

# Beschreibung des Dissertationsthemas „Nonnegative Tensor Factorization: Algorithms and Parallelization“

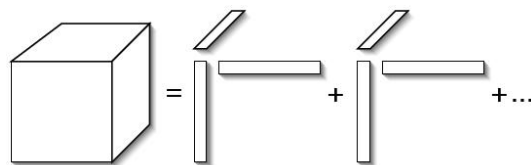
Markus Flatz

Doktoratsstudium; Technische Wissenschaften; Informatik (D 796 700 880)  
angestrebter akademischer Grad: Doktor der technischen Wissenschaften (Dr. techn.)

## Theoretischer Hintergrund

Nonnegative Tensor Factorization ist eine Verallgemeinerung der Nonnegative Matrix Factorization (NMF), mit der ich mich bereits im Rahmen meiner Masterarbeit „Nonnegative Matrix Factorization (NMF) Methods and Parallelization“ [3] beschäftigt habe. Matrizen sind Tensoren zweiter Stufe. Um auch mit Tensoren anderer Stufen arbeiten zu können, ist der Schritt von der Nonnegative Matrix Factorization zur Nonnegative Tensor Factorization notwendig.

Problemstellung: Nonnegative Tensor Factorization: Gegeben ist ein nichtnegativer Tensor und eine natürliche Zahl  $k$ . Gesucht werden  $k$  nichtnegative Tensoren mit Rang 1 (äußere Produkte von nichtnegativen Vektoren), deren Summe eine Approximation des ursprünglichen Tensors darstellt. Im Fall eines Tensors dritter Stufe kann man das folgendermaßen illustrieren (aus [5]):



Mögliche Anwendungsgebiete dieses Verfahrens sind unter anderem Data Mining, Bildverarbeitung, Bioinformatik und Materialerkennung anhand von Spektralanalysen [1]. Für praxisrelevante Problemgrößen ist die Berechnung der Faktorisierung mit einem hohen Rechenaufwand verbunden, daher ist eine Parallelisierung des verwendeten Berechnungsverfahrens wünschenswert.

# Arbeitsvorhaben

## Fragestellung

Das erste Ziel ist die Betrachtung der verschiedenen existierenden Algorithmen für Non-negative Tensor Factorization, deren Analyse und Vergleich. Im nächsten Schritt werden einige dieser Algorithmen in einer geeigneten Programmiersprache implementiert, um Tests, Experimente und Messungen durchführen zu können, wofür die Rechnersysteme des Fachbereichs Computerwissenschaften verwendet werden.

Ein besonderer Schwerpunkt wird auf der Analyse der Eignung der Verfahren für parallele Berechnung sowie deren Anpassung dafür liegen. Eine ausführliche Literaturrecherche ergab, dass zu diesem Thema erst wenig publiziert wurde. [5] zeigt Experimente auf einer Workstation mit 16 CPU-Kernen, [2] verwendet GPU-Programmierung. In beiden Fällen wurde nur ein Algorithmus implementiert und getestet. Für andere Verfahren ist noch keine parallelisierte Version zu finden, beispielsweise für das in [4] vorgestellte block-principal-pivoting-Verfahren, das als sequentielles Programm eine gute Performance zeigt und daher ein interessanter Kandidat für Parallelisierung ist. Es ist auch zu untersuchen, ob und in welchem Ausmaß die Erkenntnisse der Masterarbeit, insbesondere das parallelisierte Newton-Verfahren, auf die Nonnegative Tensor Factorization übertragen werden können.

## Methodik

Wie bereits oben erwähnt, werden mehrere Algorithmen implementiert, um Experimente durchführen zu können. Die Festlegung auf eine bestimmte Programmiersprache und ein Kommunikationsmodell erfolgt allerdings erst nach eingehender Analyse der Struktur der Algorithmen und ist auch abhängig von den verfügbaren Rechnersystemen am Fachbereich Computerwissenschaften, prinzipiell ist die Verwendung von CPUs, GPUs oder hybriden Systemen mit verschiedenen Arten der Kommunikation denkbar. Ziel ist jedenfalls, das Verhalten der Algorithmen auf größeren Systemen mit einem entsprechend hohen Grad an Parallelisierung zu untersuchen.

## Literatur

- [1] Evrim Acar, Orly Alter, Brett Bader, Zhaojun Bai, Gregory Beylkin, Lieven Delathauwer, Inderjit Dhillon, Chris Ding, Lars Eldén, Petros Drineas, Christos Faloutsos, Shmuel Friedland, Robert J. Harrison, Manal Helal, Anthony Kennedy, Dongmin Kim, Tamara Kolda, Julien Langou, Lek heng Lim, Michael Mahoney, Carla Martin, Martin Mohlenkamp, Jason Morton, Lenore Mullin, Frank Olken, Larsson Omberg, Haesun Park, Robert Plemmons, Stefan Ragnarsson, Sri Priya Ponnappalli, J. Ram Ramanujam, James Reynolds, Phillip Regalia, P. Saday Sadayappan, Berkant Savas, and Charles Van Loan. *Future Directions in Tensor-Based Computation and Modeling*. Workshop Report, National Science Foundation, Arlington, Virginia, 2009.

- [2] Jukka Antikainen, Jiří Havel, Radovan Jošth, Adam Herout, Pavel Zemčík, and Markku Hauta-Kasari. Nonnegative Tensor Factorization Accelerated Using GPGPU. *Parallel and Distributed Systems, IEEE Transactions on*, 22(7):1135–1141, July 2011.
- [3] Markus Flatz. *Nonnegative Matrix Factorization (NMF) Methods and Parallelization*. Masterarbeit, Universität Salzburg, 2012.
- [4] Jingu Kim and Haesun Park. Fast Nonnegative Tensor Factorization with an Active-Set-Like Method. In Michael W. Berry, Kyle A. Gallivan, Efstratios Gallopoulos, Ananth Grama, Bernard Philippe, Yousef Saad, and Faisal Saied, editors, *High-Performance Scientific Computing – Algorithms and Applications*, pages 311–326. Springer, 2012.
- [5] Qiang Zhang, Michael W. Berry, Brian T. Lamb, and Tabitha Samuel. A Parallel Nonnegative Tensor Factorization Algorithm for Mining Global Climate Data. In *Computational Science – ICCS 2009*, volume 5545 of *Lecture Notes in Computer Science*, pages 405–415. Springer Berlin / Heidelberg, 2009.