

Einführendes Beispiel

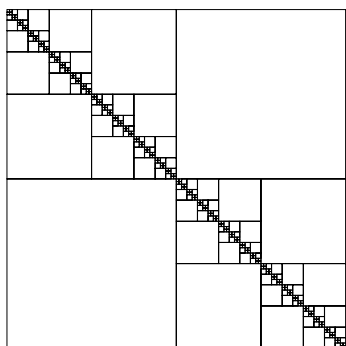


Abb. 3.1. Blockzerlegung von \mathcal{H}_7 ($n = 128$)

Dass es blockweise möglich ist, eine Matrix zu vereinfachen, ist nicht neu. Paneel-Clusterungsmethoden (vgl. [90], [125, §7]), Multipolverfahren (vgl. [122], [125, §7.1.3.2]), Mosaikapproximation (vgl. [131]) und Matrixkompressionstechniken bei Wavelets (vgl. [35]) beruhen auf dem gleichen Konzept. Allerdings ist es in keinem dieser Fälle möglich, andere Matrixoperationen als die Matrixvektormultiplikation effizient durchzuführen. Deshalb soll in diesem Kapitel herausgestellt werden, wie die Matrixoperationen ausgeführt werden und was ihr Aufwand ist. Insbesondere wird

sich herausstellen, dass alle Matrixoperationen mit fast linearem Aufwand berechnet werden können (statt $\mathcal{O}(n^2)$ oder $\mathcal{O}(n^3)$ wie bei voller Matrixdarstellung), wobei aber zu berücksichtigen ist, dass die Resultate im Allgemeinen Approximationsfehler enthalten.

In diesem Modellbeispiel ist das Matrixformat fest vorgegeben. In den praktischen Anwendungen ist das Format dagegen speziell für die jeweilige Aufgabe zu konstruieren (dies wird in §5 beschrieben werden).

3.1 Das Modellformat \mathcal{H}_p

Wir beschränken uns auf die Indexmengen $I = \{1, \dots, n\}$ mit Zweierpotenzen

$$n = 2^p \quad (3.1)$$

und definieren induktiv in p das Matrixformat \mathcal{H}_p . Um I in Abhängigkeit von p zu charakterisieren, schreiben wir auch I_p für I .

Für $p = 0$ ist $M \in \mathbb{R}^{I \times I}$ eine 1×1 -Matrix, die formal als volle Matrix dargestellt sei. Entsprechend sei \mathcal{H}_0 die Menge aller 1×1 -Matrizen in der Darstellung `Volle_Matrix(1, 1)`. Die weiteren Darstellungen \mathcal{H}_p werden rekursiv definiert.

Wir nehmen an, dass das Darstellungsformat von \mathcal{H}_{p-1} für Matrizen aus $\mathbb{R}^{I_{p-1} \times I_{p-1}}$ bekannt sei. Eine Matrix aus $\mathbb{R}^{I_p \times I_p}$ kann als Blockmatrix

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \quad M_{ij} \in \mathbb{R}^{I_{p-1} \times I_{p-1}}, \quad (3.2a)$$

dargestellt werden. Wir schränken die Menge aller $M \in \mathbb{R}^{I \times I}$ durch die folgende Forderung ein:

$$M_{11}, M_{22} \in \mathcal{H}_{p-1}, \quad M_{12}, M_{21} \in \mathcal{R}_{p-1}(k), \quad (3.2b)$$

wobei $\mathcal{R}_{p-1}(k) := \mathcal{R}_{p-1}(k, I_{p-1}, I_{p-1})$ die Rang- k -Matrixfamilie aus Definition 2.2.3 ist. Die Menge der Matrizen (3.2a) mit (3.2b) bildet die Menge \mathcal{H}_p . Der lokale Rang k ist eigentlich so zu wählen, dass eine bestimmte Approximationsgüte erreicht wird. Da in diesem Abschnitt die Approximation keine Rolle spielen soll, treffen wir in (3.2b) die einfache Wahl


$$k = 1 \quad (3.2c)$$

und schreiben kurz \mathcal{R}_{p-1} für $\mathcal{R}_{p-1}(1)$. Die rekursive Struktur von \mathcal{H}_p kann durch

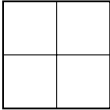
$$\mathcal{H}_p = \begin{bmatrix} \mathcal{H}_{p-1} & \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} & \mathcal{H}_{p-1} \end{bmatrix} \quad (3.2d)$$

charakterisiert werden. Löst man die Rekursion auf, erhält man die folgenden Blockzerlegungen, die in

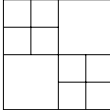
$p = 0$



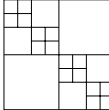
$p = 1$



$p = 2$



$p = 3$



(3.3)

für $p = 0, 1, 2, 3$ wiedergegeben sind. Abbildung 3.1 zeigt den Fall $n = 2^7 = 128$. Nach Definition von $\mathcal{R}_p(k)$ kann jeder Block b einen beliebigen Matrixblock $M|_b$ mit $\text{Rang}(M|_b) \leq k$ enthalten, wobei hier $k = 1$ gemäß (3.2c).

3.2 Zahl der Blöcke

Als Erstes sei die Zahl der Blöcke in Matrizen aus \mathcal{M}_p per Induktion bestimmt. Für $p = 0$ in (3.1) liegt eine 1×1 -Matrizen vor, d.h. $N_{\text{block}}(0) = 1$. Die Rekursion (3.2d) zeigt $N_{\text{block}}(p) = 2 + 2N_{\text{block}}(p-1)$ für $p > 1$. Diese Rekursionsgleichung wird erfüllt durch

$$N_{\text{block}}(p) = 3n - 2. \quad (3.4)$$

3.3 Speicheraufwand

Der Speicheraufwand einer \mathcal{R}_p -Matrix ($n = 2^p$) ist $S_R(p) = 2^{p+1}$ (vgl. Anmerkung 2.2.5). Sei S_p der Speicheraufwand einer Matrix aus \mathcal{H}_p . Für $p = 0$ ist nur eine 1×1 -Matrix zu speichern, d.h. $S_0 = 1$. Die Rekursion (3.2d) zeigt

$$S_p = 2S_{R1}(p-1) + 2S_{p-1} = 2^{p+1} + 2S_{p-1}.$$

Zusammen mit $S_0 = 1$ folgt $S_p = (2p+1)n$. Dies beweist das

Lemma 3.3.1. *Der Speicherbedarf einer Matrix aus \mathcal{H}_p ($n = 2^p$) beträgt*

$$S_p = n + 2n \log_2 n. \quad (3.5)$$

3.4 Matrix-Vektor-Multiplikation

Seien $M \in \mathcal{H}_p$ und $x \in \mathbb{R}^{I_p}$ mit $n = 2^p$. Die noch zu bestimmenden Kosten für $M \cdot x$ seien mit $N_{MV}(p)$ bezeichnet. Für $p \geq 1$ sei M wie in (3.2a) zerlegt: $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$. Entsprechend wird x in $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ mit $x_1, x_2 \in \mathbb{R}^{I_{p-1}}$ zerlegt. Die Multiplikation Mx reduziert sich auf die Berechnung von

$$y_{11} := M_{11}x_1, \quad y_{12} := M_{12}x_2, \quad y_{21} := M_{21}x_1, \quad y_{22} := M_{22}x_2$$

und die Additionen $y_{11} + y_{12}$ und $y_{21} + y_{22}$. Nach Anmerkung 2.3.1a kosten $M_{12}x_2$ und $M_{21}x_1$ je $3\frac{n}{2} - 1$ Operationen, während die Additionen je $\frac{n}{2}$ Operationen benötigen. Dies führt auf die Rekursion

$$N_{MV}(p) = 2N_{MV}(p-1) + 4n - 2$$

mit dem Startwert $N_{MV}(0) = 1$. Ihre Lösung lautet $N_{MV}(p) = 4np - n + 2$.

Lemma 3.4.1. *Sei $n = 2^p$. Die Matrix-Vektor-Multiplikation von $M \in \mathcal{H}_p$ mit $x \in \mathbb{R}^{I_p}$ benötigt den Aufwand*

$$N_{MV}(p) = 4n \log_2 n - n + 2. \quad (3.6)$$

Im Gegensatz zu den folgenden Operationen wird die Matrix-Vektor-Multiplikation *exakt* durchgeführt.

3.5 Matrix-Addition

Wir unterscheiden drei Typen von Additionen:

- 1) $A \oplus_1 B \in \mathcal{R}_p$ für $A, B \in \mathcal{R}_p$ mit $I = \{1, \dots, n\}$ mit den Kosten $N_{R+R}(p)$.
- 2) $A \oplus_1 B \in \mathcal{H}_p$ für $A, B \in \mathcal{H}_p$ mit den Kosten $N_{H+H}(p)$.
- 3) $A \oplus_1 B \in \mathcal{H}_p$ für $A \in \mathcal{H}_p$ und $B \in \mathcal{R}_p$ mit den Kosten $N_{H+R}(p)$.

Das Symbol \oplus_1 macht deutlich, dass nicht die exakte Addition vorliegt, sondern blockweise auf Rang-1-Matrizen gekürzt wird.

Gemäß Korollar 2.6.1b ist $N_{R+R}(p) = 18n + 29$ (später in der Form $N_{R+R}(p-1) = 9n + 29$ angewandt).

Im Falle $A, B \in \mathcal{H}_p$ verwenden wir die Blockstruktur (3.2d). Die Summe hat die Gestalt

$$\begin{bmatrix} \mathcal{H}_{p-1} & \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} & \mathcal{H}_{p-1} \end{bmatrix} + \begin{bmatrix} \mathcal{H}_{p-1} & \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} & \mathcal{H}_{p-1} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_{p-1} + \mathcal{H}_{p-1} & \mathcal{R}_{p-1} + \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} + \mathcal{R}_{p-1} & \mathcal{H}_{p-1} + \mathcal{H}_{p-1} \end{bmatrix}.$$

Die exakte Definition der Operation $\oplus_1 : \mathcal{H}_p \times \mathcal{H}_p \rightarrow \mathcal{H}_p$ lautet: Für $p = 0$ ist $\oplus_1 = +$ die exakte Addition. Ansonsten gilt die Rekursion

$$M' \oplus_1 M'' := \begin{bmatrix} M'_{11} \oplus_1 M''_{11} & M'_{12} \oplus_1 M''_{12} \\ M'_{21} \oplus_1 M''_{21} & M'_{22} \oplus_1 M''_{22} \end{bmatrix}. \quad (3.7)$$

Dabei ist in den Außerdiagonalblöcken $\oplus_1 : \mathcal{R}_{p-1} \times \mathcal{R}_{p-1} \rightarrow \mathcal{R}_{p-1}$ die schon definierte formatierte Addition von \mathcal{R}_{p-1} -Matrizen (vgl. (2.13)), während in den Diagonalblöcken die Operation $\oplus_1 : \mathcal{H}_{p-1} \times \mathcal{H}_{p-1} \rightarrow \mathcal{H}_{p-1}$ der Stufe $p-1$ vorliegt.

Gemäß (3.7) erhält man für den Aufwand die Rekursion

$$N_{H+H}(p) = 2N_{H+H}(p-1) + 2N_{R+R}(p-1) = 2N_{H+H}(p-1) + 18n + 58.$$

Zusammen mit $N_{H+H}(0) = 1$ folgt

$$N_{H+H}(p) = 18n \log_2 n + 59n - 58. \quad (3.8)$$

Im dritten Fall lässt sich B als $\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ mit $B_{ij} \in \mathcal{R}_{p-1}$ schreiben:

$$B = \begin{bmatrix} \mathcal{R}_{p-1} & \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} & \mathcal{R}_{p-1} \end{bmatrix} \text{ (vgl. Anmerkung 2.2.6). Die Summe}$$

$$A + B = \begin{bmatrix} \mathcal{H}_{p-1} + \mathcal{R}_{p-1} & \mathcal{R}_{p-1} + \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} + \mathcal{R}_{p-1} & \mathcal{H}_{p-1} + \mathcal{R}_{p-1} \end{bmatrix}$$

zeigt die Rekursion $N_{H+R}(p) = 2N_{H+R}(p-1) + 2N_{R+R}(p-1)$. Da diese und der Startwert $N_{H+R}(0) = 1$ identisch mit denjenigen für $N_{H+H}(p)$ sind, ergeben sie ebenfalls $N_{H+R}(p) = 18n \log_2 n + 59n - 58$.

Lemma 3.5.1. *Sei $n = 2^p$. Die formatierte Addition \oplus_1 zweier Matrizen aus \mathcal{H}_p benötigt ebenso wie diejenige einer \mathcal{H}_p -Matrix mit einer \mathcal{R}_p -Matrix $18n \log_2 n + 59n - 58$ Operationen.*

3.6 Matrix-Matrix-Multiplikation

Sei $n = 2^p$. Bei der Matrix-Matrix-Multiplikation sind drei verschiedene Fälle und die zugehörigen Kosten zu unterscheiden:

- | | | | |
|-----|-------------------------------|---|-------------------------------------|
| 1) | $A \cdot B \in \mathcal{R}_p$ | für $A, B \in \mathcal{R}_p$ | mit den Kosten $N_{R \cdot R}(p)$. |
| 2a) | $A \cdot B \in \mathcal{R}_p$ | für $A \in \mathcal{R}_p$ und $B \in \mathcal{H}_p$ | mit den Kosten $N_{R \cdot H}(p)$. |
| 2b) | $A \cdot B \in \mathcal{R}_p$ | für $A \in \mathcal{H}_p$ und $B \in \mathcal{R}_p$ | mit den Kosten $N_{H \cdot R}(p)$. |
| 3) | $A \odot B \in \mathcal{H}_p$ | für $A, B \in \mathcal{H}_p$ | mit den Kosten $N_{H \cdot H}(p)$. |

In den Fällen 1) und 2) sind die Resultate exakt. Im Falle 3) wird das Produkt approximativ in \mathcal{H}_p bestimmt.

Im ersten Fall lautet die Lösung $N_{R \cdot R}(p) = 3n - 1$ (vgl. Anmerkung 2.3.1c).

Im Falle $A \in \mathcal{H}_p, B \in \mathcal{R}_p$ verwendet man $A \cdot ab^\top = (Aa) \cdot b^\top$, d.h. das Resultat ist $a'b^\top \in \mathcal{R}_p$ mit $a' := Aa$ und kostet eine Matrix-Vektor-Multiplikation $A \cdot a$. Gemäß Lemma 3.4.1 sind die Kosten $N_{H \cdot R}(p) = 4n \log_2 n - n + 2$.

Für $B \in \mathcal{R}_p, A \in \mathcal{H}_p$ gilt entsprechend $BA = ab^\top \cdot A = a \cdot (A^\top b)^\top$, sodass $N_{R \cdot H}(p) = N_{H \cdot R}(p)$.

Im dritten Fall $A \odot B$ für $A, B \in \mathcal{H}_p$ hat das Produkt die Gestalt

$$\begin{aligned} & \begin{bmatrix} \mathcal{H}_{p-1} & \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} & \mathcal{H}_{p-1} \end{bmatrix} \cdot \begin{bmatrix} \mathcal{H}_{p-1} & \mathcal{R}_{p-1} \\ \mathcal{R}_{p-1} & \mathcal{H}_{p-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{H}_{p-1} \cdot \mathcal{H}_{p-1} + \mathcal{R}_{p-1} \cdot \mathcal{R}_{p-1} & \mathcal{H}_{p-1} \cdot \mathcal{R}_{p-1} + \mathcal{R}_{p-1} \cdot \mathcal{H}_{p-1} \\ \mathcal{R}_{p-1} \cdot \mathcal{H}_{p-1} + \mathcal{H}_{p-1} \cdot \mathcal{R}_{p-1} & \mathcal{R}_{p-1} \cdot \mathcal{R}_{p-1} + \mathcal{H}_{p-1} \cdot \mathcal{H}_{p-1} \end{bmatrix}. \end{aligned}$$

Auf der Stufe $p - 1$ treten alle drei Multiplikationstypen auf. Der dritte Multiplikationstyp $\mathcal{H}_{p-1} \cdot \mathcal{H}_{p-1}$ erfordert die Approximation durch \odot . Schließlich ist die Addition mittels \oplus_1 anzunähern. Abzählen der Operationen liefert die Rekursion

$$\begin{aligned} N_{H \cdot H}(p) &= 2N_{H \cdot H}(p-1) + 2N_{R \cdot R}(p-1) + 2N_{H \cdot R}(p-1) \\ &\quad + 2N_{R \cdot H}(p-1) + 2N_{H+R}(p-1) + 2N_{R+R}(p-1). \end{aligned}$$

Wir setzen die bekannten Größen

$$N_{R \cdot R}(p-1) = \frac{3}{2}n - 1, \tag{3.9}$$

$$N_{H \cdot R}(p-1) = N_{R \cdot H}(p-1) = 4\frac{n}{2} \log_2 \frac{n}{2} - \frac{n}{2} + 2 = 2n \log_2 n - \frac{5}{2}n + 2,$$

$$N_{H+R}(p-1) = 18\frac{n}{2} \log_2 \frac{n}{2} + 59\frac{n}{2} - 58 = 9n \log_2 n + \frac{41}{2}n - 58,$$

$$N_{R+R}(p-1) = 18\frac{n}{2} + 29 = 9n + 29$$

ein und erhalten $N_{H \cdot H}(p) = 2N_{H \cdot H}(p-1) + 26pn + 52n - 52$. Diese Rekursion und der Startwert $N_{H \cdot H}(0) = 1$ ergeben die Lösung $N_{H \cdot H}(p) = 13np^2 + 65np - 51n + 52$.

Lemma 3.6.1. *Die Multiplikation zweier \mathcal{H}_p -Matrizen kostet*

$$N_{H \cdot H}(p) = 13n \log_2^2 n + 65n \log_2 n - 51n + 52 \text{ Operationen.}$$

Das Produkt zwischen \mathcal{H}_p und \mathcal{R}_p erfordert

$$N_{H \cdot R}(p) = N_{R \cdot H}(p) = 4n \log_2 n - n + 2 \text{ Operationen,}$$

die Multiplikation zweier \mathcal{R}_p -Matrizen benötigt

$$N_{R \cdot R}(p) = 3n - 1 \text{ Operationen.}$$

3.7 Matrixinversion

Im Folgenden wollen wir die Inverse M^{-1} einer Matrix $M \in \mathcal{H}_p$ approximieren. Dazu wird die Inversionsabbildung $Inv : D_p \subset \mathcal{H}_p \rightarrow \mathcal{H}_p$ rekursiv definiert. Für $p = 0$ kann $Inv(M) := M^{-1}$ als exakte Inverse der 1×1 -Matrix M definiert werden, solange $M \neq 0$. Sei Inv auf $D_{p-1} \subset \mathcal{H}_{p-1}$ definiert. Die (exakte) Inverse von M mit der Blockstruktur (3.2d) ist

$$M^{-1} = \begin{bmatrix} M_{11}^{-1} + M_{11}^{-1} M_{12} S^{-1} M_{21} M_{11}^{-1} & -M_{11}^{-1} M_{12} S^{-1} \\ -S^{-1} M_{21} M_{11}^{-1} & S^{-1} \end{bmatrix}, \quad (3.10)$$

wobei das *Schur-Komplement* $S := M_{22} - M_{21} M_{11}^{-1} M_{12}$ benötigt wird. Man beachte, dass die Darstellung (3.10) und damit auch der zu beschreibende Algorithmus voraussetzen, dass M_{11} regulär ist.

- Übung 3.7.1.** a) Ist M positiv definit, so ist M_{11} regulär.
b) Sind M und M_{11} regulär, so ist auch das Schur-Komplement S regulär.

In (3.10) wird M_{11}^{-1} durch $Inv(M_{11})$ ersetzt. Die Multiplikationen mit M_{12} und M_{21} können exakt durchgeführt werden, da diese aus \mathcal{R}_{p-1} stammen. Die Additionen (zu denen auch die Subtraktion gezählt wird) werden als \oplus_1 durchgeführt. Damit können sowohl S als auch alle Matrixblöcke aus (3.10) approximativ berechnet werden, und $Inv(M)$ ist vollständig definiert. Die genaue Abfolge der Operationen ist

Matrixoperation	Kosten	approximierter Ausdruck
$M_{11} \mapsto N_{11} := Inv(M_{11}) \in \mathcal{H}_{p-1}$	N_{inv}	M_{11}^{-1}
$M_{21}, N_{11} \mapsto X_{21} := M_{21} \cdot N_{11} \in \mathcal{R}_{p-1}$	$N_{R \cdot H}$	$M_{21} M_{11}^{-1}$
$N_{11}, M_{12} \mapsto X_{12} := N_{11} \cdot M_{12} \in \mathcal{R}_{p-1}$	$N_{H \cdot R}$	$M_{11}^{-1} M_{12}$
$X_{21}, M_{12} \mapsto X_{22} := X_{21} \cdot M_{12} \in \mathcal{R}_{p-1}$	$N_{R \cdot R}$	$M_{21} M_{11}^{-1} M_{12}$
$M_{22}, X_{22} \mapsto \hat{S} := M_{22} \ominus X_{22} \in \mathcal{H}_{p-1}$	N_{H+R}	$M_{22} - M_{21} M_{11}^{-1} M_{12}$
$\hat{S} \mapsto T := Inv(\hat{S}) \in \mathcal{H}_{p-1}$	N_{inv}	S^{-1}
$T, X_{21} \mapsto Z_{21} := -T \cdot X_{21} \in \mathcal{R}_{p-1}$	$N_{H \cdot R}$	$-S^{-1} M_{21} M_{11}^{-1}$
$X_{12}, T \mapsto Z_{12} := -X_{12} \cdot T \in \mathcal{R}_{p-1}$	$N_{R \cdot H}$	$-M_{11}^{-1} M_{12} S^{-1}$
$X_{12}, Z_{21} \mapsto X_{11} := X_{12} \cdot Z_{21} \in \mathcal{R}_{p-1}$	$N_{R \cdot R}$	$-M_{11}^{-1} M_{12} S^{-1} M_{21} M_{11}^{-1}$
$N_{11}, X_{11} \mapsto Z_{11} := N_{11} \oplus_1 X_{11} \in \mathcal{H}_{p-1}$	N_{H+R}	$M_{11}^{-1} + M_{11}^{-1} M_{12} S^{-1} M_{21} M_{11}^{-1}$

Somit ist $Inv(M) = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & T \end{bmatrix}$ bestimmt. Die Addition der aufgezählten Kosten ergibt die Rekursion

$$N_{\text{inv}}(p) = 2N_{\text{inv}}(p-1) + 2N_{R.H}(p-1) + 2N_{H.R}(p-1) + 2N_{H+R}(p-1) + 2N_{R.R}(p-1)$$

(sehr ähnlich zur $N_{H.H}$ -Rekursion). Mit den Werten aus (3.9) folgt

$$N_{\text{inv}}(p) = 2N_{\text{inv}}(p-1) + 26n \log_2 n + 34n - 110.$$

Zusammen mit $N_{\text{inv}}(0) = 1$ ergibt sich $N_{\text{inv}}(p) = 13p^2n + 47pn - 109n + 110$. Dies beweist das

Lemma 3.7.2. *Die approximative Inversion einer Matrix aus \mathcal{H}_p benötigt $13n \log_2^2 n + 47n \log_2 n - 109n + 110$ Operationen.*

3.8 LU-Zerlegung

Eine LU-Zerlegung (ohne Pivotwahl) braucht nicht zu existieren. Hinreichend sind a) nichtverschwindende Hauptunterdeterminanten, b) Positivdefinitheit oder c) H-Matrix-Eigenschaft (vgl. [66, Kriterium 8.5.8]).

Die hierarchischen Matrizen für die Faktoren L und U stammen aus den folgenden Untermengen:

$$\begin{aligned}\mathcal{H}_{p,L} &:= \{M \in \mathcal{H}_p : M_{ii} = 1, M_{ij} = 0 \text{ für } j > i\}, \\ \mathcal{H}_{p,U} &:= \{M \in \mathcal{H}_p : M_{ij} = 0 \text{ für } j < i\}.\end{aligned}$$

Wie bei der üblichen Abspeicherung als volle Matrix gilt auch hier, dass der Speicherplatz für die beiden Matrizen $L \in \mathcal{H}_{p,L}$ und $U \in \mathcal{H}_{p,U}$ zusammen dem einer allgemeinen Matrix $M \in \mathcal{H}_p$ entspricht.

3.8.1 Vorwärtssubstitution

Seien eine normierte untere Dreiecksmatrix $L \in \mathcal{H}_{p,L}$ und eine rechte Seite $y \in \mathbb{R}^{I_p}$ gegeben, während die Lösung $x \in \mathbb{R}^{I_p}$ von $Lx = y$ zu bestimmen ist. Für $p \geq 1$ zerfällt L in die Blöcke

$$L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \quad \text{mit } L_{11}, L_{22} \in \mathcal{H}_{p-1,L} \quad \text{und } L_{21} \in \mathcal{R}_{p-1}.$$

Ferner seien die Vektoren blockzerlegt: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Die Lösung von $Lx = y$ geschieht mittels der Vorwärtssubstitution, die zur Rekursion

$$\text{löse } L_{11}x_1 = y_1, \quad \text{setze } z := y_2 - L_{21}x_1, \quad \text{löse } L_{22}x_2 = z$$

führt. Nach Anmerkung 2.3.1a benötigt man $3\frac{n}{2} - 1$ Operationen für die Matrixvektormultiplikation $L_{21}x_1$. Die Addition von y_2 kostet $\frac{n}{2}$ Operationen, sodass die Rekursion für den Aufwand wie folgt aussieht:

$$N_{\text{vorw}}(p) = 2N_{\text{vorw}}(p-1) + 2n - 1 \quad (3.11a)$$

Für $p = 0$ ist die Auflösung von $Lx = y$ kostenlos, da¹ $x = y$, sodass $N_{\text{vorw}}(0) = 0$. Die Lösung der Rekursion für $N_{\text{vorw}}(p)$ lautet

$$N_{\text{vorw}}(p) = 2n \log_2 n - n + 1 \quad (n = 2^p).$$

3.8.2 Rückwärtssubstitution

Die Kosten der Auflösung der Gleichung $Ux = y$ für $U \in \mathcal{H}_{p,U}$ seien mit $N_{\text{rückw}}(p)$ bezeichnet. Die Rekursionsformel ist mit N_{vorw} identisch: $N_{\text{rückw}}(p) = 2N_{\text{rückw}}(p-1) + 2n - 1$, aber der Startwert lautet $N_{\text{rückw}}(0) = 1$. Dies liefert

$$N_{\text{rückw}}(p) = 2n \log_2 n + 1 \quad (n = 2^p). \quad (3.11b)$$

Im nächsten Unterabschnitt benötigt man eine Variante der Rückwärtssubstitution: die Auflösung von $x^\top U = y^\top$ nach x . Sie ist äquivalent zu $U^\top x = y$, wobei U^\top eine obere Dreiecksmatrix ist. Da U^\top aber nicht normiert ist, ergibt sich der gleiche Aufwand $N_{\text{rückw}}(p)$ wie oben.

3.8.3 Aufwand der LU-Zerlegung

Der Ansatz $L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \in \mathcal{H}_{p,L}$, $U = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix} \in \mathcal{H}_{p,U}$ für $LU = M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \in \mathcal{H}_p$ führt auf die vier Gleichungen

$$M_{11} = L_{11}U_{11}, \quad M_{12} = L_{11}U_{12}, \quad M_{21} = L_{21}U_{11}, \quad M_{22} = L_{21}U_{12} + L_{22}U_{22}.$$

Damit stellen sich die folgenden Unteraufgaben:

- 1) bestimme die LU-Zerlegung von M_{11} (Resultat: L_{11}, U_{11}),
- 2) berechne $U_{12} := L_{11}^{-1}M_{12}$ und $L_{21} := M_{21}U_{11}^{-1}$,
- 3) berechne die LU-Zerlegung von $M_{22} - L_{21}U_{12}$ (Resultat: L_{22}, U_{22}).

Da $M_{12} \in \mathcal{R}_{p-1}$, gilt $M_{12} = ab^\top$ für geeignete $a, b \in \mathbb{R}^{I_{p-1}}$. Die Darstellung von $U_{12} \in \mathcal{R}_{p-1}$ ist gegeben durch $a'b^\top$ mit $a' = L_{11}^{-1}a$. Offenbar erhält man a' mittels Vorwärtssubstitution aus $L_{11}a' = a$ zu den Kosten $N_{\text{vorw}}(p-1)$. Analog kostet die (exakte) Berechnung von $L_{21} = M_{21}U_{11}^{-1} \in \mathcal{R}_{p-1}$ gerade $N_{\text{rückw}}(p-1)$. Es ergibt sich eine rekursive Aufwandsbeschreibung:

$$\begin{aligned} N_{LU}(p) &= 2N_{LU}(p-1) + N_{\text{vorw}}(p-1) + N_{\text{rückw}}(p-1) \\ &\quad + N_{R \cdot R}(p-1) + N_{H+R}(p-1) \\ &= 2N_{LU}(p-1) + \left[n(p-1) - \frac{n}{2} + 1 \right] + [n(p-1) + 1] \\ &\quad + \left[3\frac{n}{2} - 1 \right] + \left[9n(p-1) + 59\frac{n}{2} - 58 \right] \end{aligned}$$

¹ Hier wird ausgenutzt, dass L normiert ist, d.h. $L_{ii} = 1$ für alle $i \in I$.

$$= 2N_{LU}(p-1) + 11np + \frac{39}{2}n - 57$$

mit dem Startwert $N_{LU}(0) = 0$. Die Lösung lautet

$$N_{LU}(p) = \frac{11}{2}n \log_2^2 n + 25n \log_2 n - 57(n-1). \quad (3.11c)$$

Damit ist der Aufwand deutlich niedriger als die Berechnung der Inversen mit $N_{\text{inv}}(p) = 13n \log_2^2 n + \dots$.

Übung 3.8.1. Für positiv definite Matrizen M im \mathcal{H}_p -Format formuliere man die Cholesky-Zerlegung (vgl. (1.5b)).

3.9 Weitere Eigenschaften der Modellmatrizen und Semiseparabilität *

Die Inversionsabbildung Inv aus §3.7 wurde als approximativ beschrieben. Es gibt aber einen wichtigen Fall, in dem Inv exakt ist:

Proposition 3.9.1 (tridiagonale Matrizen). *Sei $M \in \mathbb{R}^{I_p \times I_p}$ tridiagonal.*

- a) Dann gilt $M \in \mathcal{H}_p$.*
- b) Ist M zudem regulär, so gehört auch M^{-1} zu \mathcal{H}_p .*
- c) Alle Hauptuntermatrizen $M|_{I_q \times I_q}$ ($0 \leq q \leq p$) seien regulär. Dann ist das Resultat $Inv(M)$ aus §3.7 wohldefiniert und liefert die exakte Inverse M^{-1} .*

Der Beweis wird anschließend an Korollar 3.9.9 nachgeholt.

Proposition 3.9.1 ist leicht auf Bandmatrizen mit Bandbreite $2k+1$ (k obere und k untere Außerdiagonalen, $k > 1$) zu verallgemeinern, indem man in der Definition von \mathcal{H}_p statt der Rang-1-Matrizen $\mathcal{R}_p = \mathcal{R}(1, I_p, I_p)$ die Rang- k -Matrizen $\mathcal{R}(k, I_p, I_p)$ verwendet.

Die obige Aussage macht davon Gebrauch, dass die Inverse tridiagonaler Matrizen spezielle Eigenschaften besitzt. Es gibt auch die umgekehrte Fragestellung: Unter welchen Bedingungen hat eine Matrix eine tridiagonale Inverse? Dies führt auf den Begriff der *Semiseparabilität*. Da tridiagonale Matrizen (oder etwas breitere Bandmatrizen) sich aus eindimensionalen Randwertaufgaben ergeben und in der Praxis eher zwei oder mehr Raumvariablen auftreten, ist die Anwendbarkeit der Semiseparabilitätseigenschaften allerdings beschränkt.

Weil in der Literatur die Semiseparabilität nicht einheitlich definiert ist (vgl. [132]), wird hier auf eine Definition verzichtet. Stattdessen beschreibt Definition 3.9.2 eine Menge, die hier neutral als \mathcal{S}_k bezeichnet sei und den semiseparablen Matrizen nahekommt. Für unsere Zwecke wird eine schwächere Bedingung an die Menge $\mathcal{M}_{k,\tau}$ ausreichen (vgl. Definition 3.9.5). Die hier definierten \mathcal{S}_k - und $\mathcal{M}_{k,\tau}$ -Matrizen haben interessante Invarianzeigenschaften bezüglich verschiedener Operationen.

Definition 3.9.2. Seien I angeordnet und $1 \leq k < \#I$. $M \in \mathbb{R}^{I \times I}$ gehört zu \mathcal{S}_k , wenn $\text{Rang}(M|_b) \leq k$ für jeden Block $b \subset I \times I$ gilt, der im strikten oberen Dreiecksteil $\{(i, j) : i < j\}$ oder im strikten unteren Dreiecksteil enthalten ist.

Offenbar gehört jede Matrix $M \in \mathbb{R}^{I \times I}$ zu \mathcal{S}_k mit $k = \#I - 1$.

Anmerkung 3.9.3. a) Tridiagonale Matrizen gehören zu \mathcal{S}_1 .

b) Bandmatrizen mit höchstens k oberen und k unteren Nebendiagonalen gehören zu \mathcal{S}_k .

c) Sei $D \in \mathbb{R}^{I \times I}$ diagonal. Dann gehören M und $M+D$ zu \mathcal{S}_k mit gleichem k .

d) Jede Matrix $M \in \mathcal{S}_k \cap \mathbb{R}^{n \times n}$ ($n = 2^p$) lässt sich im Format \mathcal{H}_p exakt darstellen, wenn anstelle von (3.2c) der lokale Rang k gewählt wird.

Beweis. a) Spezialfall $k = 1$ von b). b) $M|_b$ enthält höchstens k Nichtnullzeilen. c) Die Diagonale ist für \mathcal{S}_k irrelevant. d) Die Diagonalblöcke im \mathcal{H}_p -Format sind (volle) Matrizen der Größe 1×1 und enthalten ohnehin die exakten Daten. Alle anderen Blöcke b sind ganz im oberen bzw. unteren Matrixdreieck enthalten, sodass $\text{Rang}(M|_b) \leq k$. Damit können sie aber exakt im Format $\mathcal{R}(k, b)$ dargestellt werden. ■

Die folgende Übung knüpft an eine andere Definition semiseparabler Matrizen an.

Übung 3.9.4. Sei $I = \{1, \dots, n\}$. Man zeige: a) Wenn es Matrizen $M^o, M^u \in \mathcal{R}(k, I, I)$ gibt, sodass $M_{ij} = \begin{cases} M_{ij}^o & \text{für } j > i \\ M_{ij}^u & \text{für } j < i \end{cases}$, so ist $M \in \mathcal{S}_k$.

b) Für alle $1 \leq \nu \leq n - 1$ und die zugehörigen Blöcke $b = \{1, \dots, \nu\} \times \{\nu + 1, \dots, n\}$ gelte, dass die erste Spalte in $M|_b$ linear abhängig von den übrigen sei. Dann gibt es ein $M^o \in \mathcal{R}(k, I, I)$, sodass $M_{ij} = M_{ij}^o$ für $j > i$. Man formuliere eine entsprechende Bedingung, sodass auch $M_{ij} = M_{ij}^u$ für $j < i$ mit einem $M^u \in \mathcal{R}(k, I, I)$.

Im Weiteren untersuchen wir eine Matrixfamilie $\mathcal{M}_{k,\tau}$ mit schwächeren Eigenschaften. Insbesondere braucht I nicht angeordnet zu sein.

Definition 3.9.5. Seien $\emptyset \neq \tau \subset I$ eine Indexteilmenge, $\tau' := I \setminus \tau$ ihr Komplement und $k \in \mathbb{N}$. Eine Matrix A gehört zu $\mathcal{M}_{k,\tau}(I)$, wenn $\text{Rang}(A|_{\tau \times \tau'}) \leq k$ und $\text{Rang}(A|_{\tau' \times \tau}) \leq k$. Falls die Angabe von I entbehrlich ist, wird auch $\mathcal{M}_{k,\tau}$ anstelle von $\mathcal{M}_{k,\tau}(I)$ geschrieben.

Falls die Indizes so angeordnet werden, dass zuerst die Indizes aus τ kommen und dann die aus τ' folgen, erhalten wir die Blockaufteilung

$$A = \begin{array}{cc|c} & \tau & \tau' = I \setminus \tau & \\ \hline & A_{11} & A_{12} & \tau \\ A_{21} & A_{22} & & \tau' \end{array} \quad (3.12)$$

Definition 3.9.5 besagt, dass $\text{Rang}(A_{12}) \leq k$ und $\text{Rang}(A_{21}) \leq k$.

Der Zusammenhang mit \mathcal{S}_k wird gegeben durch die

Anmerkung 3.9.6. Sei $I = \{1, \dots, n\}$ angeordnet. $M \in \mathbb{R}^{I \times I}$ gehört genau dann zu \mathcal{S}_k , wenn für alle $i < n$ die Eigenschaft $M \in \mathcal{M}_{k,\tau}(I)$ für $\tau = \{1, \dots, i\}$ gilt.

Im Folgenden sind die Matrixoperationen $*,^{-1}, +$ in ihrer exakten Form, d.h. ohne jede Kürzung gemeint.

Lemma 3.9.7. a) Seien $A \in \mathcal{M}_{k_A,\tau}(I)$ und $B \in \mathcal{M}_{k_B,\tau}(I)$. Dann gilt $A * B \in \mathcal{M}_{k,\tau}(I)$ für $k = k_A + k_B$.

b) Sei $A \in \mathcal{M}_{k,\tau}(I)$ regulär. Dann gilt $A^{-1} \in \mathcal{M}_{k,\tau}(I)$ mit gleichem k .

c) Sei $A \in \mathcal{M}_{k,\tau}(I)$. Dann ist $A + D \in \mathcal{M}_{k,\tau}(I)$ für alle Diagonalmatrizen $D \in \mathbb{R}^{I \times I}$.

d) Sei $A \in \mathcal{M}_{k,\tau}(I)$ und $\emptyset \neq \tau \subset I' \subsetneq I$. Dann gehört die Hauptuntermatrix $A|_{I' \times I'}$ zu $\mathcal{M}_{k,\tau}(I')$. Die gleiche Aussage gilt für das Schur-Komplement $S_{I'} = A|_{I' \times I'} - A|_{I' \times I''} * (A|_{I'' \times I''})^{-1} * A|_{I'' \times I'}$ ($I'' := I \setminus I'$), falls dieses existiert.

Beweis. a) Mit der (3.12) entsprechenden Notation für A, B und $C := AB$ ist $C_{12} = A_{11}B_{12} + A_{12}B_{22}$. Aus $\text{Rang}(A_{11}B_{12}) \leq \text{Rang}(B_{12}) \leq k_B$ und $\text{Rang}(A_{12}B_{22}) \leq \text{Rang}(A_{12}) \leq k_A$ schließt man auf $\text{Rang}(C_{12}) \leq k_A + k_B$. Analog für $\text{Rang}(C_{21})$.

b1) Sei A_{11} als regulär angenommen. Dann ist das Schur-Komplement $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ ebenfalls regulär, und die Inverse von A aus (3.12) ist

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{bmatrix} \quad (3.13)$$

(vgl. (3.10)). Da $\text{Rang}(A^{-1}|_{\tau \times \tau'}) = \text{Rang}(-A_{11}^{-1}A_{12}S^{-1}) \leq \text{Rang}(A_{12}) \leq k$ wie auch $\text{Rang}(A|_{\tau' \times \tau}) \leq \text{Rang}(A_{12}) \leq k$, folgt $A^{-1} \in \mathcal{M}_{k,\tau}(I)$.

b2) Falls A_{11} singulär ist, wird die Matrix $A_\varepsilon := A + \varepsilon I$ für hinreichend kleines $\varepsilon \neq 0$ regulär. Da $\text{Rang}(A_\varepsilon^{-1}|_{\tau \times \tau'}) \leq \text{Rang}(A_{12})$ unabhängig von ε gilt, folgt $A_\varepsilon^{-1} \in \mathcal{M}_{k,\tau}(I)$. Der Limes $\lim_{\varepsilon \rightarrow 0} A_\varepsilon^{-1}$ ist A^{-1} , da nach Annahme A regulär ist. Der Rang erfüllt nach Übung 2.1.2 $\text{Rang}(A^{-1}|_{\tau \times \tau'}) = \text{Rang}(\lim_{\varepsilon \rightarrow 0} A_\varepsilon^{-1}|_{\tau \times \tau'}) \leq \lim_{\varepsilon \rightarrow 0} \text{Rang}(A_\varepsilon^{-1}|_{\tau \times \tau'}) \leq \text{Rang}(A_{12}) \leq k$. Zusammen mit der analogen Ungleichung $\text{Rang}(A^{-1}|_{\tau' \times \tau}) \leq k$ ergibt sich die Behauptung $A^{-1} \in \mathcal{M}_{k,\tau}(I)$.

c) Änderung der Diagonale ändert nicht die Teile $*|_{\tau \times \tau'}$ und $*|_{\tau' \times \tau}$.

d1) Restriktion der Matrix auf $I' \times I' \subset I \times I$ kann den Rang nur verkleinern: $A|_{I' \times I'} \in \mathcal{M}_{k,\tau}(I')$.

d2) Sei A als regulär angenommen. Das inverse Schur-Komplement $(S_{I'})^{-1}$ ist die Haupt- $I' \times I'$ -Untermatrix von A^{-1} . Nach Teil b) folgt $A^{-1} \in \mathcal{M}_{k,\tau}(I)$, und d1) zeigt $(S_{I'})^{-1} \in \mathcal{M}_{k,\tau}(I')$. Erneute Anwendung von b) mit $(S_{I'})^{-1}$ anstelle von A liefert die Behauptung $S_{I'} \in \mathcal{M}_{k,\tau}(I')$. Für singuläres A argumentiert man wie in b2).

Eine Konsequenz von Lemma 3.9.7 ist die folgende Aussage. ■

Lemma 3.9.8. *Sei $R(\cdot)$ eine rationale Funktion $R(x) = P^I(x)/P^{II}(x)$ mit Polynomen P^I, P^{II} der Grade $d_I, d_{II} \in \mathbb{N}_0$. Die Eigenwerte von $A \in \mathcal{M}_{k,\tau}(I)$ seien verschieden von den Polen von R . Dann gehört die Matrix² $R(A)$ zu $\mathcal{M}_{k_R,\tau}(I)$ mit $k_R = k * d_R$, wobei $d_R := \max(d_I, d_{II})$ der Grad von R ist.*

Beweis. Die Polynome P^I, P^{II} seien in $P^I(x) = a_I \prod_{i=1}^{d_I} (x - x_i^I)$ und $P^{II}(x) = a_{II} \prod_{i=1}^{d_{II}} (x - x_i^{II})$ faktorisiert³. Für $i \leq \min(d_I, d_{II})$ treten rationale Faktoren $\frac{x - x_i^I}{x - x_i^{II}} = 1 + (x_i^{II} - x_i^I) / (x - x_i^{II})$ auf. Ersetzung von x durch $A \in \mathcal{M}_{k,\tau}(I)$ liefert $R_i(A) := I + (x_i^{II} - x_i^I) (A - x_i^{II} I)^{-1} \in \mathcal{M}_{k,\tau}(I)$ gemäß Lemma 3.9.7c. Also ist $R(A)$ ein Produkt von $\min(d_I, d_{II})$ rationalen Faktoren $R_i(A)$ und $\max(d_I, d_{II}) - \min(d_I, d_{II})$ Faktoren der Form $A - x_i^I I$ für $d_I > d_{II}$ und $(A - x_i^{II} I)^{-1}$ für $d_I < d_{II}$, die alle zu $\mathcal{M}_{k,\tau}(I)$ gehören. Nach Lemma 3.9.7a liegt das Produkt in $\mathcal{M}_{k_R,\tau}(I)$ mit $k_R = kd_R$. ■

Korollar 3.9.9. *Die bisherigen Aussagen übertragen sich gemäß Anmerkung 3.9.6 entsprechend auf \mathcal{S}_k -Matrizen. Zum Beispiel ist die Inverse, wenn sie existiert, wieder in \mathcal{S}_k .*

Wir holen nun den *Beweis* zu Proposition 3.9.1 nach.

a) Teil a) folgt aus Anmerkung 3.9.3a,d.

b) Teil b) folgt aus Korollar 3.9.9.

c) Nachdem Teil b) verwendet werden kann, bleibt nur zu zeigen, dass der Algorithmus *Inv* wohldefiniert ist und keine Approximationsfehler einführt. Wir verwenden Induktion über p . Definitionsgemäß ist *Inv* auf \mathcal{H}_0 exakt. Die Aussage gelte für $p - 1$. $M \in \mathbb{R}^{I_{p-1} \times I_{p-1}}$ sei gemäß (3.2a) zerlegt. Die Untermatrizen M_{11}, M_{22} sind wieder tridiagonal, wobei M_{11} nach Voraussetzung regulär ist und nach Induktionsvoraussetzung $M_{11}^{-1} = \text{Inv}(M_{11})$ gilt. Bei der Berechnung von $M_{21} \odot \text{Inv}(M_{11}) \odot M_{12}$ sind alle Zwischenresultate vom Rang ≤ 1 und werden somit exakt ausgewertet. Lemma 3.9.7d zeigt, dass $S = M_{22} - M_{21} M_{11}^{-1} M_{12}$ ebenso wie die nach Übung 3.7.1 existierende Inverse S^{-1} im \mathcal{H}_{p-1} -Format darstellbar ist. Die exakte Inverse von M ist durch (3.10) gegeben. Für die Nebendiagonalblöcke $-M_{11}^{-1} M_{12} S^{-1}$ und $-S^{-1} M_{21} M_{11}^{-1}$ gilt ebenfalls, dass nicht nur das Gesamtprodukt, sondern auch die Zwischenresultate exakt in \mathcal{R}_{p-1} dargestellt werden. Damit ist *Inv* auch auf der Stufe p exakt.

Zusammenhänge zwischen $\mathcal{M}_{k,\tau}$ und der schwachen Zulässigkeit werden in §9.3.3 diskutiert werden.

² Zu Matrixfunktionen sei auf die späteren Definitionen in §13.1 verwiesen.

³ Entgegen der bisherigen Beschränkung auf \mathbb{R} können komplexe x_i^I, x_i^{II} auftreten.

Hierarchische Matrizen

Algorithmen und Analysis

Hackbusch, W.

2009, XX, 451 S. 65 Abb., Hardcover

ISBN: 978-3-642-00221-2