

Numerische Mathematik und Numerische Lineare Algebra in den Datenwissenschaften

Prof. Dr. rer. nat. Jens Starke
Sommersemester 2025

Inhaltsverzeichnis

1	Minimierung von Funktionen	3
1.1	Gauß-Newton-Verfahren	3
1.2	Levenberg-Marquardt-Verfahren	4

Diese Mitschrift basiert auf der gleichnamigen Vorlesung *Numerische Mathematik und Numerische Lineare Algebra in den Datenwissenschaften*, gehalten im Sommersemester 2025 an der Universität Rostock.

Alle Rechte an Inhalt und Struktur der Lehrveranstaltung liegen bei dem Modulverantwortlichen, Prof. Dr. rer. nat. Jens Starke, sowie der Universität Rostock.

Diese Mitschrift dient ausschließlich zu Lern- und Dokumentationszwecken. Eine kommerzielle Nutzung oder Weiterverbreitung ohne Zustimmung ist nicht gestattet.

Literaturempfehlungen:

1. Martin Hantu-Bourgeois, Grundlagen der Numerik und des wissenschaftlichen Rechnens, Mathematische Leitfäden, Vieweg + Teubner Verlag Wiesbaden, 2009, DOI: [10.1007/978-3-8348-9309-3](https://doi.org/10.1007/978-3-8348-9309-3)
2. Eberhard Zeidler, Nichtlineare Funktionalanalysis und ihre Anwendung, Springer Spektrum Wiesbaden, 2012, DOI: [10.1007/978-3-658-00289-3_3](https://doi.org/10.1007/978-3-658-00289-3_3)

1 Minimierung von Funktionen

Wir erinnern uns an das Gradientenabstiegsverfahren, in welchem wir die Minimierung des Funktionals

$$\phi(x) = \frac{1}{2}x^H Ax - x^H b$$

verwendet haben um die Lösung des linearen Gleichungssystems $Ax = b$ zu finden.

Analog dazu können wir diesen Ansatz auch für beliebige Nullstellenprobleme verwenden.

Für eine hinreichend glatte Funktion $f : D(f) \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $m \geq n$, wandeln wir unser Nullstellenproblem $f(x) = 0$ in ein nichtlineares Ausgleichsproblem um und minimieren

$$\phi(x) = \frac{1}{2}\|f(x)\|_2^2$$

Für eine Nullstelle \hat{x} von f gelten die notwendigen und hinreichenden Bedingungen eines lokalen Minimums:

$$\nabla\phi(\hat{x}) = J_f(\hat{x})^T \cdot f(\hat{x}) = 0$$

$$\nabla^2\phi(\hat{x}) = J_f(\hat{x})^T J_f(\hat{x}) + \sum_{i=1}^m f_i(\hat{x}) \cdot \nabla^2 f_i(\hat{x}) = J_f(\hat{x})^T J_f(\hat{x}) \succeq 0$$

Um eine potentielle Nullstelle von f zu finden können wir also auch ein Minimum von ϕ bestimmen. Das bekannte Gradientenabstiegsverfahren wäre eine Möglichkeit hierfür, konvergiert jedoch nicht besonders schnell.

Eine Alternative hierzu ist die Newton-Methode oder Newton-artige Verfahren, bei welchen wir eine in jedem Iterationsschritt Linearisierung durchführen.

Als kurze Wiederholung betrachten wir hierfür noch einmal die Idee hinter der Newton Methode. Gegeben sei erneut eine hinreichend glatte Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Für das Ziel eine Nullstelle von f zu finden, betrachten wir die Linearisierung (Taylor-Formel 1. Ordnung)

$$f_L(x) = f(x_0) + \left. \frac{\partial}{\partial x} f(x) \right|_{x=x_0} \cdot (x - x_0)$$

Da $f(x) \approx f_L(x)$ bestimmen wir nun die Nullstelle von $f_L(x)$. Als lineares Gleichungssystem ist dies leicht zu lösen und es ergibt sich

$$x = x_0 - \left(\left. \frac{\partial}{\partial x} f(x) \right|_{x=x_0} \right)^{-1} \cdot F(x_0)$$

Aus dieser Nullstelle bilden wir eine rekursive Vorschrift und erhalten die mehrdimensionale Newton-Methode:

$$x_{k+1} = x_k - \left(\left. \frac{\partial}{\partial x} f(x) \right|_{x=x_k} \right)^{-1} \cdot F(x_k), \quad x_0 \in \mathbb{R}^n$$

1.1 Gauß-Newton-Verfahren

Unter Verwendung der gleichen Idee betrachten wir nun wieder unser initiales Ausgleichsproblem, die Minimierung von $\phi(x) = \frac{1}{2}\|f(x)\|_2^2$, und approximieren f durch die Linearisierung f_L an $x^{(k)}$:

$$\phi(x) \approx \frac{1}{2} \left\| f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) \right\|_2^2$$

1.2 Levenberg-Marquardt-Verfahren

Wir haben daher nun ein lineares Ausgleichsproblem, welches leichter zu lösen ist. Das Minimum ist gegeben durch

$$x^{(k+1)} = x^{(k)} - f'(x^{(k)})^\dagger f(x^{(k)})$$

dabei beschreibt A^\dagger das Pseudoinverse einer Matrix A .

Das resultierende Iterationsverfahren heißt Gauß-Newton-Verfahren.

Bemerkung 1.1. Wenn keinen vollen Spaltenrang hat, so muss das lineare Ausgleichsproblem nicht eindeutig lösbar sein. $x^{(k+1)}$ gibt dann die Lösung an, welche zu $x^{(k)}$ den kleinsten euklidischen Abstand hat.

Bei der von uns gewählten Iterationsvorschrift kann es vorkommen, dass unsere lineare Annäherung nicht genau genug ist und wir so nicht gegen eine Lösung des ursprünglich nicht-linearen Ausgleichsproblem konvergieren. Wir beheben dieses Problem, indem wir nicht alle Werte für $x^{(k+1)}$ erlauben, sondern nur diese welche in einem bestimmten Radius um $x^{(k)}$ liegen und erhalten so ein neues Verfahren:

1.2 Levenberg-Marquardt-Verfahren

Da die Linearisierung f_L nicht global geeignet ist sondern nur um $x^{(k)}$ herum, verschärfen wir unser Ausgleichsproblem durch eine Nebenbedingung und erhalten:

$$\text{Minimiere } \frac{1}{2} \|f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)})\|_2^2 \quad \text{unter} \quad \|x - x^{(k)}\|_2 \leq \delta_k$$

Damit liegt unsere nächste Iterierte $x^{(k+1)}$ innerhalb einer kompakten Trust-Region

$$R_k = \left\{ x \in \mathbb{R}^n \mid \|x - x^{(k)}\|_2 \leq \delta_k \right\}$$

Durch die Schreibweise

$$A_k = f'(x^{(k)}), \quad h = x - x^{(k)}, \quad b_k = -f(x^{(k)})$$

erhalten wir eine uns vertrautere Notation:

$$\text{Minimiere } \Psi(x) = \frac{1}{2} \|A_k h - b_k\|_2^2 \quad \text{unter} \quad \|h\|_2 \leq \delta_k$$

eine Lösung $h^{(k)}$ liefert dann die nächste Iterierte durch $x^{(k+1)} = x^{(k)} + h^{(k)}$.

Ohne Beweis genügt eine solche Lösung $h^{(k)}$ der Gleichung

$$(A_k^H A_k + \lambda_k I) h = A_k^H b_k$$

für ein $\lambda_k \geq 0$. Weiter gilt, dass λ_k genau dann positiv ist, wenn $\|h^{(k)}\|_2^2 = \delta_k$.