

Mathematisch-algorithmische Grundlagen für Data Science

Numerische Algorithmen für Datenanalyse und Optimierung

Prof. Dr. Peter Becker

Fachbereich Informatik
Hochschule Bonn-Rhein-Sieg

Sommersemester 2023



**Hochschule
Bonn-Rhein-Sieg**
University of Applied Sciences

Allgemeines zur Vorlesung

Homepage:

<https://www2.inf.h-brs.de/~pbecke2m/dsm/>

▶ [Link zur Homepage](https://www2.inf.h-brs.de/~pbecke2m/dsm/)

Die Vorlesung wird **überwiegend folienbasiert** gehalten.

Die Folien enthalten **nur die wichtigsten Aspekte** (Definitionen, Sätze, knappe Beispiele, wichtige Bemerkung).

Alles was sonst eine Vorlesung ausmacht (Erläuterungen, ausführliche Beispiele, Beweise, Anwendungen, Querverweise auf andere Gebiete der Informatik, etc.) gibt es nur in der Vorlesung selbst.

Die Folien zur Vorlesung (Skript) stehen auf der Homepage **vor der Vorlesung** zur Verfügung.

Übungen und Vorleistung

- Wöchentlich erscheint ein Aufgabenblatt, das in der folgenden Woche in den Übungen besprochen wird.
- **Heute** Ausgabe des ersten Aufgabenblatts!
- Jedes Aufgabenblatt hat eine Sollpunktzahl.
- Zu erfüllende **Vorleistung**: 50% der gesamten Sollpunktzahl (aller Aufgabenblätter) müssen erreicht werden.
- Die Aufgabenblätter können in **Zweiergruppen** bearbeitet werden.
- teilweise ACAT-Aufgaben

Lernziele

- Grundlegende numerische Algorithmen zur Lösung von **linearen Gleichungssystemen** kennen und anwenden können,
- Verfahren der **linearen Optimierung** und deren Anwendung auf **Ausgleichsprobleme** beherrschen,
- das Konzept von **Eigenwerten** verstehen und deren Bedeutung für die **Analyse von Netzwerken** kennen,
- weitere Verfahren für die Datenanalyse kennen und anwenden können (z. B. Verfahren der **Dimensionsreduktion** oder der **Interpolation**) und
- solche Algorithmen **implementieren** können.

Inhalt (geplant)

- 1 Wiederholung: Grundlagen aus linearen Algebra
- 2 Direkte Methoden für lineare Gleichungssysteme
- 3 Ausgleichsrechnung
- 4 Optimierungsverfahren für Data Science
- 5 Eigenwerte und Netzwerkanalyse
- 6 Dimensionsreduktion
- 7 Maximum-Likelihood und der EM-Algorithmus (optional)

Organisatorisches und Formales

Inhaltliche Voraussetzungen:

- Mathematik, insbesondere **Lineare Algebra**, **Wahrscheinlichkeitstheorie**, **Analysis**, **Graphentheorie**
- **Programmieren können**

Umfang: 2V + 2Ü

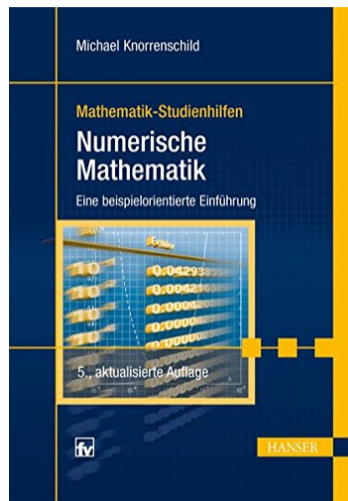
Übungen/Vorleistung: Abgabe von Lösungen zu Übungsaufgaben, $\geq 50\%$ der gesamten Sollpunktzahl

Prüfungsform: gerne mündlich

Literatur

Es gibt **nicht das Lehrbuch** zu dieser Veranstaltung!

- Bücher zu Numerik und Wissenschaftlichem Rechnen haben einen stärkeren Fokus auf die mathematischen Grundlagen der Verfahren und sind ohne Anwendungsbezug geschrieben.
- Bücher zu Data Mining Algorithmen vernachlässigen die Themen Mathematik und Optimierung.



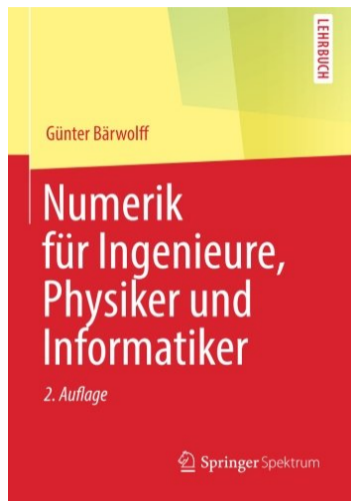
Michael Knorrenschild
*Numerische Mathematik: Eine
beispielorientierte Einführung*
Carl Hanser Verlag
2013

- knapp und anschaulich
- für die Grundlagen der Kapitel 2 und 3



Martin Hanke-Bourgeois
*Grundlagen der Numerischen Mathematik und
des Wissenschaftlichen Rechnens*
Vieweg Teubner Verlag
2008

- theoretischer, umfassender, tiefgehender als Knorrenschild
- für die Grundlagen der Kapitel 2, 3, 5 und 6



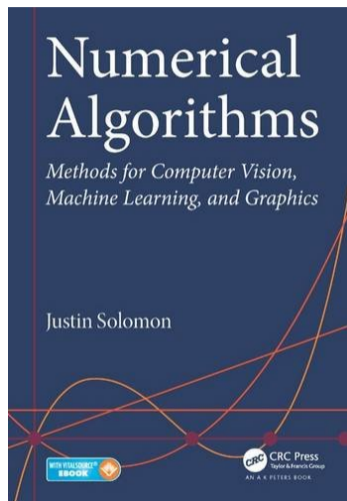
Günter Bärwolff

Numerik

Springer Spektrum

2015

- etwas ausführlicher als Knorrenschild
- für die Grundlagen der Kapitel 2, 3, 5 und 6



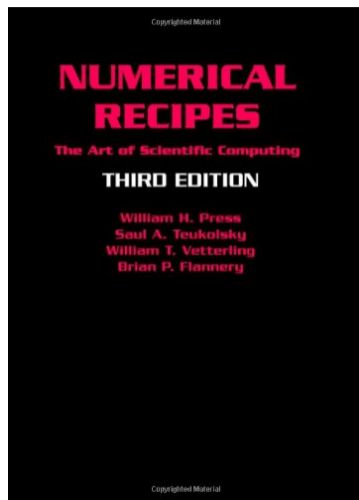
Justin Solomon

Numerical Algorithms: Methods for Computer Vision, Machine Learning, and Graphics

CRC Press

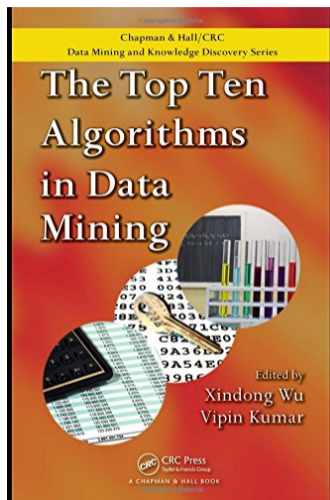
2015

- mit stärkerem Anwendungsbezug
- insbesondere Bildverarbeitung



William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery
Numerical Recipes: The Art of Scientific Computing
Cambridge University Press
2007

- Klassiker für Anwender
- mit starken Quelltextanteilen
- vollständige Quelltexte auch auf CD erhältlich
- geht weit über klassische Numerik hinaus



Xindong Wu, Vipin Kumar (Hrsg.)
The Top Ten Algorithms in Data Mining
CRC Press
2009

- Sammelband mit wichtigen Data Mining Algorithmen
- knapp und aber trotzdem gut verständlich