial_i u)(\partial_j v)\, dx \tag{2.4}$$

(see also [4]). Furthermore, for $x, y \in \Omega$ it holds that

$$|G(x, y)| \leq \frac{c(d, \kappa_C)}{\lambda_{\min}} |x - y|^{2-d}. \tag{2.5}$$

Since L is uniformly elliptic, $L^{-1} : V' \to V$ exists and $\|L^{-1}\|_{V \leftarrow V'} \leq C\lambda_{\min}^{-1}$ ($\lambda_{\min}$ is a matter of scaling). We will make use of the characteristic relation between L^{-1} and G, which is equivalent to (2.3b):

$$(L^{-1}\varphi)(x) = \int_\Omega G(x, y)\varphi(y)\, dy \qquad \text{for all } \varphi \in C_0^\infty(\Omega). \tag{2.6}$$

2.2 Approximation by Finite Dimensional Subspaces

In the following lemmata $D \subset \mathbb{R}^d$ is a domain. All distances and diameters use the Euclidean norm in $\mathbb{R}^d$ except the distance of functions which uses the $L^2(D)$-norm. The constant c_{appr} in (2.7) depends only on the spatial dimension d.

Lemma 2.1 *Let $D \subset \mathbb{R}^n$ be a convex domain and X a closed subspace of $L^2(D)$. Then for any $k \in \mathbb{N}$ there is a subspace $V_k \subset X$ satisfying $\dim V_k \le k$ so that*

$$\operatorname{dist}_{L^2(D)}(u, V_k) \le c_{appr} \frac{\operatorname{diam}(D)}{\sqrt[d]{k}} \|\nabla u\|_{L^2(D)} \qquad \text{for all } u \in X \cap H^1(D). \tag{2.7}$$

Proof. (a) First we assume $k = \ell^d$ and $D \subset Q = \{x \in \mathbb{R}^d : \|x - z\|_\infty < \frac{1}{2}\operatorname{diam}(D)\}$ for some $z \in \mathbb{R}^d$. We subdivide the cube Q uniformly into k subcubes Q_i, $i = 1, \dots, k$, and set $D_i = D \cap Q_i$, $i = 1, \dots, k$. Each of the sets D_i is convex with $\operatorname{diam}(D_i) \le \frac{\sqrt[d]{2}}{\ell} \operatorname{diam}(D)$. Let

$$W_k = \{v \in L^2(D) : v \text{ is constant on } D_i \text{ for all } i = 1, \dots, k\}.$$

Then $\dim W_k \le k$ and according to Poincaré's inequality for $u \in H^1(D)$ (in particular, we use the convex version in [20] with explicitly given constant) it holds that

$$\int_{D_i} |u - \bar{u}_i|^2 dx \le \pi^{-2} \operatorname{diam}^2(D_i) \int_{D_i} |\nabla u|^2 dx,$$

where $\bar{u}_i = \operatorname{vol}(D_i)^{-1} \int_{D_i} u \, dx$ is the mean value of u in D_i. Summation over all i yields

$$\operatorname{dist}_{L^2(D)}(u, W_k) \le \|u - \bar{u}\|_{L^2(D)} \le \frac{\sqrt[d]{2}}{\pi\ell} \operatorname{diam}(D) \, \|\nabla u\|_{L^2(D)}$$

for $\bar{u}$ defined by $\bar{u}|_{D_i} = \bar{u}_i$.

(b) For general $k \in \mathbb{N}$, choose $\ell := \lfloor \sqrt[d]{k} \rfloor \in \mathbb{N}$, i.e., $\ell^d \le k < (\ell + 1)^d$. Applying Part (a) for $k' := \ell^d$, we use the space $W_k := W_{k'}$ satisfying $\dim W_k = \dim W_{k'} \le k' \le k$. Using $\frac{1}{\ell} \le \frac{2}{\ell+1} < \frac{2}{\sqrt[d]{k}}$, we arrive at

$$\operatorname{dist}_{L^2(D)}(u, W_k) \le c_{appr} \frac{\operatorname{diam}(D)}{\sqrt[d]{k}} \, \|\nabla u\|_{L^2(D)}$$

with the constant $c_{appr} := 2\sqrt[d]{2}\, c_d$.

(c) Let $P : L^2(D) \to X$ be the $L^2(D)$-orthogonal projection onto X and $V_k = P(W_k)$. Keeping in mind that P has norm one and $u \in X$, the assertion follows from $\|u - P\bar{u}\|_{L^2(D)} = \|P(u - \bar{u})\|_{L^2(D)} \le \|u - \bar{u}\|_{L^2(D)}$. ∎

In the last proof we have restricted D_i to convex domains though Poincaré's inequality holds whenever the embedding $H^1(D_i) \hookrightarrow L^2(D_i)$ is compact (cf. [21]). This is for example true if D_i fulfils a uniform cone condition. However, in this case it is not obvious how the constant depends on the geometry.

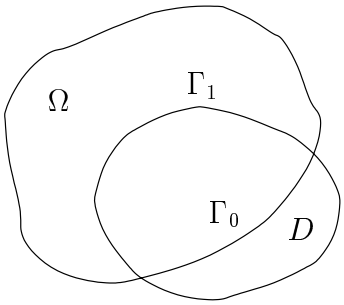
2.3 Space of L-Harmonic Functions

The Green function $G(x, \cdot)$ is a special example of an L-harmonic function in a subdomain $D_\Omega \subset \Omega$ (provided $x \notin D_\Omega$) with zero boundary values on $\partial\Omega \cap \overline{D_\Omega}$. The space X in Lemma 2.1 will be substituted by a function space $X(D) \subset L^2(D)$ which we define next. While the notation $X(D)$ will be used for different D, the underlying domain Ω is fixed.

Let D be a domain intersecting Ω: $D_\Omega := D \cap \Omega \neq \emptyset$. The boundary ∂D_Ω consists of two parts:

$$\Gamma_0(D) := D \cap \partial\Omega, \quad \Gamma_1(D) := \partial D_\Omega \backslash \Gamma_0(D) = \partial D \cap \overline{\Omega}. \tag{2.8}$$

$\Gamma_0 = \emptyset$ holds in the cases of $D \subset \Omega$ or $D \supset \Omega$. The former case may happen, whereas the latter is of no interest for us.



If D is not a subset of Ω, we require that outside of Ω functions $u \in X(D)$ are extended by zero. The functions $u \in X(D)$ are locally in $H^1(D)$ relative to $\Gamma_1(D)$ (notation: $u \in H^1_{rl,\Omega}(D)$) in the following sense:

$$H^1_{rl,\Omega}(D) := \{u \in L^2(D) : u|_{D\backslash\Omega} = 0, \, u \in H^1(K) \text{ for all } K \subset D \text{ with } \operatorname{dist}(K, \Gamma_1(D)) > 0\}. \tag{2.9}$$

The first condition is empty if $D \subset \Omega$.

The L-harmonicity[4] is required in the weak formulation of $Lu = 0$,

$$a(u,\varphi) = 0 \qquad \text{for all } \varphi \in C_0^\infty(D_\Omega) \qquad\qquad (D_\Omega = D \cap \Omega) \tag{2.10}$$

with $a(\cdot,\cdot)$ from (2.4). The final definition is

$$X(D) := \{u \in L^2(D) \cap H^1_{rl,\Omega}(D) : u \text{ satisfies (2.10)}\}. \tag{2.11}$$

The Green function $G(x,\cdot)$ can be extended to D by zero. This extension is in $H^1(D)$ and hence in $X(D)$ if $x \in \Omega \setminus \overline{D}$.

Lemma 2.2 *The space $X(D)$ is closed in $L^2(D)$.*

The proof is postponed to the next subsection, since it needs Lemma 2.4. The closeness of $X(D)$ is necessary in order to use $X(D)$ as X in Lemma 2.1.

Remark 2.3 *Consider $X(D)$ and $X(D')$ for two domains $D' \subset D$ intersecting Ω.*

(a) For any $u \in X(D)$, the restriction $u|_{D'}$ belongs to $X(D')$; hence, in short notation, $X(D)|_{D'} = X(D')$. If $\operatorname{dist}(D', \Gamma_1(D)) > 0$, even $X(D)|_{D'} = X(D') \cap H^1(D')$ holds (cf. (2.9)).

(b) The relevant parts of D and D' are $D_\Omega = D \cap \Omega$, $D'_\Omega = D' \cap \Omega$ as well as $\Gamma_0(D)$ and $\Gamma_0(D')$. D_Ω and D'_Ω are the domains of L-harmonicity, whereas $\Gamma_0(D)$ and $\Gamma_0(D')$ describe the location of zero boundary values due to the zero extension outside. As long as $D = D'$ and $\Gamma_0 = \Gamma'_0$, differences in $D \setminus \overline{\Omega}$ and $D' \setminus \overline{\Omega}$ are irrelevant, since functions from $X(D)$ and $X(D')$ vanish in these parts anyway.

2.4 The Caccioppoli Inequality

The following lemma shows that any function $u \in X(D)$ allows to estimate $\|\nabla u\|_{L^2(K_\Omega)}$ for a domain $K \subset D$ not touching $\Gamma_1(D)$ by means of the weaker norm $\|u\|_{L^2(D_\Omega)}$. Note that K may contain parts of $\Gamma_0(D)$.

Lemma 2.4 *Let $X(D)$, $\Gamma_0(D)$, $\Gamma_1(D)$ as in (2.11), (2.8), and $K \subset D$, $K_\Omega = K \cap \Omega$ with $\operatorname{dist}(K, \Gamma_1(D)) > 0$. Further, let $\kappa_C = \lambda_{\max}/\lambda_{\min}$ (cf. (2.2)). Then the so-called* Caccioppoli inequality *holds:*

$$\|\nabla u\|_{L^2(K_\Omega)} \le \frac{4\sqrt{\kappa_C}}{\operatorname{dist}(K, \Gamma_1(D))} \|u\|_{L^2(D_\Omega)} \qquad \textit{for all } u \in X(D). \tag{2.12}$$

Proof. The proof follows the lines of [6]. Let $\eta \in C^1(D)$ satisfy $0 \le \eta \le 1$, $\eta = 1$ in K, $\eta = 0$ in a neighbourhood of $\Gamma_1(D)$ and[5] $|\nabla \eta| \le 2/\delta$ in D_Ω, where we set $\delta = \operatorname{dist}(K, \Gamma_1(D))$. Since $K' := \operatorname{supp}(\eta) \subset D$ satisfies $\operatorname{dist}(K', \Gamma_1(D)) > 0$, (2.9) implies $u \in H^1(K')$. Hence, $\varphi := \eta^2 u \in H^1_0(D_\Omega)$ may be used as a test function in $a(u,\varphi) = 0$:

$$0 = \int_{D_\Omega} (\nabla u)^T C(x) \nabla(\eta^2 u) dx = 2 \int_{D_\Omega} \eta u (\nabla u)^T C(x) (\nabla \eta) dx + \int_{D_\Omega} \eta^2 (\nabla u)^T C(x) (\nabla u) dx.$$

¿From (2.2) it follows that

$$\begin{aligned}
\int_{D_\Omega} \eta^2 |C^{1/2}(x) \nabla u|^2 dx &= \left| \int_{D_\Omega} \eta^2 (\nabla u)^T C(x) (\nabla u) dx \right| = 2 \left| \int_{D_\Omega} \eta u (\nabla u)^T C(x) (\nabla \eta) dx \right| \\
&\le 2 \int_{D_\Omega} \eta\, |u|\, |C^{1/2}(x) \nabla \eta|\, |C^{1/2}(x) \nabla u|\, dx \\
&\le 4 \frac{\sqrt{\lambda_{\max}}}{\delta} \int_{D_\Omega} |u| \left(\eta |C^{1/2}(x) \nabla u| \right) dx \\
&\le 4 \frac{\sqrt{\lambda_{\max}}}{\delta} \left(\int_{D_\Omega} \eta^2 |C^{1/2}(x) \nabla u|^2 dx \right)^{1/2} \|u\|_{L^2(D_\Omega)},
\end{aligned}$$

[4] To be precise, we need L^*-harmonicity, since the Green function $G(x,y)$ is L-harmonic w.r.t. x but L^*-harmonic w.r.t. y. However, since here L consists only of the principle part (2.1), L is self-adjoined. But notice that symmetry is not at all essential.

[5] The estimate $|\nabla \eta| \le c/\delta$ can be fulfilled for all $c > 1$. Hence, in (2.12) the factor 4 may be replaced by $2c$. Since it is true for all $2c > 2$, it follows also for 2 instead of 4.

i.e., $\|\eta\, C^{1/2}(x)\nabla u\|_{L^2(D_\Omega)} \le 4\frac{\sqrt{\lambda_{\max}}}{\delta}\|u\|_{L^2(D_\Omega)}$. The estimation by

$$\|\nabla u\|_{L^2(K_\Omega)} \le \|\eta\nabla u\|_{L^2(D_\Omega)} \le \lambda_{\min}^{-1/2}\|\eta C^{1/2}(x)\nabla u\|_{L^2(D_\Omega)}$$

yields the assertion. ∎

Remark 2.5 *Since $u = 0$ in $D \setminus \Omega$, we may write the norms in the inequality of Lemma 2.4 as $\|\nabla u\|_{L^2(K)}$ and $\|u\|_{L^2(D)}$ (i.e., K instead of K_Ω and D instead of D_Ω).*

Proof of Lemma 2.2. Let $\{u_k\}_{k\in\mathbb{N}} \subset X(D)$ converge to u in $L^2(D)$. Let $K \subset D$ with $\operatorname{dist}(K, \Gamma_1(D)) > 0$. According to Remark 2.5, the sequence $\{\nabla u_k\}_{k\in\mathbb{N}}$ is bounded on K,

$$\|\nabla u_k\|_{L^2(K)} \le c\,\|u_k\|_{L^2(D)} \le C.$$

Due to the Banach-Alaoglu Theorem, a subsequence $\{u_{i_k}\}_{k\in\mathbb{N}}$ converges weakly in $H^1(K)$ to $\hat{u} \in H^1(K)$. Hence, for any $v \in L^2(K)$ we have $(u, v)_{L^2(K)} = \lim_{k\to\infty}(u_{i_k}, v)_{L^2(K)} = (\hat{u}, v)_{L^2(K)}$ proving $u = \hat{u} \in H^1(K)$. Since the functional $a(\cdot, \varphi)$ for $\varphi \in C_0^\infty(D_\Omega)$ is in $(H^1(K))'$, we see by the same argument that $a(u, \varphi) = 0$. Finally, $u_k|_{D\setminus\Omega} = 0$ leads to $u|_{D\setminus\Omega} = 0$. Hence, $u \in X(D)$ is shown. ∎

2.5 Main Theorem

First we investigate how large the dimension of a finite dimensional subspace must be to approximate a function from $X(D)$ in a subdomain D_2 of D up to a certain error.

Lemma 2.6 *Let D, $\Gamma_1(D)$, D_Ω and $X(D)$ as before (cf. Lemma 2.4) and assume that $D_2 \subset D$ is a convex domain such that*

$$\operatorname{dist}(D_2, \partial D) \ge \rho\,\operatorname{diam}(D_2) > 0.$$

Then for any $M > 1$ there is a subspace $W \subset X(D_2)$ so that

$$\operatorname{dist}_{L^2(D_2)}(u, W) \le \frac{1}{M}\,\|u\|_{L^2(D_\Omega)} \qquad \text{for all } u \in X(D) \tag{2.13}$$

and

$$\dim W \le c_\rho^d \lceil \log M\rceil^{d+1} + \lceil \log M\rceil, \quad c_\rho = 4ec_{appr}\sqrt{\kappa_C}\,\frac{1+2\rho}{\rho}. \tag{2.14}$$

Proof. (a) Consider $K(r) := \{x \in \mathbb{R}^d : \operatorname{dist}(x, D_2) \le r\}$ for $0 \le r \le \operatorname{dist}(D_2, \partial D)$. We conclude that $K(r)$ are again convex domains which are increasing with r: $K(r_1) \subset K(r_2)$ for $r_1 \le r_2$. The smallest is $K(0) = D_2$, while $K(\operatorname{dist}(D_2, \partial D))$ is the largest one which is still in D. We remark that $\operatorname{dist}(K(r_1), \partial K(r_2)) = r_2 - r_1$ for $r_1 \le r_2$ and $\operatorname{diam}(K(r)) \le \operatorname{diam}(D_2) + 2r$.

(b) Consider the sequence $r_0 > r_1 > \ldots > r_i = 0$ with $r_j := (1 - j/i)\operatorname{dist}(D_2, \partial D)$, where i is chosen later. Using $K(r)$ from Part (a) we set

$$D^j := K(r_j), \quad X^j := X(D^j) \qquad \text{(cf. (2.11))}$$

and notice that $D_2 = D^i \subset D^{i-1} \subset \ldots \subset D^0 \subset D$.

(c) Let $j \in \{1, \ldots, i\}$. Applying Lemma 2.4 (Remark 2.5) with (D^{j-1}, D^j) instead of (K, D), we obtain

$$\|\nabla v\|_{L^2(D^j)} \le \frac{4\sqrt{\kappa_C}}{\operatorname{dist}(D^j, \Gamma_1(D^{j-1}))}\,\|v\|_{L^2(D^{j-1})} \qquad \text{for all } v \in X^{j-1}$$

(we recall $\Gamma_1(D^{j-1}) = \partial D^{j-1} \cap \overline{\Omega}$). Because of $\operatorname{dist}(D^j, \Gamma_1(D^{j-1})) \ge \operatorname{dist}(D^j, \partial D^{j-1}) = r_{j-1} - r_j = r_0/i$ (see Part (a)), the resulting estimate is

$$\|\nabla v\|_{L^2(D^j)} \le \frac{4i\sqrt{\kappa_C}}{r_0}\,\|v\|_{L^2(D^{j-1})} \qquad \text{for all } v \in X^{j-1}. \tag{2.15}$$

(d) Apply Lemma 2.1 with D^j instead of D and with the choice $k := \lceil (Bi)^d\rceil$, where the factor B will be adjusted later. Then this Lemma ensures that there is a subspace $V_j \subset X^j$ satisfying $\dim V_j \le k$ and

$$\operatorname{dist}_{L^2(D^j)}(v, V_j) \le c_{appr}\,\frac{\operatorname{diam}(D^j)}{\sqrt[d]{k}}\,\|\nabla v\|_{L^2(D^j)} \qquad \text{for all } v \in X^j \cap H^1(D^j).$$

Using $\sqrt[d]{k} \geq Bi$ and $\operatorname{diam}(D^j) = \operatorname{diam}(D_2) + 2r_j \leq \operatorname{diam}(D_2) + 2r_0$ (see Part (a)), we arrive at

$$\operatorname{dist}_{L^2(D^j)}(v, V_j) \leq c_{appr} \frac{\operatorname{diam}(D_2) + 2r_0}{Bi} \|\nabla v\|_{L^2(D^j)} \qquad \text{for all } v \in X^j \cap H^1(D^j). \tag{2.16}$$

Since any $v \in X^{j-1}$ also belongs to $X^j \cap H^1(D^j)$, the estimates (2.15), (2.16) together with $r_0 \geq \rho \operatorname{diam}(D_2)$ may be combined to

$$\operatorname{dist}_{L^2(D^j)}(v, V_j) \leq \frac{1+2\rho}{\rho} \frac{4c_{appr}\sqrt{\kappa_C}}{B} \|v\|_{L^2(D^{j-1})} \qquad \text{for all } v \in X^{j-1}. \tag{2.17}$$

In particular, the factor $\frac{1+2\rho}{\rho} \frac{4c_{appr}\sqrt{\kappa_C}}{B}$ becomes $M^{-1/i}$ for the choice

$$B := B_0 M^{1/i} \quad \text{with } B_0 := 4c_{appr}\sqrt{\kappa_C}\frac{1+2\rho}{\rho}. \tag{2.18}$$

(e) For any given $u =: v_0 \in X^0$, (2.17) and (2.18) lead to $v_0|_{D^1} = u_1 + v_1$ with $u_1 \in V_1$ and

$$\|v_1\|_{L^2(D^1)} \leq M^{-1/i} \|v_0\|_{L^2(D^0)}.$$

Consequently, v_1 belongs to X^1. Similarly, for all $j = 1, \ldots, i$ we are able to find an approximant $u_j \in V_j$ so that $v_{j-1}|_{D^j} = u_j + v_j$ and $\|v_j\|_{L^2(D^j)} \leq M^{-1/i} \|v_{j-1}\|_{L^2(D^{j-1})}$. Hence, the subspace

$$W := \operatorname{span}\{V_j|_{D_2} : j = 1, \ldots, i\}$$

using the restrictions of V_j to the smallest domain $D_2 = D^i$ contains $u_j|_{D_2} \in V_j|_{D_2} \subset W$. Therefore, $v_0 = v_i + \sum_{j=1}^{i} u_j$ leads to

$$\operatorname{dist}_{L^2(D_2)}(v_0, W) \leq \|v_i\|_{L^2(D_2)} \leq \left(M^{-1/i}\right)^i \|v_0\|_{L^2(D^0)} \leq M^{-1}\|u\|_{L^2(D_\Omega)},$$

where the last inequality is due to $D^0 \subset D$ and $u|_{D\setminus\Omega} = 0$.

(f) The dimension of W is bounded by $\sum_{j=1}^{i} \dim V_j = i\lceil (Bi)^d \rceil \leq i + B^d i^{d+1}$. The choice $i := \lceil \log M \rceil$ yields

$$\dim W \leq \lceil \log M \rceil + B_0^d e^d \lceil \log M \rceil^{d+1}$$

because of $B = B_0 M^{1/i} \leq B_0 e$. Together with $c_\rho = B_0 e$, we obtain the final result. ∎

Remark 2.7 *(a) Setting $M = \exp(m)$, the dimension of W is bounded by $c_\rho^d \lceil m \rceil^{d+1} + \lceil m \rceil \sim c_\rho^d m^{d+1}$. On the other hand, if a dimension $K = \dim W$ is given, the possible improvement factor $\frac{1}{M} = \exp(-m)$ is described by $m \gtrsim (c_\rho K)^{1/(d+1)}/c_\rho$.*

(b) The factor $\frac{1+2\rho}{\rho}$ in (2.14) shows that ρ should be of order $\mathcal{O}(1)$, e.g., $\operatorname{dist}(D_2, \partial D) \geq \operatorname{diam}(D_2)$ is a reasonable choice.

Next we consider the Green functions $G(x, \cdot)$ with $x \in D_1 \subset \Omega$, which are L-harmonic in $\Omega \setminus \overline{D_1}$. Note that its approximant $G_k(x, \cdot)$ from the following theorem is of the desired form (1.2).

Theorem 2.8 *Let $D_1, D_2 \subset \Omega$ be two domains such that D_2 is convex and*

$$\operatorname{dist}(D_1, D_2) \geq \rho \operatorname{diam}(D_2) > 0.$$

Then for any $\varepsilon \in (0, 1)$ there is a separable approximation

$$G_k(x, y) = \sum_{i=1}^{k} u_i(x) v_i(y) \quad \text{with } k \leq k_\varepsilon = c_{\rho/2}^d \lceil \log \frac{1}{\varepsilon} \rceil^{d+1} + \lceil \log \frac{1}{\varepsilon} \rceil,$$

where c_ρ is defined in (2.14), so that

$$\|G(x, \cdot) - G_k(x, \cdot)\|_{L^2(D_2)} \leq \varepsilon \|G(x, \cdot)\|_{L^2(\hat{D}_2)} \qquad \text{for all } x \in D_1, \tag{2.19}$$

where $\hat{D}_2 := \{y \in \Omega : \operatorname{dist}(y, D_2) \leq \frac{\rho}{2} \operatorname{diam}(D_2)\}$.

Proof. Let $D = \{y \in \mathbb{R}^d : \operatorname{dist}(y, D_2) \le \frac{\rho}{2}\operatorname{diam}(D_2)\}$. Note that $\hat{D}_2 = D \cap \Omega$ and that because of $\operatorname{dist}(\hat{D}_2, D_1) \ge \operatorname{dist}(D, D_1) = \operatorname{dist}(D_2, D_1) - \frac{\rho}{2}\operatorname{diam}(D_2) \ge \frac{\rho}{2}\operatorname{diam}(D_2) > 0$, the right-hand side $\|G(x,\cdot)\|_{L^2(\hat{D}_2)}$ does not contain the singularity of G (cf. (2.5)).

Since $\operatorname{dist}(D_2, \partial D) = \frac{\rho}{2}\operatorname{diam}(D_2)$, we can apply Lemma 2.6 with $M = \varepsilon^{-1}$ and ρ replaced by $\rho/2$. Let $\{v_1, \ldots, v_k\}$ be a basis of the subspace $W \subset X(D_2)$ with $k = \dim W \le c_{\rho/2}^d \lceil \log \frac{1}{\varepsilon} \rceil^{d+1} + \lceil \log \frac{1}{\varepsilon} \rceil$ according to Lemma 2.6.

For any $x \in D_1$, the function $g_x := G(x,\cdot)$ is in $X(D)$. By means of (2.13), $g_x = \hat{g}_x + r_x$ holds with $\hat{g}_x \in W$ and $\|r_x\|_{L^2(D_2)} \le \varepsilon \|g_x\|_{L^2(\hat{D}_2)}$. Expressing $\hat{g}_x$ by means of the basis, we obtain

$$\hat{g}_x = \sum_{i=1}^{k} u_i(x) v_i$$

with coefficients $u_i(x)$ depending on the index x. Since x varies in D_1, the u_i are functions defined on D_1. The function $G_k(x,y) := \sum_{i=1}^k u_i(x) v_i(y)$ satisfies estimate (2.19). ∎

Remark 2.9 *Without loss of generality, we may choose $\{v_1, \ldots, v_k\}$ as an* orthogonal *basis of W. Then the coefficients $u_i(x)$ in the latter expansion equal $(G(x,\cdot), v_i)_{L^2(D_2 \cap \Omega)}$ showing that the u_i's satisfy $Lu_i = v_i$ with homogeneous Dirichlet boundary conditions. In particular, u_i is L-harmonic in $\Omega \setminus D_2$. Note that the u_i's do not depend on D_1.*

For later use, we add a trivial remark.

Remark 2.10 *Assume (2.19) and $E \subset D_1$. Then $\|G - G_k\|_{L^2(E \times D_2)} \le \varepsilon \|G\|_{L^2(E \times \hat{D}_2)}$.*

Theorem 2.8 can be easily adjusted to fundamental solutions S,

$$L_x S(x,y) = \delta(x-y) \qquad \text{for all } x, y \in \mathbb{R}^d,$$

which play a central role for example in boundary element methods (BEM). The following corollary guarantees that we are able to treat BEM matrices by $\mathcal{H}$-matrices.

Corollary 2.11 *Assume that a fundamental solution S exists for L. Let $D_1, D_2 \subset \mathbb{R}^d$ be two domains with D_2 being convex and*

$$\operatorname{dist}(D_1, D_2) \ge \rho \operatorname{diam}(D_2) > 0.$$

Then for $\varepsilon > 0$ there is $S_k(x,y) = \sum_{i=1}^k u_i(x) v_i(y)$ with $k \le k_\varepsilon = c_{\rho/2}^d \lceil \log \frac{1}{\varepsilon} \rceil^{d+1} + \lceil \log \frac{1}{\varepsilon} \rceil$, where c_ρ is defined in (2.14), so that

$$\|S(x,\cdot) - S_k(x,\cdot)\|_{L^2(D_2)} \le \varepsilon \|S(x,\cdot)\|_{L^2(\hat{D}_2)} \qquad \text{for all } x \in D_1,$$

where $\hat{D}_2 = \{x \in \mathbb{R}^d : \operatorname{dist}(x, D_2) \le \frac{1}{2}\operatorname{dist}(D_1, D_2)\}$.

3 The Discrete Green Function Integral Operator

3.1 The Finite Element Discretisation

V is the function space introduced in Section 2.1. In the (conforming) finite element discretisation, V is approximated by $V_h \subset V$. Let $n = \dim V_h$ be its dimension and $\{\varphi_i\}_{i \in I}$ a basis, where $I = \{1, \ldots, n\}$ is used as index set. The notation for the support of the finite element basis function is generalised to subsets $\tau \subset I$ as follows:

$$X_i := \operatorname{supp} \varphi_i \quad \text{for } i \in I, \qquad X_\tau := \bigcup_{i \in \tau} X_i \quad \text{for } \tau \subset I. \tag{3.1}$$

In order to avoid technical complications, we consider a *quasi-uniform* and *shape-regular* triangulation. Hence, the step size $h := \max_{i \in I} \operatorname{diam}(X_i)$ fulfils

$$\operatorname{vol}(X_i) \ge c_v h^d. \tag{3.2}$$

The supports X_i may overlap. In accordance with the standard finite element discretisation we require that each triangle belongs to the support of a bounded number of basis functions, i.e., there is a constant $c_M > 0$ so that

$$c_M \operatorname{vol}(X_\tau) \ge \sum_{i \in \tau} \operatorname{vol}(X_i). \tag{3.3}$$

We use the notation J for the natural bijection $J : \mathbb{R}^n \to V_h$ defined by $J\mathbf{x} = \sum_{i\in I} x_i\varphi_i$. For quasi-uniform and shape-regular triangulations it is known that there are constants $0 < c_{J,1} \le c_{J,2}$ (independent of h and n) such that

$$c_{J,1}\|\mathbf{x}\|_h \le \|J\mathbf{x}\|_{L^2(\Omega)} \le c_{J,2}\|\mathbf{x}\|_h \qquad \text{for all } \mathbf{x} \in \mathbb{R}^n, \tag{3.4}$$

where $\|\mathbf{x}\|_h = \sqrt{h^d \sum_{i\in I} x_i^2}$ is the (naturally scaled) Euclidean norm (cf. [13, Theorem 8.8.1]). Correspondingly, we use the scalar product $\langle \mathbf{x}, \mathbf{y}\rangle_h = h^d \sum_{i\in I} x_i y_i$.

Since J is also a function from $\mathbb{R}^n$ into V, the adjoint $J^* \in L(V', \mathbb{R}^n)$ is defined. We define the following three $n \times n$ matrices,

$$A = J^*LJ, \quad B = J^*L^{-1}J, \quad \text{and} \quad M = J^*J.$$

A is the stiffness matrix, B the Galerkin discretisation of the inverse of L, and M the mass matrix. The matrices A and M are sparse, while B as well as A^{-1} and M^{-1} are dense.

3.2 Admissible Partitions and $\mathcal{H}$-Matrices

In the following, τ and σ denote subsets of the index set I. The set $X_\tau \subset \Omega$ has already been defined in (3.1). A block of an $n \times n$ matrix is characterised by the pair product $\tau \times \sigma$ (τ contains the row indices, σ the column indices). P is a *partition* of $I \times I$, if it contains elements of the form $\tau \times \sigma$ ($\tau, \sigma \subset I$) such that (3.5a) holds[6]:

$$I \times I = \bigcup\nolimits_{\tau\times\sigma\in P} \tau \times \sigma \text{ (disjoint union)}, \tag{3.5a}$$

$$\operatorname{dist}(X_\tau, X_\sigma) \ge \rho \max\{\operatorname{diam}(X_\tau), \operatorname{diam}(X_\sigma)\} > 0 \quad \text{or} \quad \min\{\#\tau, \#\sigma\} = 1. \tag{3.5b}$$

If, in addition, (3.5b) is satisfied, P is called[7] an *admissible partition* of $I \times I$.

The desirable properties of the hierarchical matrices are based on the fact that the "clusters" τ, σ appearing in P are *hierarchically* generated. In particular, this allows the cheap computation of the minimal admissible partition (with $\mathcal{O}(n \log n)$ cost, see [1], [8], [16], [17]).

The hierarchical structure is based on a cluster tree $T(I)$ which may be assumed to be a binary tree: I is the root and each $\tau \in T(I)$ is a subset of I which either contains only one index ($\#\tau = 1$) or is the disjoint union of its two sons $\tau', \tau'' \in T(I)$. Let

$$T_\ell(I) := \{\tau \in T(I) : \text{the path from } I \text{ to } \tau \text{ has length } \ell\},$$

e.g., I is the only element of $T_0(I)$. The maximal level L with $T_L(I) \neq \emptyset$ is of the size $\mathcal{O}(\log n)$. The blocks $b = \tau \times \sigma$ are constructed such that both τ and σ belong to the same level. Therefore, P is the union of the sets $P_\ell = \{b = \tau \times \sigma : \tau, \sigma \in T_\ell(I)\}$, $0 \le \ell \le L$.

The hierarchical structure does not enter the proofs given in this paper, but for instance the following results makes use of it.

Lemma 3.1 *Let P be a partition as described above with $L = \mathcal{O}(\log n)$. Then there is a constant C_{sp} such that for any matrix $M \in \mathbb{R}^{n\times n}$ the following inequality holds between the global and the blockwise spectral norms:*

$$\|M\|_2 \le C_{sp} \sum_{\ell=0}^{L} \max_{b\in P_\ell} \|M|_b\|_2, \quad \textit{where } M|_b = (M_{ij})_{i\in\tau,\, j\in\sigma} \textit{ for } b = \tau \times \sigma. \tag{3.6}$$

Lemma 3.2 *The* exact *product*[8] *of two matrices $M_1 \in \mathcal{H}(P, k_1)$ and $M_2 \in \mathcal{H}(P, k_2)$ is in $\mathcal{H}(P, k)$ for all $k \ge k_3 = c\, L \max\{k_1, k_2\}$ with a constant c and L as in the previous lemma.*

The proofs can be found in [8] and the forthcoming paper [9].

Remark 3.3 *In the practical determination of an (minimal) admissible partition P, one uses suitable supersets $Y_\tau \supset X_\tau$ (e.g., Chebyshev spheres or bounding boxes) which are convex and satisfy* $\operatorname{dist}(Y_\tau, Y_\sigma) \ge \rho \max\{\operatorname{diam}(Y_\tau), \operatorname{diam}(Y_\sigma)\} > 0$ *if* $\min\{\#\tau, \#\sigma\} > 1$. *Note that this inequality implies (3.5b).*

[6] In practice, $\min\{\#\tau, \#\sigma\} = 1$ is replaced by $\min\{\#\tau, \#\sigma\} \le c_{\min}$ with an appropriate $c_{\min} > 1$.

[7] Either the factor ρ is fixed or we use the more precise notation "ρ-admissible partition".

[8] In the usual $\mathcal{H}$-matrix arithmetic, the exact product is replaced by a truncation of the true product (in $\mathcal{H}(P, k_3)$) to an approximation in $\mathcal{H}(P, k)$, where k is of the size of k_1 and k_2.

Having fixed the partition P and a number $k \in \mathbb{N}$, we define the set of hierarchical matrices ($\mathcal{H}$-matrices of blockwise rank at most k corresponding to the partition P) by

$$\mathcal{H}(P,k) := \{M \in \mathbb{R}^{n\times n} : \operatorname{rank}(M|_b) \le k \text{ for all } b = \tau\times\sigma \in P\}.$$

In [1], [8], [16], [17] it is shown that operations like matrix-times-vector, matrix-plus-matrix, matrix-times-matrix, matrix-inversion within $\mathcal{H}(P,k)$ cost $\mathcal{O}(nk\log^\alpha n)$ with $\alpha = 1$ (first two operations) or $\alpha = 2$ (last two operations). Also the storage amounts to $\mathcal{O}(nk\log n)$.

The next theorem shows that the Galerkin discretisation B of L^{-1} can be well approximated by $\mathcal{H}(P,k)$-matrices.

3.3 $\mathcal{H}(P,k)$-Approximation to B

Theorem 3.4 *Assume (3.5b) and let X_σ be convex for all $\tau \in T(I)$. Let P be chosen such that Lemma 3.1 can be applied. For any $\varepsilon \in (0,1)$, let $k_\varepsilon \in \mathbb{N}$ ($k_\varepsilon \sim \mathcal{O}(\log^{d+1}(\frac{1}{\varepsilon}))$) be the dimension bound from Theorem 2.8. Then for $k \ge k_\varepsilon$ there is $B_\mathcal{H} \in \mathcal{H}(P,k)$ such that the spectral norm of the difference is bounded by*

$$\|B - B_\mathcal{H}\|_2 \le \varepsilon \frac{c(\kappa_C,\rho,\operatorname{diam}(\Omega))}{\lambda_{\min}} L, \tag{3.7}$$

where $c(\kappa_C,\rho,\Omega)$ is a function depending on $\kappa_C = \lambda_{\max}/\lambda_{\min}$, ρ from (3.5b) and $\operatorname{diam}(\Omega)$. $L = \mathcal{O}(\log n)$ is the maximal level from Lemma 3.1.

Proof. (a) Let $b = \tau\times\sigma \in P$ with $\min\{\#\tau,\#\sigma\} > 1$. Hence, (3.5b) holds. Apply Theorem 2.8 with $D_1 = X_\tau$, $D_2 = X_\sigma$, and $\hat{X}_\sigma := \{x \in \Omega : \operatorname{dist}(x, X_\sigma) \le \frac{\rho}{2}\operatorname{diam}(X_\sigma)\}$. According to Remark 2.10 there is $\tilde{G}^b(x,y) = \sum_{i=1}^{k_\varepsilon} u_i^b(x)v_i^b(y)$ such that

$$\|G - \tilde{G}^b\|_{L^2(X_\tau\times X_\sigma)} \le \varepsilon \|G\|_{L^2(X_\tau\times\hat{X}_\sigma)}.$$

Let the functions u_i^b and v_i^b of $\tilde{G}^b$ be extended to Ω by zero. We define the integral operator

$$K_b\varphi = \int_\Omega \tilde{G}^b(\cdot,y)\varphi(y)\,dy \qquad \text{for } \operatorname{supp}\varphi \subset \overline{\Omega}$$

and set $B_\mathcal{H}|_b = (J^*K_bJ)|_b$ for all blocks b. The rank of $B_\mathcal{H}|_b$ is bounded by k_ε since each term $u_i^b(x)v_i^b(y)$ in $\tilde{G}^b$ produces one rank-1 matrix in $(J^*K_bJ)|_b$.

If $\min\{\#\tau,\#\sigma\} = 1$, we use the *exact* Green function, i.e., $\tilde{G}^b := G$. Since the block $B_\mathcal{H}|_b$ has rank 1 at most, $\operatorname{rank}(B_\mathcal{H}|_b) \le k$ holds again.

(b) Consider a block $b = \tau\times\sigma \in P$ with $\min\{\#\tau,\#\sigma\} > 1$. Choose any vectors $\mathbf{x} = (x_j)_{j\in\sigma}$, $\mathbf{y} = (y_i)_{i\in\tau}$ and set $u = J\mathbf{x} = \sum_{j\in\sigma} x_j\varphi_j$ and $v = J\mathbf{y}$. To see that $B_\mathcal{H}|_b$ approximates the block $B|_b$, remember the representation (2.6) of L^{-1} and use (3.4). The estimate

$$\begin{aligned}
|\langle (B|_b - B_\mathcal{H}|_b)\mathbf{y},\mathbf{x}\rangle_h| &= |\langle J^*(L^{-1}-K_b)J\beta,\alpha\rangle_h| = |((L^{-1}-K_b)v,u)_{L^2}| \\
&\le \|G-\tilde{G}^b\|_{L^2(X_\tau\times X_\sigma)}\,\|u\|_{L^2(X_\sigma)}\,\|v\|_{L^2(X_\tau)} \\
&\le \varepsilon\,\|G\|_{L^2(X_\tau\times\hat{X}_\sigma)}\,\|u\|_{L^2(\Omega)}\,\|v\|_{L^2(\Omega)} \\
&\le \varepsilon\,c_{J,2}^2\,\|G\|_{L^2(X_\tau\times\hat{X}_\sigma)}\,\|\mathbf{x}\|_h\,\|\mathbf{y}\|_h
\end{aligned}$$

proves $\|B|_b - B_\mathcal{H}|_b\|_2 \le \varepsilon\,c_{J,2}^2\,\|G\|_{L^2(X_\tau\times\hat{X}_\sigma)}$ for the spectral norm.

Although $G(\cdot,y) \in W^{1,1}(\Omega)$ for all $y \in \Omega$, $G(\cdot,\cdot)$ does not belong to $L^2(\Omega\times\Omega)$ as soon as $d \ge 4$. ¿From (2.5) it can be seen that $\|G\|_{L^2(X_\tau\times\hat{X}_\sigma)}$ may increase when the sets X_τ, $\hat{X}_\sigma$ are approaching each other. The construction of $\hat{X}_\tau$ ensures

$$\delta := \operatorname{dist}(X_\tau,\hat{X}_\sigma) = \frac{1}{2}\operatorname{dist}(X_\tau,X_\sigma) \ge \frac{\rho}{2}\operatorname{diam}(X_\tau)$$

as well as $\delta \ge \frac{\rho}{2}\operatorname{diam}(X_\sigma)$. Hence (2.5) implies

$$\|G\|_{L^2(X_\tau\times\hat{X}_\sigma)} \le \frac{c(d,\kappa_C)}{\lambda_{\min}}\delta^{2-d}\sqrt{\operatorname{vol}(X_\tau)\operatorname{vol}(\hat{X}_\sigma)}.$$

Using $\mathrm{vol}(\hat{X}_\sigma) \le \omega_d(\frac{1}{2}\operatorname{diam}(\hat{X}_\sigma))^d \le \omega_d(1+1/\rho)^d\delta^d$ and $\mathrm{vol}(X_\tau) \le \omega_d(\delta/\rho)^d$ with $\omega_d = \mathrm{vol}(B_1(0))$, we see that

$$\|G\|_{L^2(X_\tau\times\hat{X}_\sigma)} \le C_\rho \frac{c(d,\kappa_C)}{\lambda_{\min}}\delta^2 \quad \text{with } C_\rho := \omega_d \frac{(\rho+1)^{d/2}}{\rho^d}.$$

The rough estimate $\delta \le \operatorname{diam}(\Omega) = \mathcal{O}(1)$ together with Lemma 3.1 yields (3.7). ∎

Corollary 3.5 *Assume that each (possibly non-convex) set X_τ has a convex superset Y_τ satisfying the admissibility condition (cf. Remark 3.3). Then Theorem 3.4 remains true for X_τ, X_σ.*

Proof. Apply Theorem 3.4 to Y_τ and Y_σ. ∎

Remark 3.6 *(a) The factor $L = \mathcal{O}(\log n)$ in (3.7) can be avoided. Under the following two reasonable assumptions that (i) $\tau \in T(I)$ is always subdivided into its two sons $\tau', \tau'' \in T(I)$ so that the diameters are comparable, i.e. for the diameter of a cluster $X_\tau \in T_\ell(I)$ it holds that $\operatorname{diam}(X_\tau) \le cq^{-\ell}$ ($q < 1$) and (ii) the partition P is generated so that the admissibility condition is almost sharp for each admissible block $b = \tau\times\sigma \in P$, i.e., in addition to (3.5b) there is a constant $\tilde{c} > 1$, which is independent of b, so that $\operatorname{dist}(X_\tau, X_\sigma) \le \tilde{c}\rho\min\{\operatorname{diam}(X_\tau), \operatorname{diam}(X_\sigma)\}$. In this case the factor δ^2 decreases with respect to the level ℓ as $\delta^2 \le Cq^{-2\ell}$. Thus, the sum (3.6) is bounded independently of n.*

(b) Replacing ε by $\varepsilon\lambda_{\min}/(c(\kappa_C,\rho,\operatorname{diam}(\Omega))L)$, Theorem 3.4 yields $\|B - B_{\mathcal{H}}\|_2 \le \varepsilon$ with $k_\varepsilon = \mathcal{O}\left(\log^{d+1}(\frac{L}{\varepsilon})\right)$ and thanks to Part (a) even $\mathcal{O}\left(\log^{d+1}(\frac{1}{\varepsilon})\right)$.

4 Approximation of the Inverse Mass Matrix by an $\mathcal{H}$-Matrix

The inverse of the mass matrix M will arise when the inverse of the stiffness matrix is approximated by an $\mathcal{H}$-matrix. Therefore $\mathcal{H}$-matrix properties of M^{-1} are to be investigated.

For the inverse of banded matrices an exponential decay of the entries has been observed (cf. [3]). Here, $\sigma(M)$ denotes the spectrum of M.

Lemma 4.1 *Let $M = (M_{ij})_{i,j\in I}$ be a symmetric positive definite matrix with $\sigma(M) \subset [a,b]$ and denote its matrix graph by G_M (cf. [14, Subsection 6.2]). Let $i, j \in I$ and δ_{ij} be the minimal length of a path[9] in G_M from i to j. Then*

$$|(M^{-1})_{ij}| \le \hat{c}\, q^{\delta_{ij}} \quad \text{with } \hat{c} = \frac{(1+\sqrt{r})^2}{2ar},\ q = \frac{\sqrt{r}-1}{\sqrt{r}+1},\ r = \frac{b}{a}. \tag{4.1}$$

Proof. For any polynomial $p \in \Pi_k$ with $k < \delta_{ij}$ we observe $p(M)_{ij} = 0$. Furthermore, the spectral norm and the spectral radius coincide for normal matrices:

$$\|M^{-1} - p_k(M)\|_2 = \rho(M^{-1} - p_k(M)) = \max_{x\in\sigma(M)} |x^{-1} - p_k(x)|.$$

A result due to Chebyshev says that Π_k contains a polynomial p_k (cf. [19, p. 33]) so that

$$\|x^{-1} - p_k(x)\|_{\infty,[a,b]} \le \hat{c}\, q^{k+1}$$

with q, $\hat{c}$ as in (4.1). Set $k = \delta_{ij} - 1$. The previous arguments show the final result:

$$|(M^{-1})_{ij}| = |(M^{-1})_{ij} - p_k(M)_{ij}| \le \|M^{-1} - p_k(M)\|_2 \le \hat{c}\, q^{k+1} = \hat{c}\, q^{\delta_{ij}}.$$

∎

The mass matrix is by definition symmetric positive definite with $a = \|M^{-1}\|_2^{-1}$ and $b = \|M\|_2$. $(i,j) \in G_M$ implies that $X_i \cap X_j$ contains an interior point. Hence, if k is the smallest integer so that $\operatorname{dist}(X_i, X_j) \le (k-1)h$, the length of a path in G_M from i to j must be at least k, i.e., $\delta_{ij} \ge 1 + d_{ij}/h$, where $d_{ij} := \operatorname{dist}(X_i, X_j)$.

Lemma 4.2 *Let $0 < c_{J,1} \le c_{J,2}$ be the constants from (3.4). Then*

$$|(M^{-1})_{ij}| \le C\,\|M^{-1}\|_2\, q^{d_{ij}/h} \quad \text{for all } i,j \in I,$$

where $C = \frac{r-1}{2r}$ and $q = \frac{\sqrt{r}-1}{\sqrt{r}+1} \in (0,1)$ with $r = (c_{J,2}/c_{J,1})^2$ are independent of the matrix size n.

[9] If no path from i to j exists (case of a reducible matrix M), we formally set $\delta_{ij} = \infty$, because $(M^{-1})_{ij} = 0$.

Proof. Since $M = J^*J$, the spectrum of M is contained in $[a,b]$ with $a = c_{J,1}^2$ and $b = c_{J,2}^2$. Hence, the condition number of M is bounded independently of the matrix size n by $r = (c_{J,2}/c_{J,1})^2$. Applying the previous lemma and using $\delta_{ij} \geq 1 + d_{ij}/h$, we end up with the assertion. ∎

Theorem 4.3 *Assume (3.2), (3.4), (3.5b), and choose P such that Lemma 3.1 holds. For any $\varepsilon > 0$, there is $N_{\mathcal{H}} \in \mathcal{H}(P, k_\varepsilon)$ satisfying $\|M^{-1} - N_{\mathcal{H}}\|_2 \leq \varepsilon \|M^{-1}\|_2$ with $k_\varepsilon = \mathcal{O}(\log^d(\frac{L}{\varepsilon}))$ ($L = \mathcal{O}(\log n)$ from Lemma 3.1).*

Proof. (a) We use the following explicit definition of $N_{\mathcal{H}} = N_{\mathcal{H}}(k)$ depending on $k \in \mathbb{N}$. Set $N_{\mathcal{H}}|_b := M^{-1}|_b$ for $b = \tau \times \sigma \in P$ if $\#\tau\#\sigma \leq k^2$; otherwise $N_{\mathcal{H}}|_b := 0$. Since $\operatorname{rank}(N_{\mathcal{H}}|_b) \leq \min\{\#\tau, \#\sigma\} \leq k$ for all $b \in P$, $N_{\mathcal{H}}$ belongs to $\mathcal{H}(P,k)$. Let $E = M^{-1} - N_{\mathcal{H}}(k)$ be the error matrix. Due to Lemma 3.1, it remains to determine the spectral norms of $E|_b = M^{-1}|_b$ in the case of $\#\tau\#\sigma > k^2$.

(b) For $\#\tau\#\sigma > k^2$ and $i \in \tau$, $j \in \sigma$, we want to estimate $E_{ij} = (M^{-1})_{ij}$. Condition (3.5b) implies $d_{ij} = \operatorname{dist}(X_i, X_j) \geq \operatorname{dist}(X_\tau, X_\sigma) \geq \rho \max\{\operatorname{diam}(X_\tau), \operatorname{diam}(X_\sigma)\}$. We notice that $(\operatorname{diam}(X_\tau))^d \geq \operatorname{vol}(X_\tau) 2^d/\omega_d$, and from (3.3) and (3.2) we obtain that

$$\operatorname{vol}(X_\tau) \geq c_M^{-1} \sum_{i \in \tau} \operatorname{vol}(X_i) \geq \frac{c_v}{c_M} h^d \#\tau.$$

Altogether, $d_{ij} \geq C' h \sqrt[d]{\#\tau}$ follows with C' expressed by ω_d, ρ, c_M, and c_v. Similarly, $d_{ij} \geq C' h \sqrt[d]{\#\sigma}$ holds. The combination yields $d_{ij}/h \geq C' \sqrt[2d]{\#\tau\#\sigma}$. This proves

$$|E_{ij}| \leq C \|M^{-1}\|_2 \, q^{C' \sqrt[2d]{\#\tau\#\sigma}}.$$

(c) A trivial estimate of the spectral norm yields

$$\|E|_b\|_2 \leq \sqrt{\#\tau\#\sigma} \max_{i \in \tau,\, j \in \sigma} |E_{ij}| \leq C \sqrt{\#\tau\#\sigma} \, \|M^{-1}\|_2 \, q^{C' \sqrt[2d]{\#\tau\#\sigma}}.$$

We simplify the right-hand side: For a suitable $C'' > C'$, the estimate $C\ell \, q^{C' \sqrt[d]{\ell}} \leq q^{C'' \sqrt[d]{\ell}}$ holds for all $\ell \geq k_{\min}$ so that

$$\|E|_b\|_2 \leq \|M^{-1}\|_2 \, q^{C'' \sqrt[2d]{\#\tau\#\sigma}} < \|M^{-1}\|_2 \, q^{C'' \sqrt[d]{k}}.$$

Lemma 3.1 implies $\|E\|_2 \leq LC_* \|M^{-1}\|_2 \, q^{C'' \sqrt[d]{k}}$ with $C_* = CC_{sp}$. Choose $k = k_\varepsilon \geq k_{\min}$ such that $LC_* \, q^{C'' \sqrt[d]{k}} \leq \varepsilon$, i.e., $k_\varepsilon = \max\{k_{\min}, \mathcal{O}(\log^d(\frac{LC_*}{\varepsilon})\} = \mathcal{O}(\log^d(\frac{L}{\varepsilon}))$. ∎

We summarise that the simple construction used in the proof yields an $\mathcal{H}(P,k)$-approximation with $k = \mathcal{O}(\log^d(\frac{L}{\varepsilon}))$ which is asymptotically smaller than the rank k from the B-approximation in Theorem 3.4.

Remark 4.4 *The factor $L = \mathcal{O}(\log n)$ in $k_\varepsilon = \mathcal{O}(\log^d(\frac{L}{\varepsilon}))$ can be avoided by arguments as in Remark 3.6a.*

5 Approximation of the Inverse FE-Stiffness Matrix by an $\mathcal{H}$-Matrix

5.1 Projections and Smoothness Assumptions

The following two projectors will be necessary in the FE error analysis. The $L^2(\Omega)$-orthogonal projection is expressed by $Q_h := JM^{-1}J^* : L^2(\Omega) \to V_h$, i.e., $(Q_h u, v_h)_{L^2} = (u, v_h)_{L^2}$ for all $u \in V$ and $v_h \in V_h$. The related error is described by

$$e_h^Q(u) := \|u - Q_h u\|_{L^2(\Omega)}. \tag{5.1}$$

On the other hand, the finite element approximation is connected with the Ritz projection $P_h = JA^{-1}J^*L : V \to V_h$. If $u \in V$ is the solution of the variational problem $a(u,v) = f(v)$ (cf. (2.4)), $u_h = P_h u$ is its finite element solution. The FE error is

$$e_h^P(u) := \|u - P_h u\|_{L^2(\Omega)}.$$

Since the $L^2(\Omega)$-orthogonal projection is the optimal one, i.e., $e_h^Q(u) \leq e_h^P(u)$, we only need estimates of e_h^P.

The weakest form of the finite element convergence is described by

$$e_h^P(u) \leq \varepsilon_h \|f\|_{L^2(\Omega)} \quad \text{for all } u = L^{-1}f,\ f \in L^2(\Omega), \tag{5.2}$$

where $\varepsilon_h \to 0$ as $h \to 0$.

For the sake of completeness, we give a proof of the last statement: Since $I-P_h$ is an orthogonal projection with respect to the inner product $(u,v)_E := a(u,v)$, $\|u-P_hu\|_E \le \|u\|_E$ holds. The inequalities $\|\cdot\|_{L^2(\Omega)} \le C'\|\cdot\|_E \le C''\|\cdot\|_{H^1(\Omega)}$ prove **(a)** $e_h^P(u) \le C\|u\|_{H^1(\Omega)}$. Furthermore, for any fixed $u \in H^1(\Omega)$, there holds **(b)** $\lim_{h\to 0} e_h^P(u) = 0$ (cf. [13, Theorem 8.2.2]). Let $H^* := \{u \in H^1(\Omega) : u = L^{-1}f \text{ for some } f \in L^2(\Omega) \text{ with } \|f\|_{L^2(\Omega)} \le 1\}$. Since the embedding $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ is compact (cf. [13, Theorems 6.4.8 and 6.4.10]) and $L^{-1} : H^{-1}(\Omega) \to H^1(\Omega)$ is bounded, the closure of the set $H^* \subset H^1(\Omega)$ is compact. We claim **(c)** $e_h^P(u) \le \varepsilon_h$ for all $u \in H^*$ with $\varepsilon_h \to 0$ as $h \to 0$. For an indirect proof, we assume that there is a sequence $\{u_k\}_{k\in\mathbb{N}} \subset H^*$ such that $e_h^P(u_k) \ge \eta > 0$ for all $k \in \mathbb{N}$. By compactness of $\overline{H^*}$, there is a subsequence $u_{k_j} \to u^* \in H^1(\Omega)$. Note that $e_h^P(u_{k_j}) \le e_h^P(u_{k_j}-u^*) + e_h^P(u^*)$. Due to result (a), $e_h^P(u_{k_j}-u^*) \le C\|u_{k_j}-u^*\|_{H^1(\Omega)} \to 0$, while result (b) yields $e_h^P(u^*) \to 0$. Together, $e_h^P(u_{k_j}) \to 0$ contradicts the assumption (c). Hence, $\varepsilon_h := \sup_{u\in H^*} e_h^P(u) \to 0$ is proved. ∎

The standard error estimate assumes $\varepsilon_h = c_E h^\beta$:

$$e_h^P(u) \le c_E h^\beta \|f\|_{L^2(\Omega)} \quad \text{for all } u = L^{-1}f,\ f \in L^2(\Omega) \quad \text{with some } \beta > 0. \tag{5.3}$$

Usually, such an estimate is proved in two steps. By regularity assumptions, $u \in H^\alpha(\Omega)$ for some $\alpha \in (1,2]$ is established for $u = L^{-1}f$, $f \in L^2(\Omega)$, so that $\|u\|_{H^\alpha(\Omega)} \le C\|f\|_{L^2(\Omega)}$. Then by approximation properties of V_h, $\|u-P_hu\|_{H^1(\Omega)} \le C'h^{\alpha-1}\|u\|_{H^\alpha(\Omega)}$ is derived and can be generalised to $\|u-P_hu\|_{H^{2-\alpha}(\Omega)} \le C''h^{2(\alpha-1)}\|u\|_{H^\alpha(\Omega)}$. Using $\|\cdot\|_{L^2(\Omega)} \le C'''\|\cdot\|_{H^{2-\alpha}(\Omega)}$, we arrive at (5.3).

Remark 5.1 *(a) Due to our quite weak assumption $c_{ij} \in L^\infty(\Omega)$ upon the smoothness of the coefficients, one cannot ensure (5.3) for any $\beta > 0$ without further assumptions, while at best $\beta = 2$ holds.*

(b) The approximation error ε which we should choose for $\|A^{-1} - C_{\mathcal{H}}\|_2$ when we approximate A^{-1} by an $\mathcal{H}$-matrix $C_{\mathcal{H}}$, is to be adapted to the finite element error, i.e., $u - u_h = u - P_hu$ and $u_h - \tilde{u}_h$ ($\tilde{u}_h = C_{\mathcal{H}}f_h$, $f_h = J^*f$) should be of similar size.

(c) Accordingly, we take (5.2) as an assumption without specifying the behaviour with respect to $h \to 0$ and use ε_h as desirable error for $\|A^{-1} - C_{\mathcal{H}}\|_2$.

5.2 Approximation to A^{-1}

First we show that $M^{-1}BM^{-1}$ is an approximation to the inverse A^{-1} of the finite element stiffness matrix.

Lemma 5.2 *Let $c_{J,2}$ and ε_h be the quantities in (3.4) and (5.2). Then $\|MA^{-1}M - B\|_2 \le 2\,c_{J,2}^2\,\varepsilon_h$.*

Proof. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $f_h = J\mathbf{x}$, $v_h = J\mathbf{y} \in V_h$. Then, using $B = J^*L^{-1}J$ and the projections from above, we have

$$\begin{aligned}
\langle (MA^{-1}M - B)\mathbf{x}, \mathbf{y}\rangle_h &= ((MA^{-1}M - J^*L^{-1}J)M^{-1}J^*f_h, M^{-1}J^*v_h)_{L^2(\Omega)} \\
&= \big((JA^{-1}J^* - JM^{-1}J^*L^{-1}JM^{-1}J^*)f_h, v_h\big)_{L^2(\Omega)} \\
&= (P_hL^{-1}f_h - Q_hL^{-1}Q_hf_h, v_h)_{L^2(\Omega)} = (P_hL^{-1}f_h - Q_hL^{-1}f_h, v_h)_{L^2(\Omega)} \\
&= ([L^{-1}f_h - P_hL^{-1}f_h] - [L^{-1}f_h - Q_hL^{-1}f_h], v_h)_{L^2(\Omega)} \\
&\le \Big(e_h^P(L^{-1}f_h) + e_h^Q(L^{-1}f_h)\Big)\,\|v_h\|_{L^2(\Omega)} \le 2\,e_h^P(L^{-1}f_h)\|v_h\|_{L^2(\Omega)} \\
&\le 2\,\varepsilon_h\|f_h\|_{L^2}\|v_h\|_{L^2(\Omega)} \le 2\,c_{J,2}^2\,\varepsilon_h\|\mathbf{x}\|_h\|\mathbf{y}\|_h,
\end{aligned}$$

which proves $\|MA^{-1}M - B\|_2 \le 2\,c_{J,2}^2\,\varepsilon_h$. ∎

Corollary 5.3 $\|A^{-1} - M^{-1}BM^{-1}\|_2 \le 2\,c_{J,1}^{-4}\,c_{J,2}^2\,\varepsilon_h$.

Proof. Use $A^{-1} - M^{-1}BM^{-1} = M^{-1}(MA^{-1}M - B)M^{-1}$ and $\|M^{-1}\|_2 \le c_{J,1}^{-2}$. ∎

5.3 $\mathcal{H}$-Matrix Approximation to $M^{-1}BM^{-1}$

Above we have shown that B and M^{-1} can be approximated by the $\mathcal{H}$-matrices $B_{\mathcal{H}}$ and $N_{\mathcal{H}}$, respectively. Therefore, the natural approach is to use

$$C_{\mathcal{H}} := N_{\mathcal{H}}B_{\mathcal{H}}N_{\mathcal{H}}$$

as approximant of A^{-1}. Due to Lemma 3.2, the exact product $C_{\mathcal{H}} = N_{\mathcal{H}} B_{\mathcal{H}} N_{\mathcal{H}}$ belongs to $\mathcal{H}(P,k)$ provided that $k \geq k_C := c\,L \max\{c\,L \max\{k_B, k_N\}, k_N\}$ and $N_{\mathcal{H}} \in \mathcal{H}(P, k_N)$, $B_{\mathcal{H}} \in \mathcal{H}(P, k_B)$. Since $L \geq 1$ and, without loss of generality, $c \geq 1$, the latter expression becomes $k_C = c^2 L^2 \max\{k_B, k_N\}$. Assuming $k_B \geq k_N$, we arrive at

$$k_C = c^2 L^2 k_B.$$

The estimation of the spectral norm of

$$M^{-1} B M^{-1} - N_{\mathcal{H}} B_{\mathcal{H}} N_{\mathcal{H}} = (M^{-1} - N_{\mathcal{H}}) B M^{-1} + N_{\mathcal{H}} (B - B_{\mathcal{H}}) M^{-1} + N_{\mathcal{H}} B_{\mathcal{H}} (M^{-1} - N_{\mathcal{H}})$$

by

$$\|M^{-1} - N_{\mathcal{H}}\|_2 (\|B\|_2 \|M^{-1}\|_2 + \|N_{\mathcal{H}}\|_2 \|B_{\mathcal{H}}\|_2) + \|N_{\mathcal{H}}\|_2 \|M^{-1}\|_2 \|B - B_{\mathcal{H}}\|_2$$

is obvious. Let $\varepsilon_N := \|M^{-1} - N_{\mathcal{H}}\|_2$, $\varepsilon_B := \|B - B_{\mathcal{H}}\|_2$. Since $\varepsilon_N \leq \|M^{-1}\|_2$, $\varepsilon_B \leq \|B\|_2$ and $\|B\|_2, \|M^{-1}\|_2 = \mathcal{O}(1)$, we obtain

$$\|M^{-1} B M^{-1} - N_{\mathcal{H}} B_{\mathcal{H}} N_{\mathcal{H}}\|_2 \leq C_{II}(\varepsilon_N + \varepsilon_B). \tag{5.4}$$

5.4 Final Result

The combination of Corollary 5.3 and (5.4) yields

$$\|A^{-1} - N_{\mathcal{H}} B_{\mathcal{H}} N_{\mathcal{H}}\|_2 \leq C_I\, \varepsilon_h + C_{II}(\varepsilon_N + \varepsilon_B),$$

where $C_I = 2\, c_{J,1}^{-4}\, c_{J,2}^2$. For simplicity we set $k_B = k_N =: k$ and choose

$$k = \max\{\mathcal{O}(\log^{d+1}(\frac{LC_1}{\delta})), \mathcal{O}(\log^d(\frac{L\|M^{-1}\|_2}{\delta}))\}$$

with $C_1 = \frac{c(\kappa_C, \rho, \mathrm{diam}(\Omega))}{\lambda_{\min}}$ and $\delta = C_I\, \theta \varepsilon_h / (2C_{II})$, where the constants in the $\mathcal{O}(\cdot)$ expressions are detailed in Theorem 3.4 and Theorem 4.3, while $\theta \in (0,1)$. Then,

$$\|A^{-1} - N_{\mathcal{H}} B_{\mathcal{H}} N_{\mathcal{H}}\|_2 \leq C_I\, (1+\theta) \varepsilon_h \tag{5.5}$$

shows that the already existing finite element error $C_I\, \varepsilon_h$ is only slightly increased. The corresponding ranks $k_B = k_N$ behave asymptotically like

$$k_B = k_N = \mathcal{O}(\log^{d+1}(\frac{L}{\varepsilon_h})).$$

The resulting rank for $C_{\mathcal{H}} = N_{\mathcal{H}} B_{\mathcal{H}} N_{\mathcal{H}}$ is bounded by $k_C = c^2 L^2 k_B$. Thus, $C_{\mathcal{H}}$ approximates A^{-1} as described in (5.5) and belongs to $\mathcal{H}(P,k)$ for all $k \geq c^2 L^2 k_B$. This result is summarised in Part (a) of

Theorem 5.4 *(a) Let $\varepsilon_h > 0$ be the finite element error from (5.2). $L = \mathcal{O}(\log n)$ is the depth of the cluster tree (see Lemma 3.1). Then there are constants C' and C'' defining $k_C := C' L^2 \log^{d+1}(\frac{LC''}{\varepsilon_h})$ and there is an $\mathcal{H}$-matrix $C_{\mathcal{H}} \in \mathcal{H}(P, k_C)$ such that*

$$\|A^{-1} - C_{\mathcal{H}}\|_2 \leq C_I\, (1+\theta) \varepsilon_h. \tag{5.6}$$

(b) If $\varepsilon_h = O(h^\beta)$ according to (5.3), $k_C = \mathcal{O}(\log^{d+3}(n))$ holds.

Proof. As $h^{-1} = O(n^{1/d})$, the asymptotic behaviour of $\log(\frac{LC''}{\varepsilon_h}) = \log(L) + const + \log(n^{\beta/d})$ is $\mathcal{O}(\log n)$. This proves Part (b). ∎

Since $\lambda_{\min}$ in (2.2) is of size $\mathcal{O}(1)$ (without loss of generality, we may scale the problem so that $\lambda_{\min} = 1$), also $\|A^{-1}\|_2 = \mathcal{O}(1)$ holds. Hence, the absolute error (5.6) may be changed into a relative one: $\|A^{-1} - C_{\mathcal{H}}\|_2 \leq C_I^*\, \|A^{-1}\|_2 (1+\theta)\varepsilon_h$ with another constant C_I^*.

6 Computational Experiments

In the following section numerical experiments will demonstrate that the preceding results are true. At this moment we are not interested in fast numerical schemes to approximate the blocks. Instead, this section will show only the existence of low-rank approximants. Therefore CPU times are omitted.

We compare the Laplacian with operators of type (2.1). For simplicity, the following tests are performed for operators in two variables in the unit square $\Omega = [0,1]^2$. The Figures 1 and 2 show the coefficient $C(x) \in \mathbb{R}^{2\times 2}$ for $x \in \Omega$, or more precisely the spectral norm of $C(x)$ for $x \in \Omega$.

For the first example Ω is decomposed into Ω_1 and $\Omega_2 = \Omega \setminus \Omega_1$, where Ω_1 is the wall-like domain in Figure 1. Let L_a be the operator in (2.1) with the coefficient $C_a(x) = c(x)I$, where

$$c(x) = \begin{cases} a, & x \in \Omega_1 \\ 1, & x \in \Omega_2 \end{cases} .$$

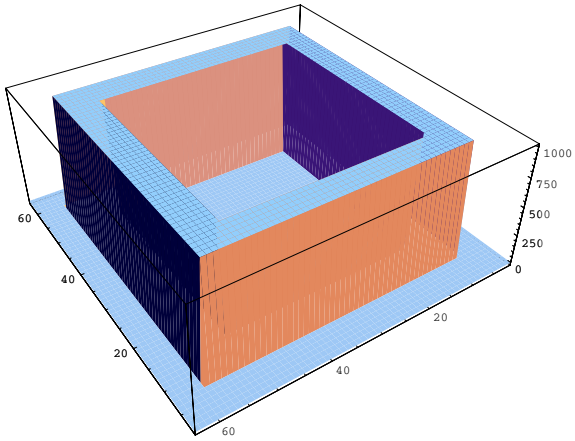


Figure 1: Coefficients of the first example

The following table shows the relative accuracy measured for different problem sizes n in Frobenius norm when approximating the inverse of the respective FEM matrix by an $\mathcal{H}$-matrix. For each admissible block the best rank-k approximant is calculated using the singular value decomposition. In the first line of each problem size the amount of storage needed for the respective $\mathcal{H}$-matrix approximant is given.

$n = 2304$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
Storage (MB)	10.2	18.9	27.6	36.2		
Δ	$4.1e{-}03$	$5.9e{-}04$	$1.1e{-}05$	$1.2e{-}06$		
L_{10^3}	$6.9e{-}03$	$9.8e{-}04$	$1.6e{-}05$	$2.1e{-}06$		
L_{10^6}	$6.9e{-}03$	$9.8e{-}04$	$1.6e{-}05$	$1.7e{-}06$		
$n = 6400$						
Storage (MB)	40.0	75.9	111.6	147.5	183.1	218.8
Δ	$3.5e{-}03$	$6.5e{-}04$	$8.8e{-}06$	$2.1e{-}06$	$4.2e{-}07$	$8.3e{-}09$
L_{10^3}	$5.5e{-}03$	$1.0e{-}03$	$1.2e{-}05$	$3.2e{-}06$	$5.5e{-}08$	$1.3e{-}08$
L_{10^6}	$5.6e{-}03$	$1.0e{-}03$	$1.2e{-}05$	$3.1e{-}07$	$4.7e{-}08$	$9.1e{-}09$
$n = 14400$						
Storage (MB)	123.4	235.7	349.6	462.0	575.9	688.2
Δ	$3.2e{-}03$	$5.9e{-}04$	$8.9e{-}06$	$2.3e{-}06$	$5.5e{-}08$	$1.5e{-}08$
L_{10^3}	$4.9e{-}03$	$8.8e{-}04$	$1.2e{-}05$	$3.3e{-}06$	$7.3e{-}08$	$1.9e{-}08$
L_{10^6}	$5.0e{-}03$	$8.8e{-}04$	$1.0e{-}05$	$3.2e{-}06$	$6.7e{-}08$	$9.1e{-}09$

The values for L_{10^3} and L_{10^6} differ only insignificantly from those for the Laplacian. Notice that L_{10^3} and L_{10^6} behave almost the same though the ratios $\kappa_C = \lambda_{\max}/\lambda_{\min}$ (see Section 2.1) differ by a factor of 1000.

While in the first example the aim was to demonstrate that jumps of arbitrary size do not affect the quality of the $\mathcal{H}$-matrix approximation, the second example is designed to show that these jumps may happen on more than few interior boundaries. Again, we decompose Ω into two domains Ω_1 and $\Omega_2 = \Omega \setminus \Omega_1$, where Ω_2 is the lower region in Figure 2.

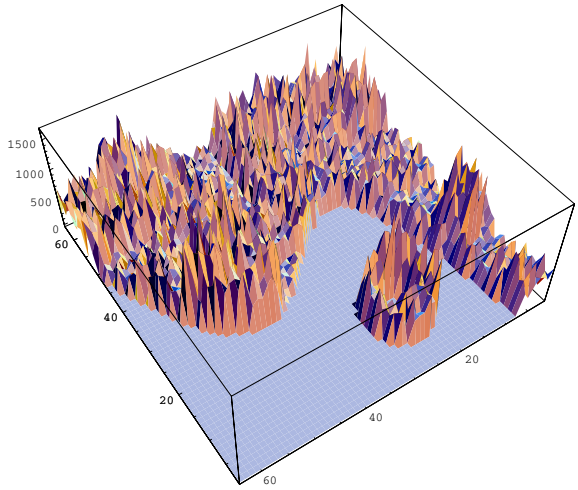


Figure 2: Coefficients of the second example

By L_a we denote an operator for which $C_a(x)$ is the identity in Ω_2 and a quadruple of random numbers from the interval $[0, a]$ in the remaining part Ω_1, so that $C_a(x)$ is always positive definite. The coefficient matrix is chosen to have a two level random structure: the first scale is $\sqrt{h}$ and the second h. The corresponding numerical results are assembled in the following table:

$n = 2304$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
Δ	$4.1e{-}03$	$5.9e{-}04$	$1.1e{-}05$	$1.2e{-}06$		
L_{10^3}	$4.4e{-}03$	$3.7e{-}04$	$1.1e{-}05$	$4.4e{-}07$		
L_{10^6}	$4.4e{-}03$	$3.4e{-}04$	$1.0e{-}05$	$2.2e{-}07$		
$n = 6400$						
Δ	$3.5e{-}03$	$6.5e{-}04$	$8.8e{-}06$	$2.1e{-}06$	$4.2e{-}07$	$8.3e{-}09$
L_{10^3}	$4.2e{-}03$	$5.3e{-}04$	$7.5e{-}06$	$1.1e{-}05$	$2.3e{-}08$	$1.7e{-}09$
L_{10^6}	$4.2e{-}03$	$5.2e{-}04$	$6.8e{-}06$	$8.9e{-}07$	$1.7e{-}08$	$5.4e{-}10$
$n = 14400$						
Δ	$3.2e{-}03$	$5.9e{-}04$	$8.9e{-}06$	$2.3e{-}06$	$5.5e{-}08$	$1.5e{-}08$
L_{10^3}	$4.0e{-}03$	$6.0e{-}04$	$1.1e{-}05$	$2.1e{-}06$	$6.0e{-}08$	$1.2e{-}08$
L_{10^6}	$4.0e{-}03$	$5.8e{-}04$	$1.0e{-}05$	$1.9e{-}06$	$5.3e{-}08$	$9.1e{-}09$

Although in Theorem 5.4 we could only proof a relative error of order ε_h the numerical results show that any prescribed accuracy can be reached by increasing the rank k of the approximation. Moreover, the accuracy does not seem to depend on the upper bound κ_C of the condition numbers of $C_a(x)$.

7 Additional Comments

There are various steps in the proofs where the constants can be improved.

In (3.5b) we have formulated the admissibility condition by the maximal diameters: $\operatorname{dist}(X_\tau, X_\sigma) \geq \rho \max\{\operatorname{diam}(X_\tau), \operatorname{diam}(X_\sigma)\}$. An interesting weaker form is obtained by changing max into min:

$$\operatorname{dist}(X_\tau, X_\sigma) \geq \rho \min\{\operatorname{diam}(X_\tau), \operatorname{diam}(X_\sigma)\}.$$

As long as the clusters of the same level are balanced in size, i.e., $\operatorname{diam}(X_\tau) \approx \operatorname{diam}(X_\sigma)$, both conditions are very similar. However, there may arise cases, where $\operatorname{diam}(X_\tau)$ and $\operatorname{diam}(X_\sigma)$ differ strongly (cf. [10]). The proof of the central Theorem 2.8 relies on $\operatorname{dist}(D_1, D_2) \geq \rho \operatorname{diam}(D_2)$. Therefore, the choice $D_2 = X_\tau$ in the proof of Theorem 3.4 is correct as long as $\operatorname{diam}(X_\tau) \leq \operatorname{diam}(X_\sigma)$. If, however, $\operatorname{diam}(X_\tau) > \operatorname{diam}(X_\sigma)$, we have to assume that X_σ is convex. Again Theorem 2.8 can be applied, now with $D_2 = X_\sigma$. The estimates in Theorem 3.4 must be slightly modified, since $\operatorname{vol}(X_\sigma)$ cannot be estimated by means of $\operatorname{diam}(X_\tau)$.

Acknowledgement. The authors wish to thank S. Müller (MPI, Leipzig) for his contribution to the proof of Theorem 2.8.

References

[1] M. Bebendorf: *Effiziente numerische Lösung von Randintegralgleichungen unter Verwendung von Niedrigrang-Matrizen.* dissertation.de, Verlag im Internet, 2001. ISBN 3-89825-183-7.

[2] S. Börm, L. Grasedyck, and W. Hackbusch: *Introduction to hierarchical matrices with applications.* 2002. To appear.

[3] S. Demko, W. F. Moss, and P. W. Smith: *Decay rates for inverses of band matrices.* Math. Comp. **43**, 491–499, 1984.

[4] G. Dolzmann and S. Müller: *Estimates for Green's matrices of elliptic systems by L^p theory.* Manuscripta Math., **88**, 261–273, 1995.

[5] I.P. Gavrilyuk, W. Hackbusch, and B. Khoromskij: *$\mathcal{H}$-matrix approximation for the operator exponential with applications.* Numer. Math., 2002. To appear.

[6] M. Giaquinta: *Multiple integrals in the calculus of variations and nonlinear elliptic systems.* Princeton University Press, Princeton, NJ, 1983.

[7] D. Gilbarg and N. S. Trudinger: *Elliptic partial differential equations of second order.* Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[8] L. Grasedyck: *Theorie und Anwendungen Hierarchischer Matrizen.* Dissertation, Universität Kiel, 2001

[9] L. Grasedyck and W. Hackbusch: *Construction and arithmetics of $\mathcal{H}$-matrices.* In preparation

[10] L. Grasedyck, W. Hackbusch, and Sabine Le Borne: *Adaptive refinement and clustering of $\mathcal{H}$-matrices.* Technical Report 106, Max-Planck-Institut für Mathematik, Leipzig, 2001.

[11] M. Grüter and K.-O. Widman: *The Green function for uniformly elliptic equations.* Manuscripta Math. **37**, 303–342, 1982.

[12] W. Hackbusch: *Multi-grid methods and applications.* Springer-Verlag, Berlin 1985

[13] W. Hackbusch: *Theorie und Numerik elliptischer Differentialgleichungen.* B. G. Teubner, Stuttgart, 1996 - English translation: *Elliptic differential equations. Theory and numerical treatment.* Springer-Verlag, Berlin, 1992.

[14] W. Hackbusch. *Iterative Lösung großer schwachbesetzter Gleichungssysteme.* Teubner, Stuttgart, 2nd edition, 1993 - English translation: *Iterative solution of large sparse systems.* Springer-Verlag, New York, 1994.

[15] W. Hackbusch: *Integralgleichungen. Theorie und Numerik.* Teubner, Stuttgart, 2nd edition, 1997 - English translation: *Integral equations. Theory and numerical treatment.* Volume 128 of *ISNM.* Birkhäuser, Basel, 1995.

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

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

[18] W. Hackbusch and Z. P. Nowak: *On the fast matrix multiplication in the boundary element method by panel clustering.* Numer. Math. **54**, 463–491, 1989.

[19] G. Meinardus: *Approximation of functions: Theory and numerical methods.* Springer-Verlag, New York, 1967.

[20] L. E. Payne, and H. F. Weinberger: *An optimal Poincaré inequality for convex domains.* Arch. Rational Mech. Anal. **5**, 286–292, 1960.

[21] J. Wloka: *Partielle Differentialgleichungen.* B. G. Teubner, Stuttgart, 1982.