Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Data-Sparse Approximation of Integral Operators

by

Boris N. Khoromskij

Lecture note no.: 17 2003

Data-Sparse Approximation of Integral Operators*

Boris N. Khoromskij
Max-Planck-Institute for Mathematics in the Sciences,
Inselstr. 22-26, D-04103 Leipzig, Germany.
bokh@mis.mpg.de

Contents

*Notes concerning the lecture course " Data-Sparse Approximation of Integral Operators" given at the University of Leipzig in the winter semester 2002/2003.

Preface

These notes are based on a lecture course given by the author in the winter semester of 2002–2003 for postgraduate students at the University of Leipzig. The purpose of this course was to provide an introduction to modern *methods of a data-sparse approximation to integral and more general nonlocal operators* based on the use of hierarchical matrices (or briefly $\mathcal{H}$-matrices [13]). Being a direct descendant of well established panel clustering, fast multipole and mosaic-skeleton methods, the $\mathcal{H}$-matrix technique allows in addition matrix-matrix operations of almost linear cost. As a consequence, it provides a constructive tool for the efficient matrix representation and fast matrix arithmetics in a wide range of applications.

We focus on the error and complexity analysis of the $\mathcal{H}$-matrix approximation to integral operators. To summarize briefly, a class of operators with asymptotically smooth kernels can be approximated up to a tolerance $O(N^{-\alpha})$, $\alpha > 0$, by data-sparse $N \times N$ $\mathcal{H}$-matrix of almost linear complexity $O(N \log^g N)$, where N is the size of a discrete problem.

We begin with a short survey on the Galerkin finite element methods for strongly elliptic variational equations. Next we give an insight to the classical polynomial approximation of multivariate functions. To proceed with, we construct a hierarchical decomposition of the product integration domain that refines adaptively towards the set of singularity points of the kernel. The polynomial interpolation allows then a patch-wise degenerate approximation to the kernel function in issue. The desired data-sparse $\mathcal{H}$-matrix approximation [13] comes up through the finite element Galerkin method applied to the operator with a modified kernel as above.

All in all, the $\mathcal{H}$-matrix format can be specified by a list of low-rank matrix blocks within a hierarchical matrix partitioning that is refined towards the main diagonal. We discuss the key building blocks in the $\mathcal{H}$-matrix construction and prove the linear-logarithmic costs for the memory requirements and for the matrix-by-vector product. Moreover, the matrix-matrix multiplication and the $\mathcal{H}$-matrix inverse can be also implemented (approximately) within the given format, and with $O(N \log^q N)$ complexity.

The basic hierarchical format can be improved and generalised in several directions as follows:

(i) Uniform and $\mathcal{H}^2$-matrices [20];

(ii) blended $\mathcal{H}$-matrix approximation [18];

(iii) coarsening of the hierarchical format using weaker admissibility criteria [19];

(iv) wire-basket approximation for $\mathcal{L}$-harmonic kernels [17];

(v) hierarchical approximation on graded meshes [16];

(vi) direct data-sparse representation for the Green function [23];

(vii) hierarchical Kronecker tensor-product formats [21].

Whereas topics (i), (ii) will be considered in Lectures 9, 10, the remaining items are addressed in [13]-[21], [23], where the comprehensive analysis of the $\mathcal{H}$-matrix techniques is presented (see also [1, 2, 11, 24]). In particular, implementational aspects are considered in details in [3]. The $\mathcal{H}$-matrix approximation to a class of operator-valued functions with applications to the elliptic, parabolic and hyperbolic problems as well as in control theory has been addressed in [7]-[10].

The author is grateful to Prof. Dr. W. Hackbusch for extensive joint works on the topic and for valuable discussions which have actually inspired this lecture course. I am thankful to Mrs. V. Khoromskaia for the help with typing the LaTeX-files.

Leipzig, March 2003.

1 Galerkin approximation to variational elliptic equations

In the present and in the following lecture, we consider the main ingredients of the finite element method (FEM) for solving the variational elliptic equations. The basic theory can be found in [5], [4], [12], [25].

1.1 Variational equations

Given a Hilbert space V and its dual V', continuous bilinear form $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ and continuous linear functional $f : V \to \mathbb{R}$, that is

$$\exists c_a > 0 : \quad a(u,v) \le c_a \|u\|_V \|v\|_V \qquad \forall u, v \in V,$$
$$\exists c_f > 0 : \quad |f(v)| \le c_f \|u\|_V \qquad \forall v \in V.$$

Problem 1. *Find $u \in V$ such that*

$$a(u,v) = f(v) \quad \forall v \in V. \tag{1.1}$$

The following theorem plays the crucial role in the analysis of finite element methods (see Ciarlet [5] for the proof).

Theorem 1.1 *(Lax-Milgram lemma). Let $a(\cdot,\cdot)$ and $f(\cdot)$ be continuous. If $a(\cdot,\cdot)$ is V-elliptic (coercive), i.e.,*

$$\exists \alpha > 0 : \quad a(v,v) \ge \alpha \|v\|_V^2 \qquad \forall v \in V,$$

then $\exists$ unique $u \in V$, solution of (1.1), such that

$$\|u\|_V \le \alpha^{-1} \|f\|_{V'}.$$

Let $a(\cdot,\cdot)$ be symmetric, $a(u,v) = a(v,u)$ $\forall u, v \in V$. Then u solves (1.1) iff u is the minimiser of $J(u)$,

$$u \in V : \quad J(u) \le J(v) := \frac{1}{2} a(v,v) - f(v) \qquad \forall v \in V.$$

In the following we deal with two principal examples.
Example 1. Given functions $F \in L^2(\Omega)$, $\Psi \in L^2(\Gamma)$ (right-hand sides) and L^∞-coefficients a_{ij}, b_j, a_0. Consider equation (1.1) in the following setting

$$a(u,v) := \int_\Omega \Big(\sum_{i,j=1}^d a_{ij} \partial_j u \partial_i v + \sum_{i=1}^d b_i \partial_i u v + a_0 u v \Big) dx,$$

$$f(v) := (F,v)_{0,\Omega} + (\psi, v)_{0,\Gamma}, \quad V = H^1(\Omega),$$

where $\Omega \in \mathbb{R}^d$ is a polygonal (polyhedral) domain, $\Gamma = \partial\Omega$, $d = 2, 3$. Equation (1.1) corresponds to the elliptic boundary value problem

$$\mathcal{L}u = F \text{ in } \Omega \in \mathbb{R}^d, \qquad \gamma_1 u := \sum_{i=1}^{d} a_{ij} n_i \partial_i u = \Psi \text{ on } \Gamma$$

with

$$\mathcal{L} := -\sum_{i,j=1}^{d} \partial_i a_{ij} \partial_j + \sum_{i=1}^{d} b_i \partial_i + a_0,$$

where γ_1 is the conormal derivative and $\mathbf{n} = (n_1, ..., n_d)^T$ denotes the external unit normal vector on Γ. In general, for second order equations we have $V \subset H^1(\Omega)$, where the Sobolev space

$$H^1(\Omega) := \{u \in L^2(\Omega) : \ |\nabla u| \in L^2(\Omega)\}$$

is equipped with the norm

$$\|u\|_{1,\Omega} := \left(\int_\Omega (|\nabla u|^2 + |u|^2) dx \right)^{1/2}.$$

In the following, we are interested in:
(i) the efficient FEM approximation to u, solution of (1.1);
(ii) the data-sparse approximation to the elliptic operator inverse $\mathcal{L}^{-1}$.

Example 2. Given the kernel function $s(x, y)$, $(x, y) \in \Sigma \times \Sigma$, associated with an integral operator on Σ, define the bilinear form

$$a(u, v) := \int_\Sigma s(x, y) u(x) v(y) dx dy, \tag{1.2}$$

where
(**a**) $\Sigma := \Omega \in \mathbb{R}^d$ for evaluation of the volume potential

$$\mathcal{A}u := \int_\Omega s(x, y) u(y) dy, \qquad x \in \Omega,$$

(**b**) $\Sigma := \Gamma = \partial\Omega$, $d = 2, 3$, for solving the boundary integral equation

$$\mathcal{A}u := \int_\Gamma s(x, y) u(y) dy = F(x), \qquad x \in \Gamma \tag{1.3}$$

by the boundary element methods (BEM).

The BEM is applied to elliptic equations with constant coefficients. We choose $s(x, y)$ as the corresponding fundamental solution. For example, with $\mathcal{L} = -\Delta$ (the Laplace operator) there holds

$$s(x, y) := \frac{1}{2\pi} \log |x - y|, \qquad d = 2,$$

$$s(x, y) := \frac{1}{4\pi} \frac{1}{|x - y|}, \qquad d = 3,$$

where $|x-y|$ is the Euclidean distance in $\mathbb{R}^d$. In case (**b**) we solve the variational equation

$$a(u,v) = f(v) := (F,v)_{0,\Gamma} \qquad \forall v \in V, \tag{1.4}$$

where $a(\cdot,\cdot)$ is given by (1.2) with the possible choice of a Hilbert space $V(\Gamma) \subset \{L^2(\Gamma),\ H^{1/2}(\Gamma),\ H^{-1/2}(\Gamma)\}$.

We are interested in solving (1.4) by low-order finite element methods with almost linear cost with respect to the discrete problem size.

1.2 Galerkin method

Let $V_h \subset V \subset H^1(\Omega)$ or $V_h \subset V(\Gamma)$, $h \to 0$, where h is the discretization parameter and $\dim V_h = N < \infty$. The finite dimensional space V_h is assumed to have the approximation property

$$\inf_{v_h \in V_h} \|v - v_h\| \to 0 \quad \text{as } h \to 0 \quad \forall v \in V.$$

The Galerkin approximation to (1.1) reads as:

$$\text{Find} \quad u_h \in V_h : \ a(u_h, v) = f(v) \qquad \forall v \in V_h. \tag{1.5}$$

Theorem 1.2 *(Cea's lemma). Under the assumptions of Theorem 1.1 (Lax-Milgram lemma), there exists a unique solution u_h to (1.5) such that*

$$\|u - u_h\|_V \le \frac{c_a}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V, \tag{1.6}$$

$$\|u_h\|_V \le \frac{1}{\alpha}\|f\|_{V'}.$$

Proof. Subtract (1.1) and (1.5), then

$$a(u - u_h, v) = 0 \quad \forall v \in V_h \quad \text{(Galerkin orthogonality)}.$$

For an arbitrary $v_h \in V_h$, setting $v = v_h - u_h$ leads to

$$a(u - u_h, v - v_h) = 0$$

that implies

$$\begin{aligned} \alpha\|u - u_h\|_V^2 &\le a(u - u_h, u - u_h) \\ &\le a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h) \\ &\le c_a\|u - u_h\|_V\|u - v_h\|_V, \end{aligned}$$

which proves (1.6). Now the stability easily follows

$$\alpha\|u_h\|_V^2 \le a(u_h, u_h) = f(u_h) \le \|f\|_{V'}\|u_h\|_V.$$

■

In the symmetric case, (1.5) is called the Ritz-Galerkin method.

Theorem 1.3 *(Refinement of Cea's lemma in symmetric case). Let $a(\cdot,\cdot)$ be symmetric, then*

$$\|u-u_h\|_a = \min_{v_h\in V_h} \|u-v_h\|_a, \tag{1.7}$$

where $\|v\|_a^2 := a(v,v)$. Moreover, (1.6) can be improved

$$\|u-u_h\|_V \le \sqrt{\frac{c_a}{\alpha}} \inf_{v_h\in V_h} \|u-v_h\|_V.$$

Proof. Define the new inner product $(v,w)_a := a(v,w)$. We then deduce the Galerkin orthogonality property

$$(u-u_h, v_h)_a = 0 \qquad \forall v_h \in V_h, \tag{1.8}$$

i.e., the error $u-u_h$ is orthogonal to V_h in $(\cdot,\cdot)_a$. By virtue of (1.8),

$$\begin{aligned} \|u-u_h\|_a^2 &= (u-u_h, u)_a - (u-u_h, u_h)_a \\ &= (u-u_h, u)_a \\ &= (u-u_h, u-v_h)_a \quad \forall v_h \in V_h. \end{aligned}$$

Hence, by the Cauchy-Schwarz inequality,

$$\|u-u_h\|_a^2 = (u-u_h, u-v_h)_a \le \|u-u_h\|_a \|u-v_h\|_a \quad \forall v_h \in V_h,$$

which implies

$$\|u-u_h\|_a \le \|u-v_h\|_a \quad \forall v_h \in V_h,$$

and consequently,

$$\|u-u_h\|_a = \min_{v_h\in V_h} \|u-v_h\|_a.$$

Thus,

$$\|u-u_h\|_V \le \frac{1}{\sqrt{\alpha}}\|u-u_h\|_a = \min_{v_h\in V_h} \frac{1}{\sqrt{\alpha}}\|u-v_h\|_a \le \sqrt{\frac{c_a}{\alpha}} \min_{v_h\in V_h} \|u-v_h\|_V.$$

■

1.3 First Strang lemma (variational crime)

Instead of (1.5) consider the generalised Galerkin method:

$$\textit{Find} \quad u_h \in V_h: \quad a_h(u_h, v) = f_h(v) \quad \forall v \in V_h. \tag{1.9}$$

(Instances: numerical integration, approximation to $s(x,y)$, collocation, etc.)

Remark 1.4 *For (1.9) the Galerkin orthogonality no longer holds*

$$a(u-u_h, v_h) \neq 0, \quad v_h \in V_h.$$

Theorem 1.5 *(First Strang lemma). Under the assumptions of Theorem 1.2, suppose that $f_h(\cdot)$ is a linear map and $a_h(\cdot,\cdot)$ is uniformly V-elliptic over $V_h \times V_h$, i.e.,*

$$\exists \alpha^* > 0: \quad a_h(v_h, v_h) \geq \alpha^* \|v_h\|_V^2 \qquad \forall v_h \in V_h,\ \forall h > 0. \tag{1.10}$$

Then there exists a unique solution u_h to (1.9), such that

$$\|u_h\|_V \leq \frac{1}{\alpha^*} \sup_{v_h \in V_h \setminus \{0\}} \frac{f_h(v_h)}{\|v_h\|_V}, \tag{1.11}$$

$$\begin{aligned} \|u - u_h\|_V &\leq \inf_{w_h \in V_h} [(1 + \frac{c_a}{\alpha^*})\|u - w_h\|_V \\ &+ \frac{1}{\alpha^*} \sup_{v_h \in V_h \setminus \{0\}} \frac{|a(w_h, v_h) - a_h(w_h, v_h)|}{\|v_h\|_V}] \\ &+ \frac{1}{\alpha^*} \sup_{v_h \in V_h \setminus \{0\}} \frac{|f(v_h) - f_h(v_h)|}{\|v_h\|_V}. \end{aligned} \tag{1.12}$$

Proof. Owing to (1.10), existence, uniqueness and (1.11) follow from the Lax-Milgram lemma. Let w_h be an arbitrary element in V_h. Setting $\sigma_h := u_h - w_h$, and using (1.10) we obtain

$$\begin{aligned} \alpha^* \|\sigma_h\|_V^2 &\leq a_h(\sigma_h, \sigma_h) \\ &= a(u - w_h, \sigma_h) + a(w_h, \sigma_h) - a_h(w_h, \sigma_h) + f_h(\sigma_h) - f(\sigma_h), \end{aligned}$$

where $a(u, \sigma_h) - f(\sigma_h) = 0$ is the exact variational equation. Assuming $\sigma_h \neq 0$, continuity of $a(\cdot,\cdot)$ leads to

$$\begin{aligned} \alpha^* \|\sigma_h\|_V &\leq c_a \|u - w_h\|_V + \frac{|a(w_h, \sigma_h) - a_h(w_h, \sigma_h)|}{\|\sigma_h\|_V} + \frac{|f_h(\sigma_h) - f(\sigma_h)|}{\|\sigma_h\|_V} \\ &\leq c_a \|u - w_h\|_V + \sup_{v_h \in V_h \setminus \{0\}} \frac{|a(w_h, v_h) - a_h(w_h, v_h)|}{\|v_h\|_V} + \sup_{v_h \in V_h \setminus \{0\}} \frac{|f_h(v_h) - f(v_h)|}{\|v_h\|_V}, \end{aligned}$$

which is also true when $\sigma_h = 0$. Combining it with the triangle inequality, that is $\|u - u_h\|_V \leq \|u - w_h\|_V + \|\sigma_h\|_V$, and taking the infimum with respect to $w_h \in V_h$, we obtain (1.12). ∎

As result, we have proved the perturbed version of Cea's lemma.

1.4 Equation in the matrix form

With given nodal basis of $V_h = span\{\varphi_i\}_{i=1}^N$, the Galerkin matrix and load vector are given by

$$A_h := \{a(\varphi_i, \varphi_j)\}_{i,j=1}^N, \qquad F_h := \{f(\varphi_i)\}_{i=1}^N \in \mathbb{R}^N.$$

Representing the solution u_h in the form $u_h = \sum u_i \varphi_i$, we arrive at the system of linear algebraic equations

$$A_h U = F_h,$$

where $U = (u_1, ..., u_N)^T$. In the FEM applications (cf. Example 1), A_h is already sparse (it contains only $O(N)$ nonzero matrix entries). Our goal is the data-sparse approximation to A_h^{-1} of the complexity $O(N \log^q N)$.

In the BEM applications (cf. Example 2), A_h is a fully populated $N \times N$ matrix. In this case our goal is the data-sparse approximation to A_h providing $O(N \log^q N)$ complexity for both the memory requirements and matrix-vector multiplication.

2 Survey on finite element methods

Let us consider an elliptic variational problem (1.1) with $H_0^1(\Omega) \subset V \subset H^1(\Omega)$ corresponding to Example 1.

2.1 Basic definitions

Definition 2.1 *A triangulation $\mathcal{T} = \{K_1, \ldots, K_M\}$ of a polygonal domain Ω is a finite collection of open triangles $\{K_i\}$ s.t.*

(1) $K_i \cap K_j = \emptyset$ *if* $i \neq j$,
(2) $\cup_i \overline{K}_i = \overline{\Omega}$, *and*
(3) *No vertex of any triangle lies in the interior of an edge of another triangle*

When $\Omega \in \mathbb{R}^d$, $d = 3$, we can substitute triangles by tetrahedra. Introduce the FE space of linear triangular elements in $C^0(\Omega)$, (so-called P_1-elements)

$$V_h := \{v \in C^0(\overline{\Omega}) : v_{|K_i} \in \mathcal{P}_1(K_i) \quad \forall K_i \in \mathcal{T}\}. \tag{2.1}$$

There holds $dim V_h = N$ - the number of nodal points $z_i \in \mathcal{N} := \{\text{set of nodal points}\}$.

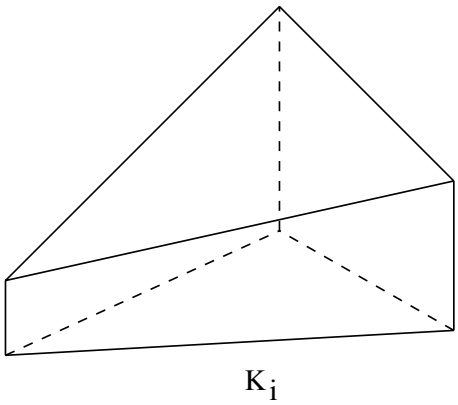

Figure 1: Linear finite element (Courant triangle).

Define the *nodal basis* in $V_h := span\{\varphi_i\}_{i=1}^N$ with $\varphi_i(z_j) = \delta_{ij}$ the Kronecker symbol. In each $K_i \in \mathcal{T}$, we have $v|_{K_i} = a + bx + cy$, and $v \in V_h$ is uniquely defined by its *nodal* values.

Definition 2.2 *A family $\mathcal{T} = \{K_i\}$ is called shape regular (κ -regular) if there exists $\kappa > 0$ such that each $K \in \mathcal{T}$ contains a circle of radius ρ_K, with*

$$\rho_K \geq \frac{h_K}{\kappa}, \qquad h_K = \frac{1}{2} diam K.$$

$\mathcal{T}$ is called quasi-uniform if the above holds with fixed $\rho_K = h$ for all $K \in \mathcal{T}$.

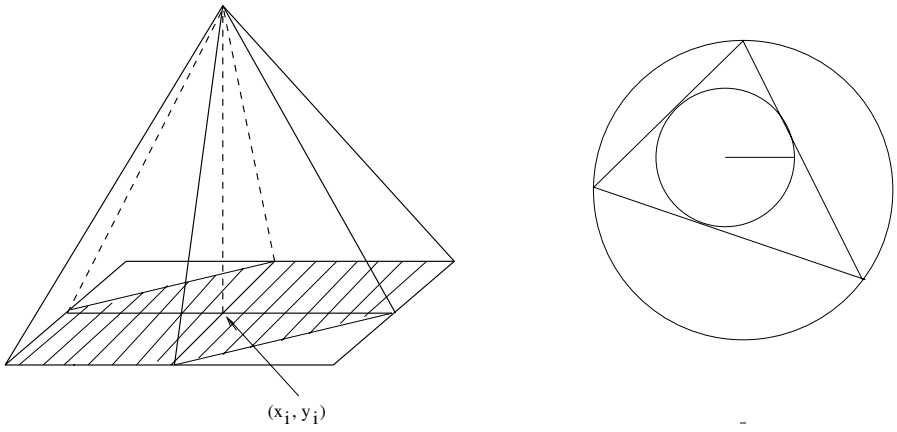

Figure 2: Piecewise linear basis function (left) and a regular triangle (right).

Remark 2.3 *There holds* $V_h \subset H^1(\Omega)$, $\|v\|^2_{1,\Omega} = \sum_{K_i \in \mathcal{T}} \|v\|^2_{1,K_i}$ *for* $v \in V_h$.

Definition 2.4 *Given* $v \in C^0(\overline{K})$, $K \in \mathcal{T}$. *We introduce the local linear interpolant* $\mathcal{I}_K v$ *by*

$$\mathcal{I}_K v := \sum_{z_i \in K} v(z_i)\varphi_i.$$

For $v \in C^0(\overline{\Omega})$ *the global linear interpolant* $\mathcal{I}_h v \in C^0(\overline{\Omega})$ *is defined on* $\overline{\Omega}$ *by*

$$\mathcal{I}_h v_{|K_i} = \mathcal{I}_{K_i} v \qquad \forall K_i \in \mathcal{T}.$$

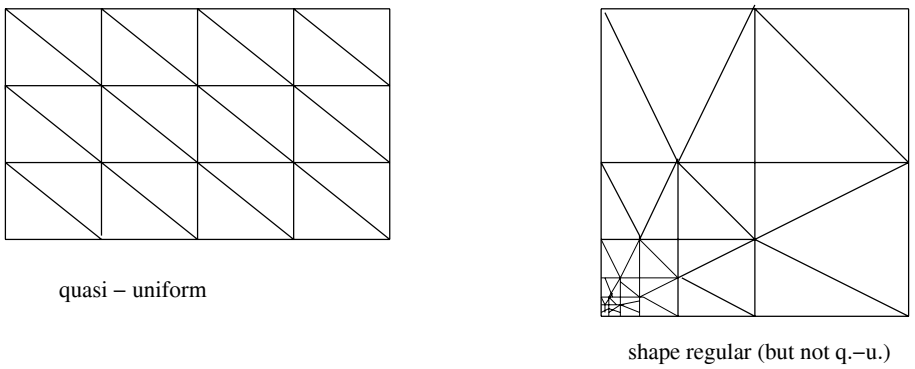

Figure 3: Quasi-uniform (left) and shape regular (right) meshes.

Our goal is to estimate

$$\|v - \mathcal{I}_h v\|_{m,\Omega} \leq C(h)\|v\|_{t,\Omega}, \quad 0 \leq m \leq t \leq 2.$$

In this way we recall that, due to Cea's lemma there holds

$$\|u - u_h\|_V \leq \frac{c_a}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V \leq \frac{c_a}{\alpha}\|u - \mathcal{I}_h u\|_V.$$

2.2 The Bramble-Hilbert lemma

Lemma 2.5 *(Interpolation by P_1 polynomials). Let $t = 2$ and Ω be a triangle. Then the interpolation operator $\mathcal{I}_h : H^t \to P_1$ is well defined. There is $C = C(\Omega, \mathcal{N})$ such that*

$$\|u - \mathcal{I}_h u\|_{t,\Omega} \le C|u|_{t,\Omega} \quad \forall u \in H^t(\Omega). \tag{2.2}$$

For $t = 1$ (2.2) holds by choosing the local interpolation

$$\mathcal{I}_K v = \frac{\int_K v dx}{\int_K 1 dx} \qquad (\mathcal{I}_K v \text{ is the constant, i.e., } \mathcal{I}_K v \notin P_1).$$

Proof. For $t = 2$ let $s = t(t+1)/2 = 3$ be the number of interpolation points in Ω. We endow $H^t(\Omega)$ with the norm

$$|||v||| := |v|_t + \sum_{i=1}^{s} |v(z_i)|,$$

and show that $|||\cdot||| \sim \|\cdot\|_t$. Then (2.2) will follow from

$$\begin{aligned} \|u - \mathcal{I}_h u\|_t &\le c|||u - \mathcal{I}_h u||| \\ &= C[|u - \mathcal{I}_h u|_t + \sum_{i=1}^{s} |(u - \mathcal{I}_h u)(z_i)] \\ &= C|u - \mathcal{I}_h u|_t = C|u|_t \end{aligned}$$

since $u(z_i) = \mathcal{I}_h u(z_i)$, and $D^\alpha \mathcal{I}_h u = 0$, $|\alpha| = t$.
Proof of the norm equivalence: (a) The embedding $H^2 \hookrightarrow C^0$ is continuous, hence

$$|v(z_i)| \le C\|v\|_t, \qquad i = 1, \dots N, \quad \text{implying } |||v||| \le (1 + Cs)\|v\|_t.$$

Other direction: (b) Suppose the converse

$$\|v\|_t \le C|||v||| \quad \forall v \in H^t(\Omega)$$

fails for $\forall C > 0$. Then $\exists$ a sequence $\{v_k\} \in H^t(\Omega)$,

$$\|v_k\|_t = 1, \quad |||v_k||| \le \frac{1}{k}, \quad k = 1, 2, \dots$$

Since $H^t(\Omega) \hookrightarrow H^{t-1}(\Omega)$ is compact, $\{v_k\}$ converges in $H^{t-1}(\Omega)$. Then $\{v_k\}$ is a Cauchy sequence in H^{t-1}. From $|v_k|_t \to 0$ and

$$\|v_k - v_\ell\|_t^2 \le \|v_k - v_\ell\|_{t-1}^2 + (|v_k|_t + |v_\ell|_t)^2$$

we see that $\{v_k\}$ is a Cauchy sequence in $H^t(\Omega)$. Since H^t is complete it follows, that $v_k \to v_*$ in H^1. By continuity, we have

$$\|v_*\|_t = 1, \qquad |||v_*||| = 0,$$

which is a contradiction, since $|v_*|_t = 0$ implies $v_* \in \mathcal{P}_{t-1}$ and thus $v_* = 0$ since $v_*(z_i) = 0$, $i = 1, \dots, N$. In case $t = 1$, the proof is similar. ∎

Let L be a bounded linear map between two normed spaces and denote $\|L\| := sup\{\|Lv\| : \|v\| = 1\}$. In our application, we shall use $L = I - \mathcal{I}_h$.

Lemma 2.6 *(Bramble-Hilbert lemma) Let $\Omega \subset \mathbb{R}^2$ be a triangle. With $t = 2$, suppose that $L : H^t(\Omega) \to Y$ is a bounded linear mapping onto normed linear space Y (in our case we use $Y = H^t(\Omega), t = 2$). If $\mathcal{P}_{t-1} \subset ker L$, then there is $C = C(\Omega)\|L\| \geq 0$, such that*

$$\|Lv\| \leq c|v|_t \qquad \forall v \in H^t(\Omega). \tag{2.3}$$

Proof. Let $\mathcal{I}_h : H^t(\Omega) \to \mathcal{P}_{t-1}$ be an interpolation operator (see Lemma 10.2). Then Lemma 10.2 and the fact $\mathcal{I}_h v \in ker L$ imply

$$\|Lv\|_Y = \|L(v - \mathcal{I}_h v)\|_Y \leq \|L\| \cdot \|v - \mathcal{I}_h v\|_t \leq C\|L\||v|_t,$$

where C appears in (2.2). ∎

2.3 Central approximation result

Theorem 2.7 *(Central approximation result). Suppose $\mathcal{T}$ is a shape regular triangulation of Ω. Then there exists a constant $C = C(\Omega, \kappa)$ such that*

$$\|u - \mathcal{I}_h u\|_{m,\Omega} \leq Ch^{2-m}|u|_{2,\Omega} \qquad \forall u \in H^2(\Omega), \;\; 0 \leq m \leq 2, \tag{2.4}$$

where $\mathcal{I}_h : H^2(\Omega) \to V_h$ is the linear interpolant.

Proof. (For the case of regular grid, i.e., all triangles are congruent). Each $K_h \subset \mathcal{T}$ is a scaled version of a reference triangle K_1,

$$K_h := hK_1 = \{x = hy : \, y \in K_1\}, \;\; h \leq 1.$$

Then we prove

$$\|u - \mathcal{I}_h u\|_{m,K_h} \leq Ch^{2-m}|u|_{2,K_h}, \quad 0 \leq m \leq 2, \tag{2.5}$$

where C is from (2.2) and $\mathcal{I}_h u \in \mathcal{P}_1(K_h)$ interpolates u at nodal points. Given $u \in H^2(K_h)$, define $v \in H^1(K_1)$ by $v(y) := u(hy)$. Then $\partial^\alpha v = h^{|\alpha|}\partial^\alpha u$ for $|\alpha| \leq 2$. Since the transformation of the area in $\mathbb{R}^2$ yields an extra factor h^{-2}, we get

$$|v|^2_{\ell,K_1} = \sum_{|\alpha|=\ell} \int_{K_1} (\partial^\alpha v)^2 dy = \sum_{|\alpha|=\ell} \int_{K_h} h^{2\ell}(\partial^\alpha u)^2 h^{-2} dx = h^{2\ell-2}|u|^2_{\ell,K_h}. \tag{2.6}$$

After summation, the smallest power dominates (if $h \leq 1$ then h^{-2m+2} dominates):

$$\|u\|^2_{m,K_h} = \sum_{\ell \leq m} |u|^2_{\ell,K_h} = \sum_{\ell \leq m} h^{-2\ell+2}|v|^2_{\ell,K_1} \leq h^{-2m+2}\|v\|^2_{m,K_1}. \tag{2.7}$$

Now inserting $u - \mathcal{I}_h u$ in place of u in (2.7) and combining the last two formulae with Lemma 10.2, we get

$$\begin{aligned} \|u - \mathcal{I}_h u\|_{m,K_h} &\leq h^{-m+1}\|v - \mathcal{I}_h v\|_{m,K_1} \\ &\overset{(m\leq 2)}{\leq} h^{-m+1}\|v - \mathcal{I}_h v\|_{2,K_1} \overset{(Lemma 2.5)}{\leq} ch^{-m+1}|v|_{2,K_1} \\ &\overset{((2.6) \text{ with } \ell=2)}{\leq} ch^{2-m}|u|_{2,K_h}, \end{aligned}$$

for $m \leq 2$ and then (2.5) follows. Now we arrive at (2.4) by squaring and summation of (2.5) over all $K_h \subset \mathcal{T}$. ∎

2.4 Error bounds for 2nd order problems. Inverse estimate

Theorem 2.8 *Let $\mathcal{T}$ be a shape-regular triangulation, and V_h is given by (2.1). Then the FE approximation*

$$u_h \in V_h : \ a(u_h, v) = (F, v)_0 \quad \forall v \in V_h$$

satisfies

$$\|u - u_h\|_{1,\Omega} \le ch|u|_{2,\Omega} \le ch\|F\|_{0,\Omega}, \tag{2.8}$$

where the latter holds for convex domains. Let Ω be convex and $F \in L^2(\Omega)$ then $u \in H^2(\Omega)$ and

$$\|u - u_h\|_{0,\Omega} \le ch\|u - u_h\|_{1,\Omega} \le ch|u|_{1,\Omega},$$
$$\|u - u_h\|_{0,\Omega} \le ch^2\|F\|_{0,\Omega}. \tag{2.9}$$

Proof. By Theorem 2.7,

$$\|u - \mathcal{I}_h u\|_{1,\Omega} \le ch|u|_{2,\Omega},$$

then Cea's lemma ensures (2.8), where for convex domains $\|u\|_2 \le c\|F\|_0$ (full elliptic regularity).

The L^2-norm estimate (2.9) is based on the Galerkin orthogonality and a duality argument often called Aubin-Nitsche trick (cf. [4, 12]). ∎

Estimate (2.4) provides a bound for coarser norm via finer norm. Bounds in the opposite direction are called *inverse estimates.*

Lemma 2.9 *Let V_h be the FE space corresponding to piecewise polynomial of degree $k = 0,\ 1$ on a quasi-uniform triangulation. There exists $C = C(\kappa, k)$ such that for $0 \le m \le t \le 2$, $k \le m$,*

$$\|v\|_t \le ch^{m-t}\|v\|_m \qquad \forall v \in V_h. \tag{2.10}$$

Proof. See [4, 12]. ∎

The pair of inequalities (2.4) and (2.10) is called *optimal* provided that we have the same order of h and h^{-1}. They play a major role in classical approximation theory. As a simple consequence of Lemma 2.9, we obtain a condition number estimate for the FE stiffness matrix A_h corresponding to second order elliptic problems

$$cond(A_h) = O(h^{-2}).$$

3 Introduction to polynomial approximation

We discuss the classical polynomial approximation to multivariate functions $f : \mathbb{R}^d \to \mathbb{R}$, $d \geq 1$, defined on the d-dimensional hyper-cube $B_d := B^d$, where $B = [-1, 1]$ is the reference interval. The function $f(x_1, \ldots, x_d)$ is supposed to have a holomorphic extension to the neighbourhood of B_d with respect to each space variable (cf. Assumption 4.1).

3.1 Best polynomial approximation of analytic functions

In the complex plane $\mathbb{C}$, we introduce the circular ring

$$\mathcal{R}_\rho := \{z \in \mathbb{C} : \ 1/\rho < |z| < \rho\}$$

with $\rho > 1$. By $\mathcal{E}_\rho = \mathcal{E}_\rho(B)$, we denote the so-called Bernstein's regularity ellipse (with foci at $z = \pm 1$ and the sum of semi-axes equal to $\rho > 1$),

$$\mathcal{E}_\rho := \{z \in \mathbb{C} : \ |z - 1| + |z + 1| \leq \rho + \rho^{-1}\}.$$

Theorem 3.1 *(Laurent's Theorem). Let $\rho > 1$, let $f : \mathbb{C} \to \mathbb{C}$ be analytic and bounded by $M > 0$ in $\mathcal{R}_\rho$ (in the following we say $f \in \mathcal{A}_\rho$), and set*

$$C_n := \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{in\theta} d\theta, \qquad n = 0, \ \pm 1, \ \pm 2, \ldots. \tag{3.1}$$

Then for all $z \in \mathcal{R}_\rho$, $f(z) = \sum_{n=-\infty}^{\infty} C_n z^n$, where the series converges to $f(z)$ for all $z \in \mathcal{R}_\rho$. Moreover $|C_n| \leq M/\rho^{|n|}$, and for all $\theta \in [0, 2\pi]$ and arbitrary integer m,

$$\left| f(e^{i\theta}) - \sum_{n=-(m-1)}^{m-1} C_n e^{in\theta} \right| \leq \frac{2M}{\rho - 1} \rho^{-m+1}. \tag{3.2}$$

Proof. See any standard book on the analysis of functions of complex variables. ∎

Let $T_n(w)$, $n = 0, 1, 2, \ldots$, be the Chebyshev polynomials, which may be defined recursively by means of the equation

$$\begin{aligned} T_0(w) &= 1, \qquad T_1(w) = w, \\ T_{n+1}(w) &= 2wT_n(w) - T_{n-1}(w), \quad n = 1, 2, \ldots. \end{aligned}$$

Note that $T_n(x) = \cos(n \arccos x), \quad x \in [-1, 1]$, which implies $T_n(1) = 1$, $T_n(-1) = (-1)^n$. It can be seen that with $w = \frac{1}{2}(z + \frac{1}{z})$, there holds

$$T_n(w) = \frac{1}{2}(z^n + z^{-n}). \tag{3.3}$$

The following theorem presents classical results on the best polynomial approximation for analytic functions.

Theorem 3.2 *Let $\rho > 1$, and let F be analytic and bounded by M in $\mathcal{E}_\rho$. Then the expansion*

$$F(w) = C_0 + 2\sum_{n=1}^{\infty} C_n T_n(w), \tag{3.4}$$

holds for all $w \in \mathcal{E}_\rho$ (Chebyshev series), and with

$$C_n = \frac{1}{\pi}\int_{-1}^{1} \frac{F(w)T_n(w)}{\sqrt{1-w^2}}dw.$$

Moreover, $|C_n| \leq M/\rho^n$ and for $w \in B$ and for $m = 1, 2, 3, \ldots,$

$$|F(w) - C_0 - 2\sum_{n=1}^{m-1} C_n T_n(w)| \leq \frac{2M}{\rho - 1}\rho^{-m+1}, \quad w \in B. \tag{3.5}$$

Proof. Let $\mathcal{A}_{\rho,s}$ denote the subclass of functions in $\mathcal{A}_\rho$ for which $C_{-n} = C_n$, then each $f \in \mathcal{A}_{\rho,s}$ has a convergent series representation (cf. Theorem 3.1)

$$f(z) = C_0 + \sum_{n=1}^{\infty} C_n(z^n + z^{-n}), \qquad z \in \mathcal{R}_\rho. \tag{3.6}$$

Furthermore, from (3.6) it is easy to see that

$$f(1/z) = f(z), \qquad z \in \mathcal{R}_\rho.$$

Let us apply the mapping $w = \frac{1}{2}(z + \frac{1}{z})$, which is a conformal transform of $\{\xi \in \mathcal{R}_\rho : |\xi| > 1\}$ onto $\mathcal{E}_\rho$ as well as it is also a conformal transform of $\{\xi \in \mathcal{R}_\rho : |\xi| < 1\}$ onto $\mathcal{E}_\rho$ (but not $\mathcal{R}_\rho$ onto $\mathcal{E}_\rho$!). This mapping satisfies the equation $w(1/z) = w(z)$ and it provides a one to one correspondence of functions F that are analytic and bounded by M in $\mathcal{E}_\rho$ with functions f in $\mathcal{A}_{\rho,s}$.

Furthermore, under this mapping we have (3.3). Hence it follows that if f defined by (3.6) is in $\mathcal{A}_{\rho,s}$, then the corresponding transformed function $F(w) = f(z(w))$ that is analytic and bounded by M in $\mathcal{E}_\rho$ is given by (3.4). Now the result follows directly by Theorem 3.1. ∎

3.2 Polynomial interpolation

Let $\mathcal{P}_N(B)$ be the set of polynomials of degree $\leq N$ on B. Define by $[\mathcal{I}_N F](x) \in \mathcal{P}_N(B)$ the interpolation polynomial of F with respect to the Chebyshev-Gauss-Lobatto (CGL) nodes

$$\xi_j = \cos\frac{\pi j}{N} \in B, \quad j = 0, 1, \ldots, N, \qquad \text{with } \xi_0 = 1,\ \xi_N = -1,$$

where ξ_j are zeroes of the polynomials $(1-x^2)T_N'(x)$, $x \in B$. In turn, the Lagrangian interpolant $\mathcal{I}_N$ of F has the form

$$\mathcal{I}_N F := \sum_{j=0}^{N} F(\xi_j) l_j(x) \in \mathcal{P}_N(B), \tag{3.7}$$

i.e. $\mathcal{I}_N(\xi_j) = F(\xi_j)$, $j = 0, \ldots, N$, with $l_j(x)$ is the set of interpolation polynomials

$$l_j := \prod_{k=0, j\neq k}^{N} \frac{x - \xi_k}{\xi_j - \xi_k} \in \mathcal{P}_N(B).$$

Clearly, $l_j(\xi_j) = 1$ and $l_j(\xi_k) = 0 \ \forall k \neq j$.

The Lagrangian interpolant can be represented in terms of the orthogonal polynomials on $[-1, 1]$. Usually Chebyshev or Legendre interpolation is applied. In these cases the coefficients have simple explicit representations via the set $\{F(\xi_j)\}$. For example, the Chebyshev interpolant has the form (see [25])

$$[\mathcal{I}_N F](x) = \sum_{k=0}^{N} F_k^* T_k(x),$$

where the expansion coefficients are given by

$$F_k^* = \frac{2}{N d_k} \sum_{j=0}^{N} \frac{1}{d_j} \cos\frac{kj\pi}{N} F(\xi_j)$$

with $d_j = 2$ for $j = 0, N$, and $d_j = 1$ otherwise. Due to its trigonometric structure, the *discrete Chebyshev transform*

$$\{F(\xi_j) | j = 0, \ldots, N\} \to \{F_k^* | k = 0, \ldots, N\}$$

can be computed by the FFT algorithm with $O(N \log_2 N)$ operations (if $N = 2^p$). The same holds for the inverse transform (back-transform)

$$\{F_k^* | k = 0, \ldots, N\} \to \{F(\xi_j) | j = 0, \ldots, N\}$$

since

$$F(\xi_j) = \sum_{k=0}^{N} F_k^* \cos\frac{kj\pi}{N}, \qquad j = 0, \ldots, N.$$

3.3 Interpolation error by the Chebyshev-Gauss-Lobatto nodes

Given the set $\{\xi_j\}_{j=0}^N$ of interpolation points on $[-1, 1]$ and the associated Lagrangian interpolation operator $\mathcal{I}_N$. The standard approximation theory for polynomial interpolation includes the so-called Lebesque constant $\Lambda_N \in \mathbb{R}_{>1}$ defined by

$$\|\mathcal{I}_N u\|_{\infty,B} \leq \Lambda_N \|u\|_{\infty,B} \qquad \forall u \in C(B). \tag{3.8}$$

In the case of Chebyshev interpolation it can be shown that Λ_N grows at most logarithmically in N, more precisely (see [6]),

$$\Lambda_N \leq \frac{2}{\pi} \log N + 1.$$

The interpolation points which produce the smallest value Λ_N^* of all Λ_N are not known, but Bernstein '54 proves that

$$\Lambda_N^* = \frac{2}{\pi} \log N + O(1).$$

Theorem 3.3 *Let $u \in C^\infty[-1,1]$ have an analytic extension to $\mathcal{E}_\rho$ bounded by $M > 0$ in $\mathcal{E}_\rho$. Then there holds*

$$\|u - \mathcal{I}_N u\|_{\infty,I} \leq (1 + \Lambda_N)\frac{2M}{\rho - 1}\rho^{-N}, \qquad N \in \mathbb{N}_{\geq 1}. \tag{3.9}$$

Proof. Due to (3.5) there holds for the best polynomial approximations to u on $[-1,1]$,

$$\min_{v \in \mathcal{P}_N} \|u - v\|_{\infty,B} \leq \frac{2M}{\rho - 1}\rho^{-N}. \tag{3.10}$$

Note that the interpolation operator $\mathcal{I}_N$ is a projection, that is, for all $v \in \mathcal{P}_N$ we have $\mathcal{I}_N v = v$. Then applying the triangle inequality with $v \in \mathcal{P}_N$,

$$\|u - \mathcal{I}_N u\|_{\infty,B} = \|u - v - \mathcal{I}_N(u - v)\|_{\infty,B} \leq (1 + \Lambda_N)\|u - v\|_{\infty,B}$$

completes the proof. ∎

Remark 3.4 *The above approximation results can be easily transformed to an arbitrary closed interval $\widetilde{B} = [a,b]$. The transformed interpolation operator is given by*

$$\mathcal{I}_{N,\widetilde{B}} u := (\mathcal{I}_N(u \circ \Psi_{\widetilde{B}})) \circ \Psi_{\widetilde{B}}^{-1}, \quad \text{where } \Psi_{\widetilde{B}} : B \to \widetilde{B},$$

with $x \mapsto \frac{b-a}{2}x + \frac{b+a}{2}$, is the affine mapping from the reference interval B to $\widetilde{B}$. The corresponding interpolation points and Lagrange polynomials are given by $\widetilde{\xi}_j := \Psi_{\widetilde{B}}(\xi_j)$ and $\widetilde{l}_j := l_j \circ \Psi_{\widetilde{B}}^{-1}$. The scaled Bernstein's regularity ellipse $\mathcal{E}_{\widetilde{\rho}}$ for the interval $B_\delta := [-\delta, \delta]$, $\delta > 0$, has the sum of semi-axes $\widetilde{\rho} = \rho/\delta$. Therefore, the corresponding approximation error is characterised by the exponential $(\rho/\delta)^{-N}$, while the constant M has the same meaning as above.

4 Degenerate approximation of the kernel function

4.1 Polynomial approximation of multivariate functions

We consider a multivariate function $f = f(x_1, \ldots, x_d) : \mathbb{R}^d \to \mathbb{R},\ d \geq 1$, defined on a box $B^d = B_1 \times B_2 \times \ldots \times B_d$ with $B_k = [a_k, b_k]$. In the following, for the ease of presentation, we set $B_k = [-1, 1]$, $k = 1, \ldots, d$.
The corresponding N-th order tensor product interpolation operator is defined by

$$\mathbb{I}_N f = \mathcal{I}_N^1 \times \mathcal{I}_N^2 \times \ldots \times \mathcal{I}_N^d f \in \mathbb{P}_N[B^d],$$

where $\mathcal{I}_N^k f$ denotes the interpolation polynomial with respect to x_k, $k = 1, \ldots, d$, at nodes $\{\xi_k\} \in B_k$ (further, we choose the Chebyshev-Gauss-Lobatto nodes). The interpolation points $\xi_\alpha \in B^d$, $\alpha = (i_1, \ldots, i_d) \in \mathbb{N}_0^d$, are obtained by the Cartesian product of one-dimensional nodes,

$$\xi_\alpha := \left(cos\frac{\pi i_1}{N}, \ldots, cos\frac{\pi i_d}{N} \right).$$

Again, $\mathbb{I}_N$ is the projection map,

$$\mathbb{I}_N : C(B^d) \to \mathbb{P}_N := \{p_1 \times \ldots \times p_d : p_i \in \mathcal{P}_N,\ i = 1, \ldots d\}$$

that implies the following estimate to the multivariate counterpart of the Lebesque constant (stability of $\mathbb{I}_N$ in the multidimensional case; see (3.8))

$$\|\mathbb{I}_N f\|_{\infty, B^d} \leq \Lambda_N^d \|f\|_{\infty, B^d}. \tag{4.1}$$

We denote by X_{-i} the $(d-1)$-dimensional subset of variables $\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d\}$ with $x_i \in B_i$. Define the product domain $\mathcal{E}_\rho^{(j)} := B_1 \times \ldots \times B_{j-1} \times \mathcal{E}_\rho(I_j) \times B_{j+1} \times \ldots \times B_d\}$.

Assumption 4.1 *Given $f \in \mathbb{C}^\infty(B^d)$, assume there is $\rho > 1$ such that for all $i = 1, \ldots, d$, and each fixed $\xi \in X_{-i}$, there exists an analytic extension of $\hat{f}(x_i) = f(x_i, \xi)$ to $\mathcal{E}_\rho(B_i) \subset \mathbb{C}$ with respect to x_i.*

Theorem 4.2 *For $f \in C^\infty(B^d)$, let Assumption 4.1 be satisfied. Then the interpolation error can be estimated by*

$$\|f - \mathbb{I}_N f\|_{\infty, B^d} \leq \Lambda_N^d \frac{2M_\rho(f)}{\rho - 1} \rho^{-N}, \tag{4.2}$$

where Λ_N is the Lebesque constant for the one-dimensional interpolant $\mathcal{I}_N^k$, and

$$M_\rho(f) := \max_{1 \leq j \leq d} \{ \max_{x \in \mathcal{E}_\rho^{(j)}} |f(x_1, \ldots, x_d)| \}. \tag{4.3}$$

Proof. The multiple use of (3.8), (3.9) and the triangle inequality lead to

$$\begin{aligned}
|f - \mathbb{I}_N f| &\leq |f - \mathcal{I}_N^1 f| + |\mathcal{I}_N^1(f - \mathcal{I}_N^2 \times \ldots \times \mathcal{I}_N^d f)| \\
&\leq |f - \mathcal{I}_N^1 f| + |\mathcal{I}_N^1(f - \mathcal{I}_N^2 f)| + \\
&+ |\mathcal{I}_N^1 \mathcal{I}_N^2(f - \mathcal{I}_N^3 f)| + \ldots + |\mathcal{I}_N^1 \times \ldots \times \mathcal{I}_N^{d-1}(f - \mathcal{I}_N^d f)| \\
&\leq [(1+\Lambda_N) \max_{x \in \mathcal{E}_\rho^{(1)}} |f(x)| + \Lambda_N(1+\Lambda_N) \max_{x \in \mathcal{E}_\rho^{(2)}} |f(x)| \\
&+ \ldots + \Lambda_N^{d-1}(1+\Lambda_N) \max_{x \in \mathcal{E}_\rho^{(d)}} |f(x)|] \frac{2}{\rho - 1} \rho^{-N} \\
&\leq \frac{(1+\Lambda_N)(\Lambda_N^d - 1)}{\Lambda_N - 1} \frac{2M_\rho}{\rho - 1} \rho^{-N}.
\end{aligned}$$

Hence (4.2) follows since for $x > 1$ there holds $\frac{(1+x)(x^n-1)}{x-1} \leq x^n$. ∎

4.2 Analyticity domain for typical fundamental solutions

Now we demonstrate how to check Assumption 4.1 in standard FEM and BEM applications. Let $\mathcal{L}$ be an elliptic operator of second order with constant coefficients. The so-called fundamental solution $S(x)$, $x \in \mathbb{R}^d$, $d = 1, 2, 3$, associated with $\mathcal{L}$ satisfies

$$\mathcal{L}S = \delta(x) \quad \text{in } \mathbb{R}^d, \quad \delta(x) \text{ is the Dirac distribution.}$$

In the BEM applications, we are interested in approximation to the volume potential

$$(A_\Omega u)(x) := \int_\Omega S(x-y)u(y)dy, \qquad x \in \Omega \in \mathbb{R}^d, \tag{4.4}$$

as well as to its "restriction" onto the boundary (the so-called single layer potential)

$$(A_\Gamma u)(x) := \int_\Gamma S(x-y)u(y)dy, \qquad x \in \Gamma = \partial\Omega. \tag{4.5}$$

The related hypersingular and double layer potentials $\mathcal{D}$ and $\mathcal{K}$ on Γ are given by

$$\mathcal{D}u(x) := -\int_\Gamma \frac{\partial^2 S(x-y)}{\partial n_x \partial n_y} u(y)dy, \quad \mathcal{K}u(x) := \int_\Gamma \frac{\partial S(x-y)}{\partial n_y} u(y)dy.$$

We consider several examples of fundamental solutions. For a given symmetric and positive definite coefficients matrix $\mathbf{a} = \{a_{jk}\} \in \mathbb{R}^{d\times d}$, define the symmetric bilinear form $\langle \cdot, \cdot \rangle_\mathbf{a}$ on $\mathbb{C}^d$ by $\langle x, y \rangle_\mathbf{a} := \langle \mathbf{a}^{-1}x, y \rangle$, where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product in $\mathbb{C}^d$. For $x \in \mathbb{R}^d$, we write $|x|_\mathbf{a}^2 = \langle x, x \rangle_\mathbf{a}$. We also introduce $\mathbf{b} = \{b_i\} \in \mathbb{R}^d$.

Example 4.3 *The Helmholtz operator ($\mathbf{a} \in \mathbb{R}^{2\times 2}$, $\mathbf{b} = 0$, $c_0 \in \mathbb{C}$, $Im\, c_0 \neq 0$).*
Let $\lambda \in \mathbb{C}\backslash(-\infty, 0]$ be chosen from $\lambda^2 = c_0$. A fundamental solution $S_\lambda(x)$ of $\mathcal{L}$ is then given by

$$S_\lambda(x) := \begin{cases} \frac{1}{\pi\sqrt{det\,\mathbf{a}}} \mathrm{i}\, H_0^{(1)}(\mathrm{i}\lambda|x|_\mathbf{a}) & \text{for } d = 2 \\ \frac{1}{4\pi\sqrt{det\,\mathbf{a}}} \frac{\mathrm{e}^{-\lambda|x|_\mathbf{a}}}{|x|_\mathbf{a}} & \text{for } d = 3, \end{cases} \tag{4.6}$$

where, for $d = 2$, $H_0^{(1)}$ is the Hankel function with $arg(\lambda) \in (-\pi, \frac{\pi}{2})$.

Example 4.4 *Advection–diffusion operator ($a_{ij} = \varepsilon\delta_{ij}$, $\varepsilon > 0$, $\mathbf{b} \neq 0$, $c_0 = 0$),*

$$S(x) := \begin{cases} \frac{i\varepsilon}{2\pi} e^{\langle \mathbf{b}, x\rangle/\varepsilon} H_0^{(1)}(\frac{i}{\varepsilon}|\mathbf{b}||x|), & \text{for } d = 2 \\ \frac{\varepsilon^2}{4\pi|x|} e^{\frac{1}{\varepsilon}(\langle \mathbf{b}, x\rangle - |\mathbf{b}||x|)} & \text{for } d = 3. \end{cases} \tag{4.7}$$

Example 4.5 *Let $\mathcal{L} = \Delta^2$ be the biharmonic operator for $d = 2$. The corresponding fundamental solution is given by $S(x) := x^2 \log x$.*

In the particular case of Example 4.3, corresponding to the Laplace operator, we have

$$s(x, y) = S(x - y) := \begin{cases} \frac{1}{2\pi} \log|x - y|, & d = 2 \\ \frac{1}{4\pi} \frac{1}{|x-y|}, & d = 3. \end{cases} \tag{4.8}$$

Given $\delta > 0$, introduce d-dimensional boxes

$$B_x := \{x : x \in [-\delta, \delta]^d\}, \quad B_y := \{y : y - y_0 \in [-\delta, \delta]^d\}, \tag{4.9}$$

located in the distance $\sigma_0 > 0$ (see Fig. 4). Note that the vertical and horizontal shift of B_y is also allowed. For $(x, y) \in B_x \times B_y$, consider the kernel function $s(x, y) :=$

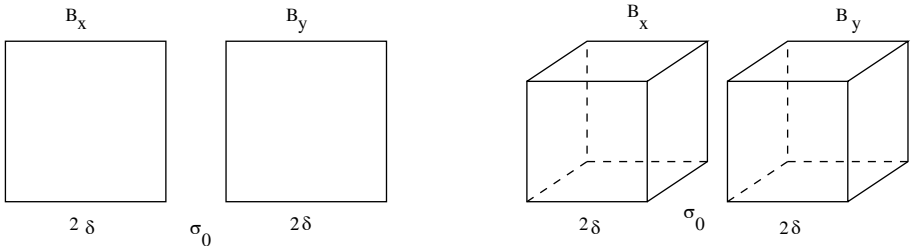

Figure 4: Bounding boxes with $d = 2, 3$.

$S(x - y)$, defined on the tensor-product domain $B_x \times B_y$ ('geometric block').

Lemma 4.6 *Let the function $s(x, y) = S(x - y)$, $(x, y) \in B_x \times B_y$ be defined as in Examples 4.3-4.5, with B_x, B_y in (4.9) such that $dist(B_x, B_y) = \sigma_0 > 0$. Introduce for $\sigma \in (0, \sigma_0)$,*

$$\rho(\sigma) = \min\{1 + \frac{\sigma}{\delta} + \sqrt{(\frac{\sigma}{\delta})^2 + 2\frac{\sigma}{\delta}}, \frac{\sigma}{\delta} + \sqrt{(\frac{\sigma}{\delta})^2 + 1}\}. \tag{4.10}$$

Then the error estimate

$$\|s(x, y) - \mathbb{I}_{N,x} s(x, y)\|_{\infty, B_x} \leq 2\Lambda_N^d \frac{M_{\rho(\sigma^*)}}{\rho(\sigma^*) - 1} \rho(\sigma^*)^{-N} \tag{4.11}$$

holds, where σ^* is defined by

$$\sigma^* = arg\min_{\sigma} \frac{M_{\rho(\sigma)}(s)}{\rho(\sigma) - 1} \rho(\sigma)^{-N}. \tag{4.12}$$

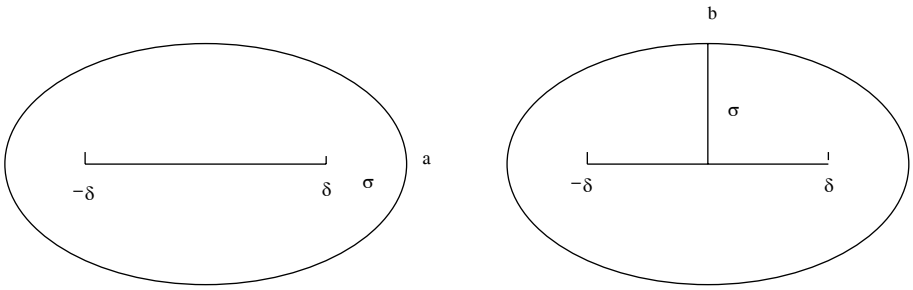

Figure 5: Bernstein's regularity ellipses.

Proof. It is easy to see that for any fixed $y \in B_y$, the function $s(\cdot, y) : B_x \to \mathbb{R}$, satisfies Assumption 4.1 with arbitrary $\rho < \rho_0$, for some $\rho_0 > 1$. To estimate the parameter $\rho_0 = \rho_0(\delta, \sigma_0)$, we construct the corresponding Bernstein's regularity ellipses: (a) with large semi-axis $a_\delta = \delta + \sigma$, $\sigma < \sigma_0$ (see Fig. 4) and (b) with small semi-axis $b_\delta = \sigma < \sigma_0$.

In case (a), we consider the ellipse $\mathcal{E}_\rho$ whose semi-axis with a length a is orthogonal to the nearest facet of B_y. By scaling $\delta \to 1$, $\sigma_0 \to \frac{\sigma_0}{\delta}$, $\sigma \to \frac{\sigma}{\delta}$, we easily obtain that

$$a = \frac{1}{2}(\rho + \frac{1}{\rho}) = 1 + \frac{\sigma}{\delta},$$

which yields for the corresponding sum of the semi-axes

$$\rho = 1 + \frac{\sigma}{\delta} + \sqrt{\left(\frac{\sigma}{\delta}\right)^2 + 2\frac{\sigma}{\delta}} > 1 \qquad \forall \sigma \in (0, \sigma_0].$$

Clearly, if the ratio $\frac{\sigma}{\delta}$ is uniformly separated from zero (say, $\frac{\sigma}{\delta} \geq c > 0$), then the parameter ρ is well separated from 1, i.e., $\rho \geq 1 + c + \sqrt{c^2 + 2c}$. On the other hand, the constant $M_\rho(s)$ in Theorem 4.2 (see (4.3)) can be easily estimated for $s(x, y)$ given in Examples 4.1-4.3. In particular, in case (4.8) we obtain

$$M_{\rho(\sigma)}(s) := \begin{cases} c\,|\log(\sigma_0 - \sigma)^{-1}|, & d = 2 \\ c\,|\sigma_0 - \sigma|^{-1}, & d = 3. \end{cases}$$

The optimal parameter $\sigma = \sigma^*$ can be chosen as the minimiser in (4.12). Indeed, the minimum in (4.12) exists since the corresponding ansatz tends to $+\infty$ as $\sigma \to 0$ or $\sigma \to \sigma_0$. Note that even the sub-optimal choice $\sigma = q\sigma_0$, $q < 1$, leads to a logarithmic bound for the polynomial degree N, i.e. $N = c \log \varepsilon^{-1}$, $\varepsilon > 0$, yielding the interpolation error estimate $|s - \mathbb{I}_{N,x} s| \leq \varepsilon$, for $(x, y) \in B_x \times B_y$.

In case (b), we consider the ellipse $\mathcal{E}_\rho$ whose semi-axis with a length b is parallel to the nearest facet of B_y. By scaling $\delta \to 1$, $\sigma_0 \to \frac{\sigma_0}{\delta}$ and $\sigma \to \frac{\sigma}{\delta}$, we obtain that

$$b = \frac{1}{2}\left(\rho - \frac{1}{\rho}\right) = \frac{\sigma}{\delta},$$

yielding

$$\rho = \frac{\sigma}{\delta} + \sqrt{\left(\frac{\sigma}{\delta}\right)^2 + 1} > 1 \qquad \forall \sigma \in (0, \sigma_0].$$

Again, if $\frac{\sigma}{\delta} \geq c > 0$ then $\rho \geq c + \sqrt{c^2+1} \geq 1 + c + \frac{c^2}{2}$. Now arguments as in case (a) complete the proof. ■

Since the coefficients of our interpolation polynomial $\mathbb{I}_{N,x}s$ depend explicitly on $y \in B_y$, we obtain the following degenerate approximation to $s(x,y)$,

$$s_k = \mathbb{I}_{N,x}s = \sum_{i=1}^{k} \Phi_i(x)\Psi_i(y), \tag{4.13}$$

where $\Phi_{i,B}(x) \in \mathbb{P}_N[B^d]$ are the polynomials in x. The number of terms k in the corresponding polynomial interpolant $\mathbb{I}_{N,x}s$ is $k = O(N^d)$.

Remark 4.7 *Based on the so-called asymptotical smoothness of $s(x,y)$, the generated kernel approximation (cf. (4.13)) can be also derived by the Taylor expansion. Usually, the asymptotical smoothness of a kernel is formulated in the following form: There exists $\gamma \geq 1$, and $g \in \mathbb{R}$ such that for all $x, y \in \mathbb{R}^d$, $x \neq y$, and all multi-indices α, β with $|\alpha| = \alpha_1 + ... + \alpha_d$, there holds*

$$|\partial_x^\alpha \partial_y^\beta s(x,y)| \leq C\, \alpha!\, \beta!\, \gamma^{|\alpha|+|\beta|} |x-y|^{-g-|\alpha|-|\beta|},$$

such that $|\alpha| + |\beta| > 0$. Here the parameter $g \in \mathbb{R}$ is related to the singularity of $s(x,y)$. In our particular case, one has to check the above condition (that contains all partial derivatives of $s(x,y)$) for the fundamental solution $S(x-y)$ (cf. Examples 4.3 - 4.5 above). In this case, the explicit estimate of the constants C and γ appears to be technically involved. Moreover, there are certain cases (e.g., the Helmholtz kernel) where the asymptotical smoothness is not valid, i.e., Taylor-based methods do not work.

4.3 Approximating functions with singularities at the endpoints of $[-1,1]$

Exponentially convergent kernel expansions of the form (4.13) can be also derived in the case of "touching" boxes B_x and B_y, i.e., with $dist(B_x, B_y) = 0$, but $B_x \cap B_y = \emptyset$. In this case one cannot expect the exponential convergence of the polynomial approximations as above since now $\rho = 1$. The idea is to apply the so-called Sinc interpolation.

To fix the idea, we consider a function $g(x,y)$ defined on the reference domain $[-1,1]^2$ which is analytic with respect to $x \in (-1,1)$ possessing a holomorphic extension to the unit disc $\mathcal{U} := \{z \in \mathbb{C} : |z| < 1\}$. Following [26], we say that a function f holomorphic in $\mathcal{U}$ is in $H^2(\mathcal{U})$ if

$$N_2(f,\mathcal{U}) := \left(\lim_{r \nearrow 1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right)^{1/2} < \infty. \tag{4.14}$$

Introduce

$$\phi(z) = \log[(1+z)/(1-z)], \quad \phi'(z) = 2/(1-z^2), \quad z_k = (e^{kh}-1)/(e^{kh}+1), \tag{4.15}$$

where the mesh parameter $h > 0$ will be specified later on. Let

$$S_{k,h}(x) = \frac{\sin[\pi(x - kh)/h]}{\pi(x - kh)/h}$$

be the kth Sinc function with step size h, evaluated at x. The following result is a direct consequence of [26, Example 4.2.9].

Theorem 4.8 *Let ϕ be defined as in (4.15). For any $y \in [-1, 1]$, assume $g(\cdot, y)$ is holomorphic in $\mathcal{U}$ and $\phi' g(\cdot, y) \in H^2(\mathcal{U})$ with a norm N_2 (cf. (4.14)) uniform in $y \in [-1, 1]$. Choose $N \in \mathbb{Z}$ and set $h = \pi\sqrt{1/N}$. Then there exists a constant C_2 independent of y and N, such that the Sinc-approximation using z_k from (4.15) provides the error estimate*

$$\sup_{x\in[-1,1]} \left| g(x, y) - \sum_{k=-N}^{N} g(z_k, y) S_{k,h}(\phi(x)) \right| \leq C_2 N_2(\phi' g(\cdot, y), \mathcal{U}) e^{-\frac{\pi}{2}\sqrt{N}}. \tag{4.16}$$

Remark 4.9 *$g_{2N+1}(x, y) = \sum_{k=-N}^{N} g(z_k, y) S_{k,h}(\phi(x))$ is a particular representation (4.13) leading to an explicit rank-$(2N+1)$ approximation of the corresponding matrix block.*

The above expansion (4.16) can be applied in a situation with the so-called weak admissibility condition (cf. §6.2), where a function g is defined as a restriction of the kernel function $s(x, y)$ of interest onto the weakly admissible "geometric block".

5 Hierarchical matrices: Main ingredients

We are interested in the data-sparse approximation to the fully populated stiffness matrix A_h arising from a FEM/BEM approximation to an integral operator of the type (4.4), (4.5) as well as to the discrete elliptic inverse A_h^{-1} (cf. Example 1).

Due to approximation results in §4.2, we know that on the appropriate (admissible) "geometric block" $B_x \times B_y \subset \Omega \times \Omega$ the singularity function $s(x, y)$ can be approximated by using the separable expansion (4.13) which converges exponentially in the expansion degree k. Due to the singular behaviour of $s(x, y)$ at $x = y$, such a separable approximation cannot be applied globally on $\Omega \times \Omega$. Therefore, we get rid of this difficulty by introducing a hierarchical decomposition of the product computational domain $\Omega \times \Omega$, where on each patch one can approximate $s(x, y)$ by a short sum of separable functions. In turn, this circumstance implies the hierarchical block decomposition of the Galerkin matrix A_h. The corresponding class of hierarchical matrices ($\mathcal{H}$-matrices) was introduced in [13] (see also [14]-[20] and [11] for further developments).

Let I be the index set of unknowns corresponding to the FE-nodal points related to certain Galerkin ansatz space. The construction of square $\mathcal{H}$-matrices defined on the product index set $I \times I$, is based on the following ingredients:

- An $\mathcal{H}$-tree (cluster tree) $T(I)$ of the index set I;
- The admissible partitioning $\mathcal{P}$ of $I \times I$ based on a block cluster tree $T(I \times I)$.
- Low rank representation of all blocks in $\mathcal{P}$.

For the ease of presentation, we shall illustrate the formal descriptions in the case of a tensor-product index set related to the model problem posed in the unit cube $\Omega = [0, 1]^d$. The kernel function $s(x, y)$ is supposed to be analytic in both arguments x and y, unless $x = y$ (see Assumption 4.1). The analysis for so-called asymptotically smooth kernels leads to the similar results (cf. Remark 4.7).

5.1 Cluster tree $T(I)$

We consider different partitionings of I into disjoint subsets including coarse and fine partitionings. The cardinality of I is denoted by $n = \#I$. The set of these partitionings is hierarchically structured and is uniquely defined by the tree $T = T(I)$, which is called the $\mathcal{H}$-tree or cluster tree. By definition, a graph G is a tree if and only if there exists exactly one path between any two vertices.

We use the following entities:
- *vertices* of T or *clusters*,
- *sons* of vertex t, denoted by $S(t)$,
- *root* of T and *leaves* of T. A leaf t of T is characterised by $S(t) = \emptyset$. The set of leaves is denoted by $T^L(I)$.

Definition 5.1 *Given an index set I. A tree T is called an $\mathcal{H}$-tree (cluster tree) if the following conditions hold:*
(i) I is the root of T and $t \subset I$ holds for all $t \in T$.
(ii) If $t \in T$ is no leaf, $S(t)$ contains disjoint subsets of I (sons) and t is the disjoint union of its sons, $t = \bigcup_{s \in S(t)} s$.
(iii) If $t \in T$ is a leaf, then $|t| \equiv \#t \leq C_l = O(1)$, is independent of n. The standard choice is $C_l = 1$.

If $\#S(t) = 2$ for all $t \in T$, t being a leaf, then T is called the binary tree.

Each index $i \in I$ is associated with the basis function ϕ_i of the Galerkin ansatz space $V_h := span\{\phi_i\}_{i \in I}$, so that the support of the basis functions is denoted by

$$X(i) := \operatorname{supp}(\phi_i) \qquad \text{for } i \in I. \tag{5.1}$$

Introducing the notation $X(\tau) := \bigcup_{i \in \tau} X(i)$, $\tau \in T$, we then define the diameter and distance of two clusters

$$diam(\tau) := \max_{x,y \in X(\tau)} ||x - y||, \quad \tau \in T,$$

$$dist(\sigma, \tau) := \min_{x \in X(\tau), y \in X(\sigma)} ||x - y||, \quad \tau, \sigma \in T,$$

where $||\cdot||$ is the Euclidean distance in $\mathbb{R}^d$.

5.2 Example of the balanced cluster tree

For example, we consider the simple model problem for $d = 1$. A Galerkin discretisation leads to the fully populated matrix

$$A = (a_{ij})_{i,j \in I} \qquad \text{with } a_{ij} = \int_0^1 s(x,y)\phi_i(x)\phi_j(y)dxdy, \tag{5.2}$$

where $\{\phi_i\}_{i \in I}$ is the finite-element basis. We may consider two different examples of $\{\phi_i\}_{i \in I}$ based on an equidistant grid $x_\nu = \nu h$ $(\nu = 0, \ldots, N)$ for the step size $h = 1/N$.

Standard examples are the *piecewise constant elements*

$$\phi_i(x) = \left\{ \begin{array}{ll} 1 & \text{if } x \in (x_i, x_{i+1}) \\ 0 & \text{otherwise} \end{array} \right\} \quad \text{for } i \in I := \{0, \ldots, N-1\}, \tag{5.3}$$

as well as the *piecewise linear elements*

$$\phi_i(x) \text{ linear on each interval } (x_{\nu-1}, x_\nu) \text{ and } \phi_i(x_\nu) = \delta_{i\nu} \quad \text{for } i, \nu \in I \tag{5.4}$$

with $I := \{0, \ldots, N\}$. The index set I is different in both cases. For its cardinality $n := \#I$ there holds $n = N$ for (5.3) and $n = N + 1$ for (5.4). To simplify the discussion, we assume that n is a power of 2, $n = 2^L$. Note that in the case of (5.3) the supports are essentially disjoint, whereas in the case of (5.4) $X(i) \cap X(i+1) = [x_i, x_{i+1}]$.

The *level* introduced in the following example may be regarded as the distance of the vertex from the root.

Example 5.2 *Let $I = \{0, ..., n-1\}$ and $n = 2^L$. I is the root. The clusters of level L are the one-element subsets*

$$\tau_1^L = \{0\},\ \tau_2^L = \{1\}, \ldots,\ \tau_n^L = \{n-1\}.$$

On level $L-1$, two subsets from level L are combined:

$$\tau_1^{L-1} = \{0,1\},\ \tau_2^{L-1} = \{2,3\}, \ldots,\ \tau_{n/2}^{L-1} = \{n-2, n-1\}.$$

Similarly, we obtain 4-element subsets τ_i^{L-2} of level $L-2$, etc. Finally, at level 0, the whole index set $\tau_1^0 = I$ is the only cluster. This defines a binary tree $T(I)$ with the vertices ('clusters')

$$\{\tau_i^\ell : 0 \le \ell \le L,\ 1 \le i \le 2^\ell\},$$

where

$$\tau_i^\ell = \left\{(i-1) * 2^{L-\ell}, (i-1) * 2^{L-\ell} + 1, \ldots, i * 2^{L-\ell} - 1\right\}.$$

The vertices at level L are leaves. The sons of τ_i^ℓ $(\ell < L)$ are $\tau_{2i-1}^{\ell+1}$ and $\tau_{2i}^{\ell+1}$.

Clearly, we have

$$X(\tau_i^\ell) = \begin{cases} \left[(i-1) * 2^{L-\ell} h, i * 2^{L-\ell} h\right] \subset [0,1] & \text{for (5.3)}, \\ \left[\left((i-1) * 2^{L-\ell} - 1\right) h, \left(i * 2^{L-\ell} + 1\right) h\right] \cap [0,1] & \text{for (5.4)}. \end{cases} \tag{5.5}$$

The upper index of τ_i^ℓ refers to the level number. This gives rise to the notation

$$T^\ell(I) = \left\{\tau_i^\ell : 1 \le i \le 2^\ell\right\}, \tag{5.6}$$

so that $T^0(I) = I$ and $T^L(I)$ forms the set of leaves.

5.3 Block cluster tree $T(I \times I)$

While the vector components are indexed by $i \in I$, the entries of a (square) matrix have indices from the index set $I \times I$. The block-cluster tree is nothing but the cluster tree for $I \times I$ instead of I, where all vertices ('blocks') $b \in T(I \times I)$ are of the form $b = \tau \times \sigma$ with $\tau, \sigma \in T(I)$.

The construction starts with stating that $I \times I$ is the root of $T(I \times I)$. Then the sons of $b = \tau \times \sigma \in T(I \times I)$ form the set of all blocks $b' := \tau' \times \sigma'$, where τ' (resp. σ') are the sons of τ (resp. σ) provided that these exist.

Construction 5.3 *(cluster tree $T(I \times I)$). Given the cluster tree $T(I)$,*

- *start with $I \times I \in T(I \times I)$ and define the sons of $b = (\tau, \sigma) \in T(I \times I)$ (with $\tau, \sigma \in T$) recursively by:*
- *If $S(\tau) \neq \emptyset$ and $S(\sigma) \neq \emptyset$ then $S(b) := \{\tau' \times \sigma' : \tau' \in S(\tau), \sigma' \in S(\sigma)\}$*
- *If $S(\tau) = \emptyset$ and $S(\sigma) \neq \emptyset$ then $S(b) := \{\tau \times \sigma' : \sigma' \in S(\sigma)\}$*

- *If* $S(\tau) \neq \emptyset$ *and* $S(\sigma) = \emptyset$ *then* $S(b) := \{\tau' \times \sigma : \tau' \in S(\sigma)\}$
- $S(b) = \emptyset$ *if* $S(\tau) = \emptyset$ *and* $S(\sigma) = \emptyset$.

Simple properties of $T(I \times I)$ are listed below:
(a) The depth of the tree $T(I \times I)$ equals to the depth of T;
(b) If all branches of T have the same length k, the third and fourth cases of Construction 5.3 do not occur;
(c) Assume the case (b). If T is a binary tree, then $T(I \times I)$ is a quad-tree.

In the case of the tree $T(I)$ from Example 5.2, the block cluster tree $T(I \times I)$ is

$$T(I \times I) = \left\{\tau_i^\ell \times \tau_j^\ell : 0 \leq \ell \leq L,\ 1 \leq i, j \leq 2^\ell\right\}. \tag{5.7}$$

The leaves of the tree $T(I \times I)$ are the 1×1-blocks $\tau_i^L \times \tau_j^L = \{(i-1, j-1)\}$.

By definition both clusters τ, σ in $b = \tau \times \sigma \in T(I \times I)$ must belong to the same level, so that $T(I \times I)$ decomposes into the sets $T^\ell(I \times I)$ of level $0 \leq \ell \leq L$:

$$T^\ell(I \times I) = \left\{\tau_i^\ell \times \tau_j^\ell : 1 \leq i, j \leq 2^\ell\right\}. \tag{5.8}$$

5.4 Admissible partitioning $\mathcal{P}$ of $I \times I$

The blocks of $T(I \times I)$ are not disjoint. A *partitioning* $\mathcal{P}$ of $I \times I$ is a *disjoint* decomposition of $I \times I$ into blocks b_α, $\alpha \in J$, such that $\bigcup_{\alpha \in J} b_\alpha = I \times I$. Moreover, we require $b_\alpha \in T(I \times I)$ for all $\alpha \in J$, i.e., $\mathcal{P} \subset T(I \times I)$.

In the case $C_L = 1$, the finest partitioning is $\mathcal{P} = \left\{\tau \times \sigma : \tau, \sigma \in T^L(I)\right\} = I \times I$, which corresponds to the format of standard 1×1 matrices. The coarsest partitioning is $\mathcal{P} = \{I \times I\}$ consisting of only one block. Since in the following we shall fill the blocks by low-rank matrices, to guarantee a sufficient approximation, the size of the block must be controlled by the so-called *'admissibility condition'*.

Definition 5.4 *The admissibility condition is some Boolean function*

$$Adm : T(I \times I) \to \{true, false\} \tag{5.9}$$

with the consistency requirement

$$Adm\,(b) \Rightarrow Adm\,(b') \quad \text{for all sons} \quad b' \quad \text{of} \quad b \in T(I \times I)$$

and the property

$$Adm\,(b) = true \quad \text{for all leaves} \quad b \in T(I \times I).$$

A partitioning $\mathcal{P}$ *is called* admissible *if* $Adm\,(b) = true$ *for all* $b \in \mathcal{P}$.

The standard admissibility criteria for $b = \tau \times \sigma$ reads as follows

$$min\{diam(\tau), diam(\sigma)\} \leq \eta\, dist(\tau, \sigma), \tag{5.10}$$

where $\eta > 0$ is some fixed admissibility parameter.

Remark 5.5 *Note that our polynomial approximation by interpolation (see §3-4) does not require any upper bound on $\eta > 0$. Of course, the convergence rate depends on η: the bigger η, the slower is the convergence.*

The admissibility condition (5.9) now takes the form

$$\begin{aligned} &Adm_\eta(b) = true \quad \text{for } b = \tau \times \sigma \in T(I \times I) \quad :\Longleftrightarrow \\ &(b \text{ is a leaf}) \quad \text{or} \quad (5.10) \quad \text{holds.} \end{aligned} \tag{5.11}$$

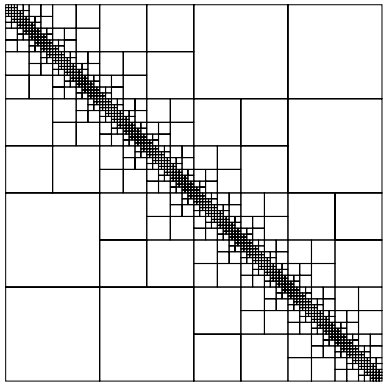

Figure 6: Partitioning by the standard admissibility condition Adm_1

A simple choice of $\eta > 0$ is $\eta = 1$ leading to Adm_1. In the 1D-model case (5.3), we have

$$\operatorname{diam}(\tau_i^\ell) = 2^{L-\ell}h, \qquad \operatorname{dist}(\tau_i^\ell, \tau_j^\ell) = \max\{0, |i-j|-1\}\, 2^{L-\ell}h. \tag{5.12}$$

Hence a block $b = \tau_i^\ell \times \tau_j^\ell$ is admissible if $\ell = L$ (i.e., b is a leaf) or if $|i-j| \geq 2$. The partitioning $\mathcal{P}$ generated by Adm_1 is shown in Figure 6.

5.5 Definition of hierarchical matrices

For an admissible partitioning $\mathcal{P}$ and a natural number k, we define the set $\mathcal{M}_{\mathcal{H},k}(I \times I, \mathcal{P}) \subset \mathbb{R}^{I \times I}$ of (real) hierarchical matrices by

$$\mathcal{M}_{\mathcal{H},k}(I \times I, \mathcal{P}) := \{M \in \mathbb{R}^{I \times I} : \operatorname{rank}(M|_b) \leq k \text{ for all } b \in \mathcal{P}\}. \tag{5.13}$$

Here, $M|_b = (m_{ij})_{(i,j)\in b}$ denotes the matrix block of $M = (m_{ij})_{i,j\in I}$ corresponding to $b \in \mathcal{P}$. The matrices from $\mathcal{M}_{\mathcal{H},k}(I \times I, \mathcal{P})$ are implemented by means of the list $\{M|_b : b \in \mathcal{P}\}$ of matrix blocks, where each $M|_b$ ($b = \tau \times \sigma$ with $\tau, \sigma \in T(I)$) is represented by the rank-k matrix

$$\sum_{\nu=1}^{k} a_\nu b_\nu^\top \quad \text{with vectors} \quad a_\nu \in \mathbb{R}^\tau, b_\nu \in \mathbb{R}^\sigma. \tag{5.14}$$

The number k is called the local rank.

We stress that besides the matrix-vector multiplication, also the matrix-matrix addition and multiplication as well as the matrix inversion can be performed approximately within the $\mathcal{H}$-matrix format.

The basic hierarchical format (5.13) can be optimised and generalised in several directions as follows:

(i) Uniform and $\mathcal{H}^2$-matrices [20];

(ii) blended $\mathcal{H}$-matrix approximation [18];

(iii) coarsening of the hierarchical format using weaker admissibility criteria [19];

(iv) wire-basket approximation for $\mathcal{L}$-harmonic kernels [17];

(v) hierarchical approximation on graded meshes [16];

(vi) generalised (factorized) data-sparse representation for the Green function [23];

Whereas topics (i), (ii) will be considered in Lectures 9, 10, the remaining items are addressed in [13]-[19], [23] where also the comprehensive analysis of the $\mathcal{H}$-matrix techniques is presented (see also [2, 1, 11, 24]). In particular, implementational aspects are well described in [3]. Results on the $\mathcal{H}$-matrix approximation to a class of operator-valued functions with applications to the elliptic, parabolic and hyperbolic problems as well as in control theory are presented in [7]-[10].

6 Complexity of the $\mathcal{H}$-matrix arithmetic

There are two contradicting issues to be satisfied. First, it must be possible to approximate the true full matrix A (e.g., $A = A_h$ from (5.2)) sufficiently well by some $A_{\mathcal{H}} \in \mathcal{M}_{\mathcal{H},k}(I \times I, \mathcal{P})$, i.e.,

$$\|A - A_{\mathcal{H}}\| \leq \varepsilon \tag{6.1}$$

must be reachable for an appropriate $k = \mathcal{O}(|\log^q \varepsilon|)$ with $q = \mathcal{O}(d)$. Second, the related costs should be almost linear in n. The most important costs are $\mathcal{N}_{st}$ (required storage) and $\mathcal{N}_{MV}$ (number of arithmetic operations for the matrix-vector multiplication). In this section, for $\mathcal{P}$ resulting from the standard admissibility condition (5.10) we obtain for both costs

$$\mathcal{N}_{st}, \mathcal{N}_{MV} = \mathcal{O}\left(kn \log n\right). \tag{6.2}$$

In addition, for kernel functions defined by the fundamental solutions we can show that $k = \mathcal{O}(|\log^{d-1} \varepsilon|)$. In the more general situation (say for asymptotically smooth kernels) one can prove $k = \mathcal{O}(|\log^d \varepsilon|)$.

6.1 $\mathcal{R}_k$-matrices

Matrices of the rank k given by (5.14) will be called $\mathcal{R}_k$-matrices, the corresponding class of matrices is denoted by $\mathcal{R}_k$. These matrices form subblocks of the $\mathcal{H}$-matrix corresponding to the admissible block $b = \tau \times \sigma$. Each $R \in \mathcal{R}_k$ can be presented in the form

$$R = A \cdot B^T, \qquad A \in \mathbb{R}^{\tau \times k}, \quad B \in \mathbb{R}^{\sigma \times k}. \tag{6.3}$$

By definition, $rank(R) \leq k$. $\mathcal{R}_k$-matrices possess the following attractive features:

1. Only $k(\#\tau + \#\sigma)$ numbers are required to store an $\mathcal{R}_k$-matrix.

2. The matrix-vector multiplication $x \mapsto y := Rx$, $x \in \mathbb{R}^{\sigma}$ can be done in two steps by $y' := B^T x \in \mathbb{R}^k$, and $y := Ay' \in \mathbb{R}^{\tau}$. The corresponding cost is $2k(\#\sigma + \#\tau)$.

3. The sum of two $\mathcal{R}_k$-matrices $R_1 = A_1 B_1^T$, $R_2 = A_2 B_2^T$ is an $\mathcal{R}_{2k}$-matrix,

$$R_1 + R_2 = [A_1|A_2][B_1|B_2]^T, \quad [A_1|A_2] \in \mathbb{R}^{\tau \times 2k}, \quad [B_1|B_2] \in \mathbb{R}^{\sigma \times k}.$$

4. The multiplication of $R \in \mathcal{R}_k$ by an arbitrary matrix M of the proper size gives again an $\mathcal{R}_k$-matrix:

$$RM = A(M^T B)^T, \qquad MR = (MA)B^T.$$

5. The best approximation of an arbitrary matrix $M \in \mathbb{R}^{\tau \times \sigma}$ by an $\mathcal{R}_k$-matrix M_k (say in the Frobenius norm, that is $\|A\|_F^2 := \sum_{(i,j) \in \tau \times \sigma} a_{ij}^2$) can be calculated by the truncated singular value decomposition (SVD) as follows. Let $M = U \Sigma V^T$ be the

SVD, i.e., U, V are unitary, $\Sigma = diag\{\sigma_1, \ldots, \sigma_n\}$ with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. Set $\Sigma_k := diag\{\sigma_1, \ldots, \sigma_k, 0, \ldots, 0\}$, then $M_k := U\Sigma_k V^T$. The complexity of such an approximation is $\mathcal{O}((\#\tau + \#\sigma)^3)$.

If $M \in \mathcal{R}_m$, then its best approximation $M_k \in \mathcal{R}_k$, $k \leq m$, can be computed by the following

Algorithm 6.1 *Given $M = AB^T \in \mathcal{R}_m$,*

(i) Calculate the (generalised) QR-decompositions $A = Q_A R_A$ and $B = Q_B R_B$ of A and B, respectively, where $Q_A \in \mathbb{R}^{\tau \times m}$, $R_A, R_B \in \mathbb{R}^{m \times m}$, $Q_B \in \mathbb{R}^{\sigma \times m}$.

(ii) Calculate a SVD, $R_A R_B^T = U\Sigma V^T$ (with the cost $O(m^3)$).

(iii) Define $M_k = A_k B_k^T$ with $A_k := Q_A U_k \Sigma_k \in \mathbb{R}^{\tau \times k}$ and $B_k := Q_B V_k \in \mathbb{R}^{\sigma \times k}$, where $U_k := [U_1, \ldots, U_k]$, $V_k := [V_1, \ldots, V_k]$ (in both cases, first k columns) and the truncated matrix Σ_k is defined as above.

The above approximation can be computed in $\mathcal{O}(m^2(\#\tau + \#\sigma) + m^3)$ operations.

6.2 Sparsity constant

In the following, we investigate the sharp estimate to the cost functions $\mathcal{N}_{st}, \mathcal{N}_{MV}$. Instead of the asymptotic description (6.2) we want information about the constant $C_{\mathcal{P}}$ in

$$\max\{\mathcal{N}_{st}, \mathcal{N}_{MV}\} \leq C_{\mathcal{P}} * knL, \tag{6.4}$$

where $L = \log_2(n)$ is the depth of the tree $T(I)$.

The key quantity is the so-called *sparsity constant*

$$C_{\mathrm{sp}}(\mathcal{P}) := \max\left\{ \begin{array}{l} \max_{\tau \in T(I) \setminus T^L(I)} \#\{\sigma \in T(I) : \tau \times \sigma \in \mathcal{P}\}, \\ \max_{\sigma \in T(I) \setminus T^L(I)} \#\{\tau \in T(I) : \tau \times \sigma \in \mathcal{P}\} \end{array} \right\}. \tag{6.5}$$

The number $\#\{\sigma \in T(I) : \tau \times \sigma \in \mathcal{P}\}$ counts how often the cluster τ is used as *row* block, while $\#\{\tau \in T(I) : \tau \times \sigma \in \mathcal{P}\}$ counts how often σ is used as *column* block.

Definition (6.5) excludes $\tau, \sigma \in T^L(I)$, since for the leaves the property $\tau \times \sigma \in \mathcal{P}$ does not require the inequality $\min\{\mathrm{diam}(\sigma), \mathrm{diam}(\tau)\} \leq \eta\, \mathrm{dist}(\sigma, \tau)$. Thus, on the level $\ell = L$, we count C_{sp} by

$$C_{\mathrm{sp}}^L(\mathcal{P}) := \max\left\{ \max_{\tau \in T^L(I)} \#\{\sigma \in T(I) : \tau \times \sigma \in \mathcal{P}\}, \max_{\sigma \in T^L(I)} \#\{\tau \in T(I) : \tau \times \sigma \in \mathcal{P}\} \right\}.$$

A look at Figure 6 shows that the number of σ with $\tau \times \sigma \in \mathcal{P}$ is 2, if we restrict to the upper triangular part; e.g., the two biggest blocks at the upper right corner share the same τ. In general (not for the biggest blocks), there are also up to 2 blocks in the same row belonging to the lower triangular part of the matrix. However, one observes that 2 blocks in the right part correspond to at most 1 block in the left part so that the sum is always ≤ 3. We summarise in

Proposition 6.2 *Consider the 1D-model case (5.3). Then the partitioning $\mathcal{P}$ generated by Adm_1 has the sparsity constant $C_{\rm sp}(\mathcal{P}) = 3$. Moreover, the L-level sparsity constant is $C^L_{\rm sp}(\mathcal{P}) = 6$.*

Proof. Take $\tau = \tau_i^\ell \in S(\tau^{\ell+1})$ with $\ell < L$ and with $\tau^{\ell+1} \in T^{\ell+1}(I)$. Each σ such that $\tau \times \sigma \in \mathcal{P}$ must belong to the same level, i.e., $\sigma = \tau_j^\ell$. We check the various cases for j. Since $\tau^{\ell+1}$ has only two non-admissible neighbouring clusters (in some cases only one), there are only four candidates for $\sigma = \tau_j^\ell$. But at most three of them appear to be admissible for τ_i^ℓ (in some cases only one or two). This proves the first assertion.

To prove the second statement, we note that in addition to the $j's$ found in the previous argument there are three more clusters on level $\ell = L$, which are admissible with τ_i^L (two closest neighbours and τ_i^L itself). ∎

In the limit case $\eta \to \infty$, the admissible partitioning $\mathcal{P}_W$ is shown in Fig. 7. This corresponds to $dist(\tau, \sigma) \to 0$. Clearly, the corresponding sparsity constant is $C_{sp} = 1$, while $C^L_{sp} = 2$. In this case the existence of a low rank approximation to the corresponding matrix blocks in $\mathcal{P}_W$ can be derived using the approximation results in §4.3.

Along with the 1D case (5.3), one can also consider model cases in 2D and 3D (then the simplest trees are no more binary, but quad-trees ($d = 2$) or octrees ($d = 3$), respectively). Due to the corresponding result from [15] for tensor-grids, we have a bound on the sparsity constant by

$$C_{sp}(\mathcal{P}, d, \eta) = (2^d - 1)(1 + 2\sqrt{d}/\eta)^d, \tag{6.6}$$

which yields $C_{sp}(\mathcal{P}, 1, \eta) = 3$, $C_{sp}(\mathcal{P}, 2, \eta) = 27$ and $C_{sp}(\mathcal{P}, 3, \eta) = 189$ for $\eta = 2\sqrt{d}$. Obviously, we recover $C_{\rm sp}(\mathcal{P}) = 3$ from Proposition 6.2. As a matter of fact, the sparsity constant increases significantly (in fact, exponentially) with the dimension d.

For the binary tree in 2D there holds $C_{sp}(\mathcal{P}, 2, \eta) = 24$ (see an example in Fig. 8 corresponding to the tensor-product index set in 2D).

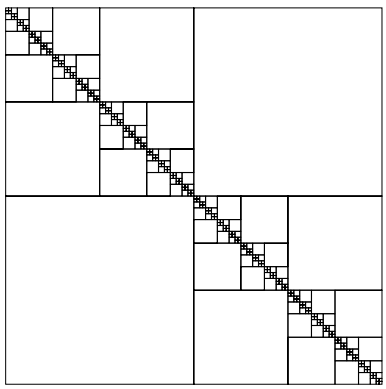

Figure 7: Partitioning $\mathcal{P}_W$ corresponding to the weak admissibility Adm_∞ in 1D.

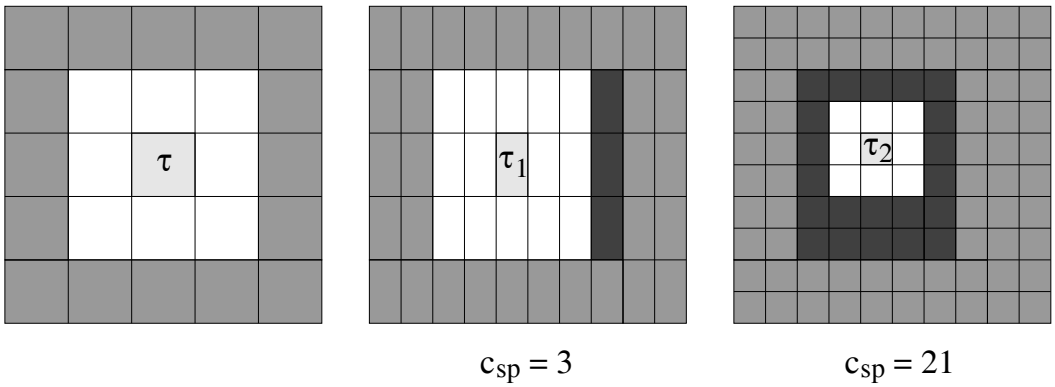

Figure 8: Sparsity constant for standard admissibility for binary tree in 2D.

6.3 Memory requirements and complexity of matrix-by-vector product

The interesting fact is that the storage cost is directly connected with the sparsity constant $C_{sp}(\mathcal{P})$. The following lemma applies to the cases $d = 1$ and $d = 2, 3$ as well.

Lemma 6.3 *Let $n = \#I$. Then there holds:*

(a) The number of blocks is bounded by $nC_{\mathrm{sp}}^{L}(\mathcal{P}) + (n-1)C_{\mathrm{sp}}(\mathcal{P})$.

(b) The storage requirements $\mathcal{N}_{st}$ for an $\mathcal{H}$-matrix $M \in \mathcal{M}_{\mathcal{H},k}(I \times I, \mathcal{P})$ is bounded by

$$\mathcal{N}_{st} \leq \left[2k\,(L-1)\,C_{\mathrm{sp}}(\mathcal{P}) + C_{\mathrm{sp}}^{L}(\mathcal{P})\right] n.$$

Whenever $C_{\mathrm{sp}}^{L}(\mathcal{P}) \leq 2kC_{\mathrm{sp}}(\mathcal{P})$, the storage can be estimated by

$$\mathcal{N}_{st} \leq 2kLC_{\mathrm{sp}}(\mathcal{P})\, n. \tag{6.7}$$

Proof. (a) The tree $T(I)$ has at most $2n-1$ vertices, in particular,

$$\#T(I)\backslash T^{L}(I) = 1 + 2 + \ldots + 2^{L-1} \leq n-1 \quad \text{and} \quad \#T^{L}(I) = n.$$

Therefore,

$$\begin{aligned}\#\mathcal{P} &= \sum_{\tau\times\sigma\in\mathcal{P}} 1 = \sum_{\tau\in T(I)} \#\{\sigma \in T(I) : \tau\times\sigma \in \mathcal{P}\} \\ &\leq \sum_{\tau\in T^{L}(I)} C_{\mathrm{sp}}^{L}(\mathcal{P}) + \sum_{\tau\in T(I)\backslash T^{L}(I)} C_{\mathrm{sp}}(\mathcal{P}) \quad \leq \quad nC_{\mathrm{sp}}^{L}(\mathcal{P}) + (n-1)\,C_{\mathrm{sp}}(\mathcal{P}).\end{aligned}$$

(b) The storage needed for a block $b = \tau\times\sigma$ is $k\,(\#\tau + \#\sigma)$, since k vectors from $\mathbb{R}^{\tau}$ and k vectors from $\mathbb{R}^{\sigma}$ are to be stored. An exception holds for level L, where only one number has to be stored.

In the following, the symbol $\sum^*$ refers to the summation restricted to $level < L$. Furthermore, we exclude the level 0, since $\mathcal{P}$ cannot contain an admissible block of level 0. We have

$$\begin{aligned}\mathcal{N}_{st}^* &= \sum\nolimits^*_{\tau\times\sigma\in\mathcal{P}} k\,(\#\tau+\#\sigma) \le k\left(\sum\nolimits^*_{\tau\times\sigma\in\mathcal{P}}\#\tau+\sum\nolimits^*_{\tau\times\sigma\in\mathcal{P}}\#\sigma\right)\\ &\le C_{\rm sp}(\mathcal{P})\,k\left(\sum\nolimits^*_{\tau\in T(I)}\#\tau+\sum\nolimits^*_{\sigma\in T(I)}\#\sigma\right) \le 2C_{\rm sp}(\mathcal{P})\,k\sum\nolimits^*_{\tau\in T(I)}\#\tau\\ &= 2C_{\rm sp}(\mathcal{P})\,k\sum_{\ell=1}^{L-1}\sum_{\tau\in T^\ell(I)}\#\tau \le 2C_{\rm sp}(\mathcal{P})\,k\,(L-1)\,n,\end{aligned}$$

where the last estimate is due to the equation $\sum_{\tau\in T^\ell(I)}\#\tau = n$ for all ℓ.

For level L, we obtain $\mathcal{N}_{st}^L = \sum_{\tau\times\sigma\in\mathcal{P}}^{level=L} 1 \le C_{\rm sp}^L(\mathcal{P})\,n$, then the result follows. ∎

Lemma 6.4 *$\mathcal{N}_{st} \le \mathcal{N}_{MV} \le 2\mathcal{N}_{st}$, where $\mathcal{N}_{MV}$ is the number of arithmetic operations for the multiplication of a matrix from $\mathcal{M}_{\mathcal{H},k}(I\times I,\mathcal{P})$ by a vector.*

Proof. In the case of a dense matrix, the matrix-vector multiplication requires one multiplication and one addition per matrix element, i.e., the cost for a matrix-vector multiplication is bounded by twice the storage cost. Evidently, the same holds for $\mathcal{R}_k$-matrices. Since an $\mathcal{H}$-matrix consists of either $\mathcal{R}_k$-blocks or full blocks, this concludes the proof. ∎

Given $M\in\mathcal{M}_{\mathcal{H},k}(I\times I,\mathcal{P})$ and $x\in\mathbb{R}^I$. To compute $y := Mx$, we use for each $\tau\times\sigma\in\mathcal{P}$, the exact algorithm for the matrix-by-vector product by an $\mathcal{R}_k$-matrix or by the fully populated matrix block (the latter arises if $C_l > 1$). With fixed τ, the results for different σ will be collected in one component by

$$y|_\tau := y|_\tau + M|_{\tau\times\sigma}x|_\sigma.$$

The final result y is obtained by agglomeration of all $y|_\tau$ within each level of $T(I)$ and then by summation of agglomerated vectors over all the levels of the tree $T(I)$.

6.4 Matrix addition

Given $M_1, M_2 \in \mathcal{M}_{\mathcal{H},k}(I\times I,\mathcal{P})$. The sum $M := M_1 + M_2$ is an $\mathcal{H}$-matrix with block-wise rank $\le 2k$. Using Algorithm 6.1, one can truncate each $\mathcal{R}_{2k}$-block into the corresponding $\mathcal{R}_k$-matrix. This leads to the formatted addition

$$M\mapsto \widetilde{M} := M_1\tilde{+}M_2 \in \mathcal{R}_k.$$

The complexity of the formatted addition is $\mathcal{O}(nk^2\log n)$.

7 $\mathcal{H}$-matrix product and inversion. Approximation issues

We mention only quite briefly the costs $\mathcal{N}_{MM}$ and $\mathcal{N}_{inv}$ for the matrix-matrix multiplication and the inversion. The dependence of these costs on $C_{\rm sp}$ is given by

$$\mathcal{N}_{MM}, \mathcal{N}_{inv} = \mathcal{O}\left(C_{\rm sp}^2 k^2 n \log^2 n\right). \tag{7.1}$$

We consider the simple example corresponding to the standard admissible partitioning $\mathcal{P}_1$ (see Fig. 9,left).

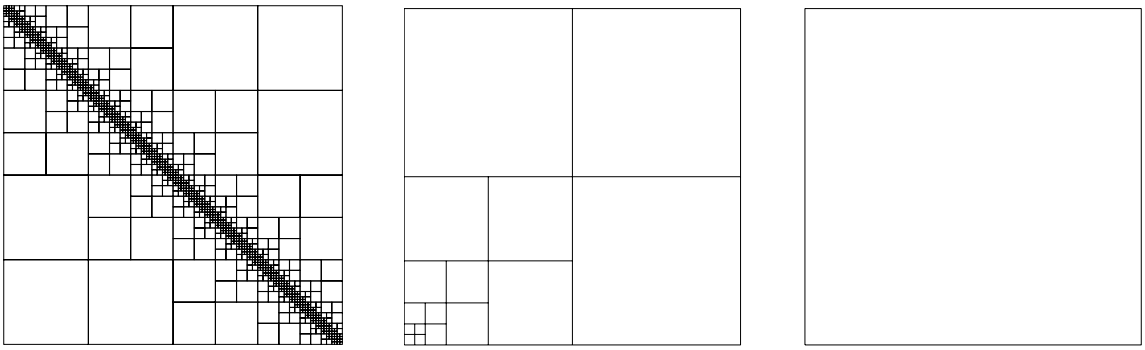

Figure 9: Left: the full $\mathcal{H}$-matrix format. Middle: the block from $\mathcal{P}_1$ on level $\ell = 1$ contained in $b = \tau_1^1 \times \tau_2^1 \in T(I \times I)$. Right: the $\mathcal{R}_k$-block.

7.1 Formatted matrix-matrix multiplication

First we note that the exact product of two $\mathcal{H}$-matrices defined on the index set $\tau \times \tau \in I \times I$ is given by

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}, \tag{7.2}$$

where

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \in \mathcal{M}_{\mathcal{H},k}(\tau \times \tau, \mathcal{P}_1|_{\tau\times\tau}) \tag{7.3}$$

with the block structure inherited from the sons $S(\tau) = \{\tau_1, \tau_2\}$. We can apply the above representation on each level including the coarsest one $\tau = I$. The diagonal blocks A_{ii}, B_{ii} belong to the full $\mathcal{H}$-matrix format $\mathcal{M}_{\mathcal{H},k}$ (further abbreviated by $\mathcal{H}_k$), while the off-diagonal blocks are specified by the simpler formats $\mathcal{W}_k^+$ (see Fig. 9,middle) and its transposed $\mathcal{W}_k^-$ (abbreviate by $\mathcal{W}^\pm$). Recall that the format of rank-k matrices is denoted by $\mathcal{R}_k$. In the following, we shall skip the parameter k in notations as long as it is fixed.

Instead of the exact product (7.2) with the cost $\mathcal{O}(n^3)$, in the $\mathcal{H}$-matrix arithmetics we are interested in the approximate (therefore much less expensive) matrix product

$A \odot B$ which is defined *recursively*. In this way, we rely upon the fast $\mathcal{R}_k$-matrix arithmetic (see §6.1) including the truncation Algorithm 6.1 as well as the formatted addition of two $\mathcal{H}$-matrices. Moreover, we also need the definition of matrix-matrix products with simpler formats, i.e., of the type $\mathcal{H} \times \mathcal{W}$, $\mathcal{W} \times \mathcal{W}$, $\mathcal{H} \times \mathcal{R}$ and $\mathcal{W} \times \mathcal{R}$. We leave the algorithms for the latter two cases as an exercise. The basic ingredients are the multiplication and formatted addition of two $\mathcal{R}$-matrices.

Now the implementation of $\mathcal{W} \times \mathcal{W}$-product can be immediately reduced to the cases $\mathcal{W} \times \mathcal{R}$ and $\mathcal{R} \times \mathcal{R}$.

Algorithm 7.1 *($\mathcal{W} \times \mathcal{W}$-matrix product).*
1. Let $A, B \in \mathcal{M}_{\mathcal{H},k}(I \times I, \mathcal{P}_{\mathcal{W}})$. Again $S(\tau) = \{\tau_1, \tau_2\}$ leads to the block structure (7.3) with $\mathcal{P}_1$ substituted by $\mathcal{P}_{\mathcal{W}}$, where the submatrices A_{ij}, B_{ij} belong to level $\ell = 1$. Therefore, each $A_{ij} \odot B_{ij} \in \mathcal{R}_k$ $(i, j = 1, 2)$ is already defined taking into account that $A_{21}, B_{21} \in \mathcal{W}_k$ while all the remaining blocks belong to $\mathcal{R}_k$-format.

2. We define the final result $A \odot B$ by

$$A \odot B := \mathit{truncation} \circ \begin{bmatrix} A_{11} \odot B_{11} + A_{12} \odot B_{21} & A_{11} \odot B_{12} + A_{12} \odot B_{22} \\ A_{21} \odot B_{11} + A_{22} \odot B_{21} & A_{21} \odot B_{12} + A_{22} \odot B_{22} \end{bmatrix}, \quad (7.4)$$

where each block of truncated matrix belongs to $\mathcal{R}_k$.

Likewise, one can derive the recursive scheme for the product of the type $\mathcal{H} \times \mathcal{W}$.

With above mentioned truncation algorithms at hands, we consider the multiplication scheme in the full $\mathcal{H}$-matrix format.

Algorithm 7.2 *($\mathcal{H} \times \mathcal{H}$-matrix product).*
1. On level $\ell = L$ the product $A \odot B$ is done exactly (since $\#\tau_i^L = 1$, i.e., level L corresponds to 1×1 matrices).

2. Assume that the product is defined for all levels $> \ell$ and let $A, B \in \mathcal{M}_{\mathcal{H},k}(\tau \times \tau, \mathcal{P}_1|_{\tau\times\tau})$ with $\tau \in T^{\ell}(I)$. Again $S(\tau) = \{\tau_1, \tau_2\}$ leads to the block structure $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ and $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$. The submatrices A_{ij}, B_{ij} belong to level $\ell+1$. Therefore, $A_{ij} \odot B_{ij}$ $(i, j = 1, 2)$ are already defined taking into account the appropriate formats of the related blocks (see Fig. 9; actually, there are multiplications of the type $\mathcal{H} \times \mathcal{H}$, $\mathcal{H} \times \mathcal{W}$ or $\mathcal{W} \times \mathcal{W}$, where the latter two are already defined for $\ell \geq 0$). We introduce the intermediate result C by

$$C := \begin{bmatrix} A_{11} \odot B_{11} + A_{12} \odot B_{21} & A_{11} \odot B_{12} + A_{12} \odot B_{22} \\ A_{21} \odot B_{11} + A_{22} \odot B_{21} & A_{21} \odot B_{12} + A_{22} \odot B_{22} \end{bmatrix}. \quad (7.5)$$

3. Next we have to project the diagonal blocks of C to the desired format $\mathcal{M}_{\mathcal{H},k}(\tau_i \times \tau_i, \mathcal{P}_1)$, $i = 1, 2$, while the off-diagonal blocks will be truncated to the target formats $\mathcal{W}_k^{\pm}$. This results in

$$A \odot B := \mathit{truncation} \circ C \in \mathcal{M}_{\mathcal{H},k}(\tau \times \tau, \mathcal{P}_1|_{\tau\times\tau}) \quad (7.6)$$

which completes the recursive definition of $A \odot B$ on level ℓ.

4. Finally, we apply the algorithm on level $\ell = 0$ with $\tau = I$.

Now we address the complexity issues.

7.2 Complexity of the matrix-matrix product

We denote the costs of different truncated operations by $\mathcal{N}_{\mathcal{H}\odot\mathcal{H}}$, $\mathcal{N}_{\mathcal{H}\odot\mathcal{W}}$, $\mathcal{N}_{\mathcal{W}\odot\mathcal{W}}$, $\mathcal{N}_{\mathcal{H}\odot\mathcal{R}}$, $\mathcal{N}_{\mathcal{W}\odot\mathcal{R}}$. We also need the cost $\mathcal{N}_{\mathcal{R}\cdot\mathcal{R}} = O(k^2 n)$, that corresponds to the exact matrix-matrix product.

Theorem 7.3 *Let $n = 2^L$ (without loss of generality), then there holds*

$$\mathcal{N}_{\mathcal{R}\cdot\mathcal{R}} = \mathcal{O}(k^2 n), \tag{7.7}$$

$$\mathcal{N}_{\mathcal{W}\odot\mathcal{R}} = \mathcal{O}(k^2 n), \quad \mathcal{N}_{\mathcal{W}\odot\mathcal{W}} = \mathcal{O}(k^2 n) + \mathcal{O}(k^3 L), \tag{7.8}$$

$$\mathcal{N}_{\mathcal{H}\odot\mathcal{R}} = \mathcal{O}(k^2 L n), \quad \mathcal{N}_{\mathcal{H}\odot\mathcal{W}} = \mathcal{O}(k^2 L n) + \mathcal{O}(k^3 n), \tag{7.9}$$

$$\mathcal{N}_{\mathcal{H}\odot\mathcal{H}} = \mathcal{O}(k^2 L^2 n) + \mathcal{O}(k^3 L n). \tag{7.10}$$

Proof. First we recall the costs $\mathcal{N}_{\mathcal{W}\cdot x} = O(kn)$, $\mathcal{N}_{\mathcal{H}\cdot x} = O(kLn)$. The cost of an exact operation $\mathcal{N}_{\mathcal{R}\cdot\mathcal{R}}$ is easy to check (see §6.1). The result for $\mathcal{N}_{\mathcal{W}\odot\mathcal{R}}$ is based on the representation $C \cdot (AB^T) = (C \cdot A)B^T$. Completely similar argument proves the cost $\mathcal{N}_{\mathcal{H}\odot\mathcal{R}}$.

Furthermore, we prove by induction the cost $\mathcal{N}_{\mathcal{W}\odot\mathcal{W}}$ using the corresponding Algorithm 7.1. Assume that the cost is proven for all levels $> \ell$ and let $A, B \in \mathcal{M}_{\mathcal{H},k}(\tau \times \tau, \mathcal{P}_{\mathcal{W}})$ with $\tau \in T^\ell(I)$. Again $S(\tau) = \{\tau_1, \tau_2\}$ leads to the block structure (7.3), where the submatrices A_{ij}, B_{ij} belong to level $\ell + 1$. Therefore, the costs of $A_{ij} \odot B_{ij}$ $(i, j = 1, 2)$ are already proven taking into account the corresponding block formats. Since the final result $A \odot B$ is given by (7.4), we obtain the recursion

$$\mathcal{N}_{\mathcal{W}\odot\mathcal{W}}(l) = 4\mathcal{N}_{\mathcal{W}\odot\mathcal{R}}(l+1) + 4\mathcal{N}_{\mathcal{R}\odot\mathcal{R}}(l+1) + O(k^3)$$

which leads to (7.8) since the number of levels is $O(L)$. The corresponding recursive relation for $\mathcal{N}_{\mathcal{H}\odot\mathcal{W}}(l)$ reads as

$$\mathcal{N}_{\mathcal{H}\odot\mathcal{W}}(l) = \mathcal{N}_{\mathcal{H}\odot\mathcal{W}}(l+1) + 3\mathcal{N}_{\mathcal{H}\odot\mathcal{R}}(l+1) + 3\mathcal{N}_{\mathcal{W}\odot\mathcal{R}}(l+1) + \mathcal{N}_{\mathcal{W}\odot\mathcal{W}}(l+1) + O(k^3 n_{\ell+1})$$

which leads to the desired bound in (7.9).

Finally, for the cost $\mathcal{N}_{\mathcal{H}\odot\mathcal{H}}$, we arrive at

$$\begin{aligned}\mathcal{N}_{\mathcal{H}\odot\mathcal{H}}(l) \;&= 2\mathcal{N}_{\mathcal{H}\odot\mathcal{H}}(l+1) + 4\mathcal{N}_{\mathcal{H}\odot\mathcal{W}}(l+1) + 2\mathcal{N}_{\mathcal{W}\odot\mathcal{W}}(l+1) + \mathcal{N}_{\mathcal{W}\odot\mathcal{W}}(l+1) \\ &\quad + O(k^2 L n_{\ell+1}) + O(k^3 n_{\ell+1}) + O(k^2 n_{\ell+1})\end{aligned}$$

which then results in (7.10). The proof is complete. ∎

Remark 7.4 *It can be seen that the exact product of two $\mathcal{H}_k$-matrices of the format $\mathcal{P}_1$ yields an $\mathcal{H}$-matrix of the rank at most $O(Lk)$.*

7.3 Matrix inversion by block Gauss elimination

Using the matrix-matrix product algorithms we describe the recursive matrix inversion process which is based on the approximate block Gauss elimination.

Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ on level ℓ. Assume that the block matrix A_{11} is invertible, too. Then the Schur complement $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ is also invertible and the exact inverse of A is given by

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{bmatrix}. \tag{7.11}$$

We are interested in the approximate implementation of (7.11) within the given hierarchical format. Again, the algorithm is defined recursively.

Algorithm 7.5 *($\mathcal{H}$-matrix inversion).*
1. On level $\ell = L$ the approximate inverse $Inv(A)$ is done exactly (since $\#\tau_i^L = 1$, i.e., level L corresponds to 1×1 matrices).

2. Assume that Inv is defined for all levels $> \ell$ and let $A \in \mathcal{M}_{\mathcal{H},k}(\tau \times \tau, \mathcal{P}_1|_{\tau\times\tau})$ with $\tau \in T^\ell(I)$. Again $S(\tau) = \{\tau_1, \tau_2\}$ leads to the block structure $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. The submatrices A_{ij} belong to level $\ell + 1$. Therefore, $X_{11} := Inv(A_{11})$ is already defined provided that the inverse exists.

3. Taking into account the appropriate formats of the related blocks (see Fig. 9; actually, there are multiplications of the type $\mathcal{H} \times \mathcal{H}$, $\mathcal{H} \times \mathcal{W}$ or $\mathcal{W} \times \mathcal{W}$, where the latter two are already defined for $\ell \geq 0$), the Schur complement is approximated by

$$\widetilde{S} = A_{22} - A_{21} \odot X_{11} \odot A_{12}.$$

We define the intermediate result C by

$$C = \begin{bmatrix} X_{11} + X_{11}A_{12}Inv(\widetilde{S})A_{21}X_{11} & -X_{11}A_{12}Inv(\widetilde{S}) \\ -Inv(\widetilde{S})A_{21}X_{11} & Inv(\widetilde{S}) \end{bmatrix}. \tag{7.12}$$

4. Next we have to substitute the exact products and addition by formatted operations and then project the diagonal blocks of C to the desired format $\mathcal{M}_{\mathcal{H},k}(\tau_i \times \tau_i, \mathcal{P}_1)$, $i = 1, 2$, while the off-diagonal blocks will be truncated to the target formats $\mathcal{W}_k^{\pm}$. This results in

$$Inv(A) := truncation \circ C \in \mathcal{M}_{\mathcal{H},k}(\tau \times \tau, \mathcal{P}_1|_{\tau\times\tau}) \tag{7.13}$$

which completes the recursive definition of $Inv(A)$ on level ℓ.

4. Finally, we apply the algorithm on level $\ell = 0$ with $\tau = I$.

Now we address the complexity issues.

Theorem 7.6 *Let $n = 2^L$, then there holds*

$$\mathcal{N}_{\mathcal{H}\odot\mathcal{H}} = \mathcal{O}(k^2L^2n) + \mathcal{O}(k^3Ln). \tag{7.14}$$

Proof. Based on the results for the $\mathcal{H}$-matrix product, the arguments are quite similar to those in the proof of Theorem 7.3. ∎

7.4 Matrix inversion method by Newton's iteration

Here we discuss an alternative algorithm to compute the inverse of an $\mathcal{H}$-matrix based on the iterative correction which only includes a formatted matrix-matrix multiplications. In general, the proper initial guess may be obtained, for example, by the approximate block Gauss elimination (cf. §7.3) with small local rank $k = O(1)$. If we deal with an approximate elliptic inverse the initial guess can be also constructed using the well developed preconditioning techniques for the discrete elliptic systems. It is important that the preconditioner should be presented in the $\mathcal{H}$-matrix format.

Assume that $A \in \mathbb{R}^{n\times n}$ is invertible. Given $X_0 \in \mathbb{R}^{n\times n}$ define the iteration (e.g., see [22])

$$X_{i+1} = X_i(2I - AX_i), \qquad i = 1, 2, \ldots. \tag{7.15}$$

Denote the error in this approximation by $E_i = I - AX_i$, $i = 0, 1, 2, \ldots$. It is easy to see that

$$X_{i+1} = X_i(I + E_i), \qquad i = 0, 1, 2, \ldots,$$

which implies

$$E_i = I - AX_{i-1}(I + E_{i-1}) = I - (I - E_{i-1})(I + E_{i-1}) = E_{i-1}^2, \quad i = 1, 2, \ldots. \tag{7.16}$$

Applying (7.16) recursively, we find that

$$E_i = E_0^{2^i}, \qquad i = 1, 2, \ldots. \tag{7.17}$$

It is also clear that

$$A^{-1} - X_i = A^{-1}E_i = A^{-1}E_0^{2^i} = X_0(I - E_0)^{-1}E_0^{2^i}.$$

Under the following assumption on the spectral radius of E_0,

$$\rho \equiv \rho[E_0] = \max_j |\lambda_j| < 1,$$

where $\lambda_j = \lambda_j(E_0)$ are the eigenvalues of E_0, we obtain that the error E_i in (7.17) vanishes like ρ^{2^i}.

Remark 7.7 *Note that the iteration (7.15) can be applied to any preconditioned matrix $B = R_0A$, where R_0 is a spectrally equivalent preconditioner to A so that $sp(B)$ is uniformly bounded with respect to n. Assuming that both R_0 and R_0A already have the $\mathcal{H}$-matrix representation, we then obtain the approximate inverse of interest from $A^{-1} = (R_0A)^{-1}R_0$. In some cases this approach provides the constructive proof for the existence of an approximate $\mathcal{H}$-matrix inverse.*

Let $E_0 = I - BX_0$. The requirement $\rho[E_0] < 1$ can be achieved under the following conditions.

Lemma 7.8 *Let B have real eigenvalues in the interval $0 < m \leq \lambda_j \leq M$, $j = 1, 2, \ldots, n$. Let $X_0(w) = wI$, then $\rho[E_0] < 1$ for all $w \in (0, \frac{2}{M})$. Moreover, if $\rho(w) = \rho[E_0(w)]$, then there holds*

$$\rho(w_*) = \min_{w \in (0, \frac{2}{M})} \rho(w) = \frac{M - m}{M + m}, \qquad w_* = \frac{2}{M + m}. \tag{7.18}$$

Proof. This lemma is a reformulation of a standard convergence result for the Richardson iteration. ∎

Implementing (7.15) in the formatted $\mathcal{H}$-matrix arithmetics one can compute the $\mathcal{H}$-matrix approximation X_i to A^{-1} with $O(\log \log \varepsilon^{-1})$ iterations, where

$$\|I - AX_i\| \leq \varepsilon.$$

8 $\mathcal{H}$-matrix approximation of integral operators

8.1 Approximation of matrix blocks associated with $\mathcal{P}$-partitioning

Given an integral operator $\mathcal{A}$ defined by the kernel function $s(x,y)$ (cf. §1.2), let $\mathbf{A} := \langle \mathcal{A}\varphi_i, \varphi_j \rangle_{i,j=1}^N$ be the corresponding exact Galerkin matrix. Let $\mathcal{P}$ be the standard admissible partitioning (see §5.4). For $\tau \times \sigma \in \mathcal{P}$, denote by B_τ, B_σ the minimal bounding boxes for $X(\tau), X(\sigma)$, respectively. These boxes have the form (4.9) such that $X(\tau) \in B_\tau$, $X(\sigma) \in B_\sigma$. We shall consider a simple approach and approximate by the low rank matrices only those blocks $b = (\tau, \sigma)$ in $\mathbf{A}$, which have non-intersecting bounding boxes. To that end, introduce the splitting $\mathcal{P} = P_{far} \cup P_{near}$, where

$$P_{far} := \{\tau \times \sigma \in \mathcal{P} : \operatorname{dist}(B_\tau, B_\sigma) > 0\}.$$

The reduction in operation count is due to the replacement of the full matrix blocks $A^{\tau\times\sigma}$, $\tau \times \sigma \in P_{far}$, by their accurate low-rank approximations

$$A_{\mathcal{H}}^{\tau\times\sigma} := \sum_{\alpha=1}^{k} \mathbf{a}_\alpha^\tau \cdot (\mathbf{c}_\alpha^\sigma)^\top,$$

where

$$\mathbf{a}_\alpha^\tau := \{\int_{X(\tau)} a_\alpha(x)\varphi_j dx\}_{j\in\tau} \in \mathbb{R}^\tau, \quad \mathbf{c}_\alpha^\sigma := \{\int_{X(\sigma)} c_\alpha(y)\varphi_j dy\}_{j\in\sigma} \in \mathbb{R}^\sigma.$$

To provide a small perturbation $A^{\tau\times\sigma} - A_{\mathcal{H}}^{\tau\times\sigma}$, the vectors $\mathbf{a}_\alpha$, $\mathbf{c}_\alpha$ are obtained from a *separable expansion*

$$s_{\tau,\sigma}(x,y) = \sum_{\alpha=1}^{k} a_\alpha(x)c_\alpha(y), \quad (x,y) \in X(\tau) \times X(\sigma) \tag{8.1}$$

with $k \ll N = \dim V_h$ approximating the kernel $s(x,y)$. In general, a_α (resp. c_α) depends on τ (resp. σ). In our particular case, analytical properties of the fundamental solution $S(x-y)$ (see Lecture 4) ensure the existence of accurate kernel approximations (on admissible geometrical blocks) providing the estimate

$$\max_{(x,y)\in X(\tau)\times X(\sigma)} |s(x,y) - s_{\tau,\sigma}(x,y)| \lesssim M_\rho[s]\rho^{-m} \qquad \text{for all } \tau \times \sigma \in P_{far} \tag{8.2}$$

with

$$\rho = \rho(\tau,\sigma) = 1 + \frac{\rho_0 dist(B_\tau, B_\sigma)}{\delta} > 1, \quad 0 < \rho_0 \le 1,$$

where in local coordinates $B_\tau = [-\delta,\delta]^d$, $B_\sigma = [-\delta,\delta]^d$. More precise estimate on $\rho(\tau,\sigma)$ depends on the location of B_τ, B_σ (cf. Lemma 4.6).

The degree m is related to the local rank by

$$k = O(m^q) \quad \text{with} \quad q = d.$$

However, in the case of $\mathcal{L}$-harmonic kernels (e.g., for $s(x,y) = S(x-y)$) the bound on k can be improved by $q = d-1$ using the so-called *wire-basket expansions* [17].

In the above estimate for $\rho > 1$, we use parameters which can be derived from the standard admissibility condition for the bounding boxes B_τ, B_σ of the size $[2\delta]^d$. Here it is worth to note that our approximation theory implies the exponential convergence in (8.2) provided that $dist(B_\tau, B_\sigma) > 0$. The uniform convergence rate now requires a uniform lower bound in

$$\frac{\rho_0 dist(B_\tau, B_\sigma)}{\delta} > C > 0 \qquad \text{for} \quad (\tau,\sigma) \in P_{far}. \tag{8.3}$$

For the one-dimensional example in §5.2 the corresponding bounding boxes coincide with $X(\tau)$ and $X(\sigma)$, respectively. With fixed $\rho_0 < 1$ in (8.3), the constant $M_\rho[s]$, which characterises the singularity degree of the kernel at $x = y$, can be bounded by

$$M_\rho[s] = \max_{x,y:\ |x-y| \geq C_1\delta} |s(x,y)|$$

with some fixed $C_1 > 0$ (cf. (8.3)). The optimal choice of ρ and $M_\rho[s]$ is based on the solution of the minimisation problem in (4.12).

8.2 Global consistency error

To simplify the exposition, we choose the FE space V_h as the space of piecewise constant functions. For each $u = \sum_{i\in I} u_i\varphi_i \in V_h$ and $\tau \in T(I)$, define $u|_\tau := \sum_{i\in\tau} u_i\varphi_i$. Introduce the bilinear form $a_{\mathcal{H}}(\cdot,\cdot)$ associated with the modified kernel $s_{\tau,\sigma}$, and defined as follows

$$a_{\mathcal{H}}(u,v) := \sum_{\tau\times\sigma\in P_{near}} \int_{X(\tau)\times X(\sigma)} u(x)|_\tau s(x,y) v(y)|_\sigma dxdy \tag{8.4}$$

$$+ \sum_{\tau\times\sigma\in P_{far}} \int_{X(\tau)\times X(\sigma)} u(x)|_\tau s_{\tau,\sigma}(x,y) v(y)|_\sigma dxdy \quad \forall u,v \in V_h,$$

where we collect contributions from the "modified kernel" $s_{\tau,\sigma}$ on each admissible block $\tau\times\sigma \in P_{far}$. Recall that $\mathcal{P}$ is the disjoint partitioning of $I\times I$, therefore $a_{\mathcal{H}}$ is well defined. Note that such an approach to define the perturbed bilinear form (due to a modification of kernel) can be applied not only in the case of piecewise constant finite elements, but also to rather general elements with overlapping supports.

By definition, the approximate $\mathcal{H}$-matrix $A_{\mathcal{H}}$ is nothing but the generalised Galerkin matrix associated with the bilinear form $a_{\mathcal{H}}$. The perturbation of the matrix $A = A_h$ by $A_{\mathcal{H}} - A$ has an exponential decay with respect to the parameter $m = O(L)$. The following lemma is not restricted to the case of piecewise constant elements.

Lemma 8.1 *Given the integral operator $\mathcal{A}$ and the associated bilinear form $a(\cdot,\cdot)$, where the kernel function can be approximated by (8.2). To control the constant M_ρ in (8.2), we assume that (cf. Remark 4.7)*

$$|s(x,y)| \leq C|x-y|^{-g}, \quad g < d. \tag{8.5}$$

Let for each admissible block in P_{far} there exists an approximation of the type (8.1), (8.2). Then the modified bilinear form $a_{\mathcal{H}}$ (cf. (8.4)) based on the block partitioning $\mathcal{P}(I \times I)$ yields

$$|a(u,v) - a_{\mathcal{H}}(u,v)| \le C_0 \rho^{-m} ||u||_0 ||v||_0 \qquad \forall u, v \in V_h. \tag{8.6}$$

The local rank k of each block in the corresponding $\mathcal{H}$-matrix $A_{\mathcal{H}}$ does not exceed $O(m^d)$, while the complexity of $A_{\mathcal{H}}$ is bounded by $O(kLN)$.

Proof. Estimate (8.6) is derived using block-by-block consideration on the base of local approximations (8.1) with (8.2). By definition, we have on each target block $\tau \times \sigma \in \mathcal{P}(I \times I)$,

$$a(u|_\tau, v|_\sigma) - a_{\mathcal{H}}(u|_\tau, v|_\sigma) = \int\limits_{X(\tau)\times X(\sigma)} u(x)|_\tau (s(x,y) - s_{\tau,\sigma}(x,y)) v(y)|_\sigma dxdy \tag{8.7}$$

with $s_{\tau,\sigma}(x,y)$ defined above. On each admissible block in P_{far} one can impose the short sum approximation to the kernel (8.2). Note that for quasi-uniform meshes and for clusters on level ℓ, there holds

$$\max_{\tau \in T^{(\ell)}(I)} |\tau| = O(2^{-d\ell}) \quad \text{and} \quad diam(B_\tau) = O(2^{-\ell}),$$

where $|\tau| := measX(\tau)$. Moreover, due to (8.5) there holds $M_\rho[s] = O(2^{g\ell})$ in (8.2).

Now we proceed with the error estimate for all $u, v \in V_h$ (without abuse of notations, we use the shortening $\|u\|_{0,\tau} = \|u\|_{0,X(\tau)}$),

$$\begin{aligned}
|a(u,v) - a_{\mathcal{H}}(u,v)| &= \left| \sum_{\ell=0}^{L} \sum_{\tau\times\sigma\in\mathcal{P}^{(\ell)}} (a(u|_\tau, v|_\sigma) - a_{\mathcal{H}}(u|_\tau, v|_\sigma)) \right| \\
&\lesssim \rho^{-m} \sum_{\ell=1}^{L-1} \sum_{\tau\times\sigma\in\mathcal{P}^{(\ell)}} 2^{g\ell} \sqrt{|\tau|\,|\sigma|} ||u|_\tau||_{0,\tau} ||v|_\sigma||_{0,\sigma} \\
&\lesssim \rho^{-m} \sum_{\ell=1}^{L-1} 2^{g\ell} \max_{\tau\times\sigma\in\mathcal{P}^{(\ell)}} \sqrt{|\tau||\sigma|} \sqrt{\sum_{\tau\times\sigma\in\mathcal{P}^{(\ell)}} ||u|_\tau||^2_{0,\tau}} \sqrt{\sum_{\tau\times\sigma\in\mathcal{P}^{(\ell)}} ||v|_\sigma||^2_{0,\sigma}} \\
&\lesssim \rho^{-m} \sum_{\ell=1}^{L-1} 2^{-d\ell+g\ell}\, a_\tau\, a_\sigma \\
&\lesssim C_{sp}(\mathcal{P}) \rho^{-m} ||u||_0\, ||v||_0,
\end{aligned}$$

with $a_\tau = \sqrt{\sum\limits_{\tau\in T^{(\ell)}(I)} ||u|_\tau||^2_{0,\tau} \sum\limits_{\sigma:\tau\times\sigma\in\mathcal{P}^{(\ell)}} 1}$, $a_\sigma = \sqrt{\sum\limits_{\sigma\in T^{(\ell)}(I)} ||v|_\sigma||^2_{0,\sigma} \sum\limits_{\tau:\tau\times\sigma\in\mathcal{P}^{(\ell)}} 1}$. Here we use the alternative equivalent definition for the sparsity constant

$$C_{sp}(\mathcal{P}(I\times I)) := \max_\ell \{ \sum_{\tau:\tau\times\sigma\in\mathcal{P}^{(\ell)}} 1, \sum_{\sigma:\tau\times\sigma\in\mathcal{P}^{(\ell)}} 1 \}.$$

The last line in the above estimate is due to stability of the L^2-norm with respect to the direct additive splitting of V_h by $u = \sum\limits_{\tau\in T^{(\ell)}} u|_\tau$, $u \in V_h$ which holds for each $\ell = 0, ..., L$. The proof is complete. ∎

8.3 Some corollaries

Corollary 8.2 *Let $V_h \in V \subset L^2(\Omega)$ be associated with a quasi-uniform mesh. Assume that the approximation of $s(x,y)$ by the separable expansion (8.1) satisfies (8.2). Then, there exists a constant $C > 0$ independent of h, such that for $\alpha \in [0,1]$,*

$$|a(u,v) - a_{\mathcal{H}}(u,v)| \lesssim h^{2(\alpha-1)} \rho^{-m} ||u||_{\alpha-1} ||v||_{\alpha-1} \qquad \forall u, v \in V_h, \tag{8.8}$$

where $a_{\mathcal{H}}$ means the bilinear form corresponding to the modified kernel (cf. (8.4)). Moreover, for $\alpha \in [1,2]$ and for the space of piecewise linear elements there holds

$$|a(u,v) - a_{\mathcal{H}}(u,v)| \lesssim h^{-g} \rho^{-m} ||u||_{\alpha-1} ||v||_{\alpha-1} \qquad \forall u, v \in V_h. \tag{8.9}$$

Proof. From Lemma 8.1 we know the estimate (8.8) for $\alpha = 0$. The consistency error bound (8.8) for $\alpha < 0$ then follows from the case $\alpha = 0$ by applying the inverse estimate

$$\|v\|_{L^2(\Omega)} \le C h^{\alpha-1} \|v\|_{H^{\alpha-1}(\Omega)} \quad \forall v \in V_h$$

with a constant C independent of h. In the case $\alpha > 1$, we just use the estimate $|M_\rho[s]| \le C h^{-g}$. ∎

Let $V_h = \operatorname{span}\{\varphi_i\}_{i \in I}$, then $A_{\mathcal{H}}$ is nothing but the exact Galerkin matrix associated with $a_{\mathcal{H}}$. Denote by $\mathcal{J} : \mathbb{R}^N \to V_h$ the natural bijection. Due to the above observation, Corollary 8.2 immediately implies that the $\mathcal{H}$-matrix approximation $A_{\mathcal{H}} \in \mathcal{M}_{\mathcal{H},k}(I \times I, \mathcal{P})$ provides the following error estimates

Corollary 8.3 *There holds*

$$\|A_h - A_{\mathcal{H}}\|_F \le C N^{-1} \rho^{-m} \le C \|A_h\|_F \rho^{-m}. \tag{8.10}$$

For the consistency error we have the bound

$$\langle (A_h - A_{\mathcal{H}}) U, V \rangle \le C N \rho^{-m} \|U\|_2 \|V\|_2 \quad \forall U, V \in \mathbb{R}^N.$$

Proof. The results follow by (8.6) due to the following norm- and matrix-norm estimates

$$\sqrt{N} \|\mathcal{J} U\|_{L^2(\Omega)} \le C \|U\|_2, \quad \|A\|_2 \le \|A\|_F \le \sqrt{N} \|A\|_2$$

with constant C independent of N. ∎

8.4 Error bound for the generalised Galerkin method

The error analysis for the generalised Galerkin method with the bilinear form $a_{\mathcal{H}}(\cdot,\cdot)$ now follows by the first Strang lemma (cf. Theorem 1.5). For ease of presentation, we assume that $V_h \subset V = L^2(\Omega)$.

Lemma 8.4 *Let $u \in V$ be the solution of (1.1), where $a(\cdot,\cdot)$ is V-elliptic with the ellipticity constant $\alpha > 0$. Under assumptions of Lemma 8.1 let*

$$\alpha^* := \alpha - C_0 \rho^{-m} > 0,$$

where C_0 is defined in (8.6). Then $a_{\mathcal{H}}(\cdot,\cdot)$ is uniformly V-elliptic over $V_h \times V_h$, i.e.,

$$A_{\mathcal{H}}(v_h, v_h) \geq \alpha^* \|v_h\|_V^2 \qquad \forall v_h \in V_h.$$

Then there exists a unique solution $u_{\mathcal{H}} \in V_h$ of the Galerkin equation

$$a_{\mathcal{H}}(u_{\mathcal{H}}, v) = (f, v)_0 \qquad \forall v \in V_h \tag{8.11}$$

such that

$$\|u - u_{\mathcal{H}}\|_V \leq \frac{C C_0 \rho^{-m}}{\alpha^*} \|u\|_V + \left(1 + \frac{C_a}{\alpha^*}\right) \inf_{w \in V_h} \|u - w\|_V,$$

where $C > 0$ does not depend on u and h.

Proof. The first assertion follows from (8.6), since

$$a_{\mathcal{H}}(v, v) = a(v, v) + a_{\mathcal{H}}(v, v) - a(v, v) \geq (\alpha - C_0 \rho^{-m}) \|v\|_V^2.$$

Now we apply first Strang lemma due to the following estimate

$$\inf_{w \in V_h} \left[\left(1 + \frac{C_a}{\alpha^*}\right) \|u - w\|_V + \frac{C_0}{\alpha^*} \rho^{-m} \|w\|_V \right]$$

$$\leq \frac{C_0}{\alpha^*} \rho^{-m} \|w^*\|_V + \left(1 + \frac{C_a}{\alpha^*}\right) \inf_{w \in V_h} \|u - w\|_V,$$

where w^* is the corresponding minimiser in the minimisation problem involved. Relying on the approximation property of V_h, the triangle inequality implies the desired stability $\|w^*\| \leq C \|u\|_V$, which completes the proof. ∎

Remark 8.5 *(Refined meshes). The results above are valid for the quasi-uniform meshes. However, the $\mathcal{H}$-matrix approximation can be successfully applied on graded meshes. The corresponding construction of the cluster tree $T(I)$ is based on a modified separation criteria. One can use either the so-called cardinality-balanced partitioning (that optimises the approximation power) or the distance-balancing strategy (that optimises the complexity). Let $d = 1$. In the first case, the two sons of τ are obtained to provide $\#\tau_1 = \#\tau_2$ (possibly, approximately). In the second case, one can use a ternary cluster tree with $N = 3^L$. Then the triple of sons τ_1, τ_2, τ_3 satisfies the separation criteria (a) $\#\tau_1 = \#\tau_3$; (b) $diam(\tau_1) = diam(\tau_2)$. For both strategies it is possible to prove the optimal approximation and complexity results for different kinds of mesh refinement (see [16] for more details).*

9 Uniform and $\mathcal{H}^2$-matrices

In the present lecture we discuss further possible improvements of hierarchical formats that lead to $O(N)$ complexities.

9.1 Uniform $\mathcal{H}$-matrices

In the 1D case, consider an integral operator with the kernel function $s(x,y)$, $(x,y) \in [0,1]^2$. Let A be the exact Galerkin matrix with respect to the space $V_h = \{\varphi_i\}_{i=1}^N$ of (say) piecewise constant finite elements on a quasi-uniform mesh. Define $T(I)$, $\mathcal{P}(I \times I)$ and $\mathcal{P}_{far}$ as above. Recall that an $\mathcal{R}_k$-approximation to the matrix block $A^{\tau\times\sigma}$, $\tau \times \sigma \in \mathcal{P}_{far}$, is based on the polynomial interpolation (cf. Remark 3.4)

$$s_{\tau,\sigma}(x,y) := \sum_{\alpha=1}^{k} s(\xi_\alpha, y) l_\alpha(x), \qquad (x,y) \in X(\tau) \times X(\sigma), \tag{9.1}$$

where $l_\alpha(x) = l^\tau_{\alpha,k-1}(x)$, $\alpha = 1,\dots,k$, are the corresponding Lagrange polynomials of degree $\le k-1$. This leads to the approximate matrix block

$$A^{\tau\times\sigma}_{\mathcal{H}} = AB^T \in \mathcal{R}_k \quad \text{with} \quad A = \{A_{i\alpha}\} \in \mathbb{R}^{\tau\times k},\ B = \{B_{j\alpha}\} \in \mathbb{R}^{\tau\times k},$$

$$A_{i\alpha} := \int_{X(\tau)} l^\tau_\alpha(x)\varphi_i(x)dx, \qquad B_{j\alpha} := \int_{X(\sigma)} s(\xi_\alpha, y)\varphi_j(y)dy.$$

Note that the matrix A does not depend on the kernel function $s(x,y)$ and thus it can be precomputed and stored *a priori*. If we use the stronger admissibility condition by substituting "max" in (5.10) instead of "min", then one can interpolate in both variables providing the tensor-product polynomial approximant (cf. §4)

$$s_{\tau,\sigma}(x,y) = \mathbb{I}_k s := \sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} s(\xi^\tau_\alpha, \xi^\sigma_\beta) l^\tau_\alpha(x) l^\sigma_\beta(y). \tag{9.2}$$

Representation (9.2) now leads to the following formulae for the matrix entries of $A^{\tau\times\sigma}_{\mathcal{H}} = \{a_{ij}\}$,

$$a_{ij} := \int_{X(\tau)\times X(\sigma)} s_{\tau,\sigma}(x,y)\varphi_i(x)\varphi_j(y)dxdy = \sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} A_{i\alpha} s(\xi^\tau_\alpha, \xi^\sigma_\beta) B_{j\beta},$$

or in matrix form

$$A^{\tau\times\sigma}_{\mathcal{H}} = A\, K^{\tau\times\sigma} B^T, \qquad A \in \mathbb{R}^{\tau\times k},\ B \in \mathbb{R}^{\sigma\times k},\ K^{\tau\times\sigma} \in \mathbb{R}^{k\times k} \tag{9.3}$$

with

$$(K^{\tau\times\sigma})_{\alpha,\beta} := s(\xi^\tau_\alpha, \xi^\sigma_\beta),$$

A as above and $B_{j\beta} = \int_{X(\sigma)} l^\sigma_\beta(y)\varphi_j(y)dy$.

Now both matrices A and B in (9.3) are independent of the kernel function $s(x,y)$, while the small $k \times k$ coefficients matrix $K^{\tau\times\sigma}$ represents the kernel function at the interpolation points. Let $A = [a_1, \ldots, a_k]$, $B = [b_1, \ldots, b_k]$ (column-wise representation) with $a_\alpha \in \mathbb{R}^\tau$, $b_\beta \in \mathbb{R}^\sigma$. Then with any given coefficients matrix $K^{\tau\times\sigma}$, we obtain

$$A_{\mathcal{H}}^{\tau\times\sigma} \in \mathcal{U}_{\tau\sigma} = \mathcal{U}_\tau \times \mathcal{U}_\sigma, \qquad \mathcal{U}_\tau = span\{a_\alpha\}_{\alpha=1}^k,\ \mathcal{U}_\sigma = span\{b_\beta\}_{\beta=1}^k,$$

i.e. the corresponding $\mathcal{R}_k$-matrices are elements of the tensor-product space of matrices $\mathcal{U}_{\tau\sigma}$. Since all the vectors a_α, b_β are predefined, only the coefficients $(K^{\tau\times\sigma})_{\alpha,\beta}$ are to be stored.

Definition 9.1 *(Uniform $\mathcal{H}$-matrix). Given the cluster tree $T(I)$, partitioning $\mathcal{P}$ of $I \times I$ and a family of matrices $\mathcal{U} = (\mathcal{U}_\tau)_{\tau\in T(I)}$ with $\mathcal{U}_\tau \in \mathbb{R}^{\tau\times k}$, $k \in \mathbb{N}$. A matrix $A \in \mathcal{M}_{\mathcal{H},k}$ is called a uniform $\mathcal{H}$-matrix with respect to the generating family $\mathcal{U}$, if for each block $A^{\tau\times\sigma}$, $\tau \times \sigma \in \mathcal{P}_{far}$, there holds*

$$A^{\tau\times\sigma} = U_\tau K^{\tau\times\sigma} U_\sigma^T, \qquad U_\tau,\ U_\sigma \in \mathcal{U}, \tag{9.4}$$

with some $K^{\tau\times\sigma} \in \mathbb{R}^{k\times k}$. We abbreviate this class of matrices as $\mathcal{U}_{\mathcal{H},k}$.

Contrary to the case of general $\mathcal{H}$-matrices, the uniform $\mathcal{H}$-matrices form a subspace of $\mathbb{R}^{I\times I}$. Therefore, the addition of two $\mathcal{U}_{\mathcal{H},k}$-matrices is exact.

Since the number of blocks in $\mathcal{P}$ is bounded by $C_{sp}(\mathcal{P})N$ (we do not count 1×1 blocks), we require about $C_{sp}k^2N$ units of memory to store an $\mathcal{U}_{\mathcal{H},k}$-matrix (cf. $C_{sp}kLN$ in the case of $\mathcal{M}_{\mathcal{H},k}$-matrices).

If we use the variable rank $k = k(\ell)$, which depends only on the level number ℓ by $k(\ell) := C(L-\ell)$, then the memory requirements can be reduced to $O(N)$. In fact, since the number of clusters on level ℓ is $O(2^{d\ell})$, there holds

$$N_{st} \le CN \sum_\ell 2^{-d(L-\ell)}(L-\ell)^2 \le CN.$$

The factorisation (9.4) gives rise to a faster algorithm to compute the product

$$y := Ax, \quad A \in \mathcal{U}_{\mathcal{H},k}.$$

The matrix-by-vector multiplication now has a complexity of the order $C_{sp}k^2N + 2kLN$. Assume that we are given the level structure of $T(I)$, i.e., $T(I) = \bigcup_\ell T^{(\ell)}(I)$ and $\bigcup_{\tau\in T^{(\ell)}(I)} \tau = I$ for all $\ell = 0, \ldots, L$. The resultant algorithm consists of three steps:

Algorithm 9.2 (Matrix-vector for uniform $\mathcal{H}$-matrices).

1. (Forward transform). For $\ell = 0, \ldots, L$, and for each $\sigma \in T^{(\ell)}(I)$ compute $y^\sigma = U_\sigma^T x^\sigma \in \mathbb{R}^k$.

2. (Inner multiplication). For all $\tau : \tau \times \sigma \in \mathcal{P}$ compute

$$\widetilde{z}^{\tau,\sigma} := K^{\tau\times\sigma} y^\sigma \in \mathbb{R}^k$$

and form the sum

$$z^\tau := \sum_{\sigma\in T(I):\,\tau\times\sigma\in\mathcal{P}} \widetilde{z}^{\tau,\sigma}.$$

3. (Backward transform). For $\ell = 0,\dots,L$, and for all $\tau \in T^{(\ell)}(I)$ compute $\widetilde{y}^\tau := U_\tau z^\tau$ and define $y^{(\ell)}$ by agglomeration such that $y^{(\ell)}_{|\tau} = \widetilde{y}^\tau$. Finally, we obtain the resultant vector y by summation over all the levels, $y = \sum_\ell y^{(\ell)}$.

Proposition 9.2 *Both Steps 1 and 3 cost $2kLN$ in the case of constant rank k, and CL^2N in the case of variable rank $k_\ell = C(L-\ell)$. Step 2 has the cost about $C_{sp}k^2N$ if $k_\ell = k$ and about $C_{sp}N$ in the case $k_\ell = C(L-\ell)$.*

Due to Proposition 9.2, we relax the dependence on both C_{sp} and L (compare with the standard complexity $C_{sp}kLN$).

9.2 $\mathcal{H}^2$-matrices

A further improvement of the $\mathcal{U}_{\mathcal{H},k}$-format is based on the introducing of a simple inter-level relations between the matrix U^τ and those associated with its sons $\tau_1,\tau_2 \in S(\tau)$, $\tau \in T(I)$. Thus, the new machinery for hierarchical representation and treatment of matrix blocks explains the name "$\mathcal{H}^2$-matrix".

Assume the constant rank k on all levels, which implies the constant polynomial degree in the Lagrange interpolant (9.2) (more precisely, there holds $k(\tau) = \min\{k, |\tau|\}$, $\tau \in T$). As a matter of fact, each $l^\tau_\alpha(x) \in \mathbb{P}_k[\tau]$ can be exactly represented (interpolated) by polynomials $l^{\tau_1}_{\alpha_1}$ and $l^{\tau_2}_{\beta_2}$ defined on $X(\tau_1)$ and $X(\tau_2)$, respectively. In particular,

$$l^\tau_\alpha(x) = \sum_{\alpha_m=1}^{k} l^\tau_\alpha(\xi^{\tau_m}_{\alpha_m}) l^{\tau_m}_{\alpha_m}(x), \qquad m = 1,2. \tag{9.5}$$

For $m = 1,2$, the latter equation implies

$$U^\tau_{i\alpha} = \sum_{\alpha_m=1}^{k} l^\tau_\alpha(\xi^{\tau_m}_{\alpha_m}) \int_{X(\tau)} l^{\tau_m}_{\alpha_m}(x)\varphi_i(x)dx = \sum_{\alpha_m=1}^{k} l^\tau_\alpha(\xi^{\tau_m}_{\alpha_m}) U^{\tau_m}_{i\alpha_m}, \tag{9.6}$$

which indeed represents U^τ in terms of U^{τ_1} and U^{τ_2}, $\tau_1,\tau_2 \in S(\tau)$. Introducing the transfer (prolongation) matrices $P^{\tau_m,\tau} \in \mathbb{R}^{k\times k}$ by

$$(P^{\tau_m,\tau})_{\alpha_m\alpha} := l^\tau_\alpha(\xi^{\tau_m}_{\alpha_m}), \qquad m = 1,2,$$

we arrive at the explicit equation

$$U^\tau = \begin{pmatrix} U^{\tau_1}P^{\tau_1,\tau} \\ U^{\tau_2}P^{\tau_2,\tau} \end{pmatrix}. \tag{9.7}$$

From (9.7) we see that the storage for U^τ, $\tau \in T(I)$, requires only the higher level generation matrices $U^{\widetilde{\tau}}$ with $\widetilde{\tau}$ from the leaves of $T(I)$ as well as the family of prolongation matrices $P^{\tau_m,\tau}$ of the small size $k\times k$.

The existence of prolongation matrices $P^{\tau',\tau} \in \mathbb{R}^{k\times k}$, $\tau' \in S(\tau)$, which imply the consistency relations (9.7), is the key property of the $\mathcal{H}^2$-matrices.

Definition 9.3 *($\mathcal{H}^2$-matrix) A uniform $\mathcal{H}$-matrix $A \in \mathcal{U}_{\mathcal{H},k}$ (with respect to a generating family $\mathcal{U}$) is called an $\mathcal{H}^2$-matrix if there is a family of prolongation matrices $P^{\tau',\tau} \in \mathbb{R}^{k\times k}$, $\tau' \in S(\tau)$ for all $\tau \in T(I)$ such that the consistency relation (9.7) holds true.*

A generalisation of the above definition to the case of variable rank $\mathcal{H}$-matrices is rather straightforward (again, we set $k_\ell = C(L-\ell)$). In this case, the corresponding error analysis for the class of integral operators under consideration is quite involved (see [20, 24] for more details).

The memory requirements are asymptotically the same as for the uniform $\mathcal{H}$-matrices, i.e. $O(N)$.

The matrix-vector multiplication algorithm is based on the corresponding modification of the forward and backward steps in Algorithm 9.2 for uniform matrices (cf. §9.1). In fact, due to (9.7), for all $\sigma \in T(I)$ and $x^\sigma \in \mathbb{R}^\sigma$, we obtain for the coefficients vector $\widehat{x}^\sigma \in \mathbb{R}^k$,

$$\widehat{x}^\sigma = (U^\sigma)^T x^\sigma = \begin{pmatrix} U^{\sigma_1} & P^{\sigma_1,\sigma} \\ U^{\sigma_2} & P^{\sigma_2,\sigma} \end{pmatrix}^T \begin{pmatrix} x^{\sigma_1} \\ x^{\sigma_2} \end{pmatrix} \tag{9.8}$$

$$= \sum_{m=1}^{2} (P^{\sigma_m,\sigma})^T (U^{\sigma_m})^T x^{\sigma_m} = \sum_{m=1}^{2} (P^{\sigma_m,\sigma})^T \widehat{x}^{\sigma_m}.$$

Therefore, again the multiplication process consists of three steps.

Algorithm 9.3 (Matrix-vector for $\mathcal{H}^2$-matrices)

1. Starting from level $\ell = L$ (i.e., with 1×1 blocks), we obtain recursively vectors $\widehat{x}^\sigma$ for all levels $\ell = L-1, \ldots, 0$.

2. The inner multiplication is performed as in Algorithm 9.2, and yields the results $\widehat{y}^\tau = K^{\tau\times\sigma}\widehat{x}^\sigma$.

3. Finally, using the representation like (9.8), where we substitute σ by τ and x^σ by $\widehat{y}^\tau$, we gather all partial results $\widehat{y}^\tau$ starting at level $\ell = 0$ and ending up at $\ell \le L$, where we multiply $\widehat{y}^\tau$ with U^τ in the leaves of $T(I)$.

Proposition 9.4 *In the variable rank case, all three steps in Algorithm 9.3 have the complexity $O(N)$ provided that the multiplication with U^σ in the leaves of $T(I)$ requires $O(N)$ work. This leads to a total cost $O(N)$.*

To complete this lecture we note that $\mathcal{H}^2$-matrix approximation requires more restrictive conditions on the approximation space V_h (e.g. quasi-uniformity of a triangulation) compared with the canonical $\mathcal{H}$-matrices. Another limitation is that it is no longer possible to implement the formatted $\mathcal{H}^2$-matrix-matrix multiplication and matrix inversion with linear complexity.

10 Blended $\mathcal{H}$-matrix formats

In the present lecture we discuss a method of coupling the uniform $\mathcal{H}$-matrices with other special matrix forms, in particular, with Toeplitz and circulant matrices (see [18] for more details). This class of matrix formats allows to utilize the attractive features of special geometries and therefore to reduce the corresponding sparsity constant, the local rank and finally the complexity of arising structured matrices.

10.1 Blended Toeplitz-$\mathcal{H}$-matrices

The standard *Toeplitz matrix* $M = \{t_{ij}\}_{i,j=1}^n$ is defined by the property that the entries t_{ij} depend only on $i-j$. Using $t_{i-j} := t_{ij}$, we introduce the notation

$$M = \{t_{i-j}\}_{i,j=1}^n =: Toepl\{t_{-n+1}, \dots, t_0, \dots t_{n-1}\} \in \mathbb{R}^{n\times n}.$$

We say that the $mn \times mn$ matrix $M = \{T_{ij}\}_{i,j=1}^n$ with $m \times m$ blocks T_{ij} has a *two-level* $n \times n$ *block-Toeplitz structure*, if $T_{ij} = T_{i-j}$ depends on $i-j$ only. We introduce the class $\mathcal{M}_{T_n,1}$ ($n \in \mathbb{N}$) of *block-Toeplitz matrices* with the notation

$$M = BToepl\{T_{-n+1}, \dots, T_0, \dots, T_{n-1}\} \quad \text{with } T_q \in \mathbb{R}^{m\times m},\ q = -n+1, \dots, n-1.$$

Matrices from $\mathcal{M}_{T_n,1}$ arise in the FE approximation of integral operators with translation invariant kernels $s(x,y) = S(x-y)$ on uniform tensor product grids.

In what follows we use the *tensor product* $A \otimes B \in \mathbb{R}^{nm\times nm}$ of two matrices $A \in \mathbb{R}^{n\times n}$ and $B = \{b_{ij}\} \in \mathbb{R}^{m\times m}$ defined by $A \otimes B := \{B_{ij}\}_{i,j=1}^m$ with block matrices $B_{ij} := b_{ij}A$. Simple examples of blended formats are

$$A \otimes B \quad \text{and} \quad B \otimes A \qquad (A \in \mathcal{R}_{k,n},\ B \in \mathcal{M}_{T_m,1}). \tag{10.1}$$

We use rank-k matrices with *fixed basis* (uniform matrices, cf. Lecture 9) which lead to sublinear memory needs and also enable a faster matrix-vector multiplication. The following definition generalises the class of rank-k matrices spanned by a fixed basis.

Definition 10.1 *Given $m, k, n \in \mathbb{N}$, and $\mathcal{R}_{k,n} = \operatorname{span}\{(\mathbf{a}_i \cdot \mathbf{c}_j^\top)\}_{i,j=1}^k$ with $\mathbf{a}_i, \mathbf{c}_j \in \mathbb{R}^n$, a matrix $M \in \mathbb{R}^{nm\times nm}$ belongs to $\mathcal{M}_{T_m\otimes\mathcal{R}_{k,n}}$ if*

$$M = \sum_{i,j=1}^k T_{ij} \otimes (\mathbf{a}_i \cdot \mathbf{c}_j^\top), \qquad T_{ij} \in \mathcal{M}_{T_m,1}.$$

A matrix M belongs to $\mathcal{M}_{\mathcal{R}_{k,n}\otimes T_m}$ if

$$M = \sum_{i,j=1}^k (\mathbf{a}_i \cdot \mathbf{c}_j^\top) \otimes T_{ij}, \qquad T_{ij} \in \mathcal{M}_{T_m,1}.$$

The following statement proves the complexity of the above defined matrix classes. First we note that the vector-spaces $\mathcal{M}_{T_m\otimes\mathcal{R}_{k,n}}$ and $\mathcal{M}_{\mathcal{R}_{k,n}\otimes T_m}$ are isomorphic. If $A\otimes B\in\mathcal{M}_{T_m\otimes\mathcal{R}_{k,n}}$, there is a permutation matrix[1] Π such that $\Pi\cdot(A\otimes B)\cdot\Pi^\top = B\otimes A\in\mathcal{M}_{\mathcal{R}_{k,n}\otimes T_m}$. Hence, the matrix-vector costs $\mathcal{N}_{MV}(M)$ are the same for $M\in\mathcal{M}_{T_m\otimes\mathcal{R}_{k,n}}$ and $M\in\mathcal{M}_{\mathcal{R}_{k,n}\otimes T_m}$. The same holds for the storage $\mathcal{N}_{st}(M)$.

Lemma 10.2 *For any $M\in\mathcal{M}_{T_m\otimes\mathcal{R}_{k,n}}$ or $M\in\mathcal{M}_{\mathcal{R}_{k,n}\otimes T_{\mathbf{m}}}$ with $m\in\mathbb{N}$ there holds*

$$\mathcal{N}_{st}(M)=2k^2m+2kn,\qquad \mathcal{N}_{MV}(M)=4knm+4c_{FFT}k^2m\log m+k^2m. \tag{10.2}$$

The summand $2kn$ in $\mathcal{N}_{st}(M)$ is due to the storage of the vectors $\mathbf{a}_i,\mathbf{c}_j$. Since these vectors are fixed, they can be stored once for all and any further matrix needs only a storage of $\mathcal{N}_{st}(M)=2k^2m$.

Proof. The cost of $2kn$ for storing $\mathbf{a}_i,\mathbf{c}_j\in\mathbb{R}^n$ is already mentioned. The k^2 matrices T_{ij} require a storage of $2k^2m$.

For the proof of $\mathcal{N}_{MV}(M)$ we describe the multiplication algorithm $\mathbf{x}\mapsto\mathbf{y}=M*\mathbf{x}$ in detail, where $M\in\mathcal{M}_{\mathcal{R}_{k,n}\otimes T_m}$. We employ the block structure

$$\mathbf{x}=\begin{pmatrix}x_1\\ \vdots\\ x_m\end{pmatrix},\quad \mathbf{y}=\begin{pmatrix}y_1\\ \vdots\\ y_m\end{pmatrix}$$

with $x_\alpha,y_\alpha\in\mathbb{R}^n$ for $\alpha=1,\dots,m$.

In **Step 1**, we compute the vector $\mathbf{g}^j=(\langle\mathbf{c}_j,x_\alpha\rangle)_{\alpha=1,\dots,m}\in\mathbb{R}^m$ of scalar products. Since j varies in $\{1,\dots,k\}$, the total cost of Step 1 is $2nmk$.

In **Step 2**, the Toeplitz matrices $T_{ij}\in\mathcal{M}_{T_m,1}$ are applied to $\mathbf{g}^j:\mathbf{h}^i:=\sum_{j=1}^m T_{ij}\mathbf{g}^j$ $(1\le i\le k)$. The corresponding cost amounts to $4c_{FFT}k^2m\log m+k^2m$, where the latter term corresponds to the summation.

In **Step 3**, the resulting vector $\mathbf{y}=M*\mathbf{x}$ is computed by means of $y_\alpha:=\sum_{i=1}^k\mathbf{h}^i_\alpha*\mathbf{a}_i$ $(\alpha=1,\dots,m)$, which requires $2nmk$ operations. Summing over Steps 1-3, we obtain the result for $\mathcal{N}_{MV}(M)$. ∎

Remark 10.3 *One can observe special block-Toeplitz structure that is inherent in the standard $\mathcal{H}$-matrix format on uniform tensor-product meshes arising in the FE approximation of integral operators with translation invariant kernels. For $d=1$, consider the format $\mathcal{M}_{\mathcal{H},k}(I\times I,\mathcal{P})$ corresponding to the admissible partitioning $\mathcal{P}$ depicted in Fig. 6 (see Lecture 5) with $n=2^L$. We take a uniform $\mathcal{H}$-matrix that is represented by coefficients $\zeta_{ij}\in\mathbb{R}^{k\times k}$. Due to the Toeplitz data-structure of blocks, on each level $\ell=2,...,L$, the only "generating" blocks $b=\tau_1^\ell\times\tau_j^\ell$ and $b=\tau_i^\ell\times\tau_1^\ell$ with the corresponding $1\le i,j\le 2^\ell$, have to be stored. This amounts in the overall computer memory $O(C_{sp}k^2L)$, so, we benefit from logarithmic storage requirements.*

[1] Π is defined by $(\Pi x)_i=x_{\pi(i)}$ for $1\le i\le n|m|$, where π is the permutation $\pi: i=\alpha+(\beta-1)n\mapsto\beta+(\alpha-1)|m|$ $(1\le\alpha\le n,\ 1\le\beta\le m)$.

It is easy to see that the similar block-Toeplitz $\mathcal{H}$-matrix structure is observed in the case of tensor-product $n_1 \times ... \times n_1$ grids in $\mathbb{R}^d$, $d \geq 2$, that are uniform in each spacial direction, such that $n = n_1^d = O(2^{dL})$. This implies the logarithmic storage requirements $O(C_0^d k^2 L)$, where $C_{sp} = O(C_0^d)$ with $C_0 \cong 6$ (cf. (6.6)).

To reduce the matrix-by-vector cost, one has to impose more structure of data. Following [17], we assume that for each block $b \in \mathcal{P}$, the representation vectors $\mathbf{a}_i, \mathbf{c}_j \in \mathbb{R}^{n_b^d}$ have a separable form $\mathbf{a}_i = a_{i1} \times ... \times a_{id}$, $\mathbf{c}_i = c_{i1} \times ... \times c_{id}$ with $a_{ik}, c_{ik} \in \mathbb{R}^{n_b}$ (cf. the so-called fully separable expansions [17]). Moreover, let the target vector $\mathbf{x} \in \mathbb{R}^{n_1^d}$ have a separable form $\mathbf{x} = x_1 \times ... \times x_d$ with $x_k \in \mathbb{R}^{n_1}$ (cf. [21]). Now it is easy to see that a cost of $M\mathbf{x}$ will be reduced to the expense of one-dimensional operations $O(C_{sp} d k^2 L n^{1/d})$, $n = n_1^d$.

10.2 Blended circulant-$\mathcal{H}$-matrices

We recall the definition of circulant matrices.

Definition 10.4 *An $n \times n$ matrix $\mathcal{C}$ is called* circulant *if it has the representation*

$$\mathcal{C} = circ\{c_1, \ldots, c_n\} := \begin{pmatrix} c_1 & c_2 & \ldots & c_n \\ c_n & c_1 & \ldots & c_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_2 & \ldots & c_n & c_1 \end{pmatrix}, \qquad c_i \in \mathbb{C}\,.$$

The set of all $n \times n$ circulant matrices is closed with respect to addition and multiplication. Any circulant matrix $\mathcal{C}$ is associated with the polynomial

$$p_c(z) := c_1 + c_2 + \ldots + c_n z^{n-1} \ (z \in \mathbb{C})$$

and it has a diagonal representation in the Fourier basis,

$$\mathcal{C} = F_n^T \Lambda_c F_n \quad \text{with } \Lambda_c = \mathrm{diag}\{p_c(1), \ldots, p_c(\omega^{n-1})\}, \qquad \omega = e^{i\pi/n}.$$

The eigenvector corresponding to the eigenvalue $p_c(\omega^{j-1})$ is given by jth column of F_n, i.e.,

$$\vec{\omega}_j = \frac{1}{\sqrt{n}}(\omega^{(k-1)(j-1)})_{k=1,\ldots,n}.$$

The matrix vector multiplication with $\mathcal{C}$ costs $2C_{FFT} n \log n + O(n)$ operations. For the block-circulant matrices we use the notation $M = Bcirc\{C_1, ..., C_n\}$, with $C_i \in \mathbb{R}^{m \times m}$.

Given $q \in \mathbb{N}_{>0}$, define the multiindex $\mathbf{n} = (n_1, \ldots, n_q) \in \mathbb{N}^q$ with $|\mathbf{n}| := n_1 n_2 \cdots n_q$. The matrix class $\mathcal{M}_{C_\mathbf{n},q}$ of q-level block-circulant matrices is given by the following definition.

Definition 10.5 *For $q \in \mathbb{N}_{>0}$, we define the matrix class $\mathcal{M}_{C_\mathbf{n},q}$ recursively.*
For $q = 1$, $M = circ(c_1, ..., c_n) \in \mathcal{M}_{C_n,1}$ denotes the standard circulant matrix $M \in$

$\mathbb{R}^{n\times n}$.
Given $q \in \mathbb{N}$, $q \geq 2$, and $\mathbf{n} \in \mathbb{N}^q$, a matrix M belongs to $\mathcal{M}_{C_{\mathbf{n}},q} \subset \mathbb{R}^{|\mathbf{n}|\times|\mathbf{n}|}$ if

$$M = Bcirc\{C_1, \ldots, C_n\} \qquad \text{with } C_j \in \mathcal{M}_{C_{\mathbf{n}'},q-1} \text{ for } 1 \leq j \leq n_1,$$

where $\mathbf{n}' := (n_2, \ldots, n_q) \in \mathbb{N}^{q-1}$.

Example 10.1 (3D BEM on a rotational surface).

Assume that the (single-layer) kernel of the boundary integral operator $\mathcal{A}$ is given by a translation invariant function $s(x,y) = S(x-y)$ $(x, y \in \mathbb{R}^3)$. Let Γ be a rotational surface $\Gamma = \Gamma_z \times [0, 2\pi] \subset \mathbb{R}^3$ obtained by means of an arc Γ_z. Consider a tensor-product ansatz space $V_h = V_z \times V_\varphi$ of piecewise constant finite elements associated with the tensor product of a uniform grid in $\varphi \in [0, 2\pi]$ (with mesh-size $h_\varphi = \frac{2\pi}{m_\varphi}$) and a quasi-uniform mesh on Γ_z. Let $V_z := \operatorname{span}\{\psi_i\}_{i\in I_z}$ with $I_z := \{i = 1, \ldots, n_z\}$ and $V_\varphi := \operatorname{span}\{\varphi_j\}_{j=1}^{m_\varphi}$. Define $W_j := V_z \otimes \varphi_j = \operatorname{span}\{v_i^j\}_{i\in I_z}$, where $v_i^j := \psi_i \otimes \varphi_j$. Note that the entries $\langle \mathcal{A}v_i^j, v_k^l\rangle$ of the exact stiffness matrix A_h depend on $i, k \in I_z$ and on the difference $j - l$ modulo m_φ. Hence, $A_h = Bcirc\{A_1, ..., A_{m_\varphi}\}$ has the block structure of $\mathcal{M}_{C_{m_\varphi},1}$ specified in Definition 10.5, where the generating blocks $A_l = \langle \mathcal{A}v_i^1, v_k^l\rangle_{i,k=1}^{n_z}$ $(l = 1, \ldots, m_\varphi)$ correspond to the product spaces $W_1 \times W_l$.

10.3 Complexity analysis

We define blended matrix formats based on block-circulant matrices. The index set I is assumed to have the product form $I = I_z \times I_c$, where $I_c = \{1, \ldots, m\}$ corresponds to the circulant part. We introduce the hierarchical tree $T(I_z)$ of depth L and the level subsets $T^{(\ell)} := \{\tau \in T(I_z) : \tau \text{ belongs to level } \ell\}$. The admissible partitioning $\mathcal{P}$ splits into the level sets $\mathcal{P}^{(\ell)} := \{\tau \times \sigma \in \mathcal{P} : \tau, \sigma \in T^{(\ell)}\}$. We recall that $\ell = 0$ corresponds to the biggest cluster $I_z \times I_z$ (root of the tree $T(I_z \times I_z)$, while $\ell = L$ corresponds to the leaves (1×1 blocks). For each block $b = \tau \times \sigma \in \mathcal{P}^{(\ell)}$, the corresponding matrix-block belongs to the vector space $\mathcal{U}(k, \tau \times \sigma) := \operatorname{span}\{\mathbf{a}_i^\tau \cdot (\mathbf{c}_j^\sigma)^\top\}_{i,j=1}^{k_\ell}$, where the rank k_ℓ depends on the level only. We denote this vector space of uniform $\mathcal{H}$-matrices by $\mathcal{U}_{\mathcal{H},k}(I_z \times I_z, \mathcal{P}, \mathcal{U})$.

Definition 10.6 *Let $I = I_z \times I_c$, with $n := \#I_z$. For a given mapping $k : \{0, \ldots, L\} \to \mathbb{N}$, the set $\mathcal{U}_{\mathcal{H},k}$ of uniform $\mathcal{H}$-matrices is defined as above. Then a matrix $M \in \mathbb{R}^{nm\times nm}$ belongs to $\mathcal{M}_{\mathcal{U}_{\mathcal{H},k}\otimes C_m}$ if*

$$M = Bcirc\{A_1, \ldots, A_m\} \qquad \text{for certain } A_p \in \mathcal{U}_{\mathcal{H},k} \subset \mathbb{R}^{n\times n},\ p \in I_c.$$

Now we introduce an algorithm for the fast matrix-vector multiplication with matrices from $\mathcal{M}_{\mathcal{U}_{\mathcal{H},k}\otimes C_m}$ based on the simultaneous use of circulant and $\mathcal{H}$-matrix structures.

Algorithm 10.7 *Given the matrices $A_1, \ldots, A_m \in \mathcal{U}_{\mathcal{H},k}$, with blocks specified by*

$$A_p^{\tau\times\sigma} = \sum\nolimits_{i,j=1}^{k_\ell} \zeta_{p,ij}^{\tau\times\sigma}\, \mathbf{a}_i^\tau \cdot (\mathbf{c}_j^\sigma)^\top \qquad (p = 1, \ldots, m),$$

and given the vector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$, *we have to compute* $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} := M\mathbf{x}$, *where* $M = Bcirc\{A_1, ..., A_m\}$. *All vector components* x_α,y_α, $\alpha \in I_c$, *belong to* $\mathbb{R}^{I_z}$, *while* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I = \mathbb{R}^{I_z \times I_c}$.

Step 1. *Compute the set of vectors formed by scalar products*

$$\mathbf{x}_j^\sigma := \left(\left\langle \mathbf{c}_j^\sigma, x_{1\sigma} \right\rangle, \ldots, \left\langle \mathbf{c}_j^\sigma, x_{m\sigma} \right\rangle\right)^\top \in \mathbb{R}^m \qquad \text{for all } 0 \le \ell \le L,\ \sigma \in T^{(\ell)},\ 1 \le j \le k_\ell,$$

where $x_{\alpha\sigma}$ *denotes the block vectors* $(x_{\alpha,i})_{i\in\sigma}$.

Step 2. *(a) Multiply the circulant matrices* $circ\{\zeta_{1,ij}^{\tau\times\sigma}, \ldots, \zeta_{m,ij}^{\tau\times\sigma}\} \in \mathbb{R}^{m\times m}$ *by vectors* $\mathbf{x}_j^\sigma$,

$$\mathbf{z}_{ij}^{\tau\times\sigma} := circ\{\zeta_{1,ij}^{\tau\times\sigma}, \ldots, \zeta_{m,ij}^{\tau\times\sigma}\}\, \mathbf{x}_j^\sigma \in \mathbb{R}^m \qquad \text{for all } 0 \le \ell \le L,\ \tau\times\sigma \in \mathcal{P}^{(\ell)},\ 1 \le i, j \le k_\ell,$$

and (b) form the sums

$$\mathbf{z}_i^\tau := \sum_{\sigma:\tau\times\sigma\in\mathcal{P}} \sum_{j=1}^{k_\ell} \mathbf{z}_{ij}^{\tau\times\sigma} \in \mathbb{R}^m \qquad \text{for all } 0 \le \ell \le L,\ \tau \in T^{(\ell)},\ 1 \le i \le k_\ell.$$

Step 3. *The summation of the intermediate results*

$$y''_{\beta\tau} := \sum_{i=1}^{k_\ell} \mathbf{a}_i^\tau \cdot (\mathbf{z}_i^\tau)_\beta \in \mathbb{R}^\tau \qquad \text{for all } 0 \le \ell \le L,\ \tau \in T^{(\ell)},\ \beta = 1, \ldots, m,$$

starts at the leaves of the tree $T(I_z)$, *where* $y'_{\beta\tau} := y''_{\beta\tau}$. *Then*

$$y'_{\beta\tau} := y''_{\beta\tau} + (y'_{\beta\tau'})_{\tau'\in S(\tau)} \in \mathbb{R}^\tau \qquad \text{for all } \tau \in T(I_z),\ \beta = 1, \ldots, m,$$

ends at the root $I_z \in T(I_z)$, *where* $y_\beta = y'_{\beta I_z}$ $(\beta = 1, \ldots, m)$ *represents the final result.*

The following statement establishes the complexity of Algorithm 10.7. Note that $N = nm$ is the dimension of the problem. Recall that under the assumptions on a construction of the admissible partitioning, $\#\mathcal{P}^{(\ell)} \le C_{sp} n 2^{\ell-L}$ holds with the sparsity constant $C_{sp}(\mathcal{P})$.

Lemma 10.8 *Assume* $L = \log n$ *(with* $n = \#I_z$*), which holds for a uniform tree* $T(I_z)$. *Concerning* k_ℓ *we assume* $k_L \le k_{L-1} \le \ldots \le k_0$. *Then for any* $M \in \mathcal{M}_{\mathcal{U}_{\mathcal{H},k}\otimes C_m}$ *the storage requirements and the matrix-vector multiplication cost are*

$$\mathcal{N}_{st}(M) \le n \left[2k_0 \log n + C_{sp} m \sum_{\ell=0}^{L} k_\ell^2 2^{\ell-L} \right], \tag{10.3}$$

$$\mathcal{N}_{MV}(M) \le nm \left[(1 + 4k_0 \log n) + 2C_{sp} C_{FFT} (1 + \log m) \sum_{\ell=0}^{L} k_\ell^2 2^{\ell-L} \right]. \tag{10.4}$$

An interesting choice of k_ℓ *is* $k_\ell = k_L + (L - \ell)\delta$ $(\delta \ge 0)$. *Then* $\sum_{\ell=0}^{L} k_\ell^2 2^{\ell-L} = O(k_L^2)$ *depends on the smallest value* k_L, *not on the maximal one* k_0.

Proof. The storage for vectors $\{\mathbf{a}_i^\tau, \mathbf{c}_j^\sigma\}$ for all $\tau, \sigma \in T(I_c)$ is estimated by

$$2\sum_{\ell=0}^{L} k_\ell \sum_{\tau\in T(\ell)} \#\tau = 2n\sum_{\ell=0}^{L} k_\ell \le 2k_0 n \log_2 n,$$

where we use that $L = \log_2 n$. The storage of the coefficients $\{\zeta_{p,ij}^{\tau\times\sigma}\}$ for all $1 \le p \le m$, $\tau\times\sigma \in \mathcal{P}^{(\ell)}$, $0 \le \ell \le L$, is bounded by

$$m\sum_{\ell=0}^{L} k_\ell^2 \#P_2^\ell = C_{sp} nm \sum_{\ell=0}^{L} k_\ell^2 2^{\ell-L}.$$

This proves (10.3). In the rest of the proof we set $\log = \log_2$.

The cost of Step 1 is $2m\sum_{\ell=0}^{L} k_\ell \sum_{\sigma\in T(\ell)} \#\sigma$ operations, since one scalar product needs $2\#\sigma$ operations. As $\sum_{\sigma\in T^{(\ell)}} \#\sigma = n$, we arrive at $2nm\sum_{\ell=0}^{L} k_\ell \le 2k_0 nm \log n$.

One multiplication by the circulant matrix in Step 2a costs $2C_{FFT}\, m\log m$. Hence, the resulting costs are $2C_{FFT}\, m\log m\sum_{\ell=0}^{L} k_\ell^2 \#\mathcal{P}^{(\ell)} = 2C_{sp}C_{FFT} nm\log m\sum_{\ell=0}^{L} k_\ell^2 2^{\ell-L}$ operations. The summation in Step 2b requires $m\sum_{\ell=0}^{L} k_\ell^2 \#\mathcal{P}^{(\ell)} \le C_{sp} nm\sum_{\ell=0}^{L} k_\ell^2 2^{\ell-L}$. Thus, the total cost of Step 2 is $C_{sp} nm(1+\log m)\sum_{\ell=0}^{L} k_\ell^2 2^{\ell-L}$

In Step 3, the computation of $y''_{\beta\tau}$ needs $m\sum_{\ell=0}^{L} 2k_\ell \sum_{\tau\in T^{(\ell)}} \#\tau = 2nm\sum_{\ell=0}^{L} k_\ell \le 2nmk_0 \log n$ operations. The summation over the tree involves $m(\#T(I_z) - n) < nm$ additions. Together, we have $nm(1 + 2k_0 \log n)$ operations. These partial costs add up to (10.4). ∎

Next we consider an example of the circulant version $\mathcal{M}_{\mathcal{U}_{\mathcal{H},k}\otimes C_m}$ of blended matrix-formats in the BEM application.

Example 10.2 (3D BEM on a rotational surface revisited)

Recall that the stiffness matrix in Example 10.1, $A_h = Bcirc\{A_1, ..., A_{m_\varphi}\}$ has the block structure of $\mathcal{M}_{C_{m_\varphi},1}$ specified in Definition 10.5, where the generating blocks $A_l = \langle \mathcal{A}v_i^1, v_k^l\rangle_{i,k=1}^{n_z}$ $(l = 1, \ldots, m_\varphi)$ correspond to the product spaces $W_1 \times W_l$. Approximating these blocks by $\mathcal{H}$-matrices from $\mathcal{U}_{\mathcal{H},k}(I_z \times I_z, \mathcal{P})$, we arrive at the desired construction. The linear-logarithmic complexity bound due to Lemma 10.8 holds with $m = m_\varphi$ and $n = n_z$.

10.4 Application to oscillatory kernels

Consider the 3D Galerkin BEM matrix with the Helmholtz kernel

$$s(x,y) = \frac{e^{i\kappa|x-y|}}{|x-y|} \qquad (x, y \in \mathbb{R}^3) \tag{10.5}$$

defined on $\Gamma = \Gamma_z \times [0, 2\pi]$ with Γ_z chosen as a polygon. In Examples 10.1, 10.2, consider an admissible block $\tau\times\sigma \in \mathcal{P}$. To avoid the use of a bounding box for τ, we simplify the construction of the cluster tree and assume that each $X(\tau) : \tau\times\sigma \in \mathcal{P}$, lies on some edge of Γ_z. This yields $B_x[\tau] = X(\tau)$.

Lemma 10.9 *Let $s(x,y)$, $(x,y) \in X(\tau) \times X(\sigma)$, be given by (10.5). Then for the polynomial interpolant of degree p (with respect to the Chebyshev nodes) there holds*

$$||s(x,y) - \mathcal{I}_p s||_{L^\infty(X(\sigma)\times X(\tau))} \leq 2(\Lambda_p + 1)\frac{M_{\rho_*}}{\rho_* - 1}\rho_*^{-p}, \tag{10.6}$$

with some $\rho_* > 1$, uniformly with respect to the block-size n, where $M_{\rho_*}[s] \leq C\exp(\kappa\delta)$, $\delta = \frac{1}{2}diamX(\tau)$.

Proof. We apply Theorem 3.3 to the univariate function $u := s(x,y)$ of x, i.e., $(x,y) \in I_\delta \times B_y[\sigma]$, identifying I_δ with a piece of Γ_z, i.e., $I_\delta := X(\tau) \cap \Gamma_z$. The corresponding regularity ellipse $\mathcal{E}_{\rho_0}$ has $\rho_0 = a_0 + b_0$, where $a_0^2 = b_0^2 + \delta^2$, and the small semi-axis b_0 given by $b_0 = \lambda_0 \operatorname{dist}(X(\tau), X(\sigma)) = 2\lambda_0\delta$, with some $\lambda < 1$ and also $a_0 = \sqrt{1+4\lambda^2}\,\delta$. With such a choice, we have

$$\rho_0 = \left(2\lambda + \sqrt{1+4\lambda^2}\right)^{-1}. \tag{10.7}$$

For the kernel given by (10.5), the constant $M_\rho[f]$ can be estimated by

$$M_\rho[f] \leq C \max_{x\in\mathcal{E}_\rho(I_\delta),\, y\in B_y[\sigma]} |s(x,y)| \leq c\frac{\exp(\kappa b_0)}{2\delta - b_0}.$$

Then (10.6) follows with $\rho_* = \rho(\sigma^*)$, where σ^* is the minimizer in (4.11). ∎

Corollary 10.10 *In the situation of Example 10.2, the approximate BEM stiffness matrix $M \in \mathcal{M}_{\mathcal{M}_{\mathcal{H},k}\otimes C_m}$ of blended type yields the following complexity estimate*

$$\mathcal{N}_{MV}(M) = O(N(L + \kappa + L\frac{\kappa^2}{n_z})\log N),$$

where $N = n_z\, m_\varphi$. The storage is estimated by

$$\mathcal{N}_{st}(M) = O(N + n_z L\,\kappa \log^2 N). \tag{10.8}$$

To complete the discussion, we note that the blended approximations of Toeplitz-$\mathcal{H}$-matrix type are similar. Blended approximations may be directly applied to 3D BEM problems on special surfaces (e.g., rotational surface, boundary of parallelepiped or L-shaped domains, etc.). In particular, this is the case for coupled FEM-BEM methods for solving elliptic problems in unbounded domains since, in this situation, an auxiliary boundary can be chosen as a special surface. It is worth to note that each fast $O(N)$-method applied to the one-dimesional problem associated with the index set I_z immediately leads to a linear-logarithmic complexity scheme in 3D with local rank k not depending on κ.

References

[1] M. Bebendorf and W. Hackbusch. Existence of $\mathcal{H}$-matrix approximations to the inverse FE-matrix of elliptic operator. Numer. Math. 2003 (to appear).

[2] S. Börm, L. Grasedyck and W. Hackbusch. Introduction to Hierarchical Matrices with Application. Preprint MPI MIS 18, Leipzig 2002.

[3] S. Börm, L. Grasedyck and W. Hackbusch. $\mathcal{H}$-Matrices. Proceedings of Winter School on the $\mathcal{H}$-Matrix Implementation. 2003.

[4] D. Braess. Finite Elemente. Springer, Berlin, 1991.

[5] P. Ciarlet, The finite element methods for elliptic problems. Amsterdam, North-Holland, 1978.

[6] R.A. DeVore and G.G. Lorentz. Constructive Approximation. Springer-Verlag, 1993.

[7] I.P. Gavrilyuk, W. Hackbusch and B.N. Khoromskij. *$\mathcal{H}$-Matrix Approximation for Elliptic Solution Operator in Cylinder Domains.* East-West J. of Numer. Math., v. 9, 1, 2001, 25-58.

[8] I.P. Gavrilyuk, W. Hackbusch and B.N. Khoromskij. *Data-Sparse Approximation to Operator-Valued Functions of Elliptic Operators.* Preprint MPI MIS 54, Leipzig, 2002; Math. Comp. (to appear).

[9] I.P. Gavrilyuk, W. Hackbusch and B.N. Khoromskij. *$\mathcal{H}$-Matrix Approximation for the Operator Exponential with Applications.* Numer. Math. (2002) 92: 83-111.

[10] L. Grasedyck, W. Hackbusch and B.N. Khoromskij. *Solution of large scale algebraic matrix Riccati equations by use of hierarchical matrices.* Preprint MPI MIS 62, Leipzig, 2002; Computing (to appear).

[11] L. Grasedyck and W. Hackbusch: *Construction and arithmetics of $\mathcal{H}$-matrices* Preprint MPI MIS 103, Leipzig, 2002.

[12] W. Hackbusch. Elliptic Differential Equations: Theory and Numerical Treatment. Springer-Verlag, Berlin, 1992.

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

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

[15] W. Hackbusch and B.N. Khoromskij: *A sparse $\mathcal{H}$-matrix arithmetic: General complexity estimates.* J. Comp. Appl. Math. **125** (2000) 479-501.

[16] W. Hackbusch and B.N. Khoromskij. *$\mathcal{H}$-Matrix Approximation on Graded Meshes.* The Mathematics of Finite Elements and Applications X, MAFELAP 1999, J.R. Whiteman (ed), Elsevier, Amsterdam, Chapter 19, 307-316, 2000.

[17] W. Hackbusch and B.N. Khoromskij. *Towards $\mathcal{H}$-Matrix Approximation of the Linear Complexity.* Operator Theory: Advances and Applications, Vol. 121, Birkhäuser Verlag, 2001, 194-220.

[18] W. Hackbusch and B.N. Khoromskij. *Blended Kernel Approximation in the $\mathcal{H}$-Matrix Techniques.* Numer. Linear Algebra Appl. 2002; 9: 281-304.

[19] W. Hackbusch, B.N. Khoromskij and R. Kriemann: *Hierarchical matrices based on a weak admissibility criterion.* Preprint MPI MIS 2, 2003, Leipzig.

[20] W. Hackbusch, B.N. Khoromskij and S. Sauter: *On $\mathcal{H}^2$-matrices.* In: Lectures on Applied Mathematics (H.-J. Bungartz, R. Hoppe, C. Zenger, eds.), Springer-Verlag, Berlin, 2000, 9-30.

[21] W. Hackbusch, B.N. Khoromskij and E. Tyrtyshnikov: *Hierarchical Kronecker tensor-product approximation*, Preprint MPI MIS 35, Leipzig 2003; Numer. Math. (submitted).

[22] E. Isaacson and H.B. Keller. Analysis of Numerical Methods. Dover Publ., Inc., NY, 1994.

[23] B.N. Khoromskij. *Data-sparse Approximate Inverse in Elliptic Problems: Green's Function Approach.* Preprint MPI MIS 79, Leipzig, 2001; J. Numer. Math., 2003 (to appear).

[24] M. Melenk, S. Börm and M. Lönsdorf. Approximation of Integral Operators by Variable-Order Interpolation. Preprint MPI MIS 82, Leipzig 2002.

[25] A. Quarteroni and A. Valli, Theory and application of Steklov-Poincaré operators for boundary-value problems. Applied and Industrial Mathematics, Kluwer AP, 1991, 179-203.

[26] F. Stenger: *Numerical methods based on Sinc and analytic functions.* Springer-Verlag, Heidelberg, 1993.