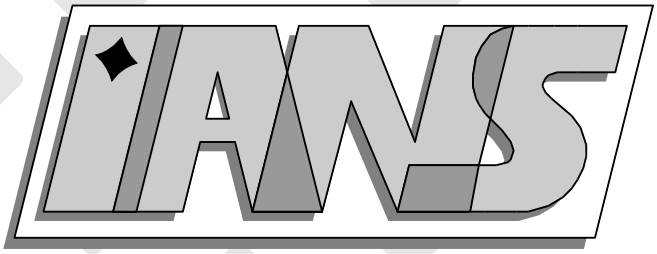
Universität Stuttgart



Vorlesung Hierarchische Matrizen

O. Steinbach

Berichte aus dem Institut für Angewandte Analysis und Numerische Simulation

Vorlesungsskript 2004/016

Universität Stuttgart

Vorlesung Hierarchische Matrizen

O. Steinbach

Berichte aus dem Institut für Angewandte Analysis und Numerische Simulation

Vorlesungsskript 2004/016

Institut für Angewandte Analysis und Numerische Simulation (IANS)
Fakultät Mathematik und Physik
Fachbereich Mathematik
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: ians-preprints@mathematik.uni-stuttgart.de
WWW: http://preprints.ians.uni-stuttgart.de

ISSN **1611-4176**

Vorwort

Dieses Vorlesungsskript *Hierarchische Matrizen* umfaßt die gleichnamige Vorlesung, welche ich im Sommersemester 2004 an der Technischen Universität Dresden im Rahmen von zwei Semesterwochenstunden ergänzt durch einstündige Übungen gehalten habe. Diese Vorlesung richtet sich sowohl an Studenten der Mathematik als auch an Studenten der Ingenieurwissenschaften. Die Darstellung ist deshalb bewußt konstruktiv gehalten, um die entscheidenden Ideen und Werkzeuge vermitteln zu können. An dieser Stelle sei auch auf den Bericht [25] zum gleichnamigen Hauptseminar an der Universität Stuttgart verwiesen, welches ebenfalls im Sommersemester 2004 in Form eines Kompaktseminars stattgefunden hat.

Mein Dank gilt dem Fachbereich Mathematik und vor allem dem Institut für Wissenschaftliches Rechnen an der TU Dresden für die Möglichkeit, im Sommersemester 2004 die Professur für Wissenschaftliches Rechnen vertreten zu können. Besonderer Dank gebührt G. Of für das Korrekturlesen des Manuskriptes.

Dresden und Stuttgart, September 2004 Olaf Steinbach

Inhaltsverzeichnis

Kapitel 1

Einführung

Weit verbreitete numerische Diskretisierungsverfahren für die näherungsweise Lösung von Randwertproblemen partieller Differentialgleichungen sind die Finite Element Methode (FEM) [6, 12, 15, 24] sowie Randelementmethoden (BEM) [21, 22, 24]. Beide Näherungsverfahren führen auf Familien von linearen Gleichungssystemen

$$A\underline{x} = \underline{f} \tag{1.1}$$

mit Steifigkeitsmatrizen $A \in \mathbb{R}^{n\times n}$. Die Dimension $n \in \mathbb{N}$ wiederspiegelt dabei die Approximationsgüte des numerischen Verfahrens, so daß insbesondere der Fall großer Dimensionen n von Interesse ist. Im folgenden wird die Matrix A stets als **invertierbar** vorausgesetzt, so daß die eindeutige Lösbarkeit des linearen Gleichungssystems (1.1) gewährleistet ist. Allgemein sei die Matrix A **vollbesetzt**, das heißt, zur Beschreibung der Matrix $A \in \mathbb{R}^{n\times n}$ sind n^2 Matrixeinträge $A[\ell, k]$ für $k, \ell = 1, \ldots, n$ abzuspeichern. Die Diskretisierung mit Finiten Elementen führt in der Regel auf **schwachbesetzte** Steifigkeitsmatrizen mit einem Speicherbedarf von nur $\mathcal{O}(n)$ Elementen. Die Beschreibung der inversen FEM Steifigkeitsmatrizen erfordert aber wiederum einen quadratischen Aufwand an Speicherplatz.

Die Verwendung eines **direkten Verfahrens** zur Lösung des linearen Gleichungssystems (1.1) wie zum Beispiel dem Eliminationsverfahren nach Gauß oder der LU–Zerlegung erfordert $\mathcal{O}(n^3)$ wesentliche Rechenoperationen. Dies bedeutet, daß eine Verdoppelung der Dimension n des linearen Gleichungssystems die Rechenzeit zur Lösung des linearen Gleichungssystems verachtfacht. Die Verwendung von **iterativen Lösungsverfahren** [11, 19, 26], wie zum Beispiel dem Verfahren konjugierter Gradienten, verlangt pro Iterationsschritt in der Regel nur eine Matrix–Vektor–Multiplikation mit der Steifigkeitsmatrix A, welche für eine vollbesetzte Matrix A mit n^2 wesentlichen Rechenoperationen durchgeführt werden kann. Wenn durch eine geeignete Vorkonditionierungsstrategie die Zahl der zum Erreichen einer vorgegebenen Genauigkeit erforderlichen Iterationsschritte unabhängig von n begrenzt werden kann, beträgt der Aufwand zum Lösen der linearen Gleichungssysteme $\mathcal{O}(n^2)$ wesentliche Rechenoperationen. Somit führt eine Verdoppelung der Dimension n zu einer Vervierfachung der Rechenzeit. Eine wesentliche Schranke für die Dimension n stellt

jedoch der verfügbare Hauptspeicher dar, so daß viele praktische Problemstellungen auf herkömmlichen Rechnern nicht behandelt werden können, da die zugehörigen vollbesetzten Steifigkeitsmatrizen nicht im Hauptspeicher aufgestellt und bereitgehalten werden können.

Deshalb besteht die Notwendigkeit, den quadratischen Aufwand zur Speicherung und Anwendung einer vollbesetzten Matrix erheblich zu verringern. Optimal wäre ein in der Zahl der Freiheitsgrade linearer Aufwand, der meist aber nur bis auf polylogarithmische Störterme erreicht werden kann. Die Aufgabe besteht deshalb darin, eine vollbesetzte Matrix $A \in \mathbb{R}^{n \times n}$ durch eine **data–sparse** Approximation $\widetilde{A} \in \mathbb{R}^{n \times n}$ anzunähern, welche folgenden Kriterien unterliegen soll:

1. Der **Speicherbedarf** zur Beschreibung der Approximation $\widetilde{A}$ ist von der Größenordnung
$$Sp(\widetilde{A}) \;=\; \mathcal{O}(n \log^{\alpha} n) \tag{1.2}$$
mit einem gewissen Parameter $\alpha \in \mathbb{N}$ unabhängig von n.

2. Die Anzahl der **wesentlichen Operationen** einer Matrix–Vektor–Multiplikation mit der Approximation $\widetilde{A}$ ist von der Größenordnung
$$Op(\widetilde{A}\underline{x}) \;=\; \mathcal{O}(n \log^{\beta} n) \tag{1.3}$$
wiederum mit einem Parameter $\beta \in \mathbb{N}$ unabhängig von n.

3. Für ein gegebenes $\varepsilon > 0$ kann die **Genauigkeit**
$$\|A - \widetilde{A}\|_M \;\leq\; \varepsilon \tag{1.4}$$
in einer gegebenen Matrix–Norm $\|\cdot\|_M$ gewährleistet werden.

Methoden zur effizienten Approximation vollbesetzter Matrizen wurden ursprünglich entwickelt für die Diskretisierung der bei Randelementmethoden auftretenden nichtlokalen Operatoren. Zu nennen sind hier die **schnelle Multipol–Methode** [10] und der **Panel–Clustering Algorithmus** [14]. Beide Verfahren beruhen auf der expliziten Kenntnis einer Reihenentwicklung der nichtlokalen Kernfunktion der Randintegraloperatoren. Beide Approximationen dienen einer effizienten Auswertung einer Matrix–Vektor–Multiplikation, ohne daß die zugehörige Steifigkeitsmatrix explizit aufgestellt wird. Diese Anwendungen entsprechen aber einer **hierarchischen Partitionierung** der Matrix, wobei die abgebrochene Reihenentwicklung der Kernfunktion eine **Niedrig–Rang–Darstellung** der zugehörigen Blockmatrizen induziert. Durch **hierarchische Matrizen** werden diese Zugänge verallgemeinert, und durch eine zugehörige **Arithmetik** wird das Rechnen mit hierarchischen Matrizen erklärt und insbesondere eine effiziente Berechnung von inversen Matrizen ermöglicht. Die hierarchische Partitionierung von Matrizen und ihre Niedrig–Rang–Approximation kann durch verschiedene Verfahren realisiert werden. Erste Arbeiten hierzu sind die **Mosaic–Skeleton–Approximation** [27] und darauf aufbauend die **Adaptive Cross Approximationsmethode** [1, 2, 4]. In [13] wurde dann ein allgemeiner Zugang

zu hierarchischen Matrizen und ihrer Arithmetik präsentiert. Dies wurde und wird in einer Reihe von Arbeiten unter verschiedenen Aspekten für unterschiedliche Anwendungen untersucht. Zu nennen sind hier insbesondere die Dissertation [9] und das Skript [5].

Hier soll zunächst ein rein algebraischer Zugang zur Definition von hierarchischen Matrizen verfolgt werden. Dieser beruht auf einer hierarchischen Block–Partitionierung von Matrizen und einer Niedrig–Rang–Darstellung der entstehenden Block–Matrizen. Die Approximierbarkeit impliziert gleichzeitig ein Zulässigkeitskriterium für die zu verwendenden hierarchisch erzeugten Block–Matrizen. Die Niedrig–Rang–Approximation basiert auf einer Faktorisierung von symmetrischen Matrizen bzw. der Singulärwertzerlegung von allgemeinen Matrizen. Sind zwei Matrizen durch dieselbe hierarchische Partitionierung gegeben, so kann die Summe aller Blockmatrizen näherungsweise wiederum als Matrix gleichen Ranges dargestellt werden. Gleiches folgt für das Produkt von hierarchischen Matrizen. Durch Bilden von Schur–Komplement–Matrizen kann schließlich die Darstellung der Inversen einer hierarchischen Matrix erklärt werden. Die Definition der Niedrig–Rang–Approximationen der Block–Matrizen auf Grundlage der Singulärwertzerlegung erfordert zunächst die Berechnung der gesamten Block–Matrix. Damit ist der Aufwand zur Generierung der hierachischen Matrix quadratisch in der Zahl ihrer Freiheitsgrade und somit nicht optimal. Für die Diskretisierung von Randintegraloperatoren und für die Berechnung der inversen Steifigkeitsmatrix bei finiten Elementen können aber a priori Kriterien für die Berechnung von Niedrig–Rang Approximationen angegeben werden.

Nach dieser Einführung werden im zweiten Kapitel Beispiele für eine effiziente Beschreibung vollbesetzter Matrizen betrachtet. Dies sind zum einen zirkulante Matrizen, deren Anwendung und Invertierung mittels der schnellen Fouriertransformation realisiert werden können. Andererseits, und wesentlich für hierarchische Matrizen, werden Niedrig–Rang–Matrizen bzw. Niedrig–Rang–Störungen regulärer Matrizen behandelt. In Kapitel 3 werden ausgehend von einer Partitionierung der zugeordneten Indexmengen Partitionierungen von Matrizen eingeführt. An einem Beispiel wird gezeigt, daß eine Tensor–Produkt–Partitionierung nicht zu einem optimalen Ergebnis führen kann, so daß die Verwendung einer hierarchisch erzeugten Partitionierung motiviert wird. Im vierten Kapitel wird die Approximierbarkeit vollbesetzter Matrizen durch Niedrig–Rang–Matrizen untersucht. Ausgehend von der Approximation symmetrischer Matrizen über ihre Faktorisierung wird die Singulärwertzerlegung für allgemeinere Matrizen eingeführt. Die Approximationsfehler werden sowohl in der Euklidischen Matrixnorm als auch in der leichter zu berechnenden Frobenius–Norm angegeben. Kapitel 5 widmet sich der Arithmetik hierarchischer Matrizen, d.h. der Addition, Multiplikation und Invertierung von $\mathcal{H}$–Matrizen. Werden die Einträge der zu approximierenden Matrix mit der Funktionsauswertung von Punktmengen identifiziert, so kann die Partitionierung der zugeordneten Indexmenge durch eine geometrische Partitionierung der zugehörigen Punktmenge definiert werden. In Kapitel 6 wird neben einer Clusterung in Boxen ein Bisektionsverfahren betrachtet. In vielen Anwendungen kann die Approximierbarkeit vollbesetzter Matrizen auf die Eigenschaften der die Matrixeinträge erzeugenden Funktion zurückgeführt werden. Neben der Darstellung mit Taylor–Reihen spielt in der Praxis insbesondere die Adaptive Cross Approximationsmethode eine wesent-

liche Rolle. Beide Zugänge werden in Kapitel 7 dargestellt. Abschließend werden im letzten Kapitel 8 hierarchische Matrizen im Zusammenhang mit finiten Elementen besprochen. Wesentlich ist dabei die Approximierbarkeit der inversen FEM Steifigkeitsmatrix.

In dieser Vorlesung wird versucht, hierarchische Matrizen aus einer rein algebraischen Sichtweise einzuführen und zu untersuchen. Bei den Anwendungen hierarchischer Matrizen bei Randelementmethoden und in der Finiten Element Methode gelingt dies nur bedingt, da die zugehörigen Fehlerabschätzungen des Approximationsfehlers in der Regel die Verwendung geeigneter Sobolev–Räume erfordern. Darauf wird hier jedoch bewußt nicht näher eingegangen, sondern auf entsprechende Referenzen verwiesen.

Kapitel 2

Data Sparse Matrizen

In diesem Kapitel sollen anhand von zwei Beispielen vollbesetzte Matrizen angegeben werden, die zu ihrer Beschreibung und somit zu ihrer Abspeicherung einen optimalen Speicherbedarf im Sinne der Forderung (1.2) benötigen. Dies sind zum einen zirkulante Matrizen [7], die allein durch die n Elemente der ersten Zeile bzw. der ersten Spalte beschrieben werden können. Durch die Berechnung der Eigenwerte von zirkulanten Matrizen kann die Matrix–Vektor–Multiplikation mit zirkulanten Matrizen bzw. die Invertierung von zirkulanten Matrizen auf Anwendungen der schnellen Fouriertransformation zurückgeführt werden.
Eine zweite und für hierarchische Matrizen wesentliche Klasse von data sparse Matrizen stellen die Niedrig–Rang–Matrizen dar, die eine optimale Speicherung und Anwendung ermöglichen. Für Niedrig–Rang–Störungen regulärer Matrizen wird hier deren Invertierbarkeit untersucht und die inversen Matrizen werden angegeben.

2.1 Zirkulante Matrizen

Eine Matrix $A \in I\!R^{n\times n}$ mit Elementen $a_{k,\ell}$ heißt **zirkulant**, falls gilt

$$\begin{aligned} a_{k+1,\ell+1} &= a_{k,\ell} && \text{für } k,\ell = 1,\ldots,n-1, \\ a_{k+1,1} &= a_{k,n} && \text{für } k = 1,\ldots,n-1. \end{aligned}$$

Damit gilt für eine zirkulante Matrix A die Darstellung

$$A_{\text{Zirkulant}} = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & a_2 & \cdots & \cdots \\ \cdots & a_{n-1} & a_0 & a_1 & a_2 & \cdots \\ & & \ddots & \ddots & \ddots & \\ a_2 & \cdots & \cdots & a_{n-1} & a_0 & a_1 \\ a_1 & a_2 & \cdots & \cdots & a_{n-1} & a_0 \end{pmatrix},$$

d.h. eine zirkulante Matrix A wird allein durch die n Elemente der ersten Zeile bzw. der ersten Spalte vollständig beschrieben. Damit gilt für den Speicherbedarf einer zirkulanten

Matrix A die Formel (1.2) mit $\alpha = 0$,

$$Sp(A_{\text{Zirkulant}}) \;=\; n\,.$$

Das Rechnen mit zirkulanten Matrizen, d.h. die Multiplikation eines gegebenen Vektors mit einer zirkulanten Matrix bzw. die Invertierung einer zirkulanten Matrix kann bei Kenntnis ihrer Eigenwerte und Eigenvektoren auf die Faktorisierung der Matrix A zurückgeführt werden. Mit der Matrix

$$J \;=\; \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ & & 0 & 1 & 0 & \\ & & & \ddots & \ddots & \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \\ 1 & 0 & \cdots & \cdots & 0 & 0 \end{pmatrix} \in I\!R^{n\times n}$$

und der Vereinbarung $J^0 = I$ erhält man für die zirkulante Matrix A eine Darstellung als Matrixpolynom,

$$A_{\text{Zirkulant}} \;=\; \sum_{\ell=0}^{n-1} a_\ell J^\ell\,. \tag{2.1}$$

Damit können die Eigenwerte und die zugehörigen Eigenvektoren der Matrix A aus denen der Matrix J berechnet werden. Die Eigenwerte der Matrix J ergeben sich aus der Eigenwertgleichung

$$0 \;=\; \det(J-\lambda I) \;=\; \det\begin{pmatrix} -\lambda & 1 & & & \\ & -\lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 1 & & & & -\lambda \end{pmatrix}$$

durch Entwicklung nach der ersten Spalte,

$$\begin{aligned} 0 \;&=\; (-\lambda)\det\begin{pmatrix} -\lambda & 1 & & & \\ & -\lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & -\lambda \end{pmatrix} + (-1)^{n+1}\det\begin{pmatrix} 1 & & & \\ -\lambda & 1 & & \\ & \ddots & \ddots & \\ & & -\lambda & 1 \end{pmatrix} \\ &=\; (-\lambda)^n + (-1)^{n+1} \;=\; (-1)^n\left[\lambda^n - 1\right]. \end{aligned}$$

Damit sind die Eigenwerte der Matrix J gerade die komplexen Einheitswurzeln

$$\lambda_k(J) \;=\; e^{i2\pi k/n} \quad \text{für } k = 0, 1, \ldots, n-1, \tag{2.2}$$

und für die Eigenwerte der zirkulanten Matrix A folgt aus der Polynomdarstellung (2.1)

$$\lambda_k(A_{\text{Zirkulant}}) \;=\; \sum_{\ell=0}^{n-1} a_\ell [\lambda_k(J)]^\ell \;=\; \sum_{\ell=0}^{n-1} a_\ell e^{i2\pi k\ell/n} \quad \text{für } k = 0, \ldots, n-1. \tag{2.3}$$

Für den zum Eigenwert $\lambda_k(J)$ gehörenden Eigenvektor $\underline{e}^k$ folgen aus der Eigenwertgleichung

$$J\underline{e}^k \;=\; \lambda_k(J)\underline{e}^k$$

und der Struktur der zirkulanten Matrix J für die Komponenten e^k_ℓ des Eigenvektors $\underline{e}^k$ die Beziehungen

$$e^k_{\ell+1} = \lambda_k(J) e^k_\ell \quad \text{für } \ell = 0, \ldots, n-2, \quad e^k_0 = \lambda_k(J) e^k_{n-1}.$$

Insbesondere gilt also

$$e^k_\ell = [\lambda_k(J)]^\ell e^k_0 \quad \text{für } \ell = 0, 1, \ldots, n-1.$$

Ohne Einschränkung der Allgemeinheit kann $e^k_0 = 1$ gewählt werden, so daß folgt

$$e^k_\ell = [\lambda_k(J)]^\ell \;=\; e^{i2\pi k\ell/n} \quad \text{für } k, \ell = 0, \ldots, n-1. \tag{2.4}$$

Die aus den Eigenvektoren von J gebildete Matrix

$$F \;=\; \left(\underline{e}^0, \ldots, \underline{e}^{n-1}\right)$$

ist gerade die **Matrix der diskreten Fouriertransformation**. Aufgrund der Polynomdarstellung (2.1) sind die Eigenvektoren von J auch die Eigenvektoren der zirkulanten Matrix A. Für diese folgt somit die Faktorisierung

$$A_{\text{Zirkulant}} \;=\; \frac{1}{n} F^* \Lambda\, F \tag{2.5}$$

mit der zu F adjungierten Matrix F^* und der durch die Eigenwerte von A gegebenen Digonalmatrix

$$\Lambda \;=\; \operatorname{diag}\left(\lambda_k(A)\right)_{k=0}^{n-1}.$$

Damit kann die Multiplikation $A\underline{x}$ einer zirkulanten Matrix A mit einem gegebenen Vektor $\underline{x}$ auf zwei komplexe Fouriertransformationen und eine Multiplikation mit einer Diagonalmatrix zurückgeführt werden. Insbesondere für $n = 2^L$ beträgt der Aufwand der **schnellen Fouriertransformation** $\mathcal{O}(n \ln n)$ wesentliche Operationen, d.h. es gilt Formel (1.3) mit $\beta = 1$,

$$Op(A_{\text{Zirkulant}}\underline{x}) \;=\; \mathcal{O}(n \,\ln n)\,.$$

Aus der Darstellung (2.5) ergibt sich auch sofort die Inverse einer zirkulanten Matrix als

$$A^{-1} \;=\; \frac{1}{n} F^* \Lambda^{-1} F\,.$$

Für eine weiterführende Behandlung zirkulanter Matrizen sei hier auf [7] verwiesen. Zur schnellen Fouriertransformation (FFT) siehe zum Beispiel [18].

2.2 Niedrig–Rang Matrizen

Matrizen $A \in \mathbb{R}^{n\times n}$ mit $\operatorname{rang} A = r \ll n$ können dargestellt werden durch

$$A_{\text{rang } r} \;=\; \sum_{k=1}^{r} \underline{a}_k \underline{b}_k^\top, \qquad \underline{a}_k, \underline{b}_k \in \mathbb{R}^n. \tag{2.6}$$

Zur Beschreibung der Matrix A sind somit die $2r$ Vektoren $\underline{a}_k$ und $\underline{b}_k$ zu speichern, d.h. der Speicherbedarf ist

$$Sp(A_{\text{rang } r}) \;=\; 2\,r\,n.$$

Für die Multiplikation einer Rang r Matrix A mit einem gegeben Vektor $\underline{x} \in \mathbb{R}^n$ ergibt sich

$$A_{\text{rang } r}\underline{x} \;=\; \sum_{k=1}^{r} \underline{a}_k \underline{b}_k^\top \underline{x} \;=\; \sum_{k=1}^{r} \left(\underline{b}_k^\top \underline{x}\right) \underline{a}_k,$$

zu berechnen sind die r Skalarprodukte $\underline{b}_k^\top \underline{x}$, wobei anschließend die Vektoren $\underline{a}_k$ mit diesem skalaren Ergebnis zu multiplizieren sind. Der Aufwand für eine Matrix–Vektor–Multiplikation beträgt also

$$Op(A_{\text{rang } r}\underline{x}) \;=\; 2\,r\,n$$

Multiplikationen. Matrizen A mit $\operatorname{rang} A = r \ll n$ können also effizient gespeichert und angewendet werden. Wegen $\operatorname{rang} A < n$ sind solche Matrizen aber **nicht** invertierbar, deshalb werden im folgenden **Niedrigrangstörungen** regulärer Matrizen B, betrachtet, d.h.

$$A \;=\; B + \sum_{k=1}^{r} \underline{a}_k\, \underline{b}_k^\top\,. \tag{2.7}$$

Es wird vorausgesetzt, daß die Matrix B effizient beschrieben und invertiert werden kann, zum Beispiel sei B eine Diagonalmatrix oder eine zirkulante Matrix. Damit kann dann auch die durch (2.7) gegebene Matrix effizient gespeichert und angewendet werden. Zu untersuchen bleibt die Lösbarkeit des linearen Gleichungssystems $A\underline{x} = \underline{f}$ sowie die Berechnung der inversen Matrix A^{-1} bzw. deren effiziente Anwendung.
Betrachtet wird zunächst der Fall $r = 1$, d.h.

$$A \;=\; B + \underline{a}\,\underline{b}^\top. \tag{2.8}$$

Für die inverse Matrix A^{-1} wird der Ansatz

$$A^{-1} \;=\; B^{-1} + \alpha B^{-1}\underline{a}\,\underline{b}^\top B^{-1}$$

mit einem noch zu wählenden reellen Parameter $\alpha \in \mathbb{R}$ betrachtet. Dann ergibt sich durch Ausmultiplizieren und Ausklammern des Skalars $\underline{b}^\top B^{-1}\underline{a}$

$$\begin{aligned} A^{-1}A &= \left[B^{-1} + \alpha B^{-1}\underline{a}\,\underline{b}^\top B^{-1}\right]\left[B + \underline{a}\,\underline{b}^\top\right] \\ &= I + B^{-1}\underline{a}\,\underline{b}^\top + \alpha B^{-1}\underline{a}\,\underline{b}^\top + \alpha B^{-1}\underline{a}\,\underline{b}^\top B^{-1}\underline{a}\,\underline{b}^\top \\ &= I + \left[1 + \alpha + \alpha\,\underline{b}^\top B^{-1}\underline{a}\right] B^{-1}\underline{a}\,\underline{b}^\top \\ &= I, \end{aligned}$$

falls

$$\alpha \;=\; -\frac{1}{1+\underline{b}^{\top}B^{-1}\underline{a}}$$

gewählt wird. Dabei ist offenbar

$$\underline{b}^{\top}B^{-1}\underline{a} \;\neq\; -1$$

vorauszusetzen. Dann ist die inverse Matrix der durch (2.8) gegebenen Matrix A gegeben durch

$$A^{-1} \;=\; B^{-1} - \frac{1}{1+\underline{b}^{\top}B^{-1}\underline{a}} B^{-1}\underline{a}\,\underline{b}^{\top}B^{-1}. \tag{2.9}$$

Diese Darstellung der inversen Matrix A^{-1} ist als **Sherman–Morrison Formel** bekannt [20]. Für $r = 1$ ergibt sich also die Lösung des linearen Gleichungssystems $A\underline{x} = \underline{f}$ aus

$$\begin{aligned}\underline{x} &= \left[I - \frac{1}{1+\underline{b}^{\top}B^{-1}\underline{a}} B^{-1}\underline{a}\,\underline{b}^{\top}\right] B^{-1}\underline{f} \\ &= \left[I - \frac{1}{1+\underline{b}^{\top}\underline{v}} \underline{v}\,\underline{b}^{\top}\right] B^{-1}\underline{f}\end{aligned}$$

mit $\underline{v} = B^{-1}\underline{a}$, insbesondere ist die inverse Matrix B^{-1} zweimal anzuwenden.
Im allgemeinen Fall $r > 1$ lautet der Ansatz für die inverse Matrix A^{-1}

$$A^{-1} \;=\; B^{-1} + \sum_{k=1}^{r}\sum_{\ell=1}^{r} \alpha_{k\ell} B^{-1}\underline{a}_k\underline{b}_{\ell}^{\top}B^{-1}.$$

Einsetzen ergibt, bei geeigneter Umbezeichnung der Indizes und unter Verwendung des Kronecker–Symbols $\delta_{k\ell} = 1$ für $k = \ell$ bzw. $\delta_{k\ell} = 0$ für $k \neq \ell$,

$$\begin{aligned}A^{-1}A &= \left[B^{-1} + \sum_{k=1}^{r}\sum_{\ell=1}^{r} \alpha_{k\ell} B^{-1}\underline{a}_k\underline{b}_{\ell}^{\top}B^{-1}\right]\left[B + \sum_{i=1}^{r}\underline{a}_i\underline{b}_i^{\top}\right] \\ &= I + \sum_{i=1}^{r} B^{-1}\underline{a}_i\underline{b}_i^{\top} + \sum_{k=1}^{r}\sum_{\ell=1}^{r}\alpha_{k\ell}B^{-1}\underline{a}_k\underline{b}_{\ell}^{\top} + \sum_{k=1}^{r}\sum_{\ell=1}^{r}\sum_{i=1}^{r}\alpha_{k\ell}B^{-1}\underline{a}_k\underline{b}_{\ell}^{\top}B^{-1}\underline{a}_i\underline{b}_i^{\top} \\ &= I + \sum_{k=1}^{r}\sum_{\ell=1}^{r}\left[\delta_{k\ell} + \alpha_{k\ell} + \sum_{i=1}^{r}\alpha_{ki}\,\underline{b}_i^{\top}B^{-1}\underline{a}_{\ell}\right] B^{-1}\underline{a}_k\underline{b}_{\ell}^{\top} \\ &= I,\end{aligned}$$

falls

$$\delta_{k\ell} + \alpha_{k\ell} + \sum_{i=1}^{r}\alpha_{ki}\,\underline{b}_i^{\top}B^{-1}\underline{a}_{\ell} \;=\; 0 \quad \text{für alle } k,\ell = 1,\ldots,r \tag{2.10}$$

erfüllt ist. Ist das lineare Gleichungssystem (2.10) eindeutig lösbar, so ergeben sich daraus die r^2 Koeffizienten $\alpha_{k\ell}$ und somit die Darstellung der inversen Matrix A^{-1}. Diese Darstellung ist allgemein als **Sherman–Morrison–Woodbury–Formel** bekannt [20].

Übungsaufgaben

2.1. Gegeben sei die Matrix

$$A = I + \underline{a}\,\underline{a}^\top + \varepsilon \underline{b}\,\underline{b}^\top \in \mathbb{R}^{3\times 3}, \quad \underline{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \underline{b} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

- Man berechne die Eigenwerte der Matrix A !
- Für die Approximation $\widetilde{A} = I + \underline{a}\,\underline{a}^\top$ gebe man den Fehler $\|A - \widetilde{A}\|$ in der Spektralnorm $\|\cdot\|_2$ und in der Frobeniusnorm $\|\cdot\|_F$ an. Dabei sind

$$\|B\|_2 = \max_{k=1,2,3} \sqrt{\lambda_k(B^\top B)}, \qquad \|B\|_F = \left[\sum_{k,\ell=1}^{3} (B[\ell,k])^2 \right]^{1/2}.$$

- Man bestimme die näherungsweise Inverse $\widetilde{A}^{-1}$.

2.2. Für eine gegebene Vektornorm $\|\cdot\|_V$ wird durch

$$\|A\|_M := \sup_{\underline{0} \neq \underline{x} \in \mathbb{R}^n} \frac{\|A\underline{x}\|_V}{\|\underline{x}\|_V}$$

eine induzierte Matrixnorm definiert.
Man zeige, daß die Frobeniusnorm $\|\cdot\|_F$ durch keine Vektornorm induziert wird, aber verträglich zur Euklidischen Vektornorm $\|\cdot\|_2$ ist, d.h. es gilt

$$\|A\underline{x}\|_2 \leq \|A\|_F \|\underline{x}\|_2 \quad \text{für alle } \underline{x} \in \mathbb{R}^n.$$

2.3. Für eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ beweise man die Spektraläquivalenzungleichungen

$$c_1 \|A\|_2 \leq \|A\|_F \leq c_2 \|A\|_2$$

mit geeignet gewählten Konstanten c_i unabhängig von $A \in \mathbb{R}^{n\times n}$.

Kapitel 3

Partitionierte Matrizen

Hierarchische Matrizen basieren auf einer geeigneten **Partitionierung** der Matrix A in **Block–Matrizen** A_{ij} und einer **Niedrig–Rang–Darstellung** dieser Blockmatrizen A_{ij}. Ein beliebiges Element $A[\ell, k]$ der Matrix $A \in I\!R^{n\times n}$ kann mit einem Element $A_{ij}[\ell_j, k_i]$ einer Blockmatrix $A_{ij} \in I\!R^{n_j\times n_i}$ identifiziert werden, wenn eine eindeutige Zuordnung der Indizees $k_i \leftrightarrow k$ und $\ell_i \leftrightarrow \ell$ vorliegt. Für $k_i = 1, \ldots, n_i$ bzw. $\ell_j = 1, \ldots, n_j$ durchlaufen die zugehörigen Indizees k bzw. ℓ gewisse Indexbereiche I_i und I_j als Teilmengen der ursprünglichen **Indexmenge**

$$I \;=\; \{1, 2, 3, \ldots, n-1, n\}\,. \tag{3.1}$$

Damit kann die Blockmatrix $A_{ij} \in I\!R^{n_j\times n_i}$ stets durch Indexmengen $I_i, I_j \subset I$ beschrieben werden. Die **Partitionierung** einer Matrix A entspricht somit einer Partitionierung der durch (3.1) gegebenen Indexmenge.
Für $p \in I\!N$ sei

$$P(I) \;=\; \{I_i\}_{i=1}^{p} \tag{3.2}$$

eine **Partitionierung** der Indexmenge I in **paarweise zueinander disjunkte** Indexmengen I_i der Dimension

$$n_i \;=\; \dim I_i$$

mit

$$I_i \cap I_j = \emptyset \quad \text{für } i \neq j.$$

Die Partitionierung (3.2) stelle eine **vollständige Zerlegung** der Indexmenge I dar, d.h. es gelte

$$I \;=\; \bigcup_{i=1}^{p} I_i, \qquad \sum_{i=1}^{p} n_i = n\,.$$

Ohne Einschränkung der Allgemeinheit kann vorausgesetzt werden, daß die Indexmengen I_i **zusammenhängende** Indizes umfaßt, d.h.

$$I_i \;=\; \left\{1 + \sum_{\nu=1}^{i-1} n_\nu, \ldots, \sum_{\nu=1}^{i} n_\nu\right\} \quad \text{für } i = 1, \ldots, p.$$

Eine solche Darstellung kann durch eine geeignete **Permutation** der Indizees stets erreicht werden.
Die Partitionierung der Indexmenge I induziert nun eine **Tensor–Produkt–Partitionierung** der Indexmenge $I \times I$ durch

$$P_2(I) \;=\; P(I) \times P(I) \;=\; \{I_j \times I_i \;:\; I_i, I_j \in P(I)\}\,. \tag{3.3}$$

Für $k \in I_i$ und $\ell \in I_j$ werden durch

$$A_{ij}[\ell_j, k_i] \;=\; A[\ell, k]$$

zugehörige Block–Matrizen

$$A_{ij} \in I\!R^{n_j \times n_i}$$

erklärt, vergleiche Abbildung 3.1 für die resultierende Partitionierung der Matrix A.

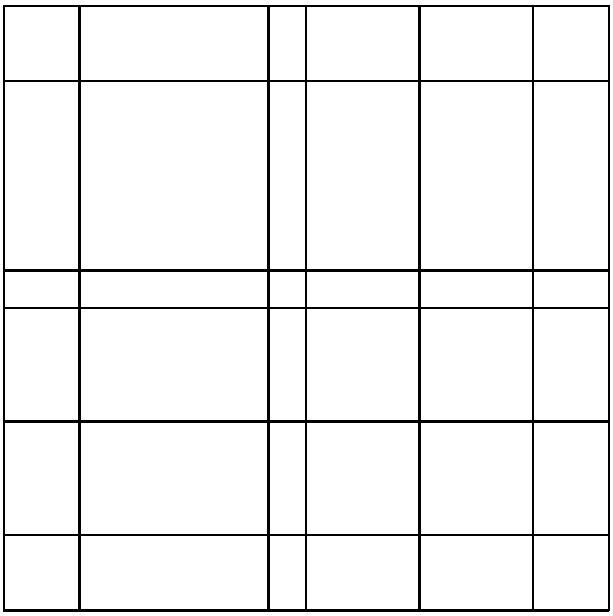


Abbildung 3.1: Tensor–Produkt–Partitionierung einer Matrix A.

Die Vorgehensweise zur Definition einer partitionierten Matrix mittels eines Tensor–Produkt–Ansatzes soll an einem einfachen Beispiel näher untersucht werden. Hierzu werde die folgende Situation betrachtet:

- Die Diagonal–Block–Matrizen $A_{ii} \in I\!R^{n_i \times n_i}$ können nur durch Matrizen vollen Ranges n_i dargestellt werden. Zur Beschreibung der Matrizen A_{ii} sind somit jeweils n_i^2 Speicherplätze erforderlich.

- Für $i \neq j$ sind die Blockmatrizen A_{ij} vom Rang 1, d.h. die Darstellung

  $$A_{ij} \;=\; \underline{a}_j \underline{b}_i^{\top}$$

  mit Vektoren $\underline{a}_j \in I\!R^{n_j}$ und $\underline{b}_i \in I\!R^{n_i}$ erfordert $n_i + n_j$ Speicherplätze.

Bestimmt werden soll der erforderliche Speicherbedarf $Sp(A)$ zur Beschreibung der Matrix A. Dieser ergibt sich aus der Summe des Speicherbedarfs $Sp(A_{ii})$ für alle vollbesetzten Diagonalblöcke A_{ii} und des Speicherbedarfs $Sp(A_{ij})$ der Rang 1 Matrizen A_{ij} der Nebendiagonalblöcke für $i \neq j$,

$$Sp(A) \;=\; \sum_{i=1}^{p} Sp(A_{ii}) + \sum_{i,j=1,i\neq j}^{p} Sp(A_{ij}) \;=\; \sum_{i=1}^{p} n_i^2 + \sum_{i,j=1,i\neq j}^{p} [n_i + n_j].$$

Für den letzten Summanden ergibt sich

$$\sum_{i,j=1,i\neq j}^{p} [n_i + n_j] \;=\; 2\sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p} n_i \;=\; 2(p-1)\sum_{i=1}^{p} n_i \;=\; 2(p-1)n\,.$$

Mit der Cauchy–Schwarz–Ungleichung ist

$$n \;=\; \sum_{i=1}^{p} n_i \;\leq\; \left(\sum_{i=1}^{p} 1^2\right)^{1/2} \left(\sum_{i=1}^{p} n_i^2\right)^{1/2} \;=\; \sqrt{p}\left(\sum_{i=1}^{p} n_i^2\right)^{1/2},$$

und somit folgt

$$\sum_{i=1}^{p} n_i^2 \;\geq\; \frac{1}{p}\,n^2\,.$$

Insgesamt gilt also

$$Sp(A) \;\geq\; \frac{1}{p}\,n^2 + 2(p-1)n\,.$$

Diese untere Schranke ist minimal für

$$p \;=\; \sqrt{n/2},$$

und somit ergibt sich für den erforderlichen Speicherbedarf der partitionierten Matrix A die Abschätzung

$$Sp(A) \;\geq\; \sqrt{2}\,n^{3/2} + (\sqrt{2}n^{1/2} - 2)n \;=\; \mathcal{O}(n^{3/2})\,.$$

Damit ist der erforderliche Speicherbedarf $Sp(A)$ nicht optimal im Sinne der Forderung (1.2), und die Verwendung eines Tensor–Produkt–Ansatzes zur Partitionierung einer Matrix führt in der Regel nicht zum gewünschten Ergebnis.

Der Ausweg besteht nun in einer **hierarchischen Partitionierung** der Indexmenge I. Ausgehend von der Partitionierung

$$P^0(I) \;=\; \left\{I_i^0\right\}_{i=1}^{p_0} \;=\; \{I\} \qquad (p_0 = 1)$$

werden für $\lambda = 0, 1, \ldots, L$ **rekursiv** Partitionierungen

$$P^\lambda(I) \;=\; \left\{I_i^\lambda\right\}_{i=1}^{p_\lambda} \tag{3.4}$$

der Indexmenge I in **paarweise zueinander disjunkte** Indexmengen I_i^λ gleicher Stufe der Dimension

$$n_i^\lambda \;=\; \dim I_i^\lambda$$

mit

$$I_i^\lambda \cap I_j^\lambda = \emptyset \quad \text{für } i \neq j$$

erklärt. Für jedes Level λ bilden diese eine **vollständige Zerlegung** der Indexmenge I,

$$I \;=\; \bigcup_{i=1}^{p_\lambda} I_i^\lambda, \qquad \sum_{i=1}^{p_\lambda} n_i^\lambda \;=\; n.$$

Insbesondere bilden die Indexmengen I_i^λ eine **Hierarchie**, d.h. für $i = 1, \ldots, p_\lambda$ und $\lambda = 1, \ldots, L$ existiert genau eine Indexmenge $I_j^{\lambda-1}$ mit

$$I_i^\lambda \;\subset\; I_j^{\lambda-1}.$$

Ohne Einschränkung der Allgemeinheit kann wieder vorausgesetzt werden, daß die Indexmengen I_i^λ **zusammenhängende Indizes** umfassen. Diese Voraussetzung kann stets durch geeignete **Permutationen** gewährleistet werden. Die Indexmengen I_i^λ bilden somit einen **Baum** T, siehe Abbildung 3.2.

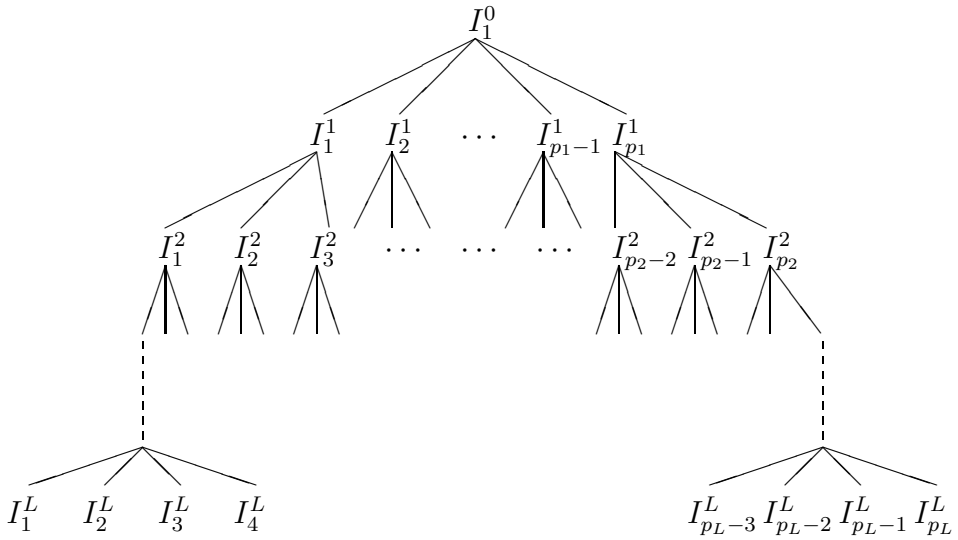


Abbildung 3.2: Baum T der Indexmengen I_i^λ.

Durch die hierarchische Partitionierung der Indexmenge I kann durch

$$P_{\mathcal{H}}(I,T) \;=\; \left\{ I_i^\lambda \times I_j^\lambda \;:\; I_i^\lambda, I_j^\lambda \in P^\lambda(I), \lambda = 0, \ldots, L \right\} \tag{3.5}$$

eine **hierarchische Partitionierung** der Indexmenge $I \times I$ erklärt werden.

Beispiel 3.1 *Für eine gegebene Indexmenge $I = I_1^0$ liege eine Unterteilung in zwei Indexmengen I_1^1 und I_2^1 vor. Dann ist*

$$P_{\mathcal{H}}(I,T) = \left\{ I_1^0 \times I_1^0, I_1^1 \times I_1^1, I_2^1 \times I_2^1, I_1^1 \times I_2^1, I_2^1 \times I_1^1 \right\}.$$

Diese induziert zugehörige Partitionierungen der Matrix A,

$$A = \left(A_{11}^0 \right), \quad A = \begin{pmatrix} A_{11}^1 & A_{12}^1 \\ A_{21}^1 & A_{22}^1 \end{pmatrix}$$

Bemerkung 3.1 *Bei der Definition* (3.5) *der hierarchischen Partitionierung $P_{\mathcal{H}}(I,T)$ der Indexmenge $I \times I$ werden hier nur Indexpaare $I_i^\lambda \times I_j^\lambda$ des gleichen Levels λ betrachtet. Allgemein können auch Indexpaare benachbarter Level betrachtet werden, zum Beispiel $I_i^\lambda \times I_j^{\lambda \pm 1}$.*

Bemerkung 3.2 *Beispiel* 3.1 *zeigt, daß ein Element $A[\ell, k]$ der Ausgangsmatrix A als Element $A_{ij}^\lambda[\ell_j, k_i]$ verschiedener Block–Matrizen A_{ij}^λ aufgefaßt werden kann. So gehört zum Beispiel das Element $A[1,1]$ zu allen Block–Matrizen A_{11}^λ für alle $\lambda = 0, 1, \ldots, L$. Damit ist für eine Beschreibung der Matrix A als eine hierarchische Block–Matrix ein* **Kriterium** *anzugeben, welche Block–Matrizen A_{ij}^λ zu verwenden sind.*

Zu finden ist also ein Kriterium für die Verwendung der Block–Matrix A_{ij}^λ anstelle der Gesamtheit aller Block–Matrizen $A_{k\ell}^\kappa$, $\kappa > \lambda$, welche durch alle Söhne I_k^κ und I_ℓ^κ der Indexmengen I_i^λ und I_j^λ erzeugt werden.

Definition 3.1 *Eine Block–Matrix A_{ij}^λ bzw. die sie erzeugenden Indexmengen I_i^λ und I_j^λ heißen zueinander* **r–zulässig**, *falls die Block–Matrix A_{ij}^λ eine Darstellung als Matrix vom Rang r ermöglicht.*

Block–Matrizen A_{ij}^L auf dem feinsten Level L, die **keine** Rang r Darstellung erlauben, werden als Matrix $A_{ij}^L \in I\!R^{n_i^L \times n_j^L}$ mit $n_i^L n_j^L$ Einträgen beschrieben. In Anlehnung an die Anwendungen werden diese Blöcke als **Nahfeld** der Matrix A bezeichnet.
Die Gesamtheit der das Nahfeld beschreibenden Indexpaare und alle zueinander zulässigen Indexpaare maximal möglicher Grösse wird mit

$$P_{\mathcal{H}}^Z(I,T) \subset P_{\mathcal{H}}(I,T) \tag{3.6}$$

bezeichnet. Für jedes Indexpaar $(k,\ell) \in I \times I$ **existiert genau ein** Paar von Indexmengen $(I_i^\lambda, I_j^\lambda) \in P_{\mathcal{H}}^Z(I,T)$ mit $(k,\ell) \in I_i^\lambda \times I_j^\lambda$.

Es kann nun ein **Algorithmus** angegeben werden, der die Erzeugung einer hierarchisch partitionierten Matrix A ermöglicht.

Algorithmus zur Erzeugung einer hierarchisch partitionierten Matrix

Starte mit dem gröbsten Level $\lambda = 0$.

1. Teste die Block–Matrizen A_{ij}^{λ} auf ihre r–Zulässigkeit.
2. Ist die Block–Matrix A_{ij}^{λ} r–zulässig, so wird dieser Block durch die entsprechende Rang r Darstellung beschrieben.
3. Ist die Block–Matrix A_{ij}^{λ} **nicht** r–zulässig, so werden zwei Fälle unterschieden:
   - 3.1. Für $\lambda = L$ ist das feinste Level erreicht und die Block–Matrix A_{ij}^{L} kann nicht durch eine Rang r Matrix dargestellt werden. Deshalb ist für diese Blöcke im Nahfeld die übliche Darstellung einer vollbesetzten Matrix zu verwenden.
   - 3.2. Für $\lambda < L$ ist das feinste Level noch nicht erreicht. Deshalb können für alle Söhne $I_k^{\lambda+1}$ und $I_\ell^{\lambda+1}$ der Indexmengen I_i^{λ} und I_j^{λ} die zugehörigen Block–Matrizen $A_{k\ell}^{\lambda+1}$ gebildet werden. Diese können dann gemäß Schritt 1. auf ihre r–Zulässigkeit untersucht werden.

Der hier beschriebene Algorithmus zur Erzeugung einer hierarchisch partitionierten Matrix A beruht im wesentlichen auf dem in Abbildung 3.2 dargestellten Baum T der Indexmengen I_i^{λ} und der dafür zugrunde liegenden hierarchischen Partitionierung $P_{\mathcal{H}}(I, T)$ der Indexmenge I. Diese erfolgt in der Regel **problemabhängig** und ist somit abhängig von der Berechnungsvorschrift der Matrixelemente $A[\ell, k]$. In den hier betrachteten Anwendungen können die Matrixeinträge $A[\ell, k]$ in Relation gesetzt werden zu geometrischen Punkten $x_k, x_\ell \in \mathbb{R}^d$. Damit kann die Partitionierung der Indexmenge I zurückgeführt werden auf eine **geometrische Partitionierung** von Punktwolken $\{x_k\}_{k=1}^{n} \subset \mathbb{R}^d$, siehe hierzu auch Kapitel 6.
Die in Definition 3.1 angegebene Bedingung der r–Zulässigkeit kann zum Beispiel **algebraisch** durch eine **Singulärwertzerlegung** der Block–Matrix $A^{ij,\lambda}$ überprüft werden. Dies erfordert aber das explizite Aufstellen der Block–Matrix $A^{ij,\lambda}$ und somit die Berechnung **aller** Matrix–Elemente $A[\ell, k]$ der ursprünglichen Matrix A. Damit führt dieses Konzept zwar auf eine im Sinne des Speicherbedarfs optimale Beschreibung der Matrix A, erfordert aber einen in der Zahl n der Freiheitsgrade quadratischen Aufwand zur Generierung dieser Beschreibung. Gesucht sind deshalb **a priori** Kriterien, die eine Generierung der vollständigen Matrix A vermeiden. Diese können wiederum nur **problemabhängig** angegeben werden.

An einem einfachen Beispiel soll abschließend gezeigt werden, wie mit einer hierarchisch partitionierten Matrix und der Niedrig–Rang–Darstellung der Block–Matrizen ein optimaler Speicherbedarf erreicht werden kann.

Beispiel 3.2 *Sei $A \in \mathbb{R}^{n\times n}$ mit $n = 2^L$. Durch Bisektion, d.h. Halbierung, werde die Indexmenge $I_1^0 = \{1,\ldots,n\}$ hierarchisch partitioniert,*

$$I_k^{\lambda-1} \;=\; I_{2k-1}^{\lambda} \cup I_{2k}^{\lambda}, \quad k = 1,\ldots,2^{\lambda-1}, \lambda = 1,\ldots,L.$$

Es wird angenommen, daß die Indexmengen I_{2k-1}^{λ} und I_{2k}^{λ} zueinander zulässig sind, und daß die zugehörige Block–Matrix $A_{2k-1,2k}^{\lambda}$ durch eine Rang r Matrix beschrieben werden kann, während die Diagonalblöcke $A_{2k-1,2k-1}^{\lambda}$ und $A_{2k,2k}^{\lambda}$ rekursiv definiert sind. Nach Konstruktion gilt $A_{ij}^{\lambda} \in \mathbb{R}^{2^{L-\lambda}\times 2^{L-\lambda}}$ für $\lambda = 0,\ldots,L$. Die resultierende hierarchische Partitionierung der Matrix A ist in Abbildung 3.3 *dargestellt.*

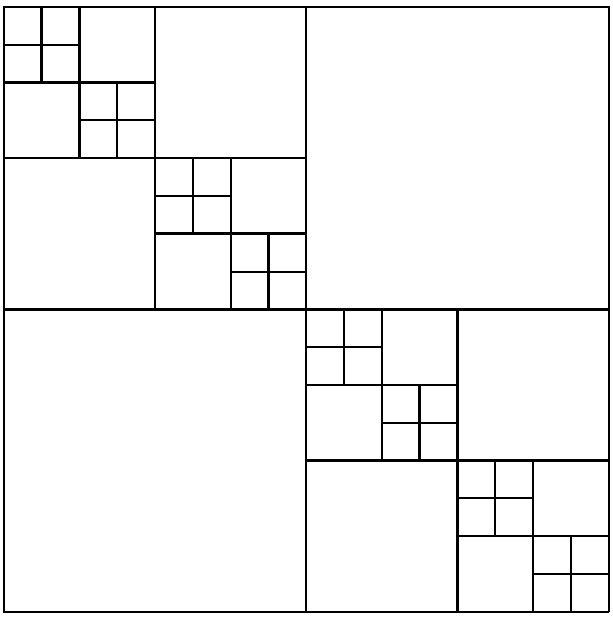


Abbildung 3.3: Hierarchisch partitionierte Matrix A.

Offenbar kann die in Abbildung 3.3 *angegebene Matrix rekursiv durch*

$$A_{kk}^{\lambda-1} \;=\; \begin{pmatrix} A_{2k-1,2k-1}^{\lambda} & A_{2k-1,2k}^{\lambda} \\ A_{2k,2k-1}^{\lambda} & A_{2k,2k}^{\lambda} \end{pmatrix}$$

für $k = 1,\ldots,2^{\lambda}-1$ und $\lambda = 1,\ldots,L$. Dabei sei für $\lambda = 1$ $A_{11}^{0} = A$.
Nach Voraussetzung sind die Nebendiagoanlmatrizen $A_{2k-1,2k}^{\lambda}$ (bzw. $A_{2k,2k-1}^{\lambda}$) Rang r Matrizen. Für den Speicherbedarf von $A_{kk}^{\lambda-1}$ ergibt sich dann

$$\begin{aligned} Sp(A_{kk}^{\lambda-1}) &= 2\,Sp(A_{ii}^{\lambda}) + 2\,Sp(A_{ij}^{\lambda}) \\ &= 2\,Sp(A_{ii}^{\lambda}) + 2\,r\,[2^{L-\lambda} + 2^{L-\lambda}] \\ &= 2\,Sp(A_{ii}^{\lambda}) + 4\,r\,2^{L-\lambda}. \end{aligned}$$

Für $\lambda = 1$ ist also

$$Sp(A) \;=\; Sp(A_{11}^0) \;=\; 2\,Sp(A_{ii}^1) + 4\,r\,2^{L-1} \;=\; 2\,Sp(A_{ii}^1) + 2\,r\,2^L$$

und durch rekursives Einsetzen folgt

$$\begin{aligned} Sp(A) &= 2\,Sp(A_{ii}^1) + 2\,r\,2^L \\ &= 2\left[2\,Sp(A_{ii}^2) + 4\,r\,2^{L-2}\right] + 2\,r\,2^L \\ &= 2^2\,Sp(A_{ii}^2) + 2\cdot 2\,r\,2^L \\ &= 2^\lambda\,Sp(A_{ii}^\lambda) + \lambda\,2\,r\,2^L \end{aligned}$$

für $\lambda = 1, \ldots, L$. Insbesondere für $\lambda = L$ ist $A_{ii}^L \in \mathbb{R}$ und somit folgt

$$Sp(A) \;=\; 2^L + L\,2\,r\,2^L \;=\; 2^L(1 + 2rL) \;=\; n(1 + 2r\log_2 n).$$

Mit Hilfe dieses Beispiels soll noch einer anderen Frage nachgegangen werden. Für Block–Matrizen $A_{ii}^\lambda \in \mathbb{R}^{n^\lambda \times n^\lambda}$ welcher Dimension $n^\lambda = 2^{L-\lambda}$ ist die Beschreibung als Rang r Matrix sinnvoll, oder wann ist die Beschreibung als vollbesetzte Matrix zu bevorzugen. Der jeweils notwendige Speicherbedarf für beide Fälle ist in Tabelle 3.1 angegeben. Es zeigt sich, daß Blockmatrizen kleiner Dimension stets als vollbesetzte Matrizen vorteilhafter zu beschreiben sind.

λ	n^λ	$Sp(A_{ii}^\lambda) = [n^\lambda]^2$	$Sp(A_{ii,r}^\lambda) = 2rn^\lambda$	$Sp(A_{ii,r}^\lambda) < Sp(A_{ii}^\lambda)$
L	1	1	2r	
$L-1$	2	4	4r	
$L-2$	4	16	8r	$r < 2$
$L-3$	8	64	16r	$r < 4$
$L-4$	16	256	32r	$r < 8$

Tabelle 3.1: Vergleich Speicherbedarf für vollbesetzte und Rang r Matrix.

Übungsaufgaben

3.1. Gegeben seien die eindimensionalen Punktmengen $\{x_k\}_{k=1}^n$ und $\{y_k\}_{k=1}^n$ sowie die zugehörige Matrix $A \in \mathbb{R}^{n\times n}$ definiert durch

$$A[\ell, k] \;=\; f(x_k, y_\ell).$$

Dabei sei

$$f(x,y) \;=\; \sum_{\alpha_1+\alpha_2\le p} a_\alpha x^{\alpha_1} y^{\alpha_2}$$

mit nichtnegativen Indizees α_i. Man bestimme in Abhängigkeit von p den Rang der Matrix A sowie den Speicherbedarf zur Beschreibung von A.

Kapitel 4

Approximation mit Niedrig–Rang–Matrizen

Eine Block–Matrix $A_{ij}^{\lambda} \in I\!R^{n_j^{\lambda} \times n_i^{\lambda}}$ heißt (r, ε)–**zulässig**, wenn sie durch eine Matrix $A_{ij,r}^{\lambda}$ gleicher Dimension mit $\operatorname{rang} A_{ij,r}^{\lambda} \leq r$ und mit einer vorgegebenen Genauigkeit ε approximiert werden kann. In diesem Kapitel soll untersucht werden, wie dieses Kriterium überprüft und wie eine solche Niedrig–Rang–Darstellung berechnet werden kann. Bevor der allgemeine Fall einer Matrix $A_{ij}^{\lambda} \in I\!R^{n_j^{\lambda} \times n_i^{\lambda}}$ mit $n_i^{\lambda} \neq n_j^{\lambda}$ betrachtet wird, soll zunächst der einfachere Fall einer symmetrischen Matrix $A_{ii}^{\lambda} \in I\!R^{n_i \times n_i}$ untersucht werden.

4.1 Approximation symmetrischer Matrizen

Gegeben sei eine **symmetrische** Matrix $A = A^{\top} \in I\!R^{n \times n}$ mit n **nichtnegativen** reellen **Eigenwerten**

$$\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A) \geq 0.$$

Die zugehörigen Eigenvektoren $\{\underline{v}_k\}_{k=1}^n$ bilden ein **Orthonormalsystem** bezüglich dem **Euklidischen Skalarprodukt**,

$$(\underline{v}_k, \underline{v}_\ell) = \delta_{k\ell} = \begin{cases} 1 & \text{für } k = \ell, \\ 0 & \text{für } k \neq \ell. \end{cases}$$

Die durch die Eigenvektoren von A gebildete Matrix

$$V = (\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n) \in I\!R^{n \times n}$$

ist **orthogonal**,

$$V^{\top} V = V V^{\top} = I\,.$$

Dann folgt durch Multiplikation von V mit der Matrix A

$$AV = (A\underline{v}_1, \ldots, A\underline{v}_n) = (\lambda_1(A)\underline{v}_1, \ldots, \lambda_n(A)\underline{v}_n) = VD$$

mit der durch die Eigenwerte von A definierten Diagonalmatrix

$$D \;=\; \mathrm{diag}\,(\lambda_k(A))_{k=1}^{n}\,.$$

Multiplikation mit $V^\top$ von links liefert

$$V^\top A V \;=\; D$$

bzw. ergibt die Multiplikation mit $V^\top$ von rechts die bekannte Faktorisierung

$$A \;=\; V D V^\top. \tag{4.1}$$

Gesucht ist nun eine symmetrische Matrix $A_r = A_r^\top \in I\!R^{n\times n}$ mit

$$\mathrm{rang}\, A_r \;\le\; r < n$$

als Lösung des **Minimierungsproblems**

$$\|A - A_r\|_2 \;=\; \min_{B=B^\top\in I\!R^{n\times n},\mathrm{rang}\, B\le r} \|A-B\|_2\,. \tag{4.2}$$

Die **Euklidische Matrixnorm** $\|A\|_2$ wird durch die **Euklidische Vektornorm** $\|\underline{x}\|_2$ induziert,

$$\|A\|_2 \;:=\; \sup_{\underline{0}\neq\underline{x}\in I\!R^n} \frac{\|A\underline{x}\|_2}{\|\underline{x}\|_2}.$$

Andererseits definiert

$$\varrho(A) \;:=\; \max_{i=1,\ldots,n} |\lambda_i(A)|$$

den **Spektralradius** der Matrix A und es gilt

$$\|A\|_2^2 \;=\; \varrho(A^\top A).$$

Insbesondere für eine symmetrische Matrix $A = A^\top$ folgt also die Gleichheit

$$\|A\|_2 \;=\; \varrho(A)\,.$$

Da der Rang einer Matrix A der maximalen Anzahl von linear unabhängigen Spalten bzw. Zeilen von A entspricht, wird dieser wegen

$$A \;=\; V D V^\top \;=\; \sum_{k=1}^{n} \lambda_k \underline{v}_k \underline{v}_k^\top$$

durch die Anzahl der nicht verschwindenden Eigenwerte $\lambda_k(A)$ bestimmt.
Die gesuchte Rang r Approximation A_r von A besitzt also maximal r nicht verschwindende Eigenwerte $\lambda_k(A_r)$. Ausgehend von der Faktorisierung $A = VDV^\top$ wird deshalb die Approximation

$$A_r \;=\; V D_r V^\top \;=\; \sum_{k=1}^{r} \lambda_k(A) \underline{v}_k \underline{v}_k^\top \tag{4.3}$$

mit der Diagonalmatrix

$$D_r = \begin{pmatrix} \lambda_1(A) & & & & & \\ & \ddots & & & & \\ & & \lambda_r(A) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

definiert, und es gilt:

Satz 4.1 *Die durch* (4.3) *definierte Rang r Matrix A_r ist Lösung der Minimierungsaufgabe* (4.2)*, und es gilt*

$$\min_{B=B^\top \in \mathbb{R}^{n\times n}, rang B \le r} \|A - B\|_2 = \|A - A_r\|_2 = \lambda_{r+1}(A).$$

Beweis: Mit

$$\|A - A_r\|_2 = \varrho(A - A_r) = \max_{k=1,\dots,n} |\lambda_k(A - A_r)| = \max_{k=r+1,\dots,n} |\lambda_k(A)| = \lambda_{r+1}(A)$$

gilt zunächst

$$\min_{B=B^\top \in \mathbb{R}^{n\times n}, \operatorname{rang} B \le r} \|A - B\|_2 \le \|A - A_r\|_2 = \lambda_{r+1}(A)\,.$$

Sei nun $B = B^\top \in \mathbb{R}^{n\times n}$ eine beliebige symmetrische Matrix mit $\operatorname{rang} B \le r$. Dann existiert ein Orthonormalsystem $\{\underline{z}_k\}_{k=1}^n$ von Eigenvektoren $\underline{z}_k \in \mathbb{R}^n$ mit

$$B\underline{z}_k = \lambda_k(B)\underline{z}_k \quad \text{für } k = 1, \dots, n.$$

Ohne Einschränkung der Allgemeinheit gelte für die Eigenwerte $\lambda_k(B)$ von B

$$\lambda_k(B) = 0 \quad \text{für } k = r+1, \dots, n.$$

Für alle Vektoren

$$\underline{u} = \sum_{k=r+1}^{n} u_k \underline{z}_k \in \operatorname{span}\{\underline{z}_{r+1}, \dots, \underline{z}_n\}$$

folgt also

$$B\underline{u} = \sum_{k=r+1}^{n} u_k B\underline{z}_k = \sum_{k=r+1}^{n} u_k \lambda_k(B)\underline{z}_k = \underline{0}.$$

Unter Benutzung der Eigenvektoren $\underline{v}_i$ von A existiert andererseits ein normierter Vektor

$$\bar{\underline{u}} \in \underbrace{\operatorname{span}\{\underline{z}_{r+1}, \dots, \underline{z}_n\}}_{\dim = n-r} \cap \underbrace{\operatorname{span}\{\underline{v}_1, \dots, \underline{v}_r, \underline{v}_{r+1}\}}_{\dim = r+1}$$

mit $\|\bar{\underline{u}}\|_2 = 1$, und es folgt

$$\|A - B\|_2 \;=\; \sup_{\underline{0}\neq\underline{x}\in I\!R^n} \frac{\|(A-B)\underline{x}\|_2}{\|\underline{x}\|_2} \;\geq\; \frac{\|(A-B)\bar{\underline{u}}\|_2}{\|\bar{\underline{u}}\|_2} \;=\; \|A\bar{\underline{u}}\|_2\,.$$

Nach Konstruktion gilt für $\bar{\underline{u}}$ eine Darstellung der Form

$$\bar{\underline{u}} \;=\; \sum_{k=1}^{r+1} \bar{u}_k \underline{v}_k$$

mit Zerlegungskoeffizienten $\bar{u}_k = (\bar{\underline{u}}, \underline{v}_k)$. Dann gilt

$$1 \;=\; \|\bar{\underline{u}}\|_2^2 \;=\; (\bar{\underline{u}}, \bar{\underline{u}}) \;=\; (\sum_{k=1}^{r+1} \bar{u}_k \underline{v}_k, \sum_{\ell=1}^{r+1} \bar{u}_\ell \underline{v}_\ell) \;=\; \sum_{k=1}^{r+1}\sum_{\ell=1}^{r+1} \bar{u}_k \bar{u}_\ell (\underline{v}_k, \underline{v}_\ell) \;=\; \sum_{k=1}^{r+1} \bar{u}_k^2.$$

Daraus folgt, unter Verwendung der Orthonormalität der Eigenvektoren $\{\underline{v}_k\}_{k=1}^n$,

$$\begin{aligned}
\|A\bar{\underline{u}}\|_2^2 \;&=\; (A\bar{\underline{u}}, A\bar{\underline{u}}) \;=\; (A\sum_{k=1}^{r+1} \bar{u}_k \underline{v}_k, A\sum_{\ell=1}^{r+1} \bar{u}_\ell \underline{v}_\ell) \;=\; \sum_{k=1}^{r+1}\sum_{\ell=1}^{r+1} \bar{u}_k \bar{u}_\ell (A\underline{v}_k, A\underline{v}_\ell) \\
&=\; \sum_{k=1}^{r+1}\sum_{\ell=1}^{r+1} \bar{u}_k \bar{u}_\ell \lambda_k(A)\lambda_\ell(A)(\underline{v}_k, \underline{v}_\ell) \;=\; \sum_{k=1}^{r+1} \bar{u}_k^2 [\lambda_k(A)]^2 \\
&\geq\; \min_{k=1,\ldots,r+1} [\lambda_k(A)]^2 \sum_{k=1}^{r+1} \bar{u}_k^2 \;=\; [\lambda_{r+1}(A)]^2
\end{aligned}$$

und somit

$$\|A - B\|_2 \;\geq\; \lambda_{r+1}(A)$$

für eine beliebige Matrix $B = B^\top \in I\!R^{n\times n}$ mit $\operatorname{rang} B \leq r$. Insgesamt gilt also

$$\lambda_{r+1}(A) \;\leq\; \min_{B=B^\top\in I\!R^{n\times n}, \operatorname{rang} B \,\leq\, r} \|A - B\|_2 \;\leq\; \|A - A_r\|_2 \;\leq\; \lambda_{r+1}(A)$$

und folglich die Gleichheit. ∎

Neben dem Minimierungsproblem (4.2) in der Euklidischen Matrixnorm $\|\cdot\|_2$, welche für ihre Auswertung die Berechnung des betragsmäßig größten Eigenwertes verlangt, ist die durch (4.3) konstruierte Rang r Matrix A_r ebenfalls Lösung der entsprechenden Minimierungsaufgabe in der leichter auszuwertenden Frobenius–Norm $\|\cdot\|_F$,

$$\|A - A_r\|_F \;=\; \min_{B=B^\top\in I\!R^{n\times n}, \operatorname{rang} B\leq r} \|A - B\|_F\,. \tag{4.4}$$

Satz 4.2 *Die durch* (4.3) *definierte Rang r Matrix A_r ist Lösung der Minimierungsaufgabe* (4.4)*, und es gilt*

$$\min_{B=B^\top\in I\!R^{n\times n}, rang B \,\leq\, r} \|A - B\|_F \;=\; \|A - A_r\|_F \;=\; \left(\sum_{k=r+1}^{n} [\lambda_k(A)]^2\right)^{1/2}.$$

Beweis: Mit (siehe auch Übungsaufgabe 4.1)

$$\|A - A_r\|_F \;=\; \|V(D - D_r)V^\top\|_F \;=\; \|D - D_r\|_F \;=\; \left(\sum_{k=r+1}^{n} [\lambda_k(A)]^2\right)^{1/2}$$

gilt zunächst

$$\min_{B=B^\top \in \mathbb{R}^{n\times n}, \operatorname{rang} B \,\le\, r} \|A - B\|_F \;\le\; \|A - A_r\|_F \;=\; \left(\sum_{k=r+1}^{n} [\lambda_k(A)]^2\right)^{1/2}.$$

Zu zeigen bleibt die Optimalität von A_r bezüglich der Frobenius–Norm $\|\cdot\|_F$. Wie im Beweis von Satz 4.1 sei $B = B^\top \in \mathbb{R}^{n\times n}$ eine Rang r Matrix mit orthonormalen Eigenvektoren $\{\underline{z}_k\}_{k=1}^n$ und zugehörigen Eigenwerten $\lambda_k(B)$, wobei $\lambda_k(B) = 0$ für $k = r+1, \ldots, n$ gelte. Für jedes $k = r+1, \ldots, n$ existiert dann ein Vektor

$$\underline{\bar{u}}_k \in \operatorname{span}\{\underline{z}_{r+1}, \ldots, \underline{z}_n\} \cap \operatorname{span}\{\underline{v}_1, \ldots, \underline{v}_r, \underline{v}_k\}$$

mit $\|\underline{\bar{u}}_k\|_2 = 1$ und $(\underline{\bar{u}}_k, \underline{\bar{u}}_\ell) = \delta_{k\ell}$ für $k, \ell = r+1, \ldots, n$. Wie im Beweis von Satz 4.1 folgen für die so konstruierten Vektoren die Abschätzungen

$$\|(A - B)\underline{\bar{u}}^k\|_2 \;\ge\; \lambda_k(A) \quad \text{für } k = r+1, \ldots, n.$$

Seien $\underline{\bar{u}}_1, \ldots, \underline{\bar{u}}_r$ so gewählt, so daß $\{\underline{\bar{u}}_k\}_{k=1}^n$ ein Orthonormalsystem bildet. Dann definiert

$$\bar{U} \;=\; (\underline{\bar{u}}_1, \ldots, \underline{\bar{u}}_n) \in \mathbb{R}^{n\times n}$$

eine orthogonale Matrix, und es folgt

$$\begin{aligned} \|A - B\|_F^2 \;&=\; \|(A - B)\bar{U}\|_F^2 \;=\; \sum_{k=1}^{n} \|(A - B)\underline{\bar{u}}^k\|_2^2 \\ &\ge\; \sum_{k=r+1}^{n} \|(A - B)\underline{\bar{u}}^k\|_2^2 \;\ge\; \sum_{k=r+1}^{n} [\lambda_k(A)]^2 \end{aligned}$$

und somit

$$\sum_{k=r+1}^{n} [\lambda_k(A)]^2 \;\le\; \min_{B=B^\top \in \mathbb{R}^{n\times n}, \operatorname{rang} B \,\le\, r} \|A - B\|_F^2 \;\le\; \|A - A_r\|_F^2 \;=\; \sum_{k=r+1}^{n} [\lambda_k(A)]^2,$$

woraus unmittelbar die Behauptung folgt. ∎

Für einen gegebenen Vektor $\underline{x} \in \mathbb{R}^n$ wird nun das exakte Matrix–Vektor–Produkt $\underline{z} = A\underline{x}$ ersetzt durch das näherungsweise Matrix–Vektor–Produkt $\widetilde{\underline{z}} = A_r\underline{x}$. Für den Fehler in der Euklidischen Vektornorm ergibt sich aus der Verträglichkeit der Euklidischen Matrixnorm mit der Euklidischen Vektornorm und der Fehlerabschätzung von Satz 4.1

$$\|\underline{z} - \widetilde{\underline{z}}\|_2 \;=\; \|(A - A_r)\underline{x}\|_2 \;\leq\; \|A - A_r\|_2 \|\underline{x}\|_2 \;=\; \lambda_{r+1}(A)\, \|\underline{x}\|_2. \tag{4.5}$$

Aus der Verträglichkeit der Frobeniusnorm zur Euklidischen Vektornorm folgt mit der Fehlerabschätzung von Satz 4.2 andererseits

$$\|\underline{z} - \widetilde{\underline{z}}\|_2 \;=\; \|(A - A_r)\underline{x}\|_2 \;\leq\; \|A - A_r\|_F \|\underline{x}\|_2 \;=\; \left(\sum_{k=r+1}^{n} [\lambda_k(A)]^2 \right)^{1/2} \|\underline{x}\|_2. \tag{4.6}$$

Die Fehlerabschätzung (4.6) ist zwar im Vergleich zu (4.5) schwächer, jedoch erlaubt die Berechnung der Frobenius–Norm $\|A - A_r\|_F$ eine einfacherere Kontrolle des Fehlers.

4.2 Approximation allgemeiner Matrizen

Nachdem im vorigen Abschnitt die Niedrig–Rang–Approximation einer symmetrischen Matrix $A \in \mathbb{R}^{n\times n}$ behandelt wurde, wird jetzt der allgemeine Fall einer Matrix $A \in \mathbb{R}^{m\times n}$ mit Rang

$$\mu \;=\; \operatorname{rang} A \;\leq\; \min\{m, n\}$$

betrachtet. Ausgangspunkt dafür ist das Matrix–Vektorprodukt $\underline{z} = A\underline{x} \in \mathbb{R}^m$ für einen beliebig gegebenen Vektor $\underline{x} \in \mathbb{R}^n$. Wird die Matrix A durch eine noch zu bestimmende Niedrig–Rang–Approximation A_r ersetzt, so erhält man das gestörte Ergebnis $\widetilde{\underline{z}} = A_r\underline{x}$. Für den Fehler in der Euklidischen Vektornorm folgt unter Verwendung der Cauchy–Schwarz–Ungleichung

$$\begin{aligned}
\|\underline{z} - \widetilde{\underline{z}}\|_2^2 &= \|(A - A_r)\underline{x}\|_2^2 \\
&= ((A - A_r)\underline{x}, (A - A_r)\underline{x})_2 \\
&= \left((A^\top - A_r^\top)(A - A_r)\underline{x}, \underline{x}\right)_2 \\
&\leq \|(A^\top - A_r^\top)(A - A_r)\underline{x}\|_2 \|\underline{x}\|_2 \\
&\leq \|(A^\top - A_r^\top)(A - A_r)\|_M \|\underline{x}\|_2^2
\end{aligned}$$

mit einer zur Euklidischen Vektornorm verträglichen Matrixnorm $\|\cdot\|_M$ der symmetrischen Fehlermatrix $(A^\top - A_r^\top)(A - A_r) \in \mathbb{R}^{n\times n}$. Dies motiviert die Betrachtung der symmetrischen Matrix $A^\top A \in \mathbb{R}^{n\times n}$ mit n reellen nichtnegativen Eigenwerten

$$\lambda_1(A^\top A) \geq \lambda_2(A^\top A) \geq \ldots \geq \lambda_n(A^\top A) \geq 0.$$

Die zugehörigen Eigenvektoren $\{\underline{v}_k\}_{k=1}^n$ der symmetrischen Matrix $A^\top A$ bilden ein Orthonormalsystem und es gilt die Faktorisierung

$$V^\top A^\top A V \;=\; \mathrm{diag}(\lambda_k(A^\top A))_{k=1}^n \;=:\; D. \tag{4.7}$$

Wegen $\lambda_k(A^\top A) \geq 0$ existieren die **Singulärwerte**

$$\sigma_k(A) \;=\; \sqrt{\lambda_k(A^\top A)} \;\geq\; 0 \quad \text{für } k = 1, \ldots, \min\{m,n\}. \tag{4.8}$$

Insbesondere gelte $\sigma_k(A) > 0$ für $k = 1, \ldots, \mu = \operatorname{rang} A \leq \min\{m,n\}$ und $\sigma_k(A) = 0$ für $k = \mu+1, \ldots, \min\{m,n\}$. Die Singulärwerte definieren eine Diagonalmatrix

$$\Sigma \;=\; \mathrm{diag}(\sigma_k(A))_{k=1}^{\min\{m,n\}} \in \mathbb{R}^{m\times n},$$

und es gilt

$$D \;=\; \Sigma^\top \Sigma \in \mathbb{R}^{n\times n}.$$

Wird durch

$$\Sigma^+ \;=\; \begin{pmatrix} \frac{1}{\sigma_1(A)} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_\mu(A)} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \in \mathbb{R}^{n\times m} \tag{4.9}$$

die **Pseudoinverse** zu Σ definiert, dann folgt aus (4.7) durch Multiplikation mit $(\Sigma^+)^\top$ von links

$$(\Sigma^+)^\top V^\top A^\top A V \;=\; \Sigma \in \mathbb{R}^{m\times n}$$

bzw.

$$U^\top A V \;=\; \Sigma \tag{4.10}$$

mit

$$U \;=\; AV\Sigma^+ \in \mathbb{R}^{m\times m}. \tag{4.11}$$

Es gilt

$$U^\top U \;=\; \Sigma^{+,\top} V^\top A^\top A V \Sigma^+ \;=\; \Sigma^+ D \Sigma^+ \;=\; \begin{pmatrix} I_\mu & \\ & 0 \end{pmatrix} \in \mathbb{R}^{m\times m}$$

mit der Einheitsmatrix $I_\mu \in \mathbb{R}^{\mu\times\mu}$, d.h. $U^\top$ ist die Pseudoinverse zu U. Damit folgt aus (4.10) die **Singulärwertzerlegung** von A,

$$A \;=\; U\Sigma V^\top \;=\; \sum_{k=1}^{\mu} \sigma_k(A)\underline{u}_k\underline{v}_k^\top \in \mathbb{R}^{m\times n}. \tag{4.12}$$

Für $r < \mu$ sei

$$\Sigma_r \;=\; \begin{pmatrix} \sigma_1(A) & & & & & \\ & \ddots & & & & \\ & & \sigma_r(A) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \in I\!R^{m\times n},$$

dann definiert

$$A_r \;=\; U\Sigma_r V^\top \in I\!R^{m\times n} \tag{4.13}$$

eine Rang r Approximation von A. Aus der Invarianz der Matrixnorm bezüglich orthogonaler Matrizen folgt dann

$$\begin{aligned} \|(A^\top - A_r^\top)(A - A_r)\|_M &= \|(V\Sigma^\top U^\top - V\Sigma_r^\top U^\top)(U\Sigma V^\top - U\Sigma_r V^\top)\|_M \\ &= \|V(\Sigma^\top - \Sigma_r^\top)U^\top U(\Sigma - \Sigma_r)V^\top\|_M \\ &= \|(\Sigma^\top - \Sigma_r^\top)(\Sigma - \Sigma_r)\|_M \end{aligned}$$

und somit

$$\|(A^\top - A_r^\top)(A - A_r)\|_2 \;=\; \lambda_{r+1}(A^\top A) \;=\; [\sigma_{r+1}(A)]^2$$

bzw.

$$\|(A^\top - A_r^\top)(A - A_r)\|_F \;=\; \left[\sum_{k=r+1}^{n} [\sigma_k(A)]^4\right]^{1/2}.$$

Für den Fehler des näherungsweisen Matrix–Vektor–Produkts folgt also

$$\|\underline{z} - \widetilde{\underline{z}}\|_2 \;\le\; \sigma_{r+1}(A)\,\|\underline{x}\|_2\,. \tag{4.14}$$

Im Fall einer symmetrischen Matrix $A = A^\top$ fällt diese Abschätzung mit (4.5) zusammen. Weiterhin gilt

$$\|A - A_r\|_M \;=\; \|U\Sigma V^\top - U\Sigma_r V^\top\|_M \;=\; \|\Sigma - \Sigma_r\|_M$$

und somit

$$\|A - A_r\|_2 \;=\; \sigma_{r+1}(A)$$

bzw.

$$\|A - A_r\|_F \;=\; \left[\sum_{k=r+1}^{n} [\sigma_k(A)]^2\right]^{1/2}.$$

Wie im symmetrischen Fall folgt, daß die durch (4.13) definierte Rang r Approximation A_r die bestmögliche Rang r Approximation von A sowohl in der Spektralnorm als auch in der Frobeniusnorm darstellt.

Übungsaufgaben

4.1. Sei $Q \in \mathbb{R}^{n\times n}$ eine orthogonale Matrix. Man zeige die Gleichheit

$$\|QA\|_F \;=\; \|A\|_F$$

für beliebige Matrizen $A \in \mathbb{R}^{n\times n}$.

4.2. Betrachtet werde das lineare Gleichungssystem

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} = \begin{pmatrix} \underline{f}_1 \\ \underline{f}_2 \end{pmatrix}$$

mit Matrizen $A \in \mathbb{R}^{n\times n}$ und der Einheitsmatrix $I \in \mathbb{R}^{n\times n}$. Die Matrix A werde durch eine Rang r Matrix A_r ersetzt, wobei die Fehlerabschätzung

$$\|A - A_r\|_2 \;\leq\; \varepsilon$$

gelte. Für die Lösung des gestörten linearen Gleichungssystems

$$\begin{pmatrix} I & A_r \\ 0 & I \end{pmatrix} \begin{pmatrix} \widetilde{\underline{x}}_1 \\ \widetilde{\underline{x}}_2 \end{pmatrix} = \begin{pmatrix} \underline{f}_1 \\ \underline{f}_2 \end{pmatrix}$$

gebe man eine Fehlerabschätzung in der Euklidischen Vektornorm an.
Wie lautet die Fehlerabschätzung bei der Voraussetzung

$$\|A - A_r\|_F \;\leq\; \varepsilon\,?$$

Wie lautet die Fehlerabschätzung für das lineare Gleichungssystem

$$\begin{pmatrix} B_1 & A \\ 0 & B_2 \end{pmatrix} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} = \begin{pmatrix} \underline{f}_1 \\ \underline{f}_2 \end{pmatrix}$$

mit regulären Matrizen $B_1, B_2 \in \mathbb{R}^{n\times n}$ und der Approximation von A durch A_r ?

4.3. Für $n \in \mathbb{N}$ seien die Vektoren $\underline{a} = (1,\ldots,1)^\top \in \mathbb{R}^n$ und $\underline{b} = (1,0,\ldots,0)^\top \in \mathbb{R}^n$ gegeben.
Man berechne die Rang 1 Approximation A_1 von $A = \underline{a}\,\underline{a}^\top + \underline{b}\,\underline{b}^\top$.
Wie lautet der Fehler $\|A - A_1\|_2$?

4.4. Sei $A \in \mathbb{R}^{n\times n}$ eine reguläre Matrix mit Singulärwerten

$$\sigma_1(A) \;\geq\; \ldots \;\geq\; \sigma_n(A) \;>\; 0.$$

Man zeige

$$\|A\|_2 \;=\; \sigma_1(A).$$

Was ergibt sich für $\|A^{-1}\|_2$?

Kapitel 5

Arithmetik von Hierarchischen Matrizen

Die Verwendung von Hierarchischen Matrizen innerhalb von Iterationsverfahren erfordert nur eine effiziente Realisierung der Matrix–Vektor–Multiplikation. Zur direkten Lösung von linearen Gleichungssystemen mit Hierarchischen Matrizen bzw. zur Konstuktion optimaler Vorkonditionierungen kann die Inverse einer Hierarchischen Matrix näherungsweise bestimmt werden. Dies erfolgt rekursiv und basiert im wesentlichen auf der Addition und Multiplikation von Hierarchischen Matrizen.
Während die arithmetischen Operationen für allgemeine hierarchische Partitionierungen erklärt werden, erfolgt die Analyse des Rechenaufwandes nur für den in Beispiel 3.2 angegebenen Spezialfall.

Für die Indexmenge

$$I = \{1, 2, 3, \ldots, n\}$$

sei die hierarchische Partitionierung (3.4) gegeben,

$$P^\lambda(I) = \left\{ I_i^\lambda \right\}_{i=1}^{p_\lambda} .$$

Dabei beschreibt $R_i^\lambda : I \to I_i^\lambda$ mit $R_i^\lambda \in \mathbb{R}^{n_i^\lambda \times n}$ gerade die Zuordnung der globalen Indexmenge I auf die lokalen Indexmengen I_i^λ. Weiterhin bezeichnet $P_\mathcal{H}(I, T)$ die zugeordnete hierarchische Partitionierung der Indexmenge $I \times I$, vergleiche (3.5). Schließlich sei

$$P_\mathcal{H}^Z(I, T) \subset P_\mathcal{H}(I, T)$$

die gemäß (3.6) erzeugte hierarchische Partitionierung aller zulässigen Indexpaare $(I_i^\lambda, I_j^\lambda)$ einschließlich der nicht approximierbaren Blöcke des Nahfeldes. Damit gilt für die zugeordnete hierarchische Matrix die Darstellung

$$A = \sum_{(I_i^\lambda, I_j^\lambda) \in P_\mathcal{H}^Z(I,T)} R_j^{\lambda,\top} A_{ij}^\lambda R_i^\lambda \tag{5.1}$$

mit Rang r Block–Matrizen

$$A_{ij}^{\lambda} \;=\; \sum_{k=1}^{r} \underline{a}_{j,\lambda,k} \underline{b}_{i,\lambda,k}^{\top}.$$

Der Einfachheit halber wird im folgenden auch für die Blockmatrizen im Nahfeld die obige Darstellung vorausgesetzt. Durch diese Annahme ergibt sich eine leichte Überschätzung des Rechenaufwandes (siehe hierzu auch Beispiel 3.2), die hier aber vernachlässigt werden soll.

5.1 Matrix–Vektor–Multiplikation

Für einen gegebenen Vektor $\underline{x} \in I\!R^n$ ist das Matrix–Vektor $\underline{z} = A\underline{x}$ zu berechnen. Einsetzen der Darstellung (5.1) ergibt

$$\underline{z} \;=\; A\underline{x} \;=\; \sum_{(I_i^{\lambda}, I_j^{\lambda}) \in P_{\mathcal{H}}^{Z}(I,T)} R_j^{\lambda,\top} A_{ij}^{\lambda} R_i^{\lambda} \underline{x} \;=\; \sum_{(I_i^{\lambda}, I_j^{\lambda}) \in P_{\mathcal{H}}^{Z}(I,T)} R_j^{\lambda,\top} \underline{z}_{ij,\lambda}$$

mit

$$\underline{z}_{ij,\lambda} \;=\; A_{ij}^{\lambda} \underline{x}_{i,\lambda}, \quad \underline{x}_{i,\lambda} = R_i^{\lambda} \underline{x}.$$

Zu realisieren sind somit die lokalen Matrix–Vektor–Multiplikationen

$$\underline{z}_{ij,\lambda} \;=\; A_{ij}^{\lambda} \underline{x}_{i,\lambda} \;=\; \left(\sum_{k=1}^{r} \underline{a}_{j,\lambda,k} \underline{b}_{i,\lambda,k}^{\top} \right) \underline{x}_{i,\lambda} \;=\; \sum_{k=1}^{r} \left(\underline{b}_{i,\lambda,k}^{\top} \underline{x}_{i,\lambda} \right) \underline{a}_{j,\lambda,k}$$

sowie die Assemblierung der Ergebnisvektoren $\underline{z}_{ij,\lambda}$.

Algorithmus zur Matrix–Vektor–Multiplikation:

Durchlaufe die Menge aller zulässigen Indexpaare $(I_i^{\lambda}, I_j^{\lambda}) \in P_{\mathcal{H}}^{Z}(I,T)$:

1. Bestimme die lokalen Vektoren $\underline{x}_{i,\lambda} = R_i^{\lambda} \underline{x}$.
2. Realisiere die lokale Matrix–Vektor–Multiplikation

   $$\underline{z}_{ij,\lambda} \;=\; \sum_{k=1}^{r} \left(\underline{b}_{i,\lambda,k}^{\top} \underline{x}_{i,\lambda} \right) \underline{a}_{j,\lambda,k}$$

   mit einem Aufwand von $r(n_i^{\lambda} + n_j^{\lambda})$ Multiplikationen.
3. Assmbliere die lokalen Anteile $R_j^{\lambda,\top} \underline{z}_{ij,\lambda}$ auf den Ergebnisvektor $\underline{z}$.

Für den Gesamtaufwand der Matrix–Vektor–Multiplikation $\underline{z} = A\underline{x}$ ergibt sich dann

$$Op(A\underline{x}) \;=\; r \sum_{(I_i^{\lambda}, I_j^{\lambda}) \in P_{\mathcal{H}}^{Z}(I,T)} \left(n_i^{\lambda} + n_j^{\lambda} \right).$$

Dieser ist wesentlich abhängig von der Struktur von $P_{\mathcal{H}}^{Z}(I,T)$ und somit von der verwendeten Zulässigkeitsbedingung. Deshalb kann diese Summe in der Regel nur im konkreten Anwendungsfall explizit ausgewertet werden.

Beispiel 5.1 *Gegeben sei die in Beispiel* 3.2 *beschriebene hierarchische Matrix* $A \in \mathbb{R}^{n\times n}$ *mit* $n = 2^L$, *wobei jeder Block* A_{ij}^λ *für* $i \neq j$ *durch eine Rang* r *Matrix dargestellt werden kann, während die Diagonalblöcke* A_{ii}^λ *rekursiv definiert sind:*

$$A_{kk}^{\lambda-1} = \begin{pmatrix} A_{2k-1,2k-1}^\lambda & A_{2k-1,2k}^\lambda \\ A_{2k,2k-1}^\lambda & A_{2k,2k}^\lambda \end{pmatrix}, \quad k = 1,\ldots,2^{\lambda-1}, \lambda = 1,\ldots,L.$$

Für die Matrix–Vektor–Multiplikation $\underline{z} = A\underline{x}$ *ergibt sich dann*

$$\underline{z}_1^\lambda = A_{2k-1,2k-1}^\lambda \underline{x}_1^\lambda + A_{2k-1,2k}^\lambda \underline{x}_2^\lambda, \quad \underline{z}_2^\lambda = A_{2k,2k-1}^\lambda \underline{x}_1^\lambda + A_{2k,2k}^\lambda \underline{x}_2^\lambda$$

mit einem Aufwand von

$$\begin{aligned} Op(A_{kk}^{\lambda-1}\underline{x}^{\lambda-1}) &= 2\,Op(A_{ii}^\lambda \underline{x}^\lambda) + 2\,Op(A_{ij}^\lambda \underline{x}^\lambda) \\ &= 2\,Op(A_{ii}^\lambda \underline{x}^\lambda) + 2[r(n_i^\lambda + n_j^\lambda)] \\ &= 2\,Op(A_{ii}^\lambda \underline{x}^\lambda) + 4\,r\,2^{L-\lambda}. \end{aligned}$$

Wie in Beispiel 3.2 folgt dann

$$Op(A\underline{x}) = n(1 + 2r\log_2 n).$$

5.2 Addition

Gegeben seien zwei hierarchische Matrizen A und B bezüglich der gleichen zulässigen hierarchischen Partitionierung $P_{\mathcal{H}}^Z(I,T)$ der Indexmenge $I \times I$,

$$A = \sum_{(I_i^\lambda, I_j^\lambda)\in P_{\mathcal{H}}^Z(I,T)} R_j^{\lambda,\top} A_{ij}^\lambda R_i^\lambda, \quad B = \sum_{(I_i^\lambda, I_j^\lambda)\in P_{\mathcal{H}}^Z(I,T)} R_j^{\lambda,\top} B_{ij}^\lambda R_i^\lambda,$$

und

$$A_{ij}^\lambda = \sum_{k=1}^r \underline{a}_{j,\lambda,k}\underline{b}_{i,\lambda,k}^\top, \quad B_{ij}^\lambda = \sum_{k=1}^r \underline{c}_{j,\lambda,k}\underline{d}_{i,\lambda,k}^\top.$$

Für die Summe folgt dann

$$A + B = \sum_{(I_i^\lambda, I_j^\lambda)\in P_{\mathcal{H}}^Z(I,T)} R_j^{\lambda,\top}\left[A_{ij}^\lambda + B_{ij}^\lambda\right] R_i^\lambda,$$

und zu berechnen sind also die Blockmatrizen

$$C_{ij}^\lambda = A_{ij}^\lambda + B_{ij}^\lambda = \sum_{k=1}^r \underline{a}_{j,\lambda,k}\underline{b}_{i,\lambda,k}^\top + \sum_{k=1}^r \underline{c}_{j,\lambda,k}\underline{d}_{i,\lambda,k}^\top$$

mit

$$\operatorname{rang} C_{ij}^\lambda \leq 2r\,.$$

Sind die die Niedrig–Rang–Matrizen A_{ij}^{λ} und B_{ij}^{λ} aufspannenden Vektorsysteme linear unabhängig, so liefert deren Summe eine Matrix mit maximalen Rang $2r$. Das Ziel ist deshalb die Definition einer geeigneten Addition, die den maximalen Rang r der Ausgangsmatrizen A_{ij}^{λ} und B_{ij}^{λ} erhält. Durch die in Kapitel 4 beschriebene Rang r Approximation der exakten Summe $A_{ij}^{\lambda} + B_{ij}^{\lambda}$ kann die Rang r Addition $A_{ij}^{\lambda} +_r B_{ij}^{\lambda}$ erklärt werden.
Diese Vorgehensweise soll nun für den Fall $r = 1$ näher untersucht werden. Gegeben seien die Rang 1 Matrizen

$$A = \underline{a}\,\underline{b}^{\top} \in I\!R^{m\times n}, \quad B = \underline{c}\,\underline{d}^{\top} \in I\!R^{m\times n}, \quad \underline{a},\underline{c} \in I\!R^{m}, \underline{b},\underline{d} \in I\!R^{n}.$$

Die Summe

$$C = A + B = \underline{a}\,\underline{b}^{\top} + \underline{c}\,\underline{d}^{\top} \in I\!R^{m\times n}$$

ist im allgemeinen eine Rang 2 Matrix, welche durch eine Rang 1 Matrix C_1 zu approximieren ist. Für die Berechnung der Singulärwerte von C sind zunächst die Eigenwerte von $C^{\top}C$ zu finden. Für die Matrix $C^{\top}C \in I\!R^{n\times n}$ ergibt sich

$$\begin{aligned} C^{\top}C &= \left(\underline{b}\,\underline{a}^{\top} + \underline{d}\,\underline{c}^{\top}\right)\left(\underline{a}\,\underline{b}^{\top} + \underline{c}\,\underline{d}^{\top}\right) \\ &= \underline{b}\,\underline{a}^{\top}\underline{a}\,\underline{b}^{\top} + \underline{b}\,\underline{a}^{\top}\underline{c}\,\underline{d}^{\top} + \underline{d}\,\underline{c}^{\top}\underline{a}\,\underline{b}^{\top} + \underline{d}\,\underline{c}^{\top}\underline{c}\,\underline{d}^{\top} \\ &= \left(\underline{a}^{\top}\underline{a}\right)\underline{b}\,\underline{b}^{\top} + \left(\underline{a}^{\top}\underline{c}\right)\underline{b}\,\underline{d}^{\top} + \left(\underline{c}^{\top}\underline{a}\right)\underline{d}\,\underline{b}^{\top} + \left(\underline{c}^{\top}\underline{c}\right)\underline{d}\,\underline{d}^{\top}. \end{aligned}$$

Für ein beliebiges $\underline{x} \in I\!R^{n}$ folgt dann

$$\begin{aligned} C^{\top}C\underline{x} &= \left(\underline{a}^{\top}\underline{a}\right)\underline{b}\,\underline{b}^{\top}\underline{x} + \left(\underline{a}^{\top}\underline{c}\right)\underline{b}\,\underline{d}^{\top}\underline{x} + \left(\underline{c}^{\top}\underline{a}\right)\underline{d}\,\underline{b}^{\top}\underline{x} + \left(\underline{c}^{\top}\underline{c}\right)\underline{d}\,\underline{d}^{\top}\underline{x} \\ &= \left(\underline{a}^{\top}\underline{a}\right)\left(\underline{b}^{\top}\underline{x}\right)\underline{b} + \left(\underline{a}^{\top}\underline{c}\right)\left(\underline{d}^{\top}\underline{x}\right)\underline{b} + \left(\underline{c}^{\top}\underline{a}\right)\left(\underline{b}^{\top}\underline{x}\right)\underline{d} + \left(\underline{c}^{\top}\underline{c}\right)\left(\underline{d}^{\top}\underline{x}\right)\underline{d} \end{aligned}$$

und somit

$$C^{\top}C\underline{x} \in \operatorname{span}\{\underline{b},\underline{d}\} \quad \text{für beliebiges } \underline{x} \in I\!R^{n}.$$

Daraus folgt $\operatorname{rang} C^{\top}C \leq 2$ bzw. $\lambda_k(C^{\top}C) = 0$ für $k = 3,\ldots,n$, und für die Eigenvektoren der nicht verschwindenden Eigenwerte gilt die Darstellung

$$\underline{x} = \alpha\underline{b} + \beta\underline{d}.$$

Einsetzen dieser Eigenvektoren in die Eigenwertgleichung $C^{\top}C\underline{x} = \lambda(C^{\top}C)\underline{x}$ und anschließender Koeffizientenvergleich von $\underline{b}$ und $\underline{d}$ ergibt das zweidimensionale Eigenwertproblem

$$\begin{pmatrix} \underline{a}^{\top}\underline{a} & \underline{a}^{\top}\underline{c} \\ \underline{c}^{\top}\underline{a} & \underline{c}^{\top}\underline{c} \end{pmatrix}\begin{pmatrix} \underline{b}^{\top}\underline{b} & \underline{b}^{\top}\underline{d} \\ \underline{d}^{\top}\underline{b} & \underline{d}^{\top}\underline{d} \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \lambda(C^{\top}C)\begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Die zugehörigen Eigenlösungen seien durch $(\lambda_1,\alpha_1,\beta_1)$ und $(\lambda_2,\alpha_2,\beta_2)$ mit $\lambda_1 \geq \lambda_2 > 0$ gegeben. Der Fall $\lambda_2 = 0$ entspricht der linearen Abhängigkeit der die Rang 1 Matrizen A und B aufspannenden Vektorsysteme, welche in der Addition wieder eine Rang 1 Matrix ergeben.

Die zugehörigen Eigenvektoren von $C^\top C$ sind

$$\widetilde{\underline{v}}_1 \;=\; \alpha_1\underline{b} + \beta_1\underline{d}, \quad \widetilde{\underline{v}}_2 \;=\; \alpha_2\underline{b} + \beta_2\underline{d}$$

bzw. normiert

$$\underline{v}_1 \;=\; \frac{\widetilde{\underline{v}}_1}{\|\widetilde{\underline{v}}_1\|_2}, \quad \underline{v}_2 \;=\; \frac{\widetilde{\underline{v}}_2}{\|\widetilde{\underline{x}}_2\|_2}.$$

Mit der Singulärwertzerlegung (4.12) folgt somit

$$C \;=\; A + B \;=\; \sigma_1(C)\underline{u}_1\underline{v}_1^\top + \sigma_2(C)\underline{u}_2\underline{v}_2^\top$$

mit $\sigma_i(C) = \sqrt{\lambda_i(C^\top C)}$ für $i = 1, 2$ und, vergleiche (4.11),

$$\underline{u}_i \;=\; \frac{1}{\sigma_i(C)} C\underline{v}_i, \quad i = 1, 2.$$

Die Rang 1 Approximation C_1 ist dann gegeben durch

$$C_1 \;=\; \sigma_1(C)\underline{u}_1\underline{v}_1^\top,$$

und für den Fehler dieser Approximation gilt

$$\|C - C_1\|_2 \;=\; \sigma_2(C)\,.$$

Für die Berechnung des normierten Eigenvektors $\underline{v}_1 \in I\!R^n$ bietet sich die folgende Alternative an. Ohne Einschränkung der Allgemeinheit sei $\alpha_1 \neq 0$. Dann ist

$$\widetilde{\underline{v}}_1 \;=\; \alpha_1\underline{b} + \beta_1\underline{d} \;=\; \alpha_1\left[\underline{b} + \frac{\beta_1}{\alpha_1}\underline{d}\right],$$

und somit folgt

$$\gamma_1 := \frac{\beta_1}{\alpha_1}, \quad \hat{\underline{v}}_1 \;=\; \underline{b} + \gamma_1\underline{d}, \quad \underline{v}_1 \;=\; \frac{\hat{\underline{v}}_1}{\|\hat{\underline{v}}_1\|_2}\,.$$

Für die Berechnung des Vektors $\underline{u}_1 \in I\!R^m$ ist

$$\begin{aligned}\underline{u}_1 \;=\; \frac{1}{\sigma_1(C)} C\underline{v}_1 \;&=\; \frac{1}{\sigma_1(C)}\left[\underline{a}\,\underline{b}^\top + \underline{c}\,\underline{d}^\top\right]\frac{\alpha_1\underline{b} + \beta_1\underline{d}}{\|\alpha_1\underline{b} + \beta_1\underline{d}\|_2} \\ &=\; \frac{1}{\sigma_1(C)}\frac{\alpha_1}{\|\alpha_1\underline{b} + \beta_1\underline{d}\|_2}\left[\underline{a}\,\underline{b}^\top + \underline{c}\,\underline{d}^\top\right][\underline{b} + \gamma_1\underline{d}]\,.\end{aligned}$$

Zu berechnen ist also

$$\begin{aligned}\bar{\underline{u}} \;&=\; \left[\underline{a}\,\underline{b}^\top + \underline{c}\,\underline{d}^\top\right][\underline{b} + \gamma_1\underline{d}] \\ &=\; \left(\underline{b}^\top\underline{b} + \gamma_1\underline{b}^\top\underline{d}\right)\underline{a} + \left(\underline{d}^\top\underline{b} + \gamma_1\underline{d}^\top\underline{d}\right)\underline{c} \\ &=\; \left[\underline{b}^\top\underline{b} + \gamma_1\underline{b}^\top\underline{d}\right]\left[\underline{a} + \frac{\underline{d}^\top\underline{b} + \gamma_1\underline{d}^\top\underline{d}}{\underline{b}^\top\underline{b} + \gamma_1\underline{b}^\top\underline{d}}\,\underline{c}\right],\end{aligned}$$

falls $\underline{b}^\top \underline{b} + \gamma_1 \underline{b}^\top \underline{d} \neq 0$ vorausgesetzt wird. Damit ergibt sich

$$\delta_1 := \frac{\underline{d}^\top \underline{b} + \gamma_1 \underline{d}^\top \underline{d}}{\underline{b}^\top \underline{b} + \gamma_1 \underline{b}^\top \underline{d}}, \quad \widetilde{\underline{u}}_1 := \underline{a} + \delta_1 \underline{c}, \quad \underline{u}_1 := \frac{\widetilde{\underline{u}}_1}{\|\widetilde{\underline{u}}_1\|_2}.$$

Insgesamt ergibt sich für die Rang 1 Addition zweier hierarchischen Matrizen mit Block–Matrizen vom Rang 1 der folgende Algorithmus.

Algorithmus zur Addition von Hierarchischen Matrizen

Durchlaufe die Menge aller zulässigen Indexpaare $(I_i^\lambda, I_j^\lambda) \in P_{\mathcal{H}}^Z(I,T)$:

Betrachte die gegebenen Rang 1 Matrizen

$$A_{ij}^\lambda = \underline{a}_{j,\lambda} \underline{b}_{i,\lambda}^\top, \quad B_{ij}^\lambda = \underline{c}_{j,\lambda} \underline{d}_{i,\lambda}^\top, \quad \underline{a}_{j,\lambda}, \underline{c}_{j,\lambda} \in I\!R^{n_j^\lambda}, \ \underline{b}_{i,\lambda}, \underline{d}_{i,\lambda} \in I\!R^{n_i^\lambda}.$$

1. Berechne die symmetrischen Gram–Matrizen

$$G_1^{j,\lambda} = \begin{pmatrix} \underline{a}_{j,\lambda}^\top \underline{a}_{j,\lambda} & \underline{a}_{j,\lambda}^\top \underline{c}_{j,\lambda} \\ \underline{c}_{j,\lambda}^\top \underline{a}_{j,\lambda} & \underline{c}_{j,\lambda}^\top \underline{c}_{j,\lambda} \end{pmatrix}, \quad G_2^{i,\lambda} = \begin{pmatrix} \underline{b}_{i,\lambda}^\top \underline{b}_{i,\lambda} & \underline{b}_{i,\lambda}^\top \underline{d}_{i,\lambda} \\ \underline{d}_{i,\lambda}^\top \underline{b}_{i,\lambda} & \underline{d}_{i,\lambda}^\top \underline{d}_{i,\lambda} \end{pmatrix}$$

   mit einem Aufwand von $3(n_i^\lambda + n_j^\lambda)$ Multiplikationen.

2. Löse das zweidimensionale Eigenwertproblem

$$G_1^{j,\lambda} G_2^{i,\lambda} \begin{pmatrix} \alpha^{ij,\lambda} \\ \beta^{ij,\lambda} \end{pmatrix} = \lambda^{ij,\lambda} \begin{pmatrix} \alpha^{ij,\lambda} \\ \beta^{ij,\lambda} \end{pmatrix}$$

   mit den Eigenlösungen

$$(\lambda_1^{ij,\lambda}, \alpha_1^{ij,\lambda}, \beta_1^{ij,\lambda}), \quad (\lambda_2^{ij,\lambda}, \alpha_2^{ij,\lambda}, \beta_2^{ij,\lambda}), \quad \lambda_1^{ij,\lambda} \geq \lambda_2^{ij,\lambda}.$$

3. Berechne den normierten Eigenvektor $\underline{v}_{i,\lambda} \in I\!R^{n_i^\lambda}$ via

$$\gamma^{ij,\lambda} = \frac{\beta_1^{ij,\lambda}}{\alpha_1^{ij,\lambda}}, \quad \widetilde{\underline{v}}_{i,\lambda} = \underline{b}_{i,\lambda} + \gamma^{ij,\lambda} \underline{d}_{i,\lambda}, \quad \underline{v}_{i,\lambda} = \frac{\widetilde{\underline{v}}_{i,\lambda}}{\|\widetilde{\underline{v}}_{i,\lambda}\|_2}$$

   mit $3n_i^\lambda + 1$ Multiplikationen sowie einer Wurzelberechnung.

4. Berechne den normierten Vektor $\underline{u}_{j,\lambda} \in I\!R^{n_j^\lambda}$ via

$$\delta^{ij,\lambda} := \frac{\underline{d}_{i,\lambda}^\top \underline{b}_{i,\lambda} + \gamma^{ij,\lambda} \underline{d}_{i,\lambda}^\top \underline{d}_{i,\lambda}}{\underline{b}_{i,\lambda}^\top \underline{b}_{i,\lambda} + \gamma^{ij,\lambda} \underline{b}_{i,\lambda}^\top \underline{d}_{i,\lambda}}, \quad \widetilde{\underline{u}}_{j,\lambda} := \underline{a}_{j,\lambda} + \delta^{ij,\lambda} \underline{c}_{j,\lambda}, \quad \underline{u}_{j,\lambda} := \frac{\widetilde{\underline{u}}_{j,\lambda}}{\|\widetilde{\underline{u}}_{j,\lambda}\|_2}$$

   mit $3n_j^\lambda + 3$ Multiplikationen (Divisionen) sowie einer Wurzelberechnung. Dabei werden die bereits in Schritt 2. berechneten Matrix $G_2^{i,\lambda}$ verwendet.

Für die Addition zweier Block–Matrizen A_{ij}^{λ} und B_{ij}^{λ} vom Rang 1 ergibt sich der Aufwand

$$Op(A_{ij}^{\lambda} +_r B_{ij}^{\lambda}) \;=\; 6(n_i^{\lambda} + n_j^{\lambda}) + \mathcal{O}(1)$$

Multiplikationen, und für den Gesamtaufwand folgt

$$Op(A +_r B) \;=\; 6 \sum_{(I_i^{\lambda}, I_j^{\lambda}) \in P_{\mathcal{H}}^{Z}(I,T)} [n_i^{\lambda} + n_j^{\lambda} + \mathcal{O}(1)].$$

Diese Abschätzung soll nun für die in Beispiel 3.2 erklärte hierarchische Matrix näher untersucht werden.

Beispiel 5.2 *Gegeben seien wie in Beispiel* 3.2 *zwei hierarchische Matrizen* $A, B \in \mathbb{R}^{n\times n}$ *mit* $n = 2^L$*, wobei jeder der Blöcke* A_{ij}^{λ} *und* B_{ij}^{λ} *für* $i \neq j$ *durch Rang* 1 *Matrizen dargestellt werden. Rekursiv gilt also*

$$A_{kk}^{\lambda-1} \;=\; \begin{pmatrix} A_{2k-1,2k-1}^{\lambda} & A_{2k-1,2k}^{\lambda} \\ A_{2k,2k-1}^{\lambda} & A_{2k,2k}^{\lambda} \end{pmatrix}, \quad B_{kk}^{\lambda-1} \;=\; \begin{pmatrix} B_{2k-1,2k-1}^{\lambda} & B_{2k-1,2k}^{\lambda} \\ B_{2k,2k-1}^{\lambda} & B_{2k,2k}^{\lambda} \end{pmatrix}$$

für $k = 1, \ldots, 2^{\lambda-1}, \lambda = 1, \ldots, L$.
Die Rang 1 *Addition* $A_{kk}^{\lambda-1} +_1 B_{kk}^{\lambda-1}$ *erfordert also die zwei Rang* 1 *Additionen*

$$A_{2k-1,2k-1}^{\lambda} +_1 B_{2k-1,2k-1}^{\lambda}, \quad A_{2k,2k}^{\lambda} +_1 B_{2k,2k}^{\lambda}$$

sowie die Rang 1 *Additionen von Rang* 1 *Matrizen*

$$A_{2k-1,2k}^{\lambda} +_1 B_{2k-1,2k}^{\lambda}, \quad A_{2k,2k-1}^{\lambda} +_1 B_{2k,2k-1}^{\lambda}.$$

Für den Aufwand ergibt sich

$$\begin{aligned} Op(A_{kk}^{\lambda-1} +_1 B_{kk}^{\lambda-1}) \;&=\; 2\,Op(A_{ii}^{\lambda} +_1 B_{ii}^{\lambda}) + 2\,Op(A_{ij}^{\lambda} +_1 B_{ij}^{\lambda}) \\ &=\; 2\,Op(A_{ii}^{\lambda} +_1 B_{ii}^{\lambda}) + 24\,2^{L-\lambda} \;+\; \mathcal{O}(1). \end{aligned}$$

Insbesondere für $\lambda = 1$ *ist also*

$$Op(A +_1 B) \;=\; 2\,Op(A_{ii}^{1} +_1 B_{ii}^{1}) + 12\,2^L + \mathcal{O}(1)$$

und rekursiv folgt

$$Op(A +_1 B) \;=\; 2^{\lambda}\,Op(A_{ii}^{\lambda} +_1 B_{ii}^{\lambda}) + 12\,\lambda\,2^L + \mathcal{O}(L).$$

Mit $\lambda = L$ *ergibt sich also*

$$Op(A +_1 B) \;=\; 2^L\,Op(A_{ii}^{L} +_1 B_{ii}^{L}) + 12\,L\,2^L + \mathcal{O}(L) \;=\; \mathcal{O}(n \log_2 n).$$

Abschließend soll die Rang r Addition $C = A + B$ zweier Rang r Matrizen A und B mit $A, B \in \mathbb{R}^{m\times n}$ betrachtet werden,

$$A \;=\; \sum_{k=1}^{r} \underline{a}_k \underline{b}_k^\top, \quad B \;=\; \sum_{k=1}^{r} \underline{c}_k \underline{d}_k^\top, \quad C \;=\; \sum_{k=1}^{r} [\underline{a}_k \underline{b}_k^\top + \underline{c}_k \underline{d}_k^\top \,.]$$

Wie im Fall $r = 1$ folgt

$$C^\top C \underline{x} \in \text{span}\,\{\underline{b}_k, \underline{d}_k\}_{k=1}^r \quad \text{für alle } \underline{x} \in \mathbb{R}^n.$$

Der Ansatz

$$\underline{x} \;=\; \sum_{j=1}^{r} \Big[\alpha_j \underline{b}_j + \beta_j \underline{d}_j\Big]$$

für die Eigenvektoren von $C^\top C$ liefert

$$\begin{aligned} C^\top C \underline{x} \;=\; & \sum_{k=1}^{r}\sum_{\ell=1}^{r}\sum_{j=1}^{r} \Big[(\underline{a}_k^\top \underline{a}_\ell)[(\underline{b}_\ell^\top \underline{b}_j)\alpha_j + (\underline{b}_\ell^\top \underline{d}_j)\beta_j] + (\underline{a}_k^\top \underline{c}_\ell)[(\underline{d}_\ell^\top \underline{b}_j)\alpha_j + \underline{d}_\ell^\top \underline{d}_j)\beta_j]\Big]\, \underline{b}_k \\ & + \sum_{k=1}^{r}\sum_{\ell=1}^{r}\sum_{j=1}^{r} \Big[(\underline{c}_k^\top \underline{a}_\ell)[(\underline{b}_\ell^\top \underline{b}_j)\alpha_j + \underline{b}_\ell^\top \underline{d}_j)\beta_j] + (\underline{c}_k^\top \underline{c}_\ell)[(\underline{d}_\ell^\top \underline{b}_j)\alpha_j + \underline{d}_\ell^\top \underline{d}_j)\beta_j]\Big]\, \underline{d}_k \end{aligned}$$

und durch Koeffizientenvergleich folgt

$$\sum_{\ell=1}^{r}\sum_{j=1}^{r} \Big[(\underline{a}_k^\top \underline{a}_\ell)[(\underline{b}_\ell^\top \underline{b}_j)\alpha_j + (\underline{b}_\ell^\top \underline{d}_j)\beta_j] + (\underline{a}_k^\top \underline{c}_\ell)[(\underline{d}_\ell^\top \underline{b}_j)\alpha_j + \underline{d}_\ell^\top \underline{d}_j)\beta_j]\Big] \;=\; \lambda\, \alpha_k$$

bzw.

$$\sum_{\ell=1}^{r}\sum_{j=1}^{r} \Big[(\underline{c}_k^\top \underline{a}_\ell)[(\underline{b}_\ell^\top \underline{b}_j)\alpha_j + \underline{b}_\ell^\top \underline{d}_j)\beta_j] + (\underline{c}_k^\top \underline{c}_\ell)[(\underline{d}_\ell^\top \underline{b}_j)\alpha_j + \underline{d}_\ell^\top \underline{d}_j)\beta_j]\Big] \;=\; \lambda\, \beta_k$$

jeweils für $k = 1, \ldots, r$. Beide Forderungen führen auf ein Eigenwertproblem der Dimension $2r$,

$$\begin{pmatrix} \underline{a}_k^\top \underline{a}_\ell & \underline{a}_k^\top \underline{c}_\ell \\ \underline{c}_k^\top \underline{a}_\ell & \underline{c}_k^\top \underline{c}_\ell \end{pmatrix}_{k,\ell=1}^{r} \begin{pmatrix} \underline{b}_k^\top \underline{b}_\ell & \underline{b}_k^\top \underline{d}_\ell \\ \underline{d}_k^\top \underline{b}_\ell & \underline{d}_k^\top \underline{d}_\ell \end{pmatrix}_{k,\ell=1}^{r} \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix}_{k=1}^{r} = \lambda \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix}_{k=1}^{r}. \tag{5.2}$$

Zu diskutieren bleibt die Aufwandsabschätzung zur Erzeugung der Rang r Approximation von $A + B$:

1. Die Generierung der symmetrischen Gram–Matrizen, zum Beispiel mit den Einträgen

   $$G_1[k, \ell] = \underline{a}_k^\top \underline{a}_\ell, \quad G_1[k, r + \ell] = \underline{a}_k \top \underline{c}_\ell, \quad G_1[r + k, r + \ell] = \underline{c}_k^\top \underline{c}_\ell$$

   für $k, \ell = 1, \ldots, r, k \le \ell$, erfordert insgesamt

   $$\frac{1}{2} r(r + 1)\,(n + m)$$

   Multiplikationen.

2. Die Lösung des Eigenwertproblems (5.2) zum Beispiel mit einem Verfahren kubischen Aufwands erfordert $\mathcal{O}((2r)^3) = \mathcal{O}(8r^3)$ wesentliche Operationen.

3. Die Berechnung der r Eigenvektoren von $C^\top C$ einschließlich ihrer Normierung erfordert $\mathcal{O}(rn)$ Multiplikationen.

4. Die Berechnung der Vektoren $\underline{u}_k = C\underline{v}_k$ und ihre Normierung verlangt $\mathcal{O}(rm)$ Multiplikationen.

Insgesamt ergibt sich für die Addition zweier Rang r Matrizen $A, B \in I\!R^{m\times n}$ ein Aufwand von

$$Op(A +_r B) \;=\; \mathcal{O}(r^2[n+m] + r^3)$$

wesentlichen Operationen.

5.3 Matrix–Matrix–Multiplikation

Gegeben sei zwei hierarchische Matrizen $A, B \in I\!R^{n\times n}$ bezüglich der gleichen zulässigen hierarchischen Partitionierung $P^Z_{\mathcal{H}}(I,T)$. Diese erlauben die Block–Darstellung

$$A \;=\; \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B \;=\; \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

mit Matrizen $A_{ij}, B_{ij} \in I\!R^{n_i\times n_j}$. Dann ist

$$A\cdot B \;=\; \begin{pmatrix} A_{11}\cdot B_{11} + A_{12}\cdot B_{21} & A_{11}\cdot B_{12} + A_{12}\cdot B_{22} \\ A_{21}\cdot B_{11} + A_{12}\cdot B_{21} & A_{21}\cdot B_{12} + A_{22}\cdot B_{22} \end{pmatrix},$$

und das Matrix–Matrix–Produkt $A\cdot B$ kann durch acht Matrix–Matrix–Multiplikationen $A_{ij}\cdot B_{jk}$ und vier Matrix–Additionen jeweils kleinerer Dimension realisiert werden. Je nach Gestalt der Matrizen A_{ij} und B_{jk} können die Matrix–Matrix–Multiplikationen $A_{ij}\cdot B_{jk}$ realisiert werden.

1. Beide Matrizen $A_{ij} \in I\!R^{n_i\times n_j}$ und $B_{jk} \in I\!R^{n_j\times n_k}$ sind als Rang r Matrizen gegeben,

$$A_{ij} \;=\; \sum_{s=1}^{r} \underline{a}_s^{ij}\,\underline{b}_s^{ij,\top}, \quad B_{jk} \;=\; \sum_{t=1}^{r} \underline{c}_t^{jk}\,\underline{d}_t^{jk,\top}.$$

   Dann ergibt sich

$$A_{ij}\cdot B_{jk} = \sum_{s=1}^{r} \underline{a}_s^{ij}\,\underline{b}_s^{ij,\top} \sum_{t=1}^{r} \underline{c}_t^{jk}\,\underline{d}_t^{jk,\top} = \sum_{t=1}^{r} \left[\sum_{s=1}^{r} \left(\underline{b}_s^{ij,\top}\underline{c}_t^{jk}\right)\underline{a}_s^{ij}\right] \underline{d}_t^{jk,\top} = \sum_{t=1}^{r} \widetilde{\underline{a}}_t^{ij}\,\underline{d}_t^{jk,\top}.$$

   Der Aufwand zur Berechnung der r Vektoren

$$\widetilde{\underline{a}}_t^{ij} \;=\; \sum_{s=1}^{r} \left(\underline{b}_s^{ij,\top}\underline{c}_t^{jk}\right)\underline{a}_s^{ij}$$

beträgt

$$r\,[rn_j + rn_i] \;=\; r^2[n_i + n_j]$$

Multiplikationen.

2. Genau eine der beiden Matrizen $A_{ij} \in I\!R^{n_i \times n_j}$ oder $B_{jk} \in I\!R^{n_j \times n_k}$ ist als Rang r Matrix darstellbar, während die andere eine hierarchische Matrix ist. Zum Beispiel für

$$B_{jk} \;=\; \sum_{t=1}^{r} \underline{a}_t^{jk} \underline{b}_t^{jk,\top}$$

folgt dann

$$A_{ij} \cdot B_{jk} \;=\; \sum_{t=1}^{r} A_{ij} \underline{a}_t^{jk} \underline{b}_t^{jk,\top} \;=\; \sum_{t=1}^{r} \widetilde{\underline{a}}_t^{ij} \underline{b}_t^{jk,\top}.$$

Die Berechnung der r Vektoren

$$\widetilde{\underline{a}}_t^{ij} \;=\; A_{ij} \underline{a}_t^{jk}.$$

erfordert dabei $r\,Op(A_{ij}\underline{a}_t^{jk})$ Multiplikationen. Für

$$A_{ij} \;=\; \sum_{s=1}^{r} \underline{a}_s^{ij} \underline{b}_s^{ij,\top}$$

gilt analog

$$A_{ij} \cdot B_{jk} \;=\; \sum_{s=1}^{r} \underline{a}_s^{ij} \underline{b}_s^{ij,\top} B_{jk} \;=\; \sum_{s=1}^{r} \underline{a}_s^{ij} \left(B_{jk}^{\top} \underline{b}_s^{ij} \right)^{\top} \;=\; \sum_{s=1}^{r} \underline{a}_s^{ij} \widetilde{\underline{b}}_s^{jk,\top}$$

mit den Vektoren

$$\widetilde{\underline{b}}_s^{jk} \;=\; B_{jk}^{\top} \underline{b}_s^{ij} \quad s = 1, \ldots, r$$

und einem Aufwand von $r\,Op(B_{jk}^{\top}\underline{b}_s^{ij})$ Multiplikationen.
Die verbleibenden Matrix–Vektor–Produkte $A_{ij}\underline{a}_t^{jk}$ bzw. $B_{jk}^{\top}\underline{b}_s^{ij}$ können im Fall hierarchischer Matrizen A_{ij} bzw. B_{jk} entsprechend realisiert werden.

3. Sind beide Block–Matrizen $A_{ij} \in I\!R^{n_i \times n_j}$ und $B_{jk} \in I\!R^{n_j \times n_k}$ als hierarchische Matrizen gegeben, so erfolgt die Matrix–Matrix–Multiplikation $A_{ij} \cdot B_{jk}$ rekursiv durch Betrachtungen der entsprechenden Teilblöcke.

Neben den Matrix–Matrix–Multiplikationen $A_{ij} \cdot B_{jk}$ sind die Ergebnis–Matrizen zu addieren. Die näherungsweise Rang r Addition $+_r$ impliziert dann durch die Rekursion eine näherungsweise Rang r Multiplikation $\cdot_r$,

$$A \cdot_r B \;:=\; \begin{pmatrix} A_{11} \cdot_r B_{11} +_r A_{12} \cdot_r B_{21} & A_{11} \cdot_r B_{12} +_r A_{12} \cdot_r B_{22} \\ A_{21} \cdot_r B_{11} +_r A_{12} \cdot_r B_{21} & A_{21} \cdot_r B_{12} +_r A_{22} \cdot_r B_{22} \end{pmatrix}.$$

Beispiel 5.3 *Gegeben seien wie in Beispiel* 3.2 *zwei hierarchische Matrizen* $A, B \in \mathbb{R}^{n\times n}$ *mit* $n = 2^L$*, wobei jeder der Blöcke* A_{ij}^λ *und* B_{ij}^λ *für* $i \neq j$ *durch Rang 1 Matrizen dargestellt werden, während* A_{ii}^λ *und* B_{ii}^λ *rekursiv gegeben sind:*

$$A_{kk}^{\lambda-1} = \begin{pmatrix} A_{2k-1,2k-1}^\lambda & A_{2k-1,2k}^\lambda \\ A_{2k,2k-1}^\lambda & A_{2k,2k}^\lambda \end{pmatrix}, \quad B_{kk}^{\lambda-1} = \begin{pmatrix} B_{2k-1,2k-1}^\lambda & B_{2k-1,2k}^\lambda \\ B_{2k,2k-1}^\lambda & B_{2k,2k}^\lambda \end{pmatrix}.$$

Für das näherungsweise Matrix–Produkt $A_{kk}^{\lambda-1} \cdot_r B_{kk}^{\lambda-1}$ *folgt*

$$A_{kk}^{\lambda-1} \cdot_r B_{kk}^{\lambda-1} = \begin{pmatrix} A_{2k-1,2k-1}^\lambda \cdot_r B_{2k-1,2k-1}^\lambda +_r A_{2k-1,2k}^\lambda \cdot B_{2k,2k-1}^\lambda & A_{2k-1,2k-1}^\lambda \cdot B_{2k-1,2k}^\lambda +_r A_{2k-1,2k}^\lambda \cdot B_{2k,2k}^\lambda \\ A_{2k,2k-1}^\lambda \cdot B_{2k-1,2k-1}^\lambda +_r A_{2k,2k}^\lambda \cdot B_{2k,2k-1}^\lambda & A_{2k,2k-1}^\lambda \cdot B_{2k-1,2k}^\lambda +_r A_{2k,2k}^\lambda \cdot_r B_{2k,2k}^\lambda \end{pmatrix}.$$

Zu realisieren sind somit zwei näherungsweise Matrix–Produkte $A_{ii}^\lambda \cdot_r B_{ii}^\lambda$*, zwei exakte Matrix–Produkte* $A_{ij}^\lambda \cdot B_{ji}^\lambda$ *von Rang* r *Matrizen* A_{ij}^λ *und* B_{ji}^λ *für* $i \neq j$*, vier exakte Matrix–Produkte* $A_{ii}^\lambda \cdot B_{ij}^\lambda$ *bzw.* $A_{ji}^\lambda \cdot B_{ii}^\lambda$ *von Rang* r *Matrizen mit hierarchischen Matrizen sowie vier Rang* r *Additionen. Es gilt also*

$$Op(A_{kk}^{\lambda-1} \cdot_r B_{kk}^{\lambda-1}) = 2\,Op(A_{ii}^\lambda \cdot_r B_{ii}^\lambda) + 2\,Op(A_{ij}^\lambda \cdot B_{ji}^\lambda) + 4\,Op(A_{ii}^\lambda \cdot B_{ij}^\lambda) + 4\,Op(+_r)$$

Auflösen der Rekursions ergibt für den Gesamtaufwand

$$Op(A \cdot_r B) = \mathcal{O}(r^2 n \log_2 n)$$

Multiplikationen.

5.4 Invertierung

Abschließend soll die näherungsweise Invertierung hierarchischer Matrizen betrachtet werden. Ausgangspunkt ist die Faktorisierung der Block–Darstellung

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} I & 0 \\ A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}$$

mit dem Schur–Komplement

$$S = A_{22} - A_{21}A_{11}^{-1}A_{12}\,.$$

Für die inverse Matrix A^{-1} folgt dann

$$\begin{aligned} A^{-1} &= \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \\ &= \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{pmatrix}. \end{aligned} \tag{5.3}$$

Die Berechnung der inversen Matrix A^{-1} erfordert also im wesentlichen

1. die Invertierung der Block–Matrix A_{11},
2. die Berechnung des Schur–Komplements $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$,
3. die Invertierung des Schur–Komplements.

Die näherungsweise Multiplikation und Addition der hierarchisch definierten Block–Matrizen ermöglicht somit die rekursive Definition aller auftretenden inversen Block–Matrizen und somit die näherungsweise Invertierung der hierachischen Matrix A. Der Gesamtaufwand beträgt dabei

$$Op(A^{-1}) \;=\; \mathcal{O}(r^2 n \ln^2 n)$$

Multiplikationen.

Beispiel 5.4 *Die Anwendung der Invertierungsformel* (5.3) *soll für eine symmetrische Matrix* $A = A^\top$ *mit* $A_{12} = \underline{a}\,\underline{b}^\top$ *betrachtet werden, zu invertieren ist also*

$$A \;=\; \begin{pmatrix} A_{11} & \underline{a}\,\underline{b}^\top \\ \underline{b}\,\underline{a}^\top & A_{22} \end{pmatrix}.$$

Nach (5.3) *ist*

$$A^{-1} \;=\; \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}\underline{a}\,\underline{b}^\top S^{-1}\underline{b}\,\underline{a}^\top A_{11}^{-1} & -A_{11}^{-1}\underline{a}\,\underline{b}^\top S^{-1} \\ -S^{-1}\underline{b}\,\underline{a}^\top A_{11}^{-1} & S^{-1} \end{pmatrix}$$

mit dem Schur–Komplement

$$S \;=\; A_{22} - \underline{b}\,\underline{a}^\top A_{11}^{-1}\underline{a}\,\underline{b}^\top \;=\; A_{22} - \alpha\,\underline{b}\,\underline{b}^\top, \quad \alpha = \underline{a}^\top A_{11}^{-1}\underline{a}.$$

Die Anwendung der Sherman–Morrison–Formel (2.9) *ergibt für die Inverse des Schur–Komplements*

$$S^{-1} \;=\; A_{22}^{-1} + \gamma\, A_{22}^{-1}\underline{b}\,\underline{b}^\top A_{22}^{-1}, \quad \gamma = \frac{\alpha}{1-\alpha\beta}, \quad \beta = \underline{b}^\top A_{22}^{-1}\underline{b}.$$

Mit

$$S^{-1}\underline{b} \;=\; \left[A_{22}^{-1} + \gamma\, A_{22}^{-1}\underline{b}\,\underline{b}^\top A_{22}^{-1}\right]\underline{b} \;=\; (1+\beta\gamma)\, A_{22}^{-1}\underline{b}$$

und

$$\underline{b}^\top S^{-1}\underline{b} \;=\; (1+\beta\gamma)\underline{b}^\top A_{22}^{-1}\underline{b} \;=\; (1+\beta\gamma)\beta$$

folgt

$$\begin{aligned} A^{-1} \;&=\; \begin{pmatrix} A_{11}^{-1} + (1+\beta\gamma)\beta A_{11}^{-1}\underline{a}\,\underline{a}^\top A_{11}^{-1} & -(1+\beta\gamma)A_{11}^{-1}\underline{a}\,\underline{b}^\top A_{22}^{-1} \\ -(1+\beta\gamma)A_{22}^{-1}\underline{b}\,\underline{a}^\top A_{11}^{-1} & A_{22}^{-1} + \gamma A_{22}^{-1}\underline{b}\,\underline{b}^\top A_{22}^{-1} \end{pmatrix} \\ &=\; \begin{pmatrix} A_{11}^{-1} + (1+\beta\gamma)\beta\widetilde{\underline{a}}\,\widetilde{\underline{a}} & -(1+\beta\gamma)\widetilde{\underline{a}}\,\widetilde{\underline{b}}^\top \\ -(1+\beta\gamma)\widetilde{\underline{b}}\,\widetilde{\underline{a}}^\top & A_{22}^{-1} + \gamma\widetilde{\underline{b}}\,\widetilde{\underline{b}}^\top \end{pmatrix} \end{aligned}$$

mit

$$\widetilde{\underline{a}} \;=\; A_{11}^{-1}\underline{a}, \quad \widetilde{\underline{b}} \;=\; A_{22}^{-1}\underline{b}.$$

Die Nebendiagonalblöcke der inversen Matrix A^{-1} sind offenbar wieder Rang 1 Matrizen, während die Hauptdiagonalblöcke mit Rang 1 Matrizen zu addieren sind. Können die Hauptdiagonalblöcke A_{11} und A_{22} rekursiv wie die ursprüngliche Matrix A dargestellt werden, so ist für die Darstellung der inversen Matrix eine (näherungsweise) Rang 1 Addition durchzuführen.
Allgemein gilt, zum Beispiel für den ersten Hauptdiagonalblock,

$$A_{11}^{-1} + (1+\beta\gamma)\beta\widetilde{\underline{a}}\,\widetilde{\underline{a}}^{\top} \;=\; A_{11}^{-1} + (1+\beta\gamma)\beta A_{11}^{-1}\underline{a}\,\underline{a}^{\top}A_{11}^{-1}.$$

Für die weiteren Betrachtungen werde vorausgesetzt, daß die inverse Matrix A_{11}^{-1} als hierarchische Matrix mit Rang 1 Matrizen in den Nebendiagonalblöcken gegeben sei. Enthält der Vektor $\underline{a}$ genau ein Nichtnullelement, so beschreibt $A_{11}^{-1}\underline{a}$ das Vielfache einer Spalte der inversen Matrix A_{11}^{-1}. Damit kann die Addition der hierarchischen Matrix A_{11}^{-1} mit einem beliebigen Vielfachen der Rang 1 Matrix $A_{11}^{-1}\underline{a}\,\underline{a}^{\top}A_{11}^{-1}$ **exakt** *realisiert werden und das Ergebnis ist wiederum eine hierarchische Matrix mit Rang 1 Matrizen in den Nebendiagonalblöcken.*

Übungsaufgaben

5.1. Gegeben seien die symmetrischen Rang 2 Matrizen

$$A \;=\; \underline{a}_1\underline{a}_1^{\top} + \underline{a}_2\underline{a}_2^{\top}, \quad B \;=\; \underline{b}_1\underline{b}_1^{\top} + \underline{b}_2\underline{b}_2^{\top}.$$

Wie lautet die Rang 2 Approximation der Summe $A + B$, d.h.

$$M \;=\; A +_2 B\,?$$

5.2. Gegeben sei die symmetrische und positiv definite Matrix

$$M \;=\; \begin{pmatrix} A & B \\ B^{\top} & D \end{pmatrix},$$

d.h. es gilt

$$(M\underline{x}, \underline{x}) \;\geq\; c_1^M \,\|\underline{x}\|_2^2 \quad \text{für alle } \underline{x} \in I\!R^n.$$

Man zeige die positive Definitheit des Schur–Komplements

$$S \;=\; D - B^{\top}A^{-1}B$$

für alle $\underline{x}_2 \in I\!R^{n_2}$.

Kapitel 6

Geometrische Partitionierungen

Die Definition hierarchischer Matrizen beruht im wesentlichen auf einer geeigneten hierarchischen Partitionierung der zugeordneten Indexmenge. Im folgenden werden für eine zunächst beliebig gegebene Funktion f Matrizen A mit Einträgen

$$A[\ell, k] \;=\; f(x_k, y_\ell) \quad \text{für } k, \ell = 1, \ldots, n$$

betrachtet, wobei $\{x_k\}_{k=1}^n, \{y_\ell\}_{\ell=1}^n \subset I\!R^d$ zwei d–dimensionale Punktmengen bezeichnen. Eine hierarchische Partitionierung der Matrix A bzw. der Indexmenge $I = \{1, \ldots, n\}$ entspricht somit einer hierarchischen Partitionierung der Punktmengen $\{x_k\}_{k=1}^n$ bzw. $\{y_\ell\}_{\ell=1}^n$.

6.1 Box–Clustering

Ohne Einschränkung der Allgemeinheit gelte

$$\{x_k\}_{k=1}^n \subset (0,1)^d \;=:\; \Omega_1^0$$

mit der die Punktmenge $\{x_k\}_{k=1}^n$ umgebenden Box Ω_1^0. Diese wird für $\lambda = 1, \ldots, L$ mit einem vorgegebenen Level L rekursiv unterteilt, siehe Abbildung 6.1 für den Fall $d = 2$,

$$\overline{\Omega}_j^{\lambda-1} \;=\; \bigcup_{i=2^d(j-1)+1}^{2^d j} \overline{\Omega}_i^{\lambda}, \quad \text{für } j = 1, \ldots, 2^{d(\lambda-1)}.$$

Aus der Clusterung der Boxen Ω_j^λ folgt nun eine Clusterung der Punktmenge $\{x_k\}_{k=1}^n$. Dabei werden zunächst alle Punkte x_k in den kleinsten Boxen zusammengefaßt,

$$\omega_j^L \;=\; \left\{ x_k \,:\, x_k \in \Omega_j^L \right\}.$$

Zu beachten ist, daß jeder Punkt x_k genau einer kleinsten Box Ω_j^L zugeordnet wird. Entsprechend ergibt sich mit

$$\omega_j^{\lambda-1} \;=\; \bigcup_{i=2^d(j-1)+1}^{2^d j} \omega_i^{\lambda} \quad \text{für } j = 1, \ldots, 2^{d(\lambda-1)}$$

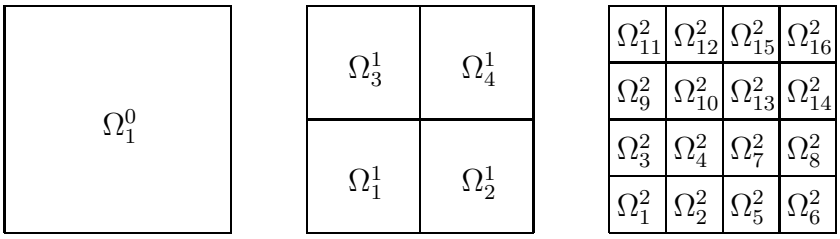


Abbildung 6.1: Hierarchie der Boxen Ω_j^λ für $\lambda = 0, 1, 2$.

eine hierarchische Clusterung der Punktmenge $\{x_k\}_{k=1}^n$, siehe Abbildung 6.2. Die zugehörigen Indexmengen sind durch

$$I_j^\lambda \;=\; \left\{k \,:\, x_k \in \omega_j^\lambda\right\}$$

gegeben.

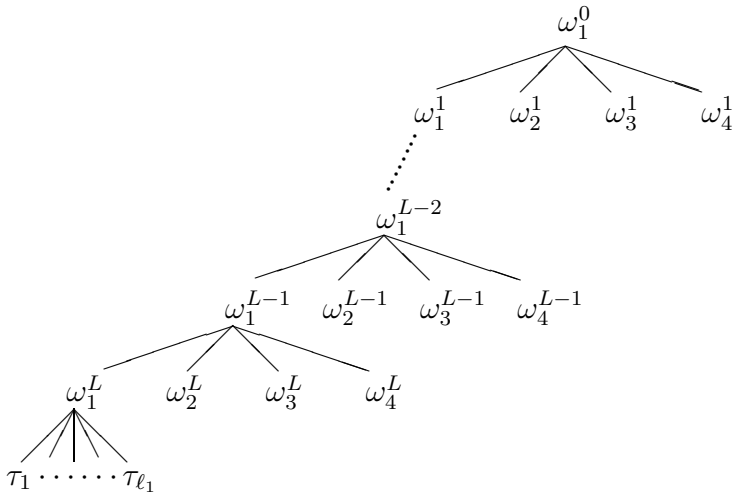


Abbildung 6.2: Baum der Cluster ω_j^λ.

Jedem Punkt x_k kann durch

$$h_k \;=\; \min_{x_\ell \neq x_k} |x_k - x_\ell|$$

eine lokale Maschenweite h_k zugeordnet werden. Dann bezeichnen

$$h_{\min} := \min_{k=1,\ldots,n} h_k, \quad h_{\max} := \max_{k=1,\ldots,n} h_k$$

die minimale bzw. maximale Maschenweite der Punktmenge $\{x_k\}_{k=1}^n$. Die Punktmenge $\{x_k\}_{k=1}^n$ heißt global gleichmäßig, falls

$$\frac{h_{\max}}{h_{\min}} \;\leq\; c_G$$

mit einer Konstanten $c_G \geq 1$ unabhängig von n erfüllt ist. Im folgenden betrachten wir eine global gleichmäßige Punktmenge mit der globalen Maschenweite $h := h_{\max}$. Für die Kantenlängen d_j^λ der Boxen Ω_j^λ ergibt sich andererseits

$$d_j^\lambda \;=\; 2^{-\lambda} \quad \text{für } \lambda = 0, \ldots, L;\; j = 1, \ldots, 2^{d\lambda}.$$

Soll den kleinsten Boxen Ω_j^L nur eine beschränkte Anzahl von Punkten x_k zugeordnet werden, so folgt

$$d_j^L \;=\; 2^{-L} \;\leq\; c\,h$$

bzw.

$$L \;\geq\; \widetilde{c}\,\ln n\,.$$

Die Anzahl der notwendigen Unterteilungen ist somit abhängig von der Dimension n der betrachteten Punktwolke.
Für $d = 2$ wird jedes Cluster Ω_i^λ in vier Söhne, für $d = 3$ in acht Söhne unterteilt. Im Gegensatz dazu wird im folgenden Abschnitt eine Unterteilung in grundsätzlich zwei Söhne eingeführt.

6.2 Bisektionsverfahren

Der Schwerpunkt einer gegebenen Punktmenge $\{x_k\}_{k=1}^n \subset I\!R^d$ ist gegeben durch

$$\hat{x} \;=\; \frac{1}{n}\sum_{k=1}^{n} x_k\,.$$

Die zugehörige Hauptrichtung $w \in I\!R^d$ ist definiert als Lösung des Maximierungsproblems

$$\sum_{k=1}^{n} |(x_k - \hat{x}, w)|^2 \;=\; \max_{v \in I\!R^d, \|v\|_2 = 1} \sum_{k=1}^{n} |(x_k - \hat{x}, v)|^2.$$

Für die Aufteilung der Indexmenge $I = \{1, \ldots, n\}$ folgt dann

$$I_1 \;:=\; \{k \in I \;:\; (x_k - \hat{x}, w) > 0\}\,, \quad I_2 \;:=\; I \backslash I_1.$$

Die Hauptrichtung $w \in I\!R^d$ ergibt sich wegen

$$\begin{aligned}
\sum_{k=1}^{n} |(x_k - \hat{x}, v)|^2 \;&=\; \sum_{k=1}^{n} \left(\sum_{i=1}^{d} [x_{k,i} - \hat{x}_i] v_i \right)^2 \\
&=\; \sum_{k=1}^{n} \sum_{i=1}^{d} \sum_{j=1}^{d} [x_{k,i} - \hat{x}_i][x_{k,j} - \hat{x}_j] v_i v_j \\
&=\; \sum_{i=1}^{d} \sum_{j=1}^{d} v_i v_j \sum_{k=1}^{n} [x_{k,i} - \hat{x}_i][x_{k,j} - \hat{x}_j] \\
&=\; \sum_{i=1}^{d} \sum_{j=1}^{d} K[j,i] v_i v_j \;=\; (Kv, v)
\end{aligned}$$

mit der durch

$$K[j,i] \;=\; \sum_{k=1}^{n} [x_{k,i} - \hat{x}_i][x_{k,j} - \hat{x}_j] \quad \text{für } i,j = 1,\ldots,d$$

definierten Kovarianzmatrix $K \in I\!R^{d\times d}$ als Lösung von

$$(Kw,w) \;=\; \max_{v\in I\!R^d, \|v\|_2=1} (Kv,v) \;=\; \max_{v\in I\!R^d, \|v\|_2=1} \frac{(Kv,v)}{(v,v)} \;=\; \lambda_{\max}(K).$$

Der zu bestimmende Vektor $w \in I\!R^d$ ist also normierter Eigenvektor zum größten Eigenwert $\lambda_{\max}(K)$ der Kovarianzmatrix K.
Der **Algorithmus** zur Partitionierung der Punktmenge $\{x_k\}_{k=1}^n$ mittels des Bisektionsverfahren lautet also:

1. Bestimmung des Schwerpunktes $\hat{x}$ der Punktmenge $\{x_k\}_{k=1}^n$.
2. Aufstellen der Kovarianzmatrix K.
3. Berechnung des maximalen Eigenwertes $\lambda_{\max}(K)$ und des zugehörigen normierten Eigenvektors w der Kovarianzmatrix K.
4. Aufteilen der Indexmenge I in zwei Teilmengen I_1 und I_2.
5. Rekursives Anwenden des Bisektionsverfahrens auf die Teilmengen I_1 und I_2, bis eine vorgegebene Dimension des Clusters erreicht ist.

Daraus resultiert ein Cluster–Baum von Indexmengen I_j^λ mit

$$I_1^0 = I, \quad I_i^{\lambda-1} \;=\; I_{2i-1}^\lambda \cup I_{2i}^\lambda \quad \text{für } i = 1,\ldots,2^{\lambda-1}, \lambda = 1,\ldots,L.$$

Übungsaufgaben

6.1. Gegeben sei die Punktmenge

$$X \;=\; \{(0,0),(1,0),(2,3)\}.$$

Man stelle die zugehörige Kovarianzmatrix auf und bestimme deren Eigenwerte und zugehörigen Eigenvektoren. Anschließend führe man einen Bisektionsschritt durch. Wie lautet der vollständige Cluster–Baum, wenn jedes Cluster auf dem feinsten Level maximal einen Punkt enthält.

6.2. Gegeben sei die Punktmenge

$$X \;=\; \{(0,0),(2,0),(0,2),(2,2)\}.$$

Wie in Aufgabe **11.** führe man einen Bisektionsschritt durch und stelle den zugehörigen Cluster–Baum dar. Was passiert?

6.3. Man beschreibe das Verfahren der einfachen Vektoriteration zur Bestimmung des maximalen Eigenwertes und wende dieses auf die Kovarianzmatrizen der beiden vorherigen Aufgaben an.

6.4. Gegeben seien $N = 4^p$ gleichmässig auf dem Einheitskreis verteilte Punkte $\{x_k\}_{k=1}^N$. Man bestimme die Anzahl L der beim Bisektionsverfahren und beim Quad–Tree–Verfahren entstehenden Level, wenn die kleinsten Cluster genau aus einem Punkt bestehen.

Kapitel 7

Niedrig–Rang–Approximation von Funktionen

Im vorherigen Kapitel 6 wurde vorausgesetzt, daß die Einträge der Matrix A durch die Auswertung einer Funktion f bezüglich zweier endlichdimensionaler Punktmengen $\{x_k\}_{k=1}^n$ bzw. $\{y_\ell\}_{\ell=1}^n$ beschrieben werden kann. Für eine einfachere Darstellung wird im folgenden die Punktmenge $\{y_\ell\}_{\ell=1}^n$ mit der Punktmenge $\{x_k\}_{k=1}^n$ identifiziert. Andernfalls sind die Partitionierungen der zugehörigen Indexmengen getrennt voneinander durchzuführen. Sei

$$A[\ell, k] \;=\; f(x_k, x_\ell) \quad \text{für } k, \ell = 1, \ldots, n.$$

Eine hierarchische Clusterung $\{I_j^\lambda\}$ der gegebenen Punktmenge $\{x_k\}_{k=1}^n$ induziert dann eine hierarchische Partitionierung der Matrix A mit Blöcken

$$A_{ij}^\lambda[\ell, k] \;=\; f(x_k, x_\ell) \quad \text{für } (k, \ell) \in I_i^\lambda \times I_j^\lambda$$

für ein gewisses Level λ. Zu bestimmen bleibt eine Niedrig–Rang–Approximation $\widetilde{A}_{ij}^\lambda$ der gegebenen Block–Matrix A_{ij}^λ. Das in Kapitel 4 beschriebene Verfahren zur Niedrig–Rang–Approximation einer gegebenen Matrix beruht auf deren Singulärwertzerlegung und erfordert somit die Kenntnis aller Matrixeinträge. Damit kann zwar eine optimale Beschreibung der vollbesetzten Matrix erreicht werden, durch die Berechnung aller Matrixeinträge bleibt der Aufwand jedoch quadratisch in der Zahl der Freiheitsgrade. In diesem Kapitel sollen deshalb Approximationsmethoden betrachtet werden, die auch eine effiziente Aufstellung der Niedrig–Rang–Approximation ermöglichen. Ausgangspunkt hierfür ist eine entsprechende Approximation der die Matrixeinträge erzeugenden Funktion $f(x, y)$.
Sei

$$f_\varrho(x, y) \;=\; \sum_{m=0}^{\varrho} f_m^{\lambda,i}(x) g_m^{\lambda,j}(y) \quad \text{für } (x, y) \in \omega_i^\lambda \times \omega_j^\lambda \tag{7.1}$$

für $\varrho \in I\!N$ eine Approximation der Funktion $f(x, y)$, welche einer Fehlerabschätzung

$$|f(x, y) - f_\varrho(x, y)| \;\leq\; c(\eta, \varrho) \quad \text{für alle } (x, y) \in \omega_i^\lambda \times \omega_j^\lambda \tag{7.2}$$

mit einem geeignet gewählten Parameter $\eta \in I\!R$ genügt. Dabei wird

$$\lim_{\varrho \to \infty} c(\eta, \varrho) \;=\; 0$$

vorausgesetzt. Dann folgt für die approximierende Matrix

$$\widetilde{A}_{ij}^{\lambda}[\ell, k] \;=\; f_{\varrho}(x_k, x_\ell) \;=\; \sum_{m=0}^{\varrho} f_m^{\lambda,i}(x_k) g_m^{\lambda,j}(x_\ell)$$

die Niedrig–Rang–Darstellung

$$\widetilde{A}_{ij}^{\lambda} \;=\; \sum_{m=0}^{\varrho} \underline{g}_m^{\lambda,j} \, \underline{f}_m^{\lambda,i,\top}$$

mit

$$f_{m,k}^{\lambda,i} \;=\; f_m^{\lambda,i}(x_k), \quad g_{m,\ell}^{\lambda,j} \;=\; g_m^{\lambda,j}(x_\ell)$$

und es gilt

$$\operatorname{rang} \widetilde{A}_{ij}^{\lambda} \;\leq\; \varrho + 1, \quad Sp(\widetilde{A}_{ij}^{\lambda}) \;=\; (\varrho+1)[n_i^{\lambda} + n_j^{\lambda}].$$

Aus (7.2) ergibt sich für Vektoren $\underline{u} \in I\!R^{n_i^{\lambda}}$ und $\underline{v} \in I\!R^{n_j^{\lambda}}$ die Fehlerabschätzung

$$\begin{aligned}
\left| ((A_{ij}^{\lambda} - \widetilde{A}_{ij}^{\lambda})\underline{u}, \underline{v}) \right| \;&=\; \left| \sum_{k=1}^{n_i^{\lambda}} \sum_{\ell=1}^{n_j^{\lambda}} \left(A_{ij}^{\lambda}[\ell, k] - \widetilde{A}_{ij}^{\lambda}[\ell, k] \right) u_k v_\ell \right| \\
&\leq\; \sum_{k=1}^{n_i^{\lambda}} \sum_{\ell=1}^{n_j^{\lambda}} |f(x_k, x_\ell) - f_{\varrho}(x_k, x_\ell)| \, |u_k| \, |v_\ell| \\
&\leq\; c(\eta, \varrho) \sum_{k=1}^{n_i^{\lambda}} |u_k| \sum_{\ell=1}^{n_j^{\lambda}} |v_\ell| \\
&=\; c(\eta, \varrho) \, \|\underline{u}\|_1 \|\underline{v}\|_1.
\end{aligned}$$

Für geeignet gewählte Parameter ϱ und η übertragen sich somit die Eigenschaften der Matrix A auf die approximierende Matrix $\widetilde{A}$. Zu klären bleibt die Konstruktion der Approximation (7.1) und die Gültigkeit der Fehlerabschätzung (7.2). Im folgenden werden einige mögliche Herleitungen diskutiert, für eine weitergehende Behandlung und entsprechende Literaturhinweise sei hier auf [24, Kapitel 14] verwiesen.

7.1 Darstellung mit Taylor–Reihen

Eine erste Möglichkeit zur Herleitung der allgemeinen Darstellung (7.1) ist die Taylor–Entwicklung der Funktion $f(x, y)$ bezüglich y. Sei $y_j^{\lambda} \in \omega_j^{\lambda}$ beliebig aber fest gegeben. Dann ist

$$f(x, y) \;=\; f\left(x, y_j^{\lambda} + t(y - y_j^{\lambda})\right) \;=:\; F(t)$$

eine skalare Funktion mit der Taylorreihe

$$F(1) \;=\; F(0) + \sum_{n=1}^{p} \frac{1}{n!} \frac{d^n}{dt^n} F(t)_{|t=0} + \frac{1}{p!} \int_0^1 (1-s)^p \frac{d^{p+1}}{ds^{p+1}} F(s) ds \,.$$

Dabei ist

$$\frac{d}{dt} F(t) \;=\; \sum_{k=1}^{d} \frac{\partial}{\partial z_k} f(x,z)_{|z=y_j^\lambda + t(y-y_j^\lambda)} (y_k - y_{j,k}^\lambda)$$

und allgemein

$$\frac{d^n}{dt^n} F(t) \;=\; \sum_{|\alpha|=n} \frac{n!}{\alpha!} (y-y_j^\lambda)^\alpha D_z^\alpha f(x,z)_{|z=y_j^\lambda + t(y-y_j^\lambda)}$$

mit Multiindizes $\alpha = (\alpha_1, \ldots, \alpha_d) \in I\!N_0^d$, $|\alpha| = \sum\limits_{i=1}^{d} \alpha_i$. Damit folgt für $p \in I\!N$ mit

$$f_\varrho(x,y) \;=\; f(x,y_j^\lambda) + \sum_{n=1}^{p} \sum_{|\alpha|=n} \frac{1}{|\alpha|} (y-y_j^\lambda)^\alpha D_z^\alpha f(x,z)_{|z=y_j^\lambda} \tag{7.3}$$

die gewünschte Darstellung (7.1), und für den Fehler der Approximation ergibt sich

$$\begin{aligned}
|f(x,y) - f_\varrho(x,y)| \;&=\; \left| \frac{1}{p!} \int_0^1 (1-s)^p \frac{d^{p+1}}{ds^{p+1}} F(s) ds \right| \\
&=\; \left| \frac{1}{p!} \int_0^1 (1-s)^p \sum_{|\alpha|=p+1} \frac{(p+1)!}{\alpha!} (y-y_j^\lambda)^\alpha D_z^\alpha f(x,z)_{|z=y_j^\lambda + s(y-y_j^\lambda)} ds \right| \\
&\leq\; \sum_{|\alpha|=p+1} \frac{1}{\alpha!} \sup_{y \in \omega_j^\lambda} |y-y_j^\lambda|^{p+1} \sup_{y \in \omega_j^\lambda, |\alpha|=p+1} \left| D_z^\alpha f(x,z)_{|z=y} \right| \\
&\leq\; \sum_{|\alpha|=p+1} \frac{1}{\alpha!} \left[\operatorname{diam} \omega_j^\lambda\right]^{p+1} \sup_{y \in \omega_j^\lambda, |\alpha|=p+1} \left| D_z^\alpha f(x,z)_{|z=y} \right| .
\end{aligned}$$

Als Beispiel wird nun für $d = 2$ die Funktion

$$f(x,y) \;=\; \log|x-y| \quad \text{für } (x,y) \in \omega_i^\lambda \times \omega_j^\lambda$$

betrachtet. Für jedes $n \in [1,p]$ existieren in der Reihendarstellung (7.3) genau $n+1$ Multiindizes $\alpha \in I\!N_0^2$ mit $|\alpha| = n$. Dann ergibt sich für die Anzahl ϱ der Terme in der Reihenentwicklung (7.3)

$$\varrho \;=\; 1 + \sum_{n=1}^{p} (n+1) \;=\; \frac{1}{2}(p+1)(p+2).$$

Für die Herleitung einer Fehlerabschätzung der Approximation (7.3) sind die partiellen Ableitungen von $f(x,y) = \log|x-y|$ bis zur Ordnung $p+1$ zu berechnen. Zunächst ist

$$\frac{\partial}{\partial y_k} f(x,y) \;=\; \frac{\partial}{\partial y_k} \log|x-y| \;=\; \frac{y_k - x_k}{|x-y|^2} \quad \text{für } k = 1,2$$

sowie

$$\frac{\partial^2}{\partial y_k^2} f(x,y) \;=\; \frac{1}{|x-y|^2} - 2\frac{(y_k-x_k)^2}{|x-y|^4}, \quad \frac{\partial^2}{\partial y_1 \partial y_2} f(x,y) \;=\; -2\frac{(y_1-x_1)(y_2-x_2)}{|x-y|^4}.$$

Allgemein gilt für $|\alpha| = q \in I\!N$ die Darstellung

$$D_y^\alpha f(x,y) \;=\; \sum_{|\beta|\le q} a_\beta^q \frac{(x-y)^\beta}{|x-y|^{|\beta|+q}} \tag{7.4}$$

mit gewissen Koeffizienten a_β^q. Daraus folgt

$$\left|D_y^\alpha f(x,y)\right| \;\le\; \sum_{|\beta|\le q} \frac{|a_\beta^q|}{|x-y|^q} \;=\; \frac{c_q}{|x-y|^q}\,.$$

Ein Vergleich mit den ersten und zweiten Ableitungen der Funktion $f(x,y)$ liefert $c_1 = 1$ und $c_2 = 3$. Eine allgemeine Abschätzung der Konstanten c_q für $q \ge 2$ folgt nun durch vollständige Induktion. Aus (7.4) ergibt sich für $i = 1, 2$ und $j \ne i$

$$\begin{aligned}\frac{\partial}{\partial y_i} D_y^\alpha f(x,y) \;&=\; \sum_{|\beta|\le q} a_\beta^q \frac{\partial}{\partial y_i} \frac{(y-x)^\beta}{|x-y|^{|\beta|+q}} \\ &=\; \sum_{|\beta|\le q} a_\beta^q \left[\beta_i \frac{(y_i-x_i)^{\beta_i-1}(y_j-x_j)^{\beta_j}}{|x-y|^{|\beta|+q}} - (|\beta|+q)\frac{(y_i-x_i)^{\beta_i+1}(y_j-x_j)^{\beta_j}}{|x-y|^{|\beta|+q+2}}\right].\end{aligned}$$

Mit

$$|y_i - x_i| \;\le\; |x-y|, \quad \beta_i \;\le\; |\beta| \;\le\; q, \quad \beta_i + \beta_j \le |\beta| \quad \text{für } i \ne j$$

folgt daraus die Abschätzung

$$\left|\frac{\partial}{\partial y_i} D_y^\alpha f(x,y)\right| \;\le\; \frac{3q}{|x-y|^{q+1}} \sum_{|\beta|\le q} \left|a_\beta^q\right| \;=\; \frac{3qc_q}{|x-y|^{q+1}} \;=\; \frac{c_{q+1}}{|x-y|^{q+1}}$$

und somit

$$c_{q+1} \;=\; 3qc_q \;=\; 3^q\, q!\,.$$

Für $|\alpha| = p \in I\!N$ ist somit

$$\left|D_y^\alpha f(x,y)\right| \;\le\; \frac{3^{p-1}(p-1)!}{|x-y|^p} \quad \text{für } (x,y) \in \omega_i^\lambda \times \omega_j^\lambda.$$

Damit ergibt sich

$$\begin{aligned}|f(x,y) - f_\varrho(x,y)| \;&\le\; \sum_{|\alpha|=p+1} \frac{1}{\alpha!} \left[\operatorname{diam} \omega_j^\lambda\right]^{p+1} \sup_{y\in\omega_j^\lambda, |\alpha|=p+1} \left|D_z^\alpha f(x,z)_{|z=y}\right| \\ &\le\; \left[\operatorname{diam} \omega_j^\lambda\right]^{p+1} \sum_{|\alpha|=p+1} \frac{1}{\alpha!} \sup_{y\in\omega_j^\lambda} \frac{3^p p!}{|x-y|^{p+1}} \\ &\le\; 3^p p! \left(\frac{\operatorname{diam} \omega_j^\lambda}{\operatorname{dist}(\omega_i^\lambda, \omega_j^\lambda)}\right)^{p+1} \sum_{|\alpha|=p+1} \frac{1}{\alpha!}\,.\end{aligned}$$

Mit der **Zulässigkeitsbedingung**

$$\operatorname{dist}(\omega_i^\lambda, \omega_j^\lambda) \geq \eta \operatorname{diam} \omega_j^\lambda, \quad \eta > 1,$$

und dem Ergebnis aus Übungsaufgabe 7.1 folgt schließlich die Fehlerabschätzung

$$|f(x,y) - f_\varrho(x,y)| \leq \frac{1}{3}\frac{1}{p+1}\left(\frac{6}{\eta}\right)^{p+1} \quad \text{für } (x,y) \in \omega_i^\lambda \times \omega_j^\lambda$$

mit der Forderung $\eta > 6$.

7.2 Explizite Reihendarstellung

Die im vorherigen Abschnitt beschriebene Taylorentwicklung erfordert die explizite Berechnung aller auftretenden Ableitungen der Funktion $f(x,y)$. In vielen Anwendungen erlauben die betrachtenden Funktionen jedoch eine explizite Reihendarstellung. Dies soll hier am Beispiel $f(x,y) = \log|x-y|$ betrachtet werden.
Für $(x,y) \in \omega_i^\lambda \times \omega_j^\lambda$ sei y_j^λ das Zentrum der Punktmenge ω_j^λ. Dann ist

$$f(x,y) = \log|x-y| = \log|(x-y_j^\lambda) - (y-y_j^\lambda)| = \operatorname{Re}(\log(z-z_0))$$

mit den komplexen Argumenten

$$z = |x-y_j^\lambda| e^{i\varphi(x-y_j^\lambda)}, \quad z_0 = |y-y_j^\lambda| e^{i\varphi(y-y_j^\lambda)}.$$

Aus der Zulässigkeitsbedingung

$$\operatorname{dist}(\omega_i^\lambda, \omega_j^\lambda) \geq \eta \operatorname{diam} \omega_j^\lambda, \quad \eta > 1,$$

folgt

$$\frac{|z_0|}{|z|} = \frac{|y-y_j^\lambda|}{|x-y_j^\lambda|} \leq \frac{\operatorname{diam}\omega_j^\lambda}{\operatorname{dist}(\omega_i^\lambda, \omega_j^\lambda)} \leq \frac{1}{\eta} < 1$$

und somit die Gültigkeit der Reihenentwicklung

$$\log(z-z_0) = \log z - \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{z_0}{z}\right)^n.$$

Für $p \in \mathbb{N}$ wird durch

$$f_\varrho(x,y) = \operatorname{Re}\left(\log z - \sum_{n=1}^{p} \frac{1}{n}\left[\frac{z_0}{z}\right]^n\right)$$

eine Approximation der Funktion $f(x,y)$ erklärt, für die die folgende Fehlerabschätzung gilt:

$$|f(x,y) - f_\varrho(x,y)| = \left|\operatorname{Re}\left(\sum_{n=p+1}^{\infty} \frac{1}{n}\left[\frac{z_0}{z}\right]^n\right)\right| \leq \sum_{n=p+1}^{\infty} \frac{1}{n}\left(\frac{1}{\eta}\right)^n \leq \frac{1}{p+1}\frac{1}{\eta-1}\left(\frac{1}{\eta}\right)^p.$$

Wegen

$$\left(\frac{z_0}{z}\right)^n \;=\; z_0^n z^{-n} \;=\; \frac{|y-y_j^\lambda|^n}{|x-y_j^\lambda|^n} e^{in\varphi(y-y_j^\lambda)} e^{-in\varphi(x-y_j^\lambda)}$$

und

$$\mathrm{Re}\left(\left[\frac{z_0}{z}\right]^n\right) \;=\; \frac{|y-y_j^\lambda|^n}{|x-y_j^\lambda|^n}\left[\cos n\varphi(y-y_j^\lambda)\cos n\varphi(x-y_j^\lambda) + \sin n\varphi(y-y_j^\lambda)\sin n\varphi(x-y_j^\lambda)\right]$$

ergibt sich

$$\begin{aligned} f_\varrho(x,y) \;&=\; \log|x-y_j^\lambda| - \sum_{n=1}^{p}\frac{1}{n}|y-y_j^\lambda|^n \cos n\varphi(y-y_j^\lambda)\frac{\cos n\varphi(x-y_j^\lambda)}{|x-y_j^\lambda|^n} \\ &\quad - \sum_{n=1}^{p}\frac{1}{n}|y-y_j^\lambda|^n \sin n\varphi(y-y_j^\lambda)\frac{\sin n\varphi(x-y_j^\lambda)}{|x-y_j^\lambda|^n} \end{aligned}$$

und somit die Darstellung (7.1) mit

$$\varrho \;=\; 2p+1.$$

7.3 Adaptive Cross–Approximation

Die bisher beschriebenen Approximationen $f_\varrho(x,y)$ einer gegebenen Funktion $f(x,y)$ erfordern entweder die Kenntnis einer geeigneten Reihenentwicklung oder die Berechnung von Ableitungen der Fundamentallösung. Ziel ist deshalb die Herleitung von Approximationen, welche nur auf der Auswertung der gegebenen Funktion $f(x,y)$ in Interpolationspunkten basieren. Eine Möglichkeit besteht dabei in der Interpolation mit Tschebyscheff–Polynomen, siehe hierzu die entsprechenden Übungsaufgaben. Der hier betrachtete Algorithmus zur Approximation der Funktion $f^(x,y)$ wurde erstmals von Tyrtyshnikov in [27] beschrieben. Die hier gegebene Darstellung orientiert sich auch an [1, 2], siehe auch [24].

Seien ω_i^λ und ω_j^λ ein Paar zueinander zulässiger Cluster. Zur Approximation der Funktion $f(x,y)$ für $(x,y) \in \omega_i^\lambda \times \omega_j^\lambda$ werden zwei Funktionenfolgen $s_k(x,y)$ und $r_k(x,y)$ konstruiert. Dabei ist $r_k(x,y)$ das jeweilige Residuum der zugehörigen Approximation $s_k(x,y)$. Zur Initialisierung werden zunächst

$$s_0(x,y) \;:=\; 0, \quad r_0(x,y) \;:=\; f(x,y)$$

gesetzt. Für $k = 1,2,\ldots,\varrho$ sei $(x_k,y_k) \in \omega_i^\lambda \times \omega_j^\lambda$ ein Punktpaar, so daß das Residuum nicht verschwindet, d.h. $\alpha_k := r_{k-1}(x_k,y_k) \neq 0$. Dann ist

$$s_k(x,y) \;:=\; s_{k-1}(x,y) + \frac{1}{\alpha_k} r_{k-1}(x,y_k) r_{k-1}(x_k,y), \tag{7.5}$$

$$r_k(x,y) \;:=\; r_{k-1}(x,y) - \frac{1}{\alpha_k} r_{k-1}(x,y_k) r_{k-1}(x_k,y). \tag{7.6}$$

Für $\varrho \in I\!N_0$ definiert schließlich

$$f_\varrho(x,y) \;=\; s_\varrho(x,y) \;=\; \sum_{k=1}^{\varrho} \frac{r_{k-1}(x,y_k)r_{k-1}(x_k,y)}{r_{k-1}(x_k,y_k)} \quad \text{für } (x,y) \in \omega_i^\lambda \times \omega_j^\lambda \tag{7.7}$$

eine Approximation der gegebenen Funktion $f(x,y)$ bezüglich dem zulässigen Clusterpaar $(\omega_i^\lambda, \omega_j^\lambda)$. Für die durch (7.5) und (7.6) gegebene Rekursion gelten die folgenden Eigenschaften.

Lemma 7.1 *Für alle* $0 \le k \le \varrho$ *und* $(x,y) \in \omega_i^\lambda \times \omega_j^\lambda$ *gilt*

$$f(x,y) \;=\; s_k(x,y) + r_k(x,y) \tag{7.8}$$

sowie

$$r_k(x,y_j) \;=\; 0 \quad \textit{für alle } 1 \le j \le k \tag{7.9}$$

bzw.

$$r_k(x_i,y) \;=\; 0 \quad \textit{für alle } 1 \le i \le k. \tag{7.10}$$

Beweis: Die Addition der Entwicklungsvorschriften (7.5) und (7.6) ergibt

$$r_k(x,y) + s_k(x,y) \;=\; r_{k-1}(x,y) + s_{k-1}(x,y)$$

für alle $k = 1, \ldots, \varrho$, und somit erhält man durch rekursive Anwendung

$$r_k(x,y) + s_k(x,y) = r_0(x,y) + s_0(x,y) = f(x,y).$$

Die zweite Behauptung ergibt sich durch vollständige Induktion nach $k = 1, \ldots, \varrho$. Für $k = 1$ ist

$$r_1(x,y_1) \;=\; r_0(x,y_1) - \frac{1}{r_0(x_1,y_1)} r_0(x,y_1) r_0(x_1,y_1) \;=\; 0.$$

Somit sei $r_k(x,y_j) = 0$ für $k = 1,2,\ldots,\varrho$ und $j = 1,\ldots,k$ erfüllt. Dann ist

$$r_{k+1}(x,y_j) \;=\; r_k(x,y_j) - \frac{1}{\alpha_{k+1}} r_k(x,y_{k+1}) r_k(x_{k+1},y_j) \;=\; 0,$$

d.h. es gilt $r_{k+1}(x,y_j) = 0$ für alle $j = 1,\ldots,k$. Schließlich ist nach (7.6)

$$r_{k+1}(x,y_{k+1}) \;=\; r_k(x,y_{k+1}) - \frac{1}{r_k(x_{k+1},y_{k+1})} r_k(x,y_{k+1}) r_k(x_{k+1},y_{k+1}) \;=\; 0.$$

Die letzte Behauptung folgt analog. ∎

Einsetzen von $x = x_i$ für $i = 1,\ldots,k$ in (7.9) liefert

$$r_k(x_i,y_j) \;=\; 0 \quad \text{für alle } i,j = 1,\ldots,k,$$

und wegen (7.8) folgt

$$s_k(x_i, y_j) \;=\; f(x_i, y_j) \quad \text{für alle } i, j = 1, \ldots, k; k = 1, \ldots, \varrho.$$

Die Approximationen $s_k(x, y)$ **interpolieren** also die gegebene Funktion $f(x, y)$ in den Stützstellen (x_i, y_j) für $i, j = 1, \ldots, k$.

Die Abschätzung des Residuums $r_k(x, y)$ für $(x, y) \in \omega_i^\lambda \times \omega_j^\lambda$ erfolgt durch Vergleich mit dem Fehler der Interpolation mit Lagrange–Polynomen, siehe hierzu [2, 24]. Die Wahl der Interpolationsknoten (x_k, y_k) erfolgt dann gemäß der Bedingung

$$|r_{k-1}(x_k, y_k)| \;\geq\; |r_{k-1}(x, y_k)| \quad \text{für alle } x \in \omega_i^\lambda.$$

Der hier beschriebene Algorithmus der adaptiven Cross–Approximation einer skalaren Funktion kann auch direkt auf die Herleitung einer Niedrigrang–Approximation einer Matrix übertragen werden [1, 4].

Übungsaufgaben

7.1. Man zeige für $\alpha \in \mathbb{N}_0^2$ und $p \in \mathbb{N}$

$$\sum_{|\alpha|=p+1} \frac{1}{\alpha!} \;=\; \frac{2^{p+1}}{(p+1)!}\,.$$

7.2. Betrachtet werden die rekursiv durch

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x) \quad \text{für } k = 1, 2, \ldots$$

definierten Tschebyscheff–Polynome.

Für $x \in [-1, 1]$ beweise man die alternative Darstellung

$$T_k(x) \;=\; \cos\left(k \,\arccos x\right).$$

Wie lauten die Nullstellen x_i^k von $T_k(x)$?

7.3. Man beweise die Orthogonalitätseigenschaft der Tschebyscheff–Polynome,

$$\int_{-1}^{1} \frac{T_k(x)T_\ell(x)}{\sqrt{1-x^2}}dx \;=\; \begin{cases} 0 & \text{für } k \neq \ell, \\ \pi/2 & \text{für } k = \ell \neq 0, \\ \pi & \text{für } k = \ell = 0. \end{cases}$$

Hinweis: Man benutze das Ergebnis aus Aufgabe **7.2**.

7.4. Für $k + \ell < 2n$ zeige man

$$\sum_{i=1}^{n} T_k(x_i^n)T_\ell(x_i^n) \;=\; \begin{cases} 0 & \text{für } k \neq \ell, \\ n/2 & \text{für } k = \ell \neq 0, \\ n & \text{für } k = \ell = 0. \end{cases}$$

Hinweis: Man benutze Aufgabe **7.3** und eine geeignete numerische Integrationsformel.

7.5. Für eine auf dem Intervall $[-1,+1]$ hinreichend glatte Funktion f bestimme man die Interpolierende $f_N = \sum_{k=0}^{N} a_k T_k(x)$ mit $f_N(x_i^{N+1}) = f(x_i^{N+1})$ und $i = 1,\ldots,N+1$. Man finde eine Darstellung der Koeffizienten a_k und beweise die Fehlerabschätzung

$$\max_{x\in[-1,1]} |f(x) - f_N(x)| \;\leq\; \frac{2^{1-N}}{N!} \max_{x\in[-1,1]} |f^{(N)}(x)|.$$

Hinweis: Man betrachte die Taylor–Entwicklung von f und den führenden Koeffizienten der Tschebyscheff–Polynome.

Kapitel 8

Anwendungen in der FEM

In diesem Kapitel werden Anwendungen von hierarchischen Matrizen für einfache Modellprobleme mit finiten Elementen in einer Raumdimension betrachtet. Diese Überlegungen lassen sich entsprechend auf zwei– und dreidimensionale Randwertprobleme übertragen.

8.1 Ansatzräume

Für $n \in \mathbb{N}$ und der Maschenweite $h := (n+1)^{-1}$ wird für das offene Intervall $\Omega = (0,1)$ eine gleichmäßige Unterteilung

$$\overline{\Omega} = [0,1] = \bigcup_{k=0}^{n} \overline{\tau}_k$$

in finite Elemente $\tau_k = (x_k, x_{k+1})$ mit Stützstellen $x_k = kh$ für $k = 0, 1, \ldots, n+1$ betrachtet. Bezüglich dieser gleichmäßigen Unterteilung wird der Raum der stückweise linearen Basisfunktionen eingeführt, d.h. für $k = 0, \ldots, n+1$ sei die zum Knoten x_k zugehörige Basisfunktion erklärt durch die Vorschrift

$$\varphi_k(x) = \begin{cases} 1 & \text{für } x = x_k, \\ 0 & \text{für } x = x_\ell \neq x_k, \\ \text{linear} & \text{sonst.} \end{cases}$$

Daraus ergibt sich

$$\varphi_k(x) = \begin{cases} \dfrac{1}{h}(x - x_{k-1}) & \text{für } x \in [x_{k-1}, x_k), \\ \dfrac{1}{h}(x_{k+1} - x) & \text{für } x \in [x_k, x_{k+1}], \\ 0 & \text{sonst.} \end{cases}$$

8.2 L_2–Projektion

Für eine im offenen Intervall $\Omega = (0,1)$ gegebene Funktion $u(x)$ ist die stückweise lineare Approximation

$$u_h(x) \;=\; \sum_{k=0}^{n+1} u_k \varphi_k(x)$$

als Lösung der Minimierungsaufgabe

$$F(\underline{u}) \;=\; \min_{\underline{v}\in I\!R^{n+2}} F(\underline{v})$$

mit dem Funktional

$$F(\underline{v}) \;=\; \int_0^1 [v_h(x) - u(x)]^2 dx$$

zu bestimmen. Es ist

$$\begin{aligned}
F(\underline{v}) &= \int_0^1 [v_h(x) - u(x)]^2 dx \\
&= \int_0^1 [\sum_{k=0}^{n+1} v_k \varphi_k(x) - u(x)]^2 dx \\
&= \int_0^1 \left[\sum_{k=0}^{n+1}\sum_{\ell=0}^{n+1} v_k v_\ell \varphi_k(x)\varphi_\ell(x) - 2\sum_{k=0}^{n+1} v_k \varphi_k(x) u(x) + [u(x)]^2 \right] dx \\
&= \sum_{k=0}^{n+1}\sum_{\ell=0}^{n+1} v_k v_\ell \int_0^1 \varphi_k(x)\varphi_\ell(x)dx - 2\sum_{k=0}^{n+1} v_k \int_0^1 u(x)\varphi_k(x)dx + \int_0^1 [u(x)]^2 dx.
\end{aligned}$$

Aus der für ein Minimum notwendigen Bedingung

$$\frac{d}{dv_j} F(\underline{v})_{|\underline{v}=\underline{u}} \;\stackrel{!}{=}\; 0 \quad \text{für } j = 0,\ldots,n+1$$

folgt

$$2\sum_{k=0}^{n+1} u_k \int_0^1 \varphi_k(x)\varphi_j(x)dx - 2\int_0^1 u(x)\varphi_j(x)dx \;\stackrel{!}{=}\; 0$$

für $j = 0,\ldots,n+1$. Der Vektor $\underline{u} \in I\!R^{n+2}$ der Zerlegungskoeffizienten u_k ergibt sich also als Lösung des linearen Gleichungssystems

$$\sum_{k=0}^{n+1} u_k \int_0^1 \varphi_k(x)\varphi_j(x)dx \;=\; \int_0^1 u(x)\varphi_j(x)dx \quad \text{für } j = 0,\ldots,n+1.$$

Die eindeutig bestimmte Näherungsfunktion $u_h = Q_h u$ wird als L_2–Projektion der gegebenen Funktion u auf den Raum der stückweise linearen Funktionen bezeichnet.

Mit den Einträgen für die Massematrix M_h und den aus der gegebenen Funktion u berechenbaren Vektor $\underline{g}$,

$$M_h[j,k] \;=\; \int_0^1 \varphi_k(x)\varphi_j(x)dx, \quad g_j \;=\; \int_0^1 u(x)\varphi_j(x)dx$$

für $k, j = 0, \ldots, n+1$, ist dies äquivalent zur Bestimmung von $\underline{u} \in I\!R^{n+2}$ als Lösung des linearen Gleichungssystems

$$M_h \underline{u} \;=\; \underline{g}.$$

Für die Diagonaleinträge $M_h[k,k]$ und $k = 1, \ldots, n$ ergibt sich

$$\begin{aligned} M_h[k,k] \;&=\; \int_0^1 [\varphi_k(x)]^2 dx \;=\; \frac{1}{h^2}\int_{x_{k-1}}^{x_k} [x - x_{k-1}]^2 dx + \frac{1}{h^2}\int_{x_k}^{x_{k+1}} [x_{k+1} - x]^2 dx \\ &=\; \frac{2}{h^2}\int_0^h s^2 ds \;=\; \frac{2}{3}h\,. \end{aligned}$$

Entsprechend ergibt sich für $k = 0$ bzw. $k = n+1$

$$M_h[0,0] \;=\; \int_0^1 [\varphi_0(x)]^2 dx \;=\; \frac{1}{h^2}\int_{x_0}^{x_1} [x_1 - x]^2 dx \;=\; \frac{1}{h^2}\int_0^h s^2 ds \;=\; \frac{1}{3}h \;=\; M_h[n+1, n+1].$$

Für die Nebendiagonalelemente mit $j = k \pm 1$ folgt

$$\begin{aligned} M_h[k+1,k] \;&=\; \int_0^1 \varphi_k(x)\varphi_{k+1}(x)dx \;=\; \frac{1}{h^2}\int_{x_k}^{x_{k+1}} (x_{k+1} - x)(x - x_{k+1})dx \\ &=\; \frac{1}{h^2}\int_0^h (h-s)s ds \;=\; \frac{1}{6}h\,. \end{aligned}$$

Damit ist die Massematrix M_h gegeben durch

$$M_h \;=\; \frac{h}{6}\begin{pmatrix} 2 & 1 & & & & \\ 1 & 4 & 1 & & & \\ & 1 & \ddots & \ddots & & \\ & & \ddots & \ddots & 1 & \\ & & & 1 & 4 & 1 \\ & & & & 1 & 2 \end{pmatrix} \in I\!R^{(n+2)\times(n+2)}.$$

Aufgrund der lokalen Basisfunktionen sind die Massematrizen M_h für alle $n \in I\!N$ Tridiagonalmatrizen, welche mittels der Cholesky–Zerlegung eine Faktorisierung in eine untere und obere Dreiecksmatrix mit Tridiagonalgestalt erlauben.

Am Beispiel der Massematrizen M_h sollen hier jedoch die Darstellung als hierarchische Matrix und die Invertierung der Massematrix als hierarchische Matrix betrachtet werden. Insbesondere entspricht die Massematrix genau der in Abschnitt 5.4 beschriebenen Situation, daß die Nebendiagonalblöcke genau ein Nichtnullelement enthalten. Damit kann die inverse Massematrix exakt als hierarchische Matrix mit Rang 1 Matrizen in den Nebendiagonalblöcken beschrieben werden.
Für das Beispiel $n = 6$ soll diese Berechnung explizit durchgeführt werden, dann ist

$$M_h \;=\; \frac{1}{48}\left(\begin{array}{cc|cc|cc} 2 & 1 & & & & \\ 1 & 4 & 1 & & & \\ \hline & 1 & 4 & 1 & & \\ & & 1 & 4 & 1 & \\ \hline & & & 1 & 4 & 1 \\ & & & & 1 & 4 \end{array}\right) \;=\; \frac{1}{48}\left(\begin{array}{cc} M_{11}^1 & M_{12}^1 \\ M_{21}^1 & M_{22}^1 \end{array}\right)$$

eine hierarchische Matrix, die der in Beispiel 3.2 betrachteten Block–Partitionierung mit Rang 1 Matrizen in den Nebendiagonalblöcken entspricht. Für die Berechnung der inversen Massematrix M_h^{-1} soll der in Abschnitt 5.4 beschriebene Algorithmus rekursiv angewendet werden.
Allgemein ist die Inverse einer durch die Block–Darstellung

$$A \;=\; \left(\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array}\right)$$

gegebenen Matrix bestimmt durch, vergleiche Abschnitt 5.4,

$$A^{-1} \;=\; \left(\begin{array}{cc} A_{11}^{-1} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{array}\right) \tag{8.1}$$

mit dem Schur–Komplement

$$S \;=\; A_{22} - A_{21}A_{11}^{-1}A_{12}.$$

Für die Invertierung von

$$M_h \;=\; M_h^0 \;=\; \left(\begin{array}{cc} M_{11}^1 & M_{12}^1 \\ M_{21}^1 & M_{22}^1 \end{array}\right)$$

ist also zunächst die Inverse der Block–Matrix

$$M_{11}^1 \;=\; \left(\begin{array}{ccc} 2 & 1 & \\ 1 & 4 & 1 \\ & 1 & 4 & 1 \\ & & 1 & 4 \end{array}\right) \;=\; \left(\begin{array}{cc} M_{11}^2 & M_{12}^2 \\ M_{21}^2 & M_{22}^2 \end{array}\right)$$

mit den Teil–Blöcken

$$M_{11}^2 = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix},\ M_{22}^2 = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix},\ M_{12}^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix},\ M_{21}^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}$$

zu berechnen. Für die Invertierung von M_{11}^1 wird wieder Formel (8.1) angewendet. Für die Inverse von M_{11}^2 ergibt sich

$$M_{11}^{2,-1} = \frac{1}{7}\begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$

und für das Schur–Komplement S_{22}^2 der Matrix M_{11}^1 folgt

$$\begin{aligned} S_{22}^2 &= M_{22}^2 - M_{21}^2 M_{11}^{2,-1} M_{12}^2 \\ &= \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix} - \frac{1}{7}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} \\ &= \frac{1}{7}\begin{pmatrix} 26 & 7 \\ 7 & 28 \end{pmatrix} \end{aligned}$$

mit der Inversen

$$S_{22}^{2,-1} = \frac{1}{97}\begin{pmatrix} 28 & -7 \\ -7 & 26 \end{pmatrix}.$$

Die Inverse von M_{11}^1 ist gemäß (8.1) gegeben durch

$$M_{11}^{1,-1} = \begin{pmatrix} M_{11}^{2,-1} + M_{11}^{2,-1} M_{12}^2 S_{22}^{2,-1} M_{21}^2 M_{11}^{2,-1} & -M_{11}^{2,-1} M_{12}^2 S_{22}^{2,-1} \\ -S_{22}^{2,-1} M_{21}^2 M_{11}^{2,-1} & S_{22}^{2,-1} \end{pmatrix}.$$

Dabei sind

$$\begin{aligned} M_{11}^{2,-1} M_{12}^2 S_{22}^{2,-1} &= \frac{1}{7 \cdot 97}\begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} 28 & -7 \\ -7 & 26 \end{pmatrix} \\ &= \frac{1}{97}\begin{pmatrix} -1 \\ 2 \end{pmatrix}\begin{pmatrix} 4 & -1 \end{pmatrix} \end{aligned}$$

und

$$\begin{aligned} B &= M_{11}^{2,-1} + M_{11}^{2,-1} M_{12}^2 S_{22}^{2,-1} M_{21}^2 M_{11}^{2,-1} \\ &= \frac{1}{7}\begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix} + \frac{1}{7 \cdot 97}\begin{pmatrix} -1 \\ 2 \end{pmatrix}\begin{pmatrix} 4 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix} \\ &= \frac{1}{7}\begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix} + \frac{4}{7 \cdot 97}\begin{pmatrix} -1 \\ 2 \end{pmatrix}\begin{pmatrix} -1 & 2 \end{pmatrix} \\ &= \frac{1}{7 \cdot 97}\begin{pmatrix} 392 & -105 \\ -105 & 210 \end{pmatrix} = \frac{1}{97}\begin{pmatrix} 56 & -15 \\ -15 & 30 \end{pmatrix}. \end{aligned}$$

Insgesamt ist also

$$M_{11}^{1,-1} = \frac{1}{97}\begin{pmatrix} \begin{pmatrix} 56 & -15 \\ -15 & 30 \end{pmatrix} & \begin{pmatrix} 1 \\ -2 \end{pmatrix}\begin{pmatrix} 4 & -1 \end{pmatrix} \\ \begin{pmatrix} 4 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & -2 \end{pmatrix} & \begin{pmatrix} 28 & -7 \\ -7 & 26 \end{pmatrix} \end{pmatrix}.$$

Die Anwendung von (8.1) zur Berechnung der Inversen von M_h verlangt nun die Auswertung des Schur–Komplements

$$\begin{aligned} S_{22}^1 &= M_{22}^1 - M_{21}^1 M_{11}^{1,-1} M_{12}^1 \\ &= \begin{pmatrix} 4 & 1 & & \\ 1 & 4 & 1 & \\ & 1 & 4 & 1 \\ & & 1 & 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} M_{11}^{1,-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 1 & & \\ 1 & 4 & 1 & \\ & 1 & 4 & 1 \\ & & 1 & 2 \end{pmatrix} - \frac{26}{97}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \widetilde{M}_{33}^2 & M_{34}^2 \\ M_{43}^2 & M_{44}^2 \end{pmatrix} \end{aligned}$$

mit den Block–Matrizen

$$\widetilde{M}_{33}^2 = \frac{1}{97}\begin{pmatrix} 362 & 97 \\ 97 & 388 \end{pmatrix}, M_{44}^2 = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}, M_{34}^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix}, M_{43}^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}.$$

Für die Berechnung der Inversen von S_{22}^1 via (8.1) ist zunächst

$$\widetilde{M}_{33}^{2,-1} = \frac{1}{1351}\begin{pmatrix} 388 & -97 \\ -97 & 362 \end{pmatrix}.$$

Für das Schur–Komplement S_{44}^2 von S_{22}^1 folgt dann

$$\begin{aligned} S_{44}^2 &= M_{44}^2 - M_{43}^2 \widetilde{M}_{33}^{2,-1} M_{34}^2 \\ &= \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} - \frac{1}{1351}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} 388 & -97 \\ -97 & 362 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} - \frac{362}{1351}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} \\ &= \frac{1}{1351}\begin{pmatrix} 5042 & 1351 \\ 1351 & 2702 \end{pmatrix}, \end{aligned}$$

und für die Inverse ergibt sich

$$S_{44}^{2,-1} = \frac{1}{8733}\begin{pmatrix} 2702 & -1351 \\ -1351 & 5042 \end{pmatrix}.$$

Die Inverse von S_{22}^1 ist dann gemäß (8.1) gegeben durch

$$S_{22}^{1,-1} = \begin{pmatrix} \widetilde{M}_{33}^{2,-1} + \widetilde{M}_{33}^{2,-1} M_{34}^2 S_{44}^{2,-1} M_{43}^2 \widetilde{M}_{33}^{2,-1} & -\widetilde{M}_{33}^{2,-1} M_{34}^2 S_{44}^{2,-1} \\ -S_{44}^{2,-1} M_{43}^2 \widetilde{M}_{33}^{2,-1} & S_{44}^{2,-1} \end{pmatrix}.$$

Mit

$$\begin{aligned} \widetilde{M}_{33}^{2,-1} M_{34}^2 S_{44}^{2,-1} &= \frac{1}{1351 \cdot 8733} \begin{pmatrix} 388 & -97 \\ -97 & 362 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 2702 & -1351 \\ -1351 & 5042 \end{pmatrix} \\ &= \frac{1}{8733} \begin{pmatrix} -97 \\ 362 \end{pmatrix} \begin{pmatrix} 2 & -1 \end{pmatrix} \end{aligned}$$

und

$$\begin{aligned} B &= \widetilde{M}_{33}^{2,-1} + \widetilde{M}_{33}^{2,-1} M_{34}^2 S_{44}^{2,-1} M_{43}^2 \widetilde{M}_{33}^{2,-1} \\ &= \frac{1}{1351} \begin{pmatrix} 388 & -97 \\ -97 & 362 \end{pmatrix} + \frac{1}{1351 \cdot 8733} \begin{pmatrix} -97 \\ 362 \end{pmatrix} \begin{pmatrix} 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 388 & -97 \\ -97 & 362 \end{pmatrix} \\ &= \frac{1}{1351} \begin{pmatrix} 388 & -97 \\ -97 & 362 \end{pmatrix} + \frac{2}{1351 \cdot 8733} \begin{pmatrix} -97 \\ 362 \end{pmatrix} \begin{pmatrix} -97 & 362 \end{pmatrix} \\ &= \frac{1}{8733} \begin{pmatrix} 2522 & -679 \\ -679 & 2534 \end{pmatrix} \end{aligned}$$

ergibt sich insgesamt

$$S_{22}^{1,-1} = \frac{1}{8733} \begin{pmatrix} 2522 & -679 & \begin{pmatrix} 97 \\ -362 \end{pmatrix} \begin{pmatrix} 2 & -1 \end{pmatrix} \\ -679 & 2534 & \\ \begin{pmatrix} 2 \\ -1 \end{pmatrix} \begin{pmatrix} 97 & -362 \end{pmatrix} & & \begin{matrix} 2702 & -1351 \\ -1351 & 5042 \end{matrix} \end{pmatrix}.$$

Mit (8.1) ergibt sich jetzt die inverse Massematrix aus

$$M_h^{-1} = 48 \begin{pmatrix} M_{11}^{1,-1} + M_{11}^{1,-1} M_{12}^1 S_{22}^{1,-1} M_{21}^1 M_{11}^{1,-1} & -M_{11}^{1,-1} M_{12}^1 S_{22}^{1,-1} \\ -S_{22}^{1,-1} M_{21}^1 M_{11}^{1,-1} & S_{22}^{1,-1} \end{pmatrix}.$$

Mit

$$\begin{aligned} M_{11}^{1,-1} A_{12}^1 S_{22}^{1,-1} &= M_{11}^{1,-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} S_{22}^{1,-1} \\ &= \frac{1}{8733} \begin{pmatrix} -1 \\ 2 \\ -7 \\ 26 \end{pmatrix} \begin{pmatrix} 26 & -7 & 2 & -1 \end{pmatrix} \end{aligned}$$

und

$$
\begin{aligned}
B &= M_{11}^{1,-1} + M_{11}^{1,-1} M_{12}^{1} S_{22}^{1,-1} M_{21}^{1} M_{11}^{1,-1} \\
&= M_{11}^{1,-1} + \frac{1}{8733} \begin{pmatrix} -1 \\ 2 \\ -7 \\ 26 \end{pmatrix} \begin{pmatrix} 26 & -7 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} M_{11}^{1,-1} \\
&= M_{11}^{1,-1} + \frac{26}{97 \cdot 8733} \begin{pmatrix} -1 \\ 2 \\ -7 \\ 26 \end{pmatrix} \begin{pmatrix} -1 & 2 & -7 & 26 \end{pmatrix} \\
&= \frac{1}{8733} \begin{pmatrix} \begin{matrix} 5042 & -1351 \\ -1351 & 2702 \end{matrix} & \begin{pmatrix} -1 \\ 2 \end{pmatrix} \begin{pmatrix} -362 & 97 \end{pmatrix} \\ \begin{pmatrix} -362 \\ 97 \end{pmatrix} \begin{pmatrix} -1 & 2 \end{pmatrix} & \begin{matrix} 2534 & -679 \\ -679 & 2522 \end{matrix} \end{pmatrix}
\end{aligned}
$$

ist insgesamt

$$
M_h^{-1} = \frac{48}{8733} \begin{pmatrix} \begin{matrix} 5042 & -1351 \\ -1351 & 2702 \end{matrix} & \begin{pmatrix} -1 \\ 2 \end{pmatrix} \begin{pmatrix} -362 & 97 \end{pmatrix} & \begin{pmatrix} 1 \\ -2 \\ 7 \\ -26 \end{pmatrix} \begin{pmatrix} 26 & -7 & 2 & -1 \end{pmatrix} \\ \begin{pmatrix} -362 \\ 97 \end{pmatrix} \begin{pmatrix} -1 & 2 \end{pmatrix} & \begin{matrix} 2534 & -679 \\ -679 & 2522 \end{matrix} & \begin{matrix} 2522 & -679 \\ -679 & 2534 \end{matrix} \begin{pmatrix} 97 \\ -362 \end{pmatrix} \begin{pmatrix} 2 & -1 \end{pmatrix} \\ \begin{pmatrix} 26 \\ -7 \\ 2 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 7 & -26 \end{pmatrix} & \begin{pmatrix} 2 \\ -1 \end{pmatrix} \begin{pmatrix} 97 & -362 \end{pmatrix} & \begin{matrix} 2702 & -1351 \\ -1351 & 5042 \end{matrix} \end{pmatrix}.
$$

Damit kann die inverse Massematrix wie behauptet **exakt** als hierarchische Matrix dargestellt werden. Dies ist unabhängig vom Diskretisierungsparameter n und dies bleibt auch richtig für nicht gleichmäßige Partitionierungen des Intervalls $\Omega = (0, 1)$. Ausmultiplikation ergibt die explizite Darstellung

$$
M_h^{-1} = \frac{16}{2911} \begin{pmatrix} 5042 & -1351 & 362 & -97 & 26 & -7 & 2 & -1 \\ -1351 & 2702 & -724 & 194 & -52 & 14 & -4 & 2 \\ 362 & -724 & 2534 & -679 & 182 & -49 & 14 & -7 \\ -97 & 194 & -679 & 2522 & -676 & 182 & -52 & 26 \\ 26 & -52 & 182 & -676 & 2522 & -679 & 194 & -97 \\ -7 & 14 & -49 & 182 & -679 & 2534 & -724 & 362 \\ 2 & -4 & 14 & -52 & 194 & -724 & 2702 & -1351 \\ -1 & 2 & -7 & 26 & -97 & 362 & -1351 & 5042 \end{pmatrix}.
$$

Mit dieser Darstellung können zwei allgemein gültige Aussagen für die inverse Massematrix M_h^{-1} verifiziert werden:

1. exponentielles Abklingen der Matrixeinträge,
2. wechselndes Vorzeichen der Matrixeinträge.

Im folgenden soll ein alternativer Zugang zur Beschreibung der Massematrix und ihrer Inversen als hierarchische Matrix betrachtet werden. Die Gestalt der Massematrix M_h entspricht der Wahl der stückweise linearen Basisfunktionen gemäß der zugeordneten Knoten $x_k = kh$ für $k = 0, \ldots, n+1$, siehe Abbildung 8.2.

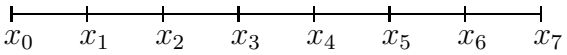


Abbildung 8.1: Standard–Numerierung der Knoten bzw. Freiheitsgrade.

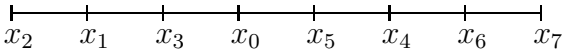


Abbildung 8.2: Permutierte Numerierung der Knoten bzw. Freiheitsgrade.

Wird die in Abbildung 8.2 angegebene Strategie zur Umnumerierung der Knoten verfolgt, so ergibt sich für die zugehörige Massematrix

$$M_h = \frac{1}{48} \left(\begin{array}{cccc|cccc} 4 & & 1 & & 1 & & & \\ & 4 & 1 & 1 & & & & \\ & 1 & 2 & & & & & \\ 1 & 1 & & 4 & & & & \\ \hline & & & & 4 & 1 & 1 & \\ 1 & & & & 1 & 4 & & \\ & & & & 1 & & 4 & 1 \\ & & & & & & 1 & 2 \end{array} \right) = \frac{1}{48} \left(\begin{array}{cc|cc} 4 & & 1 & 1 \\ & M_{11}^1 & & \\ 1 & & & \\ \hline & & & \\ 1 & & & M_{22}^1 \\ & & & \end{array} \right)$$

mit

$$M_{11}^1 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 2 & \\ 1 & & 4 \end{pmatrix}, \quad M_{22}^1 = \begin{pmatrix} 4 & 1 & 1 & \\ 1 & 4 & & \\ 1 & & 4 & 1 \\ & & 1 & 2 \end{pmatrix}.$$

Analog zu Abschnitt 5.4 kann die Inverse der Block–Matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

durch

$$A^{-1} = \begin{pmatrix} S^{-1} & -S^{-1} A_{12} A_{22}^{-1} \\ -A_{22}^{-1} A_{21} S^{-1} & A_{22}^{-1} + A_{22}^{-1} A_{21} S^{-1} A_{12} A_{22}^{-1} \end{pmatrix}$$

mit dem Schur–Komplement

$$S = A_{11} - A_{12} A_{22}^{-1} A_{21}$$

angegeben werden. Insbesondere für Matrizen der Gestalt

$$A = \begin{pmatrix} \alpha & \underline{a}_1^\top & \underline{a}_2^\top \\ \underline{a}_1 & A_1 & \\ \underline{a}_2 & & A_2 \end{pmatrix}$$

folgt für das skalare Schur–Komplement

$$S = \alpha - \underline{a}_1^\top A_1^{-1} \underline{a}_1 - \underline{a}_2^\top A_2^{-1} \underline{a}_2$$

und somit

$$A^{-1} = \frac{1}{S} \begin{pmatrix} 1 & -\underline{a}_1^\top A_1^{-1} & -\underline{a}_2^\top A_2^{-1} \\ -A_1^{-1}\underline{a}_1 & S \begin{pmatrix} A_1 & \\ & A_2 \end{pmatrix} + \begin{pmatrix} A_1^{-1}\underline{a}_1 \\ A_2^{-1}\underline{a}_2 \end{pmatrix} \begin{pmatrix} \underline{a}_1^\top A_1^{-1} & \underline{a}_2^\top A_2^{-1} \end{pmatrix} \\ -A_2^{-1}\underline{a}_2 & \end{pmatrix}.$$

Durch rekursive Anwendung kann somit die Inverse der Massematrix M_h berechnet werden. Auf eine explizite Ausführung soll an dieser Stelle jedoch verzichtet werden. Diese Strategie wurde zum Beispiel in [16, 17] bei der Invertierung der Massematrix zur Lösung zweidimensionaler Eigenwertprobleme verfolgt.

8.3 Randwertprobleme zweiter Ordnung

Neben der Approximation von gegebenen Funktionen ist die näherungsweise Lösung gewöhnlicher bzw. partieller Differentialgleichungen ein wesentliches Einsatzgebiet von finiten Elementen. Als einfachstes Modellproblem wird hier das Zweipunktrandwertproblem

$$-u''(x) = f(x) \quad \text{für } x \in \Omega = (0,1), \quad u(0) = u(1) = 0 \tag{8.2}$$

betrachtet. Sei v eine beliebige Testfunktion mit $v(0) = v(1) = 0$. Multiplikation der Differentialgleichung mit v und Integration über $\Omega = (0,1)$ ergibt

$$-\int_0^1 u''(x)v(x)dx = \int_0^1 f(x)v(x)dx.$$

Durch partielle Integration folgt dann die Variationsformulierung

$$\int_0^1 u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx.$$

Für eine konforme Galerkin–Diskretisierung dieser Variationsformulierung werden die in Abschnitt 8.1 konstruierten stückweise linearen Basisfunktionen verwendet. Unter Berücksichtigung der Randbedingungen lautet der Ansatz für die zu bestimmende Näherungsfunktion

$$u_h(x) = \sum_{k=1}^{n} u_k \varphi_k(x)$$

als Lösung der Galerkin–Variationsformulierung

$$\int_0^1 u_h'(x)\varphi_\ell'(x)dx \;=\; \int_0^1 f(x)\varphi_\ell(x)dx \quad \text{für } \ell = 1,\ldots,n.$$

Einsetzen des Ansatzes für u_h ergibt

$$\sum_{k=1}^{n} u_k \int_0^1 \varphi_k'(x)\varphi_\ell'(x)dx \;=\; \int_0^1 f(x)\varphi_\ell(x)dx \quad \text{für } \ell = 1,\ldots,n.$$

Mit den Einträgen für die Steifigkeitsmatrix A_h und den Lastvektor $\underline{f}$,

$$A_h[\ell,k] \;=\; \int_0^1 \varphi_k'(x)\varphi_\ell'(x)dx, \quad f_\ell \;=\; \int_0^1 f(x)\varphi_\ell(x)dx$$

für $k,\ell = 1,\ldots,n$ ist dies äquivalent zur Bestimmung von $\underline{u} \in I\!R^n$ als Lösung des linearen Gleichungssystems

$$A_h\underline{u} \;=\; \underline{f}\,.$$

Für die Diagonaleinträge $A_h[k,k]$ ergibt sich

$$\begin{aligned} A_h[k,k] \;&=\; \int_0^1 [\varphi_k'(x)]^2 dx \\ &=\; \int_{x_{k-1}}^{x_k} \left[\frac{d}{dx}\frac{x-x_{k-1}}{x_k-x_{k-1}}\right]^2 dx + \int_{x_k}^{x_{k+1}} \left[\frac{d}{dx}\frac{x_{k+1}-x}{x_{k+1}-x_k}\right]^2 dx \;=\; \frac{2}{h}. \end{aligned}$$

Für die Nebendiagonalelemente mit $\ell = k \pm 1$ ergibt sich

$$\begin{aligned} A_h[k+1,k] \;&=\; \int_0^1 \varphi_k'(x)\varphi_{k+1}'(x)dx \\ &=\; \int_{x_k}^{x_{k+1}} \left[\frac{d}{dx}\frac{x_{k+1}-x}{x_{k+1}-x_k}\right]\left[\frac{d}{dx}\frac{x-x_k}{x_{k+1}-x_k}\right] dx \;=\; -\frac{1}{h}\,. \end{aligned}$$

Damit ist die Steifigkeitsmatrix A_h gegeben durch

$$A_h \;=\; \frac{1}{h}\begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & 1 & & & \\ & -1 & \ddots & \ddots & & \\ & & \ddots & \ddots & -1 & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix} \in I\!R^{n\times n}.$$

Da die Struktur der Steifigkeitsmatrix A_h als Tridiagonalmatrix mit der Struktur der Massematrix M_h übereinstimmt, erlaubt auch die Inverse der Steifigkeitsmatrix A_h eine exakte Darstellung als hierarchische Matrix mit Rang 1 Matrizen in den Nebendiagonalblöcken, vergleiche hierzu auch Abschnitt 5.4.
Für $n = 8$ soll diese nun explizit berechnet werden, es ist

$$A_h \;=\; A_h^0 \;=\; 9 \left(\begin{array}{cc|cc|cc|cc} 2 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ \hline & -1 & 2 & -1 & & & & \\ & & -1 & 2 & -1 & & & \\ \hline & & & -1 & 2 & -1 & & \\ & & & & -1 & 2 & -1 & \\ \hline & & & & & -1 & 2 & -1 \\ & & & & & & -1 & 2 \end{array}\right) \;=\; 9 \begin{pmatrix} A_{11}^1 & A_{12}^1 \\ A_{21}^1 & A_{22}^1 \end{pmatrix}.$$

Für die Invertierung von A_h ist gemäß (8.1) zunächst die Inverse der Block–Matrix

$$A_{11}^1 \;=\; \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{pmatrix} \;=\; \begin{pmatrix} A_{11}^2 & A_{12}^2 \\ A_{21}^2 & A_{22}^2 \end{pmatrix}$$

mit

$$A_{11}^2 = A_{22}^2 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad A_{12}^2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad A_{21}^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \end{pmatrix}$$

zu bestimmen. Für die Inverse von A_{11}^2 ergibt sich

$$A_{11}^{2,-1} \;=\; \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

und für das Schur–Komplement S_{22}^2 der Matrix A_{11}^1 folgt

$$\begin{aligned} S_{22}^2 \;&=\; A_{22}^2 - A_{21}^2 A_{11}^{2,-1} A_{12}^2 \\ &=\; \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \\ &=\; \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \\ &=\; \frac{1}{3} \begin{pmatrix} 4 & -3 \\ -3 & 6 \end{pmatrix} \end{aligned}$$

mit der Inversen

$$S_{22}^{2,-1} \;=\; \frac{1}{5} \begin{pmatrix} 6 & 3 \\ 3 & 4 \end{pmatrix}.$$

Die Inverse von A_{11}^1 ist gemäß (8.1) gegeben durch

$$A_{11}^{1,-1} = \begin{pmatrix} A_{11}^{2,-1} + A_{11}^{2,-1} A_{12}^2 S_{22}^{2,-1} A_{21}^2 A_{11}^{2,-1} & -A_{11}^{2,-1} A_{12}^2 S_{22}^{2,-1} \\ -S_{22}^{2,-1} A_{21}^2 A_{11}^{2,-1} & S_{22}^{2,-1} \end{pmatrix}.$$

Dabei sind

$$A_{11}^{2,-1} A_{12}^2 S_{22}^{2,-1} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix}\frac{1}{5}\begin{pmatrix} 6 & 3 \\ 3 & 4 \end{pmatrix} = -\frac{1}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \end{pmatrix}$$

und

$$\begin{aligned} B &= A_{11}^{2,-1} + A_{11}^{2,-1} A_{12}^2 S_{22}^{2,-1} A_{21}^2 A_{11}^{2,-1} \\ &= \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & -1 \end{pmatrix}\frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} + \frac{2}{15}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} 4 & 3 \\ 3 & 6 \end{pmatrix}. \end{aligned}$$

Insgesamt ist also

$$A_{11}^{1,-1} = \frac{1}{5}\begin{pmatrix} \begin{pmatrix} 4 & 3 \\ 3 & 6 \end{pmatrix} & \begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \end{pmatrix} \\ \begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix} & \begin{pmatrix} 6 & 3 \\ 3 & 4 \end{pmatrix} \end{pmatrix}.$$

Die Anwendung von (8.1) zur Berechnung der Inversen von A_h verlangt jetzt die Auswertung des Schur–Komplements

$$\begin{aligned} S_{22}^1 &= A_{22}^1 - A_{21}^1 A_{11}^{1,-1} A_{12}^1 \\ &= A_{22}^1 - \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 & -1 \end{pmatrix} A_{11}^{1,-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{pmatrix} - \frac{4}{5}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \widetilde{A}_{33}^2 & A_{34}^2 \\ A_{43}^2 & A_{44}^2 \end{pmatrix} \end{aligned}$$

mit den Block–Matrizen

$$\begin{aligned} \widetilde{A}_{33}^2 &= \frac{1}{5}\begin{pmatrix} 6 & -5 \\ -5 & 10 \end{pmatrix}, \quad A_{44}^2 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \\ A_{34}^2 &= \begin{pmatrix} 0 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix}, \quad A_{43}^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & -1 \end{pmatrix}. \end{aligned}$$

Mit

$$\widetilde{A}_{33}^{2,-1} = \frac{1}{7}\begin{pmatrix} 10 & 5 \\ 5 & 6 \end{pmatrix}$$

folgt für das Schur–Komplement S_{44}^2 von S_{22}^1

$$\begin{aligned} S_{44}^2 &= A_{44}^2 - A_{43}^2 \widetilde{A}_{33}^{2,-1} A_{34}^2 \\ &= \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \end{pmatrix} \frac{1}{7} \begin{pmatrix} 10 & 5 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} - \frac{6}{7} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \\ &= \frac{1}{7}\begin{pmatrix} 8 & -7 \\ -7 & 14 \end{pmatrix}, \end{aligned}$$

und für die Inverse ergibt sich

$$S_{44}^{2,-1} = \frac{1}{9}\begin{pmatrix} 14 & 7 \\ 7 & 8 \end{pmatrix}.$$

Die Inverse von S_{22}^1 ist dann gemäß (8.1) gegeben durch

$$S_{22}^{1,-1} = \begin{pmatrix} \widetilde{A}_{33}^{2,-1} + \widetilde{A}_{33}^{2,-1} A_{34}^2 S_{44}^{2,-1} A_{43}^2 \widetilde{A}_{33}^{2,-1} & -\widetilde{A}_{33}^{2,-1} A_{34}^2 S_{44}^{2,-1} \\ -S_{44}^{2,-1} A_{43}^2 \widetilde{A}_{33}^{2,-1} & S_{44}^{2,-1} \end{pmatrix}.$$

Mit

$$\widetilde{A}_{33}^{2,-1} A_{34}^2 S_{44}^{2,-1} = \frac{1}{7}\begin{pmatrix} 10 & 5 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \frac{1}{9} \begin{pmatrix} 14 & 7 \\ 7 & 8 \end{pmatrix} = -\frac{1}{9} \begin{pmatrix} 5 \\ 6 \end{pmatrix} \begin{pmatrix} 2 & 1 \end{pmatrix}$$

und

$$\begin{aligned} B &= \widetilde{A}_{33}^{2,-1} + \widetilde{A}_{33}^{2,-1} A_{34}^2 S_{44}^{2,-1} A_{43}^2 \widetilde{A}_{33}^{2,-1} \\ &= \frac{1}{7}\begin{pmatrix} 10 & 5 \\ 5 & 6 \end{pmatrix} - \frac{1}{9} \begin{pmatrix} 5 \\ 6 \end{pmatrix} \begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \end{pmatrix} \frac{1}{7} \begin{pmatrix} 10 & 5 \\ 5 & 6 \end{pmatrix} \\ &= \frac{1}{7}\begin{pmatrix} 10 & 5 \\ 5 & 6 \end{pmatrix} + \frac{2}{9}\frac{1}{7} \begin{pmatrix} 5 \\ 6 \end{pmatrix} \begin{pmatrix} 5 & 6 \end{pmatrix} \\ &= \frac{1}{9}\begin{pmatrix} 20 & 15 \\ 15 & 18 \end{pmatrix} \end{aligned}$$

ergibt sich insgesamt Inverse von S_{22}^1

$$S_{22}^{1,-1} = \frac{1}{9}\begin{pmatrix} \begin{pmatrix} 20 & 15 \\ 15 & 18 \end{pmatrix} & \begin{pmatrix} 5 \\ 6 \end{pmatrix} \begin{pmatrix} 2 & 1 \end{pmatrix} \\ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 5 & 6 \end{pmatrix} & \begin{pmatrix} 14 & 7 \\ 7 & 8 \end{pmatrix} \end{pmatrix}.$$

Mit (8.1) ergibt sich jetzt die inverse Steifigkeitsmatrix aus

$$A_h^{-1} = \frac{1}{9}\begin{pmatrix} A_{11}^{1,-1} + A_{11}^{1,-1}A_{12}^1 S_{22}^{1,-1} A_{21}^1 A_{11}^{1,-1} & -A_{11}^{1,-1}A_{12}^1 S_{22}^{1,-1} \\ -S_{22}^{1,-1}A_{21}^1 A_{11}^{1,-1} & S_{22}^{1,-1} \end{pmatrix}.$$

Mit

$$A_{11}^{1,-1}A_{12}^1 S_{22}^{1,-1} = A_{11}^{1,-1}\begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} S_{22}^{1,-1} = -\frac{1}{9}\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}\begin{pmatrix} 4 & 3 & 2 & 1 \end{pmatrix}$$

und

$$\begin{aligned} B &= A_{11}^{1,-1} + A_{11}^{1,-1}A_{12}^1 S_{22}^{1,-1} A_{21}^1 A_{11}^{1,-1} \\ &= A_{11}^{1,-1} - \frac{1}{9}\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}\begin{pmatrix} 4 & 3 & 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 & -1 \end{pmatrix} A_{11}^{1,-1} \\ &= A_{11}^{1,-1} + \frac{4}{45}\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} \begin{pmatrix} 4 & 3 \\ 3 & 6 \end{pmatrix} & \begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \end{pmatrix} \\ \begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix} & \begin{pmatrix} 6 & 3 \\ 3 & 4 \end{pmatrix} \end{pmatrix} + \frac{4}{45}\begin{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} & \begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 3 & 4 \end{pmatrix} \\ \begin{pmatrix} 3 \\ 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix} & \begin{pmatrix} 9 & 12 \\ 12 & 16 \end{pmatrix} \end{pmatrix} \\ &= \frac{1}{9}\begin{pmatrix} \begin{pmatrix} 8 & 7 \\ 7 & 14 \end{pmatrix} & \begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 6 & 5 \end{pmatrix} \\ \begin{pmatrix} 6 \\ 5 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix} & \begin{pmatrix} 18 & 15 \\ 15 & 20 \end{pmatrix} \end{pmatrix} \end{aligned}$$

ist insgesamt

$$A_h^{-1} = \frac{1}{9}\begin{pmatrix} \begin{pmatrix} 8 & 7 \\ 7 & 14 \end{pmatrix} & \begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 6 & 5 \end{pmatrix} & \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}\begin{pmatrix} 4 & 3 & 2 & 1 \end{pmatrix} \\ \begin{pmatrix} 6 \\ 5 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix} & \begin{pmatrix} 18 & 15 \\ 15 & 20 \end{pmatrix} & \\ \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix} & & \begin{pmatrix} 20 & 15 \\ 15 & 18 \end{pmatrix} \quad \begin{pmatrix} 5 \\ 6 \end{pmatrix}\begin{pmatrix} 2 & 1 \end{pmatrix} \\ & & \begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 5 & 6 \end{pmatrix} \quad \begin{pmatrix} 14 & 7 \\ 7 & 8 \end{pmatrix} \end{pmatrix}.$$

Wie behauptet kann die inverse Steifigkeitsmatrix **exakt** als hierarchische Matrix dargestellt werden. Wie bei der Darstellung der inversen Massematrix ist dies unabhängig vom Diskretisierungsparameter n und der gewählten Partitionierung des Intervalles $\Omega = (0,1)$. Durch Ausmultiplizieren ergibt sich die explizite Darstellung

$$A_h^{-1} = \frac{1}{9}\begin{pmatrix} 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 7 & 14 & 12 & 10 & 8 & 6 & 4 & 2 \\ 6 & 12 & 18 & 15 & 12 & 9 & 6 & 3 \\ 5 & 10 & 15 & 20 & 16 & 12 & 8 & 4 \\ 4 & 8 & 12 & 16 & 20 & 15 & 10 & 5 \\ 3 & 6 & 9 & 12 & 15 & 18 & 12 & 6 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{pmatrix}.$$

Daraus können wieder zwei allgemein gültige Eigenschaften der inversen Steifigkeitsmatrix abgelesen werden:

1. Abklingverhalten der Matrix–Einträge,

2. Positivität der Matrix–Einträge.

Alternativ kann zur hierarchischen Darstellung der inversen Steifigkeitsmatrix A_h auch die in Abbildung (8.2) angegebene Permutation der Freiheitsgrade verwendet werden. Die permutierte Steifigkeitsmatrix kann dann wie im Fall der permutierten Massematrix behandelt werden.

Im folgenden soll die Existenz einer hierarchischen Darstellung der inversen Steifigkeitsmatrix aus einem anderen Blickwinkel heraus untersucht werden. Dies ermöglicht dann sofort den Übergang zu Randwertproblemen in mehreren Raumdimensionen und allgemeineren partiellen Differentialoperatoren. Dieser Zugang beruht auf der Darstellung der Lösung u des Randwertproblems (8.2) mittels der Greenschen Funktion, deren Diskretisierung ihrerseits die Darstellung als hierarchische Matrix erlaubt.
Ausgangspunkt hierfür sind die Formeln der partiellen Integration,

$$\begin{aligned} \int_a^b u'(x)v'(x)dx &= [u'(x)v(x)]_a^b + \int_a^b [-u''(x)]v(x)dx, \\ \int_a^b v'(x)u'(x)dx &= [v'(x)u(x)]_a^b + \int_a^b [-v''(x)]u(x)dx, \end{aligned}$$

die durch Gleichsetzen die zweite Greensche Formel

$$[u'(x)v(x)]_a^b + \int_a^b [-u''(x)]v(x)dx = [v'(x)u(x)]_a^b + \int_a^b [-v''(x)]u(x)dx$$

ergeben. Sei u Lösung des Randwertproblems (8.2). Für $x \in (0,1)$ und ein beliebiges $\varepsilon > 0$ lautet die zweite Greensche Formel bezüglich dem Intervall $(0, x-\varepsilon)$

$$[u'(y)v(y)]_0^{x-\varepsilon} + \int_0^{x-\varepsilon} f(y)v(y)dy \;=\; [v'(y)u(y)]_0^{x-\varepsilon} + \int_0^{x-\varepsilon} [-v''(y)]u(y)dy.$$

Für

$$v_1(y) \;=\; \alpha_1(x) + \beta_1(x)y, \quad v_1(0) = 0$$

folgt $\alpha_1(x) = 0$ sowie $v_1''(y) = 0$ und somit

$$u'(x-\varepsilon)\beta_1(x)(x-\varepsilon) + \int_0^{x-\varepsilon} f(y)v_1(y)dy \;=\; \beta_1(x)u(x-\varepsilon).$$

Für

$$v_2(y) \;=\; \alpha_2(x) + \beta_2(x)y, \quad v_2(1) = 0$$

ergibt sich entsprechend aus der zweiten Greenschen Formel bezüglich $(x+\varepsilon, 1)$

$$-u'(x+\varepsilon)v_2(x+\varepsilon) + \int_{x+\varepsilon}^1 f(y)v_2(y)dy \;=\; -\beta_2(x)u(x+\varepsilon).$$

Durch Addition der beiden Gleichungen und den Grenzübergang $\varepsilon \to 0$ folgt die Darstellungsformel

$$[\beta_1(x) - \beta_2(x)]u(x) \;=\; \int_0^x f(y)v_1(y)dy + \int_x^1 f(y)v_2(y)dy + u'(x)[\beta_1(x)x - \alpha_2(x) - \beta_2(x)x].$$

Zu lösen verbleibt

$$\beta_1(x) - \beta_2(x) = 1, \quad [\beta_1(x) - \beta_2(x)]x - \alpha_2(x) = 0, \quad \alpha_2(x) + \beta_2(x) = 0.$$

Daraus folgt

$$\alpha_2(x) = x, \quad \beta_2(x) = -x, \quad \beta_1(x) = 1 - x,$$

und somit gilt für die Lösung u des Randwertproblems (8.2) die Darstellungsformel

$$u(x) \;=\; \int_0^1 G(x,y)f(y)dy \quad \text{für } x \in (0,1) \tag{8.3}$$

mit der **Greenschen Funktion**

$$G(x,y) \;=\; \begin{cases} v_1(y) = (1-x)y & 0 < y < x, \\ v_2(y) = x(1-y) & x < y < 1. \end{cases} \tag{8.4}$$

Die durch (8.3) eindeutig bestimmte Funktion

$$u(x) = (Nf)(x) = \int_0^1 G(x,y)f(y)dy \tag{8.5}$$

ist gleichzeitig die eindeutige Lösung der Variationsformulierung

$$\int_0^1 u'(x)v'(x)dx \;=\; \int_0^1 f(x)v(x)dx$$

für alle geeignet gewählten Testfunktionen v mit $v(0) = v(1) = 0$. Diese Identifikation gilt jedoch **nur** für die kontinuierlich gegebene Lösung (8.5) des Randwertproblems (8.2), und **nicht** für zugehörige Näherungslösungen. Sei

$$u_h(x) \;=\; \sum_{k=1}^{n} u_k\varphi_k(x) \quad \text{für } x \in (0,1)$$

die eindeutig bestimmte Näherungslösung der Galerkin–Variationsformulierung

$$\int_0^1 u_h'(x)\varphi_j'(x)dx \;=\; \int_0^1 f(x)\varphi_j(x)dx \quad \text{für } j = 1,\ldots,n.$$

Die obigen Betrachtungen gelten für beliebige quadratintegrierbare Funktionen f mit

$$\int_0^1 [f(x)]^2 dx \;<\; \infty.$$

Sei $\mathrm{span}\{\psi_k\}_{k=1}^n$ ein Ansatzraum quadratintegrierbarer Basisfunktionen ψ_k. Für

$$f(x) \;=\; \sum_{k=1}^{n} f_k\psi_k(x) \quad \text{für } x \in (0,1)$$

ergeben sich die Zerlegungskoeffizienten $\underline{u} \in \mathbb{R}^n$ der Näherungslösung u_h als Lösung des linearen Gleichungssystems

$$A_h\underline{u} \;=\; \bar{M}_h^\top \underline{f}.$$

Neben der Steifigkeitsmatrix A_h bezeichnet $\bar{M}_h$ die durch

$$\bar{M}_h[j,k] \;=\; \int_0^1 \varphi_k(x)\psi_j(x)dx$$

für $k, j = 1,\ldots$ erklärte Massematrix. Es wird voraussgesetzt, daß die Näherungslösungen $u_h \to u$ für $h \to 0$ bzw. $n \to \infty$ konvergieren.

Ausgehend von der Darstellungsformel (8.5) kann durch die Lösung des Variationsproblems

$$\int_0^1 \widetilde{u}_h(x)\psi_j(x)dx \;=\; \int_0^1 \int_0^1 G(x,y)f(y)dy\psi_j(x)dx$$

für $j = 1, \ldots, n$ eine zweite Näherungslösung

$$\widetilde{u}_h(x) \;=\; \sum_{k=1}^{n} \widetilde{u}_k \varphi_k(x)$$

erklärt werden. Der Koeffizientenvektor $\underline{\widetilde{u}} \in I\!R^n$ ergibt sich als Lösung des linearen Gleichungssystems

$$\bar{M}_h \underline{\widetilde{u}} \;=\; G_h \underline{f}$$

mit der durch

$$G_h[k,j] \;=\; \int_0^1 \int_0^1 G(x,y)\psi_k(y)dy\,\psi_j(x)dx$$

für $k, j = 1, \ldots, n$ erklärten Matrix G_h.
Die durch

$$\underline{u} \;=\; A_h^{-1} \bar{M}_h^{\top} \underline{f}$$

bzw. die durch

$$\underline{\widetilde{u}} \;=\; \bar{M}_h^{-1} G_h \underline{f}$$

definierten Funktionen u_h und $\widetilde{u}_h$ sind in der Regel verschiedene Näherungsfunktionen der Lösung u des Randwertproblems (8.2). Aus der Voraussetzung der Konvergenz $u_h \to u$ bzw. $\widetilde{u}_h \to u$ folgt

$$u_h \;=\; \widetilde{u}_h + o(h),$$

d.h. bis auf einen Fehler niedrigerer Ordnung können beide Näherungslösungen miteinander identifiziert werden. Daraus folgt

$$\underline{u} \;=\; \underline{\widetilde{u}} + o(h)$$

bzw.

$$A_h^{-1} \bar{M}_h^{\top} \underline{f} \;=\; \bar{M}_h^{-1} G_h \underline{f} + o(h)$$

für beliebiges $\underline{f} \in I\!R^n$. Mit $\underline{f} = \bar{M}_h^{-\top} \underline{\bar{f}}$ folgt daraus

$$A_h^{-1} \;=\; \bar{M}_h^{-1} G_h \bar{M}_h^{-\top} + o(h).$$

Bis auf einen Fehlerterm stimmt also die inverse Steifigkeitsmatrix A_h^{-1} mit dem Matrixprodukt $\bar{M}_h^{-1} G_h \bar{M}_h^{-\top}$ überein.
Da die in (8.4) angegebene Greensche Funktion $G(x,y)$ bereits durch eine multiplikative Aufspaltung gegeben ist, folgt sofort die Darstellbarkeit von G_h als hierarchische Matrix.

Zu untersuchen bleibt die Struktur der inversen Massematrix $\bar{M}_h^{-1}$. Bei Verwendung von **biorthogonalen Basisfunktionen** mit

$$\int_0^1 \varphi_k(x)\psi_j(x)dx \;=\; \delta_{kj}$$

für $k, j = 1, \ldots, n$ folgt $M_h = I_n$ und somit

$$A_h^{-1} \;=\; G_h + o(h),$$

d.h. die Approximierbarkeit der inversen Steifigkeitsmatrix A_h^{-1} durch eine hierarchische Matrix G_h.

Bemerkung 8.1 *Die obige Herleitung kann auf allgemeinere partielle Differentialoperatoren zweiter Ordnung in mehreren Raumdimensionen, auch mit springenden Koeffizienten, übertragen werden. Dies wurde erstmals in* [3] *angegeben. Entscheidend für die Analysis ist neben der Invertierbarkeit der inversen Massematrix $\bar{M}_h$ die Fehlerabschätzung*

$$\|u_h - \widetilde{u}_h\| \;\leq\; \|u_h - u\| + \|u - \widetilde{u}_h\|,$$

die sich aus dem Approximationsfehler $\|u - u_h\|$ der FEM Variationsformulierung und dem Approximationsfehler $\|u - \widetilde{u}_h\|$ des Newtonpotentials $u = Nf$ zusammensetzt. Für einen gegebenen Ansatzraum $\{\varphi_k\}_{k=1}^n$ von Basisfunktionen φ_k ist deshalb ein Ansatzraum $\{\psi_k\}_{k=1}^n$ von biorthogonalen Basisfunktionen ψ_k zu konstruieren, der neben einer Approximationseigenschaft auch eine geeignete Stabilitätsbedingung und somit die Invertierbarkeit der Massematrix $\bar{M}_h$ gewährleistet. Hier zeigt sich insbesondere der Zusammenhang mit gemischten Diskretisierungsverfahren und hybriden Gebietszerlegungsmethoden [23]*, insbesondere mit Mortar–Methoden* [28].
Auf eine detailierte Stabilitäts– und Fehleranalysis zur Approximation der inversen FEM Steifigkeitsmatrix A_h durch hierarchische Matrizen soll an dieser Stelle jedoch verzichtet werde, siehe zum Beispiel [3].

Literaturverzeichnis

[1] Bebendorf, M.: Effiziente numerische Lösung von Randintegralgleichungen unter Verwendung von Niedrigrang–Matrizen. Dissertation, Universität des Saarlandes, Saarbrücken, 2000.

[2] Bebendorf, M.: Approximation of boundary element matrices. Numer. Math. 86 (2000) 565–589.

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

[4] Bebendorf, M., Rjasanow, S.: Adaptive low–rank approximation of collocation matrices. Computing 70 (2003) 1–24.

[5] Börm, S., Grasedyck, L., Hackbusch, W.: Hierarchical Matrices. Lecture Note 21, Max–Planck–Institut für Mathematik in den Naturwissenschaften, Leipzig, 2003.

[6] Braess, D.: Finite Elemente. Springer, Berlin, 1991.

[7] Davis, P. J.: Circulant Matrices. John Wiley & Sons, New York, 1979.

[8] Golub, G. H., van Loan, C. F.: Matrix Computations. The John Hopkins University Press, Baltimore, London, 1993.

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

[10] Greengard, L., Rokhlin, V.: A fast algorithm for particle simulations. J. Comput. Phys. 73 (1987) 325–348.

[11] Hackbusch, W.: Iterative Lösung grosser schwachbesetzter Gleichungssysteme. B. G. Teubner, Stuttgart, 1993.

[12] Hackbusch, W.: Theorie und Numerik elliptischer Differentialgleichungen. B. G. Teubner, Stuttgart, 1996.

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

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

[15] Jung, M., Langer, U.: Methode der finiten Elemente für Ingenieure. B. G. Teubner, Stuttgart, Leipzig, Wiesbaden, 2001.

[16] Lintner, M.: Lösung der 2D Wellengleichung mittels hierarchischer Matrizen. Dissertation, TU München, 2002.

[17] Lintner, M.: The eigenvalue problem for the 2D Laplacian in $\mathcal{H}$–matrix arithmetic and application to the heat and wave equation. Computing 72 (2004) 293–323.

[18] van Loan, C.: Computational Frameworks for the Fast Fourier Transform. SIAM, Philadelphia, 1992.

[19] Meister, A.: Numerik linearer Gleichungssysteme. Eine Einführung in moderne Verfahren. Vieweg, Braunschweig, 1999.

[20] Ortega, J. M., Rheinboldt, W. C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, 1970.

[21] Sauter, S. A., Schwab, C.: Randelemente. Analyse und Implementierung schneller Algorithmen. B. G. Teubner, Stuttgart, Leipzig, Wiesbaden, 2004.

[22] Schatz, A. H., Thomée, V., Wendland, W. L.: Mathematical Theory of Finite and Boundary Element Methods. Birkhäuser, Basel, 1990.

[23] Steinbach, O.: Stability estimates for hybrid coupled domain decomposition methods. Lecture Notes in Mathematics 1809, Springer, Heidelberg, 2003.

[24] Steinbach, O.: Numerische Näherungsverfahren für elliptische Randwertprobleme. Finite Elemente und Randelemente. B. G. Teubner, Stuttgart, Leipzig, Wiesbaden, 2003.

[25] Steinbach, O. (ed.): Hauptseminar Hierarchische Matrizen. Berichte aus dem Institut für Angewandte Analysis und Numerische Simulation, Universität Stuttgart, Seminarbericht 2004/013, 2004.

[26] Steinbach, O: Lösungsverfahren für lineare Gleichungssysteme. Algorithmen und Anwendungen. B. G. Teubner, Stuttgart, Leipzig, Wiesbaden, in Vorbereitung.

[27] Tyrtyshnikov, E. E.: Mosaic skeleton approximations. Calcolo 33 (1996) 47–57.

[28] Wohlmuth, B. I.: Discretization Methods and Iterative Solvers Based on Domain Decomposition. Lecture Notes in Computational Science and Engineering 17, Springer–Verlag, Berlin, 2001.

Erschienene Preprints ab Nummer 2004/001

Komplette Liste: `http://preprints.ians.uni-stuttgart.de`

2004/001 *Geis, W., Mishuris, G., Sändig, A.-M.*: 3D and 2D asymptotic models for piezoelectric stack actuators with thin metal inclusions

2004/002 *Klimke, A., Wohlmuth, B., Willner, K.*: Computing expensive multivariate functions of fuzzy numbers using sparse grids

2004/003 *Klimke, A., Wohlmuth, B., Willner, K.*: Uncertainty modeling using efficient fuzzy arithmetic based on sparse grids: applications to dynamic systems

2004/004 *Flemisch, B., Mair, M., Wohlmuth, B.*: Nonconforming discretization techniques for overlapping domain decompositions

2004/005 *Sändig, A.-M.*: Vorlesung Mathematik für Informatiker und Softwaretechniker I, WS 2003/2004

2004/006 *Bürger, R., Karlsen, K. H., Towers, J. D.*: Closed-form and finite difference solutions to a population balance model of grinding mills

2004/007 *Berres, S., Bürger, R., Tory, E. M.*: Applications of Polydisperse Sedimentation Models

2004/008 *Bürger, R., Karlsen, K. H., Towers, J. D.*: A model of continuous sedimentation of flocculated suspensions in clarifier-thickener units

2004/009 *Bürger, R., Karlsen, K. H., Towers, J. D.*: Mathematical model and numerical simulation of the dynamics of flocculated suspensions in clarifier-thickeners

2004/010 *Lehrstühle: Wendland, Wohlmuth, Abteilungen: Gekeler. Sändig*: Jahresbericht 2003

2004/011 *Sändig, A.-M. (Hrsg.), Knees, D. (Hrsg.)*: Nichtlineare Funktionalanalysis mit Anwendungen in der Festkörpermechanik

2004/012 *Wendland, W.L.*: Vorlesungsskript Partielle Differentialgleichungen

2004/013 *Steinbach, O. (ed.)*: Seminarbericht: Hierarchische Matrizen

2004/014 *Sändig, A.-M.*: Vorlesung Mathematik für Informatiker und Softwaretechniker II, SS 2004

2004/015 *Langer, U., Steinbach, O., Wendland, W. L. (eds)*: Workshop on Adaptive Fast Boundary Element Methods in Industrial Applications, Söllerhaus, 29.9.-2.10.2004.

2004/016 *Steinbach, O.*: Vorlesung Hierarchische Matrizen