Proseminar
Hierarchische Matrizen

Arithmetik der $\mathcal{H}$-Matrizen
Die Multiplikation

Crameri Remo, 4. Juni 2004

Gleiderung des Vortrages

0. Einführung und Motivation

1. Wiederholung der wichtigsten Begriffe

2. Die Multiplikation der $\mathcal{H}$-Matrizen

3. Zusammenfassung

Ziel des Vortrages

Einen Überblick in die Multiplikation der $\mathcal{H}$-Matrizen geben.

0. Einführung und Motivation

Um möglichst wenig Speicherplatz für eine Matrix zu benutzen, haben wir ein neues Format für Matrizen eingefürt. Wir erhalten die sogenannten $\mathcal{H}$-Matrizen.

In diesem Vortrag geht es darum, ein Algorithmus für eine möglichst effiziente Multiplikation zwischen zwei $\mathcal{H}$-Matrizen zu entwickeln. Dazu brauchen wir zuerst einige Vorbereitungen und Definitionen:

1. Wiederholung der wichtigsten Begriffe

Definition 1 (`Fullmatrix`-Darstellung):

Eine $n \times m$ Matrix M besitzt eine `Fullmatrix`-Darstellung, falls die Einträge M_{ij} in einem Array der Länge nm wie folgt gespeichert sind:

$$M_{11}, ..., M_{n,1}, M_{1,2}, ..., M_{n,2}, ..., M_{1,m}, ..., M_{nm}$$

Definition 2 (`rk-Matrix`-Darstellung):

Eine $n \times m$ Matrix M mit Rang höchstens k bestitzt die **`rk-matrix`**-Darstellung, falls sie folgende Form besitzt:

$$M = AB^T\text{, wobei } A \in \mathbb{R}^{n\times k}\text{, } B \in \mathbb{R}^{m\times k}$$

wobei beide Matrizen in einem Array gespeichert sind.

Definition 3 ($\mathcal{R}$(k)-Matrizen):

Sei $n, m, k \in \mathbb{N}$. Wir definieren die Menge der $n \times m$-Matrizen vom Rang $\leq k$ wie folgt:

$$\mathcal{R}(k, n, m) := \left\{ M \in \mathbb{R}^{n \times m} \mid Rang(M) \leq k. \right\}$$

Bemerkung: Betrache:

$$A \in \mathcal{R}(k, n, m), B \in \mathcal{R}(k, n, m) \Longrightarrow A + B \in \mathcal{R}(2k, n, m)$$

Definition 4 ($\mathcal{H}$-Matrix):

Sei $T_{\mathcal{I}\times\mathcal{I}}$ ein block cluster Baum für eine Index Menge $\mathcal{I}$. Die Menge der $\mathcal{H}$-Matrizen ist wie folgt definiert:

$$H(\mathcal{T}_{\mathcal{I}\times\mathcal{I}}, k) :=$$

$$\Big\{ M \in \mathbb{R}^{\mathcal{I}\times\mathcal{I}} \mid Rang(M \mid_{t\times s}) \leq k \text{ für alle zulässige Niveaus } t \times s \text{ von } \mathcal{T}_{\mathcal{I}\times\mathcal{I}} \Big\}$$

Definition 5 ($\mathcal{H}$-Matrix Darstellung):

Sei $T_{\mathcal{I}\times\mathcal{I}}$ ein block cluster Baum für eine Index Menge $\mathcal{I}$. Eine Matrix $M \in \mathcal{H}(\mathcal{T}_{\mathcal{I}\times\mathcal{I}}, k)$ besitzt die $\mathcal{H}$-Darstellung, falls die zu nichtzulässigen Niveaus gehörenden Untermatrizen eine `fullmatrix`-Darstellung haben, und die mit zulässigen Niveaus eine `rkmatrix`-Darstellung haben.

Wir möchten jetzt eine gegebene Matrix $M \in \mathbb{R}^{n\times m}$ durch eine Matrix $\tilde{M} \in \mathcal{R}(k,n,m)$ möglichst genau approximieren. Um dies zu tun, brauchen wir die Singulärwertzerlegung einer Matrix. Hier eine kurze Wiederholung:

Definition 6 (SVD und rSVD):

Sei $M \in \mathcal{R}(k,n,m)$. Eine Singulärwertzerlegung (SVD) von M ist eine Faktorisierung der Form

$$M = U\Sigma V^T$$

mit unitären Matrizen $U \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{m\times m}$ und eine diagonale Matrix $\Sigma \in \mathbb{R}^{n\times m}$, wobei die Diagonalelemente wie folgt angeordent sind:

$$\Sigma_{11} \geq \Sigma_{22} \geq ... \geq \Sigma_{kk} \geq \Sigma_{k+1,k+1} = ... = \Sigma_{min\{n,m\},min\{n,m\}} = 0$$

Die Diagonalelemente von Σ heissen die Singulärwerte von M.
Eine reduzierte Singulärwertzerlegung (rSVD) von M ist eine Faktorisierung der Form

$$M = U\Sigma V^T$$

mit Matrizen $U \in \mathbb{R}^{n\times k}$, $V \in \mathbb{R}^{m\times k}$ mit orthogonalen Zeilenvektoren und eine Diagonalmatrix $\Sigma \in \mathbb{R}^{k\times k}$ mit den Diagonalelementen

$$\Sigma_{11} \geq \Sigma_{22} \geq ... \geq \Sigma_{kk} > 0$$

Bemerkung: Eine solche Zerlegung ist nicht eindeutig!

Lemma 1 (Die beste Approximation mit fixem Rang:)

Sei $M = U\Sigma V^T$ eine SVD von $M \in \mathbb{R}^{n\times m}$. Dann:

$$\tilde{M} := \tilde{U}\tilde{\Sigma}\tilde{V}^T$$

mit Matrizen $\tilde{U} := U\mid_{n\times k}, \tilde{\Sigma} := diag(\Sigma_{11}, ..., \Sigma_{kk}), \tilde{V} := V\mid_{m\times k}$.

Dann ist $\tilde{M}$ eine beste Approximation von M, in dem Sinne, dass

$$\left\| M - \tilde{M} \right\| = min_{R\in\mathcal{R}(k,n,m)} \|M - R\|$$

in der Frobenius Norm ($\|M\|_F^2 := \sum M_{ij}^2$)

Beweis: Siehe Skript

Definition 7: Truncation $\mathcal{T}_k$ (das Kürzen $\mathcal{T}_k$):

Sei $M \in \mathbb{R}^{n\times m}$ und $k \in \mathbb{N}$. Wir definieren den truncation Operator $\mathcal{T}_k$ wie folgt:

$$\mathcal{T}_k : \qquad \mathbb{R}^{n\times m} \to \mathcal{R}(k,n,m) \qquad M \mapsto \tilde{M}$$

wobei $\tilde{M}$ eine beste Approximation von M in der Menge $\mathcal{R}(k,n,m)$ ist.

2. Die Multiplikation von $\mathcal{H}$-Matrizen

Im Folgenden werde wir uns damit beschäftigen, wie wir die Multiplikation zwischen zwei $\mathcal{H}$-Matrizen möglichst effizient bilden können.

Dazu zuerst einige Vorbereitungen und Wiederholungen der wichtigesten Begriffe, die wir brauchen werden:

Analog wie in der Definition 7 vom vorherigen Kapitel, führen wie den Operator $\tilde{T}_k$ ein:

Definition 1: Truncation $\tilde{\mathcal{T}}_k$ (das Kürzen $\tilde{\mathcal{T}}_k$):

Sei $T_{\mathcal{I}\times\mathcal{I}}$ ein block cluster Baum. Wir definieren den Operator $\tilde{T}_k$ wie folgt:

$$\tilde{\mathcal{T}}_k : \qquad \mathbb{R}^{n\times m} \to \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k) \qquad M \mapsto \tilde{M}$$

wobei für jedes Niveau $t \times s \in V(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$ gilt

$$\tilde{M} \mid_{t\times s} := \begin{cases} \mathcal{T}_k(M \mid_{t\times s}),\ \text{falls } t \times s \text{ zulässig ist} \\ \text{M}|_{t\times s} \text{ sonst.} \end{cases}$$

Lemma 1:

Der Operator $\tilde{\mathcal{T}}_k$ bildet die Matrix $M \in \mathbb{R}^{\mathcal{I}\times\mathcal{I}}$ in eine beste Approximation $\tilde{M} \in \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$ in dem Sinne, dass Folgendes gilt:

$$\left\|M - \tilde{M}\right\|_F = min_{M' \in \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}},k)} \|M - M'\|_F$$

Beweis: Siehe Skript

An dieser Stelle sind wir also im Stande, ein erstes Verfahren für die Multiplikation zwischen zwei $\mathcal{H}$-Matrizen zu definieren:

Definition 2: Die formatierte Multiplikation:

Sei $T_{\mathcal{I}\times\mathcal{I}}$ ein block cluster Baum und sei $k \in \mathbb{N}$. Wir definieren die **formatierte Multiplikation** von $\mathcal{H}$-Matrizen $A, B \in \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$ wie folgt:

$$A \odot B := \tilde{\mathcal{T}}_k(A \cdot B)$$

Bemerkung:
In Worten bilden wir also zuerst die exakte Multiplikation von A und B, und dann kürzen wir sie mit dem Operator $\tilde{T}_k$.
Das exakte Produkt wir mittels des Operators $\tilde{\mathcal{T}}_k$ auf das Format $\mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$ gekürzt. Falls aus irgendeinem Grund das Format $\mathcal{R}(k, n, m)$ praktischer wäre, sollte wir nur in der obigen Definition den Operator $\tilde{\mathcal{T}}_k$ durch den neuen Operator $\mathcal{T}_k$ ersetzen.

Das obenerwähnte Verfahren ist in manchen Fällen sehr nützlich. Sobald aber die Dimensionen der Matrizen A, B sehr gross werden, wird das Ausrechnen der exakten Multiplikation sehr teuer, und somit ist dieses Verfahren nicht mehr so praktisch.

Da unsere beiden Matrizen A, B nicht ein beliebiges Format haben, sonder $\mathcal{H}$-Matrizen sind, stellt sich die Frage, ob diese spezielle Eigenschaft doch nicht ausgenützt werden könnte.

In den nächsten Seiten beschäftigen wir uns damit.

Da $A, B \in \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$, bestehen sie aus Blöcken mit Format $\mathcal{R}(k, n, m)$ oder `fullmatrix`. Bei der Multiplikation führen wir also folgende Schritte durch:

1. Verschiedene Blöcke der Fomaten $\mathcal{R}(k, n, m)$ und `fullmatrix` werden miteinander multipliziert.

2. Wir bilden die Summe aus diesen neuen Teilprodukten und

3. wir erhalten damit die neuen Komponenten des Produktes.

Unser Ziel ist es, die Multiplikation von A, B so durchzuführen, dass das Produkt möglichst effizient ausgerechnet wird und das gewünschte Format hat, also zum Beispiel $\mathcal{H}$-Matrix oder auch $\mathcal{R}(k, n, m)$.

Bei jedem von den drei obengenannten Schritten geht aber die Struktur der $\mathcal{H}$-Matrizen verloren. Wir benötigen also für jeden dieser Schritte die Entwicklung von Algorithmen, die die Struktur unserer Matrix beibehalten. Im Folgenden werden sie der Reihe nach diskutiert:

1. Schritt:

Die Multiplikation von Blöcken der Fomaten $\mathcal{R}(k, n, m)$ und `fullmatrix` bereitet keine grosse Schwierigkeit. Wir unterscheiden 3 verschiedenen Fälle:

1. Beide Blöcke besitzen das `fullmatrix` Fomat: Das Produkt wird exakt ausgerechnet und mit Hilfe des Operators $\tilde{\mathcal{T}}_k$ ins Format $\mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$ gekürzt.

2. Ein Block besitzt das $\mathcal{R}(k,n,m)$ Format und das Andere das `fullmatrix` Format: Das Produkt besitzt dann aber trivialerweise das Format $\mathcal{R}(k,n,m)$, ohne dass man die zwei Matrizen ausmultiplizieren (Ersetze nur die Matrix R durch seine Faktorisierung AB^T!).

3. Beide Blöcke besitzen das $\mathcal{R}(k, n, m)$ Format: Durch Ausmultiplizieren des ersten Blockes erhalten wir aber eine Matrix im Format `fullmatrix` und daher sind wir wieder beim Punkt zwei!

2. Schritt:

Es geht jetzt darum (Punkt 2), eine Summe von Matrizen vom Format $\mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$ oder $\mathcal{R}(k, n, m)$ zu bilden. Weil aber im Allgemeinen das Kürzen einer Summe von m Summanden viel teuerer wird, als m mal das Kürzen von zwei Matrizen, führen wir an dieser Stelle den Operator $\mathcal{T}'_k$:

Definition 3: Fast truncation $\mathcal{T}'_k$ (Das schnelle Kürzen $\mathcal{T}'_k$):

Sei $T_{\mathcal{I}\times\mathcal{I}}$ ein block cluster Baum und $q \in \mathbb{N}_{>1}$. Wir definieren der Operator für das schnelle Kürzen wie folgt:

$$\mathcal{T}'_k : \qquad \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, qk) \to \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k) \qquad M \mapsto \tilde{M}$$

bei $q-1$ Hintereinanderdurchführung des Kürzens $\tilde{\mathcal{T}}_k$ für eine Matrix mit blockweise Rang $2k$:
Sei $M \sum_{i=1}^{q} M_i$ eine Zerlegung von M in q Matrizen $M_i \in \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$. Dann definieren wir:

$$\begin{gathered}\tilde{M}_1 = M_1 \\ \tilde{M}_i = \tilde{\mathcal{T}}_k(M_i + \tilde{M}_{i-1}) \\ \tilde{M} = \tilde{\mathcal{T}}_k(M_q + \tilde{M}_{q-1}).\end{gathered}$$

3. Schritt

Mit Hilfe des eingeführten schnellen Kürzens, besitzen nun die Blöcke unseres neuen Produktes das Format einer $\mathcal{H}$-Matrix. Falls das gewünschte Format das $\mathcal{H}$ Format ist, dann sind wir jetzt fertig.

Falls aber das definitive Format das $\mathcal{R}(k, n, m)$ sein sollte, hilft uns dabei die hierarchische Approximation:

Definition 4: Das Niveau vom Baum T:

Sei T ein Baum. Wir definieren das Niveau $l \in \mathbb{N}_o$ vom T wie folgt:

$$T^l := \{t \in V(T) \,|\, \text{Niveau}(t) = l$$

Auf jedem Niveau $l \in \mathbb{N}_o$ definieren wir weiter

$$\mathcal{L}(T, l) := T^l \cap \mathcal{L}(T).$$

Definition 5: Die Hierarchische Approximation:

Sei $\mathcal{T}_{\mathcal{I}\times\mathcal{I}}$ ein block cluster Baum und p :=depth($\mathcal{T}_{\mathcal{I}\times\mathcal{I}}$). Für jedes $M \in \mathcal{H}(\mathcal{I}\times\mathcal{I}, k)$ definieren wir die hierarchische Approximation $M_{\mathcal{H}} \in \mathcal{R}(k,n,m)$ von M in $p+1$ Schritte, wie folgt:

$$M_p\Big|_{t\times s} := \begin{cases} \mathcal{T}_k(M\,|_{t\times s}), \text{ falls } t\times s \text{ unzulässig ist} \\ \text{M}|_{t\times s} \text{ sonst} \end{cases}$$

$$M_{l-1}\Big|_{t\times s} := \begin{cases} \mathcal{T}_k(M_l\,|_{t\times s}), \text{ falls } t\times s \in T^{(l-1)}_{\mathcal{I}\times\mathcal{I}}, l = p, ..., 1, \\ M\,|_{t\times s} \text{ sonst} \end{cases}$$

$$M_{\mathcal{H}} = M_0$$

Die Matrix M wird also stufenweise aus das gewünschte Format $\mathcal{R}(k,n,m)$ gekürzt!

Lemma 2: Der Fehler bei der Hierarchischen Approximation:

Sei $T_{\mathcal{I}\times\mathcal{I}}$ ein block cluster Baum, p :=depth$(T_{\mathcal{I}\times\mathcal{I}}), M \in \mathcal{H}(T_{\mathcal{I}\times\mathcal{I}}, k)$ und $M_{\mathcal{H}}$ bezeichne die hierarchische Approximation von M. Dann gilt folgende Abschätzung für den Fehler:

$$\|M - M_{\mathcal{H}}\|_F \leq (2^{p+1} + 1)\,\|M - \mathcal{T}_k(M)\|_F\,.$$

Beweis: Siehe Skript.

3. Zusammenfassung

Die Implementierungen für die Multiplikation von $\mathcal{H}$-Matrizen sehen wir auf die folgende Folie:

THE END